Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων
και Μηχανικων Υπολογιστων

Τομεας Σηματων, Ελεγχου και Ρομποτικης

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

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΘΕΟΔΩΡΑΚΗΣ Ι. ΣΤΑΥΡΟΣ

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

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

Αθήνα, Ιούνιος 2014

Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων
και Μηχανικων Υπολογιστων

Τομεας Σηματων, Ελεγχου και Ρομποτικης

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

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

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 επταμελή επιτροπή στις 2014:

.....	.....	.....
Π. Μαραγκός	Κ. Τζαφέστας	Γ. Καραγιάννης
Καθηγητής Ε.Μ.Π.	Επίκ. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.
.....	.....	.....
Σ. Κόλλιας	Α. Σταφυλοπάτης	Γ. Ποταμιάνος
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Αν. Καθηγητής Παν/μιο Θεσσαλίας

.....
Α. Αργυρός
Καθηγητής Παν/μιο Κρήτης

Αθήνα, Ιούνιος 2014

. . .

ΘΕΟΔΩΡΑΚΗΣ Ι. ΣΤΑΥΡΟΣ
Διδάκτορας Ηλεκτρολόγος Μηχανικός & Μηχανικός Ηλεκτρονικών Υπολογιστών Ε.Μ.Π.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Γλωσσάρι

ΝΓ Νοηματικής Γλώσσας 25

ΑΝΓ Αμερικανική Νοηματική Γλώσσα 24

ΕΝΓ Ελληνική Νοηματική Γλώσσα 49

HamNoSys Hamburg Notation System 29

PDTS Posture-Detention-Transition-Steady Shift 29

MFCC Mel Frequency Cepstral Coefficients 125

ΣΕ Σχήματος-Εμφάνισης 8

ΜΣΕ Μοντέλο Σχήματος-Εμφάνισης 8

Aff-SAM Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος- Εμφάνισης 104

DS-SAM Μοντελοποίηση Σχήματος- Εμφάνισης με Απευθείας Μετασχηματισμούς Ομοιότητας 104

DTS-SAM Μοντελοποίηση Σχήματος- Εμφάνισης με Απευθείας Μετασχηματισμούς Μετατόπισης και Κλιμάκωσης 104

FD Περιγραφητές Φουριέρ 104

RB Γεωμετρικά Χαρακτηριστικά 104

M Ροπές Hu 104

SIFT Scale-invariant feature transform 27

HOG Histogram of oriented gradients 27

PCA Principal Component Analysis 26

MSSD multi-stream switching probability distribution 30

DTW Dynamic Time Warping 55

σ**.π.π.** συναρτήσεις πυκνότητας πιθανότητας 53

Πρόλογος

Βλέποντας ένα αρκετά μεγάλο κεφάλαιο στην ζωή μας να κάνει τον κύκλο του, είναι πολύ συνηθισμένο να κάνουμε έναν απολογισμό. Έτσι και εγώ. Τις τελευταίες μέρες κατά τη διάρκεια της συγγραφής της διδακτορικής διατριβής πλανιέται συνεχώς στο μυαλό μου η εξής σκέψη. Τελικά μετά από την τόσο μεγάλη προσπάθεια, τις επιτυχίες αλλά και τις αποτυχίες φυσικά, τι ήταν αυτό που άξιζε τον κόπο;

Οι δύο λέξεις που μου έρχονται στο μυαλό οι οποίες θεωρώ ότι είναι καθοριστικής σημασίας, είναι η *ομαδικότητα* και η *δημιουργία*. Είναι δύο λέξεις τις οποίες μπορεί να τις κατανοούμε και να τις αντιλαμβανόμαστε, αλλά είναι πολύ συχνό το φαινόμενο να μην τις έχουμε χωνέψει σε βάθος (όπως θα έλεγε και ο πατέρας μου). Κατά την διάρκεια του διδακτορικού μου κατάφερα να μετουσιώσω αυτά που νιώθω και αισθάνομαι σε λέξεις. Σε αυτές τις δύο απλές με τη πρώτη ματιά, αλλά πολύ ουσιαστικές και θεμελιώδεις λέξεις.

Η πλάκα είναι ότι υποτίθεται ότι η δουλειά ενός ερευνητή είναι αρκετά μοναχική. Παρόλα αυτά είχα την τύχη να έρθω σε επαφή, να μοιραστώ τις σκέψεις μου, να αλληλεπιδράσω και να συνεργαστώ με πάρα πολύ αξιόλογους ανθρώπους. Αυτό είχε ως αποτέλεσμα η συνολική προσπάθεια να είναι κάθε άλλο παρά μοναχική. Ως λάτρης του ποδοσφαίρου θα προσπαθήσω να χρησιμοποιήσω ποδοσφαιρικούς χαρακτηρισμούς.

Αρχικά θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Μαραγκό. Τον προπονητή αυτής της ομάδας, τον εμπνευστή του συνολικού μονοπατιού, τον μαέστρο της συνολικής προσπάθειας. Ήταν πάντα κοντά μου για να με συμβουλέψει και να με καθοδηγήσει. Για τον Βασίλη (ή αλλιώς Μπίλαρο) ό,τι και να πω είναι λίγο. Το δεκάρι της ομάδας, δημιουργός (όπως πρέπει να είναι ένα κλασσικό δεκάρι παλιάς κοπής που λέμε). Μαχητικός, ακούραστος, συνοδοιπόρος και πάνω από όλα φίλος. Η κουβέντα μαζί του ήταν πραγματική λύτρωση για τα πάντα. Τον Νάσο το εξάρι (παρεμπιπτόντως είναι και στην πραγματικότητα). Μαχητικός, χαμογελαστός, δημιουργικός και ευφυής. Πραγματικός κόφτης. Είναι αυτός που μου έβαλε το μικρόβιο της έρευνας όταν τον γνώρισα τυχαία σε ένα πάρτι. Τον Τάσσο το λίμπερο. Προβληματισμένο, τελειομανή με χειρουργική ακρίβεια, έτσι όπως πρέπει να είναι ένα λίμπερο για να καθαρίζει τις φάσεις. Τον Νόντα τον επιθετικό. Γρήγορος, αυθόρμητος, γεμάτος τρέλα, όρεξη και πάνω από όλα επαφή με τα δίχτυα. Τον Ισίδωρο το δεξί μπακ. Ήρεμος, δυναμικός, δραστήριος που κατοικεί στην μαρμελαδοχώρα (που θα έλεγε και ο θείος μου). Η Νάνσυ ήρεμη δύναμη, πεισματάρα, ακουστικός τύπος.

Επιπλέον, θα ήθελα να ευχαριστήσω τους παλιότερους του εργαστηρίου. Αν και δεν συνυπήρξαμε αρκετά χρονικά, η συμβολή τους ήταν καθοριστικής σημασίας. Τον Γιώργο, τον άλλο Γιώργο, τον Σταμάτη, τον Ιάσονα και τον Δημήτρη για τις συζητήσεις, κουβέντες, προβληματισμούς και για το χαβαλέ που κάναμε όποτε βρισκόμασταν. Την Βίκυ, Δέσποινα και Φωτεινή πάντα πρόσχαρες να βοηθήσουν. Ακόμα τους νεότερους, που έδωσαν νέα πνοή στο εργαστήριο. Την Αντιγόνη, που είναι από τους ανθρώπους που τα καταφέρνουν. Τον Παναγιώτη, που προσεγγίζει τα πράγματα με την δική του ματιά, και τον Πέτρο, που είναι μέσα σε όλα. Τον εργατικό και αφοσιωμένο, Γιώργο, τον Αντώνη και τον Κέβη.

Τα μέλη της τριμελούς συμβουλευτικής επιτροπής μου, τον κ. Καραγιάννη και τον κ. Τζα-

φέστα. Η επαφή μου μαζί τους ξεκίνησε από τα πρώτα εξάμηνα των προπτυχιακών μου σπουδών και συνεχίστηκε κατά τη διάρκεια του διδακτορικού. Τον κ. Αργυρό, ο οποίος μου είχε μίλησε για το ερευνητικό πεδίο της Όρασης Υπολογιστών όταν ακόμα ήμουν στα πρώτα εξάμηνα των προπτυχιακών μου σπουδών. Όπως και τα υπόλοιπα μέλη επταμελούς συμβουλευτικής επιτροπής, οι καθηγητές κ. Σταφυλοπάτης, κ. Κόλλιας και κ. Ποταμιάνος. Ήταν τιμή μου η συμμετοχή τους στην εξέταση μου και τους ευχαριστώ ιδιαίτερα.

Οι Φίλοι μου. Ο Πολυχρόνης ο καταφερτζής, ο Στέφανος ο μόντας, ο Ψάρος ο ενθουσιώδης, ο Βαγγέλης ο προβληματισμένος, ο Κώστας ο ορθολογικός, ο Χάμαλος ο καλαμπουρτζής, ο Μανώλης ο λιγομίλητος, η Αγγελική η χαμογελαστή, και ο Γιάννης ο χαβαλές. Τους ευχαριστώ πολύ γιατί ήταν εκεί για να με στηρίξουν σε κάθε προσπάθεια, να με προβληματίσουν, να τους προβληματίσω, και βασικότερο να περπατήσουμε μαζί σε αυτό το ανεξερεύνητο μονοπάτι. Η Μάγια, η οποία με άντεξε και από την καλή και από την ανάποδη. Την ευχαριστώ για όλα που μου προσέφερε, τα οποία ήταν αρκετά σε πολλαπλά επίπεδα. Η Σταυριάννα, με την ηρεμία, την σπιρτάδα και το καθαρό μυαλό.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου. Για την αμέριστη κατανόηση και συμπαράσταση όλα αυτά τα χρόνια. Τον θείο μου Γιάννη και την θεία μου Λίζα για τις αρμένικες επισκέψεις που τους έκανα και για τις πολύωρες συζητήσεις επί παντός επιστητού. Τον θείο μου Γιώργο για τους ιδιαίτερους προβληματισμούς. Την γιαγιά μου Σοφία, η οποία χαιρόταν πάντα να με ευχαριστεί. Τον αδερφό μου, για τον οποίο είμαι πολύ περήφανος. Είναι πάντα κοντά μου για να μου συμπαρασταθεί και να με συμβουλέψει. Τον πατέρα μου, τον σοφό. Με τον δικό του μαγικό και διακριτικό τρόπο, μπορεί να σου φωτίσει σκοτεινές πλευρές που δεν μπορούσες να φανταστείς ότι υπάρχουν (όπως θα έλεγε και η μητέρα μου, το ότι δεν τις βλέπουμε δεν σημαίνει ότι δεν υπάρχουν). Τέλος, την μητέρα μου, η οποία θα γίνει χίλια κομμάτια για να με βοηθήσει και θα μου υπενθυμίζει συνεχώς ότι υπάρχει κάτι βαθύτερο που αξίζει να το αναζητήσω.

Τελικά ναι άξιζε να κοπιάσεις για να δημιουργήσεις, εξερευνήσεις και να συμπρωταγωνιστήσεις μαζί με όλους αυτούς τους αξιοθαύμαστους ανθρώπους...

Θεοδωράκης Σταύρος
Αθήνα, Ιούνιος 2014

Περίληψη

Η διδακτορική αυτή έρευνα επικεντρώνεται στην αυτόματη επεξεργασία βίντεο νοηματικής γλώσσας, στην εξαγωγή χαρακτηριστικών, στη μοντελοποίηση και τελικά στην αναγνώριση νοηματικής γλώσσας συνδυάζοντας τα ερευνητικά πεδία της Αναγνώρισης Προτύπων και Όρασης Υπολογιστών. Σε αυτά τα πλαίσια αναπτύσσονται μέθοδοι για την οπτική επεξεργασία βίντεο νοηματικής γλώσσας και την εξαγωγή χαρακτηριστικών που σχετίζονται με τους αρθρωτές όπως τα χέρια και το κεφάλι. Επιπλέον αναπτύσσονται στατιστικές μέθοδοι για τη μοντελοποίηση και αναγνώριση της νοηματικής γλώσσας. Πιο συγκεκριμένα αναπτύσσονται μέθοδοι για τη μοντελοποίηση της νοηματικής γλώσσας κάνοντας χρήση δεδομενοκεντρικών υπομονάδων. Οι δεδομενοκεντρικές υπομονάδες αποτελούν τα δομικά στοιχεία που απαρτίζουν τα νοήματα και κατασκευάζονται αυτόματα καθοδηγούμενες από τα δεδομένα εκπαίδευσης. Στόχος τους είναι να μοντελοποιήσουν τα διαφορετικά είδη κίνησης, θέσης και χειρομορφής των χεριών του νοηματιστή κατά την άρθρωση ενός νοήματος. Επιπλέον η μοντελοποίηση των παραπάνω υπο-μονάδων εμπλουτίζεται από την αξιοποίηση γλωσσικής-φωνητικής πληροφορίας. Ακόμα, λόγω των πολλαπλών παράλληλων ροών πληροφορίας (π.χ. θέση, κίνηση των χεριών, είδος χειρομορφής κ.τ.λ) αναπτύσσονται μέθοδοι για την σύμμειξη των διαφορετικών καναλιών πληροφορίας με απώτερο σκοπό την αναγνώριση της νοηματικής γλώσσας. Οι μέθοδοι που προτείνονται συνδυάζονται σε ένα συνολικό σύστημα επεξεργασίας, εκπαίδευσης και αναγνώρισης και αξιολογούνται σε προτυποποιημένα σώματα βίντεο νοηματικών γλωσσών. Συγκρίσεις με σύγχρονες μεθόδους από τη βιβλιογραφία οδηγούν σε βελτιώσεις σε σχέση με βασικές υπάρχουσες μεθόδους. 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. Τέλος, αναπτύχθηκε ένα σύστημα αναγνώρισης χειρονομιών από πολυτροπικά δεδομένα, τα οποία περιλαμβάνουν οπτικές αλλά και ακουστικές ροές πληροφορίας. Πιο συγκεκριμένα αναπτύχθηκαν αλγόριθμοι για τη μοντελοποίηση και σύμμειξη των πολυτροπικών ροών πληροφορίας, με απώτερο στόχο την ανίχνευση και αναγνώριση πολυτροπικών χειρονομιών.

Abstract

This research focuses on the automatic video processing of sign language videos, feature extraction, modeling and finally sign language recognition, combining the research areas of Pattern Recognition and Computer Vision. In this context, we have developed methods, for the visual processing of sign language videos and the extraction of features related to sign articulators such as the hands and the head of the signer. Moreover we have developed statistical methods for the sign language modeling and recognition. Specifically, we propose a data-driven method for the modeling of sign language using subunits. Subunits correspond to the smallest contrastive units of SL. Each sign is represented by combining a limited number of subunits and thus subunits can be used for modeling sign languages for the purposes of automatic sign language recognition. Subunits are used for the modeling of the different type of the movements, positions and the handshapes of signer's hands during sign articulation. Further, the aforementioned modeling of the subunits is enriched using linguistic-phonetic information. In addition, we have developed methods for the integration/fusion of the multiple information cues (i.e. position, movement, handshape e.t.c.). The proposed methods are combined in an overall framework and evaluated in standardized sign language databases. Comparisons with other approaches from the state of the art, indicate that the proposed approaches lead to significant improvements. This research is expected to affect fields such as linguistic research, automatic corpora processing and the study of sign languages. Finally, we have developed a framework for multimodal gesture recognition where the information cues include both acoustic and visual cues. Specifically, we have developed algorithms for the modeling and fusion of these multimodal cues, for the automatic activity detection and multimodal gesture recognition.

Κεφάλαιο 1

Εισαγωγή

Ας σκεφτούμε το σενάριο όπου δύο άνθρωποι επικοινωνούν μέσω της νοηματικής γλώσσας.

Έστω τώρα ότι ο ένας από τους δύο ανθρώπους εκτελεί ένα νόημα ενώ ο άλλος, δεδομένου ότι έχει εκπαιδευτεί κατάλληλα (δηλαδή γνωρίζει νοηματική γλώσσα), κατανοεί την πληροφορία που θέλησε ο πρώτος να του μεταδώσει μέσω του νοήματος.

Στο παράδειγμα των δύο ανθρώπων που επικοινωνούν μέσω της νοηματικής γλώσσας, ας βάλουμε στη θέση αυτού που λαμβάνει και κατανοεί την πληροφορία που θέλησε ο πρώτος να του μεταδώσει μέσω ενός νοήματος έναν υπολογιστή.

Αντικείμενο της παρούσας διδακτορικής διατριβής είναι η εκμάθηση και εκπαίδευση ενός υπολογιστή έτσι ώστε να μπορεί να αναγνωρίζει τη νοηματική γλώσσα. Η αναγνώριση αυτή βασίζεται στην οπτική πληροφορία η οποία του παρέχεται μέσω ενός οπτικού αισθητήρα π.χ. μια κάμερα.

Για την ανάπτυξη ενός συστήματος αυτόματης αναγνώρισης νοηματικής γλώσσας είναι απαραίτητη η εξαγωγή κατάλληλων χαρακτηριστικών από την οπτική πληροφορία (βίντεο), η μοντελοποίηση της νοηματικής γλώσσας και τέλος η αναγνώριση της. Η εξαγωγή κατάλληλων χαρακτηριστικών σχετίζεται με την εξαγωγή της απαραίτητης πληροφορίας από κάθε πλαίσιο του βίντεο, όπως τη θέση των δύο χεριών του νοηματιστή πάνω στην εικόνα, το σχήμα των χεριών (χειρομορφή) κ.α. Την πληροφορία αυτή την ονομάζουμε διάνυσμα χαρακτηριστικών. Υπολογίζοντας τα διανύσματα χαρακτηριστικών για κάθε πλαίσιο του βίντεο δημιουργούμε μια ακολουθία διανυσμάτων χαρακτηριστικών χρησιμοποιώντας μεθόδους οπτικής επεξεργασίας βίντεο για την εξαγωγή τους. Στη συνέχεια, είναι απαραίτητη η εκπαίδευση κατάλληλων μοντέλων τα οποία να βασίζονται στα διανύσματα χαρακτηριστικών. Στα πλαίσια της έρευνάς μας για την εκπαίδευση και μοντελοποίηση της νοηματικής γλώσσας χρησιμοποιήσαμε στατιστικές μεθόδους και μοντέλα. Τέλος χρησιμοποιώντας τα εκπαιδευμένα μοντέλα, την ακολουθία διανυσμάτων χαρακτηριστικών και κατάλληλους αλγορίθμους επιτυγχάνουμε την αναγνώριση των νοημάτων που εμφανίζονται στο υπό ανάλυση βίντεο. Δηλαδή ουσιαστικά αναγνωρίζουμε ποιο νόημα ειπώθηκε και σε ποια χρονική στιγμή.

Από τα παραπάνω διαπιστώνουμε ότι η δημιουργία ενός συστήματος αναγνώρισης νοηματικής γλώσσας απαιτεί την συμβολή δύο μεγάλων επιστημονικών πεδίων. Πρώτον της Όρασης Υπολογιστών για την εξαγωγή των οπτικών χαρακτηριστικών από το βίντεο. Δεύτερον της Αναγνώρισης Προτύπων για τη μοντελοποίηση και αναγνώριση των νοημάτων που αναπαρίστανται από ένα σήμα δεδομένων.

1.1 Νοηματική γλώσσα

Η νοηματική γλώσσα είναι η φυσική γλώσσα της κοινότητας των Κωφών. Αποτελεί μια ολοκληρωμένη αυτόνομη γλώσσα με δικιά της γραμματική και συντακτικό. Δεν αποτελεί αναπαράσταση κάποιας από τις φωνούμενες γλώσσες ούτε είναι κάποιο είδος παντομίμας. Οι βασικές μονάδες

του λόγου ονομάζονται νοήματα. Τα νοήματα μπορούν να έχουν λεξική ή γραμματική σημασία, ακριβώς όπως τα μορφήματα και οι λέξεις στις φυσικές γλώσσες.

Στη συνέχεια, αναφέρουμε μερικές από τις λανθασμένες αντιλήψεις σε σχέση με την νοηματική γλώσσα:

- όλες οι νοηματικές γλώσσες είναι ίδιες, δηλαδή ότι υπάρχει μια διεθνής νοηματική γλώσσα,
- η νοηματική γλώσσα είναι εικονική,
- δεν έχει σύνταξη και γραμματική,
- δεν μπορεί να μεταφέρει αφηρημένες έννοιες,
- είναι υποδεέστερη της ομιλούμενης και
- το δακτυλικό αλφάβητο και η νοηματική είναι το ίδιο πράγμα.

Η μετάδοση της πληροφορίας αυτής γίνεται μέσω οπτικών-κινησιακών μοτίβων σε αντίθεση με την επικοινωνία μέσω της φωνής η οποία γίνεται μέσω ακουστικών σημάτων. Αυτά τα οπτικο-κινησιακά μοτίβα σχηματίζονται μέσω των χεριών, του σώματος και του προσώπου του ανθρώπου.

Τα χαρακτηριστικά συστατικά ενός νοήματος αποτελούν:

- η χειρομορφή,
- ο προσανατολισμός της παλάμης,
- η θέση των χεριών,
- η κίνηση των χεριών,
- η στάση του σώματος και
- οι εκφράσεις του προσώπου.

Πιο συγκεκριμένα, η χειρομορφή είναι το σχήμα που παίρνει η παλάμη και η θέση στην οποία τοποθετούνται τα δάκτυλα. Η ίδια η χειρομορφή όμως από μόνη της δεν είναι φορέας σημασίας. Για να αποκτήσει σημασία, για να δημιουργηθεί δηλαδή ένα νόημα, η χειρομορφή πρέπει να συνοδεύεται από όλα τα παραπάνω στοιχεία. Ο προσανατολισμός της παλάμης, αποτελεί την κατεύθυνση προς την οποία στρέφεται η χειρομορφή κατά τον σχηματισμό του νοήματος. Η θέση των χεριών στο χώρο ή επάνω στο σώμα έχει επίσης σημασία. Ο χώρος στον οποίο εκτελούνται τα νοήματα είναι καθορισμένος και λέγεται χώρος νοηματισμού. Ο χώρος αυτός αντιστοιχεί περίπου σε ένα τετράγωνο που ορίζεται από την κορυφή της κεφαλής ως τον άνω κορμό και εκτείνεται σε 20-30 εκατοστά δεξιά και αριστερά από τα μπράτσα. Αν χρησιμοποιήσουμε μία χειρομορφή έξω από το χώρο αυτό, π.χ. με τα μπράτσα κρεμασμένα δίπλα στο σώμα, το αποτέλεσμα δεν είναι αναγνωρίσιμο ως νόημα. Η κίνηση των χεριών είναι επίσης εξέχουσας σημασίας για την εκτέλεση ενός νοήματος. Εκτός από τη συμμετοχή της στο σχηματισμό του νοήματος, η κίνηση μπορεί να είναι και φορέας άλλων σημασιών, για παράδειγμα να δηλώνει τον αριθμό (ενικό ή πληθυντικό), το μέγεθος ενός αντικειμένου (μικρότερο ή μεγαλύτερο), ακόμα και τη συχνότητα μίας ενέργειας. Η στάση (ή κίνηση) του σώματος όπως και η έκφραση του προσώπου, αποτελούν επίσης συστατικά του νοήματος με την έννοια ότι λειτουργούν για να μεταφέρουν πληροφορία όπως αυτή που δηλώνεται από τον τόνο της φωνής στις ομιλούμενες γλώσσες. Για παράδειγμα, η έννοια του μέλλοντος διατυπώνεται στην Αμερικανική Νοηματική Γλώσσα (ΑΝΓ) συνδυάζοντας το νόημα με μία ελαφρά κλίση του σώματος προς τα εμπρός.

Ένα από τα δύο χέρια (δεξί ή αριστερό ανάλογα με τον νοηματιστή), ονομάζεται κυρίαρχο χέρι ενώ το άλλο ονομάζεται δευτερεύον. Το κυρίαρχο χέρι αρθρώνει τα βασικά φωνητικά μέρη

ενός νοήματος, ενώ το δευτερεύον λειτουργεί συμπληρωματικά, είτε εκτελώντας συμμετρικές ή αντισυμμετρικές κινήσεις σε σχέση με το κυρίαρχο χέρι, είτε λειτουργώντας ως σημείο άρθρωσης. Ως σημείο άρθρωσης αναφέρουμε την θέση του κυρίαρχου χεριού σε σχέση με το σώμα ή το δευτερεύον χέρι του νοηματιστή.

Η μεγαλύτερη δυσκολία που εμφανίζεται όταν κάποιος θέλει να μελετήσει μία νοηματική γλώσσα, είναι ¨τεχνικού¨ χαρακτήρα, με την έννοια ότι δεν υπάρχει γραφή ή μεταγραφή κάποιου είδους. Το αποτέλεσμα αυτής της κατάστασης μπορεί να συγκριθεί με αυτό που συμβαίνει σε πολλές προφορικές γλώσσες: η καταγραφή της γλώσσας είναι εξαιρετικά ελλιπής και η μελέτη της ιδιαίτερα περιορισμένη. Είναι προφανές ότι το πρόβλημα είναι εντονότερο στην περίπτωση της Νοηματικής Γλώσσας (ΝΓ), για την οποία η καταγραφή οποιασδήποτε πληροφορίας γινόταν μέχρι τώρα μόνο με φωτογραφίες ή σκίτσα, από τα οποία έλειπε ένα βασικό συστατικό των νοημάτων: η κίνηση. Επιπλέον, οι διάφορες γλωσσικές και κοινωνικές προκαταλήψεις, όπως για παράδειγμα ότι η νοηματική δεν είναι ¨ακριβώς¨ γλώσσα, έχουν εμποδίσει την ευρύτερη διάδοσή της.

1.2 Επισκόπηση σχετικής έρευνας

Σε αυτή την ενότητα κάνουμε μια εκ βαθέων ανάλυση των ερευνητικών προσεγγίσεων που έχουν προταθεί κατά καιρούς στη διεθνή βιβλιογραφία σχετικά με τις δύο μεγάλες ερευνητικές συνιστώσες με τις οποίες ασχοληθήκαμε στην παρούσα διδακτορική διατριβή. Αυτές αποτελούν την αυτόματη αναγνώριση νοηματικής γλώσσας και την αναγνώριση χειρονομιών από πολυτροπικά δεδομένα. Συγκεκριμένα για την επισκόπηση της ερευνητικής περιοχής της αυτόματης αναγνώρισης νοηματικής γλώσσας παρουσιάζουμε ερευνητικές προσεγγίσεις σχετιζόμενες με την εξαγωγή χαρακτηριστικών και μοντελοποίηση της νοηματικής γλώσσας. Σχετικά με την περιοχή της αναγνώρισης χειρονομιών από πολυτροπικά δεδομένα, κάνουμε μια γενική επισκόπηση και εστιάζουμε σε πρόσφατα δημοσιευμένες μεθόδους οι οποίες κατέκτησαν τις πρώτες θέσεις στον διαγωνισμό πολυτροπικής αναγνώρισης χειρονομιών CHALEARN.

1.2.1 Οπτική επεξεργασία και εξαγωγή χαρακτηριστικών

Η αυτόματη αναγνώριση νοηματικής γλώσσας είναι ένα σύνθετο πρόβλημα το οποίο παρουσιάζει σημαντικές προκλήσεις στην εξαγωγή κατάλληλων χαρακτηριστικών [97, 1, 27]. Οι πιο πρόσφατες ερευνητικές εργασίες βασίζονται στην οπτική επεξεργασία βίντεο νοηματικής για την εξαγωγή των κατάλληλων χαρακτηριστικών, σε αντίθεση με άλλες μεθόδους που χρησιμοποιούν datagloves [45, 47, 46], χρωματιστά γάντια [11], συστήματα καταγραφής κίνησης (motion capture) [136, 98, 72], και άλλα [148]. Η διαδικασία εξαγωγής χαρακτηριστικών μέσω της οπτικής επεξεργασίας ενός βίντεο νοηματικού λόγου αποτελείται από τρία στάδια: α) Ανίχνευση των αρθρωτών του νοηματικού λόγου (κεφάλι, χέρια και κορμός) β) Παρακολούθηση των αρθρωτών κατά την διάρκεια του νοηματικού λόγου και γ) Εξαγωγή κατάλληλων χαρακτηριστικών σχετιζόμενων με τους αρθρωτές όπως π.χ. σχήμα, θέση, κίνηση κ.α.

Ανίχνευση και παρακολούθηση αρθρωτών

Η ανίχνευση των αρθρωτών επιτυγχάνεται χρησιμοποιώντας διάφορα είδη 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 του κεφαλιού του νοηματιστή έχει εφαρμοστεί σε αρκετές ερευνητικές εργασίες [7, 117, 146, 153, 113]. Η χρήση της χρωματικής συνιστώσας ενδείκνυται λόγω της ιδιαιτερότητας 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* [123, 67]. Η πληροφορία της κίνησης επίσης έχει χρησιμοποιηθεί [33, 62], υποθέτοντας 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 [96], οι οποίοι για την ανίχνευση των χεριών βασίστηκαν στην πληροφορία του σχήματος των χεριών χρησιμοποιώντας boosted classifier tree. Σε άλλες ερευνητικές προσεγγίσεις όπως στα άρθρα [11, 10, 59], 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 για την εξαγωγή χαρακτηριστικών είναι η παρακολούθηση των αρθρωτών κατά τη διάρκεια του νοηματικού λόγου. Για την παρακολούθηση των χεριών έχουν χρησιμοποιηθεί ποικίλες μέθοδοι: blob-based [117, 122, 7], model-based [62, 19], hand contour και boundary models [23, 33]. 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 [153, 113] βασίστηκαν σε ένα πιθανοτικό πλαίσιο για την ανάθεση ετικετών (κεφάλι, χέρια κ.τ.λ) στις περιοχές όπου έχουν ανιχνευθεί οι αρθρωτές. Οι Downton και Drouet στο άρθρο [40], 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 χρησιμοποιώντας πολλαπλές κάμερες [133, 83]. Άλλος ένας παράγοντας ο οποίος αυξάνει τη δυσκολία του προβλήματος της παρακολούθησης των αρθρωτών έγκειται στην πολυπλοκότητα του φόντου του βίντεο. Στις περισσότερες ερευνητικές μελέτες χρησιμοποιήθηκαν βίντεο με ομοιόμορφο φόντο [11, 146, 153]. Ωστόσο υπάρχουν και προσεγγίσεις [23, 1] όπου εφαρμόζονται τεχνικές αφαίρεσης φόντου (background subtraction) με στόχο την ανεξαρτησία από το είδος του φόντου του βίντεο.

Εξαγωγή χαρακτηριστικών διανυσμάτων

Η εξαγωγή χαρακτηριστικών διανυσμάτων τα οποία περιγράφουν τους αρθρωτές σε κάθε πλαίσιο αποτελεί κρίσιμο στοιχείο για τα συστήματα αναγνώρισης ΝΓ. Όσο αφορά τα χέρια, απαραίτητο χαρακτηριστικό είναι οι δισδιάστατες ή τρισδιάστατες συντεταγμένες των κεντροειδών των χεριών του νοηματιστή [117, 11, 122, 33, 133]. Χαρακτηριστικά που σχετίζονται με την κίνηση των χεριών είναι επίσης απαραίτητα, όπως π.χ. η τροχιά των χεριών ή η οπτική ροή [23, 146]. Πιο περίπλοκα χαρακτηριστικά σχετίζονται με το σχήμα ή την εμφάνιση (appearance) των χεριών. Σε αρκετές ερευνητικές προσεγγίσεις χρησιμοποιήθηκαν γεωμετρικά χαρακτηριστικά βασισμένα στο σχήμα των χεριών, όπως π.χ. ροπές σχήματος [117, 122], ή αποστάσεις μεταξύ των δακτύλων και της παλάμης του χεριού [11, 10]. Άλλες προσεγγίσεις βασίστηκαν στο περίγραμμα του χεριού για την εξαγωγή χαρακτηριστικών ανεξαρτήτως μετατόπισης, κλίμακας και περιστροφής του χεριού όπως περιγραφητές Φουριέρ (Fourier descriptors) [23, 120].

Άλλες τεχνικές βασίζονται στην ανάλυση σε κύριες συνιστώσες (Principal Component Analysis (PCA)) εφαρμόζοντας κανονικοποίηση σε σχέση με το μέγεθος, την φωτεινότητα και τον προσανατολισμό [33] στις εικόνες χεριών. Το προτεινόμενο μοντέλο σχήματος-εμφάνισης [110], το οποίο παρουσιάζεται στην ενότητα 2.3.2, ακολουθεί την ίδια λογική με αυτές τις μεθόδους αλλά διαφέρει στα ακόλουθα. Χρησιμοποιεί τον αφινικό μετασχηματισμό για την ευθυγράμμιση των εικόνων των χεριών, ο οποίος επεκτείνει και τον μετασχηματισμό ομοιότητας που χρησιμοποιήθηκε στο άρθρο [14] αλλά και τον μετασχηματισμό μετακίνησης-κλίμακας όπως στα άρθρα [33, 144, 42].

Έτσι δίνεται η δυνατότητα προσέγγισης μεγαλύτερου εύρους αλλαγών στην τρισδιάστατη πόζα του χεριού. Επιπλέον, η εκτίμηση των παραμέτρων του βέλτιστου μετασχηματισμού γίνεται από κοινού με την εκτίμηση των PCA βαρών. Τέλος, σε αντίθεση με όλες τις παραπάνω μεθόδους, ενσωματώνουμε συνδυασμένη πρότερη γνώση σε σχέση με την στατική εικόνα αλλά και την δυναμική εξέλιξή της, κάνοντας τις εκτιμήσεις των παραμέτρων περισσότερο εύρωστες και επιτρέποντας την προσαρμογή του μοντέλου σε ένα νέο νοηματιστή.

Άλλες σχετικές έρευνες οι οποίες βασίζονται σε PCA, ενεργά μοντέλα σχήματος και εμφάνισης (Active Shape and Appearance Models) [29, 82] για την εξαγωγή χαρακτηριστικών και αναγνώριση χειρομορφών παρουσιάζονται στα άρθρα [2, 62, 16, 49]. Ενδεικτικά οι Huang και Jeng [62] χρησιμοποίησαν ενεργά μοντέλα σχήματος [30] για την αναπαράσταση και μοντελοποίηση του περιγράμματος του χεριού. Ενώ οι Bowden και Sahardi [16] εφάρμοσαν PCA πάνω στα περιγράμματα των χεριών και κατασκεύασαν μη γραμμικά Point Distribution μοντέλα. Το προτεινόμενο μοντέλο σχήματος-εμφάνισης ακολουθεί την ίδια λογική με αυτές τις μεθόδους αλλά διαφέρει στο ότι οι εικόνες προς μοντελοποίηση είναι εικόνες σχήματος-εμφάνισης και για τον έλεγχο της περιστροφή τους δεν χρησιμοποιούνται landmark points αλλά οι έξι παράμετροι του αφινικού μετασχηματισμού. Με αυτό τον τρόπο η αναπαράσταση δεν γίνεται μέσω landmark points και άρα αποφεύγεται η επίπονη διαδικασία επισημείωσής τους στα δεδομένα εκπαίδευσης.

Μια ακόμη δημοφιλής μέθοδος εξαγωγής χαρακτηριστικών σχήματος προκύπτει από την χρήση του Ιστογράμματος Κατευθυνόμενων Κλίσεων (Histogram of oriented gradients (HOG)) [34]. Αποτελεί ένα από κοινού ιστόγραμμα κβαντισμένων κατευθύνσεων της κλίσης της εικόνας και της μετατοπισμένης θέσης, στην γειτονιά κάθε εικονοστοιχείου. Στα άρθρα [18, 79], συναντώνται τέτοιου είδους περιγραφητές για την εξαγωγή χαρακτηριστικών για τα χέρια των νοηματιστών και συνδυάζουν τόσο την εμφάνιση όσο και το σχήμα των χεριών. Άλλος ένας δημοφιλής περιγραφητής που χρησιμοποιείται κατά κόρον για την αναγνώριση αντικειμένων είναι ο Scale-invariant feature transform (SIFT), ο οποίος χρησιμοποιήθηκε στο άρθρο [48] για την εξαγωγή χαρακτηριστικών από χειρομορφές. Μια ακόμη εργασία που χρησιμοποιεί τον περιγραφητή SIFT για την αναπαράσταση χειρομορφών είναι η μέθοδος που παρουσιάζεται στο άρθρο [124], η οποία επιπλέον λαμβάνει υπόψη γλωσσικούς περιορισμούς και εκμεταλλεύεται ένα Bayesian δίκτυο για την βελτίωση της αναγνώρισης χειρομορφών. Εκτός από μεθόδους οι οποίες επεξεργάζονται ή μοντελοποιούν δισδιάστατες εικόνες των χεριών υπάρχουν και άλλες οι οποίες βασίζονται στην τρισδιάστατη μοντελοποίηση, με στόχο την προσέγγιση των γωνιών που σχηματίζονται μεταξύ των δακτύλων ενός χεριού και της τρισδιάστατής τους πόζας [60, 49, 39, 95, 93, 94]. Αυτές οι μέθοδοι έχουν το πλεονέκτημα ότι είναι ανεξάρτητες της οπτικής γωνίας της κάμερας και της πόζας του χεριού.

1.2.2 Στατιστική μοντελοποίηση και αναγνώριση της ΝΓ

Η μοντελοποίηση των ροών πληροφορίας με σκοπό την αυτόματη αναγνώριση νοηματικής γλώσσας θέτει αρκετές προκλήσεις [97, 1, 27]. Αρκετές μέθοδοι έχουν προταθεί στην διεθνή βιβλιογραφία για την μοντελοποίηση της νοηματικής γλώσσας. Οι αρχικές προσπάθειες βασίζονταν κυρίως στη μοντελοποίηση κάθε νοήματος ξεχωριστά, με άλλα λόγια για κάθε νόημα που υπήρχε στο λεξικό εκπαιδευόταν ένα μοντέλο. Πιο πρόσφατα, μέθοδοι οι οποίες βασίζονται στην μοντελοποίηση υπομονάδων επέστησαν την προσοχή τους, εκμεταλλευόμενοι το γεγονός ότι διαφορετικά νοήματα μοιράζονται τις ίδιες υπομονάδες. Οι μέθοδοι αυτές μοντελοποιούν είτε δεδομενοκεντρικές είτε γλωσσικές-φωνητικές υπομονάδες. Οι πρώτες δεν χρειάζονται επισημειώσεις σε φωνητικό επίπεδο, αλλά χρησιμοποιούν τεχνικές αυτόματης χρονικής κατάτμησης και συσταδοποίησης. Αντιθέτως, οι δεύτερες βασίζονται σε πρότερη γλωσσική-φωνητική πληροφορία και σε φωνητικές επισημειώσεις, κατασκευάζοντας γλωσσικές-φωνητικές υπομονάδες οι οποίες είναι γλωσσικά ερμηνεύσιμες. Στις επόμενες παραγράφους συνοψίζουμε διάφορες πτυχές τής μοντελοποίησης

και αναγνώρισης της νοηματικής γλώσσας αναφερόμενοι σε σχετικές έρευνες από την διεθνή βιβλιογραφία και συσχετίζουμε αυτές τις ερευνητικές εργασίες με τη δική μας έρευνα. Επιπλέον, στον πίνακα 1.1 παρουσιάζουμε ενδεικτικές εργασίες ομαδοποιημένες βασιζόμενοι στις παραπάνω πτυχές.

Μοντελοποίηση

Η νοηματική γλώσσα εμπεριέχει πολλαπλές ροές πληροφορίας οι οποίες μεταβάλλονται δυναμικά στον χρόνο. Για την μοντελοποίησή τους απαιτούνται μοντέλα που να μπορούν να λάβουν υπόψη τους τη δυναμική μεταβλητότητά τους. Μέθοδοι που μπορούν να αντιμετωπίσουν αυτές τις πτυχές μπορεί να είναι είτε παραμετρικές π.χ. Hidden Markov Models (HMMs), Conditional Random Fields (CRFs), είτε όχι π.χ. Dynamic Time Warping (DTW). Τα HMMs αποτελούν μια πολύ δημοφιλή λύση λόγο της δυνατότητας που έχουν για την μοντελοποίηση της δυναμικής [104]. Αρχικές προσπάθειες χρησιμοποίησαν τα HMMs για την εκπαίδευση ενός Hidden Markov Model (HMM) μοντέλου ανά νόημα [116, 133], ενώ αντιθέτως πιο πρόσφατες μέθοδοι βασίστηκαν στην μοντελοποίηση σε επίπεδο υπομονάδας είτε άμεσα (explicit) [135, 11, 47] είτε έμμεσα (implicit) [54]. Μια σημαντική συνεισφορά σχετιζόμενη με τα HMMs είναι τα Parallel HMMs (PaHMMs) [136], τα οποία δίνουν τη δυνατότητα μοντελοποίησης πολλαπλών ροών πληροφορίας παράλληλα. Επίσης και άλλες υβριδικές μέθοδοι έχουν εμφανιστεί, οι οποίες συνδυάζουν τα HMMs είτε με recurrent networks [45, 137], είτε με το γνωστό tandem από την ερευνητική περιοχή της αυτόματης αναγνώρισης φωνής όπου συνδυάζουν multi-layer perceptrons με Gaussian Mixture Models (GMMs) [54]. Επιπλέον άλλες ενδεικτικά προσεγγίσεις είναι οι αλυσίδες Μαρκόφ (Markov chains), οι οποίες χρησιμοποιήθηκαν στο άρθρο [66], και ο DTW που συχνά χρησιμοποιείται σε exampler-based μεθόδους [138]. Άλλες ερευνητικές προσπάθειες εστιάζουν σε discriminative μεθόδους όπως DTW με διακριτά (discriminative) χαρακτηριστικά [76], HMMs με διακριτά τμηματικά (discriminative segmental) χαρακτηριστικά [148], multi-class Fischer kernels [6], και sequential pattern boosting με ασθενείς ταξινομητές (weak classifiers) [28]. Στην έρευνά μας εμείς χρησιμοποιήσαμε HMMs για την μοντελοποίηση υπομονάδων.

Εκτός από την αναγνώριση νοηματικής και άλλα επιμέρους προβλήματα σχετιζόμενα με την μοντελοποίηση της ΝΓ έχουν τραβήξει την προσοχή κατά καιρούς, όπως π.χ. η αντιμετώπιση της συνάρθρωσης νοημάτων κάνοντας χρήση CRFs [147], ο εντοπισμός νοημάτων σε συνεχή νοηματικό λόγο με CRFs [145] και η μοντελοποίηση των epenthesis κινήσεων, δηλαδή των κινήσεων που προκύπτουν κατά την εκφορά δύο διαδοχικών νοημάτων [133, 46]. Οι συγγραφείς στο άρθρο [18] ασχολήθηκαν με τον εντοπισμό νοημάτων εκμεταλλευόμενοι τους υπότιτλους σε βίντεο νοηματισμού, μέσω εκμάθησης με επίβλεψη από πολλαπλές επαναλήψεις του ίδιου νοήματος, ενώ στο άρθρο [89], προτείνεται μια μέθοδος για τον εντοπισμό κοινών προτύπων μεταξύ διαφορετικών νοημάτων, χρησιμοποιώντας iterative conditional modes σε πολλαπλές ακολουθίες νοημάτων.

Γλωσσικά μοντέλα και φωνητικά συστήματα επισημείωσης

Μια σημαντική ερευνητική εργασία είναι αυτή του Stokoe [118] ο οποίος εισήγαγε για πρώτη φορά την έννοια των διακριτών υπομονάδων που απαρτίζουν τη νοηματική γλώσσα. Πρότεινε μια παράλληλη αποσύνθεση των νοημάτων σε πολλαπλές παράλληλες ροές πληροφορίας: tab (θέση), dez (χειρομορφή) και sig (κίνηση). Οι Liddell and Johnson (L&J) εισήγαγαν ως φωνολογική βάση [77] την έννοια της διαδοχής διαφορετικών τύπων φωνητικών υπομονάδων στη νοηματική γλώσσα. Πρότειναν το μοντέλο Movement-Hold [78] το οποίο λάμβανε υπόψη του και την έννοια της παραλληλίας αλλά και της διαδοχής φωνητικών υπομονάδων. Συγκεκριμένα, εισήγαγαν δύο τύπους υπομονάδων: τις Movement και τις Hold. Οι Movement υπομονάδες αντιστοιχούσαν σε χρονικά τμήματα κατά τα οποία έχουμε μεταβολή τουλάχιστον σε μια από τις ροές πληροφορίας όπως π.χ. κίνηση ή αλλαγή της χειρομορφής. Αντιθέτως οι Hold υπομονάδες αντιστοιχούσαν

σε χρονικά τμήματα κατά τα οποία δεν έχουμε καμία μεταβολή. Αυτό είχε ως αποτέλεσμα τα νοήματα να απαρτίζονται από μια ακολουθία Movement και Hold υπομονάδων. Πρόσφατα οι L&J παρουσίασαν το Posture-Detention-Transition-Steady Shift (PDTS) μοντέλο [65, 64] το οποίο αντικατέστησε το παλαιότερο Movement-Hold διορθώνοντας αρκετές από τις ελλείψεις του. Επιπλέον, αρκετά συστήματα φωνητικής επισημείωσης έχουν προταθεί για τη νοηματική γλώσσα. Ενδεικτικά αναφέρουμε το σύστημα Hamburg Notation System (HamNoSys) [103] και το SignWriting [119]. Το HamNoSys είναι ένα φωνητικό σύστημα επισημείωσης το οποίο βασίζεται στην αποσύνθεση που προτάθηκε από τον Stokoe. Το SignWriting χρησιμοποιεί γραφικά σύμβολα για την αναπαράσταση της χειρομορφής, της κατεύθυνσης της παλάμης, της κίνησης, της θέσης και των εκφράσεων του προσώπου με στόχο την περιγραφή των νοημάτων. Στην έρευνα μας χρησιμοποιήσαμε PDTS φωνητικές επισημειώσης οι οποίες είχαν παραχθεί αυτομάτως από HamNoSys επισημειώσεις.

Implicit δεδομενοκεντρικές υπομονάδες

Αρκετές μέθοδοι επένδυσαν στην μοντελοποίηση των επιμέρους παράλληλων ροών πληροφορίας με βάση την αποσύνθεση που προτάθηκε από τον Stokoe. Οι Kadir et al. [66] χρησιμοποίησαν μια περιγραφή βασισμένοι στις παραμέτρους του Stokoe. Ο Ding και ο Martinez [39] μοντελοποίησαν ξεχωριστά τις τρεις βασικές ροές πληροφορίας: θέση, κίνηση και χειρομορφή. Στη συνέχεια προχώρησαν στην αναγνώριση κάθε ροής πληροφορίας ξεχωριστά και τέλος συνδύασαν τα επιμέρους αποτελέσματα μέσω μιας δενδροειδούς δομής. Οι Derpanis et al. [36] αναγνώριζαν φωνήματα κίνησης που απορέουν από την αντιστοίχιση των φωνημάτων κίνησης και των κινηματικών περιγραφητών της οπτικής κίνησης. Οι Ong και Ranganath [98] μελέτησαν τις κλίσεις των νοημάτων μοντελοποιώντας τις διάφορες παραλλαγές χρησιμοποιώντας παράλληλα-ανεξάρτητα διανύσματα χαρακτηριστικών και ένα dynamic Bayesian network. Οι Cooper et al. [28] βασίστηκαν στην εκπαίδευση πολλών ασθενών ταξινομητών και στη συνέχεια, τους συνδύασαν με στόχο τη δημιουργία ενός ταξινομητή επιπέδου νοήματος χρησιμοποιώντας Markov chains ή sequential pattern boosting. Το πρώτο σχήμα χρησιμοποιήθηκε για την μοντελοποίηση των χρονικών μεταβολών ενώ το δεύτερο για την επιλογή διακριτών χαρακτηριστικών και για την μοντελοποίηση της χρονικής εξέλιξης της πληροφορίας. Οι Han et al. [55] ασχολήθηκαν με την κατάτμηση των νοημάτων σε υπομονάδες κίνησης βασιζόμενοι στις ασυνέχειες της κίνησης. Στη συνέχεια συνδύασαν ασθενείς ταξινομητές μέσω boosting για τη δημιουργία ενός ταξινομητή επιπέδου νοήματος. Οι Yin et al. [148] ασχολήθηκαν με την κατασκευή δεδομενοκεντρικών υπομονάδων, τις οποίες αποκαλούσαν "fenemes", εφαρμόζοντας έναν αλγόριθμο για discriminative segmental feature selection. Συνοψίζοντας όλες οι παραπάνω μέθοδοι μοντελοποιούν implicit δεδομενοκεντρικές υπομονάδες σχετικές με τις παραμέτρους άρθρωσης εμπνευσμένες από τον Stokoe, και στο τέλος τις συνδυάζουν για την κατασκευή μοντέλων στο επίπεδο νοήματος.

Explicit δεδομενοκεντρικές υπομονάδες

Η άμεση (explicit) μοντελοποίηση των δεδομενοκεντρικών υπομονάδων έχει προσελκύσει το γενικότερο ενδιαφέρον. Οι Bauer & Kraiss [11] παρουσίασαν μια δεδομενοκεντρική μέθοδο για την κατάτμηση των νοημάτων σε υπομονάδες και τη μοντελοποίησή τους. Εφάρμοσαν τον αλγόριθμο συσταδοποίησης K-means με στόχο την κατασκευή ενός δεδομενοκεντρικού λεξικού, λαμβάνοντας υπόψη τα χρονικά πλαίσια του βίντεο ανεξαρτήτως δυναμικής. Επιπλέον χρησιμοποίησαν HMMs για την μοντελοποίηση κάθε υπομονάδας. Οι Fang et al. [47] χρησιμοποίησαν ένα left-right HMM τριών καταστάσεων για την κατάτμηση των νοημάτων σε υπομονάδες, λαμβάνοντας υπόψη έτσι τη δυναμική η οποία είναι θεμελιώδης στη νοηματική γλώσσα. Οι Kong & Ranganath [72] ασχολήθηκαν με την κατάτμηση των νοημάτων σε υπομονάδες εφαρμόζοντας προκατασκευασμένους κανόνες σε συνδυασμό με κατωφλιοποίηση. Χρησιμοποίησαν PCA για την εξαγωγή χαρακτηριστι-

κών και εφάρμοσαν τον αλγόριθμο συσταδοποίησης K-means για την κατασκευή των υπομονάδων. Σε όλες τις παραπάνω μεθόδους οι υπομονάδες είναι ενός τύπου. Αντίθετα οι Theodorakis et al. [126] παρουσίασαν ένα δεδομενοκεντρικό σύστημα στατιστικής μοντελοποίησης υπομονάδων το οποίο λαμβάνει υπόψη του ταυτόχρονα την έννοια και της παραλληλίας αλλά και της διαδοχής διαφορετικών φωνητικών υπομονάδων [77] εμπνευσμένοι από το μοντέλο Movement-Hold [78]. Αυτό είχε τη δυνατότητα διάκρισης χωρίς επίβλεψη μεταξύ δυναμικών και στατικών υπομονάδων, οι οποίες μοντελοποιούσαν χρονικά τμήματα κίνησης και στάσης αντιστοίχως. Συνοψίζοντας, όλες οι παραπάνω μέθοδοι βασίζονται σε δεδομενοκεντρικές υπομονάδες.

Explicit γλωσσικές-φωνητικές υπομονάδες

Οι δεδομενοκεντρικές υπομονάδες καθοδηγούνται από τα δεδομένα με αποτέλεσμα να μην υπάρχει διαίσθηση πίσω από αυτές ούτε και γλωσσολογική ερμηνεία. Από την άλλη μεριά, η χρήση γλωσσικών-φωνητικών επισημειώσεων και η κατασκευή των αντίστοιχων γλωσσικών-φωνητικών υπομονάδων μας παρέχουν αναπαραστάσεις των εσωτερικών τμημάτων ενός νοήματος οι οποίες είναι γλωσσολογικά ερμηνεύσιμες. Παρόλα αυτά μικρή πρόοδος έχει γίνει σε μεθόδους που βασίζονται σε γλωσσικές-φωνητικές υπομονάδες. Η πρώτη ερευνητική εργασία που χρησιμοποίησε φωνητικές σπισημειώσεις βασισμένες στο Movement-Hold μοντέλο παρουσιάζεται στο άρθρο [136]. Σε αυτό το άρθρο παρουσιάζεται ένα σύστημα αναγνώρισης νοηματικής γλώσσας, το οποίο βασίζεται σε γλωσσικές-φωνητικές υπομονάδες οι οποίες εκπαιδεύονται στατιστικά χρησιμοποιώντας PaHMMs. Πρόσφατα στο άρθρο [102] παρουσιάζεται μια μέθοδος η οποία εκμεταλλεύεται PDTS επισημειώσεις που είχαν εξαχθεί μέσω μιας αυτόματης συμβολικής επεξεργασίας HamNoSys επισημειώσεων, για την εκπαίδευση γλωσσικών-φωνητικών υπομονάδων. Επιπλέον εφαρμόζοντας τον αλγόριθμο Viterbi αντιστοιχίζονται οι ακολουθίες των PDTS υπομονάδων με τις πραγματικές παρατηρήσεις. Στη συνέχεια, στο άρθρο [4] οι συγγραφείς επεκτείνουν το παραπάνω σύστημα προς τις εξής κατευθύνσεις: α) Εισαγωγή του Iterative Training Algorithm (ITA), β) Παρουσίαση του πλαισίου μοντελοποίησης multi-stream switching probability distribution (MSSD) HMM, γ) Ενσωμάτωση του δευτερεύοντος χεριού και της ροής πληροφορίας της χειρομορφής και δ) Εξονυχιστικά πειράματα τα οποία περιλαμβάνουν άγνωστο νοηματιστή, συγκρίσεις με άλλες μεθόδους από την βιβλιογραφία και προσαρμογή σε νέο νοηματιστή. Μια πρόσφατη και αρκετά κοντινή ερευνητική προσπάθεια, είναι η εργασία των Koller et al. [71]. Οι συγγραφείς σε αυτό το άρθρο παρουσιάζουν ένα επαναληπτικό αλγόριθμο αναγκαστικής ευθυγράμμισης (iterative forced alignment algorithm) για την εύρεση της καλύτερης ακολουθίας γλωσσικών-φωνητικών υπομονάδων δεδομένου των οπτικών παρατηρήσεων. Οι συγγραφείς χρησιμοποιούν ένα λεξικό επιπέδου υπομονάδας το οποίο έχει κατασκευαστεί από φωνητικές επισημειώσεις, οι οποίες βασίζονται στο SignWriting σύστημα. Αυτή η μέθοδος έχει αρκετές ομοιότητες με τον προτεινόμενο αλγόριθμο ITA που παρουσιάζεται στην ενότητα 4.3. Και οι δύο προσεγγίσεις συσχετίζουν τις γλωσσικές-φωνητικές υπομονάδες με τις πραγματικές παρατηρήσεις. Παρόλα αυτά έχουν και αρκετές διαφορές. Ο αλγόριθμος ITA δίνει την δυνατότητα τροποποίησης μιας ακολουθίας υπομονάδων η οποία εμφανίζεται στο λεξικό, βασιζόμενος σε συγκεκριμένους γραμματικούς κανόνες, επιτρέποντας έτσι μεταβλητότητα στην άρθρωση ενός νοήματος. Επιπλέον γίνεται αξιοποίηση των ροών πληροφορίας και της κίνησης-θέσης αλλά και της χειρομορφής, σε αντίθεση με το άρθρο [71] όπου χρησιμοποιείται μόνο η ροή της κίνησης. Ακόμα οι υπομονάδες βασίζονται σε PDTS επισημειώσεις αντί για SignWriting επισημειώσεις. Τέλος, αυτές οι γλωσσικές-φωνητικές υπομονάδες χρησιμοποιούνται σε πειράματα αναγνώρισης νοηματικής γλώσσας.

Πειραματική αξιολόγηση: δεδομένα εκπαίδευσης και νοηματιστές

Για την αξιολόγηση ενός συστήματος αναγνώρισης νοηματικής γλώσσας δύο σημαντικές παράμετροι είναι ο αριθμός των δεδομένων προς εκπαίδευση και η αξιολόγηση σε δεδομένα από

άγνωστο νοηματιστή. Οι Wang et al. στο άρθρο [138] παρουσίασαν ένα εργαλείο αυτόματης αναζήτησης νοημάτων από ένα λεξικό νοηματικής προτείνοντας για την αναγνώριση μια exemplar-based μέθοδο, η οποία βασιζόταν στον αλγόριθμο DTW. Αυτή η μέθοδος είχε την δυνατότητα εκπαίδευσης χρησιμοποιώντας πολύ μικρό αριθμό δεδομένων. Χρησιμοποιώντας το παραπάνω σύστημα παρουσίασαν αποτελέσματα αναγνώρισης 78%, λαμβάνοντας ως σωστό ένα παράδειγμα προς αναγνώριση εάν το σωστό νόημα άνηκε στα πρώτα 10 πιο πιθανά νοήματα που επέλεγε η μέθοδος προς αξιολόγηση. Τα πειράματα πραγματοποιήθηκαν σε μια βάση δεδομένων η οποία περιείχε 1113 διαφορετικά νοήματα, δύο επαναλήψεις ανά νόημα για την εκπαίδευση του συστήματος και η αξιολόγηση έγινε σε άγνωστο νοηματιστή. Οι Kadir et al. [66] παρουσίασαν αποτελέσματα αναγνώρισης 76% σε 164 διαφορετικά νοήματα αξιολογώντας γνωστό νοηματιστή και χρησιμοποιώντας ένα παράδειγμα ανά νόημα για την εκπαίδευση του συστήματος. Άλλες ενδεικτικές ερευνητικές προσπάθειες όπου η αξιολόγηση γίνεται σε άγνωστο νοηματιστή παρουσιάζονται στα άρθρα [45, 6, 152, 28] Πιο συγκεκριμένα, οι Cooper et al. [28] παρουσίασαν αποτελέσματα αναγνώρισης 76% και 49.4%, σε λεξιλόγια 20 και 40 διαφορετικών νοημάτων αντιστοίχως. Οι Fang et al. [45] παρουσίασαν αποτελέσματα μέχρι 92% σε 208 διαφορετικά νοήματα χρησιμοποιώντας όμως dataglove για την εξαγωγή χαρακτηριστικών. Συνοψίζοντας, η αξιολόγηση σε άγνωστο νοηματιστή επιδεινώνει δραστικά τα αποτελέσματα αναγνώρισης, σε σύγκριση με την αξιολόγηση σε γνωστό νοηματιστή. Παραδείγματος χάριν: 55% για 232 νοήματα στο άρθρο [152], 16% και 10.4% για 20 και 40 νοήματα αντιστοίχως στο άρθρο [28]. Στην έρευνά μας παρουσιάζουμε πειραματικά αποτελέσματα χρησιμοποιώντας μέχρι και ένα παράδειγμα ανά νόημα για την εκπαίδευση του συστήματος και αξιολογώντας γνωστό αλλά και άγνωστο νοηματιστή.

Πίνακας 1.1: Ενδεικτικές έρευνες σχετικές με την αυτόματη αναγνώριση νοηματικής γλώσσας[†].

Works		Sensor/FE	SU Segm.	SU Constr.	Modeling	Linguistic Transcriptions	SU Seq.	Parallel Fusion	Unseen Signer
[138]	Sign-level	Vis.	✗	✗	exemplar based (DTW)	✗	✗	✗	✓
[152]		Vis.	✗	✗	HMMs	✗	✗	✗	✓
[137]		Vis.	✗	✗	HMMs/RNN	✗	✗	✗	n.a.
[116]		Vis.	✗	✗	HMMs	✗	✗	✗	n.a.
[45]		d-gloves	✗	✗	SRN/HMMs	✗	✗	✗	✓
[6]		Vis.	✗	✗	Multi-Class Fisher Score	✗	✗	✗	✓
[148]	SU-implicit	d-gloves	SBHMMs	DIST	HMMs	✗	✗	✗	n.a.
[54]		Vis.	✗	✗	MLP/HMM	✗	✗	✗	✗
[98]		MoCap	✗	✗	DBN/MH-HMM	✗	✗	✓	✗
[55]		Vis.	motion disc.	DTW	WC/Adaboost	✗	✗	✗	✗
[39]		Vis.	✗	✗	Tree-based	✗	✗	✗	✓
[28]		Vis.	rule-based	WC	SP,MC	✗	✗	✗	✓
[66]		Vis.	rule-based	WC	MC	✗	✗	✗	✗
[11]	SU-explicit	Vis.+c-gloves	K-means	K-means	HMM	✗	✗	✗	n.a.
[47]		d-gloves	LR-HMM	MKM-DTW	HMM	✗	✗	✗	n.a.
[46]		d-gloves	LR-HMM	MKM-DTW	HMM + Epenthesis	✗	✗	✗	n.a.
[72]		MoCap	rule-based	K-means	HMM	✗	✗	✗	✗
[9]		Vis.	motion disc.	DTW, hier. clust.	Adaboost	✗	✗	✗	✗
[126]		Vis.	2S-ERG HMM	K-means, DTW	MSSD-HMM, PaHMM	✗	✓	✓	✓
[136]	SU-ling.	MoCap	✗	✗	PaHMM + Epenthesis	M-H Model	✓	✓	n.a.
[71]		Vis.	✗	✗	HMMs	SignWriting	✗	✗	n.a.
[102]		Vis.	✗	✗	HMMs	PDTS	✓	✗	✗
[127]		Vis.	✗	✗	MSSD-HMM, PaHMM	PDTS	✓	✓	✓

[†] FE είναι η συντομογραφία του feature extraction, Segm. του segmentation, SU Seq. του SU sequentiality και SU-ling. του SU-linguistic. Vis. είναι η συντομογραφία του visual processing, d-gloves του datagloves, MoCap του various motion capture devices και c-gloves του color gloves. SBHMMs είναι η συντομογραφία του Segmentally Boosted HMMs, LR-HMM του left-right HMM, motion disc. του motion discontinuities και 2S-ERG HMM του two-state ergodic HMM. DIST είναι η συντομογραφία του discriminative state-space tying, MKM-DTW του modified K-means χρησιμοποιώντας dynamic time warping (DTW) και hier. clust. του hierarchical clustering. DBN είναι η συντομογραφία του dynamic bayesian network, MLP του multilayer perceptron, MH-HMM του multichannel hierarchical HMM, WC του weak classifiers, SP του sequential pattern boosting και MC του Markov chains. Τέλος χρησιμοποιούμε το σύμβολο n.a., το οποίο είναι η συντομογραφεία του non-availability, όπου η συγκεκριμένη πληροφορία δεν είναι διαθέσιμη από την αντίστοιχη δημοσίευση.

1.2.3 Πολυτροπική αναγνώριση χειρονομιών

Παρά τις ερευνητικές προσπάθειες που έχουν γίνει, η πολυτροπική αναγνώριση χειρονομιών, θεωρείται μια αρκετά ανεξερεύνητη ερευνητική περιοχή που συγκοινωνεί με πολλά ερευνητικά πεδία όπως η αναγνώριση φωνής και η αναγνώριση χειρονομιών και νοηματικής γλώσσας. Η πολυτροπική αναγνώριση χειρονομιών αποτελεί ένα πολύπλευρο πρόβλημα που θέτει σημαντικές προκλήσεις στην επεξεργασία ακουστικών και οπτικών σημάτων, στην μοντελοποίηση πολλαπλών ροών πληροφορίας και στην σύμμειξη αυτών. Στη συνέχεια αναφέρουμε ερευνητικές προσπάθειες σχετικές με τις βασικές συνιστώσες της έρευνάς μας όπως η χρονική ανίχνευση και κατάτμηση σημαντικών πολυτροπικών γεγονότων, η στατιστική μοντελοποίηση και η σύμμειξη πολλαπλών ροών πληροφορίας. Επιπλέον συνοψίζουμε μερικές ενδεικτικές προσεγγίσεις από τις ομάδες οι οποίες συμμετείχαν στα πλαίσια του διαγωνισμού αναγνώρισης χειρονομιών από πολυτροπικά δεδομένα CHALEARN [22].

Ανίχνευση και κατάτμηση σημαντικών γεγονότων

Η χρονική ανίχνευση και κατάτμηση γεγονότων σημαντικής σημασίας σε πολλαπλές ροές πληροφορίας είναι ένα αρκετά δύσκολο αλλά και ενδιαφέρον πρόβλημα. Συχνά το πρόβλημα αυτό αντιμετωπίζεται στα πλαίσια του προβλήματος εντοπισμού χειρονομιών ή νοημάτων σε ένα βίντεο ή με άλλα λόγια την προσπάθεια εύρεσης των χρονικών τμημάτων που αντιστοιχούν σε χειρονομίες ή νοήματα. Μια πολύ δημοφιλής μέθοδος που χρησιμοποιείται ευρέως είναι η ερευνητική εργασία των Wilcox και Bush [140]. Οι συγγραφείς σε αυτό το άρθρο χρησιμοποίησαν ένα μοντέλο filler με στόχο τη μοντελοποίηση των μη-σημαντικών προτύπων. Εν συνεχεία, οι Lee και Kim [74] χρησιμοποίησαν με αντίστοιχο τρόπο ένα εργοδικό μοντέλο, το οποίο ονόμασαν μοντέλο threshold, με στόχο τον καθορισμό προσαρμοσμένων κατωφλιών πιθανότητας. Η κατάτμηση μπορεί επίσης να ιδωθεί σε συνδυασμό με την αναγνώριση όπως στα άρθρα [3, 75]. Στο δεύτερο άρθρο η αρχή και το τέλος κάθε χειρονομίας καθοριζόταν από τα zero crossing της διαφοράς των πιθανοτήτων μεταξύ των μοντέλων χειρονομιών και μη-χειρονομιών. Σε άλλα σχετικά προβλήματα όπως στην αναγνώριση νοηματικής γλώσσας οι Han et al. [55] ασχολήθηκαν με την κατάτμηση νοημάτων σε υπομονάδες βασιζόμενοι στις ασυνέχειες των κινήσεων. Οι Kong και Ranganath [72] ασχολήθηκαν με την κατάτμηση των νοημάτων σε υπομονάδες εφαρμόζοντας προκατασκευασμένους κανόνες σε συνδυασμό με κατωφλιοποίηση. Στα πλαίσια της έρευνας μας ανεξαρτήτως της ύπαρξης των χρονικών επισημειώσεων των χειρονομιών χρησιμοποιούμε μια μέθοδο πολυτροπικής ανίχνευσης δράσης για κάθε ροή πληροφορίας ξεχωριστά η οποία βασίζεται σε ένα κοινό HMM πλαίσιο. Η εφαρμογή της είναι ανεξάρτητη των χρονικών επισημειώσεων και το αποτέλεσμά της χρησιμοποιείται για την στατική εκπαίδευση και μοντελοποίηση των χειρονομιών.

Μοντελοποίηση πολυτροπικών χειρονομιών

Για την πολυτροπική αναγνώριση χειρονομιών απαιτείται η διαχείριση πολλαπλών ροών πληροφορίας οι οποίες μεταβάλλονται δυναμικά. Οι χειρονομίες προς αναγνώριση μπορεί να εμφανίζονται σε διαφορετικές στιγμές και να έχουν διαφορετικές χρονικές διάρκειες σε κάθε ροή πληροφορίας. Για την μοντελοποίησή τους τα HMMs αποτελούν μια αρκετά δημοφιλή λύση λόγω της ιδιότητας που έχουν να μοντελοποιούν αποτελεσματικά χρονικές ροές πληροφορίας. Επιπλέον προσφέρουν αποτελεσματικούς αλγορίθμους για την εκπαίδευση και αξιολόγησή τους, όπως ο Baum-Welch και ο Viterbi [105]. Ενδεικτικά έχουν χρησιμοποιηθεί με σκοπό την αναγνώριση χειρονομιών [87], την ανίχνευσή τους [74] όπως επίσης για την αναγνώριση χειρονομιών με συστηματικές μεταβλητότητες στον τρόπο άρθρωσης [142]. Επιπλέον τα Parallel HMM (PaHMM) [136], μια επέκταση των κλασσικών HMM, δίνουν τη δυνατότητα μοντελοποίησης πολλαπλών παράλληλων ροών πληροφορίας. Επεκτάσεις των HMM αποτελούν επίσης τα CRFs [139]. Επιπλέον άλλες

μη-παραμετρικές μέθοδοι επίσης χρησιμοποιούνται με σκοπό την αναγνώριση πολυτροπικών χειρονομιών [21, 58]. Στα πλαίσια της έρευνάς μας χτίζουμε ένα HMM μοντέλο ανά χειρονομία για κάθε ροή πληροφορίας ξεχωριστά και επιπλέον χρησιμοποιούμε τα PaHMM με σκοπό την σύμμειξη των πολλαπλών ροών πληροφορίας.

Μέθοδοι από τον CHALEARN διαγωνισμό

Ανάμεσα στις πρόσφατες δημοσιευμένες μεθόδους οι οποίες κατέκτησαν τις πρώτες θέσεις στον διαγωνισμό αναγνώρισης χειρονομιών από πολυτροπικά δεδομένα CHALEARN αρκετές χρησιμοποίησαν με σκοπό τη μοντελοποίηση, εκπαίδευση και αναγνώριση, μεθόδους όπως HMM/GMMs, boosting, random forests, neural networks και support vector machines. Μια γενικότερη σύνοψη έχει δημοσιευθεί στο άρθρο [44]. Στη συνέχεια αναφέρουμε ενδεικτικά μερικές από αυτές. Οι Wu et al. [143], η ομάδα που κατέκτησε την πρώτη θέση, χρησιμοποίησε ως οδηγό την ακουστική πληροφορία βασιζόμενη σε end-point ανίχνευση. Στη συνέχεια συνδύασαν τους επιμέρους ταξινομητές υπολογίζοντας ένα κανονικοποιημένο σκορ εμπιστοσύνης. Οι Bayer και Thierry [12], επίσης καθοδηγούμενοι από την ακουστική πληροφορία υπολόγισαν τις πιθανότητες κάθε χειρονομίας ανά χρονικό τμήμα και ροή πληροφορίας. Για τη σύμμειξη των ροών πληροφορίας υπολόγισαν έναν σταθμισμένο μέσον όρο των επιμέρους πιθανοτήτων. Οι Nandakumar et al. [88] καθοδηγήθηκαν και από την ακουστική πληροφορία αλλά και από την πληροφορία του σκελετού. Απέρριψαν τα χρονικά τμήματα τα οποία δεν ήταν κοινά και στις δύο ροές πληροφορίας και επιπλέον χρησιμοποίησαν ένα χρονικό συντελεστή επικάλυψης για την συνένωση των χρονικών τμημάτων από τις διαφορετικές ροές. Τέλος, επέλεγαν την χειρονομία με το μεγαλύτερο συνδυασμένο σκορ. Οι Chen και Koskela [24] χρησιμοποίησαν extreme learning machine, ένα τύπο των single-hidden layer feed-forward neural network, και εφάρμοσαν ταυτόχρονα και εκ των προτέρων (early) αλλά και εκ των υστέρων (late) σύμμειξη. Κατά τη διάρκεια της εκ των υστέρων σύμμειξης των αποτελεσμάτων ταξινόμησης χρησιμοποίησαν τον γεωμετρικό μέσο. Τέλος οι Neverova et al. [92] πρότειναν έναν πολυ-κλιμακωτό αλγόριθμο εκμάθησης ο οποίος εφαρμόστηκε ταυτόχρονα στη χωρική και χρονική συνιστώσα ενώ επιπλέον χρησιμοποίησαν ένα recurrent neural network.

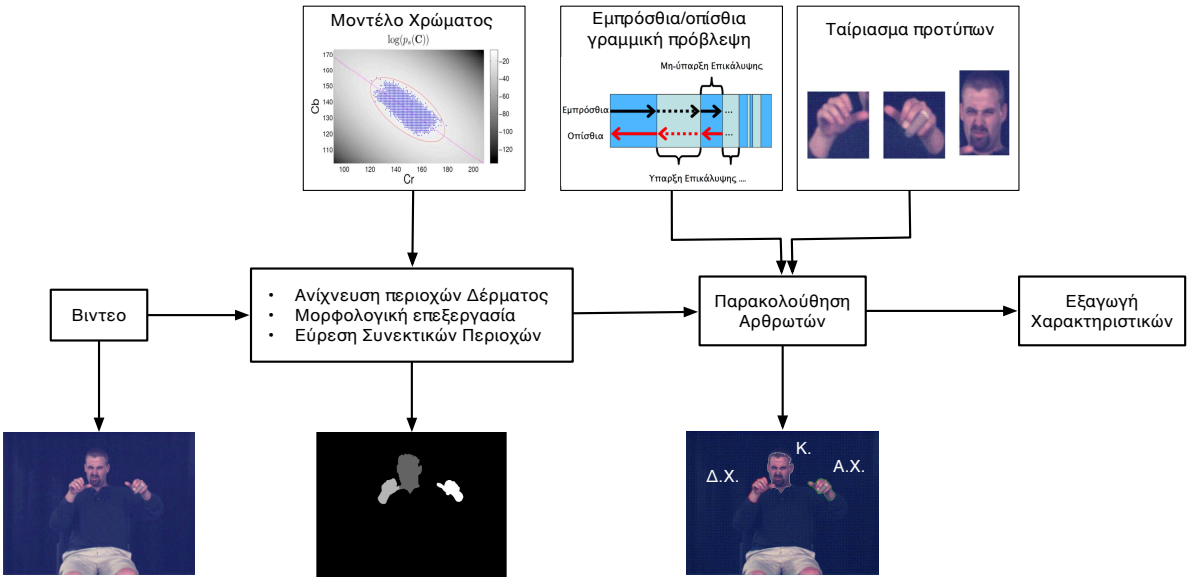
1.3 Ερευνητικές συνεισφορές

Σε αυτή την ενότητα θα παρουσιάσουμε περιληπτικά τις κύριες ερευνητικές συνεισφορές της διδακτορικής μας έρευνας. Αυτές μπορούν να συνοψισθούν στις παρακάτω κατηγορίες:

- *Εξαγωγή χαρακτηριστικών για την αναγνώριση ΝΓ,*
- *Στατιστική μοντελοποίηση της ΝΓ με δεδομενοκεντρικές υπομονάδες,*
- *Στατιστική μοντελοποίηση της ΝΓ με γλωσσικές-φωνητικές υπομονάδες,*
- *Στατιστικά μοντέλα υπομονάδας, σύμμειξη και προσαρμογή σε άγνωστο νοηματιστή,*
- *Αναγνώριση χειρονομιών από πολυτροπικά δεδομένα.*

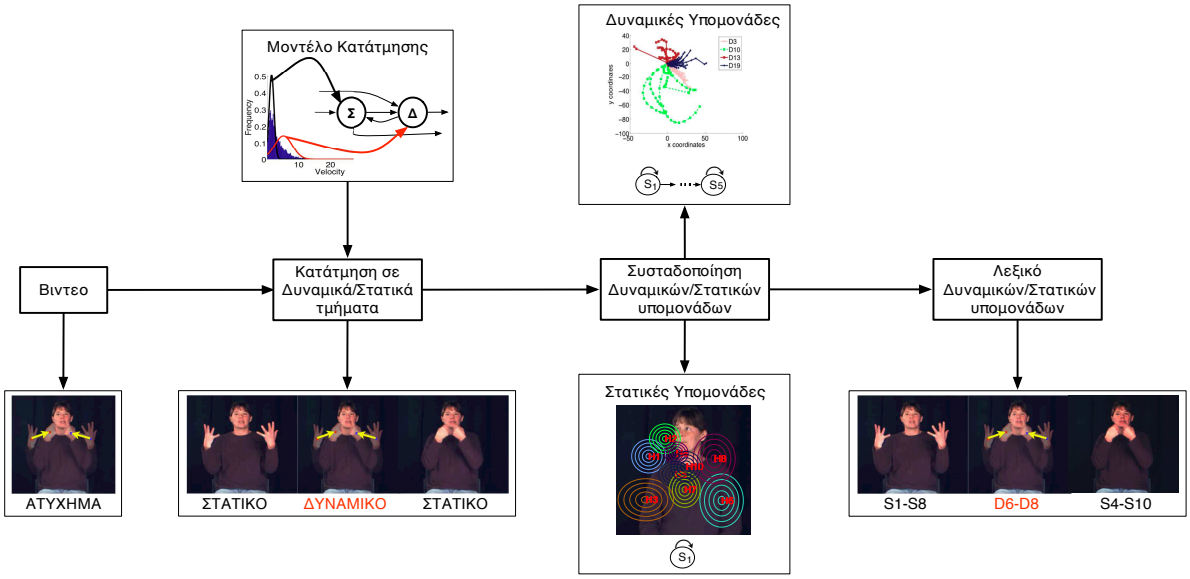
1.3.1 Εξαγωγή χαρακτηριστικών για την αναγνώριση ΝΓ

Αναπτύχθηκε ένα σύστημα εξαγωγής χαρακτηριστικών σχετιζόμενων με τους αρθρωτές που συμμετέχουν κατά τη διάρκεια άρθρωσης της νοηματικής γλώσσας. Μια σύνοψη του συστήματος απεικονίζεται στο Σχήμα 1.1. Πιο συγκεκριμένα, για την ανίχνευση των περιοχών χρώματος δέρματος και την εξαγωγή των αντίστοιχων μασκών δέρματος έγινε χρήση ενός πιθανοτικού μοντέλου χρώματος. Στη συνέχεια εφαρμόστηκε μορφολογική επεξεργασία των εξαγόμενων μασκών



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

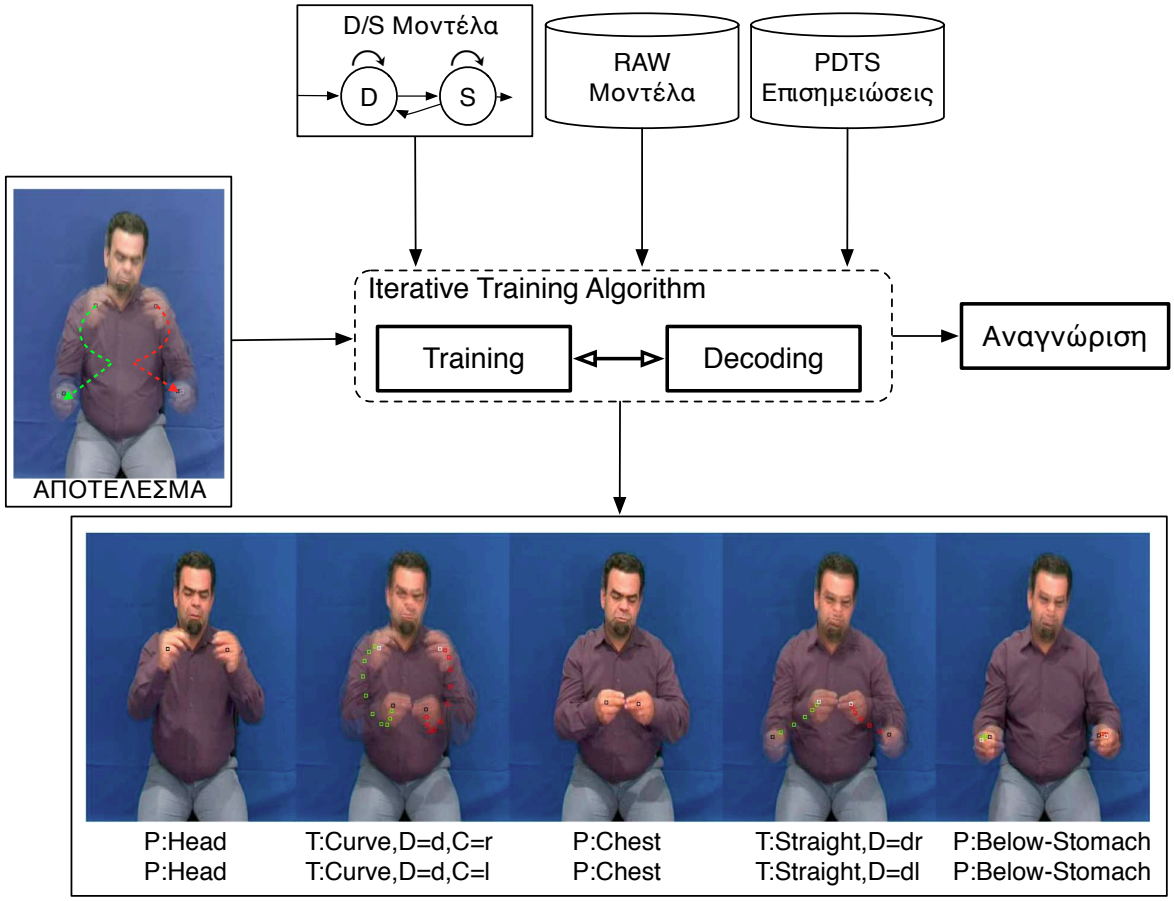
δέρματος για την ομαλοποίησή τους. Για την μορφολογική κατάτμηση των συνεκτικών περιοχών αναπτύχθηκε η μέθοδος competitive reconstruction opening η οποία βασίζεται σε αλλεπάλληλα μορφολογικά φιλτραρίσματα. Για την αντιστοίχιση των συνεκτικών περιοχών με τους αρθρωτές (χέρια, κεφάλι) συνδυάσαμε εμπρόσθια, οπίσθια γραμμική πρόβλεψη με την τεχνική ταιριάσματος προτύπου (template matching). Τέλος, εφαρμόστηκαν μέθοδοι για την εξαγωγή κατάλληλων χαρακτηριστικών διανυσμάτων που σχετίζονται με την θέση, κίνηση και το σχήμα των αρθρωτών. Στο κεφάλαιο 2 που ακολουθεί περιγράφουμε με περισσότερη λεπτομέρεια το συνολικό σύστημα.



Σχήμα 1.2: Σύνοψη συστήματος: Μοντελοποίηση της ΝΓ με δεδομενοκεντρικές υπομονάδες.

1.3.2 Στατιστική μοντελοποίηση της ΝΓ με δεδομενοκεντρικές υπομονάδες

Αναπτύξαμε ένα καινοτόμο σύστημα για την μοντελοποίηση και αναγνώριση της ΝΓ με δεδομενοκεντρικές υπομονάδες. Στο Σχήμα 1.2 παρουσιάζουμε μια σύνοψη του συνολικού συστήματος. Οι υπομονάδες στη νοηματική γλώσσα αποτελούν τις βασικές δομικές μονάδες που απαρτίζουν κάθε νόημα. Για την κατάτμηση κάθε νοήματος στις υπομονάδες από τις οποίες αποτελείται, αναπτύχθηκε ένας αλγόριθμος κατάτμησης σε **στατικά και δυναμικά (Δ/Σ)** τμήματα εμπνευσμένος από το γλωσσικό μοντέλο Movement-Hold των Liddell και Johnson. Συγκεκριμένα έγινε χρήση κρυφών Μαρκοβιανών μοντέλων (HMMs) για την μοντελοποίηση και αυτόματη κατάτμηση. Για την ομαδοποίηση και κατασκευή των Δ/Σ υπομονάδων χρησιμοποιήθηκαν αλγόριθμοι αυτόματης συσταδοποίησης (K-means, agglomerative hierarchical clustering) όπως και ο αλγόριθμος DTW. Η αποδόμηση κάθε νοήματος στις Δ/Σ υπομονάδες από τις οποίες αποτελείται έχει ως αποτέλεσμα την κατασκευή ενός δεδομενοκεντρικού λεξικού επιπέδου υπομονάδας. Για την εκπαίδευση των υπομονάδων έγινε χρήση στατιστικών HMM μοντέλων. Τέλος για την αναγνώριση της νοηματικής γλώσσας χρησιμοποιήθηκαν τα HMM μοντέλα υπομονάδας σε συνδυασμό με το αντίστοιχο λεξικό υπομονάδας. Στο κεφάλαιο 3 παρουσιάζεται αναλυτικά το συνολικό σύστημα.



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

1.3.3 Στατιστική μοντελοποίηση της ΝΓ με φωνητικές υπομονάδες

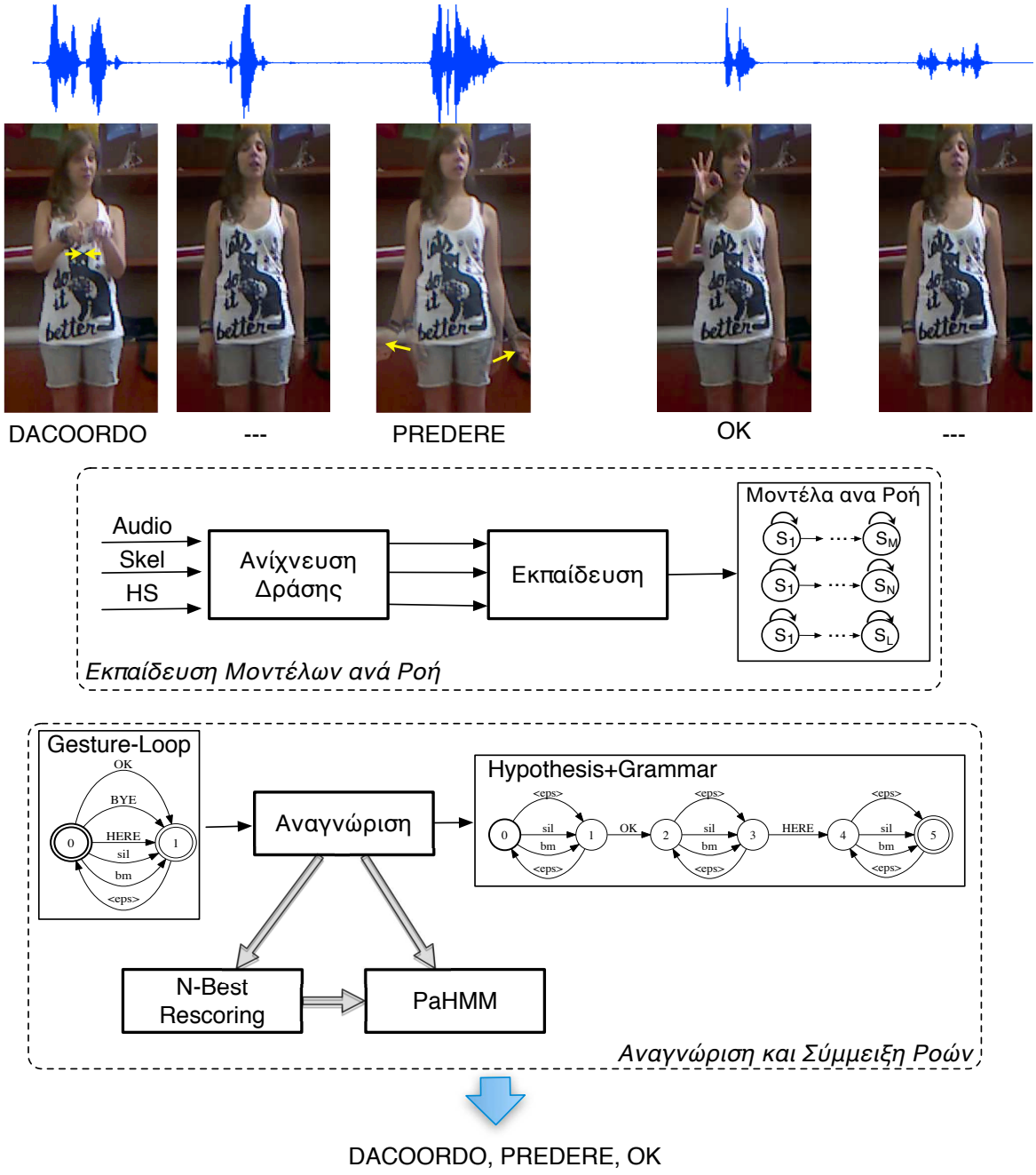
Αναπτύχθηκε ένα καινοτόμο γλωσσικό-φωνητικό πλαίσιο για την μοντελοποίηση και αυτόματη αναγνώριση της νοηματικής γλώσσας. Στο Σχήμα 1.3 παρουσιάζουμε μια σύνοψη του συστήματος. Η κύρια συνεισφορά αποτελεί την ενσωμάτωση γλωσσικής-φωνητικής γνώσης στα στατιστικά HMM μοντέλα υπομονάδων. Χρησιμοποιήθηκε πρότερη γλωσσική-φωνητική γνώση, η οποία ήταν ενσωματωμένη στις επισημειώσεις PDTS, οι οποίες εξάχθηκαν χρησιμοποιώντας ένα σύστημα αυτόματης συμβολικής επεξεργασίας επισημειώσεων HamNoSys. Με αυτόν τον τρόπο, κατασκευάζουμε υπομονάδες που είναι γλωσσικά και φωνητικά ερμηνεύσιμες. Για την εκπαίδευση των PDTS υπομονάδων χρησιμοποιήθηκαν HMMs και επιπλέον προτάθηκε ο αλγόριθμος εκπαίδευσης Iterative Training Algorithm (ITA). Αυτός είχε τη δυνατότητα να αλλάζει τις επισημειώσεις PDTS έτσι ώστε να είναι συνεπής σε σχέση με την πραγματική άρθρωση των νοημάτων, εκμεταλλευόμενος τις πραγματικές οπτικές παρατηρήσεις. Τέλος, η αναγνώριση της νοηματικής γλώσσας έγινε χρησιμοποιώντας τα PDTS HMM μοντέλα υπομονάδας και το PDTS λεξικό. Στο κεφάλαιο 4 παρουσιάζεται αναλυτικά το συνολικό σύστημα.

1.3.4 Στατιστικά μοντέλα υπομονάδων, σύμμειξη και προσαρμογή σε άγνωστο νοηματιστή

Για τη μοντελοποίηση των υπομονάδων προτάθηκε το πλαίσιο MSSD HMM. Το πλαίσιο MSSD-HMM επέτρεπε με έναν φορμαλιστικό τρόπο να χρησιμοποιούνται διαφορετικά σετ διανυσμάτων χαρακτηριστικών, ανάλογα με τον τύπο της υπομονάδας που μοντελοποιείται χωρίς να γνωρίζουμε εκ των προτέρων τη χρονική κατάτμηση στους διαφορετικούς τύπους υπομονάδων. Χαρακτηριστικό που τα συμβατικά multistream HMM δεν μπορούν να ικανοποιήσουν. Επιπλέον έγινε επέκταση των MSSD-HMM μοντέλων και χρησιμοποιήθηκαν κατάλληλοι αλγόριθμοι για την σύμμειξη πολλαπλών ροών πληροφορίας: κίνηση, θέσης και χειρομορφή του κυρίαρχου/δευτερεύοντος χεριού. Χρησιμοποιώντας ένα μικρό σύνολο δεδομένων και τους καταλλήλους αλγορίθμους, αναπτύχθηκε ένα σύστημα για την προσαρμογή των μοντέλων και του λεξικού υπομονάδας σε άγνωστο νοηματιστή. Η προσαρμογή των μοντέλων υπομονάδας έγινε χρησιμοποιώντας Maximum Likelihood Linear Regression (MLLR) και του λεξικού υπομονάδων αυξάνοντας την ποικιλία άρθρωσης που υπήρχε στο λεξικό, εισάγοντας νέες προφορές των νοημάτων. Στο κεφάλαιο 5 παρουσιάζονται αναλυτικά όλα τα παραπάνω.

1.3.5 Αναγνώριση χειρονομιών από πολυτροπικά δεδομένα

Αναπτύχθηκε ένα σύστημα για την ανίχνευση και αναγνώριση χειρονομιών από πολυτροπικά δεδομένα, τα οποία περιλάμβαναν οπτικές αλλά και ακουστικές ροές πληροφορίας. Στο Σχήμα 1.4 παρουσιάζουμε μια σύνοψη του συστήματος. Συγκεκριμένα το σύστημα πολυτροπικής αναγνώρισης χειρονομιών εκμεταλλεύεται ροές πληροφοριών που σχετίζονται με το χρώμα, το βάθος και τη φωνή όπως έχουν καταγραφεί από τον αισθητήρα Kinect. Εξάγονται χαρακτηριστικά σχετιζόμενα με την χειρομορφή, την κίνηση των χεριών και το σήμα φωνής. Η μοντελοποίηση για κάθε ροή πληροφορίας βασίστηκε σε HMM μοντέλα χειρονομιών. Επιπλέον αναπτύχθηκε ένα αυτόματο σύστημα ανίχνευσης δράσης (activity detection) για την εκπαίδευση των HMM μοντέλων. Για την αναγνώριση αρχικά παράγονται υποθέσεις αναγνώρισης για κάθε ροή πληροφορίας ξεχωριστά χρησιμοποιώντας τα διανύσματα χαρακτηριστικών και τα εκπαιδευμένα HMM μοντέλα. Στη συνέχεια, επαναξιολογούνται οι υποθέσεις αναγνώρισης χρησιμοποιώντας ένα πιθανοτικό πολυτροπικό πλαίσιο σύμμειξης (N-best rescoring). Μετά, δεδομένου της πιο πιθανής υπόθεσης από την σύμμειξη των πολυτροπικών ροών πληροφορίας και τη χρονική κατάτμηση κάθε χειρονομίας σε κάθε ροή πληροφορίας ξεχωριστά, εφαρμόζεται ένα τελικό σχήμα σύμμειξης βασιζόμενο σε PaHMM. Στο κεφάλαιο 6 παρουσιάζεται αναλυτικά το συνολικό σύστημα.



Σχήμα 1.4: Σύνοψη συστήματος: Αναγνώριση χειρονομιών από πολυτροπικά δεδομένα.

Κεφάλαιο 2

Εξαγωγή Χαρακτηριστικών για την Αναγνώριση Νοηματικής Γλώσσας

Το πρώτο βήμα για την αυτόματη αναγνώριση της νοηματικής γλώσσας είναι η εξαγωγή χαρακτηριστικών από βίντεο νοηματικού λόγου. Πιο συγκεκριμένα, είναι απαραίτητη η εξαγωγή κατάλληλων χαρακτηριστικών που σχετίζονται με τους αρθρωτές που συμμετέχουν κατά την παραγωγή της νοηματικής γλώσσας. Τα χέρια και το κεφάλι του νοηματιστή αποτελούν τους βασικούς αρθρωτές για την άρθρωση της νοηματικής γλώσσας. Σε αυτό το κεφάλαιο θα παρουσιάσουμε ένα σύστημα για την ανίχνευση και παρακολούθηση αυτών των αρθρωτών καθώς και για την εξαγωγή χαρακτηριστικών που σχετίζονται με αυτούς.

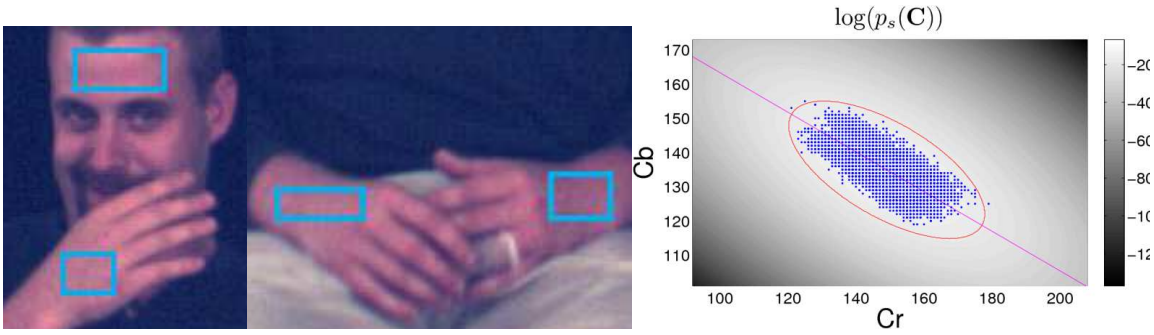
Για την εξαγωγή χαρακτηριστικών από βίντεο νοηματικής γλώσσας χρησιμοποιούμε το σύστημα που παρουσιάστηκε στα άρθρα [7, 110]. Αυτό το σύστημα αποτελείται από τα εξής: α) Ανίχνευση των περιοχών χρώματος δέρματος και κατασκευή των αντίστοιχων μασκών δέρματος, χρησιμοποιώντας ένα πιθανοτικό μοντέλο χρώματος, β) Μορφολογική επεξεργασία των εξαγόμενων μασκών δέρματος, γ) Μορφολογική κατάτμηση των εξαγόμενων μασκών δέρματος, δ) Παρακολούθηση και αντιστοίχιση των συνεκτικών περιοχών δέρματος με τους αρθρωτές και ε) Επίλυση/αποσαφήνιση των επικαλύψεων μεταξύ των αρθρωτών.

2.1 Ανίχνευση κεφαλιού και χεριών του Νοηματιστή

Για την επεξεργασία των βίντεο νοηματικής βασιζόμαστε στη πληροφορία του χρώματος εμπνευσμένοι από υπάρχουσες προσεγγίσεις [17, 7, 18]. Βασιζόμενοι στην παρατήρηση ότι το χρώμα του δέρματος του ανθρώπου έχει ιδιαίτερα 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, γεγονός όχι αρκετά περιοριστικό καθώς το χρώμα δέρματος κάθε νοηματιστή έχει αρκετά μικρή διακύμανση [20].

2.1.1 Πιθανοτικό Μοντέλο Χρώματος Δέρματος

Για την μοντελοποίηση του χρώματος δέρματος χρησιμοποιούμε μια Γκαουσιανή κατανομή με πλήρη πίνακα συμμεταβλητότητας στον χρωματικό χώρο *YCbCr* κρατώντας μόνο τις δύο χρωματικές συνιστώσες C_b,C_r. Με αυτό τον τρόπο αποκτάμε σταθερότητα σε τυχόν μεταβολές στη φωτεινότητα [20]. Υποθέτουμε ότι στα εικονοστοιχεία όπου έχουμε δέρμα οι τιμές των (C_b, C_r) ακολουθούν μια διμεταβλητή Γκαουσιανή κατανομή πυκνότητας πιθανότητας $p_s(C_b, C_r)$, η οποία



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 C_b-C_r μαζί με την κανονική κατανομή $p_s(C_b, C_r)$. Η ευθεία γραμμή αντιστοιχεί στην πρώτη ιδιοκατεύθυνση εφαρμόζοντας PCA στα δεδομένα εκπαίδευσης.

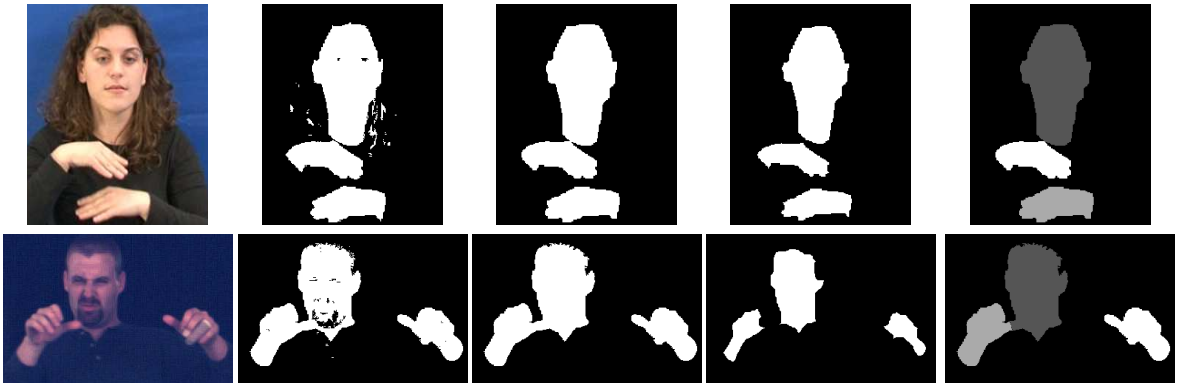
εκπαιδεύεται χρησιμοποιώντας επισημειωμένες περιοχές δέρματος. Αξίζει να σημειωθεί ότι πειραματιστήκαμε και με ένα μείγμα πολλαπλών Γκαουσιανών κατανομών επιτυγχάνοντας αρκετά αντίστοιχα αποτελέσματα. Ενδεικτικά παραδείγματα από επισημειωμένες περιοχές δέρματος απεικονίζονται στα Σχήματα 2.1(α,β). Για την ανίχνευση των περιοχών που αντιστοιχούν σε περιοχές δέρματος αρχικά υπολογίζουμε την πιθανότητα κάθε εικονοστοιχείου της εικόνας να ανήκει στην παραπάνω Γκαουσιανή κατανομή. Στη συνέχεια για την εκτίμηση της μάσκας δέρματος S_0 εφαρμόζουμε κατωφλιοποίηση της πιθανότητας $p_s(C_b(\vec{x}), C_r(\vec{x}))$ για κάθε εικονοστοιχείο $\vec{x}$ της εικόνας, όπου p_s είναι η εκπαιδευμένη κατανομή του χρώματος δέρματος Σχήμα 2.1(γ). Η τιμή της σταθεράς για την κατωφλιοποίηση υπολογίζεται έτσι ώστε ένα ποσοστό των δεδομένων εκπαίδευσης να ταξινομείται 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%. Ένα παράδειγμα της μάσκας δέρματος S_0 για ένα πλαίσιο του βίντεο απεικονίζεται στο Σχήμα 2.2(β). Η χρησιμοποίηση ενός χρωματικού μοντέλου για την ανίχνευση περιοχών δέρματος διευκολύνει την προσαρμογή σε διαφορετικούς νοηματιστές όπως και σε διαφορετικά βίντεο τα οποία έχουν γυριστεί κάτω από διαφορετικές συνθήκες καθώς μόνο οι παράμετροι της Γκαουσιανής $p_s(C_b, C_r)$ χρειάζεται να προσαρμοστούν.

2.1.2 Μορφολογική Επεξεργασία των εξαγομένων Μασκών Δέρματος

Η μάσκα δέρματος S_0 ενδέχεται να περιέχει τρύπες εσωτερικά του κεφαλιού εξαιτίας περιοχών με διαφορετικό χρώμα από το δέρμα όπως π.χ. μάτια, στόμα κ.τ.λ. Για αυτόν το λόγο προτείνουμε έναν αλγόριθμο για την ομαλοποίηση και κάλυψη των τρυπών της μάσκας S_0, ο οποίος χρησιμοποιεί εργαλεία από τη μαθηματική μορφολογία [57, 81]. Χρησιμοποιούμε την έννοια των *τρυπών* (*holes*) $\mathcal{H}(S)$ σε μια δυαδική εικόνα S [115]. Για να γεμίσουμε κάποιες περιοχές του φόντου (hole filling) οι οποίες δεν είναι τρύπες με τη 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:

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

όπου B είναι ένα δομικό στοιχείο μικρού μεγέθους, και τα σύμβολα $\oplus$, $\bullet$ αντιστοιχούν στο Minkowski dilation και closing. Στη θεωρία η καλύτερη επιλογή για το δομικό στοιχείο B είναι ένας



(α) Αρχικό πλαίσιο (β) S_0 (γ) S_2 (δ) $S_2 \ominus B_c$ (ε) Κατατμημένο S_2

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 μάσκας δέρματος S_0, (γ) Τελική εκτίμηση της μάσκας δέρματος S_2, (δ) Erosion της μάσκας S_2 με έναν μικρό δίσκο, (ε) Κατάτμηση της μάσκας δέρματος S_2, σε συνεκτικές περιοχές.

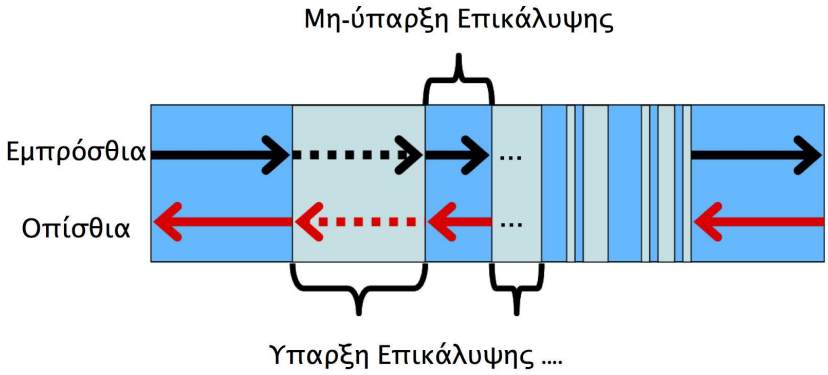
δίσκος, έτσι ώστε να υπάρχει ισοτροπική αντιμετώπιση προς όλες τις διευθύνσεις. Όμως για υπολογιστικούς λόγους στην υλοποίησή μας χρησιμοποιήσαμε ένα τετράγωνο 5×5 εικονοστοιχείων για το B αφού ο υπολογισμός του erosion/dilation με δομικό στοιχείο σχήματος τετραγώνου είναι αρκετά πιο γρήγορος.

Στη συνέχεια για να διορθώσουμε πιθανά λάθη εκμεταλλευόμαστε το γεγονός ότι οι συνεκτικές περιοχές χρώματος δέρματος μπορεί να είναι το πολύ τρεις (κεφάλι, αριστερό και δεξί χέρι). Επιπλέον θεωρούμε ότι δεν μπορούμε να έχουμε περιοχή με εμβαδόν μικρότερο από A_{min}, το οποίο αντιστοιχεί στο μικρότερο δυνατόν εμβαδόν χεριού του υπό μελέτη νοηματιστή. Ο περιορισμός αυτός προκύπτει από τη λογική υπόθεση ότι το εμβαδόν του κεφαλιού είναι πάντα μεγαλύτερο από το εμβαδόν των χεριών ενός ανθρώπου [146]. Έτσι βρίσκοντας τις συνεκτικές περιοχές της μάσκας S_1 και υπολογίζοντας τα εμβαδά τους, απορρίπτουμε όσες περιοχές έχουν εμβαδόν μικρότερο από A_{min}. Τέλος κρατώντας τις τρεις περιοχές με το μεγαλύτερο εμβαδόν καταλήγουμε στην τελική εκτίμηση της μάσκας δέρματος S_2. Ένα παράδειγμα της μάσκας S_2 απεικονίζεται στο Σχήμα. 2.2(γ).

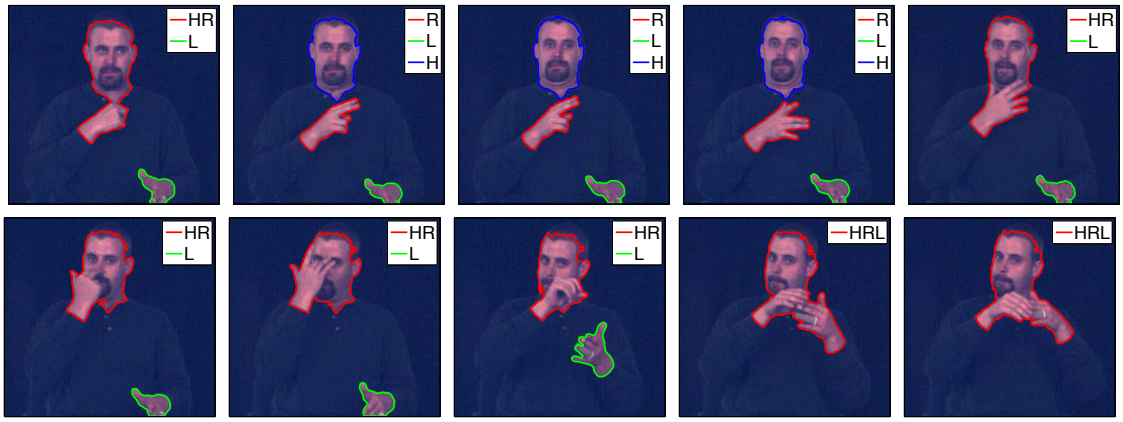
2.1.3 Μορφολογική Κατάτμηση των Μασκών Δέρματος

Η μάσκα δέρματος S_2 που υπολογίσαμε στην προηγούμενη ενότητα περιλαμβάνει τρεις συνεκτικές περιοχές που αντιστοιχούν στο κεφάλι και στα χέρια του νοηματιστή. Για τον διαχωρισμό τους ακολουθούμε την παρακάτω μέθοδο κατάτμησης βασισμένοι στη μαθηματική μορφολογία. Όταν η μάσκα S_2 περιέχει τρεις συνεκτικές περιοχές, δηλαδή δεν έχουμε επικάλυψη μεταξύ των αρθρωτών, η κατάτμηση είναι άμεση και εντοπίζει τις συνεκτικές περιοχές. Στην περίπτωση που έχουμε μικρές επικαλύψεις, δηλαδή όταν δυο ή τρεις διαφορετικές περιοχές ενώνονται με μια μικρή γέφυρα -βλέπε Σχήμα. 2.2(γ)- προτείνουμε τη μέθοδο *competitive reconstruction opening* για να τις διαχωρίσουμε. Η μέθοδος αυτή βασίζεται σε αλλεπάλληλα μορφολογικά φιλτραρίσματα.

Αν η μάσκα S_2 περιέχει N_{cc} συνεκτικές περιοχές/συνιστώσες όπου $N_{cc} < 3$ υπολογίζουμε το erosion της μάσκας S_2 με το δομικό στοιχείο B_c (δίσκος ακτίνας τριών εικονοστοιχείων) και βρίσκουμε τις συνεκτικές περιοχές της μάσκας $S_2 \ominus B_c$ απορρίπτοντας τις περιοχές με 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 τους να μην είναι εφικτός με τη μέθοδο *competitive reconstruction opening*. Σε αντίθετη περίπτωση εάν ο αριθμός



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



Σχήμα 2.4: Αποτέλεσμα από τη βάση δεδομένων BU400 της ανίχνευσης και παρακολούθησης των χεριών και του κεφαλιού του νοηματιστή σε περιπτώσεις όπου έχουμε επικαλύψεις. Παράδειγμα πλαισίων με τη 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.

των συνεκτικών περιοχών γίνει μεγαλύτερος από N_{cc} υποδεικνύεται μερική επικάλυψη. Σε αυτή την περίπτωση με αφετηρία αυτές τις συνεκτικές περιοχές εφαρμόζουμε αλλεπάλληλα conditional dilations έως ότου οι αρχικές συνεκτικές περιοχές συναντηθούν μεταξύ τους. Στη συνέχεια αφαιρούμε από όλες τις συνεκτικές περιοχές τα εικονοστοιχεία που ανήκουν σε παραπάνω από δύο περιοχές. Έτσι τελικά καταλήγουμε στις διαχωρισμένες πλέον συνεκτικές περιοχές. Στο Σχήμα 2.2(ε) μπορούμε να δούμε ένα αποτέλεσμα της εφαρμογής του αλγορίθμου *competitive reconstruction opening*.

2.2 Παρακολούθηση των Χεριών και του Κεφαλιού

Μετά την εφαρμογή της κατάτμησης στη μάσκα δέρματος S_2 εφαρμόζουμε έναν αλγόριθμο για την παρακολούθηση των χεριών και του κεφαλιού. Ο παρακάτω αλγόριθμος έχει ως αποτέλεσμα την εύρεση των αρθρωτών που συμπεριλαμβάνονται σε κάθε κατατμημένη περιοχή 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 [7]. Για την παρακολούθηση των αρθρωτών σε κάθε χρονικό πλαίσιο χρησιμοποιούμε διαφορετική προσέγγιση ανάλογα με την ύπαρξη ή μη επικάλυψης.

Μη ύπαρξη επικάλυψης: Το αποτέλεσμα της κατάτμησης της μάσκας δέρματος S_2 είναι τρεις διαφορετικές συνεκτικές περιοχές. Αντιστοιχούμε το κεφάλι στην περιοχή με το μεγαλύτερο εμβαδόν, υποθέτοντας ότι το εμβαδόν του κεφαλιού είναι πάντα μεγαλύτερο από αυτό 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 στη 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 ταιριάξει στους αρθρωτές στα τρία προηγούμενα πλαίσια εφαρμόζουμε μια εμπρόσθια γραμμική πρόβλεψη των παραμέτρων της έλλειψης για κάθε αρθρωτή για το τρέχον πλαίσιο. Λόγω της ευαισθησίας της γραμμικής πρόβλεψης σε σχέση με τον αριθμό των διαδοχικών πλαισίων όπου έχουμε επικάλυψη, εφαρμόζουμε επιπλέον οπίσθια γραμμική πρόβλεψη (βλ. Σχήμα 2.3). Επίσης υποθέτοντας 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 συνδυάζοντας τις εκτιμήσεις από την εμπρόσθια-οπίσθια γραμμική πρόβλεψη και την τεχνική ταιριάσματος προτύπου (template matching) αποκτάμε μια τελική εκτίμηση της θέσης των αρθρωτών.

Στα Σχήματα 2.4–2.7 έχουμε απεικονίσει τα αποτελέσματα από την ανίχνευση και παρακολούθηση των χεριών και κεφαλιού του νοηματιστή σε ακολουθίες από πλαίσια οι οποίες περιλαμβάνουν περιπτώσεις όπου έχουμε εμφάνιση επικαλύψεων.

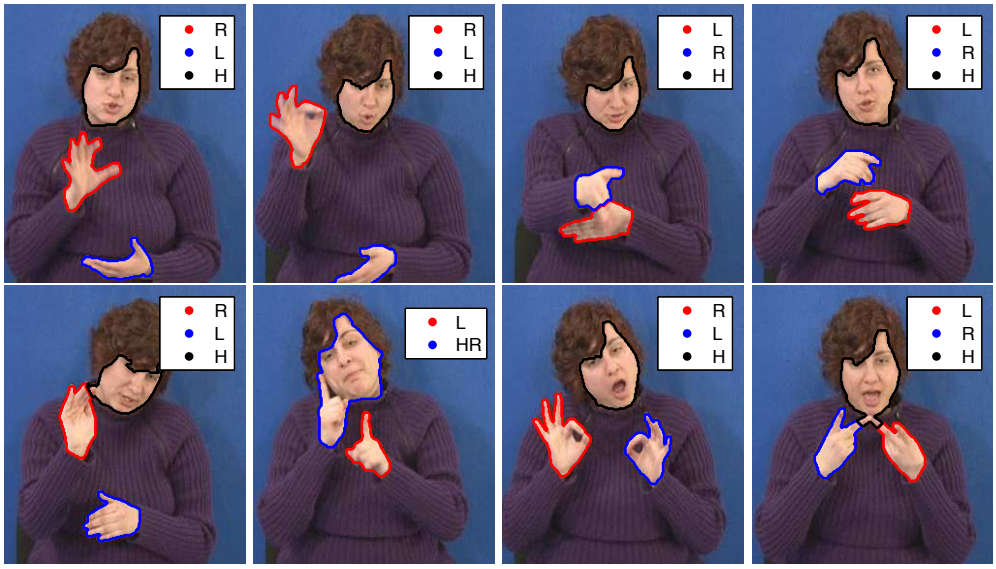
2.3 Εξαγωγή Χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών διανυσμάτων τα οποία περιγράφουν τους αρθρωτές σε κάθε πλαίσιο αποτελεί κρίσιμο στοιχείο για τα συστήματα αναγνώρισης ΝΓ. Τα χαρακτηριστικά πρέπει να σχετίζονται με τη θέση, την κίνηση και το σχήμα των χεριών. Σε αυτή την ενότητα παρουσιάζουμε τα χαρακτηριστικά διανυσμάτων που εξάγουμε για τις ροές πληροφορίας της κίνησης-θέσης των χεριών και της χειρομορφής.

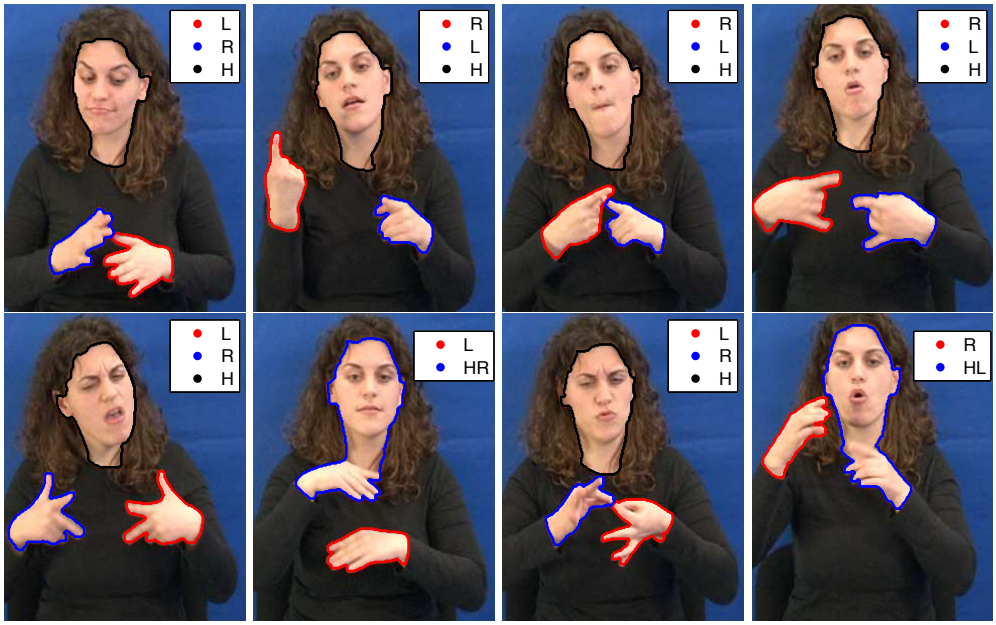
2.3.1 Ροή πληροφορίας της κίνησης-θέσης των χεριών

Τα χαρακτηριστικά διανύσματα που εξάγουμε για τη ροή πληροφορίας της κίνησης-θέσης (M-P) σχετίζονται με την κίνηση και τη θέση των χεριών του νοηματιστή στον νοηματικό χώρο. Αρχικά εξάγουμε τις συντεταγμένες των χεριών κανονικοποιημένες με τη θέση του κεφαλιού για κάθε χρονικό πλαίσιο

$$\begin{aligned} \boldsymbol{P}_R &= (x_R, y_R) - (x_H, y_H) \\ \boldsymbol{P}_L &= (x_L, y_L) - (x_H, y_H), \end{aligned} \tag{2.2}$$



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



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

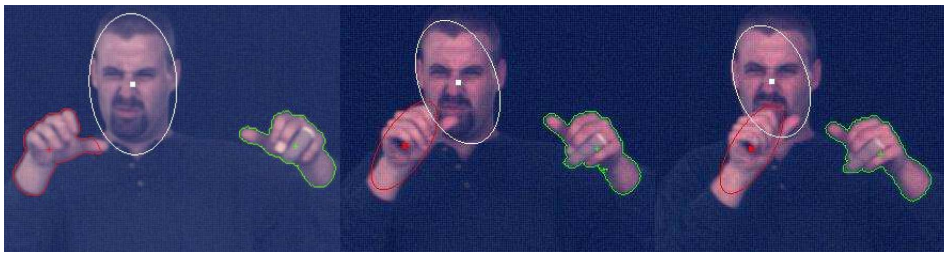
όπου (x_R, y_R), (x_L, y_L) είναι οι συντεταγμένες του δεξιού και αριστερού χεριού αντιστοίχως και (x^H, y^H) είναι οι συντεταγμένες του κεφαλιού του νοηματιστή.

Επιπλέον υπολογίζουμε την απόσταση μεταξύ των χεριών:

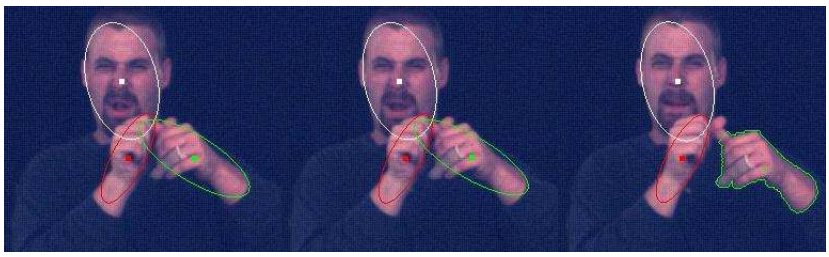
$$\boldsymbol{L} = (x_R - x_L, y_R - y_L), \tag{2.3}$$

την ταχύτητά τους:

$$\begin{aligned} \boldsymbol{V}_R &= (\dot{x_R}, \dot{y_R}), \\ \boldsymbol{V}_L &= (\dot{x_L}, \dot{y_L}), \end{aligned} \tag{2.4}$$



(α') (β') (γ')



(δ') (ε') (ϛ')

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

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, $A_0(W_{\boldsymbol{p}}^{-1}(\boldsymbol{x})) + \sum \lambda_i A_i(W_{\boldsymbol{p}}^{-1}(\boldsymbol{x}))$. Στην πάνω δεξιά γωνία δείχνουμε κάθε φορά την ανακατασκευή, αλλά στο χώρο του μοντέλου ΣΕ $A_0(\boldsymbol{x}) + \sum \lambda_i A_i(\boldsymbol{x})$, η οποία καθορίζει τα βέλτιστα βάρη.

όπου $\dot{x} = dx/dt$ συμβολίζει την χρονική παραγώγιση και τέλος την στιγμιαία κατεύθυνση της κίνησης των χεριών:

$$\begin{aligned} \boldsymbol{D}_R &= (\dot{x_R}, \dot{y_R}) / \left| \dot{x_R}, \dot{y_R} \right|, \\ \boldsymbol{D}_L &= (\dot{x_L}, \dot{y_L}) / \left| \dot{x_L}, \dot{y_L} \right|, \end{aligned} \tag{2.5}$$

2.3.2 Ροή πληροφορίας της χειρομορφής

Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης

Σε αυτή την ενότητα για λόγους συνέπειας θα παρουσιάσουμε το σύστημα της Δυναμικής Αφινικά Αναλλοίωτης Μοντελοποίησης Σχήματος-Εμφάνισης (Dynamic Affine-invariant Shape-

Appearance Model) το οποίο αναπτύχθηκε στα πλαίσια της διατριβής του Α. Ρούσσου [109, 110]. Αυτό το σύστημα προσφέρει μια αναπαράσταση των χειρομορφών. Επιπλέον, επιτρέπει την παρακολούθηση και εξαγωγή χαρακτηριστικών που σχετίζονται με το σχήμα και την εμφάνιση των χεριών ενός νοηματιστή. Τέλος, αποτελεί ένα εύρωστο σύστημα για την εξαγωγή χαρακτηριστικών που σχετίζονται με την χειρομορφή ακόμα και κατά τη διάρκεια επικαλύψεων.

Αναπαράσταση χεριών με την συνάρτηση σχήματος-εμφάνισης: Πρωταρχικός στόχος είναι η μοντελοποίηση όλων των δυνατών διαφορετικών χειρομορφών που εμφανίζονται κατά τη διάρκεια του νοηματισμού χρησιμοποιώντας δισδιάστατες εικόνες. Αυτές οι εικόνες έχουν μεγάλη ποικιλία λόγω των διαφορετικών χειρομορφών και της μεταβολής της τρισδιάστατης πόζας. Επίσης, ο αριθμός των σημείων πάνω στο επίπεδο του χεριού τα οποία είναι ορατά από την κάμερα, μεταβάλλεται 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 τη δυαδική μάσκα η οποία έχει προκύψει από την ανίχνευση του χρώματος δέρματος (βλ. ενότητα 2.1.1). Επιπλέον χρησιμοποιούμε την εμφάνιση των χεριών δηλαδή τις χρωματικές τιμές εσωτερικά της δυαδικής μάσκας. Αυτές οι τιμές εξαρτώνται από την υφή και τη σκίαση του χεριού με αποτέλεσμα να προσφέρουν πληροφορία για την τρισδιάστατη χειρομορφή. Πιο συγκεκριμένα, η συνάρτηση $f(\boldsymbol{x})$ σχήματος-εμφάνισης ισούται με

$$f(\boldsymbol{x}) = \begin{cases} g(I(\boldsymbol{x})) & \text{εάν } \boldsymbol{x} \in M \\ -c_b & \text{αλλιώς} \end{cases} \tag{2.6}$$

(2.7)

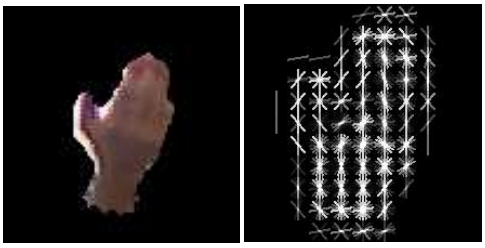
όπου $I(\boldsymbol{x})$ είναι οι τιμές (C_b, C_r) της εικόνας στον χρωματικό χώρο YCbCr και στο εικονοστοιχείο $\boldsymbol{x}$. Η συνάρτηση g είναι η προβολή του σημείου (C_b, C_r) στον πρωτεύοντα άξονα της διμεταβλητής Γκαουσιανής που μοντελοποιεί την πιθανότητα χρώματος -βλ. Σχήμα 2.1(γ)- και M η μάσκα δέρματος. Τέλος, η τιμή της σταθεράς c_b ελέγχει την ισορροπία μεταξύ σχήματος και εμφάνισης και ορίζεται πειραματικά.

Μοντελοποίηση της ποικιλίας των εικόνων σχήματος-εμφάνισης: Οι ΣΕ εικόνες των χεριών, $f(\boldsymbol{x})$, μοντελοποιούνται χρησιμοποιώντας ένα γραμμικό συνδυασμό προκαθορισμένων εικόνων μεταβολής, ακολουθούμενο από ένα αφινικό μετασχηματισμό :

$$f(W_{\boldsymbol{p}}(\boldsymbol{x})) \approx A_0(\boldsymbol{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\boldsymbol{x}),\ \boldsymbol{x} \in \Omega \tag{2.8}$$

όπου $A_0(\boldsymbol{x})$ είναι η εικόνα βάσης, $A_i(\boldsymbol{x})$ 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 $W_{\boldsymbol{p}}$ είναι ο αφινικός μετασχηματισμός με παραμέτρους $\boldsymbol{p} = (p_1 \cdots p_6)$ 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*. Ένα συγκεκριμένο Μοντέλο Σχήματος-Εμφάνισης χεριού καθορίζεται από την εικόνα βάσης $A_0(\boldsymbol{x})$ και τις ιδιοεικόνες $A_i(\boldsymbol{x})$ του γραμμικού συνδυασμού και τον αριθμό N_c. Τα διανύσματα $\boldsymbol{p}$ και $\boldsymbol{\lambda}$ είναι οι παράμετροι του μοντέλου που ταιριάζουν το δεδομένο μοντέλο σε μια εικόνα Σχήματος-Εμφάνισης σε κάθε χρονικό πλαίσιο. Οι παράμετροι αυτοί θεωρούνται ως χαρακτηριστικά της πόζας και του σχήματος του χεριού αντιστοίχως.

Εκπαίδευση του γραμμικού συνδυασμού του μοντέλου σχήματος-εμφάνισης: Για την εκπαίδευση του μοντέλου σχήματος-εμφάνισης χρησιμοποιούμε ένα αντιπροσωπευτικό σύνολο από εικόνες



(α') (β')

Σχήμα 2.9: (α) RGB εικόνα κομμένου χεριού και (β) Οπτική αναπαράσταση του HOG περιγραφητή.

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 σχήματος-εμφάνισης 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. Στη συνέχεια, οι εικόνες του γραμμικού συνδυασμού μαθαίνονται χρησιμοποιώντας PCA στο ευθυγραμμισμένο σύνολο.

Ομαλοποιημένο ταίριασμα του μοντέλου σχήματος-εμφάνισης με στατική και δυναμική πρότερη πληροφορία: Για την αντιμετώπιση των επικαλύψεων χρησιμοποιήθηκε ένα ομαλοποιημένο ταίριασμα του μοντέλου σχήματος-εμφάνισης το οποίο εκμεταλλεύεται πρότερη πληροφορία που σχετίζεται με τη 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 $\boldsymbol{p}$. Σε κάθε πλαίσιο n, βρίσκουμε τα βέλτιστα $\boldsymbol{\lambda} = \boldsymbol{\lambda}[n]$ και $\boldsymbol{p} = \boldsymbol{p}[n]$ που ελαχιστοποιούν τη συνολική ενέργεια:

$$E(\boldsymbol{\lambda}, \boldsymbol{p}) = E_{rec}(\boldsymbol{\lambda}, \boldsymbol{p}) + w_S E_S(\boldsymbol{\lambda}, \boldsymbol{p}) + w_D E_D(\boldsymbol{\lambda}, \boldsymbol{p}) \ , \tag{2.9}$$

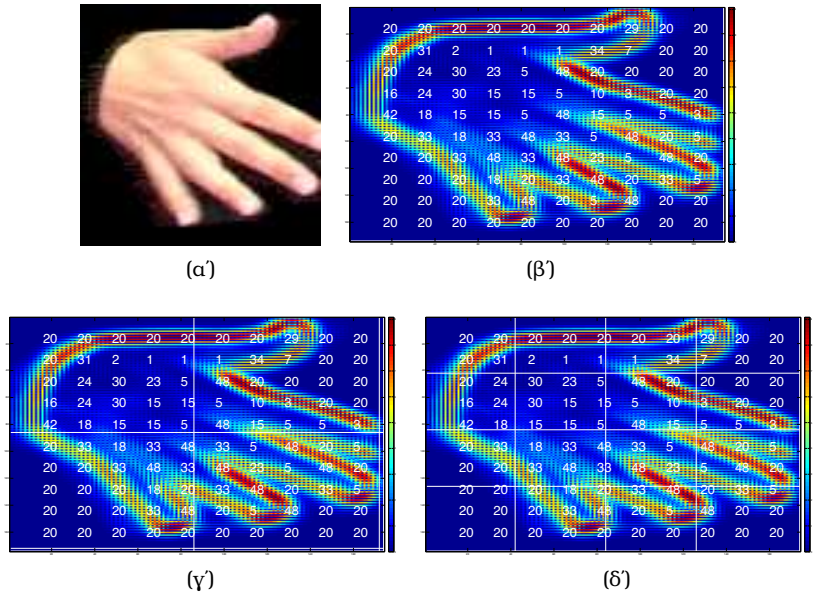
όπου w_S,w_D είναι θετικά βάρη που ελέγχουν την ισορροπία μεταξύ των τριών όρων. Οι όροι $E_S(\boldsymbol{\lambda}, \boldsymbol{p})$ και $E_D(\boldsymbol{\lambda}, \boldsymbol{p})$ αντιστοιχούν στην ενέργεια της στατικής και δυναμικής πρότερης πληροφορίας. Στο Σχήμα 2.8 έχουμε απεικονίσει τα αποτελέσματα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

Διαφορετικές αναπαραστάσεις

Για την αναπαράσταση των χειρομορφών χρησιμοποιήσαμε δύο επιπλέον διαφορετικά διανύσματα χαρακτηριστικών αρκετά δημοφιλή στα πλαίσια της ερευνητικής περιοχής αναγνώρισης αντικειμένων.

Το πρώτο είναι ο HOG [34] περιγραφητής, ο οποίος αποτελεί ένα από κοινού ιστόγραμμα κβαντισμένων κατευθύνσεων της κλίσης της εικόνας και της μετατοπισμένης θέσης, στην γειτονιά κάθε εικονοστοιχείου. Ο υπολογισμός του HOG περιγραφητή αποτελείται από τα εξής βήματα:

- Υπολογισμός των μερικών παραγώγων της εικόνας,
- Χωρική διαμέριση της εικόνας χρησιμοποιώντας ένα πλέγμα,
- Δημιουργία ιστογραμμάτων κατεύθυνσης,
- Κανονικοποίηση ιστογραμμάτων.



Σχήμα 2.10: (α) RGB εικόνα κομμένου χεριού και (β-δ) Οπτική αναπαράσταση των dense SIFT για τα τρία επίπεδα πυραμίδας. Επιπλέον απεικονίζουμε τον αριθμό της συστάδας στην οποία έχει ταξινομηθεί κάθε patch.

Στο Σχήμα 2.9 απεικονίζουμε μια οπτική αναπαράσταση του περιγραφητή HOG για μια εικόνα χεριού.

Για το δεύτερο διάνυσμα χαρακτηριστικών χρησιμοποιήσαμε την ιδέα των χωρικών πυραμίδων (spatial pyramids) [73]. Ο υπολογισμός των χαρακτηριστικών αυτών αποτελείται από τα εξής βήματα:

- Υπολογισμός των dense SIFT στην εικόνα,
- Εφαρμογή του αλγορίθμου K-means για την κατασκευή ενός οπτικού λεξιλογίου,
- Υπολογισμός των κανονικοποιημένων ιστογραμμάτων σε διαφορετικά επίπεδα πυραμίδας,
- Συνένωση των επιμέρους ιστογραμμάτων,
- Μετασχηματισμός του συνενωμένου ιστογράμματος.

Πιο συγκεκριμένα, σε κάθε εικόνα κομμένου χεριού εξάγουμε τα dense SIFT χαρακτηριστικά [80]. Στη συνέχεια βασιζόμενοι στην ιδέα των bag-of-words, εφαρμόζουμε τον αλγόριθμο συσταδοποίησης K-means σε τυχαία patches για την κατασκευή ενός οπτικού λεξιλογίου. Ο αριθμός των συστάδων που χρησιμοποιήσαμε καθορίστηκε πειραματικά σε 10. Αντίστοιχα με το άρθρο [73] υπολογίζουμε τα κανονικοποιημένα ιστογράμματα για το οπτικό λεξιλόγιο σε τρία επίπεδα πυραμίδας. Στο Σχήμα 2.10 απεικονίζουμε για μια εικόνα χεριού, μια οπτική αναπαράσταση των dense SIFT για τα τρία επίπεδα πυραμίδας μαζί με το αριθμό της συστάδας που έχει ταξινομηθεί κάθε patch. Στη συνέχεια, συνενώνουμε τα επιμέρους ιστογράμματα και μετασχηματίζουμε το συνενωμένο διάνυσμα χαρακτηριστικών σε έναν νέο χώρο χαρακτηριστικών βασιζόμενοι στη μέθοδο που παρουσιάζεται στο άρθρο [131]. Αυτός ο νέος χώρος χαρακτηριστικών έχει την ιδιαιτερότητα ότι το εσωτερικό γινόμενο μεταξύ δύο χαρακτηριστικών διανυσμάτων ισούται με την απόσταση histogram intersection των αντίστοιχων διανυσμάτων στον προηγούμενο χώρο χαρακτηριστικών των συνενωμένων ιστογραμμάτων. Τέλος, εφαρμόζουμε PCA για τη μείωση της διάστασης των χαρακτηριστικών διανυσμάτων κρατώντας τα πρώτα 100 ιδιοδιανύσματα.

Κεφάλαιο 3

Στατιστική μοντελοποίηση της Νοηματικής Γλώσσας με Δεδομενοκεντρικές Υπομονάδες

3.1 Εισαγωγή στην έννοια των υπομονάδων

Μια αδρή αντιστοίχιση μιας λέξης στον προφορικό λόγο είναι με το νόημα στη νοηματική γλώσσα. Οι δομικές υπομονάδες (φωνήματα) που απαρτίζουν μια λέξη στον προφορικό λόγο συνενώνονται διαδοχικά στον χρόνο, όπως π.χ. η αγγλική λέξη admit, η οποία μεταγράφεται φωνητικώς ως $[\text{ədmˈɪt}]$. Σε αντίθεση με τις λέξεις, τα νοήματα τείνουν να είναι μονοσυλλαβικά [32]. Η φωνητική δομή των νοημάτων στη νοηματική γλώσσα, σε αντίθεση με τις λέξεις στον προφορικό λόγο, περιέχει την έννοια της παραλληλίας [118] αλλά και της διαδοχής στον χρόνο [78]. Για τη μεταφορά αντίστοιχης ροής πληροφορίας σε σχέση με τον προφορικό λόγο, προστίθεται στην φωνητική δομή των νοημάτων η έννοια της παραλληλίας μέσω των πολλαπλών παράλληλων ροών πληροφορίας. Αυτό οφείλεται λόγω των μεγάλων αρθρωτών για την παραγωγή της νοηματικής γλώσσας, σε σχέση με τους αντίστοιχους για την παραγωγή του προφορικού λόγου, π.χ. χέρια έναντι γλώσσας [70]. Ας πάρουμε ως παράδειγμα τα νοήματα στο Σχήμα 3.1. Οι παράμετροι άρθρωσης όπως η θέση των χεριών, ο τύπος της κίνησής τους, το σχήμα τους (χειρομορφή) και οι εκφράσεις του προσώπου μπορούν να μεταβάλλονται παράλληλα στον χρόνο.

Επιπλέον, έχουν γίνει έρευνες σχετικά με την διαδοχική δομή των φωνητικών υπομονάδων στη νοηματική γλώσσα [31]. Οι Liddell και Johnson (L&J) [77] εισήγαγαν ως φωνολογική βάση για τη νοηματική γλώσσα τη διαδοχή φωνητικών υπομονάδων. Πρότειναν το μοντέλο Movement-Hold [78] το οποίο λάμβανε υπόψη του την έννοια της παραλληλίας αλλά και της διαδοχής των φωνητικών υπομονάδων. Εισήγαγαν δύο τύπους τμημάτων, τα "Movements" και τα "Holds". Τα "Movements", αντιστοιχούν σε τμήματα κατά την διάρκεια των οποίων μεταβάλλεται τουλάχιστον μια από τις παραμέτρους της άρθρωσης. Μεταβολή της θέσης των χεριών (ύπαρξη κίνησης), αλλαγή της χειρομορφής κ.α. Τα "Holds", αντιστοιχούν σε τμήματα κατά τη διάρκεια των οποίων όλες οι παράμετροι της άρθρωσης παραμένουν σταθερές. Ως αποτέλεσμα κάθε νόημα μπορεί να περιγραφεί από μια ακολουθία διαδοχικών "Movement" και "Hold" τμημάτων.

Στο Σχήμα 3.1(F) απεικονίζουμε μια εκτέλεση του νοήματος 'ΛΕΩ' στην Ελληνική Νοηματική Γλώσσα (ΕΝΓ). Αυτό το νόημα αρθρώνεται χρησιμοποιώντας μόνο το κυρίαρχο χέρι. Απαρτίζεται από μια κίνηση προς τα κάτω του κυρίαρχου χεριού ξεκινώντας από το στόμα και καταλήγοντας στον ουδέτερο χώρο νοηματισμού. Με άλλα λόγια το νόημα 'ΛΕΩ', σύμφωνα με το μοντέλο Movement-Hold μπορεί να περιγραφεί από τη διαδοχή τριών τμημάτων: 'θέση-στόμα (H)', 'κίνηση-κάτω (M)' και 'θέση-ουδέτερος χώρος (H)'. Άλλο ένα παράδειγμα απεικονίζεται στο Σχήμα 3.2 όπου

(α΄) ΕΔΩ (β΄) ΤΕΛΟΣ (γ΄) CHICAGO (δ΄) ΚΑΤΑΘΕΣΗ (ε΄) PYZI (ϛ΄) ΛΕΩ (ζ) ΥΠΟΔΕΧΟΜΑΙ (η΄) ΥΠΟΔΕΧΟΜΑΙ

Σχήμα 3.1: Νοήματα στην ΑΝΓ από τη βάση δεδομένων BU400 (α,β) και από την ASLLVD (γ,δ). (ε-θ) νοήματα στην ΕΝΓ από τη βάση δεδομένων GSL-Lem

(α΄) Hold (β΄) Movement (γ΄) Hold

Σχήμα 3.2: Κατάτμηση σε Movement και Hold τμήματα για το νόημα ADMIT στην ΑΝΓ (BU400).

το νόημα "ADMIT" στην Αμερικανική Νοηματική Γλώσσα (ΑΝΓ) μεταγράφεται σε –H M H–.

Συνοψίζοντας, παρότι έχουν προταθεί αρκετά γλωσσολογικά-φωνητικά μοντέλα η έννοια της φωνητικής υπομονάδας στη νοηματική γλώσσα δεν είναι δεδομένη και ούτε πλήρως καθορισμένη όπως στην περίπτωση του προφορικό λόγου. Είναι ένα ανοιχτό πεδίο έρευνας και από γλωσσολόγους ερευνητές [118, 78, 111], αλλά και από ερευνητές στο πεδίο της επιστήμης των υπολογιστών, με στόχο τον καθορισμό, την εύρεση και την υπολογιστική μοντελοποίηση αυτών των φωνητικών υπομονάδων [11, 136, 72, 1].

Λόγω των όσων προαναφέρθηκαν η φωνητική μοντελοποίηση με στόχο τη μοντελοποίηση και αυτόματη αναγνώριση της Νοηματικής Γλώσσας (ΝΓ) είναι ένα αρκετά δύσκολο πρόβλημα. Πρώτα από όλα υπάρχει η έλλειψη από επίσημα λεξικά, επισημειωμένα σε φωνητικό επίπεδο, βασισμένα σε πλήρως καθορισμένες φωνητικές υπομονάδες και σε ένα πρότυπο σύστημα επισημείωσης. Στο ερευνητικό πεδίο της αναγνώρισης φωνής, τέτοιου είδους πόροι είναι τυποποιημένοι, πλήρως καθορισμένοι και άμεσα προσβάσιμοι για όλη την ερευνητική κοινότητα. Στο ερευνητικό πεδίο της αυτόματης αναγνώρισης της ΝΓ οι προσεγγίσεις που χρησιμοποιούν υπομονάδες για τη μοντελοποίηση και αναγνώριση της ΝΓ είναι από τη μια πλευρά δεδομενοκεντρικές. Ορίζουν ένα σύνολο από υπολογιστικές υπομονάδες χωρίς τη χρήση επισημειώσεων σε φωνητικό επίπεδο. Ενδεικτικές προσεγγίσεις παρουσιάζονται στα άρθρα [11, 47, 72]. Από την άλλη πλευρά υπάρχουν προσεγγίσεις οι οποίες χρησιμοποιούν λεξικά βασιζόμενα σε γλωσσικά μοντέλα όπως το Movement-Hold [78] και συστήματα φωνητικής επισημείωσης της ΝΓ όπως το Stokoe σύστημα [118], το Hamburg Notation System (HamNoSys) [103] ή το SignWriting [119]. Τα παραπάνω λεξικά κατασκευάζονται είτε μέσω ανθρώπινης επισημείωσης, διαδικασία αρκετά χρονοβόρα, όπως στο άρθρο [136], είτε πιο πρόσφατα κάνοντας χρήση αυτόματων μεθόδων επεξεργασίας [102, 71]. Ανάμεσα σε αυτά τα δύο σύνολα προσεγγίσεων, έχουν προταθεί μέθοδοι [66, 17, 55, 28], οι οποίες ενσωματώνουν γλωσσικές-φωνητικές έννοιες, εμπνευσμένες κατά κύριο λόγο από την έρευνα του Stokoe [118]. Παρόλα ταύτα δεν καταλήγουν σε ευρέως επαναχρησιμοποιήσιμες υπομονάδες σύμφωνες με κάποιο από τα γνωστά συστήματα επισημείωσης ή γλωσσολογικά μοντέλα [118, 103, 119, 78].

Σε αυτό το κεφάλαιο θα παρουσιάσουμε δύο καινοτόμες μεθόδους για τη μοντελοποίηση της ΝΓ κάνοντας χρήση δεδομενοκεντρικών υπομονάδων. Αυτές τις ονομάζουμε 2-S-U και RAW. Για την

αναπαράσταση ενός νοήματος προσφέρουν μια φωνητική δομή υπομονάδων η οποία εμπεριέχει τις έννοιες της παραλληλίας αλλά και της διαδοχής διαφορετικών υπομονάδων στον χρόνο, χωρίς τη χρήση γλωσσολογικής πρότερης γνώσης. Βασίζονται στην αυτόματη κατάτμηση των νοημάτων σε δυναμικά και στατικά τμήματα, που αντιστοιχούν σε τμήματα στα οποία υπάρχει ή δεν υπάρχει κίνηση. Για την κατασκευή των **δυναμικών και στατικών (Δ/Σ)** υπομονάδων διαχειρίζονται τα δυναμικά και τα στατικά τμήματα ξεχωριστά χρησιμοποιώντας διαφορετικό είδος χαρακτηριστικών για κάθε περίπτωση. Επιπλέον, η μοντελοποίηση των Δ/Σ υπομονάδων γίνεται χρησιμοποιώντας στατιστικά μοντέλα. Τέλος ένα πολύτιμο αποτέλεσμα είναι η χωρίς επίβλεψη κατασκευή ενός λεξικού στο επίπεδο της υπομονάδας το οποίο εμπεριέχει τις παραπάνω ιδιότητες.

Παρόλο που δεν γίνεται χρήση πρότερης γλωσσολογικής γνώσης, οι μέθοδοι είναι εμπνευσμένες από το μοντέλο Movement-Hold. Συσχετίζουμε τα "Movement" και "Hold" τμήματα με τις περιπτώσεις όπου έχουμε κίνηση ή μη-κίνηση των χεριών του νοηματιστή, δηλαδή δυναμικά και στατικά τμήματα. Έτσι, κάθε νόημα αναπαριστάται από μια ακολουθία δυναμικών και στατικών υπομονάδων η οποία συνάδει μερικώς με τους φωνητικούς κανόνες του μοντέλου Movement-Hold. Για τα στατικά τμήματα χρησιμοποιούμε το διάνυσμα χαρακτηριστικών της θέσης των χεριών, και για τα δυναμικά χρησιμοποιούμε το διάνυσμα χαρακτηριστικών της κίνησής τους. Το από κοινού διάνυσμα χαρακτηριστικών το ονομάζουμε: ροή πληροφορίας της 'κίνησης-θέσης', και το χρησιμοποιούμε για την εκπαίδευση των δυναμικών και στατικών μοντέλων υπομονάδας. Τέλος, το διάνυσμα χαρακτηριστικών για την χειρομορφή χρησιμοποιείται ως ανεξάρτητη ροή πληροφορίας για την κατασκευή των υπομονάδων χειρομορφής.

3.2 Σύνοψη συστήματος

Στο Σχήμα 3.3 παρουσιάζουμε ένα διάγραμμα ροής του προτεινόμενου συστήματος. Το 2-S-U αποτελείται από τρία υποσυστήματα:

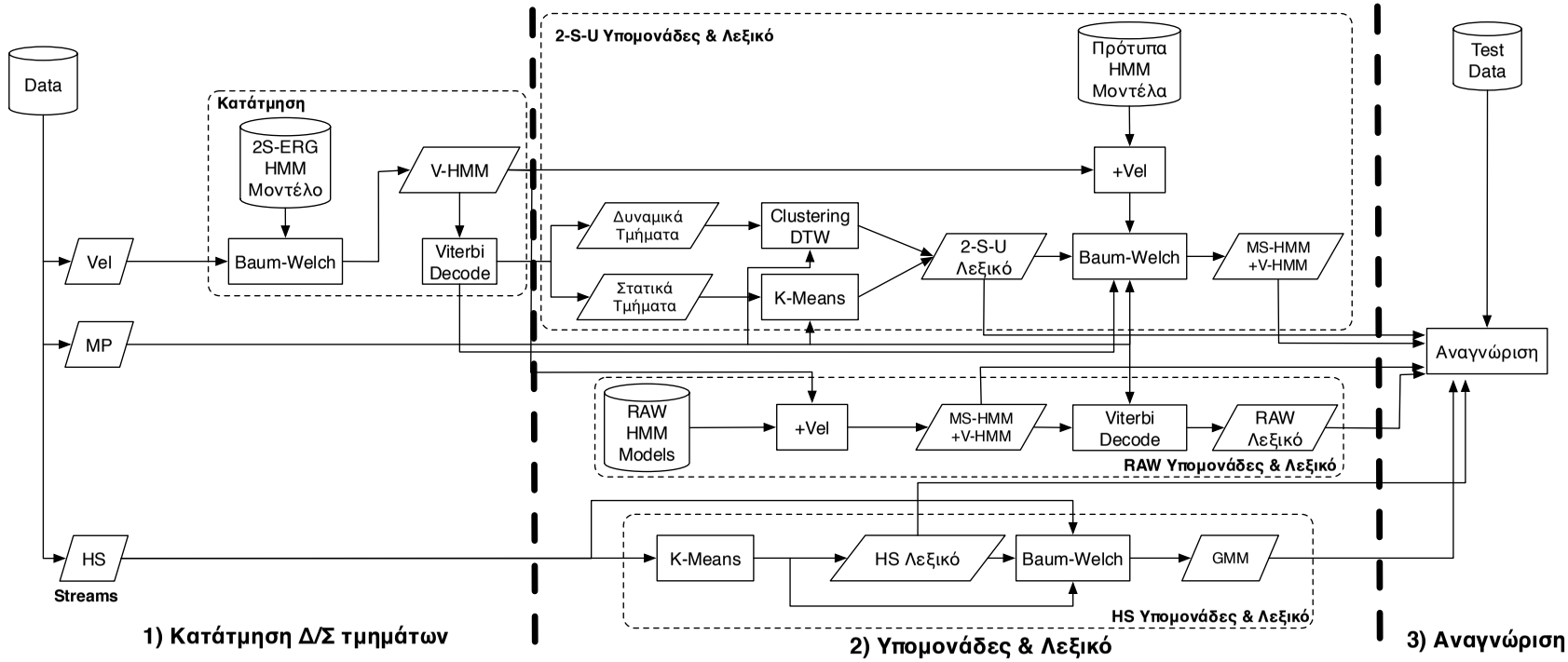
1) Κατάτμηση δυναμικών και στατικών (Δ/Σ) τμημάτων,

2) Υπομονάδες & Λεξικό και

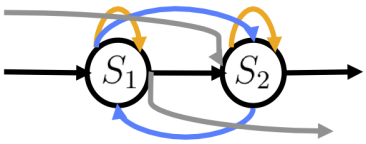
3) Αναγνώριση.

Κατάτμηση δυναμικών και στατικών (Δ/Σ) τμημάτων: Μια από τις συνεισφορές μας σχετίζεται με την αυτόματη κατάτμηση σε δυναμικά και στατικά τμήματα. Εκπαιδεύουμε δύο Γκαουσιανά μοντέλα (δυναμικό και στατικό) με διαγώνιο πίνακα συμμεταβλητότητας χρησιμοποιώντας την ταχύτητα (Vel) ως διάνυσμα χαρακτηριστικών και τα συνδυάζουμε με ένα εργοδικό HMM. Η κατάτμηση επιτυγχάνεται βρίσκοντας την πιο πιθανή ακολουθία καταστάσεων του εργοδικού μοντέλου με την χρήση του αλγορίθμου Viterbi (βλέπε ενότητα 3.3).

Υπομονάδες & Λεξικό: Μια επιπλέον συνεισφορά σχετίζεται με την αυτόματη δημιουργία ενός λεξικού σε επίπεδο υπομονάδας. Η αντιστοίχιση δηλαδή κάθε νοήματος από μια ακολουθία διαδοχικών υπομονάδων. Για την κατασκευή των Δ/Σ υπομονάδων βασιζόμαστε στην ροή πληροφορίας της κίνησης-θέσης και στην κατάτμηση σε Δ/Σ τμήματα εφαρμόζοντας δύο διαφορετικές προσεγγίσεις: 2-S-U και RAW. Για την κατασκευή των Δ/Σ 2-S-U υπομονάδων, ομαδοποιούμε τα δυναμικά και στατικά τμήματα εφαρμόζοντας αλγορίθμους συσταδοποίησης. Στη συνεχεία, συνδυάζουμε την κατάτμηση σε Δ/Σ τμήματα και την αντιστοιχία κάθε τμήματος με τη συστάδα στην οποία κατατάχθηκε για την κατασκευή του 2-S-U λεξικό (βλ. ενότητες 3.4.1 και 3.5.1). Αντίθετα, για την κατασκευή των Δ/Σ RAW υπομονάδων, γίνεται ομοιόμορφη διαμέριση των χώρων των χαρακτηριστικών που αντιπροσωπεύουν τις ροές πληροφορίας της κίνησης και θέσης αντιστοίχως. Ενώ για την κατασκευή του RAW λεξικού χρησιμοποιείται ο αλγόριθμος Viterbi για την εύρεση της πιο πιθανής ακολουθίας υπομονάδων (βλ. ενότητες 3.4.2 και 3.5.1). Τέλος για τη ροή πληροφορίας



Σχήμα 3.3: Διάγραμμα ροής του συστήματος με δεδομενοκεντρικές υπομονάδες. Τα τετράγωνα αντιπροσωπεύουν τις διαδικασίες και τα παραλληλόγραμμα, δεδομένα εισόδου ή εξόδου. 1) Κατάτμηση Δ/Σ τμημάτων: εκμεταλλευόμενοι το διάνυσμα χαρακτηριστικών της ταχύτητας, κάνουμε κατάτμηση των νοημάτων σε δυναμικά και στατικά (Δ/Σ) τμήματα. 2) Υπομονάδες & Λεξικό: Δύο διαφορετικές προσεγγίσεις για τις υπομονάδες κίνησης-θέσης (2-S-U και RAW) και μία για τις υπομονάδες χειρομορφής. 3) Αναγνώριση: Viterbi Decoding και εκ των υστέρων σύμμειξη. Σε όλες τις περιπτώσεις τα "data" αντιστοιχούν σε διανύσματα χαρακτηριστικών Ροές πληροφορίας: ταχύτητα (Vel), κίνηση-θέση (MP) και χειρομορφή (HS). V-HMM αντιστοιχεί στα εκπαιδευμένα Γκαουσιανά μοντέλα για τα (Δ/Σ) τμήματα. '+Vel' είναι η διαδικασία ενσωμάτωσης των Δ/Σ μοντέλων ταχύτητας στα HMM μοντέλα υπομονάδων.



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

της χειρομορφής δεν γίνεται κατάτμηση σε Δ/Σ τμήματα. Εφαρμόζεται ο αλγόριθμος συσταδοποίησης K-means σε όλα τα πλαίσια και το τελικό λεξικό κατασκευάζεται χρησιμοποιώντας την αντιστοίχιση κάθε πλαισίου με την συστάδα που κατατάχθηκε.

Για την μοντελοποίηση των Δ/Σ υπομονάδων χρησιμοποιούμε multistream HMMs όπως παρουσιάζεται με λεπτομέρεια στην ενότητα 5.1. Επιπλέον στα μοντέλα των Δ/Σ υπομονάδων ενσωματώνουμε τα Γκαουσιανά μοντέλα ταχύτητας (V-HMM) που εκπαιδεύτηκαν με στόχο την κατάτμηση σε Δ/Σ τμήματα. Αντίθετα, για την μοντελοποίηση των υπομονάδων χειρομορφής, χρησιμοποιούμε ένα απλό Γκαουσιανό μοντέλο.

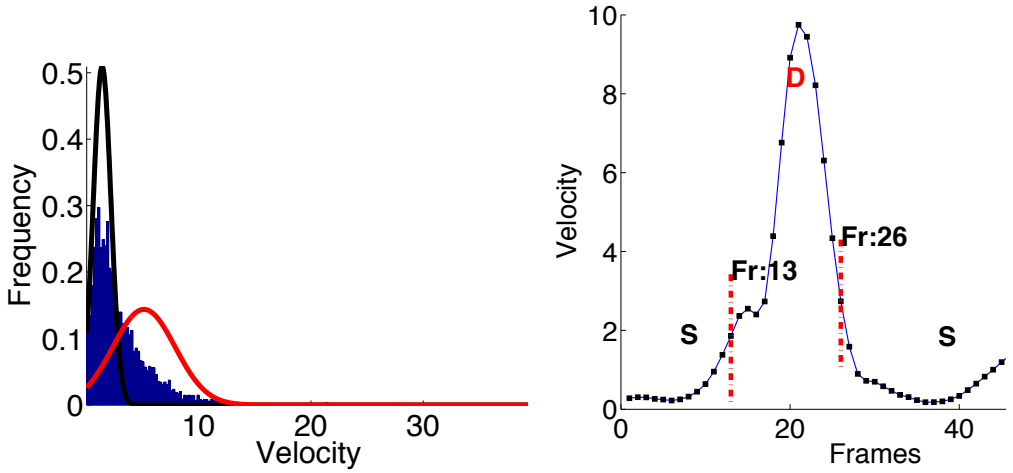
Αναγνώριση: Εκμεταλλευόμενοι τα εκπαιδευμένα μοντέλα υπομονάδων και τα λεξικά, εφαρμόζουμε τον αλγόριθμο Viterbi. Κατά την διάρκεια της αναγνώρισης βρίσκουμε ταυτόχρονα και την κατάτμηση σε Δ/Σ τμήματα, αλλά και την πιο πιθανή υπομονάδα σε κάθε τμήμα. Η τελική αναγνώριση σε επίπεδο νοήματος γίνεται εφαρμόζοντας εκ των υστέρων σύμμειξη των ροών πληροφορίας της κίνησης-θέσης και της χειρομορφής χρησιμοποιώντας PaHMM [136] (βλ. ενότητα 5.2). Κατά αυτόν τον τρόπο συνδυάζουμε τη διαδοχή των Δ/Σ υπομονάδων με την παραλληλία των πολλαπλών ροών πληροφορίας.

3.3 Αυτόματη κατάτμηση σε στατικά και δυναμικά τμήματα

Σε αυτή την ενότητα παρουσιάζουμε την κατάτμηση των νοημάτων σε διαισθητικά διαδοχικά χρονικά τμήματα και την ταξινόμησή τους σε δυναμικά ή στατικά, ανάλογα με την ύπαρξη ή μη κίνησης. Το αποτέλεσμα αυτής της κατάτμησης προσδίδει στο προτεινόμενο σύστημα τη διαδοχική φωνητική δομή των δυναμικών και στατικών (Δ/Σ) υπομονάδων.

Μουτελοποίηση δυναμικών και στατικών τμημάτων: Η κατάτμηση και ταξινόμηση σε δυναμικά ή στατικά τμήματα βασίζεται στο διάνυσμα της ταχύτητας. Υποθέτουμε ότι ένα δυναμικό τμήμα χαρακτηρίζεται από σχετικά υψηλή ταχύτητα, ενώ ένα στατικό από χαμηλή. Για τη μοντελοποίηση των στατικών και δυναμικών τμημάτων εκπαιδεύουμε δύο Γκαουσιανά μοντέλα, τα οποία τα συνδυάζουμε με ένα εργοδικό HMM δύο καταστάσεων (Σχήμα 3.4). Η μια κατάσταση του εργοδικού μοντελοποιεί τα στατικά τμήματα (χαμηλή ταχύτητα) και η άλλη τα δυναμικά (υψηλή ταχύτητα). Εκπαιδεύουμε το εργοδικό μοντέλο κάνοντας χρήση του αλγορίθμου Baum-Welch χρησιμοποιώντας όλα τα δεδομένα εκπαίδευσης. Το αποτέλεσμα είναι η εκπαίδευση των δυο Γκαουσιανών μοντέλων, ένα για τα στατικά τμήματα και ένα για τα δυναμικά. Στο Σχήμα 3.5α΄ απεικονίζουμε την κατανομή της ταχύτητας υπερθέτοντας τις συναρτήσεις πυκνότητας πιθανότητας (σ.π.π.) των εκπαιδευμένων Γκαουσιανών μοντέλων για τα Δ/Σ τμήματα. Με αυτό τον τρόπο εκτιμούμε άμεσα ένα κατώφλι ταχύτητας για τον διαχωρισμό των Δ/Σ τμημάτων.

Κατάτμηση με βάση τα μουτέλα: Μετά την εκπαίδευση του εργοδικού HMM χρησιμοποιούμε τον αλγόριθμο Viterbi για την εύρεση της πιο πιθανής ακολουθίας καταστάσεων. Αφού όμως



(α') (β')

Σχήμα 3.5: (α) Κατανομή της ταχύητας (ιστόγραμμα) υπερθέτοντας τις σ.π.π. που αντιστοιχούν στις δύο καταστάσεις του εργοδικού HMM (κόκκινη και μαύρη καμπύλη). Η μαύρη αντιστοιχεί στην κατανομή της Γκαουσιανής για τα στατικά και η κόκκινη για τα δυναμικά τμήματα. Η μονάδα μέτρησης στον x άξονα είναι εικονοστοιχεία ανά χρονικό πλαίσιο και στον y άξονα είναι η κανονικοποιημένη συχνότητα. (β) Η κατάτμηση σε Δ/Σ τμήματα πάνω στο προφίλ της ταχύτητας για το νόημα ADMIT.

οι καταστάσεις του εργοδικού HMM αντιστοιχούν στα μοντέλα των Δ/Σ τμημάτων, η ακολουθία καταστάσεων ουσιαστικά είναι η κατάτμηση σε Δ/Σ τμήματα . Στο Σχήμα 3.5β' απεικονίζουμε ένα παράδειγμα κατάτμησης για μια εκτέλεση του νοήματος ADMIT της ΑΝΓ από την βάση δεδομένων BU400. Η δομή των Δ/Σ τμημάτων είναι 'Σ Δ Σ'. Το αποτέλεσμα αυτό πρέπει να ιδωθεί σε σύγκριση με το Σχήμα 3.2, όπου με βάση το μοντέλο Movement-Hold το νόημα περιγράφεται από την ακολουθία 'Η Μ Η'.

Σύνοψη και αποτελέσματα: Η κατάτμηση σε Δ/Σ τμήματα που παρουσιάστηκε, η οποία βασίζεται στη μοντελοποίηση των Δ/Σ τμημάτων προσφέρει αρκετά πλεονεκτήματα. Πρώτον, προσφέρει από κοινού την χρονική κατάτμηση σε διαδοχικά τμήματα αλλά και την ταξινόμησή τους σε δυναμικά και στατικά τμήματα. Αυτό οφείλεται στο ότι έχουμε ενσωματώνουμε έμμεσα τα χαρακτηριστικά των Δ/Σ τμημάτων στις καταστάσεις του εργοδικού HMM μοντέλου. Δεύτερον, δεν υπάρχει η ανάγκη βελτιστοποίησης καμίας παραμέτρου ή καθορισμού κάποιου κατωφλιού. Αυτό έχει ως επακόλουθο η διαδικασία κατάτμησης συμπεριλαμβανόμενης και της εκπαίδευσης των Δ/Σ Γκαουσιανών μοντέλων να είναι άμεσα εφαρμόσιμη σε άλλες βάσεις δεδομένων. Τέλος, επειδή η υλοποίηση βασίζεται σε μοντέλα HMM ταιριάζει στο συνολικό πιθανοτικό πλαίσιο.

Τα αποτελέσματα τα εκμεταλλευόμαστε στη συνέχεια ως ακολούθως. Εφαρμόζουμε την κατάτμηση σε Δ/Σ τμήματα σε διαφορετικά διανύσματα χαρακτηριστικών, όπως π.χ. την κατεύθυνση των χεριών, με στόχο την κατάτμηση των σημάτων για την εφαρμογή των αλγορίθμων συσταδιοποίησης (ενότητα 3.4.1). Επιπλέον για την κατασκευή του λεξικού, χρησιμοποιούμε την ακολουθία των Δ/Σ τμημάτων μαζί με την αντιστοίχισή τους στις συστάδες που ταξινομήθηκαν (ενότητα 3.5). Ακόμα, οι σ.π.π. των Δ/Σ Γκαουσιανών μοντέλων ενσωματώνονται στα HMM μοντέλα υπομονάδας (ενότητα. 3.6). Τέλος, τα χρονικά όρια της κατάτμησης χρησιμοποιούνται για την εκπαίδευση των HMM μοντέλων υπομονάδας.

3.4 Μοντελοποίηση των δυναμικών και στατικών (Δ/Σ) υπομονάδων

3.4.1 2-S-U δυναμικές και στατικές υπομονάδες

Σε αυτή την ενότητα παρουσιάζουμε τη διαδικασία συσταδοποίησης των Δ/Σ τμημάτων με στόχο την κατασκευή των δυναμικών και στατικών (Δ/Σ) υπομονάδων. Παίρνουμε ως είσοδο την κατάτμηση σε Δ/Σ τμήματα που είδαμε προηγουμένως και χρησιμοποιούμε τα κατάλληλα διανύσματα χαρακτηριστικών σε κάθε περίπτωση αναλόγως με το είδος των τμημάτων, δυναμικά ή στατικά. Έτσι, καταλήγουμε σε συστάδες τμημάτων οι οποίες αντιστοιχούν σε διαφορετικές υπομονάδες.

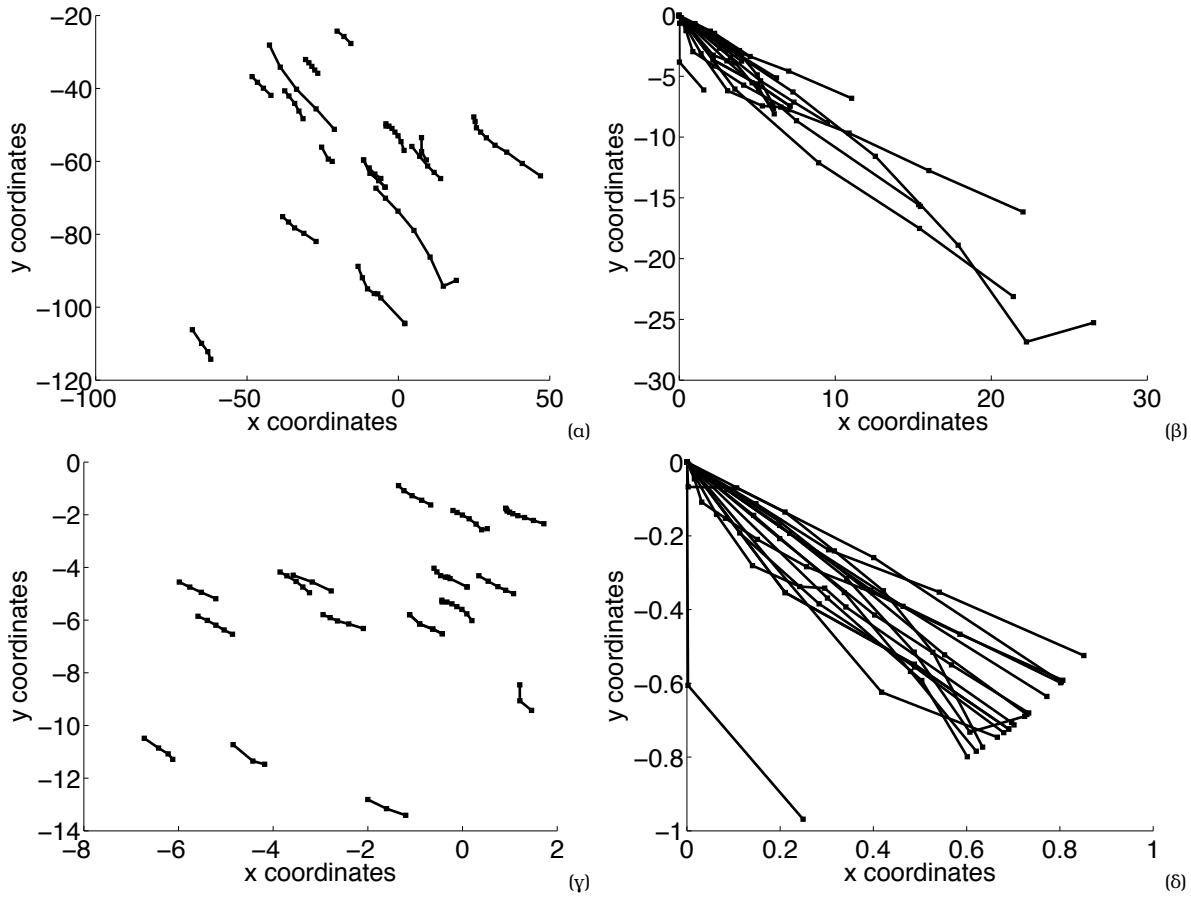
Κατασκευή δυναμικών υπομονάδων

Η δυναμική αποτελεί σημαντικό χαρακτηριστικό για τη μοντελοποίηση των κινήσεων. Έτσι για την κατασκευή των δυναμικών υπομονάδων λαμβάνουμε υπόψη τη δυναμική χρησιμοποιώντας ακολουθίες πλαισίων βίντεο. Στη συνέχεια, παρουσιάζουμε τα διανύσματα χαρακτηριστικών και τους αλγορίθμους συσταδοποίησης που χρησιμοποιήσαμε για τη μοντελοποίηση των διαφορετικών κινήσεων.

Διανύσματα χαρακτηριστικών: Τα διανύσματα χαρακτηριστικών που χρησιμοποιήσαμε για τη μοντελοποίηση των δυναμικών υπομονάδων είναι είτε η κατεύθυνση του χεριού, είτε η θέση του (τροχιά) κανονικοποιημένη σε σχέση με την αρχική θέση και το μέγεθος της κίνησης. Το διάνυσμα της κατεύθυνσης της κίνησης $\boldsymbol{D}$, έχει οριστεί στην ενότητα 2.3.1. Στη συνέχεια, περιγράφουμε τα βήματα κανονικοποίησης που εφαρμόζουμε στο διάνυσμα χαρακτηριστικών της θέσης των χεριών $\boldsymbol{P}$. Σε κάθε κίνηση η ακολουθία θέσεων (τροχιά) εξαρτάται από την αρχική θέση στην οποία βρισκόταν το χέρι. Η μοντελοποίηση και 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 αυτού του προβλήματος κανονικοποιούμε τις τροχιές με βάση την αρχική τους θέση. Αυτή η κανονικοποίηση οδηγεί στη μοντελοποίηση των κινήσεων ανεξαρτήτως της αρχικής τους θέσης. Στο Σχήμα 3.6(α) απεικονίζονται οι τροχιές ενδεικτικών δυναμικών τμημάτων πάνω στον δισδιάστατο νοηματικό χώρο. Ενώ στο Σχήμα 3.6(β) παρουσιάζονται οι ίδιες τροχιές μετά από κανονικοποίηση με βάση την αρχική θέση.

Ένας επιπλέον παράγων μεταβλητότητας αποτελεί το μέγεθος της κίνησης. Έτσι, προχωράμε σε μια ακόμα κανονικοποίηση των χαρακτηριστικών, αυτή τη φορά με βάση το μέγεθος της κίνησης -βλ. Σχήμα 3.6(γ). Τέλος, στο Σχήμα 3.6(δ) παρουσιάζουμε τις τροχιές των ίδιων δυναμικών τμημάτων εφαρμόζοντας κανονικοποίηση με βάση την αρχική θέση αλλά και το μέγεθος της κίνησης.

Ομαδοποίηση των δυναμικών τμημάτων: Χρησιμοποιώντας το αποτέλεσμα της κατάτμησης όπως είδαμε στην ενότητα 3.3 προχωράμε στην αυτόματη συσταδοποίηση των δυναμικών τμημάτων. Για την επίτευξη αυτής της ομαδοποίησης χρειαζόμαστε έναν δείκτη ομοιότητας μεταξύ δύο ακολουθιών χαρακτηριστικών διανυσμάτων. Ο αλγόριθμος Dynamic Time Warping (DTW) εφαρμόζεται για τον υπολογισμό ενός πίνακα ομοιότητας μεταξύ όλων των δυναμικών τμημάτων. Συγκεκριμένα ας θεωρήσουμε δύο δυναμικά τμήματα $X = (X_1, X_2, ..., X_{T_x})$ και $Y = (Y_1, Y_2, ..., Y_{T_y})$ όπου T_x, T_y είναι ο αριθμός πλαισίων για κάθε τμήμα. Ορίζουμε το warping μονοπάτι $W = ((\tau_1, \sigma_1), ..., (\tau_N, \sigma_N))$ όπου $1 \leq \tau_i \leq T_x$, $1 \leq \sigma_i \leq T_y$, N είναι το μήκος του μονοπατιού και η σημειογραφία για το ζευγάρι (τ_i, σ_i) σημαίνει ότι το πλαίσιο τ_i του X τμήματος αντιστοιχίζεται στο πλαίσιο σ_i του τμήματος Y. Στόχος του DTW είναι η εύρεση της μικρότερης



Σχήμα 3.6: Οι τροχιές δυναμικών τμημάτων πάνω στο δισδιάστατο χώρο νοηματισμού: (α) Χωρίς κανονικοποίηση (β) Κανονικοποίηση με βάση την αρχική θέση (γ) Κανονικοποίηση με βάση το μέγεθος της κίνησης (δ) Κανονικοποίηση με βάση και την αρχική θέση και το μέγεθος της κίνησης.

απόστασης που σχετίζεται με το warping μονοπάτι:

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

όπου το μετρικό $d(X_{\tau_i}, Y_{\sigma_i})$ που χρησιμοποιούμε είναι η ευκλείδεια απόσταση. Στη 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 τη δυναμική φύση των τροχιών των δυναμικών τμημάτων. Κάθε μια από αυτές τις συστάδες αντιστοιχεί σε μια υπομονάδα η οποία στη συνέχεια θα μοντελοποιηθεί από ένα HMM.

Δυναμικές υπομονάδες ανά χαρακτηριστικό

Η συσταδοποίηση έχει ως αποτέλεσμα την διαμέριση του χώρου των χαρακτηριστικών και την ομαδοποίηση των δυναμικών τμημάτων σε συστάδες. Κάθε συστάδα αντιστοιχεί σε μια δυναμική υπομονάδα η οποία χαρακτηρίζεται από το χαρακτηριστικό που χρησιμοποιήθηκε. Στη συνέχεια, διερευνούμε τα κατάλληλα χαρακτηριστικά για την συσταδοποιηση και μοντελοποίηση των δυναμικών υπομονάδων.

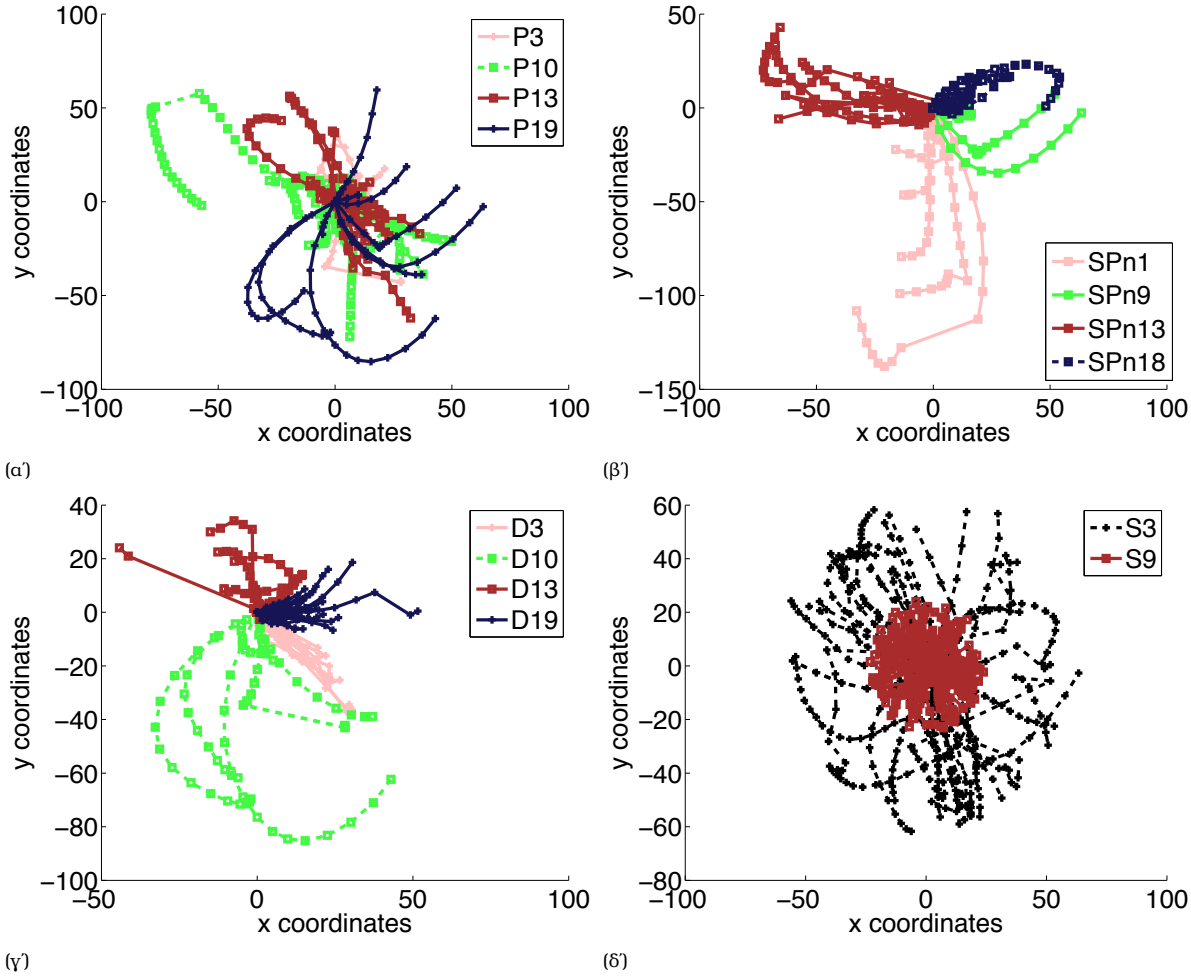
Τροχιά Κίνησης: Εφαρμόζοντας την κανονικοποίηση των τροχιών που είδαμε στην ενότητα 3.4.1 κάθε δυναμικό τμήμα αντιπροσωπεύεται από το διάνυσμα χαρακτηριστικών της κανονικοποιημένης τροχιάς. Στο Σχήμα 3.7β΄ βλέπουμε ενδεικτικές υπομονάδες οι οποίες αντιστοιχούν σε συστάδες που δημιουργήθηκαν, εφαρμόζοντας ιεραρχική συσταδοποίηση. Αυτές απεικονίζονται στον 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(β). Αντίθετα, στο Σχήμα 3.7α΄ βλέπουμε ενδεικτικά παραδείγματα υπομονάδων που προέκυψαν με εφαρμογή συσταδοποίησης και χρησιμοποιώντας ως χαρακτηριστικό την τροχιά των δυναμικών τμημάτων *χωρίς* κανονικοποίηση. Συγκρίνοντας τις υπομονάδες των Σχήματων 3.7β΄ και 3.7α΄ παρατηρούμε ότι στις υπομονάδες όπου χρησιμοποιήσαμε την τροχιά *χωρίς* κανονικοποίηση υπάρχει πολύ μεγάλη μεταβλητότητα η οποία εισάγεται λόγω της ποικιλίας των αρχικών θέσεων και του μεγέθους των κινήσεων. Επίσης, οι συστάδες που προέκυψαν είναι δύσκολο να αντιστοιχηθούν σε ενδεικτικές κινήσεις. Αντίθετα, στην περίπτωση των κανονικοποιημένων τροχιών, κάθε συστάδα αντιστοιχεί ουσιαστικά σε κίνηση διαφορετικής κατεύθυνσης. Η παραπάνω παρατήρηση ήταν αναμενόμενη καθώς η τροχιά μια κίνησης περιλαμβάνει έμμεσα ισχυρή πληροφορία σχετικά με την κατεύθυνση της κίνησης.

Κατεύθυνση και Μέγεθος Κίνησης: Οι υπομονάδες που κατασκευάστηκαν με χαρακτηριστικό διάνυσμα την κατεύθυνση, είναι αντίστοιχες με αυτές που κατασκευάστηκαν με την κανονικοποιημένη τροχιά. Κάθε υπομονάδα περιέχει κινήσεις με παρόμοια κατεύθυνση. Στο Σχήμα 3.7γ΄ 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 εμφανίζεται η υπομονάδα 'D10' είναι το νόημα 'ΕΔΩ' - βλ. Σχήμα 3.1(β). Το μέγεθος της κίνησης, που χρησιμοποιήθηκε ως παράγων κανονικοποίησης στο σημείο αυτό θα αποτελέσει παράγοντα συσταδοποίησης των κινήσεων. Στο Σχήμα. 3.7δ΄ έχουμε απεικονίσει με διαφορετικό χρώμα δύο υπομονάδες ('S9' και 'S3') χρησιμοποιώντας ως χαρακτηριστικό το μέγεθος της κίνησης. Όπως παρατηρούμε η υπομονάδα 'S9' αντιστοιχεί σε κινήσεις μικροτέρου μεγέθους από την υπομονάδα 'S3'. Ένα παράδειγμα όπου εμφανίζονται οι υπομονάδες 'S9' και 'S3' είναι τα νοήματα 'ΕΔΩ' στο Σχήμα 3.1(α) και 'ΤΕΛΟΣ' στο Σχήμα 3.1(β) αντιστοίχως.

Δυναμικές υπομονάδες πολλαπλών χαρακτηριστικών

Στην προηγούμενη ενότητα παρουσιάσαμε την κατασκευή δυναμικών υπομονάδων από τα δυναμικά τμήματα χρησιμοποιώντας ένα διάνυσμα χαρακτηριστικών κάθε φορά. Σε αυτή την ενότητα θα παρουσιάσουμε την κατασκευή υπομονάδων με την συνένωση πολλαπλών διανυσμάτων χαρακτηριστικών. Συνενώνοντας τα χαρακτηριστικά διανύσματα της κατεύθυνσης και του μεγέθους της κίνησης κατασκευάζουμε υπομονάδες που βασίζονται ταυτόχρονα τόσο στην κατεύθυνση όσο και στο μέγεθος της κίνησης. Στο Σχήμα 3.8 απεικονίζουμε τις τροχιές των δυναμικών τμημάτων που περιλαμβάνουν τέσσερις ενδεικτικές υπομονάδες. Κάθε υπομονάδα αντιπροσωπεύει κινήσεις με συγκεκριμένη κατεύθυνση αλλά και συγκεκριμένο μεγέθους, σε αντίθεση με τις single-cue υπομονάδες της προηγούμενης ενότητας. Πρέπει να αναφέρουμε ότι καθώς η συσταδοποίηση γίνεται στον κοινό χώρο χαρακτηριστικών της κατεύθυνσης και του μεγέθους της κίνησης, κάποιες από τις υπομονάδες μπορεί να είναι αρκετά σύνθετες, γεγονός που εξαρτάται από την από κοινού κατανομή των πολλαπλών ροών χαρακτηριστικών.

Μερικά παραδείγματα μπορούμε να δούμε στο Σχήμα 3.8. Επιπλέον, στο Σχήμα 3.9 παρουσιάζουμε ενδεικτικά παραδείγματα υπομονάδων υπερθέτοντας το αρχικό πλαίσιο στο τελικό και τοποθετώντας ένα βέλος το οποίο υποδεικνύει το είδος της κίνησης. Πιο συγκεκριμένα στο Σχήμα 3.9 απεικονίζουμε τις κινήσεις από τρία νοήματα τις ΑΝΓ: ΚΟΥΦΟΣ, ΑΠΟΦΑΣΙΖΩ, ΜΕ-

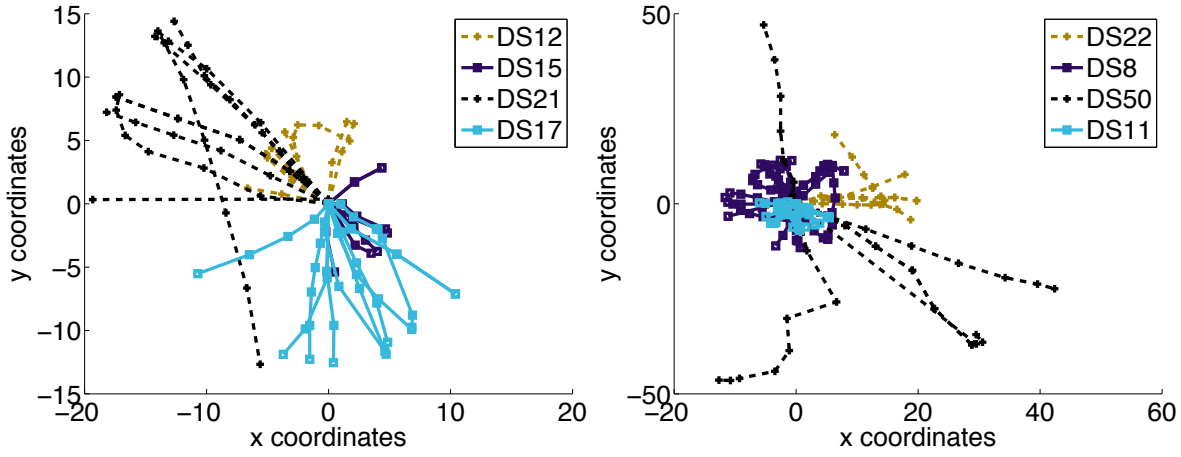


Σχήμα 3.7: Οι τροχιές διαφορετικών δυναμικών τμημάτων πάνω στον δισδιάστατο χώρο νοηματισμού μετά από κανονικοποίηση με βάση και την αρχική θέση. Το χρώμα των τροχιών ομαδοποιεί τα δυναμικά τμήματα ανάλογα με την αντίστοιχη υπομονάδα/συστάδα (subunit/cluster). Τα χαρακτηριστικά διανύσματα που χρησιμοποιήθηκαν σε κάθε περίπτωση είναι: (α) τροχιά χωρίς κανονικοποίηση, (β) τροχιά με κανονικοποίηση, (γ) κατεύθυνση της κίνησης, (δ) μέγεθος της κίνησης.

ΣΑ και τις αντίστοιχες υπομονάδες κατεύθυνσης-κλίμακας. Το Σχήμα 3.9(α) απεικονίζει την υπομονάδα κατεύθυνσης-κλίμακας D2S2 που αντιστοιχεί σε ευθεία κίνηση με κατεύθυνση D2 (πάνω-αριστερά) και κλίμακα S2 (μικρή). Τα Σχήματα 3.9(β,γ) απεικονίζουν την υπομονάδα κατεύθυνσης-κλίμακας D1S4 που αντιστοιχεί σε ευθεία κίνηση με κατεύθυνση D1 (κάτω-δεξιά) και κλίμακα S4 (μεσαία).

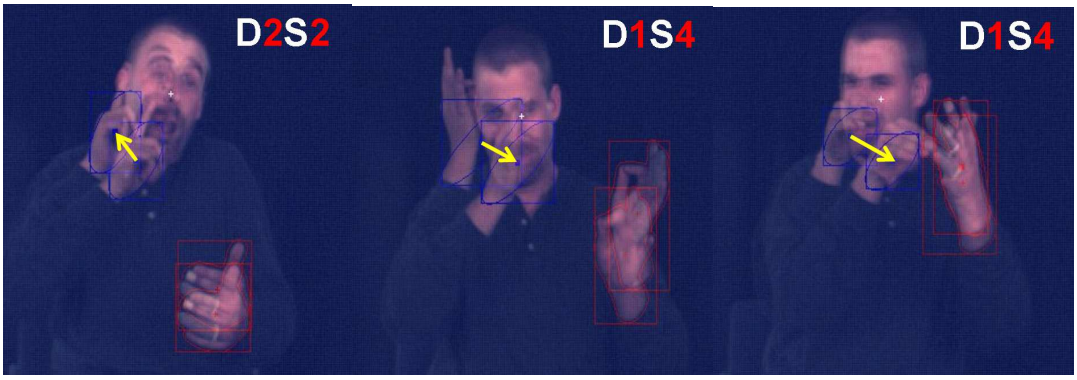
Στατικές υπομονάδες

Βασιζόμενοι στην κατάτμηση και ταξινόμηση σε Δ/Σ τμήματα, παρουσιάσουμε την κατασκευή των στατικών υπομονάδων ομαδοποιώντας τα στατικά τμήματα. Για την μοντελοποίηση τους θα χρησιμοποιήσουμε *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 συσταδοποίησης K-means με διάνυσμα χαρακτηριστικών την μη-κανονικοποιημένη θέση του χεριού $\boldsymbol{P}$. Το Σχήμα 3.10 απεικονίζει πάνω στο νοηματικό χώρο τις διαφορετικές συστάδες



(α') (β')

Σχήμα 3.8: 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.



(α') (β') (γ')

Σχήμα 3.9: Παραδείγματα υπομονάδων κατεύθυνσης-κλίμακας 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 αντιστοίχως.

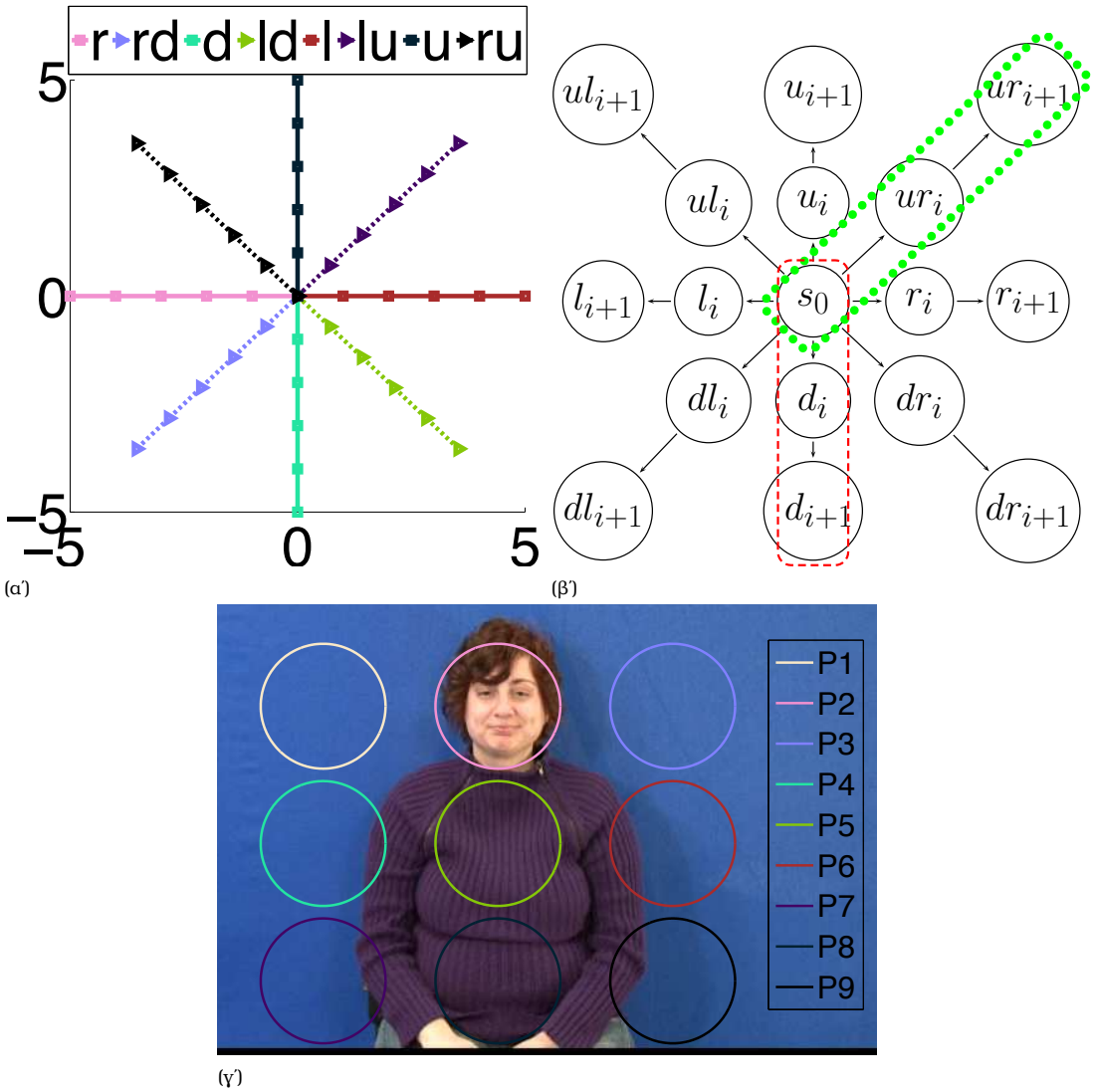
Σχήμα 3.10: Διαμέριση του δισδιάστατου νοηματικού χώρου χρησιμοποιώντας K-means για την κατασκευή των στατικών υπομονάδων.

και το αντίστοιχο κεντροειδές τους. Κάθε συστάδα αντιστοιχεί σε μια στατική υπομονάδα η οποία μοντελοποιεί μια περιοχή άρθρωσης.

3.4.2 RAW δυναμικές και στατικές υπομονάδες

Οι 2-S-U δεδομενοκεντρικές υπομονάδες που παρουσιάστηκαν στην ενότητα 3.4.1 έχουν το αρνητικό ότι κατασκευάζονται με αλγορίθμους αυτόματης συσταδοποίησης. Με άλλα λόγια γίνεται μια ομαδοποίηση του χώρου των χαρακτηριστικών, η οποία καθοδηγείται πλήρως από τα δεδομένα εκπαίδευσης. Αυτό έχει ως επακόλουθο οι παραγόμενες υπομονάδες να μοντελοποιούν συστάδες δεδομένων, οι οποίες δεν αντιστοιχούν σε υπομονάδες με βάση κάποιο γλωσσολογικό-φωνητικό σύστημα. Επιπλέον, ένα σύνηθες φαινόμενο είναι η έλλειψη επαρκών δεδομένων εκπαίδευσης για την κάλυψη του συνόλου του χώρου χαρακτηριστικών. Αυτό έχει ως αποτέλεσμα την έλλειψη ή την φτωχή εκπαίδευση μερικών μοντέλων υπομονάδας. Παραδείγματος χάριν, έστω ότι μια υπομονάδα δεν έχει εμφανιστεί στα δεδομένα εκπαίδευσης. Αλλά κατά τη διάρκεια της αναγνώρισης έχει δοθεί ένα δεδομένο προς αναγνώριση το οποίο εμπεριέχει την συγκεκριμένη υπομονάδα. Τότε ο αλγόριθμος αναγνώρισης θα αναγκαστεί να την αντιστοιχίσει στην υπομονάδα η οποία βρίσκεται πλησιέστερα με βάση το κριτήριο maximum-likelihood παρότι δεν αντιστοιχεί στο συγκεκριμένο μοντέλο. Αυτό το γεγονός επηρεάζει αρκετά την τελική αναγνώριση.

Με τα RAW μοντέλα υπομονάδας αντιμετωπίζονται τα παραπάνω προβλήματα κατασκευάζοντας ντετερμινιστικά όλα τα μοντέλα υπομονάδας τα οποία είναι απαραίτητα για την μοντελοποίηση των νοημάτων ενώ ταυτόχρονα είναι φωνητικά ερμηνεύσιμα. Παρόλα αυτά ασχολούμαστε με την πιο απλή περίπτωση όπου οι χώροι χαρακτηριστικών και οι αντίστοιχες υπομονάδες που τους μοντελοποιούν μπορούν να προκύψουν από ομοιόμορφο διαχωρισμό, όπως π.χ. οι ευθείες, καμπύλες και κυκλικές κινήσεις των χεριών, με διαφορετικές κατευθύνσεις στον νοηματικό χώρο.

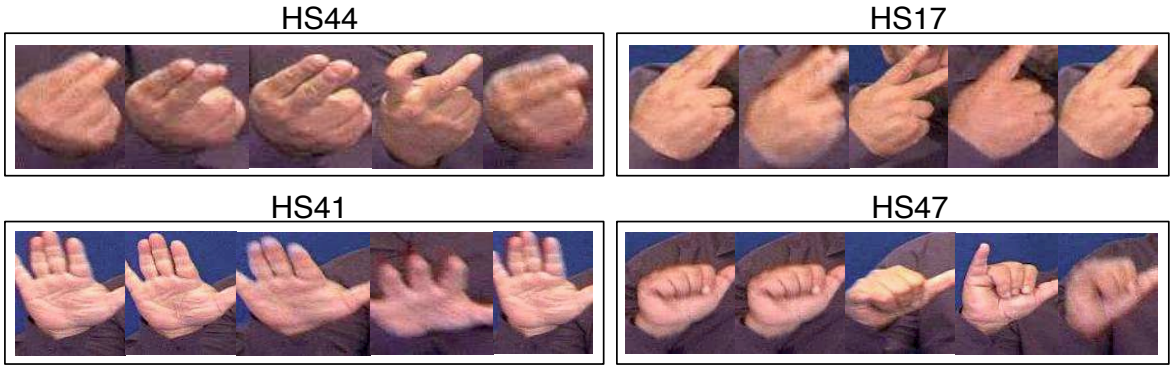


(α΄) (β΄) (γ΄)

Σχήμα 3.11: RAW δυναμικές υπομονάδες: α) Διαμέριση του χώρου χαρακτηριστικών της κατεύθυνσης κίνησης για τις ευθείες κινήσεις (right (r), left (l), up (u), down (d)). β) Τα αντίστοιχα HMM μοντέλα. RAW στατικές υπομονάδες: γ) διαμέριση του δισδιάστατου νοηματικού χώρου.

Κατασκευή των RAW υπομονάδων

RAW δυναμικές υπομονάδες: Οι φωνητικές υπομονάδες που μοντελοποιούν την κίνηση των χεριών και αντιστοιχούν σε σύμβολα του HamNoSys συστήματος χαρακτηρίζονται από συμμετρία. Παραδείγματος χάριν οι ευθείες κινήσεις διαμοιράζονται ομοιόμορφα στον δισδιάστατο χώρο της κατεύθυνσης της κίνησης. Για τη μοντελοποίηση των δυναμικών υπομονάδων λαμβάνουμε υπόψη τις ευθείες, τις καμπύλες και τις κυκλικές κινήσεις. Το σύνολο αυτών των κινήσεων μας παρέχουν αρκετά μεγάλη ποικιλία για την περιγραφή των περισσοτέρων νοημάτων παρότι υπάρχουν επιπλέον HamNoSys σύμβολα για την περιγραφή πιο περίπλοκων κινήσεων. Λαμβάνοντας υπόψη την συμμετρία, κατασκευάζουμε ντετερμινιστικά στατιστικά μοντέλα τα οποία μοντελοποιούν όλες αυτές τις διαφορετικές κινήσεις. Πιο συγκεκριμένα, κάνουμε μια ομοιόμορφη διαμέριση του χώρου χαρακτηριστικών της κατεύθυνσης της κίνησης, παράγοντας όλα τα διαφορετικά μοντέλα υπομονάδας που μοντελοποιούν ευθείες, καμπύλες και κυκλικές κινήσεις. Στο Σχήμα 3.11α΄ απεικονίζονται οι παραγόμενες ευθείες κινήσεις πάνω στον νοηματικό χώρο, κανονικοποιημένες ως προς την αρχική θέση. Στη συνέχεια, για την μοντελοποίησή τους χρησιμοποιούμε ένα HMM χρη-



Σχήμα 3.12: Παραδείγματα από διαφορετικές υπομονάδες χειρομορφής από την βάση δεδομένων GSL-Lem.

σιμοποιώντας ως διάνυσμα χαρακτηριστικών την κατεύθυνση της κίνησης. Οι μέσες τιμές σε κάθε κατάσταση αντιστοιχούν στην τιμή της κατεύθυνσης σε κάθε marker στο Σχήμα 3.11α΄. Επιπλέον επιλέγεται ίση διακύμανση σε κάθε κατάσταση, έτσι ώστε να μην υπάρχει επικάλυψη μεταξύ των διαφορετικών Γκαουσιανών μοντέλων. Στο Σχήμα 3.11β΄ απεικονίζονται τα HMM μοντέλα για όλες τις ευθείες κινήσεις. Η κατασκευή των καμπύλων και κυκλικών κινήσεων γίνεται με αντίστοιχο τρόπο.

RAW στατικές υπομονάδες: Για την κατασκευή των στατικών υπομονάδων γίνεται ομοιόμορφη διαμέριση του δισδιάστατου νοηματικού χώρου. Στο Σχήμα 3.11γ΄ απεικονίζεται η παραπάνω διαμέριση. Αυτές οι στατικές υπομονάδες δεν αντιστοιχούν σε HamNoSys σύμβολα όπως στην περίπτωση των δυναμικών υπομονάδων. Εν συνεχεία, χρησιμοποιούμε ένα GMM για την μοντελοποίηση κάθε στατικής υπομονάδας. Οι μέσες τιμές για κάθε στατική υπομονάδα αντιστοιχούν στην δισδιάστατη θέση του κέντρου του κάθε κύκλου στο Σχήμα 3.11γ΄. Επιπλέον σε κάθε μοντέλο επιλέγεται διακύμανση ίση με την ακτίνα του κάθε κύκλου.

3.4.3 Υπομονάδες Χειρομορφής

Για την κατασκευή των υπομονάδων χειρομορφής δεν χρησιμοποιούμε την κατάτμηση σε Δ/Σ υπομονάδες. Κατασκευάζονται δεδομενοκεντρικά αντίστοιχα με την προσέγγιση στο άρθρο [11]. Συγκεκριμένα, εφαρμόζεται ο αλγόριθμος συσταδοποίησης K-means σε όλα τα πλαίσια του βίντεο χρησιμοποιώντας την ευκλείδεια απόσταση. Μετά την διαμέριση του χώρου των χαρακτηριστικών της χειρομορφής, κάθε συστάδα αντιστοιχίζεται σε μια υπομονάδα χειρομορφής. Στο Σχήμα 3.12 παρουσιάζουμε τέσσερις διαφορετικές υπομονάδες χειρομορφών ('HS44','HS17','HS41' και 'HS47') όπως κατασκευάστηκαν μετά τη συσταδοποίηση. Επιπλέον, απεικονίζουμε ενδεικτικά παραδείγματα χειρομορφών που ανήκουν σε αυτές τις υπομονάδες. Όπως αναμενόταν, κάθε υπομονάδα εμπεριέχει αρκετά παρόμοιες χειρομορφές.

3.5 Λεξικό με υπομονάδες

Το λεξικό επιπέδου υπομονάδων αποτελεί μια αντιστοίχιση των νοημάτων σε ακολουθίες υπομονάδων. Με άλλα λόγια το λεξικό μάς υποδεικνύει την ακολουθία των υπομονάδων από την οποία απαρτίζεται κάθε νόημα. Κατασκευάζουμε δύο λεξικά, ένα για τη ροή πληροφορίας της κίνησης-θέσης και ένα για την πληροφορία της χειρομορφής. Για τη ροή πληροφορίας της κίνησης-θέσης παράγουμε πολλαπλά λεξικά αναλόγως με τη μέθοδο κατασκευής υπομονάδων που χρησιμοποιείται κάθε φορά. Για τις δύο κύριες μεθόδους που παρουσιάστηκαν σε αυτό το

Σχήμα 3.13: Η αποδόμηση του νοήματος 'ANY' της ΑΝΓ από τη βάση δεδομένων ASLLVD σε υπομονάδες. Σύγκριση των προτεινόμενων μεθόδου (2-S-U,RAW) με άλλες μεθόδους από την διεθνή βιβλιογραφία.

κεφάλαιο (2-S-U και RAW) το λεξικό που παράγεται κληρονομεί τη διαδοχική δομή των Δ/Σ υπομονάδων. Αυτή βασίζεται στο αποτέλεσμα της κατάτμησης σε Δ/Σ τμήματα που παρουσιάστηκε στην ενότητα 3.3. Επιπλέον, παρουσιάζουμε τα λεξικά τα οποία έχουν παραχθεί εφαρμόζοντας μεθόδους από στην διεθνή βιβλιογραφία, με στόχο τη σύγκριση των προτεινόμενων μεθόδων. Για τη ροή πληροφορίας της χειρομορφής, αφού κάνουμε χρήση της μεθόδου κατασκευής υπομονάδων χειρομορφής, παράγουμε ένα λεξικό υπομονάδας. Η μέθοδος κατασκευής υπομονάδων χειρομορφής, παρουσιάστηκε στην ενότητα 3.4.3 και βασίζεται στο άρθρο [11]. Το λεξικό αυτό δεν χρησιμοποιεί το αποτέλεσμα της κατάτμησης σε Δ/Σ τμήματα αλλά για την κατασκευή των υπομονάδων χειρομορφής βασίζεται στη συσταδοποίηση ανεξάρτητων χρονικών πλαισίων βίντεο.

3.5.1 Λεξικό για τη ροή της κίνησης-θέσης

Το λεξικό περιέχει μια εγγραφή για κάθε εκτέλεση ενός νοήματος στα δεδομένα εκπαίδευσης. Η κατασκευή του λεξικού για την 2-S-U μέθοδο βασίζεται στον συνδυασμό των Δ/Σ υπομονάδων που προκύπτουν, μετά την συσταδοποίηση των Δ/Σ τμημάτων. Αντίθετα, η κατασκευή του λεξικού για την RAW μέθοδο βασίζεται στην εφαρμογή του αλγορίθμου Viterbi, για την εύρεση της πιο πιθανής ακολουθίας RAW υπομονάδων.

Για τις μεθόδους 2-S-U και RAW οι οποίες κληρονομούν τη διαδοχική δομή των Δ/Σ υπομονάδων, κάθε υπομονάδα αντιπροσωπεύεται από ένα σύμβολο, συνενώνοντας τα εξής:

1) τον τύπο της υπομονάδας, δυναμική -Dynamic (D)- ή στατική -Static (S)- και

2) ένα αριθμό που υποδεικνύει την ταυτότητα της συστάδας που αντιστοιχήθηκε.

Επιπλέον, χρησιμοποιήσαμε και άλλες μεθόδους από τη διεθνή βιβλιογραφία για την κατασκευή δεδομενοκεντρικών λεξικών υπομονάδας με τις οποίες και συγκρίνουμε τις προτεινόμενες μεθόδους. Αυτές είναι οι Fang et al., 2004 [47] (SU-Segm), Bauer and Kraiss, 2001 [11] (SU-Frame), και η SU-noDSC. Καμία απο τις παραπάνω μεθόδους δεν διαχωρίζει σε δυναμικές και στατικές υπομονάδες. Όλες οι υπομονάδες είναι ενός τύπου. Η SU-noDSC είναι η πιο κοντινή μέθοδος στην 2-S-U αφού χρησιμοποιεί ακριβώς τον ίδιο αλγόριθμο κατάτμησης με την 2-S-U. Παρόλα αυτά δεν διαχωρίζει σε Δ/Σ υπομονάδες με αποτέλεσμα η συσταδοποίηση να γίνεται σε όλα τα τμήματα ανεξαρτήτως των δυναμικών ή στατικών ετικετών. Οι μέθοδοι SU-Segm και SU-Frame χρησιμοποιούν διαφορετικούς αλγορίθμους για την κατάτμηση και συσταδοποίηση των τμημάτων όπως περιγράφεται στις αντίστοιχες δημοσιεύσεις[47, 11]. Περισσότερες πληροφορίες για τις παραπάνω μεθόδους αναφέρονται στην ενότητα 1.2. Για τις μεθόδους SU-noDSC, SU-Segm και SU-Frame κάθε υπομονάδα χαρακτηρίζεται από έναν αριθμό που υποδεικνύει την ταυτότητα της συστάδας που αντιστοιχήθηκε. Παραδείγματος χάριν, η υπομονάδα SU10 στο Σχήμα 3.13,

υποδεικνύει ότι χρησιμοποιώντας την μέθοδο SU-noDSC, το πρώτο χρονικό τμήμα αντιστοιχήθηκε στη δέκατη συστάδα.

2-S-U και RAW αποτελέσματα: Όπως βλέπουμε στο Σχήμα 3.13 η 2-S-U και η RAW μέθοδοι έχουν ως αποτέλεσμα την ίδια χρονική κατάτμηση σε Δ/Σ τμήματα. Αυτό οφείλεται στην χρησιμοποίηση των ίδιων Γκαουσιανών μοντέλων ταχύτητας για την κατάτμηση σε Δ/Σ τμήματα. Κάθε νόημα και στις δύο μεθόδους αποδομείται σε μια ακολουθία από Δ/Σ υπομονάδες οι οποίες περιγράφουν την άρθρωση των αντίστοιχων κινήσεων και στάσεων. Οι κινήσεις μοντελοποιούνται από τις δυναμικές υπομονάδες και οι στάσεις από τις στατικές. Χρησιμοποιώντας την μέθοδο 2-S-U, το νόημα 'ΑΝΥ' της ΑΝΓ αποτελείται από μια στατική υπομονάδα S5, από δύο διαδοχικές δυναμικές D6 και D8 και μια στατική S1. Αντιθέτως, με την χρήση της μεθόδου RAW, το νόημα 'ΑΝΥ' αποτελείται από μια στατική υπομονάδα S2, από δύο διαδοχικές δυναμικές D1 και D5 και μια στατική S4. Οι υπομονάδες που χρησιμοποιούν την 2-S-U μέθοδο, αντιστοιχούν σε συστάδες των οποίων η κατασκευή έχει γίνει με αυτόματο τρόπο και καθοδηγούμενη πλήρως από τα δεδομένα εκπαίδευσης. Ενώ αντίθετα, οι υπομονάδες που χρησιμοποιούν την RAW μέθοδο, αντιστοιχούν σε συστάδες οι οποίες έχουν κατασκευαστεί με ομοιόμορφη διαμέριση του χώρου χαρακτηριστικών (ενότητα 3.4.2).

Συγκρίσεις αποτελεσμάτων: Οι μέθοδοι SU-noDSC, RAW και 2-S-Uέχουν ως αποτέλεσμα την ίδια χρονική κατάτμηση εφόσον χρησιμοποιούν το ίδιο αλγόριθμο κατάτμησης. Η ειδοποιός διαφορά τους είναι ότι η SU-noDSC δεν κάνει τον διαχωρισμό σε Δ/Σ υπομονάδες με αποτέλεσμα να συνενώνει τις ροές πληροφορίας της θέσης και κίνησης και να κατασκευάζει τις υπομονάδες εφαρμόζοντας συσταδοποίηση σε όλα τα χρονικά τμήματα ανεξαρτήτως της Δ/Σ ετικέτας στον από κοινού χώρο χαρακτηριστικών. Οι μέθοδοι SU-Segm και SU-Frame οδηγούν σε διαφορετική χρονική κατάτμηση, αφού χρησιμοποιούν διαφορετικούς αλγορίθμους κατάτμησης. Αυτή η χρονική κατάτμηση δεν χαρακτηρίζεται από την έννοια της διαδοχής Δ/Σ τμημάτων. Αυτό έχει ως συνεπακόλουθο τα χρονικά τμήματα να μην αντιστοιχούν απαραιτήτως σε κινήσεις ή στάσεις. Έτσι μια κίνηση ή μια στάση μπορεί να κατατμηθεί σε περισσότερα του ενός χρονικά τμήματα. Επιπλέον η αντιστοίχιση των χρονικών τμημάτων σε υπομονάδες βασίζεται στη συσταδοποίηση στον από κοινού χώρο χαρακτηριστικών της κίνησης-θέσης όπως συμβαίνει με την μέθοδο SU-noDSC.

3.5.2 Λεξικό για την ροή της χειρομορφής

Μετά την κατασκευή των υπομονάδων χειρομορφής, αυτές συνδυάζονται για την κατασκευή τού αντιστοίχου λεξικού υπομονάδας. Για κάθε διαφορετική προφορά ενός νοήματος εισάγεται μια νέα εγγραφή στο λεξικό η οποία αποτελείται από μια ακολουθία υπομονάδων χειρομορφής. Κάθε υπομονάδα χειρομορφής χαρακτηρίζεται από την ταυτότητα της συστάδας που αντιστοιχήθηκε. Παραδείγματος χάριν η υπομονάδα 'HS44' είναι μια υπομονάδα χειρομορφής που αντιστοιχεί στην συστάδα 44. Στο Σχήμα 3.14 απεικονίζουμε την αποδόμηση σε υπομονάδες χειρομορφής για δύο νοήματα της ΕΝΓ: το 'ΒΛΕΠΩ' και το 'ΕΞΩΤΕΡΙΚΟ' όπως εκτελέστηκαν στην βάση δεδομένων GSL-Lem. Επιπλέον, στο Σχήμα 3.12 απεικονίζονται ενδεικτικά παραδείγματα χειρομορφών που ανήκουν στις συγκεκριμένες υπομονάδες. Όπως παρατηρούμε το νόημα 'ΒΛΕΠΩ' αποτελείται από τις υπομονάδες 'HS17' και 'HS44'. Παρόλο που αυτές οι υπομονάδες αντιστοιχούν στην ίδια χειρομορφή, διαφοροποιούνται λόγω της μεταβολή της τρισδιάστατης πόζας τους. Τέλος, το νόημα 'ΕΞΩΤΕΡΙΚΟ' αποτελείται από τις υπομονάδες 'HS47' και 'HS41' οι οποίες αντιστοιχούν σε διαφορετικές χειρομορφές.

(α')

(β')

Σχήμα 3.14: Αποδόμηση σε υπομονάδες χειρομορφής για τα νοήματα 'ΒΛΕΠΩ' (α) και 'ΕΞΩΤΕΡΙΚΟ' (β) της ΕΝΓ από την βάση δεδομένων GSL-Lem.

3.6 Στατιστική μοντελοποίηση, εκπαίδευση και αναγνώριση με υπομονάδες

Τα HMM μοντέλα των RAW υπομονάδων που παρουσιάστηκαν στην ενότητα 3.4.2 έχουν παραχθεί εκ κατασκευής. Οι παράμετροι των μοντέλων αυτών έχουν καθοριστεί με βάση την διαμέριση του χώρου χαρακτηριστικών που χρησιμοποιήθηκε σε κάθε περίπτωση. Αυτό έχει ως αποτέλεσμα να μην είναι απαραίτητη η εκπαίδευσή τους. Αντίθετα, τα HMM μοντέλα των 2-S-U υπομονάδων και των υπομονάδων χειρομορφής είναι απαραίτητο να εκπαιδευτούν. Με βάση τα λεξικά που κατασκευάστηκαν όπως περιγράψαμε στην προηγούμενη ενότητα, κάθε νόημα αντιπροσωπεύεται από μια ακολουθία 2-S-U Δ/Σ υπομονάδων και υπομονάδων χειρομορφής. Επιπλέον γνωρίζουμε τα χρονικά όρια της κάθε υπομονάδας. Στόχος μας είναι να χρησιμοποιήσουμε ένα πιθανοτικό πλαίσιο για τη στατιστική εκπαίδευση των υπομονάδων και τέλος για την αναγνώριση τους. Αυτό θα πρέπει να λαμβάνει υπόψη του τη διαδοχική δομή των Δ/Σ υπομονάδων αλλά και επίσης την παραλληλία των ροών πληροφορίας. Για την εκπαίδευση θέλουμε να επιβάλουμε τη διαδοχική δομή των Δ/Σ υπομονάδων που απορρέει από την κατάτμηση σε Δ/Σ τμήματα. Επιπλέον θέλουμε να χρησιμοποιήσουμε τα τμήματα που περιέχονται σε κάθε συστάδα μετά την συσταδοποίηση για την εκπαίδευση στατιστικών μοντέλων υπομονάδας. Κατά την αναγνώριση, δεδομένης μιας ακολουθίας από παρατηρήσεις, στόχος μας είναι να βρούμε και την πιο πιθανή κατάτμηση σε Δ/Σ τμήματα, αλλά και το στατιστικό μοντέλο υπομονάδας που ταιριάζει σε κάθε Δ/Σ τμήμα αντιστοίχως. Τα παραπάνω εκπληρώνονται χρησιμοποιώντας multistream HMMs στα οποία ενσωματώνουμε τα Γκαουσιανά μοντέλα ταχύτητας για τα Δ/Σ τμήματα που εκπαιδεύτηκαν κατά τη διαδικασία της κατάτμησης (ενότητα 3.3). Περισσότερες πληροφορίες σε σχέση με τα multistream HMM μοντέλα υπομονάδων αναφέρονται στην ενότητα 5.1. Η εκπαίδευση των 2-S-U Δ/Σ υπομονάδων και των υπομονάδων χειρομορφής γίνεται ανεξάρτητα, χρησιμοποιώντας τον αλγόριθμο εκπαίδευσης που ακολουθεί.

Εκπαίδευση HMM: Δεδομένου του λεξικού υπομονάδων και των χρονικών ορίων κάθε υπομονάδας αρχικοποιούμε τα HMM μοντέλα υπομονάδων χρησιμοποιώντας μια επαναλαμβανόμενη διαδικασία. Αρχικά για κάθε υπομονάδα εφαρμόζεται ο αλγόριθμος Viterbi, με στόχο την εύρεση της πιο πιθανής ακολουθίας καταστάσεων του HMM μοντέλου υπομονάδας, και του αντίστοιχου log-likelihood για κάθε τμήμα που αντιπροσωπεύεται από την αντίστοιχη υπομονάδα. Στη συνέχεια εκτιμούμε τις παραμέτρους των HMM μοντέλων υπομονάδας, χρησιμοποιώντας για κάθε τμήματος την χρονική κατάτμηση στο επίπεδο των καταστάσεων των HMM. Η παραπάνω διαδικασία επαναλαμβάνεται μέχρις ότου να μην υπάρχει αύξηση του log-likelihood. Μετά από την παραπάνω αρχικοποίηση εφαρμόζουμε τον αλγόριθμο Baum-Welch [104]. Πιο συγκεκριμένα, για κάθε δεδομένο εκπαίδευσης, κατασκευάζουμε ένα δίκτυο HMM μοντέλων υπομονάδας. Βασιζόμαστε στο λεξικό για την αντιστοιχία κάθε νοήματος σε ακολουθία υπομονάδων. Αυτό το δίκτυο χρησιμοποιείται για τη συλλογή των απαραιτήτων στατιστικών. Όταν όλα τα δεδομένα εκπαίδευσης έχουν επεξεργαστεί, το σύνολο των συσσωρευμένων στατιστικών χρησιμοποιείται για την επανεκτίμηση των παραμέτρων των HMM μοντέλων υπομονάδας.

Αναγνώριση: Η αναγνώριση γίνεται χρησιμοποιώντας τα εκπαιδευμένα μοντέλα υπομονάδας και ένα HMM δίκτυο αναγνώρισης. Πρώτα κατασκευάζουμε ένα δίκτυο από HMM μοντέλα υπομονάδας για κάθε διαφορετική προφορά νοήματος, συμβουλευόμενοι το λεξικό για την αντιστοίχηση κάθε νοήματος σε ακολουθία υπομονάδων. Στα παραπάνω δίκτυα χρησιμοποιούμε τα εκπαιδευμένα HMMs μοντέλα υπομονάδας. Στη συνέχεια συνενώνουμε αυτά τα HMM δίκτυα με βάση μια γραμματική αναλόγως με το πρόβλημα αναγνώρισης. Παραδείγματος χάριν εάν το πρόβλημα είναι αναγνώριση μεμονωμένων νοημάτων, η γραμματική θα επιτρέπει την αναγνώριση ενός νοήματος σε κάθε δεδομένο. Έτσι καταλήγουμε σε ένα δίκτυο αναγνώρισης το οποίο αποτελείται από κόμβους συνδεδεμένους μεταξύ τους με ακμές, όπου κάθε κόμβος αντιστοιχεί σε ένα HMM μοντέλο υπομονάδας. Για ένα δεδομένο προς αναγνώριση διάρκειας T χρονικών πλαισίων βίντεο, κάθε μονοπάτι σε αυτό το δίκτυο το οποίο περνάει από T καταστάσεις του HMM είναι μια δυνητική υπόθεση αναγνώρισης. Κάθε τέτοιο μονοπάτι συνοδεύεται από μια πιθανότητα η οποία υπολογίζεται ως το γινόμενο της πιθανότητας κάθε μετάβασης στο μονοπάτι και της πιθανότητας κάθε HMM κατάσταση να παράγει την αντίστοιχη παρατήρηση. Κάθε χρονική στιγμή, βρίσκουμε το μονοπάτι που μεγιστοποιεί την παραπάνω πιθανότητα, το οποίο αντιστοιχεί στην πιο πιθανή ακολουθία υπομονάδων.

Κεφάλαιο 4

Στατιστική μοντελοποίηση της Νοηματικής Γλώσσας με Γλωσσικές Φωνητικές Υπομονάδες

Κρίσιμο συστατικό για την αυτόματη αναγνώριση ΝΓ είναι οι γλωσσικές-φωνητικές επισημειώσεις. Στα πλαίσια της αναγνώρισης της ΝΓ οι επισημειώσεις αυτές δεν είναι προτυποποιημένες όπως στην αναγνώριση φωνής. Ο λόγος είναι η μη ύπαρξη κατασταλαγμένων φωνολογικών ή γλωσσικών μοντέλων για την ΝΓ, όπως στην περίπτωση του προφορικού λόγου. Η έννοια της γλωσσικής φωνητικής υπομονάδας δεν είναι πλήρως καθορισμένη, με αποτέλεσμα την επικράτηση μεθόδων οι οποίες βασίζονται σε δεδομενοκεντρικές υπομονάδες.

Οι μέθοδοι που βασίζονται σε δεδομενοκεντρικές υπομονάδες, ορίζουν υπολογιστικά ένα σύνολο από βασικές μονάδες χωρίς τη χρήση φωνητικών επισημειώσεων. Ενδεικτικές μέθοδοι είναι [11, 47, 72, 126]. Οι βασικές μονάδες αυτές ονομάζονται ευρέως υπομονάδες (subunits) και αποτελούν τις μικρότερες μονάδες της ΝΓ. Κάθε νόημα αντιπροσωπεύεται από ένα σύνολο υπομονάδων και ως εκ τούτου οι υπομονάδες αυτές μπορούν να χρησιμοποιηθούν για την μοντελοποίηση της ΝΓ με απώτερο στόχο την αναγνώριση. Η προσέγγιση αυτή έχει το πλεονέκτημα ότι χρειάζεται μικρότερο αριθμό δεδομένων εκπαίδευσης. Παρόλα ταύτα, οι δεδομενοκεντρικές υπομονάδες καθοδηγούνται από τα δεδομένα με αποτέλεσμα να μην ερμηνεύονται γλωσσολογικά.

Από την άλλη μεριά, η χρήση γλωσσικών-φωνητικών επισημειώσεων και η κατασκευή των αντίστοιχων γλωσσικών-φωνητικών υπομονάδων μάς παρέχουν αναπαραστάσεις των εσωτερικών τμημάτων ενός νοήματος οι οποίες είναι γλωσσολογικά ερμηνεύσιμες. Οι γλωσσικές-φωνητικές υπομονάδες είναι απαραίτητες για την άμεση πρόσθεση νέων νοημάτων στο λεξιλόγιο και την προσαρμογή σε άλλες βάσεις δεδομένων ΝΓ, χωρίς τη χρήση επιπλέον δεδομένων εκπαίδευσης. Παρόλα αυτά, μικρή πρόοδος έχει γίνει σε μεθόδους που χρησιμοποιούν γλωσσικές-φωνητικές υπομονάδες με στόχο την αναγνώριση της ΝΓ. Το παραπάνω οφείλεται στην έλλειψη επίσημων λεξικών με φωνητικές επισημειώσεις βασισμένες σε πλήρως καθορισμένες φωνητικές υπομονάδες και τυπικά συστήματα επισημείωσης. Αυτό αντικατοπτρίζεται από την ποικιλία γλωσσικών μοντέλων και συστημάτων επισημείωσης που έχουν προταθεί, όπως π.χ. Movement-Hold [78], PDTS [65, 64], Stokoe system [118], Hamburg Notation System (HamNoSys) [103], SignWriting [119]. Πιο πρόσφατα, μερικές ερευνητικές προσεγγίσεις χρησιμοποιούν γλωσσικές-φωνητικές υπομονάδες με στόχο την αναγνώριση της ΝΓ. Αυτές χρησιμοποιούν είτε χειρωνακτικές φωνητικές επισημειώσεις, οι οποίες είναι αρκετά χρονοβόρες, όπως η εργασία στο άρθρο [136], είτε φωνητικές επισημειώσεις οι οποίες έχουν παραχθεί με αυτόματη επεξεργασία όπως οι εργασίες στα άρθρα [102, 71]. Άλλες εργασίες όπως τα άρθρα [66, 17, 55, 28], ενσωματώνουν γλωσσολογικές γνώσεις. Παρόλα αυτά δεν οδηγούν σε υπομονάδες ευρέως επαναχρησιμοποιήσιμες ούτε βασίζον-

ται σε κάποιο γνωστό σύστημα επισημείωσης ή γλωσσολογικό μοντέλο [118, 103, 119, 78].

Σε αυτό το κεφάλαιο προτείνουμε ένα καινοτόμο πλαίσιο για την αναγνώριση της ΝΓ βασιζόμενο σε γλωσσικές-φωνητικές υπομονάδες. Η κύρια συνεισφορά μας αποτελεί την ενσωμάτωση γλωσσικής-φωνητικής γνώσης στα στατιστικά μοντέλα υπομονάδων. Εκμεταλλευόμαστε τη γλωσσική-φωνητική γνώση η οποία είναι ενσωματωμένη στις PDTS επισημειώσεις που έχουν εξαχθεί από HamNoSys επισημειώσεις χρησιμοποιώντας ένα σύστημα αυτόματης συμβολικής επεξεργασίας [132]. Με αυτό τον τρόπο, κατασκευάζουμε υπομονάδες που είναι γλωσσικά-φωνητικά ερμηνεύσιμες. Για την εκπαίδευση των PDTS υπομονάδων προτείνουμε έναν αλγόριθμο εκπαίδευση ο οποίος ονομάζεται Iterative Training Algorithm (ITA). Αυτός αλλάζει τις PDTS επισημειώσεις έτσι ώστε να είναι συνεπής σε σχέση με την πραγματική άρθρωση των νοημάτων, εκμεταλλευόμενος τις πραγματικές οπτικές παρατηρήσεις.

4.1 Σύνοψη συστήματος

Μια σύνοψη του προτεινόμενου συστήματος απεικονίζεται στο Σχήμα 4.1. Το συνολικό σύστημα αποτελείται από τα εξής υποσυστήματα:

1) Μετατροπή HamNoSys σε PDTS,

2) Εκπαίδευση της δυναμικής,

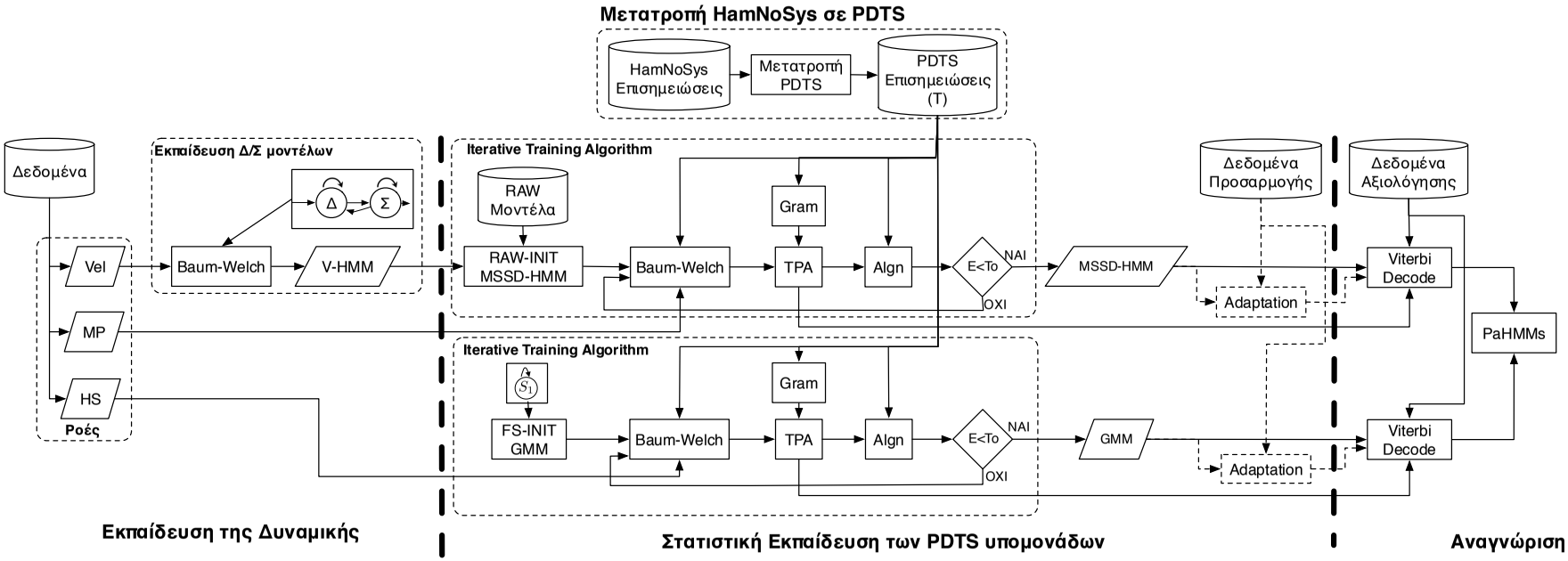
3) Στατιστική εκπαίδευση των PDTS υπομονάδων,

4) Αναγνώριση.

Μετατροπή HamNoSys σε PDTS: Χρησιμοποιούμε ένα σύστημα αυτόματης συμβολικής επεξεργασίας [132, 102] για τη μετατροπή των HamNoSys επισημειώσεων σε PDTS (βλ. ενότητα 4.2). Οι PDTS επισημειώσεις, μας προσφέρουν άμεση γλωσσική-φωνητική πληροφορία. Επιπλέον εμπεριέχουν στη φωνητική δομή την έννοια της διαδοχής υπομονάδων στον χρόνο. Τα δύο παραπάνω χαρακτηριστικά καθιστούν κατάλληλες τις PDTS επισημειώσεις για την μοντελοποίηση και αυτόματη αναγνώριση της ΝΓ χρησιμοποιώντας HMMs.

Εκπαίδευση της δυναμικής: Εκμεταλλευόμαστε το διάνυσμα χαρακτηριστικών της ταχύτητας και εκπαιδεύουμε δύο Γκαουσιανά μοντέλα (V-HMM). Ένα μοντελοποιεί τις στάσεις και ένα τις κινήσεις. Για τον συνδυασμό των Γκαουσιανών μοντέλων χρησιμοποιούμε ένα εργοδικό HMM δύο καταστάσεων. Αυτό εκπαιδεύεται χρησιμοποιώντας τον αλγόριθμο Baum-Welch και όλα τα δεδομένα εκπαίδευσης (βλ. ενότητα 3.3). Στη συνέχεια χρησιμοποιούμε αυτά τα δύο Γκαουσιανά μοντέλα για την αρχικοποίηση των σ.π.π. της ταχύτητας στα PDTS μοντέλα υπομονάδας.

Στατιστική εκπαίδευση των PDTS υπομονάδων: Η κύρια συνεισφορά μας είναι η ενσωμάτωση στα στατιστικά μοντέλα υπομονάδας, της γλωσσικής-φωνητικής πληροφορία που εμπεριέχεται στις PDTS επισημειώσεις (T). Για την εκπαίδευση των PDTS υπομονάδων προτείνουμε τον αλγόριθμο ITA. Αυτός ο αλγόριθμος διορθώνει τις PDTS επισημειώσεις ($\overline{T}$) έτσι ώστε να υπάρχει πλήρης αντιστοιχία μεταξύ των PDTS επισημειώσεων και της πραγματικής άρθρωσης ενός νοήματος (βλ. ενότητα 4.3). Για την μοντελοποίηση των PDTS υπομονάδων χρησιμοποιώντας τη ροή πληροφορίας της κίνησης-θέσης χρησιμοποιούμε MSSD-HMMs (βλ. ενότητα 5.1). Αντίθετα, για τις υπομονάδες χειρομορφής χρησιμοποιούμε ένα απλό Γκαουσιανό μοντέλο. Τέλος, για την προσαρμογή των μοντέλων υπομονάδας και του PDTS λεξικού σε ένα νέο νοηματιστή χρησιμοποιούμε το πλαίσιο προσαρμογής που παρουσιάζεται στην ενότητα 5.3.



Σχήμα 4.1: Σύνοψη προτεινόμενου συστήματος κάνοντας χρήση γλωσσικής-φωνητικής πληροφορίας. Τα τετράγωνα αντιπροσωπεύουν τις διαδικασίες, τα παραλληλόγραμμα τα δεδομένα εισόδου και εξόδου, και οι ρόμβοι τις αποφάσεις. 1) Εκπαίδευση της δυναμικής: εκμεταλλευόμαστε την ταχύτητα και εκπαιδεύουμε δύο σ.π.π.: για τις στάσεις και τις κινήσεις. 2) Στατιστική εκπαίδευση PDTS υπομονάδων: ενσωμάτωση των παραπάνω σ.π.π. και εφαρμογή του αλγορίθμου ITA για την εκπαίδευση των PDTS υπομονάδων. Επιπλέον δίνεται η δυνατότητα προσαρμογής των PDTS μοντέλων σε νέο νοηματιστή. 3) Αναγνώριση: Decoding και σύμμειξη των ροών πληροφορίας της κίνησης-θέσης και της χειρομορφής. Σε όλες τις περιπτώσεις, τα 'δεδομένα' αναφέρονται σε διανύσματα χαρακτηριστικών μετά από τη διαδικασία εξαγωγής χαρακτηριστικών. Αυτά είναι η ταχύτητα (Vel), η κίνηση-θέση (MP), και η χειρομορφή (HS). V-HMM αναφέρεται στα εκπαιδευμένα Γκαουσιανά μοντέλα ταχύτητας για κάθε τύπο υπομονάδας. RAW-INIT και FS-INIT αναφέρονται στα RAW και στα flat-start HMM μοντέλα υπομονάδας, για την αρχικοποίηση των PDTS μοντέλων υπομονάδας. Dec αναφέρεται στο Decoding, Gram στην γραμματική (grammar) και Algn στην συμβολική αντιστοίχιση (alignment). E είναι το σφάλμα μεταξύ των αρχικών PDTS επισημειώσεων T και των διορθωμένων $\overline{T}$ μετά την εφαρμογή του Dec. T_0 είναι ένα προκαθορισμένο κατώφλι. Περισσότερες πληροφορίες αναφέρονται στην ενότητα 4.3.

Αναγνώριση: Εκμεταλλευόμενοι τα εκπαιδευμένα μοντέλα υπομονάδων χρησιμοποιώντας τον ITA αλγόριθμο και το διορθωμένο PDTS λεξικό, εφαρμόζουμε τον αλγόριθμο Viterbi. Έτσι επιτυγχάνεται η εύρεση της πιο πιθανής ακολουθίας PDTS υπομονάδων, ανεξάρτητα για κάθε ροή πληροφορίας: κίνησης-θέσης και χειρομορφής. Το τελικό αποτέλεσμα αναγνώρισης προκύπτει κάνοντας εκ των υστέρων σύμμειξη των αποτελεσμάτων των ροών πληροφορίας κίνησης-θέσης και χειρομορφής χρησιμοποιώντας Parallel HMMs (PaHMMs) [136] (βλ. ενότητα 5.2). Με αυτό τον τρόπο συνδυάζουμε τη διαδοχή των PDTS υπομονάδων αλλά και την παραλληλία λόγω των πολλαπλών ροών πληροφορίας.

4.2 Νοηματική γλώσσα και γλωσσικές-φωνητικές υπομονάδες

Χρησιμοποιούμε μια μέθοδο για την μετατροπή των HamNoSys επισημειώσεων, οι οποίες περιέχουν γλωσσική-φωνητική πληροφορία για κάθε νόημα, σε σε PDTS επισημειώσεις οι οποίες είναι σύμφωνες με το γλωσσολογικό PDTS μοντέλο των Liddell και Johnson.

Το σύστημα HamNoSys [103] αποτελεί ένα σύστημα επισημείωσης της ΝΓ σε γλωσσικό-φωνητικό επίπεδο. Περιγράφει τα νοήματα επαρκώς έτσι ώστε να είναι εφικτή η αυτόματη αναπαραγωγή τους από ένα κινούμενο avatar. Η φωνητική επισημείωση βασίζεται σε ένα σύνολο από εικονικά σύμβολα, τα οποία περιγράφουν τη θέση των χεριών, την κίνησή τους, το σχήμα της χειρομορφής και τον προσανατολισμό της χειρομορφής, κατά τη διάρκεια της άρθρωσης ενός νοήματος. Η φιλοσοφία των HamNoSys επισημειώσεων είναι μινιμαλιστική, υπό την έννοια ότι αποφεύγει την καταγραφή περιττής πληροφορία. Επιπλέον βασίζεται στην περιγραφή της μεταβολής των αρθρωτών κατά την διάρκεια της άρθρωσης.

Ας πάρουμε ως παράδειγμα την επισημείωση σε HamNoSys του νοήματος 'ΚΑΡΕΚΛΑ' στην ΕΝΓ όπως απεικονίζεται στον πίνακα 4.1. Το νόημα αυτό αποτελείται από μια επαναλαμβανόμενη κίνηση προς τα κάτω και των δύο χεριών. Επιπλέον η χειρομορφή χαρακτηρίζεται από κλειστή παλάμη και δάκτυλα (γροθιά). Η φωνητική επισημείωση σε HamNoSys περιγράφει μια κίνηση προς τα κάτω του κυρίαρχου χεριού και χειρομορφή κλειστή γροθιά. Η συμμετρία των δύο χεριών και η επαναλαμβανόμενη κίνηση, υπονοούνται από το πρώτο και το τελευταίο σύμβολο αντίστοιχα. Επιπλέον, η τελική θέση των χεριών μετά την κίνηση δεν ορίζεται ρητά, αλλά μπορεί να εξαχθεί συνδυάζοντας την κίνηση προς τα κάτω από την αρχική θέση των χεριών η οποία είναι γνωστή.

Όπως παρατηρούμε στις HamNoSys επισημειώσεις αρκετή πληροφορία δεν ορίζεται ρητά αλλά υπονοείται και επιπλέον δεν υπάρχει η έννοια της χρονικής διαδοχής των φωνητικών υπομονάδων. Λόγω αυτών των δύο χαρακτηριστικών οι HamNoSys επισημειώσεις δεν είναι άμεσα χρησιμοποιούμενες σε συστήματα αυτόματης αναγνώρισης της ΝΓ, στα πλαίσια της παρούσας εργασίας. Σε συστήματα αυτόματης αναγνώρισης ΝΓ είναι απαραίτητο να παρέχεται ρητά η φωνητική πληροφορία κάθε νοήματος ανεξάρτητα εάν είναι περιττή ή όχι. Επιπλέον η έννοια της χρονικής διαδοχής των φωνητικών υπομονάδων είναι ουσιώδης ειδικά όταν η μοντελοποίηση γίνεται χρησιμοποιώντας HMMs.

Σε αντίθεση με τα HamNoSys, το PDTS σύστημα επισημείωσης [65, 64] μας παρέχει ρητά την απαραίτητη φωνητική πληροφορία κάθε νοήματος. Επιπλέον εμπεριέχει την έννοια της χρονικής διαδοχής των φωνητικών υπομονάδων. Έτσι, η αξιοποίηση των PDTS επισημειώσεων σε συστήματα αναγνώρισης της ΝΓ είναι άμεση.

Με βάση το PDTS μοντέλο, τα νοήματα αποτελούνται από τέσσερις τύπους φωνητικών υπομονάδων: postures (P), detentions (D), transitions (T), and steady shifts (S). Οι posture υπομονάδες, περιέχουν την πληροφορία σχετικά με τη θέση και τη χειρομορφή των χεριών. Οι transition υπομονάδες αντιστοιχούν στις κινήσεις των χεριών ανάμεσα σε δύο διαδοχικές posture υπομονάδες, και περιέχουν πληροφορία σχετική με την τροχιά και το είδος της κίνησης. Οι detention υπομονάδες είναι σαν τις posture, με τη διαφορά ότι τα χέρια πραγματοποιούν μια αρκετά μικρή χρονικά στάση. Τέλος, οι steady shifts υπομονάδες είναι σαν τις transition, με τη διαφορά ότι

αναφέρονται σε αργές αλλά σκόπιμες κινήσεις. Σε αυτή την εργασία δεν διαχωρίζουμε μεταξύ posture και detention υπομονάδων, ούτε μεταξύ transition και steady shift υπομονάδων.

4.2.1 Μετατροπή των HamNoSys σε PDTS φωνητικά σύμβολα

Χρησιμοποιούμε μια μέθοδο για την αυτόματη μετατροπή των HamNoSys επισημειώσεων σε PDTS επισημειώσεις [132]. Η παραπάνω μέθοδος έχει το πλεονέκτημα ότι είναι αυτόματη, με αποτέλεσμα να αποφεύγεται η ανθρώπινη εκπαίδευση και η χειρωνακτική επισημείωση σε PDTS σύμβολα. Η μετατροπή από HamNoSys σε PDTS επισημειώσεις, στόχο έχει την αποδόμηση κάθε νοήματος στις PDTS υπομονάδες από τις οποίες αποτελείται, βασιζόμενοι στην περιγραφή τους από HamNoSys σύμβολα. Η βασική δομή των HamNoSys αποτελείται από την εξής ακολουθία συμβόλων με τη συγκεκριμένη σειρά:

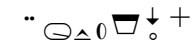
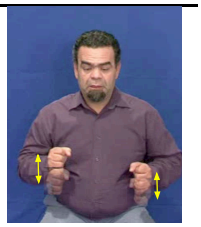
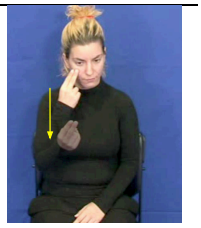
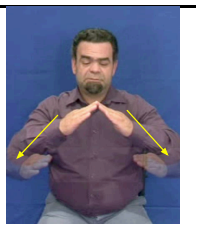
- ένδειξη συμμετρίας,
- είδος χειρομορφής,
- θέση χεριών,
- κίνηση χεριών και
- ένδειξη επανάληψης.

Η βασική ιδέα συνίσταται στη συσσώρευση των αλλαγών που περιγράφουν πως μια στάση ή κίνηση άλλαξαν σε σχέση με την προηγούμενη κατάσταση των αρθρωτών κατά τη διάρκεια εκτέλεσης ενός νοήματος. Στη συνέχεια οι διαφορές αυτές εφαρμόζονται με συγκεκριμένη σειρά για την πλήρη ανάκτηση της πληροφορίας σε κάθε χρονική κατάσταση. Η παραπάνω μέθοδος εφαρμόζεται ανεξάρτητα, σε διαφορετικές ροές πληροφορίας, όπως π.χ. τη ροή της χειρομορφής (HS), της κίνησης-θέσης (MP), είτε για το κυρίαρχο (D) είτε για το δευτερεύων χέρι (ND). Επιπλέον, μας παρέχει χρονική συσχέτιση των PDTS υπομονάδων για τις διαφορετικές ροές πληροφορίας.

Όπως έχουμε ήδη αναφέρει προηγουμένως, οι HamNoSys επισημειώσεις είναι μινιμαλιστικές. Αυτό έχει ως συνέπεια η τελική θέση των χεριών αμέσως μετά από μια κίνηση είναι άγνωστη. Η πληροφορία που έχουμε για την τελική θέση των χεριών είναι το είδος της κίνησης και την αρχική θέση των χεριών πριν από την κίνηση. Λαμβάνοντας υπόψη την πληροφορία της θέσης των χεριών ενός νοηματιστή μετά από την οπτική επεξεργασία του βίντεο, είναι δυνατόν να εξαγάγουμε την ακριβή θέση των χεριών αμέσως μετά από μια κίνηση. Παρόλα αυτά το παρόν σύστημα μετατροπής των HamNoSys σε PDTS επισημειώσεις που χρησιμοποιήθηκε δεν εξαγάγει την παραπάνω κρυμμένη πληροφορία. Αντίθετα, διαχωρίζει μεταξύ explicit και implicit θέσεις. Στις explicit θέσεις έχει αντιστοιχηθεί ένα συγκεκριμένο HamNoSys σύμβολο το οποίο περιέχει όλη την απαραίτητη πληροφορία της θέσης των χεριών κατά τη διάρκεια μιας στάσης. Σε αντίθεση, οι implicit υπομονάδες αποτελούνται από μια explicit θέση -την αρχική- ακολουθούμενη από μια ή περισσότερες κινήσεις.

Επιπλέον η μέθοδος μετατροπής χρησιμοποιεί τις epenthesis κινήσεις οι οποίες αντιστοιχούν σε κινήσεις μεταξύ δύο διαδοχικών στάσεων χωρίς να ορίζεται η ακριβής τροχιά της κίνησης. Οι epenthesis κινήσεις εμφανίζονται κυρίως στην αρχή και στο τέλος ενός νοήματος, σε νοήματα όπου έχουμε επανάληψη μιας κίνησης ή σε σύνθετα νοήματα. Μεταχειριζόμαστε αυτές τις epenthesis κινήσεις ως implicit transition υπομονάδες μιας και δεν γνωρίζουμε την ακριβή τροχιά της κίνησης. Ακόμα, η μέθοδος μετατροπής παράγει PDTS επισημειώσεις για τη ροή πληροφορίας της χειρομορφής μόνο κατά τη διάρκεια των στάσεων. Κατά τη διάρκεια των κινήσεων οι υπομονάδες χειρομορφής δεν μας παρέχουν πληροφορία σχετικά με το είδος της χειρομορφής. Έτσι μεταχειριζόμαστε αυτές τις περιπτώσεις ως implicit υπομονάδες χειρομορφής.

Πίνακας 4.1: Δύο λεξικά φωνητικού επιπέδου (HamNoSys,PDTS) για τρία νοήματα όπως εμφανίζονται στη βάση δεδομένων GSL-Lem[†]

Νόημα	HamNoSys	PDTS D/ND	PDTS M-P/HS	PDTS Ακολουθία υπομονάδων[††]	Παράδειγμα
ΚΑΡΕΚΛΑ		D	M-P	**P**–RShoulder **T**–d **P**–N/A **T**–N/A **P**–RShoulder **T**–d **P**–N/A	
		ND	M-P	**P**–LShoulder **T**–d **P**–N/A **T**–N/A **P**–LShoulder **T**–d **P**–N/A	
		D	HS	**P**–Fist **T**–N/A **P**–Fist **T**–N/A **P**–Fist **T**–d **P**–Fist	
		ND	HS	**P**–Fist **T**–N/A **P**–Fist **T**–N/A **P**–Fist **T**–d **P**–Fist	
ΒΛΕΠΩ		D	M-P	**P**–Eye **T**–do **P**–N/A	
		ND	M-P	-	
		D	HS	**P**–Finger23 **T**–N/A **P**–Finger23	
		ND	HS	-	
ΣΤΕΓΗ		D	M-P	**P**–HeadHandTouch **T**–dr **P**–N/A	
		ND	M-P	**P**–HeadHandTouch **T**–dl **P**–N/A	
		D	HS	**P**–Flat **T**–N/A **P**–Flat	
		ND	HS	**P**–Flat **T**–N/A **P**–Flat	

[†] D και ND αναφέρονται στο κυρίαρχο (dominant) και δευτερεύον (non-dominant) χέρι αντίστοιχα, M-P και HS στις ροές πληροφορίας της κίνησης-θέσης και χειρομορφής. Το σύμβολο N/A αναφέρεται στη μη διαθεσιμότητα ακριβούς επισημείωσης, η οποία χρησιμοποιείται στις implicit PDTS υπομονάδες. Τα σύμβολα P και T αντιστοιχούν σε posture και transition PDTS υπομονάδες. Οι ευθείες κινήσεις χαρακτηρίζονται από την κατεύθυνσή τους: right (r), left(l), down (d), up (u), in (i), out (o). Οι posture PDTS υπομονάδες χαρακτηρίζονται από τη θέση του χεριού στον νοηματικό χώρο, π.χ. RShoulder αντιστοιχεί στην περιοχή του δεξιού ώμου, LShoulder στην περιοχή του αριστερού ώμου και HeadHandTouch υποδεικνύει ότι και τα δύο χέρια ακουμπάνε το ένα το άλλο και βρίσκονται κοντά στο κεφάλι του νοηματιστή.

[††] Δεν απεικονίζουμε την πλήρη PDTS επισημείωση λόγω έλλειψης χώρου.

Στον πίνακα 4.1 έχουμε συμπεριλάβει τρία νοήματα της ΕΝΓ ('ΚΑΡΕΚΛΑ', 'ΒΛΕΠΩ' και 'ΣΤΕΓΗ') από τη βάση δεδομένων GSL-Lem. Επιπλέον έχουμε συμπεριλάβει τις επισημειώσεις τους σε HamNoSys και τη μετατροπή τους σε PDTS επισημειώσεις χρησιμοποιώντας την παραπάνω μέθοδο. Ας πάρουμε ως παράδειγμα το νόημα 'ΚΑΡΕΚΛΑ'. Αυτό απαρτίζεται από τέσσερις ακολουθίες PDTS υπομονάδων μια για κάθε ροή πληροφορίας. Αυτές είναι η κίνηση-θέση (MP) και η χειρομορφή (HS), για το κυρίαρχο (D) και για το δευτερεύον χέρι (ND).

Πιο συγκεκριμένα, το νόημα αυτό αποτελείται από μια μικρή επαναλαμβανόμενη κίνηση προς τα κάτω (T–d) και για τα δύο χέρια. Η αρχική θέση των χεριών είναι στο ύψος των ώμων του νοηματιστή (P–RShoulder and P–LShoulder για το D και ND χέρι αντίστοιχα). Τέλος το είδος της χειρομορφής είναι κλειστή γροθιά (P–Fist). Όπως παρατηρούμε, εμφανίζονται όλα τα είδη των implicit PDTS υπομονάδων που αναφέραμε παραπάνω. Αυτά είναι implicit posture (P–N/A), implicit transition (T–N/A) και implicit υπομονάδες χειρομορφής (T–N/A).

4.3 Εκπαίδευση γλωσσικών-φωνητικών PDTS υπομονάδων

Η PDTS επισημείωση για κάθε νόημα αντιστοιχεί σε μια πολύ συγκεκριμένη προφορά του νοήματος όπως θα εμφανιζόταν σε ένα λεξικό νοημάτων (μορφή αναφοράς -citation form-). Λόγω της ελευθερίας που υπάρχει στην άρθρωση ενός νοήματος είναι σύνηθες να διαφοροποιείται η άρθρωση από την μορφή αναφοράς. Η διαφοροποίηση αυτή οδηγεί στην εμφάνιση μιας μη αντιστοιχίας μεταξύ των PDTS επισημειώσεων και της πραγματικής εκφοράς του νοήματος. Επιπλέον,

λόγω υποθέσεων που έχουν γίνει στη μέθοδο μετατροπής (HamNoSys σε PDTS) ή ακόμα και λάθη τα οποία μπορεί να έχουν γίνει στις HamNoSys επισημειώσεις, η εμφάνιση της παραπάνω μη αντιστοιχίας αποτελεί ένα αρκετά συχνό φαινόμενο. Άλλο ένα πρόβλημα σχετικό με την αυτόματη παραγωγή των PDTS επισημειώσεων, είναι η ύπαρξη των implicit PDTS υπομονάδων. Στις συγκεκριμένες υπομονάδες δεν έχουμε σαφή PDTS επισημείωση. Ορισμένη πληροφορία είναι τελείως άγνωστη.

Στο Σχήμα 4.3 απεικονίζουμε ένα νόημα της ΕΝΓ από την βάση δεδομένων GSL-Lem όπου εμφανίζονται όλα τα παραπάνω. Η PDTS επισημείωση μεταξύ της μορφής αναφοράς (πρώτη γραμμή) και της πραγματικής άρθρωσης (τρίτης γραμμής) διαφέρουν αρκετά. Πιο συγκεκριμένα δύο PDTS υπομονάδες έχουν διαγραφεί και τρεις έχουν αντικατασταθεί από άλλες υπομονάδες. Η προαναφερθείσα μη αντιστοιχία μπορεί να χαρακτηριστεί από ένα σύνολο διαγραφών, αντικαταστάσεων ή παρεμβολών PDTS υπομονάδων από τη μορφή αναφοράς.

Στόχος μας είναι η εκπαίδευση PDTS μοντέλων υπομονάδας χρησιμοποιώντας ασθενείς PDTS επισημειώσεις σε σχέση με την πραγματική άρθρωση. Αντιμετωπίζουμε το παραπάνω πρόβλημα προτείνοντας τον αλγόριθμο Iterative Training Algorithm (ITA). Αυτός βασίζεται σε μια επαναληπτική διαδικασία μεταξύ δύο διαδοχικών βημάτων:

α) εκπαίδευση των PDTS μοντέλων υπομονάδας χρησιμοποιώντας τις PDTS επισημειώσεις

β) διόρθωση των PDTS επισημειώσεων λαμβάνοντας υπόψη τις πραγματικές παρατηρήσεις.

Στο πρώτο βήμα εφαρμόζουμε τον αλγόριθμο Baum-Welch. Στο δεύτερο βήμα εφαρμόζουμε τον αλγόριθμο Viterbi χρησιμοποιώντας μια PDTS γραμματική και τα PDTS μοντέλα υπομονάδας που εκπαιδεύτηκαν στο προηγούμενο βήμα. Το δεύτερο βήμα μπορεί να ιδωθεί ως ένα περιορισμένο πρόβλημα αναγνώρισης όπου οι επιτρεπτές ακολουθίες των PDTS υπομονάδων είναι περιορισμένες. Ο περιορισμός αυτός οφείλεται σε μια PDTS γραμματική η οποία εφαρμόζεται στις PDTS επισημειώσεις κάθε νοήματος ξεχωριστά και επιτρέπει διαγραφές, αντικαταστάσεις ή παρεμβολές των PDTS υπομονάδων. Με την εισαγωγή της PDTS γραμματικής το σύστημα εκπαίδευσης αντιμετωπίζει την ποικιλία άρθρωσης των νοημάτων και την μη αντιστοιχία της πραγματικής άρθρωσης με τη μορφή αναφοράς.

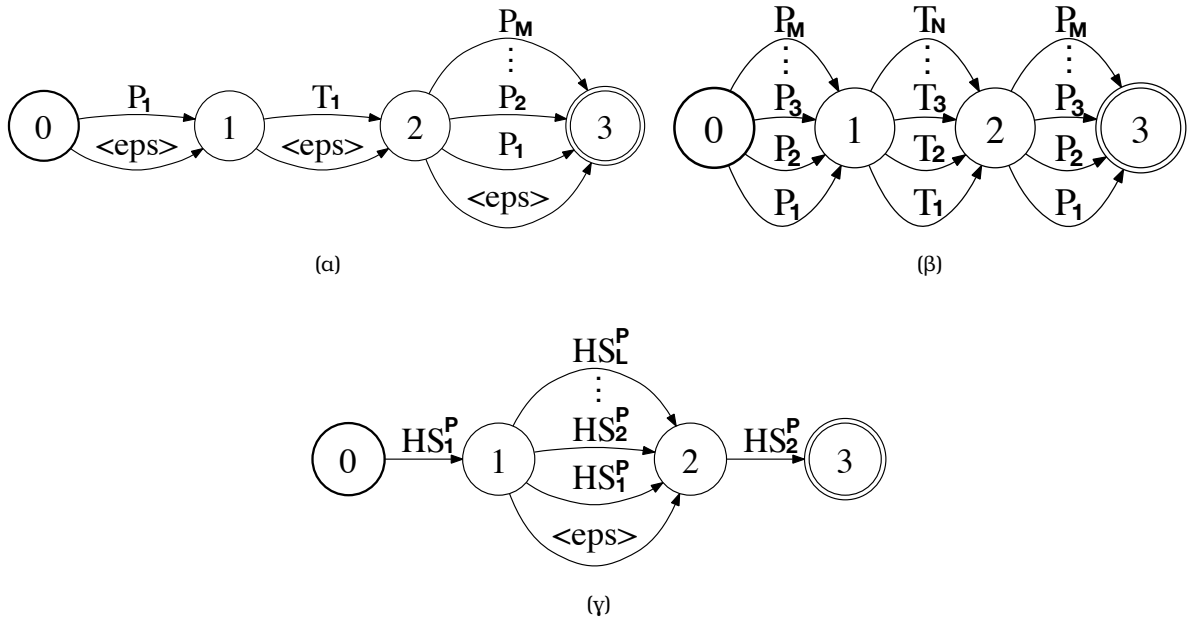
4.3.1 Iterative Training Algorithm (ITA)

Ας ορίσουμε το σύνολο $D = \{D_1, D_2, ..., D_N\}$ όπου N είναι ο αριθμός των δεδομένων εκπαίδευσης και D_i είναι το διάνυσμα χαρακτηριστικών για μια εκτέλεση ενός νοήματος. Επιπλέον το σύνολο $T = \{T_1, T_2, ..., T_N\}$ το οποίο αντιστοιχεί στις PDTS επισημειώσεις. Κάθε T_i αποτελείται από μια ακολουθία PDTS υπομονάδων η οποία περιγράφει την εκτέλεση του νοήματος με διάνυσμα χαρακτηριστικών D_i. Τέλος, έστω $\lambda = \{\lambda_1, \lambda_2, ..., \lambda_M\}$ οι παράμετροι των HMM μοντέλων (M διαφορετικές PDTS υπομονάδες).

Ο ITA αλγόριθμος αποτελείται από τα παρακάτω βήματα:

1) Αρχικοποίηση: Αρχικοποιούμε ($Init$) τα PDTS HMM μοντέλα υπομονάδας. Αυτό γίνεται είτε μέσω της flat-start διαδικασίας είτε χρησιμοποιώντας τα RAW μοντέλα που περιγράφηκαν στην ενότητα 3.4.2 (βλ. ενότητα 4.3.2).

2) Baum-Welch: Μετά την αρχικοποίηση των HMM μοντέλων (λ) εφαρμόζουμε τον αλγόριθμο Baum-Welch (BW), χρησιμοποιώντας τα δεδομένα εκπαίδευσης (D) και τις PDTS επισημειώσεις (T). Με αυτό τον τρόπο επανεκτιμούμε τις παραμέτρους όλων των PDTS HMM μοντέλων ($\overline{\lambda}$).



Σχήμα 4.2: Αναπαράσταση με finite-state-automaton (FSA) των διαφορετικών PDTS γραμματικών: (α) $G_{\{del,sub\}}$, (β) G_{sub}, (γ) G_{ins}. Το σύμβολο'eps' αντιπροσωπεύει μια ϵ μετάβαση στο FSA. M, N και L είναι ο αριθμός των διαφορετικών posture, transition και χειρομορφής υπομονάδων αντίστοιχα.

3) Viterbi Decoding: Για κάθε D_i εφαρμόζουμε Viterbi Decoding (Dec) χρησιμοποιώντας τα HMM μοντέλα $\overline{\lambda}$ και ένα δίκτυο αναγνώρισης G. Το τελευταίο κατασκευάζεται για κάθε δεδομένο εκπαίδευσης D_i συνδυάζοντας την ακολουθία των PDTS υπομονάδων T_i με μια PDTS γραμματική (βλ. ενότητα 4.3.2). Η παραπάνω διαδικασία έχει ως αποτέλεσμα την εύρεση της πιο πιθανής ακολουθίας PDTS υπομονάδων $\overline{T_i}$ η οποία ταιριάζει στην εκτέλεση του νοήματος με διάνυσμα χαρακτηριστικών D_i. Με αυτό τον τρόπο διορθώνουμε τις PDTS επισημειώσεις με βάση την πραγματική άρθρωση κάθε νοήματος.

4) Σφάλμα μεταξύ T και $\overline{T}$: Μετά τη διόρθωση των PDTS επισημειώσεων $\overline{T}$ υπολογίζουμε την απόσταση μεταξύ του T και του $\overline{T}$. Ευθυγραμμίζουμε κάθε T_i και $\overline{T_i}$ χρησιμοποιώντας έναν αλγόριθμο για την εύρεση του βέλτιστου ταιριάσματος συμβολοσειρών χρησιμοποιώντας δυναμικό προγραμματισμό. Στη συνέχεια, υπολογίζουμε τον αριθμό των υπομονάδων που αντικαταστάθηκαν (S), διαγράφηκαν (D) ή παρεμβλήθηκαν (I). Το σφάλμα υπολογίζεται ως $E = 1-(N-D-S-I)/N$ όπου N είναι ο αριθμός των δεδομένων εκπαίδευσης.

5) Επανάληψη: Τα βήματα 2-4 επαναλαμβάνονται μέχρις ότου το σφάλμα μεταξύ T και $\overline{T}$ είναι μικρότερο ενός προκαθορισμένου κατωφλιού (T_0), το οποίο ορίζεται πειραματικά.

4.3.2 Εκπαίδευση των PDTS υπομονάδων χρησιμοποιώντας τον αλγόριθμο ITA

Για την εκπαίδευση PDTS HMM μοντέλων υπομονάδας χρησιμοποιούμε τον αλγόριθμο ITA με διαφορετική αρχικοποίηση και PDTS γραμματική για τις ροές πληροφορίας της κίνησης-θέσης και χειρομορφής αντίστοιχα.

Algorithm 1 $[\overline{\lambda}, \overline{T}] = ITA(\lambda, T, D)$

$E = 1, T_0 = 0.05, \overline{T} = T$
$\lambda = Init(D, T)$
while $E > T_0$ **do**
  $\overline{\lambda} = BW(\lambda, \overline{T}, D)$
  for $i = 1$ **to** N **do**
    $G = Gram(T_i)$
    $\overline{T_i} = Dec(\overline{\lambda}, G, D_i)$
  end for
  $E = Algn(T, \overline{T})$
  $T = \overline{T}, \lambda = \overline{\lambda}$
end while

PDTS φωνητική γραμματική

Στόχος του αλγορίθμου ITA είναι η εκπαίδευση των PDTS μοντέλων υπομονάδας και η διόρθωση των PDTS επισημειώσεων έτσι ώστε να υπάρχει αντιστοιχία με την πραγματική εκφορά κάθε νοήματος στα δεδομένα εκπαίδευσης. Το τελευταίο επιτυγχάνεται εφαρμόζοντας τον αλγόριθμο Viterbi όπως περιγράφηκε προηγουμένως. Ο αλγόριθμος Viterbi εκμεταλλεύεται μια PDTS γραμματική, περιορίζοντας το πρόβλημα αναγνώρισης σε ακολουθίες των PDTS υπομονάδων οι οποίες ακολουθούν τους κανόνες που ορίζει η γραμματική που χρησιμοποιήθηκε.

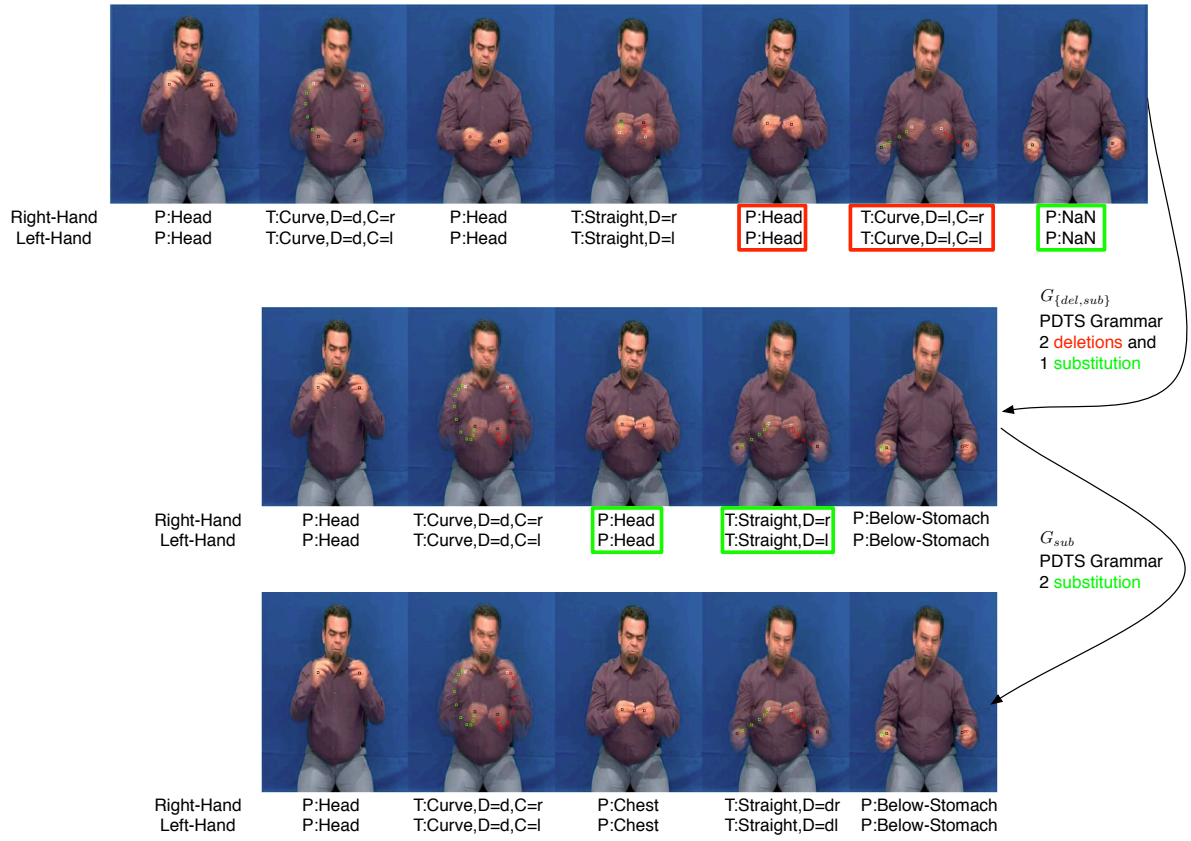
Για τη ροή πληροφορίας της κίνησης-θέσης χρησιμοποιούμε την $G_{\{del,sub\}}$ γραμματική. Αυτή επιτρέπει την διαγραφή οποιασδήποτε PDTS υπομονάδας, η οποία δεν εμφανίζεται στην πραγματική άρθρωση του νοήματος. Επιπλέον υποχρεώνει την αντικατάσταση όλων των implicit υπομονάδων με την explicit υπομονάδα που ταιριάζει καλύτερα σε κάθε περίπτωση. Για τη ροή της χειρομορφής χρησιμοποιούμε την $G_{\{ins\}}$ γραμματική. Αυτή επιτρέπει την εισαγωγή υπομονάδων χειρομορφής κατά τη διάρκεια των κινησεων, όπου οι PDTS ετικέτες για την χειρομορφή είναι άγνωστες (ενότητα. 4.2). Τέλος, και για την ροή της κίνησης-θέσης αλλά και της χειρομορφής χρησιμοποιούμε την $G_{\{sub\}}$ γραμματική. Αυτή βελτιώνει περαιτέρω τις PDTS επισημειώσεις επιτρέποντας την αντικατάσταση κάθε PDTS υπομοναδας.

Για περισσότερη κατανόηση των παραπάνω PDTS γραμματικών ας δούμε ένα παράδειγμα. Έστω ότι έχουμε ένα νόημα με PDTS επισημειώσεις:

$$T_j^{MP} = P_1\ T_1\ P_{N/A}, \quad T_j^{HS} = HS_1^P\ HS_{N/A}^T\ HS_2^P,$$

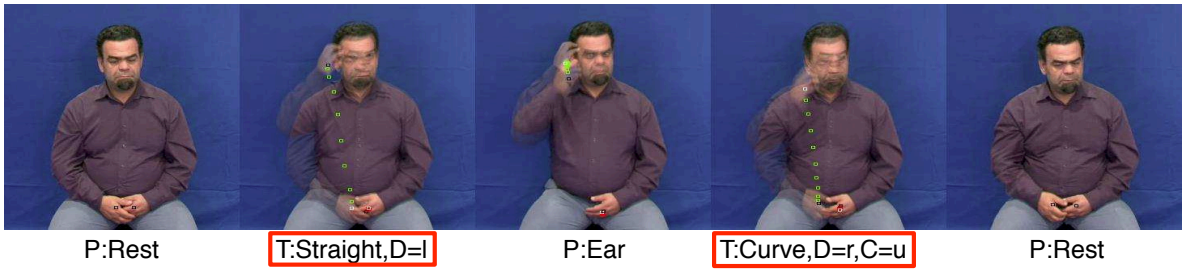
για τις ροές της κίνησης-θέσης και χειρομορφής αντίστοιχα. P_1 και T_1 είναι explicit posture και transition υπομονάδες αντίστοιχα. Επιπλέον $P_{N/A}$ είναι μια impicit posture υπομονάδα: η PDTS ετικέτα δεν είναι διαθέσιμη (N/A). HS_1^P και HS_2^P περιγράφουν την χειρομορφή και την πόζα για δύο posture (P) υπομονάδες. Ενώ $HS_{N/A}^T$ είναι μια implicit υπομονάδα χειρομορφής καθώς το είδος της χειρομορφής είναι άγνωστο κατά τη διάρκεια των κινήσεων. Στα Σχήματα 4.2(α-γ) απεικονίζουμε για το συγκεκριμένο παράδειγμα νοήματος, τις $G_{\{del,sub\}}$, G_{sub} και G_{ins} PDTS γραμματικές χρησιμοποιώντας finite-state-automaton (FSA).

Ένα πραγματικό παράδειγμα υλοποίησης του ITA αλγορίθμου χρησιμοποιώντας την ροή της κίνησης-θέσης απεικονίζεται στο Σχήμα 4.3. Πιο συγκεκριμένα, απεικονίζουμε την ευθυγράμμιση των PDTS μοντέλων υπομονάδας κατά τη διάρκεια εκπαίδευσης. Στην πρώτη γραμμή βλέπουμε την ευθυγράμμιση των PDTS μοντέλων υπομονάδας χωρίς την χρήση του αλγορίθμου ITA. Όπως παρατηρούμε υπάρχουν δύο επιπλέον PDTS υπομονάδες (επισημειωμένες με κόκκινο τετράγωνο). οι οποίες δεν εμφανίζονται στην πραγματική άρθρωση του νοήματος. Επιπλέον υπάρχει μια implicit posture υπομονάδα (επισημειωμένη με πράσινο τετράγωνο). Στη δεύτερη σειρά απεικονίζουμε την ευθυγράμμιση των PDTS μοντέλων υπομονάδας μετά την εφαρμογή του αλγορίθμου ITA

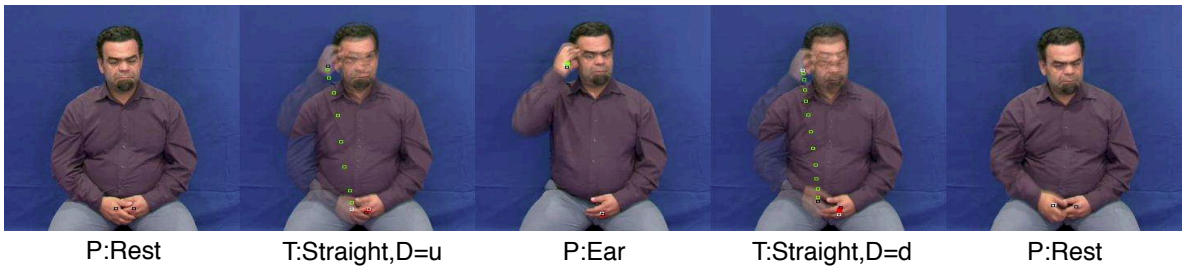


Σχήμα 4.3: Ευθυγράμμιση των PDTS μοντέλων υπομονάδας κατά τη διάρκεια εκπαίδευσης χρησιμοποιώντας την ροή κίνησης-θέσης για το νόημα 'ΑΠΟΤΕΛΕΣΜΑ' της ΕΝΓ. Πρώτη σειρά: ευθυγράμμιση των PDTS μοντέλων υπομονάδας χωρίς τη χρήση του αλγορίθμου ITA. Δεύτερη σειρά: ευθυγράμμιση των PDTS μοντέλων υπομονάδας χρησιμοποιώντας τον αλγόριθμο ITA με $G_{\{del,sub\}}$ PDTS γραμματική. Τρίτη σειρά: ευθυγράμμιση των PDTS μοντέλων υπομονάδας εφαρμόζοντας τον αλγόριθμο ITA δύο συνεχόμενες φορές, με $G_{\{del,sub\}}$ και G_{sub} PDTS γραμματικές. Το πρώτο γράμμα σε κάθε PDTS υπομονάδα χαρακτηρίζει τον PDTS τύπο της υπομονάδας, όπου είναι P για τις posture και T για τις transition υπομονάδες. Για τις posture υπομονάδες το δεύτερο όρισμα χαρακτηρίζει τη θέση του χεριού, π.χ. P:Head είναι μια στάση κοντά στο κεφάλι του νοηματιστή. Η σημειογραφία για τις transition υπομονάδες είναι: 1) Τύπος της κίνησης: καμπύλη (curve) ή ευθεία (straight), 2) η κατεύθυνσή της (D), όπου είναι right (r), left (l), up (u), down (d) και συνδυασμοί τους όπως π.χ. down-right (dr) και 3) μόνο για τις καμπύλες κινήσεις υπάρχει ένα επιπλέον όρισμα το οποίο χαρακτηρίζει την κατεύθυνση της καμπύλης (C). Για περισσότερες λεπτομέρειες σχετικά με την σημειογραφία βλέπε [56].

κάνοντας χρήση της $G_{\{del,sub\}}$ PDTS γραμματικής. Όπως παρατηρούμε οι προαναφερθείσες επιπλέον PDTS υπομονάδες διαγράφηκαν και η implicit posture υπομονάδα αντικαταστάθηκε από μια explicit posture υπομονάδα. Παρόλα αυτά, υπάρχουν ακόμα δύο λάθος PDTS υπομονάδες σε σχέση με την πραγματική άρθρωση (επισημειωμένες με πράσινο τετράγωνο). Εφαρμόζοντας τον αλγόριθμο ITA για δεύτερη συνεχόμενη φορά χρησιμοποιώντας αυτή τη φορά την G_{sub} PDTS γραμματική διορθώνονται οι λάθος PDTS υπομονάδες (τρίτη σειρά). Εφαρμόζοντας τον ITA αλγόριθμο για δεύτερη φορά χρησιμοποιώντας την G_{sub} PDTS γραμματική επιτρέπουμε την αντικατάσταση οποιασδήποτε PDTS υπομονάδας. Έτσι διορθώνονται λανθασμένες PDTS υπομονάδες οι οποίες



α) Αρχικοποίηση με flat-start (FS)



β) Αρχικοποίηση με RAW μοντέλα υπομονάδας

Σχήμα 4.4: Ευθυγράμμιση των PDTS μοντέλων υπομονάδας για το νόημα 'ΘΥΜΑΜΑΙ' της ΕΝΓ μετά την εκπαίδευση εφαρμόζοντας τον αλγόριθμο ITA και χρησιμοποιώντας τη ροή της κίνησης-θέσης. Με κόκκινα τετράγωνα υποδεικνύουμε τη λανθασμένη αντιστοίχιση των PDTS συμβόλων σε σχέση με την πραγματική άρθρωση της αντίστοιχης υπομονάδας. Για περισσότερες λεπτομέρειες σχετικά με την σημειογραφία των PDTS υπομονάδων βλέπε τη λεζάντα στο Σχήμα 4.3.

δεν μπορούσαν να διορθωθούν χρησιμοποιώντας μόνο την $G_{\{del,sub\}}$ PDTS γραμματική.

Αρχικοποίηση των HMM

Η αρχικοποίηση των παραμέτρων των HMM μπορεί να γίνει είτε μέσω της διαδικασίας flat-start είτε χρησιμοποιώντας τις παραμέτρους των RAW υπομονάδων. Στην πρώτη περίπτωση χρησιμοποιούμε τη μέση τιμή και διακύμανση από όλα τα δεδομένα εκπαίδευσης για την αρχικοποίηση όλων των HMM μοντέλων. Στη δεύτερη περίπτωση χρησιμοποιούμε τις παραμέτρους των RAW υπομονάδων που παρουσιάστηκαν στην ενότητα 3.4.2. Οι RAW δυναμικές υπομονάδες κατασκευάζονται εφαρμόζοντας ομοιόμορφη διαμέριση του χώρου χαρακτηριστικών της κατεύθυνσης της κίνησης. Λόγω της ένα προς ένα αντιστοιχίας τους με τις PDTS transition υπομονάδες χρησιμοποιούμε τις παραμέτρους τους για την αρχικοποίηση των HMM μοντέλων των PDTS transition υπομονάδων. Οι RAW στατικές υπομονάδες κατασκευάζονται επίσης εφαρμόζοντας ομοιόμορφη διαμέριση του δισδιάστατου νοηματικού χώρου. Παρότι όμως δεν υπάρχει ένα προς ένα αντιστοιχία με τις PDTS posture υπομονάδες αρχικοποιούμε κάθε PDTS posture υπομονάδα με τις παραμέτρους της πλησιέστερης RAW στατικής υπομονάδας.

Κεφάλαιο 5

Στατιστικά μοντέλα, Σύμμειξη, Προσαρμογή σε Νοηματιστή

5.1 Μοντελοποίηση υπομονάδων με κρυφά Μαρκοβιανά μοντέλα

Για την μοντελοποίηση των υπομονάδων χρησιμοποιώντας HMMs είναι απαραίτητο να θέσουμε κάποια κριτήρια στα διανύσματα χαρακτηριστικών που θα λαμβάνονται υπόψη σε κάθε τύπο υπομονάδας. Οι δύο τύποι υπομονάδας στη δικιά μας περίπτωση είναι αυτές που μοντελοποιούν κινήσεις και αυτές που μοντελοποιούν στάσεις. Οι υπομονάδες που μοντελοποιούν κινήσεις πρέπει να λαμβάνουν υπόψη τους μόνο το διάνυσμα χαρακτηριστικών που αναφέρεται στην κίνηση. Ενώ αυτές που μοντελοποιούν στάσεις να λαμβάνουν υπόψη τους μόνο το διάνυσμα χαρακτηριστικών που σχετίζεται με τη θέση. Ένα χαρακτηριστικό των υπομονάδων αυτών είναι η εναλλαγή τους κατά τη διάρκεια άρθρωσης ενός νοήματος. Η χρησιμοποίηση ενός συμβατικού multistream HMM δεν μπορεί να ικανοποιήσει τις παραπάνω προδιαγραφές, εφόσον είναι απαραίτητο όλα τα HMM μοντέλα υπομονάδας να εξαρτώνται από ένα κοινό διάνυσμα χαρακτηριστικών. Εμπνευσμένοι από το multi-space probability distribution (MSD) [129] πλαίσιο προτείνουμε το multi-stream switching probability distribution (MSSD) πλαίσιο. Αυτό το πλαίσιο επιτρέπει με ένα φορμαλιστικό τρόπο τη χρησιμοποίηση διαφορετικών σετ χαρακτηριστικών διανυσμάτων, ανάλογα με τον τύπο της υπομονάδας που μοντελοποιείται. Επιπλέον το παραπάνω μπορεί να εφαρμοστεί χωρίς να γνωρίζουμε εκ των προτέρων τη χρονική κατάτμηση στους διαφορετικούς τύπους υπομονάδων.

5.1.1 HMMs με multi-stream switching probability distribution (MSSD)

Ας θεωρήσουμε τον χώρο Ω ο οποίος αποτελείται από G streams: $\Omega = \bigcup_{g=1}^{G} \Omega_g$, όπου Ω_g είναι ένας n_g-διάστατος πραγματικός χώρος $\mathbb{R}^{n_g}$, ο οποίος καθορίζεται από το stream index g. Κάθε stream έχει μια σ.π.π. $\mathcal{N}_g(\boldsymbol{x}_g)$ όπου $\boldsymbol{x}_g \in \mathbb{R}^{n_g}$.

Ας ορίσουμε ένα γεγονός το οποίο αντιπροσωπεύεται από το τυχαίο διάνυσμα $\boldsymbol{o}$. Αυτό αποτελείται από μια διακριτή τυχαία μεταβλητή $\boldsymbol{y}$ και μια συνεχής τυχαία μεταβλητή $\boldsymbol{x}$. Η διακριτή μεταβλητή $\boldsymbol{y} \in \{l_1, l_2, \ldots, l_M\}$ και καθορίζει τον τύπο της κλάσης (M διαφορετικοί τύποι). Η συνεχής μεταβλητή $\boldsymbol{x} \in \mathbb{R}^n$ και ισούται με $\boldsymbol{x} \equiv (\boldsymbol{x}_1; \ldots; \boldsymbol{x}_G)$ όπου $n = \sum_{g=1}^{G} n_g$.

$$\boldsymbol{o} = (\boldsymbol{x}, \boldsymbol{y}) \tag{5.1}$$

Η πιθανότητα της παρατήρησης του $\boldsymbol{o}$ ορίζεται ως:

$$b(\boldsymbol{o}) = \prod_{g \in I(\boldsymbol{y})} \mathcal{N}_g(\boldsymbol{x}_g)^{w_g} \tag{5.2}$$

όπου w_g είναι το stream weight για το g stream το οποίο λειτουργεί ως εκθέτης. Η συνάρτηση $I(\boldsymbol{y})$ επιστρέφει το σετ των stream indices τα οποία θα ληφθούν υπόψη για τον συγκεκριμένο τύπο $\boldsymbol{y}$ της κλάσης. Ας μελετήσουμε το επόμενο παράδειγμα, επεκταμένο από το άρθρο [129], το οποίο περιγράφει την multi-stream switching probability distribution (MSSD) σε ένα πραγματικό πρόβλημα.

Ένας άνθρωπος ψαρεύει σε μια λίμνη. Στη λίμνη υπάρχουν δύο τύποι ψαριών και δύο τύποι χελωνών. Όταν αυτός ο άνθρωπος πιάνει είτε ένα ψάρι είτε μια χελώνα, ενδιαφέρεται να αναγνωρίσει τον τύπο του. Τα ψάρια διακρίνονται από το μήκος τους και οι χελώνες από την διάμετρό τους. Όταν ο άνθρωπος πιάσει ένα ψάρι ή μια χελώνα μετράει το μήκος και την διάμετρό του.

Σε αυτή την περίπτωση, ο χώρος Ω αποτελείται από δύο streams. Ω_1 και Ω_2 που είναι δύο μονο-διάστατα stream τα οποία αντιστοιχούν στο μήκος και τη διάμετρο είτε του ψαριού είτε της χελώνας. Περαιτέρω, $\mathcal{N}_g^1(\boldsymbol{x}_g)$ και $\mathcal{N}_g^2(\boldsymbol{x}_g)$ είναι οι σ.π.π. για τους δύο τύπους ψαριών και για το g stream. Επιπλέον, $\mathcal{N}_g^3(\boldsymbol{x}_g)$ και $\mathcal{N}_g^4(\boldsymbol{x}_g)$ είναι οι σ.π.π. για τους δύο τύπους χελωνών και για το g stream. Σε αυτό το παράδειγμα έχουμε δύο τύπων κλάσεων, ψάρια και χελώνες. Οπότε η διακριτή μεταβλητή $\boldsymbol{y} \in \{l_1, l_2\}$, όπου $I(l_1) = \{1\}$ και $I(l_2) = \{2\}$. Επιπλέον οι κλάσεις που θέλουμε να διαχωρίσουμε είναι $\omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, όπου ω_1, ω_2 είναι οι δύο τύποι των ψαριών και ω_3, ω_4 είναι οι δύο τύποι των χελωνών.

Τώρα, ας υποθέσουμε ότι ο άνθρωπος ψαρεύει κατά τη διάρκεια της ημέρας και μπορεί να διαχωρίζει εάν πιάσει ψάρι ή χελώνα. Ας πάρουμε για παράδειγμα ότι πιάνει ψάρι. Η παρατήρηση είναι $\boldsymbol{o} = (l_1, \boldsymbol{x})$ όπου $\boldsymbol{x} \equiv (\boldsymbol{x}_1; \boldsymbol{x}_2)$. Με άλλα λόγια η συνεχής μεταβλητή $\boldsymbol{x}$ είναι ένα δισδιάστατο διάνυσμα χαρακτηριστικών το οποίο αντιπροσωπεύει το μήκος και την διάμετρο. Για να αναγνωριστεί ο τύπος του ψαριού θα πρέπει να μεγιστοποιήσουμε την πιθανότητα $P(\omega_i|\boldsymbol{o})$ σε σχέση με όλες τις κλάσεις ω_i . Από εφαρμογή του κανόνα του Bayes για ισοπίθανες κλάσεις έχουμε:

$$\arg\max_i(P(\omega_i|\boldsymbol{o})) \propto \arg\max_i(P(\boldsymbol{o}|\omega_i))$$

Για λόγους απλοποίησης, κάνουμε την υπόθεση ότι οι $\boldsymbol{x}, \boldsymbol{y}$ είναι conditionally ανεξάρτητες σε σχέση με την κλάση ω_i και άρα

$$P(\boldsymbol{o}|\omega_i) = P(\boldsymbol{x}, \boldsymbol{y}|\omega_i) = P(\boldsymbol{x}|\omega_i)P(\boldsymbol{y}|\omega_i) \tag{5.3}$$

όπου

$$P(\boldsymbol{y}|\omega_i) \quad = \quad \begin{cases} 0 & \text{για } Y(\omega_i) \neq \boldsymbol{y} \\ 1 & \text{αλλιώς} \end{cases} \tag{5.4}$$

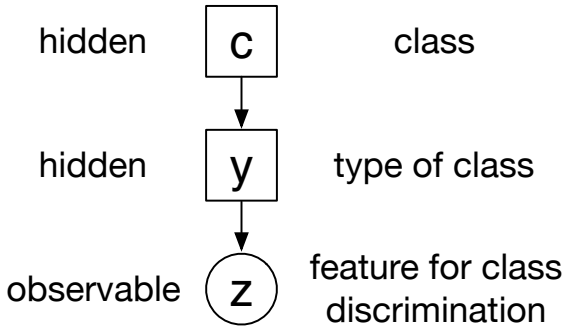
και η συνάρτηση $Y(\omega_i)$ επιστρέφει τον τύπο της κλάσης ω_i. Έτσι από της εξισώσεις (5.3) και (5.4) έχουμε

$$P(\boldsymbol{o}|\omega_i) \quad = \quad \begin{cases} 0 & \text{για } Y(\omega_i) \neq \boldsymbol{y} \\ P(\boldsymbol{x}|\omega_i) & \text{αλλιώς} \end{cases}$$

Επιπλέον, από την εξίσωση (5.5) έχουμε:

$$P(\boldsymbol{x}|\omega_i) = \prod_{g \in I(Y(\omega_i))} \mathcal{N}_g^i(\boldsymbol{x}_g)^{w_g}$$

Την παραπάνω πιθανότητα $P(\boldsymbol{x}|\omega_i)$, μπορούμε να μοντελοποιήσουμε για κάθε κλάση ω_i, χρησιμοποιώντας ένα multi-stream Γκαουσιανό μοντέλο και θέτοντας μηδέν τα stream weights ($w_g = 0$)



Σχήμα 5.1: Γραφική αναπαράσταση του σεναρίου των χαρακτηριστικών μετρήσεων, απεικονίζοντας τις κρυφές και παρατηρήσιμες μεταβλητές εσωκλείοντάς τις με τετράγωνα και κύκλους αντίστοιχα.

όπου $g \notin I(Y(\omega_i))$. Επιπλέον για να μοντελοποιήσουμε την πιθανότητα $P(\boldsymbol{o}|\omega_i)$, μπορούμε να χρησιμοποιήσουμε ένα επιπλέον stream το οποίο θα μοντελοποιεί τη διακριτή μεταβλητή $\boldsymbol{y}$ με διακριτή σ.π.π. όπως στην εξίσωση (5.4).

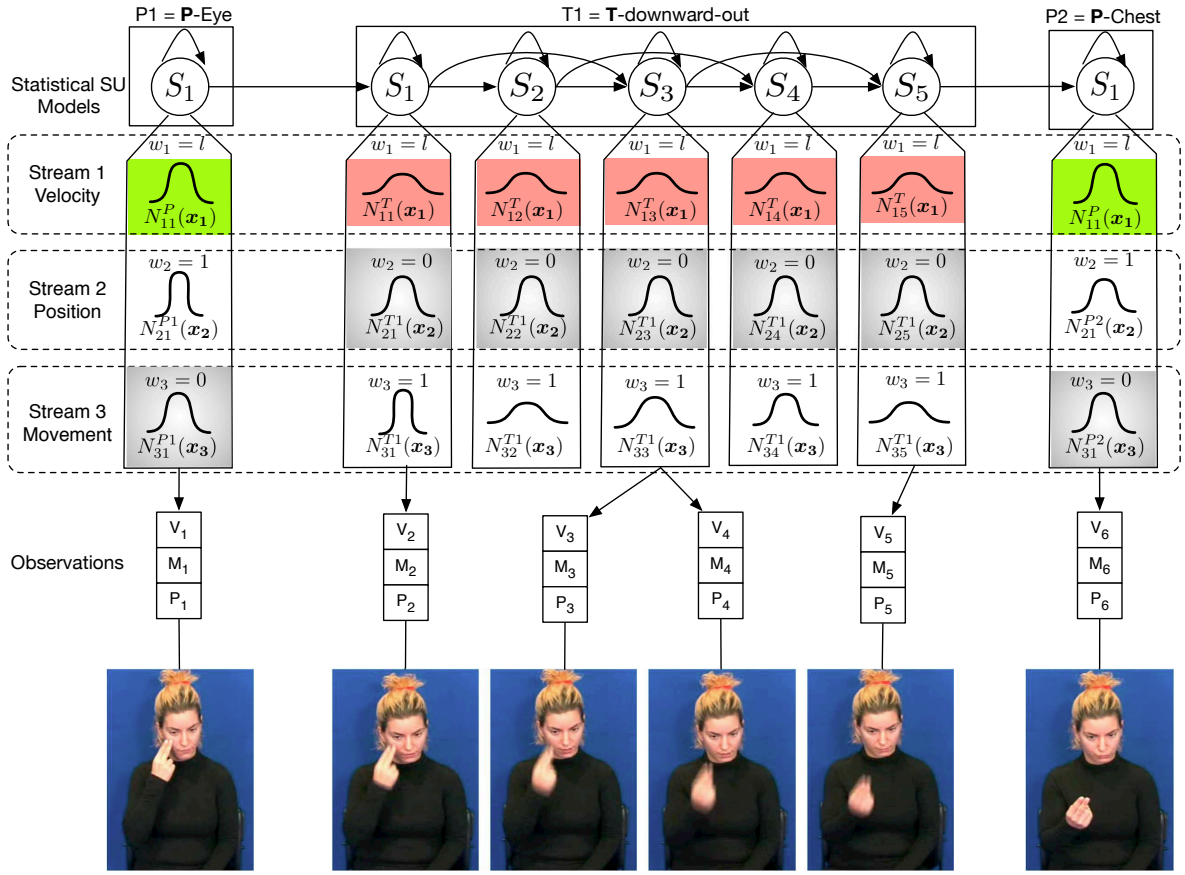
Τώρα, ας υποθέσουμε ότι ο άνθρωπος ψαρεύει κατά τη διάρκεια της νύχτας χωρίς να μπορεί να διακρίνει αν έχει πιάσει ψάρι ή χελώνα. Η μεταβλητή $\boldsymbol{y}$ είναι πλέον άγνωστη, με άλλα λόγια $\boldsymbol{y}$ είναι μια κρυφή μεταβλητή. Παρόλα αυτά, ξέρει ότι τα ψάρια και οι χελώνες έχουν διαφορετικό βάρος. Με σ.π.π. $\mathcal{N}^{\boldsymbol{y}}(\boldsymbol{z})$ για τα ψάρια και τις χελώνες αντίστοιχα, όπου $\boldsymbol{y} \in \{l_1, l_2\}$ και $\boldsymbol{z}$ είναι μια συνεχής τυχαία μεταβλητή η οποία αντιστοιχεί στη μέτρηση του βάρους. Όμως η $\boldsymbol{z}$ είναι παρατηρήσιμη. Μια γραφική αναπαράσταση των μετρήσεων φαίνεται στο Σχήμα 5.1.

Όμοια με την προηγούμενη περίπτωση, $P(\boldsymbol{o}|\omega_i) = P(\boldsymbol{x}, \boldsymbol{z}|\omega_i) = P(\boldsymbol{x}|\omega_i)P(\boldsymbol{z}|\omega_i)$. Κάνοντας integrate out την κρυφή μεταβλητή $\boldsymbol{y}$, η πιθανότητα $P(\boldsymbol{z}|\omega_i)$ γίνεται

$$P(\boldsymbol{z}|\omega_i) = \sum_y (P(\boldsymbol{z}|\boldsymbol{y})P(\boldsymbol{y}|\omega_i)) = P(\boldsymbol{z}|Y(\omega_i)) = \mathcal{N}^{Y(\omega_i)}(\boldsymbol{z})$$

Έτσι, μπορούμε να μοντελοποιήσουμε την πιθανότητα $P(\boldsymbol{o}|\omega_i)$ χρησιμοποιώντας το προηγούμενο multi-stream Γκαουσιανό μοντέλο, αντικαθιστώντας το επιπλέον stream που μοντελοποιούσε τη διακριτή μεταβλητή $\boldsymbol{y}$, με ένα stream το οποίο τώρα θα μοντελοποιεί την συνεχή μεταβλητή $\boldsymbol{z}$ με σ.π.π. $P(\boldsymbol{z}|\omega_i) = \mathcal{N}^{Y(\omega_i)}(\boldsymbol{z})$ για κάθε Γκαουσιανό μοντέλο ω_i.

Συνοψίζοντας, εισαγάγαμε ένα νέο HMM βασιζόμενο στην multi-stream switching probability distribution (MSSD) κατανομή. Αυτό μας επιτρέπει με έναν φορμαλιστικό τρόπο να λαμβάνουμε υπόψη διαφορετικό διάνυσμα χαρακτηριστικών, ανάλογα με τον τύπο των κλάσεων που μοντελοποιούμε κάθε φορά. Η κύρια ιδέα της MSSD είναι αρκετά όμοια με την multi-space probability distribution (MSD) [129]. Παρόλα αυτά, υπάρχουν και αρκετές διαφορές. Η κύρια διαφορά είναι ότι η μεταβλητή $\boldsymbol{y}$ είναι κρυφή. Απεναντίας, μοντελοποιείται μια άλλη παρατηρήσιμη συνεχής μεταβλητή $\boldsymbol{z}$ η οποία μας υποδεικνύει τον τύπο της κλάσης ($\boldsymbol{y}$). Με αυτό τον τρόπο δεν παίρνουμε εκ των προτέρων απόφαση για την τιμή της μεταβλητής $\boldsymbol{y}$. Αντιθέτως ενσωματώνουμε την απόφαση αυτή στο συνολικό πιθανοτικό πλαίσιο και η απόφαση παίρνεται κατά τη διάρκεια της αναγνώρισης. Μια ακόμα διαφορά είναι ότι η υλοποίηση γίνεται χρησιμοποιώντας multi-stream HMMs θέτοντας μηδέν τα stream weights των stream που δεν θέλουμε να λαμβάνουμε υπόψη ανάλογα με τον τύπο της κλάσης που μοντελοποιείται.



Σχήμα 5.2: Διαδοχή στατιστικών υπομονάδων για το νόημα 'ΒΛΕΠΩ' της ΕΝΓ. Η σ.π.π. $N_{ij}^{k}(\boldsymbol{x_i})$ αντιστοιχεί στο stream i, στην κατάσταση j του HMM μοντέλου και στην υπομονάδα k, π.χ. η $N_{31}^{T1}(\boldsymbol{x_3})$ σ.π.π. αντιστοιχεί στην πρώτη κατάσταση, το τρίτο stream (ροή κίνησης) και στην $T1$ PDTS υπομονάδα. Η σ.π.π. της ταχύτητας για τις κινήσεις είναι η $N_{11}^{Y}(\boldsymbol{x_1})$ και για τις στάσεις είναι η $N_{11}^{P}(\boldsymbol{x_1})$. Επιπλέον τις απεικονίζουμε με διαφορετικό χρώμα (κόκκινο και πράσινο αντίστοιχα). Τα κουτιά με την σκίαση αντιστοιχούν σε μηδενικά stream weights. Τα στατιστικά HMM ενώνονται για την δημιουργία ενός δικτύου από HMM όπως περιγράφεται από το PDTS λεξικό. Οι παρατηρήσεις των HMM ανά stream είναι V_i, M_i και P_i όπου i είναι ο αριθμός πλαισίου του βίντεο, και αντιστοιχούν στην ακολουθία εικόνων του συγκεκριμένου βίντεο για το νόημα 'ΒΛΕΠΩ'.

5.1.2 Υπομονάδες με MSSD-HMMs

Σε αυτή την ενότητα περιγράφουμε τη χρήση των MSSD-HMMs για τη μοντελοποίηση των PDTS υπομονάδων. Ακριβώς η ίδια χρήση εφαρμόζεται και για τη μοντελοποίηση των 2-S-U και RAW δεδομενοκεντρικών υπομονάδων. Η διαφορά είναι ότι οι transition και posture υπομονάδες αντικαθιστώνται από τις δυναμικές και στατικές υπομονάδες των μεθόδων 2-S-U και RAW.

Οι PDTS υπομονάδες χωρίζονται σε δύο τύπους: transition και posture υπομονάδες. Οι transition και posture υπομονάδες διαχωρίζονται μεταξύ τους από το διάνυσμα της ταχύτητας. Οι transition υπομονάδες μοντελοποιούν τμήματα με υψηλή ταχύτητα, ενώ οι posture υπομονάδες με χαμηλή ταχύτητα. Επιπλέον οι transition υπομονάδες πρέπει να εξαρτώνται μόνο από το διάνυσμα χαρακτηριστικών της κίνησης. Ενώ οι posture υπομονάδες μόνο από το διάνυσμα χαρακτηριστικών της θέσης. Ένα χαρακτηριστικό των transition και posture υπομονάδων είναι

ότι εναλλάσσονται χρονικά κατά την άρθρωση ενός νοήματος. Ενώ η χρονική κατάτμηση στις παραπάνω υπομονάδες δεν είναι γνωστή εκ των προτέρων. Η μόνη ένδειξη που έχουμε είναι η μέτρηση του διανύσματος της ταχύτητας.

Έτσι η χρήση των MSSD-HMMs είναι απαραίτητη εφόσον μας επιτρέπουν την χρησιμοποίηση διαφορετικών σετ από διανύσματα χαρακτηριστικών κατά τη διάρκεια της άρθρωσης, ανάλογα με τον τύπο των PDTS υπομονάδων. Για τη μοντελοποίηση κάθε PDTS υπομονάδας, χρησιμοποιούμε ένα MSSD-HMM με τρία streams. Αυτά αντιστοιχούν στις ροές πληροφορίας της ταχύτητας, της κίνησης και της θέσης του κυρίαρχου χεριού.

Το Σχήμα 5.2 απεικονίζει ένα παράδειγμα για μια εκτέλεση του νοήματος 'ΒΛΕΠΩ' της ΕΝΓ που συνοψίζει όλα τα παραπάνω.

- Το πρώτο stream που αντιστοιχεί στην ταχύτητα ($\boldsymbol{x_1}$), διαχωρίζει τις transition από τις posture PDTS υπομονάδες. Η σ.π.π. $N_{1j}^{y}(\boldsymbol{x_1})$ μοντελοποιεί την πιθανότητα $P(\boldsymbol{x_1}|y)$. Η μεταβλητή $y \in \{T, P\}$: transitions (T) και postures (P), και j είναι η κατάσταση του HMM.

- Το δεύτερο stream που αντιστοιχεί στη ροή πληροφορίας της θέσης ($\boldsymbol{x_2}$), πρέπει να λαμβάνεται υπόψη μόνο στις posture υπομονάδες. Αυτό το stream διαχωρίζει τις διαφορετικές posture υπομονάδες. Παραδείγματος χάριν, η σ.π.π. $N_{21}^{P_k}(\boldsymbol{x_2})$ μοντελοποιεί την πιθανότητα $P(\boldsymbol{x_2}|P_k)$ όπου P_k είναι μια posture υπομονάδα.

- Το τρίτο stream που αντιστοιχεί στη ροή πληροφορίας της κίνησης ($\boldsymbol{x_3}$), πρέπει να λαμβάνεται υπόψη μόνο στις transition υπομονάδες. Αυτό το stream διαχωρίζει τις διαφορετικές transition υπομονάδες. Παραδείγματος χάριν, η σ.π.π. $N_{3j}^{T_k}(\boldsymbol{x_3})$ μοντελοποιεί την πιθανότητα $P(\boldsymbol{x_3}|T_k)$ όπου T_k είναι μια posture υπομονάδα και j είναι η κατάσταση του HMM.

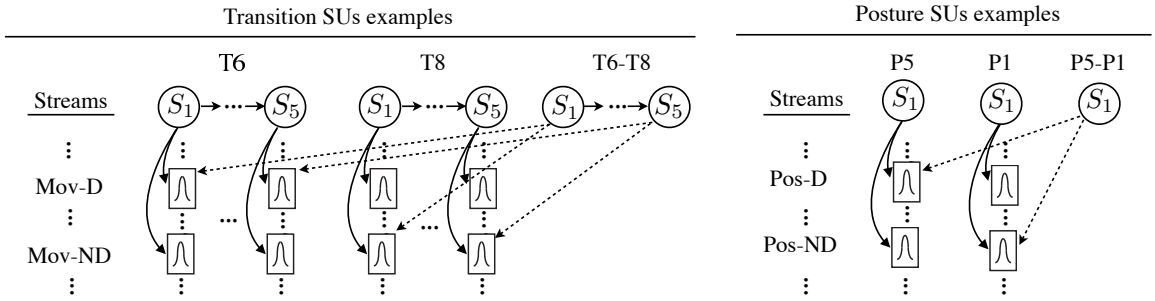
Αυτά τα HMMs ενώνονται για τη δημιουργία ενός δικτύου όπως περιγράφεται από το PDTS λεξικό. Τα κουτιά με τη σκίαση αντιστοιχούν σε μηδενικά stream weights. Επιπλέον στο stream της ταχύτητας χρησιμοποιούμε ένα διαφορετικό stream weight $w_1 = l$. Η τιμή του w_1 είναι χαρακτηριστική της σημαντικότητας του stream της ταχύτητας, σε σχέση με τα άλλα streams για τη διάκριση μεταξύ των transition και posture υπομονάδων και ορίζεται πειραματικά. Το απεικονιζόμενο δίκτυο για το νόημα 'ΒΛΕΠΩ' αποτελείται από μια posture υπομονάδα (P-Eye), μια transition (T-do) και τέλος μια posture υπομονάδα (P-chest).

5.2 Σύμμειξη πολλαπλών ροών πληροφορίας

Μια από τις μεγάλες διαφορές μεταξύ της νοηματικής και της προφορικής γλώσσας είναι ότι αρθρώνεται χρησιμοποιώντας πολλαπλές παράλληλες ροές πληροφορίας. Η σύμμειξη των παράλληλων ροών αυτών στα πλαίσια της αυτόματης αναγνώρισης της ΝΓ, αποτελεί ένα αρκετά δύσκολο και ανεξερεύνητο ως επί το πλείστον πρόβλημα [1]. Από την πλευρά της γλωσσολογία γίνεται έρευνα που σχετίζεται με την αλληλεπίδραση, τον συγχρονισμό και την σημαντικότητα των πολλαπλών παράλληλων ροών πληροφορίας [118, 78]. Σε αυτή την ενότητα, βασιζόμενοι στις υπομονάδες που έχουν χτιστεί για κάθε ροή πληροφορίας ξεχωριστά [9, 4, 7] όπως παρουσιάστηκαν στα προηγούμενα κεφάλαια, θα ασχοληθούμε με την σύμμειξή τους με στόχο την τελική αναγνώριση σε επίπεδο νοήματος.

5.2.1 Σύμμειξη κυρίαρχου και δευτερεύοντος χεριού

Σε αυτή την ενότητα περιγράφουμε την ενσωμάτωση του δευτερεύοντος χεριού χρησιμοποιώντας τις PDTS υπομονάδες. Ακριβώς με τον ίδιο τρόπο κάνουμε την ενσωμάτωση του δευτερεύοντος



Σχήμα 5.3: Tying παράδειγμα κυρίαρχου και δευτερεύοντος χεριού για transition και posture υπομονάδες.

χεριού και για τις 2-S-U και RAW δεδομενοκεντρικές υπομονάδες. Η διαφορά είναι ότι οι transition και posture υπομονάδες αντικαθιστώνται από τις δυναμικές και στατικές υπομονάδες των 2-S-U και RAW μεθόδων.

Η ενσωμάτωση του δευτερεύοντος χεριού ταιριάζει στο πλαίσιο του MSSD-HMM σχήματος που παρουσιάστηκε στην προηγούμενη ενότητα. Πιο συγκεκριμένα, προσθέτουμε τρία επιπλέον stream τα οποία αντιστοιχούν στις ροές πληροφορίας της ταχύτητας, της κίνησης και της θέσης του δευτερεύων χεριού του νοηματιστή. Με αυτό τον τρόπο, όπως περιγράφεται στην συνέχεια κάθε HMM μοντελοποιεί και τα δύο χέρια.

Όπως περιγράφηκε στην ενότητα 5.1.2, χρησιμοποιούμε δύο σ.π.π. για τη μοντελοποίηση της ταχύτητας. Η $N_{11}^{T}(\boldsymbol{x_1})$ χρησιμοποιείται στις transition υπομονάδες και η $N_{11}^{P}(\boldsymbol{x_1})$ στις posture υπομονάδες. Αυτές οι σ.π.π. χρησιμοποιούνται στα νέα μοντέλα και των δύο χεριών ως εξής:

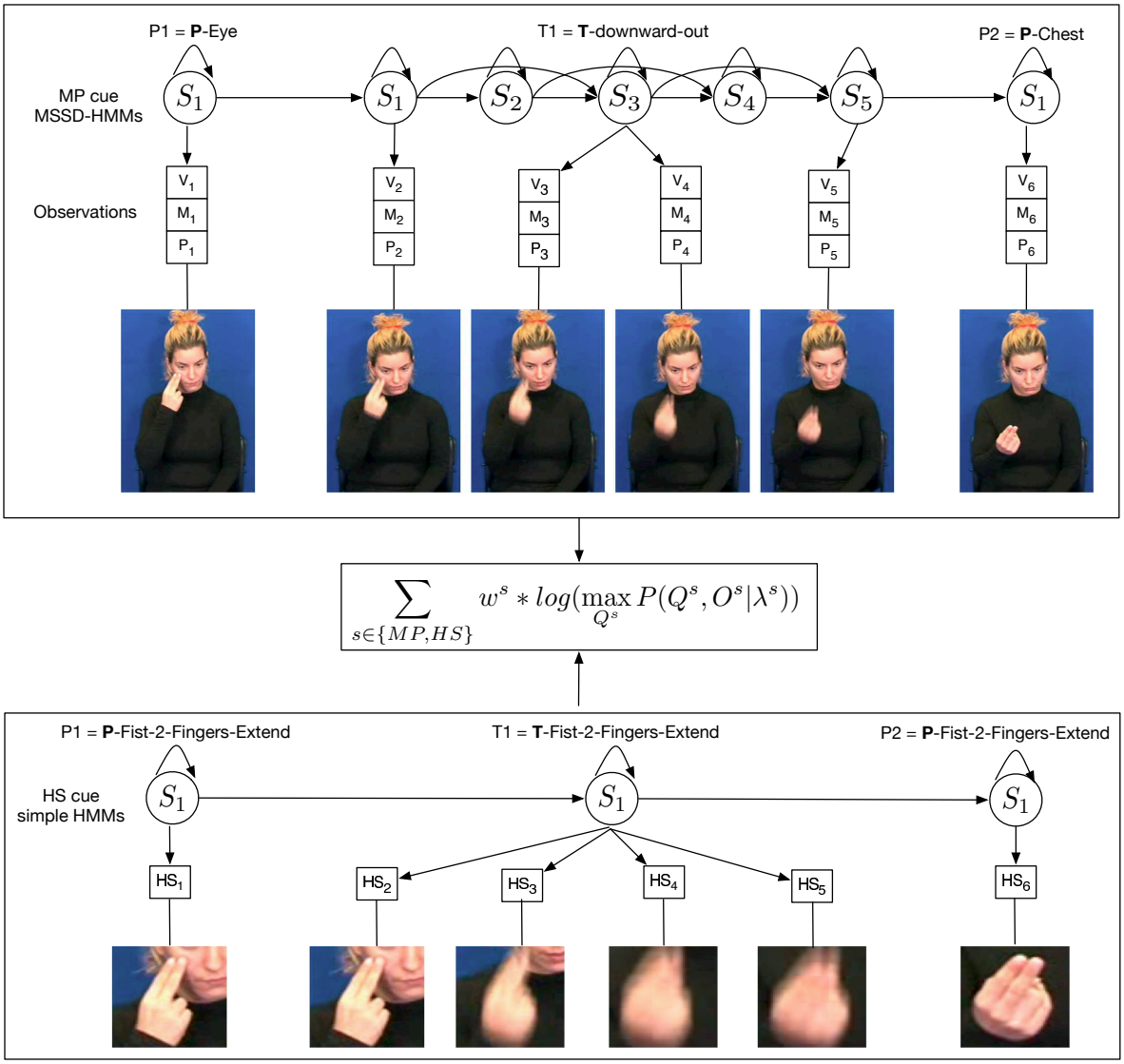
- Για όλα τα *posture* μοντέλα υπομονάδας, χρησιμοποιούμε την $N_{11}^{P}(\boldsymbol{x_1})$ σ.π.π., για τις κατανομές του stream της ταχύτητας και για τα δύο χέρια. Επιπλέον θέτουμε μηδέν τα stream weights του stream της κίνησης και ένα τα stream weights του stream της θέσης και για τα δύο χέρια.

- Για όλα τα *transition* μοντέλα υπομονάδας, που μοντελοποιούν κινήσεις μόνο από το κυρίαρχο χέρι, χρησιμοποιούμε την $N_{11}^{T}(\boldsymbol{x_1})$ σ.π.π. για τις κατανομές του stream της ταχύτητας για το κυρίαρχο χέρι. Αντιθέτως, για τις κατανομές του stream της ταχύτητας για το δευτερεύον χέρι χρησιμοποιούμε την $N_{11}^{P}(\boldsymbol{x_1})$ σ.π.π.. Επιπλέον θέτουμε μηδέν τα stream weights του stream της θέσης και για τα δύο χέρια και του stream της θέσης για το δευτερεύον χέρι. Ενώ θέτουμε ένα τα stream weights του stream της κίνησης για το κυρίαρχο χέρι.

- Για όλα τα *transition* μοντέλα υπομονάδας, που μοντελοποιούν κινήσεις και από τα δύο χέρια, χρησιμοποιούμε την $N_{11}^{T}(\boldsymbol{x_1})$ σ.π.π., για τις κατανομές του stream της ταχύτητας και για τα δύο χέρια. Επιπλέον θέτουμε μηδέν τα stream weights του stream της θέσης και ένα τα stream weights του stream της κίνησης και για τα δύο χέρια.

Τέλος κάνουμε tie τις κατανομές είτε του stream της κίνησης είτε της θέσης, του κυρίαρχου και του δευτερεύοντος χεριού, εάν μοντελοποιούν την ίδια υπομονάδα. Όταν κάνουμε "tie" κάποιες κατανομές, σημαίνει ότι χρησιμοποιούμε τις ίδιες στατιστικές παραμέτρους. Κάθε φορά που μια από αυτές επανεκτιμάται αλλάζουν αντίστοιχα όλα τα μοντέλα που είναι tied σύμφωνα με την νέα εκτίμηση.

Στο Σχήμα 5.3 απεικονίζουμε ένα παράδειγμα του "tying" πλαισίου που εφαρμόζουμε για τρεις transition υπομονάδες (T6, T8, T6-T8) και τρεις posture υπομονάδες (P5, P1 and P5-P1). Mov-D



Σχήμα 5.4: Παράδειγμα σύμμειξης ροών πληροφορίας κίνησης-θέσης και χειρομορφής, με την χρήση παράλληλων HMMs για το νόημα 'ΒΛΕΠΩ' στην ΕΝΓ.

και Mov-ND είναι τα stream της κίνησης για το κυρίαρχο και δευτερεύον χέρι αντίστοιχα. Pos-D και Pos-ND είναι τα stream της θέσης για το κυρίαρχο και δευτερεύον χέρι αντίστοιχα. Όπως παρατηρούμε η T6-T8 υπομονάδα μοιράζεται τις ίδιες κατανομές του stream της κίνησης, με τις υπομονάδες D6 και D8 για το κυρίαρχο και δευτερεύον χέρι αντίστοιχα. Όμοια, η υπομονάδα S5-S1 μοιράζεται τις ίδιες κατανομές του stream της θέσης, με τις υπομονάδες S5 και S1 για το κυρίαρχο και δευτερεύον χέρι αντίστοιχα.

5.2.2 Σύμμειξη ροών πληροφορίας: κίνησης-θέσης και χειρομορφής

Μέχρι τώρα κατασκευάζαμε υπομονάδες για τις ροές πληροφορίας της κίνησης-θέσης και χειρομορφής ανεξάρτητα. Για τη ροή της κίνησης-θέσης, κάναμε κατάτμηση των νοημάτων σε τμήματα τα οποία περιείχαν κίνηση και σε τμήματα που δεν περιείχαν. Στη συνέχεια, είτε χρησιμοποιώντας αλγορίθμους συσταδοποίησης είτε επισημειώσεις σε φωνητικό επίπεδο κατασκευάζαμε

Σχήμα 5.5: Νόημα 'ΗΣΥΧΙΑ' εκτελεσμένο από δύο νοηματιστές, μαζί με την ακολουθία των Δ/Σ δεδομενοκεντρικών υπομονάδων στην οποία αντιστοιχεί. Η διαφορά μεταξύ των δύο προφορών είναι η άρθρωση από τον 'Κώστα' μιας επιπλέον κίνησης, με άλλα λόγια η υπομονάδα D29 αντικαθίσταται από την ακολουθία υπομονάδων D21 S1 D16 S4 D21 (βλ. Πίνακα 5.1).

δεδομενοκεντρικές ή φωνητικές υπομονάδες αντίστοιχα. Αυτό είχε ως αποτέλεσμα τη δημιουργία ενός λεξικού σε επίπεδο υπομονάδας. Με άλλα λόγια για κάθε νόημα στα δεδομένα εκπαίδευσης, αντιστοιχούσε σε μια ακολουθία από υπομονάδες. Για την ροή της χειρομορφής αντίστοιχα, κατασκευάζαμε υπομονάδες χειρομορφής και το αντίστοιχο λεξικό που απέρρεε από αυτή την διαδικασία.

Σε αυτή την ενότητα θα εξετάσουμε τη σύμμειξη αυτών των δυο ροών πληροφορίας. Πιο συγκεκριμένα, προτείνουμε έναν εκ των υστέρων τρόπο σύμμειξης. Αυτός βασίζεται στα PaHMMs υποθέτοντας ότι οι ροές πληροφορίας της κίνησης-θέσης και χειρομορφής είναι τελείως ανεξάρτητες.

Σύμμειξη με παράλληλα HMMs

Τα PaHMMs είναι μια επέκταση των HMMs η οποία προτάθηκε από τον Vogler [134]. Έχουν εφαρμοστεί στην αναγνώριση της ΝΓ κάνοντας την υπόθεση ότι οι ροές πληροφοριών που συμμετέχουν είναι ανεξάρτητες η μια από την άλλη. Αυτό έχει ως αποτέλεσμα την εκπαίδευση και αξιολόγηση κάθε ροής πληροφορίας ανεξάρτητα. Ενώ η σύμμειξη γίνεται εκ των υστέρων στο επίπεδο των πιθανοτήτων, εφαρμόζοντας διαφορετικά βάρη ανάλογα με τη σημαντικότητα κάθε ροής.

Στο Σχήμα 5.4 απεικονίζουμε μια διαγραμματική απεικόνιση της σύμμειξης των ροών πληροφορίας κίνησης-θέσης και χειρομορφής, με τη χρήση PaHMMs για το νόημα 'ΒΛΕΠΩ' στην ΕΝΓ. Όπως παρατηρούμε οι ροές της κίνησης-θέσης και χειρομορφής είναι τελείως ανεξάρτητες. Σε κάθε μια από αυτές εφαρμόζεται ο αλγόριθμος Viterbi για την εύρεση της ακολουθίας των καταστάσεων του HMM (Q^s) που μεγιστοποιεί την πιθανότητα να έχουν προκύψει οι συγκεκριμένες παρατηρήσεις (O^s) από τα αντίστοιχα HMMs (λ^s). Το s αντιστοιχεί στις δύο ροές πληροφορίας, είτε κίνησης-θέσης (MP) είτε χειρομορφής (HS). Η σύμμειξη γίνεται αθροίζοντας τα log-likelihoods κάθε ροής κάνοντας χρήση ενός βάρους (w^s) ανάλογα με τη σημαντικότητα της κάθε μιας.

5.3 Προσαρμογή σε νέο νοηματιστή

Η ποικιλία στον τρόπο άρθρωσης ενός νοήματος εξαρτάται σε ένα βαθμό από τον νοηματιστή. Είναι σύνηθες, διαφορετικοί νοηματιστές να εκτελούν το ίδιο νόημα αρκετά διαφορετικά. Αυτή η ποικιλία παρατηρείται κυρίως:

- σε νοήματα τα οποία αποτελούνται από πολλαπλές επαναλήψεις μιας κίνησης, όπου ο αριθμός των επαναλήψεων μπορεί να διαφέρει,
- σε σύνθετα νοήματα, αυτά δηλαδή που αποτελούνται από την σύνθεση δύο η περισσοτέρων νοημάτων,
- σε νοήματα τα οποία αρθρώνονται με πολύπλοκες κινήσεις.

Στο Σχήμα 5.5 απεικονίζουμε το νόημα 'ΗΣΥΧΙΑ' στην ΕΝΓ όπως αρθρώνεται από δύο νοηματιστές ('Κώστας' και 'Όλγα') στη βάση δεδομένων GSL-Lem. Όπως παρατηρούμε, ο κάθε νοηματιστής εκτελεί το ίδιο νόημα με διαφορετικό τρόπο. Ο 'Κώστας' εκτελεί μια επιπλέον κίνηση σε σχέση με την 'Όλγα'. Η κίνηση αυτή αποτελεί ουσιαστικά το δεύτερο συνθετικό του νοήματος, το οποίο δεν είναι απαραίτητο να εκτελεστεί. Παρόλα αυτά το νόημα εκλαμβάνεται ως το ίδιο.

Από την πλευρά της φωνητικής μοντελοποίησης με στόχο την αυτόματη αναγνώριση ΝΓ, πρέπει να ληφθεί υπόψη αυτή η ποικιλία της άρθρωσης ενός νοήματος. Για την αντιμετώπιση του παραπάνω, τα μοντέλα υπομονάδας πρέπει να έχουν τη δυνατότητα προσαρμογής σε διαφορετικούς νοηματιστές. Επιπλέον το λεξικό να επιτρέπει ποικιλία στην άρθρωση ενός νοήματος μέσω της εισαγωγής πολλαπλών προφορών του ίδιου νοήματος.

Για αυτό χρησιμοποιούμε ένα σύστημα προσαρμογής σε νέο νοηματιστή που ταιριάζει στο ήδη υπάρχον πλαίσιο. Χρησιμοποιώντας ένα μικρό σύνολο δεδομένων από ένα νέο νοηματιστή προσαρμόζουμε τα μοντέλα υπομονάδας σε αυτόν. Επίσης αυξάνουμε την ποικιλία άρθρωσης που υπάρχει στο λεξικό, εισάγοντας νέες προφορές των νοημάτων.

Πιο συγκεκριμένα, το σύστημα προσαρμογής αυτό αποτελείται από τα παρακάτω στάδια:

- Επιλογή ενός συνόλου δεδομένων από τον νέο νοηματιστή (σύνολο δεδομένων προσαρμογής),
- Προσαρμογή των μοντέλων υπομονάδας χρησημοποιώντας Maximum Likelihood Linear Regression (MLLR) [51],
- Παραγωγή νέων προφορών για κάθε εκτέλεση νοήματος που υπάρχει στο σύνολο δεδομένων προσαρμογής,
- Εισαγωγή των νέων προφορών ως νέες εγγραφές στο λεξικό.

5.3.1 Προσαρμογή μοντέλων με χρήση MLLR

Ο αλγόριθμος MLLR υπολογίζει ένα σύνολο από γραμμικούς μετασχηματισμούς, έτσι ώστε να μειώσει την αναντιστοιχία μεταξύ των αρχικών μοντέλων και του συνόλου δεδομένων προσαρμογής από τον νέο νοηματιστή. Πιο συγκεκριμένα, ο αλγόριθμος MLLR είναι μια τεχνική προσαρμογής των μοντέλων, η οποία εκτιμά ένα σύνολο από γραμμικούς μετασχηματισμούς για τις μέσες τιμές των Γκαουσιανών μοντέλων. Το αποτέλεσμα αυτών των μετασχηματισμών είναι η μετατόπιση των μέσων τιμών των Γκαουσιανών των αρχικών HMM μοντέλων έτσι ώστε κάθε κατάσταση των προσαρμοσμένων HMM μοντέλων, να είναι πιο πιθανό να παράγει το σύνολο δεδομένων προσαρμογής από τον νέο νοηματιστή. Ο πίνακας μετασχηματισμών που χρησιμοποιείται για τη νέα εκτίμηση των προσαρμοσμένων μέσων τιμών των μοντέλων δίνεται από τον τύπο:

$$\widehat{\mu} = W\overline{\mu}, \tag{5.5}$$

Πίνακας 5.1: Νοήματα στην ΕΝΓ: 'ΗΣΥΧΙΑ', 'ΥΠΟΔΟΧΗ' και 'ΚΑΠΟΤΕ'. Αντιστοίχιση κάθε νοήματος με μια ακολουθία υπομονάδων για κάθε νοηματιστή. Ο πρώτος νοηματιστής είναι αυτός που χρησιμοποιήθηκε κατά την εκπαίδευση και ο δεύτερος κατά την προσαρμογή. Επιπλέον υποδεικνύουμε τις διαφορές μεταξύ των δύο προφορών (Map.). Στην τέταρτη στήλη έχουμε μια περιγραφή για αυτές τις διαφορές. Τέλος στα Σχήματα 3.1ζ΄ 3.1η΄ 5.5 απεικονίζονται οι εκτελέσεις των νοημάτων 'ΗΣΥΧΙΑ' και 'ΥΠΟΔΟΧΗ' και από τους δύο νοηματιστές.

Νοηματιστής	Νόημα	Προφορά Νοήματος	Περιγραφή Διαφοράς
Κώστας	ΗΣΥΧΙΑ	S5 D14 S1 D21 S1 D16 S4 D21 S5	επιπλέον κίνηση
Όλγα	ΗΣΥΧΙΑ	S5 D14 S1 D29 S5	σύνθετο νόημα
	Map:	{D21 S1 D16 S4 D21} ⟶ {D29}	
Κώστας	ΥΠΟΔΟΧΗ	S5-S3 D19 D26-D26 S3-S4 D21-D27 D27-D27 D29-D29 S5-S5 D16-D16 S5-S3	διαφορετική προφορά
Όλγα	ΥΠΟΔΟΧΗ	S5-S3 D20-D20 S4-S2 D27-D27 D22-D22 S5-S5 D16-D16 S5-S3	κίνησης
	Map:	{D19 D26-D26 S3-S4 D21-D27} ⟶ {D20-D20 S4-S2}, {D29-D29} ⟶ {D22-D22}	
Κώστας	ΚΑΠΟΤΕ	S5-S3 D14 D8 D20 S3-S3 D28 S3-S3 D28 S3-S3 D29 S5-S3	διαφορετικός αριθμός
Όλγα	ΚΑΠΟΤΕ	S5-S3 D14 D8 D20 D2 S4-S3 D2 S3-S3 D22 D29 S5-S3	επαναλήψεων &
	Map:	⟶ {D2 S4-S3 D2}, {D28 S3-S3 D28 S3-S3}⟶ {D22}	διαφορετική κίνηση

όπου W είναι ο $n \times (n+1)$ πίνακας μετασχηματισμών (n είναι η διάσταση των δεδομένων), $\overline{\mu}^\top = [1, \mu^\top]$ και $\widehat{\mu}$ είναι οι μέσες τιμές των προσαρμοσμένων Γκαουσιανών HMM μοντέλων.

Η εύρεση του πίνακα μετασχηματισμών γίνεται λύνοντας ένα πρόβλημα μεγιστοποίησης κάνοντας χρήση του αλγορίθμου Expectation-Maximisation (EM). Για πιο αποδοτική προσαρμογή των μοντέλων ομαδοποιούμε τις Γκαουσιανές κάνοντας χρήση του αλγορίθμου regression class tree και εξάγουμε έναν μετασχηματισμό για κάθε συστάδα από Γκαουσιανές. Με αυτό τον τρόπο στις Γκαουσιανές οι οποίες ανήκουν στην ίδια συστάδα (δηλαδή αρκετά κοντά στον χώρο των χαρακτηριστικών) εφαρμόζεται ο ίδιος μετασχηματισμός. Έτσι, όσο περισσότερα δεδομένα προσαρμογής έχουμε στη διάθεσή μας, τόσο είναι εφικτότερη η καλύτερη προσαρμογή των μοντέλων, αυξάνοντας τον αριθμό των συστάδων και των αντίστοιχων μετασχηματισμών. Επιπλέον, με αυτή την προσέγγιση μας δίνεται η δυνατότητα προσαρμογής Γκαουσιανών για τις οποίες δεν έχουμε καθόλου παρατηρήσεις στα δεδομένα προσαρμογής.

5.3.2 Αντιμετώπιση μη ιδωμένων προφορών από νέο νοηματιστή

Κατά την διάρκεια της εκπαίδευσης κατασκευάζουμε τις υπομονάδες χρησιμοποιώντας τα δεδομένα μόνο από τον νοηματιστή που έχουμε προς εκπαίδευση. Επιπλέον, κατασκευάζουμε και ένα λεξικό το οποίο εμπεριέχει όλες τις πιθανές προφορές για κάθε νόημα αλλά και πάλι μόνο από τον νοηματιστή προς εκπαίδευση.

Στόχος μας είναι να μπορούμε να αναγνωρίσουμε διαφορετικές προφορές των ίδιων νοημάτων από ένα νέο νοηματιστή. Για να το επιτύχουμε αυτό θα πρέπει να εισάγουμε στο λεξικό μας τις νέες προφορές από τον νέο νοηματιστή ως νέες εγγραφές. Για την κατασκευή των νέων αυτών προφορών χρησιμοποιούμε το σύνολο των δεδομένων προσαρμογής και τα προσαρμοσμένα HMM μοντέλα υπομονάδας έχοντας εφαρμόσει MLLR. Πιο συγκεκριμένα, για κάθε δεδομένο προσαρμογής από τον νέο νοηματιστή, εφαρμόζουμε τον αλγόριθμο Viterbi και βρίσκουμε την πιο πιθανή ακολουθία υπομονάδων δεδομένης της ακολουθίας των παρατηρήσεων. Αυτές οι νέες ακολουθίες υπομονάδων για κάθε νόημα ταιριάζουν στον τρόπο εκτέλεσης κάθε νοήματος από τον νέο νοηματιστή. Εάν συγκρίνουμε τις νέες προφορές με αυτές που είχαν κατασκευαστεί για τον νοηματιστή που χρησιμοποιήθηκε στην εκπαίδευση, παρατηρούμε ότι διαφέρουν.

Στον Πίνακα 5.1 απεικονίζουμε τις προφορές τριών νοημάτων ('ΗΣΥΧΙΑ', 'ΥΠΟΔΟΧΗ' και 'ΚΑΠΟΤΕ') της ΕΝΓ από δύο νοηματιστές ('Κώστας' και 'Όλγα') όπως εμφανίζονται στη βάση δεδομένων GSL-Lem. Ο νοηματιστής 'Κώστας' χρησιμοποιήθηκε κατά την εκπαίδευση και η 'Όλγα' κατά την

διαδικασία προσαρμογής. Συγκρίνοντας τις ακολουθίες υπομονάδων (προφορές) των δύο νοηματιστών παρατηρούμε ότι αυτές διαφέρουν. Έχουμε επισημειώσει τις διαφορές αυτές, μετά από εφαρμογή ενός αλγορίθμου για ευθυγράμμιση συμβολοσειρών (pairwise sequence alignment) [86]. Οι διαφορές αυτές μπορούν να συνοψίζονται ως εισαγωγές, διαγραφές ή αντικαταστάσεις υπομονάδων ή και ακολουθιών από υπομονάδες και οφείλονται σε διάφορους παράγοντες. Παραδείγματος χάριν για το νόημα 'ΗΣΥΧΙΑ' έχουμε την παρακάτω αντικατάσταση {D21 S1 D16 S4 D21}→{D29}, η οποία οφείλεται στην άρθρωση μιας παραπάνω κίνησης από τον νοηματιστή 'Κώστα' (Σχήμα 5.5). Για το νόημα 'ΚΑΠΟΤΕ' η διαφορά στην άρθρωση οφείλεται σε διαφορετικό αριθμό επαναλήψεων μιας επαναλαμβανόμενης κίνησης. Τέλος για το νόημα 'ΥΠΟΔΟΧΗ' η διαφορετική προφορά έγκειται στην εκτέλεση της ίδιας κίνησης με διαφορετικό τρόπο (Σχήματα 3.1ζ' 3.1η').

Κεφάλαιο 6

Αναγνώριση Χειρονομιών από Πολυτροπικά Δεδομένα

Η ανθρώπινη επικοινωνία και αλληλεπίδραση εκμεταλλεύεται πολλαπλές αισθητηριακές εισροές με ένα εκπληκτικό τρόπο. Παρότι λαμβάνουμε μια αρκετά μεγάλη ροή πολυτροπικών σημάτων, ιδιαίτερα ακουστικά και οπτικά, η ικανότητα που έχουμε για τη σύμμειξη αυτής της πολυτροπικής πληροφορίας μας δίνει την δυνατότητα να επικοινωνούμε αποτελεσματικά. Η σύμμειξη πολυτροπικών ροών πληροφορίας στα πλαίσια της μηχανικής μάθησης αποτελεί ένα αρκετά δύσκολο και σχετικά ανεξερεύνητο πεδίο έρευνας. Όπως αντιλαμβανόμαστε αποτελεί μια πολύ σημαντική συνιστώσα για την ανάπτυξη ευφυών συστημάτων αναγνώρισης με στόχο την επικοινωνία ανθρώπου-μηχανής.

Από την πλευρά της παραγωγής της γλώσσας, αρκετοί ερευνητές [84], υποστηρίζουν ότι οι χειρονομίες παίζουν έναν πολύ σημαντικό ρόλο στην επικοινωνία. Οι χειρονομίες σε συνδυασμό με τη φωνή κατασκευάζουν ένα ολοκληρωμένο σύστημα επικοινωνίας [13], αλληλεπιδρώντας σε πολλαπλά γλωσσικά επίπεδα. Η αλληλεπίδραση αυτή εξερευνήθηκε πρόσφατα στα πλαίσια της επικοινωνίας και κατανόησης της γλώσσας [68]. Οι άνθρωποι, προφέρουν λέξεις κάνοντας ταυτόχρονα χειρονομίες, οι οποίες μπορεί να είναι είτε περιττές είτε συμπληρωματικές. Ακόμα και άνθρωποι οι οποίοι είναι τυφλοί κάνουν χειρονομίες όταν μιλούν σε άλλους τυφλούς [63]. Από την οπτική γωνία της φυσικής εξέλιξης των ανθρώπων, σε παιδιά 6-8 μηνών οι κινήσεις των χεριών εμφανίζονται παράλληλα με τις πρώτες προσπάθειες φωνητικής επικοινωνίας [13]. Επιπλέον, η κατανόηση λέξεων σε παιδιά 8-10 μηνών συνοδεύεται από δεικτικές χειρονομίες. Όλα τα παραπάνω αρκούν για να μας παρέχουν ενδεικτικά στοιχεία για το ότι οι χειρονομίες και η φωνή φαίνεται να είναι συνυφασμένες από πολλαπλές οπτικές γωνίες.

Στα πλαίσια της επικοινωνίας ανθρώπου-μηχανής οι χειρονομίες έχουν αρχίσει να αποκτούν ολοένα και περισσότερη προσοχή [130]. Ιδιαίτερα μετά τη διάσημη ερευνητική εργασία: 'put that there' [15]. Αυτό οφείλεται εν μέρει και στις πρόσφατες τεχνολογικές εξελίξεις, όπως η ευρεία διάδοση των αισθητήρων βάθους. Η επικοινωνία ανθρώπου-μηχανής μπορεί να ενισχυθεί σημαντικά με την αξιοποίηση πολλαπλών πολυτροπικών ροών πληροφορίας. Στατικές και δυναμικές χειρονομίες, η μορφή των χεριών (χειρομορφή) όπως και η φωνή συνθέτουν ένα ελκυστικό και πλούσιο φάσμα πληροφορίας το οποίο μπορεί να προσφέρει σημαντικά πλεονεκτήματα στα συστήματα επικοινωνίας ανθρώπου-μηχανής [112]. Για την αξιοποίηση όλων των παραπάνω στα πλαίσια της πολυτροπικής αναγνώρισης χειρονομιών εμφανίζεται πλήθος ανοιχτών ερευνητικών προβλημάτων προς επίλυση. Ενδεικτικά αναφέρουμε μερικά όπως:

- η ανίχνευση της σημαντικής πληροφορίας στα οπτικά και φωνητικά σήματα,
- η εξαγωγή των κατάλληλων χαρακτηριστικών,

- η εκπαίδευση αποτελεσματικών ταξινομητών και
- η σύμμειξη των πολυτροπικών ροών πληροφορίας.

Στο παρόν κεφάλαιο θα ασχοληθούμε με την ανίχνευση και αναγνώριση πολυτροπικών χειρονομιών οι οποίες εκτελούνται ελεύθερα από πολλαπλούς χρήστες. Στα πλαίσια αυτά χρησιμοποιούμε την απαιτητική βάση δεδομένων [44] η οποία πρόσφατα βιντεοσκοπήθηκε στα πλαίσια του διαγωνισμού πολυτροπικής αναγνώρισης χειρονομιών [43]. Αυτή περιλαμβάνει πολυτροπικές χειρονομίες που εμφανίζονται στην καθημερινότητα οι οποίες εκτελούνται από πολλαπλούς χρήστες και περιλαμβάνουν χειρονομίες συνοδευόμενες από τις αντίστοιχες φωνητικές λέξεις. Επιπλέον είναι αναμεμιγμένες με τυχαίες/άσχετες κινήσεις των χεριών και του σώματος του χρήστη όπως επίσης από τυχαίες/άσχετες φωνητικές λέξεις.

Πιο συγκεκριμένα, παρουσιάζουμε ένα σύστημα πολυτροπικής αναγνώρισης χειρονομιών [100, 101] το οποίο εκμεταλλεύεται ροές πληροφορίας που σχετίζονται με το χρώμα, το βάθος και τη φωνή όπως έχουν καταγραφεί από τον αισθητήρα Kinect. Εξαγάγει χαρακτηριστικά σχετιζόμενα με τη χειρομορφή, την κίνηση των χεριών και το σήμα φωνής. Εν συνεχεία προχωράμε στην σύμμειξη των πολλαπλών ροών πληροφορίας βασιζόμενοι σε ήδη υπάρχουσα ερευνητική εργασία, η οποία λέγεται *N-best rescoring*. Η μέθοδος N-best sentence hypotheses scoring παρουσιάστηκε για πρώτη φορά σε συστήματα αναγνώρισης φωνής για την ενσωμάτωση έρευνας σχετιζόμενης με την επεξεργασία της φυσικής γλώσσας [25]. Επιπλέον, αργότερα χρησιμοποιήθηκε για την σύμμειξη διαφορετικών συστημάτων αναγνώρισης φωνής [99], ή στα πλαίσια της ευρύτερης περιοχής της οπτικό-ακουστικής αναγνώρισης φωνής [52]. Βασιζόμενοι στα χαρακτηριστικά που έχουμε εξαγάγει και στα εκπαιδευμένα HMM μοντέλα, παράγουμε υποθέσεις αναγνώρισης (recognition hypotheses) για κάθε ακολουθία χειρονομιών και για κάθε ροή πληροφορίας ξεχωριστά. Εν συνεχεία, επαναξιολογούμε τις υποθέσεις αναγνώρισης χρησιμοποιώντας ένα πιθανοτικό πολυτροπικό πλαίσιο σύμμειξης των πολλαπλών αναγνωρισμένων υποθέσεων. Μετά, δεδομένου της πιο πιθανής υπόθεσης από τη σύμμειξη των πολυτροπικών ροών πληροφορίας και τη χρονική κατάτμηση κάθε χειρονομίας σε κάθε ροή πληροφορίας ξεχωριστά, εφαρμόζουμε ένα τελικό σχήμα σύμμειξης βασιζόμενοι στα parallel HMMs [136] το οποίο το ονομάζουμε *segmental parallel fusion*. Η μοντελοποίηση βασίζεται σε HMM μοντέλα, ένα για κάθε χειρονομία και κάθε ροή πληροφορίας ξεχωριστά. Τέλος, χρησιμοποιείται επιπλέον ένα αυτόματο σύστημα ανίχνευσης δράσης (activity detection) σε κάθε ροή πληροφορίας ξεχωριστά για την αρχικοποίηση των HMM μοντέλων.

6.1 Βάση δεδομένων με πολυτροπικές χειρονομίες

Σε αυτή την ενότητα παρουσιάζουμε τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση και αξιολόγηση του συστήματος αναγνώρισης πολυτροπικών χειρονομιών. Η βάση δεδομένων έγινε στα πλαίσια του διαγωνισμού πολυτροπικής αναγνώρισης χειρονομιών [44]. Η βάση προσφέρει εικόνες βάθους και χρώματος (RGB), τη μάσκα του χρήστη, την πληροφορία του σκελετού κάθε χρήστη, τον προσανατολισμό των αρθρώσεων του χρήστη, όπως επίσης το φωνητικό σήμα που συνοδεύει την εκτέλεση κάθε χειρονομίας (βλέπε Σχήμα 6.1). Ο αριθμός των πολυτροπικών χειρονομιών προς αναγνώριση είναι 20 και η γλώσσα στην οποία λέγονται φωνητικές λέξεις που συνοδεύουν κάθε χειρονομία είναι στα Ιταλικά. Η βάση δεδομένων αποτελείται από τρία διαφορετικά μη επικαλυπτόμενα σύνολα δεδομένων, τα οποία είναι τα εξής: α) εκπαίδευσης/ανάπτυξης, β) επικύρωσης και γ) τελικής αξιολόγησης. Εμπεριέχει 39 διαφορετικούς χρήστες και 13858 εκτελέσεις πολυτροπικών χειρονομιών στο σύνολο. Μαζί με τα βίντεο δεδομένα προσφέρονται χαλαρές επισημειώσεις σε επίπεδο χειρονομίας κοινές για όλες τις ροές πληροφορίας: οπτική και φωνητική.

Υπάρχουν αρκετοί λόγοι που κάνουν τη συγκεκριμένη βάση δεδομένων αρκετά δύσκολη. Πρώτον δεν υπάρχει ένας συγκεκριμένος τρόπος για την εκτέλεση κάθε χειρονομίας. Παραδείγματος

(α΄) RGB (β΄) Βάθος (γ΄) Μάσκα (δ΄) Σκελετός

Σχήμα 6.1: Παραδείγματα των δεδομένων που εμπεριέχονται στην πολυτροπική βάση δεδομένων χειρονομιών [44].

(α΄) (β΄) (γ΄) (δ΄)

Σχήμα 6.2: (α,β) Μεταβολή της θέσης του μπράτσου του χρήστη (χαμηλά, ψηλά) για την χειρονομία 'vieni qui'. (γ,δ) Εκτέλεση της χειρονομίας 'vattene' και από τα δύο χέρια.

χάριν, η χειρονομία 'vieni qui' εκτελείται με μια επαναλαμβανόμενη κίνηση του χεριού προς το μέρος του χρήστη, με μεταβλητό αριθμό επαναλήψεων (βλέπε Σχήμα. 6.2). Όμοια, χειρονομίες που εκτελούνται με το ένα χέρι μπορεί να εκτελεστούν είτε με το αριστερό είτε με το δεξί. Επιπλέον, έχουμε την εισαγωγή είτε άσχετων χειρονομιών στην οπτική ροή είτε λέξεων εκτός λεξιλογίου στην ακουστική ροή, με στόχο την αύξηση της δυσκολίας του προβλήματος. Τέλος, σε σχέση με την οπτική ροή έχουμε εναλλαγές στο φόντο, στις συνθήκες φωτισμού, στην ανάλυση της εικόνας. Στην ακουστική ροή έχουμε πολλαπλές διαφορετικές προφορές της ίδιας λέξης και εισαγωγή τυχαίου θορύβου όπως π.χ. πληκτρολόγιο, ανεμιστήρα, άνοιγμα/κλείσιμο πόρτας κ.α. Αυτό έχει ως αποτέλεσμα το πρόβλημα αναγνώρισης των πολυτροπικών αυτών χειρονομιών σε αυτή τη βάση δεδομένων να αποτελεί ένα αρκετά δύσκολο πρόβλημα.

6.2 Προτεινόμενη Μεθοδολογία

Για την καλύτερη εξήγηση του προτεινόμενου συστήματος πολυτροπικής αναγνώρισης χειρονομιών ας αναφερθούμε σε ένα παράδειγμα. Οι πολυτροπικές χειρονομίες, ή αλλιώς χειρονομίες συνοδευόμενες από φωνή, χρησιμοποιούνται αρκετά συχνά σε πολλές διαφορετικές περιπτώσεις και κουλτούρες [85, 69]. Ένα παράδειγμα είναι η χειρονομία 'OK', η οποία εκφράζεται δημιουργώντας έναν κύκλο χρησιμοποιώντας τον αντίχειρα και τον δείκτη και κρατώντας τα άλλα δάκτυλα ανοιχτά και επιπλέον εκφέροντας ταυτόχρονα την λέξη 'Okay'. Όμοια η χειρονομία 'Come here' εκφράζεται εκτελώντας το λεγόμενο beckoning νόημα. Στη Βόρειο Αμερική γίνεται σηκώνοντας τον

δείκτη από την κλειστή παλάμη και κινώντας τον επαναλαμβανόμενα, και ταυτόχρονα εκφέροντας την φράση 'Come here'. Σε αυτό το κεφάλαιο παρουσιάζουμε ένα σύστημα για την αυτόματη ανίχνευση και αναγνώριση τέτοιου είδους χειρονομιών ακόμα και όταν αυτές είναι αναμεμειγμένες με άσχετες δράσεις. Αυτές μπορεί να είναι λεκτικές, μη-λεκτικές ή και τα δύο. Ο χρήστης μπορεί για παράδειγμα να περπατάει στο ενδιάμεσο των χειρονομιών ή να μιλάει σε κάποιον άλλο.

Στα πλαίσια αυτά, ασχολούμαστε με χειρονομίες οι οποίες είναι πάντα πολυτροπικές. Με άλλα λόγια δεν εκφράζονται μόνο λεκτικά ή μη-λεκτικά. Παρόλα αυτά δεν είναι απαραίτητο οι διαφορετικές ροές πληροφορίας να είναι συγχρονισμένες. Η μόνη υπόθεση η οποία γίνεται είναι ότι συνεχόμενες χειρονομίες πρέπει να είναι διαχωρίσιμες χρονικά. Με άλλα λόγια να απέχουν μερικά χιλιοστά του δευτερολέπτου σε όλες τις ροές πληροφορίας. Επιπλέον δεν τίθενται γλωσσικοί περιορισμοί σχετικά με την ακολουθία των χειρονομιών που θα εκτελεστεί κάθε φορά. Δηλαδή, π.χ. ότι μετά από την χειρονομία Α δεν μπορεί να εκτελεστεί η χειρονομία Β. Όλες οι πιθανές ακολουθίες χειρονομιών είναι επιτρεπτές.

Έστω $G = \{g_i\}, i = 1, \ldots, |G|$ το σύνολο των πολυτροπικών χειρονομιών προς ανίχνευση και αναγνώριση ενώ $S = \{\mathbf{O}_i\}, i = 1, \ldots, |S|$ το σύνολο των ροών πληροφορίας που παρατηρούνται ταυτόχρονα για τον σκοπό αυτό. Στη δική μας περίπτωση το τελευταίο σύνολο αποτελείται από τρεις ροές πληροφορίας:

- τα χαρακτηριστικά του φάσματος της φωνής,
- τα χαρακτηριστικά του σκελετού,
- και τα χαρακτηριστικά της χειρομορφής.

Βασιζόμενοι σε αυτές τις παρατηρήσεις, το προτεινόμενο σύστημα παράγει μια υπόθεση αναγνώρισης για την ακολουθία των χειρονομιών που εμφανίζονται σε ένα συγκεκριμένο βίντεο όπως φαίνεται παρακάτω:

$$\mathbf{h} = [bm, g_1, sil, g_5, \ldots, bm, sil, g_3]. \tag{6.1}$$

Το σύμβολο sil αντιστοιχεί στην αδράνεια/μη-δράση η οποία εμφανίζεται σε μια ή σε περισσότερες ροές πληροφορίας. Ενώ το σύμβολο bm αντιπροσωπεύει οποιαδήποτε άλλη δράση εκτός από τις προκαθορισμένες χειρονομίες G η οποία εμφανίζεται σε μια ή σε περισσότερες ροές πληροφορίας. Η παραπάνω υπόθεση αναγνώρισης για την ακολουθία των χειρονομιών που εκτελέστηκε παράγεται μέσω του προτεινόμενου αλγορίθμου σύμμειξης που συνοψίζεται στον Algorithm 2.

6.2.1 Παραγωγή των καλύτερων υποθέσεων αναγνώρισης

Χρησιμοποιώντας τα μοντέλα χειρονομιών για κάθε ροή πληροφορίας ξεχωριστά (βλ. ενότητα 6.3) και μια gesture-loop γραμματική όπως φαίνεται στο Σχήμα 6.3(α) παράγουμε μια λίστα των καλύτερων υποθέσεων αναγνώρισης για την άγνωστη ακολουθία χειρονομιών. Για αυτό χρησιμοποιούμε τον αλγόριθμο lattice N-best algorithm [114] ο οποίος υλοποιήθηκε χρησιμοποιώντας την έννοια extended token passing, όπως περιγράφεται στο [151]. Αυτός ο αλγόριθμος είναι ουσιαστικά μια παραλλαγή του αλγορίθμου Viterbi [105] ο οποίος αντί να κρατάει μόνο την καλύτερη υπόθεση αναγνώρισης, αποθηκεύει τις πρώτες N πιο πιθανές υποθέσεις αναγνώρισης. Η τιμή N δίνεται ως είσοδο στον αλγόριθμο.

Οι N-best λίστες παράγονται ανεξάρτητα για κάθε ροή πληροφορίας ξεχωριστά. Στη συνεχεία, φτιάχνουμε την τελική λίστα υποθέσεων αναγνώρισης, η οποία περιέχει όλες τις παραπάνω υποθέσεις από όλες τις ροές πληροφορίας. Αυτή μπορεί να περιλαμβάνει πολλαπλές επαναλήψεις της ίδιας υπόθεσης αναγνώρισης. Αφαιρώντας πιθανές πολλαπλές επαναλήψεις της ίδιας υπόθεσης, καταλήγουμε με L υποθέσεις και το σύνολο $H = \{\mathbf{h}_1, \ldots, \mathbf{h}_L\}$. $\mathbf{h}_i$ είναι μια ακολουθία χειρονομιών (η οποία μπορεί επίσης να περιλαμβάνει sil και bm). Στόχος μας είναι η αναδιάταξη του

Algorithm 2 Αναγνώριση Πολυτροπικών Χειρονομιών χρησιμοποιώντας Πολλαπλές Υποθέσεις Αναγνώρισης

```
% Αναγνώριση χρησιμοποιώντας μια ροή πληροφορίας κάθε φορά
for all ροές do
  παραγωγή των N-best υποθέσεων
end for
κρατάμε μόνο τη μια υπόθεση από πιθανές επαναλήψεις της ίδιας υπόθεσης

% N-best list rescoring
for all υποθέσεις do
  % Δημιουργία της γραμματικής
  κρατάμε την ακολουθία των χειρονομιών ως έχει
  αλλά επιτρέπουμε παρεμβολές/διαγραφές των sil και bm ανάμεσα στις χειρονομίες
  for all ροές do
    εφαρμόζοντας τη γραμματική:
    1) βρίσκουμε την καλύτερη ευθυγράμμιση με τις παρατηρήσεις
    2) σώζουμε το αντίστοιχο σκορ
  end for
  % Σύμμειξη, επανασκοράρισμα της κάθε υπόθεσης αναγνώρισης
  το τελικό score κάθε υπόθεσης είναι το σταθμισμένο άθροισμα όλων των score για κάθε ροή
end for
η καλύτερη υπόθεση του 1st-pass είναι αυτή με το μεγαλύτερο score

% Παράλληλο σκοράρισμα
for all ροές do
  κατάτμηση των παρατηρήσεων με βάση την καλύτερη υπόθεση
  for all τμήματα do
    εκτίμηση του σκορ για κάθε χειρονομία δεδομένου των παρατηρήσεων στο συγκεκριμένο τμήμα
    ευθυγράμμιση των τμημάτων από τις διαφορετικές ροές
    for all ευθυγραμμισμένα τμήματα do
      εκτίμηση του σταθμισμένου αθροίσματος των σκορ για όλες τις ροές και χειρονομίες
      επίλεξε την υπόθεση με το καλύτερο score (συμπεριλαμβανομένου των sil και bm)
    end for
  end for
end for
```

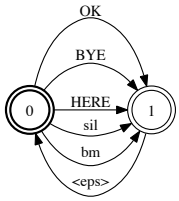
συνόλου των υποθέσεων για την εύρεση της πιο πιθανής υπόθεσης αναγνώρισης, αξιοποιώντας όλες τις ροές πληροφορίας αυτή τη φορά.

6.2.2 Πολυτροπικό σκοράρισμα και αναδιάταξη των υποθέσεων αναγνώρισης

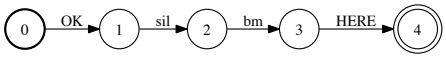
Προς αυτή την κατεύθυνση, υπολογίζουμε ένα συνδυασμένο σκόρ για κάθε υπόθεση αναγνώρισης ως το σταθμισμένο άθροισμα όλων των σκορ κάθε ροής:

$$V_i = \sum_{m \in S} w_m c_{m,i}, \quad i = 1 \dots L. \tag{6.2}$$

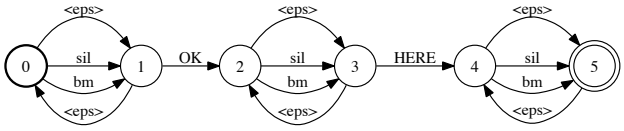
Τα βάρη w_m υπολογίζονται πειραματικά με βάση το καλύτερο ποσοστό αναγνώρισης σε ένα σύνολο δεδομένων αξιολόγησης. Τα σκορ $c_{m,i}$ για κάθε ροή είναι κανονικοποιημένες εκδοχές των $v_{m,i}$



(α′) (α)



(β′) (β)



(γ′) (γ)

Σχήμα 6.3: Αναπαράσταση με Finite-state-automaton (FSA) των γραμματικών: (α) ένα παράδειγμα της γραμματικής gesture-loop. Αυτή περιλαμβάνει τρεις χειρονομίες, την περίπτωση αδράνειας (sil) και άσχετης δράσης (bm). Η μετάβαση 'eps' αντιστοιχεί σε ϵ μετάβαση του FSA. (β) ένα παράδειγμα υπόθεσης αναγνώρισης, (γ) γραμματική δεδομένης της υπόθεσης αναγνώρισης, η οποία επιτρέπει την εισαγωγή/διαγραφή sil και bm ανάμεσα στις χειρονομίες.

σκορ τα οποία έχουν υπολογιστεί χρησιμοποιώντας τον αλγόριθμο Viterbi:

$$v_{m,i} = \max_{\mathbf{h} \in G_{h_i}} \log P(\mathbf{O}_m | \mathbf{h}, \lambda_m), \quad i = 1, \ldots, L, \ m = 1, \ldots, |S| \tag{6.3}$$

όπου $\mathbf{O}_m$ είναι η ακολουθία παρατηρήσεων για την ροή m και λ_m τα αντίστοιχα μοντέλα.

Το παραπάνω ουσιαστικά επιλύει ένα περιορισμένο πρόβλημα αναγνώρισης όπου οι επιτρεπτές ακολουθίες χειρονομιών είναι απαραίτητο να ακολουθούν μια συγκεκριμένη finite state γραμματική G_{h_i}, δεδομένης της υπόθεσης αναγνώρισης. Είναι απαραίτητο οι ακολουθίες καταστάσεων να συμπεριλαμβάνουν ακολουθίες χειρονομιών που αντιστοιχούν στην υπόθεση αναγνώρισης $\mathbf{h}_i$. Παρόλα αυτά δίνεται η δυνατότητα εισαγωγής, διαγραφής ή αντικατάστασης sil και bm μοντέλων ανάμεσα στις διαφορετικές χειρονομίες. Ένα παράδειγμα μιας υπόθεσης αναγνώρισης και της αντίστοιχης finite state γραμματικής απεικονίζεται στα Σχήματα 6.3(β,γ).

Με αυτό τον τρόπο το πλαίσιο σκοραρίσματος λαμβάνει υπόψη του περιπτώσεις μη-δράσης ή άσχετων δράσεων οι οποίες δεν είναι απαραίτητο να συμβαίνουν ταυτόχρονα από όλες τις ροές πληροφορίας. Παραδείγματος χάριν, ο χρήστης μπορεί να στέκεται ακίνητος αλλά ταυτόχρονα να μιλάει σε κάποιον άλλο ή να κάνει μια άσχετη χειρονομία χωρίς να μιλάει. Χρησιμοποιώντας την παραπάνω γραμματική επιτυγχάνουμε αύξηση του ποσοστού αναγνώρισης σε σύγκριση με το να εφαρμόζαμε απλά force-alignment δεδομένης της υπόθεσης αναγνώρισης.

Η καλύτερη υπόθεση σε αυτό το στάδιο είναι αυτή με το μεγαλύτερο συνδυασμένο σκορ όπως υπολογίστηκε από την εξ. 6.2. Η καλύτερη υπόθεση αναγνώρισης μαζί με τα αντίστοιχα χρονικά όρια κάθε χειρονομίας, της ακολουθίας χειρονομιών, δίνονται ως είσοδο στο επόμενο στάδιο. Αυτό είναι το τμηματικό παράλληλο σκοράρισμα (segmental parallel scoring). Σε αυτό το στάδιο, επιτρέπονται μόνο τοπικές βελτιώσεις εκμεταλλευόμενοι πιθανά οφέλη από τη διαδικασία ταξινόμησης χρονικών τμημάτων. Αξίζει να σημειωθεί ότι τα χρονικά όρια κάθε χειρονομίας μπορεί να είναι διαφορετικά σε κάθε ροή πληροφορίας,

6.2.3 Τμηματική παράλληλη σύμμειξη

Εδώ εκμεταλλευόμαστε τα χρονικά όρια κάθε χειρονομίας για κάθε ροή πληροφορίας ξεχωριστά, για την πιο πιθανή ακολουθία χειρονομιών, η οποία καθορίστηκε προηγουμένως. Στόχος μας είναι η μετατροπή του προβλήματος αναγνώρισης σε πρόβλημα ταξινόμησης. Πρώτα, κάνουμε κατάτμηση σε επίπεδο χειρονομιών των ροών: φωνή, σκελετός και χειρομορφή χρησιμοποιώντας τα χρονικά όρια κάθε χειρονομίας και ροής πληροφορίας. Δεδομένου ότι ενδιάμεσα των χειρονομιών μπορεί να έχουμε διαφορετικό αριθμό από sil και bm μοντέλων σε κάθε ροή πληροφορίας, είναι απαραίτητη η ευθυγράμμιση των υποθέσεων των ροών πληροφορίας. Για τους σκοπούς τη ευθυγράμμισης εφαρμόζουμε βέλτιστο ταίριασμα συμβολοσειρών κάνοντας χρήση δυναμικού προγραμματισμού. Μετά, για κάθε ευθυγραμμισμένο τμήμα t και για κάθε ροή πληροφορίας m υπολογίζουμε τη λογαριθμική πιθανότητα:

$$LL^t_{m,j} = \max_{\mathbf{q}\in Q} \log P(\mathbf{O}^t_m, \mathbf{q}|\lambda_{m,j}), \quad j = 1, \ldots, |G| + 2, \tag{6.4}$$

όπου $\lambda_{m,j}$ είναι οι παράμετροι του HMM μοντέλου για τη χειρονομία $g_j \in G \cup \{sil, bm\}$ και για τη ροή πληροφορίας $m \in S$. Επιπλέον $\mathbf{q}$ είναι μια πιθανή ακολουθία καταστάσεων του HMM. Τα παραπάνω σκορ συνδυάζονται γραμμικά για όλες τις ροές πληροφορίας με στόχο τον υπολογισμό ενός τελικού σκορ για κάθε τμήμα:

$$L^t_j = \sum_{m\in S} w'_m LL^t_{m,j}, \tag{6.5}$$

όπου w'_m είναι το βάρος για την ροή πληροφορίας m, η οποία επιλέγεται βελτιστοποιώντας το ποσοστό αναγνώρισης σε ένα σύνολο δεδομένων αξιολόγησης. Τέλος η χειρονομία με το μεγαλύτερο σκορ αντιστοιχεί στην τελική αναγνώριση για το τμήμα t.

6.3 Μοντελοποίηση ροών: φωνή, σκελετός και χειρομορφή

Για κάθε ροή πληροφορίας (φωνή, σκελετός, χειρομορφή) γίνεται ξεχωριστή μοντελοποίηση βασιζόμενη στα HMMs και στο keyword-filler paradigm. Το keyword-filler paradigm παρουσιάστηκε πρώτη φορά για τη φωνή [141, 108], σε εφαρμογές για indexing/retrieval σε ακουστικά έγγραφα [50] ή παρακολούθηση μέσω αναγνώρισης φωνής [107]. Το πρόβλημα αναγνώρισης ενός μικρού αριθμού χειρονομιών σε ένα βίντεο το οποίο μπορεί να περιλαμβάνει και άσχετες δράσεις, είτε λεκτικές είτε μη-λεκτικές, μπορεί να ιδωθεί ως ένα πρόβλημα ανίχνευσης λέξεων-κλειδιών (keyword detection). Οι χειρονομίες προς αναγνώριση αντιστοιχούν στις λέξεις-κλειδιά και όλες οι άλλες δράσεις πρέπει να αγνοούνται. Βασιζόμενοι στο παραπάνω, κάθε χειρονομία $g_i \in G$ μοντελοποιείται από ένα HMM, διαφορετικό για κάθε ροή πληροφορίας. Επιπλέον εκπαιδεύονται και δύο επιπλέον HMMs τα οποία αντιπροσωπεύουν τις μη-δράσεις (sil) και τις δράσεις (bm) για κάθε ροή πληροφορίας αντίστοιχα.

Όλα αυτά τα μοντέλα είναι left-to-right HMMs με ένα μείγμα Γκαουσιανών σε κάθε κατάσταση αντιπροσωπεύοντας την κατανομή πυκνότητας πιθανότητας των παρατηρήσεων δεδομένης της κατάστασης του HMM. Αρχικοποιούνται μέσω μια επαναληπτικής διαδικασίας η οποία θέτει τις παραμέτρους του μοντέλου, βάσει της μέσης τιμής και της διακύμανσης των παρατηρήσεων που αντιστοιχίζονται στην κάθε κατάσταση για το σύνολο των δεδομένων εκπαίδευσης. Επιπλέον επανεκτιμά τις παραμέτρους αυτές βελτιώνοντας τα χρονικά όρια της αντιστοίχισης των παρατηρήσεων με τις καταστάσεις του HMM χρησιμοποιώντας τον αλγόριθμο Viterbi [149]. Τέλος, η εκπαίδευση γίνεται χρησιμοποιώντας τον αλγόριθμο Baum-Welch [105], αυξάνοντας σταδιακά τον αριθμό των Γκαουσιανών σε κάθε μείγμα. Μολονότι ο γενικός αλγόριθμος εκπαίδευσης είναι ο παραπάνω, εξετάζονται δύο διαφορετικές εκδοχές. Αυτές περιγράφονται στη συνέχεια.

Εκπαίδευση χωρίς τη χρήση ανίχνευσης δράσης

Τα μοντέλα για κάθε ροή, μπορούν να αρχικοποιηθούν και να εκπαιδευτούν με βάση χονδροειδείς πολυτροπικές επισημειώσεις σε επίπεδο χειρονομίας. Αυτές οι επισημειώσεις είναι κοινές και ίδιες για όλες τις ροές πληροφορίας. Δεδομένου ότι δεν υπάρχει πλήρης συγχρονισμός μεταξύ των ροών, οι επισημειώσεις σε κάποιες από τις ροές μπορεί να εμπεριέχουν χρονικά τμήματα μη-δράσης ή άλλων άσχετων δράσεων στην αρχή ή/και στο τέλος της κάθε επισημειωμένης χειρονομίας. Με αυτό τον τρόπο, οι μη-δράσεις ενσωματώνονται στην αρχή και στο τέλος κάθε μοντέλου χειρονομίας. Άρα στην περίπτωση αυτή, δεν εκπαιδεύεται ξεχωριστό μοντέλο μη-δράσης (sil). Ενώ το μοντέλο δράσης εκπαιδεύεται χρησιμοποιώντας όλα τα δεδομένα εκπαίδευσης από όλες τις χειρονομίες. Το πλεονέκτημα αυτής της προσέγγισης είναι ότι μπορεί να μοντελοποιήσει έμμεσα την χρονική συσχέτιση των διαφορετικών ροών πληροφορίας. Ας πάρουμε ως παράδειγμα την εκφορά της χειρονομίας 'Bye bye'. Ο κυματισμός του χεριού μπορεί να ξεκινήσει πριν την εκφορά της λέξης. Αυτό έχει ως αποτέλεσμα πριν τη φωνητική εκφώνηση της λέξης 'Bye bye' να υπάρχει ένα χρονικό κομμάτι σιωπής (ή μιας άλλης φωνητικής δράσης), η οποία μοντελοποιείται έμμεσα.

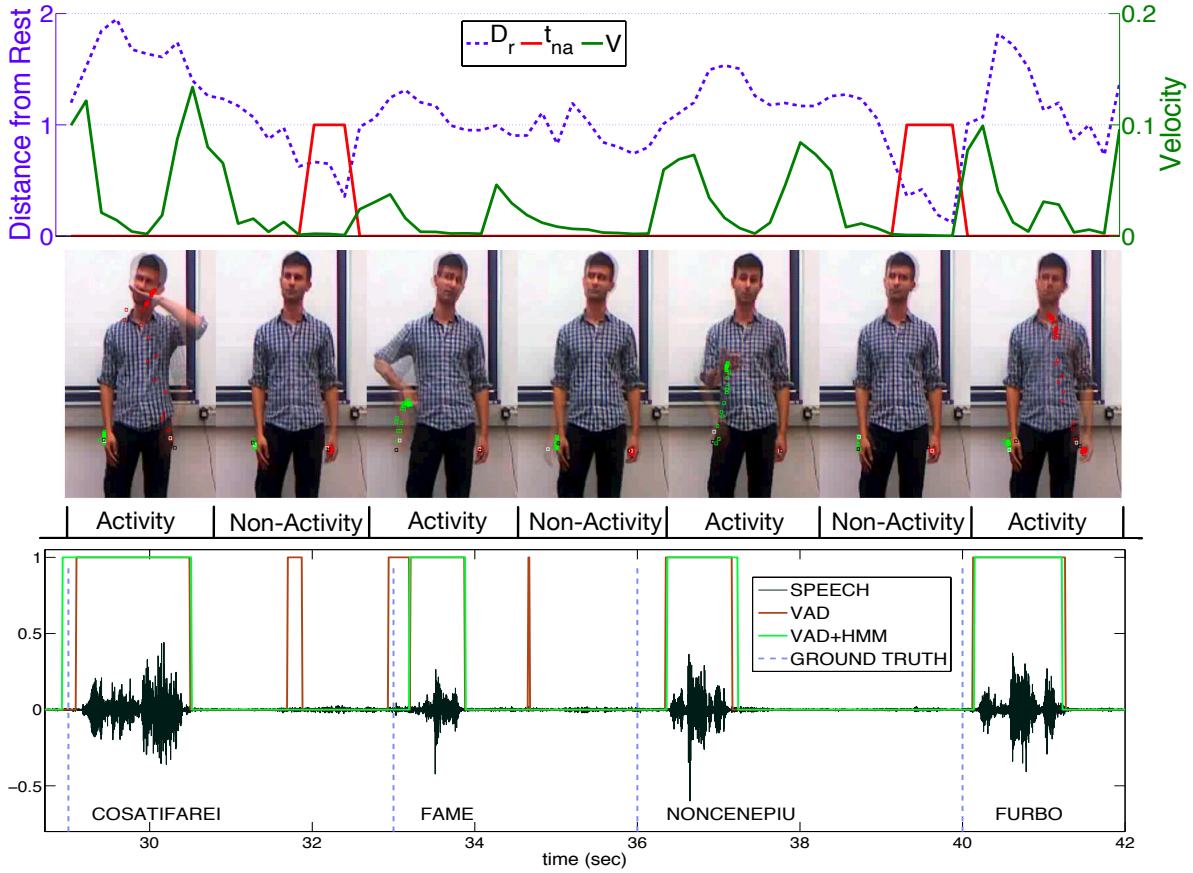
Εκπαίδευση με τη χρήση ανίχνευσης δράσης

Από την άλλη πλευρά η εκπαίδευση των μοντέλων χειρονομιών μπορεί να γίνει χρησιμοποιώντας για κάθε ροή πληροφορίας διαφορετικές επισημειώσεις επιπέδου χειρονομιών. Προς αυτή την κατεύθυνση, εφαρμόζουμε έναν αλγόριθμο ανίχνευσης δράσης activity detection ο οποίος περιγράφεται με λεπτομέρεια στην ενότητα 6.4. Βασιζόμενοι σε αυτόν, είναι εφικτό να αποκτήσουμε πιο σφιχτά χρονικά όρια κάθε χειρονομίας για κάθε ροή πληροφορίας ξεχωριστά. Έτσι τα μοντέλα χειρονομίας εκπαιδεύονται χρησιμοποιώντας αυτά τα πιο σφιχτά χρονικά όρια. Επιπλέον τα μοντέλα δράσης (bm) και μη-δράσης (sil) εκπαιδεύονται χρησιμοποιώντας τα χρονικά τμήματα που έχουν αντιστοιχηθεί σε τμήματα δράσης και μη-δράσης αντίστοιχα. Κατά αυτόν τον τρόπο τα μοντέλα χειρονομιών για κάθε ροή πληροφορίας μοντελοποιούν με μεγαλύτερη ακρίβεια κάθε χειρονομία. Ωστόσο, χάνουμε την πληροφορία σχετικά με τη χρονική συσχέτιση των διαφορετικών ροών πληροφορίας.

6.4 Πολυτροπική ανίχνευση δράσης

Για την επίτευξη της ανίχνευσης δράσης και στην ακουστική αλλά και στην οπτική ροή πληροφορίας, χρησιμοποιούμε ένα κοινό πλαίσιο βασιζόμενο σε δύο συμπληρωματικά μοντέλα. Αυτά μοντελοποιούν χρονικά τμήματα που αντιστοιχούν σε 'δράση' ή 'μη-δράση' αντίστοιχα. Παρόλα αυτά λόγω της διαφορετικής φύσης των ροών πληροφορίας, τα μοντέλα αυτά έχουν διαφορετική ερμηνεία για κάθε ροή. Για την περίπτωση της φωνής, το μοντέλο μη-δράσης μοντελοποιεί χρονικά τμήματα όπου έχουμε ησυχία, όπως και θόρυβο π.χ. θόρυβο από πληκτρολόγιο ή ανεμιστήρα κ.α. Το μοντέλο δράσης μοντελοποιεί χρονικά τμήματα όπου έχουμε φωνή, είτε αυτή αντιστοιχεί σε λέξεις εντός λεξιλογίου είτε όχι. Αντιθέτως για την περίπτωση της οπτικής ροής, το μοντέλο μη-δράσης μοντελοποιεί χρονικά τμήματα όπου ο χρήστης βρίσκεται στην θέση ξεκούρασης ενδιαμέσως της άρθρωσης δύο συνεχόμενων χειρονομιών. Αξίζει να σημειωθεί ότι η θέση ξεκούρασης δεν είναι αυστηρώς καθορισμένη: κάθε χρήστης μπορεί να έχει διαφορετική θέση ξεκούρασης. Το μοντέλο δράσης μοντελοποιεί χρονικά τμήματα όπου έχουμε την άρθρωση μιας χειρονομίας είτε εκτός, είτε εντός λεξιλογίου. Ο ανιχνευτής δράσης αρχικοποιείται και εκπαιδεύεται για κάθε ροή πληροφορίας ανεξάρτητα, όπως περιγράφεται στη συνέχεια.

Για την περίπτωση της φωνής, τα μοντέλα δράσης και μη-δράσης αρχικοποιούνται χρησιμοποιώντας χρονικά τμήματα δράσης και μη-δράσης αντίστοιχα. Αυτά προσδιορίζονται χρησιμοποιώντας μια μέθοδο για ανίχνευση φωνής (Voice Activity Detection -VAD-) που πρόσφατα



Σχήμα 6.4: Ένα παράδειγμα μια ακολουθίας χειρονομιών μαζί με την ανίχνευση δράσης και μη-δράσης και για την ακουστική αλλά και την οπτική ροή πληροφορίας. Πρώτη σειρά: Η ταχύτητα των χεριών (V), η απόσταση του σκελετού από τη θέση ξεκούρασης (D_r) και το αποτέλεσμα της αρχικής εκτίμησης των χρονικών τμημάτων που αντιστοιχούν σε μη-δράση (t_{na}). Δεύτερη σειρά: Η εκτίμηση δράσης και μη-δράσης, απεικονίζοντας τις πραγματικές εικόνες από το βίντεο. Τρίτη σειρά: Το ακουστικό σήμα συνοδευόμενο με την εκτίμηση δράσης των VAD και VAD+HMM, όπως επίσης τα επισημειωμένα χρονικά όρια κάθε χειρονομίας που περιέχει η βάση δεδομένων (ground truth).

προτάθηκε από τους Tan et al. [121]. Αυτή βασίζεται σε Likelihood Ratio Tests (LRTs). Διαχειρίζοντας τα LRT's διαφορετικά για τα έμφωνα και για τα άφωνα χρονικά πλαίσια, βελτιώνει τα αποτελέσματα σε σχέση με τις συμβατικές LRT και VAD μεθόδους. Στη συνέχεια εκπαιδεύουμε τα δύο HMM μοντέλα, δράσης και μη-δράσης, χρησιμοποιώντας μια επαναληπτική διαδικασία γνωστή ως embedded re-estimation [149] κάνοντας χρήση του αλγορίθμου Baum-Welch. Τα τελικά χρονικά όρια των τμημάτων φωνητικής δράσης και μη-δράσης, προσδιορίζονται εφαρμόζοντας τον αλγόριθμο Viterbi χρησιμοποιώντας τα ήδη εκπαιδευμένα HMM μοντέλα.

Για την οπτική ροή πληροφορίας, στόχος μας είναι η ανίχνευση τμημάτων δράσης, δηλαδή τις χρονικές περιοχές όπου έχουμε την άρθρωση μιας χειρονομίας ή απλά κίνηση των αρθρωτών του χρήστη, σε σχέση με χρονικές περιοχές όπου ο χρήστης βρίσκεται στη θέση ξεκούρασης. Για τους σκοπούς αυτούς, αρχικοποιούμε το μοντέλο μη-δράσης χρησιμοποιώντας χρονικά τμήματα μη-δράσης. Για την εύρεση των χρονικών τμημάτων αυτών, βασιζόμαστε στο ότι ο χρήστης δεν κινείται (σχεδόν ακίνητος), άρα η ταχύτητα του σκελετού του χρήστη είναι αρκετά χαμηλή, και επιπλέον είναι πολύ κοντά στη θέση ξεκούρασης $\mathbf{x_r}$. Έτσι, πρέπει να εκτιμήσουμε τη θέση ξεκού-

ρασης σε κάθε βίντεο, όπως επίσης να υπολογίσουμε την ταχύτητα των χεριών και την απόσταση του σκελετού από την εκτιμώμενη θέση ξεκούρασης, σε κάθε χρονική στιγμή. Η θέση ξεκούρασης εκτιμάται ως η μέση θέση του σκελετού όλων των χρονικών τμημάτων όπου η ταχύτητα των χεριών V είναι μικρότερη ενός κατωφλίου $V_{t_r} = 0.2 \cdot \bar{V}$, όπου $\bar{V}$ είναι η μέση ταχύτητα των χεριών σε όλο το βίντεο. Η ταχύτητα των χεριών υπολογίζεται ως $V(\mathbf{x}) = |\dot{\mathbf{x}}|$ όπου $\mathbf{x}(t)$ είναι η τρισδιάστατη θέση των χεριών την χρονική στιγμή t. Η απόσταση του σκελετού από τη θέση ξεκούρασης υπολογίζεται ως $D_r(\mathbf{x}) = |\mathbf{x} - \mathbf{x_r}|$. Τα αρχικά εκτιμώμενα χρονικά τμήματα μη-δράσης t_{na} είναι αυτά όπου ικανοποιούνται τα παρακάτω δύο κριτήρια: $\mathbf{t_{na}} = \{t : \ D_r(\mathbf{x}) < D_{t_r} and \, V(\mathbf{x}) < V_{t_r}\}$. Παίρνοντας ως είσοδο αυτά τα t_{na} τμήματα, εκπαιδεύουμε το HMM μοντέλο μη-δράσης, ενώ το μοντέλο δράσης εκπαιδεύεται χρησιμοποιώντας τα περισσευούμενα τμήματα χρησιμοποιώντας ως διάνυσμα χαρακτηριστικών τον σκελετό. Στη συνέχεια, όπως και στην περίπτωση της φωνής, επανεκπαιδεύουμε και τα δύο HMM μοντέλα χρησιμοποιώντας τον αλγόριθμο embedded re-estimation. Τέλος, τα τελικά χρονικά όρια των τμημάτων οπτικής δράσης και μη-δράσης προσδιορίζονται εφαρμόζοντας τον αλγόριθμο Viterbi χρησιμοποιώντας τα παραπάνω HMM μοντέλα.

Στο Σχήμα 6.4 απεικονίζουμε ένα παράδειγμα μιας ακολουθίας χειρονομιών μαζί με την ανίχνευση δράσης και μη-δράσης και για την ακουστική αλλά και την οπτική ροή πληροφορίας. Στην πρώτη σειρά δείχνουμε την ταχύτητα των χεριών (V), την απόσταση του σκελετού από τη θέση ξεκούρασης (D_r) και την αρχική εκτίμηση των χρονικών τμημάτων μη-δράσης (t_{na}). Παρατηρούμε ότι στα t_{na} τμήματα οι τιμές των V και D_r είναι μικρότερες των προκαθορισμένων κατωφλιών ($V_{t_r} = 0.6, D_{t_r} = 0.006$). Στη δεύτερη σειρά απεικονίζουμε τις πραγματικές εικόνες από το βίντεο, έχοντας κάνει υπέρθεση τις τροχιές των χεριών. Επιπλέον απεικονίζουμε την τελική εκτίμηση των χρονικών τμημάτων δράσης και μη-δράσης. Στο κάτω μέρος του σχήματος, απεικονίζουμε το φωνητικό σήμα μαζί με την αρχική εκτίμηση των φωνητικών δράσεων από τον VAD αλγόριθμο, την διορθωμένη χρησιμοποιώντας τα εκπαιδευμένα HMM μοντέλα (VAD+HMM) και τις επισημειώσεις σε επίπεδο χειρονομιών που περιέχει η βάση δεδομένων (ground truth). Όπως παρατηρούμε, τα διορθωμένα χρονικά όρια τα οποία εξάγονται χρησιμοποιώντας την μέθοδο VAD+HMM είναι αρκετά πιο σφιχτά και ακριβή σε σχέση με την αρχική εκτίμηση από το VAD αλγόριθμο και των ground truth επισημειώσεων.

Για να συνοψίσουμε, με την εφαρμογή των ανιχνευτών δράσης και για την οπτική αλλά και για την ακουστική ροή πληροφορίας συνενώνουμε τα αντίστοιχα αποτελέσματα με τις ground truth επισημειώσεις. Στόχος μας είναι να αποκτήσουμε πιο ακριβείς επισημειώσεις σε επίπεδο χειρονομιών για κάθε ροή πληροφορίας ανεξάρτητα. Με αυτό τον τρόπο αντιμετωπίζουμε το πρόβλημα το οποίο υπήρχε στις ground truth επισημειώσεις της βάσης δεδομένων, όπου στην αρχή και στο τέλος κάθε χειρονομίας μπορεί να είχαμε χρονικά τμήματα οπτικής ή φωνητικής μη-δράσης, ή άσχετων δράσεων.

Κεφάλαιο 7

Πειραματικά Αποτελέσματα

7.1 Ταξινόμηση χειρομορφών

7.1.1 Σώμα Δεδομένων και Επισημείωση των Χειρομορφών

Σε αυτή την ενότητα περιγράφουμε τη διαδικασία επισημείωσης των χειρομορφών που ακολουθήθηκε όπως και τα δεδομένα που χρησιμοποιήθηκαν στα πειράματα που ακολουθούν. Οι παράμετροι που επισημειώθηκαν σχετίζονται με το είδος και την πόζα των χειρομορφών.

Δεδομένα και Προδιαγραφές: Το σώμα δεδομένων BU400 Συνεχής ΑΝΓ [41] αποτελείται από 843 προτάσεις, 406 διαφορετικά νοήματα και 4 διαφορετικούς νοηματιστές. Το φόντο του βίντεο είναι ομοιόμορφο, οι εικόνες έχουν ανάλυση $648x484$ εικονοστοιχεία και η καταγραφή έχει γίνει σε 60 πλαίσια ανά δευτερόλεπτο. Στα πειράματα ταξινόμησης των χειρομορφών που θα παρουσιάσουμε στη συνέχεια, χρησιμοποιήσαμε τα δεδομένα από την εμπρόσθια κάμερα, για ένα νοηματιστή και της ιστορίας 'Accident'.

Παράμετροι Επισημείωσης Χειρομορφών: Οι παράμετροι που 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 τη σύμβαση που παρουσιάζεται στο τεχνικό κείμενο [90]. Για την επισημείωση της τρισδιάστατης πόζας χρησιμοποιήσαμε τα HamNoSys σύμβολα [56]. Πιο συγκεκριμένα οι παράμετροι που επισημειώθηκαν είναι οι εξής:

1. **Είδος Χειρομορφής -Handshape identity (HSId)-** που αντιστοιχεί στο είδος της χειρομορφής (*'A'*, *'B'*, *'1'*, *'C'* κ.α.), ανεξαρτήτως πόζας. Μερικά παραδείγματα απεικονίζονται στον Πίνακα 7.1.

2. **Τρισδιάστατη πόζα** η οποία αποτελείται από τις εξής παραμέτρους (βλ. Σχήμα. 7.1):

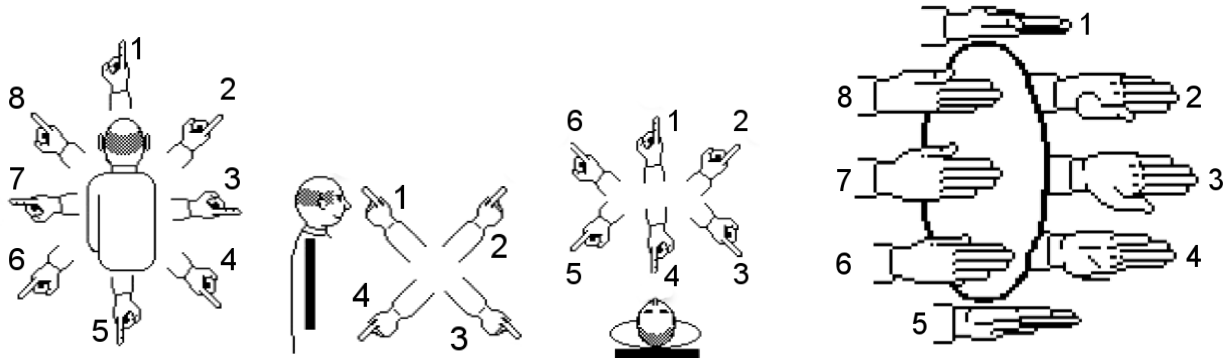
   (α') *Κατεύθυνση εκτεταμένων δακτύλων (Extended Finger Direction)* η οποία καθορίζει τον προσανατολισμό της χειρομορφής στον τρισδιάστατο χώρο και αποτελείται από την κατεύθυνση της χειρομορφής σε σχέση με τρία διαφορετικά επίπεδα: 1) Πρόσοψη -Front view (F)-, 2) Κάτοψη -Bird's view (B)- και 3) Πλάγια όψη -Side view (S)-.

   (β') *Προσανατολισμός της Παλάμης (P)* ο οποίος ορίζεται σε σχέση με την κάτοψη όπως απεικονίζεται στο Σχήμα 7.1(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. Στον Πίνακα 7.2 παρουσιάζουμε ενδεικτικά παραδείγματα από χειρομορφές που έχουν επισημειωθεί σε σχέση με το είδος τους και τον τρισδιάστατο προσανατολισμό τους.

HSId	*1*	*4*	*5*	*BL*	*cS*
					

Πίνακας 7.1: Είδος Χειρομορφής -Handshape identity (HSId)- αντιστοιχεί στο είδος της χειρομορφής ανεξαρτήτως πόζας: 5 ενδεικτικά παραδείγματα.



(α) Πρόσοψη (F) (β) Πλάγια όψη (S) (γ) Κάτοψη (B) (δ) Προσανατολισμός της Παλάμης (P)

Σχήμα 7.1: 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 του άρθρου [56] ορίζοντας αριθμητικές παραμέτρους για κάθε διαφορετικό προσανατολισμό.

7.1.2 Πειραματικά Αποτελέσματα

Σε αυτή την ενότητα παρουσιάζουμε τα πειραματικά αποτελέσματα βασιζόμενοι σε ένα στατιστικό σύστημα για την ταξινόμηση χειρομορφών. Αυτό βασίζεται στα εξής :

- Στην εξαγωγή χαρακτηριστικών που σχετίζονται με τις χειρομορφές εφαρμόζοντας τη Δυναμική Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος-Εμφάνισης την οποία περιγράψαμε στην ενότητα 2.3.2,
- Στην επισημείωση των παραμέτρων που χαρακτηρίζουν μια χειρομορφή την οποία περιγράψαμε στην ενότητα 7.1.1 και
- Στην επιλογή των δεδομένων και κλάσεων (βλ. ενότητα 7.1.1).

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 αξιολόγησή τους επαναλαμβάνοντας την παραπάνω διαδικασία 5 φορές. Τα υποσύνολα δεδομένων για την εκπαίδευση των μοντέλων και για την αξιολόγηση αποτελούνταν 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 τη μοντελοποίηση χρησιμοποιήσαμε GMMs με μια Γκαουσιανή κατανομή και διαγώνιο πίνακα συνδιακύμανσης. Τα GMMs αρχικοποιήθηκαν ομοιόμορφα και ο αλγόριθμος Baum-Welch χρησιμοποιήθηκε για την εκπαίδευσή τους [150]. Δεν χρησιμοποιήσαμε πιο σύνθετους ταξινομητές λόγω του ότι μας ενδιαφέρει η αξιολόγηση των εξαγόμενων χαρακτηριστικών για τις

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
3D hand pose S	0	0	0	3	1	0	2	2	0	2	0	0	0	0
3D hand pose B	0	0	0	6	4	0	1	1	0	6	0	0	0	0
3D hand pose 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
3D hand pose S	2	0	0	2	0	0	0	0	0	0	0	0	2	0
3D hand pose B	6	0	0	1	0	0	0	0	0	0	6	6	6	0
3D hand pose 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 7.2: Παραδείγματα από διαφορετικά είδη χειρομορφών και οι αντίστοιχες παράμετροι επισημείωσης. '# insts.' αντιστοιχεί στον αριθμό των εμφανίσεων στη βάση δεδομένων. Σε κάθε περίπτωση απεικονίζεται ένα ενδεικτικό πραγματικό παράδειγμα το οποίο αντιστοιχεί στο συγκεκριμένο είδος χειρομορφής με τον συγκεκριμένο τρισδιάστατο προσανατολισμό.

χειρομορφές και όχι η αξιολόγηση ενός πιο σύνθετου ταξινομητή. Επιπλέον τα GMMs ταιριάζουν με το συνολικό HMM-based σύστημα για την αναγνώριση νοηματικής γλώσσας [134, 2].

Τα πειράματα που υλοποιήσαμε χαρακτηρίζονται από τη 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.

Βάση Δεδομένων (DS)

Υλοποιήσαμε πειράματα σε τρεις βάσεις δεδομένων οι οποίες είναι οι εξής:

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.

Class Dependency (CD)

Το είδος εξάρτησης των κλάσεων σε σχέση με τις παραμέτρους επισημείωσης της χειρομορφής (βλ. Πίνακα 7.3). Ας πάρουμε το παράδειγμα όπου οι κλάσεις μας εξαρτώνται από την παράμετρο F. Τότε εκπαιδεύουμε διαφορετικά μοντέλα για κάθε τιμή της παραμέτρου F. Στην αντίθετη περίπτωση όπου οι κλάσεις μας *δεν* εξαρτώνται από την παράμετρο F εκπαιδεύουμε ένα μοντέλο,

το οποίο ουσιαστικά μοντελοποιεί την μεταβλητότητα της παραμέτρου F. 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.

Μέθοδος Εξαγωγής Χαρακτηριστικών

Οι μέθοδοι που χρησιμοποιήθηκαν για την εξαγωγή χαρακτηριστικών είναι οι εξής:

1. ***Αφινικά Αναλλοίωτη Μοντελοποίηση Σχήματος- Εμφάνισης (Aff-SAM)***: Αντιστοιχεί στην προτεινόμενη μέθοδο την οποία περιγράψαμε στην ενότητα 2.3.2.

2. ***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 (βλ. ενότητα 2.3.2) 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 ταίριασμα του μοντέλου εφαρμόζεται πάνω στην εικόνα ΜΣΕ η οποία περιέχει τους αρθρωτές που βρίσκονται σε επικάλυψη. Αυτή εφαρμόζεται χωρίς την χρησιμοποίηση της στατικής και δυναμικής πρότερη γνώσης των παραμέτρων του ΜΣΕ. Η παραπάνω μέθοδος είναι παρόμοια με την προσέγγιση του άρθρου [14].

3. ***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 στη δυαδική μάσκα του χεριού. Επιπλέον και σε αυτήν τη μέθοδο δεν γίνεται χρήση της στατικής και δυναμικής πρότερης γνώσης των παραμέτρων του ΜΣΕ. Η παραπάνω μέθοδος είναι παρόμοια με την προσέγγιση των άρθρων [33, 144].

4. ***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 [23, 26]. 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.

5. ***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 [61].

6. ***Γεωμετρικά Χαρακτηριστικά (RB)***: Αποτελούνται από τα γεωμετρικά χαρακτηριστικά του χεριού τα οποία είναι το εμβαδόν, η εκκεντρότητα, ο βαθμός συμπύκνωσης, μήκη του ελάσσονος και μείζονος άξονος του χεριού [1].

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

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

Αξιολόγηση του Χώρου Χαρακτηριστικών

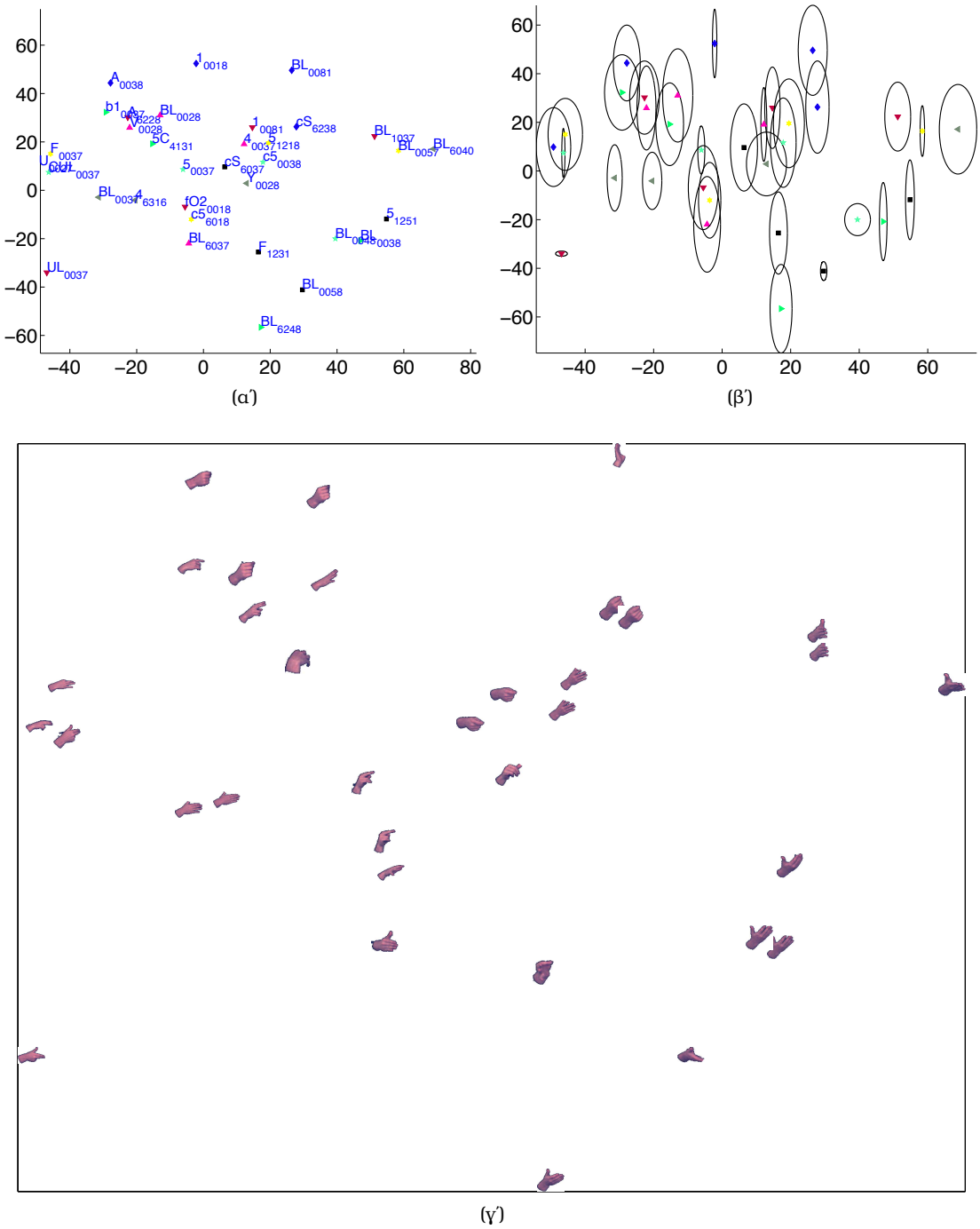
Σε αυτή την ενότητα θα ασχοληθούμε με την ανάλυση και αξιολόγηση του χώρου χαρακτηριστικών για τη μέθοδο εξαγωγής χαρακτηριστικών Aff-SAM. Για την οπτικοποίηση του χώρου χαρακτηριστικών προβάλλουμε τα διανύσματα χαρακτηριστικών στο επίπεδο $\lambda_1 - \lambda_2$ όπου λ_1 και λ_2 αντιστοιχούν στις πρώτες δύο ιδιοκατευθύνσεις. Στο Σχήμα 7.2 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 7.3). Στο Σχήμα 7.2α΄ απεικονίζουμε τα κεντροειδή των μοντέλων μαζί με την ετικέτα της κάθε κλάσης. Επιπλέον στο Σχήμα 7.2β΄ απεικονίζουμε τα ίδια κεντροειδή μαζί με τις ελλείψεις οι οποίες αντιπροσωπεύουν τις αντίστοιχες πυκνότητες πιθανότητας των μοντέλων. Τέλος, στο Σχήμα 7.2γ΄ 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 στο χώρο των χαρακτηριστικών. Αντιθέτως μοντέλα που αντιπροσωπεύουν πολύ διαφορετικές χειρομορφές βρίσκονται μακριά. Ο απεικονιζόμενος χώρος χαρακτηριστικών της μεθόδου Aff-SAM διαχωρίζει σε μεγάλο βαθμό χειρομορφές διαφορετικού είδους σε σύγκριση με τους χώρους χαρακτηριστικών από τις υπόλοιπες μεθόδους εξαγωγής χαρακτηριστικών.

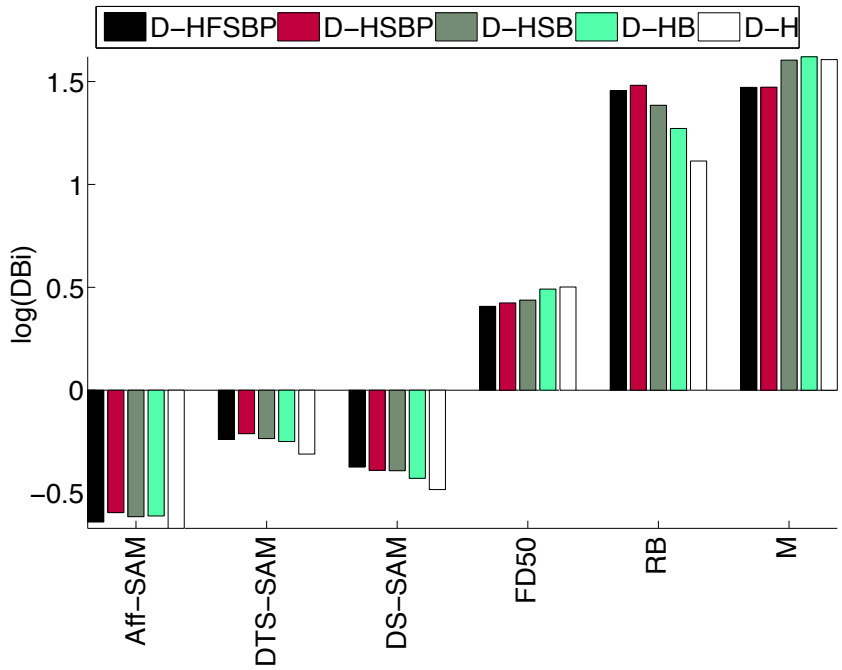
Για να υποστηρίξουμε το παραπάνω επιχείρημα ποσοτικοποιήσαμε τον διαχωρισμό διαφορετικών χειρομορφών για όλες τις μεθόδους εξαγωγής χαρακτηριστικών. Χρησιμοποιήσαμε τον δείκτη Davies-Boulding [35], ο οποίος ποσοτικοποιεί την ποιότητα και διακριτική ικανότητα ενός χώρου χαρακτηριστικών. Το Σχήμα 7.3 απεικονίζει τις μετρήσεις για κάθε μέθοδο εξαγωγής χαρακτηριστικών μεταβάλλοντας την εξάρτηση των μοντέλων στις παραμέτρους προσανατολισμού της χειρομορφής όπως περιγράψαμε στην ενότητα 7.1.2 και στον Πίνακα 7.3. Παρατηρούμε ότι οι δείκτες DB για την μέθοδο Aff-SAM είναι μικρότεροι σε σχέση με τις άλλες μεθόδους γεγονός το οποίο υποδεικνύει 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 στη μεταβολή της εξάρτησης των μοντέλων από τις παραμέτρους του προσανατολισμού. Αυτό υποδεικνύει ότι τα εκπαιδευμένα μοντέλα αντιμετωπίζουν επιτυχώς τη μεταβλητότητα στην πόζα των χειρομορφών.

7.1.3 Πειράματα Ταξινόμησης

Για την αξιολόγηση των αποτελεσμάτων ταξινόμησης εφαρμόσαμε διασταυρωμένη επικύρωση. Σε κάθε πείραμα απεικονίζουμε τη μέση τιμή και τυπική απόκλιση των πειραματικών αποτελεσμάτων. Στον Πίνακα 7.4 παρουσιάζουμε μια σύνοψη των μέσων αποτελεσμάτων από όλα τα διαφορετικά πειράματα που υλοποιήσαμε στις τρεις διαφορετικές βάσεις δεδομένων που περιγρά-



Σχήμα 7.2: Χώρος χαρακτηριστικών του πειράματος D-HFSBP όπου τα μοντέλα εξαρτώνται από όλες τις παραμέτρους επισημείωσης με τη χρήση της μεθόδου Aff-SAM. Για λόγους οπτικοποίησης προβάλουμε 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 τη διασπορά τους. (γ) Πραγματικές κομμένες εικόνες χειρομορφών για κάθε κεντροειδές.

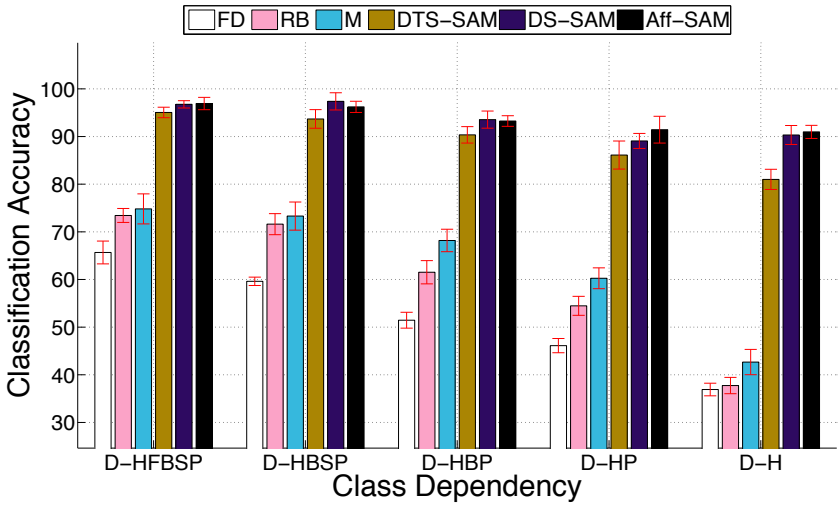


Σχήμα 7.3: Davies-Bouldin Index (DBi) σε λογαριθμική κλίμακα (άξονας y) για όλες τις μεθόδους εξαγωγής χαρακτηριστικών μεταβάλλοντας την εξάρτηση των κλάσεων στις παραμέτρους επισημείωσης.

Πείραμα	Μέθοδος	Σώμα Δεδομένων	Εξ. Κλάσεων	Επικ.	Αποτελέσματα		
					#HSIds	Avg. Acc. %	Std.
DS-1	Aff-SAM	DS-1	Πίνακας 7.3	✗	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 7.3	✓	42	**74.9**	0.9
	DS-SAM			✓	42	66.1	1.1
	DTS-SAM			✓	42	62.7	1.4

Πίνακας 7.4: Σύνοψη των πειραματικών αποτελεσμάτων, μεταβάλλοντας το σώμα δεδομένων, την εξάρτηση των κλάσεων και τη μέθοδο εξαγωγής των χαρακτηριστικών. Όπου η Εξ. Κλάσεων αντιστοιχεί στην εξάρτηση των κλάσεων σε σχέση με τις 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 περιλαμβάνονται στο σώμα δεδομένων. Avg. Acc. είναι ο μέσος όρος αναγνώρισης και Std είναι η τυπική απόκλιση.

ψαμε προηγουμένως. Οι μέσοι όροι των αποτελεσμάτων υπολογίστηκαν για τα πέντε διαφορετικά σύνολα δεδομένων εκπαίδευσης και αξιολόγησης (cross-validation). Επιπλέον μεταβάλλουμε 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 απεικονίζεται είναι για όλες τις διαφορετικές περιπτώσεις εξάρτησης των κλάσεων



Σχήμα 7.4: Πειράματα ταξινόμησης για περιπτώσεις χωρίς επικαλύψεις στη 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 7.5.

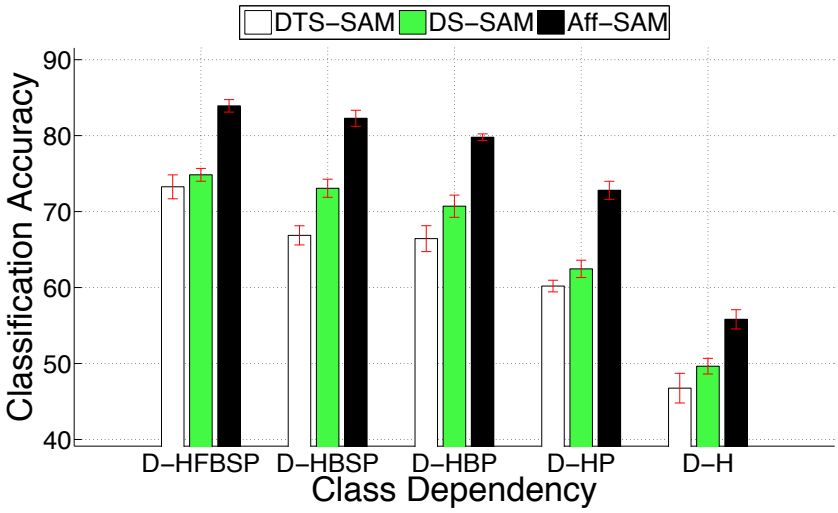
	Πόζα & Εξάρτηση των Κλάσεων				
Πειρ.	D-HFSBP	D-HSBP	D-HBP	D-HP	D-H
#	34	33	33	31	18

Πίνακας 7.5: Περιπτώσεις χειρομορφών *χωρίς* επικαλύψεις. Αριθμός κλάσεων για κάθε πείραμα στη βάση δεδομένων DS-1.

όπως περιγράφονται στον Πίνακα 7.3. Στην περίπτωση 'DS-1-extend' χρησιμοποιούμε εξάρτηση κλάσεων D-H, με στόχο να αυξήσουμε την μεταβλητότητα κάθε κλάσης.

Σώμα Δεδομένων DS-1: Στο Σχήμα 7.4 συγκρίνουμε τις διαφορετικές μεθόδους για την εξαγωγή χαρακτηριστικών, μεταβάλλοντας την εξάρτηση των μοντέλων στις παραμέτρους επισημείωσης του προσανατολισμού της χειρομοφής (άξονας x). Χρησιμοποιήσαμε τη βάση δεδομένων *DS-1* η οποία αποτελείται από 18 διαφορετικά είδη χειρομορφών από περιπτώσεις όπου δεν έχουμε επικαλύψεις. Ο αριθμός των κλάσεων μεταβάλλεται ανάλογα με το είδος εξάρτησης στις παραμέτρους επισημείωσης του προσανατολισμού και απεικονίζεται στον Πίνακα 7.5 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 7.4 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 τη μεταβολή της εξάρτησης των κλάσεων στις παραμέρους επισημείωσης του προσανατολισμού. Το παραπάνω ισχυροποιεί την παρατήρηση που έγινε στην προηγούμενη ενότητα ότι η Aff-SAM μέθοδος αντιμετωπίζει επιτυχώς την μικρή μεταβλητότητα στην τρισδιάστατη πόζα



Σχήμα 7.5: Πειράματα ταξινόμησης με και χωρίς επικαλύψεις στη 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 7.6.

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

Πίνακας 7.6: Περιπτώσεις χειρομορφών με και χωρίς επικαλύψεις. Αριθμός κλάσεων για κάθε πείραμα στη βάση δεδομένων DS-2.

των χειρομορφών. Μια σύνοψη των αποτελεσμάτων μπορούμε να δούμε στην πρώτη γραμμή του πίνακα 7.4 (DS-1).

Σώμα Δεδομένων DS-1-extend: Στη συνέχεια παρουσιάζουμε αποτελέσματα χρησιμοποιώντας τη 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. Στη δεύτερη γραμμή του πίνακα 7.4 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. Έτσι, παρατηρούμε ότι σε αρκετά πιο δύσκολα προβλήματα ταξινόμησης (μεγαλύτερο αριθμό από HSId και μεγαλύτερη μεταβλητότητα της τρισδιάστατης πόζας) η Aff-SAM υπερτερεί των μεθόδων DS-SAM και DTS-SAM.

Σώμα Δεδομένων DS-2: Στη συνέχεια παρουσιάζουμε αποτελέσματα χρησιμοποιώντας τη 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 απεικονίζεται στον πίνακα 7.6.

Στο Σχήμα 7.5 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 τη 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 να γίνεται αρκετά δύσκολο. Σε αυτό το πείραμα το εύρος της μεταβλητότητας είναι αρκετά μεγαλύτερο από αυτό που μπορεί να αντιμετωπιστεί με επιτυχία από τους αφινικούς μετασχηματισμούς του Aff-SAM μοντέλου. Επιπλέον, κατά τη διάρκεια επικαλύψεων εμφανίζονται λάθη στην οπτική μοντελοποίηση για την εκτίμηση της χειρομορφής που βρίσκεται σε επικάλυψη τα οποία μεταφέρονται και στην ταξινόμηση χειρομορφών.

7.2 Αναγνώριση νοημάτων με δεδομενοκεντρικές υπομονάδες

Σε αυτή την ενότητα παρουσιάζουμε τα πειραματικά αποτελέσματα αναγνώρισης ΝΓ κάνοντας χρήση των δεδομενοκεντρικών υπομονάδων που παρουσιάστηκαν στο κεφάλαιο 3. Πειραματιζόμαστε σε τρεις διαφορετικές βάσεις δεδομένων (GSL-Lem, ASLLVD, BU400), και συγκρίνουμε με άλλες μεθόδους που έχουν προταθεί στη διεθνή βιβλιογραφία.

Οι μέθοδοι που χρησιμοποιούμε για σύγκριση είναι οι ακόλουθες:

1. Η μέθοδος SU-noDSC είναι μια παραλλαγή της 2-S-U. Χρησιμοποιεί ένα εργοδικό HMM δύο καταστάσεων για την κατάτμηση σε χρονικά τμήματα όπως ακριβώς και η 2-S-U μέθοδος. Παρόλα αυτά δεν διαχωρίζει σε δυναμικά και στατικά τμήματα. Όλα τα χρονικά τμήματα είναι ίδιου τύπου. Συνεπώς, σε κάθε χρονικό τμήμα χρησιμοποιείται το ίδιο διάνυσμα χαρακτηριστικών. Για την κατασκευή των υπομονάδων και του αντίστοιχου λεξικού υπομονάδων, εφαρμόζεται συσταδοποίηση όλων των χρονικών τμημάτων, κάνοντας χρήση του αλγορίθμου DTW.

2. Η μέθοδος Fang et al. 2004 (SU-Segm) [47], χρησιμοποιεί ένα left-right HMM τριών καταστάσεων για την κατάτμηση σε τμήματα. Λαμβάνει υπόψη της ολόκληρα χρονικά τμήματα για την κατασκευή των υπομονάδων, όπως οι μέθοδοι SU-noDSC και 2-S-U. Για την κατασκευή των υπομονάδων και του αντίστοιχου λεξικού υπομονάδων, εφαρμόζεται συσταδοποίηση όλων των χρονικών τμημάτων, κάνοντας χρήση του αλγορίθμου DTW αντίστοιχα με την SU-noDSC μέθοδο.

3. Η μέθοδος Bauer and Kraiss 2001 (SU-Frame) [11], για την κατασκευή των υπομονάδων και του αντίστοιχου λεξικού υπομονάδων, βασίζεται στη συσταδοποίηση μεμονομένων χρονικών πλαισίων και όχι ολόκληρων χρονικών τμημάτων, χρησιμοποιώντας τον αλγόριθμο K-means.

4. Η μέθοδος Wang et al. 2010 (Sign-DTW) [138] exemplar-based, για κάθε νόημα στο λεξικό κατασκευάζει πολλαπλές τεμπλέτες (templates). Για την αναγνώριση βασίζεται στον υπολογισμό της απόστασης του νοήματος προς αναγνώριση, με όλες τις κατασκευασμένες τεμπλέτες χρησιμοποιώντας τον αλγόριθμο DTW. Η τελική αναγνώριση γίνεται επιλέγοντας το νόημα με την μικρότερη απόσταση.

Στις μεθόδους (1-3), κάθε υπομονάδα εκπαιδεύεται στατιστικά χρησιμοποιώντας HMMs, ενώ δεν γίνεται διαχωρισμός μεταξύ στατικών και δυναμικών υπομονάδων όπως στις μεθόδους 2-S-U και RAW. Περισσότερες λεπτομέρειες σε σχέση με τις μεθόδους που χρησιμοποιήσαμε για σύγκριση αναφέρονται στην ενότητα 1.2

Για την ανίχνευση και παρακολούθηση των αρθρωτών χρησιμοποιήθηκε το σύστημα που παρουσιάστηκε στο κεφάλαιο 2. Οι ροές πληροφορίας που χρησιμοποιήθηκαν σχετίζονται με την κίνηση (M) και τη θέση (P) των δύο χεριών του νοηματιστή και με τη χειρομορφή (HS) του κυρίαρχου χεριού. Η συνδυασμένη ροή πληροφορίας της κίνησης-θέσης συμβολίζεται με M-P. Για τη ροή της κίνησης-θέσης χρησιμοποιήθηκε το διάνυσμα χαρακτηριστικών που περιλαμβάνει:

- τις συντεταγμένες των χεριών κανονικοποιημένες με τη θέση του κεφαλιού ($\boldsymbol{P}$),
- την στιγμιαία κατεύθυνση της κίνησής τους ($\boldsymbol{D}$),
- την ταχύτητά τους ($\boldsymbol{V}$) και
- την απόστασή τους ($\boldsymbol{L}$).

Για τη ροή της χειροφορφής χρησιμοποιήσαμε το διάνυσμα χαρακτηριστικών που προκύπτει από την εφαρμογή της μεθόδου spatial pyramids. Περισσότερες λεπτομέρειες σε σχέση με τα διανύσματα χαρακτηριστικών αναφέρονται στο κεφάλαιο 2.

Για τις προτεινόμενες μεθόδους (2-S-U, RAW) η ροή πληροφορίας της κίνησης-θέσης χρησιμοποιείται ως εξής: στις δυναμικές υπομονάδες χρησιμοποιείται η ροή της κίνησης και στις στατικές υπομονάδες η ροή της θέσης. Επιπλέον η σύμμειξη του κυρίαρχου και δευτερεύοντος χεριού γίνεται χρησιμοποιώντας την προσέγγιση που περιγράψαμε στην ενότητα 5.2.1. Αντίθετα σε όλες τις άλλες μεθόδους υπομονάδων (SU-noDSC, SU-Segm, SU-Frame) δεν γίνεται διαχωρισμός σε στατικές και δυναμικές υπομονάδες. Έτσι, σε όλες τις υπομονάδες χρησιμοποιείται το ίδιο διάνυσμα χαρακτηριστικών. Συνενώνουμε λοιπόν τις ροές πληροφορίας της θέσης και κίνησης για το κυρίαρχο και δευτερεύον χέρι σε ένα κοινό διάνυσμα χαρακτηριστικών. Για τη μοντελοποίηση της ροής M-P εφαρμόσαμε σε κάθε μέθοδο την μοντελοποίηση που χρησιμοποιήθηκε στην κάθε δημοσίευση. Αντίθετα για την μοντελοποίηση της ροής HS χρησιμοποιήσαμε την ίδια μοντελοποίηση σε όλες τις μεθόδους όπως αυτή περιγράφεται στην ενότητα 3.4.3.

Για τη σύμμειξη των ροών πληροφορίας της κίνησης-θέσης και χειρομορφής για όλες τις μεθόδους χρησιμοποιήθηκαν τα PaHMM όπως περιγράφεται στην ενότητα 5.2.2. Η σύμμειξη των M-P και HS ροών πληροφορίας στις μεθόδους SU-Segm και SU-Frame, όπως περιγράφεται στις αντίστοιχες δημοσιεύσεις, έγινε με απλή συνένωση των χαρακτηριστικών διανυσμάτων. Όμως αυτός ο τρόπος σύμμειξης οδηγούσε σε χαμηλότερα αποτελέσματα αναγνώρισης σε σύγκριση με τη χρήση των PaHMM. Έτσι, με στόχο την δίκαια σύγκριση των μεθόδων αυτών με τις προτεινόμενες, χρησιμοποιήσαμε τελικά τα PaHMM για τη σύμμειξή τους.

7.2.1 Βάση δεδομένων GSL Lemmas Corpus (GSL-Lem)

Σε αυτή την ενότητα παρουσιάζουμε τα πειραματικά αποτελέσματα στη βάση δεδομένων GSL Lemmas Corpus (GSL-Lem). Η βάση δεδομένων GSL Lemmas Corpus (GSL-Lem) [37] αποτελείται από 1046 διαφορετικά νοήματα με πέντε επαναλήψεις το κάθε ένα από δύο διαφορετικούς νοηματιστές (τον 'Κώστα' και την 'Όλγα').

Η πειραματική αξιολόγηση αποτελείται από τα ακόλουθα σενάρια:

1) Πειραματική αξιολόγηση σε ένα νοηματιστή. Με άλλα λόγια, δεδομένα από τον ίδιο νοηματιστή έχουν χρησιμοποιηθεί και κατά τη διαδικασία εκπαίδευσης αλλά και κατά την αξιολόγηση.

2) Πειραματική αξιολόγηση σε άγνωστο νοηματιστή. Με άλλα λόγια χρησιμοποιήθηκαν δεδομένα από ένα νοηματιστή Α κατά τη διαδικασία εκπαίδευσης και δεδομένα από έναν διαφορετικό/άγνωστο νοηματιστή Β κατά τη διάρκεια της αξιολόγησης.

3) Πειραματική αξιολόγηση μετά από προσαρμογή στον άγνωστο νοηματιστή. Η κατηγορία αυτή των πειραμάτων είναι ανάμεσα στις δύο προηγούμενες. Κατά τη διάρκεια της εκπαίδευσης έχουμε χρησιμοποιήσει δεδομένα μόνο από τον νοηματιστή Α, όμοια με την περίπτωση (2). Επιπλέον όμως χρησιμοποιούμε ένα μικρό σύνολο δεδομένων από τον νοηματιστή Β προς αναγνώριση, με στόχο την προσαρμογή των μοντέλων και του λεξικού στον νοηματιστή Β. Η διαδικασία προσαρμογής που χρησιμοποιήθηκε περιγράφεται λεπτομερώς στην ενότητα 5.3.

Αριθμός υπομονάδων: Στα πειραματικά αποτελέσματα που ακολουθούν ορίσαμε τον αριθμό των υπομονάδων μεγιστοποιώντας το ποσοστό αναγνώρισης σε ένα μικρό σύνολο δεδομένων, που αποτελεί το 20% του συνόλου των δεδομένων, μη-επικαλυπτόμενο με τα δεδομένα που χρησιμοποιήθηκαν για την τελική αξιολόγηση του συστήματος. Για την μέθοδο 2-S-U χρησιμοποιήσαμε 10 στατικές υπομονάδες, 30 δυναμικές και 500 υπομονάδες χειρομορφής. Οι υπομονάδες χειρομορφής μοντελοποιούν ταυτόχρονα και το είδος της χειρομορφής αλλά και την προβολή της τρισδιάστατης πόζας τους, μιας και επεξεργαζόμαστε δισδιάστατες εικόνες. Για τις μεθόδους SU-noDSC, SU-Segm, και SU-Frame χρησιμοποιήσαμε 150, 300 και 150 υπομονάδες αντίστοιχα για την ροή της κίνησης-θέσης και 500 υπομονάδες χειρομορφής.

Πίνακας 7.7: Αξιολόγηση αναγνώρισης σε ένα νοηματιστή και 984 νοήματα από την βάση δεδομένων GSL-Lem. Τα αποτελέσματα είναι σε sign accuracy %.

2-S-U	RAW	SU-Frame	MC	SPs	SU-Segm	Sign-DTW
96.98	96.9	96.2	71.4	74.1	96.2	99

Πίνακας 7.8: Αξιολόγηση αναγνώρισης σε άγνωστο νοηματιστή και 300 νοήματα από την βάση δεδομένων GSL-Lem. Τα αποτελέσματα είναι σε sign accuracy %.

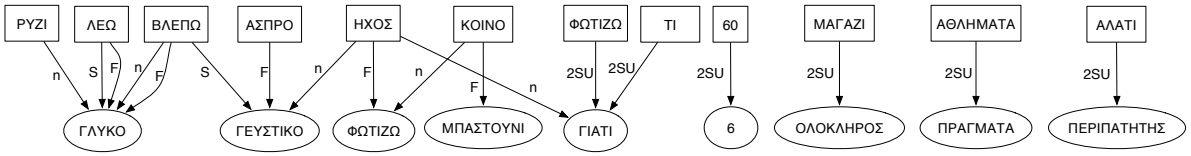
Νοηματιστής	Ροή	2-S-U	RAW	SU-noDSC	SU-Segm	SU-Frame	Sign-DTW
Όλγα	M-P	30.1	22.2	11.3	14.23	11.4	25.8
	HS	38.8	38.8	38.8	38.8	38.8	42.2
	M-P+HS	**61.2**	52.4	46.6	54.4	40.53	57.9
Κώστας	M-P	29	23.3	11.8	9.1	11.9	24.4
	HS	28.8	28.8	28.8	28.8	28.8	32.7
	M-P+HS	**50.1**	42.3	33.2	32.6	35.53	46.3

Αξιολόγηση σε ένα νοηματιστή

Εδώ παρουσιάζουμε πειραματικά αποτελέσματα χρησιμοποιώντας έναν κοινό νοηματιστή και κατά την εκπαίδευση και κατά την αξιολόγηση του συστήματος. Το λεξιλόγιο που χρησιμοποιήθηκε αποτελείται από 984 νοήματα. Η μείωση του αριθμού των νοημάτων οφείλεται σε λάθη στην παρακολούθηση των χεριών του νοηματιστή σε 62 νοήματα, με αποτέλεσμα να αφαιρεθούν από το σύνολο των δεδομένων. Τα δεδομένα διαμοιράστηκαν τυχαία σε δύο σύνολα δεδομένων. Το ένα χρησιμοποιήθηκε για την εκπαίδευση του συστήματος και το άλλο για την αξιολόγησή του. Το σύνολο δεδομένων που χρησιμοποιήθηκε κατά την εκπαίδευση περιελάμβανε τέσσερις επαναλήψεις κάθε νοήματος. Ενώ αυτό που χρησιμοποιήθηκε κατά την αξιολόγηση περιελάμβανε μια επανάληψη ανά νόημα. Περισσότερες λεπτομέρειες για τον παραπάνω διαχωρισμό αναφέρονται στην ιστοσελίδα [125].

Επιπλέον των μεθόδων που παρουσιάστηκαν προηγουμένως απεικονίζουμε αποτελέσματα από ακόμα δύο μεθόδους. Η πρώτη βασίζεται σε Markov Chains (MC) και η δεύτερη σε Sequential Patterns (SPs) [28]. Για τις δύο αυτές μεθόδους (MC και SPs) απεικονίζουμε τα αποτελέσματα αναγνώρισης που παρουσιάστηκαν στο άρθρο [28]. Οι συγγραφείς χρησιμοποίησαν την ίδια βάση δεδομένων, το ίδιο λεξιλόγιο και τα ίδια αποτελέσματα της ανίχνευσης και παρακολούθησης των χεριών του νοηματιστή. Έτσι τα αποτελέσματα είναι άμεσα συγκρίσιμα.

Ο Πίνακας 7.7 απεικονίζει τα αποτελέσματα αναγνώρισης χρησιμοποιώντας το σύνολο των ροών πληροφορίας. Όπως παρατηρούμε οι μέθοδοι 2-S-U, RAW, SU-Segm, και SU-Frame επιτυγχάνουν παρόμοια αποτελέσματα αναγνώρισης. Επιπλέον, οι προτεινόμενες μέθοδοι 2-S-U και RAW υπερτερούν των μεθόδων MC και SPs, αυξάνοντας το ποσοστό αναγνώρισης κατά 25.5% και 22.8% αντίστοιχα. Ακόμα, η Sign-DTW μέθοδος υπερτερεί κατά 2% των μεθόδων 2-S-U και RAW. Αυτό που πρέπει να έχουμε όμως κατά νου είναι ότι το συγκεκριμένο πείραμα είναι σε ένα νοηματιστή. Η χρησιμοποίηση πολλαπλών νοηματιστών αυξάνει την ποικιλία άρθρωσης των νοημάτων και άρα το πρόβλημα αναγνώρισης δυσκολεύει. Επιπλέον, αξιολογείται η δυνατότητα γενίκευσης ενός συστήματος αναγνώρισης σε άγνωστο νοηματιστή. Για τον λόγο αυτό στη συνέχεια παρουσιάζουμε πειραματικά αποτελέσματα κάνοντας αξιολόγηση σε άγνωστο νοηματιστή.



Σχήμα 7.6: Απεικονίζεται ο γράφος λαθών. Οι τετράγωνοι κόμβοι αντιστοιχούν στα νοήματα προς αναγνώριση ενώ οι ελλείψεις στο αποτέλεσμα της αναγνώρισης. Οι ακμές υποδεικνύουν ενδεικτικά λάθη αναγνώρισης χρησιμοποιώντας μία από τις μεθόδους: SU-noDSC (n), SU-Frame (F), SU-Segm (S). Τα ίδια νοήματα αναγνωρίστηκαν σωστά από τη μέθοδο 2-S-U. Οι ακμές με ταμπέλα 2SU υποδεικνύουν ενδεικτικά λάθη της μεθόδου 2-S-U.

Αξιολόγηση σε άγνωστο νοηματιστή

Εδώ παρουσιάζουμε πειραματικά αποτελέσματα κάνοντας αξιολόγηση σε άγνωστο νοηματιστή. Με άλλα λόγια, κατά τη διαδικασία εκπαίδευσης χρησιμοποιήθηκαν δεδομένα από ένα νοηματιστή Α. Ενώ κατά τη διάρκεια της αξιολόγησης χρησιμοποιήθηκαν δεδομένα από ένα διαφορετικό/άγνωστο νοηματιστή Β. Το λεξιλόγιο αποτελείται από 300 διαφορετικά νοήματα. Στον Πίνακα 7.8 απεικονίζουμε τα αποτελέσματα αναγνώρισης μεταβάλλοντας τις μεθόδους και τις ροές πληροφορίας που έχουν χρησιμοποιηθεί.

Ροή κίνησης-θέσης (M-P): Παρατηρώντας τον Πίνακα 7.8 βλέπουμε ότι η μέθοδος 2-S-U υπερτερεί της μεθόδου SU-noDSC αυξάνοντας το ποσοστό αναγνώρισης, 18% κατά μέσον όρο και για τους δύο νοηματιστές. Αυτή η στοχευμένη σύγκριση, υποδεικνύει ότι ο διαχωρισμός σε δυναμικές και στατικές υπομονάδες μαζί με την χρησιμοποίηση του MSSD -HMM πλαισίου για τη μοντελοποίηση των υπομονάδων, επηρεάζει δραστικά το ποσοστό αναγνώρισης. Επιπλέον, συγκρίνοντας τη μέθοδο 2-S-U με τις μεθόδους RAW, SU-Segm και SU-Frame, βλέπουμε ότι η 2-S-U μέθοδος επιτυγχάνει καλύτερα αποτελέσματα αναγνώρισης. Συγκεκριμένα, αυξάνει το ποσοστό αναγνώρισης κατά 6.8%, 17.8% και 17.9 αντίστοιχα και για τους δύο νοηματιστές κατά μέσο όρο. Τέλος, συγκρίνοντας τη μέθοδο 2-S-U με την Sign-DTW μέθοδο η οποία βασίζεται στη μοντελοποίηση σε επίπεδο νοήματος, το ποσοστό αναγνώρισης αυξάνει κατά 4.5% και για τους δύο νοηματιστές κατά μέσο όρο.

Υπόλοιπες Ροές (HS και M-P+HS): Όταν κάνουμε χρήση μόνο της ροής της χειρομορφής (HS) παρατηρούμε ότι όλες οι μέθοδοι που βασίζονται στην μοντελοποίηση σε επίπεδο υπομονάδας (2-S-U, RAW SU-noDSC, SU-Segm, και SU-Frame) έχουν το ίδιο ποσοστό αναγνώρισης. Αυτό οφείλεται στο ότι έχει χρησιμοποιηθεί ακριβώς η ίδια τεχνική μοντελοποίησης για τη ροή της χειρομορφής σε όλες τις μεθόδους. Τέλος, λαμβάνοντας υπόψη όλες τις ροές πληροφορίας (M-P+HS) με τη χρήση των PaHMMs παρατηρούμε ότι η μέθοδος 2-S-U υπερτερεί όλων των υπολοίπων. Συγκεκριμένα, το ποσοστό αναγνώρισης αυξάνει και για τους δύο νοηματιστές κατά μέσο όρο: 7.85% από την RAW, 15.8% από την SU-noDSC, 12.1% από την SU-Segm, 17.6% από την SU-Frame, και 3.5% από την Sign-DTW.

Λάθη κατά την αναγνώριση: Στο Σχήμα 7.6 απεικονίζουμε ενδεικτικά λάθη κατά την αναγνώριση, χρησιμοποιώντας διαφορετικές μεθόδους αναγνώρισης, σχετικά με τα πειραματικά αποτελέσματα που παρουσιάστηκαν προηγουμένως.

Αρχικά εστιάζουμε σε νοήματα όπου η μέθοδος 2-S-U αναγνώρισε σωστά, ενώ οι άλλες μέθοδοι λάθος. Όπως παρατηρούμε οι μέθοδοι που δεν κάνουν τον διαχωρισμό μεταξύ στατικών και δυναμικών υπομονάδων συγχέουν:

- Νοήματα τα οποία διαφέρουν λόγω της εισαγωγής μιας επιπλέον στάσης. Παραδείγματος χάριν τα νοήματα 'ΡΥΖΙ', 'ΛΕΩ', και 'ΒΛΕΠΩ' (βλ. Σχήματα 3.1ε', 3.14 και 3.1ς'), περιέχουν μια ρητή στάση στην ουδέτερη περιοχή του νοηματικού χώρου. Αυτά αναγνωρίζονται

Πίνακας 7.9: Προσαρμογή σε νοηματιστή χρησιμοποιώντας ένα σύνολο προσαρμογής από τον νοηματιστή προς αναγνώριση. Αποτελέσματα σε sign accuracy σε 300 νοήματα της βάσης δεδομένων GSL-Lem.

	Νοηματιστής	Όλγα			Κώστας		
	Ροή	M-P	HS	M-P+HS	M-P	HS	M-P+HS
Μέθοδοι	2-S-U	30.9	40.6	58.6	26.1	34	49.6
	2-S-U+MLLR	33	54	67.9	28.8	59.2	71.7
	2-S-U+MLLR+IP	67.3	89.1	**92.7**	63.1	88.2	**92.3**

λανθασμένα στο νόημα 'ΓΛΥΚΟ' το οποίο δεν περιέχει τη στάση αυτή. Αυτό οφείλεται στο γεγονός ότι δεν υπάρχει συγκεκριμένη υπομονάδα η οποία να μοντελοποιεί ρητά αυτήν τη στάση όπως στην περίπτωση της μεθόδου 2-S-U.

- Νοήματα τα οποία διαφέρουν λόγω της εισαγωγής μιας επιπλέον κίνησης. Παραδείγματος χάριν το νόημα 'ΗΧΟΣ' εμπεριέχει μια μικρή κίνηση η οποία δεν μοντελοποιείται ρητά από μια συγκεκριμένη υπομονάδα για τις μεθόδους SU-noDSC, SU-Segm, και SU-Frame. Αυτό έχει ως αποτέλεσμα να αναγνωρίζεται λανθασμένα στα νοήματα 'ΓΕΥΣΤΙΚΟ', 'ΦΩΤΙΖΩ' και 'ΓΙΑΤΙ' αντίστοιχα.

- Νοήματα τα οποία εκτελούνται από ένα ή δύο χέρια. Η έλλειψη διαχωρισμού σε Δ/Σ υπομονάδες μπορεί να οδηγήσει σε λανθασμένη αναγνώριση μεταξύ όμοιων νοημάτων όπου το ένα εκτελείται μόνο από το ένα χέρι και το άλλο και από τα δύο. Ένα παράδειγμα είναι το νόημα 'ΚΟΙΝΟ' το οποίο λανθασμένα αναγνωρίζεται στο νόημα 'ΦΩΤΙΖΩ' λόγω του ότι και τα δύο αρθρώνονται χρησιμοποιώντας την ίδια χειρομορφή.

Επιπλέον παρουσιάζουμε λάθη της μεθόδου 2-S-U. Συνήθη λάθη εμφανίζονται σε νοήματα τα οποία περιέχουν πολύ μικρές κινήσεις, οι οποίες δεν μπορούν να ανιχνευτούν και επομένως δεν μπορούν να μοντελοποιηθούν όπως π.χ. περιστροφή του καρπού, παίξιμο των δακτύλων. Ένα παράδειγμα είναι το νόημα 'ΑΘΛΗΜΑΤΑ'. Αυτό το νόημα περιέχει μια κίνηση περιστροφής του καρπού. Λόγω του ότι δεν υπάρχει αντιπροσωπευτική υπομονάδα για την κίνηση περιστροφής του καρπού, το νόημα αυτό αναγνωρίζεται λανθασμένα στο νόημα 'ΠΡΑΓΜΑΤΑ'. Το νόημα '60' το οποίο αναγνωρίζεται λανθασμένα στο νόημα'6'. Τα δύο αυτά νοήματα αρθρώνονται ακριβώς με τον ίδιο τρόπο με τη διαφορά ότι το νόημα '60' περιέχει μια επιπλέον κίνηση: ένα παίξιμο των δακτύλων. Επιπλέον λόγω της έλλειψης της τρισδιάστατης πληροφορίας, οι κινήσεις προβάλλονται στον δισδιάστατο χώρο. Έτσι το νόημα 'ΑΛΑΤΙ' το οποίο περιέχει μια τρισδιάστατη κυκλική κίνηση, αναγνωρίζεται λανθασμένα ως το νόημα 'ΠΕΡΙΠΑΤΗΤΗΣ'.

Προσαρμογή σε νοηματιστή

Όπως είδαμε στα πειραματικά αποτελέσματα της προηγούμενης ενότητας, όταν ο νοηματιστής είναι άγνωστος, το ποσοστό αναγνώρισης μειώνεται αισθητά σε σύγκριση με την περίπτωση όπου ο νοηματιστής είναι γνωστός. Αυτό συμβαίνει σε όλες τις μεθόδους που δείξαμε. Το σενάριο άγνωστου νοηματιστή αξιολογεί τις μεθόδους σε σχέση με τη δυνατότητα γενίκευσής τους σε έναν άγνωστο νοηματιστή.

Εδώ αξιολογούμε την 2-S-U μέθοδο μετά την εφαρμογή του σχήματος προσαρμογής που παρουσιάστηκε στην ενότητα 5.3, χρησιμοποιώντας 32 τερματικούς κόμβους στον αλγόριθμο regression class tree. Συγκεκριμένα, χρησιμοποιούμε ένα σύνολο δεδομένων προσαρμογής από τον

Πίνακας 7.10: Αξιολόγηση αναγνώρισης σε άγνωστο νοηματιστή και 97 νοήματα από την βάση δεδομένων ASLLVD. Τα αποτελέσματα είναι σε sign accuracy %.

	Νοηματιστής	M-P	HS	M-P+HS
2-S-U	Dana	40.31	44.21	**63.15**
2-S-U	Lana	38.2	40.1	**61.3**
Sign-DTW	Dana	26.3	41	55.78
Sign-DTW	Lana	33.6	35.7	53.6

νοηματιστή προς αναγνώριση το οποίο περιλαμβάνει το 20% των αρχικών δεδομένων προς αξιολόγηση, με άλλα λόγια μια επανάληψη ανά νόημα. Το υπόλοιπο 80% χρησιμοποιείται για την τελική αξιολόγηση. Τα δεδομένα προσαρμογής δεν έχουν επικάλυψη με τα δεδομένα αξιολόγησης.

Στον Πίνακα 7.9 απεικονίζουμε τα αποτελέσματα αναγνώρισης μετά την προσαρμογή στον νοηματιστή προς αξιολόγηση. Συγκρίνοντας τις μεθόδους 2-S-U vs. 2-S-U+MLLR παρατηρούμε αύξηση του ποσοστού αναγνώρισης κατά μέσον όρο και για τους δύο νοηματιστές 2.6%, 19.3% και 15.7% για τις ροές πληροφορίας M-P, HS και M-P+HS αντίστοιχα. Επιπλέον συγκρίνοντας τις μεθόδους 2-S-U+MLLR vs. 2-S-U+MLLR+IP δηλαδή προσθέτοντας επιπλέον προφορές νοημάτων από τον νοηματιστή προς αξιολόγηση, παρατηρούμε αύξηση 34.4%, 32% και 22.7% για τις ροές M-P, HS και M-P+HS αντίστοιχα. Αξίζει να σημειωθεί ότι τα αποτελέσματα του Πίνακα 7.9 για την 2-S-U μέθοδο διαφέρουν από αυτά του πινάκα 7.8 γιατί έχουν αφαιρεθεί από το σύνολο δεδομένων αξιολόγησης τα δεδομένα προσαρμογής.

Συνοψίζοντας μετά την προσαρμογή στο νοηματιστή προς αξιολόγηση κάνοντας χρήση όλων των ροών πληροφορίας, το ποσοστό αναγνώρισης αγγίζει το 92.5% κατά μέσο όρο για τους δύο νοηματιστές.

7.2.2 Βάση δεδομένων ASL Large Vocabulary Dictionary Corpus (ASLLVD)

Σε αυτή την ενότητα παρουσιάζουμε πειραματικά αποτελέσματα στην βάση δεδομένων ASLLVD [8]. Η ASL Large Vocabulary Dictionary Corpus (ASLLVD) βάση δεδομένων αποτελείται από 97 διαφορετικά νοήματα, τα οποία έχουν εκτελεστεί μια φορά από δύο νοηματιστές (Dana και Lana).

Για την εκπαίδευση των μοντέλων υπομονάδας χρησιμοποιούμε τα δεδομένα από ένα νοηματιστή δηλαδή μια επανάληψη για κάθε νόημα. Για την αξιολόγηση χρησιμοποιούμε τα δεδομένα από τον άλλο νοηματιστή. Τα δεδομένα αυτά δεν έχουν ιδωθεί κατά τη διάρκεια της εκπαίδευσης των μοντέλων.

Ο αριθμός των υπομονάδων που χρησιμοποιούνται σε κάθε ροή πληροφορίας καθορίστηκε μεγιστοποιώντας το ποσοστό αναγνώρισης σε ένα σύνολο δεδομένων ανάπτυξης. Αυτό αποτελείται από το 20% των δεδομένων του άγνωστου νοηματιστή προς αξιολόγηση. Τα δύο σύνολα δεδομένων ανάπτυξης και αξιολόγησης δεν έχουν επικάλυψη μεταξύ τους. Επιπλέον, εφαρμόζουμε τη διαδικασία 5-fold cross-validation για την επιλογή των συνόλων δεδομένων προς ανάπτυξη και αξιολόγηση. Ο μέσος αριθμός των υπομονάδων που χρησιμοποιήθηκε για κάθε ροή πληροφορίας είναι: 10 υπομονάδες για την ροή της θέσης, 10 υπομονάδες για τη ροή της κίνησης και 200 υπομονάδες για την ροή της χειρομορφής. Τα διανύσματα χαρακτηριστικών που χρησιμοποιήθηκαν όπως έχουμε ήδη αναφέρει είναι: για τη ροή της θέσης οι συντεταγμένες των χεριών κανονικοποιημένες ως προς το κεφάλι, για τη ροή της κίνησης η κατεύθυνση της κίνησης και για την ροή της χειρομορφής το διάνυσμα χαρακτηριστικών που προκύπτει από την εφαρμογή της μεθόδου spatial pyramids.

Στον Πίνακα 7.10 απεικονίζουμε τα ποσοστά αναγνώρισης. Παρατηρούμε ότι χρησιμοποιώντας

Πίνακας 7.11: Αξιολόγηση αναγνώρισης σε ένα νοηματιστή και 94 νοήματα από την βάση δεδομένων BU400. Τα αποτελέσματα είναι σε sign accuracy %.

2-S-U	SU-noDSC	SU-Frame	SU-Segm
82.5	77.8	80.9	72

τη ροή πληροφορίας της κίνησης-θέσης (M-P) και της χειρομορφής HS ξεχωριστά, η μέθοδος 2-S-U επιτυγχάνει αύξηση του ποσοστού αναγνώρισης έναντι της μεθόδου Sign-DTW κατά μέσο όρο και για τους δύο νοηματιστές 9.3% και 4.8% αντίστοιχα. Επιπλέον συνδυάζοντας όλες τις ροές πληροφορίας, M-P+HS περίπτωση, η μέθοδος 2-S-U υπερτερεί και πάλι της μεθόδου Sign-DTW, 7.5% κατά μέσο όρο και για τους δύο νοηματιστές.

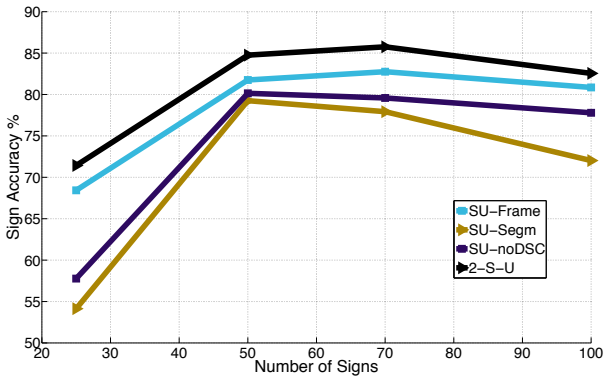
7.2.3 Βάση δεδομένων Boston University 400 Corpus (BU400)

Σε αυτή την ενότητα χρησιμοποιούμε τη βάση δεδομένων Boston University 400 Corpus (BU400) [91]. Επεξεργαζόμαστε τις ακόλουθες έξι ιστορίες: Accident, Biker-Buddy, Boston-La, Football, Lapd-story και Siblings. Τις ιστορίες αυτές τις έχει διηγηθεί ένας νοηματιστής, οπότε η εκπαίδευση και η αξιολόγηση γίνεται στον ίδιο νοηματιστή. Επιπλέον, χρησιμοποιούμε τις επισημειώσεις σε επίπεδο νοήματος για την κατάτμηση των βίντεο σε μεμονωμένα νοήματα. Το λεξιλόγιο αποτελείται από 94 διαφορετικά νοήματα και ο αριθμός των running glosses είναι 1202. Χρησιμοποιούμε το 60% των δεδομένων για την εκπαίδευση των μοντέλων, το 30% για την αξιολόγηση του συστήματος και το 10% για τον καθορισμό των συμβαλλόμενων παραμέτρων. Ακόμα, σε όλα τα πειράματα χρησιμοποιούμε 3-fold random selection των συνόλων εκπαίδευσης και αξιολόγησης και τα αποτελέσματα που απεικονίζουμε είναι ο μέσος όρος των επιμέρους αποτελεσμάτων. Παραπάνω λεπτομέρειες για την επιλογή των συνόλων εκπαίδευσης και αξιολόγησης αναφέρονται στην ιστοσελίδα [125]. Τέλος, στα πειράματα που ακολουθούν λαμβάνουμε υπόψη τις ροές πληροφορίας της θέσης και της κίνησης και για τα δύο χέρια και τη ροή της χειρομορφής για το κυρίαρχο χέρι.

Ο αριθμός υπομονάδων που χρησιμοποιήθηκε στα πειράματα που ακολουθούν, καθορίστηκε με βάση τη μεγιστοποίηση του ποσοστού αναγνώρισης στο σύνολο δεδομένων ανάπτυξης. Αυτά δεν έχουν επικάλυψη με το σύνολο δεδομένων αξιολόγησης. Για τη μέθοδο 2-S-U χρησιμοποιούμε 20, 30, και 110 υπομονάδες για τις ροές πληροφορίας της θέσης, κίνησης και χειρομορφής αντίστοιχα. Για τις μεθόδους SU-noDSC, SU-Segm και SU-Frame χρησιμοποιούμε 150, 100, 100 υπομονάδες για τη ροή της κίνησης-θέσης και 110 υπομονάδες για τη ροή της χειρομορφής. Όπως παρατηρούμε, ο αριθμός υπομονάδων που χρησιμοποιείται για τη ροή της κίνησης-θέσης στην μέθοδο 2-S-U, είναι αρκετά μικρότερος σε σχέση με τις άλλες μεθόδους. Αυτό οφείλεται στο διαχωρισμό σε δυναμικές και στατικές υπομονάδες: χρειαζόμαστε μικρότερο αριθμό υπομονάδων για την μοντελοποίηση του deconvolved χώρου χαρακτηριστικών. Τα διανύσματα χαρακτηριστικών που χρησιμοποιήθηκαν όπως έχουμε ήδη αναφέρει είναι: για τη ροή της θέσης οι συντεταγμένες των χεριών κανονικοποιημένες ως προς το κεφάλι, για τη ροή της κίνησης η κατεύθυνση της κίνησης και για τη ροή της χειρομορφής, το διάνυσμα χαρακτηριστικών που προκύπτει από την εφαρμογή της μεθόδου spatial pyramids.

Σύγκριση με άλλες μεθόδους

Στον Πίνακα 7.11 απεικονίζουμε τα αποτελέσματα αναγνώρισης για τις διαφορετικές μεθόδους χρησιμοποιώντας όλες τις ροές πληροφορίας: θέση, κίνηση και χειρομορφή. Όπως παρατηρούμε, η μέθοδος 2-S-U υπερτερεί της μεθόδου SU-noDSC. Αυτό υποδεικνύει την σπουδαιότητα του



Σχήμα 7.7: Αξιολόγηση αναγνώρισης σε ένα νοηματιστή στην βάση δεδομένων BU400, μεταβάλλοντας τον αριθμό των νοημάτων.

διαχωρισμού σε δυναμικές και στατικές υπομονάδες. Η χρήση διαφορετικών αλλά κατάλληλων χαρακτηριστικών στις δυναμικές και στατικές υπομονάδες αντίστοιχα, επηρεάζει την αναγνώριση σε σχέση με τη χρήση του συνόλου των χαρακτηριστικών σε κάθε περίπτωση. Επιπλέον, παρατηρούμε ότι η μέθοδος 2-S-U υπερτερεί όλων των άλλων, οδηγώντας σε αύξηση του ποσοστού αναγνώρισης κατά 4.7%, 1.6% και 10.5% σε σχέση με τις μεθόδους SU-noDSC, SU-Frame και SU-Segm αντίστοιχα.

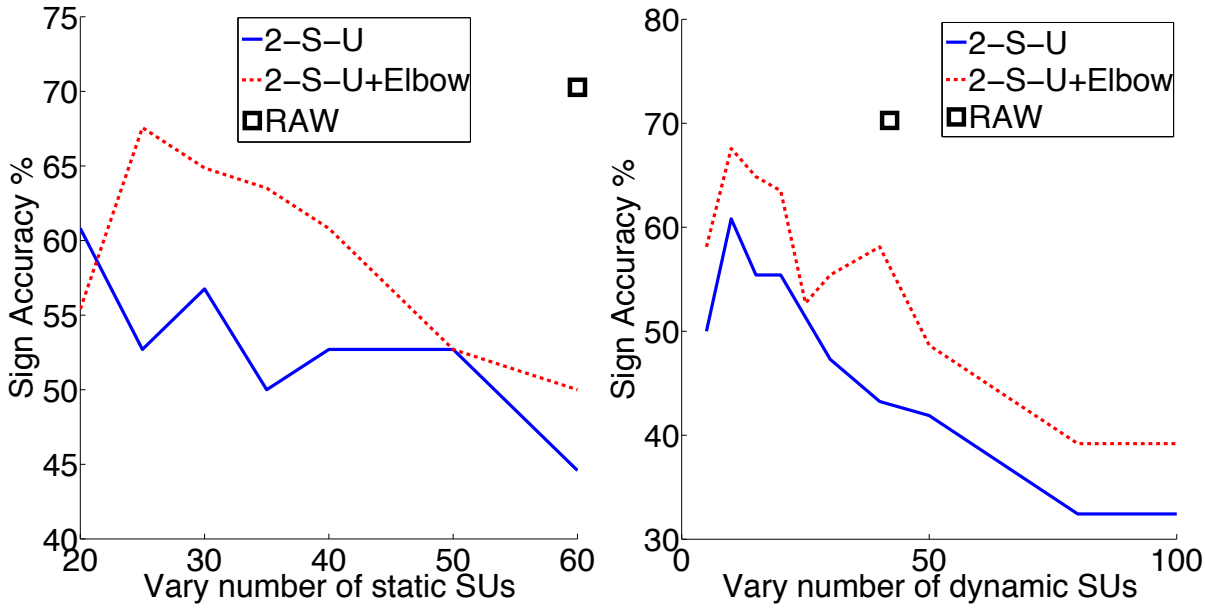
Μεταβολή του αριθμού των νοημάτων

Στο Σχήμα 7.7 συγκρίνουμε την προτεινόμενη μέθοδο 2-S-U με άλλες μεθόδους από τη διεθνή βιβλιογραφία, μεταβάλλοντας τον αριθμό των διαφορετικών νοημάτων. Επιπλέον, χρησιμοποιούμε όλες τις ροές πληροφορίας: θέση, κίνηση και χειρομορφή. Από τα αποτελέσματα παρατηρούμε ότι αυξάνοντας τον αριθμό των νοημάτων από 25 σε 50 το ποσοστό αναγνώρισης για όλες τις μεθόδους αυξάνει. Αυτό οφείλεται στο γεγονός ότι με την αύξηση των νοημάτων αυξάνει και ο αριθμός των δεδομένων που χρησιμοποιούνται για την κατασκευή και εκπαίδευση των μοντέλων υπομονάδων. Με αυτό τον τρόπο, οι υπομονάδες που προκύπτουν περιγράφουν καλύτερα την ποικιλία άρθρωσης των νοημάτων. Επιπλέον παρατηρούμε ότι η μέθοδος 2-S-U υπερτερεί όλων των άλλων για όλους τους διαφορετικούς αριθμούς νοημάτων που χρησιμοποιήθηκαν. Συγκεκριμένα η μέθοδος 2-S-U αυξάνει το ποσοστό αναγνώρισης κατά μέσο όρο, 2.7% σε σύγκριση με την SU-Frame μέθοδο, 7.3% σε σχέση με την SU-noDSC μέθοδο και 10.3% σε σχέση με την SU-Segm μέθοδο.

7.2.4 Βάση δεδομένων Continuous GSL Phrases Corpus (GSL-Phrases)

Σε αυτή την ενότητα χρησιμοποιούμε μέρος της βάσης δεδομένων Continuous GSL Phrases Corpus (GSL-Phrases) [3]. Η βάση δεδομένων έχει καταγραφεί χρησιμοποιώντας τον αισθητήρα Kinect. Έτσι η βάση προσφέρει εικόνες βάθους και χρώματος (RGB), τη μάσκα του χρήστη, την πληροφορία του σκελετού και τον προσανατολισμό των αρθρώσεων του νοηματιστή. Επεξεργαζόμαστε 30 προτάσεις συνεχούς νοηματικού λόγου στην ΕΝΓ, οι οποίες έχουν εκτελεστεί τρεις φορές η κάθε μια από ένα νοηματιστή. Το λεξιλόγιο αποτελείται από 33 διαφορετικά νοήματα. Για την εκπαίδευση των μοντέλων υπομονάδας χρησιμοποιούμε τις δύο από τις τρεις προτάσεις και η αξιολόγηση του συστήματος γίνεται χρησιμοποιώντας την τρίτη. Η ροή πληροφορίας που χρησιμοποιούμε σχετίζεται με την κίνηση-θέση των χεριών του νοηματιστή εκμεταλλευόμενοι την πληροφορία του σκελετού του νοηματιστή.

Παρουσιάζουμε πειραματικά αποτελέσματα συγκρίνοντας τις μεθόδους 2-S-U, 2-S-U+Elbow και RAW. Συγκεκριμένα, στις μεθόδους 2-S-U και RAW χρησιμοποιήσαμε ως διάνυσμα χαρακτη-



(α') (β')

Σχήμα 7.8: Αξιολόγηση αναγνώρισης των μεθόδων 2-S-U, 2-S-U+Elbow και RAW στη βάση δεδομένων GSL-Phrases, μεταβάλλοντας (α) τον αριθμό των στατικών υπομονάδων, (β) τον αριθμό των δυναμικών υπομονάδων.

ριστικών:

- τις συντεταγμένες των χεριών κανονικοποιημένες με τη θέση του κεφαλιού ($\boldsymbol{P}$),
- τη στιγμιαία κατεύθυνση της κίνησής τους ($\boldsymbol{D}$),
- την ταχύτητά τους ($\boldsymbol{V}$) και
- την απόστασή τους ($\boldsymbol{L}$).

Στην μέθοδο 2-S-U+Elbow προσθέσαμε στα παραπάνω χαρακτηριστικά διανύσματα, πληροφορία που σχετίζεται με τον αγκώνα του νοηματιστή. Πιο συγκεκριμένα, συμπεριλάβαμε: τις συντεταγμένες των αγκώνων κανονικοποιημένες α) με τη θέση του κεφαλιού και β) με τη θέση των αντίστοιχων χεριών. Αυτή η επαύξηση του διανύσματος χαρακτηριστικών δεν εφαρμόστηκε στη μέθοδο RAW γιατί η μέθοδος δεν υποστηρίζει την κατασκευή των RAW υπομονάδων χρησιμοποιώντας την επιπλέον πληροφορία από τον αγκώνα του νοηματιστή.

Πειραματιζόμαστε μεταβάλλοντας τις παραμέτρους που συμβάλλουν στην κάθε μέθοδο. Αυτές είναι ο αριθμός των στατικών (nS) και δυναμικών (nD) υπομονάδων για τις μεθόδους 2-S-U και 2-S-U+Elbow. Να σημειώσουμε ότι η μέθοδος RAWέχει σταθερό αριθμό υπομονάδων. Στο Σχήμα 7.8 παρουσιάζουμε τα αποτελέσματα αναγνώρισης μεταβάλλοντας μια από τις δύο παραμέτρους κάθε φορά. Την άλλη παράμετρο τη διατηρούμε σταθερή και ίση με την τιμή όπου επιτυγχάνεται το μεγαλύτερο ποσοστό αναγνώρισης για κάθε μέθοδο. Έτσι εξετάζουμε την μεταβολή του ποσοστού αναγνώρισης μεταβάλλοντας κάθε παράμετρο ξεχωριστά και για τις τρεις μεθόδους.

Όπως παρατηρούμε στο Σχήμα 7.8 η μέθοδος RAW υπερτερεί των μεθόδων 2-S-U και 2-S-U+Elbow επιτυγχάνοντας ποσοστό αναγνώρισης 70.2%. Το καλύτερο ποσοστό αναγνώρισης για τη μέθοδο 2-S-Uίσο με 60.8% επιτυγχάνεται με 20 στατικές υπομονάδες και 10 δυναμικές υπομονάδες. Ενώ η μέθοδος 2-S-U+Elbow επιτυγχάνει το καλύτερο ποσοστό αναγνώρισης ίσο με

Σχήμα 7.9: Απεικονίζουμε ένα παράδειγμα αναγνώρισης της πρότασης 'ΒΕΡΟΛΙΝΟ ΕΙΣΙΤΗΡΙΟ ΕΓΩ ΠΡΕΠΕΙ ΠΛΗΡΩΝΩ ΠΟΥ' για τις τρεις μεθόδους: RAW (πρώτη γραμμή), 2-S-U+Elbow (δεύτερη γραμμή) και 2-S-U (τρίτη γραμμή). Τις εισαγωγές νοημάτων τις συμβολίζουμε με κόκκινο χρώμα, τις αντικαταστάσεις με κίτρινο και τις διαγραφές με πράσινο.

67.5%, με με 25 στατικές υπομονάδες, 10 δυναμικές υπομονάδες. Παρατηρούμε ότι η μέθοδος 2-S-U+Elbow υπερτερεί της μεθόδου 2-S-U αυξάνοντας το ποσοστό αναγνώρισης κατά 6.7%. Αυτό υποδεικνύει ότι η χρησιμοποίηση της πληροφορίας του αγκώνα του νοηματιστή βοηθάει την αναγνώριση. Αυτή την επιπλέον πληροφορία την λαμβάνουν υπόψη τους μόνο οι στατικές υπομονάδες. Για την αξιοποίησή της, αυξάνεται ο αριθμός των στατικών υπομονάδων από 20 σε 25. Ενώ αντίθετα ο αριθμός των δυναμικών υπομονάδων όπου η πληροφορία του αγκώνα δεν λαμβάνεται υπόψη παραμένει σταθερός.

Στο Σχήμα 7.9 απεικονίζουμε ένα παράδειγμα αναγνώρισης της πρότασης 'ΒΕΡΟΛΙΝΟ ΕΙΣΙΤΗΡΙΟ ΕΓΩ ΠΡΕΠΕΙ ΠΛΗΡΩΝΩ ΠΟΥ' από τις τρεις παραπάνω μεθόδους. Τις εισαγωγές νοημάτων τις συμβολίζουμε με κόκκινο χρώμα, τις αντικαταστάσεις με κίτρινο και τις διαγραφές με πράσινο. Όπως παρατηρούμε, και οι τρεις μέθοδοι κάνουν κάποια λάθη στην αναγνώριση. Παρόλα αυτά και οι τρεις καταφέρουν να την αναγνωρίσουν σχεδόν σωστά παρά την ύπαρξη του φαινομένου της συνάρθρωσης διαδοχικών νοημάτων, φαινόμενο πολύ έντονο κατά τον συνεχή νοηματικό λόγο.

7.3 Αναγνώριση νοημάτων με γλωσσικές-φωνητικές υπομονάδες

Σε αυτή την ενότητα παρουσιάζουμε τα πειραματικά αποτελέσματα αναγνώρισης ΝΓ κάνοντας χρήση των γλωσσικών-φωνητικών υπομονάδων που παρουσιάσαμε στο κεφάλαιο 4. Πειραματιζόμαστε στη βάση δεδομένων GSL-Lem [37], όπου και υπήρχαν διαθέσιμες οι φωνητικές επισημειώσεις σε HamNoSys. Το λεξιλόγιο που χρησιμοποιήθηκε περιελάμβανε 300 διαφορετικά νοήματα εκτελεσμένα από δύο διαφορετικούς νοηματιστές (Κώστας και Όλγα).

Για την ανίχνευση και παρακολούθηση των αρθρωτών, χρησιμοποιήθηκε το σύστημα που παρουσιάστηκε στο κεφάλαιο 2. Οι ροές πληροφορίας που χρησιμοποιήθηκαν σχετίζονται με την κίνηση-θέση και των δύο χεριών του νοηματιστή και με την χειρομορφή του κυρίαρχου χεριού.

Για τη ροή της κίνησης-θέσης χρησιμοποιήθηκε το διάνυσμα χαρακτηριστικών που περιλαμβάνει:

- τις συντεταγμένες των χεριών κανονικοποιημένες με τη θέση του κεφαλιού ($\boldsymbol{P}$),
- τη στιγμιαία κατεύθυνση της κίνησής τους ($\boldsymbol{D}$),
- την ταχύτητά τους ($\boldsymbol{V}$) και
- την απόστασή τους ($\boldsymbol{L}$).

Για τη ροή της χειροφορφής χρησιμοποιήσαμε το διάνυσμα χαρακτηριστικών που προκύπτει από την εφαρμογή της μεθόδου spatial pyramids. Περισσότερες λεπτομέρειες σε σχέση με τα διανύσματα χαρακτηριστικών αναφέρονται στο κεφάλαιο 2.

Οι μέθοδοι από τη διεθνή βιβλιογραφία που χρησιμοποιήθηκαν για τους σκοπούς σύγκρισης είναι οι ακόλουθες: Theodorakis et al. 2014 (2-S-U) [126], Fang et al. 2004 (SU-Segm) [47] και Bauer and Kraiss 2001 (SU-Frame) [11]. Όλες αυτές οι μέθοδοι βασίζονται σε δεδομενοκεντρικές υπομονάδες σε αντίθεση με την προτεινόμενη μέθοδο SU-P η οποία βασίζεται σε γλωσσικές-φωνητικές υπομονάδες. Επιπλέον, συγκρίνουμε και με τη μέθοδο Wang et al. 2010 (Sign-DTW) [138], η οποία βασίζεται στη μοντελοποίηση σε επίπεδο νοήματος και όχι υπομονάδων. Περισσότερες λεπτομέρειες σε σχέση με τις μεθόδους που χρησιμοποιήθηκαν για σύγκριση αναφέρονται στις ενότητες 7.2 1.2. Για τη μοντελοποίηση της ροής M-P εφαρμόσαμε σε κάθε μέθοδο τη μοντελοποίηση που χρησιμοποιήθηκε αντίστοιχα στην κάθε δημοσίευση. Αντιθέτως, για τη μοντελοποίηση της ροής HS χρησιμοποιήσαμε την ίδια μοντελοποίηση σε όλες τις μεθόδους όπως αυτή περιγράφεται στην ενότητα 3.4.3. Τέλος, για τη σύμμειξη των ροών πληροφορίας της κίνησης-θέσης και χειρομορφής για όλες τις μεθόδους χρησιμοποιήθηκαν τα PaHMM όπως περιγράφεται στην ενότητα 5.2.2.

Για την αξιολόγηση χρησιμοποιούμε τη μέτρηση Sign Accuracy $= (N - S)/N \cdot 100\%$ όπου N είναι ο αριθμός των δεδομένων προς αξιολόγηση και S είναι ο αριθμός των αντικαταστάσεων. Διαγραφές και εισαγωγές δεν υπάρχουν, λόγω το ότι αναγνωρίζουμε μεμονωμένα νοήματα. Μια επιπλέον μέτρηση που χρησιμοποιούμε είναι η μέση κατάταξη (mean rank -MR-). Αυτή αντιστοιχεί στην μέση κατάταξη της σωστής κλάσης. Κατά τη διάρκεια της αναγνώρισης κατατάσσουμε όλα τα νοήματα που έχουμε στο λεξικό μας με βάση το πιο πιθανό, δεδομένης της εκτέλεσης προς αναγνώριση. Η μέτρηση της μέσης κατάταξης (MR), είναι ο μέσος όρος των θέσεων στις οποίες έχουν καταταχθεί οι σωστές ετικέτες νοημάτων για όλα τα δεδομένα προς αναγνώριση.

7.3.1 Μεταβολή των παραμέτρων του ITA

Παρουσιάζουμε πειραματικά αποτελέσματα κάνοντας αξιολόγηση του συστήματος σε άγνωστο νοηματιστή. Με άλλα λόγια, κατά τη διάρκεια της εκπαίδευσης δεν έχουμε χρησιμοποιήσει καθόλου δεδομένα από τον νοηματιστή που χρησιμοποιούμε κατά την αξιολόγηση. Εκπαιδεύουμε τα PDTS μοντέλα υπομονάδων χρησιμοποιώντας όλες τις επαναλήψεις για κάθε νόημα από ένα νοηματιστή και αξιολογούμε το σύστημά μας σε δεδομένα από ένα διαφορετικό, άγνωστο νοηματιστή.

Οι κύριες παράμετροι του αλγορίθμου ITA είναι ο τύπος της PDTS γραμματικής που χρησιμοποιήθηκε και ο τρόπος αρχικοποίησης των PDTS HMM μοντέλων υπομονάδας. Η PDTS γραμματική συμβολίζεται με G-MP και G-HS για τις ροές πληροφορίας M-P και HS αντίστοιχα. Για τη ροή M-P χρησιμοποιούμε είτε μόνο την $G_{\{del,sub\}}$ γραμματική είτε τις γραμματικές $G_{\{del,sub\}}$ και G_{sub} διαδοχικά. Για τη ροή HS χρησιμοποιούμε είτε μόνο την G_{ins} γραμματική είτε τις γραμματικές G_{ins} και G_{sub} διαδοχικά. Η αρχικοποίηση των HMM μοντέλων συμβολίζεται με I-MP και I-HS για τις ροές πληροφορίας M-P και HS αντίστοιχα. Για τη ροή M-P χρησιμοποιούμε είτε flat-start (FS) είτε RAW αρχικοποίηση, ενώ για τη ροή HS χρησιμοποιούμε FS αρχικοποίηση.

Πίνακας 7.12: Αξιολόγηση σε άγνωστο νοηματιστή μεταβάλλοντας τις παραμέτρους του αλγορίθμου ITA.

	Παράμετροι του ITA				Νοηματιστής: Όλγα			Νοηματιστής: Κώστας		
Μέθοδος	G-MP	I-MP	G-HS	I-HS	M-P	HS	M-P+HS	M-P	HS	M-P+HS
SU-P	-	-	-	-	9	35.5	34.8	10.6	31.6	37.9
SU-P+ITA	$G_{\{del,sub\}}$	FS	G_{ins}	FS	21.7	38.7	43.6	23.8	41	44.7
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	FS	G_{ins}	FS	22	38.7	60	23.2	41	57
SU-P+ITA	$G_{\{del,sub\}}$	FS	$G_{ins} + G_{sub}$	FS	21.7	42.4	46.6	23.8	42.8	46.2
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	FS	$G_{ins} + G_{sub}$	FS	22	42.4	60.7	23.2	42.8	57.4
SU-P+ITA	$G_{\{del,sub\}}$	RAW	G_{ins}	FS	21.3	38.7	47.5	20	41	47
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	RAW	G_{ins}	FS	26.2	38.7	61.6	27.8	41	58
SU-P+ITA	$G_{\{del,sub\}}$	RAW	$G_{ins} + G_{sub}$	FS	21.3	42.4	49.7	20	42.8	48.8
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	RAW	$G_{ins} + G_{sub}$	FS	26.2	42.4	**62.9**	27.8	42.8	**58.6**

Πίνακας 7.13: Αξιολόγηση σε άγνωστο νοηματιστή μεταβάλλοντας τις παραμέτρους του αλγορίθμου ITA. Μέση κατάταξη (MR) της σωστής κλάσης νοήματος για την M-P ροή.

	Παράμετροι του ITA		Νοηματιστής: Όλγα	Νοηματιστής: Κώστας
Μέθοδος	G-MP	I-MP	MR	MR
SU-P+ITA	$G_{\{del,sub\}}$	FS	42.5	50.8
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	FS	19.8	31.1
SU-P+ITA	$G_{\{del,sub\}}$	RAW	46.8	48.8
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	RAW	18.8	23.9

Επίδραση του αλγορίθμου ITA

Συγκρίνοντας τα ποσοστά αναγνώρισης των μεθόδων SU-P και SU-P+ITA (Πίνακας 7.12), παρατηρούμε ότι με τη χρησιμοποίηση του ITA επιτυγχάνουμε αύξηση του ποσοστού αναγνώρισης κατά 10.6%, 3.2% και 6.8% για τις ροές πληροφορίας M-P, HS και M-P+HS αντίστοιχα. Τα παραπάνω αποτελέσματα υποδεικνύουν ότι η χρησιμοποίηση του αλγορίθμου ITA κατά τη διάρκεια της εκπαίδευσης επηρεάζει δραστικά την αναγνώριση. Ο αλγόριθμος ITA διορθώνει τις PDTS επισημειώσεις έτσι ώστε να είναι συνεπείς με την πραγματική εκτέλεση κάθε νοήματος. Αυτό έχει ως αποτέλεσμα την εκπαίδευση εύρωστων PDTS μοντέλων υπομονάδας.

Μεταβολή της PDTS γραμματικής

Στον πίνακα 7.12 απεικονίζουμε τα πειραματικά αποτελέσματα μεταβάλλοντας την PDTS γραμματική που χρησιμοποιήθηκε στον αλγόριθμο ITA. Ας εστιάσουμε πρώτα στις περιπτώσεις όπου η αρχικοποίηση γίνεται με flat-start για τις ροές M-P και HS. Συγκρίνοντας τα αποτελέσματα χρησιμοποιώντας την $G_{\{del,sub\}}$ vs. $G_{\{del,sub\}} + G_{sub}$ PDTS γραμματικές για την M-P ροή και την G_{ins} για την HS ροή, δεύτερη και τρίτη σειρά στον πίνακα 7.12, παρατηρούμε ότι το ποσοστό αναγνώρισης για την M-P ροή μένει σχεδόν ανεπηρέαστο. Παρόλα αυτά, παρατηρώντας την πρώτη

Πίνακας 7.14: Αξιολόγηση σε άγνωστο νοηματιστή μεταβάλλοντας τις παραμέτρους του αλγορίθμου ITA. Μέση κατάταξη (MR) της σωστής κλάσης νοήματος για την HS ροή.

			Νοηματιστής	
	Παράμετροι του ITA		Όλγα	Κώστας
Μέθοδος	G-MP	I-MP	MR	MR
SU-P+ITA	G_{ins}	FS	31.6	30.8
SU-P+ITA	$G_{ins} + G_{sub}$	FS	23.9	29.8

και δεύτερη γραμμή του πίνακα 7.13, βλέπουμε ότι η μέση κατάταξη της σωστής κλάσης νοήματος μειώνεται κατά μέσο όρο και για τους δύο νοηματιστές κατά 46%. Με άλλα λόγια ενώ η πιο πιθανή κλάση νοήματος δεν επηρεάζεται, η διάταξη των υπολοίπων κλάσεων αλλάζει δραστικά. Την επιρροή τής χρησιμοποίησης της G_{sub} PDTS γραμματικής μπορούμε να τη δούμε μετά τη σύμμειξη των ροών M-P και HS, δηλαδή την M-P+HS στήλη του πίνακα 7.12. Όπως παρατηρούμε το ποσοστό αναγνώρισης αυξάνει κατά μέσο όρο για τους δύο νοηματιστές 14.3%. Τα παραπάνω υποδεικνύουν την σημαντικότητα της εφαρμογής του αλγορίθμου ITA δύο διαδοχικές φορές για την M-P ροή. Την πρώτη με την $G_{\{del,sub\}}$ γραμματική και την δεύτερη με την G_{sub} γραμματική. Συγκρίνοντας τα αποτελέσματα χρησιμοποιώντας την G_{ins} vs. $G_{ins} + G_{sub}$ PDTS γραμματική στην HS ροή και την $G_{\{del,sub\}}$ γραμματική για την M-P ροή, παρατηρούμε ότι το ποσοστό αναγνώρισης αυξάνει κατά μέσο όρο για τους δύο νοηματιστές 2.7% και 2.2% για τις ροές HS και M-P+HS αντίστοιχα. Επιπλέον, η μέση κατάταξη της σωστής κλάσης μειώνεται κατά 13.7% (πίνακας 7.14).

Για να συνοψίσουμε, η εφαρμογή του αλγορίθμου ITA με την G_{sub} PDTS γραμματική και στις δύο ροές πληροφορίας (M-P και HS) είναι κρίσιμη για την απόδοση της αναγνώρισης. Αυτό οφείλεται στο γεγονός ότι η G_{sub} PDTS γραμματική διορθώνει λανθασμένες explicit PDTS υπομονάδες οι οποίες δεν μπορούσαν να διορθωθούν μόνο με την χρησιμοποίηση της $G_{\{del,sub\}}$ PDTS γραμματικής. Έτσι, διατηρεί τη συνέπεια μεταξύ των PDTS επισημειώσεων και της πραγματικής εκτέλεσης του κάθε νοήματος. Όμοια συμπεράσματα μπορούμε να βγάλουμε σε σχέση με την εφαρμογή του αλγορίθμου ITA με την G_{sub} PDTS γραμματική, χρησιμοποιώντας αυτή τη φορά τα RAW μοντέλα για την αρχικοποίηση των PDTS μοντέλων υπομονάδας.

Αρχικοποίηση των PDTS μοντέλων

Η αρχικοποίηση των HMM μοντέλων παίζει σημαντικό ρόλο και στην αναγνώριση αλλά και στη διόρθωση των PDTS επισημειώσεων. Συγκρίνοντας την αρχικοποίηση της M-P ροής με FS ή RAW χρησιμοποιώντας τις PDTS γραμματικές $G_{\{del,sub\}} + G_{sub}$ και $G_{ins} + G_{sub}$ για τις ροές M-P και HS αντίστοιχα, παρατηρούμε ότι χρησιμοποιώντας την RAW αρχικοποίηση το ποσοστό αναγνώρισης αυξάνει κατά μέσο όρο για τους δύο νοηματιστές 4.4% και 1.7% για τις ροές M-P και M-P+HS αντίστοιχα. Στο Σχήμα 4.4 απεικονίζουμε την αντιστοίχιση των PDTS μοντέλων υπομονάδας χρησιμοποιώντας FS και RAW αρχικοποίηση. Χρησιμοποιώντας FS αρχικοποίηση παρατηρούμε ότι σε δύο transitions (επισημειωμένες με κόκκινο τετράγωνο) του νοήματος 'ΘΥΜΑΜΑΙ', έχει αντιστοιχηθεί λάθος PDTS υπομονάδα. Στο πρώτο transition, το οποίο είναι μια ευθεία κίνηση προς τα πάνω, έχει αντιστοιχηθεί σε μια PDTS υπομονάδα κίνησης ευθείας με κατεύθυνση προς τα αριστερά. Επιπλέον, το δεύτερο transition, το οποίο είναι μια ευθεία κίνηση προς τα κάτω, έχει αντιστοιχηθεί σε μια PDTS υπομονάδα κίνησης καμπύλης με κατεύθυνση προς τα δεξιά και καμπύλη προς τα πάνω. Αντίθετα όμως, με τη χρήση της RAW αρχικοποίησης παρατηρούμε ότι σε κάθε χρονικό τμήμα αντιστοιχίζεται η σωστή PDTS υπομονάδα.

Συνοψίζοντας, η χρησιμοποίηση της RAW αρχικοποίησης για την M-P ροή έχει και ποσοτι-

Πίνακας 7.15: Αξιολόγηση σε άγνωστο νοηματιστή και σύγκριση με άλλες μεθόδους από την διεθνή βιβλιογραφία. Αποτελέσματα σε sign accuracy σε 300 νοήματα της βάσης δεδομένων GSL-Lem.

	Μέθοδοι				
Νοηματιστής	SU-P+ITA	2-S-U	SU-Segm	SU-Frame	Sign-DTW
Όλγα	**62.9**	61.2	54.4	40.5	57.9
Κώστας	**58.6**	50.1	32.6	35.5	46.3

Πίνακας 7.16: Προσαρμογή σε νοηματιστή χρησιμοποιώντας ένα σύνολο προσαρμογής από τον νοηματιστή προς αναγνώριση. Αποτελέσματα σε sign accuracy σε 300 νοήματα της βάσης δεδομένων GSL-Lem.

	Νοηματιστής	Όλγα			Κώστας		
	Ροή	M-P	HS	M-P+HS	M-P	HS	M-P+HS
Μέθοδοι	SU-P+ITA	28.5	42	62.6	27.1	44.3	57
	SU-P+ITA+MLLR	37.8	60.9	76.3	34.8	66	78.4
	SU-P+ITA+MLLR+IP	73	88	94.2	66.3	89.5	94

κή επίδραση αλλά και ποιοτική επίδραση. Αύξηση του ποσοστού αναγνώρισης αλλά και σωστή αντιστοίχιση των PDTS υπομονάδων σε σχέση με την πραγματική εκφορά κάθε νοήματος.

7.3.2 Σύγκριση με άλλες μεθόδους

Σε αυτή την ενότητα συγκρίνουμε την προτεινόμενη μέθοδο SU-P+ITA χρησιμοποιώντας όλες τις ροές πληροφορίας (M-P+HS) με τέσσερις μεθόδους από την διεθνή βιβλιογραφία. Στην SU-P+ITA μέθοδο έχουμε χρησιμοποιήσει τις PDTS γραμματικές $G_{\{del,sub\}} + G_{sub}$ και $G_{ins} + G_{sub}$ για τις ροές M-P και HS αντίστοιχα. Επιπλέον, έχουμε χρησιμοποιήσει RAW και FS αρχικοποίηση για τα PDTS HMM μοντέλα υπομονάδας των ροών M-P και HS αντίστοιχα. Η πιο κοντινή μέθοδος στην SU-P+ITA από αυτές που συγκρίνουμε είναι η 2-S-U (ενότητα 1.2). Και οι δύο μέθοδοι μοιράζονται την έννοια της διαδοχής διαφορετικών τύπων υπομονάδας [77], παρόλα αυτά η 2-S-U μέθοδος βασίζεται σε δεδομενοκεντρικές υπομονάδες, ενώ η SU-P+ITA μέθοδος σε γλωσσικές-φωνητικές υπομονάδες.

Συγκρίνοντας τις μεθόδους SU-P+ITA vs. 2-S-U στον πίνακα 7.15, παρατηρούμε ότι η SU-P+ITA υπερτερεί της 2-S-U, οδηγώντας σε απόλυτη αύξηση του ποσοστού αναγνώρισης κατά μέσο όρο για τους δύο νοηματιστές 5.1%. Επιπλέον, συγκρίνουμε την SU-P+ITA μέθοδο με άλλες δύο δεδομενοκεντρικές μεθόδους υπομονάδας (SU-Segm, SU-Frame), οι οποίες δεν εμπεριέχουν την έννοια της διαδοχής διαφορετικών τύπων υπομονάδας, και μιας μεθόδου που βασίζεται στη μοντελοποίηση σε επίπεδο νοήματος (Sign-DTW). Όπως παρατηρούμε στον πίνακα 7.15, η SU-P+ITA μέθοδος υπερτερεί όλων των παραπάνω μεθόδων.

Συνοψίζοντας, έκτος από το γεγονός ότι οι γλωσσικές-φωνητικές υπομονάδες είναι γλωσσικά ερμηνεύσιμες σε αντίθεση με τις δεδομενοκεντρικές, είναι επίσης πολύ σημαντικές για την αναγνώριση της ΝΓ.

7.3.3 Προσαρμογή σε νοηματιστή

Αξιολογούμε την SU-P+ITA μέθοδο μετά την εφαρμογή του σχήματος προσαρμογής που παρουσιάστηκε στην ενότητα 5.3, χρησιμοποιώντας 32 τερματικούς κόμβους στον αλγόριθμο regression class tree. Συγκεκριμένα χρησιμοποιούμε ένα σύνολο δεδομένων προσαρμογής από τον νοηματιστή προς αναγνώριση το οποίο περιλαμβάνει το 20% των αρχικών δεδομένων προς αξιολόγηση, με άλλα λόγια μια επανάληψη ανά νόημα. Το υπόλοιπο 80% χρησιμοποιείται για την τελική αξιολόγηση. Τα δεδομένα προσαρμογής δεν έχουν επικάλυψη με τα δεδομένα αξιολόγησης.

Στον Πίνακα 7.16 απεικονίζουμε τα αποτελέσματα αναγνώρισης μετά την προσαρμογή στον νοηματιστή προς αξιολόγηση. Συγκρίνοντας τις μεθόδους SU-P+ITA vs. SU-P+ITA+MLLR παρατηρούμε ότι προσαρμόζοντας τα μοντέλα κάνοντας χρήση της τεχνικής MLLR το ποσοστό αναγνώρισης αυξάνει κατά μέσο όρο για τους δύο νοηματιστές 8.5%, 20.3% και 17.5% για τις ροές M-P, HS και M-P+HS αντιστοίχως. Επιπλέον συγκρίνοντας SU-P+ITA+MLLR vs. SU-P+ITA+MLLR+IP, δηλαδή προσθέτοντας επιπλέον προφορές νοημάτων από τον νοηματιστή προς αξιολόγηση, παρατηρούμε ότι το ποσοστό αναγνώρισης αυξάνει κατά μέσο όρο για τους δύο νοηματιστές 33.3%, 25.3% και 16.7% για τις ροές M-P, HS και M-P+HS αντιστοίχως. Πρέπει να σημειώσουμε ότι τα αποτελέσματα του Πίνακα 7.16 για την SU-P+ITA μέθοδο, διαφέρουν από αυτά των Πινάκων 7.15 και 7.12 εφόσον έχουμε αφαιρέσει από το σύνολο δεδομένων αξιολόγησης τα δεδομένα προσαρμογής.

Συνοψίζοντας μετά την προσαρμογή στο νοηματιστή προς αξιολόγηση χρησιμοποιώντας όλες τις ροές πληροφορίας και το σχήμα προσαρμογής που παρουσιάστηκε στην ενότητα 5.3 το ποσοστό αναγνώρισης αγγίζει το 94.1% κατά μέσο όρο για τους δύο νοηματιστές.

7.4 Οπτικοακουστική αναγνώριση πολυτροπικών χειρονομιών

Αρχικά παρέχουμε πληροφορίες σχετικά με τα χαρακτηριστικά διανύσματα που χρησιμοποιήθηκαν, όπως επίσης λεπτομέρειες σχετικά με τα μοντέλα για κάθε ροή πληροφορίας (ενότητα 7.4.1). Στη συνέχεια, στην ενότητα 7.4.2 παρουσιάζουμε τα σενάρια πειραματισμού και τα αποτελέσματα αναγνώρισης για την προτεινόμενη μέθοδο πολυτροπικής αναγνώρισης και των μεθόδων σύμμειξης που χρησιμοποιήθηκαν. Αυτά συγκρίνονται με προσφάτως δημοσιευμένα αποτελέσματα στην ίδια βάση δεδομένων χρησιμοποιώντας ακριβώς τον ίδιο τρόπο αξιολόγησης.

7.4.1 Χαρακτηριστικά διανύσματα και παράμετροι των HMM μοντέλων

Όπως συζητήθηκε με λεπτομέρεια στην ενότητα 6.3, εκπαιδεύουμε στατιστικά ξεχωριστά HMM για κάθε ροή πληροφορίας: σκελετός, χειρομορφή και φωνή.

Τα χαρακτηριστικά διανύσματα που χρησιμοποιήθηκαν για τη ροή του σκελετού περιέχουν: την τρισδιάστατη θέση των χεριών και των αγκώνων, την ταχύτητα των χεριών, την τρισδιάστατη κατεύθυνση της κίνησης των χεριών και την απόσταση των χεριών. Για τη ροή της χειρομορφής χρησιμοποιούμε τα χαρακτηριστικά διανύσματα που προκύπτουν από την εφαρμογή των HOG περιγραφητών. Οι περιγραφητές HOG εξάχθηκαν σε εικόνες κομμένων χεριών βάθους (depth) και χρώματος (RGB). Για την κατάτμηση των χεριών εφαρμόστηκε κατάτμηση με κατωφλιοποίηση στις εικόνες βάθους, αξιοποιώντας την πληροφορία της θέσης των χεριών που μας δίνεται από τη ροή πληροφορίας του σκελετού. Για τη ροή πληροφορίας της φωνής χρησιμοποιούμε τα Mel Frequency Cepstral Coefficients (MFCC) διανύσματα χαρακτηριστικών αποσκοπώντας στην ανάδειξη των φασματικών ιδιαιτεροτήτων του σήματος της φωνής. Παράγουμε 39 ακουστικά διανύσματα χαρακτηριστικών κάθε 10 χιλιοστά του δευτερολέπτου. Το διάνυσμα χαρακτηριστικών περιλαμβάνει τα 13 MFCC μαζί με τις πρώτες και δεύτερες παραγώγους τους.

Για όλες τις ροές πληροφορίας, εκπαιδεύουμε ένα HMM μοντέλο ανά πολυτροπική χειρονομία και επιπλέον ένα sil και bm μοντέλο όπως περιγράφεται στην ενότητα 6.3. Αυτά τα μοντέλα

ΑΔ	Μεμονωμένες ροές			Σύμμειξη ροών		
	Φωνή	*Σκελετός*	*Χειρομορφή*	Γραμματική	MHS	MHS+SPF
✗	78.4	47.6	13.3	✗	85.8	87.1
✓	87.2	49.1	20.2	✗	91.92	92.28
				✓	**93.06**	**93.33**

Πίνακας 7.17: Αξιολόγηση των μεμονωμένων ροών ανεξάρτητα από τις μεθόδους σύμμειξής τους. Τα ποσοστά αναγνώρισης είναι σε accuracy %. Η συντομογραφία ΑΔ αναφέρεται στην εφαρμογή της πολυτροπικής ανίχνευσης δράσης και η 'Γραμματική' στην εφαρμογή γραμματικής κατά τη διάρκεια του πολυτροπικού σκοραρίσματος (multimodal rescoring). MHS αναφέρεται στην εφαρμογή του πολυτροπικού σκοραρίσματος και SPF στην εφαρμογή της τμηματικής παράλληλης σύμμειξης (βλ. ενότητες 6.2.2 6.2.3).

εκπαιδεύονται είτε χρησιμοποιώντας τις επισημειώσεις που δίνονται με τη βάση δεδομένων, είτε χρησιμοποιώντας ως είσοδο τα αποτελέσματα από τη εφαρμογή της μεθόδου της πολυτροπικής ανίχνευσης δράσης που περιγράψαμε στην ενότητα 6.4. Ο αριθμός των καταστάσεων των HMM, των Γκαουσιανών ανά κατάσταση, τα stream weights και το πέναλτι εισαγωγής λέξης (word insertion penalty) για όλες τις ροές πληροφορίας καθορίζονται πειραματικά μεγιστοποιώντας το ποσοστό αναγνώρισης σε ένα σύνολο δεδομένων επικύρωσης (validation set). Για τη ροή του σκελετού εκπαιδεύουμε left-right HMMs με 12 καταστάσεις και 2 Γκαουσιανές ανά κατάσταση. Για τη ροή της χειρομορφής χρησιμοποιούμε left-right HMMs 8 καταστάσεων και 3 Γκαουσιανών ανά κατάσταση. Τέλος, για τη ροή της φωνής εκπαιδεύουμε left-right HMMs με 22 καταστάσεις και 10 Γκαουσιανές ανά κατάσταση.

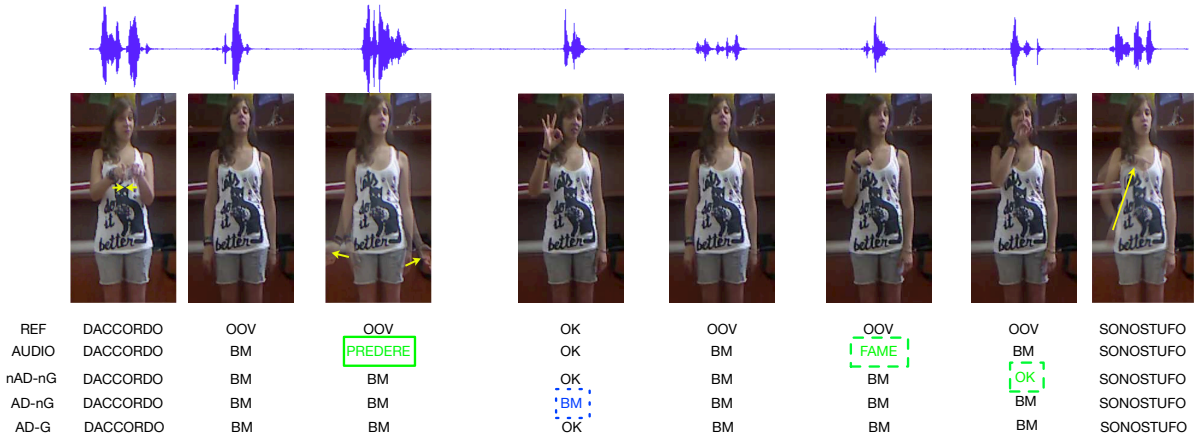
7.4.2 Αποτελέσματα αναγνώρισης

Μεμονωμένες ροές πληροφορίας

Στον Πίνακα 7.17 απεικονίζουμε τα αποτελέσματα αναγνώρισης για κάθε ροή πληροφορίας ξεχωριστά εφαρμόζονται ή όχι την μέθοδο της πολυτροπικής ανίχνευσης δράσης (AD) για την εκπαίδευση των μοντέλων όπως περιγράφεται στις ενότητες 6.3 και 6.4. Η ροή πληροφορίας της φωνής και για τις δύο περιπτώσεις, εμφανίζεται ως η κυρίαρχη ροή υπό την οπτική γωνία της απόδοσης της αναγνώρισης. Σε όλες τις ροές πληροφορίας η εφαρμογή της πολυτροπικής ανίχνευσης δράσης κατά τη διάρκεια της εκπαίδευσης διαδραματίζει σημαντικό ρόλο στην αναγνώριση. Αυτό οφείλεται στο γεγονός ότι μας προσφέρει πιο ακριβή χρονικά όρια επισημειώσεων σε επίπεδο χειρονομιών για κάθε ροή πληροφορίας ξεχωριστά. Έτσι, αντιμετωπίζεται το πρόβλημα το οποίο υπήρχε στις αρχικές επισημειώσεις της βάσης δεδομένων οι οποίες περιείχαν χρονικά τμήματα μη-δράσης στην αρχή και στο τέλος κάθε πολυτροπικής χειρονομίας. Χρησιμοποιώντας αυτά τα πιο ακριβή χρονικά όρια επιτυγχάνουμε την ακριβή μοντελοποίηση της πραγματικής άρθρωσης κάθε πολυτροπικής χειρονομίας με αποτέλεσμα να εκπαιδεύουμε πιο εύρωστα HMM μοντέλα. Αυτό απεικονίζεται και στα αποτελέσματα αναγνώρισης, όπου όπως παρατηρούμε έχουμε απόλυτη αύξηση του ποσοστού αναγνώρισης κατά 8.8%, 1.5% και 6.9% για τις ροές της φωνής, του σκελετού και της χειρομορφής αντίστοιχα.

Σύμμειξη ροών πληροφορίας

Για την αξιολόγηση του προτεινόμενου σχήματος σύμμειξης εστιάζουμε στις βασικές συνιστώσες του ξεχωριστά. Πρώτα εφαρμόζουμε το πολυτροπικό σκοράρισμα MHS. Αυτό σκοράρει όλες τις υποθέσεις αναγνώρισης χρησιμοποιώντας και τις τρεις ροές πληροφορίας και συνδυάζει γραμμικά



Σχήμα 7.10: Ένα παράδειγμα αναγνώρισης μιας ακολουθίας πολυτροπικών χειρονομιών. Στην πρώτη γραμμή φαίνεται το ακουστικό σήμα και στη δεύτερη σειρά το οπτικό σήμα με την απεικόνιση μιας ακολουθίας εικόνων που αντιστοιχεί σε διαφορετικά χρονικά πλαίσια του βίντεο. Οι επισημειώσεις σε επίπεδο χειρονομιών συμβολίζονται με 'REF'. Απεικονίζουμε τα αποτελέσματα αναγνώρισης για τη ροή της φωνής (AUDIO) και του προτεινόμενου σχήματος σύμμειξης εφαρμόζοντας ή όχι την πολυτροπική ανίχνευση δράσης (AD) και τη γραμματική (G) κατά τη διάρκεια του πολυτροπικού σκοραρίσματος. Στην περίπτωση nAD-nG δεν εφαρμόζουμε ούτε την AD ούτε την G, στην περίπτωση AD-nG εφαρμόζουμε την AD αλλά όχι την G και στην περίπτωση AD-G εφαρμόζουμε και την AD αλλά και την G. Τα λάθη επισημειώνονται ως εξής: διαγραφές (μπλε χρώμα) και εισαγωγές (πράσινο χρώμα). Το μοντέλο bm μοντελοποιεί τις πολυτροπικές χειρονομίες εκτός λεξιλογίου (out-of-vocabulary -OOV-).

τα επιμέρους σκορ για τον υπολογισμό ενός τελικού πολυτροπικού σκορ για κάθε υπόθεση. Στη συνέχεια εφαρμόζεται η τμηματική παράλληλη σύμμειξη SPF. Αυτή τροποποιεί την πιο πιθανή υπόθεση αναγνώρισης, συνδυάζοντας τις τρεις ροές πληροφορίας, χρησιμοποιώντας PaHMM, και τα χρονικά όρια κάθε χειρονομίας όπως αυτά έχουν εξαχθεί κατά το πολυτροπικό σκοράρισμα. Περισσότερες λεπτομέρειες αναφέρονται στις ενότητες 6.2.2 και 6.2.3. Στο πολυτροπικό σκοράρισμα μπορούμε είτε να υποχρεώσουμε κάθε ροή πληροφορίας να σκοράρει την κάθε υπόθεση αναγνώρισης ακριβώς όπως είναι (force alignment), είτε να αφήσουμε ένα βαθμό ελευθερίας χρησιμοποιώντας μια γραμματική η οποία επιτρέπει την εισαγωγή ή/και διαγραφή των μοντέλων bm και sil.

Όπως παρατηρούμε στον Πίνακα 7.17, εφαρμόζοντας την MHS μέθοδο τα αποτελέσματα αναγνώρισης αυξάνουν σε σχέση με την χρησιμοποίηση των ροών πληροφοριών μεμονωμένα. Το μέσο σχετικό ποσοστό μείωσης του σφάλματος (relative error reduction -RER-) [1] είναι 38%. Επιπλέον, η εφαρμογή της πολυτροπικής ανίχνευσης δράσης κατά τη διάρκεια της εκπαίδευση επηρεάζει δραστικά την αναγνώριση της MHS μεθόδου επιτυγχάνοντας σχετική μείωση του λάθους κατά 38%. Με την εφαρμογή της γραμματικής κατά το πολυτροπικό σκοράρισμα επιτυγχάνουμε μια επιπλέον σχετική μείωση του λάθους κατά 14%. Αυτό οφείλεται στο γεγονός ότι η συγκεκριμένη γραμματική λαμβάνει υπόψη της, περιπτώσεις όπου είτε δράσεις είτε μη-δράσεις δεν εμφανίζονται ταυτόχρονα σε όλες τις ροές πληροφορίας. Τέλος όπως μπορούμε να παρατηρήσουμε στον Πίνακα 7.17 το καλύτερο ποσοστό αναγνώρισης 93.33% το επιτυγχάνουμε εφαρμόζοντας τη μέθοδο SPF μετά την εφαρμογή της μεθόδου MHS.

[1] Όλα τα σχετικά ποσοστά, αναφέρονται σε σχετική μείωση του ποσοστού λάθους (RER), εκτός και αν δηλώνεται κάτι διαφορετικό.

Θέση	Μέθοδος	Lev. Dist.	Acc.%	RER
-	MHS+SPF	**0.0667**	**93.33**	-
1	iva.mm [143]	0.12756	87.244	+47.6
2	wweight	0.15387	84.613	+56.6
3	E.T. [12]	0.17105	82.895	+60.9
4	MmM	0.17215	82.785	+61.2
5	pptk	0.17325	82.675	+61.4

Πίνακας 7.18: Η προτεινόμενη μέθοδος MHS+SPF σε σύγκριση με τις πρώτες πέντε μεθόδους στον διαγωνισμό Gesture Challenge. Έχουμε συμπεριλάβει το recognition accuracy (Acc. %, την απόσταση Levenshtein (Lev. Dist.) και την σχετική μείωση λάθους (RER) από την προτεινόμενη μέθοδο MHS+SPF.

Ένα παράδειγμα αναγνώρισης

Ένα παράδειγμα αναγνώρισης απεικονίζεται στο Σχήμα 7.10. Σε αυτό το παράδειγμα απεικονίζουμε την ακουστική και οπτική ροή πληροφορίας για μια ακολουθία πολυτροπικών χειρονομιών μαζί με τις αντίστοιχες επισημειώσεις επιπέδου χειρονομίας ('REF'). Επιπλέον, έχουμε συμπεριλάβει το αποτέλεσμα αναγνώρισης για τη ροή πληροφορίας της φωνής και του προτεινόμενου σχήματος σύμμειξης εφαρμόζοντας ή όχι τις δύο βασικές συνιστώσες του:

- την πολυτροπική ανίχνευση δράσης (AD) και
- την γραμματική (G) κατά την διάρκεια του πολυτροπικού σκοραρίσματος.

Στην περίπτωση nAD-nG δεν εφαρμόζουμε ούτε την AD ούτε την G, στην περίπτωση AD-nG εφαρμόζουμε την AD αλλά όχι την G και στην περίπτωση AD-G εφαρμόζουμε και την AD αλλά και την G. Επίσης, παρατηρούμε ότι υπάρχουν αρκετές περιπτώσεις όπου χρήστης αρθρώνει μια χειρονομία εκτός λεξιλογίου (OOV). Αυτό υποδεικνύει τη δυσκολία του προβλήματος εφόσον αυτές οι περιπτώσεις πρέπει να μη ληφθούν υπόψη κατά την αναγνώριση.

Εστιάζοντας στην αναγνώριση χρησιμοποιώντας μόνο τη ροή πληροφορίας της φωνής, παρατηρούμε ότι έχουμε δύο εισαγωγές χειρονομιών ("PREDERE" και "FAME"). Κάνοντας σύμμειξη των ροών είτε στην nAD-nG είτε στην AD-nG περίπτωση οι παραπάνω εισαγωγές διορθώνονται, εφόσον η ενσωμάτωση της οπτικής ροής βοηθάει στην αναγνώριση ότι τα συγκεκριμένα τμήματα αντιστοιχούν σε χειρονομίες εκτός λεξιλογίου. Παρόλα αυτά στις nAD-nG και AD-nG περιπτώσεις έχουμε την εισαγωγή της χειρονομίας "OK" και τη διαγραφή της χειρονομίας "OK" αντίστοιχα. Αντίθετα η προτεινόμενη μέθοδος AD-G αναγνωρίζει σωστά όλη την πρόταση.

Σύγκριση με άλλες μεθόδους

Συγκρίνουμε τα αποτελέσματα αναγνώρισης της προτεινόμενης μεθόδου αναγνώρισης πολυτροπικών χειρονομιών MHS+SPF με άλλες μεθόδους [44] οι οποίες αξιολογήθηκαν στο ίδιο ακριβώς πρόβλημα [2]. Ανάμεσα στις διάφορες ομάδες που συμμετείχαν, απεικονίζουμε τα αποτελέσματα από τις πρώτες τέσσερις, όπως επίσης και το αποτέλεσμα που υποβάλαμε εμείς κατά τη διάρκεια του διαγωνισμού (ομάδα pptk).

Όπως φαίνεται στον Πίνακα 7.18 η προτεινόμενη μέθοδος υπερτερεί όλων των άλλων και οδηγεί σε σχετική μείωση του λάθους κατά 47.6% το λιγότερο. Πρέπει να σημειώσουμε ότι η προτεινόμενη

[2] Σε όλα τα αποτελέσματα που παρουσιάζουμε έχουμε χρησιμοποιήσει τους ίδιους κανόνες αξιολόγησης που τηρήθηκαν κατά τη διάρκεια του διαγωνισμού στον οποίο και συμμετείχαμε (ομάδα pptk). Στον Πίνακα 7.18 έχουμε συμπεριλάβει ως κοινό σημείο αναφοράς την απόσταση Levenshtein η οποία χρησιμοποιήθηκε ως μετρικό στα αποτελέσματα κατά τη διάρκεια του διαγωνισμού [44].

μέθοδος σε σχέση με αυτήν που υποβάλαμε κατά τη διάρκεια του διαγωνισμού επιτυγχάνει 61.4% RER. Οι διαφορές των δύο παραπάνω μεθόδων είναι ότι η προτεινόμενη μέθοδος περιλαμβάνει επιπλέον τα ακόλουθα: α) την ανίχνευση πολυτροπικής δράσης για την εκπαίδευση των HMM μοντέλων, β) την εφαρμογή της SPF μεθόδου επιπλέον της MHS, γ) την εισαγωγή της γραμματικής κατά τη διάρκεια της πολυτροπικής σύμμειξης και δ) αξιοποίηση και των δύο συνόλων δεδομένων (εκπαίδευσης, επικύρωσης) για την εκτίμηση των παραμέτρων των μοντέλων.

Κεφάλαιο 8

Σύνοψη και Κατευθύνσεις για Μελλοντική Έρευνα

Το αντικείμενο της παρούσας διδακτορικής έρευνας θα μπορούσε να συνοψιστεί ως η ανάπτυξη ενός ολοκληρωμένου συστήματος για την αναγνώριση της νοηματικής γλώσσας. Η παρούσα έρευνα επικεντρώνεται στην αυτόματη επεξεργασία βίντεο νοηματικής γλώσσας, στην εξαγωγή χαρακτηριστικών, στη μοντελοποίηση και τελικά στην αναγνώριση νοηματικής γλώσσας. Σε αυτά τα πλαίσια αναπτύχθηκαν μέθοδοι για την οπτική επεξεργασία και εξαγωγή χαρακτηριστικών από βίντεο νοηματικής γλώσσας τα οποία σχετίζονται με τους αρθρωτές όπως τα χέρια και το κεφάλι. Επιπλέον αναπτύχθηκαν στατιστικές μέθοδοι για τη μοντελοποίηση και αναγνώριση της νοηματικής γλώσσας. Τέλος, αναπτύχθηκε ένα σύστημα αναγνώρισης χειρονομιών από πολυτροπικά δεδομένα, τα οποία περιλαμβάνουν οπτικές αλλά και ακουστικές ροές πληροφορίας. Πιο συγκεκριμένα αναπτύχθηκαν αλγόριθμοι για τη μοντελοποίηση και σύμμειξη των πολυτροπικών ροών πληροφορίας, με απώτερο στόχο την ανίχνευση και αναγνώριση πολυτροπικών χειρονομιών.

8.1 Ερευνητική συνεισφορά και συμπεράσματα

Οι ερευνητικές μας συνεισφορές μπορούν να συνοψισθούν στα ακόλουθα σημεία:

- Αναπτύξαμε ένα σύστημα εξαγωγής χαρακτηριστικών σχετιζόμενων με τους αρθρωτές που συμμετέχουν κατά τη διάρκεια άρθρωσης της νοηματικής γλώσσας. Βασιστήκαμε στην ανίχνευση περιοχών χρώματος δέρματος κάνοντας χρήση ενός πιθανοτικού μοντέλου χρώματος. Στη συνέχεια εφαρμόσαμε μορφολογική επεξεργασία και κατάτμηση των εξαγόμενων μασκών δέρματος για την ομαλοποίησή τους. Για την επίλυση των επικαλύψεων, συνδυάσαμε εμπρόσθια, οπίσθια γραμμική πρόβλεψη με την τεχνική ταιριάσματος προτύπου (template matching). Τέλος, εφαρμόσαμε μεθόδους για την εξαγωγή κατάλληλων χαρακτηριστικών διανυσμάτων που σχετίζονται με τη θέση, την κίνηση και το σχήμα των αρθρωτών.

- Αναπτύξαμε ένα καινοτόμο σύστημα για τη μοντελοποίηση και αναγνώριση ΝΓ με δεδομενοκεντρικές υπομονάδες. Για την κατάτμηση κάθε νοήματος στις υπομονάδες από τις οποίες αποτελείται, εφαρμόσαμε έναν αλγόριθμος κατάτμησης σε στατικά και δυναμικά τμήματα ο οποίος βασιζόταν στη μοντελοποίηση του προφίλ της ταχύτητας με τη χρήση κρυφών Μαρκοβιανών μοντέλων. Επιπλέον εφαρμόσαμε αλγορίθμους συσταδοποίησης για τα στατικά και δυναμικά τμήματα για την κατασκευή των στατικών και δυναμικών υπομονάδων. Τέλος συνδυάσαμε τις στατικές και δυναμικές υπομονάδες για την κατασκευή ενός λεξικού επιπέδου υπομονάδας. Το συνολικό σύστημα αξιολογήθηκε σε προτυποποιημένες βάσεις δεδομένων νοηματικής γλώσσας και τα αποτελέσματα έδειξαν ότι η προσεγγιση αυτή παρου-

σιάζει σημαντικές βελτιώσεις σε σύγκριση με σύγχρονες μεθόδους που έχουν παρουσιαστεί στη διεθνή βιβλιογραφία.

- Αναπτύξαμε ένα καινοτόμο σύστημα για τη μοντελοποίηση και αναγνώριση ΝΓ κάνοντας χρήση γλωσσικής-φωνητικής πληροφορίας η οποία ήταν ενσωματωμένη σε PDTS επισημειώσεις. Έτσι, κατασκευάσαμε PDTS υπομονάδες οι οποίες ήταν γλωσσικά και φωνητικά ερμηνεύσιμες. Για την εκπαίδευση των PDTS υπομονάδων χρησιμοποιήσαμε HMMs και επιπλέον προτείναμε τον αλγόριθμο εκπαίδευσης Iterative Training Algorithm (ITA). Τέλος, η αναγνώριση της νοηματικής γλώσσας έγινε χρησιμοποιώντας τα PDTS HMM μοντέλα υπομονάδας και το PDTS λεξικό. Το προτεινόμενο σύστημα οδήγησε σε βελτιώσεις στην αναγνώριση ΝΓ σε σύγκριση με μεθόδους οι οποίες βασίζονται σε δεδομενοκεντρικές υπομονάδες, υποδεικνύοντάς μας έτσι την σπουδαιότητα της χρησιμοποίησης γλωσσικής-φωνητικής πληροφορίας.

- Προτείναμε το πλαίσιο multi-stream switching probability distribution (MSSD) HMM για τη μοντελοποίηση των υπομονάδων. Αυτό το πλαίσιο επιτρέπει με έναν φορμαλιστικό τρόπο να χρησιμοποιούνται διαφορετικά σετ διανυσμάτων χαρακτηριστικών, ανάλογα με τον τύπο της υπομονάδας που μοντελοποιείται χωρίς να γνωρίζουμε εκ των προτέρων τη χρονική κατάτμηση στους διαφορετικούς τύπους υπομονάδων. Επιπλέον επεκτείναμε τα MSSD-HMM μοντέλα με στόχο τη σύμμειξη πολλαπλών ροών πληροφορίας: κίνηση, θέσης και χειρομορφής του κυρίαρχου και δευτερεύοντος χεριού. Τέλος, αναπτύξαμε ένα σύστημα για την προσαρμογή των μοντέλων και του λεξικού υπομονάδας σε άγνωστο νοηματιστή.

- Αναπτύξαμε ένα σύστημα για την ανίχνευση και αναγνώριση χειρονομιών από πολυτροπικά δεδομένα, τα οποία περιελάμβαναν οπτικές αλλά και ακουστικές ροές πληροφορίας. Συγκεκριμένα αυτό το σύστημα εκμεταλλεύεται ροές πληροφοριών που σχετίζονται με το χρώμα, το βάθος και τη φωνή, όπως έχουν καταγραφεί από τον αισθητήρα Kinect. Η μοντελοποίηση για κάθε ροή πληροφορίας βασίστηκε σε HMM μοντέλα χειρονομιών. Για την εκπαίδευση των HMM μοντέλων αναπτύξαμε ένα αυτόματο σύστημα ανίχνευσης δράσης (activity detection). Επιπλέον, προτείναμε ένα πιθανοτικό πολυτροπικό πλαίσιο σύμμειξης βασιζόμενο στην επαναξιολόγηση των υποθέσεων αναγνώρισης και στα PaHMM. Τέλος, το προτεινόμενο σύστημα παρουσίασε σημαντικές βελτιώσεις συγκρινόμενο με προσφάτως δημοσιευμένα αποτελέσματα από μεθόδους που συμμετείχαν στον διαγωνισμό πολυτροπικής αναγνώρισης χειρονομιών CHALEARN.

Με την μέχρι τώρα έρευνά μας συνεισφέραμε στην περιοχή της αναγνώρισης νοηματικής γλώσσας και πολυτροπικών χειρονομιών. 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 στην 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 κάνοντάς την πιο φυσική και άμεση.

8.2 Μελλοντικές ερευνητικές κατευθύνσεις

Σε αυτή την ενότητα κάνουμε μια σύνοψη των μελλοντικών κατευθύνσεων στις οποίες θα μπορούσε να επεκταθεί η παρούσα έρευνά μας.

Μια αρκετά άμεση επέκταση των προτεινόμενων συστημάτων είναι η εφαρμογή τους σε προβλήματα αναγνώρισης συνεχούς νοηματικού λόγου. Στην έρευνά μας ασχοληθήκαμε σε βάθος με τη μοντελοποίηση των υπομονάδων που απαρτίζουν τη νοηματική γλώσσα και πειραματιστήκαμε

κυρίως σε προβλήματα αναγνώρισης μεμονωμένων νοημάτων. Επιπλέον εφαρμόσαμε τις προτεινόμενες μεθόδους σε ένα περιορισμένο πρόβλημα αναγνώρισης συνεχούς νοηματικού λόγου, αλλά παρόλα αυτά μια πιο συστηματική ανάλυση και μελέτη σε βάσεις δεδομένων συνεχούς νοηματικής γλώσσας αποτελεί μια σημαντική επέκταση. Ο συνδυασμός των μεθόδων σύμμειξης, που παρουσιάστηκε στα πλαίσια της αναγνώρισης πολυτροπικών χειρονομιών, με τη μοντελοποίηση των υπομονάδων για τους σκοπούς της αναγνώρισης συνεχούς νοηματικού λόγου, αποτελεί έναν πολλά υποσχόμενο μελλοντικό στόχο.

Ακόμα, υπάρχει η δυνατότητα να επεκταθούν οι κύριες πτυχές του 2-S-U συστήματος που παρουσιάστηκε και βασίζεται στις δεδομενοκεντρικές υπομονάδες. Η γενίκευση του με την χρησιμοποίηση αυτόματων τεχνικών για την επιλογή των κατάλληλων χαρακτηριστικών διανυσμάτων (feature selection) [53] σε κάθε είδος υπομονάδας αποτελεί ενδιαφέρουσα προσέγγιση. Ο συνδυασμός discriminative και generative μοντέλων αποτελεί άλλη μια πιθανή μελλοντική πορεία αυτής της έρευνας.

Άλλες μελλοντικές επεκτάσεις αφορούν στην χρησιμοποίηση σχημάτων σύμμειξης εκμεταλλευόμενοι πρότερη φωνολογική γνώση για τη συσχέτιση των εμπλεκόμενων ροών πληροφορίας, βασιζόμενοι σε γλωσσολογικές ερευνητικές εργασίες [118, 103, 119, 78]. Επιπλέον η χρησιμοποίηση και άλλων ροών πληροφορίας όπως π.χ. οι εκφράσεις του προσώπου και η στάση του σώματος του νοηματιστή [106, 5, 4]. Ακόμα, η εφαρμογή των προτεινόμενων συστημάτων με σκοπό την αυτόματη φωνητική επισημείωση μεγάλων βάσεων δεδομένων νοηματικής γλώσσας, αποτελεί επίσης μια πολύ ενδιαφέρουσα εφαρμογή της παρούσας έρευνας. Η κατεύθυνση αυτή αναμένεται να βοηθήσει την γλωσσολογική έρευνα και ανάλυση μεγάλων βάσεων δεδομένων νοηματικής γλώσσας, όπου οι φωνητικές επισημειώσεις είναι απαραίτητες.

Όσον αναφορά το σύστημα πολυτροπικής αναγνώρισης, πολλές επεκτάσεις του έχουν ενδιαφέρον. Ενδεικτικά παραδείγματα αποτελούν η γενίκευση του συστήματος χρησιμοποιώντας ένα επαναληπτικό σχήμα σύμμειξης το οποίο θα ανατροφοδοτεί τη διαδικασία εκπαίδευσης των στατιστικών μοντέλων όπως και η εκμετάλλευση των στατιστικών εξαρτήσεων μεταξύ των πολλαπλών ροών πληροφορίας. Ακόμα ιδιαίτερο ενδιαφέρον παρουσιάζει η ενσωμάτωση στα υπολογιστικά μοντέλα, πρότερης γνώσης σχετικά με τις χειρονομίες, για παράδειγμα από την πλευρά της γλωσσολογίας των χειρονομιών ή την πολυτροπική εκδοχή τους. Εν κατακλείδι, λαμβάνοντας υπόψη τις δυνατότητες του προτεινόμενου συστήματος και το διεπιστημονικό ενδιαφέρον για την αναγνώριση πολυτροπικών χειρονομιών το προτεινόμενο σύστημα καλεί για περαιτέρω επέκταση και ανάλυση.

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] T. Ahmad, C.J. Taylor, and T.F. Lanitis, A. Cootes. Tracking and recognising hand gestures, using statistical shape models. *Image and Vision Computing*, 15(5):345–352, 1997.

[3] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff. A unified framework for gesture recognition and spatiotemporal gesture segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(9):1685–1699, 2009.

[4] E. Antonakos, V. Pitsikalis, and P. Maragos. Classification of extreme facial events in sign language videos. *EURASIP Journal on Image and Video Processing*, 2014(1):14, 2014.

[5] E. Antonakos, V. Pitsikalis, I. Rodomagoulakis, and P. Maragos. Unsupervised classification of extreme facial events using active appearance models tracking for sign language videos. In *Proc. Int'l Conf. on Image Processing*, pages 1409–1412. IEEE, 2012.

[6] O. Aran and L. Akarun. A multi-class classification strategy for fisher scores: application to signer independent sign language recognition. *Pattern Recognition*, 43(5):1776–1788, 2010.

[7] A. Argyros and M. Lourakis. Real time tracking of multiple skin-colored objects with a possibly moving camera. In *Proc. European Conf. on Computer Vision*, 2004.

[8] V. Athitsos, C. Neidle, S. Sclaroff, J. Nash, A. Stefan, Q. Yuan, and A. Thangali. The american sign language lexicon video dataset. In *Proc. Conf. on Computer Vision & Pattern Recognition Workshops*, pages 1–8. IEEE, 2008.

[9] G. Awad, J. Han, and A. Sutherland. Novel boosting framework for subunit-based sign language recognition. In *Proc. Int'l Conf. on Image Processing*, pages 2729–2732. IEEE, 2009.

[10] B. Bauer, H. Hienz, and K.-L. Kraiss. Video-based continuous sign language recognition using statistical methods. *Proc. Int'l Conf. on Pattern Recognition*, 2000.

[11] B. Bauer and K. F. Kraiss. Towards an automatic sign language recognition system using subunits. In *Proc. of Int'l Gesture Workshop*, volume 2298, pages 64–75, 2001.

[12] I. Bayer and S. Thierry. A multi modal approach to gesture recognition from audio and video data. In *Proc. of the 15th ACM on International Conf. on multimodal interaction*, pages 461–466. ACM, 2013.

[13] P. Bernardis and M. Gentilucci. Speech and gesture share the same communication system. *Neuropsychologia*, 44(2):178–190, 2006.

[14] H. Birk, T.B. Moeslund, and C.B. Madsen. Real-time recognition of hand alphabet gestures using principal component analysis. In *Proc. Scandinavian Conf. Image Analysis*, 1997.

[15] R. A. Bolt. "put-that-there": Voice and gesture at the graphics interface. In *Proc. of the 7th annual Conf. on Computer Graphics and Interactive Techniques*, volume 14. ACM, 1980.

[16] 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.

[17] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady. A linguistic feature vector for the visual interpretation of sign language. In *Proc. European Conf. on Computer Vision*, 2004.

[18] 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, Jun. 2009.

[19] P. Buehler, M. Everingham, and A Zisserman. Employing signed TV broadcasts for automated learning of British Sign Language. In *"Proc. of Workshop on Representation and Processing of SL: Corpora and Sign Language Technologies"*, 2010.

[20] J. Cai and A. Goshtasby. Detecting human faces in color images. *Image and Vision Computing*, 18:63–75, 1999.

[21] S. Celebi, A. S. Aydin, T. T. Temiz, and T. Arici. Gesture recognition using skeleton data with weighted dynamic time warping. *Computer Vision Theory and Applications. Visapp*, 2013.

[22] CHALEARN. Chalearn. [Online], 2013. [Accessed 15 Aug 2013].

[23] 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 Vision Computing*, 21(8):745–758, 2003.

[24] X. Chen and M. Koskela. Online rgb-d gesture recognition with extreme learning machines. In *Proc. of the 15th ACM on Int'l Conf. on multimodal interaction*, pages 467–474. ACM, 2013.

[25] Y. L. Chow and R. Schwartz. The n-best algorithm: An efficient procedure for finding top n sentence hypotheses. In *Proc. of the Workshop on Speech and Natural Language*, pages 199–202, 1989.

[26] S. Conseil, S. Bourennane, and L. Martin. Comparison of Fourier descriptors and Hu moments for hand posture recognition. In *Proc. European Conf. on Signal Processing*, 2007.

[27] H. Cooper, B. Holt, and R. Bowden. Sign language recognition. In *Visual Analysis of Humans*, pages 539–562. Springer, 2011.

[28] H. Cooper, E.J. Ong, N Pugeault, and R. Bowden. Sign language recognition using sub-units. *Journal of Machine Learning Research*, 13:2205–2231, 2012.

[29] T.F. Cootes and C.J. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, 2004.

[30] 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.

[31] D. Corina and W. Sandler. On the nature of phonological structure in sign language. *Phonology*, 10:165–207, 2008.

[32] G. Coulter. On the nature of asl as a monosyllabic language. In *Annual Meeting of the Linguistic Society of America, San Diego, CA*, 1982.

[33] Y. Cui and J. Weng. Appearance-based hand sign recognition from intensity image sequences. *Computer Vision and Image Understanding*, 78(2):157–176, 2000.

[34] N. Dalal and B. Triggs. Histogram of oriented gradients for human detection. In *Proc. Conf. on Computer Vision & Pattern Recognition*, 2005.

[35] L. Davies, David and W. Bouldin, Donald. A cluster separation measure. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1:224 – 227, April 1979.

[36] K. Derpanis and J. Wildes, R.and Tsotsos. Definition and recovery of kinematic features for recognition of american sign language movements. *Image and Vision Computing*, 26(12):1650–1662, 2008.

[37] Dicta-Sign Project. Corpus annotations. [Online], 2012. [Accessed 2 May 2012].

[38] A. Dimou, V. Pitsikalis, T. Goulas, S. Theodorakis, P. Karioris, M. Pissaris, and S-E. Fotinea. A machine learning dedicated gsl phrases corpus: Creation, acquisition and implementation. In *Proc. Int'l Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, 2012.

[39] L. Ding and AM. Martinez. Modelling and recognition of the linguistic components in american sign language. *Image and Vision Computing*, 27:1826 – 1844, 2009.

[40] A.C. Downton and H. Drouet. Model-based image analysis for unconstrained human upper-body motion. In *Proc. Int'l Conf. on Image Processing and Its Applications*, pages 274–277, Apr. 1992.

[41] P. Dreuw, C. Neidle, V. Athitsos, S. Sclaroff, and H. Ney. Benchmark databases for video-based automatic sign language recognition. In *Proc. Language Resources Evaluation Conf.*, May 2008.

[42] W. Du and J. Piater. Hand modeling and tracking for video-based sign language recognition by robust principal component analysis. In *Proc. ECCV Workshop on Sign, Gesture and Activity*, September 2010.

[43] S. Escalera, J. Gonzàlez, X. Barò, M. Reyes, I. Guyon, V. Athitsos, H. Escalante, L. Sigal, A. Argyros, C. Sminchisescu, R. Bowden, and S. Sclaroff. Chalearn multi-modal gesture recognition 2013: grand challenge and workshop summary. In *Proc. of the 15th ACM on Int'l Conf. on multimodal interaction*, pages 365–368. ACM, 2013.

[44] S. Escalera, J. Gonzalez, X. Baro, M. Reyes, O. Lopes, I. Guyon, V. Athistos, and H.J. Escalante. Multi-modal Gesture Recognition Challenge 2013: Dataset and Results. In *15th ACM Int'l Conf. on Multimodal Interaction, ChaLearn Challenge and Workshop on Multi-modal Gesture Recognition*. ACM, 2013.

[45] G. Fang, W. Gao, X. Chen, C. Wang, and J. Ma. Signer-independent continuous sign language recognition based on SRN/HMM. pages 163–197. Springer, 2002.

[46] 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, Part A: Systems and Humans*, 37:1–9, 2007.

[47] 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*, volume 4, pages 454–457, 2004.

[48] A. Farhadi, D. Forsyth, and R. White. Transfer learning in sign language. In *Proc. Conf. on Computer Vision & Pattern Recognition*, pages 1–8. IEEE, 2007.

[49] H. Fillbrandt, S. Akyol, and K.-F. Kraiss. Extraction of 3d hand shape and posture from images sequences from sign language recognition. In *Int'l Conf. on Analysis and modeling of faces and gestures*, pages 181–186, 2003.

[50] J. Foote. An overview of audio information retrieval. *Multimedia Systems*, 7(1):2–10, 1999.

[51] M.J.F. Gales and PC Woodland. Mean and variance adaptation within the MLLR framework. *Compure Speech and Language*, 10(4):249–264, 1996.

[52] H. Glotin, D. Vergyr, C. Neti, G. Potamianos, and J. Luettin. Weighting schemes for audio-visual fusion in speech recognition. In *Int'l Conf. on Acoustics, Speech and Signal Processing*, volume 1, pages 173–176, 2001.

[53] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[54] Y. Gweth, C. Plahl, and H. Ney. Enhanced continuous sign language recognition using pca and neural network features. In *Proc. Conf. on Computer Vision & Pattern Recognition Workshops*, pages 55–60. IEEE, 2012.

[55] J. Han, G. Awad, and A. Sutherland. Modelling and segmenting subunits for sign language recognition based on hand motion analysis. *Pattern Recognition Letters*, 30:623–633, 2009.

[56] 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 4th Int'l Conf. on Language Resources and Evaluation, Lisbon, Portugal*, 2004.

[57] H.J.A.M. Heijmans. *Morphological Image Operators*. Acad. Press, Boston, 1994.

[58] A. Hernández-Vela, M. Bautista, X. Perez-Sala, V. Ponce-Lopez, S. Escalera, X. Baro, O. Pujol, and C. Angulo. Probability-based dynamic time warping and bag-of-visual-and-depth-words for human gesture recognition in rgb-d. *Pattern Recognition Letters*, 2013.

[59] K. Hienz, H. Grobel and G. Offner. Real-time hand-arm motion analysis using a single video camera. In *Int'l Conf. on Automatic Face & Gesture Recognition*, pages 323–327, 1996.

[60] 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.

[61] M-K. Hu. Visual pattern recognition by moment invariants. *IEEE Trans. on Information Theory*, 8(2):179–187, February 1962.

[62] C.-L. Huang and S.-H. Jeng. A model-based hand gesture recognition system. *Machine Vision and Application*, 12(5):243–258, 2001.

[63] J. Iverson and S. Goldin-Meadow. Why people gesture when they speak. *Nature*, 396(6708):228–228, 1998.

[64] R. E. Johnson and S. K. Liddell. *A Segmental Framework for Representing Signs Phonetically*, volume 11. 2011.

[65] R. E. Johnson and S. K. Liddell. *Toward a Phonetic Representation of Signs, I: Sequentiality and Contrast*, volume 11. 2011a.

[66] T. Kadir, R. Bowden, E. J. Ong, and A. Zisserman. Minimal training, large lexicon, unconstrained sign language recognition. In *Proc. British Machine Vision Conf.*, 2004.

[67] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122, Mar. 2007.

[68] S. D. Kelly, A. Özyürek, and E. Maris. Two sides of the same coin speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2):260–267, 2010.

[69] A. Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press, 2004.

[70] E.S. Klima and U. Bellugi. *The signs of language*. Harvard Univ. Press, 1979.

[71] O. Koller, H. Ney, and R. Bowden. May the force be with you: Force-aligned signwriting for automatic subunit annotation of corpora. In *Int'l Conf. on Automatic Face & Gesture Recognition*, 2013.

[72] W. Kong and S. Ranganath. Sign language phoneme transcription with rule-based hand trajectory segmentation. *J. Signal Processing Systems*, 59:211–222, 2010.

[73] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. Conf. on Computer Vision & Pattern Recognition*, volume 2, pages 2169–2178. IEEE, 2006.

[74] H-K. Lee and J-H. Kim. An HMM-based threshold model approach for gesture recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(10):961–973, 1999.

[75] J. Li and N. M. Allinson. Simultaneous gesture segmentation and recognition based on forward spotting accumulative hmms. *Pattern Recognition*, 40(11):3012–3026, 2007.

[76] J.F. Lichtenauer, E.A. Hendriks, and MJ Reinders. Sign language recognition by combining statistical dtw and independent classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(11):2040, 2008.

[77] S. K. Liddell. Think and believe: Sequentiality in american sign language. *Language*, 60(2):372–399, 1984.

[78] S. K. Liddell and R. E. Johnson. American sign language: The phonological base. *Sign Lang. St.*, 64:195 – 277, 1989.

[79] S. Liwicki and M. Everingham. Automatic recognition of fingerspelled words in British sign language. In *Proc. CVPR Workshop on Human Communicative Behavior Analysis*, 2009.

[80] D. G Lowe. Distinctive image features from scale-invariant keypoints. *IntUl Journal of Computer Vision*, 60(2):91–110, 2004.

[81] P. Maragos. *The Image and Video Processing Handbook*, chapter Morphological Filtering for Image Enhancement and Feature Detection. Elsevier, 2005.

[82] I. Matthews and S. Baker. Active appearance models revisited. *Int'l Journal of Computer Vision*, 60(2):135–164, 2004.

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

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

[85] D. Morris, P. Collett, P. Marsh, and O"Shaughnessy. *Gestures: their origins and distribution*. Stein and Day, 1979.

[86] EW. Myers. An O(ND) Difference Algorithm and Its Variations. *Algorithmica*, 1:251–266, 1986.

[87] Y. Nam and K. Wohn. Recognition of space-time hand-gestures using hidden Markov model. In *ACM symposium on Virtual reality software and technology*, pages 51–58. Citeseer, 1996.

[88] K. Nandakumar, K. W. Wan, S. Chan, W. Ng, J. G. Wang, and W. Y. Yau. A multi-modal gesture recognition system using audio, video, and skeletal joint data. In *Proc. of the 15th ACM on Int'l Conf. on multimodal interaction*, pages 475–482. ACM, 2013.

[89] S. Nayak, K. Duncan, S. Sarkar, and B. Loeding. Finding recurrent patterns from continuous sign language sentences for automated extraction of signs. *J. of Mach. Learn. Res.*, 13:2589–2615, 2012.

[90] C. Neidle. Signstream annotation: Addendum to conventions used for the american sign language linguistic research project. Technical report, 2007.

[91] C. Neidle and C. Vogler. A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface. In *"Proc. of 5th Workshop on Representation and Processing of SL: Interactions between Corpus and Lexicon"*, 2012.

[92] N. Neverova, C. Wolf, G. Paci, G. Sommavilla, G. Taylor, and F. Nebout. A multi-scale approach to gesture detection and recognition. In *Proc. of the IEEE Int'l Conf. on Computer Vision Wrksp*, pages 484–491, 2013.

[93] I. Oikonomidis, N. Kyriazis, and A. Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *Proc. British Machine Vision Conf.*, pages 1–11, 2011.

[94] I. Oikonomidis, N. Kyriazis, and A. Argyros. Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints. In *Proc. Int'l Conf. on Computer Vision*, pages 2088–2095. IEEE, 2011.

[95] I. Oikonomidis, N. Kyriazis, and A. Argyros. Tracking the articulated motion of two strongly interacting hands. In *Proc. Conf. on Computer Vision & Pattern Recognition*, pages 1862–1869. IEEE, 2012.

[96] E.-J. Ong and R. Bowden. A boosted classifier tree for hand shape detection. In *Int'l Conf. on Automatic Face & Gesture Recognition*, pages 889–894, 2004.

[97] S. Ong and S. Ranganath. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27:873–891, 2005.

[98] S. Ong and S. Ranganath. A new probabilistic model for recognizing signs with systematic modulations. In *Int'l Conf. on Analysis and modeling of faces and gestures*, pages 16–30, 2007.

[99] M. Ostendorf, A. Kannan, S. Austin, O. Kimball, R. M. Schwartz, and J. R. Rohlicek. Integration of diverse recognition methodologies through reevaluation of n-best sentence hypotheses. In *HLT*, 1991.

[100] G. Pavlakos, S. Theodorakis, V. Pitsikalis, A. Katsamanis, and Maragos P. Kinect-based multimodal gesture recognition using a two-pass fusion scheme. In *Proc. Int'l Conf. on Image Processing*, 2014.

[101] V. Pitsikalis, A. Katsamanis, S. Theodorakis, and Maragos P. Multimodal gesture recognition via multiple hypotheses rescoring. 2014 (under review).

[102] 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. Conf. on Computer Vision & Pattern Recognition Workshops*, 2011.

[103] S. Prillwitz, R. Leven, H. Zienert, R. Zienert, T. Hanke, and J. Henning. HamNoSys. Version 2.0. *Int'l Studies on SL and Comm. of the Deaf*, 7:225–231, 1989.

[104] L. R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[105] L.R. Rabiner and B.H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[106] 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 *9th Int'l Gesture Workshop on Gestures in Embodied Communication and Human-Computer Interaction*, pages 96–99, 2011.

[107] R. C. Rose. Discriminant wordspotting techniques for rejecting non-vocabulary utterances in unconstrained speech. In *Int'l Conf. on Acoustics, Speech and Signal Processing*, volume 2, pages 105–108. IEEE, 1992.

[108] R. C. Rose and D. B. Paul. A hidden markov model based keyword recognition system. In *Int'l Conf. on Acoustics, Speech and Signal Processing*, pages 129–132, 1990.

[109] 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. ECCV Workshop on Sign, Gesture and Activity*, 2010.

[110] A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Dynamic affine-invariant shape-appearance model for hand tracking and feature extraction in sign language handshape classification. *Journal of Machine Learning Research*, 2011.

[111] W. Sandler. *Sequentiality and simultaneity in American Sign Language phonology*. PhD thesis, Univ. of Texas at Austin, 1987.

[112] O. Sharon and P. Cohen. Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM*, 43(3):45–53, 2000.

[113] J. Sherrah and S. Gong. Resolving visual uncertainty and occlusion through probabilistic reasoning. In *Proc. British Machine Vision Conf.*, pages 252–261, 2000.

[114] R. Shwartz and S. Austin. A comparison of several approximate algorithms for finding multiple N-Best sentence hypotheses. In *Int'l Conf. on Acoustics, Speech and Signal Processing*, 1991.

[115] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer, 2004.

[116] T. Starner and A. Pentland. Real-time american sign language recognition from video using hidden markov models. In *Motion-Based Recognition*, pages 227–243. Springer, 1997.

[117] 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(12):1371–1375, Dec. 1998.

[118] W. C. Stokoe. Sign language structure. *Ann. Rev. of Anthr.*, 9:365–390, 1980.

[119] V. Sutton. *Sign writing*. Deaf Action Committee (DAC), 2000.

[120] G.J. Sweeney and A.C. Downton. Towards appearance-based multi-channel gesture recognition. In *Proc. of Int'l Gesture Workshop*, pages 7–16, 1996.

[121] L. N. Tan, B. J. Borgstrom, and A. Alwan. Voice activity detection using harmonic frequency components in likelihood ratio test. In *Int'l Conf. on Acoustics, Speech and Signal Processing*, pages 4466–4469. IEEE, 2010.

[122] N. Tanibata, N. Shimada, and Y. Shirai. Extraction of hand features for recognition of sign language words. In *Proc. of the Int'l Conf. on Vision Interface*, pages 391–398, 2002.

[123] 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 Automatic Face & Gesture Recognition*, pages 54–61, 2000.

[124] A. Thangali, J.P. Nash, S. Sclaroff, and C. Neidle. Exploiting phonological constraints for handshape inference in asl video. In *Proc. Conf. on Computer Vision & Pattern Recognition*, pages 521–528. IEEE, 2011.

[125] S. Theodorakis, V. Pitsikalis, and P. Maragos. Experiments' Data Reference Webpage. http://cvsp.cs.ntua.gr/research/sign/2su [Online], Nov. 2013. [Accessed 12 Nov. 2013].

[126] S. Theodorakis, V. Pitsikalis, and P. Maragos. Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition. *Image and Vision Computing*, 32(8):533–549, 2014.

[127] S. Theodorakis, V. Pitsikalis, and P. Maragos. Linguistic-phonetic subunits and lexicon for sign language recognition. *IEEE Trans. on Audio, Speech and Language Processing*, 2014 (under review).

[128] 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.

[129] Keiichi Tokuda, Takashi Masuko, Noboru Miyazaki, and Takao Kobayashi. Multi-space probability distribution hmm. 85(3):455–464, 2002.

[130] M. Turk. Multimodal interaction: A review. *Pattern Recognition Letters*, 36:189–195, 2014.

[131] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(3):480–492, 2012.

[132] C. Vogler. Extraction of segmental phonetic structures from hamnosys annotations. Technical Report D4.2, Institute for Language and Speech Processing, Greece, January 2011.

[133] C. Vogler and D. Metaxas. Adapting hidden markov models for asl recognition by using three-dimensional computer vision methods. In *Proc. Int'l Conf. on System, Man and Cybernetics*, volume 1, pages 156–161, 1997.

[134] C. Vogler and D. Metaxas. Parallel hidden markov models for american sign language recognition. In *Proc. Int'l Conf. on Computer Vision*, volume 1, page 116, 1999.

[135] C. Vogler and D. Metaxas. Toward scalability in ASL recognition: Breaking down signs into phonemes. *Gesture-Based Comm. in HCI*, pages 211–224, 1999.

[136] C. Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of american sign language. *Computer Vision and Image Understanding*, 81:358, 2001.

[137] Chan Wah Ng and Surendra Ranganath. Real-time gesture recognition system and application. *Image and Vision Computing*, 20(13):993–1007, 2002.

[138] H. Wang, A. Stefan, S. Moradi, V. Athitsos, C. Neidle, and F. Kamangar. A system for large vocabulary sign search. In *Proc. ECCV Workshop on Sign, Gesture and Activity*, volume 1. IEEE, 2010.

[139] S. B. Wang, A. Quattoni, L. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In *Proc. Conf. on Computer Vision & Pattern Recognition*, volume 2, pages 1521–1527. IEEE, 2006.

[140] L. D Wilcox and M. Bush. Training and search algorithms for an interactive wordspotting system. In *Int'l Conf. on Acoustics, Speech and Signal Processing*, volume 2, pages 97–100. IEEE, 1992.

[141] J. Wilpon, L. R. Rabiner, C.-H. Lee, and E. R. Goldman. Automatic recognition of keywords in unconstrained speech using hidden Markov models. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 38(11):1870–1878, 1990.

[142] A. Wilson and A. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21:884–900, 1999.

[143] J. Wu, J. Cheng, C. Zhao, and H. Lu. Fusing multi-modal features for gesture recognition. In *Proc. of the 15th ACM on Int'l conf. on multimodal interaction*, pages 453–460. ACM, 2013.

[144] 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.

[145] H. 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(7):1264–1277, 2009.

[146] M.-H. Yang, N. Ahuja, and M. Tabb. Extraction of 2d motion trajectories and its application to hand gesture recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(8):1061–1074, Aug. 2002.

[147] R. Yang and S. Sarkar. Detecting coarticulation in sign language using conditional random fields. In *Proc. Int'l Conf. on Pattern Recognition*, volume 2, pages 108–112. IEEE, 2006.

[148] 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 *Int'l Conf. on Acoustics, Speech and Signal Processing*, pages 4757–4760, 2009.

[149] S. Young, G. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland. *The HTK Book*. Entropic Cambridge Research Laboratory, Cambridge, United Kingdom, 2002.

[150] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Woodland, and P. Valtchevand. *The HTK Book*. Entropic Ltd., 1999.

[151] S J. Young, N.H. Russell, and J.H.S. Thornton. Token passing: a simple conceptual model for connected speech recognition system. Technical report, Cambridge University Electrical Engineering Department, 1989.

[152] J. Zieren and K-F. Kraiss. Robust person-independent visual sign language recognition. *Patter Recognition and Image Analysis*, pages 333–355, 2005.

[153] 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

Τα αποτελέσματα της έρευνας στην διδακτορικής μας διατριβή 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 handshape features and classification in sign language videos. *The Journal of Machine Learning Research* 14 (1), 1627-1663. 2013.

2. S. Theodorakis and V. Pitsikalis, and P. Maragos. Dynamic–Static Unsupervised Sequentiality, Statistical Subunits and Lexicon for Sign Language Recognition. *Image and Vision Computing* 32 (8), 533-549. 2014.

3. V. Pitsikalis, A. Katsamanis, S. Theodorakis, and P. Maragos. Multimodal Gesture Recognition via Multiple Hypotheses Rescoring. *Journal of Machine Learning Research* (to appear).

4. S. Theodorakis, V. Pitsikalis, and P. Maragos. Linguistic Phonetic Subunits and Lexicon for Sign Language Recognition. *IEEE Transactions on Audio Speech and Language Processing* (under submission).

Δημοσιεύσεις σε Διεθνή Συνέδρια με Κριτές

1. G. Pavlakos, S. Theodorakis, V. Pitsikalis, A. Katsamanis, and P. Maragos. Kinect-Based Multimodal Gesture Recognition Using a Two-Pass Fusion Scheme. In *Proc. Int'l Conf. on Image Processing*, 2014.

2. 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.

3. A. Dimou, V. Pitsikalis, T. Goulas, S. Theodorakis, P. Karioris, M. Pissaris, S–E. Fotinea, E. Efthimiou and P. Maragos. A machine learning dedicated GSL phrases corpus: Creation, acquisition and implementation. In *Proc. Int'l Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Satellite Workshop to the eighth International Conference on Language Resources and Evaluation, Istanbul, Turkey*, 2012.

4. 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**)

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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, Dicta-Sign, D2.1, EU, January 2010.

2. P. Maragos, V. Pitsikalis, S. Theodorakis, A. Roussos, and I. Rodomagoulakis. Progress report on multimodal fusion. Technical Report, Dicta-Sign, D2.2, EU, February 2011.

3. P. Maragos, V. Pitsikalis, S. Theodorakis, Rodomagoulakis I., and Antonakos E. Report on integrated continuous sign recognition. Technical Report, Dicta-Sign, D2.3, EU, February 2012.