Εθνικό Μετσόβιο Πολυτεχνείο

Σχολη Ηλεκτρολογων Μηχανικων
Και Μηχανικων Υπολογιστων

Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Εμψύχωση Συνθετικών Χαρακτήρων - Ανάλυση Συναισθήματος στην Αλληλεπίδραση Ανθρώπου-Μηχανής

Διδακτορική Διατριβή του

ΓΕΩΡΓΙΟΥ Λ. ΚΑΡΥΔΑΚΗ

Αθήνα, Ιούλιος 2009

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων & Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης
και Υπολογιστων

Εμψύχωση Συνθετικών Χαρακτήρων - Ανάλυση Συναισθήματος στην Αλληλεπίδραση Ανθρώπου-Μηχανής

Διδακτορική Διατριβή

του

ΓΕΩΡΓΙΟΥ Λ. ΚΑΡΥΔΑΚΗ

Συμβουλευτική Επιτροπή: Στέφανος Κόλλιας
Ανδρέας–Γεώργιος Σταφυλοπάτης
Κώστας Καρπούζης

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την ..... 2009.

...	...	...
Σ. Κόλλιας	Α.–Γ. Σταφυλοπάτης	Κ. Καρπούζης
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Ερευνητής Β Ε.Π.Ι.Σ.Ε.Υ.-Ε.Μ.Π.
...	...	...
Π. Μαραγκός	Π.Τσανάκας	Γ.Στάμου
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Λέκτορας Ε.Μ.Π.

...
Α. Ντελόπουλος
Επ. Καθηγητής Α.Π.Θ.

Αθήνα, Ιούλιος 2009

...

ΓΕΩΡΓΙΟΥ Λ. ΚΑΡΥΔΑΚΗ

στους γονείς και
την οικογένεια μου

Πρόλογος

Η διαδρομή της διδακτορικής διατριβής μου ήταν από τις πιο ενδιαφέρουσες της ζωής μου και επηρέασε σε μεγάλο βαθμό την προσωπικότητα μου, αλλά κυρίως τον τρόπο σκέψης μου, προσέγγισης και αντιμετώπισης καταστάσεων και προκλήσεων της ζωής. Ξεκινώντας αυτή τη διαδρομή δεν θα μπορούσα να είχα φανταστεί τον πλούτο γνώσεων και ιδεών που θα συναντούσα, τις προκλήσεις που θα αντιμετώπιζα αλλά και τη δημιουργική ερεύνα που οδήγησε στην παρούσα διδακτορική διατριβή.

Κατά την πορεία αυτή αμέριστη ήταν η στήριξη και το ενδιαφέρον του επιβλέποντα της διατριβής καθηγητή ΕΜΠ κ. Στέφανου Κόλλια. Προσέφερε με εξαιρετική διάθεση τις γνώσεις, την εμπειρία και την ερευνητική καθοδήγηση του όποτε του ζητήθηκε και για αυτό τον ευχαριστώ ιδιαίτερα. Ευχαριστίες οφείλω και στους καθηγητές ΕΜΠ κ. Ανδρέα Σταφυλοπάτη και κ. Πέτρο Μαραγκό. Ξεχωριστό ρόλο στην ολοκλήρωση της διατριβής διαδραμάτισε ο διδάκτωρ ΕΜΠ και Ερευνητής Β Ε.Π.Ι.Σ.Ε.Υ. κ. Κώστας Καρπούζης, ο οποίος είχε καθημερινή ενασχόληση και εμπλοκή με την ερευνητική μου εργασία και πάντα με παρότρυνε και με συμβούλευε για τις κατευθύνσεις της ερεύνας μου. Η πρακτική και φιλική προσέγγιση του, χωρίς να υστερεί σε επιστημονική ακεραιότητα, αποδείχθηκε εξαιρετικά χρήσιμη και για αυτό τον ευχαριστώ θερμά. Ο διδάκτωρ ΕΜΠ Αθανάσιος Δροσόπουλος ήταν πάντα αυστηρός κριτής των ερευνητικών ιδεών και προτάσεων μου με στόχο όχι να με αποθαρρύνει αλλά αντίθετα να με παροτρύνει να τις εξελίξω περαιτέρω, ώστε να αποκτήσουν πραγματική πρωτοτυπία και επιστημονική ολοκλήρωση. Για αυτή ακριβώς την αυστηρή κριτική του, τη διαφορετική οπτική που μου μετέδωσε, όσο αφορά στην ερευνητική δραστηριότητα, αλλά κυρίως για την παρότρυνσή του να εμπλακώ στην περιπέτεια του διδακτορικού, τον ευχαριστώ θερμά.

Αξιομνημόνευτη είναι η συνεργασία μου και ερευνητική συμβίωση μου με όλα τα μέλη του εργαστηρίου και κυρίως με τους Αμαρυλλίς Ραουζαίου, Σπύρο Ιωάννου, Χρήστο Πατερίτσα, Μηνά Περτσελάκη, Διδάκτορες ΕΜΠ, Λώρη Μαλατέστα, Παρασκευή Τζούβελη, Στέλιο Αστεριάδη, Όλγα Διαμαντή και Σταύρο Θεοδωράκη, Υποψήφιους Διδάκτορες ΕΜΠ. Επιπλέον, άριστη συνεργασία είχα με ερευνητές εκτός ΕΜΠ τόσο σε αναπτυξιακό όσο και σε ερευνητικό επίπεδο: κ. Ελένη Ευθυμίου και κ. Εβίτα Φωτεινέα διδάκτορες ΕΚΠ και ΕΜΠ αντίστοιχα και ερευνητές ΙΕΛ, την Ginevra Castellano και Loic Kessous διδάκτορες University of Genoa και Universite Paris 8 αντίστοιχα, τους καθηγητές LIMSI-CNRS Catherine Pelachaud και Jean-Claude Martin και τις ερευνητικές ομάδες τους και τους ευχαριστώ όλους ιδιαίτερα. Τέλος, θα ήταν αμέλεια να μην ευχαριστήσω όλους τους δασκάλους και καθηγητές μου, ενώ ιδιαίτερη μνεία θα ήθελα να κάνω στον καθηγητή μου Φίλιππο Άννινο που συνέβαλε τα μέγιστα στη διαμόρφωση της ακαδημαϊκής μου πορείας και ουσιαστικά με έκανε να αγαπήσω τη γνώση, αλλά δυστυχώς δεν είναι πια κοντά μας.

Γιώργος Καρυδάκης

Περίληψη

Η παρούσα διδακτορική διατριβή κινείται στο πλαίσιο της συναισθηματικής υπολογιστικής (affective computing) και της αλληλεπίδρασης ανθρώπου μηχανής (human computer interaction). Τα κεφάλαια 2 και 3 αφορούν στην συναισθηματική υπολογιστική και ειδικότερα : α) στην αναγνώριση δυναμικών συναισθηματικών καταστάσεων από πολλαπλές μορφές πληροφορίας κατά την φυσική επικοινωνία ανθρώπου μηχανής αλλά και στην διαδικασία εντοπισμού και προσαρμογής σε δυναμικές συνθήκες και αλλαγές του εννοιολογικού πλαισίου της αλληλεπίδρασης και β) στην τυποποίηση εκφραστικών παραμέτρων χειρονομιών κατά την συναισθηματικά εμπλουτισμένη αλληλεπίδραση και στην αξιολόγηση της τυποποίησης αυτής μέσω εικονικών χαρακτήρων από ανθρώπινους χρήστες. Το κεφάλαιο 4 αναφέρεται στην αναγνώριση χειρονομιών ως μέσο αλληλεπίδρασης εναλλακτικό των καθιερωμένων, αλλά και στην αναγνώριση και σύνθεση νοηματικής γλώσσας (ενότητες 4.2 και 4.3 αντίστοιχα) συνεισφέροντας στις έννοιες της καθολικής πρόσβασης (universal access) και υποστηρικτικής τεχνολογίας (assistive technology). Η αναγνώριση χειρονομιών αφορά και στην συναισθηματική υπολογιστική εξάγοντας πληροφορίες σχετικά με την έννοια και το συναισθηματικό περιεχόμενο κάποιας χειρονομίας αλλά και υποβοηθώντας την εξαγωγή ποιοτικών, συναισθηματικών εκφραστικών παραμέτρων.

Abstract

This thesis deals with the field of Affective Computing and Human Computer Interaction. Chapter 2 presents new aspects of affective computing, namely multimodal, dynamic emotion recognition focusing on naturalistic behavior and context adaptation, while chapter 3 deals with the computational formalization of gesture expressivity features, with validation of the extraction algorithm and with the derivation of a framework for multimodal and expressive synthesis on Embodied Conversational Agents. Chapter 4 deals with aspects of human computer interaction such as alternative means of interaction, Universal Access and Assistive Technology by proposing a novel scheme for automatic gesture and sign language recognition and a sign language synthesis platform. Furthermore, gesture recognition contributes to emotion analysis both by extracting primary information concerning the emotional content that accompanies a gesture and by supporting the qualitative process of expressivity features extraction.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Όσον αφορά στην συναισθηματική υπολογιστική τα προβλήματα που αντιμετωπίζει η διατριβή είναι η εκμετάλλευση της δυναμικής εξέλιξης των συναισθηματικών ενδείξεων, η συγχώνευση πολλαπλών ροών πληροφορίας, η εφαρμογή αλγορίθμων αναγνώρισης συναισθήματος σε φυσιοκρατικά δεδομένα συναισθηματικής συμπεριφοράς και η προσαρμογή σε δυναμικές συνθήκες αλληλεπίδρασης. Περιγράφεται μια δυναμική προσέγγιση βασισμένη σε πολλαπλές ενδείξεις σε μια προσπάθεια να ανιχνευτεί το συναίσθημα σε φυσικές ακολουθίες βίντεο, όπου η είσοδος λαμβάνεται σε φυσιοκρατικές συνθήκες, σε αντίθεση με ελεγχόμενες, εργαστηριακές συνθήκες καταγραφής οπτικοακουστικού υλικού που συνήθως υιοθετούν οι προσεγγίσεις στην βιβλιογραφία. Η αναγνώριση επιτελείται με την χρήση ενός αναδρομικού νευρωνικού δικτύου, του οποίου οι δυνατότητες κωδικοποίησης και αποθήκευσης πληροφοριών βραχυπρόθεσμης μνήμης βρίσκουν ιδανικό πεδίο εφαρμογής στην δυναμική εξέλιξη των εκφράσεων του προσώπου και της προσωδικής εκφραστικότητας. Αυτή η προσέγγιση διαφέρει επίσης από υπάρχουσες εργασίες δεδομένου ότι προτυποποιεί την εκφραστικότητα των χρηστών βασισμένη σε μια διαστατική αναπαράσταση, αντί της ανίχνευσης διακριτών καθολικών συναισθημάτων, που σπάνια εμφανίζονται στην καθημερινή αλληλεπίδραση ανθρώπου-μηχανής. Ο αλγόριθμος εφαρμόζεται σε μια οπτικοακουστική βάση δεδομένων που καταγράφηκε έχοντας ως πρότυπο την επικοινωνία ανθρώπου-ανθρώπου και, επομένως, περιέχει λιγότερο ακραία εκφραστικότητα και ελάχιστα διαφοροποιημένες παραλλαγές ενός αριθμού κατηγοριών συναισθήματος.

Πέραν της διερεύνησης της ενσωμάτωσης της δυναμικής φύσης των συναισθηματικών εκφράσεων, καταπιάνεται με την σε βάθος μελέτη της πολυτροπικής διάστασης της αυτόματης αναγνώρισης συναισθήματος. Τόσο ο τρόπος συγχώνευσης των αποφάσεων για την συγχώνευση σε επίπεδο απόφασης, όσο και η επιλογή χαρακτηριστικών γνωρισμάτων για την περίπτωση συγχώνευσης σε επίπεδο γνωρισμάτων αποτελούν προβλήματα που απασχολούν την ερευνητική κοινότητα της συναισθηματικής υπολογιστικής και η εργασία αυτή προσφέρει σημαντικά συμπεράσματα προς την λύση αυτών. Τέλος, η απαίτηση για προσαρμογή των αρχιτεκτονικών σε διαφορετικές συνθήκες και εκφραστικές ιδιαιτερότητες ατόμων είναι μια πτυχή που απασχολεί πολλούς ερευνητές και προς αυτή την κατεύθυνση πραγματοποιήθηκε ερευνητική εργασία ώστε να συνεισφέρει στην γενίκευση και την εφαρμοστικότητα προσεγγίσεων αυτόματης αναγνώρισης συναισθηματικής κατάστασης του χρήστη σε δυναμικά περιβάλλοντα επικοινωνίας ανθρώπου μηχανής. Μια αποτελεσματική προσέγγιση παρουσιάζεται εδώ, η οποία χρησιμοποιεί αρχιτεκτονικές νευρωνικών δικτύων για την ανίχνευση της ανάγκης για προσαρμογή της γνώσης που αποκτήθηκε με την αρχική εκπαίδευση και

την προσαρμογή της μέσω μιας αποδοτικής διαδικασίας.

Πληθώρα ερευνών από το επιστημονικό πεδίο της ψυχολογίας και της γνωστικής επιστήμης σχετικής με την συμπεριφορά και την μη λεκτική επικοινωνία αναδεικνύουν την σημασία των εκφραστικών ποιοτικών χαρακτηριστικών των μετακινήσεων του σώματος και των χειρονομιών κατά την αλληλεπίδραση ανθρώπων. Σχετικές μελέτες έχουν πραγματοποιηθεί στον χώρο της σύνθεσης εικονικών χαρακτήρων και πρακτόρων αλλά το επίπεδο της ερεύνας στον χώρο της ανάλυσης συναισθηματικής συμπεριφοράς υπό αυτό το πρίσμα είναι αρκετά χαμηλό και περιορίζεται στην ποιοτική μελέτη και όχι την υπολογιστική ανάλυση εκφραστικά εμπλουτισμένων χειρονομιών. Αυτό το κενό καλύπτει η παρούσα διατριβή με την τυποποίηση εκφραστικών παραμέτρων και την εισαγωγή υπολογιστικού αλγορίθμου εξαγωγής αυτών των παραμέτρων στο κεφάλαιο 3. Επιπλέον, στην ενότητα 3.2 ορίζεται το πλαίσιο πολυμεσικής εκφραστικής ανατροφοδότησης από Ενσαρκωμένο Πράκτορα Συνομιλητή (Embodied Conversational Agents - ECA). Στο πλαίσιο αυτό ενσωματώνεται η εκφραστική ανάλυση και αποτελεί την ικανότητα του εικονικού πράκτορα να αντιληφθεί και να ερμηνεύσει την συναισθηματικής κατάστασης του χρήστη ή έστω κάποιων ενδείξεων αυτής. Η δυνατότητα των αληθοφανών εικονικών πρακτόρων να παρέχουν εκφραστική ανατροφοδότηση στον χρήστη είναι μια σημαντική πτυχή ώστε να υποστηρίξουν τη φυσικότητα της αλληλεπίδρασης τους. Η πολυμεσική ανατροφοδότηση επηρεάζει την αληθοφάνεια της συμπεριφοράς ενός πράκτορα ως προς τον ανθρώπινο χρήστη και ενισχύει την επικοινωνιακή του εμπειρία. Αυτή η εργασία πραγματεύεται την εκφραστική και πολυμεσική ανάλυση και σύνθεση σε εικονικούς συνομιλητικούς πράκτορες (Embodied Conversational Agents - ECA) και επικεντρώνεται στις ενδιάμεσες διαδικασίες που απαιτούνται ώστε ένας πράκτορας να αντιληφθεί, ερμηνεύσει και να μιμηθεί μια σειρά από εκφράσεις του προσώπου και χειρονομίες όπως προκύπτουν από την ανάλυση των ενεργειών που εκτελέσθηκαν.

Επιπλέον, στην ενότητα 3.3 αντιμετωπίζεται το πρόβλημα της επικύρωσης χειρωνακτικού σχολιασμού εκφραστικότητας μέσω της αυτόματης εξαγωγής παραμέτρων. Η επισημείωση σωμάτων φυσιοκρατικής, πολυμεσικής, συναισθηματικής συμπεριφοράς αποτελεί πρόκληση, δεδομένου ότι περιλαμβάνει την υποκειμενική αντίληψη και απαιτεί μεγάλο χρονικό διάστημα για τον συναισθηματικό σχολιασμό σε πολλαπλά, παράλληλα επίπεδα. Αυτός ο χειρωνακτικός σχολιασμός μπορεί να ωφεληθεί από την αυτόματη εκτίμηση ποιοτικών παραμέτρων κίνησης του σώματος η οποία επικυρώνει τους χειρωνακτικούς σχολιασμούς. Οι στόχοι της παρούσας εργασίας είναι να διερευνηθεί η εφαρμοσιμότητα τεχνικών εξαγωγής εκφραστικών παραμέτρων σε χαμηλής ανάλυσης τηλεοπτικά βίντεο και τρόπους με τους οποίους αυτή η ανάλυση θα μπορούσε να χρησιμοποιηθεί για την επικύρωση του χειρωνακτικού σχολιασμού αυθόρμητης συναισθηματικής συμπεριφοράς.

Στο επιστημονικό πεδίο της αλληλεπίδραση ανθρώπου μηχανής ορίζονται συνεχώς νέες μορφές επικοινωνίας και διεπαφής με τις υπολογιστικές μηχανές και μια από αυτές τις μορφές είναι και οι χειρονομίες. Η βασισμένη σε χειρονομίες αλληλεπίδραση ανθρώπου-μηχανής (Gesture Based Human Computer Interaction) και η αναγνώριση νοηματικής γλώσσας προσελκύουν όλο και περισσότερο την προσοχή ερευνητών από ερευνητικές περιοχές όπως η μηχανική μάθηση, η αναγνώριση προτύπων, η όραση υπολογιστών, η αλληλεπίδραση ανθρώπου-μηχανής, η γλωσσολογία και η επεξεργασία φυσικής γλώσσας. Αυτός ο διεπιστημονικός ερευνητικός τομέας βρίσκει πεδίο εφαρμογής επίσης σε αρκετές περιοχές όπως πολυτροπική αλληλεπίδραση ανθρώπου υπολογιστή, συστήματα αυτομάτου ελέγχου, ρομποτική, συναισθηματική υπολογιστική

και συμπεριφορισμός, αναγνώριση νοηματικής γλώσσας, βοηθητικές τεχνολογίες απομακρυσμένης μάθησης και πλοήγηση σε εικονικά περιβάλλοντα. Ένας πρωτότυπος αλγόριθμος προτυποποίησης και αναγνώρισης χειρονομιών και λημμάτων νοηματικής γλώσσας παρουσιάζεται στο κεφάλαιο 4 που αντιμετωπίζει τις προκλήσεις που παρουσιάζονται στις προσεγγίσεις αυτόματης αναγνώρισης νοηματικής γλώσσας όπως ευρωστία έναντι σε αλλοίωση και θόρυβο κατά τον νοηματισμό, δυνατότητα γενίκευσης χωρίς εξαντλητική εκπαίδευση, αυθαίρετη αρχικοποίηση παραμέτρων και εξάρτηση από τον χρήστη και χαμηλό υπολογιστικό κόστος που καθιστά τα σχήματα αναγνώρισης ικανά να υλοποιηθούν σε εφαρμογές πραγματικού χρόνου. Αυτοοργανούμενοι χάρτες μοντελοποιούν την χωρική πληροφορία που εξάγεται μέσω της επεξεργασίας εικόνας ενώ την χρονική πτυχή αναλαμβάνουν να προτυποποιήσουν Μαρκοβιανά και κρυφά Μαρκοβιανά μοντέλα για την κίνηση και την χειρομορφή αντίστοιχα. Το σημείο εστίασης είναι η αντιμετώπιση της παρέκκλισης ή αλλοίωσης εκτέλεσης χειρονομίας τόσο από τον ίδιο χρήστη όσο και από διαφορετικούς χρήστες με την προσθήκη ευέλικτης και προσαρμοστικής διαδικασίας αποκωδικοποίησης επιτρέποντας στον αλγόριθμο να αναζητήσει την βέλτιστη διαδρομή ενώ το υπολογιστικό κόστος της διαδικασίας αναγνώρισης υποδεικνύει την προτεινόμενη αρχιτεκτονική ως κατάλληλη για εφαρμογή σε λεξικά μεγάλης κλίμακας σε πραγματικό χρόνο.

Τέλος, στην ενότητα 4.3 παρουσιάζεται ένα σύστημα σύνθεσης νοημάτων ενσωματωμένο σε εκπαιδευτική πλατφόρμα με στόχο την να υποστήριξη μαθητών των πρώτων τάξεων του σχολείου, είτε με την μορφή απομακρυσμένης μάθησης είτε με την μορφή ασύγχρονης διδασκαλίας και ο σχεδιασμός του είναι συμβατός με τις αρχές της προσβάσιμης εξ αποστάσεως εκπαίδευσης. Εκτός από τη διδασκαλία ΕΝΓ σαν πρωτεύουσα γλώσσα, στην παρούσα μορφή η πλατφόρμα μπορεί να χρησιμοποιηθεί για εκμάθηση γραπτών ελληνικών κειμένων μέσω ΕΝΓ, ενώ δυνητικά θα μπορούσε να βρεί εφαρμογή σε άλλους τομείς του σχολικού προγράμματος σπουδών αλλά και εκτός αυτού ως ενότητα σε οποιοδήποτε σύστημα υποστηρικτικής τεχνολογίας και καθολικής πρόσβασης.

Περισσότερα από 50 άρθρα περιέχουν αναφορές σε εργασίες της παρούσας διατριβής και στην ενότητα 6.4 απαριθμούνται κάποια από αυτά. Ιδιαίτερος σχολιασμός αρμόζει στο άρθρο επισκόπησης, οι συγγραφείς του οποίου αποτελούν αναγνωρισμένους ερευνητές του χώρου της συναισθηματικής υπολογιστικής, 'A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions' που δημοσιεύτηκε στο περιοδικό IEEE Transactions on Pattern Analysis and Machine Intelligence, 31/(1)-2009, το οποίο περιέχει αναφορές σε δύο εργασίες που παρουσιάζονται στην διδακτορική αυτή διατριβή. Είναι χαρακτηριστικό πως τα προβλήματα που αντιμετωπίζει η διατριβή και συγκεκριμένα το κεφάλαιο 2, παρουσιάζονται στην επισκόπηση αυτή ως ανοιχτά και φλέγοντα θέματα του ερευνητικού πεδίου. Αναλυτικά τα θέματα της επεξεργασίας φυσιοκρατικών δεδομένων, της εκμετάλλευσης της δυναμικής εξέλιξης των ροών πληροφορίας, της συγχώνευσης πολλαπλών και διαφορετικής φύσεως μορφών πληροφορίας και της εξάρτησης από το πλαίσιο της αλληλεπίδρασης αντιμετωπίζονται στα αντίστοιχα κεφάλαια της διατριβής 2.3, 2.4 και 2.5.

Κεφάλαιο 2

Αναγνώριση δυναμικών συναισθηματικών καταστάσεων από πολλαπλές μορφές πληροφορίας σε φυσική επικοινωνία ανθρώπου μηχανής

2.1 Ερευνητικό πλαίσιο

Ο συναισθηματικός και ανθρωποκεντρικός υπολογιστικός τομέας έχει προσελκύσει μεγάλο μέρος της προσοχής της ερευνητικής κοινότητας κατά τη διάρκεια των τελευταίων ετών, κυρίως λόγω της αφθονίας περιβαλλόντων και εφαρμογών ικανών να εκμεταλλευτούν και να προσαρμοστούν σε πολυμορφική είσοδο από τους χρήστες. Ο συνδυασμός πολλαπλών μορφών πληροφορίας όπως εκφράσεις του προσώπου, προσωδική πληροφορία φωνής, εκφραστικές παράμετροι χειρονομιών επιτρέπει στην συναισθηματική κατάσταση των χρηστών να γίνει αντιληπτή από τα υπολογιστικά συστήματα με έναν μη παρεισφρητικό (unintrusive) τρόπο, που είτε στηρίζεται στην πιο αποδοτική μορφή πληροφορίας σε περιπτώσεις όπου μια μορφή πάσχει από θόρυβο ή από μη ελεγχόμενο περιβάλλον καταγραφής είτε σε κατάλληλο συνδυασμό των μορφών πληροφορίας σε επίπεδο γνωρισμάτων ή σε επίπεδο απόφασης. Η εισαγωγή του όρου της συναισθηματικής υπολογιστικής (affective computing) από την Picard [193] συνοψίζεται στο γεγονός ότι ο υπολογιστικός τομέας θεωρείται ότι πλέον αφορά σε θέματα υψηλότερου επιπέδου από αριθμητικούς υπολογισμούς και ως εκ τούτου πρέπει να θεωρηθεί ως διεπαφή μεταξύ ανθρώπων και μηχανών και μερικές φορές ακόμα και μόνο μεταξύ ανθρώπων με την μεσολάβηση της μηχανής. Για να επιτευχθεί αυτό, η αρχιτεκτονική των υπολογιστικών εφαρμογών πρέπει να μπορεί να λάβει υπόψη της τη δυνατότητα των ανθρώπων να παρέχουν είσοδο με πολλαπλή μορφή στους υπολογιστές, ξεπερνώντας την απαρχαιωμένη διεπαφή τύπου παράθυρο-ποντίκι-δείκτης και χρησιμοποιώντας πιο διαισθητικά μέσα, πλησιέστερα στις ανθρώπινες συνήθειες ([120], [188]). Ένα μεγάλο μέρος της φυσιοκρατικής αλληλεπίδρασης είναι η έννοια της εκφραστικότητας [194], τόσο από την άποψη της ερμηνείας της αντίδρασης του χρήστη σε ένα συγκεκριμένο ερέθισμα όσο και της προσαρμογής της συμπεριφοράς του λαμβάνοντας υπόψη την συναισθηματική κατάσταση του χρήστη, δεδομένου ότι εξομαλύνει την καμπύλη εκμάθησης για συμβατικές διεπαφές και κάνουν τους, λιγότερο εξοικειωμένους με την τεχνολογία, χρήστες να αισθανθούν πιο άνετα. Στο πλαίσιο αυτό, η ομιλία, οι εκφράσεις του προσώπου και η εκφραστικότητα του σώ-

ματος είναι ξεχωριστής σημασίας, δεδομένου ότι συνήθως παρέχουν μια κατανοητή πτυχή της συμπεριφοράς των χρηστών. Ο Cohen , στο [39] και [40], σχολίασε την παρουσία και την σημασία των πολλαπλών μορφών πληροφοριών ενώ η Oviatt στο [172] απέδειξε ότι μια αρχιτεκτονική αλληλεπίδρασης που περιορίζεται μόνο στο ξεπερασμένο πλαίσιο παροχής εντολών με συμβατικά μέσα αποτελεί ένα πολύ μικρό μέρος των αυθόρμητων πολυμορφικών εκφράσεων στο πεδίο του καθημερινού HCI. Στο ίδιο πλαίσιο, το [122] προδιαγράφει το πολυμορφικό σύστημα ως κάποιο που 'αποκρίνεται σε είσοδο με περισσότερα από ένα κανάλια επικοινωνίας', ενώ ο Mehrabian στο [160] υπονοεί ότι οι εκφράσεις του προσώπου και ο δυναμικός τονισμός της φωνής είναι τα κύρια μέσα για τον υπολογισμό της συναισθηματικής κατάστασης ενός ατόμου [264], με το κανάλι του προσώπου να κρίνεται ακριβέστερο, ή να συσχετίζεται καλύτερα με τις βασισμένες στην πλήρη οπτικοακουστική είσοδο αποφάσεις απ'ό,τι το κανάλι της φωνής ([122], [180]). Το γεγονός αυτό ενέπνευσε διάφορες προσεγγίσεις που χρησιμοποιούν βίντεο και ήχο για να αντιμετωπίσουν την αναγνώριση συναισθήματος κατά τρόπο πολυμορφικό ([119], [57], [35], [36], [57], [257], [91]), ενώ πρόσφατα το οπτικό κανάλι επεκτάθηκε ώστε να περιλαμβάνει κινήσεις του κεφαλιού, σώματος και χεριών ([101], [30]).

Πρόσθετοι παράγοντες που ενισχύουν την πολυπλοκότητα του υπολογισμού της εκφραστικότητας σε καθημερινό HCI είναι ο συνδυασμός πληροφοριών που εξάγονται από διάφορα κανάλια [172], η ερμηνεία των δεδομένων σύμφωνα με την χρονική εξέλιξη τους καθώς και η απομάκρυνση θορύβου και αβεβαιότητας από τις φυσικές ρυθμίσεις της διαδικασίας καταγραφής ([237], [175]). Στην περίπτωση του συνδυασμού πληροφοριών πολλαπλών μορφών [252], τα συστήματα μπορούν να ενσωματώσουν τα σήματα εισόδου σε επίπεδο χαρακτηριστικών γνωρισμάτων ([221]) ή, αφού καταλήξουν σε μια απόφαση κατηγορίας για κάθε κανάλι, με τη συγχώνευση των αποφάσεων σε σημασιολογικό επίπεδο ([221] και [198]), λαμβάνοντας υπόψη οποιοδήποτε μέτρο εμπιστοσύνης και βεβαιότητας παρέχεται ή μπορεί να εξαχθεί από κάθε μορφή σε μια ρύθμιση πολύ κοντά σε μηχανή επιτροπείας (committee machine) [106]. όσον αφορά στη δυναμική φύση της εκφραστικότητας, η Littlewort [152] καταλήγει ότι ενώ οι τεχνικές βασισμένες σε μυς μπορούν να περιγράψουν επιτυχώς τη μορφολογία μιας έκφρασης του προσώπου, είναι αρκετά δύσκολο να ερμηνεύσουν κατά τρόπο μετρήσιμο (και επομένως ανιχνεύσιμο) την δυναμική πτυχή των εκφράσεων, δηλ. την χρονική εξέλιξη της ενεργοποίησης μυών και της παρατηρούμενης μετακίνησης ή παραμόρφωσης των χαρακτηριστικών γνωρισμάτων. Παρουσιάζει επίσης επιχειρήματα ως προς την άποψη πως η φυσική εκφραστικότητα είναι εγγενώς διαφορετική όσον αφορά στην χρονική εξέλιξη από την επιτηδευμένη εκφραστικότητα, βασισμένη σε ψυχολογικές μελέτες ([71] και [92]), αποδεικνύοντας την ανομοιότητα αυτή, αντιμετωπίζοντας την υλοποίηση εκφραστικότητας χρησιμοποιώντας μηχανισμούς που εκμεταλλεύονται δυναμικές ιδιότητες των εκφραστικών παραμέτρων. Ως γενικός κανόνας, τα φυσιοκρατικά δεδομένα, που επιλέγονται ως είσοδο στην εργασία αυτή, προσεγγίζουν καλύτερα την πραγματική ανθρώπινη συμπεριφορά δεδομένου ότι ο διάλογος και γενικά η αλληλεπίδραση δεν είναι υποδυόμενη και η εκφραστικότητα δεν καθοδηγείται από οδηγίες ή σενάρια. Αυτό ενισχύει τη δυσκολία στη διάκριση των εκφράσεων του προσώπου και των ακουστικών προτύπων. Εντούτοις, παρέχει ιδανικό πειραματικό πεδίο στην περιοχή της επεξεργασίας της ακολουθίας των υπό εξέταση οπτικοακουστικών δεδομένων και της συγχώνευσης συμπερασμάτων που προέρχονται από κάθε μορφή στην μονάδα του χρόνου.

Πρόσφατα, η έρευνα στην περιοχή της αλληλεπίδρασης ανθρώπου-υπολογιστή εξε-

τάζει όλο και περισσότερο την πτυχή της επικοινωνίας σχετική με το υπονοούμενο κανάλι επικοινωνίας, που είναι το κανάλι μέσω του οποίου η συναισθηματική αλληλεπιδρά με τη λεκτική πτυχή της επικοινωνίας [49]. Μια από τις προκλήσεις είναι να εμπλουτισθεί μια μηχανή με συναισθηματική νοημοσύνη. Τα συναισθηματικά ευφυή συστήματα πρέπει να είναι σε θέση να δημιουργήσουν μια συναισθηματική αλληλεπίδραση με τους χρήστες, να εμπλουτισθούν με τη δυνατότητα αντίληψης, ερμηνείας και έκφρασης συναισθημάτων [193]. Η αναγνώριση της συναισθηματικής κατάστασης των χρηστών είναι μια από τις κυριότερες απαιτήσεις για την επιτυχή αλληλεπίδραση των υπολογιστών με τους ανθρώπους. Οι περισσότερες από τις εργασίες στον τομέα της συναισθηματικής υπολογιστικής δεν συνδυάζουν πολλαπλές διαφορετικές μορφές πληροφορίας σε ένα ενιαίο σύστημα για την ανάλυση της ανθρώπινης συναισθηματικής συμπεριφοράς. Διαφορετικές ροές πληροφοριών (κυρίως εκφράσεις του προσώπου και ομιλία) εξετάζονται ανεξάρτητα. Επιπλέον, υπάρχουν λιγοστές προσπάθειες να ενσωματωθούν πληροφορίες από τη κίνηση του σώματος και των χειρονομιών. Παρ'όλα αυτά, οι Sebe κ.α. στο [212] και οι Pantic κ.α. στο [181] τονίζουν ότι ένα ολοκληρωμένο σύστημα αυτόματης ανάλυσης και αναγνώρισης ανθρώπινων συναισθηματικών συμπεριφορών πρέπει λειτουργεί πολύμορφα (λαμβάνοντας ως είσοδο πολλαπλές μορφές πληροφορίας), όπως ακριβώς συμβαίνει και με το ανθρώπινο αισθητήριο σύστημα. Επιπλέον, μελέτες από την ερευνητική περιοχή της ψυχολογίας τονίζουν την ανάγκη ενσωμάτωσης διαφορετικών, μη λεκτικών, μορφών συμπεριφοράς αντίστοιχων με αυτές που συναντώνται στην επικοινωνία μεταξύ ανθρώπων [208].

Η αναγνώριση των συναισθηματικών καταστάσεων χρηστών στην επικοινωνία ανθρώπου και μηχανής έχει αποδειχτεί ότι εξαρτάται ιδιαίτερα από μεμονωμένα ανθρώπινα χαρακτηριστικά και τρόπους συμπεριφοράς. Η είσοδος με πολλαπλές μορφές είναι βασικό ζήτημα στην επίτευξη ακριβέστερων αποτελεσμάτων. Εντούτοις, ο συνδυασμός διαφορετικών μορφών πληροφορίας είναι ένα πολύπλοκο και πολυδιάστατο θέμα στην ανάλυση συναισθήματος. Τα συστήματα αναγνώρισης συναισθήματος είναι γενικά είτε βασισμένα σε κανόνες ή εκτενώς εκπαιδευμένα μέσω συναισθηματικά εμπλουτισμένων συνόλων στοιχείων HCI. Και στις δύο περιπτώσεις, τέτοια συστήματα πρέπει να λάβουν υπόψη και να προσαρμόζουν την απόφαση τους σε συγκεκριμένους χρήστες ή στο ευρύτερο πλαίσιο αλληλεπίδρασης. Η ευρύτερη περιοχή των νευρωνικών δικτύων ταιριάζει απόλυτα στην απαίτηση προσαρμογής, με τη συλλογή και ανάλυση στοιχείων από συγκεκριμένα περιβάλλοντα εφαρμογής.

2.2 Πτυχές της αυτόματης αναγνώρισης συναισθήματος

Στην βιβλιογραφία συναντάται πληθώρα άρθρων επισκόπησης σχετικά με την ανάλυση συναισθηματικών καταστάσεων με την βοήθεια τεχνικών μηχανικής μάθησης, τεχνητής νοημοσύνης, κ.τ.λ. Μερικά από αυτά τα άρθρα επισκόπησης είναι: [203] [49] [86] [171] [181] [225] [212] και [176], δημοσιευμένα από το 1992 ως το 2007, αντίστοιχα. Ακόμα στην βιβλιογραφία συναντώνται επισκοπήσεις πρώιμης ερευνητικής εργασίας σχετικής με την ανάλυση των εκφράσεων του προσώπου: [203] [178] και [86] ενώ έρευνες σχετικές με τεχνικές αυτόματης ανάλυσης μυών του προσώπου, αναγνώρισης μονάδων δράσης (ActionUnits) και εκφράσεων του προσώπου: [225] και [176]. Τέλος, επισκοπήσεις σχετικά με τις μεθόδων αναγνώρισης συναισθήματος από πολλαπλές ροές πληροφορίας: [49] [180] [181] [212] [121] και [263].

2.2.1 Δεδομένα από φυσική αλληλεπίδραση

Ενώ η αυτόματη ανίχνευση έξι βασικών συναισθημάτων σε επιτηδευμένες και ελεγχόμενες ακουστικές και οπτικές συναισθηματικές εκφάνσεις επιτυγχάνεται με αρκετά υψηλή ακρίβεια, η ανίχνευση αυτών των εκφράσεων ή οποιασδήποτε ενδιάμεσης έκφρασης της ανθρώπινης συναισθηματικής συμπεριφοράς σε λιγότερο περιορισμένες διατάξεις παραμένει μια πρόκληση εξαιτίας του γεγονότος ότι η σκόπιμη συμπεριφορά διαφέρει από την αυθόρμητα εμφανιζόμενη συμπεριφορά ως προς την οπτική εμφάνιση, το ακουστικό σήμα και τον συγχρονισμό. Κριτικές που ασκήθηκαν από ερευνητές της γνωστικής λειτουργίας και της ψυχολογίας οδήγησε την έρευνα στον τομέα της συναισθηματικής υπολογιστικής να μετατοπιστεί στην αυτόματη ανάλυση αυθόρμητα επιδειχθείσας συναισθηματικής συμπεριφοράς και να εστιάσει σε φυσική ή φυσιοκρατική αλληλεπίδραση ανθρώπου μηχανής. Αρκετές μελέτες έχουν προκύψει πρόσφατα στην μηχανική ανάλυση αυθόρμητων εκφράσεων του προσώπου ([11] [42] [229] και [6]) αλλά και φωνητικών εκφράσεων (π.χ., [12] και [142]). Βέβαια, ο συνδυασμός της εφαρμογής αλγορίθμων αναγνώρισης συναισθήματος σε φυσική αλληλεπίδραση συνυπολογίζοντας πολλαπλές μορφές πληροφορίας καθιστά το πρόβλημα ακόμα πιο ενδιαφέρον και η παρούσα διατριβή το αντιμετωπίζει σφαιρικά και πολύπλευρα στα κεφάλαια 2.4 και 2.3.

Οι περισσότερες από τις υπάρχουσες μεθόδους για οπτικοακουστική ανάλυση συναισθήματος βασίζονται σε επιτηδευμένες και ελεγχόμενες επιδείξεις συναισθήματος (π.χ. [98] [109] [211] [215] [244] [261] [265] και [266]). Πρόσφατα, αρκετές εξαιρετικές μελέτες έχουν στραφεί προς την πολυτροπική ανάλυση συναισθήματος εφαρμοσμένη σε αυθόρμητες και φυσικές επιδείξεις συναισθήματος (π.χ. [91] [129] [174] [262] και [191]). Οι Zeng κ.α. [262] χρησιμοποίησαν δεδομένα που συλλέχθηκαν κατά την διάρκεια ψυχολογικών ερευνητικών συνεντεύξεων, οι Pal κ.α. χρησιμοποίησαν καταγραφές νηπίων [174] και οι Petridis και Pantic [191] χρησιμοποίησαν καταγραφές επαγγελματικών συναντήσεων. Αφ' ετέρου, οι Fragopanagos και Taylor [91] χρησιμοποίησαν δεδομένα που συλλέχθηκαν με την μέθοδο Wizard of Oz. Δεδομένου ότι τα διαθέσιμα δεδομένα ήταν συνήθως ανεπαρκή να εκπαιδεύσουν διεξοδικά ένα εύρωστο σύστημα μηχανικής μάθησης για την αναγνώριση συναισθηματικών καταστάσεων σε μεγάλη λεπτομέρεια (π.χ., βασικά συναισθήματα), επιχειρήθηκε η αναγνώριση συναισθηματικών καταστάσεων με μεγαλύτερο εύρος σε σχέση με τις προαναφερθείσες μελέτες. Στο [262] οι ερευνητές εστιάζουν στον διαχωρισμό συναισθημάτων σε θετικό και αρνητικό, ενώ άλλες εργασίες αναφέρουν αποτελέσματα σχετικά με την κατηγοριοποίηση οπτικοακουστικών δεδομένων σε τεταρτημόρια του συναισθηματικού χώρου αξιολόγησης-ενεργοποίησης [91]. Εντούτοις, αξίζει να σημειωθεί ότι η έρευνα που παρουσιάζεται στο [91] αναφέρει ιδιαίτερη απόκλιση στην επισημείωση των τεσσάρων επισημειωτών λόγω της υποκειμενικής κρίσης των οπτικοακουστικών συναισθηματικών ενδείξεων. Πιο συγκεκριμένα, ένας από τους επισημειωτές στηρίχθηκε κυρίως στις ακουστικές πληροφορίες, ενώ ένας άλλος στις οπτικές πληροφορίες για την επισημείωση του παρατηρούμενου συναισθήματος. Αυτό το πείραμα είναι επίσης χαρακτηριστικό της έλλειψης συγχρονισμού ανάμεσα στην ακουστική και την οπτική έκφραση της συναισθηματικής κατάστασης του χρήστη. Προκειμένου να μειωθεί αυτή η απόκλιση στην ανθρώπινη επισημείωση, στο [262] έγινε η υπόθεση ότι η έκφραση του προσώπου και η φωνητική έκφραση βρίσκονται στην ίδια προσεγγιστικά κατάσταση και σίγουρα στο ίδιο ημιεπίπεδο του συναισθηματικού χώρου και χρησιμοποίησαν την επισημείωση που αφορούσε την έκφραση του προσώπου ως χαρακτηρισμό και για το ακουστικό κανάλι πληροφορίας.

2.2.2 Αναπαράσταση συναισθήματος

Η έννοια του συναισθήματος έχει τυποποιηθεί πρωταρχικά με τρεις τρόπους στην ψυχολογία. Η έρευνα σχετικά με την βασική δομή και περιγραφή του συναισθήματος είναι σημαντική, δεδομένου ότι τα αποτελέσματα τέτοιων ερευνών παρέχουν πληροφορίες για τις εκφάνσεις της συναισθηματικής κατάστασης την οποία τα αυτόματα συστήματα αναγνώρισης συναισθήματος επιχειρούν να ανιχνεύσουν. Ο συνηθέστερος ίσως τρόπος περιγραφής του συναισθήματος από τους ψυχολόγους είναι αυτός των διακριτών κατηγοριών, μια προσέγγιση που προέρχεται από την ανθρώπινη καθημερινότητα [72] [74] [76] [227]. Το δημοφιλέστερο παράδειγμα αυτής της περιγραφής είναι οι βασικές (οικουμενικές) κατηγορίες συναισθήματος, οι οποίες περιλαμβάνουν την ευτυχία, τη θλίψη, το φόβο, το θυμό, την αποστροφή και την έκπληξη. Η περιγραφική αυτή αντιμετώπιση των βασικών συναισθημάτων στηρίχθηκε ιδιαίτερα σε διαπολιτισμικές μελέτες που πραγματοποιήθηκαν από έναν από τους πρωτοπόρους ερευνητές στην περιοχή της συναισθηματικής υπολογιστικής, τον Paul Ekman [72], αποδεικνύοντας ότι οι άνθρωποι αντιλαμβάνονται ορισμένα βασικά συναισθήματα όσον αφορά στις εκφράσεις του προσώπου με τον ίδιο τρόπο, ανεξάρτητα από τις πολιτισμικές τους επιρροές. Η επιρροή αυτής της βασικής θεωρίας αναπαράστασης συναισθημάτων έχει οδηγήσει τις περισσότερες από τις υπάρχουσες εργασίες της αυτόματης αναγνώρισης συναισθήματος να εστιάσουν στην αναγνώριση αυτών ακριβώς των βασικών συναοσθημάτων. Το κυριότερο πλεονέκτημα μιας κατηγορικής αναπαράστασης είναι ότι και οι άνθρωποι χρησιμοποιούν αυτό το κατηγορικό σχήμα για να περιγράψουν τις παρατηρηθείσες συναισθηματικές εκφάνσεις στην καθημερινή τους ζωή και ως εκ τούτου το αντίστοιχο πρωτόκολλο επισημείωσης είναι αρκετά πιο διαισθητικό και δείχνει να ταιριάζει περισσότερο με την καθημερινή διαπροσωπική επαφή των ανθρώπων. Εντούτοις, η διακριτή συναισθηματική κατηγοριοποίηση αποτυγχάνει να περιγράψει την συναισθηματική ακολουθία που συνήθως εμφανίζεται σε φυσικές ή φυσιοκρατικές συνθήκες αλληλεπίδρασης όπου τα ακραία, ευδιάκριτα και ομοιογενή συναισθήματα σπανίζουν. Παραδείγματος χάριν, αν και τα βασικά συναισθήματα αποτελούν σημεία αναφοράς ως προς την συναισθηματική κατάσταση του χρήστη, καλύπτουν ένα μάλλον μικρό εύρος της καθημερινής συναισθηματικής αλληλεπίδρασης. Η επιλογή των συναισθηματικών κατηγοριών ικανών να περιγράψουν το ευρύ φάσμα των συναισθηματικών επιδείξεων ανθρώπινης συμπεριφοράς σε συνθήκες καθημερινής διαπροσωπικής αλληλεπίδρασης πρέπει να γίνει κατά τρόπο ρεαλιστικό και σχετιζόμενο με το ευρύτερο πλαίσιο αλληλεπίδρασης [180] [181]. Εναλλακτική της κατηγορικής περιγραφής του ανθρώπινου συναισθήματος είναι η διαστατική περιγραφή [99] [200] [245], όπου μια συναισθηματική κατάσταση χαρακτηρίζεται από έναν μικρό, συνήθως, αριθμό διαστάσεων παρά από συγκεκριμένες διακριτές συναισθηματικές κατηγορίες. Αυτές οι διαστάσεις περιλαμβάνουν την αξιολόγηση, την ενεργοποίηση, τον έλεγχο, τη δύναμη, κ.λπ. Ειδικότερα, οι διαστάσεις της αξιολόγησης και της ενεργοποίησης επιχειρούν να απεικονίσουν τους κυρίως άξονες του συναισθήματος. Η διάσταση της αξιολόγησης αποτιμά πως αισθάνεται κάποιος, λαμβάνοντας τιμές θετικές και αρνητικές ενώ η διάσταση της ενεργοποίησης εάν είναι περισσότερο ή λιγότερο πιθανό προβούν σε κάποια ενέργεια ή όχι και λαμβάνει τιμές από παθητικό μέχρι ενεργητικό. Σε αντίθεση με την κατηγορική αναπαράσταση, η διαστατική επιτρέπει στους επισημειωτές να χαρακτηρίσουν ένα ευρύτερο πεδίο συναισθημάτων, αλλά η προβολή των πολυδιάστατων συναισθηματικών καταστάσεων σε διδιάστατο χώρο οδηγεί, μέχρι ενός ορισμένου βαθμού, σε απώλεια πληροφορίας. Αυτή η αναπαράσταση δεν είναι αρκετά διαισθητική και απαιτείται κάποια εξοικείωση από μέρους των επισημειωτών και ίσως και

κάποια εκπαίδευση στο λογισμικό επισημείωσης που χρησιμοποιεί την συγκεκριμένη διαστατική αναπαράσταση (π.χ., το σύστημα Feeltrace).

Στα αυτόματα συστήματα αναγνώρισης συναισθήματος που είναι βασισμένα στη διδιάστατη διαστατική αναπαράστση συναισθήματος, το πρόβλημα συχνά απλοποιείται περαιτέρω και υποβιβάζεται σε ταξινόμηση σε ημιεπίπεδα ή τεταρτημόρια του συναισθηματικού χώρου. Μια από τις επικρατέστερες σύγχρονες θεωρίες της ψυχολογίας είναι η προσέγγιση της αποτίμησης (appraisal) [209], η οποία μπορεί να θεωρηθεί ως επέκταση της διαστατικής προσέγγισης που περιγράφεται ανωτέρω. Σε αυτήν την αναπαράσταση, μια συναισθηματική κατάσταση περιγράφεται μέσω ενός συνόλου ελέγχων αποτίμησης ερεθισμάτων, συμπεριλαμβανομένης της καινοτομίας, εγγενής ευχαρίστησης, σημασίας ως προς την επίτευξη στόχων, δυνατότητας αντιμετώπισης και συμβατότητας αντίστασης. Εντούτοις, η υπολογιστική προτυποποίηση της αναπαράστασης αυτής σε ένα πλαίσιο εφαρμοσμένης μηχανικής για λόγους αυτόματης αναγνώρισης συγκίνησης παραμένει ανοιχτό πρόβλημα [204] αν και κάποιες μάλλον πρώιμες εργασίες έχουν δημοσιευτεί, κυρίως από την πλευρά της σύνθεσης [155].

2.2.3 Επισημείωση

Στην ευρύτερη διαδικασία αναγνώρισης συναισθήματος περιλαμβάνεται και η επισημείωση του οπτικοακουστικού υλικού από ειδικούς και την μετέπειτα χρήση της επισημείωσης αυτής ως σύνολο δεδομένης αλήθειας. Αρχικά, δεν είναι σαφώς καθορισμένο και ομοιόμορφο το είδος και η ποιότητα των μεταδεδομένων που πρέπει να καταγραφούν, ενώ και η χρονική παράμετρος της επισημείωσης δεν είναι τετριμμένη απόφαση. Ενώ η επισημείωση οπτικοακουστικού υλικού απλοποιείται για την περίπτωση των βασικών συναισθηματικών εκφράσεων, γίνεται αρκετά πιο πολύπλοκη διαδικασία σε περιπτώσεις διαφορετικής προσέγγισης αναπαράστασης συναισθήματος. Σε μια προσπάθεια να μειωθεί η υποκειμενικότητα των δεδομένων που καταγράφονται, γενικά ζητείται από περισσότερους του ενός ή και δύο, για την αποφυγή ισοβαθμιών, κριτών να πραγματοποιήσουν την επισημείωση ενώ δεν είναι σπάνιες οι περιπτώσεις όπου η διαφωνία μεταξύ των κριτών είναι αρκετά μεγάλη, γεγονός που καταδεικνύει και την δυσκολία σχεδιασμού και υλοποίησης ενός αυτόματου συστήματος αναγνώρισης συναισθήματος. Η κωδικοποίηση φωνητικής συμπεριφοράς παραμένει ένα ανοικτό ζήτημα καθώς οι μη γλωσσολογικές φωνήσεις όπως το γέλιο, ο βήχας, το κλάμα, κ.λπ., μπορούν εύκολα να κωδικοποιηθούν υπό αυτή τη μορφή, αλλά δεν υπάρχει κάποιο σύνολο ανεξάρτητο από την ερμηνεία για την επισημείωση συναισθηματικής ομιλίας. Μια άλλη σχετιζόμενη πτυχή είναι αυτή της πολιτισμικής εξάρτησης αλλά και του ευρύτερου πλαισίου αλληλεπίδρασης κατά την οποία λαμβάνει χώρα η συναισθηματικά εμπλουτισμένη ομιλία. Τα μεταδεδομένα για το πλαίσιο στο οποίο έγιναν οι καταγραφές, όπως τα ερεθίσματα, το περιβάλλον και η παρουσία άλλων ανθρώπων κρίνονται απαραίτητα δεδομένου ότι αυτές οι εξαρτημένες από το πλαίσιο παράμετροι μπορούν να επηρεάσουν σημαντικά τις συναισθηματικές συμπεριφορές. Επιπρόσθετα, ακόμα κι αν δεδομένα επισημείωσης είναι διαθέσιμα, ο σχεδιασμός ενός αυτοματοποιημένου συστήματος ανάλυσης συναισθηματικής συμπεριφοράς υποθέτει ότι τα στοιχεία είναι ακριβή μια υπόθεση που, όπως αναφέρθηκε νωρίτερα αλλά αναφέρεται και στην βιβλιογραφία [41] [217], δεν είναι πάντα έγκυρη. Η αξιοπιστία της κωδικοποίησης μπορεί να εξασφαλιστεί αν ζητηθεί από αρκετούς ανεξάρτητους ανθρώπινους παρατηρητές να πραγματοποιήσουν την επισημείωση και αν η συμφωνία και η συσχέτιση μεταξύ των δεδομένων επισημείωσης είναι υψηλή, η αξιοπιστία της κωδικοποίησης επιβεβαιώνε-

ται. Η αξιοπιστία μεταξύ των επισημειωτών μπορεί επίσης να βελτιωθεί με την επαρκή εκπαίδευση των επισημειωτών στο χρησιμοποιούμενο σύστημα κωδικοποίησης. Μια πιθανή μέθοδος επισημείωσης, αναφορικά με συναισθηματικές κατηγορίες, είναι η χρήση ένος πολυεπίπεδου συστήματος με μεταβλητές χρονικές κλίμακες προκειμένου να μειωθεί η υποκειμενικότητα της ανθρώπινης κρίσης και να αναπαρασταθούν καλύτερα οι ιδιότητες των συναισθηματικών συμπεριφορών [61] [141].

Επίσης, η επισημείωση ανθρώπινης συμπεριφοράς από ειδικούς είναι μια διαδικασία αρκετά χρονοβόρα. Για παράδειγμα στην περίπτωση των δεδομένων εκφράσεων του προσώπου, διαρκεί περισσότερο από 1 ώρα για να σημειωθούν χειρωνακτικά 100 εικόνες ή 1 λεπτό ακολουθίας βίντεο με την χρήση AUs [75]. Μια προσπάθεια αντιμετώπισης του φαινομένου αυτού θα μπορούσε να είναι η ημιεπιβλεπόμενη μέθοδος ενεργής εκμάθησης, που συνδυάζει την ημιεπιβλεπόμενη μάθηση [37] και την ενεργό μάθηση [88]. Ο μηχανισμός ημιεπιβλεπόμενης εκμάθησης στοχεύει στη χρησιμοποίηση μη επισημειωμένων στοιχείων, με τον μηχανισμό ενεργής εκμάθησης να στοχεύει στη εκμετάλλευση των χρήσιμων πληροφοριών που παρέχονται από την ανθρώπινη ανατροφοδότηση (επισημείωση σε αυτήν την εφαρμογή) και στην παρουσίαση στους σχολιαστές των πιό αμφιλεγόμενων δείγματων σύμφωνα με τον επιλεγμένο ταξινομητή συναισθήματος και μια μετρική εμπιστοσύνης που συνοδεύει το αποτέλεσμα ταξινόμησης. Πιό συγκεκριμένα, διάφορα πολλά υποσχόμενα πρωτότυπα συστήματα παρουσιάστηκαν τα τελευταία χρόνια [176] και [225]) που αν και δεν είναι πάντα ικανά να γενικεύσουν εύκολα σε περιπτώσεις ανεπαίσθητων εκφράσεων και πολύπλοκων συναισθηματικών συμπεριφορών που παρουσιάζονται σε πραγματικές, φυσιοκρατικές καταστάσεις, μπορούν να χρησιμοποιηθούν ως μια αρχική επεξεργασία που προηγείται της επικύρωσης και τελικής επισημείωσης, μέσω των διαδικασιών ελέγχου και διόρθωσης από ειδικούς. Ενώ μια τέτοια προσέγγιση δείχνει να μειώνει σημαντικά τον απαιτούμενο χρόνο επισημείωσης δεν υπάρχει κάποια εργασία που να επιβεβαιώνει την υπόθεση αυτή και σίγουρα απαιτείται μελλοντική έρευνα για να επιβεβαιώσει την ορθότητα και την εφαρμοσιμότητα της προσέγγισης.

2.2.4 Πολλαπλές μορφές πληροφορίας

Έχει αποδειχθεί από αρκετές πειραματικές μελέτες ότι η ενσωμάτωση πληροφοριών από πολλαπλές μορφές (οπτική, ακουστική, κ.τ.λ.) οδηγεί σε βελτιωμένη απόδοση και πιο εύρωστη λειτουργία των αρχιτεκτονικών αναγνώρισης συναισθηματικής συμπεριφοράς. Η βελτιωμένη αξιοπιστία των οπτικοακουστικών πολυτροπικών (multimodal) προσεγγίσεων σε σύγκριση με τις μονοτροπικές (unimodal) προσεγγίσεις μπορεί να ερμηνευτεί ως εξής: Οι υπάρχουσες τεχνικές ανίχνευσης και παρακολούθησης χαρακτηριστικών σημείων του προσώπου είναι ευαίσθητες στην στάση του κεφαλιού, τον θόρυβο και τις μεταβολές στις συνθήκες φωτισμού, ενώ οι υπάρχουσες τεχνικές επεξεργασίας λόγου είναι ευαίσθητες στον ακουστικό θόρυβο. Η οπτικοακουστική συγχώνευση, είτε αυτή επιτυγχάνεται σε επίπεδο χαρακτηριστικών είτε σε επίπεδο απόφασης, μπορεί να εκμεταλλευτεί συμπληρωματικά τις πληροφορίες από τα πολλαπλά κανάλια. Επιπλέον, πολλές ψυχολογικές μελέτες έχουν καταδείξει θεωρητικά και εμπειρικά τη σημασία της ολοκλήρωσης των πληροφοριών από πολλαπλές μορφές ώστε να παραχθεί μια συνεπής αναπαράσταση και να εξαχθούν αξιόπιστα συμπεράσματα σχετικά με την συναισθηματική κατάσταση του χρήστη [4] [199] [209]. Ως συνέπεια αυτών των ενδείξεων από την πλευρά των ψυχολόγων και των επιστημόνων της γνωστικής, ένας συνεχώς αυξανόμενος αριθμός ερευνών έχει προκύψει τα τελευ-

ταία χρόνια σχετικά με τον αντίκτυπο της πολυτροπικότητας (multimodality) στην αναγνώριση ανθρώπινου συναισθήματος (π.χ. [91] και [262]).

Η έρευνα για την πολυτροπική συναισθηματική ανάλυση σε φυσιοκρατικά δεδομένα είναι ακόμα σε πρώιμο στάδιο και ενώ υπάρχει γενική συμφωνία ότι η συγχώνευση πολλαπλών μορφών πληροφορίας, συμπεριλαμβανομένων οπτικοακουστικών δεδομένων, γλωσσολογικών και παραγλωσσολογικών ενδείξεων, είναι εξαιρετικά ευεργετική για την μηχανική ανάλυση ανθρώπινου συναισθήματος, κάποιες πτυχές της διαδικασίας αυτής παραμένουν ασαφείς. Μελέτες στη νευρολογία σχετικές με τον συνδυασμό αισθητήριων νευρώνων ευνοούν την πρόωρη συγχώνευση δεδομένων (συγχώνευση σε επίπεδο γνωρισμάτων) παρά την μετέπειτα συγχώνευση δεδομένων (συγχώνευση σε επίπεδο απόφασης). Ακόμα και έτσι όμως ο τρόπος δημιουργίας των διανυσμάτων χαρακτηριστικών γνωρισμάτων παραμένει ανοικτό ερευνητικό θέμα καθώς αυτά θα αποτελούνται από διαφορετικές μορφές πληροφορίας, με πιθανόν διαφορετική χρονική κλίμακα, διαφορετικά μετρικά επίπεδα και διαφορετικές χρονικές δομές. Η απλοική αντιμετώπιση της συνένωσης των μεμονωμένων διανυσμάτων χαρακτηριστικών γνωρισμάτων από κάθε μορφή πληροφορίας σε ένα ενιαίο διάνυσμα προφανώς δεν επαρκεί. Αυτό οδήγησε αρκετούς ερευνητές να ακολουθήσουν την συγχώνευση σε επίπεδο απόφασης κατά την οποία η είσοδος κάθε μορφής πληροφορίας μοντελοποιείται ανεξάρτητα και τα αποτελέσματα της μονότροπης μοντελοποίησης συνδυάζονται σε προχωρημένο στάδιο στην αλυσίδα αναγνώρισης. Η συγχώνευση σε επίπεδο απόφασης αποτελεί ανοιχτό πρόβλημα στις περιοχές της μηχανικής μάθησης και αναγνώρισης προτύπων και συνήθως αντιμετωπίζεται κατά περίσταση και πάντα σε συνάρτηση με την φύση του προβλήματος που επιχειρούν να επιλύσουν. Αρκετές μελέτες έχουν καταδείξει τα πλεονέκτηματα της συγνώνευσης αποτελεσμάτων ανεξάρτητων κατηγοριοποιητών λόγω των ασυσχέτιστων σφαλμάτων από τους διαφορετικούς ταξινομητές. Διάφορες μέθοδοι συγνώνευσης κατηγοριοποιητών (σταθεροί κανόνες και μηχανές επιτροπείας) έχουν προταθεί στη βιβλιογραφία, αλλά δεν υπάρχουν ακόμα βέλτιστες και καθολικά εφαρμόσιμες μέθοδοι. Επιπλέον, δεδομένου ότι οι άνθρωποι δέχονται ταυτόχρονα τις στενά συνδεμένες ακουστικές και οπτικές μορφές πληροφορίας, τα σήματα από πολλαπλές μορφές πληροφορίας δεν μπορούν να θεωρηθούν αμοιβαία ανεξάρτητα και ίσως αυτή η μέθοδος της μετέπειτα συγχώνευσης δεν είναι η καταλληλότερη.

Οι τεχνικές συγχώνευσης δεδομένων που χρησιμοποιούνται στις μελέτες οπτικοακουστικής αναγνώρισης συναισθήματος είναι αυτές που γίνονται σε επίπεδο γνωρισμάτων, σε επίπεδο απόφασης και συγχώνευση βασισμένη σε κάποιο μοντέλο. Χαρακτηριστικά παραδείγματα της συγχώνευσης σε επίπεδο γνωρισμάτων είναι τα [210] και [266], τα οποία συνδύασαν τα προσωδιακά χαρακτηριστικά γνωρίσματα και τα χαρακτηριστικά γνωρίσματα του προσώπου για να δημιουργήσουν κοινά διανύσματα χαρακτηριστικών γνωρισμάτων, τα οποία χρησιμοποιούνται έπειτα για να εκπαιδεύσουν ένα σύστημα αναγνώρισης συναισθήματος. Εντούτοις, διαφορετικές χρονικές κλίμακες και μετρικά επίπεδα χαρακτηριστικών γνωρισμάτων που προέρχονται από διαφορετικές μορφές, καθώς επίσης και η τελική δημιουργία διανυσμάτων γνωρισμάτων με αυξημένες διαστάσεις επηρεάζουν την απόδοση ενός συστήματος βασισμένου σε αυτή την τεχνική συγχώνευσης. Η μεγάλη πλειοψηφία των μελετών για τη αναγνώριση συναισθήματος από πολλαπλές μορφές πληροφορίας ακολουθούν την συγχώνευση σε επίπεδο απόφασης (π.χ. [98] [109] [174] [262] [265] [244] και [191]). Στην προσέγγιση αυτή η είσοδος προερχόμενη από κάθε μορφή πληροφορίας επεξεργάζεται ανεξάρτητα και αποτελέσματα της μονοτροπικής αναγνώρισης συνδυάζονται στο τέλος. Δεδο-

μένου ότι οι άνθρωποι επιδεικνύουν ακουστικές και οπτικές εκφράσεις κατά τρόπο πλεοναστικό και συμπληρωματικό, η υπόθεση της υπό όρους ανεξαρτησίας μεταξύ των ακουστικών και οπτικών ροών πληροφορίας στην συγχώνευση σε επίπεδο απόφασης είναι ανακριβής και οδηγεί σε απώλεια πληροφορίας του αμοιβαίου συσχετισμού μεταξύ των δύο μορφών. Φυσικά, αυτή η ανεξάρτητη αντιμετώπιση των ροών πληροφορίας προσδίδει ευρωστία στο σύστημα καθώς η λειτουργία του δεν επηρεάζεται από την απουσία κάποιων χαρακτηριστικών λόγω θορύβου ή κάποιας αιτίας προερχόμενης από το περιβάλλον ή τον ανθρώπινο παράγοντα. Για να αντιμετωπιστεί το πρόβλημα της απώλειας πληροφορίας στην τεχνική συγχώνευσης σε επίπεδο απόφασης προτάθηκαν μέθοδοι συγχώνευσης ροών πληροφορίας που στοχεύουν στη εκμετάλλευση του συσχετισμού μεταξύ των ακουστικών και οπτικών μορφών πληροφορίας αλλά με την χαλάρωση της απαίτησης συγχρονισμού αυτών των ροών (π.χ. [190] [91] [211] [215] [261] και [266]). Στο [266] προτείνεται ένα Κρυφό Μαρκοβιανό μοντέλο βασισμένο σε πολλαπλές ροές που συγχωνεύονται για να αναπτυχθεί μια βέλτιστη σύνδεση μεταξύ των πολλαπλών ροών σύμφωνα με τη μέγιστη εντροπία. Στο [261] οι ίδιοι ερευνητές επέκτειναν αυτό το πλαίσιο συγχώνευσης με την εισαγωγή ενός ενδιάμεσου επιπέδου εκπαίδευσης. Στο [215] παρουσιάζεται ένα τριμερές (αντίστοιχο του συζευγμένου (coupled) αλλά για τρείς ροές) κρυφό Μαρκοβιανό μοντέλο για να μοντελοποιήσει τις ιδιότητες συσχετισμού τριών συστατικών: γνωρίσματα άνω προσώπου, κάτω προσώπου και δυναμικής προσωδικής συμπεριφοράς. Στο [91] οι ερευνητές πρότειναν ένα τεχνητό νευρωνικό δίκτυο (ANN) με ανατροφοδότηση πληροφοριών αποκαλούμενο ANNA για να ενσωματώσουν πληροφορίες από το πρόσωπο, την προσωδία και το λεξικολογικό περιεχόμενο.

Εν περιλήψει, ένας μεγάλος αριθμός μελετών στην ψυχολογία και την γλωσσολογία επιβεβαιώνουν τον συσχετισμό μεταξύ μερικών συναισθηματικών επιδείξεων (ειδικά βασικών συναισθημάτων) και συγκεκριμένων ακουστικών και οπτικών σημάτων (π.χ., [4] [77] και [199]). Η ανθρώπινη συμφωνία κρίσης είναι χαρακτηριστικά υψηλότερη για τη πληροφορία προερχόμενη από την έκφραση του προσώπου απ'ό,τι για τη πληροφορία φωνητικής έκφρασης. Εντούτοις, το ποσοστό συμφωνίας μειώνεται δραματικά όταν τα ερεθίσματα είναι αυθόρμητα επιδειχθείσες εκφράσεις συναισθηματικής συμπεριφοράς και όχι επιτηδευμένες και υπερβολικές. Επιπλέον, η έκφραση του προσώπου και η φωνητική έκφραση συναισθήματος εξετάζονται συχνά χωριστά και μια τέτοια προσέγγιση, μονοτροπικής αναγνώρισης και συνδυασμού σε επίπεδο απόφασης, αποκλείει την εύρεση στοιχείων χρονικού συσχετισμού μεταξύ των μορφών πληροφορίας.

2.2.5 Δυναμική

Ένας αυξανόμενος αριθμός ερευνητών της γνωστικής επιστήμης (cognitive science) υποστηρίζει ότι η δυναμική της ανθρώπινης συμπεριφοράς είναι εξαιρετικά κρίσιμη για την ερμηνεία και την ανάλυση της (π.χ., [77] [199] [204] και [209]). Παραδείγματος χάριν, έχει αποδειχθεί ότι η δυναμική της χρονικής εξέλιξης της έκφρασης του προσώπου αντιπροσωπεύει έναν κρίσιμο παράγοντα για τη διάκριση μεταξύ αυθόρμητης και επιτηδευμένης συμπεριφοράς (π.χ., [42] [77] και [228]) και για την κατηγοριοποίηση σύνθετων συναισθημάτων όπως πόνος, ντροπή και ευθυμία (π.χ., [77] [247] και [151]). Με βάση αυτά τα συμπεράσματα, αναμένουμε ότι η γενίκευση των ευρημάτων και σε άλλες μορφές πληροφορίας έχει κάποια βάση και οι χρονικοί συσχετισμοί μεταξύ των μορφών πληροφορίας διαδραματίζει σημαντικό ρόλο στην ερμηνεία της φυσικής

ανθρώπινης συναισθηματικής συμπεριφοράς.

2.2.6 Εξάρτηση από το πλαίσιο

Μια άλλη πτυχή της έρευνας στην συναισθηματική υπολογιστική είναι αυτή της εξάρτησης από το πλαίσιο που διαδραματίζεται η ανθρώπινη συμπεριφορά ενώ είναι αποδεκτό πως η ερμηνεία της συναισθηματικής συμπεριφοράς είναι σαφώς εξαρτώμενη από το ευρύτερο πλαίσιο στο οποίο συμβαίνει [199]. Παραδείγματος χάριν, ένα χαμόγελο μπορεί να είναι μια επίδειξη ευγένειας, ειρωνείας ή ακόμα και χαιρετισμού. Για να ερμηνευθεί πλήρως μια τέτοια ένδειξη ανθρώπινης συμπεριφοράς, είναι σημαντικό να υπάρχουν πληροφορίες σχετικά με το πλαίσιο στο οποίο αυτό το σήμα έχει επιδειχθεί, δηλ., που βρίσκεται ο χρήστης (π.χ. εσωτερικό χώρο, δρόμο, αυτοκίνητο), ποιοι είναι οι βραχυπρόθεσμοι στόχοι του, η ταυτότητα του αλλά και του δέκτη του σήματος αλλά και η μεταξύ τους σχέση. Ένα κρίσιμο ζήτημα στην συναισθηματική υπολογιστική είναι η ενσωμάτωση αυτής της γνώσης και ειδικότερα πληροφοριών σχετικών με τις συνθήκες και το πλαίσιο καταγραφής, τις συναισθηματικές συνήθειες του χρήστη και την διάθεση του στο άμεσο παρελθόν, τους βραχυπρόθεσμους και μακροπρόθεσμους στόχους του, την παρουσία άλλων ανθρώπων και την μεταξύ τους σχέση, κ.τ.λ. Οι επιδειχθήσες συμπεριφορές είναι στενά συσχετιζόμενες με αυτές τις καταστάσεις και ως εκ τούτου είναι επιβεβλημένος ο συνυπολογισμός τους στην διαδικασία εξαγωγής συναισθηματικών παραμέτρων και αναγνώρισης συναισθήματος. Η απουσία πληροφοριών σχετικών με το πλαίσιο αλλά και τις ιδιαιτερότητες του χρήστη προκαλεί ασάφεια και σύγχυση τόσο στους ειδικούς που πραγματοποιούν επισημείωση συναισθηματικής συμπεριφοράς όσο και στα αυτόματα συστήματα αναγνώρισης συναισθήματος, γεγονός που αποτελεί ισχυρό κίνητρο για την ενσωμάτωση της γνώσης αυτής.

2.2.7 Εφαρμογές

Παραδείγματα συστημάτων ικανά να αντιλαμβάνονται την συναισθηματική κατάσταση του χρήστη με είσοδο από πολλαπλές μορφές πληροφορίας στο πεδίο της Επικοινωνίας Ανθρώπου Μηχανής περιλαμβάνουν:

1. το [150], που συνδυάζει εκφράσεις του προσώπου και σήματα φυσιολογίας για να αναγνωρίσει συναισθήματα του χρήστη, όπως φόβο και θυμό και έπειτα προσαρμόζει την συμπεριφορά εμψύχωσης ενός εικονικού πράκτορα ώστε να αντανακλά το συναίσθημα του χρήστη

2. το [66], που εφαρμόζει ένα πρότυπο ενσωμάτωσης γνώσης που μπορεί να θεωρηθεί ως λεπτομερής αντιστοίχηση συναισθηματικών καταστάσεων του χρήστη και τύπων διεπαφών

3. το δυναμικό εργαλείο GazeX [154], που προσαρμόζει την αλληλεπίδραση βάσει της εξαρτούμενης από το πλαίσιο, πολυτροπικής ανάλυσης της συναισθηματικής συμπεριφορά του χρήστη

4. ο αυτοματοποιημένος εκπαιδευτικός συνοδός [127], που εντοπίζει ενδείξεις σύγχυσης του μαθητή και ανάγκης για βοήθεια συνδυάζοντας πληροφορίες από κάμερες, μια καρέκλα με αισθητήρες, το ποντίκι, έναν ασύρματο αισθητήρα εφαρμοσμένο στο δέρμα και την πρόοδο της εργασίας

5. το πολυτροπικό, υποβοηθούμενο από υπολογιστή σύστημα εκμάθησης στο ίδρυμα Beckman, του πανεπιστημίου του Ιλλινόις, όπου ένας εικονικός πράκτορας προσφέρει κατάλληλη βοήθεια βασισμένη σε πληροφορίες έκφρασης προσώπου, λέξεων κλειδιών, μετακίνησης ματιών και προόδου της εργασίας.

6. AgentDysl [9], λογισμικό υποβοήθησης ανάγνωσης και προσαρμογής της διαπροσωπείας με βάση την συναισθηματική κατάσταση.

2.3 Αναγνώριση δυναμικών συναισθηματικών καταστάσεων σε φυσική επικοινωνία ανθρώπου μηχανής

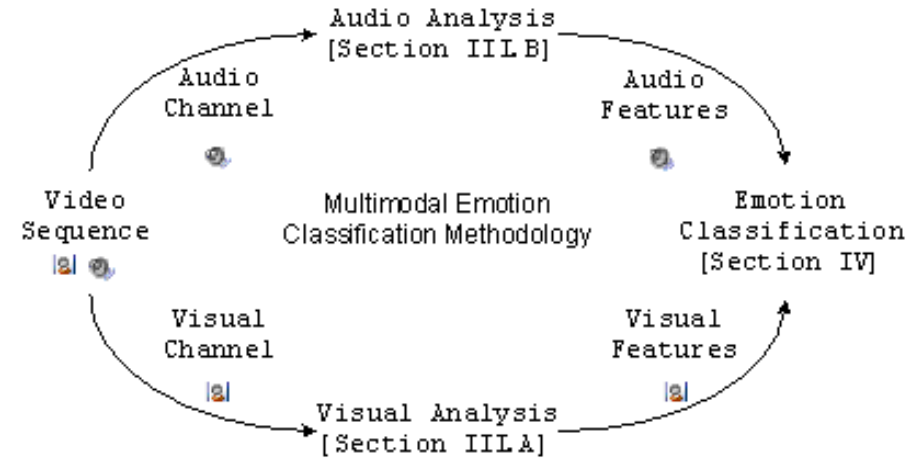
Η εργασία αυτή αφορά στην ανάλυση οπτικοακουστικών ακολουθιών με τη προτυποποίηση της συμπεριφοράς του χρήστη σε φυσικό HCI με γνώμονα την δυναμική εξέλιξη τους στον άξονα του χρόνου. Με τη χρήση ενός αναδρομικού νευρωνικού δικτύου, η βραχυπρόθεσμη μνήμη που παρέχεται μέσω της σύνδεσης ανατροφοδότησης λειτουργεί ως ενδιάμεση μνήμη και οι αποθηκευμένες τιμές λαμβάνονται υπόψη στον επόμενο χρονικό κύκλο. Η τεχνική αύτη στην τεχνητή νοημοσύνη στηρίζεται στην υπόθεση ότι τέτοιου είδους δίκτυα είναι κατάλληλα για να αναγνωρίσουν χρονικά μεταβαλλόμενα πρότυπα [79]. Πέρα από αυτό, τα αποτελέσματα δείχνουν ότι αυτή η προσέγγιση μπορεί να αποκωδικοποιήσει επιτυχώς ποικίλα πρότυπα εκφραστικότητας με την χρήση δικτύων σχετικά χαμηλής κλίμακας, το οποίο δεν συμβαίνει με άλλες εργασίες που εφαρμόζονται σε καταγραφές υποδυόμενων συμπεριφορών.

Η δομή της εργασίας είναι η εξής: η πρώτη ενότητα περιλαμβάνει την ευρύτερη αρχιτεκτονική της προτεινόμενης προσέγγισής 2.3.1 καθώς επίσης και παράδειγματα που θα χρησιμοποιηθούν στο κείμενο προκειμένου να διευκολυνθεί η παρουσίαση 2.3.2. Στην ενότητα 2.3.3 παρουσιάζουμε τους αλγορίθμους εξαγωγής χαρακτηριστικών γνωρισμάτων, για τις οπτικές και ακουστικές μορφές πληροφορίας εισόδου. Στην ενότητα 2.3.4 εξηγούμε πώς τα χαρακτηριστικά γνωρίσματα που εξάγονται, αν και πλήρως διαφορετικά όσον αφορά στην φύση και την μορφή τους, μπορούν να χρησιμοποιηθούν για να εκπαιδεύσουν ένα αναδρομικό νευρωνικό δίκτυο προκειμένου να εκτιμηθεί η συναισθηματική κατάσταση του ανθρώπου. Στην ενότητα 2.3.5 παρουσιάζουμε τα πειραματικά αποτελέσματα της υλοποίησης της μεθοδολογία μας σε φυσιοκρατικά δεδομένα και στην ενότητα 2.3.6 απαριθμούμε τα τελικά συμπεράσματα της εργασίας.

2.3.1 Εποπτική παρουσίαση αρχιτεκτονικής

Όπως αναφέρθηκε ήδη, η ευρύτερη προσέγγιση είναι βασισμένη σε μια επεξεργασία της ακολουθίας εισόδου σε επίπεδο πολλαπλών μορφών πληροφορίας. Υπάρχουν δύο διαφορετικές μεθοδολογίες που εμπίπτουν σε αυτήν την γενική κατηγορία, αυτές που το τελικό συμπέρασμα λαμβάνεται σε επίπεδο απόφασης και αυτές που λαμβάνεται σε επίπεδο γνωρισμάτων. Η προσέγγιση επιπέδου απόφασης έχει το πλεονέκτημα ότι εύκολα εφαρμόζεται σε ανεξάρτητα και πιθανόν διαφορετικής αρχιτεκτονικής, συστήματα, ήδη διαθέσιμα στη βιβλιογραφία. Η προσέγγιση συγχώνευσης σε επιπέδο γνωρισμάτων έχει το μειονέκτημα ότι συχνά είναι δύσκολο να εφαρμοστεί, αφού οι πληροφορίες στην είσοδο είναι συχνά διαφορετικής φύσης και είναι έτσι περίπλοκο να ενσωματωθούν σε ένα ενιαίο σχήμα επεξεργασίας, αλλά, όταν αυτό γίνεται επιτυχώς, παράγονται συστήματα που είναι σε θέση να επιτύχουν αρκετά καλύτερες επιδόσεις

αναγνώρισης. Η προτεινόμενη προσέγγιση ανήκει στην τελευταία κατηγορία και η αρχιτεκτονική της φαίνεται στην εικόνα 2.1



Σχήμα 2.1: Γραφική απεικόνιση της προτεινόμενης προσέγγισης

Η εξεταζόμενη ακολουθία εισόδου χωρίζεται σε ακουστικές και οπτικές ακολουθίες και επεξεργάζεται ώστε να εξαχθούν χαρακτηριστικά γνωρίσματα για κάθε ακολουθία δεδομένων. Η οπτική ακολουθία αναλύεται σε επίπεδο πλαισίου (frame) χρησιμοποιώντας την μεθοδολογία που παρουσιάζεται στην ενότητα 2.3.3.1 και εξηγείται αναλυτικά στο [118] ενώ η ακουστική ακολουθία αναλύεται σύμφωνα με την μέθοδο που παρουσιάζεται στην ενότητα 2.3.3.2. Τα οπτικά χαρακτηριστικά γνωρίσματα όλων των αντίστοιχων πλαισίων χρησιμοποιούνται ως είσοδος σε ένα αναδρομικό δίκτυο όπως εξηγείται στην ενότητα 2.3.4, όπου η δυναμική του οπτικού καναλιού χρησιμοποιείται για την ταξινόμηση της ακολουθίας σε μια από τις πέντε συναισθηματικές κατηγορίες που αναφέρονται στον πίνακα 2.23. Εξαιτίας του γεγονότος ότι τα χαρακτηριστικά γνωρίσματα που εξάγονται από το ακουστικό κανάλι είναι εντελώς διαφορετικά από εκείνα που εξάγονται από το οπτικό κανάλι, η αναδρομική δομή του δικτύου τροποποιείται κατάλληλα προκειμένου να επιτρέψει και τους δύο τύπους εισόδου στο δίκτυο, καταλήγοντας σε μια πραγματικά πολυμορφική αρχιτεκτονική ταξινόμησης, ενώ ένα πρωτότυπο σχήμα ολοκλήρωσης εξόδου αναλαμβάνει την λήψη της τελικής απόφασης.

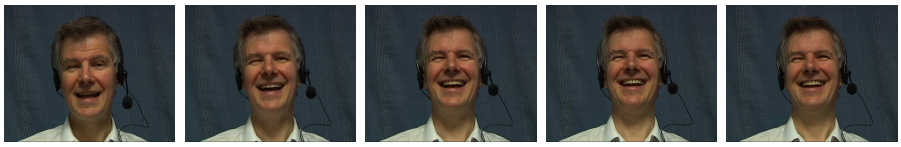
Η αξιολόγηση της απόδοσης της μεθοδολογίας περιλαμβάνει στατιστική ανάλυση των αποτελεσμάτων της εφαρμογής, ποσοτικές συγκρίσεις με άλλες προσεγγίσεις που εστιάζουν σε φυσιοκρατικά δεδομένα και ποιοτικές συγκρίσεις με άλλες γνωστές προσεγγίσεις στην αναγνώριση συναισθήματος, όπως παρουσιάζονται στην ενότητα 2.3.5.

2.3.2 Τρέχον παράδειγμα

Στην ανάπτυξη ενός συστήματος που λαμβάνει ως είσοδο πολλαπλές μορφές πληροφορίας είναι απαραίτητο να ενσωματώθουν διαφορετικά συστατικά που εξετάζουν διαφορετικές μορφές πληροφορίες. Κατά συνέπεια, η γενική αρχιτεκτονική περιλαμβάνει ποικιλία μεθοδολογιών και τεχνολογιών και συχνά είναι δύσκολο να περιγραφεί σε πλήρη έκταση. Προκειμένου να διευκολυνθεί η παρουσίαση της προσέγγισης πολλαπλών μορφών εισόδου, για την εκτίμηση της ανθρώπινης συναισθηματικής κατάστα-

σης, που προτείνεται θα χρησιμοποιήσουμε την έννοια του τρέχοντας παραδείγματος (running example).

Το παράδειγμά μας είναι ένα δείγμα από το σύνολο δεδομένων στο οποίο θα εφαρμοστεί η γενική μεθοδολογία στην ενότητα 2.3.5. Στην εικόνα 2.2 παρουσιάζονται μερικά πλαίσια από την ακολουθία του τρέχοντας παραδείγματος.



Σχήμα 2.2: Πλαίσια από το τρέχον παράδειγμα

2.3.3 Εξαγωγή χαρακτηριστικών γνωρισμάτων

2.3.3.1 Οπτική μορφή πληροφορίας

2.3.3.1.1 Τρέχον επίπεδο της επιστήμης Ο αυτόματος εντοπισμός χαρακτηριστικών σημείων του προσώπου είναι ένα δύσκολο πρόβλημα και, αν και πολλές εργασίες έχουν δημοσιευτεί για τον εντοπισμό και την παρακολούθηση τους [226], σχετικά μικρός αριθμός εργασιών αναφέρεται στο απαραίτητο, και συχνά χειροκακτικό ή ημιαυτόματο, βήμα αρχικοποίησης των αλγορίθμων παρακολούθησης [147], το οποίο απαιτείται στο πλαίσιο της εξαγωγής χαρακτηριστικών γνωρισμάτων του προσώπου και αναγνώρισης έκφρασης του προσώπου. Τα περισσότερα συστήματα αναγνώρισης εκφράσεων του προσώπου χρησιμοποιούν το πρότυπο κωδικοποίησης δράσης του προσώπου (Facial Action Coding System) που προτάθηκε από τους Ekman και Friesen [73] για την προτυποποίηση των εκφράσεων του προσώπου. Το FACS περιγράφει τις εκφράσεις χρησιμοποιώντας 44 μονάδες δράσης (AU) που αφορούν συστολές συγκεκριμένων μυών του προσώπου.

Επιπλέον του FACS, οι μετρικές MPEG–4 [222] χρησιμοποιούνται συνήθως για την προτυποποίηση εκφράσεων του προσώπου και των σχετιζόμενων συναισθημάτων. Αυτές ορίζουν έναν εναλλακτικό τρόπο προτυποποίησης των εκφράσεων του προσώπου και των υπονοούμενων συναισθημάτων, ο οποίος επηρεάζεται έντονα από νευροψυχολογικές και ψυχολογικές μελέτες. Το MPEG–4, εστιάζοντας κυρίως στη σύνθεση εκφράσεων του προσώπου και την εμψύχωση εικονικών χαρακτήρων, ορίζει τις παραμέτρους εμψύχωσης του προσώπου (Facial Animation Parameters) που είναι έντονα επηρεασμένες από τις μονάδες δράσης (Action Units), τον πυρήνα του FACS. Μια σύγκριση και αντιστοίχηση μεταξύ FAPs και AUs αναφέρεται στο [82].

Οι περισσότερες υπάρχουσες προσεγγίσεις στην εξαγωγή χαρακτηριστικών γνωρισμάτων του προσώπου είτε σχεδιάζονται ώστε να αντιμετωπίσουν περιορισμένη ποικιλομορφία βίντεο χαρακτηριστικών ή απαιτούν χειροκίνητη αρχικοποίηση ή παρέμβαση. Συγκεκριμένα η διαδικασία εξαγωγής χαρακτηριστικών σημείων του προσώπου στο [147] εξαρτάται από την οπτική ροή, στο [146] βασίζεται σε είσοδο βίντεο υψηλής ανάλυσης ή απαλλαγμένη από θόρυβο, στο [213] εξαρτάται μόνο από την πληροφορία χρώματος, στο [46] απαιτούνται δύο κάμερες τοποθετημένες στο κεφάλι (head-mounted) και στο [179] απαιτείται χειρωνακτική επισημείωση των χαρακτηριστικού σημείων στο πρώτο πλαίσιο κάθε βίντεο. Επιπλέον, πολύ λίγες υλοποιήσεις προσεγγίσεων μπορούν να εφαρμοστούν σε πραγματικό ή σχεδόν πραγματικό χρόνο. Στην

εργασία αυτή συνδυάζουμε ένα υποσύνολο από τις μεθοδολογίες ανίχνευσης χαρακτηριστικών γνωρισμάτων προκειμένου να παραχθεί ένα εύρωστο σύστημα εκτίμησης FAP.

2.3.3.1.2 Εντοπισμός προσώπου Το πρώτο βήμα στο στάδιο της ανίχνευσης των χαρακτηριστικών γνωρισμάτων του προσώπου είναι αυτό της ανίχνευσης του προσώπου. Σε αυτό το βήμα ο στόχος είναι να προσδιοριστεί εάν υπάρχουν ή όχι πρόσωπα στην εικόνα και, εάν ναι, να επιστραφεί η θέση τους στην εικόνα και το μέγεθος κάθε προσώπου [255]. Η ανίχνευση και ο εντοπισμός του προσώπου μπορεί να εκτελεσθεί με ποικίλες μεθόδους [182] [232] [87]. Στην εργασία αυτή επιλέχτηκε μια μη παραμετρική διακριτική ανάλυση με μια Μηχανή Διανυσμάτων Υποστήριξης (Support Vector Machine) να ταξινομεί τις περιοχές της εικόνας σε προσώπο και μη-πρόσωπο, μειώνοντας κατά συνέπεια τη διάσταση του προβλήματος εκπαίδευσης σε ένα μέρος του αρχικού, με αμελητέα απώλεια απόδοσης ταξινόμησης [93].

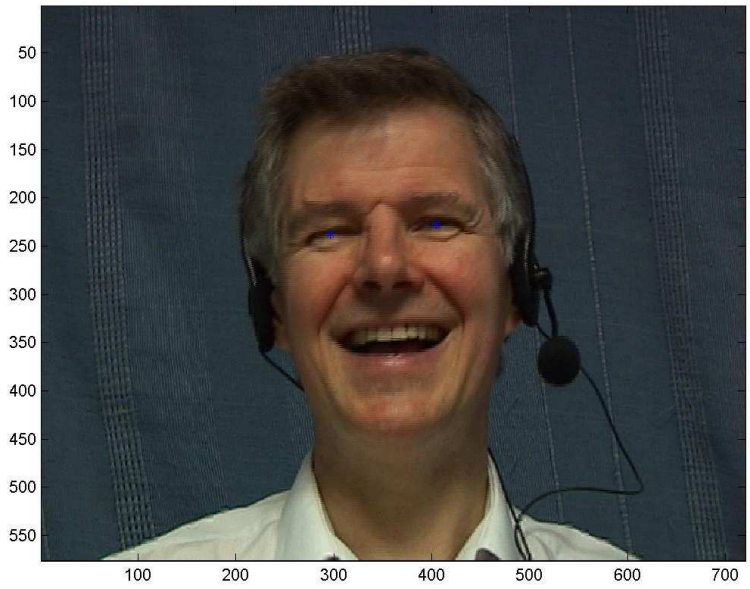
Προκειμένου να εκπαιδευθεί το SVM και να ρυθμιστούν οι λεπτομέρειες της διαδικασίας χρησιμοποιήθηκαν 800 παραδείγματα προσώπου από τη NIST ειδική βάση δεδομένων [192]. Όλα αυτά τα παραδείγματα ευθυγραμμίστηκαν πρώτα όσον αφορά στις συντεταγμένες των ματιών και του στόματος και υπέστησαν αλλαγή κλίμακας στο απαιτούμενο μέγεθος και έπειτα το σύνολο επεκτάθηκε με την τεχνητή εισαγωγή θορύβου που επιτεύχθηκε με την εφαρμογή διαταραχών μεταφοράς και περιστροφής μικρής κλίμακας σε όλα τα δείγματα, με συνέπεια το τελικό σύνολο εκπαίδευσης να αποτελείται από 16695 δείγματα.

Η ακρίβεια του βήματος εξαγωγής χαρακτηριστικών γνωρισμάτων που θα ακολουθήσει εξαρτάται σε μεγάλο βαθμό από την θέση του κεφαλιού και έτσι οι περιστροφές του προσώπου πρέπει να αντιμετωπιστούν πριν από την περαιτέρω επεξεργασία εικόνας. Αντιμετωπίστηκε μόνο η περιστροφή γύρω από τον άξονα που ορίζει η ευθεία κάθετη στο επίπεδο της εικόνας (+Z), δεδομένου ότι αυτός είναι και ο συχνότερος τύπος περιστροφής που εμφανίζεται σε τηλεοπτικές ακολουθίες πραγματικής ζωής. Έτσι, είναι απαραίτητο να υπολογιστεί αρχικά η θέση του κεφαλιού και έπειτα να περιστραφεί η εικόνα του προσώπου αναλόγως. Προκειμένου να υπολογιστεί η γωνία περιστροφής του κεφαλιού αρχικά εντοπίζονται τα δύο μάτια στην ανιχνευμένη περιοχή κεφαλιού.

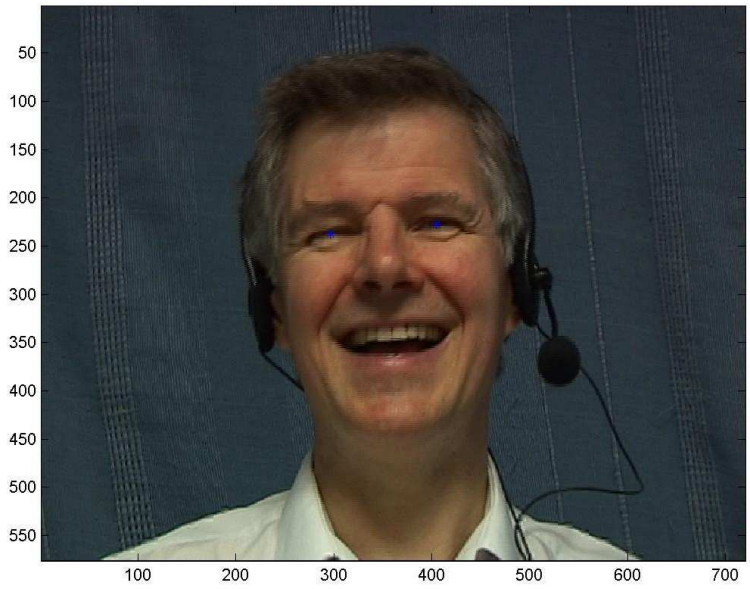
Για την διαδικασία αυτή χρησιμοποιούμε έναν πολυστρωματικό νευρώνα (Multi Layer Perceptron). Σαν συνάρτηση ενεργοποίησης επιλέγουμε μια sigmoidal συνάρτηση και για την μάθηση υιοθετούμε τον αλγόριθμο εκμάθησης Marquardt-Levenberg [102]. Για να εκπαιδευτεί το δίκτυο έχουμε χρησιμοποιήσει περίπου 100 τυχαίες εικόνες διαφορετικής ποιότητας, ανάλυσης και συνθηκών φωτισμού από τη βάση δεδομένων ERMIS [81], στις οποίες οι μάσκες των ματιών προσδιορίστηκαν χειρωνακτικά. Το δίκτυο έχει 13 νευρώνες εισόδου, οι οποίοι είναι οι τιμές φωτεινότητας, οι χρωματικές συνιστώσες Cb και Cr και οι 10 σημαντικότεροι συντελεστές DCT (με zigzag επιλογή) της γειτονικής 8x8 περιοχής εικονοστοιχείου. Οι έξοδοι είναι 2, μια για εικονοστοιχείο που ανήκει σε περιοχή ματιού και μια για περιοχές μη ματιών. Η αρχιτεκτονική του MLP έχει βελτιστοποιηθεί για να περιλάβει δύο κρυμμένα επίπεδα 20 νευρώνων.

Οι περιοχές των δύο ματιών στο πρόσωπο υπολογίζονται αρχικά κατά προσέγγιση χρησιμοποιώντας ανθρωπομετρικούς κανόνες που παρουσιάζονται στον πίνακα 2.1 και έπειτα το MLP εφαρμόζεται χωριστά σε κάθε εικονοστοιχείο στις δύο περιοχές ενδιαφέροντος. Για περιστροφές μέχρι 30 μοίρες, η μεθοδολογία αγγίζει ένα ποσοστό

επιτυχίας κοντά στο 100% στον εντοπισμό της κόρης οφθαλμού.



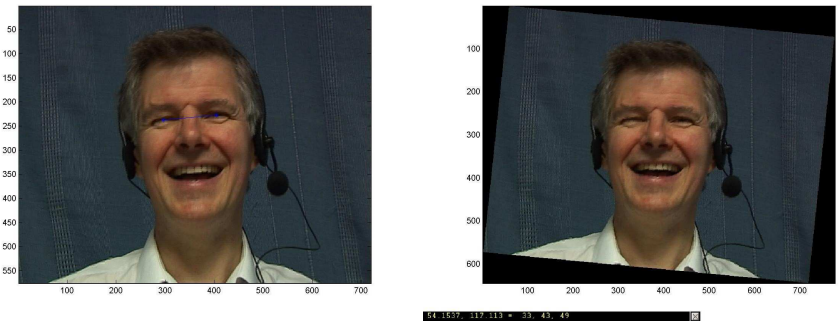
Σχήμα 2.3: Εντοπισμός ματιών με την χρήση του MLP



Σχήμα 2.4: Λεπτομέρεια από τον εντοπισμό ματιών με την χρήση του MLP

Στις εικόνες 2.3 και 2.4 βλέπουμε το αποτέλεσμα της εφαρμογής του MLP στο πρώτο πλαίσιο του τρέχοντας παραδείγματος. Μόλις εντοπίσουμε τις κόρες των ματιών, μπορούμε να υπολογίσουμε την περιστροφή του κεφαλιού με τον υπολογισμό της γωνίας μεταξύ του οριζόντιου επιπέδου και της ευθείας καθορισμένης από τα κέντρα των ματιών. Μπορούμε έπειτα να περιστρέψουμε το πλαίσιο εισόδου προκειμένου η θέση του κεφαλιού να εναρμονιστεί με την οριζόντια (βλ. εικόνα 2.5). Τέλος,

μπορούμε έπειτα να χωρίσουμε κατά προσέγγιση το πλαίσιο σε τρείς επικαλυπτόμενες ορθογώνιες περιοχές ενδιαφέροντος που περιλαμβάνουν τόσο τα χαρακτηριστικά γνωρίσματα του προσώπου όσο και περιοχές παρασκηνίου του προσώπου. Αυτές οι τρεις υποψήφιες περιοχές γνωρισμάτων είναι η περιοχή του αριστερού και δεξιού ματιού/φρυδιού και το στόμα (βλ. εικόνα 2.6). Η κατάτμηση είναι βασισμένη στους κατά προσέγγιση ανθρωπομετρικούς κανόνες που παρουσιάζονται στον πίνακα 2.1.



Σχήμα 2.5: Περιστροφή πλαισίου βάσει της θέσης των ματιών.

Πίνακας 2.1: Ανθρωπομετρικοί κανόνες για την εξαγωγή των υποψήφιων περιοχών γνωρισμάτων

Χαρακτηριστικό	Θέση	Πλάτος	Ύψος
Αριστερό μάτι και φρύδι	Πάνω αριστερά	0.6 x (πλάτος προσώπου)	0.5 x (ύψος προσώπου)
Δεξί μάτι και φρύδι	Πάνω δεξιά	0.6 x (πλάτος προσώπου)	0.5 x (ύψος προσώπου)
Στόμα και μύτη	Κάτω κέντρο	πλάτος προσώπου	0.5 x (ύψος προσώπου)

2.3.3.1.3 Εντοπισμός μύτης Τα χαρακτηριστικά σημεία της μύτης δεν χρησιμοποιούνται για την εκτίμηση της έκφρασης, αλλά είναι σταθερά σημεία που χρησιμοποιούνται στις μετρήσεις αποστάσεων αναφοράς για την εκτίμηση FAP (βλ. εικόνα 2.20). Κατά συνέπεια, αρκεί να βρεθεί η άκρη της μύτης και δεν είναι απραίτητο να βρεθούν ακριβώς τα όριά της. Η πιο κοινή προσέγγιση στον εντοπισμό μύτης ξεκινά με τον εντοπισμό των ρουθουνιών, επειδή τα ρουθούνια ανιχνεύονται εύκολα βάσει της χαμηλής έντασης φωτεινότητας τους. Για να προσδιοριστούν οι ακριβείς θέσεις των ρουθουνιών εφαρμόζουμε ένα κατώφλι t_n στο κανάλι φωτεινότητας της στοματικής περιοχής:

$$t_n = \frac{\overline{L^n} + 2\min\left(L^n\right)}{3} \tag{2.1}$$

όπου L^n είναι η μήτρα φωτεινότητας για την υπό εξέταση περιοχή και $\overline{L^n}$ είναι η μέση φωτεινότητα της περιοχής. Το αποτέλεσμα της κατωφλίωσης αυτής παρουσιάζεται στην εικόνα 2.7. Συνδεδεμένα αντικείμενα σε αυτόν τον δυαδικό χάρτη αριθμούνται και θεωρούνται ως υποψήφια ρουθούνια. Υπό κακές συνθήκες φωτισμού, σκιές μπορούν να εμφανιστούν εκατέρωθεν της μύτης, με συνέπεια περισσότερα από δύο υποψήφια ρουθούνια να εμφανίζονται στη μάσκα. Με την χρήση στατιστικών ανθρωπομετρικών στοιχείων για την απόσταση του αριστερού και του δεξιού ματιού

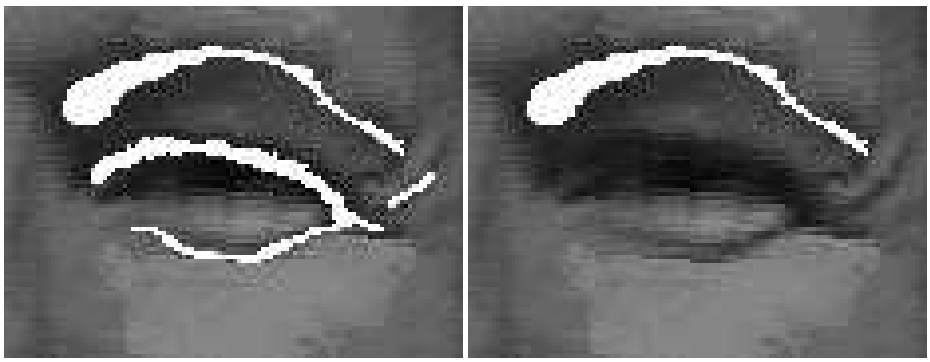
Σχήμα 2.6: Περιοχές ενδιαφέροντος για εξαγωγή χαρακτηριστικών γνωρισμάτων προσώπου

(bipupil width, D_{bp}) μπορούμε να αφαιρέσουμε τα άκυρα υποψήφια αντικείμενα και να προσδιορίσουμε τα πραγματικά ρουθούνια. Το κέντρο της μύτης ορίζεται ως το σημείο ανάμεσα στα ρουθούνια.

Σχήμα 2.7: Υποψήφιες θέσεις ρουθουνιών

2.3.3.1.4 Εντοπισμός φρυδιών Οι περιοχές των φρυδιών εξάγονται μέσω μιας μεθόδου βασισμένης στο γεγονός ότι έχουν απλή κατευθυντική μορφή και ότι βρίσκονται στο μέτωπο, το οποίο λόγω της μορφολογίας του, έχει γενικά ομοιόμορφο φωτισμό. Το πρώτο βήμα στην ανίχνευση φρυδιών είναι η κατασκευή ενός χάρτη ακμών της, διαβαθμίσεων του γκρι (gray-scaled), περιοχής ενδιαφέροντος ματιών και φρυδιών. Αυτός ο χάρτης κατασκευάζεται μετά από την αφαίρεση της διαστολής και της διάβρωσης της εικόνας διαβαθμίσεων του γκρι που χρησιμοποιεί μια ευθεία ώς δομικό στοιχείο. Ο εν λόγω μηχανισμός ανίχνευσης ακμών είναι κατάλληλος για τα φρύδια επειδή μπορεί να είναι κατευθυντικός, διατηρεί τα αρχικά χαρακτηριστικά γνωρίσματα όπως αυτό του μεγέθους και μπορεί να συνδυαστεί με ένα κατώφλι για την αφαίρεση μικρότερων ανωμαλιών του δέρματος όπως ρυτίδες. Αυτή η διαδικασία μπορεί να θεωρηθεί ως ειδική περίπτωση ενός υψιπερατού, μη γραμμικού φίλτρου.

Κάθε συνδεδεμένη περιοχή στον χάρτη ακμών εξετάζεται έπειτα ενάντια σε ένα σύνολο κριτηρίων φιλτραρίσματος που έχουν διαμορφωθεί μέσω της στατιστικής ανάλυσης χαρακτηριστικών των περιοχών των φρυδιών, όπως μήκος και θέση, σε 20 άτομα της βάσης δεδομένων ERMIS [81]. Τα αποτελέσματα αυτής της διαδικασίας



Σχήμα 2.8: Βήματα εντοπισμού φρυδιών

για το αριστερό φρύδι παρουσιάζονται στην εικόνα 2.8. Η ίδια διαδικασία εφαρμόζεται επίσης για το δεξί φρύδι.

2.3.3.1.5 Εντοπισμός ματιών Μεγάλο εύρος μεθοδολογιών έχουν προταθεί στη βιβλιογραφία για εξαγωγή διαφορετικών χαρακτηριστικών του προσώπου και ειδικά για τα μάτια, τόσο σε ελεγχόμενα όσο και σε ανεξέλεγκτα περιβάλλοντα. Το κοινό στοιχείο μεταξύ τους είναι ότι, ανεξάρτητα από το γενικό ποσοστό επιτυχίας που έχουν, όλες αποτυγχάνουν σε κάποιο σύνολο περιπτώσεων, λόγω των ιδιαίτερων δυσκολιών και εξωτερικών προβλημάτων που συνδέονται με το πρόβλημα. Κατά συνέπεια, δεν είναι λογικό να επιλεχθεί μια ενιαία μεθοδολογία και να εφαρμοστεί κατά βέλτιστο τρόπο σε όλες τις περιπτώσεις. Προκειμένου να ξεπερασθεί αυτό, επιλέγουμε να χρησιμοποιήσουμε πολλαπλές διαφορετικές τεχνικές προκειμένου να ανιχνευθούν τα χαρακτηριστικά γνωρίσματα του προσώπου που παρουσιάζουν την μεγαλύτερη πρόκληση, δηλ. τα μάτια και το στόμα και να συνδυάσουμε το μεμονωμένα αποτελέσματα τους.

Μάσκα βασισμένη σε MLP

Αυτή η προσέγγιση καθορίζει τις θέσεις ματιών που εξάγονται από το νευρωνικό δίκτυο MLP που χρησιμοποιήθηκε, παραπάνω, προκειμένου να προσδιοριστούν οι κόρες των ματιών στη φάση ανίχνευσης περιοχών των ματιών. Στηρίζεται στο γεγονός ότι τα βλέφαρα εμφανίζονται συνήθως σκοτεινότερα από το δέρμα λόγω των βλεφαρίδων και είναι σχεδόν πάντα γειτονικά της ίριδας. Κατά συνέπεια, συμπεριλαμβάνοντας σκοτεινά αντικείμενα κοντά στο κέντρο ματιών, προσθέτουμε τις βλεφαρίδες και την ίριδα στην μάσκα ματιών. Το αποτέλεσμα απεικονίζεται στην εικόνα 2.9.

Σχήμα 2.9: Μάσκα υπολογιζόμενη από το MLP δίκτυο

Μάσκα βασισμένη σε ακμές

Αυτή είναι μια μάσκα που περιγράφει την περιοχή μεταξύ των άνω και κάτω βλεφάρων. Δεδομένου ότι το κέντρο του ματιού ανιχνεύεται σχεδόν πάντα σωστά από το MLP, οι οριζόντιες ακμές των βλεφάρων στην περιοχή γύρω από το μάτι χρησιμοποιούνται για να περιορίσουν τη μάσκα ματιών κατά την κάθετη κατεύθυνση. Για την ανίχνευση των οριζόντιων ακμών χρησιμοποιούμε τον Canny τελεστή ακμών λόγω της ιδιότητας του καλού εντοπισμού. Από όλες τις ακμές που ανιχνεύονται στην

εικόνα επιλέγουμε αυτές ακριβώς πάνω από και κάτω από το ανιχνευμένο κέντρο ματιών και συμπληρώνουμε την περιοχή μεταξύ τους προκειμένου να υπολογισθεί η τελική μάσκα ματιών. Το αποτέλεσμα απεικονίζεται στην εικόνα 2.10.

Σχήμα 2.10: Η μάσκα ματιών βασισμένη στις ακμές

Μάσκα βασισμένη σε εξάπλωση περιοχών

Αυτή η μάσκα δημιουργείται χρησιμοποιώντας μια τεχνική επέκτασης περιοχής. Η τελευταία συνήθως δίνει πολύ καλά αποτελέσματα κατάτμησης που αντιστοιχούν καλά στις παρατηρηθείσες ακμές. Η κατασκευή αυτής της μάσκας στηρίζεται στο γεγονός ότι η υφή του προσώπου είναι πιό σύνθετη και σκοτεινότερη μέσα στην περιοχή ματιών και ειδικά στα σύνορα βλεφάρων-χιτώνα-ίριδας, απ'ό,τι στις περιοχές γύρω τους. Αντί της χρησιμοποίησης ενός κριτηρίου πυκνότητας ακμών, χρησιμοποιούμε μια απλή όμως αποτελεσματική νέα μέθοδο υπολογισμού του κέντρου και της μάσκας των ματιών.

Για κάθε εικονοστοιχείο στην περιοχή του κέντρου του ματιού υπολογίζουμε την σταθερή απόκλιση του καναλιού φωτεινότητας σε μια γειτονιά 3x3 και έπειτα εφαρμόζουμε ένα κατώφλι στο αποτέλεσμα χρησιμοποιώντας ως τιμή του κατωφλίου την φωτεινότητα του ίδιου του εικονοστοιχείου. Αυτή η διαδικασία οδηγεί σε μια ευρύτερη περιοχή του κέντρου του ματιού ώστε να περιληφθούν μερικά από τα παρακείμενα χαρακτηριστικά του προσώπου. Η διαδικασία επαναλαμβάνεται για γειτονιές 5x5 και με τη χρησιμοποίηση μεγαλύτερης γειτονιάς ενισχύουμε την ευρωστία της διαδικασίας ενάντια σε παραλλαγές της ποιότητας της εικόνας και τον θόρυβο. Τα δύο αποτελέσματα συγχωνεύονται ώστε να παραχθεί η τελική μάσκα που απεικονίζεται στην εικόνα 2.11. Η διαδικασία αποτυγχάνει συχνότερα από τις άλλες χρησιμοποιούμενες προσεγγίσεις, αλλά αποδίδει πολύ καλά σε εικόνες χαμηλής ανάλυσης και ποιότητας χρώματος. Η γενική διαδικασία είναι αρκετά παρόμοια με αυτή μιας μορφολογικής λειτουργίας bottom hat, με τη διαφορά ότι η τελευταία είναι πιο ευαίσθητη στο μέγεθος των δομικών στοιχείων.

Σχήμα 2.11: Μάσκα των ματιών βασισμένη στην απόκλιση της φωτεινότητας

Μάσκα βασισμένη στην φωτεινότητα

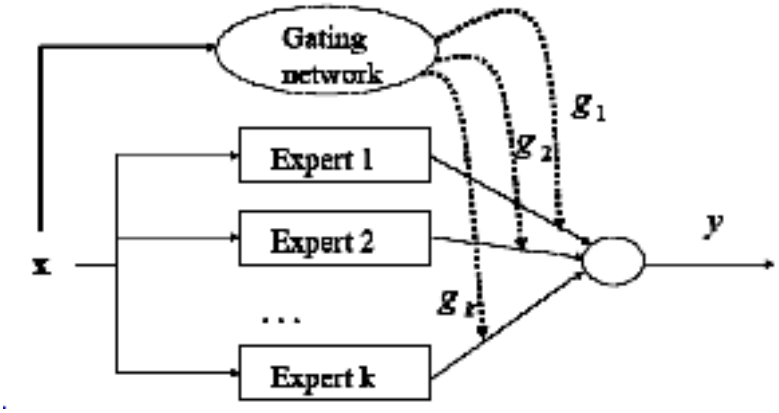
Τέλος, κατασκευάζεται μια δεύτερη μάσκα βασισμένη στην φωτεινότητα για εξαγωγή συνόρων του ματιού και του βλεφάρου, που χρησιμοποιεί την κανονική πιθανότητα της φωτεινότητας και ένα απλό προσαρμοστικό κατώφλι στην περιοχή των ματιών. Το αποτέλεσμα είναι συνήθως μια απεικόνιση περιοχών στα όρια του ματιού. Σε μερικές περιπτώσεις, εν τούτοις, οι τιμές φωτεινότητας γύρω από το μάτι είναι πολύ χαμηλές εξαιτίας σκιών από τα φρύδια και τον άνω μέρος της μύτης. Για την βελτίωση του αποτελέσματος σε τέτοιες περιπτώσεις, η ανιχνευμένη περιοχή περικόπτεται κάθετα στα λεπτύτερα σημεία εκατέρωθεν του κέντρου ματιών. Το προκύπτον κυρτό κέλυφος (convex hull) της μάσκας απεικονίζεται στην εικόνα 2.12.

Σχήμα 2.12: Μάσκα των ματιών βασισμένη στην φωτεινότητα

Συγχώνευση μασκών

Ο λόγος που έχουμε επιλέξει να χρησιμοποιήσουμε τέσσερις διαφορετικές μάσκες είναι ότι δεν υπάρχει στη βιβλιογραφία κάποια τυποποιημένη μέθοδος επιλογής της ιδανικής μεθοδολογίας εντοπισμού ματιών για μια δεδομένη εικόνα προσώπου. Συνεπώς, έχοντας τέσσερις ανιχνευμένες μάσκες δεν είναι εύκολο να καταλήξουμε ποια είναι η καταλληλότερη και να την επιλέξουμε ως τελικό αποτέλεσμα της γενικής ενότητας εντοπισμού ματιών. Αντ' αυτού, επιλέγουμε να συνδυάσουμε τις διαφορετικές μάσκες χρησιμοποιώντας μια μηχανή επιτροπείας (committee machine).

Λαμβάνοντας υπόψη το γεγονός ότι καθεμία από τις διαφορετικές μεθοδολογίες που χρησιμοποιούνται έχει μερικά γνωστά ισχυρά και αδύνατα σημεία, η μηχανή επιτροπείας είναι η καταλληλότερη προσέγγιση συνδυασμού των μασκών γιατί στόχος είναι η δυναμική δομή του συνόλου εμπειρογνωμόνων, που τροποποιείται κατάλληλα για να προσαρμοστεί στις απαιτήσεις της εφαρμογής [106]. Η γενική δομή αυτής της μεθοδολογίας παρουσιάζεται στην εικόνα 2.13. Αποτελείται από k επιβλεπόμενες ενότητες, γνωστοί και ως εμπειρογνώμονες και ένα δίκτυο με πύλες που λειτουργεί ως μεσολαβητής μεταξύ των εμπειρογνωμόνων. Η κύρια υπόθεση είναι ότι καθένας από τους εμπειρογνώμονες λειτουργούν καλύτερα σε διαφορετικές περιοχές του διαστήματος εισόδου, σύμφωνα με ένα πιθανολογικό πρότυπο που είναι πρότερα γνωστό και ως εκ τούτου προκύπτει η ανάγκη χρήσης δίκτυου με πύλες.



Σχήμα 2.13: Αρχιτεκτονική συνδυασμού εμπειρογνωμόνων

Ο ρόλος του δικτύου πυλών είναι να υπολογίσει, βασισμένο στην είσοδο,την πιθανότητα g_i κάθε μεμονωμένος εμπειρογνώμονας i να λειτουργεί σωστά και να παρέχει τις εκτιμήσεις αυτές στην ενότητα συνδυαστών. Το δίκτυο πυλών αποτελείται από ένα επίπεδο νευρώνων softmax. Η επιλογή της softmax ως συνάρτηση ενεργοποίησης για τους νευρώνες έχει τις σημαντικές ιδιότητες:

$$0 \leq g_i \leq 1, \forall i \in 1..k$$

$$\sum_{i=1}^{k} g_i = 1$$

δηλ. επιτρέπει τις εκτιμήσεις να ερμηνευτούν ως πιθανότητες. Στην εφαρμογή μας έχουμε $k = 4$ ειδικούς, τις αντίστοιχες υλοποιήσεις μεθοδολογιών ανίχνευσης ματιών που παρουσιάστηκαν νωρίτερα. Το δίκτυο με πύλες ευνοεί τις μεθόδους εξαγωγής χαρακτηριστικών γνωρισμάτων βασισμένες στο χρώμα σε εικόνες υψηλής ανάλυσης και ποιότητα χρώματος, ενσωματώνοντας κατά συνέπεια τις πρότερα γνωστές πιθανότητες της επιτυχίας για τους εμπειρογνώμονές μας στη διαδικασία συνδυασμού.

Επιπλέον, η ενότητα συνδυασμού που λειτουργεί κανονικά ως $y = \bar{g} \cdot \bar{e}$, όπου $\bar{e}$ είναι το διάνυσμα των εκτιμήσεων των ειδικών, τροποποιείται στην εργασία μας για να λειτουργήσει ως:

$$y = \frac{\bar{g} \cdot \bar{f} \cdot \bar{e}}{|\bar{f}|}$$

όπου f είναι το διάνυσμα των τιμών εμπιστοσύνης που συνδέονται με την έξοδο κάθε εμπειρογνώμονα, βελτιώνοντας περαιτέρω την ποιότητα διαδικασίας συνδυασμού μασκών. Οι τιμές εμπιστοσύνης υπολογίζονται με τη σύγκριση της θέσης, της μορφής και του μεγέθους των ανιχνευμένων μασκών με εκείνες που βασίζονται σε ανθρωπομετρικές στατιστικές μελέτες.

Η τροποποιημένη μονάδα συνδυάζει τις τέσσερις μάσκες σε επίπεδο απόφασης εικονοκυττάρου. Το αποτέλεσμα της διαδικασίας για το αριστερό μάτι του πλαισίου του τρέχοντας παραδείγματος απεικονίζεται στην εικόνα 2.14.

Σχήμα 2.14: Η τελική μάσκα για το αριστερό μάτι

2.3.3.1.6 Εντοπισμός στόματος Ομοίως με την περίπτωση των ματιών, το στόμα είναι ένα από τα χαρακτηριστικά γνωρίσματα του προσώπου που δεν ανιχνεύεται και εντοπίζεται πάντα επιτυχημένα, κυρίως λόγω του εύρους των παραμορφώσεων που παρατηρούνται σε αυτό κατά την διάρκεια ακολουθιών όπου ο άνθρωπος μιλά, το οποίο είναι και η συνηθέστερη περίπτωση στην εφαρμογής μας. Σε αυτή την εργασία χρησιμοποιούμε τις ακόλουθες μεθοδολογίες προκειμένου να υπολογιστεί η θέση και τα όρια του στόματος:

Μάσκα βασισμένη σε MLP

Ένα νευρωνικό δίκτυο MLP εκπαιδεύεται για να προσδιορίσει την στοματική περιοχή με την χρήση μιας ουδέτερης εικόνας. Το δίκτυο έχει παρόμοια αρχιτεκτονική με αυτήν που χρησιμοποιείται στην περίπτωση των ματιών. Τα δεδομένα εκπαίδευσης προέρχονται από ουδέτερες εικόνες. Δεδομένου ότι το στόμα είναι κλειστό στην ουδέτερη εικόνα, υπάρχει μια μακριά περιοχή χαμηλής φωτεινότητας μεταξύ των χειλιών. Κατά συνέπεια, η υποψήφια περιοχή ενδιαφέροντος φιλτράρεται αρχικά με εναλλασσόμενο διαδοχικό φιλτράρισμα μέσω ανακατασκευής (ASFR Alternating Sequential Filtering by Reconstruction - ASFR) ώστε να απλοποιηθούν και να δημιουργηθούν συνδεδεμένες περιοχές παρόμοιας φωτεινότητας. Κατωφλιοποίηση της φωτεινότητα εφαρμόζεται έπειτα για να εντοπιστεί η περιοχή μεταξύ των χειλιών. Αυτή η περιοχή διαστέλλεται κάθετα και τα στοιχεία που απεικονίζονται από αυτήν την περιοχή χρησιμοποιούνται για να εκπαιδευτεί το δίκτυο.

Το δίκτυο MLP που έχει εκπαιδευθεί στο ουδέτερο πλαίσιο έκφρασης είναι αυτό που χρησιμοποιείται για να παραχθεί μια εκτίμηση της στοματικής περιοχής στα υπόλοιπα πλαίσια της ακολουθίας. Η έξοδος του νευρωνικού δικτύου στη στοματική περιοχή ενδιαφέροντος υπόκειται κατωφλιοποίηση προκειμένου να διαμορφωθεί ένας δυαδικός χάρτης που περιέχει διάφορες μικρές υποπεριοχές. Το κυρτό περίβλημα των περιοχών αυτών υπολογίζεται για να παραχθεί η τελική μάσκα για το στόμα. Το αποτέλεσμα αυτής της διαδικασίας απεικονίζεται στην εικόνα 2.15.

Σχήμα 2.15: Μάσκα του στόματος βασισμένη στο MLP δίκτυο

Μάσκα βασισμένη σε ακμές

Σε αυτήν την δεύτερη προσέγγιση, το κανάλι φωτεινότητας της στοματικής περιοχής φιλτράρεται πάλι χρησιμοποιώντας ASFR φίλτρα για την απλοποίηση της εικόνας. Η οριζόντια μορφολογική κλίση της στοματικής περιοχής ενδιαφέροντος υπολογίζεται έπειτα. Από τη θέση της μύτης που έχει ανιχνευθεί ήδη, και, όπως έχουμε εξηγήσει ήδη, η διαδικασία ανίχνευσης της μύτης σπάνια αποτυγχάνει, μπορούμε να χρησιμοποιήσουμε τη θέση της μύτης για την καθοδήγηση της διαδικασίας ανίχνευσης της στοματικής περιοχής. Κατά συνέπεια, τα συνδεδεμένα στοιχεία που είναι κοντά στο κέντρο της μύτης και είναι υποψήφια να ανήκουν στην περιοχή του στόματος αφαιρούνται. Από το υπόλοιπο της μάσκας, πολύ μικρά αντικείμενα αφαιρούνται επίσης. Ένα μορφολογικό κλείσιμο εκτελείται έπειτα και το μεγαλύτερο, κατά τον οριζόντιο άξονα, των υπόλοιπων αντικειμένων επιλέγεται ως τελική στοματική μάσκα. Το αποτέλεσμα αυτής της διαδικασίας απεικονίζεται στην εικόνα 2.16.

Σχήμα 2.16: Στοματική μάσκα βασισμένη στις ακμές

Μάσκα βασισμένη στις γωνίες των χειλιών

Το κύριο πρόβλημα των περισσότερων μεθόδων βασισμένων στην φωτεινότητα για την ανίχνευση του στόματος είναι η ύπαρξη των άνω δοντιών, τα οποία τείνουν να αλλάξουν την ομοιομορφία έντασης της περιοχής. Η τελευταία προσέγγισή μας στην ανίχνευση της στοματικής περιοχής εκμεταλλεύεται τη σχετικά χαμηλή φωτεινότητα των άκρων των χειλιών και συμβάλλει στον σωστό προσδιορισμό της οριζόντιας έκτασης του στόματος που δεν ανιχνεύεται πάντα με τις προηγούμενες μεθόδους. Η εικόνα αρχικά κατωφλιοποιείται παρέχοντας μια εκτίμηση του εσωτερικού της στοματικής περιοχής, ή της περιοχής μεταξύ των χειλιών σε περίπτωση κλειστού στόματος. Κατόπιν, διακρίνουμε δύο διαφορετικές περιπτώσεις:

1. κανένα δόντι δεν είναι εμφανές και η στοματική περιοχή δείχνεται από μια συνεκτική σκοτεινή περιοχή

2. τα δόντια είναι εμφανή και έτσι δύο σκοτεινές περιοχές εμφανίζονται εκατέρωθεν των δοντιών

Στην πρώτη περίπτωση η ανίχνευση της έκτασης του στόματος είναι απλή. Στην δεύτερη αξιολογείται η εγγύτητα του στοματικού κέντρου κάθε αντικειμένου και τα κατάλληλα αντικείμενα επιλέγονται. Το κυρτό περίβλημα του αποτελέσματος συγχωνεύεται έπειτα με μορφολογική ανακατασκευή με έναν χάρτη οριζόντιων ακμών για να περιλάβει τα άνω και κάτω χείλια.

Προκειμένου να ταξινομηθεί η στοματική περιοχή σε μια από τις δύο περιπτώσεις και να εφαρμοστεί αντίστοιχη μεθοδολογία στοματικής ανίχνευσης αρχίζουμε με την επιλογή του μεγαλύτερου συνδεδεμένου αντικειμένου της εικόνας εισόδου και εύρεση του κέντρου του. Εάν η οριζόντια θέση του κέντρου του είναι κοντά στην οριζόντια θέση της μύτης, υποθέτουμε ότι το αντικείμενο αυτό είναι πραγματικά το εσωτερικό του στόματος και έχουμε την πρώτη περίπτωση όπου δεν υπάρχει κανένα εμφανές δόντι. Εάν το κέντρο δεν είναι κοντά στην οριζόντια θέση της μύτης υποθέτουμε ότι έχουμε τη δεύτερη περίπτωση όπου υπάρχουν προφανή δόντια και το αντικείμενο που εξετάζεται είναι η σκοτεινή περιοχή μιας από τις δύο πλευρές των δοντιών.

Το αποτέλεσμα της εφαρμογής αυτής της μεθοδολογίας στο τρέχον παράδειγμα απεικονίζεται στην εικόνα 2.17.

Σχήμα 2.17: Η στοματική μάσκα βασισμένη στις άκρες των χειλιών

Συγχώνευση μασκών

Ο συνδυασμός των μασκών εκτελείται χρησιμοποιώντας ένα σύστημα εμπειρογνωμόνων παρόμοιο με αυτό που χρησιμοποιείται για τον συνδυασμό των διαφορετικών μασκών για τα μάτια. Η κύρια διαφορά εδώ είναι ότι δεν μπορούμε να αξιολογήσουμε την πιθανότητα επιτυχίας για κάθε μια εκ των μεθόδων χρησιμοποιώντας πληροφορίες εύκολα διαθέσιμες στην είσοδο, όπως η ανάλυση ή το βάθος χρώματος, και επομένως το δίκτυο πυλών έχει τον τετριμμένο ρόλο ανάθεσης και στους τρεις εμπειρογνώμονες της ίδιας τιμής εμπιστοσύνης.

Αυτό δεν σημαίνει ότι η ενότητα συνδυασμού εξόδων είναι επίσης τετριμμένη. Το αντίθετο, χρησιμοποιούμε ακόμα την τροποποιημένη έκδοση της ενότητας, όπου οι ανθρωπομετρικές στατιστικές χρησιμοποιούνται για να επικυρώσουν τις τρεις μάσκες και ο βαθμός επικύρωσης χρησιμοποιείται στο στάδιο του συνδυασμού των μασκών. Η προκύπτουσα μάσκα για το στόμα στο υπό εξέταση πλαίσιο απεικονίζεται στην εικόνα 2.18.

Σχήμα 2.18: Η τελική μάσκα του στόματος

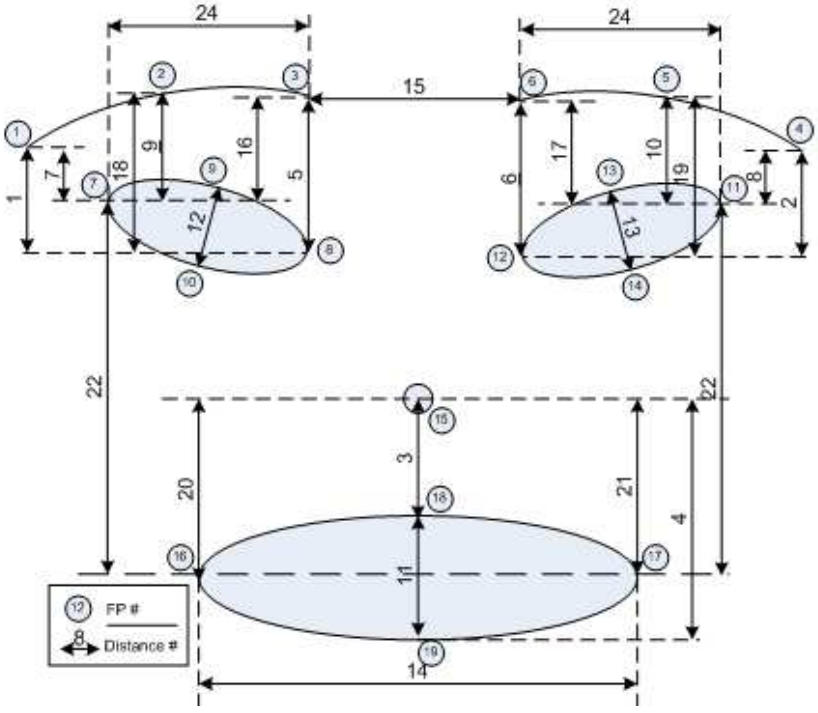
2.3.3.1.7 Εξέχοντα σημεία και FAPs Οι μάσκες χαρακτηριστικών γνωρισμάτων του προσώπου που ανιχνεύονται στην προηγούμενη ενότητα δεν χρησιμοποιούνται άμεσα για τη διαδικασία αναγνώρισης συναισθήματος. Αποτελούν την βάση για άλλες μορφές πληροφορίας. Συγκεκριμένα, χρησιμοποιούμε τις μάσκες προκειμένου να ανιχνευθούν τα χαρακτηριστικά οριακά σημεία των στοιχείων του προσώπου. Ο πίνακας 2.2 παρουσιάζει τον πλήρη κατάλογο σημείων που ανιχνεύονται στο ανθρώπινο πρόσωπο, ως ένα υποσύνολο του πλήρους καταλόγου σημείων χαρακτηριστικών γνωρισμάτων του προσώπου που καθορίζονται στο πρότυπο MPEG–4 [222]. Παραδείγματος χάριν, η εικόνα 2.19 απεικονίζει τα σημεία χαρακτηριστικών γνωρισμάτων που ανιχνεύονται στο πλαίσιο του τρέχοντας παραδείγματος.

Πίνακας 2.2: Χαρακτηριστικά σημεία

Σημείο	**MPEG–4**	**Περιγραφή**
1	4.5	Εξωτερικό σημείο αριστερού φρυδιού
2	4.3	Μεσαίο σημείο αριστερού φρυδιού
3	4.1	Εσωτερικό σημείο αριστερού φρυδιού
4	4.6	Εξωτερικό σημείο δεξιού φρυδιού
5	4.4	Μεσαίο σημείο δεξιού φρυδιού
6	4.2	Εσωτερικό σημείο δεξιού φρυδιού
7	3.7	Εξωτερικό σημείο αριστερού ματιού
8	3.11	Εσωτερικό σημείο αριστερού ματιού
9	3.13	Άνω σημείο αριστερής βλεφαρίδας
10	3.9	Κάτω σημείο αριστερής βλεφαρίδας
11	3.12	Εξωτερικό σημείο δεξιού ματιού
12	3.8	Εσωτερικό σημείο δεξιού ματιού
13	3.14	Άνω σημείο δεξιάς βλεφαρίδας
14	3.10	Κάτω σημείο δεξιάς βλεφαρίδας
15	9.15	Μύτη
16	8.3	Αριστερή γωνία στόματος
17	8.4	Δεξιά γωνία στόματος
18	8.1	Άνω σημείο στόματος
19	8.2	Κάτω σημείο στόματος

Δεδομένου ότι όταν οι άνθρωποι αλλάζουν την έκφρασή του προσώπου τους μεταβάλλουν και τη θέση μερικών εκ αυτών των σημείων (βλ. πίνακα 2.3) η κύρια μονάδα πληροφορίας που θα εξεταστεί κατά τη διάρκεια της ταξινόμησης συναισθήματος θα είναι το σύνολο FAPs που χαρακτηρίζουν ένα πλαίσιο. Προκειμένου να παραχθεί αυτό το σύνολο αρχίζουμε με τον υπολογισμό ενός διανύσματος αποστάσεων d (διάστασης 25) που περιέχει τις κάθετες και οριζόντιες αποστάσεις μεταξύ των 19 FPs που εντοπίστηκαν, όπως φαίνεται στην εικόνα 2.20 . Η μονάδα μέτρησης δεν είναι το εικονοκύτταρο, αλλά η κανονικοποιημένη σταθερή κλίμακα μονάδων MPEG–4, δηλαδή. ENS, MNS, MW, IRISD και ES [222]. Οι μοναδιαίες ποσότητες μετριούνται άμεσα από τις αποστάσεις FP στην ουδέτερη εικόνα, παραδείγματος χάριν το ES υπολογίζεται ως η απόσταση μεταξύ FP_9 και FP_13 (απόσταση μεταξύ των κόρων ματιών). Το πρώτο βήμα είναι να δημιουργηθεί το διάνυσμα απόστασης αναφοράς $\bar{d}_n$ με την επεξεργασία του ουδέτερου πλαισίου και τον υπολογισμό των αποστάσεων που περιγράφονται στην εικόνα 2.20 και έπειτα δημιουργείται ένα παρόμοιο διάνυσμα από-

Σχήμα 2.19: Χαρακτηριστικά σημεία εντοπισμένα στο πλαίσιο εισόδου



Σχήμα 2.20: Αποστάσεις χαρακτηριστικών σημείων

στασης $\bar{d}_i$ για κάθε υπό εξέταση πλαίσιο i. Τα FAPs υπολογίζονται με τη σύγκριση $\bar{d}_n$ και $\bar{d}_i$.

2.3.3.2 Ακουστική μορφή πληροφορίας

2.3.3.2.1 Τρέχον επίπεδο της επιστήμης Η ομιλία θεωρείται ως ροή πληροφορίας που μπορεί να χρησιμοποιηθεί ως είσοδο προκειμένου να αναγνωριστεί το ανθρώ-

Πίνακας 2.3: Χαρακτηριστικά σημεία πλαισίων από διαφορετικές ακολουθίες

πινο συναίσθημα [13], [171]. Σε αυτή την πρόωρη εργασία [12] η ομιλία χρησιμοποιήθηκε για να διαχωρίσει δυαδικά μεταξύ αρνητικού (καλύπτοντας τις καταστάσεις χρηστών όπως ο θυμός, η ενόχληση, ή η απογοήτευση) και των υπολοίπων καταστάσεων, δηλ. ουδέτερος, απογοητευμένος, κουρασμένος, χαρούμενος, κ.α. Ο κύριος λόγος για την αντιστοίχηση αυτήν στο αρνητικό ημιεπίπεδο έναντι του ουδέτερου/θετικού ημιεπίπεδου αξιολόγησης ήταν στην συγκεκριμένη εφαρμογή, να ανιχνευτεί το πρόβλημα στην επικοινωνία. Πιο πρόσφατες μελέτες έχουν κατορθώσει να επεκτείνουν τις κατηγορίες ταξινόμησης σε τρείς [2] ή ακόμα και οκτώ [166], δείχνοντας κατά συνέπεια οτι η ομιλία είναι μια μορφή πληροφορίας που μπορεί πραγματικά να παρέχει σημαντικές ενδείξεις σχετικά με την συναισθηματική κατάσταση.

Το σύνολο των χαρακτηριστικών γνωρισμάτων που χρησιμοποιούνται για να ποσοτικοποιήθουν οι προσωδικές παραλλαγές στην ομιλία, που χρησιμεύει ως βάση για την ταξινόμηση, επίσης εξελίσσεται συνεχώς. Ενώ παλαιότερες μελέτες έκλιναν προς την εκτίμηση της $F0$ ως κεντρικό φορέα συναισθηματικού περιεχομένου, πρόσφατες μελέτες χρησιμοποιούν ένα πολυποίκιλο σύνολο χαρακτηριστικών γνωρισμάτων, βασισμένο στον τόνο, την ένταση, την διάρκεια, το φάσμα, τις μετρικές σταθερότητας και τις λεξικολογικές ιδιότητες. Εντούτοις, δεν υπάρχει κάποια τυποποίηση στο θέμα αυτό, με διαφορετικούς ερευνητές να πειραματίζονται με αρκετά διαφορετικά σύνολα χαρακτηριστικών γνωρισμάτων. Οι περισσότερες μελέτες περιλαμβάνουν τις βασικές στατιστικές του $F0$ και της καμπύλης έντασης όπως ανώτατο, κατώτατο, εύρος, απόκλιση [5] [12] [60], αν και ακόμη και εδώ οι λεπτομέρειες των σχημάτων κανονικοποίησης μπορούν να διαφέρουν. Μελέτες συχνά αποκλίνουν σε περισσότερο πολύπλοκα χαρακτηριστικά γνωρίσματα που λαμβάνονται από αυτές τις καμπύλες, υιοθετώντας διάφορες υψηλότερου βαθμού ροπές (higher order moments), προσαρμογή καμπυλών, κ.λπ. Μέχρι τώρα πολύ λίγες μελέτες έχουν χρησιμοποιήσει χαρακτηριστικά γνωρίσματα που λαμβάνονται από τα αντιληπτικά ή πρότυπα παραγωγής.

Συνολικά, δεν έχει δημοσιευτεί κάποια σύγκριση μεγάλης κλίμακας μεταξύ των διαφορετικών ομάδων χαρακτηριστικών γνωρισμάτων, αξιολογώντας την σχετική σημασία τους, εν τούτοις μερικές πρόσφατες εργασίες αμφισβήτησαν την σημαντικότητα των χαρακτηριστικών γνωρισμάτων βασισμένων στην $F0$ έναντι άλλων.

2.3.3.2.2 Εξαγωγή χαρακτηριστικών γνωρισμάτων Μια σημαντική διαφορά μεταξύ των οπτικών και ακουστικών μορφών πληροφορίας σχετίζεται με την διάρκεια της ακολουθίας που πρέπει να παρατηρήσουμε προκειμένου να είμαστε σε θέση να αποκτήσουμε μια κατανόηση του περιεχομένου της ακολουθίας. Στην περίπτωση του βίντεο, ένα πλαίσιο είναι συχνά αρκετό για να συμπεράνουμε το περιεχόμενο και σε κάθε περίπτωση ικανό να επεξεργαστεί και να εξαχθούν πληροφορίες σχετικά με αυτό. Από την άλλη, ένα ακουστικό σήμα πρέπει να έχει μια ελάχιστη διάρκεια ώστε οποιοδήποτε είδος επεξεργασίας να μπορεί να εφαρμοσθεί σε αυτό.

Επομένως, αντί της επεξεργασίας των στιγμιαίων τιμών του ακουστικού σήμα-

τος, όπως κάναμε με την οπτική μορφή πληροφορίας, πρέπει να επεξεργαστούμε τις ηχητικές καταγραφές κατά ομάδες. Προφανώς, ο καθορισμός αυτών των ομάδων θα διαδραματίσει σημαντικό ρόλο στην απόδοση του προκύπτοντος συστήματος. Εξετάζουμε ηχητικά δείγματα που ομαδοποιούνται σε ηχητικά τμήματα (τόνους), ακολουθίες που οριοθετούνται από μικρές παύσεις. Το σκεπτικό πίσω από αυτή την απόφαση είναι ότι αν και οι εκφράσεις μπορούν να αλλάξουν μέσα σε έναν ενιαίο τόνο, το υποκείμενο ανθρώπινο συναίσθημα δεν αλλάζει σημαντικά ώστε να μετατοπιστεί από το ένα τεταρτημόριο στο άλλο. Για τον λόγο αυτόν, ο τόνος δεν είναι μόνο η ακουστική μονάδα επάνω στην οποία εφαρμόζουμε τις τεχνικές εξαγωγής ακουστικών χαρακτηριστικών γνωρισμάτων αλλά και η μονάδα αναφοράς κατά τη διάρκεια της λειτουργίας του ευρύτερου συστήματος ταξινόμησης συναισθήματος.

Όπως προκύπτει από σχετικές εργασίες είναι αρκετά προφανές ότι η επιλογή του σωστού συνόλου ακουστικών χαρακτηριστικών γνωρισμάτων που εξετάζονται για την ταξινόμηση απέχει πολύ από μια τετριμμένη διαδικασία. Προκειμένου να ξεπερασθεί αυτό στην παρούσα εργασία, αρχικά εξάγεται ένα εκτεταμένο σύνολο από 377 ακουστικά χαρακτηριστικά γνωρισμάτα. Αυτό περιλαμβάνει χαρακτηριστικά γνωρίσματα βασισμένα στην ένταση, τον τόνο, MFCC (Mel Frequency Cepstral Coefficient), τις φασματικές ζώνες Bark, τα χαρακτηριστικά ακουστικού τμήματος και το μήκος παύσης.

Αναλύσαμε κάθε τόνο υιοθετώντας μια μέθοδο προσωδικής αναπαράστασης που βασίστηκε στην μέθοδο Prosogram [161]. Το Prosogram βασίζεται σε μια τυποποίηση των θεμελιωδών στοιχείων συχνότητας (καμπύλη) για φωνητικούς (ή συλλαβικούς) πυρήνες. Συνολικά δίνει για κάθε εκφρασμένο πυρήνα μια μέτρηση της έντασης και του μήκους. Σύμφωνα με ένα 'glissando threshold' σε μερικές περιπτώσεις δεν λαμβάνουμε μια σταθερή ένταση αλλά μια ή περισσότερες γραμμές που καθορίζουν την εξέλιξη της έντασης για αυτόν τον πυρήνα. Αυτή η αναπαράσταση είναι ένας τρόπος παρόμοιος με την αναπαράσταση 'ρόλων πιάνου' που χρησιμοποιείται σε μουσικές ακολουθίες. Αυτή η μέθοδος, υλοποιημένη στο περιβάλλον Praat, προσφέρει τη δυνατότητα της αυτόματης κατάτμησης βασισμένη και στα φωνητικά μέρη αλλά και στα ενεργειακά μέγιστα. Σύμφωνα με αυτό το πρότυπο τυποποίησης αναπαράστασης εξαγάγαμε διάφορους τύπους χαρακτηριστικών γνωρισμάτων βασισμένα στο διάστημα έντασης, μήκους πυρήνων και αποστάσεις μεταξύ των πυρήνων.

Λαμβάνοντας υπόψη ότι το μοντέλο κατηγοριοποίησης που χρησιμοποιείται σε αυτήν την εργασία, όπως θα δούμε σε παρακάτω ενότητα, είναι βασισμένο σε ένα νευρωνικό δίκτυο, παράγοντες που επηρεάζουν αρνητικά τον χρόνο εκπαίδευσης είναι το ευρύ φάσμα χαρακτηριστικών γνωρισμάτων και το μέγεθος του σχολιασμένου συνόλου δεδομένων. Προκειμένου να ξεπερασθεί αυτό είναι εμφανής η ανάγκη για στατιστική επεξεργασία των ακουστικών χαρακτηριστικών γνωρισμάτων, ώστε να βρεθούν τα πιο προεξέχοντα χαρακτηριστικά, μειώνοντας κατά συνέπεια τον αριθμό των χαρακτηριστικών γνωρισμάτων. Στην εργασία μας το επιτυγχάνουμε αυτό με τον συνδυασμό δύο ευρέως γνωστών τεχνικών: ανάλυση της διακύμανσης (ANOVA) και Pearson συντελεστής συσχετισμού (PMCC). Η πρώτη μέθοδος χρησιμοποιείται αρχικά ώστε να διερευνυθεί η διαχωριστική δυνατότητα κάθε χαρακτηριστικού γνωρίσματος. Καταλήγοντας σε ένα μειωμένο σύνολο χαρακτηριστικών γνωρισμάτων, που περιέχει περίπου τα μισά από τα χαρακτηριστικά γνωρίσματα. Για να μειωθεί περαιτέρω η διάσταση των χαρακτηριστικών γνωρισμάτων υπολογίσαμε τους PMCC για όλα τα υπόλοιπα ζευγάρια χαρακτηριστικών γνωρισμάτων. PMCC είναι ένα μέτρο της τάσης δύο μεταβλητών σχετικά με το ίδιο αντικείμενο να αυξάνονται ή να μειώ-

νονται από κοινού. Ομάδες χαρακτηριστικών γνωρισμάτων με αυξημένο συσχετισμό ($> 90\%$) διαμορφώθηκαν ως ένα ενιαίο χαρακτηριστικό γνώρισμα.

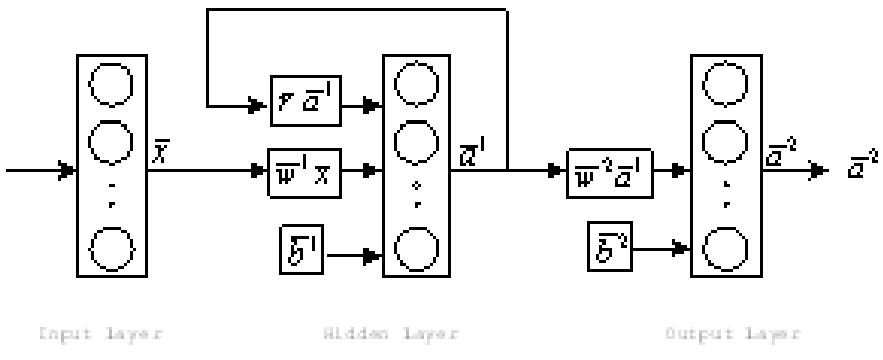
Η γενική διαδικασία οδηγεί στη τελική μείωση του αριθμού των ακουστικών χαρακτηριστικών γνωρισμάτων εξεταζόμενου κατά τη διάρκεια της ταξινόμησης από 377 σε 32. Τα επιλεγμένα χαρακτηριστικά γνωρίσματα παρατίθεται στον πίνακα 2.4. Όλα τα επιλεγμένα χαρακτηριστικά γνωρίσματα είναι αριθμητικά και συνεχή.

Πίνακας 2.4: Ακουστικά χαρακτηριστικά γνωρίσματα μετά την διαδικασία επιλογής

ptsegno	pfl2	ifqrange	ifmins
pfmean	pfmaxs	ifstart	ittmax
pfstd	pfmins	ifend	vfract
pfmax	pfmicro1	ifl1	vshimapq3
pfrange	pfmicro2	ifl2	vnhr
pfqrange	ifmean	ifpar3	vhnr
pfstart	ifstd	ifdct2	ltasslp
pfend	ifmax	ifmaxs	ltasfmax

2.3.4 Πολυμεσική αναγνώριση έκφρασης

Προκειμένου να εξεταστεί η δυναμική των εκφράσεων είναι απαραίτητη η χρήση ένος σχήματος ταξινόμησης που είναι σε θέση να προτυποποιήσει και να μάθει τη δυναμική, όπως είναι ένα κρυφό μαρκοβιανό πρότυπο ή ένα επαναληπτικό νευρωνικό δίκτυο. Σε αυτήν την εργασία χρησιμοποιήθηκε ένα επαναληπτικό νευρωνικό δίκτυο (εικόνα 2.21). Αυτός ο τύπος δικτύου διαφέρει από τα συμβατικά feed-forward δίκτυα στο ότι το πρώτο επίπεδο έχει μια επαναληπτική σύνδεση. Η καθυστέρηση σε αυτή την σύνδεση αποθηκεύει τις τιμές από το προηγούμενο χρονικό βήμα ώστε να μπορεί χρησιμοποιηθεί στο τρέχον χρονικό βήμα, παρέχοντας κατά συνέπεια το στοιχείο της μνήμης.



Σχήμα 2.21: Το επαναληπτικό νευρωνικό δίκτυο

Από όλες τις πιθανές υλοποιήσεις επαναληπτικών δικτύων έχουμε επιλέξει το Elman δίκτυο για την εργασία μας [79] [80]. Αυτό είναι ένα δίκτυο με δύο επίπεδα με την ανατροφοδότηση από την έξοδο του πρώτου επιπέδου στην είσοδο του πρώτου κρυμμένου. Αυτή η αναδρομική σύνδεση επιτρέπει στο δίκτυο Elman να κατασκευάσει και να αναγνωρίσει πρότυπα όπου η χρονική εξέλιξη των εισόδων μεταφέρουν

κρίσιμη πληροφορία. Αν και ακολουθούμε μια προσέγγιση που περιλαμβάνει μόνο ένα επίπεδο που περιέχει αναδρομικές συνδέσεις, στην πραγματικότητα το δίκτυο έχει τη δυνατότητα να εκπαιδευτεί σε πρότυπα μεγαλύτερου μήκους, αφού οι τρέχουσες τιμές επηρεάζονται και από προηγούμενες τιμές και όχι μόνο από το τελευταίο στιγμιότυπο της εισόδου. Οι συναρτήσεις μεταφοράς των νευρώνων που χρησιμοποιούνται στο Elman δίκτυο είναι tan-sigmoid για το κρυμμένο (αναδρομικό) επίπεδο και αμιγώς γραμμικές για το επίπεδο εξόδου. Πιο συγκεκριμένα:

$$a_i^1 = \tan sig(k_i^1) = \frac{2}{1 + \mathrm{e}^{\text{-}2k_i^1}}\text{-}1$$

$$a_j^2 = k_j^2$$

όπου a_i^1είναι η ενεργοποίηση του νευρώνα i στο πρώτο (κρυμμένο) επίπεδο, k_i^1 είναι το επαγωγικό τοπικό πεδίο ή δυνατότητα ενεργοποίησης του νευρώνα στο πρώτο επίπεδο, a_i^2 είναι η ενεργοποίηση του j νευρώνα στο δεύτερο επίπεδο (εξόδου) και k_i^2 είναι το επαγωγικό τοπικό πεδίο ή δυνατότητα ενεργοποίησης του νευρώνα j στο δεύτερο επίπεδο.

Το επαγωγικό τοπικό πεδίο στο πρώτο επίπεδο υπολογίζεται ως:

$$k_i^1 = \bar{w}_i^1 \cdot \bar{x} + \bar{r}_i \cdot \bar{a}^1 + b_i^1$$

όπου $\bar{x}$ είναι το διάνυσμα εισόδου, $\bar{w}_i^1$ είναι το διάνυσμα βαρών εισόδου για τον i νευρώνα, $\bar{a}^1$ είναι το διάνυσμα εξόδου του πρώτου επιπέδου για το προηγούμενο χρονικό βήμα, $\bar{r}_i$ είναι το αναδρομικό διάνυσμα βαρών και b_i^1 είναι το διάνυσμα πόλωσης (bias). Το τοπικό πεδίο στο δεύτερο επίπεδο υπολογίζεται με τον συμβατικό τρόπο ως εξής:

$$k_j^2 = \bar{w}_j^2 \cdot \bar{a}^1 + b_j^2$$

όπου $\bar{w}_i^2$ είναι το βάρος εισόδου και b_j^2 είναι η πόλωση.

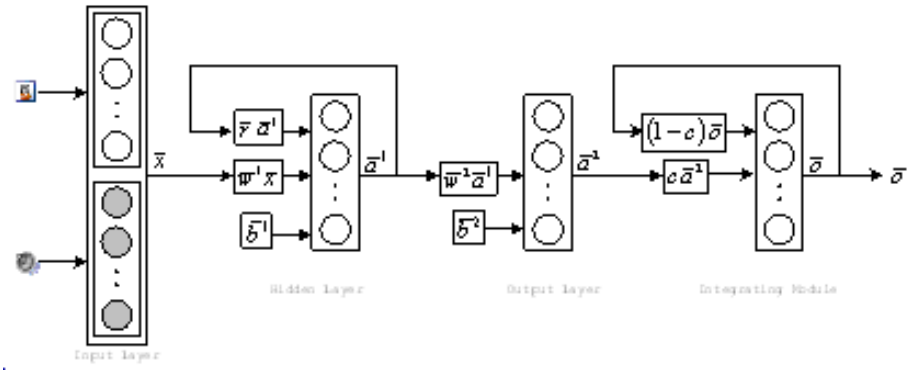
Αυτός ο συνδυασμός συναρτήσεων ενεργοποίησης, με αυτή την αρχιτεκτονική και με αυτές τις συναρτήσεις μεταφοράς μπορούν να προσεγγίσουν οποιαδήποτε συνάρτηση (με πεπερασμένο αριθμό ασυνεχειών) με αυθαίρετη ακρίβεια. Η μόνη απαίτηση είναι το κρυμμένο επίπεδο να έχει αρκετούς νευρώνες ([207] και [103]). Όσον αφορά στην εκπαίδευση, χρησιμοποιείται αλγόριθμος χρονικά περικομμένου back-propagation (περικομμένο BPTT) [106].

Το επίπεδο εισόδου του χρησιμοποιημένου δικτύου έχει 49 νευρώνες (17 για τα FAPs και 32 για τα ακουστικά χαρακτηριστικά γνωρίσματα). Το κρυμμένο επίπεδο έχει 20 νευρώνες και αυτό της εξόδου έχει 5 νευρώνες, έναν για καθεμία από τις πέντε πιθανές κατηγοριές: Ουδέτερο, Q1 (πρώτο τεταρτημόριο του Feeltrace [48]), Q2, Q3 και Q4. Το δίκτυο εκπαιδεύεται για να παράγει τιμή 1 (on) στην έξοδο που αντιστοιχεί στο επισημειωμένο τεταρτημόριο του τόνου και τιμή 0 για τα άλλα αποτελέσματα.

Κατά την λειτουργία του δικτύου παρέχουμε ως εισόδους τις τιμές από τα εξεταζόμενα χαρακτηριστικά γνωρίσματα για κάθε πλαίσιο. Καθώς εισάγεται κάθε πλαίσιο το δίκτυο εκπαιδεύεται στην δυναμική που περιγράφεται από τον τρόπο που τα χαρακτηριστικά γνωρίσματα αλλάζουν και καταφέρνει έτσι να παρέχει μια σωστή κατηγοριοποίηση στην έξοδο του.

Ένα ζήτημα που χρίζει εξέτασης, εν τούτοις, είναι ότι δεν εξελίσσονται όλες οι εξεταζόμενες εισόδους δυναμικά με τον χρόνο. Συγκεκριμένα, όπως έχουμε ήδη αναφέρει, όσον αφορά στην ακουστική μορφή πληροφορίας ο τόνος θεωρείται και υποβάλλεται προς επεξεργασία ως ενιαία μονάδα. Κατά συνέπεια, οι τιμές χαρακτηριστικών

γνωρισμάτων αναφέρονται στο σύνολο του τόνου και δεν είναι διαθέσιμες ανά πλαίσιο. Κατά συνέπεια, ένα επαναληπτικό νευρωνικό δίκτυο δεν μπορεί να επεξεργαστεί με την αρχική της μορφή τα στοιχεία της ακουστικής μορφής πληροφορίας.



Σχήμα 2.22: Το τροποποιημένο δίκτυο Elman με ολοκληρωτή εξόδου

Προκειμένου να αντιμετωπισθεί η ιδιαιτερότητα αυτή των ακουστικών δεδομένων, τροποποιούμε την απλή δομή του δικτύου της αρχιτεκτονικής που φαίνεται στην εικόνα 2.21 όπως φαίνεται στην εικόνα 2.22. Σε αυτή την τροποποιημένη έκδοση χρησιμοποιούνται δύο διαφορετικοί τύποι κόμβων εισόδου:

1. για τα οπτικά χαρακτηριστικά γνωρίσματα μορφής διατηρούμε τους συμβατικούς νευρώνες εισαγωγής που συναντιούνται σε όλα τα νευρωνικά δίκτυα

2. για τα ακουστικά χαρακτηριστικά γνωρίσματα μορφής χρησιμοποιούμε νευρώνες στατικής τιμής. Αυτοί διατηρούν την ίδια τιμή σε όλη τη λειτουργία του νευρικού δικτύου.

Οι τιμές για τα ακουστικά χαρακτηριστικά γνωρίσματα που έχουν υπολογιστεί για έναν τόνο εισάγονται στο δίκτυο ως τιμές που αντιστοιχούν στο πρώτο πλαίσιο. Στα επόμενα χρονικά βήματα, ενώ τα οπτικά χαρακτηριστικά γνωρίσματα που αντιστοιχούν στα επόμενα πλαίσια εισάγονται στους πρώτους νευρώνες εισόδου του δικτύου, οι στατικοί νευρώνες εισόδου διατηρούν τις αρχικές τιμές για την ακουστική μορφή πληροφορίας, επιτρέποντας κατά συνέπεια στο δίκτυο να λειτουργήσει κανονικά.

Κάποιος μπορεί εύκολα να παρατηρήσει ότι αν και το δίκτυο έχει τη δυνατότητα να λάβει την δυναμική που υπάρχει στην είσοδο του, δεν μπορεί να εκπαιδευτει στην αναγνώριση της δυναμικής στην ακουστική μορφή δεδομένου ότι εκπαιδεύεται μόνο με στατικές τιμές. Ακόμα, πρέπει να σχολιάσουμε ότι η δυναμική αυτής της μορφής πληροφορίας δεν αγνοείται. Το αντίθετο, οι στατικές τιμές χαρακτηριστικών γνωρισμάτων που υπολογίζονται για αυτήν την μορφή, όπως έχει εξηγηθεί, βασίζονται, και υπό μια έννοια εμπεριέχουν, τη δυναμική του ακουστικού καναλιού της καταγραφής.

Οι πιό κοινές εφαρμογές των επαναληπτικών νευρωνικών δικτύων περιλαμβάνουν σύνθετα προβλήματα όπως η προτυποποίηση, η προσέγγιση, η παραγωγή και η πρόβλεψη δυναμικών ακολουθιών γνωστών ή άγνωστων στατιστικών χαρακτηριστικών. Σε αντίθεση με απλούστερες δομές δικτύων, η χρήση τους για φαινομενικά ευκολότερες εργασίες ταξινόμησης δεν είναι εξίσου απλή ή προφανής. Ο λόγος είναι ότι ενώ τα απλά νευρωνικά δίκτυα δίνουν μια απάντηση της μορφής ενός διανύσματος τιμών στην έξοδο τους αφού εξετάσουν την είσοδο, τα επαναληπτικά νευρωνικά δίκτυα παρέχουν τέτοιες τιμές μετά από κάθε διαφορετικό χρονικό στάδιο. Βέβαια, το θέμα

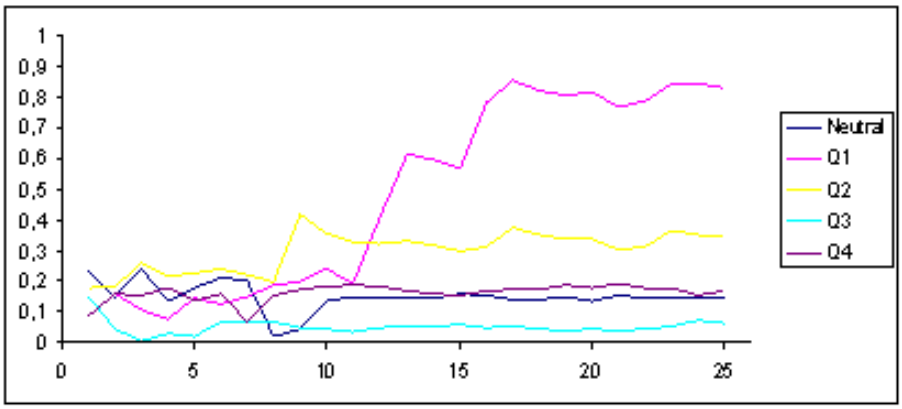
που προκύπτει είναι σε ποιο βήμα είναι βέλτιστο να γίνει η δειγματοληψία της έξοδος του δικτύου για να επιτευχθεί βέλτιστη απόφαση ταξινόμησης.

Κατά γενικό κανόνα, τα πρώτα αποτελέσματα ενός επαναληπτικού νευρωνικού δικτύου δεν είναι πολύ αξιόπιστα. Ο λόγος είναι ότι ένα επαναληπτικό νευρωνικό δίκτυο εκπαιδεύεται για να προσομοιώσει τη δυναμική που υπάρχει στις ακολουθίες δεδομένων και επομένως πρέπει να παρουσιαστεί σε αυτό μια ακολουθία δεδομένων ικανοποιητικού μήκους προκειμένου να είναι ικανό να ανιχνεύσει και να ταξινομήσει την δυναμική αυτή. Αφ' ετέρου, δεν είναι πάντα ασφαλές να λάβουμε υπ'όφη την έξοδο στο τελικό χρονικό βήμα ως αποτέλεσμα ταξινόμησης του δικτύου επειδή:

1. η διάρκεια των δεδομένων εισόδου μπορεί να είναι μερικά χρονικά βήματα πιό μακροχρόνια από τη διάρκεια της κυρίαρχης δυναμικής συμπεριφοράς και έτσι η λειτουργία του δικτύου κατά τη διάρκεια των τελευταίων βημάτων μπορεί να είναι τυχαία

2. ένα προσωρινό λάθος μπορεί να εμφανιστεί σε οποιοδήποτε χρονικό βήμα της λειτουργίας του δικτύου

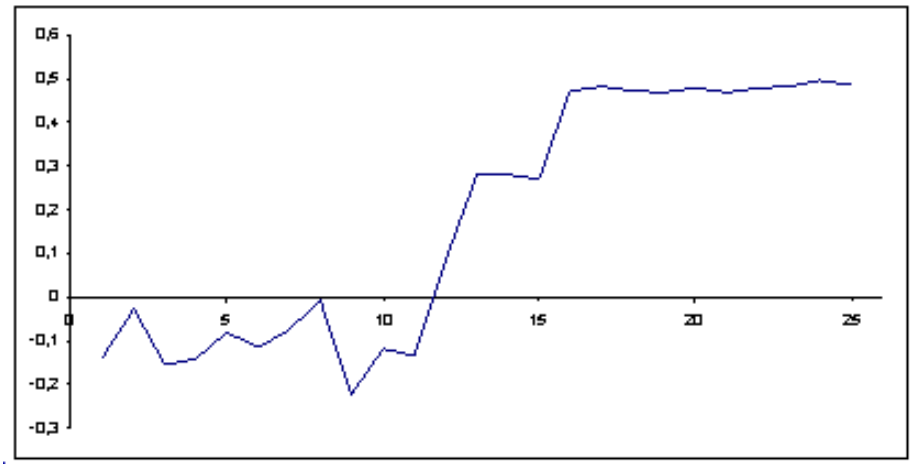
Παραδείγματος χάριν, στην εικόνα 2.23 παρουσιάζουμε την έξοδο του δικτύου μετά από κάθε πλαίσιο κατά την επεξεργασία του τόνου στο τρέχον παράδειγμα. Μπορούμε να δούμε ότι κατά τη διάρκεια των πρώτων πλαισίων η έξοδος του δικτύου είναι αρκετά τυχαία και αλλάζει με ταχύ ρυθμό. Όταν αρκετά ικανοποιητικό μήκος της ακολουθίας έχει παρουσιασθεί στο δίκτυο, έτσι ώστε η δυναμική να μπορεί να αναπαρασταθεί, τα αποτελέσματα αρχίζουν να συγκλίνουν στις τελικές τιμές τους. Αλλά ακόμα και τότε μικρές αλλαγές στα επίπεδα εξόδου μπορούν να παρατηρηθούν μεταξύ διαδοχικών πλαισίων.



Σχήμα 2.23: Μεμονωμένες έξοδοι του δικτύου μετά από κάθε πλαίσιο

Αν και για αυτό το παράδειγμα όπου η ταξινόμηση στην κλάση Q1 είναι σαφής, αυτές οι αστάθειες δεν είναι αρκετές ώστε να αλλάξουν την απόφαση ταξινόμησης (βλ. εικόνα 2.24) υπάρχουν περιπτώσεις στις οποίες το περιθώριο ταξινόμησης είναι μικρότερο και ακόμα και αυτές οι μικρές διαφοροποιήσεις οδηγούν σε προσωρινή αλλαγή απόφασης ταξινόμησης.

Προκειμένου να θωρακίσουμε με ευρωστία το σχήμα ταξινόμησής μας προσθέσαμε έναν μηχανισμό σταθμισμού στην ενότητα εξόδου του νευρωνικού δικτύου που αυξάνει τη ευστάθεια του (βλ. εικόνα 2.22). Συγκεκριμένα, τα τελικά αποτελέσματα της αρχιτεκτονικής υπολογίζονται όπως παρακάτω:



Σχήμα 2.24: Διαφορά μεταξύ επιθυμητής και μη επιθυμητής εξόδου με την μεγαλύτερη τιμή

$$o_j(t) = c \cdot a_j^2 + (1-c) \cdot o_j(t-1)$$

όπου $o_j(t)$ είναι η τιμή που υπολογίζεται για την έξοδο j κατά το χρονικό βήμα t, $o_j(t-1)$ είναι η τιμή εξόδου που υπολογίζεται κατά το προηγούμενο χρονικό βήμα και c είναι μια παράμετρος με πεδίο τιμών (0,1] που ελέγχει την ευαισθησία/σταθερότητα του σχήματος ταξινόμησης. Όταν το c είναι πιό κοντά στο μηδέν το σύστημα γίνεται πολύ σταθερό και μια ακολουθία μεγάλου μήκους με διαφορετικές τιμές k_j^2 απαιτείται για να επηρεαστούν τα αποτελέσματα ταξινόμησης ενώ όταν το c προσεγγίζει την μονάδα το σύστημα γίνεται πιό ευαίσθητο στις αλλαγές στην έξοδο του δικτύου. Ενώ όταν $c = 1$ η υποενότητα ενσωμάτωσης απενεργοποιείται και η έξοδος του δικτύου παράγεται ως γενικό αποτέλεσμα ταξινόμησης. Στην εργασία μας, κατόπιν πειραματισμού για διαφορετικές τιμές c, έχουμε επιλέξει $c = 0.5$.

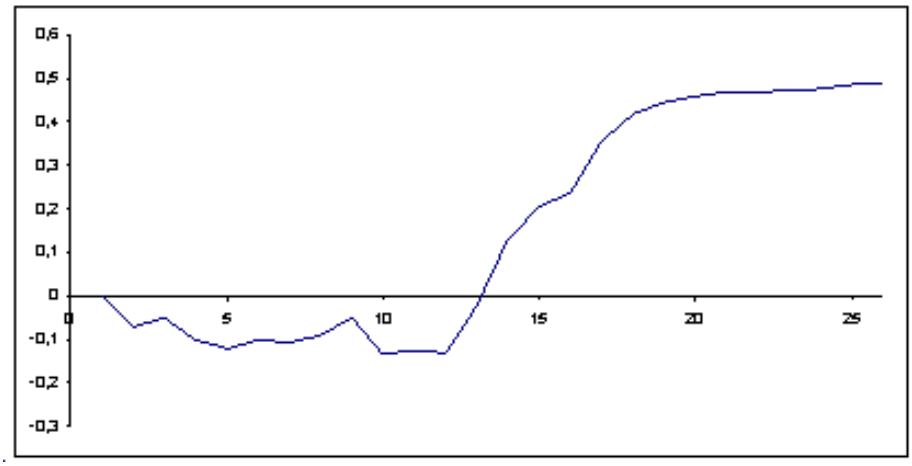
Στην εικόνα 2.25 μπορούμε να δούμε το περιθώριο απόφασης με την χρησιμοποίηση της ενότητας ολοκλήρωσης στάθμισης στην έξοδο του δικτύου. Όταν συγκριθεί με την εικόνα 2.24 μπορούμε σαφώς να δούμε ότι η εξέλιξη του περιθωρίου είναι ομαλότερη, το οποίο δείχνει ότι έχουμε πράγματι πετύχει απόδοση ταξινόμησης του δικτύου περισσότερο σταθερή και λιγότερο εξαρτώμενη από το πλαίσιο που επιλέγεται ως τέλος ενός τόνου.

Φυσικά, για να λειτουργήσει αυτός ο σταθμισμένος ολοκληρωτής, χρειάζεται να καθοριστούν οι τιμές εξόδου του δικτύου για το χρονικό βήμα 0, δηλ. πριν από το πρώτο πλαίσιο. Εύκολα φαίνεται ότι λόγω του τρόπου που επιδρούν τα προηγούμενα αποτελέσματα με την εξέλιξη του χρόνου λόγω της παραμέτρου c, αυτή η αρχικοποίηση είναι σχεδόν αδιάφορη για τόνους επαρκούς μήκους. Από την άλλη, αυτή η αρχικοποίηση μπορεί να έχει επηρεάσει σημαντικά τους τόνους που είναι πολύ σύντομοι. Σε αυτή την εργασία, έχουμε επιλέξει να αρχικοποίησουμε τις εξόδους:

$$\bar{o}(0) = 0$$

Εναλλακτικά μπορούν να αρχικοποιηθούν τα $\bar{o}(0)$ βασιζόμενοι στα ποσοστά των διαφορετικών κατηγοριών εξόδου στα επισημειωμένα στοιχεία αναφοράς που χρησιμοποιούνται για να εκπαιδεύσουν τον ταξινομητή. Αποφύγαμε αυτή την λύση για να μην προστεθεί μια προκατάληψη (bias) των αποτελεσμάτων προς οποιαδήποτε κατηγορία, δεδομένου ότι θέλαμε να είμαστε βέβαιοι ότι η απόδοση αναγνώρισης κατά τη

διάρκεια της δοκιμής οφείλεται μόνο στην δυναμική και πολυμορφική προσέγγιση που προτείνεται σε αυτήν την εργασία.



Σχήμα 2.25: Περιθώριο απόφασης με την χρήση της υποενότητας ενσωμάτωσης

Είναι άξιο αναφοράς ότι από άποψη μοντελοποίησης θα ήταν εφικτό να περιλάβουμε τον ολοκληρωτή στη δομή του δικτύου παρά να τον ενσωματώσουμε ως εξωτερική ενότητα, απλά με την προσθήκη ενός επαναληπτικού βρόχου στο επίπεδο εξόδου. Αποφασίσαμε να το αποφύγουμε αυτό, για να μη επηρεάσουμε την συμπεριφορά του δικτύου κατά την διάρκεια της εκπαίδευσης, αφού ο πρόσθετος επαναληπτικός βρόχος θα αύξανε σημαντικά τον χρόνο εκπαίδευσης και το μέγεθος και το μέσο μήκος των απαιτούμενων δεδομένων εκπαίδευσης.

2.3.5 Πειραματικά αποτελέσματα

2.3.5.1 Σημεία αναφοράς

Δεδομένου ότι ο στόχος αυτής της εργασίας είναι να αποδείξει την δυνατότητα ταξινόμησης ακολουθιών με φυσικές εκφράσεις, έχουμε επιλέξει να χρησιμοποιήσουμε τη SAL (Sensitive Artificial Listener) βάση δεδομένων [114]. Οι καταγραφές βασίστηκαν στην έννοια του ευαίσθητου τεχνητού ακροατή, όπου το SAL μιμείται την συμπεριφορά του ατόμου που κατευθύνει την αλληλεπίδραση και του καλού ακροατή, δηλ. εμπλέκει ένα πρόθυμο πρόσωπο σε συναισθηματικά εμπλουτισμένη αλληλεπίδραση. Αν και ο τελικός στόχος είναι να εκτιμήσει το SAL αυτόματα το περιεχόμενο και την μέθοδο, από την πλευρά του συναισθηματικού εμπλουτισμού, της αλληλεπίδρασης, κάτι τέτοιο δεν ήταν πλήρως υλοποιημένο κατά την διάρκεια της δημιουργίας της βάσης δεδομένων SAL και έτσι χρησιμοποιήθηκε μια προσέγγιση τύπου Wizard of Oz για την διαδικασία ερωταποκρίσεων του SAL [131].

Ένα σημείο που διαδραματίζει σημαντικό ρόλο στην συναισθηματική κατάσταση του ανθρώπου κατά την φυσική ανθρώπινη αλληλεπίδραση είναι ο χαρακτήρας, οι μακροπρόθεσμες συναισθηματικές συνήθειες και η βραχυπρόθεσμη διάθεση κάθε ατόμου. Διαφορετικά άτομα μπορεί να έχουν διαφορετικές συναισθηματικές αντιδράσεις σε παρόμοια ερεθίσματα. Συνεπώς, ο σχολιασμός και η επισημείωση των καταγραφών δεν πρέπει να βασιστεί στο προτιθέμενο συναίσθημα αλλά στο πραγματικό, τελικό αποτέλεσμα της αλληλεπίδρασης με το SAL. Λαμβάνοντας αυτό υπόψη, χρησιμοποιήθηκε το εργαλείο συναισθηματικής επισημείωσης FeelTrace για τον σχολιασμό των καταγραφών του SAL [48]. Αυτό είναι ένα εργαλείο επισημείωσης που έχει αναπτυχθεί

στο Queen's University Belfast χρησιμοποιώντας διαστατική αναπαράσταση συναισθήματος και παρέχει χρονικά εξαρτώμενες διαστατικές αναπαραστάσεις. Επιτρέπει στους ειδικούς να επισημάνουν το συναισθηματικό περιεχόμενο ενός χρονικά μεταβαλλόμενου ερεθίσματος όπως το αντιλαμβάνονται. Η εικόνα 2.26, δείχνει το γραφικό περιβάλλον του FeelTrace.

Σχήμα 2.26: Η διαστατική αναπαράσταση του FeelTrace [246]

Ο συναισθηματικός χώρος αναπαρίσταται από έναν κύκλο στην οθόνη, που χωρίζεται σε τέσσερα τεταρτημόρια από τους δύο κύριους άξονες. Ο κάθετος άξονας αντιπροσωπεύει την ενεργοποίηση και λαμβάνει τιμές από πολύ ενεργό σε πολύ παθητικό και τον οριζόντιο άξονα που αντιπροσωπεύει την αξιολόγηση, που λαμβάνει τιμές από πολύ θετικό σε πολύ αρνητικό. Απεικονίζει την δημοφιλή άποψη ότι ο συναισθηματικός χώρος είναι κατά προσέγγιση κυκλικός. Το κέντρο του κύκλου χαρακτηρίζει κάποιο ουδέτερο συναίσθημα και τοποθετώντας τον δείκτη σε αυτή την περιοχή ο χρήστης υπονοεί ότι δεν εκφράζεται κάποιο ιδιαίτερο συναίσθημα την συγκεκριμένη χρονική στιγμή ή ότι το συναίσθημα που εκφράζεται είναι χαμηλής έντασης και συναισθηματικής χροιάς. Ο χρήστης χρησιμοποιεί το ποντίκι και διατρέχει το συναισθηματικό διάστημα, έτσι ώστε κάθε στιγμή η θέση του να επισημαίνει τα επίπεδα ενεργοποίησης και αξιολόγησης που αντιλαμβάνεται, με το σύστημα να καταγράφει αυτόματα τις συντεταγμένες κάθε στιγμή.

Οι συντεταγμένες της επισημείωσης στον διδιάστατο χώρο αντιστοιχείται σε πέντε συναισθηματικές κατηγορίες που παρουσιάζονται στον πίνακα 2.23. Εφαρμόζοντας μια τυπική μέθοδο ανίχνευσης μικρής διακοπής στο ακουστικό κανάλι των υπό εξέταση καταγραφών, η βάση δεδομένων έχει χωριστεί σε 477 τόνους, με το μήκος κάθε τμήματος να κυμαίνεται από 1 μέχρι 174 πλαίσια. Το Q1 είναι η κυρίαρχη κατάσταση που υπάρχει στη βάση δεδομένων, αφού 42.98% των τμημάτων ταξινομούνται ως Q1, όπως φαίνεται στον πίνακα 2.5.

Πίνακας 2.5: Κατανομή κλάσεων στην βάση δεδομένων SAL

	Ουδέτερο	Q1	Q2	Q3	Q4	Σύνολο
Τόνοι	47	205	90	63	72	477
Ποσοστά	9,85%	42,98%	18,87%	13,21%	15,09%	100,00%

2.3.5.2 Στατιστικά αποτελέσματα

Από την εφαρμογή της προτεινόμενης μεθοδολογίας, χρησιμοποιώντας το σύνολο των επισημειωμένων δεδομένων ως αλήθεια εδάφους, καταλήγουμε σε μια μέτρηση του ποσοστού αναγνώρισης του συστήματος που φτάνει σε ποσοστό 81,55%. Συγκεκριμένα, 389 τόνοι κατηγοριοποιήθηκαν σωστά, ενώ 88 λάθος. Σαφώς, αυτό το είδος πληροφοριών, αν και ενδεικτικό, είναι ανεπαρκής για να κατανοηθεί πλήρως και να αξιολογηθεί η απόδοση της μεθοδολογίας μας.

Κινούμενοι προς αυτή την κατεύθυνση, στον πίνακα 2.6 φαίνεται ο πίνακας σύγχυσης για το πείραμα. Στον πίνακα, οι σειρές αντιστοιχούν στην κλάση αναφοράς και οι στήλες στην απάντηση του συστήματος. Κατά συνέπεια, παραδείγματος χάριν, υπήρχαν 5 τόνοι που χαρακτηρίστηκαν ως ουδέτεροι στην επισημείωση αναφοράς αλλά κατηγοριοποιήθηκαν λανθασμένα ως Q2 από το σύστημά μας.

Πίνακας 2.6: Συνολικός πίνακας σύγχυσης

	Ουδέτερο	Q1	Q2	Q3	Q4	Σύνολο
Ουδέτερο	**34**	1	5	3	0	43
Q1	1	**189**	9	12	6	217
Q2	4	3	**65**	2	1	75
Q3	4	6	7	**39**	3	59
Q4	4	6	4	7	**62**	83
Σύνολο	47	205	90	63	72	477

Πίνακας 2.7: Συνολικός ποσοστιαίος πίνακας σύγχυσης

	Ουδέτερο	Q1	Q2	Q3	Q4	Σύνολο
Ουδέτερο	**79,07%**	2,33%	11,63%	6,98%	0,00%	100,00%
Q1	0,46%	**87,10%**	4,15%	5,53%	2,76%	100,00%
Q2	5,33%	4,00%	**86,67%**	2,67%	1,33%	100,00%
Q3	6,78%	10,17%	11,86%	**66,10%**	5,08%	100,00%
Q4	4,82%	7,23%	4,82%	8,43%	**74,70%**	100,00%
Σύνολο	9,85%	42,98%	18,87%	13,21%	15,09%	100,00%

Λαμβάνοντας υπόψη το γεγονός ότι το σημείο αναφοράς μας είναι πολωμένο προς το τεταρτημόριο Q1, παρέχουμε επίσης στον πίνακα 2.7 τον πίνακα σύγχυσης υπό μορφή ποσοστών έτσι ώστε να αντιμετωπιστεί αυτή η προκατάληψη. Εκεί μπορούμε να δούμε ότι η προτεινόμενη μεθοδολογία αποδίδει εύλογα καλά για τις περισσότερες περιπτώσεις, με εξαίρεση το Q3, για το οποίο το ποσοστό αναγνώρισης είναι αρκετά χαμηλό. Αυτό που αξίζει περαιτέρω διερεύνησης είναι ότι περισσότερο από 10% των τόνων Q3 ταξινομήθηκαν στο ακριβώς διαμετρικό τεταρτημόριο, το οποίο είναι βεβαίως ένα σημαντικό λάθος.

Ακόμα, στην ανάλυσή των πειραματικών αποτελεσμάτων μέχρι τώρα δεν έχουμε λάβει υπόψη έναν πολύ σημαντικό παράγοντα: αυτό του μήκους των τόνων. Όπως συζητήθηκε παραπάνω, η αρχιτεκτονική δικτύου Elman απαιτεί επαρκή αριθμό πλαισίων ως είσοδο ώστε να αναγνωρίσει την δυναμική. Επιπλέον, υπάρχουν αρκετοί τόνοι στα δεδομένα αναφοράς που είναι πάρα πολύ μικρά σε μήκος μη επιτρέποντας

στο δίκτυο να φθάσει σε σημείο όπου η έξοδος του να μπορεί να διαβαστεί με υψηλή εμπιστοσύνη.

Προκειμένου αυτό να γίνει πιο εμφανές παρουσιάζουμε ξεχωριστούς πίνακες σύγχυσης για μικρούς τόνους (πίνακες 2.10 και 2.11) και για κανονικού μήκους (πίνακες 2.8 και 2.9). Σε αυτό το πλαίσιο θεωρούμε ως κανονικούς τόνους αυτούς που περιλαμβάνουν τουλάχιστον 10 πλαίσια και ως σύντομους αυτούς με μήκος από 1 μέχρι 9 πλαίσια.

Καταρχήν, μπορούμε να δούμε ξεκάθαρα ότι η απόδοση του συστήματος, όπως αναμενόταν είναι αρκετά διαφορετική στις δύο περιπτώσεις. Συγκεκριμένα, υπάρχουν 83 λάθη σε 131 σύντομους τόνους ενώ υπάρχουν μόνο 5 λάθη σε 346 κανονικούς τόνους. Επιπλέον, δεν υπάρχει κάποιο σοβαρό λάθος στην περίπτωση μεγάλων τόνων, δηλ. δεν υπάρχει καμία περίπτωση στην οποία ένας τόνος ταξινομείται στο ακριβώς αντιδιαμετρικό τεταρτημόριο απ'ό,τι η κλάση αναφοράς.

Συνολικά, η λειτουργία του συστήματός μας υπό κανονικές συνθήκες (τέτοιες θεωρούμε τις περιπτώσεις όπου ο τόνος έχει μήκος τουλάχιστον 10 πλαισίων) επιτυγχάνει ένα ποσοστό ταξινόμησης 98,55%, το οποίο είναι αρκετά ενθαρρυντικό, ακόμη και για δεδομένα που καταγράφονται υπό ελεγχόμενες συνθήκες, πόσο μάλλον για φυσιοκρατικές καταγραφές.

Πίνακας 2.8: Πίνακας σύγχυσης για κανονικούς τόνους

	Ουδέτερο	Q1	Q2	Q3	Q4	Σύνολο
Ουδέτερο	**29**	0	0	0	0	29
Q1	0	**172**	3	0	0	175
Q2	1	1	**54**	0	0	56
Q3	0	0	0	**30**	0	30
Q4	0	0	0	0	**56**	56
Σύνολο	30	173	57	30	56	346

Πίνακας 2.9: Ποσοστιαίος πίνακας σύγχυσης για κανονικούς τόνους

	Ουδέτερο	Q1	Q2	Q3	Q4	Σύνολο
Ουδέτερο	**100,00%**	0,00%	0,00%	0,00%	0,00%	100,00%
Q1	0,00%	**98,29%**	1,71%	0,00%	0,00%	100,00%
Q2	1,79%	1,79%	**96,43%**	0,00%	0,00%	100,00%
Q3	0,00%	0,00%	0,00%	**100,00%**	0,00%	100,00%
Q4	0,00%	0,00%	0,00%	0,00%	**100,00%**	100,00%
Σύνολο	8,67%	50,00%	16,47%	8,67%	16,18%	100,00%

Φυσικά, παραμένει το ερώτημα εάν είναι φυσιολογικό για ένα σύστημα να αποτυγχάνει τόσο πολύ στην περίπτωση των σύντομων τόνων, ή εάν οι πληροφορίες που περιλαμβάνονται σε αυτά είναι ικανοποιητικά για μια αρκετά καλύτερη απόδοση και το σύστημα πρέπει να βελτιωθεί. Προκειμένου να απαντηθεί αυτή η ερώτηση χρησιμοποιήσαμε τον δείκτη Williams [248]. Αυτός ο δείκτης σχεδιάστηκε αρχικά για να μετρήσει την κοινή συμφωνία αρκετών ειδικών με έναν άλλο ειδικό. Συγκεκριμένα, ο δείκτης στοχεύει να απαντήσει στην ερώτηση: 'λαμβάνοντας υπόψη ένα σύνολο ειδικών και ενός άλλου ειδικού, συμφωνεί ο μεμονωμένος ειδικός με το σύνολο των

ειδικών τόσο συχνά όσο ένα μέλος του συνόλου συμφωνεί με ένα άλλο μέλος του;'. Αυτή είναι μια άλλη διατύπωση του ερωτήματος που θέσαμε παραπάνω. Στο πλαίσιό των πειραμάτων μας, ο απομονωμένος ειδικός είναι το προτεινόμενο σύστημα. Προκειμένου να έχει επαρκή αριθμό ειδικών το σύνολο ειδικών για την εφαρμογή της μεθοδολογίας του δείκτη Williams ζητήσαμε τρεις επιπλέον ανθρώπους να ταξινομήσουν τους 131 σύντομους τόνους σε μια από τις πέντε συναισθηματικές κατηγορίες. Στην περίπτωσή μας, ο δείκτης Williams για το σύστημα σε σχέση με τους τέσσερις ανθρώπινους σχολιαστές τροποποιείται ως εξής:

Η συνδυασμένη συμφωνία μεταξύ των ειδικών του συνόλου ορίζεται ως:

$$P_g = \frac{2}{4\left(4-1\right)} \sum_{a=1}^{3} \sum_{b=a+1}^{4} P\left(a,b\right)$$

όπου $P\left(a,b\right)$ είναι ο λόγος των συμφωνιών μεταξύ των ειδικών a και b

$$P\left(a,b\right) = \frac{\left|\left\{s \in S : R_a\left(s\right) = R_b\left(s\right)\right\}\right|}{131}$$

Στην παραπάνω εξίσωση S είναι το σύνολο των 131 επισημειωμένων τόνων και $R_a\left(s\right)$ είναι η κατηγορία που επισημείωσε ο ειδικός a για τον τόνο s. Η παρατηρούμενη ευρύτερη συμφωνία του συνόλου αναφοράς με το προτεινόμενο σύστημα μετριέται ως:

$$P_0 = \frac{\sum_{a=1}^{4} P\left(0,a\right)}{4}$$

χρησιμοποιούμε το 0 για να διακρίνουμε τον υπό εξέταση ειδικό, το σύστημα μας. Ο δείκτης Williams για το σύστημα μας είναι ο λόγος:

$$I_0 = \frac{P_0}{P_g}$$

Η τιμή I_0 μπορεί να ερμηνευθεί ως εξής: Έστω τυχαία επελεγμένος τόνος και εκτιμημένος από έναν τυχαία επιλεγμένο σχολιαστή αναφοράς. Αυτή η εκτίμηση θα συμφωνούσε με την εκτίμηση του συστήματος σε ποσοστό $\frac{I_0}{100}$ με την απόφαση που θα λαμβανόταν από έναν δεύτερο τυχαία επιλεγμένο σχολιαστή αναφοράς. Εφαρμόζοντας αυτή την μεθοδολογία για τους 131 σύντομους τόνους στο σύνολο δεδομένων αναφοράς σε σχέση με αυτό του αρχικού ειδικού και των τριών πρόσθετων ειδικών υπολογίσαμε $I_0 = 1,12$. Ένα ποσοστό I_0 μεγαλύτερο της μονάδας, όπως είναι αυτό στο παράδειγμά μας, δείχνει ότι το σύστημα συμφωνεί με τους ανθρώπινους σχολιαστές συχνότερα από ότι συμφωνούν μεταξύ τους. Δεδομένου ότι το σύστημά μας δεν διαφωνεί με τους ανθρώπινους σχολιαστές περισσότερο από ότι διαφωνούν μεταξύ τους, μπορούμε να καταλήξουμε στο συμπέρασμα ότι το σύστημα αποδίδει τουλάχιστον το ίδιο καλά με τους ανθρώπους σε αυτή την δύσκολη και αμφίσημη απόφαση ταξινόμησης σύντομων τόνων. Συνεπώς, η χαμηλή επίδοση του συστήματος στην περίπτωση των σύντομων τόνων είναι απολύτως κατανοητή και δεν πρέπει να ληφθεί σαν ένδειξη μιας συστηματικής αποτυχίας ή αδυναμίας.

2.3.5.2.1 Ποσοτική συγκριτική μελέτη Σε προηγούμενη εργασία από μέλη του εργαστηρίου [119] προτείνεται μια διαφορετική μεθοδολογία για την επεξεργασία φυσικών δεδομένων με στόχο την εκτίμηση της συναισθηματικής κατάστασης του ανθρώπου. Σε εκείνη την εργασία ακολουθήθηκε μια παρόμοια προσέγγιση στην ανάλυση

Πίνακας 2.10: Πίνακας σύγχυσης για τόνους μικρού μήκους

	Ουδέτερο	Q1	Q2	Q3	Q4	Σύνολα
Ουδέτερο	**5**	1	5	3	0	14
Q1	1	**17**	6	12	6	42
Q2	3	2	**11**	2	1	19
Q3	4	6	7	**9**	3	29
Q4	4	6	4	7	**6**	27
Σύνολα	17	32	33	33	16	131

Πίνακας 2.11: Ποσοστιαίος πίνακας σύγχυσης για τόνους μικρού μήκους

	Ουδέτερο	Q1	Q2	Q3	Q4	Σύνολο
Σύνολο	**35,71%**	7,14%	35,71%	21,43%	0,00%	100,00%
Q1	2,38%	**40,48%**	14,29%	28,57%	14,29%	100,00%
Q2	15,79%	10,53%	**57,89%**	10,53%	5,26%	100,00%
Q3	13,79%	20,69%	24,14%	**31,03%**	10,34%	100,00%
Q4	14,81%	22,22%	14,81%	25,93%	**22,22%**	100,00%
Σύνολο	12,98%	24,43%	25,19%	25,19%	12,21%	100,00%

της οπτικής πληροφορίας του βίντεο με σκοπό τον εντοπισμό των χαρακτηριστικών γνωρισμάτων του προσώπου. Οι τιμές FAP τροφοδοτούνται έπειτα σε ένα σύστημα βασισμένο σε κανόνες που παράγει μια απάντηση σχετικά με την συναισθηματική κατάσταση του ανθρώπου.

Σε μεταγενέστερη έκδοση αυτής της εργασίας, αξιολογείται η πιθανότητα των ανιχνευμένων περιοχών να είναι πράγματι τα επιθυμητά χαρακτηριστικά γνωρίσματα του προσώπου με τη βοήθεια ανθρωπομετρικών στατιστικών προερχόμενα από το [258] και τον βαθμό εμπιστοσύνης των εξόδων του συστήματος. Η εκτίμηση βασισμένη σε κανόνες διαφοροποιείται επίσης στο ότι εμπλουτίζεται με τη δυνατότητα συνεκτίμησης των βαθμών εμπιστοσύνης που συνδέονται με κάθε FAP προκειμένου να ελαχιστοποιηθεί η διάδοση λάθους της διαδικασίας εξαγωγής χαρακτηριστικών γνωρισμάτων στο τελικό αποτέλεσμα [237].

Συγκριτικά με την τρέχουσα εργασία μας, αυτά τα συστήματα έχουν τα πρόσθετα πλεονεκτήματα:

1. ενσωματώνουν την γνώση που κατέχουν οι ειδικοί με την μορφή κανόνων στην διαδικασία ταξινόμησης

2. μπορούν να αντιμετωπίσουν ατέλειες προερχόμενες από το βήμα της εξαγωγής χαρακτηριστικών

Από την άλλη υστερούν στα εξής:

1. αγνοούν την δυναμική εξέλιξη του εκδηλωμένου συναισθήματος

2. δεν εξετάζουν τις υπόλοιπες μορφές πληροφορίες πέρα της οπτικής

Κατά συνέπεια, τα παραπάνω συστήματα είναι άριστοι υποψήφιοι για να συγκρίνουμε την τρέχουσα εργασία μας ώστε να αξιολογηθεί η συνεισφορά της προτεινόμενης δυναμικής και πολύμορφης προσέγγισης. Στον πίνακα 2.12 παρουσιάζουμε

αποτελέσματα από τις δύο προηγούμενες και την τρέχουσα προσέγγιση. Δεδομένου ότι η δυναμική εξέλιξη δεν εξετάζεται, κάθε πλαίσιο αντιμετωπίζεται ανεξάρτητα στα προγενέστερα συστήματα. Επομένως, το ποσοστό αναγνώρισης υπολογίζεται ως τον λόγο των σωστά ταξινομημένων πλαισίων προς το σύνολο τους. Κάθε πλαίσιο θεωρείται πως ανήκει στο ίδιο τεταρτημόριο με το τμήμα του οποίου είναι μέρος.

Αξίζει να σημειωθεί ότι τα αποτελέσματα είναι από τμήματα του συνόλου δεδομένων που επιλέχτηκε ως εκφραστικό για κάθε μεθοδολογία. Ενώ για την τρέχουσα εργασία αυτό αναφέρεται στο 72,54% του συνόλου δεδομένων και το κριτήριο της επιλογής είναι το μήκος του τόνου, στις προηγούμενες εργασίες μόνο 20% των πλαισίων επιλέχθηκαν με κριτήριο την σαφήνεια με την οποία παρατηρείται η έκφραση, αφού τα πλαίσια κοντά στην αρχή ή το τέλος του τόνου είναι συχνά κοντά στο ουδέτερο συναίσθημα, ώστε η οπτική είσοδος που παρέχεται σε ένα σύστημα να έχει νόημα. Η απόδοση όλων των προσεγγίσεων στο πλήρες σύνολο δεδομένων παρουσιάζεται στον πίνακα 2.13, όπου είναι προφανές ότι η δυναμική και πολύμορφη προσέγγιση είναι κατά πολύ ανώτερη.

Πίνακας 2.12: Ποσοστό αναγνώρισης σε τμήματα του συνόλου φυσικών δεδομένων

Μεθοδολογία	Βάσει κανόνων	Πιθανοτική/κανόνες	Δυναμικά/πολύμορφα
%	78,4%	65,1%	98,5%

Πίνακας 2.13: Ποσοστό αναγνώρισης επί του συνόλου φυσικών δεδομένων

Μεθοδολογία	Βάσει κανόνων	Πιθανοτική/κανόνες	Δυναμικά/πολύμορφα
%	27,8%	38,5%	81,5%

2.3.5.2.2 Ποιοτική συγκριτική μελέτη Όπως έχουμε αναφέρει ήδη, πρόσφατα συναντάται μεγάλος αριθμός δημοσιεύσεων στο ερευνητικό πεδίο της εκτίμησης της ανθρώπινης έκφρασης και συναισθήματος. Αν και η μεγάλη πλειοψηφία αυτών των εργασιών εστιάζει στις έξι καθολικές εκφράσεις και σε ακολουθίες με ακραίες εκφράσεις, θα ήταν παράλειψη εάν δεν γινόταν μια, έστω ποιοτική, σύγκριση με το τρέχων επίπεδο επιστήμης. Στον πίνακα 2.14 παρουσιάζουμε τα ποσοστά αναγνώρισης μερικών σημαντικών εργασιών. Βεβαίως, δεν είναι δυνατό ή δίκαιο να συγκριθούν τα αποτελέσματα άμεσα, δεδομένου ότι προέρχονται από εφαρμογή σε διαφορετικά πειραματικά σύνολα. Ακόμα και έτσι όμως, είναι δυνατό να γίνει ποιοτική σύγκριση:

1. Η [38] είναι μια βάση δεδομένων με θέματα που καθοδηγήθηκαν ώστε να επιδείξουν εκφράσεις του προσώπου που αντιστοιχούν στις έξι βασικές συναισθηματικές κατηγορίες. Τα θέματα της βάσης δεδομένων καθοδηγήθηκαν από έναν ερευνητή ώστε να εκτελέσουν μια σειρά 23 εκφράσεων του προσώπου περιλαμβάνοντας μία ή και συνδυασμούς μονάδων δράσης AUs.

2. Στη βάση δεδομένων MMI τα θέματα κλήθηκαν να επιδείξουν 79 ακολουθίες εκφράσεων που περιελάμβαναν μία ή και συνδυασμούς ενός ελάχιστου αριθμού

AUs και έναν πρωτότυπο συνδυασμό AUs. Καθοδηγήθηκαν από έναν εμπειρογνώμονα (κωδικοποιητή FACS) στο πώς να επιδείξουν τις απαραίτητες εκφράσεις του προσώπου και κλήθηκαν να περιλάβουν μια σύντομη ουδέτερη κατάσταση στην αρχή και στο τέλος κάθε έκφρασης.

3. Στο [35] θέματα κλήθηκαν να επιδείξουν 11 διαφορετικές εκφράσεις συναισθήματος, με την κάθε ακολουθία έκφρασης συναισθήματος να διαρκεί από 2 έως 6 δευτερόλεπτα και το μέσο μήκος των ακολουθιών έκφρασης να είναι 4 δευτερόλεπτα. Ακόμη και οι πιό σύντομες ακολουθίες σε αυτό το σύνολο δεδομένων είναι κατά πολύ μεγαλύτερες από τους σύντομους τόνους της βάσης SAL. Η αρχική οδηγία που δίνεται στα θέματα έχει λάβει ως πραγματική επιδειχθείσα έκφραση σε όλες τις προαναφερθείσες βάσεις δεδομένων, το οποίο σημαίνει ότι υπάρχει μια ελλοχεύουσα υπόθεση ότι δεν υπάρχει καμία διαφορά μεταξύ φυσικής και υποδυόμενης έκφρασης.

Όπως μπορούμε να δούμε, το κοινό χαρακτηριστικό μεταξύ των συνόλων δεδομένων που χρησιμοποιούνται συνηθέστερα στην βιβλιογραφία για την αξιολόγηση της έκφρασης του προσώπου και την αναγνώριση συναισθήματος είναι πως οι εκφράσεις είναι ακραίες, εξεζητημένες και υποδυόμενες. Κατά συνέπεια, αυτές οι εκφράσεις επιδεικνύονται με σαφήνεια και σε ακραία μορφή. Στην περίπτωση της φυσικής ανθρώπινης αλληλεπίδραση, από την άλλη, οι εκφράσεις είναι χαρακτηριστικά διακριτικότερες και συχνά όχι εύκολα διαχωρίσιμες. Επίσης, το στοιχείο της ομιλίας προσθέτει έναν σημαντικό βαθμό παραμόρφωσης στα χαρακτηριστικά γνωρίσματα του προσώπου που δεν σχετίζονται με την επιδειχθείσα έκφραση και μπορεί να είναι παραπλανητικό για ένα αυτοματοποιημένο σύστημα ανάλυσης εκφράσεων προσώπου.

Συνεπώς, υποστηρίζουμε ότι το γεγονός ότι η επίδοση της προτεινόμενης μεθοδολογίας όταν εφαρμόζεται σε ένα φυσικό σύνολο δεδομένων είναι συγκρίσιμη με την επίδοση άλλων εργασιών στο τρέχων επίπεδο της επιστήμης όταν εφαρμόζονται σε σκηνοθετημένες ακολουθίες είναι μια ένδειξη της επιτυχίας της. Επιπλέον, μπορούμε να παρατηρήσουμε ότι όταν οι εξαιρετικά σύντομοι τόνοι αφαιρούνται από το σύνολο στοιχείων της ταξινόμησης η επίδοση της προτεινόμενης προσέγγισης υπερβαίνει το 98%, το οποίο, σύμφωνα με την βιβλιογραφία, είναι εξαιρετικά υψηλό για ένα σύστημα αναγνώρισης συναισθήματος.

Η Multistream Hidden Markov Model προσέγγιση είναι πιθανώς αυτή που είναι αμεσότερα συγκρίσιμη με την εργασία που παρουσιάζεται εδώ. Αυτή είναι μια εναλλακτική δυναμική πολύμορφη προσέγγιση, όπου χρησιμοποιούνται HMMs αντί RNNs για τη προτυποποίηση της δυναμικής των εκφράσεων. Αν και έχει χρησιμοποιηθεί ένα διαφορετικό σύνολο στοιχείων για την πειραματική αξιολόγηση της προσέγγισης των MHMMs, η μεγάλη διαφορά των ποσοστών απόδοσης των δύο προσεγγίσεων δείχνει ότι η χρησιμοποίηση των RNNs για τη δυναμική πολύμορφη ταξινόμηση ανθρώπινου συναισθήματος είναι μια πολλά υποσχόμενη κατεύθυνση.

2.3.6 Συμπεράσματα

Στην εργασία αυτή έχουμε εστιάσει στο πρόβλημα της αναγνώρισης ανθρώπινου συγαισθήματος για την περίπτωση φυσικών, σε αντίθεση με υποδυόμενες και ακραίες εκφράσεις. Τα κύρια στοιχεία της προσέγγισής μας είναι ότι χρησιμοποιούμε πολλαπλούς αλγορίθμους για την εξαγωγή των χαρακτηριστικών γνωρισμάτων του προσώπου ώστε η γενικά δύσκολη διαδικασία να γίνει πιό εύρωστη έναντι λαθών που

Πίνακας 2.14: Ποσοστά αναγνώρισης στην ευρύτερη βιβλιογραφία [38], [177], [264]

Μεθοδολογία	Ποσοστό αναγνώρισης	Σύνολο δεδομένων
TAN	83,31%	Cohen2003
Multi-level HMM	82,46%	Cohen2003
TAN	73,22%	Cohn–Kanade
Χρονικοί κανόνες	86,6%	MMI
Multistream HMM	72,42%	Chen2000
Προτεινόμενη μεθοδολογία	81,55%	SAL Βάση
Προτεινόμενη μεθοδολογία	**98,55%**	Τμήμα βάσης SAL

διαδίδονται από την υποενότητα της επεξεργασίας εικόνας, ενώ από την πλευρά της κατηγοριοποίησης εστιάζουμε στην δυναμική των εκφράσεων του προσώπου παρά στις ακριβείς παραμορφώσεις του προσώπου που σχετίζονται με αυτές, όντας κατά συνέπεια σε θέση να χειριστούμε ακολουθίες στις οποίες η αλληλεπίδραση είναι φυσική ή φυσιοκρατική παρά σκηνοθετημένη ή ακραία και τέλος, ακολουθούμε μια πολύμορφη προσέγγιση όπου οι ακουστικές και οπτικές μορφές συνδυάζονται, ενισχύοντας κατά συνέπεια την απόδοση και τη σταθερότητα του συστήματος.

Από τεχνικής άποψης, οι συνεισφορές μας περιλαμβάνουν το τροποποιημένο επίπεδο εισόδου που επιτρέπει στο Elman δίκτυο να επεξεργαστεί τόσο δυναμικές όσο και στατικές εισόδους ταυτόχρονα. Αυτό χρησιμοποιείται ώστε να συνδυάζονται επιτυχώς διαφορετική φύσεως είσοδοι προκειμένου να οδηγήσουν σε μια πραγματικά πολύμορφη αρχιτεκτονική. Επιπλέον, η τροποποίηση του επιπέδου εξόδου επιτρέπει στο δίκτυο Elman να ενσωματώνει προηγούμενες τιμές, με το βάρος σημασίας τους να μειώνεται εκθετικά με τον χρόνο και ενισχύει την σταθερότητα του συστήματος.

2.4 Αναγνώριση συναισθηματικών καταστάσεων από πολλαπλές μορφές πληροφορίας

Παραμένοντας στο πεδίο της συναισθηματικής υπολογιστικής αλλά εστιάζοντας περισσότερο στην επεξεργασία πολλαπλών μορφών πληροφορίας, στις μεθόδους συγχώνευσης τους και την συνεισφορά κάθε καναλιού στο τελικό αποτέλεσμα πραγματοποιήσαμε την μελέτη της διαδικασίας και την καταγραφή ενός πειραματικού συνόλου συναισθηματικού περιεχομένου, την επεξεργασία του για την εξαγωγή χαρακτηριστικών γνωρισμάτων, την επιλογή χαρακτηριστικών με στόχο την μείωση της διάστασης του προβλήματος αλλά και την βελτίωση της απόδοσης του κατηγοριοποιητή και τέλος την εφαρμογή διαφορετικών σχημάτων συγχώνευσης πολλαπλών μορφών πληροφορίας είτε σε επίπεδο γνωρισμάτων είτε σε επίπεδο απόφασης.

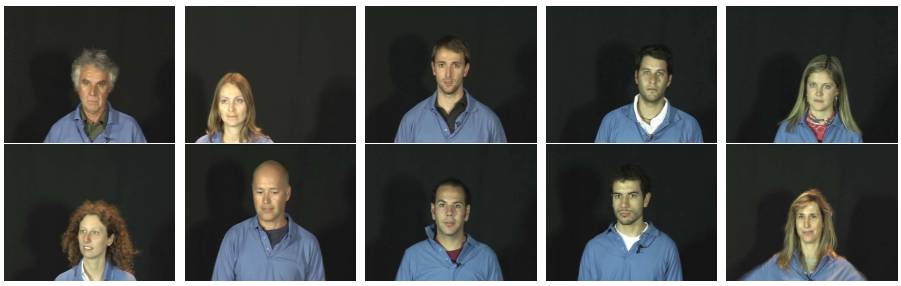
2.4.1 Συλλογή πολύμορφων δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε σε αυτήν την μελέτη συλλέχθηκε κατά τη διάρκεια του τρίτου θερινού σχολείου του προγράμματος HUMAINE, που διοργανώθηκε στην Γένοβα της Ιταλίας τον Σεπτέμβριο του 2006. Η γενική διαδικασία καταγραφής βασίστηκε στην βάση δεδομένων GEMEP [10], μια συλλογή απεικονίσεων συναισθηματικών εκφράσεων από πολλαπλές μορφές πληροφορίας, όπου έγινε

καταγραφή δεδομένων όσον αφορά στις εκφράσεις του προσώπου, στην κίνηση του σώματος και τις χειρονομίες και στην ομιλία.

2.4.1.1 Συμμετέχοντες

Οι δέκα συμμετέχοντες ήταν κατανεμημένοι όσο το δυνατόν ομοιόμορφα σχετικά με το φύλο τους (εικόνα 2.4.1.1). Τα θέματα εκπροσωπούσαν πέντε διαφορετικούς πολιτισμούς αφού προέρχονταν από: Γαλλία, Γερμανία, Ελλάδα, Ισραήλ και Ιταλία.



Σχήμα 2.27: Οι συμμετέχοντες των καταγραφών

2.4.1.2 Τεχνικές ρυθμίσεις

Δύο DV κάμερες (με ρυθμό καταγραφής 25 πλαίσια/δευτερόλεπτο) κατέγραψαν τα θέματα από εμπρόσθια όψη. Μια κάμερα κατέγραφε το σώμα του θέματος και η άλλη εστίαζε στο πρόσωπο του θέματος. Έχουμε επιλέξει μια τέτοια ρύθμιση επειδή η απαιτούμενη ανάλυση εικόνας για την εξαγωγή χαρακτηριστικών γνωρισμάτων του προσώπου είναι πολύ μεγαλύτερη από αυτή για την ανίχνευση κινήσεων σώματος ή την παρακολούθηση χειρονομιών. Αυτή η απαίτηση θα μπορούσε να ικανοποιηθεί μόνο εάν μια κάμερα εστίαζε στο πρόσωπο του δράστη. Οι ροές βίντεο συγχρονίστηκαν χειρωνακτικά μετά από την διαδικασία καταγραφής. Υιοθετήσαμε μερικούς περιορισμούς σχετικά με την συμπεριφορά και τον ρουχισμό των θεμάτων. Μακριά μανίκια και καλυμμένος λαιμός προτιμήθηκαν αφού η πλειοψηφία των αλγορίθμων ανίχνευσης χεριών και κεφαλιού βασίζονται στην ανίχνευση χρώματος του δέρματος και επιπλέον, χρησιμοποιήθηκε ομοιόμορφο φόντο που καθιστά την διαδικασία αφαίρεσης φόντου ευκολότερη. Όσον αφορά στην διαδικασία εξαγωγής χαρακτηριστικών γνωρισμάτων του προσώπου εφαρμόσαμε μερικούς περιορισμούς όπως η απουσία γυαλιών όρασης, τριχοφυίας στο πρόσωπο και η αποφυγή χτενισμάτων που πιθανόν να επικαλύπτουν κάποια χαρακτηριστικά του προσώπου.

Για τις ακουστικές καταγραφές χρησιμοποιήσαμε ένα σύστημα άμεσης καταγραφής σε δίσκο υπολογιστή με την χρήση λογισμικού. Χρησιμοποιήσαμε μια εξωτερική κάρτα ήχου που συνδέθηκε με τον υπολογιστή μέσω του IEEE 1394 πρωτοκόλλου μεταφοράς δεδομένων υψηλής ταχύτητας (επίσης γνωστό ως FireWire ή i.Link). Ένα μικρόφωνο τοποθετήθηκε στο πουκάμισο των συμμετεχόντων και συνδέθηκε με ένα ασύρματο HF πομπό και ο δέκτης συνδέθηκε στην κάρτα ήχου με την χρήση ενός προσαρμοστήρα XLR (ισορροπημένος ακουστικός προσαρμοστήρας για υψηλής ποιότητας μικρόφωνα και διασύνδεση εξοπλισμών). Η εξωτερική κάρτα ήχου περιέλαβε έναν προενισχυτή (για δύο εισόδους XLR) που χρησιμοποιήθηκε προκειμένου να ρυθμίσει το κέρδος εισόδου και για να ελαχιστοποιήσει τον αντίκτυπο του θορύβου που εισάγει το σύστημα καταγραφής. Ο ρυθμός δειγματοληψίας της καταγραφής ήταν 44,1 kHz και ο κβαντισμός στα 16 bit.

2.4.1.3 Διαδικασία

Οι συμμετέχοντες κλήθηκαν να υποδυθούν οκτώ συναισθηματικές καταστάσεις: θυμός, απελπισία, ενδιαφέρον, ευχαρίστηση, θλίψη, ενόχληση, χαρά και υπερηφάνεια, κατανεμημένες εξίσου στο συναισθηματικό διάστημα ενεργοποίησης/εκτίμησης (βλ. εικόνα 2.15). Κατά τη διάρκεια της διαδικασίας καταγραφής ένας ερευνητής είχε το ρόλο του σκηνοθέτη και καθοδηγούσε τα θέματα κατά την διάρκεια της διαδικασίας. Οι συμμετέχοντες κλήθηκαν να εκτελέσουν συγκεκριμένες χειρονομίες που αντιστοιχούν σε κάθε συναίσθημα. Οι επιλεγμένες χειρονομίες παρουσιάζονται στην εικόνα 2.15.

Πίνακας 2.15: Τα υποδυόμενα συναισθήματα και οι αντίστοιχες χειρονομίες

Συναίσθημα	Εκτίμηση	Ενεργοποίηση	Χειρονομία
Θυμός	Αρνητική	Υψηλή	Βίαιη κάθοδος χεριών
Απόγνωση	Αρνητική	Υψηλή	Παράτα με
Ενδιαφέρον	Θετική	Χαμηλή	Σήκωμα χεριού
Ικανοποίηση	Θετική	Χαμηλή	Άνοιγμα χεριών
Λύπη	Αρνητική	Χαμηλή	Απαλή κάθοδος χεριών
Ενόχληση	Αρνητική	Χαμηλή	Άφησε με
Χαρά	Θετική	Υψηλή	Κυκλική κίνηση
Περηφάνια	Θετική	Υψηλή	Κλείσιμο χεριών στο στήθος

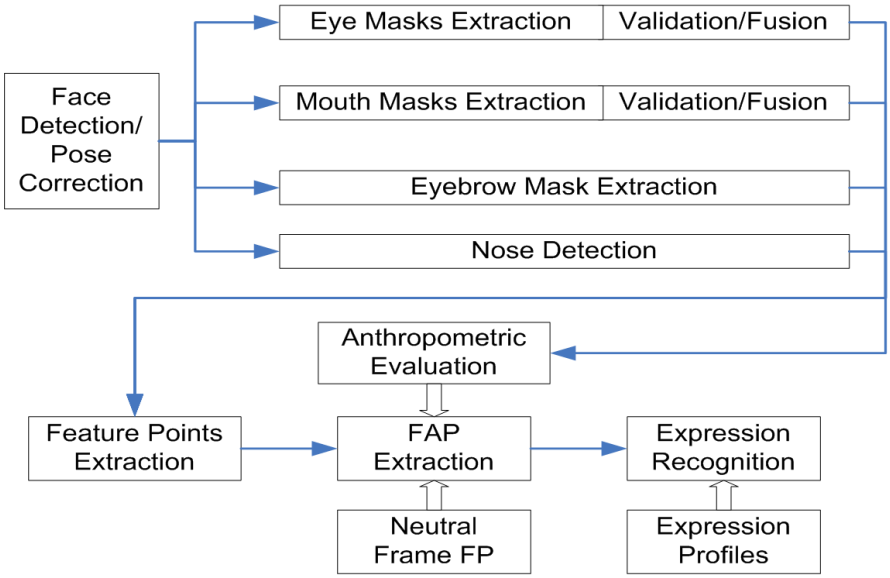
Όπως και στην περίπτωση του συνόλου δεδομένων GEMEP [10], προφέρθηκε μια ψευδόγλωσσική πρόταση από τους ηθοποιούς κατά τη διάρκεια εκτέλεσης των υποδυόμενων συναισθηματικών καταστάσεων. Η πρόταση 'Toko, damato ma gali sa' επιλέχθηκε με σκοπό να εξυπηρετήσει διαφορετικές ανάγκες. Κατ' αρχάς, εφόσον οι διαφορετικοί ομιλητές έχουν διαφορετικές μητρικές γλώσσες, η χρήση μιας συγκεκριμένη γλώσσας δεν θα ήταν επαρκής για αυτήν την ενώ ιδιαίτερη προσοχή δόθηκε ώστε να περιλάβουμε στην πρόταση φωνήματα που υπάρχουν σε όλες τις γλώσσες των ομιλητών. Επίσης, οι λέξεις στην πρόταση αποτελούνται από απλά δίφωνα ('ma' και 'sa'), διπλά δίφωνα ('toko' και 'gali') ή τριπλά δίφωνα ('damato'). Κατόπιν, τα συμπεριλαμβανόμενα φωνήεντα ('o', 'a', 'i') είναι φωνήεντα που είναι σχετικά απόμακρα στο διάστημα φωνηέντων, π.χ. το τρίγωνο φωνηέντων και προφέρονται συνήθως παρόμοια σε όλες τις γλώσσες της ομάδας ομιλητών. Προτείναμε στους ομιλητές μια ερμηνεία της πρότασης. Ο 'toko' είναι το όνομα ενός υποτιθέμενου 'πράκτορα', δηλ., ένα πραγματικό ή συνθετικό άτομο, με τον οποίο οι χρήστες είναι σε αλληλεπίδραση. Έπειτα, το 'damato magali sa' υποτίθεται ότι σημαίνει κάτι σαν 'μπορείς να το ανοίξεις'. Κάθε συναίσθημα επαναλήφθηκε τρεις φορές από κάθε ηθοποιό, με αποτέλεσμα να συλλέξουμε 240 υποδημένες χειρονομίες, εκφράσεις του προσώπου και λεκτικά δείγματα.

2.4.2 Εξαγωγή χαρακτηριστικών γνωρισμάτων

2.4.2.1 Εξαγωγή χαρακτηριστικών γνωρισμάτων του προσώπου

Μια επισκόπηση της προτεινόμενης μεθοδολογίας φαίνεται στην εικόνα 2.4.2.1. Το πρόσωπο εντοπίζεται αρχικά, έτσι ώστε να μπορούν να υπολογιστούν οι κατά προσέγγιση θέσεις περιοχών χαρακτηριστικών γνωρισμάτων του προσώπου από την θέση και

την περιστροφή του κεφαλιού. Η περιστροφή του προσώπου κατά τον +Z άξονα υπολογίζεται, διορθώνεται η θέση του κεφαλιού και το κεφάλι κατατέμνεται στις ακόλουθες περιοχές του προσώπου: αριστερό μάτι/φρύδι, δεξί μάτι/φρύδι, μύτη και στόμα. Κάθε μια από αυτές τις περιοχές, αποκαλούμενες περιοχές υποψήφιων γνωρισμάτων, περιέχουν τα χαρακτηριστικά γνωρίσματα των οποίων τα όρια πρέπει να εξαχθούν για τους σκοπούς της εργασίας μας. Μέσα σε κάθε υποψήφια περιοχή η ακριβής εξαγωγή χαρακτηριστικών γνωρισμάτων εξάγεται με ένα συνδυασμό μεθόδων για κάθε χαρακτηριστικό γνώρισμα του προσώπου, δηλ. μάτια, φρύδια, στόμα και μύτη, που χρησιμοποιώντας μια προσέγγιση με πολλές ενδείξεις, παράγει έναν μικρό αριθμό ενδιάμεσων μασκών χαρακτηριστικών σημείων. Οι παραγόμενες μάσκες για κάθε χαρακτηριστικό του προσώπου συγχωνεύονται για να παράγουν μια τελική μάσκα. Η διαδικασία συγχώνευσης μασκών χρησιμοποιεί ανθρωπομετρικά κριτήρια [258] για την επικύρωση και την ανάθεση βάρους σε κάθε ενδιάμεση μάσκα αντίστοιχες της εμπιστοσύνης που την συνοδεύει. Οι σταθμισμένες μάσκες των χαρακτηριστικών γνωρισμάτων συγχωνεύονται και παράγεται μια τελική μάσκα μαζί με την συνολική εκτίμηση εμπιστοσύνης.



Σχήμα 2.28: Εποπτική άποψη της διαδικασίας εξαγωγής χαρακτηριστικών γνωρισμάτων του προσώπου

Δεδομένου ότι η παραπάνω διαδικασία εντοπίζει και παρακολουθεί σημεία στην περιοχή του προσώπου, επιλέξαμε να εργαστούμε με MPEG˜4 FAPs (παράμετροι εμψύχωσης του προσώπου) και όχι με μονάδες δράσης (AUs), δεδομένου ότι τα πρώτα ορίζονται ρητά στην μέτρηση της παραμόρφωσης αυτών των σημείων χαρακτηριστικών γνωρισμάτων. Επιπρόσθετα, διακριτά σημεία είναι ευκολότερο να παρακολουθηθούν σε περιπτώσεις ακραίων περιστροφών και η θέση τους μπορεί να υπολογιστεί βάσει στατιστικών ανθρωπομετρικών δεδομένων σε περιπτώσεις επικάλυψης, πράγμα που δεν συμβαίνει συνήθως σε περίπτωση ολόκληρων χαρακτηριστικών γνωρισμάτων του προσώπου. Ένα άλλο χαρακτηριστικό των FAPs που αποδείχθηκε χρήσιμο είναι η τιμή τους, η οποία είναι κρίσιμη στο να διαφοροποιήσει περιπτώσεις ποικίλης ενεργοποίησης του ιδίου συναισθήματος (π.χ. χαρά και ευθυμία) [196] και εκμεταλλεύεται την ασάφεια σε συστήματα βασισμένα σε κανόνες [119]. Η μέτρηση FAPs απαιτεί τη διαθεσιμότητα ενός πλαισίου όπου η έκφραση του θέματος είναι ουδέτερη. Αυτό το

πλαίσιο καλείται ουδέτερο πλαίσιο και επιλέγεται χειρωνακτικά από τις υπό ανάλυση τηλεοπτικές ακολουθίες. Οι τελικές μάσκες χαρακτηριστικών γνωρισμάτων χρησιμοποιούνται για να εξάγουν 19 χαρακτηριστικά σημεία του προσώπου (FPs) [196]. Τα σημεία χαρακτηριστικών γνωρισμάτων που λαμβάνονται από κάθε πλαίσιο συγκρίνονται με τα αντίστοιχα του ουδέτερου πλαισίου για να υπολογιστούν οι παραμορφώσεις του προσώπου και τα FAPs. Τα επίπεδα εμπιστοσύνης στην εκτίμηση των FAP προέρχονται από τα αντίστοιχα επίπεδα εμπιστοσύνης των σημείων χαρακτηριστικών γνωρισμάτων. Τα FAPs χρησιμοποιούνται μαζί με τα επίπεδα εμπιστοσύνης τους για να εκτιμηθεί η έκφραση του προσώπου.

Σε συμφωνία με τις άλλες μορφές πληροφορίας, τα χαρακτηριστικά γνωρίσματα του προσώπου πρέπει να επεξεργαστούν ώστε να δημιουργήσουν ένα διάνυσμα τιμών ανά τόνο. Τα FAPs αρχικά αντιστοιχούν σε κάθε πλαίσιο του τόνου. Δύο προσεγγίσεις θεωρήθηκαν. Η πρώτη προσέγγιση βασίζεται στην εξαγωγή του πιό προεξέχοντος πλαισίου μέσα σε έναν τόνο. Κατά τη διάρκεια αυτής της διαδικασίας, μια μέση τιμή υπολογίζεται για κάθε FAP στον τόνο και έπειτα το πλαίσιο με την υψηλότερη απόκλιση επιλέγεται. Από την άλλη ένας τρόπος να αποτυπωθεί η χρονική εξέλιξη των τιμών FAP είναι να υπολογιστεί ένα σύνολο στατιστικών χαρακτηριστικών γνωρισμάτων από αυτές τις τιμές και τα παράγωγά τους. Αυτή η τελευταία διαδικασία εμπνεύστηκε από την ισοδύναμη διαδικασία που εκτελείται στα ακουστικά χαρακτηριστικά γνωρίσματα, όπως θα φανεί και αργότερα. Επιλέξαμε πειραματικά την δεύτερη καθώς το ποσοστό αναγνώρισης με αυτό το σύνολο εισόδου ήταν ανώτερο. Περισσότερο περίπλοκες τεχνικές για να εξαχθεί το πιο προεξέχον πλαίσιο συμπεριλαμβάνονται στη μελλοντική εργασία.

2.4.2.2 Εξαγωγή χαρακτηριστικών γνωρισμάτων του σώματος

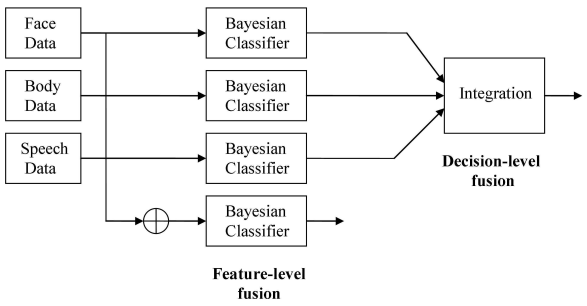
Η παρακολούθηση του σώματος και των χεριών των θεμάτων έγινε χρησιμοποιώντας την πλατφόρμα EyesWeb [27]. Εκκινώντας από τη σιλουέτα και τις περιοχές χεριών των ηθοποιών, εξαγάγαμε πέντε κύριες εκφραστικές ενδείξεις κίνησης, χρησιμοποιώντας την βιβλιοθήκη επεξεργασίας εκφραστικών χειρονομιών του EyesWeb [28]: ποσότητα κίνησης, δείκτης συστολής του σώματος, ταχύτητα, επιτάχυνση και ομαλότητα χεριού. Τα στοιχεία κανονικοποιήθηκαν σύμφωνα με τη συμπεριφορά που απέδειξε κάθε χρήστης και εξετάζοντας τις μέγιστες και τις ελάχιστες τιμές κάθε εκφραστικής παραμέτρου κίνησης για κάθε θέμα, ήταν δυνατόν να συγκριθούν ομοιόμορφα τα στοιχεία από όλα τα θέματα. Η αυτόματη εξαγωγή επιτρέπει την λήψη της χρονικής σειράς των επιλεγμένων ενδείξεων κίνησης, ανάλογα με τον ρυθμό πλαισίων του βίντεο. Για κάθε κατηγορία ενδείξεων κίνησης επιλέξαμε ένα υποσύνολο χαρακτηριστικών γνωρισμάτων που περιγράφουν την δυναμική των παραμέτρων εκφραστικότητας. Με βάση το πρότυπο που προτάθηκε στο [32] υπολογίσαμε τους ακόλουθους δυναμικούς δείκτες κίνησης: αρχική και τελική κλίση, αρχική και τελική κλίση στην κύρια κορυφή, μέγιστη τιμή, μέση τιμή, λόγος μεταξύ της μέγιστης και της μέσης τιμής, λόγος της μέγιστης και της αμέσως επόμενης τιμής, ενεργειακό κέντρο, απόσταση μεταξύ μέγιστης τιμής και ενεργειακού κέντρου, δείκτης συμμετρίας, αριθμός τοπικών μεγίστων, αριθμός τοπικών μεγίστων πρίν το ολικό μέγιστο, λόγος της διάρκειας του ολικού μεγίστου και της συνολικής διάρκειας. Αυτή η διαδικασία ακολουθήθηκε για κάθε σύνολο ενδείξεων κίνησης όλων των βίντεο της βάσης δεδομένων, έτσι ώστε κάθε χειρονομία να χαρακτηρίζεται από ένα σύνολο 80 χαρακτηριστικών γνωρισμάτων κίνησης.

2.4.2.3 Εξαγωγή ακουστικών χαρακτηριστικών γνωρισμάτων

Το σύνολο χαρακτηριστικών γνωρισμάτων που χρησιμοποιήσαμε περιέχει χαρακτηριστικά γνωρίσματα βασισμένα στην ένταση, τον μουσικό τόνο, τους MFC (Mel-frequency cepstrum) συντελεστές, φασματικές ζώνες Bark, χαρακτηριστικά τμήματος φωνής και μήκος παύσης. Το πλήρες σύνολο περιέχει 377 χαρακτηριστικά γνωρίσματα. Τα χαρακτηριστικά γνωρίσματα από την καμπύλη της έντασης και του μουσικού τόνου εξάγονται χρησιμοποιώντας ένα σύνολο 32 στατιστικών γνωρισμάτων. Αυτό το σύνολο χαρακτηριστικών γνωρισμάτων εφαρμόζεται τόσο στην καμπύλη της έντασης, του μουσικού τόνου όσο και στις παραγώγους τους. Στα δεδομένα δεν εφαρμόστηκε κανονικοποίηση πριν από την εξαγωγή χαρακτηριστικών γνωρισμάτων. Συγκεκριμένα, δεν κανονικοποιούμε με βάση τον χρήστη ή το γένος του όπως συχνά γίνεται προκειμένου να ενσωματωθεί η μοναδικότητα κάθε χρήστη. Εξετάσαμε τα ακόλουθα 32 χαρακτηριστικά γνωρίσματα: το μέγιστη, μέση και ελάχιστη τιμή, πιο συχνά εμφανιζόμενη τιμή, ενδοτεταρτημοριακό πεδίο (διαφορά μεταξύ του 75ου και 25ου εκατοστημορίου), κύρτωση, τρίτη κεντρική ροπή δειγμάτων, πρώτος και δεύτερος συντελεστής τεταρτημοριακής οπισθοδρόμησης, εκατοστημόρια 2.5 %, 25 %, 50 %, 75 %, and 97.5 %, ασσυμετρία καμπυλότητας (skewness), απόκλιση, διακύμανση. Έτσι προκύπτουν 64 χαρακτηριστικά γνωρίσματα του μουσικού τόνου και 64 χαρακτηριστικά γνωρίσματα βασισμένα στην έντασης. Αυτό το σύνολο χαρακτηριστικών γνωρισμάτων χρησιμοποιήθηκε αρχικά για την μελέτη περιγράμματος αλλά ταιριάζουν και στην περίπτωση της χρονικής εξέλιξης ή κατά τον φασματικό άξονα. Πράγματι, εξαγάγαμε επίσης παρόμοια χαρακτηριστικά γνωρίσματα για τις φασματικές ζώνες όπως γίνεται στο [135]. Έτσι προέκυψαν 13 συντελεστές MFC χρησιμοποιώντας μέσο όρο σε χρονικό παράθυρο. Αντίστοιχα ένα σύνολο από 35 χαρακτηριστικά γνωρίσματα εξάχθηκαν και από τον φωνητικό τόνο και το μήκος των τμημάτων με φωνή. Τέλος 35 χαρακτηριστικά γνωρίσματα εξάχθηκαν βάσει του μήκους των μικρών διακοπών (ή σιωπών) και άλλα 35 βάσει του μήκους των μη-διακοπών.

2.4.3 Μονόμορφη και πολύμορφη αναγνώριση συναισθήματος

Προκειμένου να συγκριθούν τα αποτελέσματα των προσεγγίσεων από μοναδική και από πολλαπλές μορφές πληροφορίας, χρησιμοποιήσαμε μια ευρέως χρησιμοποιούμενη προσέγγιση βασισμένη σε έναν Bayesian ταξινομητή (BayesNet) υλοποιημένο από το λογισμικό Weka, μια ελεύθερα προσβάσιμη εργαλειοθήκη που περιέχει μια συλλογή από αλγόριθμους μηχανικής μάθησης για εξόρυξη δεδομένων [250]. Στο σχήμα 2.4.3 παρουσιάζουμε το προτεινόμενο πλαίσιο:



Σχήμα 2.29: Επισκόπηση του πλαισίου αναγνώρισης συναισθήματος

Όπως φαίνεται στο αριστερό μέρος του διαγράμματος, ένας ταξινομητής εφαρ-

μόστηκε σε κάθε μορφή (πρόσωπο, χειρονομίες, ομιλία). Όλα τα σύνολα δεδομένων ήταν κανονικοποιημένα. Εφαρμόστηκε διακριτοποίηση χαρακτηριστικών γνωρισμάτων βασισμένη στο κριτήριο MDL Kononenko (ελάχιστου μήκους περιγραφής) [138] για να μειώσει την πολυπλοκότητα εκπαίδευσης. Μια προσέγγιση περιβλήματος (wrapper) ακολουθήθηκε για την επιλογή υποσυνόλου χαρακτηριστικών γνωρισμάτων (που επιτρέπει την αξιολόγηση των συνόλων με τη χρήση ενός σχήματος εκμάθησης) προκειμένου να μειωθεί ο αριθμός των εισόδων στους ταξινομητές και να εντοπιστούν συγκεκριμένα χαρακτηριστικά γνωρίσματα που μεγιστοποιούν την απόδοση του ταξινομητή. Μια μέθοδος αναζήτησης του καλύτερου πρώτου με εμπρόσθια κατεύθυνση χρησιμοποιήθηκε και τα αποτελέσματα φαίνονται στους πίνακες 2.16 και 2.17. Επιπλέον, σε όλα τα πειράματα, το σύνολο δεδομένων χρησιμοποιήθηκε για εκπαίδευση και δοκιμή με την μέθοδο της διαγώνιας επικύρωσης (cross-validation).

Αξιολογήσαμε δύο διαφορετικά πρωτόκολλα: (1) ένα πρωτόκολλο εκπαιδευμένο και δοκιμασμένο σε όλα τα 240 δείγματα δεδομένων, ακόμα και όταν κάποια μορφή πληροφορίας δεν ήταν διαθέσιμοι για μερικά δείγματα εξαιτίας δυσκολιών στην διαδικασία εξαγωγής χαρακτηριστικών ή ανακολουθίας στο σενάριο καταγραφής του σώματος και (2) ένα πρωτόκολλο χρησιμοποιώντας διανύσματα μόνο από δεδομένα διαθέσιμα για όλες τις μορφές πληροφορίας. Όπως φαίνεται στα αποτελέσματα της επόμενης ενότητας, ενώ το πρώτο πρωτόκολλο επιτρέπει την επεξεργασία δεδομένων με ελλιπή δείγματα, είναι λιγότερος ακριβές, το δεύτερο είναι λιγότερο προσαρμοστικός, αλλά ακριβέστερο στην ταξινόμηση.

Για την συγχώνευση των εκφράσεων του προσώπου, των χειρονομιών και ακουστικών πληροφοριών, δύο διαφορετικές προσεγγίσεις ακολουθήθηκαν (δεξί μέρος του σχήματος 2.4.3): βασισμένη είτε στα χαρακτηριστικά γνωρίσματα, όπου χρησιμοποιείται ένας ενιαίος ταξινομητής με τα χαρακτηριστικά γνωρίσματα και των τριών μορφών είτε βασισμένα στην απόφαση, όπου χρησιμοποιείται ένας ταξινομητής για κάθε μορφή και τα αποτελέσματα συνδυάζονται σε αργότερο βήμα. Στη δεύτερη προσέγγιση η έξοδος υπολογίστηκε συνδυάζοντας τις προκύπτουσες πιθανότητες των μονόμορφων συστημάτων. Πραγματοποιήσαμε δύο πειράματα για δύο διαφορετικές προσεγγίσεις για την βασισμένη στην απόφαση συγχώνευση. Η πρώτη προσέγγιση περιλαμβάνει την επιλογή του συναισθήματος στο οποίο ανατέθηκε η μεγαλύτερη συνδυασμένη πιθανότητα από τις τρεις μορφές πληροφορίας. Η δεύτερη προσέγγιση περιλαμβάνει την επιλογή του συναισθήματος που αντιστοιχεί σε πλειοψηφική απόφαση (majority voting) όπου κάθε ψήφος αντιστοιχεί σε μια από τις τρεις μορφές. Εάν δεν προκύψει 'αυτοδυναμία' μέσω αυτής της μεθόδου το συναίσθημα με την υψηλότερη πιθανότητα σε οποιαδήποτε από τις τρεις μορφές επιλέγεται.

2.4.4 Αποτελέσματα

2.4.4.1 Αναγνώριση συναισθήματος από εκφράσεις του προσώπου

Στον πίνακα 2.18 παρουσιάζεται ο πίνακας σύγχυσης του συστήματος αναγνώρισης συναισθήματος που βασίζεται μόνο στις εκφράσεις του προσώπου όταν χρησιμοποιούνται όλα τα δείγματα (πρώτο πρωτόκολλο). Η γενική απόδοση αυτού του ταξινομητή είναι 48.3% (αυξάνεται μέχρι 59.6% όταν χρησιμοποιούνται μόνο τα διαθέσιμα δείγματα για όλες τις μορφές). Τα πιο αναγνωρίσιμα συναισθήματα είναι ο θυμός (56.67%), η ενόχληση, η χαρά και η ευχαρίστηση (53.33%). Η περηφάνεια συγχέεται με με την ευχαρίστηση (20%), ενώ η θλίψη με την ενόχληση (20%), ένα συναίσθημα στο ίδιο τεταρτημόριο του διαστήματος αξιολόγησης-ενεργοποίησης και

γενικά τα συναισθήματα αυτά ανήκουν στην ίδια οικογένεια (περιοχή στην διαστατική αναπαράσταση συναισθήματος) συναισθημάτων. Αξίζει να σημειωθεί πάντως πως η εκφραστικότητα των συμμετεχόντων, σε μια ανεπίσημη εκτίμηση, όσον αφορά στο κανάλι πληροφορίας των εκφράσεων του προσώπου ήταν πολύ φτωχή και γενικά τα θέματα μάλλον παραμελούσαν να εκφράσουν μέσω των εκφράσεων του προσώπου το ζητούμενο συναίσθημα.

2.4.4.2 Αναγνώριση συναισθήματος από εκφραστικές παραμέτρους του σώματος

Στον πίνακα 2.19 παρουσιάζεται η απόδοση του συστήματος αναγνώρισης συναισθήματος που βασίζεται μόνο στις χειρονομίες όταν χρησιμοποιούνται όλα τα δείγματα (πρώτο πρωτόκολλο). Η γενική απόδοση αυτού του ταξινομητή είναι 67.1% (αυξάνεται μέχρι 83.2% όταν χρησιμοποιούνται μόνο τα διαθέσιμα δείγματα για όλες τις μορφές). Τα πιο αναγνωρίσιμα συναισθήματα είναι θυμός και περηφάνια με αρκετά μεγάλα ποσοστά αναγνώρισης (80 και 96.67% αντίστοιχα. Η θλίψη συγχέεται μερικώς με την περηφάνια (36.67%) καθώς και με τα υπόλοιπα συναισθήματα εκτός από τον θυμό.

2.4.4.3 Αναγνώριση συναισθήματος από ακουστικές ενδείξεις

Ο πίνακας 2.20 παρουσιάζει την απόδοση του συστήματος αναγνώρισης συγκίνησης που βασίζεται μόνο στην ακουστική πληροφορία όταν χρησιμοποιούνται όλα τα δείγματα (πρώτο πρωτόκολλο). Η γενική απόδοση αυτού του ταξινομητή είναι 57.1% ενώ αυξάνεται μέχρι 70.8% όταν χρησιμοποιούνται μόνο τα διαθέσιμα δείγματα για όλες τις μορφές. Ο θυμός και η θλίψη αναγνωρίζονται ακριβέστερα ((93.33% και 76.67% αντίστοιχα). Η απόγνωση συγχέεται με την ευχαρίστηση (23.33%) και γενικά δεν λαμβάνει πολύ καλά ποσοστά αναγνώρισης.

2.4.4.4 Συγχώνευση σε επίπεδο χαρακτηριστικών γνωρισμάτων

Ο πίνακας 2.21 επιδεικνύει τον πίνακα σύγχυσης της αναγνώρισης συναισθήματος από πολλαπλές μορφές πληροφορίας όταν όλα τα δείγματα χρησιμοποιούνται (πρώτο πρωτόκολλο). Η γενική απόδοση αυτού του ταξινομητή ήταν 78,3%, το οποίο είναι πολύ υψηλότερο από την απόδοση που επιτυγχάνεται από το επιτυχέστερο σύστημα με είσοδο μια μόνο μορφή πληροφορίας. Η διαγώνιος του πίνακα αποκαλύπτει ότι όλα τα συναισθήματα, εκτός από την απελπισία, μπορούν να αναγνωριστούν με ακρίβεια ανώτερη του 70%. Ο θυμός αναγνωρίστηκε με την υψηλότερη ακρίβεια, όπως και σε όλα τα μονόμορφα συστήματα.

2.4.4.5 Συγχώνευση σε επίπεδο αποφάσεων

Η προσέγγιση βασισμένη στην συγχώνευση απόφασης έλαβε χαμηλότερα ποσοστά αναγνώρισης από αυτή βασισμένη στην συγχώνευση χαρακτηριστικών γνωρισμάτων. Είναι χαρακτηριστικό επίσης πως η απόδοση του ταξινομητή ήταν 74,6% τόσο στην περίπτωση της καλύτερης πιθανότητας όσο και της πλειοψηφίας, γεγονός που καταδεικνύει πως σπάνια οι αποφάσεις των μεμονωμένων ταξινομητών διαφωνούσαν. Η απόδοση του ταξινομητή αυξάνει μέχρι 85,1% για την πρώτη προσέγγιση και φθάνει 88,20% για τη δεύτερη προσέγγιση όταν μόνο τα δείγματα διαθέσιμα για όλες τις μορφές πληροφορίας χρησιμοποιούνται. Αυτή η μικρή διαφορά στην απόδοση δείχνει πως

ακόμα και σε περιβάλλοντα θορυβώδη ή ανεξέλεγκτα αυτή η αρχιτεκτονική διαθέτει ευρωστία και μπορεί να αποδώσει σε ικανοποιητικά επίπεδα.

2.4.5 Συμπεράσματα

Παρουσιάσαμε ένα πλαίσιο για την ανάλυση και την αναγνώριση συναισθήματος από τις εκφράσεις προσώπου, τις χειρονομίες και την ομιλία. Εκπαιδεύσαμε και επιβεβαιώσαμε την δυνατότητα γενίκευσης του προτύπου με έναν Bayesian ταξινομητή, μέσω ενός συνόλου δεδομένων πολλαπλών μορφών με οκτώ υποδυόμενα συναισθήματα και δέκα θέματα πέντε διαφορετικών υπηκοοτήτων.

Ως σύνολο πειραματισμού χρησιμοποιήσαμε ένα σύνολο δεδομένων 240 δειγμάτων για κάθε μορφή (πρόσωπο, σώμα, ομιλία), εξετάζοντας επίσης και τις περιπτώσεις με πιθανή απώλεια τιμών. Εκτιμήσαμε επίσης ένα πρότυπο που δημιουργήθηκε μη λαμβάνοντας υπόψη τις περιπτώσεις με λιγότερες από τρείς μορφές πληροφορίας. Το πρώτο πρότυπο έλαβε χαμηλότερα ποσοστά αναγνώρισης για τα οκτώ συναισθήματα από το δεύτερο, τόσο στα μονόμορφα συστήματα όσο και στο πολύμορφο σύστημα, αλλά δεν πρέπει να αγνοήσουμε την ιδιότητα του να διαχειρίζεται σύνολα δεδομένων με απώλεια τιμών, περίπτωση αρκετά συχνή σε πραγματικές συνθήκες. Εξετάζοντας τις αποδόσεις των συστημάτων αναγνώρισης συναισθήματος από μοναδική μορφή πληροφορίας, αυτό που βασίζεται στις χειρονομίες εμφανίζεται να είναι επιτυχέστερο, ακολουθούμενο από αυτό της ομιλίας και αυτό των εκφράσεων του προσώπου. Σημειώνουμε πως σε αυτήν την μελέτη χρησιμοποιήσαμε χειρονομίες συγκεκριμένες για κάθε συναίσθημα: αυτές είναι χειρονομίες που επιλέγονται ώστε να αρμόζουν σε κάθε συγκεκριμένο συναίσθημα. Μια εναλλακτική προσέγγιση που μπορεί επίσης να είναι ενδιαφέρουσα θα ήταν να αναγνωρισθεί το συναίσθημα από την διαφορετική εκφραστικότητα της ίδιας χειρονομίας (που δεν θα συνδέεται απαραίτητα με κάποιο συγκεκριμένο συναίσθημα). Αυτό θα επέτρεπε καλύτερη σύγκριση με τα συστήματα βασισμένα στις εκφράσεις του προσώπου και την ομιλία και θα εξεταστεί μελλοντικά. Η συγχώνευση πολύμορφων δεδομένων αύξησε σημαντικά τα ποσοστά αναγνώρισης σε σύγκριση με τα μονόμορφα συστήματα: η πολύμορφη προσέγγιση παρουσίασε βελτίωση μεγαλύτερης του 10% σχετικά με την απόδοση του συστήματος βασισμένου στις χειρονομίες, όταν χρησιμοποιούνται όλα τα 240 δείγματα. Περαιτέρω, η συγχώνευση που εκτελείται σε επίπεδο χαρακτηριστικών γνωρισμάτων παρουσιάζει καλύτερα αποτελέσματα από αυτή που εκτελείται σε επίπεδο απόφασης, που φανερώνει πως η επεξεργασία των δεδομένων εισόδου σε ένα κοινό διάστημα χαρακτηριστικών γνωρισμάτων είναι επιτυχέστερη. Βέβαια κάτι τέτοιο δεν είναι πάντα εφικτό σε συνθήκες πραγματικών καταγραφών καθώς μια ή και περισσότερες μορφές πληροφορίας μπορεί να απουσιάζουν ή να παράγουν χαρακτηριστικά γνωρίσματα χαμηλής εμπιστοσύνης.

2.5 Προσαρμογή νευρωνικών δικτύων στην αναγνώριση συναισθηματικής κατάστασης

Μια αποτελεσματική προσέγγιση παρουσιάζεται εδώ, η οποία χρησιμοποιεί αρχιτεκτονικές νευρωνικών δικτύων για την ανίχνευση της ανάγκης για προσαρμογή της γνώσης που αποκτήθηκε με την αρχική εκπαίδευση και την προσαρμογή της μέσω μιας αποδοτικής διαδικασίας προσαρμογής. Επίσης παρουσιάζεται μια πειραματική μελέτη, με τα συναισθηματικά εμπλουτισμένα σύνολα δεδομένων που παράχθηκαν

στο πλαίσιο του HUMAINE Network of Excellence.

2.5.1 Αρχιτεκτονική προσαρμοστικού νευρωνικού δικτύου

Έστω ότι θέλουμε να κατηγοριοποιήσουμε, σε μία από τις p διαθέσιμες κατηγορίες συναισθήματος ω, κάθε διάνυσμα εισόδου $\overline{x}_i$ που περιέχει χαρακτηριστικά από το σήμα εισόδου. Ένα νευρωνικό δίκτυο παράγει ένα διάνυσμα εξόδου, με διάσταση p, $\overline{y}(\overline{x}_i)$.

$$\overline{y}(\overline{x}_i) = \left[p^i_{\omega_1} p^i_{\omega_2} \dots p^i_{\omega_p}\right]^T \tag{2.2}$$

Όπου $p^i_{\omega_j}$, η πιθανότητα η είσοδος i να ανήκει στην κλάση j. Έστω ότι το δίκτυο έχει εκπαιδευτεί βάσει ενός συνόλου δεδομένων, $S_b = \left\{ \left(\overline{x'}_1,\ \overline{d'}_1\right), \cdots, \left(\overline{x'}_{m_b},\ \overline{d'}_{m_b}\right) \right\}$, όπου τα διανύσματα $\overline{x'}_i$ και $\overline{d'}_i$ με $i = 1, 2, \cdots, m_b$ δηλώνουν το διάνυσμα εισόδου i και το αντίστοιχο επιθυμητό διάνυσμα εξόδου με p στοιχεία.

Έπειτα, έστω $\overline{y}(\overline{x}_i)$ η έξοδος του δικτύου όταν παρουσιαστεί στο δίκτυο ένα νέο σύνολο εισόδων και έστω η είσοδος i μέσα σε αυτό το σύνολο, που πιθανόν ανήκει σε διαφορετικό πρόσωπο ή διαφορετικές συνθήκες από αυτές που επικρατούσαν κατά την αρχική εκπαίδευση. Βασισμένοι στα παραπάνω, ελαφρώς τροποποιημένα βάρη πρέπει να εκτιμηθούν σε τέτοιες περιπτώσεις μέσα από μια διαδικασία προσαρμογής.

Το διάνυσμα $\overline{w}_b$ περιλαμβάνει όλα τα βάρη του δικτύου πριν την εφαρμογή της διαδικασίας προσαρμογής και $\overline{w}_a$ τις νέες τιμές βαρών που διαμορφώνονται μετά την προσαρμογή. Για να εφαρμοστεί η διαδικασία προσαρμογής, ένα σύνολο εκπαίδευσης S_c εξάγεται από την τρέχουσα περίπτωση, που αποτελείται από m_c εισόδους. $S_c = \left\{ \left(\overline{x}_1,\ \overline{d}_1\right), \cdots, \left(\overline{x}_{m_c},\ \overline{d}_{m_c}\right) \right\}$ όπου $\overline{x}_i$ και $\overline{d}_i$ με $i = 1, 2, \cdots, m_c$ ανταποκρίνονται στην i είσοδο και επιθυμητή έξοδο των δεδομένων που χρησιμοποιούνται για προσαρμογή. Ο αλγόριθμος προσαρμογής που ενεργοποιείται, όποτε ανιχνευτεί η ανάγκη για προσαρμογή, υπολογίζει τα βάρη $\overline{w}_a$, ελαχιστοποιώντας τα ακόλουθα κριτήρια λάθους σε σχέση με τα βάρη:

$$E_a = E_{c,a} + \eta E_{f,a}$$

$$E_{c,a} = \frac{1}{2} \sum_{i=1}^{m_c} \left\| \overline{z}_a(\overline{x}_i) - \overline{d}_i \right\|_2$$

$$E_{f,a} = \frac{1}{2} \sum_{i=1}^{m_b} \left\| \overline{z}_a(\overline{x'}_i) - \overline{d'}_i \right\|_2 \tag{2.3}$$

Όπου $E_{c,a}$ είναι το λάθος του δικτύου επί του συνόλου S_c (τρέχουσα γνώση) και το αντίστοιχο λάθος $E_{f,a}$ για το σύνολο S_b (πρότερη γνώση). $\overline{z}_a(\overline{x}_i)$ και $\overline{z}_a(\overline{x}'_i)$ είναι οι έξοδοι του προσαρμοσμένου δικτύου, στα διανύσματα εισόδου $\overline{x}_i$ και $\overline{x}'_i$ αντίστοιχα, με βάρη $\overline{w}_a$. Παρομοίως $\overline{z}_b(\overline{x}_i)$ είναι η έξοδος δικτύου, με βάρη $\overline{w}_b$, στην είδοδος $\overline{x}_i$. Όταν εφαρμόζεται η προσαρμογή για πρώτη φορά η τιμή $\overline{z}_b(\overline{x}_i)$ είναι ίση με $\overline{y}(\overline{x}_i)$. Η παράμετρος η είναι παράγοντας σταθμισμού της σημασίας του τρέχοντος συνόλου εκπαίδευσης συγκριτικά με το πρότερο και $\|\cdot\|_2$ είναι η L_2 νόρμα.

Στόχος της εκπαίδευσης είναι να ελαχιστοποιηθεί το λάθος $E_{f,a}$ και να εκτιμήσει τα νέα βάρη $\overline{w}_a$. Ο αλγόριθμος που υιοθετήθηκε προτάθηκε στο [64]. Υποθέτουμε πως μια μικρή αλλαγή στις τιμές των βαρών συνάψεων $\overline{w}_b$ (πριν την διαδικασία προσαρμογής) είναι αρκετή για να επιτευχθούν ικανοποιητικά αποτελέσματα κατηγοριοποίησης. Τότε,

$$\overline{w}_a = \overline{w}_b + \Delta\overline{w}$$

όπου $\Delta\overline{w}$ είναι μικρές μεταβολές. Βάσει της υπόθεσης αυτής οδηγούμαστε σε μια αναλυτική λύση για την εκτίμηση των $\overline{w}_a$, αφού επιτρέπει την γραμμική λύση της μη γραμμικής συνάρτηση ενεργοποίησης ενός νευρώνα, με την χρήση σειρών Taylor πρώτου βαθμού. Η εξίσωση (2.3) δείχνει ότι τα νέα βάρη του δικτύου εκτιμούνται λαμβάνοντας υπόψη και την τρέχουσα και την προηγούμενη γνώση του δικτύου. Για να τονιστεί, εντούτοις, η σημασία των τρεχόντων δεδομένων εκπαίδευσης, κάποιος μπορεί να αντικαταστήσει τον πρώτο όρο με τον περιορισμό ότι τα πραγματικά αποτελέσματα του δικτύου είναι ίσα με τα επιθυμητά, δηλαδή:

$$z_a(\overline{x}_i) = d_i, i = 1, \cdots, m_c, \forall\overline{x} \in S_c \tag{2.4}$$

Η γραμμική λύση της εξίσωσης (2.4) σε σχέση με τις αλλαγές βάρους είναι ισοδύναμη με ένα σύνολο γραμμικών εξισώσεων:

$$\overline{c} = A \cdot \Delta\overline{w} \tag{2.5}$$

όπου $\overline{c}$ και A εκφράζονται σε σχέση με τα προηγούμενα βάρη. Συγκεκριμένα:

$$\overline{c} = [d_1 \cdots d_{m_c}]^T - [z_b(\overline{x}_1) \cdots z_b(\overline{x}_{m_c})]^T \tag{2.6}$$

Επιπλέον, η ελαχιστοποίηση του δεύτερου όρου στην (2.3), ο οποίος εκφράζει την επίδραση των νέων βαρών δικτύου στο σύνολο στοιχείων S_b, μπορεί να θεωρηθεί ως ελαχιστοποίηση της απόλυτης διαφοράς του λάθους πέρα από τα στοιχεία S_b με τα προηγούμενα και τρέχοντα βάρη δικτύων. Αυτό σημαίνει ότι οι αυξήσεις βαρών τροποποιούνται ελάχιστα, σύμφωνα μετά το ακόλουθο κριτήριο λάθους:

$$E_S = \|E_{f,a} - E_{f,b}\|_2 \tag{2.7}$$

Το $E_{f,b}$ ορίζεται παρόμοια με το $E_{f,a}$ με το $\overline{z}_a$ να αντικαθίσταται από το $\overline{z}_b$ στην (2.3). Δείχνεται [184] ότι η (2.7) παίρνει την μορφή:

$$E_S = \frac{1}{2}(\Delta\overline{w})^T \cdot K^T \cdot K \cdot \Delta\overline{w} \tag{2.8}$$

όπου τα στοιχεία του πίνακα K εκφράζονται με όρους των προηγούμενων βαρών w_b του δικτύου και των δεδομένων εκπαίδευσης στο S_b. Η συνάρτηση λάθους που ορίζεται από την (2.8) είναι κυρτή δεδομένου ότι είναι τετραγωνικής μορφής. Κατά συνέπεια, οι μεταβολές των βαρών μπορούν να υπολογιστούν με την λύση της (2.8). Η μέθοδος προβολής κλίσης χρησιμοποιήθηκε στο [64] για να υπολογίσει τις μεταβολές των βαρών.

Κάθε φορά που εξακριβώνει ο μηχανισμός απόφασης ότι απαιτείται προσαρμογή, ένα νέο σύνολο εκπαίδευσης S_c δημιουργείται, το οποίο αντιπροσωπεύει τις τρέχουσες συνθήκες. Κατόπιν, τα νέα βάρη δικτύων υπολογίζονται λαμβάνοντας υπόψη τόσο την τρέχουσα όσο και την πρότερη γνώση. Αφού το σύνολο το S_c έχει βελτιστοποιηθεί μόνο για την τρέχουσα κατάσταση, δεν μπορεί να θεωρηθεί κατάλληλο για μελλοντικές ακολουθίες ή καταστάσεις του περιβάλλοντος. Αυτό οφείλεται στο γεγονός ότι στοιχεία που λαμβάνονται από μελλοντικές καταστάσεις του περιβάλλοντος μπορεί να έρθουν σε σύγκρουση με στοιχεία που λαμβάνονται από το παρόν. Αντίθετα, υποθέτουμε ότι το σύνολο εκπαίδευσης S_b, που γενικά είναι βασισμένο σε

εκτενή πειραματισμό, είναι σε θέση να προσεγγίσει την επιθυμητή έξοδο του δικτύου σε οποιαδήποτε πιθανή μελλοντική κατάσταση του περιβάλλοντος. Συνεπώς, σε κάθε φάση προσαρμογής δικτύου, ένα νέο σύνολο εκπαίδευσης S_c δημιουργείται και το προηγούμενο απορρίπτεται, ενώ τα νέα βάρη υπολογίζονται βάσει της τρέχουσας και της παλαιότερης γνώσης, που παραμένει σταθερή καθ'όλη τη λειτουργία του δικτύου.

2.5.2 Εντοπίζοντας την ανάγκη για προσαρμογή

Ο σκοπός του μηχανισμού αυτού είναι να ανιχνεύσει την ακαταλληλότητα της εξόδου του νευρωνικού δικτύου ταξινόμησης και συνεπώς να ενεργοποιήσει τον αλγόριθμο προσαρμογής σε εκείνα τα χρονικά στιγμιότυπα που εντοπίζεται η μεταβολή των παραγόντων του περιβάλλοντος.

Ας υποθέσουμε αρχικά ότι μια προσαρμογή δικτύου έχει πραγματοποιηθεί και ας εστιάσουμε στα οπτικά γνωρίσματα εισόδου. Έστω $\overline{x}(k)$ το διάνυσμα χαρακτηριστικών γνωρισμάτων της k εικόνας ή πλαισίου, μετά από το χρόνο που εφαρμόστηκε η προσαρμογή. Ο δείκτης k επομένως επαναρχικοποιείται κάθε φορά που εφαρμόζεται η προσαρμογή, με $\overline{x}(0)$ το αντίστοιχο διάνυσμα χαρακτηριστικών γνωρισμάτων της εικόνας όπου έγινε η προσαρμογή του δικτύου. Στην είσοδο αυτή, η απόδοση του δικτύου επιδεινώθηκε, δηλ., η έξοδος του δικτύου παρέκκλινε από την επιθυμητή πέρα ενός αποδεκτού ορίου. Το διάνυσμα $\overline{c}$ στην εξίσωση (2.6) εκφράζει την διαφορά μεταξύ των επιθυμητών και των πραγματικών εξόδων του δικτύου βάσει των βαρών $\overline{w}_b$ και εφαρμόζεται στο τρέχον σύνολο δεδομένων. Κατά συνέπεια, εάν το μέτρο του διανύσματος $\overline{c}$ αυξηθεί, η απόδοση του δικτύου παρεκκλίνει από την επιθυμητή έξοδο και το δίκτυο πρέπει να προσαρμοστεί. Αντιθέτως, εάν το διάνυσμα $\overline{c}$ λαμβάνει μικρές τιμές, δεν απαιτείται προσαρμογή. Ακολούθως χρησιμοποιούμε τη διαφορά μεταξύ της εξόδου του προσαρμοσμένου δικτύου και της εξόδου του αρχικά εκπαιδευμένου ταξινομητή για να προσεγγίσουμε την τιμή του $\overline{c}$. Επιπλέον, υποθέτουμε ότι η διαφορά που υπολογίζεται κατά την επεξεργασία της εισόδου $\overline{x}(0)$ αποτελεί μια καλή εκτίμηση του επιπέδου βελτίωσης που μπορεί να επιτευχθεί από τη διαδικασία προσαρμογής. Έστω $e(0)$ αυτή η διαφορά και $e(k)$ η διαφορά μεταξύ των αντίστοιχων εξόδων των δύο ταξινομητών, όταν σε αυτά παρουσιάζεται το $\overline{x}(k)$. Αναμένεται ότι το επίπεδο βελτίωσης εκφρασμένο ως $e(k)$ θα είναι κοντά στο $e(0)$ αν τα αποτελέσματα της ταξινόμησης είναι καλά. Αυτό προϋποθέτει ότι οι εικόνες εισόδου είναι παρόμοιες με αυτές που χρησιμοποιούνται κατά τη διάρκεια της φάσης προσαρμογής. Ένα λάθος $e(k)$, αρκετά διαφορετικό από το $e(0)$, οφείλεται γενικά στην αλλαγή περιβάλλοντος. Κατά συνέπεια, η ποσότητα $a(k) = |e(k) - e(0)|$ μπορεί να χρησιμοποιηθεί για την ανίχνευση της αλλαγής του περιβάλλοντος ή ισοδύναμα τα χρονικά σημεία όπου η προσαρμογή είναι απαραίτητη. Κατά συνέπεια, καμία προσαρμογή δεν απαιτείται εάν:

$$a(k) < T \tag{2.9}$$

όπου T είναι ένα κατώφλι που εκφράζει την μέγιστη ανοχή, πέραν της οποίας είναι απαραίτητη η προσαρμογή για την βελτίωση της απόδοσης του δικτύου. Μια τέτοια προσέγγιση ανιχνεύει με ακρίβεια τις χρονικές στιγμές ανάγκης προσαρμογής σε περιπτώσεις απότομων αλλά και βαθμιαίων αλλαγών του περιβάλλοντος δεδομένου ότι η σύγκριση εκτελείται μεταξύ της τρέχουσας διαφοράς λάθους $e(k)$ και αυτής που λαμβάνεται αμέσως μετά από την προσαρμογή, δηλ.$e(0)$. Στην περίπτωση απότομης λειτουργικής αλλαγής, το λάθος $e(k)$ δεν θα είναι κοντά στο $e(0)$ και συνεπώς, το $a(k)$ θα υπερβαίνει το κατώφλι T και η προσαρμογή ενεργοποιείται. Σε περίπτωση που

συμβαίνει μια βαθμιαία αλλαγή, το λάθος $e(k)$ θα παρεκκλίνει βαθμιαία αλλά σταθερά από το $e(0)$ έτσι ώστε η ποσότητα $a(k)$ να αυξάνεται βαθμιαία και η προσαρμογή να ενεργοποιηθεί στο πλαίσιο που $a(k) > T$.

Η προσαρμογή δικτύων μπορεί να εκτελεσθεί στιγμιαία κάθε φορά που το σύστημα τίθεται σε λειτουργία από τον χρήστη. Κατά συνέπεια, η ποσότητα $a(0)$ υπερβαίνει αρχικά το κατώφλι T και η προσαρμογή αναγκάζεται να πραγματοποιηθεί.

2.5.3 Πειραματική μελέτη

2.5.3.1 Σώμα πειραμάτων

Στην παρακάτω ενότητα παρουσιάζονται αποτελεσμάτα εκτενούς πειραματισμού, βασισμένου στον ανωτέρω περιγεγραμμένο προσαρμοστικό νευρωνικό δίκτυο. Δεδομένου ότι στόχος αυτής της εργασίας είναι να υπογραμμίσει την δυνατότητα να ταξινόμησης ακολουθιών με φυσιοκρατικές εκφράσεις, επιλέξαμε να χρησιμοποιήσουμε την SAL βάση δεδομένων για εκπαίδευση και δοκιμές γενίκευσης [114].

Ένα σημείο που χρήζει εξέτασης στη φυσική ανθρώπινη αλληλεπίδραση είναι ότι ο χαρακτήρας κάθε ατόμου διαδραματίζει σημαντικό ρόλο στην συναισθηματική κατάσταση του ανθρώπου. Διαφορετικά άτομα μπορεί να έχουν διαφορετικές συναισθηματικές αποκρίσεις σε παρόμοια ερεθίσματα. Επομένως, η επισημείωση των καταγραφών δεν πρέπει να βασιστεί στο σχεδιασμένο ή προτιθέμενο συναίσθημα αλλά στο πραγματικό αποτέλεσμα της αλληλεπίδρασης με το SAL.

Ο συναισθηματικός χώρος αντιπροσωπεύεται από έναν κύκλο στην οθόνη, που χωρίζεται σε τέσσερα τεταρτημόρια από δύο κύριους άξονες. Ο κάθετος άξονας αντιπροσωπεύει την ενεργοποίηση, διατρέχει το διάστημα από πολύ ενεργό σε πολύ παθητικό και τον οριζόντιο άξονα που αντιπροσωπεύει την αξιολόγηση, που διατρέχει το διάστημα από πολύ θετικό σε πολύ αρνητικό. Απεικονίζει την δημοφιλή άποψη ότι ο συναισθηματικός χώρος είναι κατά προσέγγιση κυκλικός. Το κέντρο του κύκλου χαρακτηρίζει την ουδέτερη κατάσταση, που είναι και αρχικά προεπιλεγμένη και η τοποθέτηση του δρομέα σε αυτή την περιοχή μαρτυρά ότι δεν υπάρχει κάποιο πραγματικό εκφραζόμενο συναίσθημα. Ο χρήστης μετακινεί το ποντίκι στον κυκλικό συναισθηματικού χώρο, έτσι ώστε κάθε στιγμή η θέση του να επισημαίνει τα αντιληπτά επίπεδα ενεργοποίησης και αξιολόγησης και το σύστημα καταγράφει αυτόματα τις συντεταγμένες του δρομέα οποιαδήποτε στιγμή, ώστε να υπάρχει αντιστοίχηση και προς τις δυο κατευθύνσεις.

Οι συντεταγμένες των μετακινήσεων του ποντικιού στο δισδιάστατο γραφικό περιβάλλον του χρήστη αντιστοιχείται στις πέντε συναισθηματικές κατηγορίες που παρουσιάζονται στον πίνακα 2.23. Εφαρμόζοντας μιας τυπική μέθοδο ανίχνευσης μικρών διακοπών στο κανάλι ακουστικής πληροφορίας των καταγραφών, η βάση δεδομένων χωρίστηκε σε 477 τόνους, με τα μήκη τους να κυμαίνονται από 1 εως 174 πλαίσια. Μια προκατάληψη υπάρχει προς το Q1 στη βάση δεδομένων, καθώς 42,98% από τους τόνους είναι ταξινομημένοι ως Q1, όπως φαίνεται στον πίνακα 2.24. Ο πίνακας 2.25 παρουσιάζει τέσσερα διαφορετικά θέματα κατά την επίδειξη συναισθημάτων από ποικίλους τόνους και τεταρτημόρια.

2.5.3.2 Πειράματα

Τα πειράματά μας στοχεύουν στην διερεύνηση της πρακτικής ορθότητας της προτεινόμενης διαδικασίας προσαρμογής. Η κύρια ιδέα της πειραματικής μελέτης είναι να

διερευνήσει την απόδοση των προσαρμοσμένων δικτύων σε εισόδους που ανήκουν στο ίδιο τόνο, αλλά δεν χρησιμοποιήθηκαν για την προσαρμογή, καθώς επίσης και σε τόνους του ίδιου συναισθηματικού τεταρτημόριου με αυτό που χρησιμοποιήθηκε για λόγους προσαρμογής.

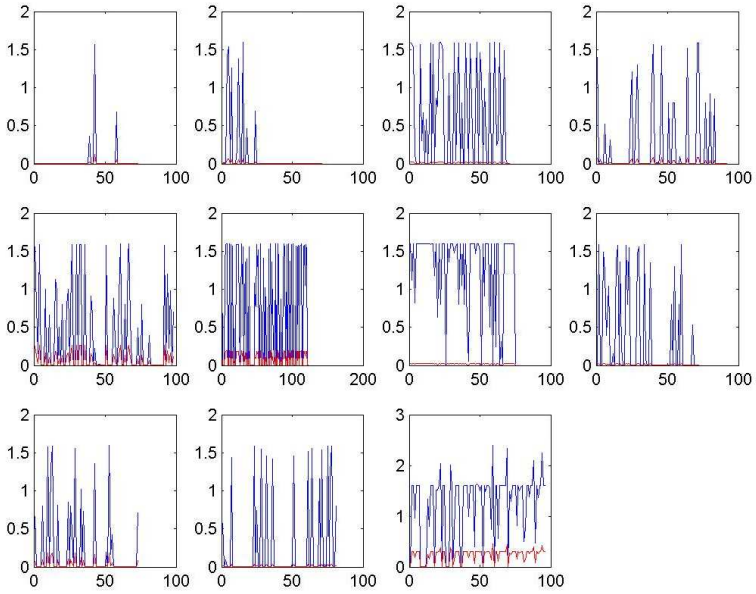
Από ένα σύνολο περίπου 35.000 πλαισίων, από 477 τόνους της SAL βάσης δεδομένων, επιλέξαμε μόλις 500 πλαίσια, και από τα τέσσερα θέματα, για λόγους εκπαίδευσης ενός εμπροσθοτροφοδοτούμενου δικτύου οπισθοδιαδοσης (feed-forward back-propagation network) που θα αναφέρεται στο εξής ως NetProm. Οι λεπτομέρειες αρχιτεκτονικής για το NetProm είναι: τρία επίπεδα, που αποτελούνται από 10 και 5 νευρώνες στο πρώτο και δεύτερο κρυμμένο επίπεδο αντίστοιχα και 5 νευρώνες του επιπέδου εξόδου. Τα διανύσματα στόχου (επιθυμητής εξόδου) σχηματοποιήθηκαν ως διανύσματα 5x1 για κάθε πλαίσιο ώστε μόνο μια, από τις 5 υποψήφιες κατηγορίες, να ισούται με την μονάδα, ενώ όλες οι υπόλοιπες να είναι μηδέν. Έτσι παραδείγματος χάριν εάν το πλαίσιο άνηκε στο πρώτο τεταρτημόριο, το διάνυσμα εξόδου θα ήταν [1 0 0 0 0]. Η πέμπτη κατηγορία του προβλήματος ταξινόμησης αντιστοιχεί στην ουδέτερη συναισθηματική κατάσταση και οι άλλες τέσσερις στα τέσσερα τεταρτημόρια του κύκλου του Whissel.

Η επιλογή των 500 πλαισίων που χρησιμοποιήθηκαν για την κατάρτιση του δικτύου NetProm έγινε βάσει ένος κριτηρίου χαρακτηριστικότητας. Πιό συγκεκριμένα, για κάθε πλαίσιο, ορίστηκε μια μετρική που χαρακτήριζε την απόσταση των τιμών FAPs του συγκεκριμένου πλαισίου σε σχέση με τις μέσες τιμές FAPs άλλων πλαισίων της ίδιας κατηγορίας. Αυτή η μετρική της διακύμανσης FAP ήταν η παράμετρος ταξινόμησης για τα πλαίσια. Με τον περιορισμό ότι κάθε κατηγορία πρέπει να εκπροσωπηθεί όσο το δυνατόν εξίσου, επιλέξαμε τα 500 πιό προεξέχοντα πλαίσια και τα χρησιμοποιήσαμε ως είσοδο για την εκπαίδευση του δικτύου NetProm.

Όσον αφορά στη φάση προσαρμογής επιλέξαμε έντεκα τόνους που αποτελούνται από τον μεγαλύτερο αριθμό πλαισίων. Αυτή η επιλογή βασίστηκε στην ιδέα ότι δεν θα είχε πολύ νόημα να επιλέγαμε πολύ σύντομους τόνους, επειδή τα δεδομένα προσαρμογής θα ήταν πολύ λίγα, όπως θα εξηγηθεί καλύτερα αργότερα. Επίσης ελέγξαμε ότι κανένα από τα πλαίσια αυτών των έντεκα τόνων δεν χρησιμοποιήθηκε για την εκπαίδευση του NetProm. Κάθε ένας από τους έντεκα τόνους χωρίστηκε σε δύο μέρη, το υποσύνολο προσαρμογής και το υποσύνολο εξέτασης που περιέχουν 30% και 70% αντίστοιχα του συνολικού αριθμού πλαισίων του τόνου.

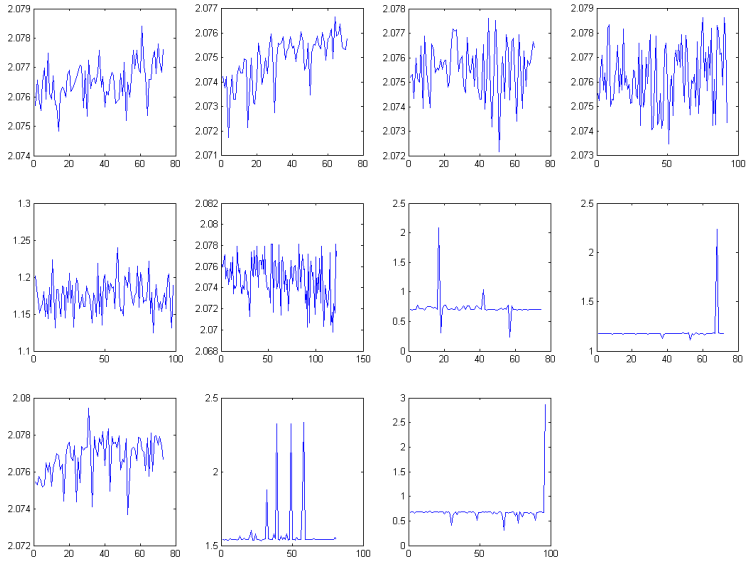
Το NetProm προσαρμόστηκε χρησιμοποιώντας υποσύνολα προσαρμογής των έντεκα τόνων και προέκυψαν έντεκα νέα δίκτυα Net_i $i = 1..11$. Κάθε Net_i εξετάστηκε έπειτα με το υποσύνολο εξέτασης του αντίστοιχου τόνου και τα αποτελέσματα μπορούν να φανούν στην εικόνα 2.30. Είναι σαφές ότι η διαδικασία προσαρμογής λειτουργήσε ευεργετικά και μείωσε κατά πολύ το MSE σε κάθε τόνο που εφαρμόστηκε.

Επιπλέον, εξετάσαμε την διαδικασία ανίχνευσης της ανάγκης για προσαρμογή που περιγράφηκε νωρίτερα. Συγκεκριμένα χρησιμοποιήσαμε τα έντεκα δίκτυα που προέκυψαν και συγκρίναμε την απόδοση τους με αυτή του δικτύου NetProm βάσει του κριτηρίου που περιγράφεται στην εξίσωση 2.9 και τα αποτελέσματα φαίνονται στην εικόνα 2.31. Στα πρώτα έξι πειράματα και το ένατο δεν υπήρχε διαφορά στο εκφρασμένο συναίσθημα. Μπορεί να δειχθεί πως οι τιμές του $e(k)$ σ'αυτές τις περιπτώσεις είναι κοντά σε αυτές του $e(0)$ και έτσι δεν υφίσταται ανάγκη για προσαρμογή. Αντίθετα στις περιπτώσεις 7,8,10 και 11 περιείχαν ένα ή περισσότερα πλαίσια από διαφορετικά θέματα που έδειχναν παρόμοιο συναίσθημα. Στις περισσότερες από αυτές τις περιπτώσεις οι τιμές του $\alpha(k)$ ήταν αυξημένες εξαιτίας της αδυναμίας του εκπαιδευμένου



Σχήμα 2.30: Μέσο τετραγωνικό λάθος του NetProm (μπλέ) and Neti (κόκκινο)

δικτύου να προσαρμοστεί στις αλλαγές περιβάλλοντος, δηλαδή την αλλαγή θέματος. Συνεπώς, εντοπίζεται η ανάγκη για προσαρμογή μέσω του κριτηρίου της εξίσωσης 2.9 και τα αποτελέσματα είναι πολύ ενθαρρυντικά δείχνοντας ότι η προτεινόμενη διαδικασία αποτελεί αποτελεσματικό προσαρμοστικό εργαλείο στην αναγνώριση συναισθήματος/έκφρασης.



Σχήμα 2.31: Ανιχνεύοντας την ανάγκη για προσαρμογή χρησιμοποιώντας το κριτήριο της εξίσωσης 2.9

2.5.4 Συμπεράσματα

Η αναγνώριση των εκφράσεων του προσώπου και των χειρονομιών χεριών διαδραματίζουν πολύ σημαντικό ρόλο στην προσαρμογή της αλληλεπίδρασης ανθρώπου-υπολογιστή στις ανάγκες και της ανατροφοδότησης από τους χρήστες, ειδικά αφού ψυχολογικές έρευνες έχουν δείξει ότι το πρόσωπο είναι ζωτικής σημασίας συστατικό της ανθρώπινης εκφραστικότητας. Εντούτοις, στην καθημερινή αλληλεπίδραση ανθρώπου-υπολογιστή, τα συναισθήματα είναι συνήθως διακριτικά και όχι ακραία, ως εκ τούτου είναι δύσκολη η χρησιμοποίηση ενός μικρού συνόλου καθολικών κατηγοριών. Για να αντιμετωπιστεί αυτό, κάποιος πρέπει να θεωρήσει τις πολλαπλές μορφές πληροφορίας σαν λύση ενίσχυσης. Κάτι τέτοιο μπορεί να επιτευχθεί με ενδείξεις αποτυχίας ή σημαντική μείωση της απόδοσης αναγνώρισης από ένα κανάλι πληροφορίας. Πέρα από αυτό, η εξατομικευμένη εκφραστικότητα και η εξάρτηση από το πλαίσιο κάνουν τη γενίκευση των τεχνικές μηχανικής μάθησης εξαιρετικά δύσκολη εργασία. Σε αυτή την εργασία προτείναμε μια επέκταση στην διαδικασία προσαρμογής νευρωνικών δικτύων. Κατόπιν της εκπαίδευσης σε ένα συγκεκριμένο θέμα, το δίκτυο προσαρμόζεται χρησιμοποιώντας προεξέχοντα δείγματα από στιγμιότυπα με άλλο θέμα, έτσι ώστε να προσαρμοστεί και να βελτιώσει τη δυνατότητά γενίκευσης του. Τα αποτελέσματα που παρουσιάζονται εδώ δείχνουν ότι η απόδοση του δικτύου βελτιώνεται χρησιμοποιώντας αυτήν την προσέγγιση, χωρίς την ανάγκη να εκπαιδευθεί ένα δίκτυο για κάθε θέμα ή να επανεκπαιδευτεί σε εκτενές σύνολο δεδομένων, το οποίο θα εξάλειφε την σημαντική ιδιότητα γενίκευσης του δικτύου.

Πίνακας 2.16: Περιγραφή των 10 χαρακτηριστικότερων γνωρισμάτων ανά μορφή πληροφορίας που προέκυψαν από την διαδικασία επιλογής γνωρισμάτων για μονόμορφη αναγνώριση

Maximum-QoM	χειρονομία	ποσότητα κίνησης
Mean-QoM	χειρονομία	ποσότητα κίνησης
MeanMax-QoM	χειρονομία	ποσότητα κίνησης
MaxFollMax-QoM	χειρονομία	ποσότητα κίνησης
InSlope-CI	χειρονομία	δείκτης συστολής
NPeaks-CI	χειρονομία	δείκτης συστολής
MeanMax-CI	χειρονομία	δείκτης συστολής
MaxCentroid-CI	χειρονομία	δείκτης συστολής
PeakDurGestDur-CI	χειρονομία	δείκτης συστολής
FinalSlope-VEL	χειρονομία	ταχύτητα
Pitch-min	ήχος	βήμα
Pitch-p2c1	ήχος	βήμα
Pitch-q875	ήχος	βήμα
pv-skew	ήχος	φωνητικό τμήμα
Pause-p2c1	ήχος	παύση
Segment-max	ήχος	διάρκεια τμήματος
Segment-mean	ήχος	διάρκεια τμήματος
Segment-kurt	ήχος	διάρκεια τμήματος
mean-mfcc-06	ήχος	MFCC
mean-mfcc-10	ήχος	MFCC
open-jaw-p2c1	πρόσωπο	κατακόρυφη μετακίνηση σαγονιού
open-jaw-q975	πρόσωπο	κατακόρυφη μετακίνηση σαγονιού
open-jaw-q90	πρόσωπο	κατακόρυφη μετακίνηση σαγονιού
lower-top-midlip-range	πρόσωπο	κατακόρυφη μετακίνηση σαγονιού άνω κεντρικού χείλους
raise-bottom-midlip-extdt	πρόσωπο	κατακόρυφη μετακίνηση σαγονιού κάτω κεντρικού χείλους
widening-mouth-range	πρόσωπο	οριζόντια μετακίνηση έσω άκρων των χείλων
widening-mouth-std	πρόσωπο	οριζόντια μετακίνηση έσω άκρων των χείλων
widening-mouth-kurt	πρόσωπο	οριζόντια μετακίνηση έσω άκρων των χείλων
widening-mouth-range2	πρόσωπο	οριζόντια μετακίνηση έσω άκρων των χείλων
close-left-eye-max	πρόσωπο	κατακόρυφη μετακίνηση άνω και κάτω αριστερού βλεφάρου

Πίνακας 2.17: Επιλεγμένα χαρακτηριστικά γνωρίσματα για κατηγοριοποίηση από πολλαπλές μορφές πληροφορίας

Symmetry-QoM	χειρονομία	Ποσότητα κίνησης	Συμμετρία
MeanMax-CI	χειρονομία	Δείκτης συστολής	Αναλογία μέσου και μεγίστου
FinalSlope-VEL	χειρονομία	Ταχύτητα	Τελική κλίση
FinalSlope-ACC	χειρονομία	Επιτάχυνση	Τελική κλίση
FinalSlope-FL	χειρονομία	Ρευστότητα	Τελική κλίση
Intens-mad0	ήχος	Ένταση	MAD
Pitch-p1c2	ήχος	Βήμα	Παλινδρόμηση πρώτης τάξης
Pitch-p2c1	ήχος	Βήμα	Δεύτερης τάξης
Pitch-range2	ήχος	Βήμα	Ενδοτεταρτημοριακή απόσταση
pv-p1c2	ήχος	Φωνητικό τμήμα	Παλινδρόμηση πρώτης τάξης
Pause-tmax	ήχος	Παύση	Χρόνος μεγίστου
Segmt-tmax	ήχος	Τμήμα	Χρόνος μεγίστου
BarkTL-sgxt	ήχος	Φάσμα Bark	Χρονική γραμμή
BarkSL-kurt	ήχος	Φάσμα Bark	Κύρτωση φασματικής γραμμής
open-jaw-range	πρόσωπο	Κατακόρυφη μετακίνηση σαγονιού	Περιοχή τιμών
widening-mouth-kurt	πρόσωπο	Οριζόντια μετακίνηση έσω άκρων των χείλων	Κύρτωση

Πίνακας 2.18: Πίνακας σύγχυσης της αναγνώρισης συναισθήματος βασισμένης στην ανάλυση της έκφρασης του προσώπου

a	**b**	**c**	**d**	**e**	**f**	**g**	**h**		
56.67	3.33	3.33	10	6.67	10	6.67	3.33	**a**	**Θυμός**
10	**40**	13.33	10	0	13.33	3.33	10	**b**	**Απόγνωση**
6.67	3.33	**50**	6.67	6.67	10	16.67	0	**c**	**Ενδιαφέρον**
10	6.67	10	**53.33**	3.33	6.67	3.33	6.67	**d**	**Ενόχληση**
3.33	0	13.33	16.67	**53.33**	10	0	3.33	**e**	**Χαρά**
6.67	13.33	6.67	0	6.67	**53.33**	13.33	0	**f**	**Ευχαρίστηση**
6.67	3.33	16.67	6.67	13.33	20	**33.33**	0	**g**	**Περηφάνια**
3.33	6.67	3.33	20	0	13.33	6.67	**46.67**	**h**	**Λύπη**

Πίνακας 2.19: Πίνακας σύγχυσης της αναγνώρισης συναισθήματος βασισμένης στην ανάλυση της έκφρασης χειρονομιών

a	**b**	**c**	**d**	**e**	**f**	**g**	**h**		
80	10	0	3.33	0	0	6.67	0	**a**	**Θυμός**
3.33	**56.67**	6.67	0	0	0	26.67	6.67	**b**	**Απόγνωση**
3.33	0	**56.67**	0	6.67	6.67	26.67	0	**c**	**Ενδιαφέρον**
0	10	0	**63.33**	0	0	26.67	0	**d**	**Ενόχληση**
0	10	0	6.67	**60**	0	23.33	0	**e**	**Χαρά**
0	6.67	3.33	0	0	**66.67**	23.33	0	**f**	**Ευχαρίστηση**
0	0	0	3.33	0	0	**96.67**	0	**g**	**Περηφάνια**
0	3.33	0	3.33	0	0	36.67	**56.67**	**h**	**Λύπη**

Πίνακας 2.20: Πίνακας σύγχυσης της αναγνώρισης συναισθήματος βασισμένης στην ανάλυση της ομιλίας

a	**b**	**c**	**d**	**e**	**f**	**g**	**h**		
93.33	0	3.33	3.33	0	0	0	0	**a**	**Θυμός**
10	**23.33**	16.67	6.67	3.33	23.33	3.33	13.33	**b**	**Απόγνωση**
6.67	0	**60**	10	0	16.67	3.33	3.33	**c**	**Ενδιαφέρον**
13.33	3.33	10	**50**	3.33	3.33	13.33	3.33	**d**	**Ενόχληση**
20	0	10	13.33	**43.33**	10	3.33	0	**e**	**Χαρά**
3.33	6.67	6.67	6.67	0	**53.33**	6.67	16.67	**f**	**Ευχαρίστηση**
3.33	10	3.33	13.33	0	13.33	**56.67**	0	**g**	**Περηφάνια**
0	6.67	3.33	10	0	3.33	0	**76.67**	**h**	**Λύπη**

Πίνακας 2.21: Πίνακας σύγχυσης της πολύμορφης αναγνώρισης συναισθήματος

a	**b**	**c**	**d**	**e**	**f**	**g**	**h**		
90	0	0	0	10	0	0	0	**a**	**Θυμός**
0	**53.33**	3.33	16.67	6.67	0	10	10	**b**	**Απόγνωση**
6.67	0	**73.33**	13.33	0	3.33	3.33	0	**c**	**Ενδιαφέρον**
0	6.67	0	**76.67**	6.67	3.33	0	6.67	**d**	**Ενόχληση**
0	0	0	0	**93.33**	0	6.67	0	**e**	**Χαρά**
0	3.33	3.33	13.33	3.33	**70**	6.67	0	**f**	**Ευχαρίστηση**
3.33	3.33	0	3.33	0	0	**86.67**	3.33	**g**	**Περηφάνια**
0	0	0	16.67	0	0	0	**83.33**	**h**	**Λύπη**

Πίνακας 2.22: Ολοκλήρωση σε επίπεδο απόφασης με την μέθοδο της βέλτιστης πιθανότητας

a	**b**	**c**	**d**	**e**	**f**	**g**	**h**		
96,67	0	0	0	0	0	3,33	0	**a**	**Θυμός**
13,33	**53,33**	6,67	0	0	3,33	13,33	10	**b**	**Απόγνωση**
3,33	0	**60**	3,33	10	13,33	6,67	3,33	**c**	**Ενδιαφέρον**
13,33	6,67	6,67	**60**	0	3,33	0	10	**d**	**Ενόχληση**
0	0	10	3,33	**86,67**	0	0	0	**e**	**Χαρά**
6,67	3,33	0	0	0	**80**	6,67	3,33	**f**	**Ευχαρίστηση**
3,33	0	6,67	0	0	10	**80**	0	**g**	**Περηφάνια**
3,33	3,33	0	10	0	3,33	0	**80**	**h**	**Λύπη**

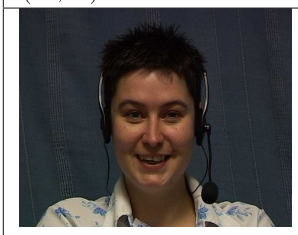
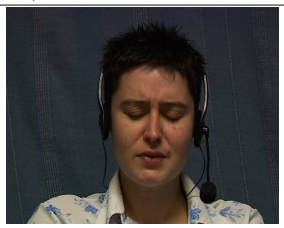
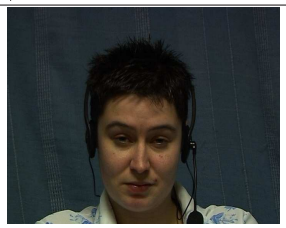
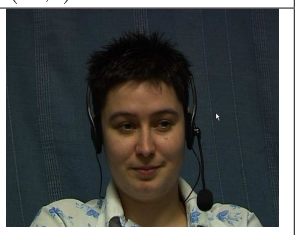
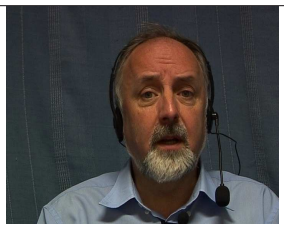
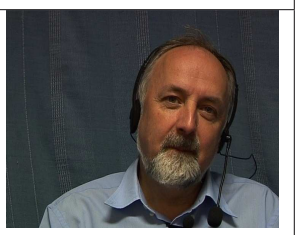
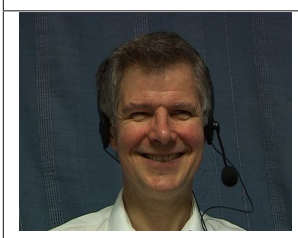
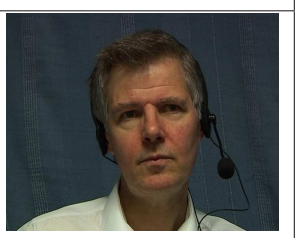
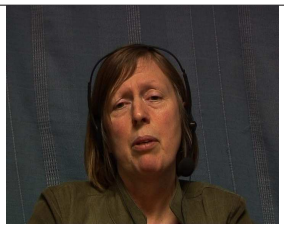
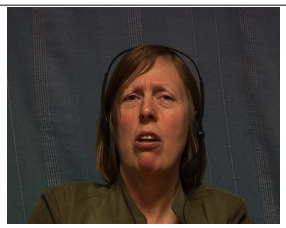
Πίνακας 2.23: Κατηγορίες συναισθήματος

Κατηγορία	**Θέση στο FeelTrace [49] διάγραμμα**
Q1	θετική ενεργοποίηση, θετική αξιολόγηση (+/+)
Q2	θετική ενεργοποίηση, αρνητική αξιολόγηση (+/-)
Q3	αρνητική ενεργοποίηση, αρνητική αξιολόγηση (-/-)
Q4	αρνητική ενεργοποίηση, θετική αξιολόγηση (-/+)
Ουδέτερο	κοντά στο κέντρο

Πίνακας 2.24: Κατανομή των συναισθηματικών κατηγοριών στην βάση δεδομένων SAL

	Ουδέτερο	**Q1**	**Q2**	**Q3**	**Q4**	**Σύνολο**
Τόνοι	47	205	90	63	72	477
Ποσοστά	9,85%	42,98%	18,87%	13,21%	15,09%	100,00%

Πίνακας 2.25: Κατανομή κατηγοριών στην βάση SAL

1^o τεταρτημόριο (+,+)	2^o τεταρτημόριο (-, +)	3^o τεταρτημόριο (-, -)	4^o τεταρτημόριο (+,-)
			
			
			
			

Κεφάλαιο 3

Εκφραστική και πολυμεσική ανάλυση και σύνθεση σε εικονικούς πράκτορες

3.1 Ερευνητικό πλαίσιο

Πληθώρα ερευνών από το επιστημονικό πεδίο της ψυχολογίας και της γνωστικής επιστήμης σχετικής με την συμπεριφορά και την μη λεκτική επικοινωνία αναδεικνύουν την σημασία των εκφραστικών ποιοτικών χαρακτηριστικών των μετακινήσεων του σώματος και των χειρονομιών κατά την αλληλεπίδραση ανθρώπων. Σχετικές μελέτες έχουν πραγματοποιηθεί στον χώρο της σύνθεσης εικονικών χαρακτήρων και πρακτόρων αλλά το επίπεδο της ερεύνας στον χώρο της ανάλυσης συναισθηματικής συμπεριφοράς υπό αυτό το πρίσμα είναι αρκετά χαμηλό και περιορίζεται στην ποιοτική μελέτη και όχι την υπολογιστική ανάλυση εκφραστικά εμπλουτισμένων χειρονομιών. Η δυνατότητα των αληθοφανών εικονικών πρακτόρων να παρέχουν εκφραστική ανατροφοδότηση στον χρήστη είναι μια σημαντική πτυχή ώστε να υποστηρίξουν τη φυσικότητα της αλληλεπίδρασης τους. Τόσο η ανάλυση όσο και η σύνθεση των ενδείξεων από πολλαπλές μορφές πληροφορίας αποτελούν σημαντικές πτυχές της επικοινωνίας ανθρώπου-μηχανής. Η πολυμεσική ανατροφοδότηση επηρεάζει την αληθοφάνεια της συμπεριφοράς ενός πράκτορα ως προς τον ανθρώπινο χρήστη και ενισχύει την επικοινωνιακή του εμπειρία. Κατά γενική ομολογία, η μιμητικότητα είναι ένα αναπόσπαστο, αν και συχνά ασυναίσθητο, μέρος της αλληλεπίδρασης ανθρώπου-ανθρώπου [140] [33]. Στο πλαίσιο αυτό, μπορεί να οριστεί ένας 'βρόχος', όπου η σύνθεση οδηγεί στην ασυνείδητη μίμηση της στάσης σώματος, των χειρονομίες ακόμη και των εκφράσεων προσώπου, του έτερου συμβαλλόμενου μέρους, ο οποίος συμβάλει στην βελτιώση της αλληλεπίδρασης των δύο μερών [231]. Επεκτείνοντας την ιδέα αυτή στο πεδίο της επικοινωνίας ανθρώπου-υπολογιστή, είναι ασφαλές να περιμένει κάποιος ότι χρήστες που αλληλεπιδρούν με μια διεπαφή, τύπου Ενσαρκωμένου Πράκτορα Συνομιλητή (Embodied Conversational Agent - ECA), εμπλουτισμένη με συναίσθημα, και η είσοδος του χρήστη λαμβάνεται μέσω φυσικών τρόπων επικοινωνίας (εκφράσεις προσώπου, χειρονομιών, στάση σώματος κ.α.) αισθάνονται πιο άνετα και φυσικά συγκριτικά με τους καθιερωμένους (ποντίκι, πληκτρολόγιο, κ.α.) τρόπους επικοινωνίας [172].

Ενώ η συναισθηματική διέγερση διαμορφώνει όλα τα ανθρώπινα επικοινωνιακά σήματα [70], το οπτικό κανάλι (εκφράσεις προσώπου, στάση σώματος και χειρονομίες) θεωρείται το πιο σημαντικό στην ανθρώπινη ερμηνεία συμπεριφορών [4], δεδομένου ότι οι ανθρώπινοι παρατηρητές είναι ακριβέστεροι όταν η κρίση τους βασίζεται στο πρόσωπο και το σώμα απ'ό,τι όταν βασίζεται μόνο στην φωνή. Το γεγονός αυτό δείχνει

ότι οι άνθρωποι στηρίζονται στις εκφράσεις προσώπου και την γλώσσα του σώματος για να ερμηνεύσουν την διάθεση και την συμπεριφορά κάποιου και σε μικρότερο βαθμό στις ακουστικές εκφράσεις. Εντούτοις, αν και αρκετοί ερευνητές δεν έχουν προσδιορίσει ένα σύνολο χαρακτηριστικών φωνής ικανό να κατηγοριοποιήσει τα συναισθήματα, οι ανθρώπινοι ακροατές φαίνεται να είναι πιο ακριβείς στην αποκωδικοποίηση των συναισθημάτων από τα χαρακτηριστικά φωνής [125]. Κατά συνέπεια, τα σύνολα δεδομένων που χρησιμοποιούνται στην ανάλυση συναισθήματος πρέπει να περιλαμβάνουν τουλάχιστον τις εκφράσεις του προσώπου και να φροντίσουν να υπάρχει μια αντίληψη είτε των χαρακτηριστικών του σώματος είτε της λεκτικής προσωδίας. Τέλος, ενώ ο υπερβολικός αριθμός πληροφορίων από διαφορετικά κανάλια δείχνει να συγχέει την ανθρώπινη κρίση [172], με συνέπεια να καταλήγουν σε λιγότερο ακριβείς κατηγοριοποιήσεις της παρουσιασμένης συμπεριφοράς, όταν είναι διαθέσιμα περισσότερα κανάλια παρατήρησης (π.χ. πρόσωπο, σώμα και ομιλία), ο συνδυασμός πολλαπλών μορφών πληροφορίας (συμπεριλαμβανομένης της ομιλίας και της φυσιολογίας) μπορεί να αποδειχθεί κατάλληλος για την αυτοματοποιημένη ανάλυση ανθρώπινου συναισθήματος, όπως φάνηκε και στο προηγούμενο κεφάλαιο της διατριβής.

Έχουν υπάρξει πολλές ψυχολογικές μελέτες για το συναίσθημα και τη μη λεκτική επικοινωνία [69] καθώς και για τις εκφραστικές μετακινήσεις και στάσεις του σώματος [20, 55, 168, 238]. Αυτές οι μελέτες βασίστηκαν κυρίως σε υποδυόμενα, βασικά συναισθήματα (θυμός, αποστροφή, φόβος, χαρά, θλίψη, έκπληξη). Στον συναισθηματικό υπολογιστικό τομέα, πρόσφατες μελέτες μη λεκτικής συμπεριφοράς κατά τη διάρκεια έκφρασης συναισθημάτων είναι επίσης περιορισμένες όσον αφορά στον αριθμό των τρόπων έκφρασης (modalities) ή τον αυθορμητισμό του συναισθήματος: σημαδευτές στο σώμα αναγνωρίζουν τα τέσσερα βασικά συναισθήματα [128], καταγραφή κίνησης της στάσης του σώματος κατά τη διάρκεια δύο αποχρώσεων των τεσσάρων βασικών συναισθημάτων [58], επεξεργασία βίντεο των εκφράσεων του προσώπου και κινήσεις του κορμού κατά τη διάρκεια έξι υποδυόμενων συναισθηματικών συμπεριφορών [101]. Οι περισσότερες από αυτές τις μελέτες εξετάζουν υποδυόμενα, βασικά συναισθήματα, ενώ είναι λίγες οι περιπτώσεις που εξετάζονται συναισθήματα εκφραζόμενα με περισσότερες της μιας μορφές πληροφορίας, βασισμένα σε αυθόρμητες εκφράσεις, παρά τη γενική εκτίμηση ότι η συλλογή οπτικοακουστικών βάσεων δεδομένων που δίνουν έμφαση σε φυσικές εκφράσεις συναισθημάτων είναι απαραίτητη [63].

Πράγματι, η οικοδόμηση ενός σώματος πραγματικών/φυσιοκρατικών συναισθημάτων με πολλαπλές μορφές έκφρασης αποτελεί πρόκληση, δεδομένου ότι περιλαμβάνει την υποκειμενική αντίληψη κατά την επισημείωση του και απαιτεί μεγάλο χρονικό διάστημα για τον σχολιασμό του συναισθήματος σε πολλαπλά, παράλληλα επίπεδα. Αυτός ο χειρωνακτικός σχολιασμός μπορεί να ωφεληθεί από την αυτόματη ανίχνευση συναισθηματικά σχετικών τμημάτων του βίντεο μέσω της επεξεργασίας εικόνας των συνολικών τηλεοπτικών τμημάτων. Η εκτίμηση της ποσότητας κίνησης από την αυτόματη επεξεργασία εικόνας θα μπορούσε να επικυρώσει τους χειρωνακτικούς σχολιασμούς των μετακινήσεων του συνεντευξιαζόμενου κατά τη διάρκεια του σχολιασμού του βίντεο, καθώς επίσης και της συνολικής συναισθηματικής ενεργοποίησης σε επίπεδο ολόκληρου του βίντεο. Τέλος, ο αυτόματος σχολιασμός μπορεί να διευκολύνει την χειροκίνητη διαδικασία σχολιασμού με την παροχή της κατάτμησης μετακίνησης και των ακριβών τιμών εκφραστικών παραμέτρων όπως η ταχύτητα, η χωρική έκταση ή η ρευστότητα μιας χειρονομίας. Παρ'όλα αυτά, ο χειρωνακτικός σχολιασμός και η επεξεργασία εικόνας παρέχουν πληροφορίες σε διαφορετικά επίπεδα αφαίρεσης και ο συνδυασμός τους δεν αποτελεί μια τετριμμένη διαδικασία. Επιπλέον, το μεγαλύ-

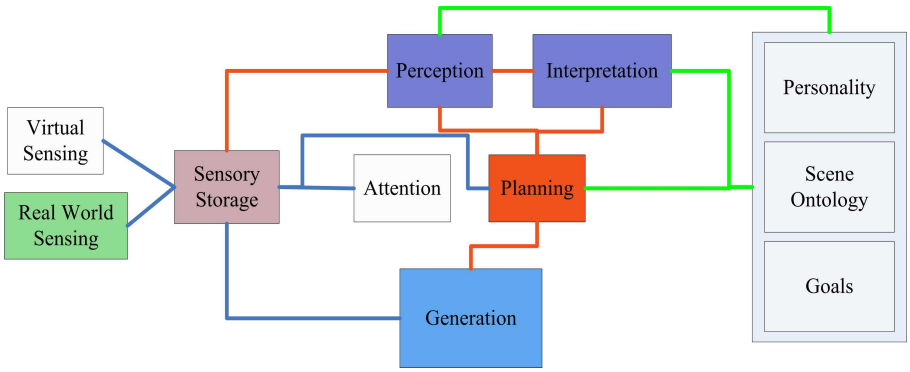
τερο μέρος των εργασιών στην επεξεργασία εικόνων συναισθηματικής συμπεριφοράς έχει γίνει σε υψηλής ποιότητας βίντεο που καταγράφονται σε εργαστηριακές συνθήκες όπου τα συναισθήματα είναι λιγότερο αυθόρμητα απ'ότι κατά τη διάρκεια μη προγραμματισμένων τηλεοπτικών συνεντεύξεων.

3.2 Εκφραστική μιμητικότητα ανθρώπων από εικονικούς χαρακτήρες

Αυτή η εργασία πραγματεύεται την εκφραστική και πολυμεσική σύνθεση σε εικονικούς συνομιλητικούς πράκτορες (Embodied Conversational Agents - ECA), βασισμένη στην ανάλυση των ενεργειών που εκτελέσθηκαν από ανθρώπινους χρήστες. Σαν είσοδο θεωρούμε την ακολουθία εικόνων καταγεγραμμένης ανθρώπινης συμπεριφοράς. Τεχνικές επεξεργασίας εικόνας επιστρατεύονται προκειμένου να εντοπιστούν ενδείξεις απαραίτητες για την εξαγωγή χαρακτηριστικών γνωρισμάτων εκφραστικότητας. Η πολυμεσική φύση της προσέγγισης έγκειται στο γεγονός ότι αναλύονται και επεξεργάζονται οι πτυχές της συμπεριφοράς του χρήστη που σχετίζονται με το πρόσωπο και τις χειρονομίες. Η μιμητικότητα συνίσταται στην αντίληψη, ερμηνεία, προγραμματισμό και την τελική εμψύχωση των εκφράσεων που παρουσιάζονται από τον άνθρωπο, καταλήγοντας όχι σε ένα ακριβές αντίγραφο παρά σε ένα εκφραστικό παραπλήσιο της αρχικής συμπεριφοράς του χρήστη.

Η εργασία μας επικεντρώνεται στις ενδιάμεσες διαδικασίες που απαιτούνται ώστε ένας πράκτορας να αντιληφθεί, ερμηνεύσει και να αντιγράψει μια σειρά από εκφράσεις του προσώπου και χειρονομίες από έναν άνθρωπο στον πραγματικό κόσμο όπως μπορεί να φανεί στην εικόνα 3.1. Οι ακολουθίες εικόνων υποβάλλονται σε επεξεργασία ώστε να εξαχθούν προεξέχοντα σημεία του προσώπου (Facial Definition Parameters - FDPs), η παραμόρφωσή τους (Facial Animation Parameters - FAPs) [196] και η θέση των χερίων του χρήστη [196]. Τα FDPs και FAPs ορίζονται στο πλαίσιο του MPEG–4 προτύπου και συνιστούν τυποποιημένα μέσα μοντελοποίησης της γεωμετρίας του προσώπου και της εκφραστικότητας του και έχουν επηρεαστεί έντονα από τις Μονάδες Δράσης (Action Units) που ορίζονται στις νευροφυσιολογικές και ψυχολογικές μελέτες των Ekman και Friesen [73]. Η υιοθέτηση της συμβολικής εμψύχωσης στο πλαίσιο του MPEG–4 ευνοεί τον καθορισμό των συναισθηματικών καταστάσεων, αφού η εξαγωγή απλών, συμβολικών παραμέτρων είναι κατάλληλή για ανάλυση και σύνθεση χειρονομιών και εκφράσεων του προσώπου. Αυτές οι πληροφορίες χρησιμοποιούνται για τον υπολογισμό παραμέτρων εκφραστικότητας και την κατηγοριοποίηση σε συναισθηματικές πτυχές της κατάστασης ενός χρήστη. Υποβάλλονται σε περαιτέρω επεξεργασία σε ένα πλαίσιο για την αντίληψη, την ερμηνεία, τον προγραμματισμό και την παραγωγή συμπεριφορών πρακτόρων αντίστοιχων με εκείνές που παρατηρούνται στον άνθρωπο. Με την έννοια αντίληψη, εννοούμε την συμπεριφορά που μπορεί να μην είναι ακριβές αντίγραφο της συμπεριφοράς του ανθρώπου που κατανοεί ο πράκτορας, αλλά βασίζεται σε κάποια διαδικασία ερμηνείας της συμπεριφοράς [157]. Η δυνατότητα του πράκτορα μας να αντιληφθεί την εκφραστικότητα του προσώπου και των χεριών παρουσιάζεται στην παρακάτω ενότητα.

Η επεξεργασία των δεδομένων αυτών περιλαμβάνει μια συμβολική και σημασιολογική επεξεργασία, διαδικασίες αναπαράστασης υψηλού επιπέδου και προγραμματισμό μακροπρόθεσμου σχεδιασμού. Επιπλέον, υποδηλώνει μια ερμηνεία της αντιλαμβανόμενης έκφρασης με συγκεκριμένα πεδία τιμών FAPs, η οποία μπορεί να προσαρμοστεί



Σχήμα 3.1: Εποπτική εικόνα της προτεινόμενης προσέγγισης. Οι υποενότητες σε έγχρωμο πλαίσιο έχουν υλοποιηθεί και στην πειραματική πλατφόρμα που παρουσιάζεται στην ενότητα 3.2.6.

από τον πράκτορα (π.χ. επίδειξη πιο έντονου θυμού) και να αναπαρασταθεί με έναν τρόπο που είναι μοναδικός στον πράκτορα (θυμός σε ένα άλλο σύνολο FAPs). Η ενότητα παραγωγής [187] [105], που συνθέτει την τελική επιθυμητή συμπεριφορά του πράκτορα περιγράφεται στην ενότητα 3.2.5, ενώ η ικανότητα του ECA μας να αντιληφθεί τις εκφράσεις του προσώπου και τις χειρονομίες που εκτελούνται από έναν πραγματικό χρήστη παρουσιάζεται στην ενότητα 3.2.4 με τη βοήθεια ενός απλού σεναρίου όπου ο ECA αντιλαμβάνεται και αναπαράγει την συμπεριφορά του χρήστη. Το χαρακτηριστικό αυτό της αντίληψης εξασφαλίζει ότι η προκύπτουσα εμψύχωση (animation) δεν είναι ένα πιστό αντίγραφο. Στο μέλλον, στοχεύουμε στην εκμετάλλευση της ικανότητας αυτής στην υλοποίηση ενός πιο σύνθετου μοντέλου απόφασης, το οποίο θα αναλαμβάνει την επιλογή των ενεργειών που θα εκτελέσει ο ECA, σύμφωνα επίσης με την τρέχουσα συμπεριφορά του χρήστη και να αξιολογήσουμε την ορθότητα της προτεινόμενης προσέγγισης χρησιμοποιώντας τον σχεδιασμό που συζητείται στην ενότητα 3.2.7.

3.2.1 Επισκόπηση συστήματος

Τα εκφραστικά χαρακτηριστικά, χρήσιμα κατά τη διάρκεια της ενότητας σύνθεσης, είναι βασισμένα στην εξαγωγή πληροφοριών του προσώπου και των χειρονομιών. Για το πρόσωπο, FAP τιμές [222] εξάγονται χρησιμοποιώντας τη μεθοδολογία που περιγράφεται στην υποενότητα 3.2.2 [118]. Όσον αφορά στην ανάλυση χειρονομίων, υπολογίζονται οι σχετικές αποστάσεις των χεριών και του κεφαλιού, κανονικοποιημένες σύμφωνα με το μέγεθος του κεφαλιού. Το σχήμα 3.3 είναι ενδεικτικό της διαδικασίας για τον εντοπισμό και την παρακολούθηση των χεριών.

Προκειμένου να πετύχουμε την απαραίτητη ανάλυση εικόνας για να λειτουργήσει αποτελεσματικά ο αλγόριθμος εξαγωγής των χαρακτηριστικών του προσώπου και ταυτόχρονα να ικανοποιούνται οι χωρικές απαιτήσεις της επεξεργασίας χειρονομιών, δύο μεμονωμένες ροές βίντεο καταγράφηκαν από διαφορετικές συσκευές. Η επιλογή μιας τέτοιας ρύθμισης έγινε επειδή η ανάλυση εικόνας που απαιτείται για την εξαγωγή χαρακτηριστικών του προσώπου είναι πολύ μεγαλύτερη από αυτή για την παρακολούθηση χειρονομιών. Αυτό θα μπορούσε να επιτευχθεί μόνο εάν μια συσκευή εστίαζε στο πρόσωπο του θέματος. Οι δύο ροές βίντεο συγχρονίστηκαν χειρωνακτικά πριν επεξεργαστούν.

3.2.2 Εξαγωγή χαρακτηριστικών γνωρισμάτων προσώπου

Η ανάλυση του προσώπου περιλαμβάνει διάφορα βήματα επεξεργασίας που αποσκοπούν στην ανίχνευση ή παρακολούθηση του προσώπου, ώστε να εντοπίστουν χαρακτηριστικές περιοχές του προσώπου όπως μάτια, στόμα και μύτη σε αυτό, και στην συνέχεια να εξαχθούν γνωρίσματα και να καταγραφεί η μετακίνηση των χαρακτηριστικών γνωρισμάτων, όπως τα χαρακτηριστικά σημεία σε αυτές τις περιοχές. Αν και τα FAPs παρέχουν όλα τα απαραίτητα στοιχεία για εμψύχωση με βάση το πρότυπο MPEG–4, δεν μπορούμε να τα χρησιμοποιήσουμε για την ανάλυση των εκφράσεων από καταγεγραμμένες σκηνές, λόγω της απουσίας σαφούς ποσοτικού πλαισίου καθορισμού. Προκειμένου να υπολογιστούν τα FAPs από πραγματικές ακολουθίες εικόνων, πρέπει να καθορίσουμε μια αντιστοίχηση μεταξύ αυτών και της μετακίνησης των συγκεκριμένων παραμέτρων καθορισμού του προσώπου (FDPs), που χαρακτηρίζει τα χαρακτηριστικά σημεία προσώπου (FPs), τα οποία αντιστοιχούν σε εμφανή σημεία στο ανθρώπινο πρόσωπο. Η αντιστοίχηση αυτή πραγματοποιείται με την μέθοδο που περιγράφεται στην ενότητα 2.3.3.1.

Σχήμα 3.2: Ενδεικτικά χαρακτηριστικά σημεία του προσώπου

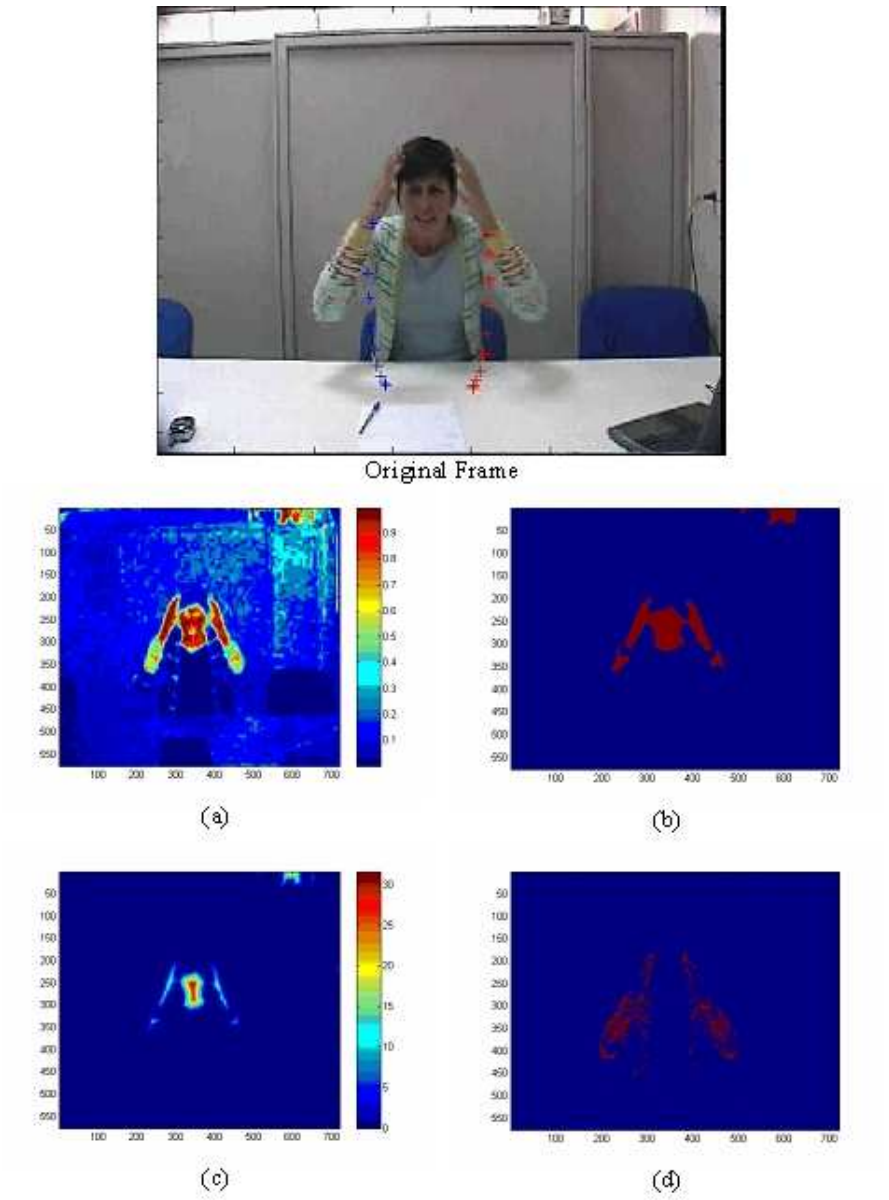
3.2.3 Εντοπισμός και παρακολούθηση χεριών

Όσον αφορά στην ανάλυση χειρονομιών, διάφορες προσεγγίσεις εξετάστηκαν για την υποενότητα παρακολούθησης κεφαλιού-χεριών όπου όλες τους αναφέρονται στα [252] και [170]. Από αυτές εξετάστηκαν λεπτομερέστερα οι μέθοδοι βασισμένες σε βίντεο είσοδο δεδομένου ότι η καταγραφή κίνησης (motion capture) ή άλλες παρεμφερείς τεχνικές παρεμβαίνουν περισσότερο στην συναισθηματική κατάσταση του αντικειμένου. Σημαντικότεροι παράγοντες που ελήφθησαν υπόψη ήταν το υπολογιστικό κόστος και η ευρωστία, καταλήγοντας σε μια ακριβής, υλοποιημένη να εκτελείται σε πραγματικό χρόνο υποενότητα εντοπισμού και παρακολούθησης χρωματικών περιοχών δέρματος.

Η διαδικασία περιλαμβάνει τη δημιουργία κινούμενων μασκών δέρματος, δηλαδή περιοχές με χρωματικά χαρακτηριστικά κοντά σε ένα πρότυπο δέρματος που παρακολουθούνται μεταξύ διαδοχικών πλαισίων. Με την παρακολούθηση των κεντροειδών αυτών των μασκών δερμάτων, παράγουμε μια εκτίμηση των μετακινήσεων των χεριών του χρήστη. Πρότερη γνώση σχετικά με το ανθρώπινο σώμα, τις συνθήκες

καταγραφής των χειρονομιών και ενδείξεις των διαφορετικών μερών σωμάτων (κεφάλι, δεξί χέρι, αριστερό χέρι) ενσωματώθηκαν στην υποενότητα. Για κάθε πλαίσιο παράγεται μια μήτρα πιθανότητας χρώματος δέρματος με τον υπολογισμό της ένωσης της πιθανότητας ομοιότητας των καναλιών Cb/Cr της εικόνας με ένα δυναμικό πρότυπο δερματικής περιοχής. Η μάσκα χρώματος δέρματος λαμβάνεται έπειτα από τη εφαρμογή ενός κατωφλιού στην μήτρα πιθανότητας δέρματος. Πιθανές κινούμενες περιοχές βρίσκονται μετά από εφαρμογή κατωφλίου στην διαφορά των εικονοστοιχείων του τρέχοντος και του επόμενου πλαισίου, παράγοντας τη μάσκα πιθανής-κίνησης. Αυτή η μάσκα δεν περιέχει πληροφορίες σχετικά με την κατεύθυνση ή το μέγεθος της μετακίνησης, αλλά είναι ενδεικτική της ύπαρξης κίνησης και χρησιμοποιείται για να επιταχύνει τον αλγόριθμο επικεντρώνοντας την παρακολούθηση σε περιοχές πιθανής κίνησης της εικόνας. Χρωματικές μάσκες και μάσκες κίνησης περιέχουν έναν μεγάλο αριθμό μικρών αντικειμένων λόγω της παρουσίας θορύβου και αντικειμένων με χρωματικά χαρακτηριστικά παρόμοια με το δέρμα. Για να το ξεπεράσουμε αυτό, εφαρμόζουμε μορφολογικό φιλτράρισμα και στις δύο μάσκες για να εξαλειφθούν αυτά τα μικρά αντικείμενα. Ως παράμετρο στους μορφολογικές τελεστές χρησιμοποιείται ένα κυκλικό δομικό στοιχείο με ακτίνα 1% του πλάτους εικόνας ή και σχετικό με το μέγεθος του ανιχνευμένου κεφαλιού και τελικά επιβιώνουν μόνο τα αντικείμενα/περιοχές μεγαλύτερα από κάποιο επιθυμητό μέγεθος. Ο μετασχηματισμός απόστασης της μάσκας χρώματος υπολογίζεται αρχικά. Αυτά τα αντικείμενα χρησιμοποιούνται ως σημαδευτές για την μορφολογικό ανακατασκευή της αρχικής μάσκας χρώματος. Στην προκύπτουσα μάσκα εφαρμόζεται ένας μορφολογικός τελεστής κλεισίματος για καλύτερο υπολογισμό των κεντροειδών των χεριών. Στο επόμενο πλαίσιο, μια νέα μάσκα κινούμενου δέρματος δημιουργείται και εκτελείται μια ένα προς ένα αντιστοίχηση αντικειμένων με την ευρετική μέθοδο που περιγράφεται παρακάτω. Η αντιστοίχηση των αντικειμένων μεταξύ δύο πλαισίων εκτελείται στη μάσκα χρώματος και είναι βασισμένη στην απόσταση των αντικειμένων παρόμοιας επιφάνειας. Το συστατικό του διανύσματος μετακίνησης των χεριών παράλληλο στο Z άξονα δεν λαμβάνεται υπόψη δεδομένου ότι κάτι τέτοιο θα απαιτούσε πληροφορία βάθους από την επεξεργασία εικόνας και αυτό θα επιδείνωνε σημαντικά την απόδοση του προτεινόμενου αλγορίθμου και πιθανόν να απαιτούσε μια δεύτερη συσκευή καταγραφής και παράλληλη επεξεργασία των δύο ροών εισόδου. Η υλοποίηση του περιγεγραμμένου αλγορίθμου επιτρέπει έναν ρυθμό επεξεργασίας 12 fps σε έναν συνηθισμένο υπολογιστή κατά τη διάρκεια των πειραμάτων μας, ο οποίος είναι αρκετός για τη συνεχή παρακολούθηση χειρονομιών, ενώ μια υλοποίηση που εκμεταλλεύεται την δυνατότητα επεξεργασίας στο υλικό της κάρτας γραφικών του υπολογιστή αυξάνει ακόμα περισσότερο τον ρυθμό επεξεργασίας. Η ευρετική αντιστοίχηση αντικειμένων το καθιστά ικανό να παρακολουθήσει τα τμήματα των χεριών με ακρίβεια, τουλάχιστον κατά τη διάρκεια συνηθισμένων ακολουθιών χειρονομίας. Επιπλέον, ο συνδυασμός των πληροφοριών χρώματος και κίνησης απομακρύνει οποιοδήποτε θόρυβο ή ψεγάδια, ενισχύοντας κατά συνέπεια την ευρωστία της προτεινόμενης προσέγγισης.

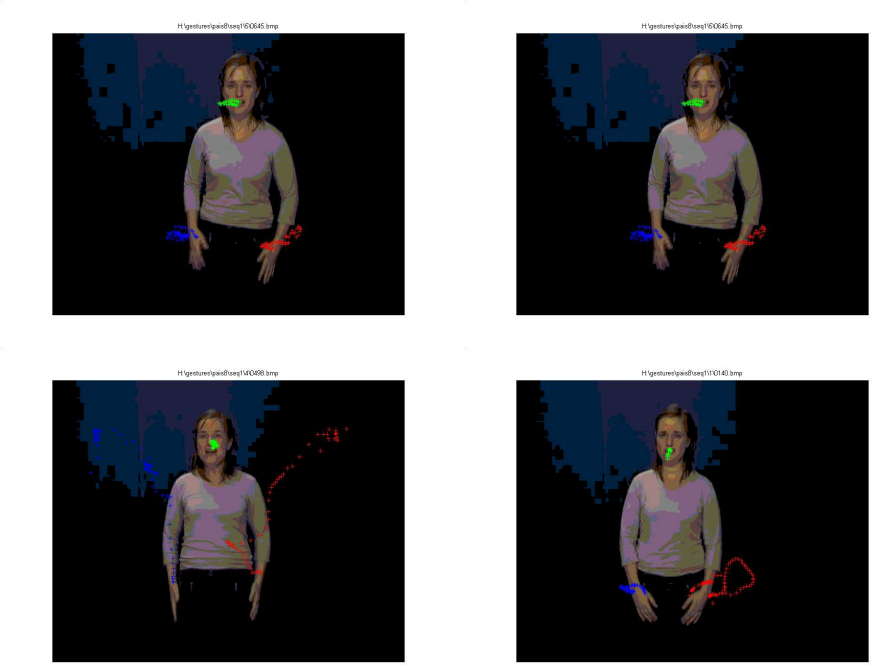
Ο αλγόριθμος παρακολούθησης είναι υπεύθυνος για την αντιστοίχηση των περιοχών κινούμενου δέρματος, μέσα στην ακολουθία εικόνων της υπό εξέτασης χειρονομίας, που παρήχθησαν από την παραπάνω περιγεγραμμένη μέθοδο. Το μέγεθος των περιοχών δερμάτων, η απόσταση με σημείο αναφοράς την θέση της περιοχής στο προηγούμενο πλαίσιο, η συνέχεια των διανυσμάτων κίνησης ως προς την κατεύθυνση και χωρικοί περιορισμοί είναι κριτήρια που εξασφαλίζουν ότι η επόμενη περιοχή που επιλέγεται για να αντικαταστήσει την τρέχουσα είναι περίπου το ίδιο μέγεθος, κοντά



Σχήμα 3.3: Βήματα του αλγόριθμου εντοπισμού και παρακολούθησης χεριών (α) Πιθανότητα δέρματος (β) Εφαρμογή κατωφλίου (γ) Μετασχηματισμός απόστασης (δ) Διαφορά πλαισίων

στην τελευταία θέση και κινείται κατά προσέγγιση κατά μήκος της ίδιας κατεύθυνσης με την προηγούμενη εφ' όσον η στιγμιαία ταχύτητα είναι επάνω από ένα ορισμένο κατώφλι. Κατά συνέπεια σε κάθε υποψήφια περιοχή απονέμεται ένα αντίτιμο για την ικανοποίηση αυτών των κριτηρίων ή τιμωρείται για την αποτυχία συμμόρφωσης με τους παραπάνω περιορισμούς. Η νικήτρια περιοχή ορίζεται ως περιοχή αναφοράς για το επόμενο πλαίσιο. Τα κριτήρια δεν έχουν επίδραση αποκλεισμού, που σημαίνει ότι εάν μια περιοχή αποτύχει να ικανοποιήσει ένα από τα κριτήρια δεν αποκλείεται από τη διαδικασία, ενώ το θετικό ή αρνητικό αντίτιμο που αποδίδεται στην περιοχή είναι ανάλογο με την ταύτιση κάθε περιοχής με κάθε κριτήριο. Στην συνολική βαθμολογία της τελικά επιλεγμένης περιοχής εφαρμόζεται ένα κατώφλι ώστε να αποκλείονται περιοχές που σημειώσαν χαμηλές βαθμολογίες. Στην περίπτωση αυτή η θέση του χεριού

παραμένει αμετάβλητη σε σχέση με το προηγούμενο πλαίσιο. Αυτό το χαρακτηριστικό γνώρισμα είναι ιδιαίτερα χρήσιμο σε περιπτώσεις επικάλυψης όταν και η θέση του χεριού παραμένει ίδια όπως αμέσως προτού παρουσιαστεί η επικάλυψη. Η επικάλυψη κεφαλιού-χεριών ή χεριών-χεριών αντιμετωπίζεται με την ακόλουθη απλοϊκή παρ'όλα αυτά αποδοτική μέθοδος. Υποθέτουμε ότι επικάλυψη συμβαίνει στο n πλαίσιο και παύει να υπάρχει στο k πλαίσιο. Η θέση του χεριού κατά τη διάρκεια της φάση επικάλυψης (πλαίσια $n \ldots k$) θεωρείται η θέση του χεριού στο πλαίσιο $n-1$. Μετά από το πλαίσιο k ο αλγόριθμος εντοπισμού και παρακολούθησης για το συγκεκριμένο χέρι συνεχίζεται κανονικά. Μετά από έναν ορισμένο αριθμό πλαισίων ολόκληρη η διαδικασία επανεκινείται ώστε πιθανό λάθος να μην διαδίδεται. Οι εικόνες 3.4 και και 3.7 δείχνουν τις τροχιές των χεριών και του κεφαλιού που παράγονται από τον προτεινόμενο αλγόριθμο όταν εφαρμόζεται στην βάση δεδομένων GEMEP [10] και στα θέματα του πειραματικού σώματός μας αντίστοιχα.



Σχήμα 3.4: Αποτελέσματα εφαρμογής του αλγορίθμου εντοπισμού και παρακολούθησης χεριών και κεφαλιού

3.2.4 Υπολογισμός παραμέτρων εκφραστικότητας χειρονομιών

Η εκφραστικότητα της συμπεριφοράς είναι ένα αναπόσπαστο τμήμα της διαδικασίας επικοινωνίας αφού μπορεί να παρέχει πληροφορίες για την επικρατούσα συναισθηματική κατάσταση, διάθεση και προσωπικότητα ενός προσώπου [239]. Πολλοί ερευνητές έχουν ερευνήσει τα χαρακτηριστικά ανθρώπινης κινήσης και τα έχουν κωδικοποιήσει σε δυαδικές κατηγορίες όπως αργά/γρήγορα, μικρά/ευρύα, αδύναμα/ενεργητικά, μικρά/μεγάλα, δυσάρεστα/ευχάριστά. Για να μοντελοποιήσουμε την εκφραστικότητα, στην εργασία μας, χρησιμοποιούμε έξι διαστάσεις εκφραστικότητας που περιγράφονται στο [105], ως τον πιο ολοκληρωμένο τρόπο περιγραφής της εκφραστικότητας, δεδομένου ότι αντιμετωπίζει όλο το φάσμα παραμέτρων έκφρασης συναισθήματος. Έχουν οριστεί, σε επίπεδο σύνθεσης, πέντε παράμετροι που διαμορφώνουν την εκφρα-

στικότητα της συμπεριφοράς, ως υποσύνολο των έξι διαστάσεων της εκφραστικότητας συμπεριφοράς

- Καθολική Ενεργοποίηση (Overall activation)
- Χωρική Έκταση (Spatial extent)
- Χρονική (Temporal)
- Ρευστότητα (Fluidity)
- Ενέργεια (Power)

Κατά την διαδικασία εξαγωγής εκφραστικών παραμέτρων χειρονομιών θεωρούμε μια χειρονομία G ως μια ακολουθία, μήκους T πλαισίων, συντεταγμένων του αριστερού και δεξιού χεριού (x_{li}^G, y_{li}^G) και (x_{ri}^G, y_{ri}^G), αντίστοιχα, με το $i \in [1, T]$. Οι συντεταγμένες των χεριών είναι σχετικές με το κέντρο του παραλληλόγραμμου που περικλείει την περιοχή του κεφαλιού που θεωρείται ως θέση του κεφαλιού και κανονικοποιημένες με την διαγώνιο του ίδιου παραλληλόγραμμου που θεωρείται ως το μέγεθος του κεφαλιού. Οι μετασχηματισμοί αυτοί είναι απαραίτητοι προκειμένου οι συντεταγμένες να είναι αμετάβλητες σε σχέση με την θέση του χρήστη στο επίπεδο της εικόνας αλλά και την απόσταση του από την κάμερα καθώς αυτές οι τιμές αυτών των παραμέτρων δεν είναι εκ των προτέρων γνωστές. Έτσι τυπικά η χειρονομία ορίζεται ως:

$$G = [((x_{l1}^G, y_{l1}^G), (x_{r1}^G, y_{r1}^G)), ((x_{l2}^G, y_{l2}^G), (x_{r2}^G, y_{r2}^G)), \cdots, ((x_{lT}^G, y_{lT}^G), (x_{rT}^G, y_{rT}^G))] \quad (3.1)$$

Για λόγους απλότητας τα ζεύγη (x_{li}^G, y_{li}^G) και (x_{ri}^G, y_{ri}^G) αντιστοιχούνται στις εκφράσεις R_i^G και L_i^G αντίστοιχα. Επίσης ορίζεται η ποσότητα μετακίνησης D_i μεταξύ των πλαισίων ως το μέτρο του διανύσματος που ορίζεται από τα σημεία (x_i, y_i) και (x_{i+1}, y_{i+1}), $D_i = \left|\overrightarrow{(x_i, y_i)(x_{i+1}, y_{i+1})}\right|$.

Η καθολική ενεργοποίηση θεωρείται ως η ποσότητα μετακίνησης κατά τη διάρκεια μιας διαλογικής αλληλεπίδρασης. Ποιοτικά κάποιος θα μπορούσε να την αντιστοιχήσει στον άξονα της ενεργοποίησης, αλλά μια τέτοια προσέγγιση θα ήταν αρκετά απλοϊκή καθώς η διάσταση της ενεργοποίησης περιλαμβάνει περισσότερες της μιας έννοιες και σίγουρα δεν μπορεί να περιγραφεί μονοσήμαντα από την εκφραστική παράμετρο της καθολικής ενεργοποίησης. Τυπικά, την ορίζουμε ως το άθροισμα των ποσοτήτων μετακίνησης:

$$OA_G = \sum_{i=1}^{T-1} D_{li}^G + D_{ri}^G \quad (3.2)$$

Η χωρική έκταση εκφράζεται με την επέκταση ή την σύμπτυξη του χρησιμοποιούμενο χώρου μπροστά από τον πράκτορα/άνθρωπο και όπως και η καθολική ενεργοποίηση σχετίζεται, όχι όμως αποκλειστικά, με τον άξονα της ενεργοποίησης, καθώς κάποιος που γενικά βρίσκεται στο αρνητικό ημιεπίπεδο της ενεργοποίησης, είναι δηλαδή μάλλον παθητικός είναι απίθανο να επεκτείνει σημαντικά τον χώρο χειρονομίας (gesturing space) μπροστά του. Προκειμένου να οριστεί σαφώς ο χώρος έκτασης κατά την διάρκεια της χειρονομίας ορίζουμε την στιγμιαία χωρική έκταση e_i ως το μέτρο

του διανύσματος που ορίζουν τα σημεία (x_{li}, y_{li}) και (x_{ri}, y_{ri}) την στιγμή i. Έτσι η εκφραστική παράμετρος της χωρικής έκτασης αντιστοιχεί στην μέγιστη ποσότητα της στιγμιαίας χωρικής έκτασης κατά την διάρκεια της χειρονομίας:

$$\begin{aligned} SE_G &= \max e_i, i \in [1, T] \\ e_i &= \left| \overrightarrow{(x_{ri}, y_{ri})(x_{li}, y_{li})} \right| \end{aligned} \tag{3.3}$$

Η χρονική παράμετρος εκφραστικότητας της χειρονομίας δηλώνει την ταχύτητα της μετακίνησης κατά την διάρκεια της χειρονομίας και διαχωρίζει γρήγορες με αργές χειρονομίες. Εφόσον η ποσότητα D_i ορίζει την στιγμιαία ταχύτητα για την στιγμή i η χρονική εκφραστική παράμετρος ορίζεται ως τον μέσο όρο της ποσότητας αυτής και δεδομένου πως η OA αντιστοιχεί στο διακριτό ολοκλήρωμα της:

$$TE_G = \frac{OA}{T} \tag{3.4}$$

Ενώ, η παράμετρος εκφραστικότητας της ενέργειας αναφέρεται στη μετακίνηση των χεριών κατά τη διάρκεια φάση κτυπήματος χειρονομίας (π.χ., γρήγορης/συγρατημένης ενέργειας). Οι χειρονομίες αποτελούνται από τρεις φάσεις: προετοιμασία, κτύπημα και απόσυρση. Το ουσιαστικό μήνυμα τους συγκεντρώνεται στην φάση του κτυπήματος, ενώ τα στοιχεία προετοιμασιών και απόσυρσης αποτελούνται από την κίνηση των χεριών σε και από την ουδέτερη θέση της χειρονομίας. Η τυποποίηση της ενέργειας της χειρονομίας σύμφωνα με αυτόν τον ορισμό όμως είναι εξαιρετικά δύσκολο να πραγματοποιηθεί στην αυτόματη εξαγωγή εκφραστικών παραμέτρων χειρονομιών εφόσον οι φάσεις της χειρονομίας είναι δεν είναι σαφώς διαχωρίσιμες σε πραγματικές συνθήκες αλληλεπίδρασης, ενώ ακόμα και σε ελεγχόμενες συνθήκες με ελεύθερες χειρονομίες αποτελεί μεγάλη πρόκληση και σίγουρα θα έπρεπε να ενσωματώνει γνώση σχετικά με την φύση αν όχι την ακριβή κατηγορία της χειρονομίας. Εναλλακτικά επιλέξαμε να αντιστοιχήσουμε ποιοτικά την παράμετρο αυτή στην πρώτη παράγωγο του μέτρου της D που παραπέμπει στην επιτάχυνση των χεριών κατά την διάρκεια της χειρονομίας:

$$PO = |D|' \tag{3.5}$$

Η ρευστότητα διαφοροποιεί τις ομαλές/κομψές από τις ξαφνικές/απότομες χειρονομίες. Αυτή η έννοια επιδιώκει να καταγράψει την συνέχεια μεταξύ των μετακινήσεων και είναι κατάλληλη να μοντελοποιήσει τις αλλαγές στην επιτάχυνση και επιβράδυνση των άκρων. Υπό αυτό το πρίσμα, ορίζουμε την ρευστότητα μιας χειρονομίας ως την διακύμανση της ενέργειας όπως τελικά ορίστηκε στην προηγούμενη παράγραφο:

$$FL = var(PO) \tag{3.6}$$

Σημαντική σημείωση είναι πως η ποσότητα που FL αντιστοιχεί σε μια ποσότητα αντιστρόφως ανάλογη της έννοιας της ρευστότητας. Έτσι μια χειρονομία με υψηλή τιμή της εκφραστικής παραμέτρου FL επιδεικνύει χαμηλή ρευστότητα και επομένως κατηγοριοποιείται στις ξαφνικές/απότομες χειρονομίες. Η μετατροπή της σε μια έκφραση ανάλογης της έννοιας της ρευστότητας δεν αποτελεί τετριμμένη διαδικασία καθώς δεν είναι γνωστές οι τιμές του άνω και κάτω ορίου της ρευστότητας συνολικά για όλες τις κατηγορίες χειρονομιών. Το ίδιο πρόβλημα παρουσιάζεται και στην ενότητα της σύνθεσης καθώς δεν είναι πάντα εύκολο να καθοριστούν οι μέγιστες

και ελάχιστες τιμές για κάθε εκφραστική παράμετρο και συνεπώς δεν μπορεί να γίνει ομοιόμορφη κανονικοποίηση, αλλά η διαδικασία εξαρτάται κάθε φορά από το υπό εξέταση σύνολο χειρονομιών και τις αντίστοιχες τιμές των εκφραστικών παραμέτρων.

3.2.5 Σύνθεση

Οι επικοινωνιακές ικανότητες των πρακτόρων με ευχέρεια στην συνομιλία θα μπορούσαν να βελτιωθούν σημαντικά εάν είχαν την ικανότητα να μεταβιβάσουν παράλληλα τα εκφραστικά συστατικά της φυσικής συμπεριφοράς. Εμπνευσμένοι από τα αναφερόμενα αποτελέσματα στο [239], στο [105] ορίστηκε και υλοποιήθηκε ένα σύνολο πέντε παραμέτρων που επιδρούν στην ποιότητα της συμπεριφοράς του πράκτορα, οι οποίες είναι η χωρική έκταση (SPC), χρονική παράμετρος (TMP), ενέργεια (PWR), ρευστότητα (FLT) και επαναληπτικότητα (REP). Κατά συνέπεια, οι ίδιες χειρονομίες ή εκφράσεις του προσώπου εκτελούνται από τον εικονικό πράκτορα με έναν ποιοτικά διαφορετικό τρόπο ανάλογα με αυτό το σύνολο παραμέτρων.

Ο πίνακας 3.1 παρουσιάζει την επίδραση που έχει κάθε παράμετρος εκφραστικότητας στην εμψύχωση των κινήσεων του κεφαλιού, τις εκφράσεις του προσώπου και τις χειρονομίες. Η παράμετρος χωρικής έκτασης (SPC) διαμορφώνει το εύρος της μετακίνησης των χεριών, και του κεφαλιού. Επηρεάζει το κατά πόσο ευρεία ή περιορισμένη θα είναι η μετατόπισή τους κατά τη διάρκεια της τελικής εμψύχωσης. Παραδείγματος χάριν ας εξετάσουμε την ανόρθωση των φρυδιών στην έκφραση της έκπληξης: εάν η τιμή της παραμέτρου χωρικής έκτασης είναι πολύ υψηλή η τελική θέση των φρυδιών θα είναι πολύ ψηλά στο μέτωπο. Η χρονική παράμετρος (TMP) βραχύνει ή επιμηκύνει την φάση κίνησης της προετοιμασίας και απόσυρσης της χειρονομίας καθώς επίσης και την διάρκεια της έναρξης και λήξης της έκφρασης του προσώπου. Η επίδραση της στο πρόσωπο είναι να επιταχύνει ή να επιβραδύνει την ανόρθωση/τον υποβιβασμό των φρυδιών. Η εμψύχωση του πράκτορα υλοποιείται με τον ορισμό μερικών πλαισίων κλειδιών (key frames) και τον υπολογισμό των καμπυλών παρεμβολής που διέρχονται μέσω αυτών των πλαισίων. Η ρευστότητα (FLT) και η ενέργεια (PWR) επιδρούν στις καμπύλες παρεμβολής. Η ρευστότητα αυξομειώνει τη συνέχεια των καμπυλών επιτρέποντας στο σύστημα να παράγει περισσότερο/λιγότερο ομαλές παραλλαγές εμψύχωσης. Ας εξετάσουμε την επίδρασή της στο κεφάλι: εάν η τιμή της παραμέτρου ρευστότητας είναι πολύ χαμηλή η προκύπτουσα καμπύλη θα εμφανιστεί σαν να προκύπτει μέσω της γραμμικής παρεμβολής. Κατά συνέπεια, κατά τη διάρκεια της τελικής εμψύχωσης το κεφάλι θα έχει μια σπασμωδική μετακίνηση. Η ενέργεια εισάγει μια υπέρβαση της χειρονομίας/έκφρασης, η οποία είναι μια μικρή χρονική περίοδος κατά την οποία το μέρος του σώματος που συμμετέχει στην χειρονομία/έκφραση φθάνει σε ένα σημείο πέρα από το τελικό. Παραδείγματος χάριν το συνοφρύωμα που επιδεικνύεται στην έκφραση του θυμού θα είναι ισχυρότερο για μια μικρή χρονική περίοδο και έπειτα τα φρύδια θα φθάσουν στην τελική θέση. Η τελευταία παράμετρος, επαναληπτικότητα (REP), ασκεί μια επιρροή στις χειρονομίες και τις κινήσεις του κεφαλιού. Αυξάνει των αριθμό των κτυπημάτων των χειρονομιών για να αναπαραστήσει την επανάληψη στην εμψύχωση της χειρονομίας. Ας εξετάσουμε την περίπτωση της χειρονομίας 'χαιρετισμός', μια υψηλή τιμή της επαναληπτικής παραμέτρου θα αυξήσει τον αριθμό των πάνω-κάτω μετακινήσεων. Αφ' ετέρου αυτή η παράμετρος μειώνει το χρονικό διάστημα των νευμάτων του κεφαλιού ώστε να παραχθούν περισσότερα νεύματα και κουνήματα στην ίδια χρονική περίοδο.

Ο παραπάνω πίνακας μπορεί να γίνει περισσότερο κατανοητός με δύο διαισθητικά

Πίνακας 3.1: Επίδραση των εκφραστικών παραμέτρων στο κεφάλι, τις εκφράσεις του και τις χειρονομίες

	Κεφάλι	**Εκφράσεις Προσώπου**	**Χειρονομίες**
SPC	ευρύτερη/περιορισμένη κίνηση	αυξημένη/μειωμένη συναισθηματική διέγερση	ευρύτερη/περιορισμένη κίνηση
TMP	αυξημένη/μειωμένη ταχύτητα κίνησης	πρόωρη/καθυστερημένη έναρξη και λήξη	αυξημένη/μειωμένη ταχύτητα της φάσης προετοιμασίας και απόσυρσης
FLT	αυξημένη/μειωμένη συνέχεια των κινήσεων του κεφαλιού	αυξημένη/μειωμένη συνέχεια κίνησης	αυξημένη/μειωμένη συνέχεια μεταξύ διαδοχικών χειρονομιών
PWR	αυξημένη/μειωμένη υπέρβαση κεφαλιού	αυξημένη/μειωμένη επιτάχυνση κίνησης	αυξημένη/μειωμένη επιτάχυνση της φάσης κτυπήματος
REP	περισσότερα/λιγότερα νεύματα	δεν έχει υλοποιηθεί ακόμα	περισσότερες/λιγότερες επαναλήψεις της φάσης κτυπήματος

παραδείγματα. Η παράμετρος SPC επιδρά στο εύρος των εκφράσεων του προσώπου και στις κινήσεις του κεφαλιού και του σώματος του πράκτορα: εάν επιλεχθεί μια υψηλή τιμή του SPC και ο πράκτορας πρέπει να εκτελέσει ένα χαμόγελο, οι γωνίες των χειλιών του θα διευρυνθούν και θα εμφανιστούν μέγιστες. Η παράμετρος TMP επιδρά στην ταχύτητα των κινήσεων του κεφαλιού, εμφάνιση και εξαφάνιση των εκφράσεων του προσώπου και στην προετοιμασία και απόσυρση χειρονομιών. Παραδείγματος χάριν, εάν μια χαμηλή τιμή TMP επιλεχθεί και ο πράκτορας πρέπει να γνέψει, να παρουσιάσει ένα συνοφρύωμα και να εκτελέσει την χειρονομία 'κτύπημα', το νεύμα του κεφαλιού θα είναι νωθρό, τα φρύδια θα ζαρώσουν αργά και θα κινήσει τον βραχίονά αργά προτού να εκτελέσει το 'κτύπημα' στην χειρονομία.

Η υποενότητα σύνθεσης είναι σε θέση να αναπαράγει ένα μεγάλο σύνολο βασικών εκφράσεων προσώπου, που προτείνονται από τον Ekman [69] αλλά και πολλούς άλλους που θεωρούνται συνδυασμός τους. Οι χειρονομίες υπολογίζονται μέσω της παρεμβολής ακολουθίας στατικών θέσεων ορισμένων ως περιστροφές ώμων και βραχιόνων (θέση βραχιόνων), χειρομορφή (που επιλέγονται από ένα σύνολο προκαθορισμένων μορφών) και προσανατολισμός παλάμης [104]. Έτσι η ενότητα σύνθεσης μπορεί να αναπαράγει επιτυχώς εικονικές και χειρονομίες χτυπήματος ενώ κυκλικές χειρονομίες δεν εκτελούνται προς το παρόν.

Από την πλευρά της υλοποίησης, το σύστημα πράκτορα παράγει τα δεδομένα εμψύχωσης σε συμβατά με MPEG–4, FAP/BAP δομή, τα οποία καθοδηγούν στη συνέχεια το σκελετικό και το πρότυπο του προσώπου σε OpenGL. Ένα σύνολο παραμέτρων,

καλούμενοι παράμετροι εμψύχωσης προσώπου (FAPs) και παράμετροι εμψύχωσης σώματος (BAPs), χρησιμοποιούνται για να εμψυχωθεί το πρόσωπο και το σώμα αντίστοιχα. Με ορισμό τιμών FAPs και BAPs, μπορούμε να ορίσουμε τις εκφράσεις του προσώπου και τις θέσεις του σώματος. Η εμψύχωση ορίζεται από μια ακολουθία πλαισίων κλειδιών (keyframes). Το πλαίσιο κλειδί αποτελείται από ένα σύνολο FAP ή BAP τιμές που πρέπει να υλοποιηθούν. Η εμψύχωση λαμβάνεται με την παρεμβολή μεταξύ αυτών των πλαισίων κλειδιών. Η παρεμβολή εκτελείται χρησιμοποιώντας TCB (Tension, Continuity, Bias) εύκαμπτους κανόνες χάραξης καμπυλών (splines) [137].

Οι παράμετροι εκφραστικότητας εφαρμόζονται με την τροποποίηση των TCB παραμέτρων των κανόνων χάραξης καμπυλών παρεμβολής, των τιμών των παραμέτρων εμψύχωσης και του χρονισμού των πλαισίων κλειδιών. Για παράδειγμα, η παράμετρος SPC επηρεάζει το πλαίσιο κλειδι διαφοροποιώντας την τιμή των FAPs και BAPs. Όσο μεγαλύτερη τιμή έχει το SPC, τόσο ευρύτερες θα είναι οι καμπύλες παρεμβολής και τόσο οι εκφράσεις του προσώπου θα είναι πιό ορατές και οι χειρονομίες ευρύτερες. Η παράμετρος FLD θα διαμορφώσει τις παραμέτρους συνοχής (Continuity parameters) των κανόνων χάραξης καμπυλών, κάνοντάς τις ομαλότερες (υψηλό FLD) ή σπασμωδικότερες (χαμηλό FLD).

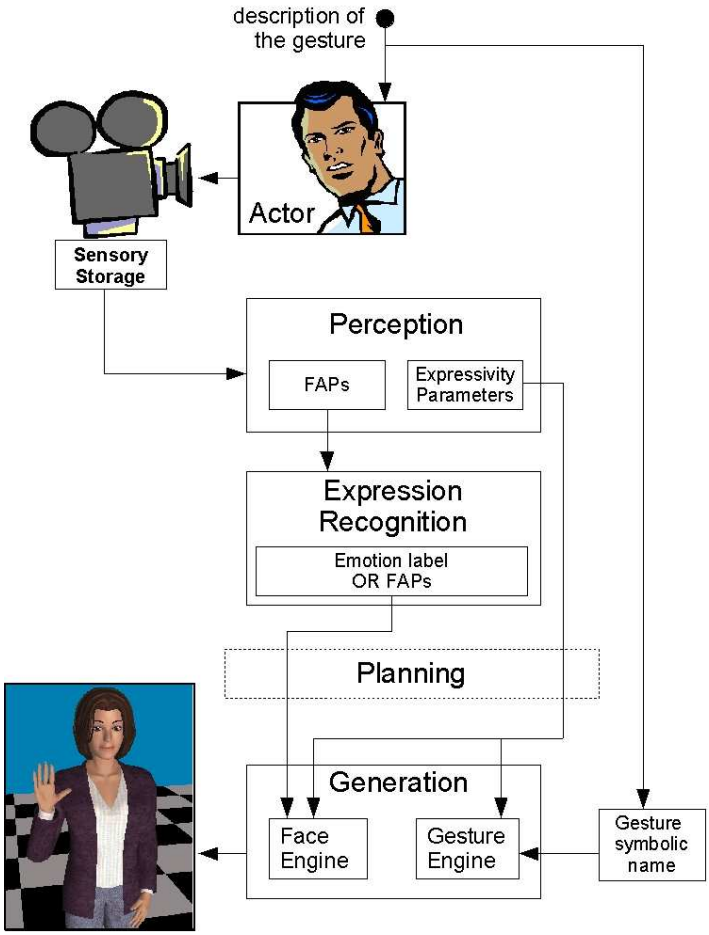
3.2.6 Υλοποίηση

Στην εισαγωγική ενότητα περιγράφηκε το γενικό πλαίσιο του συστήματος ικανού να αναλύσει μια πραγματική σκηνή και να αναπαράγει τη εκφραστική εμψύχωση ενός εικονικού πράκτορα. Εδώ παρουσιάζετε ένα σενάριο που είναι μια μερική υλοποίηση του πλαισίου αυτού. Το παρόν σύστημά μας (εικόνα 3.5) είναι σε θέση να εξάγει στοιχεία από το βίντεο, να τα επεξεργαστεί και να τα εμψυχώσει στον εικονικό πράκτορα. Η τελική εμψύχωση στοχεύει στην αναπαραγωγή χειρονομιών και εκφράσεων προσώπου που πραγματοποιεί ο χρήστης εμπλουτισμένη με τις παραμέτρους εκφραστικότητας που υπολογίζονται από την πραγματική ακολουθία.

Πίνακας 3.2: Κατανομή χειρονομιών σε τεταρτημόρια

Κλάση Χειρονομίας	**Τεταρτημόριο στον χώρο του Whissel**
εξηγώ	(0,0), (+, +), (-, +), (-, -)
ωχ θεέ μου	(+, +), (-, +)
άφησέ με μόνο	(-, +), (-, -)
ανόρθωση χεριού	(0,0), (+, +), (-, -)
βαριέμαι	(-, -)
χαιρετώ	(0,0), (+, +), (-, +), (-, -)
χειροκροτώ	(0,0), (+, +), (-, +), (-, -)

Η είσοδος του συστήματος προέρχεται από υποδυόμενες ενέργειες εκτελεσμένες από τα αντικείμενα του πειράματος. Η ενέργεια αποτελείται από μια χειρονομία συνοδευόμενη από μια έκφραση του προσώπου. Τόσο ο τύπος της χειρονομίας όσο και ο τύπος της έκφρασης ζητούνται ρητά από τον δράστη και περιγράφονται σε εντόν σε φυσική γλώσσα (παραδείγματος χάριν ζητείται από τον χρήστη να 'κυματίσει το δεξί χέρι του μπροστά από τη κάμερα παρουσιάζοντας χαρούμενο πρόσωπο'). Η



Σχήμα 3.5: Υλοποιημένο σενάριο

Σχήμα 3.6: Τα αντικείμενα των πειραμάτων

δημιουργία ενός σώματος από συμβάντα πραγματικής ζωής, χωρίς ο χρήστης να υποδύεται, εξετάστηκε αλλά τελικά δεν επιλέχτηκε επειδή οι εκφράσεις που προέρχονται από συμβάντα πραγματικής ζωής συμβαίνουν υπό διαφορετικές, αυθαίρετες και μη ελεγχόμενες καταστάσεις και συχνά είναι δύσκολο να αναπραχθούν. Το λεκτικό τους περιεχόμενό και η γενική ποιότητα των καταγραφών είναι μη ελεγχόμενες και συνήθως ένα άτομο καταγράφεται μόνο σε μια ή σε πολύ λίγες από τις διαφορετικές συναισθηματικές καταστάσεις, ενώ η επισημείωση είναι δύσκολη, υποκειμενική και συχνά ασαφής. Οι ακολουθίες εικόνων εισόδου του προτεινόμενου συστήματος είναι βίντεο που καταγράφονται κατά την διάρκεια μιας συνεδρείας που περιελάμβανε 7 χρήστες, εικόνα 3.6, όπου κάθε ένας από αυτούς εκτελεί 7 χειρονομίες, πίνακας 3.2. Κάθε χειρονομία εκτελέσθηκε αρκετές φορές με τον χρήστη να βρίσκεται σε διαφορετική συναισθηματική κατάσταση κάθε φορά. Οι κατηγορίες των χειρονομιών είναι: 'εξηγώ', 'ω θεέ μου' (δύο χέρια στο κεφάλι),'αφησέ με μόνο', 'ανόρθωση χεριού' (ζητώ προσοχή), 'βαριέμαι', 'χαιρετώ' και 'χειροκρότημα'. Ο πίνακας 3.2 προσδιορίζει ποιες συναισθηματικές επαναλήψεις εκτελέσθηκαν για κάθε συγκεκριμένη χειρονομία. Παραδείγματος χάριν η χειρονομία 'χαιρετώ' εκτελέσθηκε 4 φορές μια για το ουδέτερο

συναίσθημα και μια για κάθε ένα από για τα συγκεκριμένα τεταρτημόρια του χώρου Whissel ((+,+), (-,+), (+,-)) . Μερικοί συνδυασμοί δεν συμπεριλήφθηκαν στο σενάριο αφού δεν είχε νόημα η εκτέλεση, για παράδειγμα, της χειρονομίας 'βαριέμαι' με χαρούμενη διάθεση (+,+).

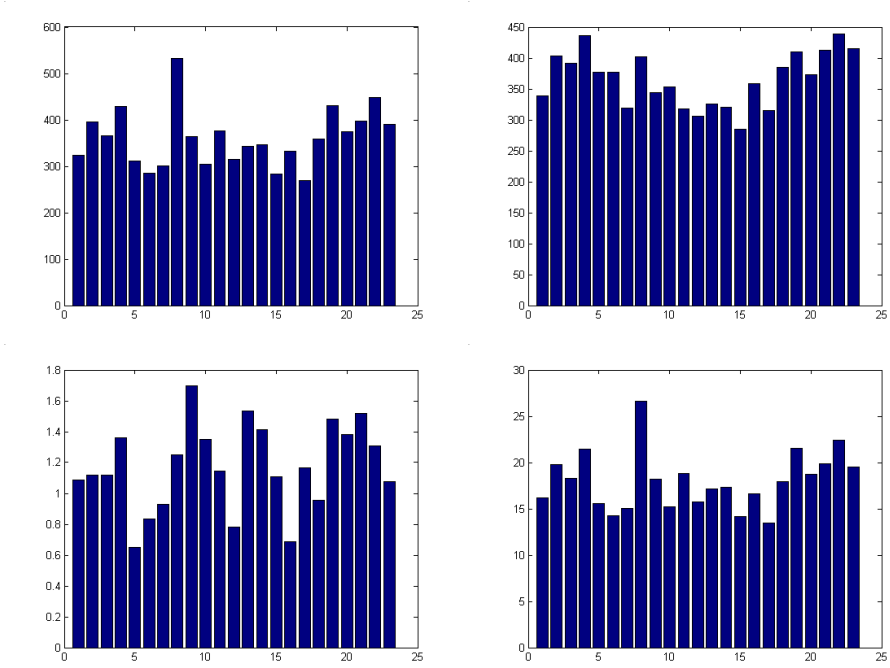
Η υποενότητα της Αντίληψης (perception) αναλύει το προκύπτον βίντεο εξάγοντας τις παραμέτρους εκφραστικότητας της χειρονομίας και την μετατόπιση των χαρακτηριστικών του προσώπου που χρησιμοποιούνται για να προσδιορίσουν τις τιμές των FAP που αντιστοιχούν στην διενεργηθείσα έκφραση του προσώπου. Οι τιμές FAP και οι παράμετροι εκφραστικότητας τροφοδοτούν την υποενότητα αναγνώρισης έκφρασης. Εάν η έκφραση του προσώπου αντιστοιχεί σε ένα από τα πρότυπα των συναισθηματικών εκφράσεων του προσώπου, αυτή η υποενότητα είναι σε θέση να παραγάγει την συμβολική κατηγορία του (ετικέτα συναισθήματος) από τις τιμές FAP που εισάγονται. Εάν όχι οι τιμές FAP χρησιμοποιούνται ακέραιες. Επίσης όσον αφορά στην αναγνώριση χειρονομιών μπορεί να χρησιμοποιηθεί το σύστημα που περιγράφεται στην ενότητα 4.2. Στο εγγύς μέλλον, σχεδιάζουμε επίσης να εφαρμόσουμε μια υποενότητα προγραμματισμού (εμφανίζεται ως το κουτί με διακεκομμένες γραμμές στην εικόνα 3.5 για την προσαρμογή είτε των παραμέτρων εκφραστικότητας είτε του συναισθήματος. Η τελική μορφή της εμψύχωσης, αποτελούμενη από την χρονική εξέλιξη των FAPs και BAPs, διαμορφώνεται στην υποενότητα Παραγωγής (generation) που χωρίζεται σε μηχανή προσώπου και χειρονομίας. Η μηχανή προσώπου (που υπολογίζει επίσης τις κινήσεις του κεφαλιού του πράκτορα) δέχονται ως είσοδο τον χαρακτηρισμό της κατηγορίας συναισθήματος (ή μια ακολουθία FAP) και ένα σύνολο παραμέτρων εκφραστικότητας. Ο τρόπος με τον οποίο οι εκφράσεις του προσώπου εμφανίζονται και οι κινήσεις του κεφαλιού εκτελούνται διαμορφώνεται από τις παραμέτρους εκφραστικότητας όπως εξηγείται στην ενότητα 3.2.5, πίνακας 3.1. Με τον ίδιο τρόπο, η υποενότητα σύνθεσης εκφραστικών χειρονομιών λαμβάνει ως είσοδο τον χαρακτηρισμό της κατηγορίας χειρονομίας και ένα σύνολο παραμέτρων εκφραστικότητας. Έτσι οι χειρονομίες που παράγονται από την υποενότητα σύνθεσης χειρονομίας επηρεάζονται από το σύνολο παραμέτρων εκφραστικότητας, όπως εξηγείται στην ενότητα 3.2.5, πίνακας 3.1.

Το σύστημα δεν λειτουργεί ακόμα σε πραγματικό χρόνο, αλλά στοχεύουμε να διερευνήσουμε τις δυνατότητες που υπάρχουν προς την κατεύθυνση αυτήν στο μέλλον. Μέσα στα άμεσα σχέδια μας είναι να αξιολογήσουμε την ορθότητα του συστήματος μας μέσω δοκιμών αντιληπτικότητας προκειμένου να εκτιμηθεί η ποιότητα των συναισθηματικών εμψυχώσεων (Ενότητα 3.2.7).

Η εικόνα 3.8 παρουσιάζει τις τιμές των παραμέτρων εκφραστικότητας που υπολογίστηκαν χρησιμοποιώντας τον παραπάνω περιγεγραμμένο αλγόριθμο στις χειρονομίες του πειράματος. Στην εικόνα 3.10 αποδεικνύεται η ορθότητα της υποενότητας για τον υπολογισμό της εκφραστικότητας αφού δείχνεται το αναμενόμενο αποτέλεσμα. Αυτό είναι ότι χειρονομίες που ανήκουν στο θετικό ημιεπίπεδο ενεργοποίησης έχουν υψηλότερες τιμές στην παράμετρο της Καθολικής Ενεργοποίησης και Ενέργειας σε σύγκριση με εκείνες που ανήκουν στο αρνητικό ημιεπίπεδο ενεργοποίησης.

Η εικόνα 3.9 καταδεικνύει τρείς περιπτώσεις μίμησης συμπεριφοράς. Η χειρονομία που μιμείται στην περίπτωση (α) είναι ουδέτερη, στην (β) χαρούμενη (+/+ τεταρτημόριο) και στην (γ) λυπημένη (-/- τεταρτημόριο). Δεν είναι όλα τα χαρακτηριστικά γνωρίσματα εκφραστικότητας ευδιάκριτα από το στιγμιότυπο της εμψύχωσης, αλλά η χωρική έκταση γίνεται πολύ εύκολα αντιληπτή τόσο στην χειρονομία όσο και στην έκφραση του προσώπου.

Σχήμα 3.7: Αποτελέσματα εφαρμογής του αλγορίθμου εντοπισμού και παρακολούθησης χεριών και κεφαλιού

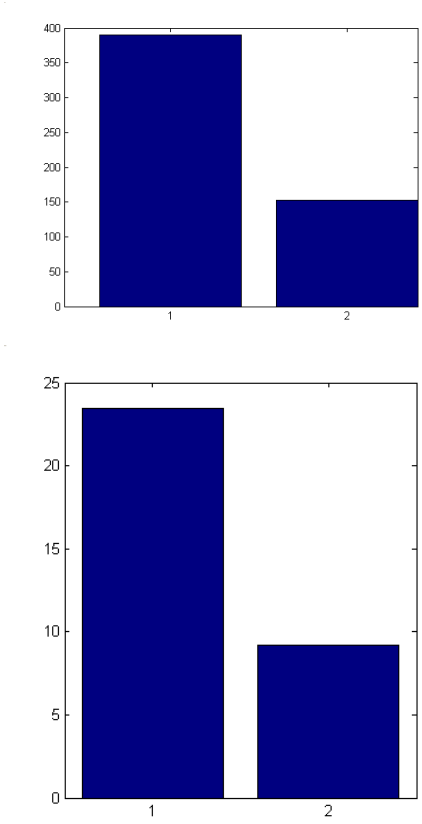


Σχήμα 3.8: Μέσες τιμές των τιμών των παραμέτρων εκφραστικότητας για κάθε κατηγορία χειρονομίας (α)Καθολική Ενεργοποίηση (β)Χωρική Έκταση (γ)Ρευστότητα (δ)Ενέργεια

3.2.7 Σχήμα αξιολόγησης του συστήματος

Μια αξιολόγηση αντιληπτικότητας του μέχρι τώρα υλοποιημένου συστήματος κρίνεται ιδιάζουσας σημασίας προκειμένου να διερευνηθεί πώς η συνθετική συμπεριφορά γίνε-

Σχήμα 3.9: Επίδειξη της μιμητικότητας συμεριφοράς



Σχήμα 3.10: Μέσες τιμές των τιμών των παραμέτρων εκφραστικότητας (α)Καθολική Ενεργοποίηση και (β)Ενέργεια για ενεργητικές και παθητικές χειρονομίες αντίστοιχα

ται αντιληπτή από τους τελικούς χρήστες και να προσδιορίστουν οι τρόποι βελτίωσης της τρέχουσας εξόδου του συστήματος. Η προσαρμογή της εμψύχωσης της εκφραστικότητας σύμφωνα με τα αναμενόμενα αποτελέσματα έχει νόημα και πέρα από το

πλαίσιο αυτό [105]. Σε αυτό το σημείο πρόκειται να παρουσιάσουμε ένα σχήμα που προτείνουμε ώστε να αξιολογηθεί η τρέχουσα εργασία.

Υπάρχουν διάφορες ερωτήσεις που αναζητούν απάντηση, στις δοκιμές αξιολόγησης, που αφορούν στην εκτίμηση της αντίληψης για τις συνθετικές συναισθηματικές εκφράσεις προκειμένου να συνεχιστεί η σε βάθος ανάλυση των αναφερόμενων παραμέτρων. Στην περίπτωσή μας, μια πρώτη ερώτηση είναι η αντίληψη και η ταξινόμηση της συντεθειμένης παραγωγής από ανθρώπινους κριτές. Σχεδιάζουμε να πραγματοποιήσουμε μια αξιολόγηση όπου είκοσι μεταπτυχιακοί φοιτητές θα κληθούν να εκτιμήσουν σειριακά μια ακολουθία βίντεο. Θα εμφανιστεί ένα ερωτηματολόγιο που θα αποτελείται από ερωτήσεις κλίμακας σχετικά με το ποσοστό κάθε κατηγορίας που θεωρούν ότι ανήκει το τμήμα βίντεο που παρακολούθησαν. Η επιλογή των κατηγοριών ταυτίζεται με τις κατηγορίες που ζητήθηκε να υποδυθούν τα αντικείμενα του πειράματος.

Οι συμμετέχοντες θα ερωτηθούν πώς οι συνθετικές εμψυχώσεις γίνονται αντιληπτές με την παρακολούθηση των βίντεο μιμητικής συμπεριφοράς. Κατά συνέπεια, θα διαιρεθούν σε δύο ομάδες. Στην πρώτη ομάδα θα παρουσιαστούν τα συνθετικά βίντεο ενώ στην δεύτερη θα παρουσιαστούν και τα αρχικά και τα αντίστοιχα συνθετικά βίντεο. Η διάταξη της παρουσίασης του συνόλου των βίντεο θα γίνει τυχαία ώστε να αποφευχθούν φαινόμενα που σχετίζονται με την διάταξη των συναισθημάτων που παρουσιάζονται. Ως εννοιολογική βασική γραμμή μας θα θεωρήσουμε μια παρουσίαση του πράκτορα μας με ουδέτερη έκφραση.

Τα ερωτηματολόγια θα χρησιμοποιηθούν επίσης για να συλλεχθούν πληροφορίες σχετικά με το ποίο συναίσθημα αναγνώρισαν σε κάθε ακολουθία που παρουσιάστηκε. Οι πιθανές απαντήσεις θα περιοριστούν σε συγκεκριμένες κατηγορίες. Η εμπιστοσύνη των συμμετεχόντων σχετικά με αυτή την κατηγοριοποίηση θα είναι μετρήσιμη αφου θα συμπεριληφθεί ερώτημα σχετικά με την ένταση του αντιλαμβανόμενου συναισθήματος.

Τα αποτελέσματα από αυτήν την πρώτη δοκιμή αξιολόγησης μπορούν να παρέχουν χρήσιμα στοιχεία όσον αφορά στην αντίληψη και αναγνώριση των συντεθειμένων εκφράσεων, καθώς επίσης και πληροφορίες για την επίδραση του παρεχομένου (αρχικές ακολουθίες βίντεο) στην αντίληψη συναισθηματικής κατάστασης για τις αντίστοιχες συνθετικές εκδόσεις. Η μέτρηση εμπιστοσύνης θα βοηθήσει στον σχηματισμό συμπερασμάτων σχετικά με τον ρόλο των παραμέτρων εκφραστικότητας και λεπτομερέστερους χειρισμούς των παραμέτρων αυτών και από κοινού με τη αξιολόγηση αντίληψης να βοηθήσουν στην αποκωδικοποίηση του ρόλου των παραμέτρων αυτών στην αντίληψη για συντεθειμένες εκφράσεις.

3.2.8 Συσχετισμός αυτόματης εξαγωγής παραμέτρων εκφραστικότητας και επισημείωσης

Πάντα στο πλαίσιο της πειραματικής αξιολόγησης του προτεινόμενου συστήματος πραγματοποιήθηκε, χρησιμοποιώντας το πειραματικό σύνολο που χρησιμοποιήθηκε στην ενότητα 3.2.5 και τις χειρονομίες που απαριθμούνται στον πίνακα 3.2, μια μελέτη του συσχετισμού των εκφραστικών παραμέτρων όπως αυτοί εξάχθηκαν με την μέθοδο που περιγράφεται στην ενότητα 3.2.4 και της επισημείωσης που προέκυψε με την επίδειξη των βίντεο με τις χειρονομίες σε χρήστες. Επιλέχθηκαν 16 αντιπροσωπευτικά βίντεο με δείγματα χειρονομιών όπου η περιοχή του προσώπου του χρήστη έχει υποστεί επεξεργασία ώστε να μην μπορεί είναι ευδιάκριτη η έκφραση του ενώ ο ήχος αποκόπηκε ώστε ο επισημειωτής να επηρεάζεται μόνο από την κίνηση των χεριών. Η σειρά των χειρονομιών ήταν τυχαία ώστε να μην επηρεασθεί ο χρήστης

και η όλη διαδικασία επίδειξης των βίντεο και επισημείωσης ήταν προσβάσιμη από το διαδίκτυο. Ο αριθμός των συμμετεχόντων στην διαδικασία της επισημείωσης ήταν 20, μισοί άνδρες και μισές γυναίκες, η ηλικία τους ήταν στο φάσμα 25 με 35 και ίδιας εθνικότητας. Αρχικά έγινε ποιοτική επεξήγηση των εκφραστικών παραμέτρων στους συμμετέχοντες και αργότερα τους ζητήθηκε αποτιμήσουν την εκφραστικότητα της χειρονομίας αναθέτοντας μια τιμή από -3 εως $+3$ σε κάθε παράμετρο για όλα τα βίντεο, τα οποία είχαν την δυνατότητα να παρακολουθήσουν όσες φορές επιθυμούσαν. Οι χρήστες ανέφεραν ότι μόνο μετά από την επίδειξη και αξιολόγηση μερικών χειρονομιών μπορούσαν πραγματικά να εκτιμήσουν σωστά την εκφραστικότητα σε κάθε περίπτωση. Έχει επίσης ενδιαφέρον να αναφερθεί πως συνάντησαν αρκετή δυσκολία στον ποιοτικό διαχωρισμό των παραμέτρων καθώς και την τελική απόφαση για την ανάθεση τιμής σε κάθε παράμετρο.

Όσον αφορά στην επιτηδευμένη συναισθηματική εκφραστικότητα υπάρχουν τρείς άξονες στους οποίους θα μπορούσε να γίνει ανάλυση σχετικά με την εξέλιξη του συναισθήματος, την συσχέτιση με μεθόδους αυτόματης αναγνώρισης συναισθηματικών παραμέτρων, απόκλιση στην εκφορά του ζητούμενου συναισθήματος, κ.α.:

- Το συναίσθημα το οποίο ζητήθηκε να γίνει είτε με ρητές οδηγίες είτε με κάποια μέθοδο εκμαίευσης, αλλά σε κάθε περίπτωση μια συγκεκριμένη συναισθηματική κατηγορία ή κάποια περιοχή στον συναισθηματικό χώρο.

- Η αυτόματη επεξεργασία είτε με την μορφή εξαγωγής χαρακτηριστικών γνωρισμάτων είτε με την μορφή της εξόδου κάποιου κατηγοριοποιητή που δέχεται ώς είσοδο τα χαρακτηριστικά αυτά.

- Το συναίσθημα που τελικά γίνεται αντιληπτό είτε από ειδικούς είτε από απλούς χρήστες. Ενώ μπορεί να είχε ζητηθεί να εκφρασθεί κάποιο συναίσθημα δεν είναι βέβαιο πως αυτό μεταβιβάσθηκε επιτυχώς και έγινε αντιληπτό από τους παρατηρητές και έτσι ενώ οι προηγούμενοι άξονες ορίζονται σαφώς ο τελευταίος παρουσιάζει κάποια υποκειμενικότητα καθώς επηρεάζεται από πολλούς ανεξέλεγκτους παράγοντες. Αυτό όμως δεν μειώνει την σημασία του για μια πλήρης και ολοκληρωμένη μελέτη στον τομέα της συναισθηματικής υπολογιστικής.

Αρχικά διερευνήθηκε η συσχέτιση των εκφραστικών παραμέτρων που προέκυψαν μέσω της αυτόματης ανάλυσης και της επισημείωσης από τους συμμετέχοντες στο πείραμα για το σύνολο των 16 βίντεο και μετέπειτα των τελευταίων 12 όπως φαίνεται και στον πίνακα 3.3. Αυτό έγινε εξαιτίας της παρατήρησης πως οι χρήστες έδειχναν πιο εξοικειωμένοι με την διαδικασία επισημείωσης αφού είχαν ολοκληρώσει μερικές χειρονομιές. Παράλληλα, καταγράφηκε η συσχέτιση τόσο της μέσης όσο και της ενδιάμεσης τιμής των επισημειωμένων τιμών των παραμέτρων εκφραστικότητας.

Στις τιμές παρατηρείται ισχυρή συσχέτιση μόνο για την εκφραστική παράμετρο της χωρικής έκτασης, ενώ για τα υπόλοιπα παρατηρείται καλή συσχέτιση. Η αρνητική συσχέτιση για την ρευστότητα οφείλεται, όπως αναφέρεται και παραπάνω (ενότητα 3.2.4), στον τρόπο υπολογισμού της παραμέτρου αυτής καθώς ουσιαστικά εκτιμά την απόκλιση από τον μέσο όρο, ποσότητα αντιστρόφως ανάλογη με αυτό που γίνεται αντιληπτό ως ρευστότητα. Άξιο αναφοράς είναι επιπλέον πως οι χρήστες ανέθεσαν τιμές που συσχετίζονται κατά 0.95 για τις παραμέτρους της δύναμης και της ταχύτητας, γεγονός που ενισχύει τον ισχυρισμό τους για την δυσκολία διαχωρισμού μερικών παραμέτρων.

Πίνακας 3.3: Συσχέτιση αυτόματα υπολογισμένων και επισημειωμένων εκφραστικών παραμέτρων

	Ενεργοποίηση	Έκταση	Ταχύτητα	Ρευστότητα	Ενέργεια
Avg/16	0.38	0.78	0.53	-0.20	0.57
Median/16	0.38	0.76	0.50	-0.21	0.61
Avg/12	0.36	0.74	0.50	-0.22	0.52
Median/12	0.38	076	0.48	-0.18	0.55

3.2.9 Συμπεράσματα

Έχουμε παρουσιάσει ένα γενικό πλαίσιό αποτελούμενο από διάφορες αλληλοσυνδεόμενες ενότητες και ένα από τα πιθανά σενάριά υλοποίησης του σύμφωνα με το οποίο ένας πράκτορας αντιλαμβάνεται, ερμηνεύει και μιμείται μια σειρά από εκφράσεις του προσώπου και χειρονομίες από ένα χρήστη στον πραγματικό κόσμο. Η τελική εμψύχωση του πράκτορα προέρχεται από δεδομένα διαφόρων τύπων: ακατέργαστες τιμές παραμέτρων, ετικέτες κατηγορίας συναισθήματος, παράμετροι εκφραστικότητας και συμβολική αναπαράσταση χειρονομιών. Το σύστημα είναι σε θέση να αντιληφθεί και να ερμηνεύσει τις εκφράσεις του προσώπου και τις χειρονομίες που γίνονται από έναν χρήστη, ενώ μια πιθανή επέκταση περιλαμβάνει συναισθηματικές ενδείξεις προσωδίας από το κανάλι της ομιλίας. Στο μέλλον, στοχεύουμε στην εκμετάλλευση της ικανότητας αυτής στην υλοποίηση ενός πιο σύνθετου μοντέλου απόφασης, το οποίο θα αναλαμβάνει την επιλογή των ενεργειών που θα εκτελέσει ο ECA, σύμφωνα επίσης με την τρέχουσα συμπεριφορά του χρήστη και να αξιολογήσουμε την ορθότητα της προτεινόμενης προσέγγισης χρησιμοποιώντας τον σχεδιασμό που συζητείται στην ενότητα 3.2.7.

3.3 Επικύρωση χειρωνακτικού σχολιασμού εκφραστικότητας μέσω της αυτόματης εξαγωγής παραμέτρων

Οι στόχοι της παρούσας εργασίας είναι να διερευνηθεί η εφαρμοσιμότητα τεχνικών επεξεργασίας εικόνας σε χαμηλής ανάλυσης τηλεοπτικά βίντεο και πώς η επεξεργασία εικόνας θα μπορούσε να χρησιμοποιηθεί για την επικύρωση του χειρωνακτικού σχολιασμού αυθόρμητης συναισθηματικής συμπεριφοράς. Η πρώτη ενότητα περιγράφει το σώμα των τηλεοπτικών συνεντεύξεων που έχει συλλεχθεί και τους χειρωνακτικούς σχολιασμούς που έχουν καθοριστεί. Ενώ, τέλος διερευνώνται διάφοροι τρόπους σύγκρισης των χειρωνακτικών σχολιασμών και των αποτελεσμάτων της επεξεργασίας εικόνας επί ενός μεγαλύτερου συνόλου αρχείων βίντεο.

3.3.1 Χειρωνακτική επισημείωση πολύμορφων συναισθηματικών συμπεριφορών

Το σώμα EmoTV περιέχει 50 τηλεοπτικά δείγματα συναισθηματικών τηλεοπτικών συνεντεύξεων [1]. Τα βίντεο κωδικοποιούνται χρησιμοποιώντας τον κωδικοποιητή

(codec) Cinepak (720x576 εικονοστοιχεία, 25 πλαίσια/sec). Στόχος του EmoTV σώματος είναι να παρέχει γνώση σχετικά με το συντονισμό μορφών πληροφορίας κατά τη διάρκεια αυθόρμητων συμπεριφορών εμπλουτισμένων με συναίσθημα. Κατά συνέπεια, ένα πολυεπίπεδο σχήμα κωδικοποίησης σχεδιάστηκε ώστε να επιτρέπει την αναπαράσταση συναισθήματος σε διάφορα χρονικά επίπεδα και επίπεδα αφαίρεση [59]. Σε ευρύτερο επίπεδο υπάρχει ο σχολιασμός συναισθήματος (κατηγορικός και διαστατικός) συμπεριλαμβανομένης της καθολικής ενεργοποίησης. Παρόμοιοι σχολιασμοί είναι διαθέσιμοι στο επίπεδο συναισθηματικών τμημάτων του βίντεο.

Σε επίπεδο πολυμεσικών συμπεριφορών [157] υπάρχουν επίπεδα για κάθε ορατή μορφή: κορμός, κεφάλι, ώμοι, εκφράσεις του προσώπου, βλέμμα και χειρονομίες. Το κεφάλι, ο κορμός και οι τροχιές των χεριών περιέχουν μια περιγραφή σχετικά με την θέση και την μετακίνηση των μορφών αυτών. Έτσι σχολιασμοί θέσης και μετακίνησης εναλλάσσονται. Όσον αφορά στον σχολιασμό των συναισθηματικών μετακινήσεων, εμπνευστήκαμε το σχήμα σχολιασμού μας από το πρότυπο εκφραστικότητας που προτείνεται στο [105] το οποίο περιγράφει την εκφραστικότητα με ένα σύνολο έξι διαστάσεων: χωρική και χρονική έκταση, ενέργεια, ρευστότητα, επανάληψη και καθολική ενεργοποίηση. Η ποιότητα μετακίνησης επισημειώνεται σύμφωνα με αυτό το πρότυπο για τον κορμό, το κεφάλι, τους ώμους και τις χειρονομίες.

Για το σχολιασμό χειρονομιών, διατηρήσαμε τις κλασσικές ιδιότητες σχολιασμού [136] [159]. Το σχήμα κωδικοποίησής μας επιτρέπει έτσι όχι μόνο τον σχολιασμό της εκφραστικότητας μετακίνησης αλλά και τον σχολιασμό της δομικής περιγραφής ('φάσεις') των χειρονομιών, αφού τα χρονικά πρότυπα των χειρονομιών πιθανόν να σχετίζονται με την έκφραση του συναισθήματος: προετοιμασία (φέρνοντας το χέρι και την παλάμη στην θέση του κτύπηματος), κτύπημα (το πιό ενεργητικό μέρος της χειρονομίας), ακολουθία χτυπημάτων (διαδοχικά κτυπήματα), αναμονή (μια φάση παύσης ακριβώς πριν από ή αμέσως μετά από το κτύπημα) και απόσυρση (μετακίνηση πίσω στην αρχική θέση). Όσον αφορά στο σύνολο λειτουργιών χειρονομιών, όπως αυτά παρατηρήθηκαν στο σώμα, η επιλογή ήταν η εξής: χειριστική (επαφή με το σώμα ή το αντικείμενο), χτυπήματος (συγχρονισμένος με την έμφαση στην ομιλία), δεικτική (το χέρι χρησιμοποιείται για να δείξει σε ένα υπαρκτό ή φανταστικό αντικείμενο), διευκρινιστική (αντιπροσωπεύει τις ιδιότητες, ενέργειες, σχέσεις σχτιζόμενες με αντικείμενα και χαρακτήρες), εμβληματικές (κίνηση με μια σαφώς ορισμένη σημασία για μια κοινωνική ομάδα). Προς το παρόν, η χειρομορφή δεν σχολιάζεται δεδομένου ότι δεν θεωρείται ως κύριο χαρακτηριστικό γνώρισμα της συναισθηματικής συμπεριφοράς στην βιβλιογραφία, αν εξαιρεθούν κάποιες ειδικές περιπτώσεις διάταξης της χειρομορφής.

Ενώ η επισημείωση των συναισθημάτων έχει γίνει από 3 κριτές και οδήγησαν σε υπολογισμό της συμφωνίας [59], το παρόν πρωτόκολλο που χρησιμοποιείται για την επικύρωση των σχολιασμών των πολύμορφων συμπεριφορών προβλέπει έναν δεύτερο έλεγχο των σχολιασμών ακολουθούμενο από συζήτηση. Αν και εξετάζουμε επίσης την επικύρωση των σχολιασμών από τον αυτόματο υπολογισμό των συμφωνιών των κριτών η αυτόματη επεξεργασία εικόνας παρέχει ένα εναλλακτικό μέσο για την επικύρωση του χειρωνακτικού σχολιασμό.

3.3.2 Αυτόματη επεξεργασία βίντεο συναισθηματικών συμπεριφορών

Η επεξεργασία εικόνας χρησιμοποιείται για να παρέχει εκτίμηση των μετακινήσεων κεφαλιού και χεριών συνδυάζοντας πληροφορίες σχετικά με την θέση περιοχών δερμάτων και την εκτίμησης κίνησης. Ο εντοπισμός κεφαλιού και χεριών σε ακολουθίες εικόνων είναι βασισμένος στην ανίχνευση συνεχών περιοχών χρωματικής πληροφορίας κοντά σε αυτή του δέρματος. Για τη δεδομένη εφαρμογή, ένα πολύ γενικό πρότυπο δέρματος είναι ικανοποιητικό, δεδομένου ότι δεν υπάρχει καμία ανάγκη για την αναγνώριση της χειρομορφής. Όπως αναφέρεται και παραπάνω το υπό εξέταση σώμα αποτελείται από βίντεο καταγεγραμμένα σε πραγματικές συνθήκες και επομένως η αρχική στάση του ατόμου είναι αυθαίρετη και δεν υπόκειται σε χωρικούς περιορισμούς όπως 'το δεξί χέρι στη δεξιά πλευρά του κεφαλιού' ενώ είναι συχνές οι επικαλύψεις χεριών-κεφαλιού. Επιπρόσθετα, μερικές περιοχές παρόμοιας χρωματικής χροιάς με το δέρμα μπορούν να αποπροσανατολίσουν τον αλγόριθμο αυτόματης ανίχνευσης και παρακολούθησης. Για να αντιμετωπιστούν τα ανωτέρω προβλήματα απαιτείται μια χειροκίνητη αρχικοποίηση του αλγορίθμου παρακολούθησης. Κατά τη διάρκεια αυτής της διαδικασίας ο χρήστης επιβεβαιώνει τις περιοχές που προτείνονται από το σύστημα ως χέρια και κεφάλι του ατόμου που συμμετέχει στην τηλεοπτική συνέντευξη. Κατόπιν, δεδομένου ότι οι συνθήκες φωτισμού παραμένουν σταθερές κατά την διάρκεια της συνέντευξης, η ανίχνευση και παρακολούθηση εκτελούνται αυτόματα. Τέλος όπως συνήθως συμβαίνει στην επεξεργασία εικόνας είναι απαραίτητο να διαχωριστεί η καθεαυτή κίνηση του ατόμου από την φαινόμενη που προκαλείται από την αλλαγή θέσης της συσκευής λήψης. Στην προσέγγιση μας αντιμετωπίζουμε το πρόβλημα αυτό λαμβάνοντας υπόψη την σχετική θέση των χεριών και του κεφαλιού, εφόσον αυτή θα παραμείνει σχετικά σταθερή για μικρές αλλαγές στην θέση της κάμερας.

Η μέτρηση της μετακίνησης σε διαδοχικά πλαίσια υπολογίζεται ως το άθροισμα των κινούμενων εικονοστοιχείων που ανήκουν στις μάσκες δέρματος, κανονικοποιημένες με το εμβαδόν των περιοχών δέρματος. Η κανονικοποίηση εκτελείται προκειμένου να απαλειφθεί ο παράγοντας εστίασης της συσκευής, ο οποίος μπορεί να κάνει τις κινούμενες περιοχές δερμάτων να εμφανιστούν μεγαλύτερες χωρίς να συμβαίνει πραγματικά ζωηρή δραστηριότητα. Οι πιθανές κινούμενες περιοχές βρίσκονται έπειτα από κατωφλιοποίηση της διαφοράς των τιμών των εικονοστοιχείων μεταξύ του τρέχοντος και του επόμενου πλαισίου. Αυτή η μάσκα δεν περιέχει πληροφορίες σχετικά με την κατεύθυνση ή το μέγεθος της μετακίνησης, αλλά είναι ενδεικτική μόνο της κίνησης και χρησιμοποιείται για να επιταχύνει τον αλγόριθμο επικεντρώνοντας την παρακολούθηση σε κινούμενες περιοχές της εικόνας. Τόσο οι μάσκες χρώματος όσο και αυτές της κίνησης περιέχουν μεγάλο αριθμό μικρών αντικειμένων λόγω της παρουσίας θορύβου και αντικειμένων με χρώμα παρόμοιο με του δέρματος. Για να ξεπεραστεί αυτό, εφαρμόζουμε μορφολογικό φιλτράρισμα και στις δύο μάσκες για την αφαίρεση μικρών αντικειμένων. Ακολούθως, η κινούμενη μάσκα δέρματος παράγεται με τον συνδυασμό των επεξεργασμένων μασκών δέρματος και κίνησης και την μορφολογική ανακατασκευή της μάσκας χρώματος που χρησιμοποιεί τη μάσκα κινήσεων ως σημαδευτή.

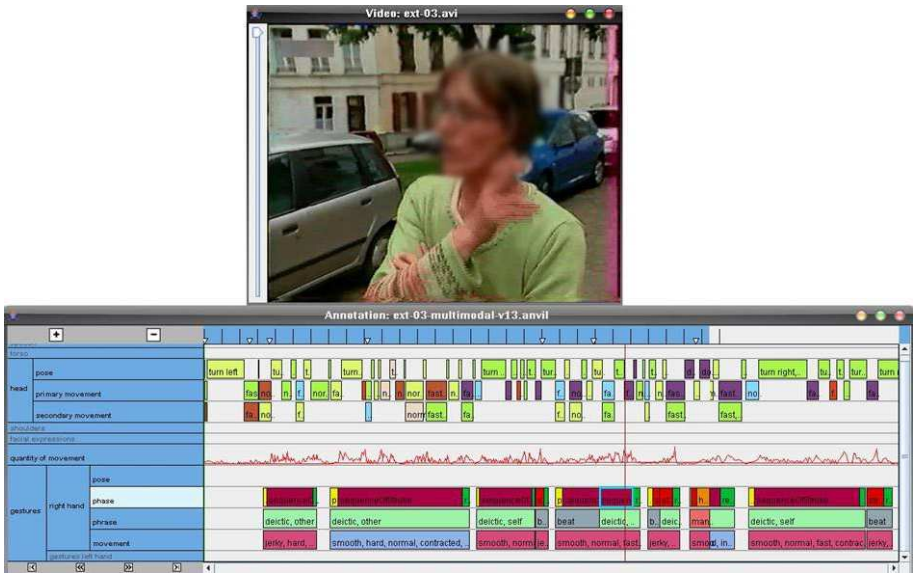
Η εξαγωγή των εκφραστικών παραμέτρων γίνεται με το σύνολο εξισώσεων που περιγράφονται παραπάνω στην ενότητα 3.2.4. Βέβαια λόγω της φύσης του τηλεοπτικού σώματος συνεντεύξεων και ανεξέλεγκτων παραμέτρων όπως το παρασκήνιο, τον αριθμό των ατόμων στην εικόνα, την στάση του σώματος και τον ρουχισμό έγιναν

κάποιες παραδοχές, υποθέσεις και συμβιβασμοί που αφορούσαν την κάθε περίπτωση συνέντευξης και όχι το σύνολο του σώματος. Όπως φαίνεται και από τον πίνακα 3.4 κάθε περίπτωση απαιτούσε διαφορετική προσέγγιση και αντιμετώπιση με τον παραπάνω περιγεγραμμένο αλγόριθμο να αποτελεί κοινό σημείο αναφοράς για όλες τις συνεντεύξεις. Χαρακτηριστικά παραδείγματα που δείχνουν την πολυπλοκότητα του τηλεοπτικού σώματος, με την λίστα να μπορούσε να είχε πολύ μεγαλύτερο αριθμό παραδειγμάτων, είναι:

- ext-03: Η συνεντευξιαζόμενη έχει τα χέρια σταυρωμένα και καλυμμένα για την μεγαλύτερη διάρκεια της συνέντευξης, ενώ κρατάει ένα παλτό με χρωματικά χαρακτηριστικά πολύ κοντά στο πρότυπο χρώματος που χρησιμοποιούμε
- ext-22: Η συνεντευξιαζόμενη φοράει μαγιό και έχει εκτεθειμένο το δέρμα σε όλο το μέρος του σώματος της που είναι εμφανές στην κάμερα, ενώ η στάση του σώματος της είναι πλευρική.
- ext-36: Στο παρασκήνιο εμφανίζονται και εξαφανίζονται άτομα μη σχετικά με την συνέντευξη καθώς γίνεται σε ένα μπαρ!

3.3.3 Σύγκριση χειρονακτικής και αυτόματης επεξεργασίας

Στην ενότητα αυτή επεξηγούμε την σύγκριση χειρωνακτικής και αυτόματης επεξεργασίας σε 10 βίντεο από το σώμα EmoTV (πίνακας 3.4).



Σχήμα 3.11: Ενσωμάτωση στο εργαλείο Anvil [136] των χειρωνακτικών σχολιασμών (χρωματισμένα τμήματα) των κινήσεων του κεφαλιού (ανώτερες γραμμές) και των χειρονομιών (χαμηλότερες γραμμές) και αποτελέσματα της επεξεργασίας εικόνας για την εκτίμηση της ποσότητας μετακίνησης (συνεχής γραμμή στο μέσον).

3.3.3.1 Γενική ενεργοποίηση συναισθηματικών συμπεριφορών σε κάθε βίντεο

Στον πίνακα 3.5 οι τιμές που λαμβάνονται για (1) τον χειρωνακτικό σχολιασμό της συναισθηματικής ενεργοποίησης (μέσος όρος των σχολιασμών από 3 ειδικούς), (2)

Πίνακας 3.4: Άτυπη περιγραφή των 10 βίντεο της μελέτης

Βίντεο	**Άτυπη περιγραφή**
01	Δρόμος ; κίνηση κεφαλιού, εκφράσεις προσώπου ; άνθρωποι κινούνται στο παρασκήνιο
02	Δρόμος ; κίνηση (κορμού, χεριών, κεφαλιού) ; άνθρωποι κινούνται στο παρασκήνιο
03	Δρόμος ; κίνηση (κορμού, χεριών, κεφαλιού) ; εκφράσεις προσώπου ; δερματική περιοχή στον κορμό
22	Παραλία ; κίνηση (κορμού, χεριών, κεφαλιού) ; εκφράσεις προσώπου ; πολλαπλές δερματικές περιοχές
36	Εσωτερικός χώρος ; κίνηση (χεριών, κεφαλιού) ; εκφράσεις προσώπου ; άνθρωποι κινούνται στο παρασκήνιο
41	Εσωτερικός χώρος ; κίνηση κεφαλιού ; εκφράσεις προσώπου
44	Εξωτερικός χώρος ; εκφράσεις προσώπου ; κίνηση (χεριών, κεφαλιού)
49	Εξωτερικός χώρος ; κίνηση (χεριών, κεφαλιού) ; εκφράσεις προσώπου ; άνθρωποι κινούνται στο παρασκήνιο
71	Εσωτερικός χώρος ; κίνηση (χεριών, κεφαλιού) ; εκφράσεις προσώπου ; άνθρωποι κινούνται στο παρασκήνιο
72	Εξωτερικός χώρος ; κίνηση (χεριών, κεφαλιού)

την αυτόματη εκτίμηση της ποσότητας μετακίνησης στο επίπεδο ολόκληρου του τηλεοπτικού βίντεο και (3) το ποσοστό των δευτερολέπτων για κάθε βίντεο που υπάρχει τουλάχιστον ένας χειρωνακτικός σχολιασμός μετακίνησης (κεφάλι, χέρι ή κορμός).

Αυτές οι τρεις μετρήσεις παρέχουν τις διαφορετικές εκτιμήσεις της ποσότητας ενεργοποίησης σχετιζόμενης με το συναίσθημα. Η ανάλυση συσχετισμού δείχνει ότι τα μέτρα (1) και (2) συσχετίζονται σημαντικά ($r = 0.64, p < 0,05$). Αυτό δείχνει ότι η αυτόματη επεξεργασία 10 βίντεό μας επικυρώνει το χειρωνακτικό σχολιασμό της ενεργοποίησης στο καθολικό επίπεδο κάθε βίντεο.

Η ανάλυση συσχετισμού επίσης δείχνει ότι οι μετρήσεις (1) και (3) μπορούν να επίσης να συσχετιστούν ($r = 0.49$). Τέλος, η ανάλυση συσχετισμού δείχνει ότι οι μετρήσεις (2) και (3) είναι μέτρια συσχετισμένες ($r = 0.42$). Εντούτοις, λόγω του μικρού αριθμού δειγμάτων, αυτά τα δύο τα μέτρα δεν αρκούν για να εξηγήσουν την στατιστική συσχέτιση μεταξύ των δύο προσεγγίσεων. Περισσότερα στοιχεία απαιτούνται για να επιβεβαιώσουν τα παραπάνω αποτελέσματα.

Πίνακας 3.5: Χειρωνακτική μέτρηση (1)(3) και αυτόματη (2) της καθολικής ενεργοποίησης στα 10 επιλεγμένα βίντεο

Βίντεο #	**(1) Χειρωνακτική 1: χαμηλή 5: υψηλή**	**(2) Αυτόματη**	**(3) Χειρωνακτική % sec με > 1 χειρωνακτικής στο Anvil**
Αναλυτές	3 ειδικοί	Σύστημα	1 ειδικός και επικύρωση από δευτερο
01	4	3398,50	81,2
02	3	269,64	72,9
03	4,33	1132,80	92,6
22	4,33	3282,80	81,1
36	4,66	2240,50	94,4
41	3	959,60	73,6
44	3,33	1771,30	92,3
49	4,33	1779,00	91,2
71	2,67	904,73	86,1
72	3,33	330,92	56,7

3.3.3.2 Εκτίμηση και επισημείωση κίνησης σε χρονικό επίπεδο

Στο επίπεδο του χρόνου, σχεδιάσαμε μια μέθοδο ώστε να συγκρίνουμε τον χειρωνακτικό σχολιασμό (των μετακινήσεων του κεφαλιού, των χεριών και του κορμού) με την αυτόματη εκτίμηση των μετακινήσεων. Το σχήμα 3.11 επιδεικνύει πώς και οι δύο τύποι σχολιασμών ενσωματώνονται στο εργαλείο Anvil [136].

Η τρέχουσα ενότητα επεξεργασίας εικόνας επιτρέπει να παραχθεί μια εκτίμηση μετακίνησης για κάθε πλαίσιο του βίντεο. Δεν παρέχει χωριστές εκτιμήσεις της μετακίνησης για τα διαφορετικά μέρη του σώματος για τους λόγους που εξηγήθηκαν στην ενότητα 3.3.2. Κατά συνέπεια, συγκρίναμε την ένωση των χειρωνακτικών σχολιασμών των μετακινήσεων στο κεφάλι, χέρια και κορμό με την αυτόματη εκτίμηση μετακίνησης για ολόκληρο το πλαίσιο. Όταν η ενότητα επεξεργασίας εικόνας ανιχνεύσει κίνηση, ελέγχουμε αν υπάρχει συμφωνία με το χειρωνακτικό σχολιασμό με τον έλεγχο ύπαρξης σε αυτόν μετακίνησης τουλάχιστον σε ένα από τα τρία μέρη του σώματος.

Οι συνεχείς τιμές εκτίμησης κίνησης που παρέχονται από την ενότητα επεξεργασία εικόνας υποβάλλονται σε κατωφλίωση προκειμένου να παραχθεί αυτόματος σχολιασμός των μετακινήσεων σε δυαδική μορφή και να μπορεί να συγκριθεί με τους χειρωνακτικούς σχολιασμούς. Ο ορισμός διαφορετικών τιμών κατωφλίου στην παραπάνω διαδικασία οδηγεί σε διαφορετικές τιμές συμφωνίας μεταξύ των χειρωνακτικών σχολιασμών και αυτόματης ανίχνευσης κίνησης. Η τιμή αυτού του κατωφλίου το οποίο προσδιορίζει την ύπαρξη μετακίνηση από την ενότητα επεξεργασίας εικόνας πρέπει να είναι η ελάχιστη τιμή σύμφωνα με την οποία μια μετακίνηση είναι αντιληπτή και έχει επισημειωθεί. Αξιολογήσαμε τη συμφωνία μεταξύ της ένωσης των

χειρωνακτικών σχολιασμών των μετακινήσεων και της εκτίμησης της μετακίνησης με αρκετές τιμές αυτού του κατωφλίου επάνω από το οποίο η ενότητα επεξεργασίας εικόνας αποφασίζει ότι μια μετακίνηση ανιχνεύεται. Οι πειραματικές τιμές για αυτό το κατώφλι ήταν μεταξύ 0.1% και 40% της μέγιστης τιμής της εκτίμησης καθολικής κίνησης σε κάθε βίντεο. Χρησιμοποιούμε ένα χρονικό παράθυρο 0.04 δευτερολέπτων για τον υπολογισμό της συμφωνίας μεταξύ των χειρωνακτικών και των αυτόματων σχολιασμών δεδομένου ότι αυτό είναι το διάστημα μεταξύ δύο διαδοχικών πλαισίων. Ο προκύπτων πίνακας σύγχυσης φαίνεται στον πίνακα 3.6. Ο βαθμός συμφωνίας είναι υψηλότερος για τα βίντεο 22 και 3 που χαρακτηρίζονται από πολλές μετακινήσεις (κεφάλι, χέρι) και κατά την διάρκεια των οποίων το δέρμα είναι ορατό στην ανώτερη περιοχή του κορμού και στα οποίο δεν υπάρχουν άλλα άτομα να κινούνται στο παρασκήνιο της εικόνας. Η χαμηλότερη συμφωνία λαμβάνεται για τα βίντεο 36 και 71 όπου εμφανίζονται άνθρωποι που κινούνται στο παρασκήνιο, η μετακίνηση των οποίων δεν έχει σχολιαστεί χειρωνακτικά δεδομένου ότι ο σχολιασμός αυτός εστιάζει αποκλειστικά στον συνεντευξιαζόμενο. Ενδιάμεσες τιμές λαμβάνονται για το βίντεο 41 που χαρακτηρίζεται μόνο από μικρές μετακινήσεις του κεφαλιού και περιστασιακές μετακινήσεις του κορμού. Γενικά τα βίντεο που περιλαμβάνουν συνεντεύξεις καταγραμμένες σε εξωτερικό χώρο επιτυγχάνουν υψηλότερες τιμές συμφωνίας παρά εκείνα που καταγράφονται σε εσωτερικό χώρο, αποκαλύπτοντας τον αντίκτυπο της τηλεοπτικής ποιότητας και των συνθηκών φωτισμού. Συμπερασματικά, δεν προκύπτει κάποια συστηματική σχέση από την ανάλυση των διαφωνιών πέρα από την παρατήρηση ότι για 6 βίντεο, ο αριθμός διαφωνιών τύπου 'αυτόματη 0 - χειρωνακτική 1' είναι υψηλότερος από τον αριθμό διαφωνιών τύπου 'αυτόματη 1 - χειρωνακτική 0'.

Πίνακας 3.6: Πίνακας σύγχυσης των συμφωνιών μεταξύ του χειρωνακτικού σχολιασμού της μετακίνησης και της αυτόματης εκτίμησης της ποσότητας μετακίνησης. Το κατώτατο όριο είναι ποσοστό της μέγιστης τιμής της εκτίμησης μετακίνησης.

Βίντεο	**Κατώφλι**	**Συμφωνία**		
		Αυτόματη 0 Χειρωνακτική 0	**Αυτόματη 1 Χειρωνακτική 1**	**Σύνολο**
01	0,004	0,050	0,799	0,849
02	0,004	0,203	0,611	0,814
03	0,001	0,009	0,892	0,901
22	0,016	0,113	0,799	0,912
36	0,001	0,039	0,449	0,489
41	0,002	0,186	0,483	0,669
44	0,001	0,063	0,550	0,613
49	0,003	0,013	0,858	0,871
71	0,042	0,139	0,307	0,446
72	0,047	0,340	0,355	0,695
ΜΟ	**0,012**	**0,115**	**0,610**	**0,726**

Η επιλογή των 10 βίντεο από το σώμα του EmoTV είναι πλούσια σε χειρωνακτικό σχολιασμό μετακινήσεων του χεριών, του κορμού και του κεφαλιού (παραδείγματος χάριν, το ποσοστό των πλαισίων για το οποίο δεν υπάρχει κανένας χειρωνακτικός σχολιασμός των μετακινήσεων είναι μόνο 26% για το βίντεο 41, 7% για το βίντεο

Πίνακας 3.7: Πίνακας σύγχυσης των διαφωνιών μεταξύ του χειρωνακτικού σχολιασμού της μετακίνησης και της αυτόματης εκτίμησης της ποσότητας μετακίνησης. Το κατώτατο όριο είναι ποσοστό της μέγιστης τιμής της εκτίμησης μετακίνησης.

Βίντεο	**Κατώφλι**	**Διαφωνία**		
		Αυτόματη 0 Χειρωνακτική 1	**Αυτόματη 1 Χειρωνακτική 0**	**Σύνολο**
01	0,004	0,014	0,138	0,151
02	0,004	0,118	0,068	0,186
03	0,001	0,034	0,065	0,099
22	0,016	0,013	0,075	0,088
36	0,001	0,494	0,017	0,511
41	0,002	0,254	0,077	0,331
44	0,001	0,373	0,014	0,387
49	0,003	0,054	0,075	0,129
71	0,042	0,554	0,000	0,554
72	0,047	0,213	0,092	0,305
ΜΟ	**0,012**	**0,212**	**0,062**	**0,274**

3 και 5% για βίντεο 36). Κατά συνέπεια, προκειμένου να είμαστε σε θέση να υπολογίσουμε τις στατιστικές τιμές της συμφωνίας μεταξύ των χειρωνακτικών και των αυτόματων σχολιασμών, εξισορροπήσαμε τον αριθμό πλαισίων με και αυτόν χωρίς χειρωνακτικό σχολιασμό με τον ακόλουθη διαδικασία: 1) τον υπολογισμό του αριθμού πλαισίων χωρίς οποιοδήποτε χειρωνακτικό σχολιασμό της μετακίνησης και 2) μια τυχαία επιλογή του ίδιου αριθμού πλαισίων με τουλάχιστον έναν χειρωνακτικό σχολιασμό της μετακίνησης. Ο προκύπτων πίνακας σύγχυσης φαίνεται στον πίνακα 3.8. Η νέα μέση τιμή συμφωνίας είναι υψηλότερη (0.794) από αυτή που λαμβάνεται στον πίνακα 3.6 χωρίς ισορροπημένο αριθμό πλαισίων (0.726). Ο πίνακας 3.9 επίσης αποκαλύπτει ότι οι διαφωνίες δεν είναι ισορροπημένες πια: ο αριθμός πλαισίων για τον οποίο υπήρξε τουλάχιστον ένας χειρωνακτικός σχολιασμός της μετακίνησης και για τον οποίο καμία μετακίνηση δεν ανιχνεύθηκε από την αυτόματη επεξεργασία είναι μεγαλύτερος από την αντιστροφή για 8 από τα 10 βίντεο.

Οι μέγιστες τιμές kappa και το κατώφλι για τις οποίες λήφθηκαν αυτές απαριθμούνται στον πίνακα 3.10, στήλη (1). Η προκύπτουσα σειρά τιμών kappa ποικίλλει μεταξύ 0.422 και 0.833 ανάλογα με το βίντεο. Αυτές οι τιμές μπορούν να θεωρηθούν μάλλον καλές λαμβάνοντας υπόψη την ποιότητα των βίντεο μας. Στα αποτελέσματα που περιγράφονται στον πίνακα 3.10 στήλη (1), επιλέξαμε τα κατώφλια σαν τιμές που παρέχουν τις μέγιστες τιμές kappa. Οι διαφορές στις τιμές των κατωφλίων με τα αντίστοιχα κατώφλια που λαμβάνονται για τα διαφορετικά βίντεο δείχνουν πως αυτή η τιμή πρέπει να προσαρμόζεται για κάθε βίντεο, πιθανώς λόγω των διαφορών στις τοποθεσίες και τις συνθήκες καταγραφής. Ερευνήσαμε τη χρήση των χρονικών διαστημάτων κάθε βίντεο κατά τη διάρκεια των οποίων ελάχιστη (ή καθόλου) μετακίνηση έγινε αντιληπτή. Υπολογίσαμε τη μέση εκτίμηση μετακίνησης που παράγεται από την αυτόματη ενότητα επεξεργασίας κατά τη διάρκεια καθενός από αυτά τα διαστήματα. Επιλέξαμε τον μέσο όρο των τιμών αυτών ως κατώφλι και η μέση τιμή του kappa μειώθηκε (πίνακας 3.10 στήλη (2)). Περαιτέρω πειράματα είναι κατά συνέπεια

Πίνακας 3.8: Πίνακας σύγχυσης των συμφωνιών μεταξύ του χειρωνακτικού σχολιασμού της μετακίνησης και της αυτόματης εκτίμηση της ποσότητας μετακίνησης για ένα ισορροπημένο σύνολο πλαισίων με και χωρίς χειρωνακτικό σχολιασμό της μετακίνησης

Βίντεο	**Κατώφλι**	**Συμφωνίες**		
		Αυτόματη 0 Χειρωνακτική 0	**Αυτόματη 1 Χειρωνακτική 1**	**Σύνολο**
01	0,047	0,353	0,358	0,711
02	0,013	0,418	0,407	0,825
03	0,076	0,442	0,423	0,865
22	0,071	0,450	0,467	0,917
36	0,034	0,500	0,300	0,800
41	0,008	0,421	0,283	0,704
44	0,010	0,460	0,316	0,776
49	0,084	0,476	0,357	0,833
71	0,048	0,500	0,283	0,783
72	0,044	0,385	0,345	0,730
ΜΟ	**0,043**	**0,440**	**0,354**	**0,794**

Πίνακας 3.9: Πίνακας σύγχυσης των διαφωνιών μεταξύ του χειρωνακτικού σχολιασμού της μετακίνησης και της αυτόματης εκτίμηση της ποσότητας μετακίνησης για ένα ισορροπημένο σύνολο πλαισίων με και χωρίς χειρωνακτικό σχολιασμό της μετακίνησης

Βίντεο	**Κατώφλι**	**Διαφωνίες**		
		Αυτόματη 0 Χειρωνακτική 1	**Αυτόματη 1 Χειρωνακτική 0**	**Σύνολο**
01	0,047	0,142	0,147	0,289
02	0,013	0,093	0,082	0,175
03	0,076	0,077	0,058	0,135
22	0,071	0,033	0,050	0,083
36	0,034	0,200	0,000	0,200
41	0,008	0,216	0,080	0,296
44	0,010	0,185	0,039	0,224
49	0,084	0,143	0,024	0,167
71	0,048	0,217	0,000	0,217
72	0,044	0,155	0,115	0,270
ΜΟ	**0,043**	**0,146**	**0,059**	**0,206**

απαραίτητα ώστε να μελετηθεί πώς αυτή η τιμή κατωφλίου μπορεί να τεθεί.

Πίνακας 3.10: Τιμές kappa που λαμβάνονται για τον ίδιο αριθμό πλαισίων που περιλαμβάνουν έναν τουλάχιστον χειρωνακτικό σχολιασμό της μετακίνησης και του αριθμού πλαισίων που δεν περιλαμβάνουν έναν χειρωνακτικό σχολιασμό της μετακίνησης: (1) Το κατώφλι για το οποίο η τιμή kappa είναι μέγιστη, (2) το κατώφλι που προέκυψε με τον υπολογισμό του μέσου όρου της αυτόματης εκτίμησης της μετακίνησης 2 δευτερολέπτων του βίντεο για τα οποία καμία το μετακίνηση δεν μπορεί να γίνει αντιληπτή.

	(1) Κατώφλι που αντιστοιχεί στο μέγιστο kappa		**(2) Κατώφλι που προκύπτει από 2s χωρίς κίνηση**	
Βίντεο #	**Μέγιστο kappa**	**Κατώφλι**	**Kappa**	**Κατώφλι**
01	0,422	0,047	0,275	0,056
02	0,649	0,013	0,547	0,016
03	0,731	0,076	0,57	0,059
22	0,833	0,071	0,633	0,079
36	0,600	0,034	0,6	0,037
41	0,407	0,008	0,342	0,039
44	0,553	0,01	0,19	0,066
49	0,667	0,084	0,428	0,089
71	0,565	0,048	0,304	0,008
72	0,459	0,044	0,327	0,034
ΜΟ	**0,589**	**0,043**	**0,421**	**0,048**

Κεφάλαιο 4

Αναγνώριση και σύνθεση χειρονομιών και Ελληνικής Νοηματικής Γλώσσας

4.1 Ερευνητικό πλαίσιο

4.1.1 Επισκόπηση Μεθόδων Αναγνώρισης Χειρονομιών

Η αναγνώριση χειρονομιών και η βασισμένη σε χειρονομίες αλληλεπίδραση ανθρώπου-μηχανής (Gesture Based Human Computer Interaction) προσελκύουν όλο και περισσότερο την προσοχή ερευνητών από ερευνητικές περιοχές όπως η μηχανική μάθηση, η αναγνώριση προτύπων, η όραση υπολογιστών, η αλληλεπίδραση ανθρώπου-μηχανής, η γλωσσολογία και η επεξεργασία φυσικής γλώσσας. Αυτός ο διεπιστημονικός ερευνητικός τομέας βρίσκει πεδίο εφαρμογής επίσης σε αρκετές περιοχές όπως πολυτροπική αλληλεπίδραση ανθρώπου υπολογιστή, συστήματα αυτομάτου ελέγχου, ρομποτική, συναισθηματική υπολογιστική και συμπεριφορισμός, αναγνώριση νοηματικής γλώσσας, βοηθητικές τεχνολογίες απομακρυσμένης μάθησης και πλοήγηση σε εικονικά περιβάλλοντα. Η αλληλεπίδραση ανθρώπου-μηχανής καθορίζει συνεχώς νέες μορφές επικοινωνίας και διεπαφής με τις υπολογιστικές μηχανές [27]. Οι χειρονομίες μπορούν να μεταβιβάσουν πληροφορίες για τις οποίες άλλες μορφές αλληλεπίδρασης (π.χ. ομιλία) δεν είναι αποδοτικές ή κατάλληλες. Στα πλαίσια της φυσικής και φιλικής προς το χρήστη αλληλεπίδρασης, οι χειρονομίες μπορούν να χρησιμοποιηθούν, μονοτροπικά, ή να συνδυαστούν με άλλες μορφές πληροφορίας σε πολυτροπικές αρχιτεκτονικές αλληλεπίδρασης που περιλαμβάνουν ομιλία, ή ακόμα και κείμενο [23]. Η αναγνώριση συναισθήματος και η εξαγωγή εκφραστικών παραμέτρων, όπως φαίνεται στα αντίστοιχα κεφάλαια της διατριβής, είναι μια άλλη περιοχή όπου η ανάλυση χειρονομιών αποδεικνύεται χρήσιμη και παρέχει σημαντικές ενδείξεις στο πλαίσιο της αναγνώρισης συναισθήματος από πολλαπλές μορφές πληροφορίας σε φυσική αλληλεπίδραση του χρήστη με το υπολογιστικό σύστημα [49].

Μια χειρονομία είναι μια κίνηση του σώματος που μεταβιβάζει πληροφορία. Μια ταξονομία χειρονομιών μπορεί να διατυπωθεί σε μια συνεχής σειρά: ελεύθερη χειρονομία, συνδεδεμένη με την ομιλία, παντομίμα, συμβολικές και τέλος, νοηματικές γλώσσες όπως προτείνεται από τον Kendon στο [132]. Μια εναλλακτική ταξονομία χειρονομιών μπορεί να οριστεί σύμφωνα με τη λειτουργία τους:

- συμβολικές χειρονομίες: χειρονομίες που, σε κάθε πολιτισμό, έχουν αποκτήσει μια ιδιαίτερη και ενιαία έννοια.

- δεικτικές χειρονομίες: οι τύποι χειρονομιών που εμφανίζονται συχνότερα σε HCI περιπτώσεις και είναι χειρονομίες υπόδειξης οντότητας ή κατεύθυνσης

- εικονικές χειρονομίες: χειρονομίες που χρησιμοποιούνται για να μεταβιβάσουν πληροφορίες σχετικές με το μέγεθος, την χωρική σχέση, την ενέργεια, τη μορφή ή τον προσανατολισμό του αντικειμένου στην συνομιλία.

- χειρονομίες μιμητικές: χειρονομίες που χρησιμοποιούνται τυπικά για να μιμηθούν μια δράση, ένα αντικείμενο ή μια έννοια

Υπάρχει αφθονία προσεγγίσεων και μεθοδολογιών για την αναγνώριση χειρονομιών που παρουσιάζονται επαρκώς στο [164], [170] και [253]. Οι Mitra και Acharya εστιάζουν στην αναγνώριση χειρονομιών, ενώ οι Ong και Ranganath επεκτείνουν την έρευνά τους στην αυτόματη αναγνώριση νοηματικής γλώσσας. Και οι δύο επισκοπήσεις εξετάζουν τεχνικές εξαγωγής χαρακτηριστικών γνωρισμάτων και ζητήματα κατηγοριοποίησης σχετικά με την αυτόματη ανάλυση χειρονομιών. Οι Wu και Huang εστιάζουν περισσότερο στην προτυποποίηση του χεριού (ανάλυση σχήματος, αλυσίδα κινηματικής και δυναμική) και ζητημάτων όρασης υπολογιστών και αναγνώρισης προτύπων που σχετίζονται με τον εντοπισμό και την παρακολούθηση των χεριών αλλά και την εξαγωγή χαρακτηριστικών γνωρισμάτων από ακολουθίες εικόνων. Ενώ στο [3] παρουσιάζεται ένα ενοποιημένο πλαίσιο που αντιμετωπίζει τόσο το πρόβλημα της χρονικής και χωρικής κατάτμησης των χειρονομιών και των χεριών αντίστοιχα αλλά και την αναγνώριση των χειρονομιών.

Η είσοδος που χρησιμοποιείται σε κάθε προσέγγιση αναγνώρισης χειρονομιών μπορεί ευρέως να ομαδοποιηθεί σε δύο κυρίαρχες κατηγορίες: την καταγραφή κίνησης με συσκευές άμεσης μέτρησης όπως γάντια δεδομένων και την είσοδος βίντεο ενώ κάμερες μέτρησης χρόνου πτήσης (time of flight cameras) ή επιταχυνσιόμετρα έχουν επίσης χρησιμοποιηθεί. Ενώ τα γάντια καταγραφής δεδομένων είναι μια αρκετά πολυδάπανη και παρεισφρητική λύση παρέχουν μια πιο εύρωστη, ακριβής, λεπτομερή και αποδοτική προσέγγιση στην καταγραφή της τριδιάστατης θέσης των χεριών και κάμψης των αρθρώσεων των δάχτυλων σε πραγματικό χρόνο όταν συγκρίνονται με βασισμένες στην εικόνα προσεγγίσεις. Δημοφιλείς υποβοηθούμενες από συσκευές λύσεις στην καταγραφή πληροφοριών των χεριών είναι τα Virtual Realities CyberGloves για μετρήσεις γωνιών κάμψης των αρθρώσεων των δαχτύλων και οι Polhemus συσκευές καταγραφής τριδιάστατης θέσης. Άλλες συσκευές περιλαμβάνουν τα VPL DataGlove, Mattel Power-Glove, AcceleGlove, EMI-Gloves, and the Flock of Birds. Από την προοπτική της εφαρμοσιμότητας και οι δύο βασικές προσεγγίσεις μπορούν να θεωρηθούν κατάλληλες αν και οι βασισμένες στην εικόνα προσεγγίσεις έχουν ευρύτερο πεδίο εφαρμογής. Οι βασισμένες σε γάντι καταγραφής προσεγγίσεις μπορούν να εγκατασταθούν εύκολα σε εσωτερικούς χώρους ή σε περιπτώσεις όπου η ακρίβεια καταγραφής είναι κρίσιμη. Η είσοδος βίντεο από την άλλη θα μπορούσε να χρησιμοποιηθεί στην πλειονότητα των περιπτώσεων με τον χρήστη να μην είναι υποχρεωμένος να εξοπλιστεί εξειδικευμένες συσκευές που πιθανόν να επηρεάσουν την συμπεριφορά του και να προκαλέσουν αφύσικες αντιδράσεις.

Ενώ το πλεονέκτημα της εισόδου από εικόνα είναι πως δεν είναι παρεισφρητική, συχνά αρκετές υποθέσεις πρέπει να γίνουν ή περιορισμοί να εφαρμοστούν κατά την διαδικασία καταγραφής που αφορούν είτε στο περιβάλλον, είτε στον χρήστη είτε στην ρύθμιση της κάμερας. Συνηθέστερα όταν χρησιμοποιείται κάποιο πρότυπο χρωματικής χροιάς δέρματος ο χρήστης καλείται να φορέσει μακριά μανίκια και κατάλληλο

ρουχισμό ώστε πολλές περιοχές δέρματος, που δεν είναι πρόσωπο ή χέρια, να μην είναι εκτεθειμένες (π.χ. ντεκολτέ, ρούχα παρόμοιας χρωματικής χροιάς με το δέρμα, κ.α.). Η παρουσία τέτοιων περιοχών μπορεί να προκαλέσει σύγχυση ή ακόμα και κατάρρευση στους αλγορίθμους ανίχνευσης χεριών και να παρεμποδίσει τον εντοπισμό και την παρακολούθηση των χεριών. Επιπλέον, το παρασκήνιο είναι εξαιρετικής σπουδαιότητας σε κάποιες προσεγγίσεις δεδομένου ότι εκτελείται αφαίρεση υποβάθρου ή κάποιος άλλος τελεστής επεξεργασίας εικόνας που απαιτεί ομοιόμορφο ή/και στατικό παρασκήνιο.

Σπάνια όμως και σε περιπτώσεις ολιστικής αντιμετώπισης, τα πρωτογενή δεδομένα καταγραφής χρησιμοποιούνται άμεσα στην διαδικασία κατηγοριοποίησης αλλά αντίθετα υπόκεινται περαιτέρω επεξεργασία και χαρακτηριστικά γνωρίσματα εξάγονται από αυτά. Σχεδόν όλες οι τεχνικές εξαγωγής χαρακτηριστικών γνωρισμάτων περιλαμβάνουν την θέση των χεριών. Συνήθως όταν υιοθετείται η καταγραφή κίνησης η τριδιάστατη θέση περιλαμβάνεται στο σύνολο χαρακτηριστικών γνωρισμάτων ενώ στις περιπτώσεις της οπτικής εισόδου μόνο η διδιάστατη προβολή της θέσης χεριών εξάγεται και η τριδιάστατη θέση μπορεί μόνο να υπολογιστεί με την χρήση στερέοψης. Η θέση του χεριού στις περισσότερες περιπτώσεις είναι σχετική με κάποιο σημείο αναφοράς π.χ. το κεφάλι του χρήστη ή της πλάτης του στην περίπτωση καταγραφής δεδομένων, όπου τοποθετείται ένας πρόσθετος αισθητήρας στην πλάτη του χρήστη. Οι προσεγγίσεις που χρησιμοποιούν οπτική είσοδο υπολογίζουν προσεγγιστικά το κέντρο του χεριού με την χρήση του κέντρου βάρους της περιοχής που έχει εξαχθεί ως χέρι (center of gravity) και το σημείο αυτό θεωρείται ως θέση του χεριού. Ο χαρακτηρισμός της χειρομορφής επιτυγχάνεται στην πλειοψηφία των περιπτώσεων με δύο τρόπους. Για προσεγγίσεις που βασίζονται σε συσκευή καταγραφής (γάντι) το σύνολο των γωνιών κάμψης των αρθρώσεων των δαχτύλων θεωρείται επαρκές σύνολο ικανό να περιγράψει την διενεργηθείσα χειρομορφή, ενώ στις προσεγγίσεις με οπτική είσοδο, είτε επιχειρείται να υπολογιστεί το μήκος και ο άξονας των διανυσμάτων μεταξύ του κέντρου βάρους του χεριού ή του καρπού του χεριού και του άκρου κάθε δακτύλου είτε επιχειρείται να χαρακτηριστεί η χειρομορφή μέσω χαρακτηριστικών περιοχής ή περιγραφείς Fourier ή/και καμπυλότητας. Τα σύνολα αυτά μπορούν να θεωρηθούν ως διαφοροποιήσιμα σύνολα χαρακτηριστικών γνωρισμάτων. Τα εξαγχθέντα χαρακτηριστικά γνωρίσματα μπορούν να υποβληθούν σε περαιτέρω επεξεργασία, μια διαδικασία που προηγείται της φάσης της κατηγοριοποίησης, εφαρμόζοντας Eigen ανάλυση [235, 216, 268], κανονικοποίηση, συσταδοποίηση, ανάλυση πρωτευουσών συνιστωσών, μετασχηματισμό Fourier, κ.λπ. Επιπλέον, στις περισσότερες προσεγγίσεις που δέχονται ως είσοδο εικόνα, όπου εντοπίζονται περιοχές χεριών, υπολογίζονται επίσης χαρακτηριστικά περιοχής (area features). Αυτά τα χαρακτηριστικά γνωρίσματα περιλαμβάνουν το μέγεθος, αναλογία μηκών και γωνία αξόνων, εκκεντρικότητα, στερεότητα (solidity), απόκλιση, παραμόρφωση και καμπυλότητα.

Στην ευρύτερη αρχιτεκτονική αναγνώρισης χειρονομιών η επιλογή της μεθόδου ταξινόμησης θεωρείται αποφασιστικότερος παράγοντας μεταξύ των άλλων συστατικών της αρχιτεκτονικής και σε πολλές περιπτώσεις καθορίζει ή επηρεάζει σε μεγάλο βαθμό τον σχεδιασμό και των υπόλοιπων συστατικών. Αν και ένας μεγάλος αριθμός άρθρων υιοθετεί κάποια καθιερωμένη και δοκιμασμένη σε άλλα προβλήματα λύση (π.χ. HMM) υπάρχει σημαντικός αριθμός προσεγγίσεων που χρησιμοποιούν έναν συνδυασμό σχημάτων ταξινόμησης είτε για την προεπεξεργασία της εισόδου ή των εξαγόμενων χαρακτηριστικών γνωρισμάτων είτε για την μετέπειτα επεξεργασία της απόφασης κάθε ταξινομητή σε οποιαδήποτε διάταξη έχει επιλεγεί (παράλληλη είτε σειριακή) στη

συνολική αρχιτεκτονική του γενικού ταξινομητή.

Στο [43] οι ερευνητές παρουσιάζουν μια διαδεδομένη αρχιτεκτονική που περιλαμβάνει (βασισμένη σε χρωματικό πρότυπο) ανίχνευση χεριών, PCA ταξινόμηση για την χειρομορφή και διακριτά HMMs για την πληροφορία θέσης. Ενώ αναφέρουν ένα ικανοποιητικό ποσοστό αναγνώρισης (94.5%) για τη αναγνώριση στατικής χειρομορφής, στα αντίστοιχα πειράματα σε δύο άτομα για δυναμική χειρονομία, το ποσοστό αναγνώρισης κυμαίνεται μεταξύ 83% και 98.6% ανάλογα με την αναλογία μεγέθους συνόλων εκπαίδευσης/επαλήθευσης. Αν και ο συνδυασμός χειρομορφής και θέσης των χεριών παρουσιάζει αρκετό ενδιαφέρον ως προσέγγιση τα διακριτά HMMs δεν δείχνουν ικανά να μοντελοποιήσουν επαρκώς την χειρομορφή και να αντιμετωπίσουν τις παραλλαγές στην εκτέλεση της χειρονομίας είτε από τον ίδιο είτε από διαφορετικούς χρήστες. Η προσέγγιση που προτείνεται στο [256] είναι αρκετά παρόμοια και διαφέρει μόνο στην διαδικασία προεπεξεργασίας, όπου υιοθετείται ο k-means αλγόριθμος συσταδοποίησης, στο σύνολο χαρακτηριστικών γνωρισμάτων και στον τύπο των HMMs, όπου είναι συνεχή και αριστερά-προς-δεξιά (continuous, left-to-right). Εισάγουν επίσης έναν αλγόριθμο εντοπισμού χειρονομιών που χωρίζει την τροχιά χεριών σε ουσιαστικά και χωρίς περιεχόμενα τμήματα. Συνεχή HMM χρησιμοποιούνται επίσης από [112] αλλά η ανίχνευση και η παρακολούθηση των χεριών βασίζονται στα Μοντέλα Ενεργού Σχήματος. Στο [110], σε μία προσπάθεια να διαχωριστούν χειρονομίες 'αναζήτησης προσοχής' (attention seeking), παρουσιάζεται μια παραλλαγή HMM, κωδικοποίησης ρητής και υποννοούμενης χρονικής πληροφορίας (Implicit/Explicit Temporal Information Encoded), προκειμένου να παραμετροποιηθεί η πιθανότητα εκπομπής (emission probability) στις καταστάσεις που ανήκουν στο κρυφό επίπεδο. Το [156] εστιάζει περισσότερο στην εφαρμοσιμότητα, της μάλλον απλοικής προσέγγισης της χρήσης SOM για στατικές χειρονομίες και HMM για δυναμικές, σε φορητές συσκευές και χαρακτηριστικά γνωρίσματα που καταγράφονται από επιταχυνσιόμετρα. Στο [162] οι χειρονομίες αποσυντίθενται σε πρωτογενείς μονάδες σε μία προσπάθεια να βελτιωθεί η εφαρμοσιμότητα της αρχιτεκτονικής σε μεγάλες κλίμακες μειώνοντας την πολυπλοκότητα. Αυτές οι πρωτογενείς μονάδες χειρνονομίας (gesture primitives) συνθέτουν έννοιες και συνδέονται με μια βάση γνώσης χρησιμοποιώντας μια προσεγγιστική εννοιολογική τεχνική ταύτισης γράφων. Τα HMMs χρησιμοποιούνται πάλι για τον προσδιορισμό των πρωτογενών μονάδων αλλά η προσθήκη ενός επιπέδου αναπαράστασης και επεξεργασίας γνώσης σε συνδυασμό με την συμπερασματολογία στη ευρύτερη αρχιτεκτονική είναι μια ενδιαφέρουσα κατεύθυνση και αναμφίβολα χρήζει περαιτέρω διερεύνησης. Άλλη μια δημοσίευση προσανατολισμένη σε εφαρμογή της αυτόματης αναγνώρισης χειρονομιών είναι και η [189] που μελετά την αναγνώρισης χειρονομιών χρησιμοποιώντας ώς είσοδο την κίνηση του δαχτύλου μέσω ενός κυκλικού πλέγματος λέιζερ και της κατηγοριοποίησης με την χρήση HMM με είσοδο την ταχύτητα, την επιτάχυνση και την κατεύθυνση του δαχτύλου. Μια τριδιάστατη επέκταση στο υλοποιημένο σύστημα συζητείται επίσης. Στο [202] επιχειρείται να περιγραφούν HMM που αντιλαμβάνονται και προσαρμόζονται σε παραμέτρους που μεταβάλλονται ανάλογα με τις συνθήκες (context aware) όπου ο έλεγχος πλαισίου επιτυγχάνεται με την προσθήκη μιας τιμής του σχετικού πλαισίου στο διάνυσμα χαρακτηριστικών γνωρισμάτων. Η αβεβαιότητα στην διαδικασία εξαγωγής χαρακτηριστικών που προκαλείται από ποικίλες αλλαγές στις συνθήκες φωτισμού ή υποβάθρου συζητείται στο [206] προτείνοντας μια επέκταση στα HMM σύμφωνα με την οποία τιμές βεβαιότητας συνοδεύουν τα υποψήφια διανύσματα χαρακτηριστικών γνωρισμάτων και επιλέγεται το διάνυσμα που μεγιστοποιεί την πιθανότητα παρατή-

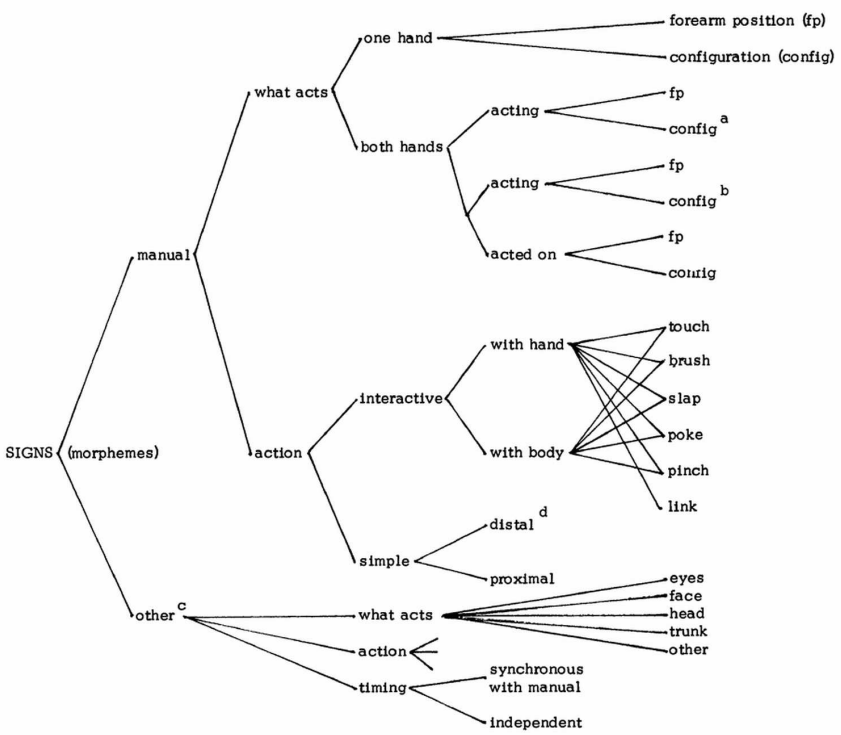
ρησης την δεδομένη περίπτωση επιλέγεται. Το [249] αντιμετωπίζει το πρόβλημα της συστηματικής απόκλιση στην έξοδο αισθητήρων ή τις συνθήκες του περιβάλλοντος, περιπτώσεις όπου συμβατικά HMM υστερούν, με την παρουσίαση της παραμετρικής παραλλαγής HMM όπου η εκπαίδευση υποθέτει ότι κάθε διάνυσμα χαρακτηριστικών γνωρισμάτων εισόδου συνοδεύεται με την τιμή της ελεύθερης παραμέτρου.

4.1.2 Επισκόπηση Μεθόδων Αναγνώρισης Νοηματικής Γλώσσας

Η νοηματική γλώσσα είναι το γλωσσικό σύστημα που χρησιμοποιείται, από την ομάδα ατόμων με ακουστική αναπηρία, προκειμένου να επικοινωνήσουν μεταξύ τους αλλά και με ομιλούντες ανθρώπους. Αντίθετα με τις προφορικές γλώσσες, οι νοηματικές γλώσσες (ΝΓ) είναι ισχυρά βασισμένες στην εικονικότητα προκειμένου να μεταβιβάσουν έννοιες. Μια μορφοσυντακτική δομή υιοθετείται για να εκφράσει γλωσσικές σχέσεις στον τριδιάστατο χώρο (χώρο νοηματισμού) και οργανώνεται πολύ διαφορετικά από τις προφορικά αρθρωμένες γλώσσες. Οι έννοιες αναπαριστώνται από τα νοήματα, την βασική γραμματική μονάδα μιας νοηματικής γλώσσας, που διαμορφώνει μια φυσική οπτική γλώσσα.

Αν και οι νοηματικές γλώσσες χρησιμοποιούνται από έναν σημαντικό αριθμό ανθρώπων, ακριβείς στατιστικές είναι δύσκολο να καταγραφούν λόγω των ανομοιόμορφων κριτηρίων που χρησιμοποιούνται στις διαφορετικές προσπάθειες καταγραφής των χρηστών νοηματικής γλώσσας σε κάθε χώρα [163]. Οι χρήστες νοηματικής γλώσσας δεν πρέπει να συγχέονται με τον πληθυσμό με ακουστική αναπηρία δεδομένου ότι τα δύο σύνολα δεν είναι ταυτόσημα, παρά την ύπαρξη σημαντικής επικάλυψης μεταξύ των δύο. Υπάρχει αρκετά μεγάλη ασάφεια σχετικά με τα στατιστικά δεδομένα των κωφών δεδομένου ότι άνθρωποι που παρουσίασαν προβλήματα κώφωσης σε προχωρημένο στάδιο στην ζωή τους δεν θεωρούνται πάντα κωφοί και ανάλογα το που τίθεται η διαχωριστική γραμμή, τα ποσοστά ποικίλουν σημαντικά. Επίσης, ο τρόπος στατιστικής παρουσίασης ποικίλλει σημαντικά και πολλές μελέτες εκθέτουν δεδομένα μόνο για τα αστικά κέντρα και αγνοούν ή γενικεύουν για τις υπόλοιπες περιοχές προσθέτοντας αβεβαιότητα στις συγκεντρωτικές στατιστικές. Παρά τις ασάφειες αυτές, οι αριθμοί προσεγγίζουν την τάξη των 30 εκατομμυρίων παγκοσμίως, ενώ μελέτες του πανεπιστημίου Gallaudet αναφέρουν ένα ποσοστό 2–5% του γενικού πληθυσμού. Οι εκ γενετής κωφοί (σε αντίθεση με αυτούς που παρουσίασαν προβλήματα κώφωσης αργότερα στην ζωή τους) υπολογίζονται στο 1% του πληθυσμού της Ευρώπης. Οι περισσότεροι κωφοί χρησιμοποιούν την νοηματική γλώσσα ως πρωταρχική γλώσσα τους και η ικανότητα τους στην προφορική γλώσσα είναι συχνά εξαιρετικά περιορισμένη, ενώ η ικανότητα τους στην ανάγνωση και την γραφή είναι ισοδύναμες με αυτές ενός παιδιού που παρακολουθεί τις πρώτες τάξεις του δημοτικού σχολείου [173].

Στο σύνολο τους τα νοήματα, με λίγες εξαιρέσεις συνήθως εξαρτώμενες από τις συνθήκες, αρθρώνονται σε έναν νοητό κύβο μπροστά από το κεφάλι και το σώμα του νοηματιστή, τον αποκαλούμενο χώρο νοηματισμού. Η διερεύνηση των δυνατοτήτων του χώρου νοηματισμού, την εικονικότητα καθώς επίσης και την παραγωγική χρήση των λεξικολογικών χαρακτηριστικών γνωρισμάτων όπως οι ταξινομητές, πολλές έννοιες μπορούν να μεταβιβαστούν χωρίς την χρήση καθιερωμένων λημμάτων. Οι νοηματικές γλώσσες μπορούν επομένως να αντεπεξέλθουν ικανοποιητικά στις απαιτήσεις των περισσοτέρων προφορικών γλωσσών με αρκετά μικρότερο αριθμό λημμάτων. Συνεπώς, τα μοντέλα και η γραμματική των νοηματικών γλωσσών πρέπει να είναι σε θέση να περιγράψουν επαρκώς τα αντωνυμικά συστήματα εκμεταλλευόμενα τον



Σχήμα 4.1: Δομή συστατικών νοηματικής από τον Stokoe [218]

προσδιορισμό της θέσης στον τριδιάστατο χώρο μπροστά από τον νοηματιστή και έτσι να καθοδηγούν τη σύνθεση εικονικών πρακτόρων και τα συστατικά συστημάτων αυτόματης αναγνώρισης αυτής της μοναδικής πτυχής της νοηματικής γλώσσας. Τα βασικά χαρακτηριστικά της νοηματικής γλώσσας, όπως φαίνεται στο σχήμα 4.1, περιλαμβάνουν:

- την χειρομορφή: το προφανέστερο χαρακτηριστικό ενός νοήματος είναι η μορφή του χεριού στην αρχή αλλά και κατά τη διάρκεια του νοήματος

- τον προσανατολισμό της παλάμης: η κατεύθυνση του διανύσματος κάθετου στο επίπεδο της παλάμης με κατεύθυνση προς το εσωτερικό της

- η θέση: υπάρχουν μόνο ορισμένες περιοχές πάνω ή κοντά στο σώμα του νοηματιστή όπου τα χέρια μπορούν να βρεθούν κατά την εκτέλεση ενός νοήματος

- την κίνηση: ένα σημαντικό ποσό της έννοιας του νοήματος εκφράζεται μέσω της μετακίνησης των χεριών. Η μετακίνηση μπορεί να μεταβιβάσει πληροφορίες για τον δράστη ή τον παραλήπτη μιας δράσης, ή ακόμα και τη σημασιολογική κατηγορία του αντικειμένου που περιλαμβάνεται σε αυτή. Η επανάληψη της μετακίνησης μπορεί περαιτέρω να δηλώσει συχνότητα, πολλαπλότητα ή γραμματική διαφοροποίηση κατηγορίας, π.χ. μεταξύ του ρήματος και του ουσιαστικού (ένα νόημα με απλή μετακίνηση μπορεί να υποδείξει τη λειτουργία ρήματος, ενώ η επανάληψη της ίδιας κίνησης σε ένα νόημα μπορεί να δείξει το αντίστοιχο ουσιαστικό). Η επέκταση της μετακίνησης μπορεί να υποδείξει μέγεθος ή όγκο, ενώ η ταχύτητα ή η δύναμη σε συνδυασμό με τα κατάλληλα μη χειρωνακτικά χαρακτηριστικά, μπορεί να εκφράσει μια σειρά επιρρηματικών ιδιοτήτων.

- τα μη χειρωνακτικά χαρακτηριστικά: η θέση του σώματος και η έκφραση του προσώπου μπορούν να λειτουργήσουν αφ' ενός ως δείκτης ρόλου όπως δράστης, δέκτης, κ.λπ. ή αντίστοιχα να υποδείξουν επιρρηματική και συντακτική συμφωνία. Αφ' ετέρου, μπορούν να εμπλουτίσουν το νόημα με, κατά μια έννοια, προσωδική πληροφορία. Στη δεύτερη περίπτωση, προσθέτουν σημασιολογικές ιδιότητες όπως η έγκριση, ο θαυμασμός, η απόρριψη, κ.λπ. στις γραμματικά δομημένες ουδέτερες ακολουθίες νοημάτων. Οι σχηματισμοί του στόματος και οι εκφράσεις του προσώπου, κυρίως μετακίνηση φρυδιών και βλέμματος, μαζί με τη μετακίνηση σώματος και ώμων συνιστούν το πολυεπίπεδο σύστημα πληροφορίας, τα χαρακτηριστικά γνωρίσματα του οποίου, εάν απαιτούνται, συμμετέχουν υποχρεωτικά στο σχηματισμό νοημάτων.

Σχήμα 4.2: Ενδεικτικό παράδειγμα των συστατικών νοηματικής γλώσσας

Τα γλωσσικά συστατικά της νοηματικής γλώσσας συνεπάγονται ένα κλειστό σύνολο ζευγαριών γνωρίσματος-τιμής, τα οποία συνιστούν τη γλωσσική φωνολογία της νοηματικής. Οι διάφοροι τύποι συστατικών τμημάτων νοηματικής μπορούν να αναπαράγουν κάθε πιθανό υπάρχον ή νέο σημάδι. Κατά συνέπεια είναι κρίσιμο οι μορφολογικές προσεγγίσεις επισημείωσης να είναι αρκετά πλούσιες σε χαρακτηριστικά γνωρίσματα κατάλληλα να αποτελέσουν πρωτόκολλο επικοινωνίας μεταξύ τεχνολογιών εικονικής σύνθεσης, αυτόματης αναγνώρισης και γλωσσολογικής επεξεργασίας. Ένα διακριτικό χαρακτηριστικό των νοηματικών γλωσσών είναι η εκτενής χρήση ταξινομητών, ενός συνόλου δεικτών κατηγορίας, σχήματος, κατάταξης, κ.λπ. (π.χ. άνθρωπος, ζώο, όχημα, στρογγυλός, τετραγωνικός), οι οποίοι ολοκληρώνουν ή τροποποιούν την έννοια που μεταβιβάζεται από κάποιο νόημα. Οι ταξινομητές παρέχουν έναν ισχυρό μηχανισμό για αναπαράσταση νέων εννοιών αλλά και για την τροποποίηση της έννοιας υπαρχόντων λημμάτων. Τα νοήματα διαμορφώνονται περαιτέρω μέσω της εκτέλεσης είτε με το ένα είτε με δύο χέρια. Στην περίπτωση των νοημάτων με δύο χέρια, υπάρχουν τρεις επιλογές:

- και τα δύο χέρια διαμορφώνουν την ίδια χειρομορφή και διαγράφοντας την ίδια κίνηση
- και τα δύο χέρια διαμορφώνουν την ίδια χειρομορφή αλλά μόνο το κυρίαρχο χέρι εκτελεί κίνηση

- τα δύο χέρια διαμορφώνουν διαφορετικές χειρομορφές αλλά μόνο το κυρίαρχο χέρι κινείται, με το μη κυρίαρχο χέρι να μπορεί να λάβει έναν περιορισμένο αριθμό βασικών χειρονομιών

Μια περισσότερο υπολογιστική ταξονομία συστατικών είναι η HA/TAB/SIG/DEZ, όπως προτείνεται στο [25]:

- HA: οι σχετικές θέσεις των χεριών
- TAB: η θέση των χεριών από την άποψη της εγγύτητάς τους σε καίρια σημεία του σώματος
- SIG: η μετακίνηση των χεριών
- DEZ: η χειρομορφή
- ORI: προσανατολισμός παλάμης

4.1.2.1 Προκλήσεις στην αυτόματη αναγνώριση νοηματικής γλώσσας

Οι χειρονομίες που χρησιμοποιούνται στις νοηματικές γλώσσες συχνά θεωρούνται ανεξάρτητες από άλλες μορφές χειρονομιών δεδομένου ότι διαθέτουν γλωσσολογικό υπόβαθρο και εκτελούνται χρησιμοποιώντας μια σειρά μεμονωμένων κινήσεων ή χειρονομιών που συνδυάζονται προκειμένου να διαμορφώσουν γραμματικές δομές. Σε μερικές περιπτώσεις όπως τον δακτυλοσυλλαβισμό (fingerspelling), οι νοηματικές γλώσσες μπορούν να θεωρηθούν σηματοφορικής φύσης. Εντούτοις, οι χειρονομίες στις νοηματικές γλώσσες βασίζονται στα γλωσσολογικά συστατικά τους και αν και επικοινωνιακής φύσης, διαφέρουν από την επικοινωνία με χειρονομιές δεδομένου ότι οι χειρονομίες σε τέτοιες περιπτώσεις αντιστοιχούν απόλυτα στα σύμβολα που επιχειρούν να αναπαραστήσουν. Οι νοηματικές γλώσσες είναι γραμματικά και λεξικολογικά πλήρεις και συγκρίνονται συχνά με την ομιλία από την άποψη της επεξεργασίας που απαιτείται για την αναγνώρισή τους, αν και η πολυεπίπεδη δομή τους τις καθιστά ως εξαιρετικά μεγάλη πρόκληση ως προς την αναγνώριση τους.

Η νοηματική γλώσσα βρίσκεται στην κορυφή της ταξονομίας χειρονομιών που προτείνεται στο [159] όπως φαίνεται στο σχήμα 4.3. Ως τέτοια, γίνεται ευρέως αποδεκτό ότι αποτελεί την απόλυτη πρόκληση από την άποψη της αναγνώρισης μεταξύ των κατηγοριών χειρονομιών ενώ έχουν εξαιρετικά μεγάλο συνομιλητικό ρυθμό έναντι άλλων επικοινωνιακών μορφών. Είναι ενδεικτικό ότι ο ρυθμός σε συνθήκες νοηματισμού εκ γενετής κωφών είναι της τάξης 175–225 λήμματα ανά λεπτό, ενώ ο ρυθμός γραφής συνήθως κυμαίνεται από 15–25 λέξεις [90, 17].

Σχήμα 4.3: Ιεραρχία χειρονομιών του Kendon [159]

Τα στιγμιότυπα των νοημάτων ποικίλλουν σε πολλές πτυχές. Ακόμα κι αν ο ίδιος νοηματιστής προσπαθήσει να εκτελέσει το ίδιο νόημα, μικρές αλλαγές στην ταχύτητας και την θέση των χεριών θα εμφανιστούν αναπόφευκτα μεταξύ των στιγμιοτύπων του ίδιου νοήματος. Κάθε νόημα ποικίλλει τόσο χρονικά όσο και χωρικά και η ταχύτητα και η διάρκεια νοηματισμού μπορεί να διαφέρουν σημαντικά. Οι κινήσεις του

νοηματιστή, όπως η μετατόπιση σε κάποια κατεύθυνση ή η περιστροφή γύρω από τον άξονα του σώματος, πρέπει να ληφθούν υπόψη. Τα δάχτυλα κατά τον νοηματισμό μπορούν να επικαλυφθούν, δεδομένου ότι βρίσκονται πίσω από άλλα δάκτυλα ή άλλα μέρη των χεριών ή και των μπράτσων. Σε αντίθεση με την αναγνώριση μεμονωμένων νοημάτων, όπου τα χρονικά σημεία έναρξης και τέλους των νοημάτων είναι εκ των προτέρων γνωστές, το σύστημα πρέπει επιπροσθέτως να ανιχνεύσει αυτά τα χρονικά σημεία καθώς και τις μεταβάσεις μεταξύ των νοημάτων κατά την αναγνώριση συνεχούς νοηματισμού. Για μια ακολουθία συνδεδεμένων νοημάτων όπου η εκτέλεση κάθε νοήματος επηρεάζεται από τα χρονικά γειτονικά (προηγούμενο και επόμενο) νοήματα το φαινόμενο της συνάρθρωσής (co-articulation). Η δομή μιας πρότασης στην προφορική γλώσσα είναι γραμμική, μια λέξη ακολουθείται από άλλη, ενώ στη νοηματική γλώσσα υπάρχει μια παράλληλα εξελισσόμενη δομή. Το νόημα μπορεί να αρχίσει και να τελειώσει σε οποιαδήποτε χρονική στιγμή μιας ακολουθίας, δεδομένου ότι δεν υπάρχει χρονικός περιορισμός στην εκτέλεση ενός νοήματος και ο αριθμός των νοημάτων σε μια φράση δεν είναι προκαθορισμένος. Η επεξεργασία μεγάλου όγκου δεδομένων είναι μια χρονοβόρα διαδικασία, γεγονός που καθιστά την αυτόματη αναγνώριση νοηματικής γλώσσας σε πραγματικό χρόνο δύσκολη τόσο από πλευράς εξαγωγής χαρακτηριστικών γνωρισμάτων όσο και από πλευράς κατηγοριοποίησης ειδικά για λεξιλόγια μεγάλης κλίμακας. Επιπλέον, στην περίπτωση της εμπρόσθιας μονοκάμερης καταγραφής, ο τριδιάστατος χώρος προβάλλεται σε διδιάστατο επίπεδο, με συνέπεια την απώλεια της πληροφορίας βάθους και η ανακατασκευή της τριδιάστατης τροχιάς του χεριού είναι μια επίπονη διαδικασία και αρκετές φορές αδύνατη. Τέλος, η θέση του νοηματιστή μπροστά από την συσκευή καταγραφής μπορεί να διαφέρει και ως εκ τούτου οι θέσεις των χεριών και γενικά ότι αναφέρεται σε θέση δεν πρέπει να επεξεργάζεται σε απόλυτη τιμή αλλά να είναι κανονικοποιημένη και σχετική με κάποιο σημείο αναφοράς (π.χ. το κεφάλι του νοηματιστή).

Όλες οι προαναφερθείσες προκλήσεις μπορούν να συνοψιστούν στα εξής:

- χωροχρονική παρέκκλιση εκτέλεσης νοήματος
- ανεξέλεγκτο περιβάλλον
  - θέση νοηματιστή
  - απόσταση νοηματιστή από την κάμερα
  - ύψος και ανατομικές ιδιαιτερότητες νοηματιστή
  - συνθήκες φωτισμού
  - μη στατικό παρασκήνιο
- διδιάστατη προβολή για μονοκάμερες αρχιτεκτονικές
- χρονικές αλλοιώσεις
  - συχνότητα
  - διάρκεια
  - επανάληψη
- σειριακές αλλοιώσεις
  - χρονικά τμήματα κίνησης

  - χρονικά τμήματα παύσης
- τροποποιήσεις τροχιάς
  - επένθεση κίνησης
  - μορφή
  - ομαλότητα
  - ρυθμός
  - ένταση
- γραμματικά φαινόμενα
  - εμφατικές τροποποιήσεις
  - παραγωγή ουσιαστικών από ρήματα
  - αριθμητική ενσωμάτωση
  - σύνθετα νοήματα
  - μη χειρωνακτικά γνωρίσματα

4.1.2.2 Πτυχές Αυτόματης Αναγνώρισης Νοηματικής Γλώσσας

4.1.2.2.1 Νοηματικές Γλώσσες Σχετικά με τις εθνικές νοηματικές γλώσσες στις οποίες έχουν γίνει απόπειρες αναγνώρισης υπάρχει επαρκής, αλλά όχι πλήρης, αντιπροσώπευση παγκοσμίως. Η Αμερικανική και η Κινέζικη νοηματική γλώσσα έχουν την μεγαλύτερη εκπροσώπηση, ενώ η Βρετανική έχει σημαντικό αριθμό άρθρων που πραγματεύονται την αναγνώριση της, όπως φαίνεται στον πίνακα 4.1. Η Γερμανική, Ιαπωνική και Ταιβανική νοηματική γλώσσα εξετάζονται επίσης στα [14, 16, 100, 108], [94, 115, 116, 201, 220] και [111, 144, 148, 219] αντίστοιχα. Αραβική [7, 214], Αυστραλιανή [230], Γαλλική [53], Ελληνική [185], Ιταλική [117], Ισπανική [47] και οι Ολλανδική [8] συναντώνται περιστασικά και σίγουρα δεν έχουν μελετηθεί σε βάθος. Υπάρχει μια ασαφής υπόθεση που συσχετίζει τον αριθμό άρθρων που εξετάζουν μια συγκεκριμένη νοηματική γλώσσα με το βάθος και το εύρος της γλωσσικής και γραμματικής έρευνας που συνδέεται με τη συγκεκριμένη γλώσσα, αν και αυτό δεν είναι ευδιάκριτο από τα αποτελέσματα των ερευνών. Αν και κάθε χώρα έχει τη δική της εθνική νοηματική γλώσσα υπάρχουν παρόμοιοι γλωσσικοί και γραμματικοί κανόνες που διέπουν όλες τις νοηματικές γλώσσες και κάποιος θα μπορούσε να υποστηρίξει πως ένα σύστημα που εφαρμόζεται και επαληθεύεται χρησιμοποιώντας πειραματικό σύνολο μια νοηματικής γλώσσας μπορεί να γενικεύσει καλά όταν εφαρμοστεί σε άλλες νοηματικές γλώσσες, εφόσον θέματα όπως η ενσωμάτωση γραμματικών κανόνων και η προσαρμοστικότητα σε νέες συνθήκες έχουν αντιμετωπιστεί επιτυχώς και έχει προβλεφθεί κάποια διαδικασία επαναπροσδιορισμού των ελεύθερων παραμέτρων.

Αν και η γλωσσική επιλογή κλίνει προς τη χώρα προέλευσης των ερευνητών, είναι θεμιτό να υποτεθεί πως η γλωσσική ομοιότητα μεταξύ των νοηματικών γλωσσών καθιστά την πειραματική γλωσσική επιλογή ένα δευτερεύον ζήτημα και δεν επηρεάζει σε μεγάλο βαθμό την δυνατότητα γενίκευσης κάθε προσέγγισης. Αφ' ετέρου η βιβλιογραφία σε δίγλωσσες ή πολύγλωσσες μελέτες είναι σχεδόν ανύπαρκτη. Τέτοιες έρευνες παρουσιάζουν εξαιρετικό ενδιαφέρον, τόσο από την πλευρά της διερεύνησης των δυνατοτήτων γενίκευσης των προσεγγίσεων αναγνώρισης αλλά και από την γλωσσολογική πτυχή της καθώς χρήσιμα συμπεράσματα μπορούν να εξαχθούν για την δομή και τις ιδιαιτερότητες κάθε γλώσσας.

Γλώσσα	Εργασία
Αμερικανική	[24, 52, 56, 107, 216, 235, 233, 236, 241, 255, 260]
Κινέζικη	[83, 84, 85, 96, 97, 95, 123, 153, 242, 240, 243, 268, 269, 267]
Βρετανική	[22, 15, 45, 44, 126, 169, 270]

Πίνακας 4.1: Εργασίες αυτόματης αναγνώρισης ανά νοηματική γλώσσα

4.1.2.2.2 Μέγεθος λεξιλογίου Ένα άλλο σημαντικό ζήτημα στον ευρύ ερευνητικό τομέα της αυτόματης αναγνώρισης νοηματικής γλώσσας είναι το μέγεθος λεξιλογίου των πειραματικών σωμάτων που χρησιμοποιούνται για να ελέγξουν την ευρωστία και τις ικανότητες γενίκευσης των προτεινόμενων συστημάτων, όπως μπορεί να φανεί στον πίνακα 4.2. Οι περισσότερες μελέτες χρησιμοποιούν πειραματικά σύνολα δεδομένων με αρκετά περιορισμένο αριθμό νοημάτων, της τάξης 10 έως 65, ενώ άλλες επεκτείνουν το λεξιλόγιο τους, σε περισσότερο αντιπροσωπευτικότερο δείγμα της αντίστοιχης γλώσσας, αλλά ακόμα και έτσι το μέγεθος κυμαίνεται μεταξύ 164 και 274 λημμάτων.

Οι εργασίες που προσεγγίζουν ένα ολοκληρωμένο σύστημα αναγνώρισης είναι εκείνες που αγγίζουν τα εντυπωσιακά μεγέθη λεξιλογίου, απαριθμώντας περισσότερα από 5000 νοήματα. Οι μελέτες που ανήκουν στην τελευταία ομάδα εστιάζουν κυρίως στη αναγνώριση λεξιλογίου μεγάλης κλίμακας και τα ζητήματα που σχετίζονται με ένα ολοκληρωμένο σύστημα, λειτουργώντας σε πραγματικό χρόνο και ικανό να υποστηρίξει αυτόματη γλωσσική αναγνώριση. Ο αριθμός επαναλήψεων που εκτελούνται για κάθε εγγραφή λεξιλογίου είναι επίσης σημαντική παράμετρος, όπως είναι και η αναλογία δειγμάτων εκπαίδευσης/επαλήθευσης. Τυπικά κάθε νόημα επαναλαμβάνεται 5-10 φορές π.χ.[22, 45, 44, 126, 153, 242, 240, 270], ενώ υπάρχουν περιπτώσεις με περισσότερες [144, 107, 116] ή λιγότερες [8, 83, 117, 219, 243] επαναλήψεις εκτελούνται για κάθε λήμμα στο περιορισμένο, πειραματικό λεξιλόγιο.

Μέγεθος	Εργασία
$10 < \& < 65$	[7, 14, 15, 16, 22, 24, 47, 52, 107, 108, 111, 117, 144, 185, 214, 216, 219, 220, 230, 235, 236, 241, 243, 249, 255, 260]
$164 < \& < 274$	[8, 45, 44, 83, 100, 126, 201, 242, 270]
> 5000	[56, 84, 85, 97, 96, 95, 94, 123, 153, 240, 268, 267]

Πίνακας 4.2: Μέγεθος λεξιλογίου

4.1.2.2.3 Εξάρτηση από νοηματιστή Η εξάρτηση του συστήματος αναγνώρισης από τον νοηματιστή είναι μια ακόμα κρίσιμη πτυχή της αναγνώρισης νοηματικής γλώσσας, υπό το πρίσμα της γενίκευσης της προτεινόμενης αρχιτεκτονικής σε ένα υλοποιήσιμο σύστημα. Πολλά άρθρα [8, 14, 22, 45, 100, 108, 126, 148, 242, 240, 243, 255, 268, 269] προτείνουν προσεγγίσεις που έχουν εκπαιδευτεί και επαληθευτεί στον ίδιο νοηματιστή. Αυτός ο περιορισμός της εξάρτησης από τον νοηματιστή, δεν μπορεί να ισχύει για ένα γενικό αυτόματο σύστημα αναγνώρισης νοηματικής γλώσσας δεδομένου ότι δεν είναι δυνατό να λαμβάνονται τα δεδομένα εκπαίδευσης από όλους τους υποψηφίους χρήστες του συστήματος. Η ανεξαρτησία από τον χρήστη/νοηματιστή

και οι τρόποι αντιμετώπισης της παρέκκλισης κατά την εκτέλεση νοημάτων και η αλλοίωση λόγω γραμματικών ιδιωματισμών είναι ζωτικής σημασίας ώστε να επιτυγχάνονται ικανοποιητικά ποσοστά αναγνώρισης σε μια αυθαίρετη ρύθμιση από έναν μη εγγεγραμμένο χρήστη. Αρκετές εργασίες [47, 107, 115, 214, 260] έχουν δοκιμάσει τα σχήματα αναγνώρισης σε πολλαπλούς νοηματιστές, αλλά αυτό είναι επαρκές μόνο στην περίπτωση που η επαλήθευση γίνεται σε χρήστες που δεν έχουν περιληφθεί στο σύνολο δεδομένων εκπαίδευσης, ώστε να επιτευχθεί πραγματική ανεξαρτησία από τον νοηματιστή. Ο πίνακας 4.3 απαριθμεί τις εργασίες που εξετάζουν μη εγγεγραμμένους νοηματιστές, ενώ ο πίνακας 4.4 παρουσιάζει την αντίστοιχη μείωση του ποσοστού αναγνώρισης.

Εργασία	Αριθμός νοηματιστών	Εγγεγραμμένοι/Μη Εγγεγραμμένοι
[44]	12	9/3
[83]	7	5/2
[84]	6	5/1
[95]	6	5/1
[230]	7	7/10
[267]	7	4/3
[270]	6	varying

Πίνακας 4.3: Εργασίες με μη εγγεγραμμένους νοηματιστές

Εργασία	% για Εγγεγραμμένους	% για Μη Εγγεγραμμένους
[83]	95.3	88.2
	96.6	90.1
[84]	90.5	82.9
	91.6	83.7
[95]	91.3	86.3
[230]	96.2	89.9
[270]	94.1-98.9	3.7-79.1

Πίνακας 4.4: Μείωση ποσοστού αναγνώρισης για μη εγγεγραμμένους χρήστες

Συμπερασματικά, τόσο το μέγεθος λεξιλογίου όσο και η ανεξαρτησία από τον νοηματιστή είναι κρίσιμα ζητήματα στην προοπτική της εύρωστης και εφαρμόσιμης αναγνώρισης νοηματικής γλώσσας. Αν και εργασίες που χρησιμοποιούν περιορισμένο λεξιλόγιο λημμάτων, μεγάλη αναλογία εκπαίδευσης/επαλήθευσης και ταυτόσημους νοηματιστές για πειράματα επαρκούν για να εννοιολογική απόδειξη, τέτοιες υποθέσεις και συμβιβασμοί δεν ισχύουν σε πραγματικές εφαρμογές. Οι ταξινομητές βασισμένοι σε γλωσσικά συστατικά νοηματικής γλώσσας (signemes σε αντιστοιχία με τα phonemes της προφορικής γλώσσας) δείχνουν να αντιμετωπίζουν ικανοποιητικά λεξιλόγια μεγάλης κλίμακας, δεδομένου ότι απαιτούν περιορισμένο αριθμό signemes για να χαρακτηρίσουν το σύνολο των νοημάτων και νέα νοήματα μπορούν να προτυποποιηθούν εύκολα με τον καθορισμό γλωσσικών συστατικών, χωρίς ρητή επανεκπαίδευση, η ανεξαρτησία από τον νοηματιστή παραμένει τροχοπέδη στη διαδικασία γενίκευσης. Οι προσεγγίσεις που απαιτούν εξαιρετικά μικρό σύνολο εκπαίδευσης δείχνουν να υποφέρουν από την ίδια αδυναμία απαλοιφής της εξάρτησης από τον χρήστη.

Μια ενδιαφέρουσα λύση, θα ήταν η δημιουργία ενός πρόσθετου επιπέδου προσωποποίησης και προσαρμογής του εκπαιδευμένου συστήματος, έτσι ώστε η υπάρχουσα γνώση να ενισχυθεί με την τρέχουσα γνώση χωρίς την ανάγκη χρονοβόρας επανεκπαίδευσης.

4.1.2.2.4 Γλωσσικά Σώματα Ένα γλωσσικό σώμα (corpus) είναι μια συλλογή γλωσσικών οντοτήτων που επιλέγονται και διατάσσονται σύμφωνα με ρητά γλωσσικά κριτήρια προκειμένου να χρησιμοποιηθούν ως αντιπροσωπευτικό δείγμα της υπό εξέταση γλώσσας. Επιπλέον, ο ορισμός ψηφιακού γλωσσικού σώματος προϋποθέτει την κωδικοποίηση και επισημείωση κατά τυποποιημένο και ομοιογενή τρόπο προκειμένου να είναι συμβατός με διαδικασίες ανάκτησης. Ο σχεδιασμός και η υλοποίηση ψηφιακών γλωσσικών σωμάτων νοηματικής, όπως συμβαίνει και για τα περισσότερα γλωσσικά σώματα, δεν αποτελεί τετριμμένη διαδικασία. Πτυχές σχεδιασμού και υλοποίησης πρέπει να ληφθούν υπόψη ώστε το τελικό αποτέλεσμα να αποδειχθεί χρήσιμο και επαναχρησιμοποιήσιμο για την ανάλυση και την εξαγωγή συμπερασμάτων σχετικά με την ίδια τη νοηματική γλώσσα. Η εφαρμοστικότητα και η δυνατότητα επαναχρησιμοποίησης του σώματος σε πολλαπλά σχήματα εξαγωγής χαρακτηριστικών γνωρισμάτων και κατηγοριοποίησης είναι αναγκαίο να διασφαλιστεί. Κατάλληλη μεταγραφή και σχολιασμός επίσης αποτελούν διαδικασίες απαραίτητες και προαπαιτούμενες για εποπτευόμενη μάθηση (supervised learning) και για πολυεπίπεδες προσέγγισεις αντίστοιχα.

Όπως αναφέρεται και στην παράγραφο 4.1.2.3, σχετικά με το ποσοστό αναγνώρισης διάφορων κατηγοριοποιητών, μόνο με τη εφαρμογή αρχιτεκτονικών αναγνώρισης σε ομοιόμορφα και πολλαπλά σύνολα δεδομένων και σώματα μπορεί να επαληθευτεί η ευρωστία και η δυνατότητα γενίκευσης κάθε συστήματος. Αρκετές προσπάθειες να δημιουργηθούν τέτοια σώματα αναφέρονται στη λογοτεχνία, μερικές από τις οποίες είναι:

- Phoenix: έχει καταγραφεί από το καθημερινό δελτίο ειδήσεων 'Tagesschau' του γερμανικού τηλεοπτικού καναλιού Phoenix. Σε αυτό το πρόγραμμα ένας διερμηνέας νοηματίζει τις ειδήσεις στη γερμανική νοηματική γλώσσα ταυτόχρονα με τον εκφωνητή στην κάτω δεξιά γωνία της οθόνης. Αυτή η βάση δεδομένων μεταγράφεται στη Γερμανική νοηματική γλώσσα και τη γερμανική γλώσσα. Οι καταγραφές δεν γίνονται σε ελεγχόμενο περιβάλλον, αλλά αντίθετα ο νοηματιστής παρουσιάζεται μπροστά από ένα έντονα ανομοιογενές, μη-σταθερό παρασκήνιο.

- ECHO: αποτελείται από τρία γλωσσικά σώματα: Βρετανικής [251], Σουηδικής [18] και Ολλανδικής [50] νοηματικής γλώσσας. Αυτά τα σώματα περιέχουν παιδικά παραμύθια και ποίηση νοηματισμένα από έναν νοηματιστή. Εντούτοις, διαθέτουν εξαιρετικά μεγάλο και ποικιλόμορφο λεξιλόγιο, γεγονός που καθιστά την αυτόματη εκμάθηση δύσκολη.

- Boston: η ερευνητική ομάδα Αμερικανικής νοηματικής γλώσσας στο πανεπιστήμιο της Βοστώνης δημιούργησε ένα σύνολο βίντεο, εν μέρει διαθέσιμο στον ιστοτόπο τους και περγράφεται στο [167]. Το γλωσσικό σώμα είναι επισημειωμένα και έχει καταγραφεί από τρείς οπτικές γωνίες. Στο [259] δημοσιεύονται αποτελέσματα αναγνώρισης νοηματικής γλώσσας για το σώμα αυτό, αν και στο συγκεκριμένο σώμα η έμφαση έχει δοθεί σε γλωσσολογικά θέματα.

- I6-Boston201: αποτελείται από 201 προτάσεις της Αμερικανικής Νοηματικής Γλώσσας και έχει καταγραφεί σε ελεγχόμενο περιβάλλον. Τα νοήματα έχουν καταγραφεί από τέσσερις κάμερες και είναι υποσύνολο του σώματος που έχει δημιουργήσει το πανεπιστήμιο της Βοστόνης [259].

- Signs of Ireland: γλωσσικό σώμα που δημιουργήθηκε στο κέντρο μελετών κωφών του Δουβλίνου [145] περιέχει δεδομένα βίντεο από περίπου 40 κωφούς χρήστες Ιρλανδικής νοηματικής γλώσσας συλλεχθέντα σε χρονικό διάστημα 3 ετών. Οι συμμετέχοντες αφηγούνται μια παιδική ιστορία και νοηματίζουν αποσπασματικά προτάσεις και το προκύπτον γλωσσικό σώμα επισημειώνεται χειροκίνητα.

- Ένα σώμα 2468 προτάσεων της γερμανικής νοηματικής γλώσσας από το πεδίο των δελτίων καιρού παρουσιάζεται στο [26] όπου επίσης παρουσιάζεται και ένα γλωσσικό σώμα με πληροφορίες αεροπορικού ταξιδιού (Air Travel Information System) που περιλαμβάνει 595 προτάσεις σε πέντε γλώσσες.

- RWTH-BOSTON-104: παρουσιάζεται στο [65] και περιέχει 201 προτάσεις Αμερικανικής νοηματικής γλώσσας. Αυτό το σώμα χρησιμοποιείται κυρίως για αυτόματη αναγνώριση νοηματικής γλώσσας ενώ το σώμα RWTH-BOSTON-400 αποτελείται από 843 προτάσεις, από διάφορους νοηματιστές.

- Purdue RVL-SLLL: συνίσταται από 2576 βίντεο που αντιστοιχούν σε 14 διαφορετικούς εκ γενετής κωφούς της Αμερικανικής νοηματικής γλώσσας (184 βίντεο ανά υπογράφοντα) και περιγράφεται στο [158].

- Greek Sign Language Corpus (GSLC): το Ελληνικό γλωσσικό σώμα νοηματικής όπως περιγράφεται στο [68]. Το πρώτο μέρος περιλαμβάνει έναν κατάλογο λημμάτων ενώ το δεύτερο μέρος αποτελείται από σύνολα ελεγχόμενων εκφράσεων, οι οποίες διαμορφώνουν παραδείγματα ικανά να αποκαλύψουν μηχανισμούς της Ελληνικής νοηματικής γλώσσας και συγκεκριμένα φαινόμενα γραμματικών πυρήνων. Τέλος το τρίτο μέρος περιέχει ελεύθερες ακολουθίες αφήγησης. Η διαδικασία επισημείωσης πραγματοποιήθηκε σε επίπεδο μορφήματος και φράσης/πρότασης χρησιμοποιώντας το σύστημα επισημείωσης ELAN και στην συνέχεια ακολούθησε ποιοτικός έλεγχος της επισημείωσης.

Οι περισσότερες γλωσσικές βάσεις δεδομένων που χρησιμοποιούνται στην επεξεργασία νοηματικων γλωσσών μέχρι πρόσφατα δεν παρέχουν ή δεν περιλαμβάνουν απαραίτητα στοιχεία για την αξιολόγηση αλγορίθμων επεξεργασίας νοηματικών γλωσσών [259], αν και σημαντικά βήματα προς την επίλυση του προβλήματος έχουν γίνει [68]. Η ανάγκη για δημιουργία βάσεων δεδομένων συγκριτικής μέτρησης επιδόσεων (benchmark) που να μπορεί να χρησιμοποιηθούν για την διερεύνηση των γλωσσικών προβλημάτων και της αξιολόγησης των αυτόματων συστημάτων γλωσσικής αναγνώρισης νοημάτων ή στατιστικών συστημάτων αυτόματης μετάφρασης συμπεριλαμβανομένων μεμονωμένων εκφράσεων, αφηγημάτων και πληροφοριών προφοράς διαλόγων είναι κρίσιμη. Άλλα χαρακτηριστικά ειδικά για περιπτώσεις ανάλυσης εκφράσεων προσώπου είναι ο νοηματισμός να καταγράφεται από πολλαπλές γωνίες, συμπεριλαμβανομένης και μιας εμπρόσθιας εστιασμένης στο πρόσωπο. Οι βάσεις δεδομένων μέχρι τώρα δεν έχουν παραχθεί με στόχο την χρήση τους στην αναγνώριση νοηματικών γλωσσών [65]. Για να χρησιμοποιήσουν τα δεδομένα για εκπαίδευση ή επαλήθευση

της απόδοσης συστημάτων αναγνώρισης νοηματικών γλωσσών, πρέπει να δημιουργηθούν οι απαραίτητες μεταγραφές, η οποία είναι μια δαπανηρή διαδικασία κυρίως από την άποψη ανθρώπινων πόρων.

4.1.2.2.5 Είσοδοι Σχετικά με την είσοδο κάθε προσέγγισης υπάρχουν τρείς κυρίαρχες κατηγορίες: αυτή της καταγραφής κίνησης (συσκευές άμεσης μέτρησης), τα γάντια δεδομένων (datagloves) και αυτή της οπτικής εισόδου όπως φαίνεται στον πίνακα 4.5. Ενώ οι συσκευές καταγραφής κίνησης και τα γάντια δεδομένων είναι αρκετά δαπανηρές και πραγματικά παρεισφρητικές παρέχουν έξοδο υψηλής ακρίβειας και ευρωστίας για την κάμψη των αρθρώσεων των δάχτυλων (και σε μερικές περιπτώσεις των χεριών) και την τριδιάστατη θέση των χεριών σε πραγματικό χρόνο σε σχέση με τις προσεγγίσεις που βασίζονται στην οπτική είσοδο. Η καταγραφή κίνησης με σημαδευτές χρησιμοποιείται επίσης στα [235, 236] ενώ μια συσκευή μέτρησης του 'χρόνου πτήσης' (time-of-flight camera) χρησιμοποιείται στο [94]. Τέλος, στο [24] συνδυάζονται οι δύο κυρίαρχες προσεγγίσεις όσον αφορά στην είσοδο της αρχιτεκτονικής αναγνώρισης νοηματικής γλώσσας.

Είσοδος	Εργασία
γάντια δεδομένων	[83, 84, 85, 97, 96, 95, 107, 123, 148, 201, 219, 230, 242, 240, 241, 267]
εικόνα	[7, 8, 14, 16, 15, 22, 45, 44, 47, 52, 53, 56, 94, 100, 108, 111, 115, 126, 144, 214, 216, 220, 243, 243, 249, 255, 260, 268, 270]
καταγραφή κίνησης	[235, 236]
κάμερα tof	[94]
εικόνα και επιταχυνσιόμετρο	[24]

Πίνακας 4.5: Τύποι εισόδων

Δημοφιλείς λύσεις στην καταγραφή πληροφορίας των χεριών από συσκευή είναι τα Virtual Realities CyberGloves για μετρήσεις γωνιών των αρθρώσεων [123, 201, 230, 236, 242, 241, 267] και οι καταγραφείς θέσης Polhemus [83, 240, 230, 267]. Άλλες συσκευές περιλαμβάνουν τα DataGlove από την VPL, το Power-Glove της Mattel, το AcceleGlove, τα EMI-Gloves και το Flock of Birds καταγραφέα κίνησης.

Από την άποψη της εφαρμοστικότητας και οι δύο προσεγγίσεις μπορούν να θεωρηθούν κατάλληλες, με την οπτική είσοδο να έχει ευρύτερο πεδίο εφαρμογής. Οι βασισμένες σε γάντι προσεγγίσεις μπορούν εύκολα να εγκατασταθούν σε περιβάλλοντα εσωτερικού χώρου ή σε περιπτώσεις όπου η ακρίβεια εισόδου είναι κρίσιμη. Τέτοιες περιπτώσεις θα ήταν σε βιομηχανικό ρομποτικό έλεγχο. Η οπτική είσοδος αφ' ετέρου θα μπορούσε να χρησιμοποιηθεί για υπαίθρια περιβάλλοντα όπου ο χρήστης δεν είναι εξοπλισμένος με εξειδικευμένες συσκευές και θα νοημάτιζε φυσικά όπως επικοινωνεί με ένα άλλο πρόσωπο με ακουστικά προβλήματα. Φανταστείτε παραδείγματος χάριν έναν νοηματιστή που εισέρχεται σε μια τράπεζα και τον υπάλληλο της τράπεζας να επικοινωνούν μέσω ενός συστήματος αυτόματης αναγνώρισης νοηματικής γλώσσας για μονόδρομη επικοινωνία ή ακόμα, με την χρήση ενός εικονικού νοηματιστή, για αμφίδρομη αλληλεπίδραση.

4.1.2.2.6 Περιορισμοί Οι βασισμένες σε βίντεο προσεγγίσεις έχουν το πλεονέκτημα της μη παρεισφρητικότητας αλλά συχνά αρκετές υποθέσεις πρέπει να γίνουν ή περιορισμοί να εφαρμοστούν κατά τη διάρκεια της καταγραφής που αφορούν είτε στο περιβάλλον, στον νοηματιστή είτε στη ρύθμιση των συσκευών καταγραφής. Συνήθως όταν υιοθετείται κάποιο πρότυπο χρωματικής χροιάς δέρματος ο χρήστης καλείται να φορέσει μακριά μανίκια και κατάλληλο ρουχισμό ώστε να μην είναι εκτεθειμένες πολλές περιοχές δέρματος που δεν είναι κεφάλι ή χέρια. Η ύπαρξη τέτοιων περιοχών μπορεί να οδηγήσει τους αλγορίθμους ανίχνευσης χεριών σε κατάρρευση και να εμποδίσει τον εντοπισμό των χεριών του νοηματιστή. Επιπλέον, το παρασκήνιο είναι εξαιρετικά σημαντικό σε κάποιες περιπτώσεις όταν πραγματοποιείται αφαίρεση στατικού υποβάθρου ή άλλος τελεστής επεξεργασίας εικόνας που υποθέτει ομοιόμορφο ή/και στατικό παρασκήνιο. Τέτοιες περιπτώσεις είναι οι [47, 220, 270]. Αρκετές εργασίες [243, 8, 100, 268, 7, 14] πραγματοποιούν τις καταγραφές με τον νοηματιστή να φορά βαμβακερά γάντια με διαφορετικό χρώμα σε περιοχές όπως δάχτυλα, ακροδάχτυλα και παλάμη διευκολύνοντας έτσι την κατάτμηση της εικόνας σύμφωνα με το διακριτό χρώμα κάθε περιοχής. Επιπλέον περιορισμοί που εφαρμόζονται είναι:

- στάση σώματος
- σχετική θέση χεριών αλλά και χεριού και κεφαλιού
- στατική θέση κεφαλής
- συνεχής κίνηση χεριών (για κατάτμηση με βάση το χρώμα αλλά υποβοηθούμενη από κίνηση)
- συγκεκριμένη αρχική θέση χεριών (ουδέτερη)
- αποφυγή επικαλύψεων

4.1.2.2.7 Εφαρμοσιμότητα Μια κρίσιμη πτυχή για την εφαρμογή σε πραγματικό χρόνο των προτεινόμενων αρχιτεκτονικών είναι ο χρόνος επεξεργασίας που απαιτείται τόσο για την εξαγωγή χαρακτηριστικών γνωρισμάτων όσο και για την διαδικασία αναγνώρισης. Πέρα από τη δυνατότητα εφαρμογής σε πραγματικό χρόνο η δυνατότητα εφαρμογής σε μεγάλης κλίμακας λεξιλόγια απαιτεί μικρούς χρόνους επεξεργασίας για την αξιολόγηση κάθε προτύπου/νοήματος. Εναλλακτικά του σχήματος πρότυπο ανά νόημα υιοθετείται η χρήση πιο απλών συστατικών της νοηματικής γλώσσας σε μία προσπάθεια μείωσης του αριθμού των κλάσεων στο πρόβλημα αναγνώρισης. Η έμφαση σε πολλά άρθρα [45, 96, 107, 126, 148, 235, 240] έχει δοθεί στην εφαρμογή σε μεγάλης κλίμακας λεξιλόγια ενώ άλλα [47, 84, 242, 268] στοχεύουν να μειώσουν τον υπολογιστικό φόρτο σε επίπεδα αποδεκτά για εφαρμογή πραγματικού χρόνου. Ο απαιτούμενος χρόνος εκπαίδευσης/επαλήθευσης δεν αναφέρεται σε όλες τις εργασίες και σε πολλές που γίνεται αναφορά σε αυτόν τον χρόνο δεν αποσαφηνίζεται αν περιλαμβάνεται και ο χρόνος εξαγωγής χαρακτηριστικών γνωρισμάτων ή μόνο ο χρόνος κατηγοριοποίησης. Ο πίνακας 4.6 παρουσιάζει ενδεικτικές τιμές χρόνων επεξεργασίας:

4.1.2.2.8 Αναγνώριση μεμονωμένων νοημάτων και συνεχούς νοηματισμού Ενώ η πλειοψηφία των άρθρων πραγματεύεται αναγνώριση μεμονωμένων λημμάτων, υπάρχει ακόμα σημαντικός αριθμός εργασιών που εστιάζει στην αναγνώριση συνεχούς

Εργασία	Χρόνος Επεξεργασίας
[84]	0.268
[94]	1–4
[111]	10
[123]	0.137
[242]	0,04
[240]	0,22
[268]	0,135
[267]	1,372
[270]	11.79/4.15/3.08/2.92

Πίνακας 4.6: Αναφερόμενοι χρόνοι επεξεργασίας (δευτερόλεπτα)

νοηματισμού ή αναγνώριση προτάσεων νοηματικής γλώσσας, όπως φαίνεται στον πίνακα 4.7. Επιπλέον, μερικοί ερευνητές μελετούν τόσο την μεμονωμένη όσο και την συνεχή αναγνώριση, που επεκτείνει μεθοδολογίες που λειτουργούν σε επίπεδο νοήματος σε αναγνώριση σε επίπεδο προτάσεων. Το τελευταίο είναι σημαντικά πιό σύνθετο και εμπεριέχει περισσότερες και δυσκολότερες προκλήσεις, δεδομένου ότι τα χρονικά όρια των νοημάτων πρέπει να ανιχνευθούν αυτόματα και κάθε νόημα που περιέχεται στην πρόταση επηρεάζεται από το προηγούμενο και το επόμενο νόημα, το φαινόμενο της συνάρθρωσης. Τέλος, υπάρχει και εργασίες που ασχολούνται με αναγνώριση δακτυλικού συλλαβισμού [185, 144] με αναγνώριση μεμονωμένων γραμμάτων νοηματικής.

Η απλοϊκή προσέγγιση στο πρόβλημα της αναγνώρισης συνεχούς νοηματισμού είναι η επέκταση της μεμονωμένης αναγνώρισης νοημάτων σε συνδυασμό με την αυτόματη αναγνώριση των χρονικών ορίων κάθε νοήματος επιτρέποντας έτσι στο σύστημα να αντιμετωπίσει την νοηματική πρόταση ως ακολουθία μεμονωμένων νοημάτων. Η ανίχνευση των χρονικών ορίων πραγματοποιείται με διάφορους τρόπους:

- τοπικό ελάχιστο σε:
  - ταχύτητα του χεριού σε προσεγγίσει με οπτική είσοδο ή τριδιάστατη παρακολούθηση με την βοήθεια συσκευών
  - κάμψη των αρθρώσεων των δακτύλων για προσεγγίσεις με γάντι
- μέγιστα στην παράγωγο γωνιών τροχιάς κίνησης
- αναλογία μεταξύ ελάχιστης επιτάχυνσης και μέγιστης ταχύτητας
- HMM που εκπαιδεύεται για κατάτμηση νοημάτων
- HMM που εκπαιδεύεται για την προτυποποίηση μεταβάσεων
- μοντελοποίηση και εντοπισμός επένθεσης
- μείωση πιθανότητας ταύτισης

Τύπος αναγνώρισης	Εργασία
Συνεχής	[14, 15, 16, 53, 117, 201, 216, 230, 235, 233, 236, 240, 241]
Μεμονωμένη & Συνεχής	[8, 97, 96, 95, 108, 148]

Πίνακας 4.7: Μεμονωμένη & Συνεχής αναγνώριση

4.1.2.2.9 Γλωσσολογικά Θέματα Εκτός από τα βασικά γλωσσικά συστατικά της θέσης, μετακίνησης, χειρομορφής και προσανατολισμού παλάμης η νοηματική γλώσσα είναι εμπλουτισμένη με μη χειρωνακτικά χαρακτηριστικά και πλήρη γραμματική δομή. Και οι δύο πτυχές έχουν διερευνηθεί ελάχιστα όσον αφορά στην εμπλοκή τους στην διαδικασία αυτόματης αναγνώρισης. Για μερικές νοηματικές γλώσσες το ίδιο ακριβώς συμβαίνει και για την καθεαυτό μελέτη των φαινομένων αυτών και της γραμματικής δομής αφού η γραμματική ανάλυση είναι ελλιπής και οι εκφράσεις του προσώπου και τα υπόλοιπα μη χειρωνακτικά χαρακτηριστικά που χρησιμοποιούνται από κοινού και παράλληλα με τα χειρωνακτικά χαρακτηριστικά δεν έχουν καταγραφεί και μελετηθεί πλήρως. Τα γραμματικά φαινόμενα συνήθως εκλαμβάνονται ως θόρυβος ή παρέκκλιση του νοηματιστή αφού η αρχιτεκτονική δεν έχει εκπαιδευτεί να τα αναγνωρίζει και να τα επεξεργάζεται ανεξάρτητα. Φυσικά τα γραμματικά φαινόμενα πρέπει να ενσωματωθούν στην αλυσίδα αναγνώρισης μέσω κάποιας ενότητας επεξεργασίας φυσικής γλώσσας και της γνώσης που εξάγεται από αυτή.

Ως τώρα οι περισσότερες εργασίες στην αυτόματη αναγνώριση νοηματικής γλώσσας συνήθως αγνοούν τις εκφράσεις του προσώπου που προκύπτουν ως τμήμα φυσικού νοηματισμού, ακόμα κι αν μεταφέρουν σημαντικές γραμματικές και 'προσωδικές' πληροφορίες. Η σαφής αντιστοιχία μεταξύ των γωνιών περιστροφής του κεφαλιού και του χρονικού προσδιορισμού στον νοηματισμό είναι μια πολλά υποσχόμενη κατεύθυνση για μελλοντικά συστήματα αναγνώρισης μη χειρωνακτικών χαρακτηριστικών [29]. Στο [234] παρουσιάζεται μια ακόμα ελπιδοφόρα προσπάθεια ενσωμάτωσης τέτοιων χαρακτηριστικών και των αντίστοιχων γραμματικών φαινομένων μέσω της παρακολούθησης παραμέτρων του προσώπου.

4.1.2.2.10 Χαρακτηριστικά γνωρίσματα Ανεξάρτητα από τον τύπο εισόδου κάθε προτεινόμενης προσέγγισης τα αριθμητικά δεδομένα υπόκεινται περαιτέρω επεξεργασία μέσω της εξαγωγής χαρακτηριστικών γνωρισμάτων ικανών να περιγράψουν το ιδιαίτερο χαρακτηριστικό που επιχειρούν να μοντελοποιήσουν. Σχεδόν όλες οι τεχνικές εξαγωγής χαρακτηριστικών γνωρισμάτων περιλαμβάνουν την θέση του κυρίαρχου χεριού (δεξί). Συνήθως, όταν υιοθετείται η καταγραφή κίνησης περιλαμβάνεται η τριδιάστατη θέση στα χαρακτηριστικά γνωρίσματα ενώ για τις προσεγγίσεις βασισμένες στην οπτική είσοδο μόνο η διδιάστατη προβολή της θέσης χεριών μπορεί να εξαχθεί χωρίς την βοήθεια στερέοψης. Η θέση των χεριών σχεδόν πάντα υπολογίζεται σχετικά με κάποιο σημείο αναφοράς π.χ. το κεφάλι του χρήστη ή το κάτω μέρος της πλάτης του στην περίπτωση καταγραφής δεδομένων με την βοήθεια συσκευών, τοποθετώντας έναν πρόσθετο αισθητήρα στην πλάτη του νοηματιστή. Οι προσεγγίσεις που βασίζονται σε είσοδο βίντεο υπολογίζουν συνήθως το κέντρο βάρους (Center Of Gravity) της περιοχής και θεωρούν αυτό το σημείο ως θέση του χεριού. Ο χαρακτηρισμός της χειρομορφής επιτυγχάνεται στην πλειοψηφία των περιπτώσεων με δύο τρόπους. Για βασισμένη σε γάντι είσοδο η κάμψη των αρθρώσεων των δαχτύλων θεωρείται επαρκής ώστε να περιγράψει την διενεργηθείσα χειρομορφή, ενώ στις προσεγγίσεις με οπτική

είσοδο, ειδικά αυτές που υποβοηθούνται με τη χρήση χρηματιστών γαντιών, το μήκος και η γωνία των διανυσμάτων που εκκινούν από το κέντρο βάρους του χεριού ή τον καρπό του χεριού και καταλήγουν στα ακροδάχτυλα μπορούν να θεωρηθούν ως σύνολο χαρακτηριστικών γνωρισμάτων ικανά να περιγράψουν την πιθανή ρύθμιση των δαχτύλων. Στο [185] το τελικό σημείο των διανυσμάτων δεν είναι τα ακροδάχτυλα αλλά το όριο του χεριού με αποτέλεσμα να λαμβάνονται περισσότερα των πέντε διανύσματα. Ενώ, είναι μάλλον λίγες οι περιπτώσεις που τα χαρακτηριστικά γνωρίσματα της χειρομορφής εξάγονται βάσει του περιγράμματος του χεριού που συνήθως προκύπτει με την μέθοδο των ενεργών περιοχών [62].

Μια ακόμα σημαντική πτυχή της χειρομορφής είναι ο προσανατολισμός της παλάμης. Αυτό καταγράφεται εύκολα με γυροσκοπικές μετρήσεις και χρησιμοποιείται περισσότερο στις βασισμένες σε γάντι προσεγγίσεις, ενώ δεν περιλαμβάνεται στο σύνολο χαρακτηριστικών σχεδόν σε όλα τα συστήματα βασισμένα στην επεξεργασία εικόνας και την όραση υπολογιστών καθώς είναι εξαιρετικά δύσκολο να προκύψει από την εικόνα χωρίς την χρήση κάποιου προτύπου του χεριού και την μέθοδο ταύτισης προτύπων. Η αποσύνθεση σε διανύσματα Eigen χρησιμοποιείται συχνά [235, 216, 268] ενώ και άλλες τεχνικές επεξεργασίας δεδομένων υιοθετούνται σε άλλες περιπτώσεις και περιλαμβάνουν κανονικοποίηση, συσταδοποίηση, ανάλυση πρωτευουσών συνιστωσών, μετασχηματισμό Fourier, κ.λπ. Επιπλέον, στις περισσότερες προσεγγίσεις που δέχονται ως είσοδο εικόνα, όπου εντοπίζονται περιοχές χεριών, υπολογίζονται επίσης χαρακτηριστικά περιοχής (area features). Αυτά τα χαρακτηριστικά γνωρίσματα περιλαμβάνουν το μέγεθος, αναλογία μηκών και γωνία αξόνων, εκκεντρικότητα, στερεότητα (solidity), απόκλιση, παραμόρφωση και καμπυλότητα. Αρκετά άρθρα [243, 44] υιοθετούν μια τελείως διαφορετική προσέγγιση στην εξαγωγή χαρακτηριστικών γνωρισμάτων υιοθετώντας ομογραφία και ογκομετρικά χαρακτηριστικά γνωρίσματα αντίστοιχα, ενώ στο [47] χρησιμοποιείται ένα αρκετά απλό χαρακτηριστικό όπως είναι τα κινούμενα τμήματα (moving blocks) της εικόνας. Τέλος, είναι αρκετά κοινό [22, 45, 126] να χρησιμοποιήται η επισημείωση HA/TAB/SIG/DEZ που παρέχει έναν περιγραφέα υψηλού επιπέδου που διευκρινίζει τα χαρακτηριστικά υπό ένα πιο ευρύ πρίσμα, όπως περιγράφεται στο [25].

4.1.2.2.11 Ποσοστά Αναγνώρισης Το ποσοστό αναγνώρισης είτε μεμονωμένης είτε συνεχούς νοηματισμού, αν και θα μπορούσε να θεωρηθεί το σημαντικότερο κριτήριο σύγκρισης για τις προτεινόμενες προσεγγίσεις, στην πραγματικότητα, εξαιτίας των διαφορετικών συνθηκών και ανεξέλεγκτων μεταβλητών, είναι απλά ενδεικτικό και αρκετά υποκειμενικό κριτήριο. Η κυρίαρχη διαφορά είναι ότι χρησιμοποιούνται διαφορετικά σύνολα δεδομένων στα οποία δοκιμάζονται οι αλγόριθμοι και μόνο οι ίδιοι ερευνητές ή ερευνητές από το ίδιο εκπαιδευτικό ίδρυμα ή ερευνητική ομάδα συγκρίνουν τις προτεινόμενες αρχιτεκτονικές στο ίδιο γλωσσικό σώμα και στην πλειοψηφία των περιπτώσεων τα γλωσσικά σώματα δεν είναι ελεύθερα διαθέσιμα ή διατίθενται κάτω από αρκετούς περιορισμούς πνευματικών δικαιωμάτων. Επιπλέον, η ποιότητα του πειραματικού συνόλου δεδομένων είναι μια παράμετρος όχι άμεσα συγκρίσιμη και δεν διευκρινίζεται πάντα εάν το γλωσσικό σώμα χρησιμοποιήθηκε στο σύνολο του ή υπέστη κάποια διαδικασία επιλογής πριν την επεξεργασία.

Εργασία	Ποσοστό αναγνώρισης	Εργασία	Ποσοστό αναγνώρισης
[243]	71.8–92.1	[260]	93.0
[22]	73.0–84.0 (97.6 restricted)	[267]	93.1
[45]	74.3	[51]	93.2
[201]	80.2	[236]	93.2 RH 94.5 RH&LH
[148]	80.4 C 94.8 I	[255]	93.4
[185]	80.5–97.2	[153]	93.4–95.2
[16]	80.8	[52]	93.2
[97]	80.0–93.2 I 82.0–96.3 C	[7]	93.4
[117]	83.3	[100]	94.0
[84]	83.7(unr) 91.6 (reg)	[230]	94.0 (reg) 85.0(unr)
[56]	86.0(auto) 97.1(manual)	[14]	94.0–2.2
[95]	86.3	[235]	94.2 I 84.8 C
[123]	87.3	[8]	94.6 I (reg) 47.6 I (unr) / 72.8 C
[126]	89.1–92.0	[241]	95.0
[83]	90.1 (unr) 96.6 (reg)	[108]	95.0
[85]	90.5	[169]	97.4
[111]	91.0	[107]	98.0 (95.0 no retrain)
[219]	91.2–94.1	[242]	97.4–98.2
[220]	91.4–98.3	[270]	99.3 (reg) 44.1 (unr)
[144]	91.9–96.58	[240]	99.7 I 92.8 C
[216]	92 desktop 98 hat-mounted	[47]	99.5
[268]	92.5	[24]	90.4

Πίνακας 4.8: Ποσοστά αναγνώρισης

4.1.2.3 Σχήματα Κατηγοριοποίησης

Στο κέντρο κάθε προσπάθειας αυτόματης αναγνώρισης βρίσκεται η επιλογή ταξινομητή η οποία θεωρείται αποφασιστικότερη μεταξύ των άλλων συστατικών της συνολικής αρχιτεκτονικής και συνήθως καθορίζει ή επηρεάζει σε σημαντικό βαθμό τις σχεδιαστικές αποφάσεις των υπόλοιπων συστατικών. Αν και πληθώρα των άρθρων ακολουθεί κάποια καθιερωμένη και δοκιμασμένη σε άλλα προβλήματα λύση από την πλευρά της κατηγοριοποίησης (π.χ. Hidden Markov Models) υπάρχει ένας σημαντικός αριθμός προσεγγίσεων που υιοθετούν έναν συνδυασμό αρχιτεκτονικών κατηγοριοποίησης είτε σε σειριακή διάταξη είτε τροποποιώντας ή/και εμπλουτίζοντας την εσωτερική λειτουργία είτε της διαδικασίας εκπαίδευσης ή/και της διαδικασίας αξιολό-

γησης μιας γνωστής αρχιτεκτονικής μηχανικής μάθησης ή τεχνητής νοημοσύνης.

Μια ευρεία ομαδοποίηση των αρχιτεκτονικών ταξινόμησης θα μπορούσε να ήταν: Νευρωνικά δίκτυα, HMM και παραλλαγές, Γραμμική ανάλυση, Δεντρικές δομές, Συσταδοποίηση και Σύγκριση ακολουθίας καταστάσεων.

Τεχνικές κατηγοριοποίησης		
Νευρωνικά Δίκτυα	[111]	3D Hopfield
	[117]	Hierarchical SOM
	[144]	Backpropagation
	[219]	Composite Hyperrectangular
	[230]	4 Backpropagation
	[255]	Time Delay
HMM		[8, 14, 15, 16, 24, 47, 97, 100, 108, 148, 185, 216, 220, 260, 270]
Παραλλαγές HMM	[83]	SOFM/HMM
	[84]	GMM, FSM, SOFM/HMM
	[95]	SOFM/SRN/HMM
	[96]	HMM/DTW
	[235, 236]	Parallel
	[240]	Clustered Gaussian
	[241]	Multidimensional
	[249]	Parametric
	[268, 269]	Tied-Mixture Density (TMDHMM)
Ενίσχυσης (Boosting)	[45]	AdaBoost
	[44]	AdaBoost/AdaPlusBoost
	[22, 126, 169]	Boosted Cascades
Διάφορες	[7]	Polynomial
	[22, 126]	Markov/ICA
	[51, 52]	Recursive Partition Tree
	[53]	High level
	[56]	Euclidean Distance
	[85]	k-means/DTW
	[94]	Karhunen-Loeve Decomposition (PCA)
	[107]	Linear
	[115]	3D LUT
	[123]	ISODATA/DTW
	[153]	Parallel multistream model
	[201, 214]	KNN/Bayesian
	[243]	Nearest Neighbor
	[242]	Semi-Continuous Dynamic Gaussian Mixture Model
	[267]	Maximum Variance Criterion

4.1.2.3.1 Νευρωνικά δίκτυα Οι Huang και Huang στο [111] παρουσιάζουν μια πρωτότυπη εργασία για παρακολούθηση χεριών, εξαγωγή χαρακτηριστικών γνωρισμάτων και αναγνώριση χειρονομιών χρησιμοποιώντας ένα τριδιάστατο νευρωνικό δίκτυο Hopfield (HNN). Το τριδιάστατο HNN χρησιμοποιείται μόνο στη φάση της αναγνώρισης, πραγματοποιώντας ταίριασμα γράφων ανάμεσα στο στιγμιότυπο του νοήματος εισόδου και των εκπαιδευμένων προτύπων, ενώ η μετρική απόστασης Hausdorff χρησιμοποιείται για την παρακολούθηση της κίνησης των χεριών. Μετασχηματισμός Fourier εφαρμόζεται στα χαρακτηριστικά γνωρίσματα κίνησης με στόχο τα γνωρίσματα να παραμένουν αναλλοίωτα σε κλιμάκωση και περιστροφή. Πλαίσια κλειδιά εξάγονται για εκπαίδευση και επαλήθευση και οι πληροφορίες κίνησης και σχήματος αυτών των πλαισίων εξετάζονται έναντι των εκπαιδευμένων προτύπων και το πρότυπο που βρίσκεται πιο κοντά στο στιγμιότυπο θεωρείται πως είναι η κλάση του νοήματος. Για 15 διαφορετικές χειρονομίες επιτεύχθηκε ποσοστό αναγνώρισης πάνω από 91%, ενώ ο χρόνος επεξεργασίας είναι περίπου 10 δευτερόλεπτα ανά νόημα.

Οι Infantino κ.α. στο [117] ενσωματώνουν μια μηχανή κοινής λογικής (commonsense engine), προσομοιώνοντας τη διαδικασία της ανθρώπινης σκέψης, σε αναγνώριση προτάσεων νοηματικής. Οι συντεταγμένες χεριών, εξαγόμενες με διαδικασίες επεξεργασίας εικόνας, διαμορφώνουν διανύσματα χαρακτηριστικών γνωρισμάτων για κάθε πλαίσιο που περιλαμβάνονται στο νόημα, τα οποία εκπαιδεύουν ένα πολυεπίπεδο, ιεραρχικό αυτοοργανούμενο χάρτη (SOM) για την μεμονωμένη κατηγοριοποίηση λημμάτων που ακολουθείται από χρονική κατάτμηση βασισμένη στην κίνηση των χειλιών του νοηματιστή που υποβοηθά την επέκταση σε συνεχή νοηματισμό. Τα αναγνωρισμένα σημάδια συνδυάζονται λαμβάνοντας υπόψη το σημασιολογικό πλαίσιο και την ορθότητα της πρότασης που προκύπτει. Η μηχανή κοινής λογικής, υλοποιημένη στην πλατφόρμα OpenCyc, επιλέγει τη σωστή πρόταση ανάλογα με το εννοιολογικό πλαίσιο της πρότασης. Σε δύο πειράματα, σε δύο σύνολα δεδομένων 30 βίντεο 20 νοημάτων και 80 βίντεο 40 νοημάτων, τα αντίστοιχα ποσοστά επιτυχημένης χρονικής κατάτμησης ήταν 96.7% και 95.5% αντίστοιχα και τα ποσοστά κατηγοριοποίησης ήταν 83.3% και 82.5% αντίστοιχα.

Οι Lee και Tsai στο [144] χρησιμοποιούν το σύστημα ανάλυσης κίνησης VICON για καταγράψουν την τριδιάστατη θέση, η οποία υποβάλλεται σε επεξεργασία παράγοντας χαρακτηριστικά γνωρίσματα αμετάβλητα σε επικαλύψεις, περιστροφή, κλιμάκωση και μεταφορά, τα οποία στη συνέχεια εκπαιδεύουν ένα τυποποιημένο οπίσθιας διάδοσης (back-propagation) νευρωνικό δίκτυο. Το πειραματικό σώμα αποτελείται από καταγραφές δέκα φοιτητών, που εκτελούν κάθε μια από τις είκοσι στατικές χειρονομίες δέκα φορές και χωρίζεται σε σύνολο εκπαίδευσης με 1350 στιγμιότυπα και σύνολο δοκιμών με τα υπόλοιπα 1438 στιγμιότυπα. Ο διαφορετικός αριθμός νευρώνων στα δύο κρυμμένα επίπεδα είχαν ως αποτέλεσμα διαφορετικά αποτελέσματα αναγνώρισης, που ποικίλλουν από 91.90% σε 96.58% για 25 και 250 νευρώνες ανά κρυμμένο επίπεδο αντίστοιχα.

Ο Su στο [219] παρουσιάζει μια προσέγγιση βασισμένη σε ασαφείς κανόνες για την χωροχρονική αναγνώριση χειρονομιών χεριών. Οι απόλυτοι αν-τότε (if-then) κανόνες εξάγονται από τις τιμές των βαρών των συνάψεων των αντίστοιχων εκπαιδευμένων υπερορθογώνιων σύνθετων νευρωνικών δικτύων (HRCNNs) και έπειτα ασαφοποιούνται. Γάντια καταγραφής δεδομένων EMI-Gloves χρησιμοποιήθηκαν για να καταγράψουν ένα διάνυσμα χαρακτηριστικών γνωρισμάτων με διάσταση 20 που αντιπροσωπεύει τις γωνίες των αρθρώσεων των δάχτυλων. Υποθέτοντας ότι κάθε νόημα χρησιμοποιεί κατά μέγιστο δύο χειρομορφές τα πρώτα και τελευταία δέκα διανύσματα

χρησιμοποιούνται για να αντιπροσωπεύσουν την πρώτη και την δεύτερη βασική χειρομορφή και η η κατηγοριοποίηση πραγματοποιείται με την εξέταση κάθε ασαφούς κανόνα. Δύο γλωσσικά σώματα από 90 νοήματα της νοηματικής γλώσσας της Ταϊβάν από τέσσερα άτομα (2 για κάθε σώμα), τέσσερις επαναλήψεις για κάθε άτομο και κάθε μια από τις 34 βασικές χειρονομίες. Ένα ποσοστό αναγνώρισης 94.1% επιτεύχθηκε για το πρώτο γλωσσικό σώμα, το οποίο χρησιμοποιήθηκε επίσης για την εξαγωγή των κανόνων και για να επαληθευτούν οι δυνατότητες γενίκευσης του συστήματος εξετάστηκαν ενάντια στη δεύτερη βάση δεδομένων, η οποία δεν χρησιμοποιήθηκε στην εκπαίδευση, επιτυγχάνοντας ποσοστό αναγνώρισης 91.2%.

Οι Vamplew και Adams στο [230] παρουσιάζουν το SLARTI, μια αρθρωτή αρχιτεκτονική αποτελούμενη από πολλαπλά νευρωνικά δίκτυα και έναν ταξινομητή πλησιέστερου γείτονα. Η είσοδος λαμβάνεται από γάντια καταγραφής δεδομένων, ενός CyberGlove και ενός Polhemus IsoTrak με συνολικά 18 αισθητήρες οι τιμές των οποίων εκπαιδεύουν τέσσερα πλήρως συνδεδεμένα, εμπρόσθιας διάδοσης, με ένα κρυφό επίπεδο νευρωνικά δίκτυα για χειρομορφή, προσανατολισμό παλάμης, θέση και κίνηση. Ο αριθμός των κόμβων του επιπέδου εισόδου, του κρυμμένου επιπέδου και της εξόδου ποικίλλει για κάθε σύνολο χαρακτηριστικών και ειδικά το δίκτυο με είσοδο την θέση ένα αναδρομικό δίκτυο υιοθετείται ως βήμα προεπεξεργασίας. Η συγχώνευση των αποτελεσμάτων των τεσσάρων υποενοτήτων πραγματοποιείται είτε από έναν αλγόριθμο αναζήτησης πλησιέστερου γείτονα είτε από το C4.5 δέντρο απόφασης. Επτά εγγεγραμμένοι και τρεις μη εγγεγραμμένοι χρήστες πραγματοποιούν 52 νοήματα που διαιρέθηκαν τυχαία σε 13 ακολουθίες τεσσάρων νοημάτων και η αρχιτεκτονική SLARTI εξετάστηκε στα τόσο σε εκτελέσεις από τους εγγεγραμμένους νοηματιστές όσο και σε παραδείγματα από μη εγγεγραμμένους νοηματιστές, επιτυγχάνοντες έναν μέσο όρο 94% και 85% αντίστοιχα.

Οι Yang κ.α. εισάγουν έναν αλγόριθμο εξαγωγής χαρακτηριστικών γνωρισμάτων βασισμένο στην κατάτμηση πολλαπλών κλιμάκων (multiscale segmentation) και χρησιμοποιούν τις προκύπτουσες τροχιές για να κατηγοριοποιήσουν τις χειρονομίες ένα νευρωνικό δίκτυο χρονικής καθυστέρησης (time delay neural network - TDNN). Εστιάζουν κυρίως στην κατάτμηση, κατά τη διάρκεια της οποίας οι ομοιογενείς περιοχές μεταξύ διαδοχικών πλαισίων αντιστοιχούνται, υπολογίζονται αφινικοί μετασχηματισμοί και τελικά αντιστοιχίες εικονοστοιχείων συνδυάζονται με δερματική πληροφορία να υπολογιστεί η τροχιά του χεριού στις ακολουθίες πλαισίων. Σε 40 νοήματα της Αμερικανικής νοηματικής γλώσσας, διάρκειας τριών ως πέντε δευτερολέπτων και με την χρήση της διασταυρωμένης επικύρωσης με πέντε επαναλήψεις (five-fold cross validation) προέκυψαν 98.14% σε εκπαιδευμένες τροχιές και 93.42% σε τροχιές δοκιμής για μοναδική τροχιά ενώ χρησιμοποιώντας πλλαπλές τροχιές και συγχώνευση με ψηφοφορία προέκυψαν 99.02% και 96.21% αντίστοιχα.

4.1.2.3.2 HMM Τα κρυφά Μαρκοβιανά μοντέλα (Hidden Markov Model - HMM) αποτελούν μια διπλά στοχαστική διαδικασία αποτελούμενη από πεπερασμένο αριθμό καταστάσεων και ένα σύνολο τυχαίων συναρτήσεων συσχετισμένων με τις καταστάσεις. Σε διακριτές χρονικές στιγμές, η διαδικασία βρίσκεται σε μια από τις καταστάσεις και παράγει ένα σύμβολο παρατήρησης σύμφωνα με την σχετική τυχαία συνάρτηση που αντιστοιχεί στην τρέχουσα κατάσταση. Έχει αποδειχθεί πως τα HMM μοντελοποιούν αποτελεσματικά προβλήματα με χωροχρονική πληροφορία. Η μοντελοποίηση αποκαλείται 'κρυμμένη' επειδή το μόνο που είναι εμφανές στον εξωτερικό παρατηρητή είναι μια ακολουθία παρατηρήσεων. Συνήθως, ένα σύνολο χαρακτηρι-

στικών γνωρισμάτων εισόδου εκπαιδεύουν ένα HMM και ένα διαφορετικό σύνολο δεδομένων επαληθεύει την ορθότητα της εκπαίδευσης. Η ιδιότητα των HMM να αντισταθμίσει παρεκκλίσεις είτε σε χρονικό επίπεδο είτε σε επίπεδο τιμών των γνωρισμάτων τα καθιστά εξαιρετικά αποτελεσματικά για προβλήματα όπως αναγνώριση ομιλίας και χαρακτήρων. Αυτά ακριβώς τα χαρακτηριστικά των HMM τα καθιστούν μια κατάλληλη αλλά και ιδιαίτερα δημοφιλή λύση, όπως θα φανεί αργότερα, για το πρόβλημα της αυτόματης αναγνώρισης νοηματικής γλώσσας. Το κυριότερο μειονέκτημα τους είναι η ανάγκη μεγάλου όγκου δεδομένων απαραίτητων για την εκπαίδευση των παραμέτρων του μοντέλου, γεγονός που δυσχεραίνει την γενίκευση των μοντέλων για την επίτευξη ανεξαρτησίας νοηματιστή.

Οι Assan και Grobel στα [100, 8] πραγματοποιούν πειράματα σε 262 διαφορετικά νοήματα από την Ολλανδική νοηματική γλώσσα και πέρα από την αναγνώριση μεμονωμένων νοημάτων επεκτείνουν την αρχιτεκτονική για την αναγνώριση νοηματικής σε επίπεδο πρότασης με ένα σχήμα εισαγωγής/διαγραφής. Μελετούν την επιρροή των αποκλίσεων του διανύσματος εισόδου στο ποσοστό αναγνώρισης και μια προσέγγιση για αναγνώριση συνδεδεμένων νοημάτων. Μονοκάμερη καταγραφή βίντεο, χρωματιστά γάντια και τεχνικές επεξεργασίας εικόνας χρησιμοποιούνται για την εξαγωγή των κέντρων βάρους (COGs) και των γωνιών ως απόλυτη τιμή του μέσου όρου των τεσσάρων γωνιών των δαχτύλων για το επίπεδο της κάμερας. Συνεχή HMMs εφαρμόζονται, με ένα μίγμα δύο Γκαουσιανών και τον αρχικό αριθμό καταστάσεων να ταυτίζεται με τον αριθμό καταστάσεων στην μικρότερη ακολουθία της κατηγορίας. Δύο νοηματιστές πραγματοποιούν τα 262 νοήματα που αποτελούν το λεξιλόγιο αναγνώρισης παράγοντας ένα γλωσσικό σώμα 3930 δειγμάτων, που χωρίζεται σε 3 σύνολα εκπαίδευσης/δοκιμής με τα δύο πρώτα να περιλαμβάνουν τον ίδιο νοηματιστή για εκπαίδευση και επαλήθευση και το τρίτο να χρησιμοποιείται για επικύρωση της των δυνατοτήτων διανοηματιστικής αναγνώρισης. Το ποσοστό αναγνώρισής τους κυμαίνεται από 51.8% – 91.1% για τα διαφορετικά σύνολα χαρακτηριστικών γνωρισμάτων, 56.2% και 47.6% για τα πειράματα ανεξαρτησίας από τον νοηματιστή και 72.8% για την αναγνώριση σε επίπεδο πρότασης 14 διαφορετικών προτάσεων που περιέχουν συνολικά 26 νοήματα.

Οι Bauer και Hienz στα [15, 108] επεκτείνουν την αρχιτεκτονική που προτείνεται στο [8] για αναγνώριση συνεχούς νοηματισμού με την ενσωμάτωση ένος ευρετικού αλγορίθμου αναζήτησης που αποτελεί βελτιστοποίηση της αναζήτησης του πρώτου καλύτερου (best-first). Αυτή η δεντρική αναζήτηση ακτίνας (beam tree search) αντί της αναζήτησης όλων των πιθανών διαδρομών, χρησιμοποιεί ένα κατώφλι ώστε να εξετάζει μόνο μια ομάδα πιθανών υποψηφίων. Η εξαγωγή χαρακτηριστικών γνωρισμάτων που χρησιμοποιείται στην εργασία είναι πανομοιότυπη με αυτήν που υιοθετείται στο [8] και η το πειραματικό σύνολο αποτελείται από 3.5 ώρες καταγραφών εκπαίδευσης και 0.5 ώρα δεδομένων επαλήθευσης, καταγεγραμμένα από έναν χρήστη. Οι 97 κατηγορίες νοημάτων αναγνωρίστηκαν με ακρίβεια που κυμαίνεται από 94% ως 2.2% ανάλογα με την επιλογή του συνόλου των γνωρισμάτων ενώ οι απαρατήρητες αρκετά καλές επιδόσεις παρατηρήθηκαν και σε ακολουθίες νοημάτων οι οποίες δεν είχαν παρουσιαστεί στο σύστημα κατά την εκπαίδευση. Ενώ σε άλλη εργασία [14] η ίδια ερευνητική ομάδα επεκτείνει το προτεινόμενο σύστημα εμπλουτίζοντας το με γλωσσικό πρότυπο και ένα δίγραμμο πρότυπο βελτιώνοντας ελαφρώς το ποσοστό αναγνώρισης. Το ενδιαφέρον σημείο σε αυτή την εργασία είναι πώς η αναγνώριση μεμονωμένων νοημάτων επεκτείνεται σε συνεχή μέσω της αυτόματης ανίχνευσης των ορίων των νοημάτων. Οι Bauer και Kraiss στο [16] παρουσιάζουν μια στατιστική προσέγγιση, βασισμένη στον κανόνα

απόφασης Bayes και την ενσωμάτωση αυτοοργανούμενων υπομονάδων, όπου καθορίζεται ένα περιορισμένο σύνολο υπομονάδων νοημάτων απαιτούνται προκειμένου να εκπαιδευτούν HMM για αυτά, παρά για ολόκληρο το νόημα. Αυτή η προσέγγιση παρουσιάζει αρκετά πλεονεκτήματα σχετικά με τη συμβατική διαμόρφωση HMM δεδομένου ότι μειώνει τον όγκο των δεδομένων κατάρτισης και κάθε νόημα μοντελοποιείται ως μια ακολουθία πεπερασμένων υπομονάδων. Αυτή η ευέλικτη αρχιτεκτονική μπορεί ευκολότερα να προσαρμοστεί σε απαιτήσεις λεξιλόγια μεγάλης κλίμακας και ανεξαρτησίας από τον χρήστη. Η συσταδοποίηση ακολουθείται από την αυτοοργανούμενη προτυποποίηση. Τα πρώτα αποτελέσματα για αυτήν την προσέγγιση δεν είναι πολύ ενθαρρυντικά δεδομένου ότι αναφέρουν ποσοστό αναγνώρισης 80.8% σε 12 νοήματα και 10 καθορισμένες υπομονάδες. Τέλος, στο [108] πειράματα σε γλωσσικό σώμα της Γερμανικής νοηματικής γλώσσας με 52 νοήματα με την χρήση HMM και πιθανότητες βάσει μονό/δίγραμμο προτύπου για μεμονωμένα και συνδεδεμένα νοήματα οι ίδιοι ερευνητές επιτυγχάνουν ποσοστά αναγνώρισης 92.2-95.0% ανάλογα με το κατώφλι της αναζήτησης ακτίνας και το πρότυπο πιθανότητας (μονό ή δίγραμμο).

Στο [47] παρουσιάζεται ένα εναλλακτικό σύνολο χαρακτηριστικών γνωρισμάτων σε μια προσπάθεια να μειωθεί η υπολογιστική πολυπλοκότητα και διευρυνθούν οι κατηγορίες συσκευών που μπορούν να υλοποιήσουν τέτοιους αλγορίθμους, π.χ. κινητές συσκευές με περιορισμένους πόρους υλικού. Υιοθετούν την απόσταση κινούμενων τμημάτων (Moving Block Distance - MBD), βασισμένη στην οπτική ροή. Ένα πλαίσιο διαιρείται σε τμήματα, τα οποία αριθμούνται και συντάσσονται ένα μονοδιάστατο διάνυσμα. Η διαφορά μεταξύ των τιμών διαδοχικών (από αριστερά προς τα δεξιά και πάνω προς κάτω) κινούμενων τμημάτων υπολογίζεται και διαμορφώνει το χαρακτηριστικό διάνυσμα κάθε πλαισίου. Τα χαρακτηριστικά των HMM ήταν αριστερά-προς-τα-δεξιά, 15 καταστάσεις, συνεχές με (1, 5, 25) Γκαούσιαν συναρτήσεις κατανομής πιθανότητας και εφαρμόστηκαν σε ένα λεξιλόγιο 33 λημμάτων επιτυγχάνοντας το εντυπωσιακό ποσοστό αναγνώρισης 99.5%.

Οι Liang και Ouhyoung στο [148], μια αρκετά πρώιμη εργασία, παρουσιάζουν ένα σύστημα σε πραγματικό χρόνο για αναγνώριση λεξιλογίων μεγάλης κλίμακας συνεχούς νοηματικής γλώσσας ένα χρησιμοποιώντας DataGlove ως είσοδο και HMM ως κατηγοριοποιητή. Η ανίχνευση ασυνέχειας για την χρονική κατάτμηση επιτυγχάνεται με τον έλεγχο χρονικά μεταβλητών παραμέτρων παραμέτρων, σε μια ενδιαφέρουσα προσέγγιση και για ένα λεξιλόγιο 250 νοημάτων αποτελούμενο από 51 θεμελιώδεις στάσεις σώματος, 6 προσανατολισμούς και 8 πρωτεύουσες κινήσεις επιτυγχάνουν ποσοστό αναγνώρισης 80.4% για συνεχή, εξαρτούμενο από τον χρήστη νοηματισμό και 94.8% για μεμονωμένα νοήματα.

Οι Pashaloudi και Margaritis στο [185] προσπαθούν να αναγνωρίσουν γράμματα της Ελληνικής νοηματικής γλώσσας, λαμβάνοντας ως είσοδο εικόνες χεριών και εξάγοντας γεωμετρικές ιδιότητες, που κατασκευάζουν τα διανύσματα χαρακτηριστικών γνωρισμάτων. Τα αντίστοιχα διανύσματα έχουν ως αφετηρία το κέντρο μάζας του χεριού και τελικό σημείο σημεία στα όρια του χεριού, τα οποία εξάγουν με έναν αλγόριθμο παρακολούθησης ορίων. Για 16 επαναλήψεις των 24 γραμμάτων επιτυγχάνουν 80.56% – 97.22% για διαφορετικό αριθμό σημείων στα όρια του χεριού,ενώ εισάγοντας τεχνητό θόρυβο 5–10% το ποσοστό αναγνώρισης παραμένει σε υψηλά επίπεδα (90.20–86.52%) γεγονός που επιβεβαιώνει την ανοχή σε τυχαίο σφάλμα της μεθόδου.

Σε μια από τις δημοσιεύσεις με τις περισσότερες αναφορές στον ερευνητικό τομέα της αυτόματης αναγνώρισης νοηματικής γλώσσας είναι η [216] όπου οι Starner κ.α. προτείνουν πραγματοποιούν δύο πειράματα για αναγνώριση προτάσεων Αμερικανικής

νοηματικής γλώσσας. Στο πρώτο πείραμα χρησιμοποιείται μια κάμερα τοποθετημένη στο γραφείο ενώ στο δεύτερο η μικροσκοπική κάμερα εφαρμόζεται σε καπέλο που φορά ο χρήστης. Τα χέρια ανιχνεύονται, χρησιμοποιώντας δερματικό πρότυπο και μια τεχνική επέκτασης περιοχών (region growing) και δημιουργείται ένα διάνυσμα χαρακτηριστικών γνωρισμάτων δεκαέξι στοιχείων από τις συντεταγμένες κάθε χεριού x,y, το διάνυσμα κίνησης μεταξύ πλαισίων, την έκταση της περιοχής χεριών (σε εικονοστοιχεία), τη γωνία του άξονα λιγότερης αδράνειας, το μήκος αυτού του διανύσματος Eigen και τέλος η εκκεντρότητα των ορίων της έλλειψης. Το πειραματικό σώμα αποτελείται από 500 προτάσεις και τα δύο συστήματα επιτυγχάνουν μια ακρίβεια, που μετριέται ως συνάρτηση του συνολικού αριθμού λέξεων και του αριθμού διαγραφών, αντικαταστάσεων και εισαγωγών, για διαφορετικούς σύνολα χαρακτηριστικών γνωρισμάτων και γραμματικών περιορισμών 74.5–91.9% και 96.8–98.2% αντίστοιχα. Οι Brashear κ.α. στο [24] επεκτείνουν την προηγούμενη εργασία τους, που περιγράφεται στο [216], συμπεριλαμβάνοντας στα χαρακτηριστικά εισόδου μετρήσεις από επιταχυνσιόμετρο. Τρία επιταχυνσιόμετρα με τρεις βαθμούς ελευθερίας τοποθετούνται στον αριστερό και στον δεξί καρπό και στον κορμό στοχεύουν να συλλέξουν πληροφορίες τις οποίες το σύστημα με την οπτική είσοδο δυσκολεύεται να εξάγει. Ενώ το αρχικό σύστημά τους [216] ήταν ένα σύστημα απόδειξη της ορθότητας σκέψης αυτή η εργασία [24] είναι απόδειξη της μεταφερσιμότητας και της δυνατότητας του συστήματος να 'φορεθεί' από τον χρήστη. Διαφορετικές παράμετροι των HMM για σύντομα και μεγαλύτερα νοήματα επιλέγονται, ενώ το σύστημα εκτελείται σε πραγματικό χρόνο (10 πλαίσια ανά δευτερόλεπτο και 8–12 πακέτα δεδομένων επιταχυνσιόμετρου ανά δευτερόλεπτο) εξετάστηκε σε 71 προτάσεις, 7 νοημάτων, σε ένα σύνολο 497 λημμάτων. Η βελτίωση σε σχέση με το σύστημα βασισμένο στην οπτική είσοδο (52.38%) και τα επιταχυνσιόμετρα (65.87%) και το συνδυασμένο σύνολο (90.48%).

Οι Tanibata κ.α. στο [220] παρουσιάζουν μια αρκετά δημοφιλή προσέγγιση όπου αλγόριθμοι ανίχνευσης δέρματος ανιχνεύουν και παρακολουθούν τα χέρια και το κεφάλι του χρήστη, εξάγονται χαρακτηριστικά γνωρίσματα και τελικά HMMs εκπαιδεύονται για μεμονωμένη αναγνώριση νοημάτων. Ένα χρωματικό πρότυπο δέρματος χρησιμοποιείται για να εντοπίσει το πρόσωπο και τα χέρια ενώ παρακολουθούνται επίσης και οι αγκώνες με την τεχνική της ομοιότητας προτύπων με την ταύτιση ενός προτύπου αγκώνα ενώ με παρόμοιο τρόπο αντιμετωπίζεται και το φαινόμενο της επικάλυψης. Ο γενικός αλγόριθμος ανίχνευσης χεριών υποθέτει στατικό παρασκήνιο και ως εκ τούτου η αφαίρεση υποβάθρου πραγματοποιείται αρκετά εύκολα. Στα χαρακτηριστικά γνωρίσματα που εξάγονται και χρησιμοποιούνται για την εκπαίδευση των HMM περιλαμβάνουν χαρακτηριστικά γνωρίσματα βασισμένα στην περιοχή του χεριού όπως η ομαλότητα και η έκταση, την κατεύθυνση κίνησης των χεριών, τις σχετικές συντεταγμένες θέσης και τον αριθμός προεξοχών. Ένα γλωσσικό σώμα της Ιαπωνικής νοηματικής γλώσσας τριών δειγμάτων 70 λέξεων, εκ των οποίων 65 επεξεργάσθηκαν επιτυχώς και από αυτά 64 ταξινομήθηκαν σωστά. Κατά τη διάρκεια της κατηγοριοποίησης ένα στιγμιότυπο θεωρείται υποψήφιο μόνο εάν ο αλγόριθμος Viterbi οδηγηθεί σε τελική κατάσταση.

Οι Zahedi κ.α. στηρίζουν την μέθοδο εξαγωγής χαρακτηριστικών γνωρισμάτων τους σε απλά χαρακτηριστικά γνωρίσματα βασισμένα στην εμφάνιση (appearance), εξαλείφοντας την ανάγκη για κατάτμηση ή παρακολούθηση χεριών και κεφαλιού, ενώ για ακόμη μια φορά απλά HMM χρησιμοποιούνται για την κατηγοριοποίηση. Το σύνολο χαρακτηριστικών γνωρισμάτων περιλαμβάνει την αρχική εικόνα (original image - OI), κατώφλι έντασης δερματικής χρωματικής χροιάς (skin intensity thresholding -

SIT), πρώτη και δεύτερη παράγωγος (FD και SD αντίστοιχα) διαδοχικών πλαισίων, θετική πρώτη παράγωγο (positive first derivative - PFD), που περιέχει εικονοστοιχεία που δεν άνηκαν σε δερματικές περιοχές στο προηγούμενο πλαίσιο και ανήκουν στο τρέχον πλαίσιο, αρνητική πρώτη παράγωγος (NFD), το αντίστροφο του PFD και την ένωση PFD και NFD και την απόλυτη πρώτη παράγωγο (AFD). Σε 110 εκτελέσεις 10 νοημάτων της Αμερικάνικης νοηματικής γλώσσας από τρεις νοηματιστές και την χρήση της μεθόδου παράλειψης ενός στιγμιοτύπου (leave one out) το ποσοστό λάθους ήταν 7%. Ένα σημείο που εγείρει ερωτήσεις και σίγουρα αξίζει περαιτέρω διερεύνησης είναι ότι όταν προστίθεται και δεύτερη πλευρική κάμερα το ποσοστό λάθους κατηγοριοποίησης δεν μειώνεται.

Οι Zieren και Kraiss στο [270] πειραματίζονται με διαφορετικά περιβάλλοντα, την εξάρτηση από τον νοηματιστή και λεξιλόγιο Βρετανικής νοηματικής γλώσσας (British Sign Language - BSL). Οι υποενότητες επεξεργασίας εικόνας περιλαμβάνουν μοντελοποίηση υποβάθρου, εντοπισμό και παρακολούθηση χεριών βάσει δερματικού χρωματικού προτύπου δερμάτων, ενώ οι επικαλύψεις επιλύονται με την βοήθεια βιομηχανικού προτύπου σώματος και η κανονικοποίηση ανά νοηματιστή φροντίζει για την ανεξαρτησία από τον χρήστη και την ευρωστία του συστήματος. Τέλος, υιοθετούνται HMM για την κατηγοριοποίηση. Τα χαρακτηριστικά γνωρίσματα που χρησιμοποιούνται για την εκπαίδευση και τη επαλήθευση περιλαμβάνουν τις συντεταγμένες του κέντρου βάρους των χεριών και τις παραγώγους τους, την έκταση περιοχής και την παράγωγο της, την πυκνότητα και την εκκεντρότητα περιοχής, την αναλογία αξόνων αδράνειας και ορθογώνιου στον κύριο άξονα και τον προσανατολισμό του κυρίως άξονα. Ένα γλωσσικό σώμα 232 μεμονωμένων νοημάτων με πέντε επαναλήψεις από 6 νοηματιστές σε διαφορετικά περιβάλλοντα, αποτέλεσε το πειραματικό σύνολο και ένα ποσοστό αναγνώρισης μεταξύ 94.1% και 98.9% επετεύχθη για εξαρτώμενο σενάριο προσώπων και 3.7% και 44.1% για σενάριο ανεξάρτητο προσώπων, ενώ για μια ιδανική ρύθμιση και ένα μειωμένο λεξιλόγιο 18 νοημάτων επιτυγχάνεται ένα ποσοστό αναγνώρισης 99.3%. Ο χρόνος επεξεργασίας ποικίλλει ανάλογα με την ανάλυση και το μέγεθος του συνόλου χαρακτηριστικών γνωρισμάτων, αλλά ενδεικτικές τιμές είναι 11.79s/4.15s/3.08s/2.92s ανά νόημα.

Ενώ τα HMM χρησιμοποιούνται εκτεταμένα, υπάρχουν υπόνοιες πως, τουλάχιστον στην τυποποιημένη μορφή τους, μοιάζουν ελλιπή στην αντιμετώπιση των περισσοτέρων προκλήσεων που αναφέρονται στην κατηγοριοποίηση. Στην εκπαίδευση των HMM, κάθε πρότυπο νοήματος υπολογίζεται χωριστά χρησιμοποιώντας τις αντίστοιχες ακολουθίες παρατήρησης εκπαίδευσης χωρίς την εξέταση στοιχείων που είναι αρκετά γειτονικά αλλά δεν ταιριάζουν ακριβώς με τις καταστάσεις παρατήρησης. Η Δυναμική Χρονική Στρέβλωση (Dynamic Time Warping - DTW) και τα HMM συσχετίζονται αρκετά μεταξύ τους, αν και κάθε μέθοδος έχει τις δικές τις ιδιότητες. Η DTW αναζητεί το βέλτιστο μονοπάτι ενώ η συνάρτηση πιθανότητας των HMM αθροίζει την πυκνότητα κατά μήκος όλων των πιθανών μονοπατιών. Είναι επίσης γνωστό πως τα HMM δεν είναι ανεκτικά στην παρουσία μη χαρακτηριστικών δεδομένων (outliers) τόσο στο σύνολο εκπαίδευσης που χρησιμοποιείται για την εκπαίδευση των παραμέτρων τους όσο στην προσπάθεια εκτίμησης της πιθανότητας συμμετοχής [34].

4.1.2.3.3 Παραλλαγές HMM Οι Fang, Gao κ.α. συνιστούν μια ερευνητική ομάδα εξαιρετικά ενεργή στο πεδίο της αυτόματης αναγνώρισης νοηματικής γλώσσας με την δημοσίευση αρκετών άρθρων [83, 84, 85, 96, 95, 153, 268, 269]. Αρχικά στο [97]

συνδύασαν τεχνητά νευρωνικά δίκτυα και δυναμικό προγραμματισμό για την αυτόματη κατάτμηση συνεχούς νοηματισμού λεξιλογίου μεγάλης κλίμακας Κινέζικης νοηματικής γλώσσας σε υπονοήματα. Με την εφαρμογή μιας μεθόδου ταχείας αντιστοίχησης κατασκευάζει έναν κατάλογο υποψηφίων λέξεων. Χαρακτηριστικές θέσεις χεριών, προσανατολισμοί και χειρομορφές, αποκαλούμενα υπονοήματα, χρησιμοποιούνται ως καταστάσεις. Η επένθεση μετακίνησης αντιμετωπίζεται από πρόσθετα HMM εξαρτώμενα από το πλαίσιο. Δύο γάντια δεδομένων με 36 αισθητήρες και δύο συσκεύες παρακολούθησης τοποθετούνται στα γάντια διαμορφώνουν ένα διάνυσμα εισόδου με διάσταση 48. Το πειραματικό σύνολο δεδομένων αποτελείται 82 χειρομορφές, 50 θέσεις στο σώμα και 50 προσανατολισμούς παλάμης και τα πειράματα εκτελέσθηκαν τόσο για μεμονωμένη όσο και για συνεχή αναγνώριση νοηματισμού. Για την αναγνώριση μεμονωμένων νοημάτων, 1065 βασικά νοήματα που εκτελέσθηκαν οκτώ φορές από έναν καθηγητή νοηματικής γλώσσας, 7 επαναλήψεις χρησιμοποιήθηκαν για την εκπαίδευση και η τελευταία επανάληψη για επαλήθευση, με την διαδικασία της διαγώνιας επικύρωσης τα υπονοήματα κατατμήθηκαν επιτυχώς σε ποσοστό 80–93.2%. Για τη περίπτωση συνεχούς νοηματισμού, ένα σώμα με 80 προτάσεις από 220 νοήματα χρησιμοποιήθηκε και το ποσοστό αναγνώρισης είναι 82–96.3% και 82–98.2% χωρίς και με κανονικοποίηση χαρακτηριστικών γνωρισμάτων αντίστοιχα. Επιπλέον, χρησιμοποιώντας τα πρότυπα εξαρτώμενα από το πλαίσιο (Context Dependent Models) το μέσο ποσοστό αναγνώρισης φθάνει σε 95.2%, ενώ χωρίς αυτά τα πρότυπα το ποσοστό ήταν αρκετά κατώτερο 73.1%.

Επιπλέον, στο [153] προτείνεται ένα παράλληλο πρότυπο πολλαπλής ροής (parallel multistream) για τη ενσωμάτωση της κίνησης των χειλιών στην γλωσσική αναγνώριση, επιτυγχάνοντας ποσοστά αναγνώρισης μεταξύ 93.4% και 95.2% για διαφορετικά μεγέθη λεξιλογίων. Στο [83], εισήγαγαν ένα υβριδικό σύστημα SOFM/HMM για να λύσουν το πρόβλημα εξάρτησης από τον χρήστη για τα πρακτικά εφαρμόσιμα σχήματα αναγνώρισης. Συνδυάζοντας την ισχυρή απόδοση αυτοοργάνωσης των SOFM με τις εξαιρετική ικανότητα επεξεργασίας χρονικών προτύπων των HMM βελτιώνουν το ποσοστό αναγνώρισής τους ενός απλού HMM κατά 5%. Κάθε κέντρο κόμβου του SOFM θεωρείται ως κατάσταση του HMM και ένας νέος αλγόριθμος αναγνώρισης προτείνεται για να βελτιώσει την απόδοση των HMM μετά από προσεκτική ανάλυση της συνάρτησης πυκνότητας πιθανότητας HMM και την ενσωματώση της μετέπειτα (posteriori) πιθανότητας αυτής της κατηγορίας σε ολόκληρο το σύνολο. Αυτή η διαδικασία βελτιώνει περαιτέρω το ποσοστό αναγνώρισης κατά 1.9%. Βέβαια στην εργασία αυτή το SOFM επιτελεί μια απλή διαδικασία συσταδοποίησης και οι συγγραφείς δεν εκμεταλλεύονται πλήρως τις δυνατότητες του καθώς στην θέση του θα μπορούσε να χρησιμοποιηθεί οποιοσδήποτε αλγόριθμος συσταδοποίησης χωρίς ιδιαίτερη διαφορά. Ένα από τα χαρακτηριστικά του SOFM είναι πως η σχέση γειτνίασης μεταξύ των κόμβων αποτυπώνεται καλύτερα από οποιονδήποτε άλλον αλγόριθμο συσταδοποίησης, ιδιότητα που παραμένει ανεκμετάλλευτη. Για τη συλλογή δεδομένων, δύο CyberGloves και τρεις καταγραφείς θέσης χρησιμοποιήθηκαν ως συσκευές εισόδου. Δύο καταγραφείς θέσης τοποθετήθηκαν στους καρπούς και ένας τρίτος τοποθετήθηκε στην πλάτη του νοηματιστή ώστε να χρησιμοποιηθεί ως σημείο αναφοράς. Ένα πειραματικό σύνολο δεδομένων 7 νοηματιστών, 3 επαναλήψεων, 208 νοημάτων χωρίστηκε σε ένα εγγεγραμμένο και ένα μη εγγεγραμμένο σύνολο. Τα μέσα ποσοστά αναγνώρισης για HMM, SOFM/HMM και το ενισχυμένο σύστημα ήταν 90.7%, 95.3%, 96.6% και 83.2%, 88.2%, 90.1% για το εγγεγραμμένο και μη εγγεγραμμένο σύνολο αντίστοιχα. Σε μία προσπάθεια [84, 96] να αντιμετωπιστούν οι δυσκολίες στη αναγνώριση λεξι-

λογίων μεγάλης κλίμακας και στοχεύοντας στην μείωση του απαιτούμενου χρόνου αναγνώρισης χωρίς απώλεια ακρίβειας η ερευνητική ομάδα επεκτείνει το προηγούμενο σύστημα με ένα ασαφές δέντρο απόφασης με ετερογενείς κατηγοριοποιητές. Αυτό το ασαφές δέντρο απόφασης έχει δύο επίπεδα, ένα που διαχωρίζει τα νοήματα που περιλαμβάνουν και τα δύο χέρια και έναν κατηγοριοποιητή χειρομορφής με μικρό υπολογιστικό κόστος και επιτελεί μια προεπεξεργασία ώστε οι αδύναμοι υποψήφιοι να αποκλειστούν σταδιακά από την υπόλοιπη διαδικασία αναγνώρισης. Το παραπάνω σύστημα SOFM/HMM χρησιμοποιείται μόνο επί των κόμβων φύλλα του δέντρου ώστε να ληφθεί η τελική απόφαση. Το μοντέλο Γκαουσσιανών μιγμάτων (Gaussian mixture model - GMM) υιοθετείται πρώτα ως ταξινομητής για το πλήθος των χεριών και έπειτα η μέθοδος μηχανών πεπερασμένων καταστάσεων προτείνεται ως κατηγοριοποιητής χειρομορφής. Ένα εκτεταμένο σύνολο δεδομένων που περιλαμβάνει τον εντυπωσιακό αριθμό των 61356 στιγμιοτύπων από 5113 νοήματα, έξι νοηματιστών και δύο επαναλήψεων χρησιμοποιείται. Ο χρόνος επεξεργασίας για την αναγνώριση των νοημάτων βελτιώθηκε σημαντικά κατά ένα λόγο 11 φορές γρηγορότερο από το προηγούμενο σύστημα SOFM/HMM και επίσης το ποσοστό αναγνώρισής βελτιώθηκε ελαφρώς κατά 0.95%. Το ίδιο γλωσσικό σώμα χρησιμοποιήθηκε στα δύο πειράματα του [85], όπου το πρώτο στοχεύει στην επικύρωση της ικανότητας να συσταδοποιηθούν παρόμοια υπονοήματα νοημάτων στην ίδια κατηγορία και το δεύτερο εξαγωγή αυτών των υπονοημάτων για την Κινέζικη νοηματική γλώσσα. Κάθε τμήμα του νοήματος αναπαριστάται από καταστάσεις στο HMM και έτσι τα νοήματα μπορούν να χωριστούν κάτω σε διάφορα τμήματα. 238 υπονοήματα εξήχθησαν αυτόματα από 5113 νοήματα και μπορούν να χρησιμοποιηθούν ως βασικές μονάδες για την αναγνώριση λεξιλογίων μεγάλης κλίμακας με αποδεκτή απόδοση.

Ένας αλγόριθμος χρονικής συσταδοποίησης βασισμένος στον k-means αλγόριθμο προτείνεται στο [96]. Η δυναμική χρονική στρέβλωση υιοθετείται ως μετρική απόστασης επειδή μπορεί να υπολογίσει την απόσταση μεταξύ δύο χρονικών ακολουθιών με την ευθυγράμμιση διαφορετικών χρονικών σημάτων ελαχιστοποιώντας την συνολική απόσταση και το σχετικό μονοπάτι. 309 από τις 317 προκαθορισμένες συστάδες προσδιορίστηκαν αυτόματα και ένα μέσο ποσοστό αναγνώρισης 90.5% επιτεύχθηκε. Τέλος, βασιζόμενοι στο σύστημα SOFM/HMM που παρουσιάστηκε παραπάνω, μια αρχιτεκτονική SOFM/SRN/HMM προτείνεται στο [95] για αναγνώριση συνεχούς νοηματισμού ανεξαρτήτως χρήστη. Σύμφωνα με αυτή την αρχιτεκτονική ένα απλό αναδρομικό δίκτυο (simple recurrent network - SRN) για να οριοθετήσει χρονικά τα νοήματα κατά τον συνεχή νοηματισμό και τα αποτελέσματα της εφαρμογής του SRN λαμβάνονται ως καταστάσεις των HMM στα οποία ο αλγόριθμος Viterbi υιοθετείται για αναζήτηση της βέλτιστης ακολουθίας νοημάτων. Τα πειραματικά αποτελέσματα καταδεικνύουν ότι η προτεινόμενη αρχιτεκτονική έχει καλύτερη απόδοση έναντι του συμβατικού HMM και επιτυγχάνει ποσοστό αναγνώρισης σε επίπεδο νοήματος 82,9% και 86,3% συνεχή νοηματισμό ανεξαρτήτως χρήστη.

Το [240] εξετάζει το ζήτημα της κλιμάκωσης του συστήματος αναγνώρισης σε λεξιλόγια που πλησιάζουν το πλήρες λεξικλεξιλόγιο της νοηματικής προτείνοντας την χρήση φωνημάτων έναντι ολόκληρων νοημάτων ως βασικές μονάδες στη διαδικασία αναγνώρισης. Ένα HMM εκπαιδεύεται για κάθε φώνημα και οι συναρτήσεις Γκάους των καταστάσεων συσταδοποιούνται και δομείται ένα δίκτυο δεντρικής δομής το οποίο αργότερα υπόκειται σε διαδικασία κλαδέματος (pruning) στοχεύοντας στην μείωση του διαστήματος αναζήτησης και η μέθοδος N-Best-pass χρησιμοποιείται για να εγγυηθεί ακρίβεια αναγνώρισης υποβοηθούμενο από ένα δίγραμμο γλωσσικό πρότυπο.

Ένας νοηματιστής εκτέλεσε 2439 φωνήματα πέντε φορές και 5119 νοήματα που περιέχουν 200 προτάσεις της νοηματικής γλώσσας ποικίλου μήκους (2–10 νοήματα ανά πρόταση) της Κινέζικης νοηματικής γλώσσας διαμορφώνοντάς ένα πειραματικό σύνολο δεδομένων που παράγει ποσοστό αναγνώρισης φωνήματος 100%, απαιτώντας 3.2 δευτερόλεπτα/φώνημα ενώ η τεχνική με την συσταδοποίηση των Γκαουσιανών συναρτήσεων 99,7% και 0,22 δευτερόλεπτα/φώνημα αντίστοιχα. Η διαδικασία με το δίκτυο δενδρικής δομής αποδίδει με ποσοστό 92,8% για αναγνώριση σε επίπεδο νοήματος.

Η ίδια ομάδα στο [243] αντιμετωπίζει το θέμα της εξάρτησης των χαρακτηριστικών γνωρισμάτων από την γωνία λήψης της κάμερας και χρησιμοποιεί ένα σχήμα αναγνώρισης με στοιχεία ομογραφίας όπου κάθε νόημα αναπαριστάται ως σειρά μικροσκοπικών κινήσεων των χεριών και χωρίζεται σε ατομικές μονάδες τριών διαδοχικών πλαισίων. Τεχνικές στερέοψης και η μέθοδος του πλησιέστερου γείτονα χρησιμοποιούνται καθοριστεί η ταυτοποίηση με την χρήση ομογραφίας. 64 νοήματα της Κινεζικής νοηματικής γλώσσας εκτελέσθηκαν από έναν νοηματιστή φορώντας χρωματιστά γάντια και καταγράφηκαν από την εμπρόσθια όψη και από μια διαφορετική όψη που κυμαίνεται από $-30°$ ως $+30°$. Η χρονική κατάτμηση και η ο χαρακτηρισμός των σημείων εκτελέσθηκε χειρωνακτικά και τα ποσοστά αναγνώρισης ήταν μεταξύ 71,8% και 92,1%. Στο [268] και στο [269] χρωματιστά γάντια υιοθετούνται για να βοηθήσουν τις τεχνικές όρασης υπολογιστών. Εφαρμόζεται ανάλυση πρωτευουσών συνιστωσών (PCA) για καλύτερη αναπαράσταση των χαρακτηριστικών των ακροδακτύλων και HMM συνδεδεμένης πυκνότητας μιγμάτων (Tied-Mixture Density Hidden Markov Models - TMDHMM) επιταχύνουν την αναγνώριση χωρίς σημαντική απώλεια ακρίβειας αναγνώρισης. 1756 δείγματα εκπαίδευσης από 439 νοήματα διαμόρφωσαν το πειραματικό σώμα επιτυγχάνοντας 88,6%, 89,7% και 92,5% για τρία επίπεδα περιγραφής ενώ συγκρινόμενα με τα συμβατικά HMM επιτυγχάνουν ένα ελαφρώς χειρότερο ποσοστό αναγνώρισης (1,2%), αλλά ο χρόνος επεξεργασίας είναι περίπου ο μισός (0.285–0.135s).

Ο Vogler, ένας από τους πιό ενεργούς ερευνητές στην περιοχή, εστιάζει στην μοντελοποίηση βάσει φωνημάτων, παράλληλα HMM και εφαρμογή σε λεξιλόγια μεγάλης κλίμακας στη διατριβή του [233]. Στο [235] οι Vogler και Metaxas αντιμετωπίζουν το πρόβλημα του της συνάρθρωσης στις νοηματικές γλώσσες. Τα νοηματικά φωνήματα (signemes) μπορούν να εμφανιστούν ταυτόχρονα κατά τον νοηματισμό και ο αριθμός πιθανών συνδυασμών φωνημάτων μετά από επιβολή γλωσσικών περιορισμών είναι της τάξεως του 10^8, ένας αριθμός απαγορευτικός για μοντελοποίηση με συμβατικά HMM. Παραγοντικά (Factorial) HMMs και συζευγμένα HMMs (coupled) είναι δύο παραλλαγές των HMM που προσπαθούν μοντελοποιήσουν πολλαπλές διαδικασίες που συμβαίνουν παράλληλα, αλλά απαιτούν ακόμα την εκτίμηση των συνδυασμών κατά την εκπαίδευση. Ενώ τα παράλληλα HMM (PaHMM), που μοντελοποιούν παράλληλες διαδικασίες ανεξάρτητα, παρόμοιος με FaHMMs αλλά η έξοδος και οι καταστάσεις των διαδικασιών είναι ανεξάρτητες. Οι χρόνοι κατάρτισης είναι πολυωνυμικοί σε σχέση με τον αριθμό των καταστάσεων και γραμμικοί με τον αριθμό των παράλληλων διαδικασιών ενώ ο αλγόριθμος αποκωδικοποίησης είναι πέρασμα δειγμάτων (token passing) αντί του τυποποιημένου αλγορίθμου Viterbi. Το σύστημα παρακολούθησης Ascension Technologies MotionStar χρησιμοποιήθηκε για να καταγράψει 400 προτάσεις εκπαίδευσης, 99 προτάσεις δοκιμής από ένα λεξιλόγιο 22 νοημάτων που περιελάμβαναν 30 χειρομορφές, 8 προσανατολισμούς χεριών, 20 σημαντικές θέσεις του σώματος και 40 μετακινήσεις. Το πρότυπο μετακίνηση-παύσης Liddell και Johnson υιοθετήθηκε και τα ποσοστά αναγνώρισης ήταν 84.85% σε επίπεδο πρότασης

και 94.23% για την σε επίπεδο νοημάτων, τα οποία είναι ανώτερα σε σχέση με τις αντίστοιχες επιδόσεις των συμβατικών HMM που ήταν 80.81% και 93.27% αντίστοιχα. Σε πιό πρόσφατη εργασία τους [236] ενσωματώνουν επίσης την χειρομορφή (γωνίες αρθρώσεων) με την χρήση συσκευής Virtual Technologies Cyberglove. Δύο τύποι πειραμάτων πραγματοποιήθηκαν, ο πρώτος στόχευε στην επικύρωση της ορθότητας της επιλογής χαρακτηριστικών γνωρισμάτων χειρομορφής και ο δεύτερος στην μοντελοποίηση ανεξάρτητων καναλιών πληροφορίας. Το πειραματικό σύνολο δεδομένων ήταν 499 προτάσεις, με διάρκεια μεταξύ 2 και 7 νοημάτων και συνολικά 1604 νοήματα από ένα λεξιλόγιο 22 νοημάτων. Στο πρώτο πείραμα η περιγραφή βασισμένη σε τετράπλευρο της χειρομορφής αποδείχθηκε πιό εύρωστη από τις ακατέργαστες γωνίες αρθρώσεων επιτυγχάνοντας 95.21% και 83.15% αντίστοιχα για την κατηγοριοποίηση της χειρομορφής. Στο δεύτερο πείραμα τέσσερις αρχιτεκτονικές εξετάστηκαν, δηλαδή συμβατικά HMM, PaHMM για μετακίνηση των δύο χεριών, PaHMM για μετακίνηση και χειρομορφή του δεξιού χεριού και PaHMMs και τα τρία κανάλια και αναφέρονται ποσοστά αναγνώρισης 80.81%, 84.85%, 88.89% και 87.88% για την αναγνώριση σε επίπεδο πρότασης και 93.27%, 94.55%, 96.15% και 95.51% για την αναγνώριση σε επίπεδο νοήματος.

Οι Wilson και Bobick στο [249] αντιμετωπίζουν το πρόβλημα της συστηματικής παρέκκλισης στην έξοδο των αισθητήρων και των συμφραζόμενων πληροφοριών, περιπτώσεις όπου η συμβατική εφαρμογή HMM υστερεί. Προτείνουν δύο πλαίσια βασισμένα στα HMM σχεδιασμένα να προτυποποιήσουν και να αναγνωρίσουν χειρονομίες που παρεκκλίνουν συστηματικά. Αρχικά, παραμετρικά HMM (PHMM), όπου η συστηματική παραλλαγή υποτίθεται ότι είναι εκ φύσεως επικοινωνιακή, η χειρονομία εισόδου υποτίθεται ότι ανήκει σε μια οικογένεια χειρονομιών χρησιμοποιούν μια γραμμική σχέση μεταξύ της παραμετρικής ποσότητας χειρονομίας (π.χ. έκταση) και τα κέντρα των συναρτήσεων πυκνότητας πιθανότητας του PHMM. Διατυπώνουν μια μέθοδο προσδοκίας-μεγιστοποίησης (Expectation-Maximization - EM) για το PHMM με την μορφή της κατανομής πιθανότητας εξόδουνα εξαρτώνται από την παράμετρο χειρονομίας, η τιμή της οποίας υποτίθεται ότι συνοδεύει το κάθε διάνυσμα εισόδου. Τρία πειράματα πραγματοποιήθηκαν για να επικυρώσουν την δυνατότητα παραμετροποίησης της εκπαίδευσης, την χρησιμότητα των PHMM στην αναγνώριση χειρονομιών και την απόδοση της τεχνικής PHMM κάτω από διαφορετικά ποσοστά θορύβου.

Οι Wang κ.α. στο [241] εφαρμόζουν μια αρκετά καθιερωμένη μέθοδο αναγνώρισης νοηματικής όπου γάντι καταγραφής δεδομένων Cyberglove και η συσκευή παρακολούθησης Flock of Birds χρησιμοποιούνται για να καταγράψουν μετρήσεις, που υπόκεινται κβαντισμό και συσταδοποίηση (vector quantization clustering) και τελικά εκπαιδεύουν πολυδιάστατα HMM. Πέντε χρήστες νοηματίζουν 26 γράμματα της Αμερικανικής νοηματικής γλώσσας με την χρονική κατάτμηση να εκτελείται μέσω της ανίχνευσης χαμηλής ταχύτητας των χεριών αποτελούν το πειραματικό σύνολο και επιτυγχάνει ποσοστό αναγνώρισης 95%.

Στην εκπαίδευση των HMM κάθε πρότυπο αποτιμάται ανεξάρτητα χρησιμοποιώντας τις αντίστοιχες χαρακτηρισμένες ακολουθίες παρατήρησης εκπαίδευσης χωρίς να λαμβάνονται υπόψη ακολουθίες παραπλήσιες αλλά όχι ταυτόσημες. Οι μέθοδοι DTW και HMM έχουν εγγενείς ομοιότητες αλλά αφ'ετέρου έχουν τα δικά τους μοναδικά χαρακτηριστικά. Οι Juang κ.α. [124] παρουσιάζουν μια ενοποιημένη άποψη των δύο μεθόδων και σχολιάζουν πως η DTW αναζητά την βέλτιστη διαδρομή ενώ τα HMM με την συνάρτηση πιθανότητας αθροίζει την πυκνότητα πιθανοτήτων κατά μήκους όλων των πιθανών και καταλήγουν πως η DTW προσφέρει υψηλότερο επίπεδο

αφαιρετικότητας στην αναζήτηση της βέλτιστης διαδρομής σε σχέση με τα HMM. Ο αλγόριθμος δυναμικής χρονικής στρέβλωσης παρέχει επιδρά ως μια μη γραμμική διαδικασία κανονικοποίησης προκειμένου να λειτουργήσει επιτυχώς η μετρική ομοιότητας. Ο αλγόριθμος DTW λειτουργεί επεκτείνοντας την ακολουθία πρότυπο και καταγράφοντας την επιμήκυνση που απαιτείται για αυτή την διαδικασία, με ακολουθίες που παρουσιάζουν αρκετή ομοιότητα να απαιτούν την λιγότερη επιμήκυνση. Ένα πρόβλημα που συνδέεται με τον αλγόριθμο DTW είναι ότι συνήθως απαιτεί σημαντικό υπολογιστικό κόστος για καταλήξει στην βέλτιστη διαδρομή. Αυτό είναι και το σημαντικότερο μειονέκτημα της συμβατικής τεχνικής ταύτισης προτύπων που ενσωματώνεται στον αλγόριθμο DTW. Από την άλλη τα HMM ενώ έχουν ικανοποιητική απόδοση σε χωροχρονικά προβλήματα το μεγαλύτερο μειονέκτημα τους είναι πως απαιτούν σημαντικό όγκο δεδομένων εκπαίδευσης ενώ η απόδοση τους φαίνεται να επηρεάζεται σημαντικά από την επιλογή των σχεδιαστικών παραμέτρων π.χ. αριθμός καταστάσεων, πλήθός μιγμάτων κ.τ.λ.

4.1.2.3.4 Ενισχυτικές Μέθοδοι Η μέθοδος της ενίσχυσης (boosting) είναι μια γενική μέθοδος που μπορεί να χρησιμοποιηθεί για τη βελτίωση της απόδοσης ενός δεδομένου αλγορίθμου μηχανικής μάθησης. Πιό συγκεκριμένα, βασίζεται στην αρχή πως ένας αποδοστικός αλγόριθμος κατηγοριοποίησης μπορεί να προκύψει μέσω του γραμμικού συνδυασμού πολλών αδύναμων κατηγοριοποιητών των οποίων η απόδοση μπορεί να είναι ακόμα και ελαφρώς καλύτερη της τυχαίας ταξινόμησης.

Οι Cooper και Bowden στο [45] παρουσιάζουν μια προσέγγιση για αναγνώριση λεξιλογίων μεγάλης κλίμακας χωρίς την απαίτηση της παρακολούθησης και της εξαγωγής πληροφορίας χειρομορφής εφαρμόζοντας ένα σχήμα με δύο επίπεδα. Το πρώτο επίπεδο ταξινόμησης είναι ένα σύνολο ταξινομητών συστατικών της νοηματικής, συγκεκριμένα θέση (tab), μετακίνηση (sig) και ρύθμιση χεριών (ha), βασισμένο στον αλγόριθμο AdaBoost που έχει ως στόχο την ανίχνευση παρουσίας υπομονάδων νοημάτων. Στο δεύτερο επίπεδο αυτές οι υπομονάδες συγκεντρώνονται σε επίπεδο νοήματος με την χρήση Μαρκοβιανών μοντέλων πρώτης τάξης. Το πειραματικό σύνολο δεδομένων τους αποτελείται από δέκα επαναλήψεις 164 νοημάτων που χρησιμοποιείται επίσης στο [126]. Τα ενδιάμεσα αποτελέσματα ταξινόμησής τους για τα συστατικά νοημάτων είναι 33.2%, 31.7% και 29.4% για tab, sig και ha αντίστοιχα. Το γενικό ποσοστό αναγνώρισής τους είναι 74.3% που είναι κατώτερο από αυτό που παρουσιάζεται στο [126] (79.2%), αν και οι συντάκτες δηλώνουν ως μελλοντική ερευνητική εργασία την πρόθεσή να περιλάβουν έναν ταξινομητή χειρομορφής (dez) προκειμένου να βελτιωθεί το γενικό ποσοστό αναγνώρισής τους. Οι ίδιοι ερευνητές στο [44] πειραματίζονται με τους διαφορετικές αρχιτεκτονικής ενίσχυσης, συγκεκριμένα AdaBoost και AdaPlusBoost, χρησιμοποιώντας ογκομετρικά χαρακτηριστικά γνωρίσματα ως επέκταση των haar γνωρισμάτων στο πεδίο του χρόνου, συσσωρεύοντας διαδοχικά πλαίσια σε έναν εννιαίο όγκο. Κατά συνέπεια, ένα νόημα μπορεί να θεωρηθεί ως υποσύνολο αυτού του όγκου. Το σύνολο δεδομένων τους αποτελείται από 12 νοηματιστές (9 εκπαίδευσης) που εκτελούν 5 επαναλήψεις 5 διαφορετικών νοημάτων.

Οι Bowden, Ong, Kadir κ.α. δημοσίευσαν αρκετά άρθρα [22, 169, 126] σχετικά με τον εντοπισμό χεριών, χειρομορφής και αναγνώρισης νοηματικής γλώσσας και εστιάζοντας στην μέθοδο της ενίσχυσης, την ελάχιστη εκπαίδευση και τη γενίκευση σε λεξιλόγια μεγάλης κλίμακας. Στο [22] εισάγουν μια πρωτότυπη αρχιτεκτονική δύο βημάτων όπου ένα αρχικό στάδιο κατηγοριοποίησης εξάγει μια περιγραφή υψηλού επιπέδου της χειρομορφής και της κίνησης και ακολουθείται από έναν συνδυασμό

ταξινομητών Μαρκοβιανών αλυσίδων και ανάλυσης ανεξάρτητων συνιστωσών (Independent Component Analysis - ICA). Το πλεονέκτημα αυτής της προσέγγισης είναι η ελάχιστη εκπαίδευση που απαιτείται για την μοντελοποίηση των νοημάτων. Η περιγραφή υψηλού επιπέδου, βασισμένη στην νοηματική γλωσσολογία, περιγράφει τις ενέργειες σε εννοιολογικό επίπεδο και αποτελείται από 4 σύνολα χαρακτηριστικών γνωρισμάτων, 5 ha 13 tab 10 sig και 6 dez. Βέλτιστη αναπαράσταση γνώρισματος σε σύμβολο επιτυγχάνεται μέσω ICA, η οποία μετασχηματίζει δυαδικά ψηφία σε χώρο όπου μπορεί να χρησιμοποιηθεί η Ευκλείδια μετρική απόστασης ενώ μετέπειτα η μοντελοποίηση της χρονικής μετάβασης υλοποιείται με αλυσίδες Markov. Ο αριθμός των καταστάσεων και οι πιθανές μεταβάσεις ($2^{28} \times 6$ και 2.6×10^{17} αντίστοιχα) μειώνονται με την πρόσθεση καταστάσεων όταν αυτές εμφανίζονται κατά την εκπαίδευση. Η απόδοση της τεχνικής δύο βημάτων ποικίλλει από 73% εως 84% για ένα λεξικλεξιλόγιο 43 λέξεων, όπου μόνο ένα στιγμιότυπο νοήματος επιλέχτηκε για την εκπαίδευση. Το ποσοστό κατηγοριοποίησης αυξάνεται εντυπωσιακά (97.67%) όταν εξετάζεται σε επιλεγμένο υποσύνολο του λεξιλογίου. Ενώ στο [126], χρησιμοποιείται δενδρική δομή σειράς μεθόδων ενίσχυσης (boosted cascades) για την ανίχνευση χεριών και την αναγνώριση χειρομορφής, που εισήχθησαν στο [169], αναφέροντας το εντυπωσιακό ποσοστό επιτυχίας 99.8% στην ανίχνευση χεριών και 97.4% στην κατηγοριοποίηση. Το ενδιαφέρον σημείο σε αυτήν την προσέγγιση είναι οι ελάχιστες απαιτήσεις εκπαίδευσης, χρησιμοποιώντας μόνο ένα στιγμιότυπο εκπαίδευσης ανά κατηγορία.

4.1.2.3.5 Λοιπές προσεγγίσεις Οι Assaleh και Rousan στο [7] παρουσιάζουν μια αρχιτεκτονική βασισμένη σε πολυωνυμικούς ταξινομητές, που δεν απαιτούν επαναληπτική εκπαίδευση, είναι ιδιαίτερα ευέλικτοι υπολογιστικά και όταν συγκριθούν με προηγούμενη εργασία τους (ANFIS) για το το ίδιο σύνολο δεδομένων παρουσιάζουν σημαντικά βελτιωμένη απόδοση. Το πειραματικό σύνολο τους είναι 42 νοήματα (30 σύμβολα) του Αραβικού νοηματικού γλωσσικού αλφάβητου που εκτελείται από τριάντα εκ γενετής κωφούς συμμετέχοντες. Χρησιμοποιούν τα χρωματιστά γάντια και εξάγουν τριάντα χαρακτηριστικά γνωρίσματα. Εκπαιδεύοντας έναν πολυωνυμικό κατηγοριοποιητή δεύτερης τάξης ανά κατηγορία, δημιουργούνται 42 δίκτυα. 2323 δείγματα διαιρούνται σε σύνολο εκπαίδευσης και δοκιμής που περιέχουν 1625 και 698 δείγματα αντίστοιχα. Το αρχικό ποσοστό αναγνώρισης 84,5% που επετεύχθη χρησιμοποιώντας το ANFIS βελτιώθηκε σε 93,4%. Σε μια πιό πρόσφατη δημοσίευση [214] οι Shanableh κ.α. χρησιμοποιούν τον αλγόριθμο του K κοντινότερου γείτονα (K nearest neighbor - KNN) και τον Μπεϋζιανό ταξινομητή για να αναγνωρίσουν μεμονωμένα νοήματα. Σε σύνολο δεδομένων με τρεις νοηματιστές, 50 επαναλήψεις και 23 νοήματα επιτυγχάνει 93%.

Οι Cui και Weng στο [52] συνδυάζουν κίνηση και χωρική πληροφορία με ένα σχήμα πρόβλεψης και επαλήθευσης για τον εντοπισμό των χεριών σε σύνθετο παρασκήνιο, που περιγράφεται λεπτομερέστερα στο [51] και κατασκευάζοντας εικόνες και διανύσματα προσοχής (attention). Τελικά ένα δέντρο επαναληπτικής διχοτόμησης (recursive partition tree) είναι υπεύθυνο για την κατηγοριοποίηση και όταν συγκρίνεται με την ταξινόμηση κοντινότερου γείτονα 28 διαφορετικά νοήματα το ξεπερνά σε απόδοση επιτυγχάνοντας ποσοστό αναγνώρισης 93,2%. Ο Dalle στο [53] παρουσιάζει ένα υπολογιστικό πρότυπο υψηλού επιπέδου της νοηματικής γλώσσας που χρησιμοποιεί το διάστημα νοηματισμού ως αναπαράσταση τόσο της έννοιας όσο και της χωρικής δομής της πρότασης. Εισάγει σημαντική ποσότητα γνώσης σχετικά με το γλωσσικό

συντακτικό και την γραμματική της νοηματικής που μπορεί να χρησιμοποιηθεί για πρόβλεψη και έλεγχο συνέπειας της ερμηνείας της πρότασης αλλά και επαλήθευση των υπόλοιπων συστατικών της αρχιτεκτονικής π.χ. επεξεργασία εικόνας.

Ο Derpanis κ.α. στο [56] εισάγουν μια προσέγγιση αποσύνθεσης, διατηρώντας μια ισορροπία ανάμεσα στην γλωσσολογική θεωρία και τις μεθόδους όρασης υπολογιστών, για την αναγνώριση των χειρονομιών από μονοφθαλμική χρονική ακολουθία εικόνων. Η αποσύνθεση έγκειται σε τρία σύνολα: χειρομορφή, θέση και μετακίνηση χεριών. Για τη διαδικασία αναγνώρισης υπολογίζεται η Ευκλείδια απόσταση μεταξύ των χειρονομιών εισόδου και των αντίστοιχων αποθηκευμένων προτύπων και σταθμίζονται με γνώμονα την αμοιβαία σταθερά των αντίστοιχων διανυσμάτων χαρακτηριστικών γνωρισμάτων τους. Από την άποψη των πειραμάτων οι ερευνητές κατέγραψαν 592 ακολουθίες νοημάτων από 15 νοηματιστές και επιτυγχάνουν ένα ποσοστό 86.00% για την πλήρως αυτοματοποιημένη επεξεργασία και 97,13% για αρχικοποιημένη χειρωνακτικά επεξεργασία. Οι Fillbrandt κ.α. [89] ενδιαφέρονται περισσότερο για τη διαδικασία εξαγωγής χαρακτηριστικών γνωρισμάτων και την παροχή πληροφοριών για την ρύθμιση των δαχτύλων και της τρισδιάστατης στάσης χεριών, μοντελοποιώντας το χέρι ως ένα σύνολο διδιάστατων προτύπων εμφάνισης, διασυνδεδεμένων μέσω μεταβάσεων. Οι Fujimura και Liu στο [94] εκμεταλλεύονται την πληροφορία βάθους που παρέχεται από μια κάμερα χρόνου πτήσης (time-of-flight camera) για να ανιχνεύσουν τον νοηματιστή, το χέρι και την παλάμη του του και επίσης την χειρομορφή, τη μετακίνηση, τον προσανατολισμό και την ταχύτητα. Εφαρμόζουν Karhunen-Loeve αποσύνθεση στα τρισδιάστατα χαρακτηριστικά του άνω σώματος του χρήστη για να διαμορφώσουν πρότυπα βάθους. Ο κατηγοριοποιητής βασίζεται σε πίνακα επιτυγχάνοντας απόδοση σχεδόν πραγματικού χρόνου (1–4 δευτερόλεπτα ανά νόημα). Αναλυτικές πληροφορίες για τα ποσοστά αναγνώρισης δεν είναι διαθέσιμες αν και υποστηρίζουν ότι αναγνωρίζουν πάνω από την 100 νοήματα Ιαπωνικής νοηματικής γλώσσας.

Οι Hernandez κ.α. στο [107] χρησιμοποιούν γραμμικό ταξινομητή για να προσδιορίσουν μοναδικές ακολουθίες φωνημάτων οι οποίες εξάγονται από τις κατηγορίες νοημάτων. Χρησιμοποιείται το γάντι AcceleGlove για να καταγραφούν 30 νοήματα, με 42 στάσεις σώματος, 6 προσανατολισμούς, 11 θέσεις και 7 μετακινήσεις και αναφέρουν ακρίβεια αναγνώρισης 98% για το ίδιο σύνολο εκπαίδευσης και επαλήθευσης ενώ 95% για διαφορετικά.

Οι Jiang κ.α. στο [123] εισάγουν μια πολυεπίπεδη αρχιτεκτονική αναγνώρισης συνεχούς νοηματισμού, σύμφωνα με την οποία η ακολουθία νοημάτων τοποθετείται αρχικά σε ένα σύνολο λέξεων που είναι εύκολο να προκαλέσουν σύγχυση (σύνολο σύγχησης) μέσω μιας καθολικής γρήγορης αναζήτησης και αναγνωρίζεται μέσω μιας τελευταίας τοπικής αναζήτησης. Ο καθορισμός των συνόλων σύγχυσης πραγματοποιείται από τον αλγόριθμο DTW/ISODATA. Τα πειραματικά αποτελέσματά τους, για 29652 δείγματα σε 4942 νοήματα από τρεις νοηματιστές που επαναλαμβάνει την εκτέλεση δύο φορές, αναφέρουν ποσοστά αναγνώρισης 87,39% έναντι 82,73% ενός τυποποιημένου HMM, ενώ ο χρόνος αναγνώρισης μειώνεται σε 0,137 δευτερόλεπτα έναντι 2.364 δευτερόλεπτα αντίστοιχα.

Οι Sagawa και Takeuchi στο [201] διερευνούν την επιρροή επείσακτων χειρονομιών και του φαινομένου της επένθεσης και προτείνουν μια μέθοδο χρονικής κατάτμησης της Ιαπωνικής νοηματικής γλώσσας σε νοήματα και μεταβάσεις. Η είσοδος βασίζεται σε γάντι δεδομένων (CyberGlove) και οι χρονικές στιγμές όπου η ταχύτητα των χεριών είναι ελάχιστη ή η αλλαγή κατεύθυνσης είναι σημαντική προσδιορίζονται ως υποψήφια χρονικά όρια. Στα πειράματά τους 100 δείγματα προτάσεων περιελάμβαναν

575 νοήματα και 571 μεταβάσεις 461 (80,2%) νοήματα αναγνωρίστηκαν σωστά και 64 (11,2%) μεταβάσεις εκτιμήθηκαν εσφαλμένα ως νοήματα.

Οι Wang και Gao στο [242], εξετάζουν την ταχύτητα εκτέλεσης συστήματος αναγνώρισης μεμονωμένων νοημάτων και παρουσιάζουν μια τεχνική ημισυνεχών δυναμικών Γκαουσσιανών μιγμάτων (Semi-Continuous Dynamic Gaussian Mixture Model - SCDGMM) αναγνώρισης που ενισχύεται με ένα δενδρικό σχήμα αναζήτησης βασισμένο στην εντροπία καταφέρνοντας εντυπωσιακή μείωση του χρόνου αναγνώρισης κατά έναν λόγο 15, διατηρώντας παράλληλα αποδεκτό ποσοστό αναγνώρισης. Με την χρήση δύο CyberGloves έγινε η καταγραφή ενός λεξιλογίου 274 νοημάτων, 10 επαναλήψεις (8 για εκπαίδευση) επιτυγχάνοντας ένα ποσοστό αναγνώρισης 98,2%, αλλά ο μέσος χρόνος αναγνώρισης ανά νόημα ήταν 0,61 δευτερόλεπτα, χρόνος απαγορευτικός για εφαρμογή πραγματικού χρόνου ενός τέτοιου συστήματος. Η προσθήκη ενός A-δέντρου ο χρόνος αναγνώρισης ενός νοήματος μειώνεται σε 0,04 δευτερόλεπτα ανά νόημα διατηρώντας το ποσοστό αναγνώρισης σε 97,4%. Οι Zhang κ.α. στο [267] παρουσιάζουν μια μέθοδος εξαγωγής χαρακτηριστικών γνωρισμάτων πολλαπλής ανάλυσης μειώνοντας περαιτέρω την διάστασης των διανυσμάτων εισόδου με το κριτήριο της μέγιστης διαφοράς (Maximum Variance Criterion - MVC) και τελικά μειώνοντας τον χρόνο αναγνώρισης κατά 0,992 δευτερόλεπτα (2.364s–1.372s) κατά μέσο όρο αυξάνοντας παράλληλα την ακρίβεια αναγνώρισης κατά 3,42% (88.73%–92.15%), έναντι ενός συμβατικού HMM συστήματος.

4.1.2.4 Σύνοψη

Συνοπτικά η αναγνώριση νοηματικής γλώσσας δεν πρέπει να αντιμετωπίζεται ως απλά ένα υποσύνολο της αναγνώρισης χειρονομιών ή ως αναγνώριση χωροχρονικών προτύπων. Οι γλωσσολογικές πτυχές της νοηματικής γλώσσας δεν πρέπει να αγνοηθούν, ειδικά κατά την αναγνώριση συνεχούς νοηματισμού όπου τα γραμματικά φαινόμενα είναι πλουσιότερα και οι ροές πληροφορίας δεν περιορίζονται σε χειρωνακτικά χαρακτηριστικά γνωρίσματα. Είναι επιτακτική ανάγκη να διερευνηθεί σε βάθος πώς η δομή της νοηματικής γλώσσας, το συντακτικό και τα γραμματικά φαινόμενα θα μπορούσαν να ενσωματωθούν στη ευρύτερη αρχιτεκτονική αναγνώρισης δεδομένου ότι θα ήταν εξαιρετικά ευεργετικό στην ολοκλήρωση του ερευνητικού πεδίου αλλά θα προσέφερε αρκετά και στην επαλήθευση της γλωσσολογικής ανάλυσης. Τα μη χειρωνακτικά χαρακτηριστικά γνωρίσματα πρέπει επίσης να ληφθούν υπόψη και στη διαδικασία εξαγωγής χαρακτηριστικών γνωρισμάτων.

Το χαρακτηριστικό της εφαρμοσιμότητας κάθε αρχιτεκτονικής αποτελεί ένα από τα βασικότερα κριτήρια κατά την αξιολόγηση ενός συστήματος καθώς ενώ χρήσιμα συμπεράσματα μπορούν να προκύψουν με την πραγματοποίηση θεωρητικών ερευνητικών εργασιών που επαληθεύονται με πειράματα περιορισμένης κλίμακας και με αρκετές υποθέσεις και συμβιβασμούς η εφαρμογή τους σε πραγματικές καταστάσεις πρέπει να είναι πάντα ένας από τους στόχους κάθε ερευνητικής προσπάθειας. Οι προκλήσεις που πρέπει να αντιμετωπιστούν όσον αφορά την εφαρμοστικότητα των αρχιτεκτονικών είναι τα λεξιλόγια μεγάλης κλίμακας, η ανεξαρτησία από τον νοηματιστή και το υπολογιστικό κόστος. Προτείνοντας ένα σχήμα που απαιτεί εξαντλητική εκπαίδευση, αποτυγχάνει να γενικεύσει ή απαιτεί μη ρεαλιστικό χρόνο επεξεργασίας φαίνεται να είναι άσκοπο, εκτός εάν πρόκειται για μια εξαιρετικά πρωτότυπη ιδέα και εξετάζεται από την πλευρά της επικύρωση της γενικής ιδέας. Τα ίδια ισχύουν και για τη διαδικασία εξαγωγής χαρακτηριστικών γνωρισμάτων αλλά και τους αντίστοιχους περιορισμούς σχετικά με το περιβάλλον και τον χρήστη. Εύκολα μπορούν να γίνουν υποθέσεις

ή συμβιβασμοί που θα μπορούσαν να οδηγήσουν στην κατάρρευση ενός αλγορίθμου όταν αυτός εφαρμόζεται σε πραγματικές συνθήκες. Συμπερασματικά, θα μπορούσαμε να αναφέρουμε πως αν και ο εν λόγω ερευνητικός τομέας είναι αρκετά ενεργός, αφθονία δημοσιεύσεων περιορίζεται σε μια απλοϊκή προσέγγιση προσπαθώντας να επιλύσει όλα τα σχετικά υποπροβλήματα καταλήγοντας να αντιμετωπίζει επιφανειακά τις επιμέρους πτυχές του προβλήματος. Οι ερευνητές σπάνια μελετούν εξαντλητικά την ουσία των σχετικών προβλημάτων και απαιτείται μια ερευνητική κατεύθυνση προκειμένου να ωριμάσει συνολικά η ερευνητική περιοχή.

Εργασία	Γλώσσα	Λεξιλόγιο	Είσοδος	Μέθοδο κατηγοριοποίησης	Γνωρίσματα	Ποσοστό αναγνώρισης
[7]	Arabic	42	video	polynomial	fingertip positions and orientations	93.41
[8]	Netherlands	262	video	HMM	size. COG, angles	94/73 I 56,2/47,6 C
[14]	German	52/97	video	HMM	size. COG, angles	2.2–94
[15]	German	52/97	video	HMM	size. COG, angles	91.7
[16]	German	12	video	HMM+k-Means	size. COG, angles	80.8
[22]	British	43	video	2 stage (high level+MM/ICA)	5 HA, 13 TAB, 10 SIG and 6 DEZ	73–84
[24]	American	71	vision + accelometer	HMM	hand coords, area features	90.48
[45]	British	164	video	AdaBoost viseme + MM	Ha Tab Sig	74.3
[44]	British	164	video	AdaBoost and AdaPlusBoost	volumetric information	
[47]	Spanish	33	video	HMM	Moving Block Distance	99.50
[52]	American	28	video	Recursive Partition Tree	appearance-based	93.2
[53]	French		video	3 subsystems		
[56]	American	592	video	model - instance distance	shape, location and movement	86(auto) 97,13 (manual)
[83]	Chinese	208	gloves	SOFM/HMM	48 dimensional shape, position and orientation	90,1 (unreg) 96,6 (reg)
[84]	Chinese	5113	gloves	fuzzy decision tree on GMM, FSM, SOFM/HMM	48 dimensional shape, position and orientation	83,7(unreg) 91,6 (reg)
[85]	Chinese	5113	gloves	k-means + DTW	48 dimensional shape, position and orientation	90,5
[94]	Japanese	100	video + tof camera	rules	location, hand shape, movement	

[97]	Chinese	1065/80(220)	gloves	HMM	Position Orientation Shape	80–93.2 I 82–96,3 C
[96]	Chinese	5113/750	gloves	HMM/DTW	48 dimensional shape, position and orientation	
[95]	Chinese	5113	gloves	SOFM/SRN/HMM	48 dimensional shape, position and orientation	86,3
[100]	German	262	video	HMM	location orientation handshape	94
[107]	American	30	gloves	linear classifier	shape, orientation, direction, trajectory	98 (95 without retraining)
[108]	German	52	video	HMM	size. COGs, angles	95%
[111]	Taiwan	15	video	3D Hopfield	fourier descriptors	91
[115]	Japanese		video	3-D Look Up Table	position (COGs)	
[116]	Japanese		video	3-D Look Up Table	shape and orientation	
[117]	Italian	40	video	Hierarchical SOM	position	83,3
[123]	Chinese	4942	gloves	ISODATA confusion set + DTW	position and orientation	87.39
[126]	British	164	video	like bowden 2 stage	40 dimensional HA/TAB/SIG/DEZ	89,1–92
[144]	Taiwan	20	motion analysis	back-propagation NN	15 dimensional fingertips distances	91,9–96,58
[148]	Taiwan	250/196	glove	HMM	posture, position, orientation, and motion	80,4 C 94,8 I
[153]	Chinese	5177	video	parallel multistream model	handshape, position orientation	93,4–95,2
[169]	British		video	tree structure of boosted cascades	shape	97,4
[185]	Greek	24	image	HMM	37 dimensional fingertip vectors	80,56 – 97,22
[201]	Japanese	200	glove	K nearest neighbor (KNN) and Bayesian	position	80,2
[214]	Arabic	23	video	KNN + Bayessian	position, movement	93–96

[216]	American	40	video	HMM	16 dimensional, moments, position, direction, area, eigenvectors	92–98
[219]	Taiwan	34	gloves	hyperrectangular NN + fuzzy	handshpape	91,2–94,1
[220]	Japanese	65	video	HMM	flatness, area, COGs, direction, protrusions	91,4–98,3
[230]	Australian	13	glove	4 back-propagation NN	handshape, orientation, motion	94 (reg) 85(unreg)
[235]	American	22/500	motion capture	Parallel HMM	3d position, eigenvalues	94,23 I 84,85 C
[236]	American	22/500	motion capture	HMM	motion, handshape	94,55
[242]	Chinese	274	gloves	Semi-Continuous Dynamic Gaussian Mixture Model	handshape	97,4–98,2
[240]	Chinese	5119 signs 200 sentence	gloves	clustered Gaussian HMM	handshape	99,7 I 92,8 C
[241]	American	26 letters 36 hand-shapes	gloves	multidimensional HMM	clustered position and handshape	95
[243]	Chinese	64	video / homography	nearest neighbor	homography	71,8–92,1
[249]		50	video	Parametric HMM	3d hand position	
[255]	American	40	video	time-delay NN	trajectories	93,42
[260]	American	10	video	HMM	image derivatives	93
[268]	Chinese	439	video	Tied-Mixture Density Hidden Markov Models	distances, area, coords	92,5
[267]	Chinese	4942	gloves	Maximum Variance Criterion	coords	93,15
[270]	British	232	video	HMM	Coordinates, derivatives, area	99,3 (reg) 44,1 (unreg)

Πίνακας 4.9: Συνοπτικά όλες οι προσεγγίσεις αυτόματης αναγνώρισης νοηματικής γλώσσας

4.2 Προτεινόμενη Αρχιτεκτονική Αναγνώρισης Χειρονομιών

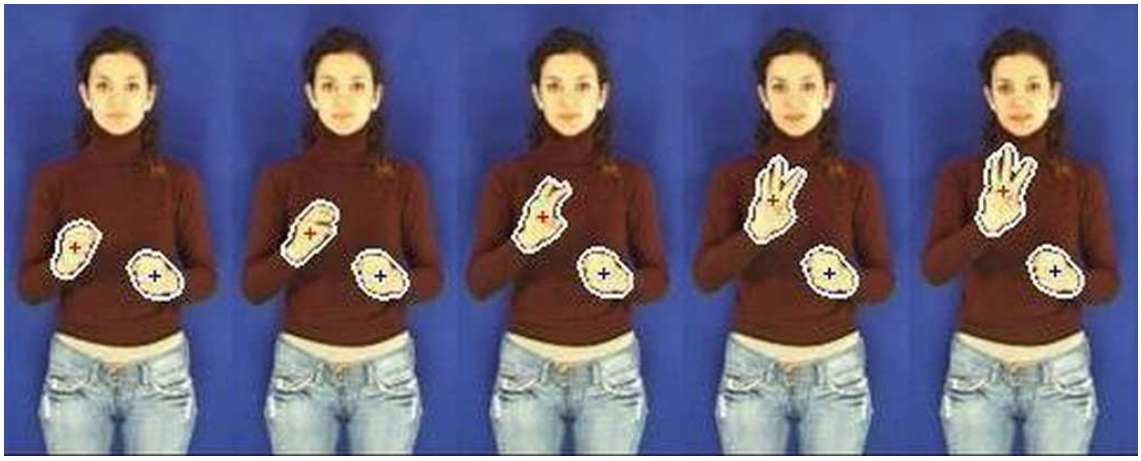
Η προτεινόμενη εργασία εισάγει ένα πιθανοτικό σχήμα αναγνώρισης για χειρονομίες και λήμματα νοηματικής γλώσσας. Αυτοοργανούμενοι χάρτες μοντελοποιούν την χωρική πληροφορία που εξάγεται μέσω της επεξεργασίας εικόνας ενώ την χρονική πτυχή αναλαμβάνουν να προτυποποιήσουν Μαρκοβιανά και κρυφά Μαρκοβιανά μοντέλα για την κίνηση και την χειρομορφή αντίστοιχα. Πολλαπλά πρότυπα δημιουργούνται για κάθε κατηγορία χειρονομιών και, συνδυαζόμενα κατάλληλα, παράγουν έναν επικυρωμένο μηχανισμό κατηγοριοποίησης που αποδίδει με συνέπεια. Το σημείο εστίασης της τρέχουσας εργασίας είναι να αντιμετωπιστεί η παρέκκλιση εκτέλεσης χειρονομίας τόσο από τον ίδιο χρήστη όσο και από διαφορετικούς χρήστες με την προσθήκη ευέλικτης και προσαρμοστικής διαδικασίας αποκωδικοποίησης επιτρέποντας στον αλγόριθμο να αναζητήσει την βέλτιστη διαδρομή ενώ το υπολογιστικό κόστος της διαδικασίας αναγνώρισης υποδεικνύει την προτεινόμενη αρχιτεκτονική ως κατάλληλη για εφαρμογή σε λεξιλόγια μεγάλης κλίμακας σε πραγματικό χρόνο.

4.2.1 Εξαγωγή Χαρακτηριστικών Γνωρισμάτων

Η ευρύτερη αρχιτεκτονική αποτελείται από δύο κύριες ενότητες, όπως και η πλειοψηφία των προσεγγίσεων που παρουσιάζονται στην σχετική ενότητα 4.1.2, συγκεκριμένα την εξαγωγή χαρακτηριστικών γνωρισμάτων και την κατηγοριοποίηση. Η προτεινόμενη διαδικασία εξαγωγής χαρακτηριστικών γνωρισμάτων της προσέγγισης βασίζεται σε μονοκάμερη οπτική είσοδο στην οποία εφαρμόζεται η μέθοδος των γεωδαιτικών ενεργών περιοχών (Geodesic Active Regions), ενισχυμένα με δυνάμεις δερματικού χρώματος και κίνησης, που εξελίσσονται ελαχιστοποιώντας μια συνάρτηση λάθους, ώστε να εντοπίσουν περιοχές χεριών. Χαρακτηριστικά γνωρίσματα σχετικά με τη θέση, την περιοχή και την χειρομορφή εξάγονται και χρησιμοποιούνται ως είσοδο για τους ταξινομητές που συζητούνται στο 4.2.2.

Για την δημιουργία του διανύσματος χαρακτηριστικών γνωρισμάτων απαιτείται επεξεργασία της εικόνας που αποτελεί πλαίσιο του βίντεο καταγραφής των εκτελέσεων των νοημάτων. Κατά την διαδικασία αυτή κάθε πλαίσιο υπόκειται κατάτμηση ώστε να να απομονωθούν οι περιοχές των χεριών του χρήστη, οι οποίες θα επεξεργασθούν περαιτέρω για την εξαγωγή γνωρισμάτων που στοχεύουν στον χαρακτηρισμό της θέσης και της χειρομορφής. Για σκόπους κατάτμησης χρησιμοποιούνται Γεωδαιτικές Ενεργές Περιοχές (Geodesic Active Regions) [183] που βασίζονται στα Γεωδαιτικά Ενεργά Περιγράμματα (Geodesic Active Contours - GAC) [31]. Τα GAC είναι παραμορφώσιμα διδιάστατα περιγράμματα, τα οποία εξελίσσονται με στόχο σε κάθε βήμα να ελαχιστοποιήσουν μια κατάλληλη συνάρτηση ενέργειας, ώστε να ικανοποιήσουν συγκεκριμένες ανάγκες κατάτμησης. Η διαδικασία GAC έχει ενισχυθεί περαιτέρω ώστε να ενσωματώσει χρωματικές πληροφορίες δέρματος και κίνησης και η συνολική διαδικασία, που περιγράφεται λεπτομερώς στο [62] και φαίνεται εποπτικά στην εικόνα 4.4, διαμορφώνει μια εύρωστη και αξιόπιστη ανίχνευση και παρακολούθηση χεριών.

Το σύνολο χαρακτηριστικών γνωρισμάτων που εξάγεται από την ανωτέρω διαδικασία περιλαμβάνει συντεταγμένες χεριών (τροχιές χεριών), περιγραφείς χειρομορφής και περιοχής. Προκειμένου να χαρακτηριστεί η χειρομορφή χρησιμοποιούνται περιγραφείς περιγράμματος (περιγραφείς Fourier, γνωρίσματα καμπυλότητας) και περιγραφείς περιοχής (ροπές (moments), έκταση, εκκεντρότητα, πυκνότητα, λόγος αξόνων, προ-



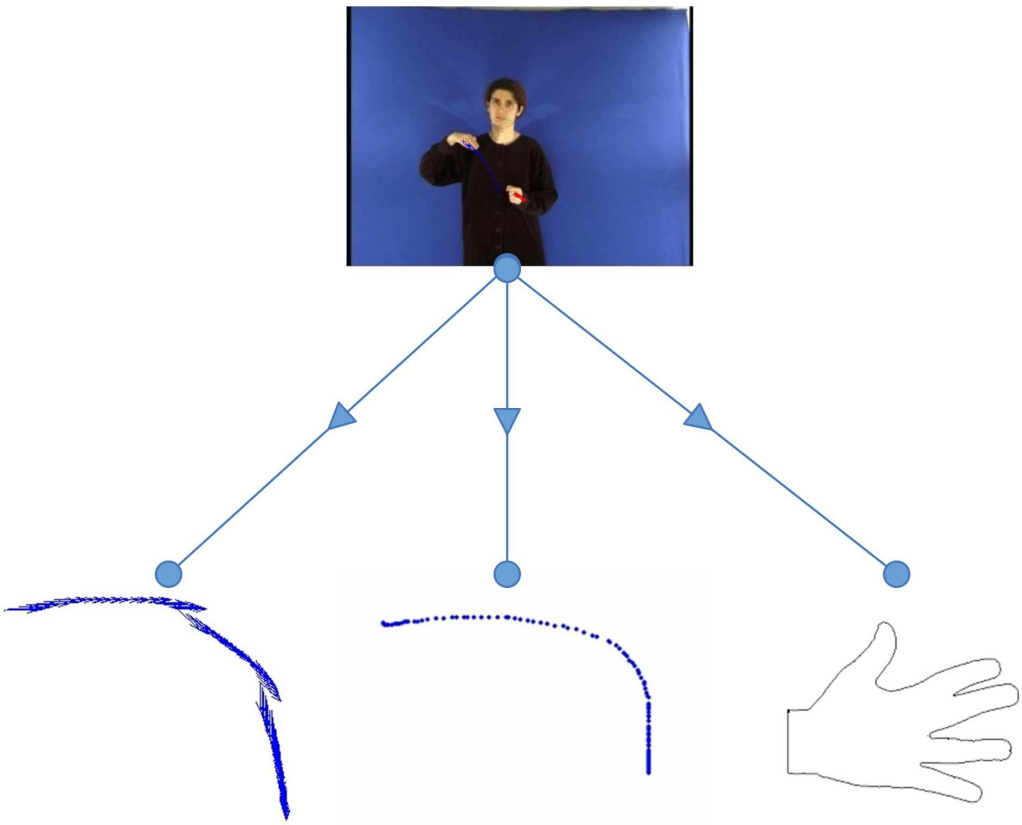
Σχήμα 4.4: Αποτελέσματα κατάτμησης εικόνας και παρακολούθησης χεριών [62]

σανατολισμός).

4.2.2 Προτυποποίηση Χειρονομιών

Η προτεινόμενη αρχιτεκτονική μοντελοποίησης χειρονομιών είναι βασισμένη στον μετασχηματισμό ενός στιγμιοτύπου σε συμβολική μορφή η οποία ακολούθως χρησιμοποιείται για να εκπαιδεύσει αντίστοιχα πιθανοτικά πρότυπα. Ο πρώτος μετασχηματισμός είναι βασισμένος στην σχετική θέση των χεριών, με σημείο αναφοράς το κεφάλι, κατά τη διάρκεια της χειρονομίας και επιτυγχάνεται με την αναπαράσταση μέσω ενός αυτοοργανούμενου χάρτη (Self Organizing Maps - SOM) [186]. Παρά γεγονός ότι οι κόμβοι χαρτών αντιμετωπίζονται ως σύμβολα, η συνάρτηση γειτνίασης του χάρτη παρέχει μια μετρική απόστασης μεταξύ τους, η οποία χρησιμοποιείται κατά τη διάρκεια της κατηγοριοποίησης ενός νέου νοήματος. Επιπλέον, επιτρέπει τη χρήση της απόστασης Levenshtein για τη σύγκριση ακολουθιών συμβόλων και τον καθορισμός μιας 'μέσης' ακολουθίας, του γενικευμένου μέσου συνόλου. Ένας πρόσθετος μετασχηματισμός βασίζεται στην κίνηση των χεριών κατά την εκτέλεση των χειρονομιών στοχεύει στην περιγραφή των αλλαγών κατεύθυνσης των χεριών κατά τη διάρκεια του νοηματισμού. Τα σύμβολα αυτού του μετασχηματισμού αποτελούν το σύνολο των διακριτών γωνιών της τροχιάς του χεριού. Αυτό το σύνολο περιλαμβάνει τις κβαντισμένες τιμές των γωνιών προκειμένου να χρησιμοποιηθούν ως καταστάσεις ενός πρόσθετου συνόλου Μαρκοβιανών προτύπων πρώτης τάξης.

Έστω c ο αριθμός των κατηγοριών που περιλαμβάνονται στο σύνολο D. Έτσι, το σύνολο δεδομένων $D = \{D_1, D_2, \ldots, D_c\}$ θα αποτελείται από c υποσύνολα και κάθε τέτοιο υποσύνολο D_j περιέχει n_j στιγμιότυπα χειρονομιών $D_j = \{G_{1_j}, G_{2_j}, \ldots, G_{n_j}\}$, όπου n_j ο αριθμός επαναλήψεων για την κατηγορία j. Κάθε στιγμιότυπο G_{j_i} περιέχει l_{j_i} συντεταγμένες ώστε $G_{j_i} = \{(x_{1_{j_i}}, y_{1_{j_i}}), (x_{2_{j_i}}, y_{2_{j_i}}), \ldots, (x_{l_{j_i}}, y_{l_{j_i}})\}$, όπου l_{j_i} ο αριθμός των συντεταγμένων που ανήκουν στην τροχιά για την επανάληψη i της κατηγορίας j. x, y είναι συντεταγμένες του χεριού σχετικές με την θέση του κεφαλιού στο συγκεκριμένο πλαίσιο. Η θέση των χεριών είναι σχετική δεδομένου ότι η θέση του χρήστη κατά τη διάρκεια της καταγραφής δεν είναι εκ των προτέρων γνωστή και περαιτέρω κανονικοποίηση εφαρμόζεται για κάθε χρήστη δεδομένου των ανατομικών χαρακτηριστικών του χρήστη (ύψος, μήκος χεριών) που θα μπορούσαν να δημιουργήσουν πρόβλημα κατά την δημιουργία των προτύπων όπως θα φανεί στο ακόλουθο κεφάλαιο 4.2.2.1. Αυτή η διαδικασία εξασφαλίζει ότι οι πληροφορίες θέσης παραμέ-



Σχήμα 4.5: Διαισθητική εξαγωγή χαρακτηριστικών

νουν αναλλοίωτες τόσο σε σχέση με την θέση του χρήστη μπροστά από την κάμερα όσο και σε σχέση με χαρακτηριστικά όχι πάντα κοινά για όλους τους χρήστες.

4.2.2.1 Πρότυπο θέσης

Αν και οι αυτοοργανούμενοι χάρτες χρησιμοποιούνται ως τεχνική απεικόνισης δεδομένων ή μείωσης διάστασης επιλέγουμε να χρησιμοποιήσουμε τον SOM ως εργαλείο συσταδοποίησης ώστε να παραχθεί μια πιο αφηρημένη αναπαράσταση του χώρου νοηματισμού και να επιτρέψουμε στα ίδια τα δεδομένα να διαμορφώσουν τις σχέσεις γειτνίασης μεταξύ των κόμβων του χάρτη. Μια απλοϊκότερη προσέγγιση θα ήταν να κβαντιστεί σαφώς (crisp) ο νοηματικός χώρος σε περιοχές και να ανατεθούν αυθαίρετα σχέσεις γειτνίασης μεταξύ των περιοχών. Αυτή η προσέγγιση, ενώ δοκιμάστηκε, αποδείχθηκε πως δεν είναι ικανή να γενικεύσει ικανοποιητικά, λόγω του απόλυτου διαχωρισμού των περιοχών ενώ η ιδιότητα των SOM, του ανταγωνισμού των δειγμάτων για αντιπροσώπευση, δείχνει περισσότερο προσαρμοστική προσέγγιση. Τα βάρη των νευρώνων επιτρέπεται να αλλάξουν με την εκπαίδευση σε μια προσπάθεια να γίνουν περισσότερο όμοια με τα δείγματα και την ελπίδα να επικρατήσουν στην επόμενη σύγκριση έχοντας μικρότερη απόσταση, σύμφωνα με κάποια μετρική και είναι αυτή η διαδικασία επιλογής και εκμάθησης που επηρεάζει τις τιμές των βαρών ώστε να οργανωθεί πιο αντιπροσωπευτικά ο χάρτης. Κατά τη διάρκεια της εκπαίδευσης οι τιμές των βαρών των γειτονικών κόμβων επηρεάζονται ώστε να αναπαριστούνται και οι σχέσεις γειτνίασης μέσα από τα κέντρα των κόμβων. Αυτό το χαρακτηριστικό γειτνίασης είναι κρίσιμο για την διαδικασία αποκωδικοποίησης όπως θα φανεί στην ενότητα 4.2.3.

Οι συντεταγμένες (x, y) όλων των σημείων από όλες τις επαναλήψεις του συνόλου

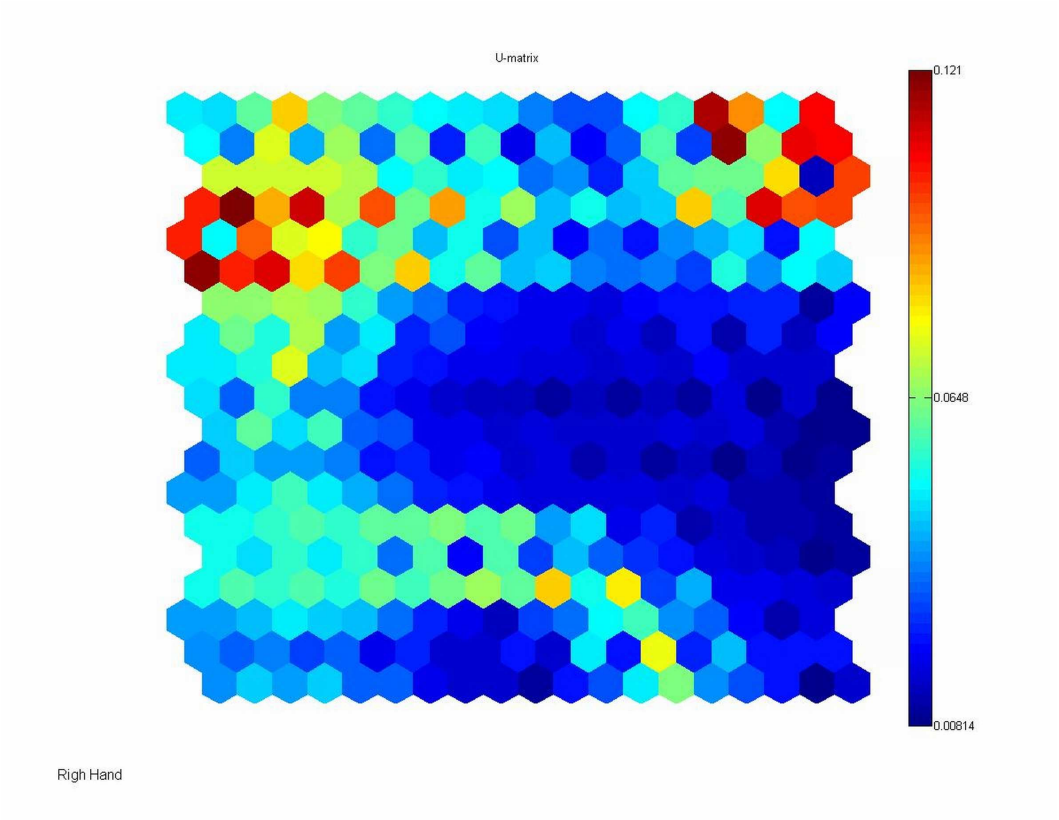
εκπαίδευσης όλων των κατηγοριών χρησιμοποιούνται για να εκπαιδεύσουν ένα εξαγωνικό, διδιάστατο SOM πλέγματος με τη διαδικασία της εκμάθησης δέσμης (batch). Οι συντεταγμένες προεπεξεργάζονται προκειμένου να κανονικοποιηθούν και να είναι αναλοίωτες της θέσης του χρήστη, όπως συζητήθηκε νωρίτερα. Η κανονικοποίηση για κάθε χρήστη εξασφαλίζει ότι οι χρήστες που τείνουν να χρησιμοποιούν μεγαλύτερο ή μικρότερο χώρο νοηματισμού ή είναι ψηλότεροι ή κοντύτεροι δεν εισάγουν θόρυβο ή προκαταλαμβάνουν κατά κάποιο τρόπο την διαδικασία εκπαίδευσης. Επιπλέον, οι θέσεις χεριών είναι σχετικές με τη θέση του κεφαλιού ώστε η πληροφορία της θέση να παραμένει αμετάβλητη της πραγματικής θέσης του χρήστη κατά την καταγραφή. Τα σημεία εισάγονται προς εκπαίδευση χωρίς διάταξη, ανεξάρτητα της κατηγορίας που ανήκουν αλλά και της διατεταγμένης θέσης τους κατά την εκτέλεση του νοήματος. Η εκπαίδευση του SOM εκτελείται για ολόκληρο το σύνολο δεδομένων D και όχι για κάθε κατηγορία. Αυτό το χαρακτηριστικό παρουσιάζει αρκετά πλεονεκτήματα καθώς μειώνει τον απαιτούμενο χρόνο εκπαίδευσης και ενισχύει την απλότητα της αρχιτεκτονικής. Επιπλέον και δεδομένου ότι ένας επαρκής αριθμός στιγμιοτύπων έχει χρησιμοποιηθεί κατά την εκπαίδευση, η εισαγωγή κάποιας νέας κατηγορίας στο λεξικλεξιλόγιο δεν απαιτεί εκ νέου εκπαίδευση καθώς μπορεί να υποτεθεί ότι ο χώρος νοηματισμού έχει αναπαρασταθεί επαρκώς. Αφετέρου, η χρήση ελάχιστων (1-2) δειγμάτων από κάθε κατηγορία δεν επηρεάζει σημαντικά την αναπαράσταση όλων των περιοχών του νοηματικού χώρου.

Η επιτυχής αναπαράσταση γίνεται εμφανέστερη στις εικόνες 4.6 και 4.7 που απεικονίζουν τα u-matrix του χάρτη του δεξιού και του αριστερού αντίστοιχα. Η ενοποιημένη μήτρα απόστασης (Unified distance matrix - umatrix) είναι ίσως η δημοφιλέστερη μέθοδος απεικόνισης SOM όπου στο πλέγμα των νευρώνων και απεικονίζει την απόσταση μεταξύ των παρακείμενων μονάδων του SOM. Εάν τα δείγματα είναι αρκετά διαφορετικά, η απόσταση μεταξύ των αντίστοιχων μονάδων χαρτών παρουσιάζεται με θερμά χρώματα, τα οποία αντιπροσωπεύουν μεγάλη μέση απόσταση μεταξύ των παρακείμενων μονάδων του SOM και αντίστροφα. Η ενοποιημένη μήτρα απόστασης για το αριστερό χέρι έχει υποστεί αντικατοπτρισμό δεδομένου ότι έτσι γίνεται πιο διαισθητική αναπαράσταση των θέσεων. Αξίζει να σημειωθεί πως ο χώρος νοηματισμού έχει απεικονιστεί ομοιόμορφα, ειδικά σε περιοχές όπου η εμφάνιση του χεριού είναι συχνή και συμπίπτει με τις περιοχές μεγάλης ομοιομορφίας (μπλε περιοχές).

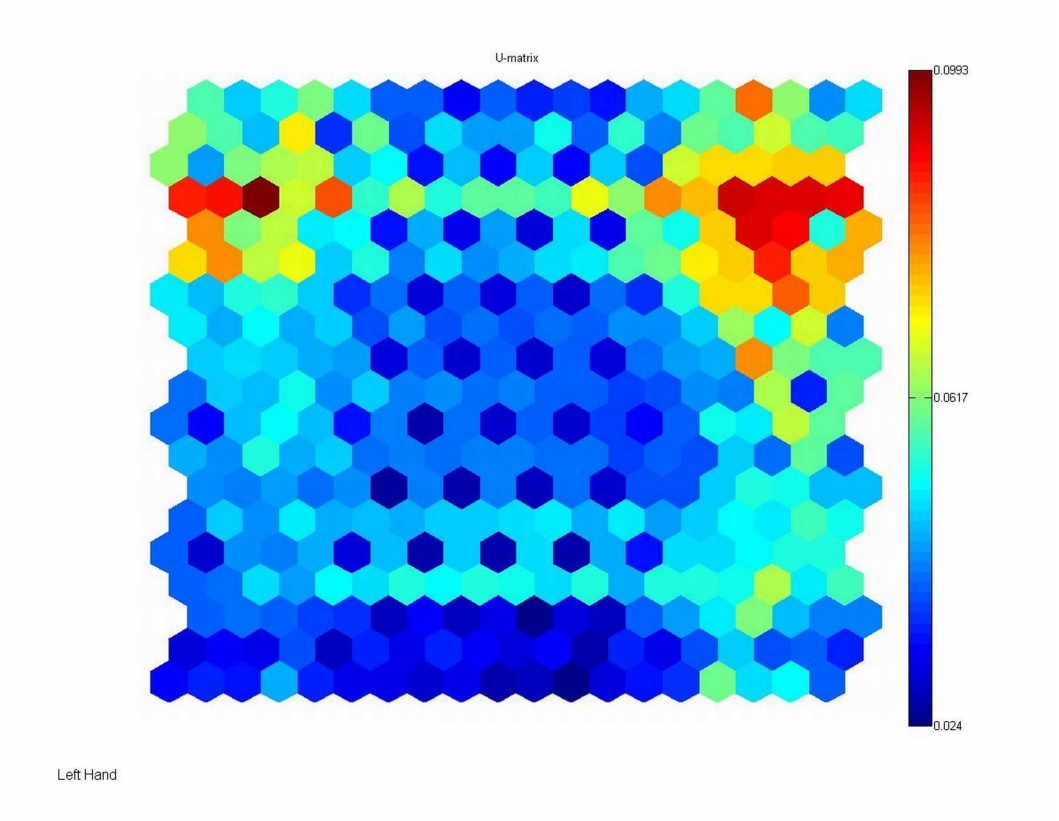
Μετά την εκπαίδευση, κάθε σημείο αντιστοιχείται στην μονάδα μέγιστης ταύτισης (Best Matching Unit - BMU) στο χάρτη, δηλ. την μονάδα του χάρτη που σύμφωνα με την Ευκλείδια απόσταση είναι εγγύτερα στο υπό εξέταση σημείο. Κατά συνέπεια, ένα στιγμιότυπο G_i μπορεί να μετασχηματιστεί από μια σειρά σημείων σε μια σειρά μονάδων του χάρτη σύμφωνα με τη συνάρτηση μετασχηματισμού T. Χάριν απλότητας θα αντικαταστήσουμε την παράσταση $G_{j_i} = \{(x_{1_{j_i}}, y_{1_{j_i}}), (x_{2_{j_i}}, y_{2_{j_i}}), \ldots, (x_{l_{j_i}}, y_{l_{j_i}})\}$ με $G = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$. Έτσι το στιγμιότυπο G θα αναπαρασταθεί στον χάρτη SOM:

$$\begin{aligned} T(G) &= (u_1, u_2, \ldots, u_l) \\ &: u_i = W_r(BMU(x_i, y_i)), i \in [1, l] \end{aligned} \tag{4.1}$$

Η συνάρτηση $BMU(x_i, y_i)$ επιστρέφει τον δείκτη της μονάδας BMU για το σημείο (x_i, y_i), ενώ η W_r αποτελεί ένα φίλτρο ενδιάμεσης τιμής που εφαρμόζεται στις τιμές για ένα σταθερό μήκος παραθύρου r γύρω από το i. Δεδομένου ότι u_i είναι ο δείκτης ενός κόμβου του χάρτη, αυτή η συνάρτηση ορίζεται ως $BMU : R^2 \rightarrow S$, όπου S είναι το σύνολο δεικτών όλων των μονάδων του χάρτη και μπορεί να αντιμετωπιστεί ως



Σχήμα 4.6: U-matrix για το δεξί χέρι



Σχήμα 4.7: U-matrix για το αριστερό χέρι

σύνολο συμβόλων. Σε πολλές περιπτώσεις, η τιμή u_i των προηγούμενων αλλά και των επόμενων σημείων κατά την εκτέλεση ενός νοήματος παραμένει ίδια, αν και η συνεχής μετακίνηση του χεριού αντιπροσωπεύεται από διακριτά σημεία, διαδοχικά σημεία είναι γενικά αρκετά κοντά στο διάστημα εισόδου. Η αντικατάσταση των διαδοχικών ίσων τιμών u_i με μια μοναδική τιμή οδηγεί, μετά την εφαρμογή του φίλτρου ενδιάμεσης τιμής, στον ακόλουθο ορισμό:

$$\begin{aligned} G^{'} = N(T(G)) = \{u_1^{'}, u_2^{'}, \ldots, u_m^{'}\} \\ : m \leq l, u_t^{'} \neq u_{t-1}^{'} \forall t \in [2, m] \end{aligned} \tag{4.2}$$

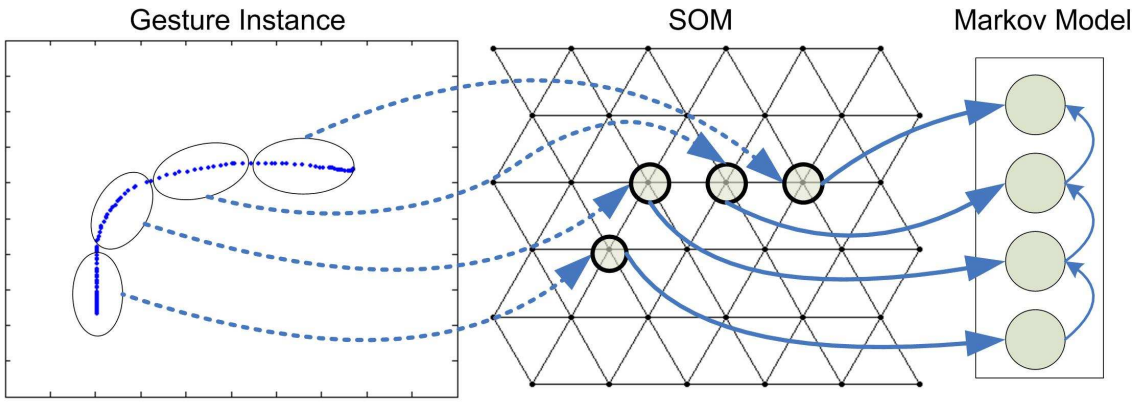
όπου N είναι μια συνάρτηση που αφαιρεί τις διαδοχικές ίσες τιμές u_i και $G^{'}$ είναι η μετασχηματισμένη μορφή του στιγμιοτύπου του νοήματος. Ο μετασχηματισμός αυτός μπορεί να θεωρηθεί ως μετασχηματισμός συνεχούς σήματος σε m διακριτά σύμβολα τα οποία καθορίζουν τις πεπερασμένες καταστάσεις του Μαρκοβιανού πρώτης τάξης. Με την αφαίρεση διαδοχικών ίσων τιμών για τα σύμβολα u, οι πιθανότητες αυτομετάβασης στη μήτρα πιθανοτήτων μετάβασης είναι αρχικά μηδενικές. Με την εφαρμογή του ίδιου μετασχηματισμού $N(T)$ στην φάση αποκωδικοποίησης, όπως θα εξηγηθεί λεπτομερώς στο κεφάλαιο 4.2.3, οι πιθανότητες αυτομετάβασης θα μηδενιστούν επίσης για το άγνωστο στιγμιότυπο. Αυτή η διαδικασία οδηγεί σε απώλεια πληροφορίας σχετικά με τη διάρκεια μιας κατάστασης αλλά αυτή η πληροφορία δεν θεωρείται κρίσιμη για την αυτόματη αναγνώριση νοηματικής γλώσσας και χειρονομιών.

Ένα Μαρκοβιανό πρότυπο, για κάθε μια από τις c κατηγορίες, δημιουργείται. Η ακολουθία των τιμών u_i στην μετασχηματισμένη μορφή του νοήματος $G^{'}$ του συνόλου $D_j^{'}$, χρησιμοποιείται για τον υπολογισμό των πιθανοτήτων μετάβασης και την πιθανότητα πρώτης κατάστασης π_j^{som} του πρότυπου MM_j^{som} που επιχειρεί να περιγράψει την κατηγορία j. Το αποτέλεσμα είναι ένα σύνολο c Μαρκοβιανών μοντέλων MM^{som}. Μεγάλο πλεονέκτημα σε αυτή την προσέγγιση είναι πως η διαδικασία σχεδιασμού του συστήματος απαλλάσσεται από τον προβληματισμό της επιλογής του αριθμού των καταστάσεων που θα απαρτίζουν την Μαρκοβιανή αλυσίδα καθώς αυτή καθορίζεται αυτόματα σε ένα ακόμα δείγμα αυτοργάνωσης και προσαρμοστικότητας. Αντίστοιχες επιλογές, όπως και ο αριθμός των Γκαουσιανών κατανομών σε περιπτώσεις συνεχών προτύπων, στα HMM δείχνουν να επηρεάζουν δραματικά την απόδοση τους και τέτοιου τύπου ελεύθερες παράμετροι, που συχνά καθορίζονται πειραματικά και σχεδιαστικές αποφάσεις αποτελούν ένα από τα μειονεκτήματα της συγκεκριμένης μεθόδου.

$$\begin{aligned} MM^{som} = \{MM_1^{som}, MM_2^{som}, \ldots, MM_c^{som}\} \\ : D_j^{'} = \{G_1^{'}, G_2^{'}, \ldots, G_{n_j}^{'}\} \rightarrow MM_j^{som} \end{aligned} \tag{4.3}$$

Αυτά τα πρότυπα χρησιμοποιούνται αργότερα για την αξιολόγηση μιας αχαρακτήριστης χειρονομιάς προκειμένου να κατηγοριοποιηθεί σε μια από τις c κατηγορίες. Το σχήμα 4.8 απεικονίζει τον μετασχηματισμό για ένα στιγμιότυπο χειρονομιάς με έναν περισσότερο διαισθητικό τρόπο.

Κατά τη διάρκεια της φάσης εκπαίδευσης και πιο συγκεκριμένα κατά τον υπολογισμό της μήτρας πιθανοτήτων μετάβασης η σχέση γειτνίασης μεταξύ των καταστάσεων λαμβάνεται επίσης υπόψη. Αν υποθέσουμε πως ένα στιγμιότυπο νοήματος που χρησιμοποιείται για την εκπαίδευση περιλαμβάνει μια μετάβασης μεταξύ u_i και u_j, πολλαπλές συνάψεις μεταβάσεων δημιουργούνται. Δημιουργούνται συνάψεις από u_i στο u_j και όλους τους γειτονικούς κόμβους του. Το σχήμα 4.9 επεξηγεί αυτήν την



Σχήμα 4.8: Αντιστοίχηση σημείων τροχιάς χεριών σε κόμβους του SOM, που αποτελούν καταστάσεις των Μαρκοβιανών μοντέλων

διαδικασία, όπου με πράσινο χρώμα εμφανίζεται ο κόμβος u_i, κόκκινο u_j και πορτοκαλί οι γείτονες του u_j. Το βάρος των συνάψεων αυτών, που διαμορφώνουν τελικά την πιθανότητα για αυτή τη μετάβαση, είναι ανάλογο προς τη γειτονική σχέση μεταξύ u_j και του αντίστοιχου γείτονα. Έτσι για την πραγματική μετάβαση $u_i \rightarrow u_j$ η μήτρα μεταβάσεων ενημερώνεται σύμφωνα με την ακόλουθη εξίσωση:

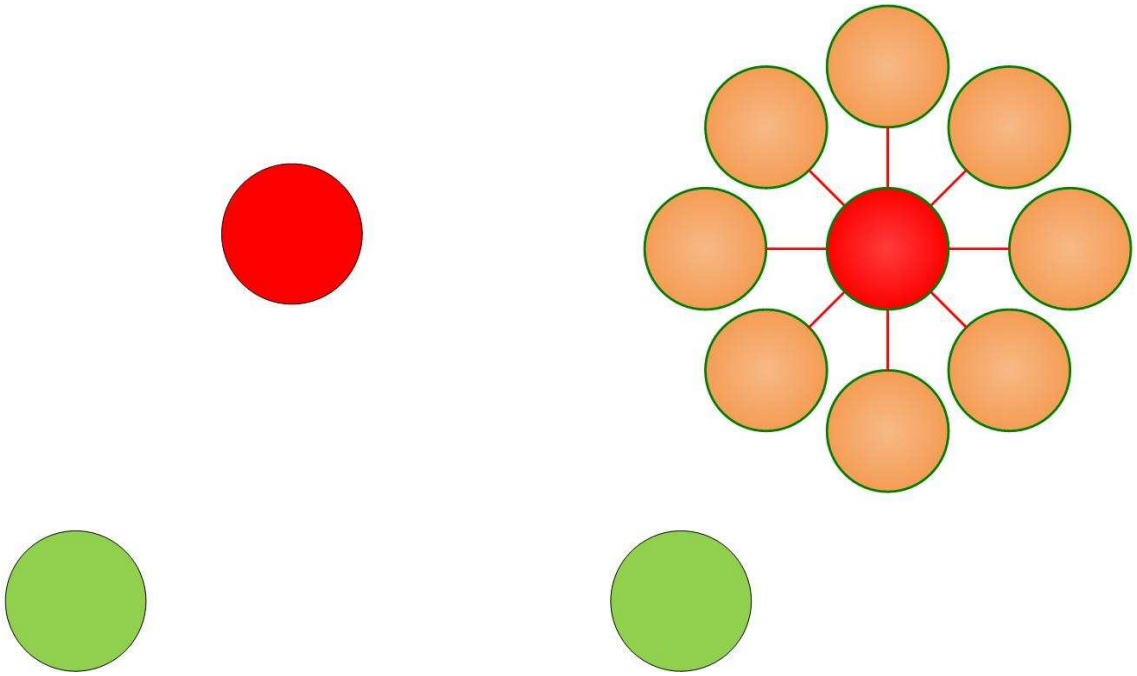
$$\begin{gathered} TM(i,x) = TM(i,x) + NF_{u_j}(x) \\ \forall x \in [1,N] : NF_{u_j}(x) > 0 \end{gathered} \tag{4.4}$$

όπου TM η μήτρα μεταβάσεων για το αντίστοιχα Μαρκοβιανό πρότυπο και N ο αριθμός καταστάσεων. Η TM ενημερώνεται επαναληπτικά ώστε να υπολογισθούν οι πιθανότητες μετάβασης από τους γείτονες του u_j σε κάθε κόμβο $u_k : TM(j,k) > 0$ και όλους τους γείτονες τους $u_l : NF_{u_k}(l) > 0$. Η ποσότητα αύξησης αυτή τη φορά για τη μετάβαση $TM(x,l)$ είναι $NF_{u_j}(x)NF_{u_k}(l)$, λαμβάνοντας υπόψη τη σχέση γειτνίασης τόσο του u_j με τον u_x όσο και του u_l με τον u_k. Η TM τελικά κανονικοποιείται ώστε οι γραμμές, που αντιστοιχούν σε καταστάσεις, που περιλαμβάνονται σε ακολουθίες εκπαίδευσης, να αθροίζουν στην μονάδα και τον μηδενισμό της διαγωνίου (αυτομετάβαση).

Αυτή η διαδικασία διασφαλίζει ότι σε περίπτωση που η ακολουθία καταστάσεων που αντιστοιχεί στο αχαρακτήριστο στιγμιότυπο περιέχει μετάβαση που δεν περιέχεται σε κανένα από τα στιγμιότυπα που χρησιμοποιήθηκαν κατά την εκπαίδευση, αυτή η μετάβαση δεν θα αποκλειστεί από την αποκωδικοποίηση της ακολουθίας και θα περιληφθεί στην αναζήτηση της βέλτιστης διαδρομής με μη μηδενική πιθανότητα. Αυτό ενισχύει την ευρωστία της αρχιτεκτονικής ενάντια στις διακυμάνσεις εκτέλεσης από τους χρήστες και ελαχιστοποιεί την ανάγκη για εξαντλητική εκπαίδευση με πολυάριθμα στιγμιότυπα εκπαίδευσης, καθιστώντας την αρχιτεκτονική κατάλληλη για ένα σύστημα ανεξάρτητο από τον χρήστη με ελάχιστη εκπαίδευση (minimal training user independent). Ένα επιπλέον χαρακτηριστικό της προτεινόμενης αρχιτεκτονικής που υπερτερεί άλλων προσεγγίσεων (HMM) που απαιτούν εκτεταμένα σύνολα εκπάιδευσης.

4.2.2.2 Πρότυπο Κατεύθυνσης

Με σκοπό την περιγραφικότερη αναπαράσταση κάθε κατηγορίας νοημάτων, εισάγεται ένας πρόσθετος μετασχηματισμός, βασισμένος στην κίνηση των χεριών κατά την εκτέ-



Σχήμα 4.9: Μεταβάσεις βασισμένες σε σχέσεις γειτνίασης

λεση της χειρονομίας. Αυτός περιγράφει την ακολουθία διαφορετικών κατευθύνσεων κατά την τροχιά των χεριών σε αντίθεση με την θέση στον χώρο που περιγράφηκε νωρίτερα. Προκειμένου να επιτευχθεί μια τέτοια αναπαράσταση, υπολογίζονται τα διανύσματα κατεύθυνσης από τα διαδοχικά σημεία της τροχιάς του χεριού σύμφωνα με την εξίσωση 4.5. Αυτές οι γωνίες κβαντίζονται σε οχτώ διαφορετικές συμβολικές τιμές όπως απεικονίζονται στο σχήμα 4.10. Τα τμήματα συντεταγμένων στο σχήμα 4.8 και 4.10 θεωρούνται ως ένα σύνολο συντεταγμένων που ανήκουν στην ίδια συστάδα (BMU και κβαντισμένη γωνία για το σχήμα 4.8 και 4.10 αντίστοιχα). Υπό αυτή την έννοια, ορίζουμε το μετασχηματισμό ενός στιγμιοτύπου G χρησιμοποιώντας την συνάρτηση OF ακολούθως:

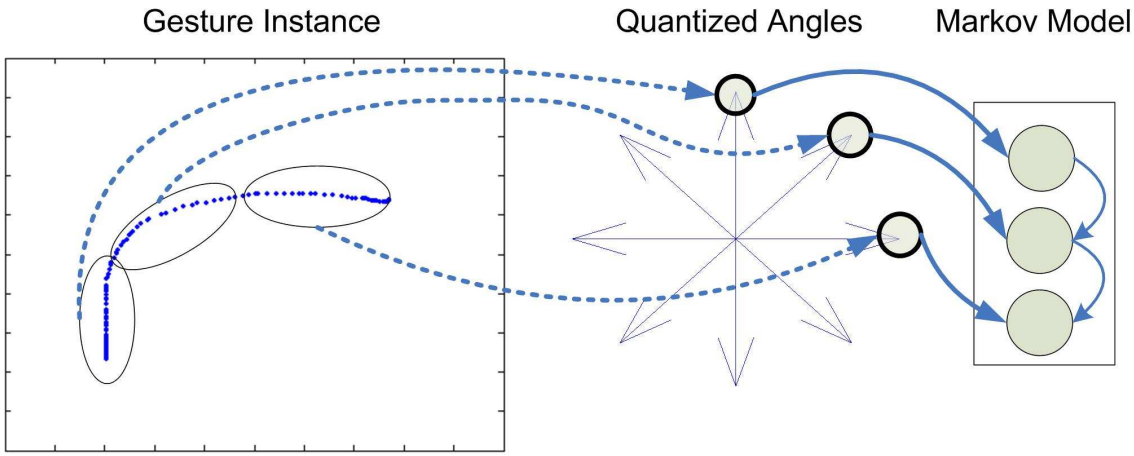
$$\begin{aligned} OF(G) &= \{v_1, v_2, ..., v_m\} \\ : v_i &= W_r(Q(\arctan(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}))) \end{aligned} \tag{4.5}$$

όπου v_i είναι τα επίπεδα κβαντισμού, Q η συνάρτηση κβαντισμού και W_r ένα φίλτρο ενδιάμεσης τιμής που εφαρμόζεται στις τιμές για ένα σταθερό παράθυρο μήκους r γύρω από την τρέχουσα τιμή. Ο στόχος του τελευταίου φίλτρου είναι η εξομάλυνση των κβαντισμένων τιμών απομακρύνοντας πιθανές αστάθειες του χεριού κατά τη διάρκεια της εκτέλεσης του νοήματος ή ατέλειες που οφείλονται στις μεθόδους επεξεργασίας εικόνας. Εφαρμόζοντας τη συνάρτηση μετασχηματισμού σε συνδυασμό με την συνάρτηση N απαλειφής διαδοχικών ίσων τιμών, έχουμε:

$$G_i^{''} = N(OF(G)) = \{v_1, v_2, ..., v_m\} \tag{4.6}$$

όπου οι τιμές v_i καθορίζουν τις καταστάσεις για ένα νέο σύνολο Μαρκοβιανών μοντέλων MM^{of} που εκπαιδεύεται χρησιμοποιώντας το μετασχηματισμένο σύνολο $D_j^{''}$. Η πιθανότητα πρώτης κατάστασης π_j^{of} υπολογίζεται επίσης.

$$MM^{of} = \{MM_1^{of}, MM_2^{of}, ..., MM_c^{of}\}$$
$$: D_j^{''} = \{G_1^{''}, G_2^{''}, ..., G_n^{''}\} \rightarrow MM_i^{of} \tag{4.7}$$



Σχήμα 4.10: Δημιουργία Μαρκοβιανού μοντέλου από την πληροφορία κατεύθυνσης της χειρονομίας

4.2.2.3 Γενικευμένος μέσος και Απόσταση Levenshtein

Ένα πρόσθετο πρότυπο δημιουργείται ανά κατηγορία νοημάτων, ο γενικευμένος μέσος του συνόλου $D_j^{'}$. Ο γενικευμένος μέσος (generalized median) ενός συνόλου ακολουθιών S ορίζεται ως η ακολουθία, που αποτελείται από έναν συνδυασμό, όλων ή μερικών συμβόλων από αυτά που χρησιμοποιούνται στο σύνολο, που ελαχιστοποιεί συνολικά την απόσταση από όλες τις ακολουθίες του συνόλου S [143]. Στην περίπτωση που ο γενικευμένος μέσος ανήκει στο σύνολο S καλείται γενικευμένος μέσος συνόλου (Generalized Set Median).

$$M_j = generalized\_median(D_j^{'})$$
$$= \arg\min_{g} \sum_{G' \in D_j'} L(g, G') \tag{4.8}$$

Έστω M_j ο γενικευμένος μέσος του συνόλου $D_j^{'}$, όπως υπολογίστηκε χρησιμοποιώντας μια τροποποιημένη έκδοση της απόστασης Levenshtein L, μια ευρέως διαδεδομένη μετρική απόστασης. Αυτή η παραλλαγή της απόστασης Levenshtein ενσωματώνει τη γειτονική σχέση μεταξύ κόμβων του SOM που είναι και τα σύμβολα των δύο ακολουθιών που συγκρίνονται για την ανάθεση κόστους αντικατάστασης συμβόλων και υιοθετείται επίσης κατά τη διάρκεια της αποκωδικοποίησης συζητείται στο κεφάλαιο 4.2.3. Ο αρχικός αλγόριθμος ανάθεσης κόστους παρουσιάζεται στον αλγόριθμο 1 και πραγματοποιείται κατά τη διάρκεια σύγκρισης κάθε συμβόλου συμβόλου $(str1[i], str2[j])$ των ακολουθιών $str1$ και $str2$ αντίστοιχα ώστε να λάβει την απόφαση σχετικά με το ποια ενέργεια επιτυγχάνει το ελάχιστο κόστος.

Εν τούτοις ένας τέτοιος αλγόριθμος δεν λαμβάνει υπόψη πόσο παρόμοια είναι τα σύμβολα $str1[i], str2[j]$, εάν υπάρχει πραγματικά κάποια σχέση ομοιότητας ή γειτνίασης μεταξύ των συμβόλων του συνόλου. Στην περίπτωση του SOM μια τέτοια σχέση

Algorithm 1 Αρχικός αλγόριθμος υπολογισμού απόστασης Levenshtein

if $str1[i] = str2[j]$ **then**
  $cost := 0$
else
  $cost := 1$
end if
$d[i,j] := minimum($
$d[i-1,j]+1, //deletion$
$d[i,j-1]+1, //insertion$
$d[i-1,j-1]+cost//substitution$
)

υπάρχει μεταξύ των κόμβων που αποτελούν τα σύμβολα του συνόλου που απαρτίζουν κάθε ακολουθία. Το κόστος που ανατίθεται κατά την αντικατάσταση θα έπρεπε να είναι μικρότερο εάν τα δύο σύμβολα που συμμετέχουν στην αντικατάσταση είναι αρκετά κοντά σύμφωνα με κάποια μετρική ομοιότητας (συνάρτηση γειτνίασης στην περίπτωσή μας) και αντίστοιχα μεγαλύτερο εάν οι δύο κόμβοι είναι απομακρυσμένοι. Έτσι ο αλγόριθμος 1 τροποποιείται αναλόγως:

Algorithm 2 Τροποποιημένος αλγόριθμος υπολογισμού Levenshtein απόστασης για αντικατάσταση συμβόλων

if $str1[i] = str2[j]$ **then**
  $cost := 0$
else
  $cost := 1 - NF_{str1[i]}(str2[j])$
end if

Κατά συνέπεια το κόστος για μια αντικατάσταση είναι ανάλογο της γειτονικής σχέσης των δύο κόμβων (σύμβολα ακολουθιών) που συμμετέχουν στην αντικατάσταση. Παρόμοια παραλλαγή μπορεί να εφαρμοστεί στην μετρική απόστασης Damerau–Levenshtein για την περίπτωση της αντιμετάθεσης δύο συμβόλων, μια περίπτωση αρκετά σπάνια στην νοηματική γλώσσα. Τέλος, η σχέση γειτνίασης των κόμβων επηρεάζει και τις άλλες δύο ενέργειες στον υπολογισμό απόστασης Levenshtein: συγκεκριμένα της διαγραφής και της εισαγωγής. Το κόστος κάθε ενέργειας είναι ίσο με τη γειτονική σχέση του κόμβου i που παρεμβάλλεται ή διαγράφεται και του προηγούμενου $i-1$ και επόμενου $i+1$: $1-\frac{NF_i(i-1)+NF_i(i+1)}{2}$. Η μέση απόσταση Levenshtein ML_j μεταξύ των μελών του συνόλου $D_j^{'}$ και της M_j υπολογίζεται επίσης και αποτελεί έναν άτυπο τρόπο μέτρησης της διακύμανσης στα μέλη του συνόλου και θα χρησιμοποιηθεί αναλόγως στο στάδιο αποκωδικοποίησης (κεφάλαιο 4.2.3).

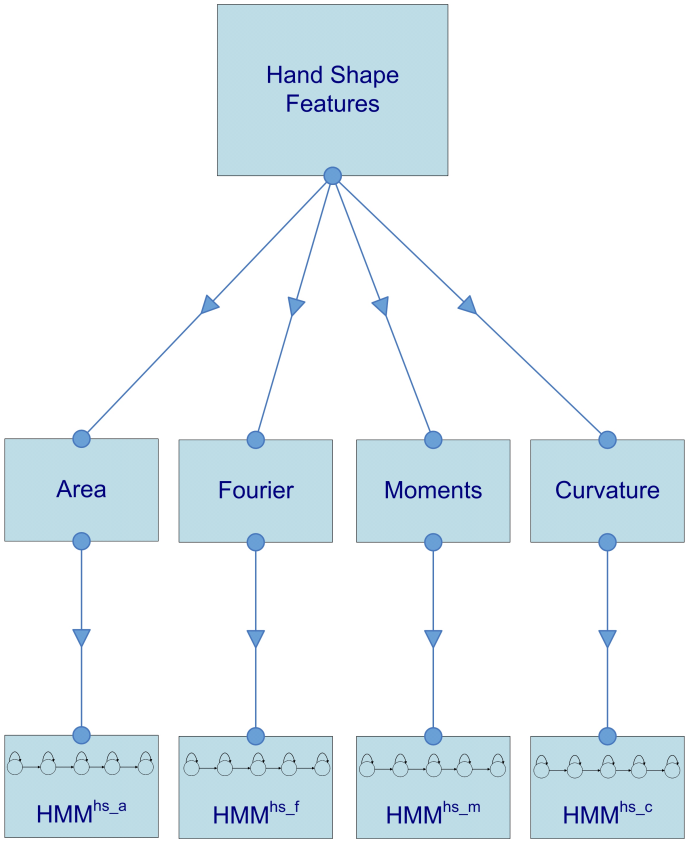
$$ML_j = \frac{\sum_{i=1}^{n_j} L(G_i^{'}, M_j)}{n_j} \tag{4.9}$$

4.2.2.4 Πρότυπα Χειρομορφής

Η προτυποποίηση της χειρομορφής κατά τη διάρκεια του νοηματισμού μπορεί να είναι αρκετά πολύπλοκη διεργασία ειδικά όταν τα χαρακτηριστικά γνωρίσματα που την περιγράφουν προέρχονται από επεξεργασία μονοκάμερης καταγραφής. Το σύνολο

χαρακτηριστικών γνωρισμάτων που περιγράφει την χειρομορφή επιχειρεί να μοντελοποιήσει τρείς διαφορετικές ροές πληροφορίας που εκτυλίσσονται παράλληλα: τον προσανατολισμό της παλάμης, την κατεύθυνση των ακροδακτύλων και τις γωνίες κάμψης των αρθρώσεων των δακτύλων. Το αποτέλεσμα της συνύπαρξης αυτών των ροών πληροφορίας είναι μια εικόνα, η οποία επεξεργάζεται και παράγονται χαρακτηριστικά γνωρίσματα ικανά μόνο να περιγράψουν τον τελικό συνδυασμό των πτυχών αυτών.

Προκειμένου να επιτευχθεί η προτυποποίηση της χειρομορφής χρησιμοποιούμε τέσσερα συνεχή HMM [223] [224]. Όπως φαίνεται και στην εικόνα 4.11 χαρακτηριστικά περιοχής του χεριού HMM^{hs_a}, Fourier περιγραφείς HMM^{hs_f}, ροπές HMM^{hs_m} και Cepstrum συντελεστές καμπυλότητας HMM^{hs_c} χρησιμοποιούνται για να μοντελοποιήσουν διαφορετικές πτυχές του συνδυασμού των ροών πληροφορίας που περιγράφηκαν παραπάνω σε μια προσπάθεια ανάλυσης συνιστωσών.



Σχήμα 4.11: Hidden Markov Models βασισμένα σε γνωρίσματα χειρομορφής [223] [224]

4.2.3 Αποκωδικοποίηση Χειρονομίας

Για την κατηγοριοποίηση ενός αχαρακτήριστου στιγμιοτύπου νοήματος, υποβάλλεται προς εξέταση σε όλα τα προαναφερθέντα εκπαιδευμένα πρότυπα και οι πιθανότητες συμμετοχής συνδυάζονται, σταθμίζοντας τις χρησιμοποιώντας βάρη που υπολογίζονται σύμφωνα με τα μεμονωμένα ποσοστά αναγνώρισης, επιτυγχάνοντας κατά συνέπεια ένα εύρωστο σχήμα αναγνώρισης ενάντια σε περιπτώσεις χαμηλής εμπιστοσύνης ή αμφιβολίας.

Η κατηγοριοποίηση ενός νοήματος εισόδου βασίζεται στα δύο σύνολα Μαρκοβιανών προτύπων, την Levenshtein απόστασης με την ακολουθία γενικευμένου μέσου και τα HMM της χειρομορφής όπως περιγράφονται στις εξισώσεις 4.3, 4.7 και στα κεφάλαια 4.2.2.3, 4.2.2.4 αντίστοιχα. Έστω G_k ένα στιγμιότυπο νοήματος άγνωστης κατηγορίας και $G_k^{'}$ και $G_k^{''}$ οι μετασχηματισμοί του, σύμφωνα με τις εξισώσεις 4.2 και 4.6. Χρησιμοποιώντας το σύνολο προτύπων MM^{som}, η πιθανότητα του στιγμιοτύπου να ανήκει στην κατηγορία j μπορεί να υπολογιστεί ως εξής:

$$P(G_k^{'}|MM_j^{som}) = \prod_{i=1}^{q} S_i^{som} \tag{4.10}$$

Η παραπάνω εξίσωση υπολογίζει το γινόμενο των τιμών S_i^{som}, οι οποίες αντιπροσωπεύουν έναν παράγοντα αξιολόγησης για κάθε κατάσταση $u_i : i \in [1, q] : q = \left|G_k^{'}\right|$ του μετασχηματισμού $G_k^{'}$ σε σχέση με το πρότυπο MM_j^{som}. Αυτές οι τιμές υπολογίζονται ως εξής:

$$\begin{aligned} S_1^{som} &= \max_z(NF_{u_1}^{som}(z)\pi_j^{som}) \\ S_i^{som} &= \max_z(NF_{u_i}^{som}(z)P(z|u_{i-i}, MM_j^{som})) \end{aligned} \tag{4.11}$$

Για την αξιολόγηση της πρώτης κατάστασης, εκτελείται απλά μια αναζήτηση για τον κόμβο που έχει την μεγαλύτερη συνδυασμένη πιθανότητα του:

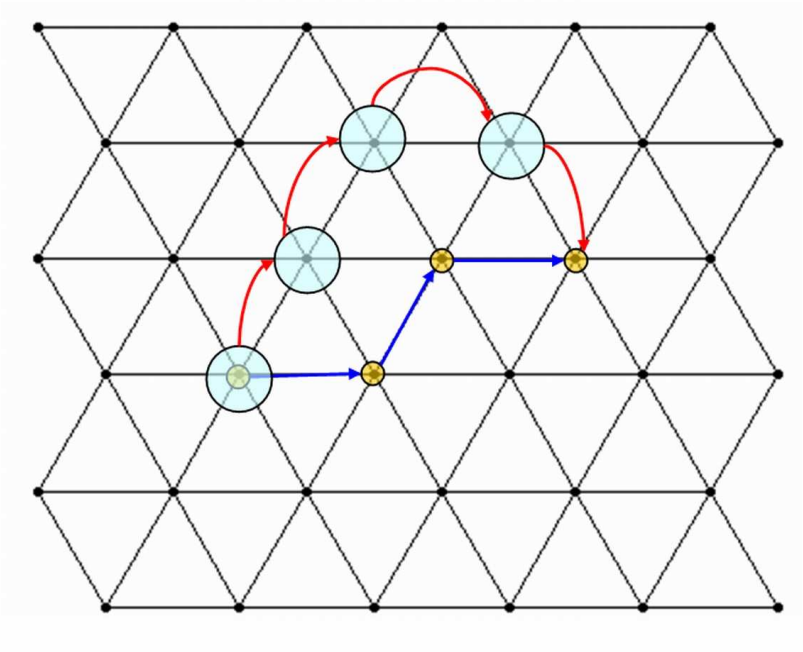
- να είναι γειτονικό στο u_1 και συγκεκριμένα $NF_{u_1}^{som}(z)$
- να είναι η πρώτη κατάσταση της Μαρκοβιανής αλυσίδας MM_j^{som} και συγκεκριμένα π_j^{som}

Για τους κόμβους που $\in [2, q]$, μια παρόμοια αναζήτηση εκτελείται αλλά η δεύτερη πιθανότητα δεν είναι αυτή της πρώτης κατάστασης αλλά μια πιθανότητα μετάβασης $P(z|u_{i-i}, MM_j^{som})$. $NF_{u_i}^{som}(z)$ είναι η απόσταση του κόμβου κατάστασης z με τον u_i όπως ορίζεται από την συνάρτηση γειτνίασης SOM με τον δεύτερο κόμβο ως κέντρο του. Δεδομένου ότι το z διατρέχει όλους τους κόμβους του χάρτη, η ποσότητα αυτή συνδυάζει την πιθανότητα μετάβασης από την προηγούμενη κατάσταση σε μια σχετικά μικρή απόσταση επάνω στο πλέγμα του χάρτη από την τρέχουσα κατάσταση. Αυτή η μονάδα θα χρησιμοποιηθεί επίσης ως προηγούμενη κατάσταση στο επόμενο βήμα:

$$u_i = \arg\max_z(S_i^{som}) : i \in [1, q] \tag{4.12}$$

Αυτή η διαδικασία αναζήτησης εξασφαλίζει την ελαχιστοποίηση της απόκλισης στον χώρο κατά τη διάρκεια της εκτέλεσης του νοήματος, ενισχύοντας τη συνολική αρχιτεκτονική με ευρωστία και προσαρμοστικότητα. Αυτό μπορεί να γίνει καλύτερα κατανοητό με ένα επεξηγηματικό παράδειγμα, όπως αυτό που φαίνεται στο σχήμα 4.12. Έστω μια ακραία περίπτωση όπου ο πίνακας μετάβασης του προτύπου MM_j^{som} έχει μοναδιαίες τιμές για τις μεταβάσεις που εμφανίζονται με μπλε βέλη και μηδενικές για όλες τις άλλες μεταβάσεις και ένα στιγμιότυπο νοήματος με τροχία (ανοικτό μπλε) και μεταβάσεις που χρωματίζονται με κόκκινα βέλη. Διαισθητικά ένας ανθρώπινος χρήστης θα κατέληγε στο συμπέρασμα ότι το στιγμιότυπο αυτό είναι παρόμοια με το πρότυπο MM_j^{som} αλλά όχι ακριβώς ίδιο. Το σύστημα θα συμπέραινε ακριβώς ότι η ομοιότητα του στιγμιοτύπου με την κατηγορία είναι τόσο μεγάλη όσο εγγύτερα είναι οι αντίστοιχοι κόμβοι. Πραγματικά, θα ανέθετε μια ποινή κατά τη διάρκεια

της αποκωδικοποίησης ίση με το γινόμενο των τιμών $NF_a^{som}(a^{'})$, όπου a ο εκάστοτε κόμβος του προτύπου και $a^{'}$ ο αντίστοιχος κόμβος του στιγμιοτύπου.



Σχήμα 4.12: Αποκωδικοποίηση βάση θέσης: ένα επεξηγηματικό παράδειγμα

Μια σχεδόν πανομοιότυπη διαδικασία αποκωδικοποίησης εκτελείται και για την περίπτωση της κατεύθυνσης κίνησης. Η μόνη διαφορά είναι πως η συνάρτηση γειτνίασης NF^{of} ορίζεται αυθαίρετα και πιο συγκεκριμένα ανατίθεται τιμή $1/2$ για την άμεσα γειτονική κατεύθυνση, $1/4$ για την δεύτερη γειτονική και μηδενική τιμή για όλες τις υπόλοιπες. Κατά συνέπεια οι αντίστοιχες εξισώσεις είναι:

$$P(G_k^{''}|MM_j^{of}) = \prod_{i=1}^{q} S_i^{of} \tag{4.13}$$

$$\begin{aligned} S_1^{of} &= \max_z(NF_{v_1}^{of}(z)\pi_j^{of}) \\ S_i^{of} &= \max_z(NF_{v_i}^{of}(z)P(z|v_{i-i}, MM_j^{of})) \end{aligned} \tag{4.14}$$

$$v_i = \arg\max_z(S_i^{of}) : i \in [1,q], : q = \left|G_k^{''}\right| \tag{4.15}$$

Τα σύντομα (μικρές τιμές q) στιγμιότυπα νοημάτων τείνουν να αποκομίσουν πλεονέκτημα αφού διαθέτουν λιγότερες καταστάσεις, άρα λιγότερες μεταβάσεις. Για να αντιμετωπίσουμε αυτό το πρόβλημα εισάγουμε μια πρόσθετη μετρική ομοιότητας βασισμένη στον γενικευμένο μέσο κάθε κατηγορίας M_j σύμφωνα με την τροποποιημένη απόσταση Levenshtein, που περιγράφεται στο κεφάλαιο 4.2.2.3. Αυτό αντιμετωπίζει επίσης το πρόβλημα της υπόχειρονομιάς, όπου εάν ένα στιγμιότυπο αποτελεί το αρχικό μέρος μιας κατηγορίας νοημάτων λαμβάνει υψηλή αξιολόγηση χρησιμοποιώντας μόνο MM^{som} και MM^{of}.

$$P(G_k^{'}|M_j) = \frac{ML_j}{L(G_k^{'}, M_j)} \tag{4.16}$$

$$P(HS_k|HMM_j) = \sum_{h\in[a,f,m,c]} w_h P(HS_k^h|HMM_j^{hs_i}) \tag{4.17}$$

Ας σημειωθεί πως η ποσότητα $P(G_k^{'}|M_j)$ είναι μια μετρική ομοιότητας και όχι πιθανότητα, δεδομένου ότι μπορεί να λάβει τιμές άνω της μονάδας. Σχετικά με την αποκωδικοποίηση κάθε νοήματος σύμφωνα με τα HMM που περιγράφονται στο κεφάλαιο 4.2.2.4, υπολογίζεται ως σταθμισμένο άθροισμα των μεμονωμένων μορφών πληροφορίας της χειρομορφής, όπως φαίνεται στην εξίσωση 4.17. Τα βάρη του σταθμισμένου αθροίσματος προκύπτουν από τα ποσοστά αναγνώρισης βασισμένα σε μεμονωμένη ροή πληροφορίας Τέλος, η νικήτρια κατηγορία αποφασίζεται σύμφωνα με την εξίσωση 4.18 για το κυρίαρχου και το μη-κυρίαρχου χέρι.

$$\begin{gathered} G_k \in D_j \Leftrightarrow \arg\max_j (w_{som} P(G_k^{'}|MM_j^{som}) \\ +w_{of} P(G_k^{''}|MM_j^{of}) + w_L P(G_k^{'}|M_j) + P(HS_k|HMM_j)) \end{gathered} \tag{4.18}$$

Κάθε ένας από τους τέσσερις όρους της εξίσωσης 4.18 αντιπροσωπεύει τη συμμετοχή κάθε χωριστής μορφής πληροφορίας στη λήψη της τελικής απόφασης για την κατηγοριοποίηση του στιγμιοτύπου G_k σε μια από τις κατηγορίες του λεξιλογίου. Σε κάθε ένα από τα κανάλια πληροφορίας ανατίθεται ένα βάρος $(w_{som}, w_{of}, w_L, w_a, w_f, w_m, w_c,)$ που προέρχεται από το ποσοστό αναγνώρισης που επιτυγχάνει κάθε κανάλι χωριστά. Στο αδύνατο (μη κυρίαρχο) χέρι ανατίθεται ένα επιπλέον βάρος $(w_{weak} < 1)$ δεδομένου ότι η συμμετοχή του στη διαδικασία λήψης απόφασης είναι μικρότερη από αυτήν του κυρίαρχου χεριού.

Περαιτέρω κριτήρια εξασφάλισης ποιότητας μπορούν να εφαρμοστούν υπό μορφή κατωφλίου είτε στη συνολική αξιολόγηση είτε τμηματικά, σε όρους της εξίσωσης 4.18. Επιπλέον, περιπτώσεις ασάφειας μπορούν να εντοπιστούν όταν n κατηγορίες έχουν όλες υψηλά αποτελέσματα αξιολόγησης. Η ασάφεια μπορεί να επιλυθεί με τον έλεγχο της διαφοράς επίδοσης των κατηγοριών με την υψηλότερη απόδοση.

4.2.4 Ανάλυση Λάθους

Γενικά ο ορισμός της συνάρτησης λάθους για τη $f(x, y)$, για ανεξάρτητο λάθος $\delta x, \delta y$ είναι:

$$\delta f^2 = (\frac{\partial f}{\partial x}\delta x)^2 + (\frac{\partial f}{\partial y}\delta y)^2 \tag{4.19}$$

Το SOM, σε αντίθεση με παραδοσιακές τεχνικές ανάλυσης λάθους, χρησιμοποιεί διάφορες μετρικές ανάλυσης λάθους για να καθορίσει την ποιότητα αναπαράστασης/χαρτογράφησης, καθοριστικό κριτήριο για τον επιτυχημένο σχεδιασμό ενός χάρτη. Η απλούστερη και συνηθέστερα χρησιμοποιούμενη μετρική είναι αυτή του συσσωρευμένου λάθους κβάντισης e (cumulative quantization error), το οποίο είναι η μέση απόσταση μεταξύ των δειγμάτων εισόδου d_i και των αντίστοιχων μονάδων του SOM BMU.

$$e = \frac{E}{N} : E = \sum_1^N distance(d_i, w_{bmu}) \tag{4.20}$$

Μια εναλλακτική μέτρηση λάθους, που περιλαμβάνει επίσης την γειτονιά, είναι η διαστρέβλωση (distortion) και ορίζεται στην εξίσωση 4.21. Όπως φαίνεται από την

τελευταία εξίσωση δύο είναι οι όροι που καθορίζουν την διαστρέβλωση: η συνάρτηση γειτνίασης $h(bmu(i), j)$ και η διανυσματική απόσταση προτύπου-δεδομένων $w_j - x(i)$. Η ποσότητα αυτή είναι καταλληλότερη στην εργασία μας καθώς περιλαμβάνει και την συνάρτηση γειτνίασης, που διαδραματίζει καθοριστικό ρόλο τόσο κατά την εκπαίδευση των προτύπων όσο και κατά την αποκωδικοποίηση στιγμιοτύπων χειρονομιών. Σε ειδικές περιπτώσεις, σταθερή γειτνίαση και διακριτά δεδομένα, η διαστρέβλωση μπορεί να ερμηνευθεί ως συνάρτηση ενέργειας του SOM την οποία ελαχιστοποιεί προσεγγιστικά η φάση εκπαίδευσης. Οι Wu και Takatsuka στο [254] προτείνουν μια μετρική λάθους βασισμένη στην εντροπία του SOM EH. Η εξίσωση 4.22 ορίζει την εντροπία του SOM, όπου k δείγματα αντιστοιχούνται σε n νευρώνες του SOM και m είναι ο συνολικός αριθμός νευρώνων.

$$D^{SOM} = \sum_i \sum_j h(bmu(i), j) \parallel w_j - x(i) \parallel^2 \tag{4.21}$$

$$\begin{aligned} EH &= -\sum_{i=1}^{m} e_i \log e_i \\ E_n &= \sum_{j=1}^{k} distance(d_j, w_n) \\ e_n &= \frac{E_n}{E} \end{aligned} \tag{4.22}$$

Προεκειμένου να μελετήσουμε την διάδοση λάθους στην προτεινόμενη αρχιτεκτονική διεξάγεται μια μελέτη λάθους του και εστιάζουμε περισσότερο στο στάδιο αποκωδικοποίησης SOM της ευρύτερης αρχιτεκτονικής ως πιο χαρακτηριστική και ενδεικτική μονάδα της ευρύτερης αρχιτεκτονικής. Κατά συνέπεια περιορίζουμε τη μελέτη μας στην αξιολόγηση της εξίσωσης 4.10 $P(G_k^{'}|MM_j^{som})$ και $\prod_{i=1}^{q} S_i^{som}$. Ερευνήσαμε την επίδραση ενός τυχαίου λάθους $\delta x, \delta y$ του σημείου x, y στην τροχιά G_k έτσι ώστε τελικά το σημείο τροχιάς να είναι $x + \delta x, y + \delta y$ κατά την αξιολόγηση της $S_i^{som} = \max_z(NF_{u_i}^{som}(z)P(z|u_{i-i}, MM_j^{som}))$. Χάριν απλότητας υποθέτουμε ότι κάθε σημείο τροχιάς σε G_k αντιστοιχείται σε διαφορετικό BMU στο SOM έτσι ώστε $u_t^{'} \neq u_{t-1}^{'} \forall t \in [2, m]$ in $G_k^{'}$. Η υπόθεση αυτή δεν επηρεάζει την συνολική ανάλυση λάθους, αφού χωρίς αυτή το πιθανό λάθος θα εμφανιζόταν σε επόμενο βήμα του αλγορίθμου και παρέχει έναν απλούστερο τρόπο εισαγωγής του τυχαίου λάθους.

Σε περίπτωση που το λάθος είναι αρκετά μικρό κανένα λάθος δεν εισάγεται στο στάδιο αποκωδικοποίησης και το λάθος απορροφάται από την εφαρμογή του SOM χωρίς να διαδίδεται στα επόμενα βήματα του αλγορίθμου αποκωδικοποίησης. Αυτό συμβαίνει επειδή γενικά οι συντεταγμένες χεριών με σχετικά μικρή απόκλιση αντιστοιχούνται στον ίδιο κόμβο του SOM αφού $BMU(x+\delta x, y+\delta y) = BMU(x, y)$. Με απλά λόγια το λάθος απορροφάται από την λειτουργία συσταδοποίησης του SOM και εξαλείφεται από τη υπόλοιπη διαδικασία. Από την άλλη όταν $\delta x, \delta y \gg$: $BMU(x+\delta x, y+\delta y) \neq BMU(x, y)$:

$$\begin{aligned} u_i &\neq u^{'} \\ u^{'} &= BMU(x, y) \\ u_i &= BMU(x + \delta x, y + \delta y) \end{aligned} \tag{4.23}$$

και το $\delta x, \delta y$ επηρεάζει τελικά την ποσότητα S_i. Έστω $u^{'}$ η πιθανότερη μετάβαση από τον κόμβο u_{i-1}, στο πρότυπο MM_j^{som}, όπως προέκυψε από το στάδιο της εκπαίδευσης και οι υπόλοιπες πιθανότητες μετάβασης από u_{i-1} αμελητέες:

$$\begin{aligned} P(u^{'} \mid u_{i-1}, MM_j^{som}) &\rightarrow 1^{-} \\ P(u \mid u_{i-1}, MM_j^{som}) &\ll \forall u \neq u^{'} \end{aligned} \tag{4.24}$$

Κατά συνέπεια ισχύει $\delta S_i^{SOM} \approx NF_{u_i}^{som}(u^{'})$, όπου $NF_{u_i}^{som}(u^{'})$ είναι η σχέση γειτνίασης $u_i, u^{'}$ και είναι ανάλογη του τυχαίου λάθους $\delta x, \delta y$. Εύκολα συμπεραίνεται ότι εφόσον $u^{'}$ γίνεται το νέο u_i, σύμφωνα με την εξίσωση 4.12, το λάθος δεν διαδίδεται στα επόμενα βήματα της διαδικασίας αναγνώρισης.

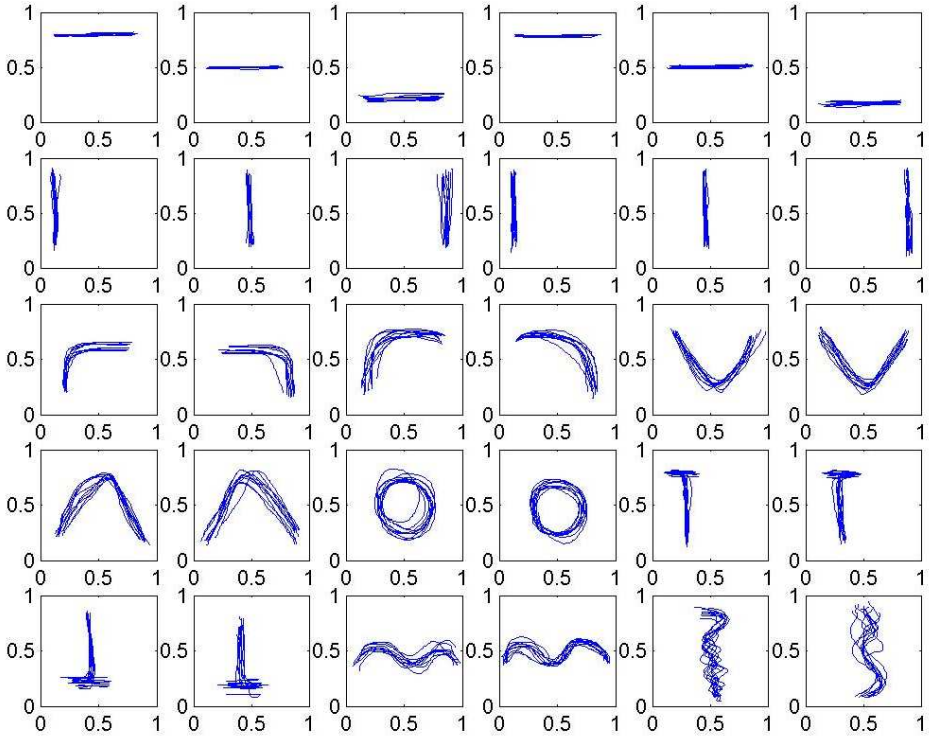
Σε ένα αφαιρετικό επίπεδο, η αποκωδικοποίηση λειτουργεί ως αλγόριθμος μεγιστοποίησης ενέργειας, επιδιώκοντας σε κάθε βήμα να μεγιστοποιήσει την ποσότητα S_i^{som} με πιθανό κόστος την επιλογή ενός διαφορετικού από τον αρχικό, αλλά πιθανότερο κόμβο BMU. Το κόστος εντοπίζεται στην ποσότητα $NF_{u_i}^{som}(u^{'})$, δεδομένου ότι με την επιλογή διαφορετικού κόμβου η αξιολόγηση ενός συγκεκριμένου ζευγαριού στιγμιοτύπου/πρότυπου τιμωρείται από αυτήν την ποσότητα προκειμένου να αντισταθμιστεί η αντικατάσταση κόμβων. Ο αλγόριθμος, σε μια πιο διαισθητική περιγραφή, προσπαθεί να συγκλίνει στο πιθανότερο μονοπάτι του προτύπου, τιμωρώντας κάθε απόκλιση από αυτήν την πορεία με τη σχέση γειτνίασης των αντίστοιχων κόμβων. Το χαρακτηριστικό αυτό καθιστά την προτεινόμενη αρχιτεκτονική εύρωστη και προσαρμοστική. Αντιμετωπίζει τόσο τα τυχαία λάθη στην πληροφορία εισόδου όσο και τη απόκλιση κατά την εκτέλεση του νοήματος από τους χρήστες. Ο συνυπολογισμός του χαρακτηριστικού γειτνίασης στη ευρύτερη διαδικασία αποκωδικοποίησης βοηθά στην αντιμετώπιση προβλημάτων που μπορεί να προκύψουν από δυναμικά παρασκήνια, ποικιλομορφία εκτέλεσης νοημάτων από τους χρήστες αλλά και ανατομικές ή εργονομικές ιδιαιτερότητες του χρήστη.

4.2.5 Πειραματικά Αποτελέσματα

Προκειμένου να επικυρωθεί η ορθότητα του προτεινόμενου μηχανισμού αναγνώρισης χειρονομιών και νοηματικής γλώσσας εκτελέσθηκαν πειράματα σε δύο σύνολα δεδομένων: ένα συνθετικό σύνολο που περιέχει μόνο διδιάστατες χωρικές πληροφορίες του ενός χεριού και ενός πραγματικού γλωσσικού σώματος Ελληνικής Νοηματικής Γλώσσας που περιγράφεται στο [68]. Το πρώτο σύνολο δεδομένων αποτελείται από 10 επαναλήψεις 30 χειρονομιών και όπως φαίνεται στο σχήμα 4.13 διαμορφώθηκε ένα σύνολο δεδομένων που περιέχει κατηγορίες που ποικίλλουν σε πολυπλοκότητα από πολύ απλές χειρονομίες σε αρκετά σύνθετες. Το δεύτερο σύνολο δεδομένων αποτελείται από 3 εκ γενετής νοηματιστές που επαναλαμβάνουν 3 φορές 118 αντιπροσωπευτικά λημμάτα της Ελληνικής Νοηματικής Γλώσσας κάτω από ελεγχόμενες συνθήκες καταγραφής.

4.2.5.1 Συνθετικό Σύνολο

Πραγματοποιήθηκαν πειράματα, χρησιμοποιώντας το συνθετικό σύνολο δεδομένων, προκειμένου να αξιολογηθεί η απόδοση αναγνώρισης της προτεινόμενης μεθόδου βασισμένης μόνο στη θέση ενός χεριού. Χρησιμοποιώντας όλα τα στιγμιότυπα, τόσο για σκοπούς εκπαίδευσης όσο και επαλήθευσης, σε μία προσπάθεια να επικύρωσης των ικανοτήτων μάθησης του συστήματος, λάβαμε ποσοστό αναγνώρισης 100%, γεγονός που



Σχήμα 4.13: Το συνθετικό πειραματικό σύνολο

επιβεβαιώνει πως η διαδικασία μοντελοποίησης προτυποποιεί επαρκώς τις κατηγορίες χειρονομιών και αυτή της αποκωδικοποίησης είναι ικανή τις αναγνωρίσει αποτελεσματικά. όσον αφορά στην αξιολόγηση των ικανοτήτων γενίκευσης της προτεινόμενης μεθόδου, ένα άλλο πείραμα πργαματοποιήθηκε χρησιμοποιώντας τη μέθοδο της διεπικύρωσης με 10 επαναλήψεις (10 fold cross validation) και σε αυτήν την περίπτωση το μέσο ποσοστό αναγνώρισης ήταν 93%. Επιπλέον πειράματα εκτελέσθηκαν σχετικά με την ακτίνα και την συνάρτηση γειτνίασης του SOM και καταλήξαμε στο συμπέρασμα ότι μια ακτίνα 2 και ο Γκαουσιανός τύπος συνάρτησης απέδωσαν καλύτερα. Προκειμένου να συγκριθούν τα αποτελέσματα του συστήματός μας με την συνηθέστερα χρησιμοποιημένη προσέγγιση στη βιβλιογραφία υλοποιήθηκε ένας κατηγοριοποιητής βασισμένος σε HMM, εκπαιδεύοντας ένα HMM ανά κατηγορία. Χρησιμοποιήσαμε συνεχή, αριστερά προς δεξιά πρότυπα και ένα μίγμα τριών γκαουσσιανών συναρτήσεων πυκνότητας πιθανότητας. Κατά τη διάρκεια της φάσης αποκωδικοποίησης κάθε στιγμιότυπο εξετάστηκε ενάντια σε όλα τα πρότυπα και αυτό με την υψηλότερη πιθανότητα συμμετόχής επιλέχτηκε ως νικητής αποδίδοντας ένα μέσο ποσοστό αναγνώρισης του 86,36%.

Η πειραματική αυτή μελέτη δείχνει ότι η προτεινόμενη αρχιτεκτονική παράγει ενθαρρυντικά αποτελέσματα και συγκρινόμενη με μια από τις δημοφιλέστερες προσεγγίσεις αποδεικνύεται αποδοτικότερη κάτι που οφείλεται κυρίως στο χαρακτηριστικό της προσαρμοστικότητας που την καθιστούν ικανή να αντιμετωπίσει φαινόμενα απόκλισης και θορύβου στην είσοδο. Αυτό επιτυγχάνεται μέσω της ενσωμάτωσης των σχέσεων γειτνίασης μεταξύ των κόμβων θέσης που παρέχονται από το SOM, την αντιμετώπιση του προβλήματος χρονικής στρέβλωσης με την τροποποίηση του αλγορίθμου υπολο-

γισμού απόστασης Levenshtein και την διασπορά των πιθανοτήτων μετάβασης στην Μαρκοβιανή μήτρα.

4.2.5.2 Greek Sign Language Corpus

Το γλωσσικό σώμα που χρησιμοποιήθηκε ως δεύτερο σύνολο πειραματισμού ήταν το Γλωσσικό Σώμα Ελληνικής Νοηματικής (Greek Sign Language Corpus - GSLC). Όλες οι πτυχές του σώματος συμπεριλαμβανομένων των επιλογών σχεδιασμού, του καθορισμού του περιεχομένου, των συνθηκών καταγραφής, του ποιοτικού ελέγχου, της επισημείωσης, κ.λπ. περιγράφονται λεπτομερώς στο [68].

Αρχικά, αντιπαραθέτουμε την προτεινόμενη αρχιτεκτονική (Self Organizing Markov Models - SOMM) ενάντια σε τυποποιημένα, αριστερά προς δεξιά, συνεχή HMM με τον αριθμό των καταστάσεων να καθορίζονται πειραματικά ώστε να μεγιστοποιηθεί το ποσοστό αναγνώρισης τους. Σε αυτό το σημείο είναι άξιο αναφοράς πως η προτεινόμενη προσέγγιση δεν απαιτεί τέτοιες αυθαίρετες αποφάσεις σχεδιασμού όπως ο πειραματικά καθορισμένος αριθμός καταστάσεων μιας και η τοπολογία και η μήτρα πιθανοτήτων μετάβασης των Μαρκοβιανών προτύπων καθορίζεται αυτόματα κατά την εκπαίδευση χωρίς την μέθοδο δοκιμής και σφάλματος. Αυτό το σύνολο δεδομένων είναι σημαντικά πιο περίπλοκο από το πρώτο συνθετικό σύνολο δεδομένων αφού πολλά από τα νοήματα διαφέρουν μόνο στην κυρίαρχη χειρομορφή και έχουν τα ίδια χαρακτηριστικά θέσης και μετακίνησης. Τα αντίστοιχα ποσοστά αναγνώρισης για τις δύο προσεγγίσεις φαίνονται στον πίνακα 4.10.

Χαρακτηριστικά	HMM	SOMM
Θέση κυρίαρχου χεριού	61.12	73.41
Θέση και κατεύθυνση δύο χεριών	79.18	91.10

Πίνακας 4.10: Ποσοστά αναγνώρισης βάσει της θέσης του χεριού

Η ανάλυση κάθε μορφής πληροφορίας στο τελικό ποσοστό αναγνώρισης απεικονίζεται στον πίνακα 4.11, ενώ η ανάλυση των ποσοστών αναγνώρισης με HMM και χαρακτηριστικά χειρομορφής παρουσιάζεται στον πίνακα 4.13. Επιπλέον, ο πίνακας 4.12 δείχνει τα ποσοστά αναγνώρισης σε ένα υποσύνολο του γλωσσικού σώματος (πάνω από 90%).

Χαρακτηριστικά	Ποσοστό
Θέση (κυρίαρχο χέρι)	73.41
Θέση (δύο χέρια)	83.30
Θέση + Κατεύθυνση + Levenshtein	91.10
Θέση + Κατεύθυνση + Levenshtein + Χειρομορφή	97.80

Πίνακας 4.11: Ανάλυση ποσοστών αναγνώρισης ανά σύνολο χαρακτηριστικών γνωρισμάτων

Κατά τη διάρκεια μιας άτυπης αξιολόγησης του συνόλου δεδομένων από την άποψη της ευκολίας αναγνώρισης κάθε κατηγορίας, δύο ειδικοί, ένας από την κωφή κοινότητα και ο άλλος ερευνητής που έχει καλή κατανόηση της λειτουργίας της προτεινόμενης προσέγγισης, επιθεώρησαν με προσοχή όλες τις επαναλήψεις των διενεργηθέντων

νοημάτων αλλά και μια αποτύπωση των θέσεων των εντοπισμένων χεριών κατά τη διαδικασία εξαγωγής χαρακτηριστικών γνωρισμάτων. Έτσι σε κάθε νόημα ανατέθηκε μια τιμή που αντιπροσωπεύει τη δυσκολία αναγνώρισης που προκαλείται είτε από την ασυνεπή απόδοση των νοηματιστών είτε από λάθη που εισάγονται από τη διαδικασία εξαγωγής χαρακτηριστικών γνωρισμάτων. Αξίζει να σημειωθεί πως τα 12 σημάδια που αναγνωρίστηκαν αυτόματα με το χαμηλότερο ποσοστό άνηκαν στα 15 πιο δύσκολα σύμφωνα με την ταξινόμηση που προέκυψε από την προηγούμενη διαδικασία επισημείωσης από ειδικούς. Τα αντίστοιχα ποσοστά αναγνώρισης για το μειωμένο σύνολο 106 $(118 - 12)$ κατηγοριών παρουσιάζεται στον πίνακα 4.12.

Χαρακτηριστικά	Ποσοστό
Θέση (κυρίαρχο χέρι)	74.77
Θέση (δύο χέρια)	88.10
Θέση + Κατεύθυνση + Levenshtein	95.05
Θέση + Κατεύθυνση + Levenshtein + Χειρομορφή	99.54

Πίνακας 4.12: Ανάλυση ποσοστών αναγνώρισης σε ένα υποσύνολο του GSLC

Χαρακτηριστικά	Ποσοστό
Περιοχή	47.44
Fourier	36.67
Ροπές	36.82
Καμπυλότητα	26.35
Σύνολο	55.1

Πίνακας 4.13: Ανάλυση ποσοστών αναγνώρισης για την χειρομορφή [223] [224]

Επιπλέον, εξετάστηκε πως η προτεινόμενη διαδικασία συγχώνευσης αποδίδει ενάντια σε δημοφιλείς παραλλαγές HMM όπως τα Multi-Stream, Parallel and Product HMM που σχεδιάστηκαν με σκοπό την έπεξεργασία πολλαπλών ροών πληροφορίας. Τα αποτελέσματα για αυτές τις παραλλαγές HMM καθώς επίσης και όλα τα πειράματα και αποτελέσματα που εξετάζουν HMMs στο GSLC προέρχονται από το [223] [224].

Σχήμα	Ποσοστό
Multi-Stream HMM	92.27
Parallel HMM	92.45
Product HMM	93.64
SOMM	97.80

Πίνακας 4.14: Απόδοση προσεγγίσεων (αποδόσεις HMM από [223] [224]

Τα πειράματα εκτελέσθηκαν χρησιμοποιώντας Matlab σε έναν τυπικό προσωπικό υπολογιστή (2GHz, 3GB RAM), με την μέθοδο της διεπικύρωσης αφήνοντας κάθε φορά ένα δείγμα εκτός εκπαίδευσης και ο χρόνος επεξεργασίας που απαιτείται για κάθε βήμα παρουσιάζεται στον πίνακα 4.15. Η εκπαίδευση του SOM είναι η πιο απαιτητική διαδικασία από την άποψη του χρόνου επεξεργασίας, αλλά αυτή η διαδικασία εκτελείται μόλις μια φορά ανεξάρτητα από τον αριθμό κατηγοριών και χωρίς

σημαντική απώλεια ακρίβειας όσον αφορά τις ικανότητες αναπαράστασης μπορεί να εκτελεσθεί δειγματοληψία επί των σημείων των τροχιών των χεριών. Αυτή η διαδικασία θα μείωνε σημαντικά τον απαραίτητο χρόνο εκπαίδευσης SOM. Ο χρόνος που απαιτείται για το στάδιο αποκωδικοποιήσης ποικίλλει ανάλογα με το μήκος της ακολουθίας. Η προτεινόμενη αρχιτεκτονική αποδεικνύεται σημαντικά ταχύτερη από άλλες κυρίαρχες προσεγγίσεις όπως φαίνεται στον πίνακα 4.16 και καθίσταται κατάλληλη για εφαρμογές πραγματικού χρόνου. Αξίζει να σημειωθεί πως τα Product HMM, που αποδεικνύονται η αποτελεσματικότερη παραλλαγή HMM (πίνακας 4.14), απαιτεί 15 φορές μεγαλύτερο χρόνο επεξεργασίας για την αποκωδικοποίηση.

Διαδικασία	Δεξί	Αριστερό	Σύνολο
Εκπαίδευση SOM	5.6151	3.2025	8.8176
Πρότυπο θέσης	1.3345	2.0800	3.4145
Πρότυπο κατεύθυνσης	1.1311	0.9159	2.0470
Αποκωδικοποίηση	0.0655	0.0803	0.1458

Πίνακας 4.15: Απαιτούμενοι χρόνοι ανά διαδικασία (δευτερόλεπτα)

Σχήμα	Εκπαίδευση	Επαλήθευση
SOMM	14.279	0.145
Multistream HMM	28.416	0.870
Product HMM	53.938	2.280

Πίνακας 4.16: Απαιτούμενοι χρόνοι εκπαίδευσης και επαλήθευσης για SOMM, Multistream HMM και Product HMM

4.2.6 Συμπεράσματα

Στην παρούσα εργασία προτείνουμε μια αρχιτεκτονική αυτόματης αναγνώρισης νοηματικής γλώσσας σε επίπεδο λήμματος ενσωματώνοντας αυτοοργανούμενους χάρτες, αλυσίδες Markov και τροποποιημένη μετρική απόστασης Levenshtein. Τα εξαγόμενα χαρακτηριστικά γνωρίσματα εκπαιδεύουν μεμονωμένους ταξινομητές, οι οποίοι στη συνέχεια συνδυάζονται σε επίπεδο απόφασης, μια ενισχυτική τεχνική βασισμένη σε αδύναμους κατηγοριοποιητές, ενισχύοντας την προτεινόμενη αρχιτεκτονική με πολλαπλές μορφές πληροφορίας και ευρωστία έναντι σε αλλοιώσεις στιγμιοτύπων και θορυβώδη και ανεξέλεγκτα περιβάλλοντα. Η διακύμανση κατά την απόδοση των χειρονομιών αντιμετωπίζεται μέσω της ευελιξίας τόσο της διαδικασίας εκπαίδευσης όσο και της αποκωδικοποίησης που βασίζεται στο χαρακτηριστικό γειτνίασης SOM και του αλγορίθμου αναζήτησης βέλτιστης τροχιάς που εκτελείται κατά τη διάρκεια της ταξινόμησης. Επιπλέον, το υπολογιστικό κόστος και η ταχύτητα επεξεργασίας της διαδικασίας αναγνώρισης αποδεικνύει ότι η προτεινόμενη αρχιτεκτονική είναι κατάλληλη για εφαρμογές πραγματικού χρόνου αφού ικανοποιεί όλες τις απαιτήσεις που συνοδεύουν τέτοια σενάρια.

Μελλοντική αλλά και τρέχουσα εργασία περιλαμβάνει την διερεύνηση της αναγνώρισης συστατικών νοηματικής γλώσσας (signeme). Εμπνευσμένοι από το σύστημα μεταγραφής HamNoSys μια αρχιτεκτονική θα μπορούσε να συμπεράνει σχετικά με

διαφορετικές πτυχές της νοηματικής γλώσσας όπως την χειρομορφή, την κατεύθυνση και τον προσανατολισμό της παλάμης, την θέση και την μετακίνηση των χεριών, ενώ ακόμα και μη χειρωνακτικά χαρακτηριστικά γνωρίσματα όπως οι εκφράσεις του προσώπου [118], η κίνηση των χειλιών, την παρακολούθηση του βλέμματος [9] κ.λπ. καθώς και τρόποιι ενσωμάτωσης τους στην διαδικασία αναγνώρισης πρέπει να μελετηθούν δεδομένου ότι είναι εξαιρετικά κρίσιμα στη γλωσσική ανάλυση νοηματικής. Η επέκταση της παρούσας εργασίας σε συνεχή νοηματισμό θα μπορούσε να εξεταστεί απλοϊκά με την προσθήκη ενός μηχανισμού εντοπισμού νοήματος ή μιας διαδικασίας χρονικής κατάτμησης στην υπάρχουσα αρχιτεκτονική. Μια τέτοια προσέγγιση στο παρελθόν έχει αποδειχθεί ανεπαρκής και η ανάγκη συνυπολογισμού της γλωσσικής και γραμματικής ανάλυσης έχει γίνει όλο και περισσότερο προφανής. Η ενσωμάτωση γλωσσικής γνώσης είτε στη διαδικασία συγχώνευσης ροών πληροφορίας για αναγνώριση μεμονωμένων λημμάτων είτε στην αναγνώριση σε επίπεδο πρότασης θα ενίσχυε σημαντικά την ευρωστία της διαδικασίας. Η προσθήκη αυτού του τελικού επιπέδου γλωσσικών ή γραμματικών φαινομένων υποβοηθούμενου από γνώση (knowledge assisted). Τέλος, η διερεύνηση πιθανών βελτιστοποιήσεων του αλγορίθμου, η πρόβλεψη νοημάτων και χειρονομιών πριν την ολοκλήρωση τους και δενδρικές δομές απόφασης αποτελούν μελλοντικές κατευθύνσεις της προτεινόμενης αρχιτεκτονικής.

4.3 Σύνθεση Ελληνικής Νοηματικής Γλώσσας

Η Ελληνική Νοηματική Γλώσσα (ΕΝΓ) είναι μια φυσική οπτική γλώσσα που χρησιμοποιείται από τα μέλη της Ελληνικής Κοινότητα Κωφών, που αριθμεί μερικές χιλιάδες μέλη, εκ γενετής ή όχι, νοηματιστές. Η έρευνα για τη γραμματική της ΕΝΓ είναι καθ' εαυτή περιορισμένη, ενώ κάποια μελέτη έχει γίνει ήδη σε μεμονωμένες πτυχές του συντακτικού της, καθώς επίσης και στην εφαρμοσμένη και εκπαιδευτική γλωσσολογία. Είναι ασφαλές να πει κανείς ότι ΕΝΓ στην παρούσα μορφή της, είναι ένας συνδυασμός του παλαιότερου τύπου ελληνικών νοηματικών γλωσσικών διαλέκτων με επιρροές από την Γαλλική νοηματική γλώσσα. Εστιάζοντας στον πυρήνα του λεξιλογίου παρατηρούμε μεγάλες ομοιότητες με νοηματικές γλώσσες γειτονικών χωρών, ενώ μοιράζεται και κοινές διαγλωσσικές τάσεις και μορφολογία συντακτικού όπως έχουν δείξει μελέτες πάνω στις γλώσσες αυτές [17] [149].

Η ΕΝΓ έχει αναπτυχθεί σε ένα κοινωνικό και γλωσσολογικό πλαίσιο παρόμοιο με τις περισσότερες νοηματικές γλώσσες. Χρησιμοποιείται ευρέως στην ελληνική κοινότητα κωφών και οι εκ γενετής χρήστες ΕΝΓ υπολογίζονται στους 40.600. Υπάρχει επίσης ένας μεγάλος αριθμός ακουόντων νοηματιστών ΕΝΓ, κυρίως σπουδαστές ΕΝΓ και συγγενείς κωφών. Αν και ο ακριβής αριθμός ακουόντων σπουδαστών ΕΝΓ στην Ελλάδα είναι άγνωστος, η πιο πρόσφατη απογραφή της Ομοσπονδίας Κωφών Ελλάδας (ΟΜΚΕ) δείχνει ότι, στο έτος 2003 περίπου 300 άνθρωποι εγγράφηκαν σε μαθήματα ΕΝΓ ως δεύτερη γλώσσα. Η πρόσφατη αύξηση των κωφών σπουδαστών στην βασική εκπαίδευση, καθώς επίσης και ο αριθμός των κωφών σπουδαστών που φοιτούν σε άλλα ιδρύματα, εκπαιδευτικές μονάδες απομακρυσμένων πόλεων και στην ιδιωτική εκπαίδευση μπορεί να οδηγήσουν σε διπλασιασμό του συνολικού αριθμού δευτερευόντων χρηστών ΕΝΓ στην Ελλάδα. Επίσημα ιδρύματα όπου χρησιμοποιείται η ΕΝΓ περιλαμβάνουν 11 συλλόγους κωφών στα ελληνικά αστικά κέντρα και ένα 14 εκπαιδευτικά ιδρύματα κωφών.

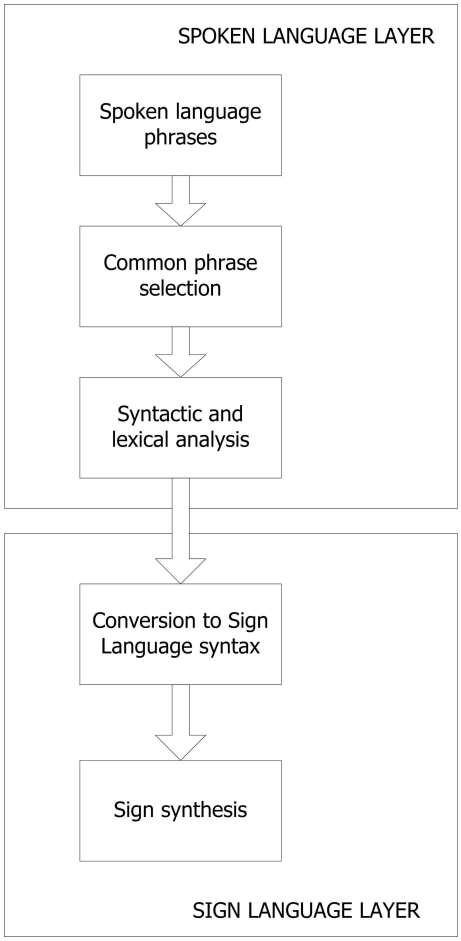
Οι περισσότεροι άνθρωποι, που δεν έχουν προσωπικές ή κοινωνικές επαφές με κωφούς, έχουν την τάση να πιστεύουν ότι οι νοηματικές γλώσσες είναι ένα είδος

παντομίμας ή αναπαράστασης κάποιας από τις φωνούμενες γλώσσες που μιλάνε οι ίδιοι. Η αλήθεια όμως είναι πολύ διαφορετική από αυτή την ευρέως διαδεδομένη αντίληψη. Οι νοηματικές γλώσσες, που είναι πολλές και εντελώς διαφορετικές μεταξύ τους, διαφέρουν από τις υπόλοιπες φυσικές γλώσσες ως προς το ότι μας είναι λιγότερο γνωστές, όχι όμως και ως προς τις γλωσσολογικές αρχές που διέπουν την επικοινωνιακή τους λειτουργία.

Η Ελληνική Νοηματική Γλώσσα (ΕΝΓ) είναι η φυσική γλώσσα της κοινότητας των Κωφών στην Ελλάδα. Όπως συμβαίνει και με τις υπόλοιπες νοηματικές, η ιδιαιτερότητά της σε σχέση με αυτό που ο περισσότερος κόσμος έχει συνηθίσει να ονομάζει 'γλώσσα' είναι ότι η γραμματική της, δηλαδή το σύστημα των κανόνων βάσει των οποίων διαρθρώνεται ο λόγος και επιτυγχάνεται η επικοινωνία, δεν είναι προφορικό αλλά οπτικο-κινησιακό. Η ΕΝΓ λέγεται 'ελληνική' γιατί χρησιμοποιείται στην Ελλάδα από Έλληνες νοηματιστές, αυτό όμως δεν σημαίνει σε καμία περίπτωση ότι απεικονίζει την ελληνική γλώσσα ή ότι προέρχεται από αυτήν. Αντίθετα, πρόκειται για ένα αυτόνομο γλωσσικό σύστημα που μπορεί να μελετηθεί και να αναλυθεί όπως και κάθε άλλη φυσική γλώσσα.

Το προτεινόμενο, σε αυτή την εργασία, σύστημα σύνθεσης νοημάτων, που βασίζεται στην γλώσσα μοντελοποίησης εικονικού κόσμου (Virtual Reality Modeling Language), είναι ενσωματωμένο σε εκπαιδευτική πλατφόρμα και σκοπεύει να υποστηρίξει νεαρούς μαθητές των πρώτων τάξεων του σχολείου να αποκτήσουν το κατάλληλο γλωσσικό υπόβαθρο ώστε να απορροφήσουν καλύτερα περαιτέρω εκπαιδευτικό υλικό. Η προσέγγιση που επιλέχθηκε δίνει στους σπουδαστές την δυνατότητα της συστηματικής και δομημένης εκμάθησης ΕΝΓ, είτε με την μορφή απομακρυσμένων ιδιαίτερων μαθημάτων είτε με την μορφή εικονικών μαθημάτων μέσω ασύγχρονης διδασκαλίας και ο σχεδιασμός της είναι συμβατός με τις αρχές της προσβάσιμης από απόσταση εκμάθησης. Εκτός από τη διδασκαλία ΕΝΓ σαν πρωτεύουσα γλώσσα, στην παρούσα μορφή η πλατφόρμα μπορεί να χρησιμοποιηθεί για εκμάθηση γραπτών ελληνικών κειμένων μέσω ΕΝΓ, ενώ δυνητικά θα μπορούσε να βρεί εφαρμογή σε άλλους τομείς του σχολικού προγράμματος σπουδών. Η διαδικασία διδασκαλίας γραπτών ελληνικών κειμένων στους νέους κωφούς μαθητές είναι εξαιρετικής σπουδαιότητας αν λάβουμε υπόψη τη δυσκολία των κωφών ανθρώπων να συσχετίσουν τις έννοιες και με τις γραπτές μορφές τους. Αυτό συμβαίνει επειδή η γραπτή αναπαράσταση μιας έκφρασης είναι μια σύμβαση για την αναπαράσταση των ήχων, η οποία είναι ακατανόητη στην περίπτωση όπου καμία αντίληψη για τον ήχο δεν είναι δυνατή. Σύμφωνα με στατιστικές από το ελληνικό παιδαγωγικό ινστιτούτο [139], η μέση αναγνωστική ικανότητα των ενηλίκων κωφών αντιστοιχεί στο μέσο επίπεδο μαθητών πρωτοβάθμιας εκπαίδευσης.

Ο εικονικός νοηματιστής εκτελεί το αποτέλεσμα μιας διαδικασίας μετατροπής κείμενο-σε-νόημα. Περιγράφουμε τις διαδικασίες που ακολουθούνται κατά τη διάρκεια της σύνταξης από το εκπαιδευτικό υλικό και την υλοποίηση της ενότητας σύνθεσης της εκπαιδευτικής πλατφόρμας. Σε αυτήν την διαδικασία χρησιμοποιήσαμε υπάρχοντα τμήματα λογισμικού για την διαδικτυακή ενσάρκωση ενός h–anim εικονικού χαρακτήρα. Η υιοθέτηση ευρέως αποδεκτών προτύπων ορισμού και εμψύχωσης εικονικών χαρακτήρων καθιστούν δυνατή την επαναχρησιμοποίηση και επεκτασιμότητα των πόρων του συστήματος και του περιεχόμενού του. Η εικόνα 4.14 περιγράφει την εποπτική αρχιτεκτονική του συστήματος και την ροή πληροφοριών μεταξύ των συστατικών λογισμικού.



Σχήμα 4.14: Εποπτική εικόνα της προτεινόμενης αρχιτεκτονικής

4.3.1 Γλωσσολογικά θέματα

Η υλοποίηση τόσο εφαρμογών απομακρυσμένης μάθησης όσο και εργαλείων περιληπτικής παρουσίασης γραπτού κειμένου της εκπαιδευτικής πλατφόρμας απαιτεί τη συλλογή εκτενών ηλεκτρονικών γλωσσικών πόρων για την ΕΝΓ όσον αφορά στο λεξιλόγιο και των δομικών συντακτικών κανόνων της γλώσσας [205]. Τα πειραματικά δεδομένα της μελέτης είναι βασισμένα στη βασική έρευνα για την γλωσσολογική ανάλυση ΕΝΓ που επιχειρείται από το 1999 καθώς επίσης και για την εμπειρία που αποκτήθηκε από τα προγράμματα NOEMA και PROKLISI [67]. Τα δεδομένα αποτελούνται από τις ψηφιοποιημένες γλωσσικές καταγραφές των εκ γενετής κωφών νοηματιστών ΕΝΓ και των υπαρχουσών βάσεων δεδομένων των δίγλωσσων λεξιλογίων ΕΝΓ, επικυρωμένα από την Ελληνική Ομοσπονδία Κωφών. Η ανάπτυξη των γλωσσικών πόρων ακολουθεί κατάλληλες μεθοδολογικές αρχές στη συλλογή και ανάλυση δεδομένων που απεικονίζουν την ιδιαιτερότητα της γλώσσας.

Πολλοί από τους γραμματικούς κανόνες ΕΝΓ προέρχονται από την ανάλυση ενός ψηφιακού σώματος που έχει δημιουργηθεί καταγράφοντας σε βίντεο εκ γενετής νοηματιστές σε περιβάλλον συζήτησης ή κατά τον νοηματισμό μιας αφήγησης. Αυτή η διαδικασία απαιτείται, επειδή υπάρχει περιορισμένη προηγούμενη επίσημη ανάλυση ΕΝΓ και η εξαγωγή κανόνων πρέπει να βασιστεί σε πραγματικές παραγωγές δεδομένων εκ γενετής νοηματιστών. Ο σχεδιασμός του συστήματος, εκτός από το εκπαιδευτικό περιεχόμενο που υποστηρίζει προς το παρόν, εστιάζει στη δυνατότητα να παραχθούν

οι νοηματικές φράσεις, οι οποίες τηρούν τους γραμματικούς κανόνες ΕΝΓ σε έναν βαθμό ακρίβειας που επιτρέπει την αναγνώριση τους από εκ γενετής νοηματιστές ως σωστές εκφράσεις της γλώσσας.

Στη παρούσα φάση υλοποίησης, το υποσύστημα για τη διδασκαλία της γραμματικής ΕΝΓ καλύπτει περιορισμένο λεξιλόγιο και γραμματική ικανή να αναλύσει έναν περιορισμένο αριθμό γραμματικών φαινομένων ΕΝΓ. Η σύνθεση ΕΝΓ απαιτεί την ανάλυση των λημμάτων ΕΝΓ στα φωνολογικά μέρη τους και τη σημασιολογία τους. Συμφωνήθηκε ότι μόνο μονομορφικά λήμματα που χρησιμοποιούν μόνο μία χειρομορφή επρόκειτο να αναλυθούν αρχικά, έτσι ώστε η αξιολόγηση από την τεχνική ομάδα θα καθόριζε περαιτέρω βήματα. Σε δεύτερο στάδιο, πιο περίπλοκες ακολουθιακές δομές νοημάτων εξετάζονται (π.χ. νοήματα σύνθετης λέξης) και μόλις μετασχηματιστούν τα μεμονωμένα νοήματα και αποθηκευτούν σε βάση δεδομένων, θα ερευνηθεί πώς πρόσθετα επίπεδα όπως τα βασικά μη χειρωνακτικά χαρακτηριστικά γνωρίσματα μπορούν να προστεθούν με τη λιγότερη τεχνική δυσκολία.

Επιπλέον, μια ενδιαφέρουσα παράμετρος ενός εικονικού νοηματιστή είναι η δυνατότητα δακτυλοσυλλαβισμού (fingerspelling). Τα νοήματα δεν πρέπει να συγχέονται με το δακτυλικό αλφάβητο, το οποίο είναι απλώς ένας τρόπος μεταγραφής του ελληνικού αλφαβήτου. Αυτή η τεχνική είναι χρήσιμη σε περιπτώσεις μετασχηματισμού ουσιαστικών, ακρωνύμιων, ορολογίας ή γενικά όρων για τους οποίους δεν υπάρχει κάποιο συγκεκριμένο νόημα. Ο δακτυλοσυλλαβισμός χρησιμοποιείται εκτενώς σε άλλες νοηματικές γλώσσες όπως η αμερικανική νοηματική γλώσσα (ASL) ή η βρετανική νοηματική γλώσσα (BSL), ενώ τα στοιχεία μας στην ΕΝΓ δείχνουν ότι χρησιμοποιείται μόνο περιστασιακά και σπάνια ενσωματώνεται στον πυρήνα της γλώσσας. Από τεχνικής άποψης, είναι αρκετά απλό για ένα εικονικό χαρακτήρα να νοηματίσει γράμματα δεδομένου ότι αυτή η διαδικασία δεν περιλαμβάνει κανένα συντακτικό, μετακίνηση στον νοηματικό χώρο ή χρήση μη χειρωνακτικών στοιχείων γραμματικής. Προηγούμενες προσπάθειες εμψύχωσης της νοηματικής γλώσσας περιορίζονταν σε αυτό το επίπεδο ή έφταναν στην εμψύχωση διαδοχικών δομών της γραπτής ή προφορικής γλώσσας. Από τότε σημειώθηκε ανάπτυξη στην γλωσσική περιγραφή των γλωσσικών δομών νοηματικής. Από την άλλη, λίγοι κωφοί στην Ελλάδα χρησιμοποιούν νοηματισμό γραμμάτων. Για τον λόγο αυτόν η παρούσα εργασία στοχεύει να σχηματίσει μια μορφή της ΕΝΓ όσο πιο κοντά στη φυσική ρέουσα χρήση της και χρησιμοποιεί περιστασιακά τον νοηματισμό γραμμάτων, π.χ. σε γλωσσικά παιχνίδια, που εστιάζει στην διδασκαλία των γραπτών ελληνικών κειμένων.

Το εργαλείο που υιοθετήθηκε για τη μεταγραφή και τη σημειογραφία των λεξικολογικών νοημάτων είναι το HamNoSys, ένα πικτογραφικό σύστημα σημειογραφίας που αναπτύχθηκε από το πανεπιστήμιο του Αμβούργο για την περιγραφή της φωνολογίας των νοημάτων [195]. Αυτή η σημειογραφία χαρακτηρίζει μεμονωμένα λήμματα ΕΝΓ, ενώ για τον χαρακτηρισμό διαδοχικών δομών, σε επίπεδο φράσης, υιοθετήθηκε ο γλωσσικός σχολιαστής ELAN [78]. Θεωρήσαμε αυτά τα δύο συστήματα ως τα καταλληλότερα για τον μετασχηματισμό κειμένου σε λήμματα της νοηματικής, κρίνοντας από τεχνογνωσία σε πρόσφατα προγράμματα. Το κλασικό πρότυπο Stokoe χρησιμοποιείται για τη μορφοφωνολογική περιγραφή, με ένα επιπλέον επιπέδο με τα σημασιολογικά ισοδύναμα των εκφράσεων. Στόχος είναι να προστεθούν περισσότερα επίπεδα για την καλύτερη περιγραφή χαρακτηριστικών όπως μη χειρωνακτικά χαρακτηριστικά γνωρίσματα και πραγματολογία. Το Sign-writing ήταν ένα άλλο εργαλείο μεταγραφής υπό εξέταση, αλλά δεν επιλέχτηκε, αφού το HamNoSys αναμένεται να ενσωματωθεί, ως επίπεδο, στο προκαθορισμένο περιβάλλον του ELAN στο εγγύς μέλ-

λον.

Τα γλωσσικά μέσα που χρησιμοποιεί η ΕΝΓ (όπως και άλλες νοηματικές γλώσσες) για να διατυπώσει τις έννοιες και για να δημιουργήσει μορφολογία και σύνταξη, βασίζονται στην κίνηση των χεριών, στην στάση ή στην κίνηση του σώματος και στην έκφραση του προσώπου. Οι βασικές μονάδες του λόγου (τις οποίες η επιστήμη της γλωσσολογίας ονομάζει γλωσσικά σημεία) της ΕΝΓ ονομάζονται νοήματα. Τα νοήματα μπορούν να έχουν λεξική ή γραμματική σημασία, ακριβώς όπως τα μορφήματα και οι λέξεις στις φυσικές γλώσσες.

Το χαρακτηριστικότερο συστατικό ενός νοήματος λέγεται χειρομορφή. Η χειρομορφή είναι το σχήμα που παίρνει η παλάμη και η θέση στην οποία τοποθετούνται τα δάκτυλα τη στιγμή που αρχίζει να σχηματίζεται ένα νόημα. Η ίδια η χειρομορφή όμως από μόνη της δεν είναι φορέας σημασίας. Για να αποκτήσει σημασία, για να δημιουργηθεί δηλαδή ένα νόημα, η χειρομορφή πρέπει να συνοδεύεται και από τα παρακάτω στοιχεία:

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

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

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

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

4.3.2 Τεχνικά Θέματα

Από την πληθώρα διαθέσιμων τεχνολογιών εικονικών χαρακτήρων και μεθόδων εμψύχωσης τους για την αναπαράσταση της νοηματικής γλώσσας υιοθετήθηκε μια από τις πιο προεξέχουσες τεχνολογικές λύσεις. Οι μετατοπίσεις ενός συνθετικού, τρισδιάστατου μοντέλου νοηματιστή πρέπει να καταγραφούν σε ένα υψηλό και επαναχρησιμοποιήσιμο επίπεδο περιγραφής, προτού μετασχηματιστούν σε παραμέτρους της κίνησης

μερών του σώματος. Στο πεδίο εμψύχωσης νοημάτων έχουν υπάρξει αρκετές παρόμοιες προσπάθειες (VISICAST, Thetos, SignSynth και VSIGN) που χρησιμοποιήσαμε ως υπόβαθρο.

4.3.2.1 Το πρότυπο h–anim

Η έκρηξη στον τομέα των υπολογιστικών δυνατοτήτων και την δικτυακή προσβασιμότητα έχει προκαλέσει ανανεωμένο ενδιαφέρον στο πεδίο των τρισδιάστατων γραφικών τα τελευταία χρόνια, με αποτέλεσμα την σταθερή εμφάνιση εφαρμογών που αφορούν τη προτυποποίηση και εμψύχωση τρισδιάστατων ανθρώπινων μορφών. Ένα σύνηθες μειονέκτημα των εφαρμογών αυτών είναι η έλλειψη δυνατοτήτων επαναχρησιμοποίησης και μεταφερσιμότητας. Η απουσία τυποποιημένης σκελετικής αναπαράστασης οδηγεί συνήθως στην ανάπτυξη αυθαίρετων λύσεων που δεν βοηθούν στην ομαλή μετάβαση μεταξύ συστημάτων λογισμικού.

Το πρότυπο ISO h–anim (ISO/IEC 19774) παρέχει μια συστηματική προσέγγιση στην αναπαράσταση ανθρωποειδών μοντέλων σε ένα τρισδιάστατο, πολυμεσικό περιβάλλον. Στο πλαίσιο αυτό, κάθε ανθρωποειδές διαμορφώνεται αφαιρετικά από την άποψη της δομής όπως ένας αρθρωμένος χαρακτήρας, ο οποίος μπορεί να ενσωματωθεί σε διαφορετικά συστήματα αναπαράστασης και να εμψυχωθεί χρησιμοποιώντας τις λειτουργικότητες που παρέχονται από το σύστημα. Βάσει των παραπάνω, το πρότυπο h–anim καθορίζει την εμψύχωση ως μια βασισμένη στον χρόνο, τρισδιάστατη, διαδραστική, λειτουργική συμπεριφορά πολυμεσικών, σαφώς ορισμένων χαρακτήρων, που αφήνοντας τον προσδιορισμό της γεωμετρίας στον χρήστη. Στόχοι του προτύπου είναι η συμβατότητα, οδηγώντας στην ύπαρξη μοντέλων που μπορούν να υλοποιηθούν σε οποιαδήποτε εφαρμογή, η ευελιξία, διαχωρίζοντας τον ορισμό του μοντέλου από την συμπεριφορά του, η απλότητα και η επεκτασιμότητα. Για να επιτευχθούν οι στόχοι αυτοί το πρότυπο επιτρέπει την άμεση πρόσβαση στην ιεραρχία των αρθρώσεων του σκελετού αλλά και στην γεωμετρία που συνιστούν τα ξεχωριστά μέλη του σώματος με τέτοιο τρόπο που να κάνει διακριτό τον ορισμό του μοντέλου και την εμψύχωση του.

Η σκελετική περιγραφή ενός μοντέλου h–anim αποτελείται από ένα δέντρο οντοτήτων, που αντιστοιχούν στις αρθρώσεις του ανθρώπινου σώματος, οι οποίες καθορίζουν τους μετασχηματισμούς από την άρθρωση ρίζας (HumanoidRoot) έως το φύλλο της ιεραρχίας του δέντρου. Μόνο η άρθρωση HumanoidRoot είναι απαραίτητη, ενώ όλα τα άλλα αντικείμενα είναι προαιρετικά. Βέβαια όσα περισσότερα κοινά αντικείμενα βρίσκονται σε ένα ορισμό του μοντέλου, τόσο πιό εύκαμπτο γίνεται το μοντέλο κατά την διάρκεια της εμψύχωσης. Κατά συνέπεια, το πρότυπο καθορίζει προκαθορισμένα σύνολα αρθρώσεων ως επίπεδα άρθρωσης (levels of articulation LOA): ένα σύνολο με δεκατέσσερις αρθρώσεις ορίζεται ως ένα χαμηλό επίπεδο άρθρωσης, ενώ ένα πρότυπο ανθρωποειδούς με 72 αρθρώσεις χαρακτηρίζεται ως υψηλό επίπεδο άρθρωσης.

Όσον αφορά τη διαδικασία μοντελοποίησης, τα ανθρωποειδή h–anim δημιουργούνται έχοντας κατά νου τον μέσο άνθρωπο (κατά προσέγγιση ύψος 1,75m) και με τέτοια στάση σώματος ώστε να στρέφεται προς τον άξονα +Z, με τον άξονα +Y πάνω και τον άξονα +X στα αριστερά του ανθρωποειδούς. Η αρχή των αξόνων βρίσκεται στο επίπεδο του δαπέδου, μεταξύ των ποδιών του ανθρωποειδούς. Στη προεπιλεγμένη θέση σώματος, τα χέρια του είναι ευθεία και παράλληλα στα πλευρά του σώματος με τις παλάμες των χεριών να στρέφονται προς το σώμα, ενώ τα χαρακτηριστικά του προσώπου περιλαμβάνουν τα φρύδια σε ανάπαυση, το στόμα κλειστό και τα μάτια διάπλατα ανοικτά.

Για την καταγραφή και τον ορισμό της χειρομορφής και των χειρονομιών, τεχνικές καταγραφής κίνησης και απτικές συσκευές (όπως CyberGrasp ή γάντια με αισθητήρες επιτάχυνσης) εξετάστηκαν αρχικά εντούτοις, συμφωνήθηκε ότι, εάν τα σύμβολα σημειογραφίας HamNoSys παρείχαν αποδεκτή ποιότητα οι τεχνικές αυτές δεν θα ακολουθηθούν. Σε κάθε περίπτωση πάντως η σημασιολογική επισημείωση είναι μια πολύ πιο εύκαμπτη και επαναχρησιμοποιήσιμη λύση από τα αρχεία βίντεο ή την καταγραφή κίνησης, δεδομένου ότι ένας h–anim εικονικός χαρακτήρας μπορεί να εκμεταλλευτεί την δυναμική φύση των φωνολογικών και συντακτικών κανόνων της νοηματικής γλώσσας.

4.3.2.2 Υλοποίηση

Η αλληλεπίδραση του σχεδιαστή περιεχομένου με τον εικονικό χαρακτήρα επιτυγχάνεται μέσω μιας γλώσσας σεναρίου (scripting). Στην εφαρμογή μας, επιλέξαμε τη γλώσσα STEP (Scripting Technology for Embodied Persona) [113] ως ενδιάμεσο επίπεδο μεταξύ του τελικού χρήστη και του εικονικού χαρακτήρα. Ένα σημαντικό πλεονέκτημα που διαθέτουν αυτού του τύπου οι γλώσσες είναι ότι διαχωρίζουν την περιγραφή μεμονωμένων νοημάτων από τον ορισμό της γεωμετρίας και της ιεραρχία του εικονικού χαρακτήρα, κατά συνέπεια κάποιος μπορεί να αλλάξει την περιγραφή κάποιου νοήματος, χωρίς απαραίτητα να αναδιαμορφωθεί ο εικονικός νοηματιστής. Το μοντέλο εικονικού χαρακτήρα που χρησιμοποιήθηκε είναι συμβατό με το h–anim πρότυπο, έτσι μπορεί εύκολα να να αντικατασταθεί με οποιοδήποτε από τα ευρέως διαθέσιμα μοντέλα ή να οριστεί κάποιο νέο.

Η εμψύχωση με σενάριο (scripted animation) είναι μια μεταφέρσιμη και επεκτάσιμη εναλλακτική λύση εμψύχωσης έναντι της εμψύχωσης βασισμένη σε τεχνικές καταγραφής κίνησης. Κάποιος μπορεί να παρομοιώσει τη σχέση μεταξύ αυτών των δύο προσεγγίσεων με αυτή μεταξύ συνθετικής αναπαράστασης και σε αναπαράσταση μέσω βίντεο αρχείων. Η καταγραφή κίνησης μπορεί να είναι εξαιρετικά λεπτομερής όσον αφορά την ποσότητα και το βάθος των πληροφοριών, αλλά είναι δύσκολο να προσαρμοστεί κατά την αναπαραγωγή και απαιτεί τεράστιους χώρους αποθήκευσης και εύρος ζώνης καναλιού στην μετάδοση. Από την άλλη, η εμψύχωση με σενάριο απαιτεί συνήθως χειρωνακτική επέμβαση κατά την σύνταξη και κατά συνέπεια ο τρόπος αναπαράστασης είναι απλός και αφηρημένος. Έτσι, τα σενάρια αυτά απαιτούν μερικές εκατοντάδες χαρακτήρες για να περιγράψει ένα λήμμα και μπορεί να επαναχρησιμοποιηθεί για να παραγάγει παραλλαγές του ίδιου νοήματος [90]. Αυτό φαίνεται στο τμήμα κώδικα στην εικόνα 4.15, η οποία επεξηγεί τους απαιτούμενους μετασχηματισμούς για τις αρθρώσεις του δεξιού χεριού για να εμψυχώσει την χειρομορφή D. Όπως γίνεται εύκολα αντιληπτό, το ίδιο τμήμα κώδικας μπορεί να παράγει τους μετασχηματισμούς του αριστερού χεριού για την ίδια χειρομορφή αν εφαρμοσθεί μια τεχνική καθρεπτισμού, ενώ μια πιο περίπλοκη χειρομορφή μπορεί να εκκινήσει με αυτήν την αναπαράσταση και να τροποποιηθούν μόνο μερικά συστατικά της.

Ένας συντακτικός αναλυτής [21] αποκωδικοποιεί τα δομικά σχέδια των γραπτών ελληνικών και τα αντιστοιχεί στα ισοδύναμα συστατικά της ΕΝΓ. Αυτά τροφοδοτούνται στο αυτοματοποιημένο σύστημα που αποκωδικοποιεί τις ακολουθίες σημειογραφίας HamNoSys για κάθε λήμμα. Το σύστημα αυτό ουσιαστικά μετασχηματίζει τα μεμονωμένα ή συνδυασμένα σύμβολα HamNoSys σε ακολουθίες εντολών προς τον εικονικό νοηματιστή. Μια ενδεικτική ακολουθία σημειογραφίας HamNoSys αποτελείται από σύμβολα που περιγράφουν τη διαμόρφωση αφετηρίας ενός νοήματος και τις ενέργειες που ακολουθούν. Τα σύμβολα που περιγράφουν την αρχική διαμόρφωση

```
par([
turn(humanoid,r_thumb1,rotation(1.9,1,1.4,0.6),very_fast),
turn(humanoid,r_thumb2,rotation(1,0.4,2.2,0.8),very_fast),
turn(humanoid,r_thumb3,rotation(1.4,0,0.2,0.4),very_fast),
turn(humanoid,r_index1, rotation(0,0,0,0),very_fast),
turn(humanoid,r_index2,rotation(0,0,0,0),very_fast),
turn(humanoid,r_index3,rotation(0,0,0,0),very_fast),
turn(humanoid,r_middle1,rotation(0,0,1,1.5999),very_fast),
turn(humanoid,r_middle2,rotation(0,0,1,1.5999),very_fast),
turn(humanoid,r_middle3,rotation(0,0,1,1.5999),very_fast),
turn(humanoid,r_ring1,rotation(0,0,1,1.7999),very_fast),
turn(humanoid,r_ring2,rotation(0,0,1,1.5999),very_fast),
turn(humanoid,r_ring3,rotation(0,0,1,0.6000),very_fast),
turn(humanoid,r_pinky1,rotation(0,0,1,1.9998),very_fast),
turn(humanoid,r_pinky2,rotation(0,0,1,1.5999),very_fast),
turn(humanoid,r_pinky3,rotation(0,0,1,0.7998),very_fast)
])
```

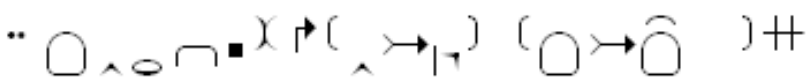
Σχήμα 4.15: Κώδικας STEP για μία χειρομορφή

αναφέρονται στην χειρομορφή που χρησιμοποιείται κατά τη διάρκεια του νοήματος και την αρχική θέση και τον προσανατολισμό του χεριού που εκτελεί το νοήμα. Εάν και το δευτερεύον χέρι συμμετέχει στο νόημα, όπως συμβαίνει στην περίπτωση του νοήματος 'Γιατρός', σημασία έχει η σχετική θέση των δύο χεριών, π.χ. 'το κύριο χέρι αγγίζει τον αγκώνα του δευτερεύοντος χεριού'. Επιπλέον πληροφορίες περιλαμβάνουν την συμμετρία, εάν και τα δύο χέρια ακολουθούν το ίδιο πρότυπο μετακίνησης και άλλα μη χειρωνακτικά συστατικά. Η εικόνα 4.17 παρουσιάζει ένα στιγμιότυπο νοηματισμού του λήμματος 'Γάιδαρος'. Το μοντέλο του εικονικού χαρακτήρα είναι το 'yt', από τον Matthew T. Beitler και διαθέσιμο στην διεύθυνση http://www.cis.upenn.edu/~beitler. Μια δοκιμαστική έκδοση του συστήματος με περιορισμένο λεξιλόγιο και μερικά παραδείγματα φράσεων βρίσκεται στον δικτυακό τόπο http://www.image.ece.ntua.gr/~gcari/gslv.

Η εικόνα 4.18 παρουσιάζει την ακολουθία HamNoSys για ένα συγκεκριμένο νόημα, που εμφανίζεται στην κορυφή της διεπαφής εικονικού νοηματιστή. Το πρώτο σύμβολο δείχνει ότι και τα δύο χέρια εκτελούν την ίδια μετακίνηση, εκκινώντας από συμμετρικές αρχικές, σχετικές με τον κορμό του νοηματιστή, θέσεις. Το δεύτερο σύμβολο δείχνει την χειρομορφή που χρησιμοποιείται αρχικά, η οποία είναι μια ανοικτή παλάμη, αποκαλούμενη χειρομορφή D στην ΕΝΓ, ενώ το επόμενο σύμβολο υποδεικνύει τον προσανατολισμό της παλάμης. Τα υπόλοιπα σύμβολα αναφέρονται στην αρχική θέση της παλάμης, που στην περίπτωση μας σχεδόν αγγίζει την κορυφή του κεφαλιού. Σύμβολα που περιλαμβάνονται εντός παρενθέσεων περιγράφουν σύνθετες μετακινήσεις, ενώ ο τελευταίος χαρακτήρας δηλώνει οτι η κίνηση είναι επαναλαμβανόμενη.

Η εικόνα 4.20 παρουσιάζει τον εικονικό νοηματιστή να εμψυχώνει το λήμμα 'παιδί', ενώ η εικόνα 4.22παρουσιάζει ένα στιγμιότυπο του νοήματος 'παιδιά'. Η υιοθέτηση της προσέγγισης του αυτοματοποιημένου συστήματος παραγωγής εμψυχώσεων μας επιτρέπει να χρησιμοποιήσουμε την περιγραφή του νοήματος (εικόνα 4.21) για να κατασκευάσουμε τον πληθυντικό του. Σε αυτήν την περίπτωση, η μορφή πληθυντικού παρουσιάζεται με την επανάληψη του ίδιου νοήματος, μετακινώντας σε κάθε επανάληψη το χέρι ελαφρώς προς το δεξιά του νοηματιστή. Η κατεύθυνση υποδεικνύεται

Σχήμα 4.16: Στιγμιότυπο του ψηφιακού νοηματιστή κατά την διάρκεια του νοήματος ΓΑΙΔΑΡΟΣ



Σχήμα 4.17: Η συμβολοσειρά HamNoSys για το νόημα ΓΑΙΔΑΡΟΣ

από το σύμβολο που προηγείται της παρένθεσης, ενώ το περιεχόμενο της παρένθεσης περιγράφει αυτήν την δευτερεύουσα μετακίνηση. Κατά συνέπεια, ο αναλυτής περιορίζεται στο να υποδείξει μόνο την τροποποίηση του αρχικού νοήματος για να παράγει την πληθυντική έκδοση του λήμματος. Στην ΕΝΓ, οι μεμονωμένες περιπτώσεις νοημάτων είναι περιορισμένες, ενώ η επαναχρησιμοποίηση, με μικρές παραλλαγές, του ίδιου νοήματος-βάσης συναντάται συχνά, επιτρέποντας την εύρεση αποδοτικών κανόνων παραγωγής, όπως αυτός που περιγράφεται παραπάνω. Μια άλλη δυνατότητα είναι να αλλαχτεί η χειρομορφή για ένα νόημα, ειδικά όταν ο νοηματιστής θέλει να προσδιορίσει μια αριθμητική ποσότητα. Η εικόνα 4.23 παρουσιάζει τον εικονικό νοηματιστή να εκτελεί την έκδοση ΕΝΓ της 'ημέρας', ενώ η εικόνα 4.24 παρουσιάζει την έκδοση ΕΝΓ 'δύο ημερών'. Η διαφορά στην τελευταία περίπτωση συνίσταται στην χρήση διαφορετικής χειρομορφής, χειρομορφή 'δύο-δάχτυλά', αντί αυτής του 'ευθύ-δείκτη', για να εκτελέσει την ίδια μετακίνηση, που αρχίζει από την ίδια αρχική θέση. Αυτή η διαφορά είναι εμφανέστερη σε στην εικόνα 4.25, η οποία παρουσιάζει την πρόσοψη του εικονικού νοηματιστή. Αυτό είναι ένα πραγματικά ενδιαφέρον χαρακτηριστικό γνώρισμα του λογισμικού Blaxxun Contact 5 [19], που χρησιμοποιεί το σύστημα. Παρά το γεγονός ότι η προεπιλογμένη άποψη είναι αυτή που έχει ορίσει ο σχεδιαστής, ο χρήστης έχει την δυνατότητα να να επιλέξει διαφορετική όψη, όπως μετωπική και πλάγια όψη του νοήματος, χαρακτηριστικό ιδιαίτερα κρίσιμο στα μαθησιακά περιβάλλοντα, δεδομένου ότι φροντίζει για την επίδειξη των διαφορών μεταξύ παρόμοιων νοημάτων

Σχήμα 4.18: Η ΕΝΓ έκδοση του νοήματος ΠΑΙΔΙ

```
par([
turn(humanoid,r_elbow,rotation(-0.722,0.2565,0.1206,1.5760),fast),
turn(humanoid,r_shoulder,rotation(-0.722,0.2565,0.1206,0.0477),fast),
turn(humanoid,r_wrist,rotation(0,1.5,-1,1.570),fast)
]),
sleep(500),
par([
turn(humanoid,r_shoulder,rotation(-0.598,0.2199,0.1743,0.0812),fast),
turn(humanoid,r_elbow,rotation(-0.598,0.2199,0.1743,1.2092),fast)
])
```

Σχήμα 4.19: Ο κώδικας STEP του νοήματος ΠΑΙΔΙ

και την ανάδειξη των χωρικών χαρακτηριστικών του νοήματος [133, 134].

4.3.3 Περιορισμοί της εκπαιδευτικής πλατφόρμας

Οι κύριοι περιορισμοί της μελέτης περιγράφονται κατωτέρω. Αυτοί χωρίζονται σε γλωσσικούς, εκπαιδευτικούς και τεχνικούς. Οι περισσότεροι από τους περιορισμούς είναι διαδεδομένοι σε προσπάθειες εμψύχωσης νοηματικής γλώσσας και αναφέρονται σε όλες σχεδόν τις αντίστοιχες βιβλιογραφικές αναφορές.

Τα σημαντικότερα προβλήματα σχετικά με την υλοποίηση του εικονικού νοηματιστή περιλαμβάνουν την ομαλή μετάβαση μεταξύ διαδοχικών νοημάτων και χειρομορφών έτσι ώστε διαδοχικά νοήματα σε μια πρόταση να αρθρώνονται όσο το δυνατόν φυσικότερα. Τα παραπάνω προβλήματα γίνονται εντονότερα όταν οι κινήσεις είναι κυκλικές ή κυματιστές όπου τα πλαίσια κλειδιά (key frames) πολλές φορές δεν αρκούν για να αποδώσουν την κίνηση φυσικά και απρόσκοπτα. Επιπρόσθετα, ένας σημαντικός παράγοντας στη σύνθεση νοημάτων είναι η γραμματική χρήση των μη

Σχήμα 4.20: Η ΕΝΓ έκδοση του νοήματος ΠΑΙΔΙΑ

Σχήμα 4.21: Η ΕΝΓ έκδοση του νοήματος ΗΜΕΡΑ

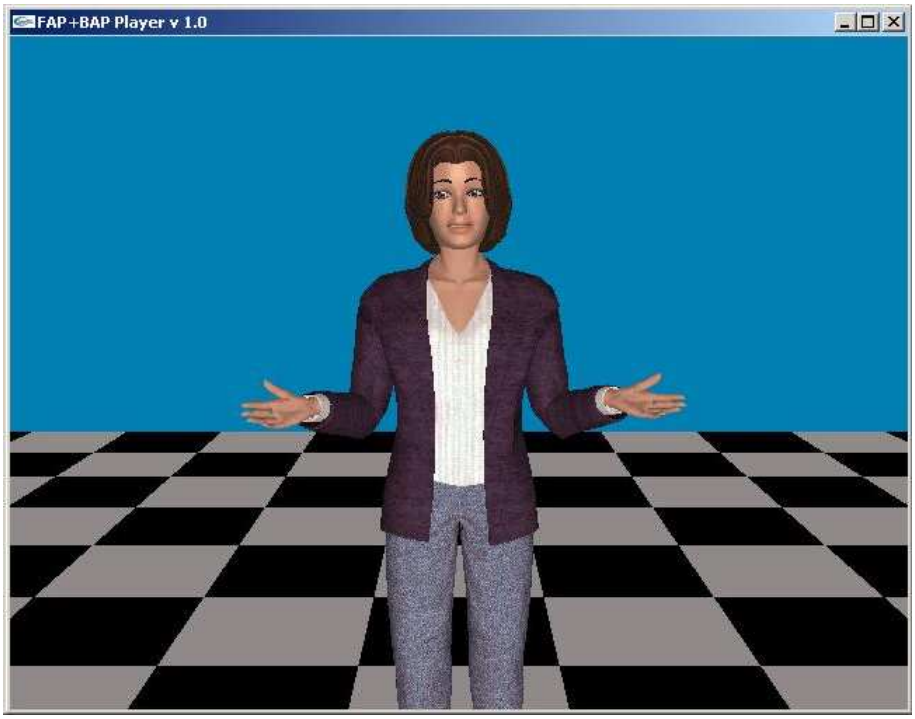
χειρωνακτικών ενδείξεων, όπως η αυθόρμητη έκφραση του προσώπου και το βλέμμα των ματιών [130], ιδιαίτερα όταν το βλέμμα ματιών πρέπει να ακολουθήσει τη διαδρομή των μετακινήσεων των χεριών. Παρόμοια προβλήματα προσδοκούνται στις στοματικές παραμορφώσεις και στα προσωδικά χαρακτηριστικά γνωρίσματα της φωνολογίας νοημάτων. Το λογισμικό STEP δεν υποστηρίζει παραμορφώσεις του προσώπου, έτσι εξετάζεται η μεταφορά του συστήματος στο πρότυπο MPEG–4 [165]. Ένα ενδεικτικό

Σχήμα 4.22: Η ΕΝΓ έκδοση του νοήματος ΔΥΟ ΗΜΕΡΕΣ

Σχήμα 4.23: Η εμπρόσθια όψη της ΕΝΓ έκδοσης του νοήματος ΔΥΟ ΗΜΕΡΕΣ

παράδειγμα της ωρίμανσης της MPEG–4 συνθετικής τεχνολογίας είναι ο εικονικός χαρακτήρας 'Greta' [54] ο όποιος υποστηρίζει όλα τα απαραίτητα χειρωνακτικά και μη χειρωνακτικά συστατικά.

Σχήμα 4.24: Η προβληματική ΕΝΓ έκδοση του νοήματος ΒΑΡΚΑ



Σχήμα 4.25: Η πλατφόρμα εμψύχωσης MPEG4 παραμέτρων GRETA νοηματίζει χειρωνακτικά και μη χειρωνακτικά μέσα

4.3.4 Συμπεράσματα

Το παρόν σύστημα, προ πάντων σαν εκπαιδευτικό εργαλείο, στοχεύει στο να προσφέρει ένα φιλικό προς το χρήστη περιβάλλον για τους νέους κωφούς μαθητές ηλικίας 6 έως 9 ετών, έτσι ώστε αυτοί να μπορέσουν να έχουν την οπτική αναπαράσταση γραπτών λέξεων και φράσεων. Για τους νεαρούς σπουδαστές, ως άτομα με ειδικές ανάγκες, η πλατφόρμα ξεπερνά μερικά από τα εμπόδια προσβασιμότητας και η δυνατότητα της χρήσης από το σπίτι την κάνει ακόμη πιο προσιτή στην οικογένεια, ενθαρρύνοντας κατά συνέπεια την επικοινωνία σε ΕΝΓ, αλλά και την καθολική πρόσβαση στην πλειοψηφία της ελληνικής γλώσσας. Νέα γραπτά κείμενα μπορούν να κωδικοποιηθούν και η πλατφόρμα μπορεί να λάβει σε γενικές γραμμές απεριόριστο εκπαιδευτικό περιεχόμενο, πέραν του υπάρχοντος. Αφ' ετέρου, απεριόριστες σχολικές μονάδες, όπως οι αυξανόμενες ειδικές μονάδες με κωφούς σπουδαστές σε απομακρυσμένες περιοχές μπορούν να συνδεθούν μέσω της πλατφόρμας.

Επιπλέον, η μετάφραση κειμένου σε νόημα μπορεί να επεκταθεί και να εφαρμοστεί σε διαφορετικά περιβάλλοντα όπως η διδασκαλία ελληνικής γλώσσας σε κωφούς σπουδαστές μεγαλύτερων τάξεων, διδασκαλία ΕΝΓ σε ακούοντες σπουδαστές κ.λπ. Σε αυτό το πλαίσιο, η γραμματική της ΕΝΓ μπορεί να αναλυθεί περαιτέρω, περισσότερα λήμματα μπορούν να περιγραφούν και να αποκωδικοποιηθούν, καθιστώντας την συνθετική παραγωγή πιο κοντά στις φυσικές νοηματικές εκφράσεις. Αυτό είναι μια πρόκληση όχι μόνο για τη θεωρητική έρευνα, αλλά και για την πληροφορική και την ισχύουσα γλωσσική έρευνα.

Όσον αφορά στις γλωσσικές και εκπαιδευτικές πτυχές του συστήματος, ένα από τα σημαντικότερα ζητήματα που πρέπει να αντιμετωπιστεί είναι το γεγονός ότι σε μερικούς τομείς της γλώσσας δεν υπάρχουν τυποποιημένα νοήματα και μπορούν να προκύψουν θεωρητικές αντιρρήσεις ως προς τη αναπαράσταση τους. Εντούτοις, μια πλατφόρμα όπως αυτή που περιγράφεται, επιτρέπει πολλαπλές παραλλαγές νοημάτων και δεν έχει περιορισμούς ως προς το μέγεθος των αρχείων όπως συμβαίνει σε προσεγγίσεις βασισμένες σε αρχεία βίντεο. Επιπλέον, η πλατφόρμα είναι ανοικτή σε αναπροσαρμογές μέσω της διαδικασίας δημιουργίας ακολουθιών κώδικα. Ένα άλλο ζήτημα είναι η επιλογή των καταχωρήσεων που περιλαμβάνονται σε κάθε στάδιο της ανάπτυξης του συστήματος ανάλογα με την πολυπλοκότητα των φωνολογικών χαρακτηριστικών τους. Όπως αναφέρεται στην ενότητα καθορισμού περιεχομένου γραμματικής, συμφωνήθηκε να περιληφθούν μόνο μονομορφικές καταχωρήσεις σε πρώτο στάδιο. Σε επόμενα στάδια υπάρχει πρόβλεψη για πολυμορφικά και σύνθετα νοήματα, λειτουργικά μορφήματα, συντακτική χρήση μη χειρωνακτικών στοιχείων, διαδοχικά και ταυτόχρονα συντακτικά φαινόμενα, κ.α. Τέλος, τα διαθέσιμα στοιχεία που αφορούν την ΕΝΓ είναι σημαντικά λιγότερα σε σχέση με αυτά της αντίστοιχης γραπτής γλώσσας κάνοντας την μελέτη τους ιδιαίτερα δύσκολη.

Η απόλυτη πρόκληση βέβαια, όπως σε όλα τα παρόμοια συστήματα, παραμένει η πλήρως αυτόματη μετάφραση της γλώσσας. Είναι ακόμα πάρα πολύ δύσκολο να παραχθούν αποδεκτές προτάσεις στην αυτόματη μετάφραση οποιασδήποτε γλώσσας προς το παρόν, πόσο μάλλον σε μια λιγότερη μελετημένη γλώσσα όπως η ΕΝΓ.

Κεφάλαιο 5

Συμπεράσματα και Μελλοντικές επεκτάσεις

5.1 Συμπεράσματα

5.1.1 Συναισθηματική υπολογιστική

Όπως συζητήθηκε και στο εισαγωγικό κεφάλαιο της διατριβής η ερευνητική εργασία πραγματεύεται, όσον αφορά στο ερευνητικό πεδίο της ανάλυσης συναισθήματος, αρκετά από τα ανοιχτά θέματα της περιοχής. Τόσο στην ενότητα 2.3 όσο και στην 2.5 το συναισθηματικά εμπλουτισμένο πειραματικό σύνολο δεδομένων αποτελείται από φυσιοκρατικές καταγραφές, σε αντίθεση με υποδυόμενες, ακραίες και ελεγχόμενες εκφράσεις που χρησιμοποιούνται στην συντριπτική πλειοψηφία των εργασιών στην αναγνώριση συναισθηματικών καταστάσεων. Επιπλέον, η απαίτηση για συνυπολογισμό της δυναμικής της ανθρώπινης συμπεριφοράς κατά την αυτόματη συναισθηματική ανάλυση, όπως επιτάσσουν μελέτες της γνωστικής επιστήμης, ικανοποιείται με την προσαρμογή του αναδρομικού νευρωνικού δικτύου τύπου Elman. Εστιάζουμε στην δυναμική των συναισθηματικών ενδείξεων παρά στις στατικές τιμές που σχετίζονται με αυτές, όντας κατά συνέπεια σε θέση να χειριστούμε ακολουθίες στις οποίες η αλληλεπίδραση είναι φυσική ή φυσιοκρατική παρά σκηνοθετημένη και ακραία. Από τεχνικής άποψης, οι συνεισφορές κατά την προσαρμογή του νευρωνικού δικτύου τύπου Elman περιλαμβάνουν το τροποποιημένο επίπεδο εισόδου που επιτρέπει στο Elman δίκτυο να επεξεργαστεί τόσο δυναμικές όσο και στατικές εισόδους ταυτόχρονα, απαίτηση για συγχώνευση σε επίπεδο γνωρισμάτων πολλαπλών ροών πληροφορίας διαφορετικής μορφής και φύσης και το τροποποιημένο επίπεδο εξόδου επιτρέποντας στο δίκτυο να ενσωματώνει προηγούμενες τιμές εξόδου και να αντιμετωπίζει αρχικές ή στιγμιαίες αστάθειες στην έξοδο του, καθώς η βέλτιστη χρονική στιγμή δειγματοληψίας της εξόδου του δικτύου δεν είναι προφανής απόφαση.

Στις ενότητες 2.3 και 2.4 η διατριβή συνεισφέρει στην μελέτη και επίλυση προκλήσεων που αφορούν στην πτυχή της συναισθηματικής υπολογιστικής της συγχώνευσης πολλαπλών και διαφορετικής φύσεως μορφών πληροφορίας. Προτείνεται ένα πλαίσιο για την ανάλυση και την αναγνώριση συναισθήματος από τις εκφράσεις προσώπου, τις χειρονομίες και την ομιλία εξετάζοντας και αντιμετωπίζοντας μια σειρά από προκλήσεις και προβλήματα. Η πιθανή απώλεια κάποιας μορφής πληροφορίας κατά την επεξεργασία είναι αρκετά συχνό φαινόμενο στις φυσικές καταγραφές είτε λόγω τεχνικών δυσκολιών και αδυναμίας εξαγωγής χαρακτηριστικών είτε επειδή αυτό καθορίζε-

ται από την συμπεριφορά του χρήστη και αντιμετωπίστηκε επιτυχώς. Η συγχώνευση πολύμορφων δεδομένων αύξησε σημαντικά τα ποσοστά αναγνώρισης σε σύγκριση με τα μονόμορφα συστήματα ενώ από τους δύο τύπους συγχώνευσης αυτή που εκτελείται σε επίπεδο χαρακτηριστικών γνωρισμάτων παρουσιάζει καλύτερα αποτελέσματα από αυτή που εκτελείται σε επίπεδο απόφασης, που φανερώνει πως η επεξεργασία των δεδομένων εισόδου σε ένα κοινό διάστημα χαρακτηριστικών γνωρισμάτων είναι επιτυχέστερη αλλά αρκετά πιο απαιτητική από άποψη πολυπλοκότητας και ευαισθησίας σε απώλεια δεδομένων ή σε χαρακτηριστικά γνωρίσματα χαμηλής εμπιστοσύνης. Στο πλαίσιο της συγχώνευσης σε επίπεδο χαρακτηριστικών γνωρισμάτων προτείνεται ένα σχήμα επιλογής χαρακτηριστικών από κάθε μορφή πληροφορίας ώστε να μειωθεί η υπολογιστική πολυπλοκότητα των τεχνικών μηχανικής μάθησης που χρησιμοποιούνται για την προτυποποίηση και αναγνώριση πολυμορφικών συναισθηματικών καταστάσεων. Επιπλέον, ο σχεδιασμός της διαδικασίας καταγραφής αποτελεί παράπλευρη συνεισφορά καθώς, ενώ έχει βασιστεί στην μέθοδο [10], δεν έχει προταθεί αντίστοιχο πρωτόκολλο στην βιβλιογραφία για πολυμορφική καταγραφή υποδυόμενων χειρονομιών. Είναι χαρακτηριστικό πως η προτεινόμενη διαδικασία καταγραφής αποτέλεσε την βάση για πειραματικό σώμα εκφραστικών χειρονομιών η καταγραφή του οποίου γίνεται στα πλαίσια του ευρωπαϊκού προγράμματος Callas και είναι εν εξελίξει.

Στην καθημερινή αλληλεπίδραση ανθρώπου-υπολογιστή συχνά οι συνθήκες που συνιστούν το πλαίσιο αλληλεπίδρασης διαφέρουν κατά πολύ από αυτές που ίσχυαν κατά την εκπαίδευση του συστήματος αυτόματης ανάλυσης συναισθήματος. Αυτό είναι γνωστό ως το πρόβλημα της εξάρτησης από το πλαίσιο της αλληλεπίδρασης και της εξατομικευμένης εκφραστικότητας και καθιστούν τη γενίκευση τεχνικές μηχανικής μάθησης εξαιρετικά δύσκολη διαδικασία. Στην διατριβή προτείνεται ένα σχήμα εντοπισμού διαφορετικών συνθηκών από αυτές που ίσχυαν κατά την εκπαίδευση και μια επέκταση στην διαδικασία προσαρμογής της λειτουργίας νευρωνικών δικτύων σε αυτές τις διαφορετικές συνθήκες. Το σχήμα εντοπισμού της ανάγκης για προσαρμογή βασίζεται σε πολυμεσικά δεδομένα εισόδου και στηρίζεται στην σημαντική μείωση της απόδοσης αναγνώρισης από ένα κανάλι πληροφορίας. Τα αποτελέσματα που παρουσιάζονται εδώ δείχνουν ότι η απόδοση του δικτύου βελτιώνεται χρησιμοποιώντας αυτήν την προσέγγιση, χωρίς την ανάγκη να εκπαιδευθεί ένα δίκτυο εκ νέου για κάθε θέμα ή να επανεκπαιδευτεί σε εκτενές σύνολο δεδομένων, το οποίο θα εξάλειφε την σημαντική ιδιότητα γενίκευσης του δικτύου. Είναι χαρακτηριστικό πως η αρχική γνώση που του δικτύου που χρησιμοποιήθηκε στην πειραματική επικύρωση της αρχιτεκτονικής προσαρμογής αποκτήθηκε εκπαιδεύοντας το δίκτυο με λιγότερο από 1% του συνολικά διαθέσιμου συνόλου δεδομένων.

5.1.2 Υπολογιστική προτυποποίηση εκφραστικότητας χειρονομιών

Οι ποιοτικές παράμετροι των χειρονομιών που συνδέονται με την εκφραστικότητα έχουν διερευνηθεί επαρκώς από την πλευρά της κινησιολογίας και της σύνθεσης εικονικών χαρακτήρων ικανών να μεταβιβάσουν συναισθηματικό περιεχόμενο. Ανεπαρκής κρίνεται όμως η διερεύνηση της υπολογιστικής πτυχής αυτών των εκφραστικών παραμέτρων από την οπτική γωνία της ανάλυσης. Το κενό αυτό καλύπτει η τρέχουσα διατριβή με την υπολογιστική τυποποίηση εκφραστικών παραμέτρων από την πλευρά της ανάλυσης και την εισαγωγή αλγορίθμου εξαγωγής αυτών των παραμέτρων στο κεφάλαιο 3 ο οποίος επικυρώνεται με την μέθοδο του ελέγχου αντίληψης και σύγκρισης με χειρωνακτικό σχολιασμό. Επιπρόσθετα, στην ενότητα 3.2 ορίζεται πλαίσιο πο-

λυμεσικής εκφραστικής ανατροφοδότησης από Ενσαρκωμένο Πράκτορα Συνομιλητή (Embodied Conversational Agent - ECA) στο οποίο ενσωματώνεται η εκφραστική ανάλυση και αποτελεί την ικανότητα του εικονικού πράκτορα να αντιληφθεί και να ερμηνεύσει την συναισθηματικής κατάστασης του χρήστη ή έστω κάποιων ενδείξεων αυτής. Η δυνατότητα των εικονικών πρακτόρων να παρέχουν εκφραστική ανατροφοδότηση στον χρήστη είναι μια σημαντική πτυχή ώστε να υποστηρίξουν τη φυσικότητα και την αληθοφάνεια της αλληλεπίδρασης τους με στόχο να ενισχυθεί η επικοινωνιακή εμπειρία του χρήστη. Παράλληλα, προτείνεται εύρωστος αλγόριθμος εντοπισμού και παρακολούθησης των χεριών και του κεφαλιού του χρήστη με χαρακτηριστικό το χαμηλό υπολογιστικό κόστος και την εφαρμοσιμότητα σε πραγματικό χρόνο και σε εικόνες χαμηλής ανάλυσης και δυναμικών συνθηκών καταγραφής. Άξιο αναφοράς είναι το γεγονός πως υλοποιήσεις τόσο του αλγορίθμου εντοπισμού και παρακολούθησης χεριών όσο και εξαγωγής εκφραστικών παραμέτρων χειρονομιών ήδη εφαρμόζονται σε ευρωπαϊκά ερευνητικά έργα ανάλυσης συναισθηματικά εμπλουτισμένων συμπεριφορών.

Τέλος, στην ενότητα 3.3 αντιμετωπίζεται το πρόβλημα της επικύρωσης χειρωνακτικού σχολιασμού εκφραστικότητας μέσω της αυτόματης εξαγωγής εκφραστικών παραμέτρων. Πραγματοποιείται σύγκριση και ανάλυση συσχέτισης των δύο επισημειώσεων (αυτόματης και χειρωνακτικής) και προτείνεται αλγόριθμος αυτόματου εντοπισμού χρονικών τμημάτων ενεργοποίησης των χρηστών στο σώμα υπό εξέταση. Η επισημείωση σωμάτων φυσιοκρατικής, πολυμεσικής, συναισθηματικής συμπεριφοράς αποτελεί πρόκληση, δεδομένου ότι περιλαμβάνει την υποκειμενική αντίληψη και απαιτεί μεγάλο χρονικό διάστημα για τον συναισθηματικό σχολιασμό σε πολλαπλά, παράλληλα επίπεδα. Αυτός ο χειρωνακτικός σχολιασμός ωφελείται από την αυτόματη εκτίμηση ποιοτικών παραμέτρων κίνησης του σώματος η οποία επικυρώνει τους χειρωνακτικούς σχολιασμούς.

5.1.3 Αναγνώριση χειρονομιών και νοηματικής γλώσσας

Στο ερευνητικό πεδίο της αναγνώρισης χειρονομιών και νοηματικής γλώσσας υπάρχουν προκλήσεις και ανοιχτά θέματα που αφορούν στην εφαρμοστικότητα των προτεινόμενων προσεγγίσεων σε λεξικά μεγάλης κλίμακας, στην ανεξαρτησία από τον χρήστη και στο υπολογιστικό κόστος. Στο κεφάλαιο 4 προτείνεται ένα σχήμα αναγνώρισης χειρονομιών που αντιμετωπίζει τα παραπάνω ζητήματα. Μια πρωτότυπη αρχιτεκτονική αυτόματης αναγνώρισης νοηματικής γλώσσας σε επίπεδο λήμματος προτείνεται ενσωματώνοντας αυτοοργανούμενους χάρτες, αλυσίδες Μαρκώφ και τροποποιημένη μετρική απόστασης Levenshtein. Τα εξαγόμενα χαρακτηριστικά γνωρίσματα εκπαιδεύουν μεμονωμένους ταξινομητές, οι οποίοι στη συνέχεια συνδυάζονται σε επίπεδο απόφασης, μια ενισχυτική τεχνική βασισμένη σε αδύναμους κατηγοριοποιητές, ενισχύοντας την προτεινόμενη αρχιτεκτονική με πολλαπλές μορφές πληροφορίας και ευρωστία έναντι σε αλλοιώσεις στιγμιοτύπων και θορυβώδη και ανεξέλεγκτα περιβάλλοντα. Η διακύμανση κατά την απόδοση των χειρονομιών αντιμετωπίζεται μέσω της ευελιξίας τόσο της διαδικασίας εκπαίδευσης όσο και της αποκωδικοποίησης που βασίζεται στο χαρακτηριστικό γειτνίασης SOM και του αλγορίθμου αναζήτησης βέλτιστης τροχιάς που εκτελείται κατά τη διάρκεια της ταξινόμησης. Το χαρακτηριστικό αυτό της γειτνίασης επιτρέπει την διασπορά των πιθανοτήτων μετάβασης σε γειτονικούς κόμβους-σύμβολα και με αυτή την μέθοδο μειώνεται σημαντικά η απαίτηση για εξαντλητική εκπαίδευση. Η προσαρμογή του αλγορίθμου υπολογισμού απόστασης

Levenshtein ώστε να συνυπολογίζει την ομοιότητα των συμβόλων της ακολουθίας αντιμετωπίζει το πρόβλημα πιθανών ασταθειών ή θορύβου που οφείλονται είτε στην απόδοση της χειρονομίας είτε στην τεχνική εξαγωγής χαρακτηριστικών γνωρισμάτων. Αντίθετα με ανταγωνιστικές τεχνικές, δεν απαιτείται αυθαίρετη ή πειραματικά καθορισμένη αρχικοποίηση σχεδιαστικών παραμέτρων ενώ η μέθοδος συγχώνευσης διαφορετικών ροών πληροφορίας έχει αρκετές ομοιότητες με τεχνικές ενίσχυσης αδύναμων κατηγοριοποιητών. Το υπολογιστικό κόστος και η ταχύτητα επεξεργασίας της διαδικασίας αναγνώρισης αποδεικνύουν ότι η προτεινόμενη αρχιτεκτονική είναι κατάλληλη για εφαρμογές πραγματικού χρόνου αφού ικανοποιεί όλες τις απαιτήσεις που συνοδεύουν τέτοιου είδους σενάρια.

Επιπρόσθετα, προτείνεται πλαίσιο προσβάσιμης μέσω διαδικτύου σύνθεσης νοηματικής γλώσσας από εικονικό νοηματιστή. Το παρόν σύστημα, προ πάντων σαν εκπαιδευτικό εργαλείο, στοχεύει στο να προσφέρει ένα φιλικό προς το χρήστη περιβάλλον οπτική αναπαράσταση γραπτών λέξεων και φράσεων. Το σύστημα ξεπερνά μερικά από τα εμπόδια προσβασιμότητας και παρέχει την δυνατότητα εξ αποστάσεως εκπαίδευσης αλλά και της καθολικής πρόσβασης στην ελληνική νοηματική γλώσσα. Νέα γραπτά κείμενα μπορούν να κωδικοποιηθούν και η πλατφόρμα μπορεί να λάβει σε γενικές γραμμές απεριόριστο εκπαιδευτικό περιεχόμενο ενώ συστήματα υποβοηθητικών τεχνολογιών αποτελούν εξαιρετικό πεδίο εφαρμογής της.

5.2 Μελλοντικές επεκτάσεις

5.2.1 Αναγνώριση δυναμικών συναισθηματικών καταστάσεων από πολλαπλές μορφές πληροφορίας σε φυσική επικοινωνία ανθρώπου μηχανής

Σχετικά με την μελλοντική εργασία όσο αφορά στην ανάλυση δυναμικής συναισθηματικής αλληλεπίδρασης σκοπεύουμε να επεκτείνουμε περαιτέρω την εργασία μας σε πολύμορφη φυσιοκρατική αναγνώριση έκφρασης με την εξέταση περισσότερων μορφών όπως στάση σώματος και εκφραστικά χαρακτηριστικά χειρονομιών και με την ενσωμάτωση της αβεβαιότητας που συνοδεύει τα χαρακτηριστικά γνωρίσματα προκειμένου να μεγιστοποιηθεί η απόδοση και η ευρωστία του συστήματος σε ανεξέλεγκτα περιβάλλοντα. Η αναγνώριση συναισθήματος από πολλαπλές μορφές πληροφορίας είναι επεκτάσιμη ως προς την χρήση πιο εξελιγμένων και καταλληλότερων τεχνικών κατηγοριοποίησης μοναδικής αλλά και πολλαπλής πληροφορίας και ο περαιτέρω πειραματισμός στο πλαίσιο συνδυασμού χαρακτηριστικών γνωρισμάτων και αποφάσεων. Μια εναλλακτική προσέγγιση που μπορεί επίσης να είναι ενδιαφέρουσα θα ήταν να αναγνωρισθεί το συναίσθημα από εκφραστικά χαρακτηριστικά της ίδιας χειρονομίας ή κανονικοποίηση των τιμών των εκφραστικών παραμέτρων σύμφωνα με το είδος της χειρονομίας και προς αυτή την κατεύθυνση σίγουρα θα μπορούσε να ενσωματωθεί η αρχιτεκτονική αναγνώρισης χειρονομιών που παρουσιάστηκε στο αντίστοιχο κεφάλαιο της διατριβής.

Πιθανή επέκταση της εργασίας σχετικά με την διαδικασία προσαρμογής νευρωνικών δικτύων με γνώμονα την αναγνώριση συναισθηματικής κατάστασης του χρήστη περιλαμβάνει την ενσωμάτωση περαιτέρω μορφών πληροφορίας, επέκταση σε διαφορετικά φυσιοκρατικά πλαίσια και εισαγωγή μηχανισμών αντιμετώπισης της αβεβαιότητας διαφόρων μορφών πληροφορίας που αποφασίζουν ποιά από αυτές τις μορφές

είναι πιο αξιόπιστη για να βασιστούν στην συνεκπαίδευση. Ακόμα, η διερεύνηση της εφαρμογής των αλγορίθμων ανίχνευσης ανάγκης για προσαρμογή και προσαρμογής σε αναδρομικά δίκτυα όπως αυτό που περιγράφηκε στην σχετική ενότητα αποτελεί εξαιρετικά ενδιαφέρουσα προοπτική.

5.2.2 Εκφραστική και πολυμεσική ανάλυση και σύνθεση σε εικονικούς πράκτορες

Πιθανή επέκταση του πλαισίου εκφραστικής ανάλυσης και σύνθεσης από εικονικούς πράκτορες περιλαμβάνει την ενσωμάτωση συναισθηματικών ενδείξεων προσωδίας από το κανάλι της ομιλίας. Στο μέλλον, στοχεύουμε στην εκμετάλλευση ενός πιο σύνθετου μοντέλου απόφασης, το οποίο θα αναλαμβάνει την επιλογή των ενεργειών που θα εκτελέσει ο ECA, σύμφωνα επίσης με την τρέχουσα συμπεριφορά του χρήστη και να αξιολογήσουμε την ορθότητα της προτεινόμενης προσέγγισης χρησιμοποιώντας τον σχεδιασμό που συζητείται στην ενότητα 3.2.7. Επιπλέον, μια ενδιαφέρουσα επέκταση στο προτεινόμενο πλαίσιο εκφραστικής και πολυμεσικής σύνθεσης με την χρήση εικονικών πρακτόρων, βασισμένη στην ανάλυση των ενεργειών που εκτελέσθηκαν από ανθρώπινους χρήστες, είναι αυτή της αντίληψης ενδείξεων οπτικής προσοχής (visual attention) από το χρήστη [197]. Η οπτική προσοχή μπορεί να ενσωματωθεί στην υπάρχουσα αρχιτεκτονική για να επιλέξει ορισμένες μόνο πληροφορίες στα στάδια της αισθητήριας αποθήκευσης, αντίληψης ή ερμηνείας για πρόσβαση στα περαιτέρω στάδια επεξεργασίας, καθώς επίσης και διαμορφώνοντας τον προγραμματισμό και για κάποια παραγωγή συμπεριφοράς, όπως ο προσανατολισμός του βλέμματος των πρακτόρων. Ένα σύστημα οπτικής προσοχής, εφαρμόσιμο στο πραγματικό και εικονικό περιβάλλον, σε ένα ενοποιημένο πλαίσιο, είναι μια ενδιαφέρουσα προοπτική. Περαιτέρω, πληροφορίες για το πλαίσιο αλληλεπίδρασης είναι ένας εξαιρετικά σημαντικός παράγοντας όταν γίνεται προσπάθεια να αναλυθούν οι σημασιολογικές πτυχές της ανθρώπινης συμπεριφοράς. Η οπτική προσοχή μαζί με την προσωποποίηση και την εξατομίκευση είναι διαδικασίες που ένα τμήμα λογισμικού που αναλαμβάνει τον σχεδιασμό των εμψυχώσεων πρέπει να λάβει υπόψη. Η χρήση του βρόχου ανάλυσης-σύνθεσης ως φάση εκμάθησης που εκλεπτύνει το πρότυπο σύνθεσης της εκφραστικότητας και της συμπεριφοράς πιθανόν να αποδειχθεί χρήσιμη στην συνολική διαδικασία. Τέλος, προφανώς η ολοκλήρωση της υλοποίησης των υποενοτήτων εκείνων που δεν είχαν υλοποιηθεί στο αρχικό πρωτότυπο είναι αναγκαία για την πλήρη και συνολική επικείμενη αξιολόγηση του συστήματος.

Ακόμα, η αυτόματη επεξεργασία εικόνας που θα μπορεί να επικυρώσει τον χειρωνακτικό σχολιασμό της συναισθηματικής κατάστασης σε τηλεοπτικές συνεντεύξεις μπορεί μελλοντικά να επεκταθεί με τη χωριστή εκτίμηση της ποσότητας μετακίνησης για διαφορετικά μέρη σώματος και πιθανόν πολλών ανθρώπων ώστε να αντιμετωπιστεί η ύπαρξη ανθρώπων που ενώ κινούνται στο υπόβαθρο αλλά δεν συμμετέχουν ουσιαστικά στην συνέντευξη, την αυτόματη εξαγωγή και των υπολοίπων εκφραστικών παραμέτρων όπως η χωρική έκταση, την επικύρωση του χειρωνακτικού σχολιασμού της ενεργοποίησης σε επίπεδο συναισθηματικών τμημάτων του βίντεο, την συσχέτιση της εκτίμησης της ποσότητας μετακίνησης και τις φάσεις της χειρονομίας (προετοιμασία, κτύπημα, απόσυρση), την χρήση χρονικών φίλτρων για τη βελτίωση της αυτόματης ανίχνευσης κινήσεων και τελικά τον συνυπολογισμός του σχολιασμού του κορμού του σώματος.

5.2.3 Αναγνώριση και σύνθεση χειρονομιών και Ελληνικής Νοηματικής Γλώσσας

Μελλοντική αλλά και τρέχουσα εργασία περιλαμβάνει την διερεύνηση της αναγνώρισης συστατικών νοηματικής γλώσσας (signeme). Εμπνευσμένοι από το σύστημα μεταγραφής HamNoSys μια αρχιτεκτονική θα μπορούσε να συμπεράνει σχετικά με διαφορετικές πτυχές της νοηματικής γλώσσας όπως την χειρομορφή, την κατεύθυνση και τον προσανατολισμό της παλάμης, την θέση και την μετακίνηση των χεριών, ενώ ακόμα και μη χειρωνακτικά χαρακτηριστικά γνωρίσματα όπως οι εκφράσεις του προσώπου [118], η κίνηση των χειλιών, την παρακολούθηση του βλέμματος [9] κ.λπ. καθώς και τρόποι ενσωμάτωσης τους στην διαδικασία αναγνώρισης πρέπει να μελετηθούν δεδομένου ότι είναι εξαιρετικά κρίσιμα στη γλωσσική ανάλυση νοηματικής. Οι γλωσσολογικές πτυχές της νοηματικής γλώσσας δεν πρέπει να αγνοηθούν κατά την διαδικασία αναγνώρισης, αλλά παράλληλα είναι ανάγκη να ωριμάσουν και οι υπόλοιπες εμπλεκόμενες ερευνητικές περιοχές, όπως η γλωσσολογία και επεξεργασία φυσικής γλώσσας. Η επέκταση της παρούσας εργασίας σε συνεχή νοηματισμό θα μπορούσε να εξεταστεί απλοϊκά με την προσθήκη ενός μηχανισμού εντοπισμού νοήματος ή μιας διαδικασίας χρονικής κατάτμησης στην υπάρχουσα αρχιτεκτονική. Μια τέτοια προσέγγιση στο παρελθόν έχει αποδειχθεί ανεπαρκής και η ανάγκη συνυπολογισμού της γλωσσικής και γραμματικής ανάλυσης έχει γίνει όλο και περισσότερο προφανής. περιορίζονται σε χειρωνακτικά χαρακτηριστικά γνωρίσματα.

Είναι επιτακτική ανάγκη να διερευνηθεί σε βάθος πώς η δομή της νοηματικής γλώσσας, το συντακτικό και τα γραμματικά φαινόμενα θα μπορούσαν να ενσωματωθούν στη ευρύτερη αρχιτεκτονική αναγνώρισης δεδομένου ότι θα ήταν εξαιρετικά ευεργετικό στην ολοκλήρωση του ερευνητικού πεδίου αλλά θα προσέφερε αρκετά και στην επαλήθευση της γλωσσολογικής ανάλυσης. Η ενσωμάτωση γλωσσικής γνώσης είτε στη διαδικασία συγχώνευσης ροών πληροφορίας για αναγνώριση μεμονωμένων λημμάτων είτε στην αναγνώριση σε επίπεδο πρότασης θα ενίσχυε σημαντικά την ευρωστία της διαδικασίας. Η προσθήκη αυτού του τελικού επιπέδου γλωσσικών ή γραμματικών φαινομένων υποβοηθούμενου από γνώση (knowledge assisted). Τέλος, η διερεύνηση πιθανών βελτιστοποιήσεων του αλγορίθμου, η πρόβλεψη νοημάτων και χειρονομιών πριν την ολοκλήρωση τους καθώς και δενδρικές δομές απόφασης αποτελούν μελλοντικές κατευθύνσεις της προτεινόμενης αρχιτεκτονικής.

Πιθανές επεκτάσεις του συστήματος σύνθεσης νοηματικής γλώσσας με την χρήση εικονικού χαρακτήρα περιλαμβάνουν την ολοκλήρωση του αυτόματου τρόπου μετάφρασης κειμένου σε σύμβολα νοηματικής γλώσσας, εμπλουτισμένο με γλωσσολογική επεξεργασία, στοιχείο που θα δώσει την δυνατότητα σε νέα γραπτά κείμενα να κωδικοποιηθούν και την επίλυση των περιορισμών που προκύπτουν είτε από την μεριά της σύνθεσης είτε από αυτή της κωδικοποίησης. Σε επόμενα στάδια υπάρχει πρόβλεψη για πολυμορφικά και σύνθετα νοήματα, λειτουργικά μορφήματα, συντακτική χρήση μη χειρωνακτικών στοιχείων, διαδοχικά και ταυτόχρονα συντακτικά φαινόμενα, κ.α. Τέλος, τα διαθέσιμα στοιχεία που αφορούν την ΕΝΓ είναι σημαντικά λιγότερα σε σχέση με αυτά της αντίστοιχης γραπτής γλώσσας κάνοντας την μελέτη τους ιδιαίτερα δύσκολη. Τέλος, πιθανή διασύνδεση του συστήματος αναγνώρισης και σύνθεσης νοηματικής θα αποτελούσε την βάση ενός αμφίδρομου συστήματος επικοινωνίας νοηματιστή-υπολογιστή αλλά και μεταξύ νοηματιστών με την χρήση υπολογιστή.

Κεφάλαιο 6

Κατάλογος δημοσιεύσεων

6.1 Περιοδικά

1. G. Caridakis, A. Raouzaiou, E. Bevacqua, M. Mancini, K. Karpouzis, L. Malatesta and C. Pelachaud, Virtual agent multimodal mimicry of humans, Language Resources and Evaluation 41 (3–4), Special issue on Multimodal Corpora, pp. 367–388, Springer, 2007

2. K. Karpouzis, G. Caridakis, S-E. Fotinea, E. Efthimiou, Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture, Computers & Education, Special Issue in Web3D Technologies in Learning, Education and Training, Elsevier, Volume 49, Issue 1, pp. 54–74, August 2007.

3. S. Ioannou, G. Caridakis, K. Karpouzis, S. Kollias, Robust Feature Detection for Facial Expression Recognition EURASIP Journal on Image and Video Processing, Article ID 29081, pp. 1–22, 2007.

4. G. Caridakis, K. Karpouzis, S. Kollias, User and Context Adaptive Neural Networks for Emotion Recognition, Neurocomputing, Elsevier, Volume 71, Issue 13-15, pp. 2553–2562, 2008.

5. S.E. Fotinea, E. Efthimiou, K. Karpouzis, G. Caridakis, A Knowledge-based Sign Synthesis Architecture, Universal Access in the Information Society 6 (4), pp. 405–418, Springer, 2008.

6. J.-C. Martin, G. Caridakis, L. Devillers, K. Karpouzis, S. Abrilian, Manual annotation and automatic image processing of multimodal emotional behaviors: validating the annotation of TV interviews, Personal and Ubiquitous Computing, Special issue on Emerging Multimodal Interfaces, Volume 13, Number 1 / January, pp. 69–76, 2009.

7. G. Caridakis, K. Karpouzis, N. Drosopoulos, S. Kollias, Adaptive Sign Language Recognition using Self Organizing Markov Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, (υπό κρίση).

8. G. Caridakis, K. Karpouzis, N. Drosopoulos, S. Kollias, SOMM: Self Organizing Markov Map for gesture recognition, Pattern Recognition Letters, (υπό κρίση).

9. G. Caridakis, K. Karpouzis, N. Drosopoulos, S. Kollias, A critic review of Sign Language Recognition Approaches, ACM Computing Surveys, (υπό κρίση).

10. L. Kessous, G. Castellano, G. Caridakis, Multimodal Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis, Real-Time Affect Analysis and Interpretation: Closing the Affective Loop in Virtual Agents and Robots, Special issue of the Journal on Multimodal User Interfaces, (υπό κρίση).

11. G. Caridakis, K. Karpouzis, M. Wallace, L. Kessous, N. Amir, Multimodal user's affective state analysis in naturalistic interaction, Journal of Multimodal User Interfaces, Special Issue AFFINE, (υπό κρίση).

6.2 Κεφάλαια σε βιβλία

12. K. Karpouzis, G. Caridakis, L. Kessous, N. Amir, A. Raouzaiou, L. Malatesta, S. Kollias, Modeling naturalistic affective states via facial, vocal, and bodily expressions recognition, T. Huang, A. Nijholt, M. Pantic, A. Pentland (eds.), Lecture Notes in Artificial Intelligence, Special Volume on AI for Human Computing, pp. 91–112, Springer, 2007.

13. G. Castellano, L. Kessous, G. Caridakis, Multimodal emotion recognition from expressive faces, body gestures and speech, C. Peter, R. Beale (eds), Affect and Emotion in Human-Computer Interaction, Lecture Notes in Computer Science, Springer, 2007.

14. R. Cowie, E. Douglas-Cowie, K. Karpouzis, G. Caridakis, M. Wallace, S. Kollias, Recognition of Emotional States in Natural Human-Computer Interaction, D. Tzovaras (ed.), Multimodal User Interfaces, pp. 119–153, Springer Berlin Heidelberg, 2008.

6.3 Συνέδρια

15. K. Karpouzis, G. Caridakis, S-E. Fotinea, E. Efthimiou, Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture, International Workshop on Web3D Technologies in Learning, Education and Training, Udine, Italy, 2004.

16. S. E. Fotinea, E. Efthimiou, K. Karpouzis, G. Caridakis, Dynamic GSL synthesis to support access to e-content, HCI International, Las Vegas, Nevada, USA, 2005.

17. G. Caridakis, K. Karpouzis, G. Sapountzaki, S-E. Fotinea, E. Efthimiou, A dynamic environment for Greek Sign Language Synthesis using virtual characters, ACM Web3D Symposium, March 29 – April 1, Bangor, UK, 2005.

18. S. Ioannou, L. Kessous, G. Caridakis, K. Karpouzis, V. Aharonson, S. Kollias, Adaptive On-Line Neural Network Retraining for Real Life Multimodal Emotion Recognition, International Conference on Artificial Neural Networks (ICANN), Athens, Greece, September 2006.

19. J.-C. Martin, G. Caridakis, L. Devillers, K. Karpouzis, S. Abrilian, Manual Annotation and Automatic Image Processing of Multimodal Emotional Behaviours: Validating the Annotation of TV Interviews, 5th Conference on Language Resources and Evaluation (LREC), Genoa, Italy, May 2006.

20. G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Raouzaiou, K. Karpouzis, Modeling naturalistic affective states via facial and vocal expressions recognition, International Conference on Multimodal Interfaces (ICMI 2006), Banff, Alberta, Canada, November 2–4, 2006

21. G. Caridakis, A. Raouzaiou, K. Karpouzis, S. Kollias, Synthesizing Gesture Expressivity Based on Real Sequences, Workshop on multimodal corpora: from multimodal behaviour theories to usable models, LREC Conference, Genoa, Italy, 24–26 May, 2006.

22. E. Bevacqua, A. Raouzaiou, C. Peters, G. Caridakis, K. Karpouzis, C. Pelachaud, M. Mancini, Multimodal Sensing, Interpretation and Copying of Movements by a Virtual Agent, Perception and Interactive Technologies, Kloster Irsee, Germany, June 19–21, 2006

23. Lori Malatesta , George Caridakis, Amaryllis Raouzaiou, Kostas Karpouzis, Agent Personality Traits in Virtual Environments Based on Appraisal Theory Predictions, AISB: Artificial and Ambient Intelligence, Language, Speech and Gesture for Expressive Characters, Newcastle upon Tyne, UK, April 2–4, 2007

24. G. Caridakis, C. Pateritsas, A. Drosopoulos, A. Stafylopatis, S. Kollias, Probabilistic Video-Based Gesture Recognition Using Self-Organizing Feature Maps, 17th International Conference on Artificial Neural Networks (ICANN),September 9–13, Porto, Portugal, 2007.

25. G. Caridakis, O. Diamanti, K. Karpouzis, P. Maragos, Automatic Sign Language Recognition: vision based feature extraction and probabilistic recognition scheme from multiple cues, 1st ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA), July 15–19, Athens, Greece, 2008.

26. G. Caridakis, K. Karpouzis, C. Pateritsas, A. Drosopoulos, A. Stafylopatis, S. Kollias, Hand Trajectory-based Gesture Recognition using Self-Organizing Feature Maps and Markov Models, IEEE International Conference on Multimedia & Expo (ICME), June 23–26, Hannover, Germany, 2008.

27. G. Caridakis, K. Karpouzis, N. Drosopoulos, S. Kollias, Adaptive gesture recognition in Human Computer Interaction, International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), London, UK, 2009

28. S. Fotinea, E. Efthimiou, G. Caridakis, O. Diamanti, N. Mitsou, K. Karpouzis, C. Tzafestas, P. Maragos, DIANOEMA: Visual analysis and sign recognition for GSL modelling and robot teleoperation, 8th International Gesture Workshop, GW2009, Bielefeld University, Germany, 2009.

6.4 Επιλεγμένες Αναφορές

1. Aharonson, V., Nehmadi, N., Messer, H., 2007. Automatic emotional stimulus identification from facial expressions. In: Proceedings of the Fourth conference on IASTED International Conference: Signal Processing, Pattern Recognition, and Applications table of contents. ACTA Press Anaheim, CA, USA, pp. 333–337.

2. Asteriadis, S., Tzouveli, P., Karpouzis, K., Kollias, S., 2009. Estimation of behavioral user state based on eye gaze and head pose: application in an e-learning environment. Multimedia Tools and Applications 41(3), 469–493.

3. Barra-Chicote, R., Montero, J., Macias-Guarasa, J., Lufti, S., Lucas, J., Fernandez-Martinez, F., Dharo, L., San-Segundo, R., Ferreiros, J., Cordoba, R., 2008. Spanish Expressive Voices: corpus for emotion research in Spanish. In: Programme of the Workshop on Corpora for Research on Emotion and Affect. p.66.

4. Boukis, C., Pnevmatikakis, A., Polymenakos, L., 2007. Artificial intelligence and innovations 2007 from theory to applications. In: 4th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2007). Springer.

5. Busso, C., Bulut, M., Lee, C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J., Lee, S., Narayanan, S., 2008. IEMOCAP: interactive emotional dyadic motion capture database. Language Resources and Evaluation 42(4), 335–359.

6. Busso, C., Narayanan, S., 2008. Recording audio-visual emotional databases from actors: a closer look. In: Programme of the Workshop on Corpora for Research on Emotion and Affect. p.17.

7. Buttussi, F., Chittaro, L., Coppo, M., 2007. Using Web3D technologies for visualization and search of signs in an international sign language dictionary. In: Proceedings of the Twelfth International Conference on 3D Web Technology 2007, Web3D 2007. Vol. 2007. pp. 61–70.

8. Castellano, G., Aylett, R., Dautenhahn, K., Paiva, A., McOwan, P., Ho, S., 2008. Long-term affect sensitive and socially interactive companions. In: Fourth International Workshop on Human-Computer Conversation.

9. Castellano, G., Aylett, R., Paiva, A., McOwan, P., 2008. Affect Recognition for Interactive Companions. In: Workshop on Affective Interaction in Natural Environments (AFFINE), ACM International Conference on Multimodal Interfaces (ICMI'08).

10. Chittaro, L., Ranon, R., 2007. Web3D technologies in learning, education and training: Motivations, issues, opportunities. Computers and Education 49(1), 3–18.

11. Efthimiou, E., Fotinea, S., 2007. An environment for deaf accessibility to educational content. In: Proceedings of The First International Conference on Information and Communication Technology and Accessibility, Hammamet, Tunisia. pp. 125–130.

12. Efthimiou, E., Fotinea, S., 2007. GSLC: Creation and annotation of a greek sign language corpus for hci. Lecture Notes In Computer Science 4554, 657.

13. Fotinea, S., Efthimiou, E., 2008. Tools for Deaf Accessibility to an eGOV Environment. In: Proceedings of the 11th international conference on Computers Helping People with Special Needs. Springer, pp. 446–453.

14. Garcia-Rojas, A., Vexo, F., Thalmann, D., Raouzaiou, A., Karpouzis, K., Kollias, S., 2006. Emotional body expression parameters in virtual human ontology. In: Proceedings of 1st Int. Workshop on Shapes and Semantics. pp. 63–70.

15. Kollias, S., Stafylopatis, A., Duch, W., 2008. International Conference on Artificial Neural Networks (ICANN 2006). Neurocomputing 71(13–15), 2409–2410.

16. Maat, L., Pantic, M., 2007. Gaze-X: Adaptive, affective, multimodal interface for single-user office scenarios. Vol. 4451 LNAI.

17. Mancini, M., Castellano, G., Bevacqua, E., Peters, C., 2007. Copying Behaviour of Expressive Motion. Lecture Notes in Computer Science 4418, 180.

18. Martin, J., Devillers, L., 2008. A Multimodal Corpus Approach for the Study of Spontaneous Emotions. Affective Information Processing, 267.

19. Martin, J., Paggio, P., Kuehnlein, P., Stiefelhagen, R., Pianesi, F., 2008. Introduction to the special issue on multimodal corpora for modeling human multimodal behavior. Language Resources and Evaluation 42(2), 253–264.

20. McIntyre, G., Göcke, R., 2008. The Composite Sensing of Affect. Lecture Notes In Computer Science, 104–115.

21. Pelachaud, C., 2008. Studies on gesture expressivity for a virtual agent. Speech Communication.

22. Peter, C., Beale, R., 2008. Affect and Emotion in Human-Computer Interaction: From Theory to Applications. Springer-Verlag Berlin, Heidelberg.

23. Ratliff, M., Patterson, E., 2008. Emotion recognition using facial expressions with active appearance models. In: Proceedings of the 3rd IASTED International Conference on Human-Computer Interaction, HCI 2008. pp. 138–143.

24. Rehm, M., Vogt, T., Wissner, M., Bee, N., 2008. Dancing the night away: controlling a virtual karaoke dancer by multimodal expressive cues. In: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems, Volume 3. International Foundation for Autonomous Agents and Multiagent Systems Richland, SC, pp. 1249–1252.

25. Simou, N., Athanasiadis, T., Kollias, S., Stamou, G., Stafylopatis, A., 2008. Semantic adaptation of neural network classifiers in image segmentation. Vol.5163 LNCS.

26. Tzovaras, D., 2008. Multimodal User Interfaces: From Signals to Interaction. Springer.

27. Wang, X., Jiang, F., Yao, H., 2008. Sign language synthesis of individuation based on data model. In: Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP'08 International Conference on. pp. 354–357.

28. Zeng, Z., Pantic, M., Roisman, G., Huang, T., 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 31(1), 39–58.

29. Zeng, Z., Tu, J., Jr., B.P., Huang, T., 2008. Audio-visual affective expression recognition through multistream fused HMM. IEEE Transactions on Multimedia 10(4), 570–577.

Βιβλιογραφία

[1] S. Abrilian, L. Devillers, S. Buisine, and J.-C. Martin. Emotv: Annotation of real-life emotions for the specification of multimodal affective interfaces. In *11th International Conference on Human-Computer Interaction, HCII'2005*, 2005.

[2] H. Ai, D. Litman, K. Forbes-Riley, M. Rotaru, J. Tetreault, and A. Purandare. Using system and user performance features to improve emotion detection in spoken tutoring dialogs. In *Interspeech ICSLP*, 2006.

[3] J. Alon, V. Athitsos, Q. Yuan, and S Sclaroff. A unified framework for gesture recognition and spatiotemporal gesture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear(3), 2009.

[4] N. Ambady and R. Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2):256–274, 1992.

[5] J. Ang, R. Dhilon, A. Krupski, E. Shriberg, and A. Stolcke. Prosody based automatic detection of annoyance and frustration in human computer dialog. In *International Conference on Spoken Language Processing*, 2002.

[6] A.B. Ashraf, S. Lucey, J.F. Cohn, T. Chen, Z. Ambadar, K. Prkachin, P. Solomon, and B.J. Theobald. The painful face: Pain expression recognition using active appearance models. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 9–14. ACM New York, NY, USA, 2007.

[7] K. Assaleh and M. Al Rousan. Recognition of arabic sign language alphabet using polynomial classifiers. *Journal on Applied Signal Processing*, 2005(13):2136–2145, 2005.

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

[9] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. Estimation of behavioral user state based on eye gaze and head poseϑapplication in an e-learning environment. *Multimedia Tools and Applications, Springer*, 2008.

[10] T. Baenziger, H. Pirker, and K. Scherer. Gemep - geneva multimodal emotion portrayals: a corpus for the study of multimodal emotional expressions. In *LREC'06 Workshop on Corpora for Research on Emotion and Affect*, 2006.

[11] MS Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: machine learning and application to spontaneous behavior. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, volume 2, 2005.

[12] A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. How to find trouble in communication. *Speech Communication*, 40(1-2):117–143, 2003.

[13] A. Batliner, R. Huber, H. Niemann, E. Noeth, J. Spilker, and K. Fischer. The recognition of emotion. In Wahlster and Verbmobil, editors, *Wahlster and Verbmobil*, Foundations of Speech-to-Speech Translations. Springer, 2000.

[14] B. Bauer and H. Hienz. Relevant features for video-based continuous sign language recognition. *Automatic Face and Gesture Recognition*, 00:440, 2000.

[15] B. Bauer, H. Hienz, and KF. Kraiss. Video-based continuous sign language recognition using statistical methods. *International Conference on Pattern Recognition*, 02:2463, 2000.

[16] B. Bauer and KF. Kraiss. Video-based sign recognition using self-organizing subunits. *International Conference on Pattern Recognition*, 02:20434, 2002.

[17] U. Bellugi and S. Fischer. A comparison of sign language and spoken language: Rate and grammatical mechanisms. *Cognition*, 1:173–200, 1972.

[18] B. Bergman and J. Mesch. *ECHO Data Set for Swedish Sign Language (SSL).* Department of Linguistics, University of Stockholm, http://www.let.ru.nl/, 2004.

[19] Blaxxun. Blaxxun contact 5.

[20] R. T. Boone and J. G. Cunningham. Children's decoding of emotion in expressive body movement: The development of cue attunemen. *Developmental Psychology*, 34:1007–1016, 1998.

[21] S. Boutsis, P. Prokopidis, V. Giouli, and S. Piperidis. Robust parser for unrestricted greek text. In *2nd Language Resources and Evaluation Conference*, 2000.

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

[23] A. Braffort, R. Gherbi, S. Gibet, J. Richardson, and D. Teil. *Gesture-based communication in human-computer interaction.* Springer New York, 1999.

[24] H. Brashear, T. Starner, P. Lukowicz, and H. Junker. Using multiple sensors for mobile sign language recognition. In *Proceedings. Seventh IEEE International Symposium on Wearable Computers*, pages 45–52, 2005.

[25] D. Brien and M. Brennan. *Dictionary of British Sign Language.* Faber and Faber, Boston, 1992.

[26] J. Bungeroth, D. Stein, P. Dreuw, H. Ney, S. Morrissey, A. Way, and L. van Zijl. The atis sign language corpus. In *International Conference on Language Resources and Evaluation*, Marrakech, Morocco, May 2008.

[27] A. Camurri, P. Coletta, A. Massari, B. Mazzarino, M. Peri, M. Ricchetti, A. Ricci, and G. Volpe. Toward real-time multimodal processing: Eyesweb 4.0. In *Artificial Intelligence and Simulation of Behaviour Convention: Motion, Emotion and Cognition*, 2004.

[28] A. Camurri, B. Mazzarino, and G. Volpe. Analysis of expressive gesture: The eyesweb expressive gesture processing library. In G.Volpe A. Camurri, editor, *Gesture-based Communication in Human-Computer Interaction*. Springer Verlag, 2004.

[29] U. Canzler and T. Dziurzyk. Extraction of non manual features for videobased sign language recognition. *International Association of Pattern Recognition*, pages 318–321, 2002.

[30] G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Raouzaiou, and K. Karpouzis. Modeling naturalistic affective states via facial and vocal expressions recognition. In *International Conference on Multimodal Interfaces*, 2006.

[31] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[32] G. Castellano, A. Camurri, B. Mazzarino, and G. Volpe. A mathematical model to analyse the dynamics of gesture expressivity. In *Artificial Intelligence and Simulation of Behaviour Convention: Artificial and Ambient Intelligence*, 2007.

[33] T.L. Chartrand, W. Maddux, and J. Lakin. Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. Hassin, J. Uleman, and J.A. Bargh, editors, *The New Unconscious*, pages 334–361. NY: Oxford University Press, 2005.

[34] S.P. Chatzis, D.I. Kosmopoulos, and T.A. Varvarigou. Robust Sequential Data Modeling Using an Outlier Tolerant Hidden Markov Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 2008.

[35] L. Chen and T. S. Huang. Emotional expressions in audiovisual human computer interaction. In *International Conference on Multimedia & Expo*, 2000.

[36] L.S. Chen, T. S. Huang, Miyasato T., and Nakatsu R. Multimodal human emotion / expression recognition. In *Automatic Face and Gesture Recognition*, 1998.

[37] I. Cohen, FG Cozman, N. Sebe, MC Cirelo, and TS Huang. Semisupervised learning of classifiers: Theory, algorithms, and their application to human-computer interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1553–1566, 2004.

[38] I. Cohen, N. Sebe, A. Garg, L.S. Chen, and T.S. Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, 91:160–187, 2003.

[39] P. Cohen. Multimodal interaction: A new focal area for ai. In *International Joint Conferences on Artificial Intelligence*, 2001.

[40] P. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Clow, and I. Smith. The efficiency of multimodal interaction: A case study. In *International Conference on Spoken Language Processing*, 1998.

[41] J.F. Cohn. Foundations of human computing: Facial expression and emotion. *Lecture Notes in Computer Science*, 4451:1, 2007.

[42] JF Cohn and K. Schmidt. The timing of facial motion in posed and spontaneous smiles. *Active Media Technology*, page 57, 2003.

[43] T. Coogan, G. Awad, J. Han, and A. Sutherland. Real time hand gesture recognition including hand segmentation and tracking. *Advances in Visual Computing*, pages 495–504, 2006.

[44] H M Cooper and R Bowden. Large lexicon detection of sign language. *International Conference on Computer Vision Workshop Human Computer Interaction*, 4796:88–97, 2007.

[45] H M Cooper and R Bowden. Sign language recognition using boosted volumetric features. In *International Association for Pattern Recognition Conference on Machine Vision Applications*, 2007.

[46] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *Pattern Analysis and Machine Intelligence*, 23:681–685, 2001.

[47] G. Cortes, L. Garcia, Carmen Benitez, and Jose C. Segura. Hmm-based continuous sign language recognition using a fast optical flow parameterization of visual information. In *INTERSPEECH-2006*, 2006.

[48] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schroeder. Feeltrace: An instrument for recording perceived emotion in real time. In *International Speech Communication Association Workshop on Speech and Emotion*, 2000.

[49] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and JG Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

[50] O. Crasborn, E. van der Kooij, A. Nonhebel, and W. Emmerik. *ECHO Data Set for Sign Language of the Netherlands (NGT)*. Department of Linguistics, Radboud University Nijmegen, 2004.

[51] Y. Cui and J. Weng. A learning-based prediction-and-verification segmentation scheme for hand sign image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):798–804, 1999.

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

[53] P. Dalle. High level models for sign language analysis by a vision system. *Language*, pages 17–20, 2006.

[54] F. De-Rosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. De Carolis. From greta's mind to her face: modeling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59:81–118, 2003.

[55] M. DeMeijer. The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13:247–268, 1989.

[56] K. Derpanis, R. Wildes, and J. Tsotsos. Hand Gesture Recognition within a Linguistics-Based Framework. In *8th European Conference on Computer Vision*, Prague, Czech Republic, May 11-14 2004. Springer.

[57] L.C. DeSilva and N. Pei Chi. Bimodal emotion recognition. In *Face and Gesture Recognition Conference*, 2000.

[58] P. R. DeSilva, A. Kleinsmith, and N. Bianchi-Berthouze. Towards unsupervised detection of affective body posture nuances. In *1st International Conference on Affective Computing and Intelligent Interaction*, 2005.

[59] L. Devillers, S. Abrilian, and J.-C. Martin. Representing real life emotions in audiovisual data with non basic emotional patterns and context features. In *First International Conference on Affective Computing and Intelligent Interaction*, 2005.

[60] L. Devillers and L. Vidrascu. Real-life emotion recognition human-human call center data with acoustic and lexical cues. In Christian Mller Susanne, editor, *Speaker characterization*, chapter Lecture Notes in Computer Science, pages 34–42. Springer,Verlag, 2007.

[61] L. Devillers, L. Vidrascu, and L. Lamel. Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 18(4):407–422, 2005.

[62] O. Diamanti and P. Maragos. Geodesic Active Regions For Segmentation and Tracking of Human Gestures in Sign Language Videos. In *15th IEEE International Conference on Image Processing, 2008.*, pages 1096–1099, 2008.

[63] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach. Emotional speech: towards a new generation of databases. *Speech Communication*, 40:33–60, 2003.

[64] N. Doulamis, A. Doulamis, and S. Kollias. On-line retrainable neural networks: Improving performance of neural networks in image analysis problems. *IEEE Transactions on Neural Networks*, 11:1–20, 2000.

[65] P. Dreuw, C. Neidle, V. Athitsos, S. Sclaroff, and H. Ney. Benchmark databases for video-based automatic sign language recognition. In *International Conference on Language Resources and Evaluation*, Marrakech, Morocco, May 2008.

[66] Z. Duric, W.D. Gray, R. Heishman, F. Li, A. Rosenfeld, M.J. Schoelles, C. Schunn, and H. Wechsler. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. *Proceedings of the IEEE*, 90(7):1272–1289, 2002.

[67] E. Efthimiou, A. Vacalopoulou, S-E. Fotinea, and G. Steinhauer. A multipurpose design and creation of gsl dictionaries. In *Workshop on the Representation and Processing of Sign Languages, From SignWriting to Image Processing, LREC-2004*, 2004.

[68] Eleni Efthimiou and Stavroula-Evita Fotinea. *Universal Acess in Human Computer Interaction. Coping with Diversity*, chapter GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI, pages 657–666. Springer Berlin / Heidelberg, 2007.

[69] P. Ekman. Basic emotions. In T. Dalgleish and M. J. Power, editors, *Handbook of Cognition and Emotion*, pages 301–320. John Wiley, New York, 1999.

[70] P. Ekman and W. Friesen. The repertoire of nonverbal behavioral categories: origins, usage and coding. *Semiotica*, 1:49–98, 1969.

[71] P. Ekman and W. Friesen. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6:238–252, 1982.

[72] P. Ekman, W. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, WA LeCompte, T. Pitcairn, and PE Ricci-Bitti. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4):712, 1987.

[73] P. Ekman and W. V. Friesen. *The facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, 1978.

[74] P. Ekman, W.V. Friesen, and P.C. Ellsworth. *Emotion in the human face.* Cambridge University Press Cambridge, 1982.

[75] P. Ekman, W.V. Friesen, and J.C. Hager. *Facial action coding system.* Consulting Psychologists Press Palo Alto, CA, 1978.

[76] P. Ekman and H. Oster. Facial expressions of emotion. *Annual Review of Psychology*, 30(1):527–554, 1979.

[77] P. Ekman and E.L. Rosenberg. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, USA, 2005.

[78] ELAN. Max planck institute for psycholinguistics.

[79] J. L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[80] J. L. Elman. Distributed representations, simple recurrent networks, and grammatical structure. , 7, 195-224. *Machine Learning*, 7:195–224, 1991.

[81] ERMIS. Emotionally Rich Man-machine Intelligent System IST-2000-29319.

[82] I.A. Essa and A.P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19:757–763, 1997.

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

[84] G. Fang, W. Gao, and D. Zhao. Large vocabulary sign language recognition based on fuzzy decision trees. *IEEE Transactions on Systems, Man and Cybernetics*, 34:305–314, 2004.

[85] G. Fang, X. Gao, W. Gao, and Y Chen. A novel approach to automatically extracting basic units from chinese sign language. *Pattern Recognition*, pages 454–457, 2004.

[86] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003.

[87] I. Fasel, B. Fortenberry, and J. R. Movellan. A generative framework for real-time object detection and classification. *Computer Vision and Image Understanding*, 98:182–210, 2005.

[88] C.N. Fiechter. Efficient reinforcement learning. In *Proceedings of the seventh annual conference on Computational learning theory*, pages 88–97. ACM New York, NY, USA, 1994.

[89] H. Fillbrandt, S. Akyol, and K.-F. Kraiss. Extraction of 3d hand shape and posture from image sequences for sign language recognition. In *International Workshop on Analysis and Modeling of Faces and Gestures*, pages 181–186, 2003.

[90] R.A. Foulds. Biomechanical and perceptual constraints on the bandwidth requirements of sign language. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering]*, 12(1):65–72, March 2004.

[91] N. Fragopanagos and J. G. Taylor. Emotion recognition in human computer interaction. *Neural Networks*, 18:389–405, 2005.

[92] M. G. Frank and P. Ekman. Not all smiles are created equal: The differences between enjoyment and other smiles. *The International Journal for Research in Humor*, 6:9–26, 1993.

[93] R. Fransens and Jan De Prins. Svm-based nonparametric discriminant analysis, an application to face detection. In *9th IEEE International Conference on Computer Vision*, 2003.

[94] K.and Xia Liu Fujimura. Sign recognition using depth image streams. *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 381–386, 2006.

[95] G. Gao, W.and Fang, D. Zhao, and Y. Chen. A chinese sign language recognition system based on sofm/srn/hmm. *Pattern Recognition*, 37(12):2389–2402, 2004.

[96] G. Gao, W.and Fang, D. Zhao, and Y. Chen. Transition movement models for large vocabulary continuous sign language recognition. *Automatic Face and Gesture Recognition*, pages 553–558, 2004.

[97] W. Gao, J.Y. Ma, J.Q. Wu, and C.L. Wang. Sign language recognition based on HMM/ANN/DP. *International Journal on Pattern Recognition and Artificial Intelligence*, vol. 14, no. 5:587–602, 2000.

[98] H.J. Go, K.C. Kwak, D.J. Lee, and M.G. Chun. Emotion recognition from the facial image and speech signal. In *SICE 2003 Annual Conference*, volume 3, 2003.

[99] M.K. Greenwald, E.W. Cook, and P.J. Lang. Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology*, 3(1):51–64, 1989.

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

[101] H. Gunes and M. Piccardi. Fusing face and body display for bi-modal emotion recognition: Single frame analysis and multi-frame post integration. In *1st International Conference on Affective Computing and Intelligent Interaction*, 2005.

[102] M. T. Hagan and M. Menhaj. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5:989–993, 1994.

[103] B. Hammer and P. Tino. Recurrent neural networks with small weights implement definite memory machines. *Neural Computation*, 15:1897–1929, 2003.

[104] B. Hartmann, M. Mancini, and C. Pelachaud. Formational parameters and adaptive prototype instantiation for mpeg-4 compliant gesture synthesis. In *Computer Animation*, 2002.

[105] B. Hartmann, M. Mancini, and C. Pelachaud. Implementing expressive gesture synthesis for embodied conversational agents. In *Gesture Workshop*, 2005.

[106] S. Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall International, 1999.

[107] J.L. Hernandez-Rebollar, N. Kyriakopoulos, and R.W. Lindeman. A new instrumented approach for translating american sign language into sound and text. *Automatic Face and Gesture Recognition, 2004*, pages 547–552, 2004.

[108] H. Hienz, B. Bauer, and K.F. Kraiss. Hmm-based continuous sign language recognition using stochastic grammars. In *Gesture Workshop*, page 185, 1999.

[109] S. Hoch, F. Althoff, G. McGlaun, and G. Rigoll. Bimodal fusion of emotional data in an automotive environment. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, 2005.

[110] M. Hossain and M. Jenkin. Recognizing hand-raising gestures using hmm. In *CRV '05: Proceedings of the 2nd Canadian conference on Computer and Robot Vision*, pages 405–412, Washington, DC, USA, 2005.

[111] CL Huang and WY Huang. Sign language recognition using model-based tracking and a 3d hopfield neural network. *Machine Vision and Applications*, Volume 10, Numbers 5-6 / April, 1998:292–307, 1998.

[112] CL Huang, MS Wu, and SH Jeng. Gesture recognition using the multi-pdm method and hidden markov model. *Image and Vision Computing*, 18,11:865–879, 2000.

[113] Z. Huang, A. Eliens, and C. Visser. Step: A scripting language for embodied agents. In *Workshop on Lifelike Animated Agents*, 2002.

[114] Humaine. Network of excellence on emotions.

[115] K. Imagawa, S. Lu, and S. Igi. Color-based hands tracking system for sign language recognition. In *3rd. International Conference on Face & Gesture Recognition*, volume 00, page 462, Los Alamitos, CA, USA, 1998.

[116] K. Imagawa, H. Matsuo, R. Taniguchi, D. Arita, S. Lu, and S. Igi. Recognition of local features for camera-based sign language recognition system. *International Conference on Pattern Recognition*, 04:4849, 2000.

[117] I. Infantino, R. Rizzo, and S. Gaglio. A framework for sign language sentence recognition by commonsense context. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(5):1034–1039, 2007.

[118] S. Ioannou, G. Caridakis, K. Karpouzis, and S. Kollias. Robust feature detection for facial expression recognition. *EURASIP Journal on Image and Video Processing*, 2007.

[119] S. Ioannou, A. Raouzaiou, V. Tzouvaras, T. Mailis, K. Karpouzis, and S. Kollias. Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Networks*, 18(4):423–435, May 2005.

[120] A. Jaimes. Human-centered multimedia: Culture, deployment, and access. *IEEE Multimedia Magazine*, 13, 2006.

[121] A. Jaimes and N. Sebe. Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*, 108(1-2):116–134, 2007.

[122] A. Jaimes and N. Sebe. Multimodal human computer interaction: A survey. *Computer Vision and Image Understanding journal, special issue on Human-Computer Interaction*, 2007.

[123] F. Jiang, H. Yao, and G. Yao. Multilayer architecture in sign language recognition system. In *6th international conference on Multimodal interfaces*, 2004.

[124] B. Juang, L. Rabiner, S. Levinson, and M. Sondhi. Recent developments in the application of hidden markov models to speaker-independent isolated word recognition. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85.*, 10:9–12, Apr 1985.

[125] P. Juslin and K. Scherer. Vocal expression of affect. In J. Harrigan and K. Rosenthal, R.and Scherer, editors, *The New Handbook of Methods in Nonverbal Behavior Research*. Oxford University Press, 2005.

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

[127] A. Kapoor, W. Burleson, and R.W. Picard. Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8):724–736, 2007.

[128] A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. F. Driessen. Gesture-based affective computing on motion capture data. In *1st International Conference on Affective Computing and Intelligent Interaction, ACII'2005*, 2005.

[129] K. Karpouzis, G. Caridakis, L. Kessous, N. Amir, A. Raouzaiou, L. Malatesta, and S. Kollias. Modeling naturalistic affective states via facial, vocal, and bodily expressions recognition. *Lecture Notes in Computer Science*, 4451:91, 2007.

[130] K. Karpouzis, A. Raouzaiou, and S. Kollias. Moving avatars: emotion synthesis in virtual worlds. *Human – Computer Interaction: Theory and Practice, Lawrence Erlbaum Associates*, 2:503–507, 2003.

[131] J. Kelley. *Natural Language and computers: Six empirical steps for writing an easy-to-use computer application*. PhD thesis, The Johns Hopkins University, 1983.

[132] A Kendon. *Conducting Interaction*. Cambridge, University Press, 1990.

[133] R. Kennaway. Synthetic animation of deaf signing gestures. In *International Gesture Workshop*, 2001.

[134] R. Kennaway. Experience with, and requirements for, a gesture description language for synthetic animation. In *5th International Workshop on Gesture and Sign Language based Human-Computer Interaction*, 2003.

[135] L. Kessous and N. Amir. Comparison of feature extraction approaches based on the bark time/frequency representation for classification of expressive speech. In *Interspeech*, 2007.

[136] M. Kipp. *Gesture Generation by Imitation. From Human Behavior to Computer Character Animation.* Dissertation.com, 2004.

[137] D.H.U. Kochanek and R.H. Bartels. Interpolating splines with local tension, continuity, and bias control. In *Computer Graphics, SIGGRAPH'84*, 1984.

[138] I. Kononenko. On biases in estimating multi-valued attributes. In *14th International Joint Conference on Articial Intelligence*, 1995.

[139] V. Kourbetis. Noima stin ekpaidefsi (in greek), 1999. Hellenic Pedagogical Institute, Athens.

[140] J. Lakin, V. Jefferis, C. Cheng, and T. Chartrand. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27:145–162, 2003.

[141] K. Laskowski and S. Burger. Annotation and analysis of emotionally relevant behavior in the ISL Meeting Corpus. *Proc. LREC, Genoa, Italy*, 2006.

[142] C.M. Lee and SS Narayanan. Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, 13(2):293–303, 2005.

[143] Y. Lee and S. Kassam. Generalized median filtering and related nonlinear filtering techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(3):672–683, 1985.

[144] Yung-Hui Lee and Cheng-Yueh Tsai. Taiwan sign language (tsl) recognition based on 3d data and neural networks. *Expert Systems with Applications*, 2007.

[145] L. Leeson and B. Nolan. Digital Deployment of the Signs of Ireland Corpus in Elearning. In *Proceedings of the Language Resources in Education Conference (LREC), Marrakesh, May*, 2008.

[146] S. H. Leung, S. L. Wang SL, and W. H. Lau. Lip image segmentation using fuzzy clustering incorporating an elliptic shape function. *IEEE Trans. on Image Processing*, 13, 2004.

[147] Ying li Tian, Takeo Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on PAMI*, 23, 2001.

[148] RH Liang and M. Ouhyoung. A real-time continuous gesture recognition system for sign language. In *Automatic Face and Gesture Recognition*, pages 558–567, 1998.

[149] S. Liddell. *American Sign Language Syntax.* The Hague: Mouton, 1980.

[150] C.L. Lisetti and F. Nasoz. MAUI: a multimodal affective user interface. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 161–170. ACM New York, NY, USA, 2002.

[151] G. Littlewort, M. Bartlett, and K. Lee. Faces of pain: automated measurement of spontaneousallfacial expressions of genuine and posed pain. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 15–21. ACM New York, NY, USA, 2007.

[152] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24:615–625, 2006.

[153] J. Ma, W. Gao, and R. Wang. A parallel multistream model for integration of sign language recognition and lip motion. In *ICMI*, pages 582–589, 2000.

[154] L. Maat and M. Pantic. Gaze-X: Adaptive, Affective, Multimodal Interface for Single-User Office Scenarios. *Lecture Notes in Computer Science*, 4451:251, 2007.

[155] L. Malatesta, A. Raouzaiou, K. Karpouzis, and S. Kollias. Towards modelling embodied conversational agent character profiles using appraisal theory predictions in expression synthesis. *Applied Intelligence*, 30:58–64, 2009.

[156] V.-M. Mantyla, J. Mantyjarvi, T. Seppanen, and E. Tuulari. Hand gesture recognition of a mobile device user. In *IEEE International Conference on Multimedia and Expo*, 2000.

[157] J.-C. Martin, S. Abrilian, L. Devillers, M. Lamolle, M. Mancini, and C. Pelachaud. Levels of representation in the annotation of emotion for the specification of expressivity in ecas. In *International Working Conference on Intelligent Virtual Agents*, 2005.

[158] A.M. Martinez, R.B. Wilbur, R. Shay, and A.C. Kak. Purdue rvl-slll asl database for automatic recognition of american sign language. *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, pages 167–172, 2002.

[159] D. McNeill. *Hand and mind - what gestures reveal about thoughts*. University of Chicago Press, 1992.

[160] A. Mehrabian. Communication without words. *Psychol.Today*, 2:53–56, 1968.

[161] P. Mertens. The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In *Speech Prosody*, 2004.

[162] B.W. Miners, O.A. Basir, and M. Kamel. Knowledge-based disambiguation of hand gestures. *Systems, Man and Cybernetics, 2002 IEEE International Conference on*, 5:6 pp. vol.5–, Oct. 2002.

[163] R. Mitchell, T. Young, B. Bachleda, and M. Karchmer. How many people use asl in the united states? why estimates need updating. *Sign Language Studies*, 6, 2006.

[164] S. Mitra and T. Acharya. Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(3):311–324, 2007.

[165] MPEG-4. Mpeg-4 home page.

[166] D. Neiberg, K. Elenius, I. Karlsson, and K. Laskowski. Emotion recognition in spontaneous speech. In *Fonetik*, 2006.

[167] C.J. Neidle. *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. MIT Press, 2000.

[168] J. Newlove. *Laban for actors and dancers*. New York: Routledge, 1993.

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

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

[171] P.Y. Oudeyer. The production and recognition of emotions in speech: features and algorithms. *International Journal of Human-Computer Studies*, 59(1-2):157–183, 2003.

[172] S. Oviatt. Ten myths of multimodal interaction. *Communications of the ACM*, 42:74–81, 1999.

[173] C. Padden and C. Ramsey. *Language Acquisition by Eye*, chapter American Sign Language and Reading Ability in Deaf Children, pages 165–189. Lawrence Erlbaum Associates, 2000.

[174] P. Pal, AN Iyer, and RE Yantorno. Emotion detection from infant facial expressions and cries. In *2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings*, volume 2, 2006.

[175] M. Pantic. Face for interface. In M. Pagani and Ed. Hershey, editors, *The Encyclopedia of Multimedia Technology and Networking*, pages 208–314. Idea Group, 2005.

[176] M. Pantic and MS Bartlett. Machine analysis of facial expressions. *Face recognition*, 2007.

[177] M. Pantic and I. Patras. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man and Cybernetics*, 36:433–449, 2006.

[178] M. Pantic and L.J.M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[179] M. Pantic and L.J.M. Rothkrantz. Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 18:881–905, 2000.

[180] M. Pantic and L.J.M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.

[181] M. Pantic, N. Sebe, J. Cohn, and T.S. Huang. Affective multimodal human-computer interaction. *ACM Multimedia*, 20:669–676, November 2005.

[182] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *International Conference on Computer Vision*, 1998.

[183] N. Paragios and R. Deriche. Geodesic active regions: A new framework to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1/2):249–268, March 2002.

[184] D. Park, M. A. EL-Sharkawi, and R. J. Marks II. An adaptively trained neural network. *IEEE Transactions on Neural Networks*, 2:334–345, 1991.

[185] V. Pashaloudi and K. Margaritis. A performance study of a recognition system for greek sign language alphabet letters. In *International Conference "Speech and Computer"*, 2004.

[186] C. Pateritsas, M. Pertselakis, and A. Stafylopatis. A SOM-based classifier with enhanced structure learning. In *2004 IEEE International Conference on Systems, Man and Cybernetics*, volume 5, 2004.

[187] C. Pelachaud and M. Bilvi. Computational model of believable conversational agents. *Communication in Multiagent Systems, Lecture Notes in Computer Science*, 2650:300–317, 2003.

[188] A. Pentland. Socially aware computation and communication. *IEEE Computer*, 38:33–40, 2005.

[189] S. Perrin, A. Cassinelli, and M. Ishikawa. Gesture recognition using laser-based tracking system. *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 541–546, May 2004.

[190] M. Pertselakis and A. Stafylopatis. Dynamic modular fuzzy neural classifier with tree-based structure identification. *Neurocomputing*, 71(4-6):801–812, 2008.

[191] S. Petridis and M. Pantic. Audiovisual discrimination between laughter and speech. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pages 5117–5120, 2008.

[192] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1090–1104, 2000.

[193] R. Picard. *Affective Computing*. The MIT Press, Cambridge, 1997.

[194] R. W. Picard. Towards computers that recognize and respond to user emotion. *IBM Syst. Journal*, 39:705–719, 2000.

[195] S. Prillwitz, R. Leven, H. Zienert, T. Hanke, and J. Henning. *HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide*, 1989.

[196] A. Raouzaiou, N. Tsapatsoulis, K. Karpouzis, and S. Kollias. Parameterized facial expression synthesis based on mpeg-4. *EURASIP Journal on Applied Signal Processing*, 2002:1021–1038, 2002.

[197] K. Rapantzikos and Y. Avrithis. An enhanced spatiotemporal visual attention model for sports video analysis. In *International Workshop on content-based Multimedia indexing, CBMI*, 2005.

[198] A. Rogozan. Discriminative learning of visual data for audiovisual speech recognition. *International Journal Artificial Intelligent Tools*, 8:43–52, 1999.

[199] J.A. Russell, J.A. Bachorowski, and J.M. Fernandez-Dols. Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54(1):329–349, 2003.

[200] J.A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294, 1977.

[201] H. Sagawa and M. Takeuchi. A method for recognizing a sequence of sign language words represented in a japanese sign language sentence. *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 434–439, 2000.

[202] K. Sage, J. Howell, and H. Buxton. *Gesture-Based Communication in Human-Computer Interaction*, chapter Developing Context Sensitive HMM Gesture Recognition, page 3853. Springer Berlin / Heidelberg, 2004.

[203] A. Samal and P.A. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern recognition*, 25(1):65–77, 1992.

[204] D. Sander, D. Grandjean, and K.R. Scherer. A systems approach to appraisal mechanisms in emotion. *Neural networks*, 18(4):317–352, 2005.

[205] G. Sapountzaki, E. Efthimiou, C. Karpouzis, and V. Kourbetis. Open-ended resources in greek sign language: Development of an e-learning platform. In *Workshop on the Representation and Processing of Sign Languages, From SignWriting to Image Processing, LREC-2004*, 2004.

[206] Y. Sato, T. Ogawa, and T. Kobayashi. Extension of Hidden Markov Models for Multiple Candidates and Its Application to Gesture Recognition. *IEICE Trans Inf Syst*, E88-D(6):1239–1247, 2005.

[207] A. M. Schaefer and H. G. Zimmermann. Recurrent neural networks are universal approximators. In *ICANN*, 2006.

[208] K. R. Scherer and H. G.: Wallbott. *Analysis of Nonverbal Behavior*, chapter Handbook of discourse analysis. Handbook of Discourse: Analysis. Academic Press London, 1985.

[209] K.R. Scherer. Appraisal theory. *Handbook of cognition and emotion*, pages 637–663, 1999.

[210] B. Schuller, R. Müeller, B. Höernler, A. Höethker, H. Konosu, and G. Rigoll. Audiovisual recognition of spontaneous interest within conversations. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 30–37. ACM New York, NY, USA, 2007.

[211] N. Sebe, I. Cohen, T. Gevers, and T.S. Huang. Emotion recognition based on joint visual and audio cues. In *Proceedings of the 18th International Conference on Pattern Recognition*, pages 1136–1139, 2006.

[212] N. Sebe, I. Cohen, and T.S. Huang. Multimodal emotion recognition. *Handbook of Pattern Recognition and Computer Vision*, pages 981–256, 2005.

[213] N. Sebe, M.S. Lew, I. Cohen, Y. Sun, T. Gevers, and T.S. Huang. Authentic facial expression analysis. In *International Conference on Automatic Face and Gesture Recognition*, 2004.

[214] T. Shanableh, K. Assaleh, and M. Al-Rousan. Spatio-temporal feature-extraction techniques for isolated gesture recognition in arabic sign language. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, 37(3):641–650, 2007.

[215] M. Song, J. Bu, C. Chen, and N. Li. Audio-visual based emotion recognition-a new approach. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, 2004.

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

[217] S. Steidl, M. Levit, A. Batliner, E. Noth, and H. Niemann. Of All Things the Measure Is Man: Automatic Classification of Emotions and Inter-Labeler Consistency. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05)*, volume 1, 2005.

[218] W. Stokoe. Sign language structure. *Annual Review of Anthropology*, 9:365–390, 1980.

[219] Mu-Chun Su. A fuzzy rule-based approach to spatio-temporal hand gesture recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 30(2):276–281, 2000.

[220] N. Tanibata, N. Shimada, and Y. Shirai. Extraction of hand features for recognition of sign language words. In *The 15th International Conference on Vision Interface*, 2002.

[221] P. Teissier, J. Robert-Ribes, and J. L. Schwartz. Comparing models for audiovisual fusion in a noisy-vowel recognition task. *IEEE Trans. Speech Audio Processing*, 7:629–642, 1999.

[222] A. Tekalp and J. Ostermann. Face and 2-d mesh animation in mpeg-4. *Signal Processing: Image Communication*, 15:387–421, 2000.

[223] S. Theodorakis. Isolated greek sing language recognition using hidden markov models. Master's thesis, School of Electrical and Computer Engineering of the National Technical University of Athens, 2008.

[224] S. Theodorakis, A. Katsamanis, and P. Maragos. Product-hmms for automatic sign language recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2009), Taipei, Taiwan, Apr.*, 2009.

[225] Y.L. Tian, T. Kanade, and J.F. Cohn. Facial expression analysis. *Handbook of face recognition*, pages 247–276, 2005.

[226] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical report, Carnegie Mellon University, 1991.

[227] S.S. Tomkins and B.P. Karon. *Affect, imagery, consciousness*. Springer New York, 1962.

[228] M.F. Valstar, H. Gunes, and M. Pantic. How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 38–45. ACM New York, NY, USA, 2007.

[229] M.F. Valstar, M. Pantic, Z. Ambadar, and J.F. Cohn. Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 162–170. ACM New York, NY, USA, 2006.

[230] P. Vamplew and A. Adams. Recognition of sign language gestures using neural networks. *Australian Journal of Intelligent Information Processing Systems*, 5:94–102, 1998.

[231] L. van Swol. The effects of nonverbal mirroring on perceived persuasiveness, agreement with an imitator, and reciprocity in a group discussion. *Communication Research*, 30:461–480, 2003.

[232] P. Viola and M. Jones. Robust real-time face detection. *IEEE International Conference on Computer Vision*, 02:747, 2001.

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

[234] C. Vogler and S. Goldenstein. Facial movement analysis in asl. *Universal Access in the Information Society*, 6:363–374, 2008.

[235] C. Vogler and D. Metaxas. Parallel hidden markov models for american sign language recognition. *International Conference on Computer Vision*, 01:116, 1999.

[236] Christian Vogler and Dimitris N. Metaxas. Handshapes and movements: Multiple-channel american sign language recognition. In *Gesture Workshop*, pages 247–258, 2003.

[237] M. Wallace, S. Ioannou, A. Raouzaiou, K. Karpouzis, and S. Kollias. Dealing with feature uncertainty in facial expression recognition using possibilistic fuzzy rule evaluation. *International Journal of Intelligent Systems Technologies and Applications*, 1, 2006.

[238] H. G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.

[239] H. G. Wallbott and K. R. Scherer. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 514:690–699, 1986.

[240] C. Wang, W. Gao, and S. Shan. An approach based on phonemes to large vocabulary chinese sign language recognition. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 393–398, 2002.

[241] H. Wang, C. Ming, and C. Oz. American sign language recognition using multi-dimensional hidden markov models. *Journal of Information Science and Engineering*, 22, 5:1109–1123, 2006.

[242] J. Wang and W. Gao. A fast sign word recognition method for chinese sign language. *Lecture Notes in Computer Science, Advances in Multimodal Interfaces ŷ ICMI 2000*, 1948/2000:599–606, 2000.

[243] Q. Wang, X. Chen, C. Wang, and W. Gao. Sign language recognition from homography. In *IEEE International Conference on Multimedia and Expo*, pages 429–432, 2006a.

[244] Y. Wang and L. Guan. Recognizing human emotion from audiovisual information. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05)*, volume 2, 2005.

[245] D. Watson, LA Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6):1063, 1988.

[246] C. Whissel. The dictionary of affect in language. In R. Plutchnik and H. Kellerman, editors, *Emotion: Theory, Research and Experience: The Measurement of Emotions*, Emotion: Theory, Research and Experience: The Measurement of Emotions, pages 113–131. Academic Press, 1989.

[247] A.C.C. Williams. Facial expression of pain: An evolutionary account. *Behavioral and brain sciences*, 25(04):439–455, 2003.

[248] G.W. Williams. Comparing the joint agreement of several raters with another rater. *Biometrics*, 32:619–627, 1976.

[249] A. Wilson and A. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(9), 1999.

[250] I.H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques, 2nd Edition*. Morgan Kaufmann, San Francisco, CA, 2005.

[251] B. Woll. Sutton-Spence, and D. Waters, 2004. *ECHO Data Set for British Sign Language (BSL)*, 2004.

[252] L. Wu, S. Oviatt, and P. Cohen. Multimodal integration: A statistical view. *IEEE Transactions on Multimedia*, 1:334–341, 1999.

[253] Y. Wu and T. Huang. Hand modeling, analysis, and recognition for vision-based human computer interaction. *IEEE Signal Processing Magazine*, 18:51–60, 2001.

[254] Y. Wu and M. Takatsuka. The Geodesic Self-Organizing Map and its error analysis. In *Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38*, pages 343–351. Australian Computer Society, Inc. Darlinghurst, Australia, Australia, 2005.

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

[256] Ho-Sub Yoon, Younglae J. Baea Jung Soha, and Hyun Seung Yangb. Hand gesture recognition using combined features of location, angle and velocity. *Pattern Recognition*, 24,7:1491–1501, 2001.

[257] Y. Yoshitomi, S. Kim, T. Kawano, and T. Kitazoe. Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face. In *ROMAN*, 2000.

[258] J. W. Young. Head and face anthropometry of adult u.s. civilians. Technical report, FAA Civil Aeromedical Institute, 1993.

[259] M. Zahedi, P. Dreuw, D. Rybach, T. Deselaers, and H. Ney. Continuous sign language recognition - approaches from speech recognition and available data resources. In *LREC Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios*, pages 21–24, Genoa, Italy, May 2006.

[260] M. Zahedi, D. Keysers, and H. Ne. Appearance-based recognition of words in american sign language. In *Pattern Recognition and Image Analysis*, 2005.

[261] Z. Zeng, Y. Hu, M. Liu, Y. Fu, and T.S. Huang. Training combination strategy of multi-stream fused hidden Markov model for audio-visual affect recognition. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 65–68. ACM New York, NY, USA, 2006.

[262] Z. Zeng, Y. Hu, G.I. Roisman, Z. Wen, Y. Fu, and T.S. Huang. Audio-visual spontaneous emotion recognition. *Lecture Notes in Computer Science*, 4451:72, 2007.

[263] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.

[264] Z. Zeng, J. Tu, M. Liu, T.S. Huang, B. Pianfetti, Roth D., and S. Levinson. Audio-visual affect recognition. *IEEE Transactions on Multimedia*, 9:424–428, 2007.

[265] Z. Zeng, J. Tu, M. Liu, T. Zhang, N. Rizzolo, Z. Zhang, T.S. Huang, D. Roth, and S. Levinson. Bimodal HCI-related affect recognition. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 137–143. ACM New York, NY, USA, 2004.

[266] Z. Zeng, J. Tu, B. Pianfetti, M. Liu, T. Zhang, Z. Zhang, TS Huang, and S. Levinson. Audio-visual affect recognition through multi-stream fused HMM for HCI. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, volume 2, 2005.

[267] Chen-Xi Zhang, Hong-Xun Yao, Feng Jiang, De-Bin Zhao, and Xiao-Ting Sun. Multilayer method based on multi-resolution feature extracting and mvc dimension reducing method for sign language recognition. *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, 7:4452–4457 Vol. 7, 2005.

[268] Liang-Guo Zhang, Yiqiang Chen, Gaolin Fang, Xilin Chen, and Wen Gao. A vision-based sign language recognition system using tied-mixture density hmm. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 198–204, 2004.

[269] Liangguo Zhang, Gaolin Fang, Wen Gao, Xilin Chen, and Yiqiang Chen. Vision-based sign language recognition using sign-wise tied mixture hmm. *Advances in Multimedia Information Processing PCM*, pages 1035–1042, 2004a.

[270] J. Zieren and KF. Kraiss. Robust person-independent visual sign language recognition. In *Pattern recognition and image analysis*, 2005.