ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
Μεταπτυχιακό Ηλεκτρονικού Αυτοματισμού

Γιαννόπουλος Εμμανουήλ - Ρίνης Εμμανουήλ
Α.Μ. 97507 Α.Μ. 97530

1. Πολυτροπική Διεπαφή Ανθρώπου-Υπολογιστή

2. Ομιλούντα Πρόσωπα και Αναγνωριστές Ομιλίας Που Μπορούν να Δουν: Η Επεξεργασία Εικόνας Ομιλίας με Υπολογιστή

Εργασίες στο μάθημα: Επικοινωνία με Ομιλία
Διδάσκων: Γεώργιος Κουρουπέτογλου

Αθήνα 1999

Περιεχόμενα Εργασίας 1

ΠΟΛΥΤΡΟΠΙΚΗ ΔΙΕΠΑΦΗ ΑΝΘΡΩΠΟΥ-ΜΗΧΑΝΗΣ

Jean Caelen

Η δημοσίευση αυτή παρουσιάζει προβλήματα, έννοιες και αρχές που αφορούν στη σχεδίαση μίας πολυτροπικής διεπαφής ανθρώπου-μηχανής η οποία πραγματοποιεί αναγνώριση και σύνθεση ομιλίας σε πραγματικά πλαίσια: Τροπικές σχέσεις στην περιοχή των ανθρωπίνων λειτουργιών, γενικό πλαίσιο αλληλεπίδρασης (αποκλειστικό, ταυτόχρονο, εναλλασόμενο και συνεργητικό), διαχείριση γεγονότων σε σχέση με τη συνοχή, τη χρονολόγηση, τον πλεονασμό των συμβάντων, την ερμηνεία της διασταυρωμένης πληροφορίας (φαινόμενα κοινής παραπομπής), και την επάρκεια της αναπαράστασης της σημασίας. Προτείνονται ταξινομήσεις των διεπαφών και των εφαρμογών με σκοπό να συζητηθούν μερικά από αυτά τα θέματα, και θα προτείνουμε μια πιθανή λύση του προβλήματος της κοινής παραπομπής μεταξύ των τρόπων. Οι έννοιες παρουσιάζονται μέσω ενός παραδείγματος: Η ICP σχεδίαση είναι μια εφαρμογή σχεδίασης η οποία έχει πολυτροπική διεπαφή (φωνή + χειρονομία). Η αρχιτεκτονική της ICP σχεδίασης περιγράφεται με λεπτομέρεια, ειδικά σε σχέση με τη διαχείριση των γεγονότων και την ερμηνεία της πολυτροπικής πληροφορίας. Η πολλαπλών στρωμάτων οργάνωση της, δομείται ως ακολούθως: Υλικό αφιερωμένο για αναγνώριση και σύνθεση ομιλίας, όπως επίσης και για χειρονομίες (εάν υπάρχουν), servers χαμηλού επιπέδου γεγονότων (ομιλία , ποντίκι και πληκτρολόγιο), ένας διαχειριστής γεγονότων για ανάμιξη πολυτροπικής πληροφορίας, ένας διαλογικός ελεγκτής, και μία υψηλού επιπέδου διεπαφή επικοινωνίας με την εφαρμογή.

Υπάρχουν πολλές περιπτώσεις όπου οι υπολογιστές βοηθούν τους ανθρώπους στα καθήκοντα τους ή στην “επικοινωνία” τους (ο τελευταίος όρος θα πρέπει να χρησιμοποιείται με προσοχή):

1. Η μηχανή ενεργεί ως ένας *μεσολαβητής* -καθιστά δυνατές μεγάλης απόστασης επικοινωνίες μεταξύ ανθρώπων που εργάζονται μαζί ή συνεργάζονται για κάποιο αντικείμενο. Σε αυτή την περίπτωση, τα έγγραφα που ανταλλάσονται ή χειρίζονται θα πρέπει να χρησιμοποιούν πολυμέσα ώστε να είναι πλήρως πληροφοριακά.
2. Η μηχανή αναπαριστά μια *εικονική πραγματικότητα* επεκτείνοντας την ανθρώπινη δημιουργικότητα ή τις δυνατότητες έκφρασης. Ο χρήστης είναι απορροφημένος από ένα κόσμο με τον οποίο όμως αλληλεπιδρά. Σήμερα αυτή η αλληλεπίδραση

γίνεται κυρίως με χειρονομίες. Στο κοντινό όμως μέλλον θα μπορούμε να χρησιμοποιούμε την γλώσσα εξίσου καλά.

3. Η μηχανή δρα σαν σύντροφος -*συνεργάζεται* με τον χρήστη πάνω στα δεδομένα προβλήματα του, χρησιμοποιώντας το διάλογο για να κατανοήσει τους αντικείμενικούς σκοπούς και προθέσεις του χρήστη. Αυτό μπορεί να αυξήσει την αποτελεσματικότητα της εργασίας.

Σε όλες αυτές τις περιπτώσεις, η αλληλεπίδραση μεταξύ ανθρώπων -μηχανών μπορεί να διανοηθεί ως πολυαισθητηριακή. Μια τέτοια αλληλεπίδραση θα λέγεται πολυτροπική αν ικανοποιούνται δύο συνθήκες. Πρώτον, η αλληλεπίδραση θα πρέπει να έχει σχεδιαστεί πάνω σε διάφορες ανθρώπινες αισθήσεις και μηχανικούς τρόπους, όπως όραση, λόγος (που εκφέρεται ή ακούγεται) και χειρονομία (κίνηση, εστίαση, γράψιμο,σχεδίαση) ταυτόχρονα και συνεργάσιμα. Δεύτερον, η μηχανή πρέπει να καταλαβαίνει και να ερμηνεύει τις πληροφορίες που μεταφέρονται από διάφορες συσκευές εισόδου-εξόδου που ονομάζονται μέσα.

Εδώ πρέπει να παρατηρήσουμε ότι μια απλή παρουσίαση των ανεξαρτήτων τρόπων, δεν αποφέρει και μια επαρκή αλληλεπίδραση. Το λογισμικό πρέπει να είναι έτσι σχεδιασμένο ώστε να επιτρέπει συνεργασία μεταξύ αυτών των τρόπων. Αυτό σημαίνει ότι το λογισμικό πρέπει να μπορεί να δέχεται ένα τυπικό πολυτροπικό μύνημα, όπως "βάλε αυτό εκεί",όπου ο λόγος χρησιμοποιείται για να δοθεί η εντολή ("βάλε"), και οι κινήσεις του ποντικιού για να "σχεδιαστεί" ένα αντικείμενο ("αυτό"), καθώς και η τοποθεσία του ("εκεί"). Αυτός ο τύπος μηνύματος, δηλαδή με πολυαισθητηριακό τρόπο, μπορεί να ερμηνευτεί μόνο συγχωνεύοντας την πληροφορία που προήλθε από την ομιλία, και την χειρονομία (κίνηση του ποντικιού). Αυτή η συγχώνευση αυξάνει το πρόβλημα των διατροπικών χρονικών και χωρικών αναφορών. Το πρόβλημα αυτό μπορεί να λυθεί μόνο μέσω μιας περίπλοκης ερμηνείας των πολυτροπικών συμβάντων.

Για τον σχεδιαστή της διεπαφής ανθρώπου- υπολογιστή, οι καταστάσεις 1, 2 και 3 που ακολουθούν δείχνουν κοινά χαρακτηριστικά, και τα ανταποκρινόμενα συστήματα εκεί, θα πρέπει να χρησιμοποιούν κοινές ρυθμίσεις. Για παράδειγμα το επίπεδο των πολυμέσων που εμπλέκεται στη *μεταφορά της πληροφορίας* είναι το χαμηλότερο επίπεδο, και είναι κοινό για όλες τις καταστάσεις. Πάντως, πάνω από το επίπεδο μεταφοράς, οι

καταστάσεις διαφέρουν τόσο στη διαχείριση τους, όσο και στην επεξεργασία των πληροφοριών εισόδου-εξόδου.

Στην κατάσταση 1 (μεσολάβηση), η διαχείριση της πληροφορίας μπορεί να είναι σύγχρονη (πχ για ομάδα επεξεργασίας κειμένου), ή ασύγχρονη (πχ για μια εφαρμογή ηλεκτρονικού ταχυδρομείου). Εδώ, το δεύτερο επίπεδο αφορά διάλογο ανθρώπου υπολογιστή ο οποίος μπορεί να είναι περισσότερο ή λιγότερο εξεζητημένος. Το τρίτο επίπεδο αφορά στη διαχείριση των κανόνων συμμετοχής, οι οποίοι με τη σειρά τους καθορίζουν τα δικαιώματα και τα προνόμια που προσδίδονται στα αντικείμενα που αποτελούν μέρος της εφαρμογής.

Στην κατάσταση 2 (εικονική πραγματικότητα), η συμπεριφορά του χρήστη δείχνει μια προτίμηση στον ίδιο τον διάλογο, ο οποίος μετατρέπεται σε ένα βασικό σχήμα δράσης-αντίδρασης.

Στην κατάσταση 3 (συνεργασία καθήκοντος), συμβαίνει το αντίθετο. Εδώ η σχέση χρήστη συστήματος (ή ακριβέστερα, η σχέση χειριστή-εργασίας) δείχνει μια προτίμηση,όπως στην περίπτωση όπου ο χρήστης είναι μόνος στην αλληλεπίδραση με τη μηχανή. Οι ικανότητες επικοινωνίας της μηχανής θα πρέπει να μοντελοποιηθούν πάνω στην ανθρώπινη επικοινωνία για να αυξήσουν την αποδοτικότητα και την αξιοπιστία κατά την εκτέλεση της εργασίας.

Θα δούμε ότι η βελτίωση της ερμηνείας των πολυτροπικών αυτών πληροφοριών εξαρτάται από την τελευταία αυτή κατάσταση. Ωστόσο, παρόλης της μεταβλητότητας των δύο αυτών καταστάσεων και της ποιότητας του συστήματος, μια πολυτροπική διεπαφή θα πρέπει να εμπλέκει τρεις διαδικασίες: Διαχείριση του τρόπου, συγχώνευση των πληροφοριών εισόδου, και διαχωρισμός των πληροφοριών εξόδου.

Η δημοσίευση αυτή επικεντρώνεται στις τρεις αυτές όψεις της πολυτροπικότητας. Για λόγους απλότητας, οι τρόποι περιορίζονται στη φωνή και στη χειρονομία. Η πολυτροπικότητα εγγίζεται από δύο πλευρές: Του χρήστη και της μηχανής (ή ακριβέστερα, της τεχνολογίας). Πράγματι, η διεπαφή ανθρώπου μηχανής είναι το σημείο συνάντησης αυτών των δύο, και η σχεδίαση της είναι ένας συμβιβασμός μεταξύ των απαιτήσεων για εργονομία και των τεχνολογικών περιορισμών.

Στοιχεία Εργονομίας

Η πολυτροπική διεπαφή ανθρώπου υπολογιστή, αυξάνει μερικά νέα εργονομικά προβλήματα σχετικά με τους τρόπους επικοινωνίας, οι οποίοι είναι η ομιλία, η χειρονομία, η όραση, κ.τ.λ. Συγκεκριμένα αυξάνει το ζήτημα σχετικά με την *καταλληλότητα* μεταξύ ενός δοσμένου τρόπου, και της αντικειμενικότητας και των συλλογισμών του χρήστη, ένα ζήτημα το οποίο προστίθεται στα παραδοσιακά προβλήματα διεπαφής, της καταλληλότητας της παρουσίασης και της επεξεργασίας. Ειδικότερα, το ζήτημα της καταλληλότητας των τρόπων που αφορούν στα αισθητήρια όργανα και στις μηχανιστικές τροπολογίες, καθώς επίσης και τη βέλτιστη χρήση για μια δέδομένη εφαρμογή.

Από τη σκοπιά ενός συνολικού και γενικού σχήματος, οι απαιτήσεις της διεπαφής μπορούν να παρουσιαστούν ακόλουθα ως:

H	<->	W	+	m
Μηχανιστικές τροπολογίες		κείμενο, εικόνες		επεξεργασία τρόπων

Οπου το **H** αναφέρεται στον άνθρωπο χρήστη (ή χειριστή),το **m** στη μηχανή της οποίας ο χρήστης έχει μόνο μία συνοπτική αναπαράσταση, και το **W** τον κόσμο όπου αυτός ή αυτή αντιλαμβάνεται (μεταφορικά,πραγματικά ή εικονικά), το οποίο δίνει νόημα στην αναπαράσταση του\της. Η σημαντική απαίτηση, είναι ότι οι αισθητηριακοί και μηχανιστικοί τρόποι του **H** πρέπει να συμπίπτουν με αυτούς της **m,** και ότι οι συλλογιστική του/της πρέπει να συμπίπτει με τις επεξεργασίες της **m.**

Αυτό σημαίνει ότι ένα σχέδιο για πολυτροπική διεπαφή ανθρώπου-μηχανής - όπως αυτό προέρχεται από την ανθρώπινη πολυαισθητηριακή ικανότητα - πρέπει να παίρνει υπόψη τους ακόλουθους παράγοντες:

- Η χρήση των τρόπων και η καταλληλότητά τους για την εργασία,
- Οι στατηγικές αλληλεπίδρασης προσαρμοσμένες στην ικανότητα και την απόδοση του χρήστη,

- Η διαχείριση των γεγονότων χαμηλού επιπέδου (όπως συνοχή, χρονολόγηση,πλεονασμός, κ.τ.λ.), δεδομένων των ορίων της ικανότητας αντίληψης του ανθρώπου και των ικανοτήτων της μηχανής,
- Κοινά επίπεδα αφηρημένων εννοιών και αναπαραστάσεων,
- Πολυτροπικές παρουσιάσεις και απόψεις,
- Και γενικότερα, ένα συγγενικό μοντέλο βασισμένο στο χρήστη (όχι στη μηχανή).

Με μία τέτοια πολυτροπική διεπαφή, ο χρήστης είναι δυνατόν να αντιλαμβάνεται (στο επίπεδο της διεπαφής) μία αντανάκλαση των δεδομένων της μηχανής και των δομών της δουλειάς. Στο παρελθόν, οι δομές αυτές επιβάλλονταν από τεχνολογικούς περιορισμούς. Σήμερα πάντως, η εστίαση γίνεται πάνω στον χρήστη, και οι εργονομικοί περιορισμοί τείνουν να αντικαταστήσουν τους τεχνολογικούς. Τώρα είμαστε στο άλλο άκρο του φάσματος, και το τρέχον ζήτημα αφορά στο βαθμό στον οποίο η δομή της διεπαφής ανθρώπου μηχανής πρέπει να μοντελοποιηθεί στη συμπεριφορά του χρήστη. Απευθυνόμενοι στο ζήτημα της σύνθεσης και της αναγνώρισης της ομιλίας, η ουσία του ζητήματος είναι πόσο βέλτιστη είναι η ομιλία για διεπαφή ανθρώπου μηχανής σε μια δεδομένη εργασία.

Μερικά στοιχεία απάντησης σε αυτό το ερώτημα προέρχονται από το γεγονός ότι ο χρήστης έχει ένα αντικειμενικό σκοπό, και η δραστηριότητα του\της σχεδιάζεται σε κάποιο βαθμό (εξαρτάται από την εκπαίδευση, την εμπειρία στη δουλειά, την πρακτική ή τεχνική γνώση), και στο ότι μπορεί να επανοργανωθεί σύμφωνα με τους περιορισμούς που επιβάλλονται από τη μηχανή. Ο χρήστης έχει επίσης προτιμήσεις, συνήθειες και ιδιοσυγκρασίες. Τρέχων παρατηρήσεις δείχνουν ότι νέες "συνήθειες χειρονομίας" αναδύονται από τους χρήστες ποντικιού. Αυτές οι συνήθειες είναι συχνά μακριά από το βέλτιστο και σε μερικές περιπτώσεις, η γλώσσα μπορεί να γίνει πολύ πιο κατάλληλο επικοινωνιακό μέσο από ότι η χειρονομία. Ας γυρίσουμε τώρα σε αυτά τα ζητήματα, και ας τα αναλύσουμε σε τμήματα δύο τύπων λειτουργικών τρόπων: Τρόποι βασισμένοι σε γλώσσα και τρόποι που δεν βασίζονται στη γλώσσα.

Τρόποι βασισμένοι στη γλώσσα

Συνολικές μελέτες του τρόπου χρήσης που να υπολογίζουν την πολυτροπική διεπαφή ανθρώπου μηχανής με όρους όπως η αποτελεσματικότητα, η αξιοπιστία και η ευλυγισία δεν είναι διαθέσιμες, αφού τέτοιες διεπαφές δεν είναι ακόμα διαθέσιμες εκτός από λίγες εργαστηριακές περιπτώσεις. Πάντως, μερικά πειράματα δείχνουν ότι ο λόγος είναι επιθυμητός σε σίγουρες καταστάσεις. Ακολουθούν μερικά συμπεράσματα που θεωρούν ότι η διεπαφή ανθρώπου μηχανής χρησιμοποιεί τον λόγο σαν είσοδο:

- Η ικανοποίηση του χρήστη εξαρτάται από την κοινωνικοεπαγγελματική τάξη.
- Η εκμάθηση της διεπαφής είναι συνήθως γρηγορότερη.
- Η διόρθωση λαθών είναι συνήθως πιο αποτελεσματική. Αλλά:
- Το πλαίσιο μπορεί να είναι περιοριστικό (θόρυβος, εμπιστευτικότητα, κ.τ.λ.).
- Το γλωσσολογικό επίπεδο της μηχανής (δηλαδή το επίπεδο της γλώσσας όπου αντιλαμβάνεται η μηχανή) απαιτεί μία προσαρμογή από τον χρήστη.

Η σχεδίαση μιας "διαλέκτου" προερχόμενης από φυσική γλώσσα μπορεί να είναι μια διαθέσιμη λύση για την διεπαφή (σε αντίθεση με μία υπο-γλώσσα, μία επίσημη ή τεχνητή γλώσσα), με σκοπό να διευκολύνει το χρήστη να μάθει τις οντότητες και τις λειτουργίες. Μία τέτοια "διάλεκτος" θα μπορούσε να διατεθεί για να ενεργοποιήσει τη μηχανή, αφού αυτή μπορεί να είναι κατασκευασμένη έτσι ώστε το λεξικό της να είναι καλά ορισμένο και η σύνταξη της να παραμένει περιορισμένη.

Τέτοιες "διάλεκτοι" έχουν τα ίδια χαρακτηριστικά όπως και οι ανθρώπινες "*ισχύουσες γλώσσες*" - γλώσσες που χρησιμοποιούνται σε αλληλεπιδράσεις που βοηθούν να πραγματοποιηθεί μία δεδομένη εργασία - οι οποίες, σε εξαιρετικές περιπτώσεις, δεν έχουν σχεδόν καθόλου σύνταξη, και έχουν ένα πολύ περιορισμένο και ειδικευμένο λεξικό. Οι ισχύουσες γλώσσες, είναι άμεσα συνδεδεμένες με τη φύση της εφαρμογής. Αντιστρόφως, δημόσιες εφαρμογές για βάση δεδομένων φωνητικών απορριών εμπλέκουν μεγάλη γλωσσολογική ποικιλία και ελλειπτικά φαινόμενα (έλλειψη: παράλειψη μίας ή περισοτέρων λέξεων οι οποίες γίνονται κατανοητές από το περιεχόμενο). Οι χρήστες προσαρμόζονται στη μηχανή απλοποιώντας την προφορά τους. Αυτό οδηγεί σε λιγότερες ελλείψεις ή αναφορές, και σε σωστή σύνταξη (ακόμα και αν αυτό δεν απαιτείται). Το ίδιο φαινόμενο λαμβάνει χώρα σε σχέση με την προσωδία (επιτονισμός

ομιλίας και ρυθμός) όπου άτομα διαβάζουν κείμενα δυνατά λόγω περιορισμών κατανόησης. Σε αυτές τις περιπτώσεις είναι παρατηρημένο, ότι τα άτομα τονίζουν το κενό μεταξύ των λέξεων, ακόμη και μεταξύ των συλλαβών. Από την άλλη μεριά, η προφορική παραγωγή είναι υποβαθμισμένη είτε από την αύξηση του φόρτου εργασίας, είτε σε περιπτώσεις όπου υπάρχει έντονη έντονη εστίαση σε συγκεκριμένους.

Αυτές οι μελέτες δείχνουν ότι τέτοιοι περιορισμοί στη χρήση της ομιλίας δεν είναι αναμενόμενοι μόνο στην ανεπαρκή απόδοση των συστημάτων αναγνώρισης ομιλίας, από τότε που οι αναγνωριζόμενες από τη μηχανή γλώσσες αντιστοιχούν σε κατηγορίες λειτουργικών γλωσσών. Στην πραγματικότητα, αυτά τα όρια φαίνεται να είναι συνδεδεμένα με τα χαρακτηριστικά του ίδιου του τρόπου ομιλίας. Ας παρατηρήσουμε επίσης ότι ο τρόπος ομιλίας είναι ανώτερος από τον γραπτό, σε ότι αφορά στα όρια ταχύτητας εισόδου από το πληκτρολόγιο και ότι κινητοποιεί τους αισθητηριοκινητικούς πόρους του χρήστη – αφού είναι πιθανόν η χρησιμοποίηση του αριθμητικού μπλοκ για είσοδο, να μπορεί επίσης να τροποποιήσει την κατάσταση.

Τρόποι μη βασισμένοι στη γλώσσα

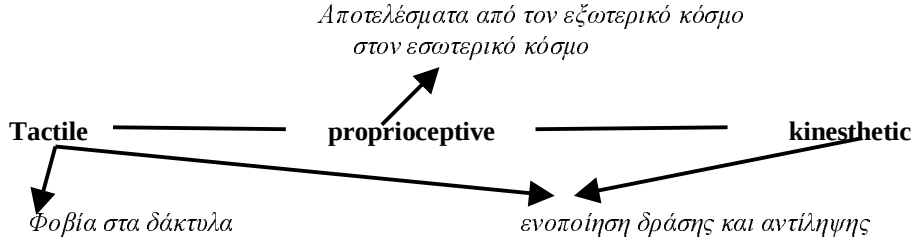
Χειρονομία

Η ανθρώπινη επικοινωνία σχεδόν πάντα περιλαμβάνει μια γλώσσα, η οποία είναι μια ανταλλαγή συμβόλων (ή σχετικών νοημάτων) μέσω ενός κώδικα ανάμεσα στους συνδιαλλεγόμενους. Η αλληλεπίδραση από την άλλη μεριά, είναι μια μορφή μη-συμβολικής επικοινωνίας. Το να αλληλεπιδράς, είναι το να δίνεις εντολές ή να λαμβάνεις πρωτοβουλίες και να δέχεσαι ερεθίσματα.Η χειρονομία είναι μια μορφή αλληλεπίδρασης αλλά όχι απαραίτητα μια μορφή επικοινωνίας με την αυστηρή έννοια του όρου.

Το γεγονός ότι η χειρονομία δεν προτάθηκε νωρίτερα ως τρόπος αλληλεπίδρασης ανθρώπου-υπολογιστή μπορεί να αποδοθεί σε σε δυο λόγους. Πρώτον, η επιβεβαίωση του γεγονότος ότι μπορεί να αλληλεπιδρά με άλλους τρόπους είναι πρόσφατη. Στην πραγματικότητα, ο έλεγχος της ανθρώπινης χειρονομίας είναι πολύτροπος και περιλαμβάνει ακοή, όραση κτλ. Δεύτερον, η σύλληψή της απαιτεί μάλλον πολύπλοκους μηχανισμούς.

Στη χειρονομία, η δράση και η αντίληψη είναι βαθιά συνδεδεμένες. Κάθε ενέργεια είναι επίσης και ένα αντιλήψιμο γεγονός που αφορά στη χειρονομία και στα αποτελέσματά της. Για αυτό, η χειρονομία είναι δύσκολο να μοντελοποιηθεί. Οι άνθρωποι έχουν περίπου 700 μύες, 110 αρθρώσεις, και περίπου 100 βαθμούς ελευθερίας. Όπως και στην όραση, υπάρχει ένα είδος "φοβίας χειρονομίας", εντοπισμένη στα ακροδάκτυλα, μέσω της οποίας η ενέργεια ανταλλάσεται με το περιβάλλον. Προσδοκώντας να κατανοήσουμε αυτή την περιπλοκή ανθρώπινη λειτουργία, μπορούμε να ξεκινήσουμε διακρίνοντας τρεις κύριες λειτουργίες για τις χειρονομίες:

1. Την "*εργοτική*" λειτουργία, όπου είναι μια μετατροπή είτε ενέργειας είτε ύλης στο περιβάλλον. Εδώ η χειρονομία εκλαμβάνεται σαν ενέργεια ή δύναμη.
2. Την "*επιστημική*" λειτουργία, η οποία επιτρέπει την απόκτηση γνώσης για το περιβάλλον (μέσω αγγίγματος). Αυτή η λειτουργία έχει τρεις παραμέτρους (T,P,K) T = tactile(απτός), P = proprioceptive (αντιληπτικός) ,και K = Kinaesthetic (κινηταισθητικός). Αυτοί οι τρεις όροι είναι μη διαχωρίσιμοι, αφού για να αναγνωρίσουμε για παράδειγμα το σχήμα ενός αντικειμένου, πρέπει να ενεργήσουμε χρησιμοποιώντας και τις τρεις.
3. Την "*σημειωτική*" (επικοινωνιακή) λειτουργία, η οποία μπορεί από μόνη της να χρησιμοποιείται για την επικοινωνία (πχ η γλώσσα με νοήματα των κουφών), ή μπορεί να συμπληρώνει άλλες γλώσσες (πχ η ομιλία) για να περιγράψει, να υποδείξει ρυθμό κτλ. Αυτή η λειτουργία μπορεί να απαντηθεί σε πολλούς τομείς όπως ιδεογραφήματα, ζωγραφική, γραπτός λόγος, σαν γλωσσικό υποβοήθημα, οργανική χειρονομία ή σε μορφή γλωσσική επικοινωνίας(πχ γλώσσες νοημάτων ή τεχνητές γλώσσες όπως αυτές που χρησιμοποιούνται στις άμεσου-χειρισμού διεπαφές ανθρώπου-υπολογιστή (HIC-ΔΑΥ).



Σχήμα 1*. Επιστημική λειτουργία της χειρονομίας*

Μερικές χειρονομίες επιτελούν αρκετές λειτουργίες ταυτόχρονα. Για παράδειγμα, οι χειρονομίες ενός διευθυντή ορχήστρας εξάρουν ρυθμό (έμφαση), υποδεικνύουν μια μουσική έκφραση (γλώσσα), και λένε στους μουσικούς πότε να αρχίσουν να παίζουν (εντολή).

Από αυτήν την άποψη, το γράψιμο και το σημείωμα δεν λογίζονται ως χειρονομίες, γιατί αυτά που είναι σημαντικά είναι μόνο τα αποτελέσματα και όχι ο τρόπος που παράγονται. Γεννιέται λοιπόν το ερώτημα για το ποιο έχει "σημειωτική" λειτουργία, η ίδια η χειρονομία ή το ίχνος της στη συσκευή εισόδου; Ανάλογα, θα έπρεπε, οι ενέργειες στο πληκτρολόγιο ή με το ποντίκι να εκλαμβάνονται ως χειρονομίες;

Καταλήγωντας, μια ΔΑΥ μπορεί να χρησιμοποιεί τη χειρονομία με πολύ διαφορετικό και παράδοξο τρόπο. Ο χρήστης μπορεί να μεταπηδήσει από μια οργανικού τρόπου διεπαφή (πχ πληκτρολόγιο ή ποντίκι), όπου η χειρονομία είναι ταυτόχρονα ενέργεια και αντίληψη, σε μια διεπαφή όπου η χειρονομία δεν είναι πια σημαντική από μόνη της αλλά γίνεται "ορατή" και αναγνωρίζεται, απλώς σαν το ίχνος μιας άλλης σημασίας.

Όραση

Η όραση (σύλληψη εικόνας) μέσω κάμερας επιτρέπει τη σύλληψη χειρονομιών, και συγκεκριμένα εκφράσεων προσώπου. Αυτές οι εκφράσεις αναγνωρίζονται χρησιμοποιώντας τεχνικές αναγνώρισης προτύπων. Η όψη της κίνησης των χειλιών μπορεί να υποστηρίξει την αναγνώριση ομιλίας. Γενικά, όταν χρησιμοποιούμε την όραση για την σύλληψη χειρονομίας έχουμε να αντιμετωπίσουμε μόνο τεχνικά ζητήματα και δεν αλλάζει ο ρόλος της χειρονομίας. Παρόλ'αυτά η όραση στη ρομποτική λαμβάνεται σαν μια μηχανική διεπαφή με το περιβάλλον. Σ`αυτή την περίπτωση η όραση γίνεται ένας λήπτης χωρικής κίνησης και εντοπισμού στο χώρο, και χρησιμοποιείται για τον έλεγχο της χειρονομίας.

Επάρκεια τρόπων

Το παράδειγμα του άμεσου χειρισμού στις γραφικές διεπαφές, φαίνεται να έχει φτάσει στα όριά του. Μόνο ορατά αντικείμενα μπορούν να περιγραφούν, και η ακολουθία εντολών (*επιλογή* που ακολουθείται από *λειτουργία*) είναι αντίθετη από αυτή που υπαγορεύεται απ᾽το ένστικτό μας (αντίθετη απ᾽αυτή που συναντάται στη φυσική γλώσσα) και συχνά ανεπαρκής. Γι᾽αυτό μια πολυτροπική ΔΑΥ είναι χρήσιμη. Παρά το γεγονός ότι πολύ λίγες εργονομικές μελέτες έχουν διεκπεραιωθεί πάνω σ᾽αυτές τις διεπαφές (αφού δεν είναι ακόμα διαθέσιμες), μπορούν να γίνουν μερικές γενικές παρατηρήσεις σχετικά με την επάρκεια και εφαρμοσιμότητα κάθε τρόπου:

- Τρόπος ομιλίας:
  Είσοδος ως: εντολές, μακρο-εντολές (απομονωμένες λέξεις, συνεχής ομιλία)
  Έξοδος ως: βοήθεια, παραδείγματα, αιτήσεις, επεξήγηση, σύσταση (σύνθεση, μαγνητοφωνημένες προτάσεις).
- Γραπτός τρόπος:
  Είσοδος ως: αναγνωριστές, ψηφία (πληκτρολόγιο, μπλοκ γραφικών)
  Έξοδος ως : Λεπτομερής εξήγηση (οθόνη).
- Τρόπος χειρονομίας:
  Είσοδος ως : περιγραφή στις 2 ή 3 διαστάσεις (ποντίκι, οθόνη επαφής), γλώσσα νοημάτων (κάμερα), “εργοτική” δράση (πληκτρολόγιο με αλληλεπίδραση).
- Οπτικός τρόπος:
  Είσοδος ως: προσανατολισμός χρήστη, έκφραση προσώπου χρήστη
  Έξοδος ως : γραφικά, εικόνες (γραφικά υπολογιστών).

Κανένας τρόπος δεν υπερισχύει συστηματικά : η επικράτηση κάποιου τρόπου σε μια ΔΑΥ κατάσταση, εξαρτάται από την ανταγωνιστικότητα και την απόδοση, το πεδίο αλληλεπίδρασης και επικοινωνίας κτλ.

Προβλήματα της ΔΑΥ

Τα προβλήματα του πολυτροπικού ΔΑΥ, σε σχέση με αυτά του παραδοσιακού ΔΑΥ, προκύπτουν από την ποικιλία των τρόπων εισόδου και εξόδου. Η πληροφορία των

διάφορων τρόπων είναι ανεξάρτητη και πρέπει να αναλυθεί, να ερμηνευτεί και να παραχθεί.

Αυτά τα προβλήματα αφορούν σε τρεις κύριες περιοχές. Στο επίπεδο της διαχείρισης τρόπων, τα ερωτήματα προκύπτουν όσον αφορά στη χρονολόγηση και στο συγχρονισμό των γεγονότων, στις ενέργειες που συνιστούν την πληροφορία που ανταλλάσεται, και στο πλαίσιο μέσα στο οποίο γίνεται η αλληλεπίδραση. Στο επίπεδο στο οποίο λαμβάνει χώρα η *συγχώνευση* και ο *διαχωρισμός* της πληροφορίας, προκύπτουν ερωτήματα που αφορούν στη μορφοσύνταξη των αλληλεπιδράσεων που βασίζονται στη γλώσσα , όπως επίσης στη σημασιολογία και στην πραγματολογία (πχ προβλήματα συναναφοράς) κάθε είδους αλληλεπιδράσεων. Καταλήγοντας, τα προβλήματα πρέπει να εμφανίζονται στον *τρόπο με τον οποίο η πληροφορία ανταλλάσεται* μεταξύ των διαφόρων ΔΑΥ, όπως επίσης και μεταξύ της ΔΑΥ και του λειτουργικού πυρήνα της εφαρμογής.

Σε κάθε τρόπο αντιστοιχίζεται ένα αντιπροσωπευτικό μοντέλο της πληροφορίας που μεταφέρει. Για παράδειγμα, το μοντέλο *χειρονομίας* μπορεί να είναι διανύσματα χωρικών συντεταγμένων στο χρόνο, ενώ το μοντέλο *ομιλίας* αποτελείται από αλυσίδες χαρακτήρων που αντιστοιχούν σε αναγνωρισμένες λέξεις, προτάσεις ή ακόμα μη επεξεργασμένα σήματα. Η συχνότητα δειγματοληψίας μπορεί να αλλάζει από μέσο σε μέσο.

Σχετικά με την είσοδο (το πρόβλημα είναι συμμετρικό για την έξοδο), μια λειτουργιστική ματιά δείχνει διαφορετικά στρώματα στις ΔΑΥ, ξεκινώντας από το συμπαγές επίπεδο των σημάτων, μέχρι το ποιο αφηρημένο επίπεδο των δράσεων. Αυτά τα στρώματα είναι η απόκτηση σήματος, η αυτόματη αναγνώριση σήματος, η κατανόηση των συμβόλων που μεταφέρουν, κατανόηση των σημάτων με συσχετιστικό τρόπο, και κατασκευή ενεργειο-στραφούς πολυτροπικού μηνύματος.

Μέσω αυτών των διαφορετικών φάσεων, η πληροφορία αρχικά σχηματοποιείται, και μετά μετατρέπεται σε μια αφηρημένη αναπαράσταση (που μπορεί να διαφέρει όχι μόνο από τρόπο σε τρόπο αλλά και μέσα στον ίδιο τρόπο, από στρώμα σε στρώμα). Τελικά εκπέμπεται στο υψηλότερο επίπεδο, το στρώμα διαλόγου. Στα επόμενα, θα εξετάσουμε λεπτομερώς καθ᾽ένα απ᾽αυτά τα στρώματα της επεξεργασίας της πληροφορίας.

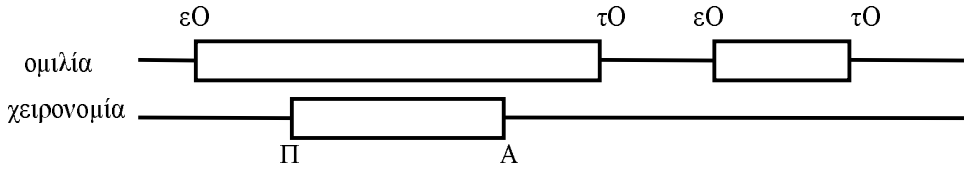
Διαχείριση τρόπων

Για να ξεκινήσουμε, θα ήταν χρήσιμο να θέσουμε ένα διαχωρισμό μεταξύ των γεγονότων (τα διάφορα γεγονότα που ανακλούν τη φυσική οργάνωση των ενεργειών) και της πληροφορίας (μεταξύ των συνιστωσών μονάδων της).

Γεγονότα και πληροφορία

Ορισμός *γεγονότος*: Ένα γεγονός είναι η αρχή του τέλους ενός εξωτερικού σήματος προς τη μηχανή. Ένα γεγονός σηματοδοτεί μια αντιλήψιμη αλλαγή σε ένα μέσο. Αυτός ο ορισμός επικεντρώνεται στη μηχανή, ή ακριβέστερα στα κανάλια εισόδου-εξόδου, τα οποία αποκαλούμε μέσα, και όχι στο χρήστη.

Παραδείγματα:
-Γεγονότα ποντικιού: κλικ = κ, πίεσε = Π, άφησε = Α, μετακίνησε = 2-δ τροχιά=(αΤ,τΤ)(αρχική τροχιά , τελική τροχιά)
-Γεγονότα ομιλίας : έναρξη ομιλίας=εΟ, τέλος ομιλίας=τΟ, αρχική λέξη= αΛ, τελική λέξη= τΛ



Ομιλία και χειρονομία μπορούν να συμπίμπτουν εν μέρει κατά μήκος του άξονα χρόνου

Ορισμός *μονάδας φορέα-πληροφορίας*: Μια μονάδα φορέας-πληροφορίας είναι μια μονάδα που περιέχει σημασία, αλλά η σημασία της είναι διαφορετική για το χρήστη και διαφορετική για τη μηχανή. Από τη μεριά του χρήστη, η μονάδα φορέας πληροφορίας είναι μια σημειωτική μονάδα (μια μονάδα που μεταφέρει σημασία και διακρισιμότητα),

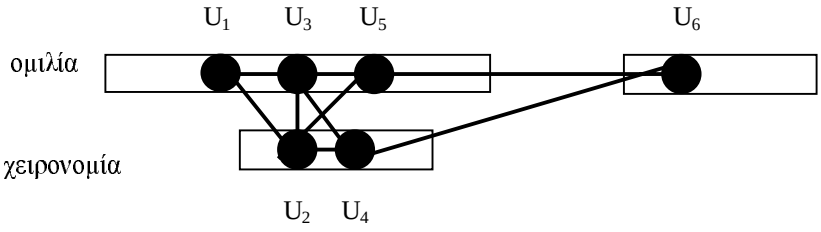
ενώ από τη μεριά της μηχανής, αυτή η μονάδα είναι μια αναφορική μονάδα, δηλαδή αναφέρεται σε ένα γεγονός ή ενέργεια.

Ορισμός *ενέργειας* και *δράσης:* Μία *ενέργεια(πράξη),* είναι μια ακολουθία σημειωτικών μονάδων οι οποίες παράγονται ή λαμβάνονται από το χρήστη. Αυτή η ακολουθία μεταφέρεται από ένα σήμα (ομιλία ή χειρονομία), το οποίο οροθετείται από flags, ειδικά για κάθε τρόπο (παύσεις για την ομιλία, πάτημα κουμπιού για το ποντίκι κ.τ.λ.). Η χρονική οργάνωση αυτής της ακολουθίας, ορίζεται από μια σύνταξη. Για την ομιλία, μια ενέργεια είναι μια έννοια που αφορά στην ενέργεια (πράξη) ομιλίας. Μια *δράση*, είναι μια λειτουργία που γίνεται από τη μηχανή, και εκφράζεται από μια αλλαγή κατάστασης η οποία μπορεί (ή δεν μπορεί) να είναι αντιληπτή από το χρήστη.

Παραδείγματα:

-Πληροφορία χειρονομίας: ενέργεια = κ (τετράγωνο).Π(τοποθεσία).Τροχιά.Α(σημαίνει "ζωγράφισε τετράγωνο, εδώ, με αυτό το μέγεθος"). Μονάδες = κ(τετράγωνο),Π (τοποθεσία), Τροχιά .Α` Α = μετα-χειρονομία (flag).

-Γλωσσολογική πληροφορία: Ενέργεια = ενέργεια ομιλίας όπως έχει ορισθεί από τον Searle. Μονάδες = λεξικές, συντακτικές, σημασιολογικές, προσωδιακές, μετα-ασυναρτησιακές.



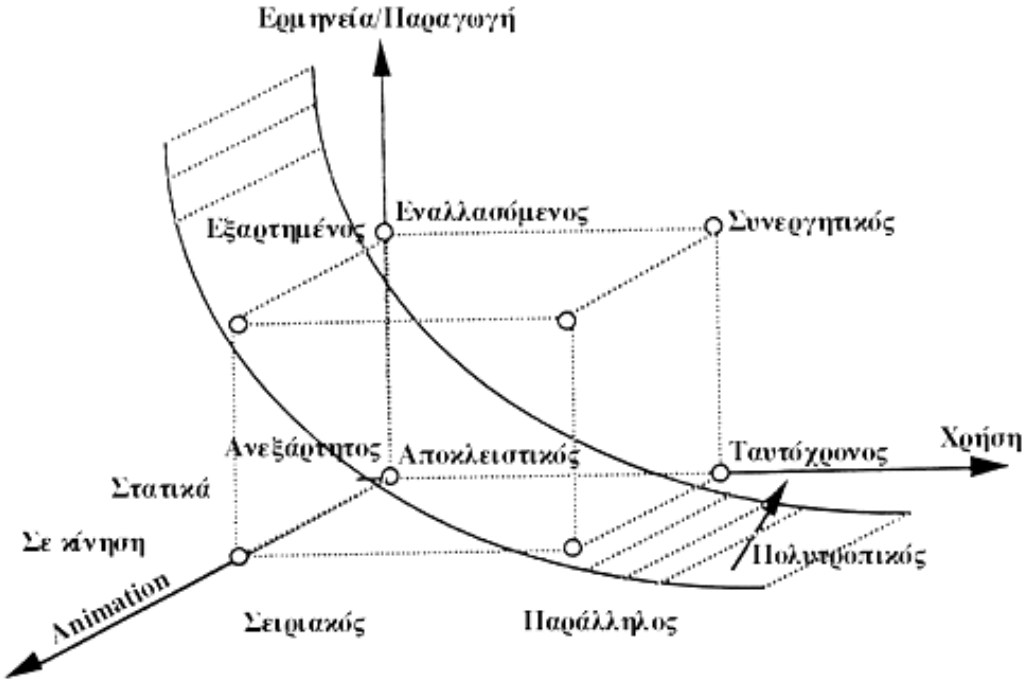
Υπάρχουν σημασιολογικές και χρονικές σχέσεις μεταξύ των μονάδων. Αυτό είναι ένα γενικότερο πρόβλημα των *πολύτροπικών συν-αναφορών*.

Το πλαίσιο αλληλεπίδρασης

Ορισμός: Το πλαίσιο αλληλ/σης ορίζεται από την τριπλέτα:

{*χρήση τρόπων, εξάρτηση πληροφορίας, κίνηση*}

όπου η *χρήση τρόπων* υποδεικνύει αν η ερμηνεία τρόπου είναι σειριακή ή παράλληλη, η *εξάρτηση πληροφορίας* υποδεικνύει αν η πληροφορία που μεταφέρεται από τα διάφορα μέσα εξαρτάται από αυτά, και η *κίνηση* υποδεικνύει την δυναμική, δηλαδή αν οι πράξεις είναι συνεχείς ή στιγμιαίες.

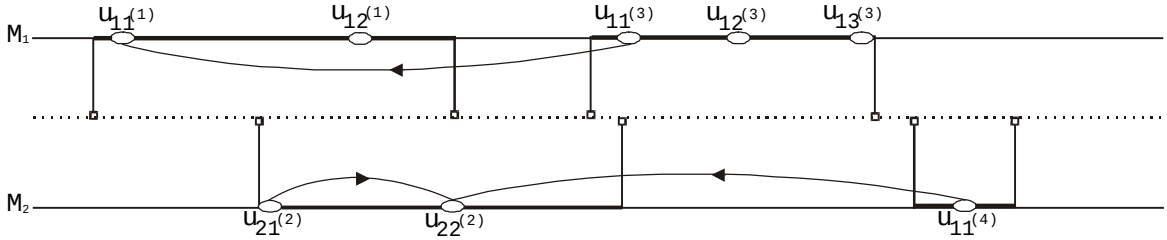


Σχήμα 2. Το πεδίο των πολυτροπικών συστημάτων: Εναλλασόμενο, ταυτόχρονο(σύγχρονο) και συνεργητικό πλαίσιο.

Σε αυτό το κείμενο, μόνο οι δύο πρώτοι όροι θα ληφθούν υπ᾽όψη. Οι συνδιασμοί τους ορίζουν τέσσερεις τύπους πλαισίων αλληλεπίδρασης όπου τους καλούμε αποκλειστικό, σύμφωνο, εναλλασόμενο και συνεργητικό.

Το αποκλειστικό πλαίσιο χαρακτηρίζει κάθε σύστημα με τουλάχιστον δυο εισόδους ή εξόδους όπου χρησιμοποιούνται ανεξάρτητα.Αυτή η περίπτωση δεν είναι στην ουσία πολυτροπική οπότε δεν εξετάζεται. Παρακάτω εξετάζονται τα υπόλοιπα τρία πλαίσια.

Α. Το σύμφωνο (ταυτόχρονο) πλαίσιο



Αυτό το το πλαίσιο ορίζεται στο πεδίο της *χρήσης τρόπων* με την απουσία χρονικών περιορισμών και στο πεδίο της εξάρτησης πληροφορίας με την απουσία αλληλοαναφοράς μεταξύ μονάδων διαφορετικών μέσων. Οι ιδιότητες αυτών των πλαισίων φαίνονται στο παρακάτω παράδειγμα, όπου χρησιμοποιούνται αναφορικά και δείκτες. Οι ενέργειες του χρήστη δίνονται από το Α1 εώς το Α3, και η αποκρίσεις της μηχανής παριστάνονται στο γράφημα. Η αναφορική ανάλυση είναι λάθος όταν η αναφορά εκφράζεται μέσω άλλου τρόπου και/ή οι δείκτες δεν μπορούν να αναλυθούν.

Παράδειγμα λανθασμένης αναφορικής ανάλυσης:

Α1: "Ζωγράφισε ένα κύκλο" + κ(πράσινο)	ομιλία + χειρονομία
Α2: κ(τετράγωνο).Π(τοποθεσία).Τροχιά .Α	χειρονομία
Α3: "σβήσε αυτό" (το "αυτό" είναι αναφορική αντωνυμία)	ομιλία

1 2 3

ροζ κύκλος πράσινο τετράγωνο αυτό=κύκλος

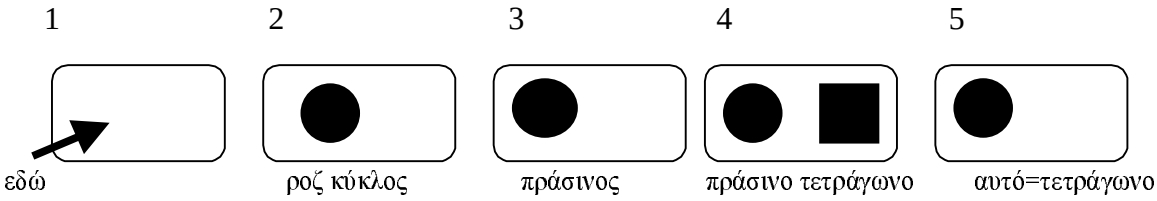
Β. Το εναλλασόμενο πλαίσιο

Ορίζεται στο πεδίο της *χρήσης τρόπων,* με Έναρξηu_i(k)>=Τέλοςu_i(k-1) με τα μέσα Mi διαφορετικό του Mi΄, και στο πεδίο της *εξάρτησης τρόπων,* με απουσία αλληλοαναφορικών περιορισμών μεταξύ των μονάδων. Οι ιδιότητες του εναλλακτικού πλαισίου που αφορούν στην ανάλυση αναφορών και δεικτών φαίνονται στο παράδειγμα. Εδώ η αναφορική ανάλυση είναι σωστή όταν η αναφορά γίνεται μέσω άλλου τρόπου και

οι δείκτες μπορούν να αναλυθούν. Όμως αυτό το πλαίσιο όντας περίπλοκο, μπορεί να μειώσει την αντίληψη και τον μηχανικό προσανατολισμό του χρήστη.

Παράδειγμα αναφορικής και δεικτικής ανάλυσης:

A1:"Ζωγράφισε κύκλο εδώ"	ομιλία
A2: k(τοποθεσία)	χειρονομία
A3: k(πράσινο)	χειρονομία
A4: k(τετράγωνο).Π(τοποθεσία).Τροχιά A	χειρονομία
A5: "σβήσε αυτό"	ομιλία

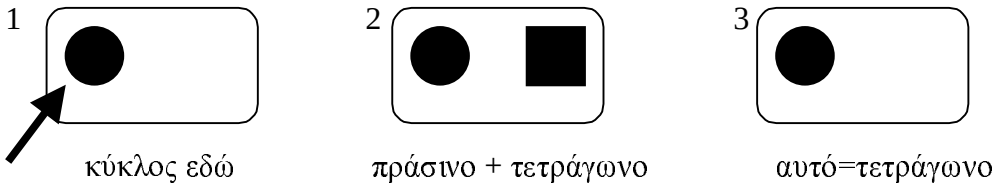


Γ. Το συνεργητικό πλαίσιο

Αυτό το πλαίσιο ορίζεται στο πεδίο της χρήσης τρόπων με την απουσία περιορισμών και στο πεδίο της εξάρτησης πληροφορίας με αλληλοαναφορικούς περιορισμούς μεταξύ μονάδων. Οι ιδιότητες του πλαισίου φαίνονται παρακάτω. Εδώ οι δείκτες και οι αναφορές είναι σωστές όταν έχουμε αναφορές με άλλους τρόπους και οι δείκτες μπορούν να αναλυθούν. Το συνεργητικό πλαίσιο είναι το πιο οικονομικό όσον αφορά την αντίληψη και το μηχανικό προσανατολισμό του χρήστη. Αν και φαίνεται η πιο καλή λύση, το πλαίσιο αυτό είναι προβληματικό στην επεξεργασία αναμενόμενων ή καθηστερημένων ενεργειών.

Παράδειγμα αναφορικής και δεικτικής ανάλυσης

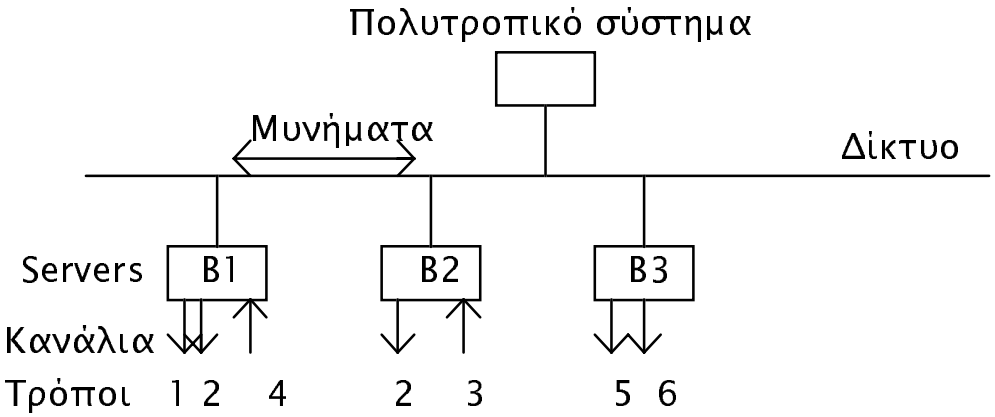
A1: "Ζωγράφισε κύκλο εδώ" + k(τοποθεσία)	ομιλία + χειρονομία
Οι αλληλοαναφορικές δράσεις σημειώνονται με "+"	
A2: "πράσινο" - k(τετράγωνο).A(τοποθεσία).Τροχιά A	ομιλία + χειρονομία
Δράσεις όπου δεν σχετίζονται απαραίτητα με αλληλοαναφορά σημειώνονται με "-"	
A3: "σβήσε αυτό"	ομιλία



Φορμαλισμός

Στην γενικότερη περίπτωση, ένα σύστημα πολυμέσων δεν είναι απαραίτητα επικεντρωμένο. Υποθέτουμε ότι χρησιμοποιεί κατανεμημένους πόρους, που λέγονται servers μέσων (βλέπε σχήμα 3). Αυτοί οι servers μπορεί να περιέχουν υλισμικό για αναγνώριση ομιλίας ή για σύνθεση ομιλίας. Μπορεί επίσης να υπάρχει λογισμικό για αναγνώριση χειρονομίας, και να μην υπάρχει άλλο ειδικό υλισμικό, παρά ένα ποντίκι.

Σε αυτή την περίπτωση, το πολυτροπικό σύστημα από μόνο του γίνεται ένας server χωρίς μέσα. Αυτός ο server μπορεί με τη σειρά του να είναι κατανεμημένος. Σε αυτή την περίπτωση οι λειτουργίες του είναι να διαχειρίζεται τρόπους, γεγονότα και υπηρεσίες, και να συγχωνεύει την πληροφορία σε μερικά κοινά επίπεδα, έτσι ώστε να μπορέσει να το μεταφέρει στην εφαρμογή από μόνος του (ο τρόπος διαλόγου, κατανεμημένος τρόπος, κτλ).



Σχήμα 3. *'Ενα πολυτροπικό σύστημα οργανωμένο γύρω από servers μέσων. (Εξ ορισμού, ένα μέσο είναι ένα φυσικό κανάλι. 'Ενα μέσο διαφέρει από έναν τρόπο).*

Τυπικά, η διαφορά μεταξύ ενός γεγονότος και μιας πληροφορίας -- κατευθυνόμενης μονάδας διατηρείται. Ας ορίσουμε:

Δομές Γεγονότων

1. 'Εστω ότι $\alpha_i(\kappa)$ είναι η i-στή ενέργεια του τρόπου κ που λαμβάνεται (ή εκπέμπεται) από ένα πολυτροπικό σύστημα από (ή προς) ένα σύνολο servers {S}.

'Εχουμε:

γεγονός-από-ενέργεια: αποδίδεται-στο $\alpha_i(\kappa)$

τύπος: $\varepsilon_i(\kappa)$ = {Αρχή $\alpha_i(\kappa)$, Τέλος $\alpha_i(\kappa)$}

τρόπος: κ

χρονολόγηση: $\tau(\varepsilon_i(\kappa))$

σειρά: i

πηγή / προορισμός: {S}

τέλος-γεγονότος

2. 'Εστω ότι $\upsilon_{i\lambda}(\kappa)$ είναι η λ-οστή μονάδα που περιέχεται στο $\alpha_i(\kappa)$

'Εχουμε:

γεγονός-από-μονάδα: αποδίδεται-στο $\upsilon_{i\lambda}(\kappa)$

τύπος: $\varepsilon(\kappa)$ = {Αρχή $\upsilon_{i\lambda}$ $\lambda(\kappa)$, Τέλος $\upsilon_{i\lambda}(\kappa)$}

ενέργεια: α_i $i(\kappa)$

χρονολόγηση: τ ($\varepsilon_{i\lambda}(\kappa)$)

σειρά: λ

τέλος γεγονότος

Δομές Γεγονότων

1. Χρονολόγηση (σημειώνεται $\leq$), μονοτροπική

$\varepsilon_{i\lambda\text{-}\pi}(\kappa) \leq \varepsilon_{i\lambda}(\kappa)$ εαν και μόνον εάν ® $\pi \geq 1$, $\tau(\varepsilon_{i\lambda\text{-}\pi}(\kappa)) \leq \tau(\varepsilon_{i\lambda}(\kappa))$

2. *Συγχρονισμός (σημειώνεται $\approx$), πολυτροπικός*

®$\kappa \neq \kappa'$, $\varepsilon_{i\lambda}(\kappa) \approx \varepsilon_{\iota\lambda'}(\kappa')$ <u>εάν και μόνον εάν</u> $\varepsilon_{i\lambda}(\kappa) \in$ [Αρχή $\upsilon_{\iota\lambda'}(\kappa')$, Τέλος $\upsilon_{\iota\lambda'}(\kappa')$] ή

$$\varepsilon_{i'\lambda'}(\kappa') \in [\text{Αρχή } \upsilon_{i\lambda}(\kappa), \text{Τέλος } \upsilon_{i\lambda}(\kappa)]$$

με

$$\varepsilon_{i\lambda}(\kappa) \in [\text{Αρχή } \upsilon_{i'\lambda'}(\kappa'), \text{Τέλος } \upsilon_{i'\lambda'}(\kappa')] \text{ εάν και μόνον εάν } \tau(\text{Αρχή } \upsilon_{i'\lambda'}(\kappa') \leq \tau(\varepsilon_{i\lambda}(\kappa)) \leq \tau(\text{Τέλος } \upsilon_{i'\lambda'}(\kappa')]$$

όπου (≤) δείχνει μερική σειρά και

(≈) δείχνει ισοδύναμες σχέσεις

Αυτές οι σχέσεις μπορούν επίσης να απευθύνονται στα γεγονότα των πράξεων.

3. Σύγχρονες μονάδες (ενέργειες)

Δύο μονάδες (ενέργειες) είναι σύγχρονες εάν περιέχουν δύο σύγχρονα γεγονότα.

$$\circledR \kappa \neq \kappa', \upsilon_{i'\lambda'}(\kappa') \approx \upsilon_{i'\lambda'}(\kappa') \text{ εάν και μόνον εάν } \exists\ \varepsilon_{i\lambda}(\kappa) \approx \varepsilon_{i'\lambda'}(\kappa') \text{ -ίδιο- για τις πράξεις } \alpha_i$$

Η διάρκεια δύο σύγχρονων μονάδων (ενεργειών) είναι:

$$\delta(\upsilon_{i\lambda}(\kappa) \approx \upsilon i'\lambda'(\kappa')) = \max[\tau(\varepsilon_{i\lambda}(\kappa)), \tau(\varepsilon_{i'\lambda'}(\kappa'))] - \min[\tau(\varepsilon_{i\lambda}(\kappa)), \tau(\varepsilon_{i'\lambda'}(\kappa'))] \text{-ίδιο για τις πράξεις } \alpha_i$$

Δύο Ορισμοί του Παρόντος -- χρόνου

1.*Ακαριαίος παροντικός-χρόνος*: Η διάρκεια της μικρότερης μονάδας σε μία δεδομένη στιγμή.

2.*"Πυκνότητα" του παροντικού–χρόνου*: Το χρονικό διάστημα ορίζεται από τη διάρκεια κάθε σύγχρονης ενέργειας σε μία δεδομένη στιγμή. Η πυκνότητα του παροντικού -- χρόνου ποικίλλει με τον χρόνο.

Ειδικές περιπτώσεις:

- Σε συστήματα εναλλασόμενου τύπου δεν υπάρχει σύγχρονη ενέργεια (πράξη) ή μονάδα.
- Σε συστήματα ταυτόχρονου τύπου, η διαχείριση του τρόπου είναι η ίδια όπως και στα συστήματα συνεργητικού τύπου, εκτός από το επίπεδο συγχώνευσης της πληροφορίας το οποίο δεν υπάρχει.

Γενικό Πλαίσιο Αλληλεπίδρασης σε ένα Δυναμικό Σύστημα

Ενα σύστημα λέγεται ότι είναι δυναμικό, εάν μπορεί να διαχειριστεί διάφορα γενικά πλαίσια αλληλεπίδρασης. Κάθε τύπος γενικού πλαισίου αλληλεπίδρασης έχει περιγραφτεί πιο κάτω σαν μία τριπλέτα:

{χρήση του τρόπου, εξάρτηση πληροφορίας**,** κίνηση/αναπαράσταση}.

1.Η χρήση των τρόπων είναι καθορισμένη από τη δράση - αντίληψη του βρόγχου και από τους μηχανικούς περιορισμούς του συστήματος.

Παράδειγμα: βάλε (αντικείμενο, τοποθεσία)

" βάλε αυτό εδώ" < πχ (αυτό) < πχ (εδώ) => **εναλλασόμενος**

("βάλε αυτό εδώ" ~ πχ (αυτό)) < πχ (εδώ) => **συνεργητικός(ο+)**

("βάλε αυτό" ~ πχ (αυτό)) < ("εδώ" ~ πχ (εδώ)) **=> συνεργητικός**

("βάλε" < ("αυτό" ~ πχ (αυτό)) < ("εδώ" ~ πχ (εδώ)) **=> συνεργητικός (χ+)**

όπου

" " = ενέργεια ομιλίας

πχ = χειρονομική περιγραφή ενέργεια

ο+ = επικράτηση του τρόπου της ομιλίας

χ+ = επικράτηση του τρόπου της χειρονομίας

Σε αυτή την τελευταία περίπτωση, η χειρονομία βάζει σημεία στίξης στην ομιλία και καθορίζει προσωρινή υφή**.** Εδώ, τα γεγονότα είναι σύγχρονα και η πληροφορία είναι εξαρτώμενη. Συμπεραίνουμε λοιπόν ότι το γενικό πλαίσιο αληλλεπίδρασης είναι *συνεργητικό,* με *επικράτηση της χειρονομίας*.

2.Η εξάρτηση της πληροφορίας καθορίζεται από τις εννοιολογικές / πραγματικές σχέσεις μεταξύ των μονάδων.

Παράδειγμα:

πχ(τρίγωνο)~ "μετακίνησε τον κύκλο" => *ταυτόχρονο* γενικό πλαίσιο

Εδώ, και οι δύο πράξεις είναι σύγχρονες και ανεξάρτητες, καθώς το καθορισμένο τρίγωνο δεν παραπέμπει στον κύκλο από την πράξη της ομιλίας. Εδώ συμπεραίνουμε ότι το γενικό πλαίσιο αλληλεπίδρασης είναι "ταυτόχρονο".

Αυτά τα λίγα παραδείγματα δείχνουν ότι το γενικό πλαίσιο αλληλεπίδρασης μπορεί να εξαχθεί σαν συμπέρασμα από την οργάνωση και τα περιεχόμενα των πράξεων. Αυτό σημαίνει ότι μπορεί να προσδιοριστεί μόνο έμμεσα.

Λειτουργίες Διαχείρισης

Συνοψίζοντας, η διαχείριση του τρόπου αποτελείται από:

- Τη σύλληψη των συμβάντων από τους βοηθούς μέσων (αντίστροφα προς τους βοηθούς μέσων για έξοδο).
- Την κατασκευή των δομών του συμβάντος και της πληροφορίας.
- Την διαχείριση του γενικού πλαισίου αλληλεπίδρασης σαν μία λειτουργία τύπου πληροφορίας και γνώσης μεταβιβαζόμενης από γειτονικά επίπεδα (μέτρο συγχώνευσης ή διαλόγου, για παράδειγμα).
- Ο εκσυγχρονισμός μιας ιστορίας αυτού του γενικού πλαισίου.
- Η χρησιμοποίηση της γνώσης πάνω στα αισθητηριο-κινητικά χαρακτηριστικά του χρήστη (χρόνος αντίδρασης, τροπικές προτιμήσεις, κ.τ.λ.).

Η Συγχώνευση και ο Διαχωρισμός της Πληροφορίας

Ο πυρήνας του προβλήματος σε μία πολυτροπική διεπαφή ανθρώπου μηχανής αφορά την συγχώνευση (για είσοδο) και ο διαχωρισμός (για έξοδο) της πληροφορίας. Το μέτρο συναλλαγής της συγχώνευσης (ή, αντιστοίχως, του διαχωρισμού) συνδέει το μέτρο διαχείρησης του τρόπου (που είναι χαμηλότερο) με το μέτρο συναλλαγής μέσω διαλόγου

(το οποίο είναι υψηλότερο). Το τελευταίο έχει καλά ορισμένες λειτουργίες σε μια πολυτροπική διεπαφή ανθρώπου μηχανής.

Αυτές οι λειτουργίες είναι για να:

- Κατασκευάσουν έναν σημειωτικό κόσμο επικοινωνίας (αλληγορίες, γλώσσες, κ.τ.λ.).
- Δομήσουν και οργανώσουν την επικοινωνία.
- Διαχειριστούν και να ελέγχουν δυναμικά την συνεργάσιμη αλληλεπίδραση.
- Διορθώσουν επικοινωνιακά λάθη.
- Εξασφαλίσουν βοήθεια στην μάθηση, καθοδήγηση στην εργασία, κ.τ.λ.

Η ανάγκη για σαφή μέτρο συγχώνευσης έχει τεθεί σαν ζήτημα. Βέβαια, κάθε μία από τις λειτουργίες του μπορεί να τοποθετηθεί αντί γι'αυτό στο μέτρο του διαλόγου, στην περίπτωση όπου το μέτρο του διαλόγου θα μπορούσε να περικλύει την ανάλυση της χαμηλού επιπέδου πληροφορίας και της συγχώνευσης του. Ποιά είναι τα επιχειρήματα που ευνοούν το σαφές μέτρο συγχώνευσης για πολυτροπική διεπαφή ανθρώπου μηχανής; Αυτή είναι μία πολύ πλατιά ερώτηση. Προτείνουμε να την περιορίσουμε στα ακόλουθα σημεία.

Στρατηγική συγχώνευσης. *Πότε* πρόκειται να λάβει χώρα η συγχώνευση; 'Οποτε είναι δυνατό (σε τμήματα μιας βάθους ύψους στρατηγικής); 'Οσο αργότερα είναι δυνατό (δηλαδή καθυστέρηση χρόνου); Σε στάδια; Πώς εκτελείται η συγχώνευση; Γύρω από μια κοινής νοηματικής παρουσίασης δομή και ενός επικρατέστερου τρόπου; Αυτό θα μπορούσε να εμπλέκει μια "ενοποιημένη γραμματική", δηλαδή μία γραμματική που να εξασφαλίζει καλο σχηματισμό και γλωσσολογική επάρκεια της συνδυασμένης εισόδου. 'Η χωρίς επικρατέστερο τρόπο; Αυτό θα μπορούσε να συνεπάγεται τη χρήση μιας πολυτροπικής γραμματικής, η οποία θα χρησιμοποιεί γενική θεωρία δράσης χωρίς κοινή δομή. *Πού* (δηλαδή σε πιο μέτρο) εκτελείται η συγχώνευση; Είναι συγκεντρωμένη σε ένα διαλογικό μέτρο ή εκτελείται με μία απλωμένη και προοδευτική τεχνοτροπία διαμέσω των διαφορετικών μέτρων;

Κριτήρια συγχώνευσης. 'Ενα άλλο ερώτημα αφορά τα κριτήρια για συγχώνευση πληροφοριακών συμβάντων. Θα έπρεπε αυτά να συγχωνεύονται σε σχέση με την προσωρινή αμεσότητα (π.χ. σύμφωνα με αισθητηριοκινητικούς κανόνες); Θα έπρεπε το

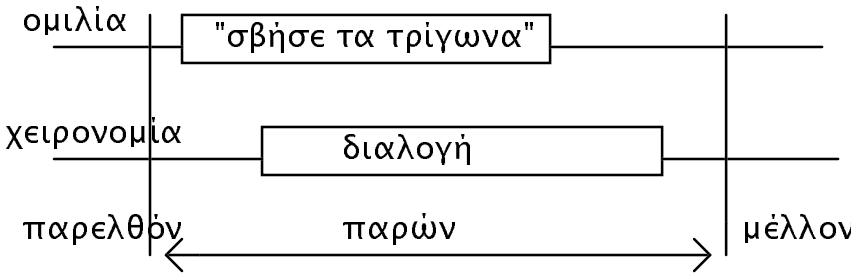
καθοριστικό κριτήριο να έχει δομημένη συνέπεια ή σημασιολογική πληρότητα; Θα έπρεπε να είναι σημασιολογική ομοιότητα / ταυτότητα; 'Η θα έπρεπε να προέρχεται από μία πράξη λογικής ή από μία λογική σκοπιμότητας; Θα έπρεπε η συγχώνευση να αναδύεται σαν μια λειτουργία της αλληλεπίδρασης του γενικού πλαισίου ή σαν μια λειτουργία της προσπάθειας του χρήστη ;

Ο σκοπός αυτής της δημοσίευσης δεν μας επιτρέπει μια λεπτομερή ανασκόπηση κάθε ενός από αυτά τα σημεία. Μόνο λίγα από τα πιο σχετικά θέματα εξετάζονται παρακάτω, πιο πολύ για επεξηγηματικούς σκοπούς. Για να εξηγήσουμε με παραδείγματα λύση αναφοράς, με σκοπό να ξεκινήσουμε τη συζήτηση, ας εξετάσουμε λίγες τυπικές περιπτώσεις απλών πολυτροπικών εντολών.

Τύποι Αναφοράς

1. Αναφορά σε ένα Σύνολο Αντικειμένων

Ας αφήσουμε την πράξη(ενέργεια) ομιλίας "σβήσε τα τρίγωνα" να γίνει ταυτόχρονα με την πράξη χειρονομίας της διαλογής (περικυκλώνοντας) διάφορα γραφικά αντικείμενα πάνω σε μία οθόνη (όπως επιδυκνείεται πιο κάτω).



Χρησιμοποιώντας τον ανώτερο ορισμό της πυκνότητας του παροντικού χρόνου, μία σωστή ερμηνεία αυτών των δύο πράξεων εξαρτάται από την πρόθεση του χρήστη και από το γενικό πλαίσιο στο οποίο οι πράξεις παρήχθησαν. Το γενικό αυτό πλαίσιο, μπορεί με τη σειρά του, να υποδιαιρεθεί σε γενικό πλαίσιο αλληλεπίδρασης, σε γλωσσολογικό γενικό πλαίσιο, σε ασυνάρτητο ή διαλογικό γενικό πλαίσιο, και σε γενικό πλαίσιο δράσης ή εργασίας.

'Ολα αυτά τα πλαίσια εξαρτώνται το ένα από τα άλλα. Αυτή η εξάρτηση παρουσιάζεται στο ακόλουθο παράδειγμα.

- Εάν το γενικό πλαίσιο αλληλεπίδρασης = *συνεργητικό*, το μήνυμα πρέπει να ερμηνευτεί σαν μία σημαντική οντότητα (σβήσε "όλα τα" τρίγωνα από αυτά που διαλέχθηκαν). Το άρθρο "τα" ερμηνεύεται σαν ένας δείκτης και πρέπει να συγχωνευτεί με την πληροφορία που προέρχεται από την χειρονομία. Αντιστρόφως, εάν η πράξη της ομιλίας ήταν "σβήσε *αυτά* τα τρίγωνα", το γλωσσολογικό γενικό πλαίσιο θα έχει επιβάλλει ένα *συνεργητικό* γενικό πλαίσιο διαμέσω του δείκτη "αυτά" (ρισκάροντας την αναμονή για ένα μελλοντικό κομάτι πληροφορίας προερχόμενης από χειρονομία).

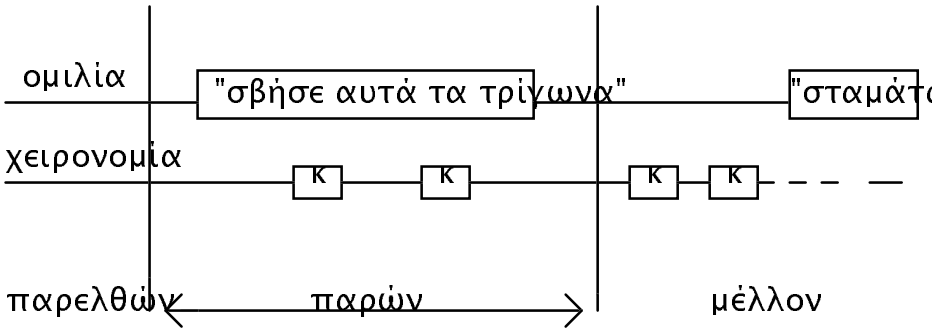
- Εάν το γενικό πλαίσιο αλληλεπίδρασης = *ταυτόχρονο*, η ερμηνεία αποδίδει δύο μυνήματα: Από τη μία μεριά, η εντολή για σβήσιμο "των" τριγώνων αναφερόταν στο παρελθόν, και από την άλλη μεριά, η επιλογή των αντικειμένων σε κάποια μελλοντική ενέργεια. Το άρθρο "τα" τότε, ερμηνεύεται σαν μία αναφορά, και δεν πρέπει να συγχωνευτεί με την από χειρονομίας προερχόμενη πληροφορία.

- Εάν υπάρχει μία διατροπική διαμάχη που να αφορά το τρέχον γενικό πλαίσιο αλληλεπίδρασης, τότε αυτό το πλαίσιο θα μπορούσε να τεθεί σαν ζήτημα, και για τα άλλα γενικά πλαίσια θα μπορούσε να γίνει προσπάθεια να ερμηνευτούν. Αυτή θα μπορούσε να ήταν η περίπτωση όπου, για παράδειγμα, το σύνολο των αντικειμένων που σχεδιάστηκε από την χειρονομία δεν περιελάμβανε τρίγωνο. 'Οπως δείχνουν τα παραδείγματα, η ασάφεια του γλωσσολογικού γενικού πλαισίου και το μη καθόρισμένο πλαισίου αλληλεπίδρασης, μπορεί να αυξήσει τα σημαντικά προβλήματα ερμηνείας.

2. Αναφορά σε μία Σειρά Αντικειμένων

Στην δεύτερη από τις δύο περιπτώσεις που μόλις συζητήθηκαν, η πράξη ομιλίας "σβήσε αυτά τα τρίγωνα" δεν ερμηνεύτηκε σαν ασάφεια όσον αφορά στη χειρονομία. Προσέξτε, ωστόσο, την ταυτόχρονη ερμηνεία των αποτελεσμάτων της σχεδίασης της χειρονομίας, σε ένα γεμάτο επαναλήψεις τρόπο λειτουργιών, όπως ο ορισμός μιας σειράς αντικειμένων, όπου κάθε αντικοίμενο πρέπει να δείχνει στο επόμενο του, π.χ. κινήσεις του ποντικιού.

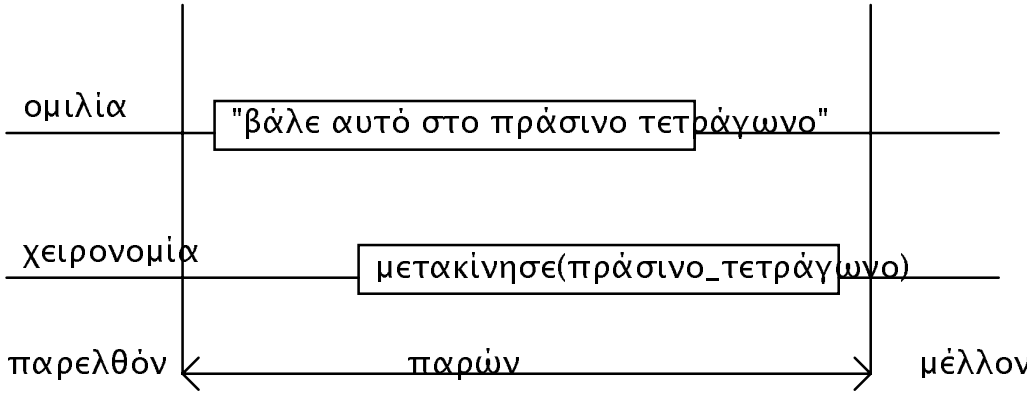
Εναλακτικά μπορούμε να δεσμεύσουμε πληροφορία στον παροντικό χρόνο (όπως ορίστηκε πιο πάνω) σε ένα συνεργητικό γενικό πλαίσιο:



Εδώ, ένα "τέλος του τωρινού δείκτη" ή "τέλος της ενέργειας του δείκτη", όπως ένα διπλό χτύπημα του ποντικιού ή μια προφορική διαταγή "σταμάτα", θέτει τα όρια του "παροντικού χρόνου". Το παράδειγμα αυτό παρουσιάζει ότι η ασάφεια που δημιουργείται από το γενικό πλαίσιο εργασίας μπορεί να επιλυθεί από το το διαλογικό γενικό πλαίσιο.

3. Αναφορά σε ένα Κινούμενο Αντικείμενο

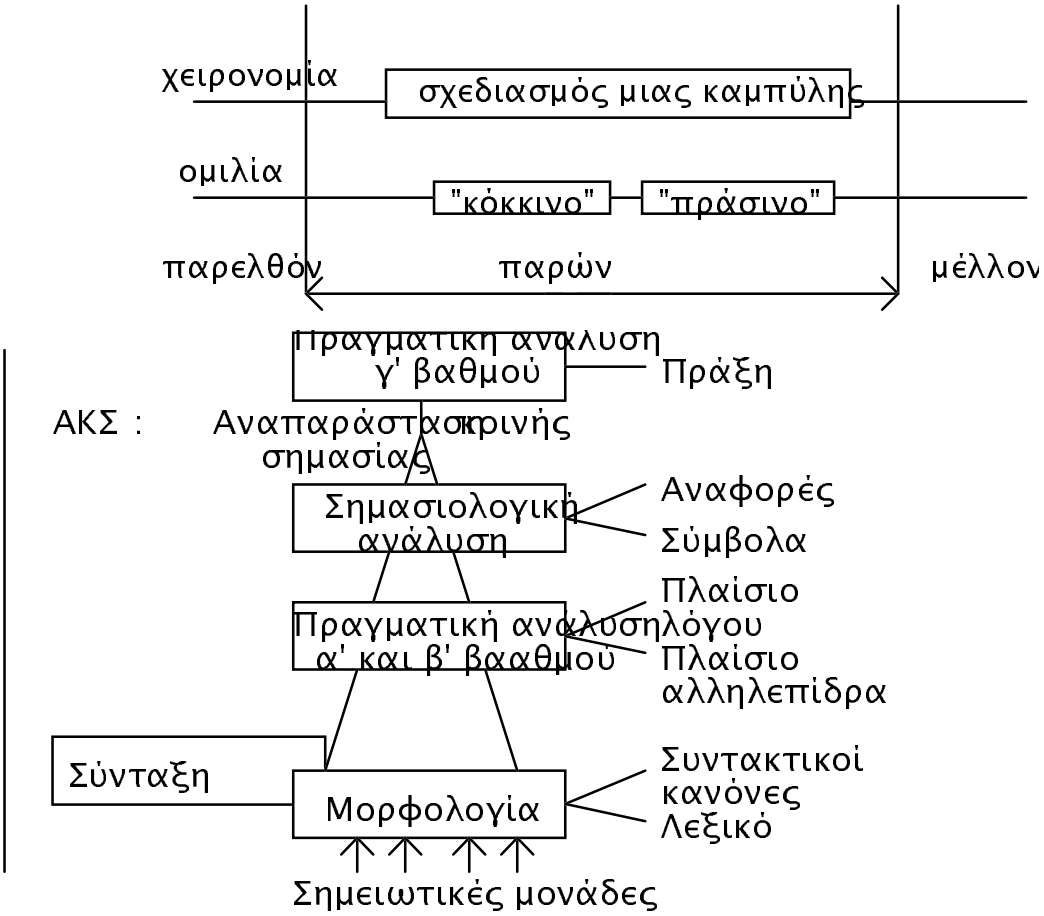
Στην επόμενη περίπτωση, η πράξη ομιλίας "βάλε αυτό στο πράσινο τετράγωνο" αναφέρεται σε ένα κινούμενο αντικείμενο, το οποίο έχει εικονικά οριστεί, από τη στιγμή που επιλέχτηκε χρησιμοποιώντας το ποντίκι.Εδώ η ερώτηση είναι η ακόλουθη: Είναι το αντικείμενο στο οποίο αναφερόμαστε με την πράξη της ομιλίας ίδιο με το αντικείμενο που επιλέχτηκε;



Μία παρόμοια περίπτωση έχει παρατηρηθεί σε πραγματικές καταστάσεις, όταν ο χρήστης θέλει να βελτιώσει τις δράσεις του εκμεταλευόμενος παράλληλους τρόπους. Για παράδειγμα, ο χρήστης θα μπορούσε να πει "βάλε αυτό στο πράσινο τρίγωνο", την ίδια στιγμή όπου άλλαζε το χρώμα του τριγώνου, καταστρέφοντας ως εκ τούτου την αναφορά. Εδώ, και το γενικό πλαίσιο της εργασίας και το γενικό πλαίσιο της αλληλεπίδρασης δημιουργούν την ασάφεια.

4. Απευθείας Αναφορά

Σε αυτή την περίπτωση, ο χρήστης θα μπορούσε, για παράδειγμα, να σχεδιάσει μια καμπύλη χρησιμοποιώντας το ποντίκι, και ταυτόχρονα να αλλάξει το χρώμα τμημάτων της καμπύλης χρησιμοποιώντας ομιλία. Η αναφορά του "χρώματος" ισχύει μόνο κατά τη διάρκεια της χειρονομίας και πέρνει έμμεσα την σημασία του τρέχων αντικειμένου κατά τη διάρκεια της χειρονομίας.



Σχήμα 4. *Τα επίπεδα της συγχώνευσης σε ένα πολυτροπικό σύστημα.*

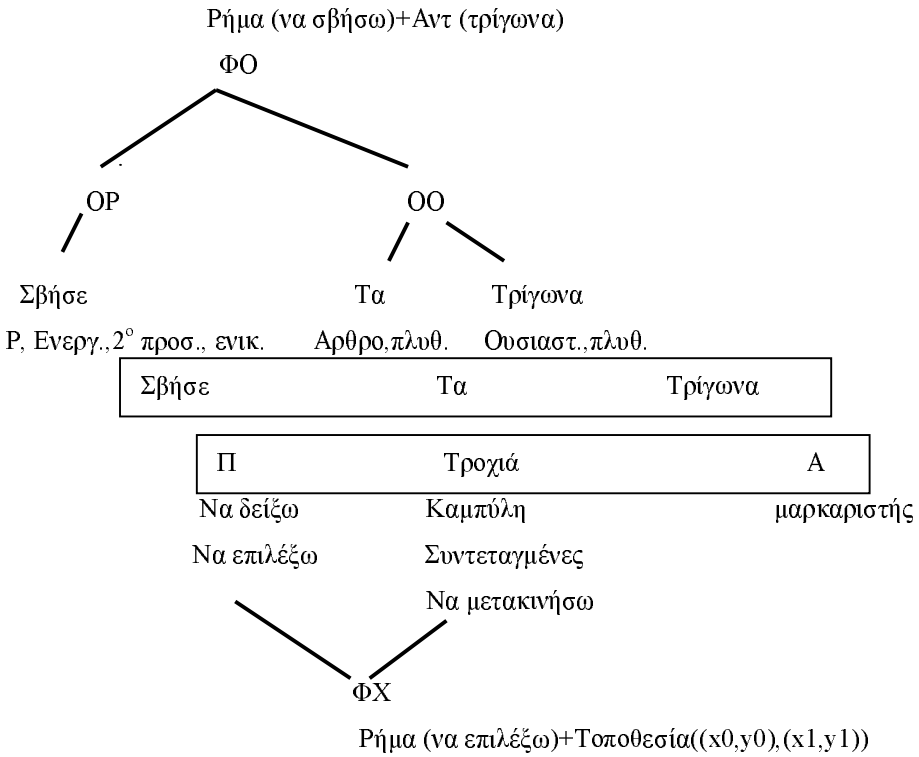
Επίπεδα Συγχώνευσης

Τα ακόλουθα παραδείγματα παρουσιάζουν τον ρόλο του μέτρου συγχώνευσης, ο οποίος έχει δύο πτυχές. Πρώτον, πρέπει να κάνει την ερμηνεία όσο ανεξάρτητη είναι δυνατό από το γενικό πλαίσιο, και τότε πρέπει να επιτρέπει ένα βαθμιαίο διαχωρισμό των αναφορών στις περιπτώσεις ασάφειας. Επίσης,το μέτρο της συγχώνευσης όπως ορίστηκε εδώ μπορεί να επιτρέψει νέους τρόπους χωρίς σημαντική τροποποίηση του μέτρου του διαλόγου. Αυτός ο διπλός ρόλος μας οδηγεί στο να προτείνουμε μια βαθμωτή συγχώνευση της πληροφορίας, ξεκινώντας από το μορφο-συντακτικό επίπεδο και τελειώνοντας στο σημασιολογικό επίπεδο. Αυτό παρουσιάζεται στο Σχήμα 4.

Σε αυτό το σχήμα, η συγχώνευση εκτελείται πάνω σε μονάδες συγκεντρωμένες από την πυκνότητα του παροντικού χρόνου. Επιφέρει αφηρημένες δομές αναπαράστασης οι οποίες δεν περιλαμβάνουν πια ένα συστατικό τρόπου (ΑΚΣ-αναπαράσταση κοινής σημασίας, CMR-common meaning representation). Τότε αυτές οι δομές μεταβιβάζονται στον ελεγκτή του διαλόγου. Ακολουθεί μία λεπτομερής επιθεώρηση κάθε σταδίου της διαδικασίας συγχώνευσης.

Μορφο-συντακτική Τροπική Ανάλυση

Αυτός ο τύπος ανάλυσης εκτελείται για κάθε πράξη που ανιχνέυεται στον παροντικό χρόνο. Αποδίδει μια αναπαράσταση που προσαρμόζεται σε κάθε τρόπο, ο οποίος περιγράφει και τη δομή των στοιχείων τη λειτουργική δομή. Για παράδειγμα, στην περίπτωση της εντολής: ("σβήσε τα τρίγωνα" + χειρονομική διαλογή αντικειμένων), αυτή η ανάλυση αποδίδει την ακόλουθη αναπαράσταση:



ΦΟ Φράση Ομιλίας, ΟΡ Ομάδα Ρημάτων, ΟΟ Ομάδα ουσιαστικών, ΦΧ Φράση Χειρονομίας, ΟΠ Ομάδα Προθέσεων

Πραγματική Ανάλυση του Πρώτου και του Δευτέρου Βαθμού

Αυτό το στάδιο περιέχει το διατροπικό δέσιμο των δεικτών και των πραγματολογικών μαρκαριστών. Αυτή η ανάλυση καταλήγει στην δημιουργία συνδέσμων μεταξύ ελεύθερα αναφερόμενων στοιχείων από έναν τρόπο, και αναφερόμενων στοιχείων από άλλους τρόπους. Η ανάλυση επίσης επιτρέπει τη δημιουργία συνδέσμων μεταξύ πράξεων.

Γυρνώντας στο ίδιο παράδειγμα, υποθέστε ότι οι ακόλουθες αξίες μεταφέρονται στα γειτονικά μέτρα: γενικό πλαίσιο αλληλεπίδρασης = συνεργητικό, γενικό πλαίσιο διαλόγου = οδηγία, γενικό πλαίσιο εργασίας = τρέχουσα δράση.

Ανάλυση του δείκτη "τα": Το δέσιμο της λέξης "τα" με τα αντικείμενα τα οποία ορίζονται με χειρονομία, είτε σαν δείκτης είτε σαν αναφορά, χρησιμοποιεί τους ακόλουθους λογικούς κανόνες:

® (αντικείμενα): αντικείμενο ∈ Περιοχή (καθορισμένη)

® (αντικείμενα): αντικείμενο ∈ Ιστορία των ορατών αντικειμένων

Περιοχή = Εσωτερικό (Τροχιά).

Η ανάλυση αποδίδει μια λίστα από αντικείμενα, αδιαφορώντας για την σημασιολογική κατηγορία τους (αυτό σημαίνει ότι αδιαφορεί για το γεγονός αν είναι τρίγωνα ή όχι).

Ανάλυση των πραγματολογικών μαρκαριστών: Προσωδία, "Π" ενεργεί. Η Προσωδιακή ανάλυση (η οποία είναι εκτός του σκοπού αυτής της δημοσίευσης) βοηθά στο να κατηγοριοποιήσεις μια πράξη ομιλίας. Στην περίπτωση του παραδείγματος, μια συνεχόμενα κατερχόμενη μελωδική καμπύλη θα μπορούσε να υποδηλώνει μια κατηγορηματική πράξη ομιλίας. "Α", αναλύεται σαν ένας μαρκαριστής που υποδηλώνει το τέλος της πράξης χειρονομίας. Αυτές οι δύο μονάδες πληροφορίας - κατεύθυνσης υποδηλώνουν ότι η πράξη έχει ολοκληρωθεί. Είναι επομένως λογικό να υποθέσουμε ότι συνιστούν μια οντότητα. Θεωρώντας ένα συνεργητικό γενικό πλαίσιο αλληλεπίδρασης, συνεχίζουμε με τη χωρο-χρονική πραγματολογική ανάλυση.

Χωρο-χρονική σημασιολογική ανάλυση

Εδώ μπορούν να παραχθούν παραδείγματα σχημάτων δράσης και υποκειμένου χρησιμοποιώντας μια ΑΚΣ (Αναπαράσταση Κοινής Σημασίας). Σε αυτό το επίπεδο εμπλέκονται περίπλοκοι μηχανισμοί για σημασιολογική ερμηνεία της φυσικής γλώσσας. Αυτοί οι μηχανισμοί χρησιμοποιούν γνώση βασισμένη σε δράσεις και υποκείμενα, καθώς επίσης και κανόνες εξαγωγής συμπεράσματος έτσι ώστε να προσαρμοστούν τα σχήματα στη τρέχουσα περίπτωση. Αυτοί οι ποικίλοι μηχανισμοί συνδέονται με τη εκάστοτε εφαρμογή, ανάλογα με το βαθμό γενικότητας. Ο φορμαλισμός για την αναπαράσταση γνώσης χρησιμοποιεί πολυτροπική γραμματική που βασίζεται στην περίπτωση. Το $ υποδυκνύει ένα προτότυπο ή μια τάξη (Σ για σημασιολογία Ν για συντακτικό). Ο "Σύνδεσμος (Link)" είναι μια ιδιότητα που χρησιμοποιείται για την ένωση δυο πολυτροπικών πληροφοριών.

Η γνωσιολογική βάση των Δράσεων

Δράση : Να σβήσω
Ενεργοποίηση = διπλό κλικ ($ΣΑντ) | Ρήμα($ΣΣβήσε)
ΑΝΤ = ΟΟ($Επικρατέστερα=$ΣΑντ) | κλικ($ΣΑντ)
Χρόνος=ΟΠ(προθ($ΝΧρόνος).ΟΟ) | Επι($ΝΧρόνος)

Δράση : Να τοποθετήσω
Ενεργοποίηση=mvt-κλικ($ΣΑντ) | Ρήμα ($ΣΜετακινήσου)
ΑΝΤ = ΟΟ($Επικρατέστερα=$ΣΑντ) | κλικ($ΣΑντ)
Τοποθεσία= ΟΠ(προθ($ΝΤοποθεσία).ΟΟ) | Επι($ΝΤοποθεσία) | κλικ($ΣΤοποθεσία)
Χρόνος = ΟΠ(προθ($ΝΧρόνος).ΟΟ) | Επι($ΝΧρόνος)
κτλ.

Η γνωσιολογική βάση των Αντικειμένων

Τρίγωνο: είδος γεωμετρικού-αντικειμένου
Μέγεθος: ΟΕ($Επικρατέστερα = $ΣΑντ) | mvt-κλικ($ΣΑντ)
Χρώμα: ΟΟ($Επικρατέστερα = $ΣΑντ) | κλικ($ΣΠαλέττα)
Συντεταγμένες: (x,y)

Δράσεις: {Να σβήσω, Να τοποθετήσω)

Κόσμος-Αναφοράς: Τεχνητός

Σημασιολογικός-Σύνδεσμος = Συνώνυμο($Πυραμίδα)

κτλ.

Λεξικό για " Γραμματικές Λέξεις" και " Ενδεικτικούς"

Εκείνα:

Παραστατικός= κλικ($Αντ) | Παραστατικός($Εκείνα)

Πραγματολογική-Σύνδεσμος = Δείκτης($Αντ)

Αυτό:

Προσωπική αντωνυμία, πληθυντικός = Π-Π($Αυτό)

Πραγματολογική-Σύνδεσμος = Αναφορά($Αντ)

Τα:

Άρθρο, πληθυντικός = Π-Π($Το)

Πραγματολογική- Σύνδεσμος = Δείκτης($Αντ)

Εδώ:

Τοπικό επίρρημα = Επι – Τοπ($Εδώ)

Πραγματολογικός-Σύνδεσμος = Δείκτης($Τοπ)

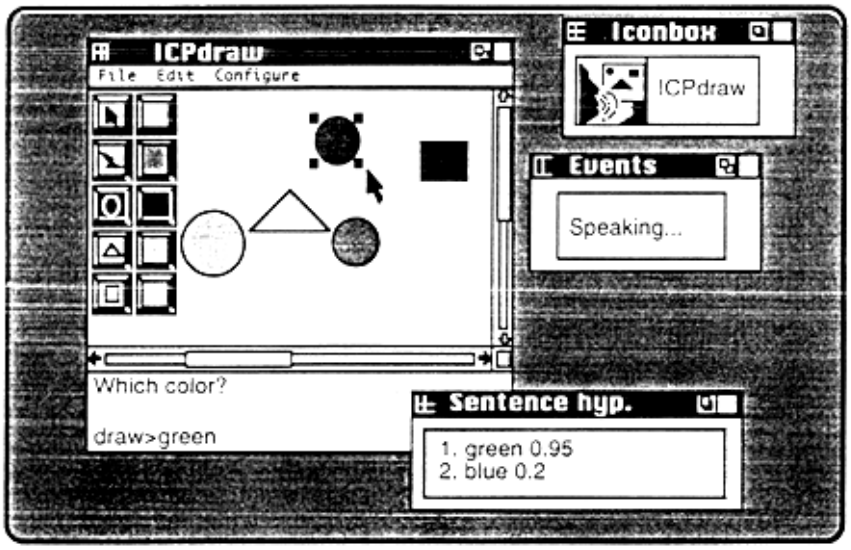
κτλ.

Μπορούν να εισαχθούν διάφοροι κανόνες σε σχέση με τα σχήματα δράσης και αντικειμένου:

- Κανόνες για την ενεργοποίηση των δράσεων χρησιμοποιώντας μια έρευνα για την υποστήριξη (με την ευρεία έννοια, δηλαδή ή μια γλωσσολογική λέξη ή μια ορισμένη μονάδα χειρονομίας). Σε περιπτώσεις όπου βρίσκουμε δυο υποστηρίξεις, χρησιμοποιούνται άλλοι κανόνες για να επεξεργαστούν είτε τον πλεονασμό (όταν δυο σχήματα χρησιμοποιούνται) είτε τη σύγκρουση (όπου μόνο ένα σχήμα χρησιμοποιείται).
- Κανόνες για την ενεργοποίηση των αντικειμένων χρησιμοποιώντας ειδικές λειτουργίες στον κατάλογο των αντικειμένων του προηγούμενου βήματος.
- Κανόνες για τη βελτίωση των ιδιοτήτων του αντικειμένου και την ενεργοποίηση του ιδεατού κόσμου.

- Κανόνες για καθορισμό των χωρο-χρονικών ιδιοτήτων των δράσεων, χρησιμοποιώντας μια έρευνα μέσω του λεξικού και του συντακτικού.

Λύσεις: Οι(Η) υποθέσεις(η) μεταδίδονται στον ελεγκτή διαλόγου, ο οποίος δρα ως συνδετικός κρίκος με τα ανώτερα στρώματα. Αυτές οι υποθέσεις αντιπροσωπεύονται από μια αλυσιδωτή λίστα από σχήματα.



Σχήμα 5 Παράδειγμα μιας ICPdraw οθόνης. Αποτελείται από τέσσερα παράθυρα. Το πρώτο (λέγεται ICPdraw) διαχωρίζεται σε μια ζώνη εργασίας, μια ζώνη γραφής (για γραπτές εντολές) και καταλόγους επιλογών. Το δεύτερο παράθυρο (λέγεται Iconbox) δείχνει ένα λογότυπο. Το τρίτο παράθυρο (λέγεται Events) δείχνει την κατάσταση των καναλιών επικοινωνίας. (Για παράδειγμα, αυτό το παράθυρο δείχνει πότε ο χειριστής μπορεί να χρησιμοποιήσει ομιλία ως είσοδο. Το τέταρτο παράθυρο (λέγεται sentence hyp) μας δείχνει τα αποτελέσματα της αναγνώρισης ομιλίας. (Οι τέσσερεις καλύτερες υποθέσεις καταγράφονται και εμφανίζονται με φθίνουσα σειρά αποτελέσματος).

ICPdraw: Ένα παράδειγμα πολυτροπικής ΔΑΜ

Είναι μια αντικειμενοστραφής εφαρμογή που χρησιμοποιεί πολυτροπικότητα. Προσφέρει ένα tool-box γραφικών και καταλόγους λειτουργιών, οι οποίοι ενεργοποιούνται χρησιμοποιώντας ομιλία, γραφή ή χειρονομία (μέσω του ποντικιού). Περιλαμβάνει τις συνήθεις λειτουργίες για σχεδίαση: Επιλογή αντικειμένου είτε δείχνοντας, περικυκλώνοντας ή προφορικά, μετακινώντας ένα αντικείμενο χρησιμοποιώντας είτε ομιλία είτε το ποντίκι, αλλάζοντας το χρώμα ενός αντικειμένου κτλ. Τα αντικείμενα μπορούν να ομαδοποιηθούν και να αποκρυφθούν. Τα σχήματά τους μπορούν να αλλαχθούν χρησιμοποιώντας “μοχλούς”. Το πλαίσιο αλληλεπίδρασης είναι συνεργητικό.

Αρχιτεκτονική Λογισμικού

Το σύστημα περιλαμβάνει:

Α. Το λειτουργικό πυρήνα της ICPDraw εφαρμογής

Β. Τα επίπεδα συγχώνευσης

Γ. Διαχείριση τρόπων

Δ. Servers πολυμέσων

Αυτά τα μοντέλα μεταχειρίζονται σαν UNIX™ διαδικασίες, οι οποίες τρέχουν παράλληλα και επικοινωνούν χρησιμοποιώντας ένα πρωτόκολο (ICP). Στην πραγματικότητα, το γεγονός ότι οι εντολές χειρονομίας και ομιλίας μπορούν να ταυτόχρονες, απαιτεί μια κατανεμημένη αρχιτεκτονική βασισμένη στην ιδέα των ανεξάρτητων "εξυπηρετήσεων". Ένα δευτερεύον κίνητρο για αυτήν την αρχιτεκτονική είναι το κόστος του υλισμικού: εδώ χρειάζεται μόνο ένας server ομιλίας για πολλές μηχανές.

Η ΔΑΜ βασίζεται στο πρότυπο X-Windows , το οποίο διευθύνει το πληκτρολόγιο και το ποντίκι αλλά όχι την ομιλία. Για να γίνει αυτό αναπτύχθηκε μια "εξυπηρέτηση ομιλίας". Περιλαμβάνει έναν server σημάτων (το σήμα είναι διαθέσιμο σε όλο το δίκτυο) και δυο client –διαδικασίες, την αναγνώριση ομιλίας και τη σύνθεση ομιλίας. Οι clients και οι servers μοιράζονται μια κοινή μνήμη και ανταλλάσουν δεδομένα με σύγχρονο τρόπο (η διαχείρηση γίνεται με σηματοφορείς). Όπως βλεπουμε στο σχήμα 6, όλο το περιβάλλον μπορεί να διανεμηθεί σε πολλές μηχανές.

Ο server ομιλίας είναι ένα υποπρόγραμμα με δυο κύριες λειτουργίες. Η πρώτη είναι να λάβει το σήμα ομιλίας. Αυτό γίνεται είτε όταν το σήμα περιμένει είτε μέσω ενός ανακλαστικού που πυροδοτείται όταν το επίπεδο ήχου είναι αρκετά υψηλό. Η δεύτερη λειτουργία του server ομιλίας είναι να στέλνει το σήμα σε όλο το δίκτυο (αυτές οι λειτουργίες είναι συμμετρικές για έξοδο ομιλίας).

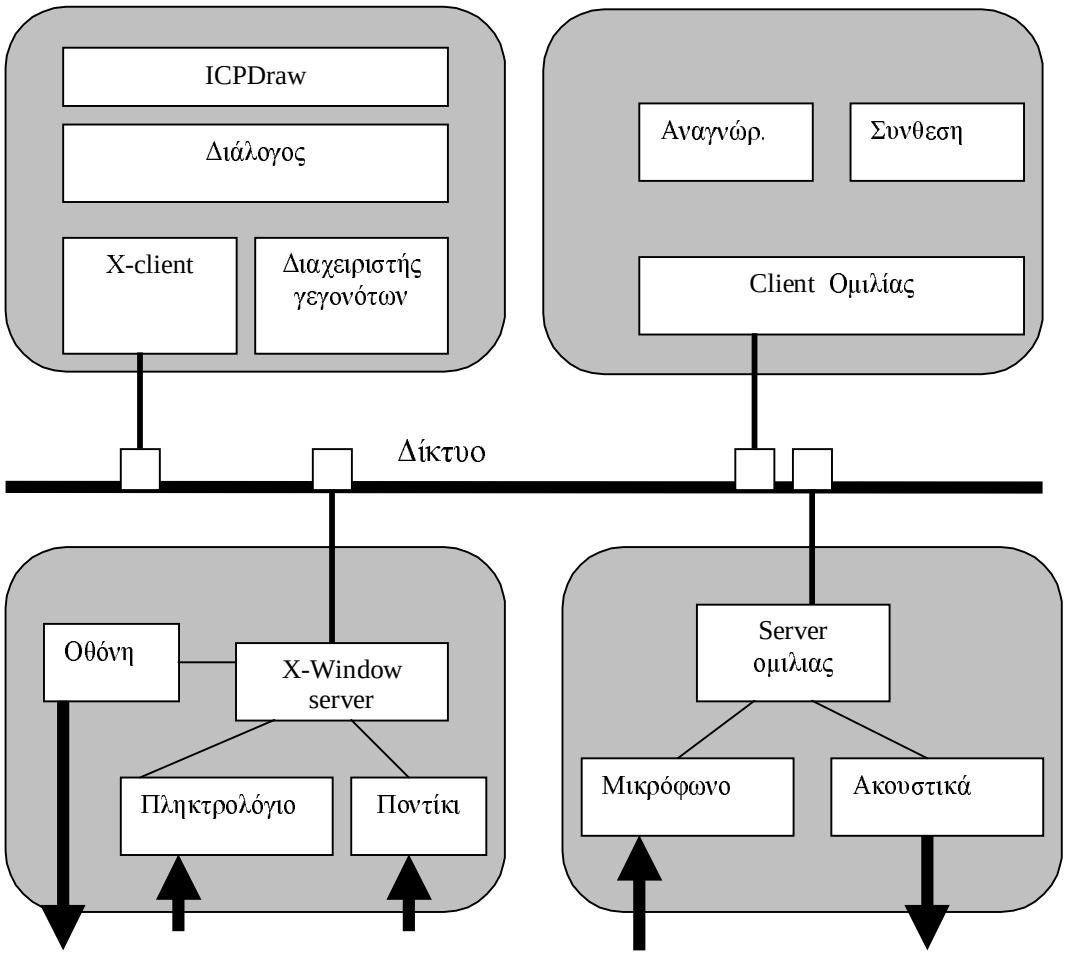
Οι client-διαδικασίες (αναγνώριση και σύνθεση) είναι γενικά απενεργοποιημένες. Ο server ομιλίας τις ενεργοποιεί τη διαδικασία αναγνώρισης μόλις το σήμα είναι διαθέσιμο και έτοιμο για ανάλυση. Η εφαρμογή ενεργοποιεί τη διαδικασία σύνθεσης μόλις είναι έτοιμο το σήμα εξόδου. Όταν ολοκληρωθεί αυτό οι clients εκπέμπουν τα αποτελέσματα και δίνουν τον έλεγχο στον server. Ο server τότε τοποθετεί τα αποτελέσματα (με τη σειρά που τα λαμβάνει) ως εξωτερικά γεγονότα στη σειρά γεγονότων του X server. Ο X

client είναι τότε έτοιμος να λάβει την ομάδα πολυτροπικών γεγονότων που είναι διαθέσιμη από το μοντέλο διαλόγου.

Γλώσσες χειρισμού

Η αφηρημένη γλώσσα του ICPDraw, για χειρισμό αντικειμένου ορίζεται από το συντακτικό εντολών " Δράση (<arg1><arg2>…<argn>)", στο οποίο προσκολλάται ένα κομμάτι χειρονομίας ή ομιλίας. Εδώ,

- Η δράση αντιπροσωπεύει μια στοιχειώδη εργασία. Υποδεικνύεται από το "κατηγόρημα" (υποδεικνύεται από το ρήμα) στην πρόταση.
- Τα arg_i είναι τα ορίσματα της δράσης είτε τύπου ΟΟ είτε ΟΠ. Η ΟΟ είναι συνήθως αντικείμενο της εφαρμογής. Στην περίπτωση της ομιλίας τα επίθετα είναι ιδιότητες του αντικειμένου (όπως χρώμα, μέγεθος κτλ.), ενώ στην περίπτωση των χειρονομιών τα επίθετα είναι κατευθύνσεις ή τροχιές.



***Σχήμα 6**. Συστατικά του ICPDraw σε μια κατανεμημένη αρχιτεκτονική σε UNIX .*

1.Προφορική γλώσσα

Κάθε συστατικό αυτής της γλώσσας είναι προαιρετικό. Για παράδειγμα, οι εντολές "ζωγράφησε πράσινο κύκλο" και "όχι…πράσινο" αποδεκτές μορφές. Η γραμματική ορίζεται ακολούθως:

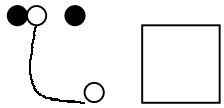
Action -> V.NG1.Location2
Action -> V.Pr
Reit -> NG1.Location2
Rectif -> no.NG1.Location2
Rectif -> more.AdjT
NG1 -> Det2.AdjT.N.AdjC
Pr -> {The}
V -> {Draw, move, errase, change, undo, select, duplicate, quit, etc.}
Det -> {the, one, two, three, four, thiw, these, etc.}
AdjT -> {large, small}
N -> Obj
Obj -> {square, circle, triangle, etc.}
AdjC -> { white, black, blue, yellow, red, green}
Location -> NG2 | LocP2 | LocA
NG2 -> LocP1.N.AdjC
LocP1 -> {under the, under this, on the, on this, besides the, besides this, etc.}
LocP2 -> {to the right, to the left, above, under, in the center, etc.}
LocA -> {here, there, over there, towards here, towards there, around here, etc.}

2.Γλώσσα με χειρονομίες

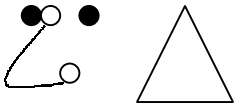
Μια τεχνητή γλώσσα χειρονομιών προστίθεται στις συνηθισμένες χειρονομίες που συναντώνται στις άμεσου χειρισμού ΔΑΜ. Αυτή η τεχνητή γλώσσα λειτουργεί ως εξής:

“ζωγράφισε ένα τετράγωνο”

κ(Τοποθεσία).Π(Τοποθεσία).Τροχιά(τετράγωνο).Α

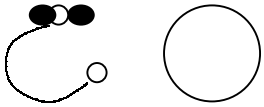


“ζωγράφισε ένα τρίγωνο”



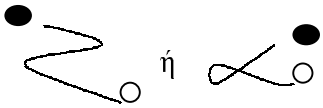
κ(Τοποθεσία).Π(Τοποθεσία).Τροχιά(τρίγωνο).Α

“ζωγράφισε ένα κύκλο”



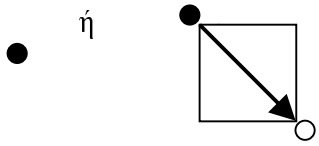
κ(Τοποθεσία).Π(Τοποθεσία).Τροχιά(κύκλος).Α

“σβήσε (αντ)”



Π(Αντ).(Τροχιά(Ζ) | Τροχιά(α)).Α

“επέλεξε (ένα ή πολλά) αντικείμενα(s)”

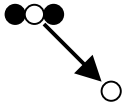


κ(Αντ) | Π(Τοποθεσία).Τροχιά.Α

“μετακίνησε(αντ)”

Π(Αντ).Τροχιά.Α

“αντέγραψε(αντ)”



κ(Αντ).Π(Αντ).Τροχιά.Α

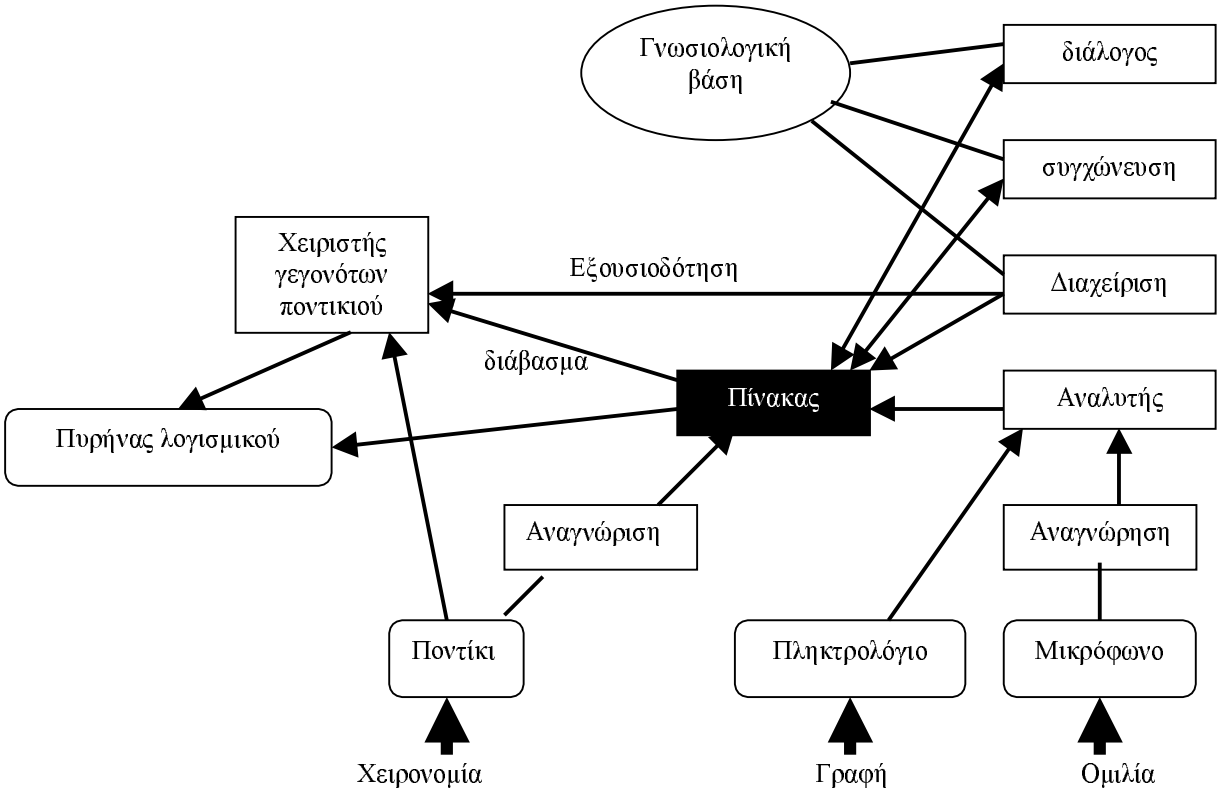
“προσδιόρισε (Τοποθεσία)” = k(τοποθεσία)

Οι εντολές μπορούν να τυπωθούν, να εκφωνηθούν ή να εκφραστούν με χειρονομίες (μέσω του ποντικιού). Το μοντέλο, το οποίο επεξεργάζεται είσοδο φυσικής γλώσσας (τυπωμένη ή εκφωνημένη), χρησιμοποιεί υπομοντέλα για τη γλωσσολογική ανάλυση της εντολής. Αυτοί οι αναλυτές αποδίδουν τη δομή των συστατικών της εντολής (σ-δομή) και τη δομή της λειτουργίας των συστατικών της εντολής (λ-δομή). Επίσης παράγουν τις τέσσερεις καλύτερες λύσεις (ως αλυσίδες χαρακτήρων). Οι αναλυτές χρησιμοποιούν μία λεξική λειτουργική γραμματική (ΛΛΓ-LFG) και το σύστημα αναγνώρισης ομιλίας εξελίχθηκε ως μέρος του Esprit Multiworks project (ESPRIT II project no.2105). Το σύστημα αναγνώρισης Ομιλίας χρησιμοποιεί μοντέλα Markov για αλυσιδωτές λέξεις.

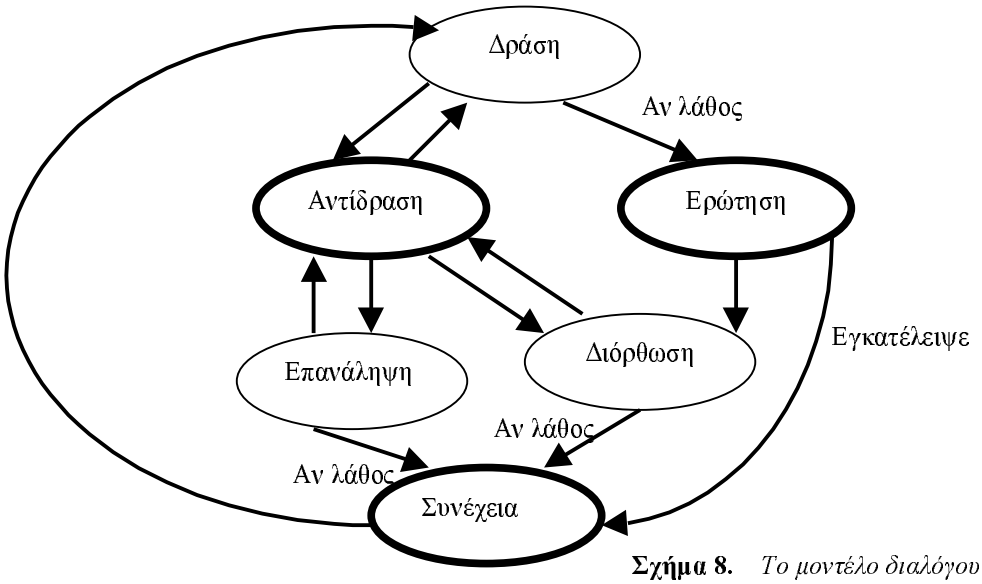
Δουλεύοντας με το ICPDraw

Το ICPDraw χρησιμοποιεί τις μεθόδους που περιγράφηκαν προηγουμένως. Η συγχώνευση είναι πολύ πιο απλοποιημένη, αφού τα προβλήματα που συνδέονται με στις αναφορές είναι περιορισμένα. Η διαχείριση γεγονότων γίνεται μέσω ενός πίνακα ποόπου καταχωρούνται τα γεγονότα, τις μονάδες, εις ενέργειες και τις ΑΚΣ. Ένας βρόγχος αναμονής χρησιμοποιείται για παρατήρηση και για τη διαχείρηση του πίνακα. Τα γεγονότα που σχετίζονται με το ποντίκι, απαιτούν δυο είδη διαχείρησης. Από τη μία, η ανάδραση σε κάποιες δράσεις πρέπει να είναι άμεση, ενώ από τη άλλη ειδικές διαδικασίες αναγνώρισης απαιτούν ολοκληρωμένες εντολές. Η δειγματοληψία πρέπει να προσαρμόζεται έτσι ώστε να επεξεργάζεται περιπτώσεις όπως “ματακίνησε αυτό το τρίγωνο”, όπου η προφορά της λέξης “αυτό” συγχρονίζεται με ένα κλικ του ποντικιού.

Τα πολυτροπικά γεγονότα διευθύνονται μέσω ενός πίνακα που τον χειρίζεται ο διαχειριστής. Ο διαχειριστής έχει πρόσβαση στη γνώση της εφαρμογής. Διατηρεί ένα ιστορικό των γεγονότων, ένα ιστορικό αντικειμένων (ως αλυσίδα δομών δεδομένων), και προειδοποιεί την εφαρμογή όταν η εντολή είναι έτοιμη.



Σχήμα 7. *Λειτουργία του ICPDraw*



Σχήμα 8. *Το μοντέλο διαλόγου*

Η στρατηγική του διαλόγου στο ICPDraw είναι εξ'ολοκλήρου αποκριτική (Σχήμα 8).

- Αν η ερμηνεία (η μετάφραση) μιας ενέργειας είναι σωστή, και αν η δράση είναι εκτελέσιμη, τότε το σύστημα βρίσκεται σε κατάσταση "απόκρισης", οπότε και εκτελείται η εντολή του χρήστη. Έπειτα το σύστημα περιμένει για μια νέα εντολή όπου μπορεί να είναι "Δράση", "Επανάληψη" ή "Διόρθωση".
- Αν η ενέργεια οδηγήσει σε μια ανεπιθύμητη κατάσταση, το σύστημα κάνει ερώτηση και μπαίνει σε μια κατάσταση "αναμονής" (αναμένει για μια εντολή διόρθωσης). Το σύστημα δεν επιτρέπει δυο συνεχόμενα λάθη και εισάγεται σε μια κατάσταση "συνέχειας" (περιμένει για νέα δράση). Η αποκλίσεις αποφεύγονται σκόπιμα, και το σύστημα τις μεταχειρίζεται σαν "εγκατάλειψη" (επιστροφή σε καινούρια δράση).

Παρακάτω παρατίθεται ο αλγόριθμος λειτουργίας:

```
Διάλογος
    {
     Εισαγωγή (Επιλογή(d))
     Ερμηνεία (Ενέργεια, δράση)
     Ενώ Ενέργεια "Απόρριψη" ΚΑΝΕ
          {
            Αν Ενέργεια = λάθος τότε Ερώτηση(d)
               {
                 Ερμηνεία (Ενέργεια, διόρθωση)
                      {
                        Αν λάθος τότε Συνέχεια
                        Αλλοιώς Απόκριση
                       }
             αλλοιώς Απόκριση
      Απόκριση:Ερμηνεία(Ενέργεια (Δράση Ή Διόρθωση Ή Επανάληψη))
                 Αν λάθος τότε Συνέχεια
                 αλλοιώς Απόκριση
                }
      Συνέχεια : Ερμηνεία(Ενέργεια, Δράση)
      Τέλος-Ενώ
         }}Τέλος Διαλόγου
```

Όπου:

Ερμηνεία(X,Y) : διαδικασία για κατασκευή κειμένου X στο πλαίσιο Y.

Ερώτηση : Υπο-διάλογος για αίτηση.
Απόκριση : Εκτέλεση μιας ενέργειας και αναμονή νέας
Συνέχεια : Ανάμονή νέας δράσης.

Η ερμηνεία διεκπαιρεώνεται εφαρμόζοντας δράση και σχήματα αντικειμένων. Η εντολή εκτελείται μόνο εάν είναι συμπληρωμένη (και το αποτέλεσμα ορατό ευθύς αμέσως). Αν δεν είναι, τότε το σύστημα δεν αποκρίνεται.

Αποτελέσματα και συζήτηση

Το ICPDraw είναι μια εφαρμογή πλατφόρμας που πληρείται επάνω σε ένα σταθμό εργασίας. Προσφέρει αναγνώριση ομιλίας αλυσιδωτής λέξης. Το λεξικό περιλαμβάνει περίπου πενήντα λέξεις, και η σύνταξη είναι περιορισμένη. Ο σταθμός είναι αρκετά ισχυρός να επιτρέψει διάλογο σε πραγματικό χρόνο (παρά το γεγονός ότι παράλληλα τρέχει το σύστημα αναγνώρισης ομιλίας).

Το ICPDraw μπορεί να ελέγξει και να αξιολογήσει ιδέες που αναφέρονται τόσο σε πολυτροπικό διάλογο όσο και σε κατανεμημένη αρχιτεκτονική. Όσον αφορά στην τελευταία, το ICPDraw συνέβαλλε ουσιαστικά στο σχεδιασμό της επεξεργασίας και κατανομής σήματος σε διαφορετικούς σταθμούς εργασίας (εφαρμογή client-server) οι οποίοι δεν συμπεριλάμβαναν κάποιο ειδικό υλισμικό για τη λήψη του σήματος ή για αναγνώριση ομιλίας. Όσον αφορά στο πολυτροπικό διάλογο, το ICPDraw συνέβαλε στο να αναγνωρισθούν προβλήματα που έχουν να κάνουν με ταυτόχρονη χρησιμοποίηση ομιλίας και ποντικιού, ειδικά μάλιστα για κινούμενα αντικείμενα (για παράδειγμα, όταν μια φωνητική εντολή περιγράφει ένα αντικείμενο, ενώ ταυτόχρονα μια εντολή του ποντικιού αλλάζει αυτό το αντικείμενο).

Επι προσθέτως, το ICPDraw μπορεί να χρησιμοποιηθεί στη μελέτη της χρήσης της πολυτροπικότητας. Τέτοιες μελέτες είναι απαραίτητες για να πιστοποιηθεί αφενός αν σε κάποιες συγκεκριμένες περιπτώσεις, είναι σκόπιμη η χρήση κάποιου τρόπου αντί για κάποιον άλλο, αφετέρου αν οι χρήστες αλλαζουν τις συνήθειές τους ώστε να επωφελούνται πραγματικά από την πολυτροπικότητα. Μια περιορισμένη έρευνα με δέκα περίπου φοιτητές, δείχνει ότι η πολυτροπικότητα αυξάνει την αποδοτικότητα του χρήστη,

αφού αρκετές εντολές μπορούν να εκτελούνται παράλληλα. Ειδικότερα, αυτή η παραλληλία χρησιμοποιείται για να προβλέπει και να προετοιμάζει τα σκίτσα αντικειμένων. Αυτό σημαίνει ότι μπορούν να βρεθούν νέες στρατηγικές εξ'αιτίας των νέων δυνατοτήτων που προσφέρει η πολυτροπικότητα. Για παράδειγμα, μια κοινή στρατηγική είναι η δημιουργία αντικειμένων οπουδήποτε στην οθόνη μέσω ομιλίας, με ταυτόχρονη τοποθέτησή τους σε συγκεκριμένα σημεία μέσω του ποντικιού.

Απο την άλλη μεριά, αυτή η παραλληλία μπορεί να επιφέρει και κάποια προβλήματα, γιατί οι εντολές μπορεί να συμπίπτουν χρονικά και να γόνονται έτσι διφορούμενες. Φαινόμενα όπως αποσυγχρονισμένη φωνή και χειρονομία, επανάληψη κτλ αυξάνουν με την αύξηση της ταχύτητας εκτέλεσης. Πάντως, αυτό που συνηθίζεται *δεν* είναι η χρήση ενός τρόπου, όπως ίσως ήταν αναμενόμενο. Αντιθέτως, οι χρήστες έχουν την τάση να εξιδεικεύουν τη χρήση των διαφόρων τρόπων Για παράδειγμα, η φωνή χρησιμοποιείται, για επανάληψη ή για εντολές που δεν απαιτούν να κοιτάς την οθόνη ή επακριβή τοποθέτηση, και η χειρονομία χρησιμοποιείται για στοιχειώδεις και αποκριτικού τύπου δράσεις και επίσης για να επιτευχθεί άμεση ανάδραση. Αυτό δείχνει ότι οι δυο τρόποι συνυπάρχουν και συμπληρώνουν ο ένας τον άλλον.

Το μονοπάτι προσφέρεται για και για άλλες πιο συστηματικές εκτιμήσεις οι οποίες ήδη δείχνουν να δικαιολογούν περαιτέρω εξερεύνηση του πολυτροπικού παραδείγματος χρησιμοποιώντας ΔΑΜ που είναι όλο και πιο ρεαλιστικές.

Συνόψη

Μια ΔΑΜ είναι μια ένωση μεταξύ δύο δομών. Από τη μία, υπάρχουν τα διάφορα επίπεδα ενός κόσμου αναφορών (νοημάτων), και από την άλλη, υπάρχουν τα διάφορα επίπεδα της αφηρημένης αρχιτεκτονικής του λογισμικού. Αλλάζοντας επίπεδα (μεταξύ αναπαραστάσεων, ιδεών και συμβόλων) μπορεί να μοντελοποιηθεί σαν μια διαδικασία σε δύο άξονες : τον συνταγματικό άξονα (ο οριζόντιος άξονας ορισμένος ως ο συνδυασμός νοημάτων σαν συνάρτηση του χρόνου) ο οποίος χρησιμοποιεί "διάλογο", και τον παραδειγματικό άξονα (οι συνδυασμοί νοημάτων στον κάθετο άξονα) ο οποίος χρησιμοποιεί "έλεγχο". Η "Αλληλεπίδραση" εισάγεται με μια πιο άμεση σχέση στο υλικό σύστημα. Αυτό σημαίνει ότι κατά την αλληλεπίδραση, τόσο ο αριθμός των συνταγματικών συνδυασμών όσο και το βάθος των κόσμων είναι περιορισμένα σε σχέση

με το διάλογο. Επίσης μια ΔΑΜ είνα ένας συνδετικός κρίκος μεταξύ του ανθρώπινου περιβάλλοντος, του περιβάλλοντος της μηχανής και του κοινού τους περιβάλλοντος. Αυτό σημαίνει ότι μια ΔΑΜ δρα ως αιχμαλωτίζων, επιδρών, (πολυτροπικού) καθρέφτη της μηχανής, (διατροπικού) καθρέφτη του χρήστη και (μετα-τροπικού) σημείου συνάντησης και των δύο.

Ο σχεδιαστής μιας ΔΑΜ πρέπει να λάβει υπ'όψη τα αισθητικά και κινητικά (μηχανικά) χαρακτηριστικά του χρήστη, κυρίως κατά το στάδιο της διαχείρησης τρόπων. Η συγχώνευση και ο διαχωρισμός της πληροφορίας εισάγει τα τυπικά προβλήματα της πολυτροπικής ΔΑΜ.

Αρκετές πτυχές της πολυτροπικότητας, δεν συζητήθηκαν σε αυτή τη δημοσίευση. Μία από αυτές αφορά σε θέματα εκτίμησης, τα οποία είναι τόσο βασικά όσο τα θέματα σχεδίασης για τις πραγματικές εφαρμογές. Αυτά τα ζητήματα εξάρουν τη σημασία της ταυτοποίησης διαφόρων τύπων σφαλμάτων κατανόησης (επικοινωνίας). Αυτά τα σφάλματα δεν οφείλονται μόνο στη χαμηλή απόδοση που προσφέρουν τα συστήματα αναγνώρισης ομιλίας, αλλά επίσης σε πιο βασικά φαινόμενα που σχετίζονται με κινητική πρόβλεψη/συμφωνία έναντι σε καθυστέρηση/δισταγμό, σύγκρουση μεταξύ διαφορετικών τρόπων και απρόσμενων ενεργειών. Οπωσδήποτε απαιτούνται περισσότερες μελέτες σε αυτά τα θέματα.

Περίληψη

Υπάρχουν αρκετές περιπτώσεις όπου ένας υπολογιστής βοηθά τις ανθρώπινες δραστηριότητες και επικοινωνία. Η μηχανή μπορεί να δράσει ως μεσολαβητής, ως γεννήτορας μιας εικονικής πραγματικότητας ή ως συνεργάτης. Η κατασκευή μιας πολυτροπικής διεπαφής, δηλαδή ενός μέσου που χρησιμοποιεί περισσότερους από έναν τρόπο επικοινωνίας μεταξύ ανθρώπου και μηχανής, γίνεται λαμβάνοντας υπόψη την καταλληλότητα του τρόπου, τις στρατηγικές και την ταχύτητα αλληλεπίδρασης εφαρμοσμένα στις ιδιαιτερότητες του χρήστη. Οι τρόποι αλληλεπίδρασης είναι ο τρόπος ομιλίας, ο γραπτός, η χειρονομία και ο οπτικός τρόπος. Στο κείμενο εξετάζεται η συγχώνευση ομιλίας και χειρονομίας. Η τελευταία, χαρακτηρίζεται από τρεις λειτουργίες: την σημειωτική (επικοινωνιακή), την επιστημική (αντιληπτική) και την εργοτική. Το πλαίσιο αλληλεπίδρασης ορίζεται από τη χρήση τρόπων, δηλαδή το αν η ερμηνεία του τρόπου είναι σειριακή ή παράλληλη και την εξάρτηση πληροφορίας, δηλαδή το αν η πληροφορία εξαρτάται από τα κανάλια εισόδου-εξόδου. Έτσι ορίζονται τρία πλαίσια αλληλεπίδρασης το εναλλασόμενο, το συνεργητικό και το ταυτόχρονο. Ένα πολυτροπικό σύστημα περιλαμβάνει τα κανάλια εισόδου-εξόδου (μέσα), τους servers οι οποίοι μεταφέρουν τα πολυτροπικά μηνύματα από τα κανάλια εισόδου (ή προς τα κανάλια εξόδου). Η κύρια λειτουργία του πολυτροπικού συστήματος είναι η συγχώνευση της (διτροπικής) πληροφορίας στην είσοδο. Αντίστοιχα γίνεται και ο διαχωρισμός της στην έξοδο του συστήματος. Η διτροπική αναφορά σε αντικείμενα μπορεί να γίνει: ταυτόχρονα αν αναφερόμαστε σε ομάδα αντικειμένων, σε διαφορετικά χρονικά πλαίσια αν αναφερόμαστε σε ακολουθία αντικειμένων, ταυτόχρονα αλλά με προσοχή για κινούμενο αντικείμενο και ταυτόχρονα (όσο διαρκεί η χειρονομία) αν η αναφορά είναι άμεση. Τα επίπεδα της συγχώνευσης είναι (από τα χαμηλότερα προς τα υψηλότερα) η μορφολογία και η σύνταξη των σημειωτικών μονάδων, τα πραγματολογικά 1ου και 2ου βαθμού, η σημασιολογία, και τα πραγματολογικά 3ου βαθμού (αφορούν) στις δράσεις της μηχανής. Το ICPDraw είναι ένα παράδειγμα πολυτροπικής διεπαφής. Προσφέρει τη δυνατότητα επικοινωνίας με ομιλία και χειρονομία (με το ποντίκι). Χρησιμοποιεί μια κατανεμημένη αρχιτεκτονική λογισμικού που περιλαμβάνει servers ομιλίας που συλλαμβάνουν το σήμα και δύο client διαδικασίες (με μικρόφωνο για αναγνώριση και

ακουστικά για τη σύνθεση). Η εφαρμογή χρησιμοποιεί συγκεκριμένο συντακτικό και γραμματική για τους δύο τρόπους.

ΛΕΞΙΚΟ ΑΓΓΛΙΚΩΝ ΟΡΩΝ ΚΑΙ ΣΥΝΤΜΗΣΕΩΝ

Act	Ενέργεια, πράξη	Ακολουθία μονάδων που εκπέμπονται ή λαμβάνονται από το χρήστη
Action	Δράση	Λειτουργία της μηχανής
Alternate	Εναλλασόμενο	Γενικό πλαίσιο αλληλεπίδρασης, κατά το οποίο μπορεί να υπάρξει συν-αναφορά μεταξύ των μονάδων, και υπάρχουν χρονικοί περιορισμοί
Bimodal	Διτροπικός	
Concurrent	Ταυτόχρονο, σύγχρονο	Γενικό πλαίσιο αλληλεπίδρασης, κατά το οποίο δεν υπάρχει συν-αναφορά μεταξύ των μονάδων και δεν υπάρχουν χρονικοί περιορισμοί
Deictics	Δείκτες	
Domain	Περιοχή	Καθορισμένη περιοχή που ανήκει κάποιο αντικείμενο
Epistemic	Επιστημική	Λειτουργία κατά την οποία αποκτάται γνώση για το περιβάλλον, συνήθως μέσω της αφής
Ergotic	Εργοτική	Λειτουργία της χειρονομίας κατά την οποία μεταφέρεται ύλη ή ενέργεια στο περιβάλλον
Event	Γεγονός	Σήμα που στέλνεται προς τη μηχανή
Fission	Διαχωρισμός	Η διαδικασία διαχωρισμού της πληροφορίας κατά την έξοδο
Fusion	Συγχώνευση	Η διαδικασία αφομοίωσης της πληροφορίας κατά την είσοδο
Gesture	Χειρονομία	
Information dependance	Εξάρτηση πληροφορίας	
Intermodal	Διατροπικός	
Marker	Μαρκαριστής	
Media	Μέσο	Συσκευή εισόδου-εξόδου
Metamodal	Μετατροπικός	
Mode	Τρόπος	Ο τρόπος αλληλεπίδρασης με τον υπολογιστή
Morphosyntax	Μορφοσυντακτικό	
Multimodality	Πολυτροπικότητα	Η χρήση περισσότερων από ένα τρόπων αλληλεπίδρασης με τον υπολογιστή
Pragmatics	Πραγματολογία	
Prosody	Προσωδία	Μουσικός τόνος διάρκεια και ένταση του ήχου
Semantics	Σημασιολογία	
Semiotic	Σημειωτική	Λειτουργία η οποία είναι από μόνη της επικοινωνιακή (πχ η γλώσσα των κωφών) ή υποδυκνείει, ορίζει, δίνει ρυθμό κτλ.
Stimuli	Ερεθίσματα	

Synergistic	Συνεργητικό	Γενικό πλαίσιο αλληλεπίδρασης κατά το οποίο μπορεί να υπάρξει συν-αναφορά και δεν υπάρχουν χρονικοί περιορισμοί
Syntax	Συντακτικό	
Trajectory	Τροχιά	Πορεία που δείχνει τη θέση που θα τοποθετηθεί κάποιο αντικείμενο
Unit	Μονάδα	Μονάδα πληροφορίας που μεταφέρει σημασία (διαφορετική για τη μηχανή και για το χρήστη)
CMR	ΑΚΣ	Αναπαράσταση κοινής σημασίας
GS	ΦΧ	Φράση χειρονομίας
HCI	ΔΑΥ	Διεπαφή ανθρώπου-υπολογιστή
NG	ΟΟ	Ομάδα ουσιαστικών
PG	ΟΠ	Ομάδα προθέσεων
SP	ΦΟ	Φράση ομιλίας
VG	ΟΡ	Ομάδα ρημάτων

Ομιλούντα πρόσωπα και αναγνωριστές ομιλίας που μπορούν να δουν: η επεξεργασία σημάτων εικόνας ομιλίας με υπολογιστή

N.Michael Brooke

Ç ãíùóôÞ óçìáóßá των ορατών χαρακτηριστικών όôçí áíèñþðένς åðéêïéíùíßá ìå ïìéëßá, äåß÷íáé üôé ìðïñåß íá åêìåôáëìåõôåß êáé óå óõóôÞìáôá âáóéóìÝíá óå õðïëïãéóôÝò. Áðü ôç ìéá ìåñéÜ, ç áíÜðôυîç ôης διτροπικής, "οπτικοακουστικής αναγνώρισης ομιλίας" èá ìðïñïýóå íá âåëôéþóåé ôçí áðüäïóç ôçò áíáãíþñéóçò ïìéëßáò, åéäéêÜ óå èïñôâþäç ðåñéâÜëëïíôá. Áðü ôçí Üëëç ìåñéÜ, ç åöáñìïãÞ ôης "σύνθεσης εικόνας ομιλίας", Þ ãñáöéêþí áðü ððïëïãéóôÝò ðïõ åðéäειêνύïõí áðïìéìÞóåéò ïìéëïýíôùí ðñïóþðùí, èá ìðïñïýóáí ãéá ðáñÜäåéãìá, να παρέχουν οπτικά ερεθίσματα ãéá íá ìðïñÝóåé íá äéäá÷ôåß êáé íá ðñïóäéïñéóôåß ç éêáíüôçôá äéáâÜóìáôïò ôçò ïìéëßáò ðñïò üöåëïò áõôþí ðïõ Ý÷ïõí êáôåóôñáìÝíç áêïÞ. Ç áíÜðôõîç της οπτικοακουστικής αναγνώρισης και της σύνθεσης εικόνας ομιλίας Ý÷åé äéáóõíäèåß ìå : ôá ðëåïíåêÞìáôá ôçò ôå÷íïëïãßáò åðåíåñãáóßáò åéêüíáò, ôçí Üöéîç óýã÷ñïíùí õðïëïãéóôéêþí ôå÷íéêþí üðùò íåõñùíéêÜ äßêôõá êáé στοχαστικά μοντέλα, και της ισχύος των μοντέρνων επεξεργαστών γραφικών. ÌåñéêÜ áðü áõôÜ ôá óçìáíôéêÜ âÞìáôá óå áõôÞ ôçí åîåëéêôéêÞ äéåñãáóßá θα υπογραμμιστούν, êáôáëÞãïíôáò στην κατάσταση όπως αυτή έχει σήμερα. Åðßóçò èá óõæçôçèåß η Üðïøς για τς óýíèåóς êáé τçν áíáãíþñéός, ότι αποτελούν äýï äéáöïñåôéêÝò ðëåõñÝò ôïõ ßäéïõ íïìßóìáôïò.

Ορολογία: Επεξεργασία εικόνας προσώπου, κινούμενες εικόνες με γραφικά υπολογιστή, σύνθεση εικόνας ομιλίας, οπτικοακουστική αναγνώριση ομιλίας, εξαγωγή χαρακτηριστικών του προσώπου, data-driven τεχνικές, ανάλυση σε πρωτεύουσες συντεταγμένες, κρυφά μοντέλα Markov.

1. Åéóáãùãþ

Εßíáé ãíùóôü åäþ êáé êáéñü üôé παρακολουθώντας ôï ðñüóùðï åíüò áíèñþïõ ðïõ ìéëÜåé, üðùò êáé áêïýãïíôáò ôç öùíÞ ôïõ ìðïñåß íá âåêôéþóåé την κατανόηση της ομιλίας, åéäéêÜ üðïõ õðÜñ÷åé èüñõâïò [Erber (1975)]. Ç âëÜâ ôçò áêïÞò åßíáé åßíáé ìéá åéäéêÞ ðåñßðôùóç èïñõâþäïôò ïìéëßáò óôçí ïðïßá ç áêïõóôéêÞ åßóïäïò åßíáé õðïâéâáóìÝíç, óéåöèåßðå ôçí óðáíßùò ïëïêëçñùôéêÞ êáôáóôñïöÞ áðü ôçí âëÜâ. Áêùíá êáé ãéá ôçí êáíïíéêÞ áêïÞ, ôï íá êÜíåéò ÷ñÞóç ìç áêïõóôéêþí χαρακτηριστικών ôá ïðïßá óõíïäåýïõí ôçí ðáñáãùãÞ ïìéëßáò, ìðïñεί íá âåëôéþσει ðïëý ôçí

åðéêïéíùíßá ìå ïìéëßá óε êáèçìåñéíÝò êáôáóôÜóåéò. Γéá ôçí ðåñßðôùóç ôçò âëÜâçò ôçò áêïÞò, åðéêåíôñþíåôáé óôçí éêáíüôçôá åðéêïéíùíßáò óôçí êáèçìåñéíÞ ôïõò æùÞ. Ïé κινήσεις ôïõ óþìáôïò üðùò ôï êïýíçìá êáé ç êßíçóç ôïõ êåöáëéïý, êáèþò êáé εκφράσεις ôïõ ðñïóþðïõ üðùò ïé êéíÞóåéò ôùí öñõäéþí [Dittmann (1972), Ekman (1979)] ìðïñïýν íá áðïäþóïõí óçìáíôéêά χαρακτηριστικά ομιλίας, ðáñüëï ðïõ äåí åìðëÝêïíôáé Üìåóá ìå åõêñéíÞò Þ÷ïõò ïìéëßáò. ÐÜíôùò, åßíáé ïé ïñáôÝò αρθρωτικές êéíÞóåéò ôïõ ðñïóþðïõ αυτές οι οποίες óõíäÝïíôáé ìå ôç ñýèìóç ôçò öùíçôéêÞò Ýκôáóçò οι ïðïßες ãåíéêÜ Ý÷ïõí êõñßùò ó÷Ýóç ìå ôá ÷åßëéá-, Þ, ðéï óùóÜ, ìå ôç äéáäéêáóßá äéáâÜóìáôïò ôçò ïìéëßáò. Ðáñüëï üôé ï ôåëåõôáßïò ôñüðïò ðñïôéìÜôáé áêñéâþò åðåéäÞ ÷áñáêôçñéóôéêÜ äéáöïñåôéêÜ áðü ôá ÷åßëç ìðïñïýí íá åìðëÝêïíôáé, ðïëëÝò áðü ôéò óýã÷ñïíåò ìåëÝôåò, ðåñéÝ÷ïõí ðïëëÜ óôïé÷åßá áðü áõôÜ ðïõ èá ðåñéãñáöèïýí óå áõôü ôï Üñèñï, ôá ïðïßá áöïñþíôáé êéò ïñáôÝò κινήσεις των αρθρωτών ôïõ ðñïóþðïõ êáé êõñßùò áõôÝò ãýñù áðü ôç óôïìáôéêÞ ðåñéï÷Þ.

ÌÝñïò áðü ôï üöåëïò, ðñïêýðôåé áðü ôï üôé ç ÷ñÞóç οπτικών χαρακτηριστικών ομιλίας óõìðëçñώνει τα ακουστικά χαρακτηριστικά êáé õðÜñ÷ïõí êáëïß ëüãïé ãéá áõôü. Ãéá ðáñÜäéãìá, ακουστικά χαρακτηριστικά κατά την άρθρωση óõìöþíùí, ôåßíïõí íá óõó÷åôéóôïýí ìå ÷áìçëÞò Ýíôáóçò, ìéêñÞò äéÜñêåéáò, êáé ìÜëëïí πολύ ëåðôïìåñή ÷áñáêôçñéóôéêÜ ôçò ìÝóçò ðñïò õøçëÞς ðåñéï÷Þς óõ÷íïÞôùí ôïõ öÜóìáôïò, τα οποία åßíáé éäéáéôÝñùò åõáßóèçôá óôéò ðáñåìâïëÝò ôïõ èïñýâïõ. Ôçí ßäéá þñá, ôï áêïõóôéêü ìÝñïò óõ÷íÜ áðïêáëýðôåôáé êáèáñÜ áðü áðü ôéò ευδιάκριτες κινήσεις ôùí ÷åéëéþí, ôùí äïíôéþí êáé ôçò ãëþóáò. Ôá ìç ïñáôÜ áêïõóôéêÜ ôìÞìáôá üðùò ï ëÜñõããáò êáé η σταμφύλη, äéÝðïõí ôï ðáñïõóéáóôéêü ôçò ðáñáâùãÞò ôçò ïìéëßáò, üðùò ç åíñÑéíüçôá, ï ôïíéóìüò êáé ç Ýíôáóç, ôåßíïõí íá åðéâÜëëïõí áñãÜ ìåôáâáëëùìåíåò êáé Ýíôïíåò áëëáãÝò óôï ÷áìçëÞò óõ÷íüçôáò ìÝñïò ôïõ áêïõóôéêïý öÜóìáôïò. ÁõôÜ åßíáé áíèåêôéêÜ óôçí ðáñåìâïëÞ ôïõ èïñýâïõ [Summerfield (1987), Summerfield et al. (1989)]. Ôï üöåëïò ôï íá âëÝðåéò ôï ðñüóùðï ôïõ ïìéëéôÞ ìðïñåß íá åßíáé óçìáíôéêü êáé Ý÷åé ððïëïãéóôåß üôé éóïäõíáìåß ìå ìéá áýîçóç ôùí 8-10 dB óôïí ëüãï óÞìáôïò ðñïò èüñõâï üôáí ïé ïìéëïýìåíåò ðñïÜóåéò ðáñïõóéÜæïíôáé óå èÝìáôá ìå èïñõâþäåò õðüâáèñï. Áõôü éóïäõíáìåß ìå ìéá áëëáãÞ στο ποσοστό αναγνώρισης λέξεων, áðü ðåñßðïõ 20% óå ðåñßðïõ 80% åÜí ï ëüãïò óÞìáôïò ðñïò èüñõâï ôçò ïìéëßáò ìå èüñõâï, åßíáé -6dB [MacLeod & Summerfield (1987)].

Ãéá ôïõò ëüãïõò ðïõ ðåñéëÞöèçόáí ðéï ðÜíù, η εκμετάλευση ορατών σημάτων ομιλίας και των χαρακτηριστικών που περιέχουν είναι αναμενόμενα να βελτιώσουν την κατανόηση της ομιλίας, εéäéêÜ üðïõ õðÜñ÷åé èüñõâïò. ÕðÜñ÷åé Ýíáò áñéèìüò áðü äõíáôÝò ðåñéï÷Ýò åöáñìïãþí ãéá óõóôÞìáôá áõôùìáôçò ïìéëßáò, ôá ïðßá ìðïñïýí åßôå íá áíáëýóïõí åßôå íá óõíèÝóïõí τους ορατούς αρθρωτές ομιλίας. ÐÜíôùò, äýï åéäéêÜ ðáñáäåßãìáôá èá áñêÝóïõí åäþ ãéá íá ðáñïõóéÜóïõí ôéò ãåíéêÝò áñ÷Ýò της åνίσχυσης της επεξεργασίας ôçò ïìéëßáò, ðåñélαμβάνοντας ένα οπτικό

óôïé÷åßï. Ðñþôïí, õðÜñ÷ïõí ðåñéâÜëëïíôá üðùò ôá ÷áéñéóôçñéá ôùí áåñïóôáöþí üðïõ ï Ýëåã÷ïò êáé ôá üñãáíá åßíáé ôüóï ðïëýðëïêá, þóôå ï ÷áéñïêßíçôïò ÷áéñéóìüò åßíáé áíåðáñêÞò êáé ç áëëçëåðßäñáóç ôçò öùíÞò åßíáé áðáñáßôçôç. ÓõíÞèùò ïé áíáãíùñéóôÝò áêïõóôéêÞò ïìéßáò äåí áðïäßäïõí êáëÜ óôá åðßðåäá õøçëïý èïñýâïõ áõôïý ôïõ ðåñéâÜëëïíôïò. ÐÜíôùò, åÜí óôçí åßóïäï ôïõò ìðïñïýóå íá ìðåé åðéðñüóèåôá Ýíá êáíÜëé åéêüíáò, ç áðüäïóç ôïõò èá ìðïñïýóå íá åðáõîçèåß. Ôå÷íéêÝò óõíäéáëåãþò ðáñïõóßá èïñõâùäþí áêïõóôéêþí óçìÜôùí ðØÜñ÷ïõí, áëëÜ óõ÷íÜ åßíáé ìåñéêþò ìüíï áðïôåëåóìáôéêÝò êáé áêñéâÝò óôï íá ðõïëïãéóôïýí [Mellor & Vagra (1993)]. Η οπτική επεξεργασία ðñïóöÝñåé ðñüóèåôç áíïóá óôï èüñõâïìå ôï äõíáôüí ÷áìçëüôåñï êüóôïò. Äåýôåñïí, η εκτίμηση êáé ç âåëôßùóç ôùí éêáíïôÞôùí “äéáâÜóìáôïò” ôçò ïìéëßáò áõôþí ðïõ Ý÷ïõí âëÜç óôçí áêïÞ, åßíáι πολύ æùôéêü ãéá ôçí áðïêáôÜóôáóç ôïõò. `Ενας συνθέτης εικόνας ομιλίας, δηλαδή ένα πακέτο γραφικών κινουμένων εικόνων, που θα μπορούσε να εξομοιώσει ρεαλιστικά τις ορατές αρθρώσεις του ðñïóþðïõ åíüò ïìéëçôÞ, èá Þôáí äõíáôüí íá åîáóöáëßóåé ìéá ìåãÜëç åìâÝëåéáò åëåãîéìüôçôá êáé óõãêåêñéìÝíï ïðôéêü åñÝèéóìá ãéá óêïðïýò åêôßìçóçò êáé åêðáßäåõóç. Êáôáãñáöåßò ïìéëçôþí ÷ñçóéìïðïéïýíôáé óõ÷íÜ, áëëÜ åßíáé ëéãüôåñï åý÷ñçóôïé åðåéäÞ ðñÝðåé íá êáèïñéóôïýí êáé íá ðñïöáéñéóôïýí áðü ðñéí. Åðßçò, äåí ìðïñïýí íá ïñéóèïýí ðëÞñùò êáé Ýôóé åßíáé äýóêïëï íá åëÝãîåéò ìå áêñßâåéá ìåôáâïëÝò åßôå ôων συνθηκών απεικόνισης åßôå ôùí ïïñöáóìþí ôïõ ïìéëçôÞ. `Η σύίèåόç ìå ôá÷ýôçôá êáé áêñßâåéá, ìðïñåß åðßçò íá ðñïóöÝñåé ôçí ðéèáíüôçôá óõóôçìÜôùí áõôüìáôçò áëëçëåðßäñáóçò ãéá áíáãíþóôåò ïìéëßáò, óôçí ïðïßá το ερέθισμα μπορεί να εξελιχθεί σε σχέση με ôçν απόδοση ôùí èåìÜôùí óå συγκεκριμένους ôýðïõò õëéêού.

Ç åêìåôÜëåõóç ïðôéêþí óçìÜôùí óå áõôüìáôá óõóôÞìáôá, ðáñïõóéÜæåé äýï êύρια ðñïâëÞìáôá. ÁõôÜ åßíáé üôé:

á) Åìðëýêïõí ìåãÜëο όγκο äåäïìÝíùí. Ãéá ðáñÜäéãìá, áêüìá êáé åÜí ïé áéêüíåò ôçò óôïìáôéêÞò ðåñéï÷Þò åíüò ïìéëçôÞ ðáñáãóíôïõóáí ìïíï÷ñùìáôéêÜ ìå 256 åðßðåäá ôïõ ãêñßæïõ, êáé óå ìéá ìåôñéïðáèÞ ÷ùñéêÞ ανάëυός ôùí 100 x100 óçìåßùí, ôüôå èá åß÷áìå 50 áéêüíåò ôï äåõôåñüëåðôï (ο ρυθμός με τον οποίο καταγράφονται πεδία εικόνας, υπάρχουν δύο πεδία ανά πλαίσιο σε μία τηλεοπτική εικόνα), áõôü èá ðáñÞãáãå ìéóü âéáôìÝñεï bytes äåäïìÝíùí ãéá êÜèå äåõôåñüëåðôï ïìéëßáò. Åö'üóïí ôá åýñç æþíçò ôùí åðéêïíùíéáêþí óõóôçìÜôùí üðùò ôá äçìüóéá ôçëåòùíéêÜ äßêôõá äåí õðïóôçñßæïõí ôüóï õøçëÜ åðßðåäá äåäïìÝíùí, äåí åßíáé äõíáôüí, ãéá ðáñÜäéãìá, íá êáôáóêåõáóôåß Ýíá “âéíôåüòùíï” áðëÜ ìåôáöÝñïíôáò áð'åõèåßáò ôá áêïõóôéêÜ êáé ôá ïðôéêÜ äåäïìÝíá ðáñÜëëçëá. ÊÜðïéåò ìïñöÝò óõìðåóçò êáé êùäéêïðïßçóçò äåäïìÝíùí, êõñßùò ôùí ïðôéêþí äåäïìÝíùí, åßíáé áðáñáßôçôåò. Ìéá åéäéêÞ ïïñöÞ êùäéêïðïßçóçò åßíáé ôïõ ôýðïõ üðïõ èá ðáñÞãáãå Ýëåã÷ï äåäïìÝíùí ãéá êáèïäÞãçóç åíüò ìïíôÝëïõ üðïõ èá óõíèÝôåé ôçí åéêüíá ôïõ ðñïýðïõ ôïõ ïìéëçôÞ, óôï äÝêôç. Áí êáé ôï âéíôåüòùíï åßíáé Ýôóé ìéá äõíáôÞ åöáñìïãÞ ãéá

σύνθεση εικόνας ομιλίας, óýã÷ñïíá ó÷Þìáôá êùäéêïðïßçóçò åéêüíáò σήμερα, åðéôñÝðïõí ôçí áðïôåëåóìáôéêÞ, óå ðñáãìáôéêü ÷ñüíï ìåôáöïñÜ ôùí äåäïìÝíùí ôçò åéêüíáò óå ÷áìçëά åýñη ζώνης. Ìå óêïðü ôç äéáôÞñçóç ÷áìçëþí åυñών ζώνης, Ýíáò áñéèìüò áðü áõôÜ ôá ó÷Þìáôá åêìåôáëëåýïíôáé ÷ñïíéêÝò ëýóåéò áíôß ãéá ÷ùñéêÝò ëýóåéò [Pearson & Robinson (1985), Frowein et al. (1991)]. Óôçí ðåñßðôùóç ôçò áõôüìáôçò áíáãíþñéóçò εικόνας ïìéëßáò, ïé åéêüíåò ðñÝðåé íá óõëëáìâÜíïíôáé êáé íá åðåîåñãÜæïíôáé óå ÷ñïíéêÝò êëßìáêåò êïíôÜ óôïí ðñáãìáôéêü ÷ñüíï, ôï ïðïßï åðßóçò õðïäçëþíåé êÜðïéá ìïñöÞ ìåßùóçò ôùí äåäïìÝíùí.

Ç êùäéêïðïßçóç êáé ç óõìðåñάóç ôùí äåäïìÝíùí ôùí åéêïíþí ôùí ðñïóþðùí ôùí ïìéëçôþí, ó÷åôßæåôáé ìå ôï èÝìá ôçò åðéëïãÞò ìéáò êáôÜëëçëçò ðåñéãñáöÞò ôùí ïñáôþí ìïñöáóìþí ôïõ ðñïóþðïõ. Áõôü ìðïñåß íá ãßíáé êáôáíïçôü ëáìâÜíïíôáò õð'üøéí ôïí ðáñáëëçëéóìü ôùí ðåñéãñáöþí ôùí óçìÜôùí ïìéëßáò ìå ôην ðáñáäïóéáêή áêïõóôéêή περιοχή, äåßôå ãéá ðáñÜäåéãìá [O'Shaugnessy (1987)]. Óå áõôή ôïí περιο÷ή, åßíáé äõíáôüí íá ðåñéãñáöåß Ýíá óÞìá ïìéëßáò ìå üñïõò ôçò äéáäéêáóßáò παραγωγής, ïñßæïíôáò ôéò öõóéêÝò êείÞóåéò êáé èÝóåéò ôùí áñèñùôþí ðïõ áðïôåëïýí ôï öùíçôéêü óýóôçìá. ÅíáëëáêôéêÜ, ç áêïõóôéêÞ ïìéëßá ìðïñåß íá ðåñéãñáöåß ìå üñïõò ôïõ ÷ñïíéêÜ ìåôáâáëëόμåíïõ öÜóìáôïò éäéïôÞôùí ôïõ óÞìáôïò, üðùò ãéá ðáñÜäåéãìá óôéò ìåèüäïõò áíÜëõóçò êáé óýíèåóçò των φωνοσυντονισμός. Åäþ, ç ðáñïõóßáóç ôçò ïìéëßáò ìå üñïõò èÝóåùí êáé êείÞóåùí φωνοσυντονισμού, ðáñάìÝíåé ðåñéóóüôåñï Þ ëéãüôåñï Ïììåóά óõó÷åôéόìÝíç ìå ôç óõìðåñéöïñÜ ôùí áñèñùôþí. Óå Ýíá ðéï áöçñçìÝíï åßßðåäï, ðáñóëï που ç ðáñïõóßáóç ðñïÝñ÷åôáé áðü ôï áêïõóôéêü óÞìá, ç ó÷Ýóç ìå ôï áêïõóôéêü óýóôçìá åßíáé ðéï ðïëýðëïêç, üðùò óôçí êùäéêïðïßçóç ãñάìéêÞò ðñüâëåøçò, Þ LPC. ÅðéóôñÝöïíôáò ôþñá óôïí ÷þñï åéêïíéêÞò ïìéëßáò, τα οπτικά ÷αρακτηριστικά ìðïñïýí íá óõó÷åôéóôïýí ìå ôçí áíáôïìßá ôïõ êåöáëéïý êáé óôéò äõíáôüçôåò êßíçóçò ðïõ áðïññÝïõí áðü ôçí êáôáóêåõÞ ôïõ, äåßôå ãéá ðáñÜäéãìá [Platt & Badler (1981), Waters & Terzopoulos (1992), Wang & Forsey (1994)]. ÁõôÞ ç ðåñéãñáöÞ ìðïñåß íá ðåñéêëýåé τις μύες. ÅíáëëáêôéêÜ ìðïñïýí íá âáóéόôïýí ðÜíù óôéò êείÞóåéò êáé ôéò ôïðïèåόßåò ôùí åðéöáíåáêâþí óçìåßùí êáé äïìþí ôïõ ðñïóþðïõ ïé ïðïßåò åîáíáãêÜæïíôáé áðü ôéò διακρίνουσες óêåëåôôéêÝò êáé ìõéêÝò äïìÝò, áëëÜ äåί åßíáé ôáõôüóçìåò ìå áõôÝò, ãéá ðáñÜäéãìá [Brooke & Summerfield (1983), LeGoff et al. (1994)]. Óôï ðéï áöçñçìÝíï åðßðåäï, ïé ïïñöáóìïß ôïõ ðñïóþðïõ ìðïñïýí íá ðáñïõóéáóôïýí ìå üñïõò ôùí éäéïôÞôùí ôçò ÷ñïíéêÜ ìατáâáëëüìåíçò óõìðåñéöïñÜò ôùí åéêüíùí ôïõ ðñïóþðïõ, ÷ùñßò íá ëáìâÜíïõí ôð'üøéí ôç äïìÞ και ôá ÷áñáêôçñéóôéêÜ ôùí åéêüíùí Þ ôç ëåðôïìåñÞ ó÷Ýóç ìå ôçí áíáôïìßá ôïõ ðñïóþðïõ [Brooke &Templeton (1990), Brooke & Tomlinson (1994), Brooke & Scott (1994a)]. Ç áîéïðïßçóç ôçò õðïëïãéóôéêÞò ôå÷íïëïãßáò êáé ôùí ôå÷íéêþí, êõñßùò áõôþí ôïõ ÷åéñéóìïý åéêüíáò êáé ôùí ãñáöéêþí, Ý÷åé ìéá äõíáôÞ åðßäñáóç ðÜíù êáé óôçí êùäéêïðïßçóç åéêüíáò êáé óôçí åðéëïãÞ ðåñéãñáöéêÞò ïñãÜíùóçò ãéá ôçí åðåîåñãάσßá εικόνας ïìéëßáò.

â) Åðåéäþ τα οπτικά χαρακτηριστικά ομιλίας óõíÞèùò ÷ñçóéìïðïéïýíôáé ãéá íá ενισχύσουν ôα áêïõóôéêά χαρακτηριστικά, ôï äéÜâáóìá ôçò ïìéëßáò åßíáé μία ôõðéêή διτροπική διαδικασία êáé θέτονται óηìáíôéêÜ æçôÞìáôá ãéá ôçí ïëïêëÞñùóç ôçò εικόνας êáé ôου Þ÷ου ôου ëüãïõ óôçí ó÷åäßáóç êÜèå áõôùìáôïõ ïðôéêï-ακουστικού óõóôÞìáôïò ïìéëßáò. Επιπñüóèåôá, áíáæçôþíôáò Ýíáí õðïëïãéóôéêÜ áðïôåëåóìáôéêü ôñüðï íá ÷åéñéóôïýí ïé äýï τρόποι, ç ïëïêëÞñùóç èá Ýðñåðε íá åßíáé ôÝôïéá þóôå ç êáëýôåñç äõíáôÞ ÷ñÞóç íá ãßíåé êáé áðü ôους äýï τρόπους ìáæß [Robert-Ribes et al. (1993), Robert-Ribes, this volume]. Óôïé÷åßá äåß÷íïí üôé óôïõò áíèñþïïò, ç óõã÷þíåõóç ëáìâÜíåé ÷þñá óå Ýíá åðßðåäï õøçëüôåñï áðü åêåßíï ôïõ ðåñéöåñéáêïýí íåõñéêïýí óõóôÞìáôïò áëëÜ åßíáé ðñéí áðü ôï åðßðåäï ðïõ ñçôÜ ðáßñíåé ôéò áðïöÜóåéò [Summerfield (1987), Summerfield (1991)]. Aõôü åßíáé åðßóçò óýìöùíï ìå ôï fuzzy logic ìïíôÝëï íüçóçò ôïõ Massaro [Massaro, this volume].

Áõôüìáôá óõóôÞìáôá, Ý÷ïõí êÜíåé ðïëëÝò äéáöïñåôéêÝò ðñïóåããßóåéò óôçí διτροπική óõã÷þíåõóç. Ãéá ðáñÜäéãìá, óå Ýíáí πρώιμο áíáãíùñéóôÞ ïìéëßáò [Petajan et al. (1988a)], ç äéåñãáóßá áíáãíþñéóçò åöáñìüóôçêå áíåîÜñôçôá êáé ðáñÜëëçëá óôá áêïõóôéêÜ êáé óôá ïðôéêÜ äåäïìÝíá. Ïé äýï Ýñïäïé ôüôå åîåôÜóôçêáí ÷ñçóéìïðïéþíôáò ìéá ιεραρχική ðñïóÝããéóç, ãéá íá ïñßóïõí ðïý Þôáí áõôÝò óýìöùíåò. Το ποσοστό áíáãíþñéóçò Þôáí õøçëüôåñï áðü áõôïýò ôùí îå÷ùñéóôþí áíáãíùñéóôþí, áëëÜ äåí õðÞñ÷áí óôïé÷åßá ðïõ íá äåß÷íïõí üôé ï âáèüò áíáãíþñéóçò Þôáí ï ìÝãéóôïò. Áíôßèåôá, óôçí video óýíèåóç ïìéëßáò, Ýíáò åëêõóôéêüò ôñüðïò íá ðáñÜåéò äõíáìéêÜ åéêüíåò ðñïóþðïõ èá Þôáí íá ôéò áíôëÞóåéò áðü ôï ßäéï ôï áêïõóôéêü óÞìá ôçò ïìéëßáò. ÐÜíôùò áí êáé åßíáé äõíáôüí íá õðïëïãéóôåß ç áêïõóôéêÞ Ýñïäïò áðü ôï öùíçôéêü óýóôçìá ìå äåäïìÝíç ìéá óõãêåêñéìÝíç äéáìüñöùóç ôùí áñèñùôþí, ôï áíôßóôñïoöï ðñüâëçìá åßíáé ðéï ðïëýðëïêï. Ìå äåäïìÝíï Ýíá óõãêåêñéìÝíï áêïõóôéêü óÞìá, ððÜñ÷ïí ðïëëÝò áñèñùôéêÝò äéáìïñöþåéò ðïõ ìðïñïýí íá õðïëïãéóôïýí ãéá áõôü [Atal et al. (1978)].

2. Êùäéêïðïßçóç, óõìðåóç êáé ðåñéãñáöÞ äåäïìÝíùí åéêüíáò

Ôï ìüíï êéíïýìåíï ïóôü óôï êåöÜëé åßíáé ôï ÷áìçëüôåñï mandible êáé ðáñüëï ôçí áíύψωση êáé ôï ÷áìÞëùìá ôïõ óáãùíéïý, åßíáé Ýíá êýñéï ïñáôü ÷áñáêôçñéóôéêü êáôÜ ôç äéÜñêεéá ôçò ðáñáãùãÞò ôçò ïìéëßáò, üëïé ïé Üëëïé ïñáôïß ìïñöáóìïß åßíáé áðïÝëåóìá ôçò äñÜóçò ôùí ìõþí ðÜíù óôïõò ìáëáêïýò éóôïýò. Το μυικό σύστημα ôïõ ðñïóþðïõ åßíáé ðïëýðëïêï, και σôéò êéíÞóåéò ôùí ÷åëéþí, εμπλέκονται όχι λιγότερες από δεκατρείς ομάδες μυών [Hardcastle (1976)]. Åöüóïí Ýôïéìåò çëåêôñïìïãñáößåò áðïêáëýðôïõí ôï ëåðôïìåñÞ ÷ñïíéóìü ôçò ìõéêÞò äñáóôçñéüôçôáò, ãéá ðáñÜäéãìá [Tuller, Harriw & Keslo (1982)], η áðåõèåßáò ìÝôñçóç ôïõ âáèïýò ôçò ìõéêÞò óýóðáóçò äåí åßíáé ðñáêôéêÞ. Εφόσον η åðåîåñãáóßá ôçò åéêόνας ïìéëßáò âáóßæåôáé óå

áðåõèåßáò ðáñïõóßáóç ôùí áñèñùôéêþí äñáóôçñéïÞôùí êáé ôùí áíáôïìéêþí äïìþí, ðáñáìÝíåé ðñïò ôï ðáñóí áñêåâÜ ìç ðñáêôéêÞ. Ïé Ýñåõíåò ðñÝðåé íá óôñáöïýí óå ðáñáôçÞóåéò ôùí åíåñãåéþí ðïõ ðñïÝñ÷ïíôáé áðü ôçí ìõéêÞ äñáóôçñéüôçôá ôùí ïñáôþí ÷áñáêôçñéóôéêþí.

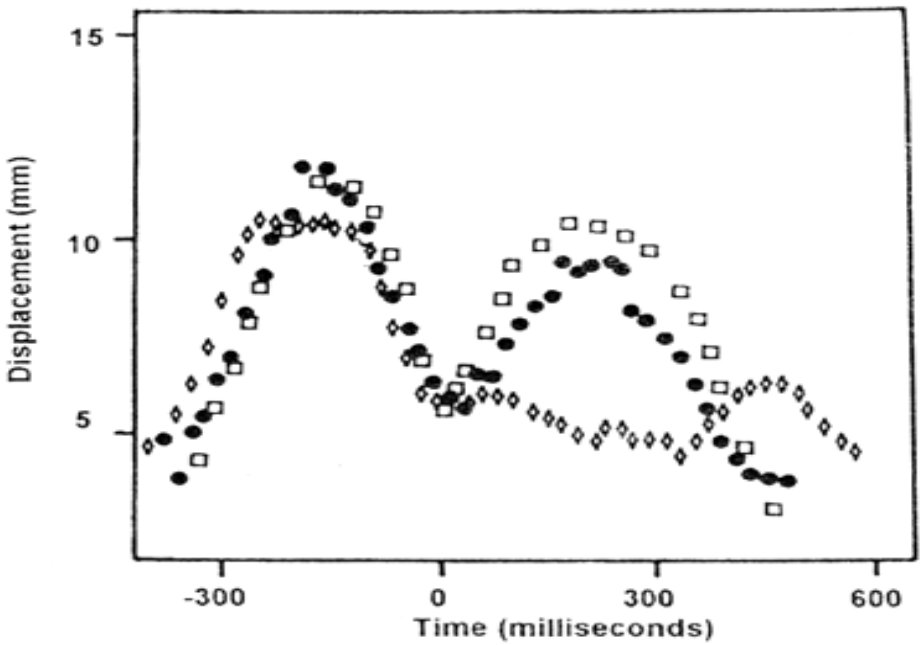
2.1 ÌÝèïäïé åîáãùãÞò ÷áñáêôçñéóôéêþí

Ðáëáéüôåñåò åîåëßîåéò óôçí ìåôá÷åßñçóç ôùí åéêüíùí ôùí ðñïóþðùí ïìéëéôþí Ý÷ïõí åêôåèåß áëëïý [Ârooke (1992c), Brooke &Tomlinson (1994)]. Ìðïñïýí íá ðåñéëçöèïýí ùò áêïëïýèùò. Ç âáóéêÞ ðáñáÞñçóç åßíáé íá ìåéùèåß ο όγκος ôùí äåäïìÝíùí þóôå íá ÷áéñéóôïýí óå Ýíá ðñïçãïýìåíï óôÜáéï, Ýôóé þóôå ìéá åðüìåíç åðåîåñãáóßá íá ãßíáé ãñÞãïñá êáé áðïôåëåóìáôéêÜ, áëëÜ ôçí ßäéá óôéãìÞ íá äéáôçñïýí ôá ïõóéώδη οπτικά χαρακτηριστικά. Ïé ðñïçãïýìåíåò ìÝèïäïé óõíÞèùò ðåñéåëÜìâáíáí ôç óýλëçøç ôùí åéêüíùí ôïõ ðñïóþðïõ åíüò ïìéëçôÞ, ãéá ðáñÜäåéãìá óå öéêì Þ âéíôåïãáóÝôá êáé ôüôå åðåîåñãáæüíôõóáí ôéò åéêüíåò -ìßá êÜèå öïñÜ óôç óåéñÜ- ìå óêïðü íá êáôáãñÜøïõí ôéò ÷ñïíéêÜ ìåôáâáëüìåíåò èÝóéò ôùí êõñßùò áñèñùôþí, óõíÞèùò óôç ïïñöÞ μαρκαρισμένων óçìåßùí ãýñù áðü ôá ÷åßëç êáé ôï óáãþíé ôïõ ïìéëéôÞ. Ìéá ðáëéüôåñç ìåëÝôç ÷ñçóéìïðïéïýóå óçìåßá ìÝôñçóçò ãéá íá åîåñåõíÞóåé ðáñÜìåôñïõò ðïõ èá ìðïñïýóáí íá ëçöèïýí ðð'üøη [Finn (1986)]. Åöüóïí στις ðáëéüôåñåò Ýñåõíåò ìåñéêÝò öïñÝò ÷ñçóéìïðïéïýóáí ÷áéñïêßíçôç êáôáãñáöÞ, ãéá ðáñÜäåéãìá [Brooke & Summerfield (1983)], Þôáí êáé åßíáé áêüìá óýíçèåò íá ÷ñçóéμïðïéïýíται μαρκαρισμένα ÷áñáêôçñéóôéêÜ ôïõ ðñïóþðïõ ãéá áõôüìáôç êáôáãñáöÞ äéáìÝóïõ óåéñþí áðü åéêüíåò, ãéá ðáñÜäåéãìá [Bothe, Lindner &Rieger (1993), Stork, Wolff & Levine (1992)]. Ôáõôü÷ñïíåò êáôáãñáöÝò áðü ôçí ìðñïóôéíÞ êáé ôéò ðëåõñéêÝò ïðôéêÝò ãùíßåò ôïõ ðñïóþðïõ, ìðïñïýí íá ãßíïõí åßôå ÷ñçóéìïðïéþíôáò äýï êÜìåñåò, ìðñïóôÜ êáé óôá ðëÜãéá ôïõ áíôéêεéìÝíïõ, åßôå ÷ñçóéìïðïéþíôáò êáèñÝðôåò ððü ãùíßá, Ýôóé þóôå íá åîáöüáëßæïõí ôñιóäéÜóôáôåò óõíôåôáãìÝíåò ëåðôïìÝñåéåò, ãéá ðáñÜäåéãìá, ôης ðñïåîï÷Þς ôùí ÷åëëéþí.

ÐïëëÝò áðü áõôïý ôïõ åßäïõò ôéò ðñïóåãßóåéò åß÷áí áíáðôõ÷èåß óçìáíôéêÜ. Ìßá ÷ñçóéìïðïßçóå åéäéêïý óêïðïý õëéêü ãéá íá åíôïðßóåé παράθυρο, êáèáñßóåé êáé óõëëÜåé κατώφλι, ìáýñåò êáé Üóðñåò åéêüíåò ôçò óôììáôéêÞò ðåñéï÷Þò óå ðñáãìáôéêü ÷ñüíï [Petajan (1984)]. Mßá Üëëç áðïóðïýóå ôçí ðåñéï÷Þ ôùí ÷åëëéþí êáé ôïõ óôùìáôïìò áðü Ýã÷ñùìåò åéêüíåò, óôéò ïðïßåò ôá ÷åßëç Þôáí ÷ñùìáôéóìÝíá ìðëå êõáíü Ýôóé þóôå íá ìðïñïýáí íá áíáãíùñéóôïýí áõôüìáôá [Benoit et al. (1992)]. ÌåñéêÝò ìÝèïäïé ÷ñçóéìïðïéïýóáí ðáñéïñéóìïýò ãéá íá óôåñåþíôí ôï êåöÜëé ôïõ ïìéëéôÞ Üêáìðôá óå ìßá èÝóç, Üëëåò åðÝôñåðáí ôçí åëåõèåñßá ôïõ ïìéëçôÞ ãéá íá êÜíáé ôéò öõóéêÝò (ãëùóóïïëïãéêÜ ðëçñïöïñéáêÝò) êéíÞóåéò ôïõ êåöáëéýí êáé ôïõ óþìáôïò êáé ôï ìåôáêéíïýóáíá áíáëõôéêÜ [Brooke &Summerfield (1983)]. Ïé ôåëéêÝò Ýſïïäïé áðü üëåò áõôÝò ôéò ôå÷íéêÝò Þôáí ôõðéêÜ ÷ñïíéêÜ ìåôáâáëüìåíá ó÷åéáãñÜìôá ðïõ Ýääé÷íáí, ãéá ðáñÜäåéãìá, ôçí

åîÜñôçóç ôçò ÷ùñéêÞò èÝóçò óçìåßùí êëåéäéþí ôïõ ðñïóþðïõ Þ ôïõ ðëÜôïõò, ýøïõò, ðåñéìÝôñïõ êáé Ýêôáóçò ôçò óôïìáôéêÞò êïéëüôçôáò ê.ο.κ [Brooke &Summerfield (1983), Bothe, Rieger &Tackmann (1993), Petajan et al. (1988a)]. ÁõôÜ ôá ó÷åäéáãñÜìáôá ìðïñýí íá ðáñáôçñçèïýí ïõóéáóôéêÜ óáí óõãêåêñéìÝíç Ýêöñáóç áñèñùôéêþí ôñï÷éþí. `Eíá ðáñÜäåéãìá öáßíåôáé óôï ó÷Þìá 1.

ÏðùóäÞðïôå, ïé ëåðôïìÝñåéò ôùí óõãêåêñéìÝíùí ìåèüäùí ãéá ôçí ðáñáãùãÞ ôïõò, Ý÷ïõí üëåò óáí êïéíü üôé åíÜãïõí áðïôåëåóìáôéêÜ ÷áñáêôçñéóôéêÜ áðü ôéò åéêüíåò êáé ç óõìðßåóç ôùí äåäïìÝíùí áðïêôáôáé áðü ôçí áíÜêôçóç ìüíï ôùí ÷ñïíéêÜ ìåôáâáëëùìåíùí ôéìþí áõôþí ôùí ÷áñáêôçñéóôéêþí óôçí êùäéêïðïéçìÝíç ìïñöÞ. ÐÜíôùò, åöüóoí ìåñéêÜ ÷áñáêôçñéóôéêÜ, üðùò ï ÷ùñéóìüò ôùí ÷åëéêþí êáé ôï ðëÜôïò ôùí ÷åëéêþí, åßíáé ãíùóôü üôé åßíáé óçìáíôéêÜ αντιλήψιμα χαρακτηριστικά [Plant (1980), Montgomery & Jackson (1983)], ç ôáýôιός οπτικών χαρακτηριστικών δεν είναι με κανένα τρόπο θεμελιωμένη αποτελεσματικά. Ðñüòåôá, ìåñéêÜ óçìáíôéêÜ χαρακτηριστικά, üðùò ç όψη ôùí äïíôéþí êáé ôçò ãëþóóáò [McGrath, Summerfield & Brooke (1984)], Þ ç åðéóêßáóç êáé ç õöÞ ôïõ äÝñìáôïò ôïõ ðñïóþðïõ, äåí Ý÷ïõí áêüìá åìöùëéáóôåß óå áðëÝò ðáñáìåôñéêÝò ìïñöÝò. Åðßóçò, ÷áñáêôçñéóôéêÜ üðùò ôá äüíôéá êáé ç ãëþóóá åßíáé ìüíï ìåñéêþò Þ óôéãìéáßá ïñáôÜ, Ýôóé þóôå íá ìçí åßíáé åýêïëç ç åîáóöÜëéóç ïëïêëçñùìÝíùí ìåôñÞóåùí ôùí êéíÞóåùí ôïõò áêüìá êáé áðü áêñéâÝò êáé ðïëýðëïêò ôå÷íéêÝò üðùò áêôßíåò-× cineradiography [Perkell (1969)]. ' Åôóé áõôÝò ïé ìåôñÞóåéò Ý÷ïõí ðåñéïñéóôôåß, áêüìá áí êáé ìðïñýí íá öáíåñþóïõí ÷ñÞóéìá χαρακτηριστικά ομιλίας [Summerfield et al. (1989),McGrath, Summerfield & Brooke (1984)].



Σχήμα 1*: Τροχιές των αρθρωτών, ή γραφικές παραστάσεις των χρονικά μεταβαλλόμενων κινήσεων ειδικών χαρακτηριστικών του προσώπου. Αυτό το γραφικό παράδειγμα δείχνει της μετρημένες κάθετες κινήσεις της γωνίας του εξωτερικού μέρος του χειλιού ενός ομιλητή κατά τη διάρκεια εκφορών των συλλαβών /aba/ (ανοικτά τετράγωνα), /abi/ (μαύροι κύκλοι) και /abu/ (ρόμβοι).*

2.2. ÌÝèïäïé áðåõèåßáò åðåîåñãáóßáò åéêüíáò

Ðéï ðñüöáôá, οι åîåëßîåéò óôï õëéσμικό ôùí õðïëïãéóôþí, êáé ç äéáèåóéìüôçôá øçöéáêþí ðëáéóßùí, Ý÷ïõí êÜíáé ðéï åýêïëç ôçí åðåîåñãáóßá êáé ôçí άμεση ìåôá÷åßñçóç ôùí åéêïíþí. Ãéá ðáñÜäåéãìá, Ýãéíå äõíáôÞ ç áðïèÞêåõóç êáé επανάληψη ακολουθιών áðü øçöéïðïéçìÝíåò ìïíü÷ñωìåò åéêüíåò åíüò ïìéëçôÞ óå ðñáãìáôéêü ÷ñüíï. Εκμετάλευση αυτού, έγινε σε μια πρωτότυπη μελέτη κατά την οποία έγινε χρήση /hVd/ εκφορών πέντε μακρών βρετανικών φωνηέντων, που αντιστοιχούν στις λέξεις 'hard', 'heard', 'heed', 'hoard' και 'who' d' [Brooke & Templeton (1990)]. ØçöéïðïéçìÝíåò ακολουθίες åéêüíùí ôçò óôïìáôéêÞò ðåñéï÷Þò åíüò ïìéëçôÞ åß÷áí áðïèçêåõôåß êáé åðåîåñãáóôåß Ýôóé þóôå íá ìðïñïýóáí íá åêôåèïýí ìå ìåôáâáëëüìåíç ÷ùñéêÜ áíÜëõóç ìåôáîý 256 x 256 óçìåßùí êáé 8 x 8 óçìåßùí êáé ìåôáâáëëüìåíç êëßìáêá ôïõ ãêñßæïõ áíÜëõóç ìåôáîý 256 êáé 2 åðéÝäων ôïõ ãêñßæïõ. Ïé óåéñÝò åß÷áí åêôåèåß óôá 25 πλαίσια ανά äåõôåñüëåðôï ãéá üëåò ôéò êáôáóôÜóåéò. 'Åíá ôõ÷áßï óýíïëï ερεθισμάτων áðü üëåò ôéò êáôáóôÜóåéò åß÷å ðáñïõóéáóôåß óå άτομα üðïõ τους εß÷å æçôçèåß ç áíáãíþιός ôçò

εκφερόμενης ëÝÎçò ðÜíù óôç âÜóç ôùí ïðôéêþí äåäïìÝíùí ìüíï. Ôá áðïôåëÝóìáôá Ýäåéîáí üôé ç ðñïóðÜèåéá áíáãíþñéóçò öùíÞåíôïò ðáñÝìåíå ïõóéáóôéêÜ óôï ßäéï åðßðåäï (ðåñßðïõ 85% óùóôÞ) êÜôù áðü ÷ùñéêÞ áíÜëõóç 32 x 32 óçìåßùí, áëëÜ Üñ÷éóå íá ðéöôåé óå ÷áìçëüôåñåò áíáëýóåéò. Ç áíÜëõóç ôων επιπέδων ôïõ ãêñßæïõ, èá ìðïñïýóå íá ìåéùèåß óå êáèáñü ìáýñï êáé Üóðñï ìå ìéêñÞ åðßäñáóç óôçí ðñïóðÜèåéá áíáãíþñéóçò ôïõ öùíÞåíôïò. Áõôü äåí Þôáí ìç áíáìåíüìåíï áöïý ôï åðßðåäï öùôåéíüôçôáò Þôáí ñõèìéóìÝíï Ýôóé, þóôå ç ðåñéï÷Þ ôçò óôïìáôéêÞò êïéëüôçôáò íá öáßíåôáé ìáýñç êáé ç õðüëïéðç ðåñéï÷Þ óáí Üóðñç óôïí ðåñéïñéóìü ôùí äýï åðéðÝäùí. Ïé åíäåßìåéò áõôïý ôïõ ðåéñÜìáôïò Þôáí ôé ç ÷ùñéêÞ áíÜëõóç ôçò ôÜîçò ôùí 32 x 32 óçìåßùí Þôáí åðáñêåßò ãéá íá åìöùëéÜóåé ðïëëά από τα οπτικά χαρακτηριστικά óôçí áíáãíþñéóç öùíÞåíôïò. ÐÜíôùò, áõôÜ ôá áðïôåëÝóìáôá ÷ñåéÜæïíôáé ðñïóå÷ôéêÞ ìåôá÷åßñióç. Ìüíï Ýíá öùíçôéêü ðëåßóéï äïêéìÜóôçêå. Ôñßá áðü ôá ðÝíôå ìáêñÜ öùíÞåíôá âñßóêïíôáé åðßóçò óôéò ãùíßåò ôïõ öùíçôéêïý ôñéãþíïõ, âëÝðå [Ladefoged (1975)] êáé ôá äýï áêüìá êåíôñéêÜ öùíÞåíôá, ãéá ôá ïðïßá åßíáé ãíùóôü üôé åßíáé ðéï äýóêïëï íá áíáãíùñéóôïýí [McGrath, Summerfield & Brooke (1984)], Þôáí ðéï óõ÷íÜ συνγκεχιμένα. Ç åêôßìçóç åäþ èá ìðïñïýóå íá åßíáé ìüíï Ýíá ðåñéïñéóìÝíï êÜôù öñÜìá.

Ïé åíåëßîåéò ôïõ ëïãéóìéêïý, Ý÷ïõí åðßçòò âåëôéþóåé ôç ìÝôñçóç ôùí áñèñùôéêþí ôñï÷éþí êáé Ýíáò áñéèìüò ðåñßðëïêùí ìåèüäùí åããñáöÞò, ðïõ ðåñéëáìâÜíåé öéáùôÜ Þ ðáñáìïñöùìÝíá ðåñéãñÜììáôá, Ý÷ïõ áíáðôõ÷èåß. Ìðïñïýí íá ÷ñçóéìïðïéçèïýí ãéá áõôüìáôç åããñáöÞ ðåñéãñáììÜôùí, üðùò ôá ðåñéèþñéá ôùí ÷åëéêþí, ãéá ðáñÜäåéãìá [Terzopoulos & Waters (1990), Bregler & Omohundro (1994), Luettin, Thacker & Beet, this volume, Blake, Curwen & Zisserman (1993), Blake & Isard (1994), Dilton, Kaucic &blake, this volume].

2.3 Νευρωνικά δίκτυα και data driven μέθοδοι

Η κατασκευή ισχυρότερων επεξεργαστών επέτρεψε όχι μονο τη διαχείριση δεδομένων ψηφιακής εικόνας, αλλά οδήγησε επίσης σε μία εκ νέου στροφή προς τις μεθόδους παράλληλης επεξεργασίας και άλλων, στατιστικές και data driven τεχνικές επεξεργασίας. Το κύριο πλεονέκτημα αυτών των τεχνικών, είναι ότι δεν χρειάζεται να τροφοδοτηθούν με *α priori* γνώση της λεπτομερής δομής των δεδομένων που χρησιμοποιούν. Η έρευνα στον τομέα της ακουστικής αναγνώρισης, μεταχειρίστηκε επιτυχώς πολλές από αυτές τις

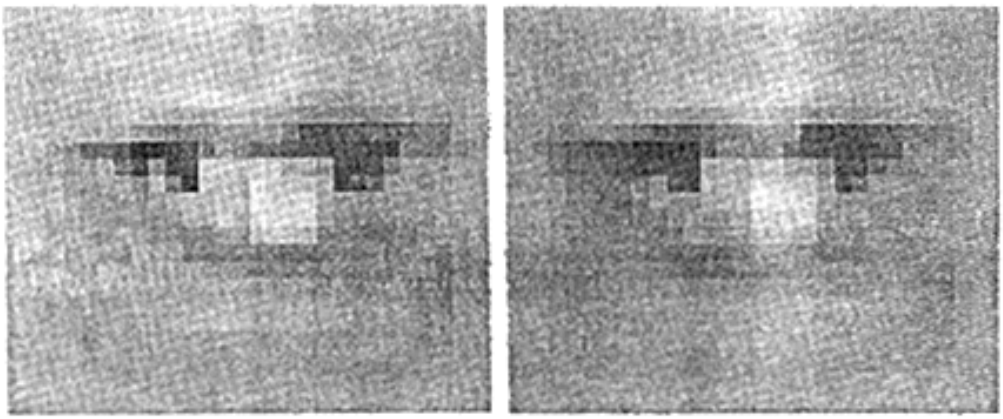
εξελίξεις και η επεξεργασία εικόνας-ήχου μπόρεσε να ωφεληθεί εισάγοντας τις ίδιες τεχνικές-με κατάλληλη προσαρμογή όπου ήταν αναγκαίο. Τα νευρωνικά δίκτυα είναι μηχανισμοί που χαρακτηρίζονται από πίνακες απλών στοιχείων με στενή σχέση μεταξύ τους. Οι ιδιότητες του δικτίου, προσδιορίζονται από τις παραμέτρους των στοιχείων και τις μεταξύ τους διασυνδέσεις. Από τις πολλές αρχιτεκτονικές δικτύων, η πιο διαδεδομένη στον τομέ της επεξεργασίας εικόνας ήχου, είναι αυτή του πέρσεπτρον πολλών επιπέδων (Multi Layer Perceptron, MLP). Στα MLPs, είναι δυνατόν να υπάρχουν ένα ή και περισσότερα επίπεδα 'κρυφών' στοιχείων μεταξύ των στοιχείων εισόδου και εξόδου. Τα δεδομένα εκμάθησης μπορεί να εφαρμοστούν στα στοιχεία εισόδου και υπάρχουν αλγόριθμοι για την εύρεση ενός βέλτιστου αριθμού τιμών των παραμέτρων, για τις οποίες για συγκεκριμένες τιμές εισόδων, το δίκτυο θα δώσει συγκεκριμένες τιμές εξόδων. Τα στοιχεία του δικτίου λειτουργούν παράλληλα και η γνώση που αποκτάται κατά την εκμάθηση, διανέμεται σε όλα μαζί. Τα εκπαιδευμένα δίκτυα χρησιμοποιούνται για την ταξινόμηση εισόδων και μπορεί να ταξινομήσουν πρότυπα εισόδου τα οποία είναι παρόμοια, αλλά όχι ίδια με αυτά των δεδομένων εκμάθησης. Μπορούν επίσης να λειτουργήσουν ακόμα και όταν μερικά από τα στοιχεία είναι ελλατωματικά. Το δίκτυο δεν καταστρέφεται με την αύξηση της δυσκολίας. Ένα πρώτο παράδειγμα χρήσης των MLP στην επεξεργασία ομιλίας, ήταν ένα πείραμα αναγνώρισης φωνηέντων που έγινε χρησιμοποιώντας εικόνες ενός πλαισίου (τμήματος) της πρόσοψης της στοματικής περιοχής ενός ομιλητή, κατά την εκφώνηση φωνηέντων. Χρησιμοποιήθηκαν μονόχρωμες εικόνες στα 64 επίπεδα γκρίζου, ανάλυσης 256x256 σημείων, οι οποίες **(α)** τέθηκαν σε μια μέση ένταση γκρίζου, **(β)** συμπιέστηκαν στα 16x12 σημεία και **(γ)** η αντίθεση αυξήθηκε με γραμμική ανακατανομή των επιπέδων γκρίζου στα μέσα 5/8 των επιπέδων γκρίζου, προς όλο το φάσμα γκρίζου. Τα MLP είχαν 192 στοιχεία εισόδου (ένα για κάθε σημείο), έξι ενδιάμεσα στοιχεία στο κρυφό επίπεδο, και έντεκα στοιχεία στο εξωτερικό επίπεδο, ένα για κάθε ένα από τα μη δίφθογγα φωνήεντα του αγγλοσαξωνικού λεξιλογίου. Για την υπό επιτήρηση εκμάθηση των MLP, χρησιμοποιήθηκαν πλαίσια δειγμάτων εκφώνησης φωνηέντων, στα οποία τα φωνήεντα κατηγοριοποιήθηκαν και οι παράμετροι των MLP επαναπροσδιορίστηκαν με την μέθοδο οπίσθιας διάδοσης. Στη συνέχεια τοποθετήθηκαν δείγματα φωνηέντων του ίδιου ομιλητή ως είσοδοι στο εκπαιδευμένο MLP και παρατηρήθηκε η ταξινόμηση στις

μονάδες εξόδου. Τα αποτελέσματα έδειξαν ότι ακόμα και σε αυτή τη χαμηλή ανάλυση, τα φωνήεντα ταξινομήθηκαν σωστά σε ποσοστό 91% κατά μέσο όρο και 84% στη χειρότερη περίπτωση. Ακόμα και όταν η ανάλυση μειώθηκε στα 4 επίπεδα γκρίζου, τα αντίστοιχα ποσοστά ήταν 87% και 72%. Αυτά τα αποτελέσματα ήταν σε αρμονία με ανάλογα με αυτά πειραμάτων που βασίζονταν στην αντίληψη, αν και δεν ήταν διαθέσιμα δυναμικά χαρακτηριστικά του προσώπου στο MLP. Έδειξαν επίσης ότι ουσιώδη οπτικά χαρακτηριστικά μπορούσαν να αποκτηθούν και σε χαμηλά επίπεδα ανάλυσης. Το γεγονός ότι το κρυφό επίπεδο του MLP είχε μόνο 6 στοιχεία, οι απεικονίσεις οι οποίες προσέφεραν, έδειχναν ειδικά χαρακτηριστικά της ταυτότητας του φωνήεντος, ήταν λίγων διαστάσεων και προσέφεραν μια τεχνική συμπίεσης εικόνας. Αυτό μπορούσε να γίνει με τη χρήση ενός σετ από εικόνες ομιλητή ως είσοδο και υπολογίζοντας αναδρομικά μια απεικόνιση που παράγει εξόδους ταυτόσημες με τις εισόδους, δηλαδή εφαρμόζοντας εκπαίδευση (εκμάθηση) χωρίς επιτήρηση στα MLP για να παράγουν μια απεικόνιση 'ταυτότητας' από τις εικόνες εκπαίδευσης. Όντας εκπαιδευμένες, οι έξοδοι των στοιχείων του κρυφού επιπέδου θα κωδικοποιούσαν μια εικόνα εισόδου η οποία διατηρούσε την μέγιστη διακρισιμότητα μεταξύ των των εικόνων. Παρά το γεγονός ότι έχει χρησιμοποιηθεί επιτυχώς, η κωδικοποίηση εικόνων με τη χρήση των MLP, στην πράξη παρουσιάζει έναν αριθμό δυσκολιών. Ένα από τα κύρια μειονεκτήματα, των MLP είναι ότι οι απεικονίσεις δεν αποκαλύπτουν την φύση των χαρακτηριστικών που περικλείουν οι κινήσεις του προσώπου και λόγω του γεγονότος ότι η πληροφορία διαμοιράζεται σε όλο το δίκτυο, δεν είναι εφικτό να καταταχθούν ανάλογα με τη σημασία τους, τα χαρακτηριστικά των εισόδων εικόνας τα οποία συνεισφέρουν στη εικόνα του προσώπου.

Τα MLP ίσως να μην προσφέρουν σημαντικά πλεονεκτήματα έναντι του εναλλακτικού εργαλείου της ανάλυσης σε πρωτεύουσες συντεταγμένες (Principal Component Analysis, PCA), το οποίο υπερβαίνει μερικά από τα μειονεκτήματα των MLP. Η PCA μπορεί να περιγραφεί ως ένα εργαλείο συμπίεσης εικόνας, ως εξής: Η ένταση κάθε ενός από τα σημεία μιας (n x m)-σημείων μονόχρωμης εικόνας, μπορεί να απεικονισθεί σαν την τιμή κατά μήκος ενός από τους άξονες ενός συστήματος (n x m)-αξόνων. Τότε λοιπόν, μια εικόνα είναι ένα σημείο στον (n x m)-άστατο χώρο. Από ένα σύνολο εικόνων εκμάθησης, η PCA μπορεί να υπολογίσει ένα μετασχηματισμό του συστήματος αξόνων

σε ένα άλλο όπου οι άξονές του, ή οι πρωτεύουσες συντεταγμένες του (principal components), ορίζουν ένα νέο χώρο όπου οι μπορεί να γίνει απεικόνιση. Οι πρωτεύουσες συντεταγμένες κατατάσσονται ανάλογα με την απόκλιση του συνόλου εκμάθησης στο οποίο αναφέρονται. Γι αυτό, η πρώτη συντεταγμένη αφορά στο μεγαλύτερο κλάσμα της απόκλισης, η δεύτερη στο μεγαλύτερο κλάσμα της υπόλοιπης απόκλισης και ούτω καθεξής. Τα σημεία των εικόνων μπορούν να αναπαρασταθούν από τα σημεία στον μετασχηματισμένο χώρο από τις τιμές των πρωτευουσών συντεταγμένων. Αν το μεγαλύτερο μέρος της απόκλισης των δεδομένων αντιπροσωπεύεται από μικρό αριθμό πρωτευουσών συντεταγμένων, τα τελευταία μπορούν ανα χρησιμοποιηθούν για την κωδικοποίηση των εικόνων, προσφέροντας έτσι συμπίεση των δεδομένων. Αυτή η τεχνική χρησιμοποιείται ευρέως σήμερα για μετασχηματισμό της εικόνας, περιλαμβάνοντας κωδικοποίηση ολόκληρου του προσώπου. Η εκτίμηση της απόδοσης των PCA κωδικοποιητών έχει πρόσφατα εξεταστεί με χρήση λεξιλογίου από τριπλέτες ψηφίων (πχ ‘έξι μηδέν τέσσερα’). Έγιναν βιντεοσκοπήσεις του προσώπου ενός ομιλητή ο οποίος πρόφερε τριπλέτες ψηφίων από τις πρότυπες λίστες NATO RCG-10. Το πρόσωπο του ομιλητή τέθηκε σε προκαθορισμένη θέση και προσδιορίστηκε μια συγκεκριμένη περιοχή γύρω από το στόμα. Η επεξεργασία της έγινε με τη μέθοδο PCA και παρήγαγε εικόνες ανάλυσης 32 x 24 σημείων. Διακόσιες τριπλέτες ψηφίων οι οποίες αποτελούσαν συνολικά περίπου 17000 εικόνες, χρησιμοποιήθηκαν ως δεδομένα εκμάθησης και τέθηκαν στην PCA. Δεδομένου ότι οι κινήσεις του στόματος είναι ανατομικώς περιορισμένες, οι εικόνες ομιλίας περιορίζονται σε μικρό χωρικό διάστημα και είναι υψηλής δόμησης. Η συμπίεση των δεδομένων λοιπόν είναι υψηλού βαθμού και βρέθηκε ότι το 82% της απόκλισης του συνόλου των εικόνων μπορούσε να αναπαρασταθεί με χρήση δεκαπέντε πρωτευουσών συντεταγμένων. Στο **σχήμα 3** βλεπουμε τον τρόπο με τον οποίο σχετίζονται η απόκλιση και ο αριθμός των πρωτευουσών συντεταγμένων (ο άξονας της απόκλισης δείχνει την απόκλιση που απομένει αφαιρώντας την πρώτη συντεταγμένη, αντιπροσωπεύοντας τη ‘μέση εικόνα προφοράς’). Αυξάνοντας το μέγεθος του συνόλου εκμάθησης, δεν μειώνεται σημαντικά η απόκλιση για τον ίδιο αριθμό πρωτευουσών συντεταγμένων. Για το δεδομένο λεξιλόγιο που χρησιμοποιήθηκε, αυτό δείχνει ότι ένα μικρό σύνολο εκμάθησης είναι επαρκώς αντιπροσωπευτικό για όλο το εύρος των ορατών χαρακτηριστικών. Οι εικόνες

του συνόλου εκμάθησης και αυτές των 100 τριπλετών ψηφίων που πληρούσαν ένα ξεχωριστό σύνολο ελέγχου, όλες κωδικοποιήθηκαν χρησιμοποιώντας 15 τιμές στοιχείων. Η σύγκριση της τετραγωνικής ρίζας της διαφοράς μέσου τετραγώνου στο επίπεδο γκρίζου (σε όλα τα σημεία) μεταξύ προτύπων εικόνων και αυτών που ανακατασκευάστηκαν από τον PCA κωδικοποιητή των 15 καναλιών, έδειξε μικρή μόνο διαφορά για τις εικόνες ελέγχου και εκμάθησης. Αυτό επιβεβαίωσε την ακρίβεια του μοντέλου κωδικοποίησης για γνωστές και μη γνωστές εικόνες του ίδιου λεξιλογίου. Τα σφάλματα ανακατασκευής, ήταν γενικά διασκορπισμένα σε ξεχωριστά τυχαία σημεία και δεν έδειχναν συστηματικές ασυμφωνίες σε συγκεκριμένες περιοχές των εικόνων, αν και περιοχές όπως οι άκρες των δοντιών προκαλούσαν κάποια τοπικά σφάλματα. Το **σχήμα 2**, απεικονίζει τις διαφορές μεταξύ ενός πρότυπου πλαισίου και της ανακατασκευής του.

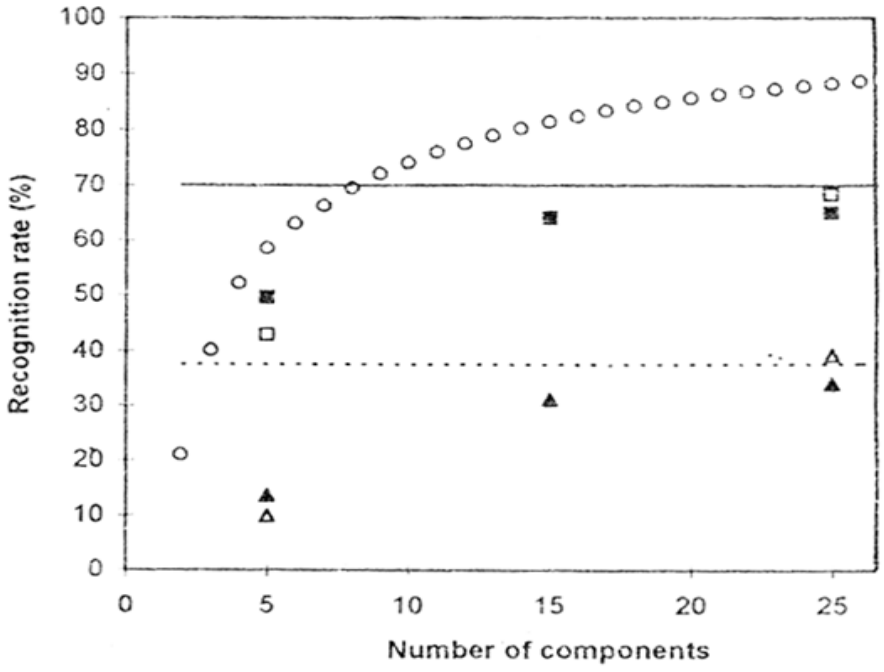


Σχήμα 2: *Ποιότητα εικόνων ανακατασκευασμένες από την PCA. Η αριστερή εικόνα είναι μια πρότυπη προεπεξεργασμένη στα 32x24* σημεία. *Η δεξιά είναι ανακατασκευή της ίδιας εικόνας από PCA, χρησιμοποιώντας τις δεκαπέντε πρώτες πρωτεύουσες συντεταγμένες.*

Ένα κύριο χαρακτηριστικό της PCA είναι ότι είναι σχετικά σταθερή και δεν παρουσιάζει υπερ-ευαισθησίες. Η αλλαγές στις εικόνες, είναι ανάλογες με αυτές στις τιμές του κώδικα και οι χρονικές αποκλίσεις είναι σε συμφωνία με τα επίπεδα κίνησης των αρθρωτών (**σχήμα 2**).

Για τον έλεγχο της αποτελεσματικότητας της PCA κωδικοποίησης, διεξήχθησαν πειράματα οπτικής αναγνώρισης επάνω στις πρότυπες ακολουθίες εικόνων και σε αυτές

που ανακατασκευάστηκαν από PCA, των δειγμάτων ελέγχου και εκμάθησης, με χρήση 5, 15 και 25 πρωτευουσών συντεταγμένων. Τα αποτελέσματα φαίνονται στο **σχήμα 3.**



Σχήμα 3: *Εκτίμηση της PCA κωδικοποίησης για εικόνες 32x24 σημείων από εκφωνήσεις τριπλετών ψηφίων. Φαίνονται τα ποσοστά αναγνώρισης με διάφορους αριθμούς πρωτευουσών συντεταγμένων. Τα ανοικτά σύμβολα είναι ανακατασκευές εικόνας από το σύνολο εκμάθησης. Τα μαύρα σύμβολα είναι ανακατασκευές από το σύνολο ελέγχου (δεν χρησιμοποιήκαν στην σχεδίαση των κωδικοποιητών). Τα τετράγωνα δείχνουν αναγνώριση απλού ψηφίου, τα τρίγωνα τριπλέτας ψηφίων. Η συμπαγής οριζόντια γραμμή δείχνει το ποσοστό αναγνώρισης απλού ψηφίου και η διακεκομμένη οριζόντια δείχνει τα ποσοστά τριπλέτας ψηφίων για τις πρότυπες εικόνες στην ίδια ανάλυση. Οι ανοικτοί κύκλοι τον τρόπο που μεταβάλλεται η απόκλιση με τον αριθμό των πρωτευουσών συντεταγμένων που χρησιμοποιήθηκαν για την κωδικοποίηση των εικόνων.*

Όταν χρησιμοποιήθηκαν 15 και 25 πρωτεύουσες συντεταγμένες δεν υπήρξε σημαντική διαφορά στην απόδοση στα σύνολα ελέγχου και εκμάθησης, και οι πίνακες σύγχισης ήταν παρόμοιοι και για τους δύο κωδικοποιητές. Οι κυριότερες συγχίσεις ήταν μεταξύ των ‘μηδέν (zero)’ και ‘επτά (seven)’, και ‘οκτώ (eight)’ και ‘εννέα (nine)’. Το ποσοστό αναγνώρισης αυξήθηκε σημαντικά όταν αυξήθηκε και ο αριθμός των πρωτευουσών συντεταγμένων από 5 σε 15, αλλά λίγο όταν ο αριθμός τους έγινε 25 από 15. Στην κωδικοποίηση 25 καναλιών, το ποσοστό αναγνώρισης απλού ψηφίου ήταν λίγο μόνο

μικρότερο (65-68%) από αυτό για τις πρότυπες εικόνες στην ίδια ανάλυση (70%). Τελικά λοιπόν, μια κωδικοποίηση 15 καναλιών μάλλον είναι επαρκής για την αναπαράσταση δεδομένων ελέγχου. Τα ποσοστά αναγνώρισης έδειξαν μια ισχυρή συσχέτιση με την απόκλιση της PCA επάνω στο σύνολο των εικόνων εκμάθησης, πράγμα το οποίο προσφέρεται ως άμεσος τρόπος αντικειμενικής εκτίμησης της απόδοσης της PCA. Παρόλο που χρησιμοποιήθηκε περιορισμένο λεξιλόγιο κκαι ένα ευρύτερο θα απαιτούσε μεγαλύτερο σύνολο εκμάθησης, οι φυσικοί και ανατομικοί περιορισμοί των σχημάτων του στόματος δείχνουν ότι ένα PCA μοντέλο κωδικοποίησης θα ήταν εφικτό και λογικά συμπαγές.

3.Αυτόματα συστήματα για επεξεργασία εικόνας ομιλίας.

Τα συστήματα επεξεργασίας εικόνας ομιλίας που είναι βασισμένα σε υπολογιστή, έχουν γενικά αναπτυχθεί παράλληλα με τις μεθόδους καταγραφής και ανάλυσης που περιγράφηκαν στις προηγούμενες παραγράφους. Για αυτό το λόγο, στα πρώτα συστήματα αναγνώρισης, οι φόρμες αναφοράς για τα στοιχεία του λεξιλογίου, αποτελούνταν από σετ αρθρωτικών τροχιών για χαρακτηριστικά του προσώπου όπως την περίμετρο και το εμβαδό της στοματικής κοιλότητας, το πλάτος του στόματος, και το διαχωρισμό των χειλιών. Η αναγνώριση γινόταν βρίσκοντας το σετ από φόρμες αναφοράς που ήταν πιο κοντά στην άγνωστη φόρμα, αφού ευθυγραμίζονταν οι φόρμες υπό αναγνώριση και αυτές τις αναφοράς με τη μέθοδο της Δυναμικής Χρονικής Στρέβλωσης (Dynamic Time-Warping). Στις επόμενες προσπάθειες, χρησιμοποιήθηκε μια αναπαράσταση του περιγράματος του εσωτερικού χείλους η οποία έκανε δυνατή την εκτίμηση της απόστασης μεταξύ των σχημάτων των χειλιών. Έγινε έτσι δυνατή η κατασκευή ενός κωδικού βιβλίου με μεγέθη των χειλιών και η σύγκριση των φορμών υπό αναγνώριση και αναφοράς με διανυσματική κβάντιση, δηλαδή αντικαθιστώντας τα πραγματικά σχήματα του στόματος με τα πιο κοντινά τους, που υπήρχαν μέσα στο κωδικό βιβλίο. Ένα συγκεκριμένο πείραμα αναγνώρισης με πλαίσιο διανυσματικής κβάντισης που χρησιμοποιούσε ένα κωδικό βιβλίο με 256 στοιχεία, είχε ως αποτέλεσμα

ποσοστά αναγνώρισης 74% για κανονικές λέξεις και 96% για λέξεις ψηφίων (πχ ‘ένα’, ‘δύο’ κτλ.) Η αναγνώριση στις αμέσως επόμενες προσπάθειες γινόταν με τη βοήθεια γρηγορότερων, μεγαλύτερων και ισχυρότερων επεξεργαστών για την καταχώρηση και ανάλυση των δεδομένων εικόνας προσώπου, καθώς και την οδήγηση νευρωνικών δικτύων και άλλων μοντέλων αναγνώρισης βασισμένων σε στατιστικές μεθόδους.

Αντιστρόφως οι πρώιμοι συνθέτες εικόνας ομιλίας χρησιμοποιούσαν τα διαθέσιμα διανυσματικά συστήματα γραφικών, για να δημιουργήσουν διαγράμματα περιγραμμάτων δύο διαστάσεων, όπου έδειχναν τα ουσιώδη χαρακτηριστικά της τοπογραφίας του προσώπου. Η Animation επετεύχθηκε ψευδο-κινηματογραφικά, υπολογίζοντας σετ από διανύσματα των σχημάτων του προσώπου σε μικρά χρονικά διαστήματα (20-40 millisecond). Τα πλαίσια αυτά τοποθετούνταν στην οθόνη και γινόταν εξομοίωση της κίνησης στην ίδια χρονική κλίμακα. Τα συστήματα που έχουν εξελιχθεί τελευταία, έχουν βασιστεί στην διάθεση καλύτερων γραφικών και επεξεργαστών για να κατασκευάσουν βελτιωμένες έγχρωμες εικόνες προσώπου. Περιλαμβάνουν χαρακτηριστικά όπως το δέρμα, τα δόντια και τη γλώσσα, τα οποία είναι δύσκολο να αναπαρισταθούν μέσω του περιγράμματός τους.

Η σύνθεση και η αναγνώριση όσο αφορά στην εικόνα, είναι περίπλοκες λόγω των φαινομένων συνάρθρωσης, δηλαδή γνωστών διακυμάνσεων στα σχήματα του στόματος για μερικά φωνήματα, όταν αυτά εμφανίζονται σε διαφορετικά φωνητικά πλαίσια. Αυτά τα φαινόμενα είναι σημαντικά αν θέλουμε η σύνθεση εικόνας ομιλίας να είναι φυσική και κατανοητή. Είναι ζωτικής σημασίας σε ένα σύστημα αυτόματης αναγνώρισης εικόνας ομιλίας, γιατί εκτός και αν το λεξιλόγιο είναι αυστηρά περιορισμένο, οι συναρθρωτικές αποκλίσεις πρέπει να αναγνωρίζονται χωρίς να καταχωρούνται πάρα πολλά πρότυπα αναφοράς.

3.1 Αναγνώριση εικόνας προσώπου και αναγνώριση εικόνας ομιλίας.

Ακόμα και όταν χρησιμοποιούνται εξελιγμένες μέθοδοι επεξεργασίας εικόνας, η αναγνώριση εικόνας ομιλίας παραμένει ένα πρόβλημα ταύτισης προτύπων, όπου τα

χαρακτηριστικά που μεταβάλονται με το χρόνο πρέπει να ‘παγιδευτούν’ και να συγκριθούν με κάποιας μορφής βιβλιοθήκη αναφοράς. Οι χρονικές αποκλίσεις, έχουν ‘παγιδευτεί’ από συστήματα επεξεργασίας τα οποία ανιχνεύουν τις διαφορές μεταξύ αλλεπάληλων εικόνων συγκεκριμένων περιοχών γύρω από το στόμα με τη χρήση τεχνικών οπτικής ροής. Για την ταύτιση χρησιμοποιήθηκαν φόρμες με αλλαγές τέτοιου είδους. Τα νευρωνικά δίκτυα προσφέρουν ένα δρόμο για αναγνώριση και ταύτιση προτύπων τα οποία δεν είναι πανομοιότυπα με αυτά στη φάση της εκμάθησης. Για αυτό το λόγο μπορούν να ανταπεξέλθουν στο πρόβλημα της συνάρθρωσης και στην ποικιλία των ατομικών εκφράσεων κάθε ομιλητή, ακόμα και σε στενά φωνητικά πλαίσια. Ένα σύστημα βασισμένο σε MLP χρησιμοποιήθηκε για την εκμάθηση μιας απεικόνισης μονόχρωμης εικόνας της στοματικής περιοχής ενός ομιλητή, σε ένα φάσμα χαμηλής εμβέλειας. Η διτροπική ολοκλήρωση επετεύχθηκε συνδιάζοντας αυτό το φάσμα με το ανάλογο που προήλθε κατευθείαν από το αντίστοιχο ηχητικό σήμα, παράγοντας έτσι ένα μοναδικό σήμα εικόνας-ήχου. Τα MLPs δέν είναι ιδανικά για την αναπαράσταση χρονικών αποκλίσεων, αλλά μια παραλλαγή τους, τα νευρωνικά δίκτυα χρονικής καθυστέρησης (Time Delay Neural Networks,TDNN), μπορούν να δεχθούν εισόδους μέσα σε κάποιο χρονικό παράθυρο. Έχει περιγραφεί ένα τέτοιο πρωτότυπο σύστημα, το οποίο χρησιμοποίησε δέκα συγκεκριμένα σημεία γύρω από τα χείλια, με σκοπό να εκτιμήσει πέντε στοματικές παραμέτρους. Αυτές χρησιμοποιήθηκαν ως οπτικές είσοδοι σε ένα συνδιασμένο TDNN εικόνας-ήχου κατα τη διάρκεια ενός πειράματος με δέκα λέξεις. Η επιπλέον χρησιμοποίηση εικόνας, βελτίωσε σημαντικά το ποσοστό αναγνώρισης του ηχητικού σήματος παρουσία θορύβου. Ένα παρόμοιο σύστημα χρησιμοποίησε ξεχωριστά ηχητικά και οπτικά TDNN για την φωνηματική ταξινόμηση. Οι ξεχωριστές έξοδοι συνδιάζονταν στο επόμενο επίπεδο, σε ένα άλλο TDNN. Και αυτό το σύστημα έδειξε τα οφέλη της χρήσης εισόδων εικόνας στην αναγνώριση ομιλίας, παρουσίας θορύβου.

Τα κρυφά μοντέλα Markov (Hidden Markov Models, HMM), έχουν εξελιχθεί σε μια πολύ διαδεδομένη μέθοδο αντιμετώπισης των αποκλίσεων στην άρθρωση και στο χρόνο κατά την παραγωγή ομιλίας. Ένα HMM μοντελοποιεί την παραγωγή ομιλίας ως μία μηχανή πεπερασμένων καταστάσεων (Finite State Mashine, FSM), της οποίας η ακολουθία καταστάσεων είναι ορισμένη με πιθανότητες αλλαγής κατάστασης. Κάθε

κατάσταση έχει μια συνάρτηση πυκνότητας πιθανότητας (probability density function,PDF), η οποία εκφράζει την πιθανοφάνεια ενός συγκεκριμένου προτύπου, μέσα από πολλά πιθανά πρότυπα εξόδου, να παραχθεί στην συγκεκριμένη κατάσταση. Οι πιθανότητες αλλαγής κατάστασης και οι PDF εξάγονται από δεδομένα πραγματικής ομιλίας, τα οποία αναπαρίστανται με χρονικά μεταβαλλόμενα διανύσματα παραμέτρων των χαρακτηριστικών του προσώπου. Γι αυτό τα HMM προσφέρουν όχι μόνο ένα τρόπο συνδιασμού ηχητικών και οπτικών δεδομένων σε ένα μοναδικό διάνυσμα, αλλά και και ένα φυσικό τρόπο αντιμετώπισης της ποικιλίας παραγωγής ομιλίας. Για την αναγνώριση, υπολογίζεται για κάθε ένα από τα στοιχεία του λεξιλογίου (λέξεις ή φωνήματα), ένα σετ από HMM. Τα φωνήματα μπορεί να είναι στη μορφή 'τρίφωνου', δηλαδή φωνητικοί φθόγγοι, μέσα στο πλαίσιο ενός προηγούμενου και ενός επόμενου φθόγγου. Η ακολουθία από HMM η οποία είναι η πιο πιθανή να έχει 'γεννήσει' ένα μη αναγνωρισμένο σετ από διανύσματα χαρακτηριστικών, αντιπροσωπεύει το 'αναγνωρισμένο' λεγόμενο.

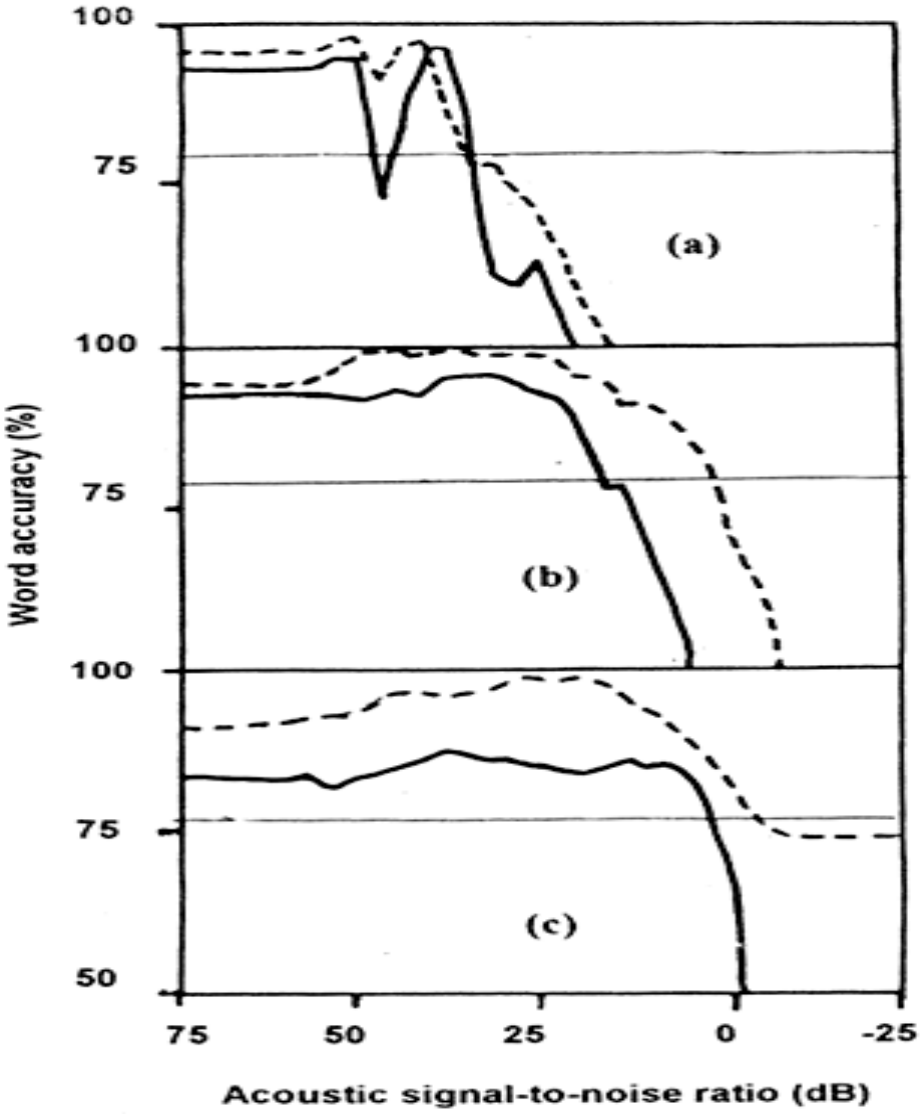
Αναφέρεται ένας αριθμός από πρόσφατες έρευνες, κατά τις οποίες χρησιμοποιούνται ταυτόχρονα οπτικά και ακουστικά χαρακτηριστικά σε αυτόματη αναγνώριση ομιλίας, και είναι βασισμένες σε HMM. Ένα σημαντικό ζήτημα είναι η ολοκλήρωση των ακουστικών και οπτικών δεδομένων. Τα δεδομένα από τους δύο τρόπους μπορούν να συνδυαστούν με πολλούς τρόπους, σε διαφορετικά στάδια της διαδικασίας αναγνώρισης και αυτό μπορεί να επηρεάσει την απόδοση της αναγνώρισης. Πολλές από τις έρευνες χρησιμοποιούν χαρακτηριστικά που εξάγονται από εικόνες ομιλητή, για να παραχθούν τα οπτικά στοιχεία του διανύσματος χαρακτηριστικών.

Χρησιμοποιώντας PCA κωδικοποίηση εικόνας, μπορεί να κατασκευαστεί ένα οπτικό μοντέλο που επεξεργάζεται τις ιδιότητες της απόκλισης που περιγράφηκαν παραπάνω. Έχει αναφερθεί ένα πρωτότυπο τέτοιο σύστημα. Σε αντίθεση με τα ιεραρχικά συστήματα τα οποία ταξινομούν γεγονότα ομιλίας σε διαφορετικά σημεία, μέσα σε ανεξάρτητες διαδικασίες και μπορούν να απαλλαγούν από σχετικά δεδομένα, αυτό το πρωτότυπο επεξεργάστηκε όλα τα δεδομένα διτροπικά. Για την εκμάθηση και τον έλεγχο ενός εξαρτώμενου από ομιλητή αναγνωριστή, χρησιμοποιήθηκε ένα λεξιλόγιο από τις λίστες του NATO RSG-10. Προστέθηκαν επίσης διαφορετικά επίπεδα θορύβου στην ακουστική είσοδο των τριπλετών ελέγχου για να ελεγθεί το κέρδος στην απόδοση

από τη χρήση οπτικών δεδομένων. Μαγνητοσκοπήθηκαν μονόχρωμες εικόνες του προσώπου και καταχωρήθηκαν ως εικόνες 64x64 σημείων σε πραγματικό χρόνο (25 πλαίσια το δευτερόλεπτο). Η λήψη του ακουστικού σήματος γινόταν ταυτόχρονα, σε πραγματικό χρόνο, χρησιμοποιώντας μία 26-καναλιών αναπαράσταση, των 100 πλαισίων ανά δευτερόλεπτο. Η 64x64 σημείων ανάλυση μειώθηκε στα 16x16 και επιλέχθηκε μια περιοχή 10x6 σημείων από κάθε πλαίσιο. Αυτή η ανάλυση είναι λίγο πάνω από το κάτω όριο στο οποίο τα οπτικά χαρακτηριστικά χάνονται. Διακόσιες τριπλέτες ψηφίων χρησιμοποιήθηκαν σαν δεδομένα εκμάθησης για το PCA, με σκοπό να φιαχθούν κωδικοποιητές για τα δεδομένα εικόνας με ένα, τρία και δέκα κανάλια ή πρωτεύουσες συντεταγμένες, για την κωδικοποίηση κάθε πλαισίου εικόνας. Οι ν-καναλιών κωδικοποιημένες εικόνες, σχημάτιζαν το οπτικό συστατικό ενός συνδιασμένου οπτικο-ακουστικού διανύσματος χαρακτηριστικών και οι έξοδοι των 26 καναλιών σχημάτιζαν το ακουστικό συστατικό. Η ομιλία παρίστατο διτροπικά στα 100 διανύσματα το δευτερόλεπτο, χρησιμοποιώντας τέσσερα αντίγραφα από κάθε πλαίσιο εικόνας. Για τα πειράματα αναγνώρισης, χρησιμοποιήθηκαν HMMs τριφώνων με σκοπό τη μοντελοποίηση της συνάρθρωσης. Οι PDFs των καταστάσεων ορίσθηκαν ως κανονικές (Γκαουσιανές) κατανομές και οι παράμετροι των HMM προσδιορίστηκαν με τη χρήση των 200 τριπλετών που χρησιμοποιήθηκαν και στους PCA κωδικοποιητές. Εκατό από τις RSG-10 τριπλέτες που δεν είχαν χρησιμοποιηθεί για την εκμάθηση, έγιναν δεδομένα ελέγχου για τα πειράματα αναγνώρισης. Τα αποτελέσματα της λεξικής ακρίβειας, φαίνονται στο **σχήμα 4**. Για περισσότερη ευκρίνεια, είναι σχεδιασμένα για αναγνωριστές με : 26 κανάλια ακουστικής εισόδου (μόνο ήχος), 26 κανάλια ακουστικής εισόδου και 10 κανάλια οπτικής εισόδου (πλήρως οπτικο-ακουστικά), και 10 κανάλια οπτικής εισόδου (μόνο εικόνα). Το **(α)** μέρος του σχήματος, δείχνει την ακρίβεια αναγνώρισης όταν χρησιμοποιήθηκαν HMM με αποκλίσεις κατάστασης, δηλαδή όταν οι μέσες τιμές και οι αποκλίσεις κάθε παραμέτρου σε κάθε κατάσταση υπολογίστηκαν ξεχωριστά. Ουσιαστικά, το οπτικό συστατικό προσφέρει μικρό όφελος, εκτός από τα επίπεδα υψηλού ακουστικού θορύβου. Ακόμα και εκεί, ο οπτικο-ακουστικός αναγνωριστής, αποδίδει χειρότερα από έναν αμιγώς οπτικό. Αυτό συμβαίνει γιατί η συνεισφορά των αλλοιωμένων ακουστικών δεδομένων καταστρέφει την απόδοση του οπτικο-ακουστικού αναγνωριστή. Για την απόκτηση μιας πιο ρεαλιστικής βάσης για την

εκτίμηση του ρόλου των ακουστικών δεδομένων, υιοθετήθηκε ένας πρότυπος αλγόριθμος ανίχνευσης και μάσκας, για την επανόρθωση του προβλήματος του ακουστικού θορύβου. Τα αποτελέσματα αυτού, φαίνονται στο μέρος **(β)**. Υπάρχει μια βελτίωση στην ακουστική-μόνο αναγνώριση, και μια θετική συνεισφορά από την πρόσθεση δεδομένων εικόνας. Το κέρδος είναι σημαντικό στα υψηλά επίπεδα του ακουστικού λόγου σήματος προς θόρυβο. Παρόλα αυτά, η απόδοση σε αυτά τα επίπεδα είναι πιο φτωχή από αυτή του αναγνωριστή εικόνας-μόνο, στα ίδια επίπεδα. Αυτό ήταν αποτέλεσμα της αμεροληψίας των καταστάσεων των HMM με χαμηλές αποκλίσεις στα επίπεδα υψηλού θορύβου. Για την αντιμετώπιση του προβλήματος, τα HMM υπολογίσθηκαν με αποκλίσεις αφορούσαν μόνο στις παραμέτρους του διανύσματος χαρακτηριστικών και δεν εξαρτόνταν από τις καταστάσεις. Τα αποτελέσματα φαίνονται στο μέρος **(γ)**. Όπως ήταν αναμενόμενο, η συνολική ακρίβεια λέξης μειώθηκε, αλλά η συνεισφορά του οπτικού συστατικού είναι σημαντική στο συνολικό εύρος του λόγου σήματος προς θόρυβο, και δεν καταστρέφεται από την 'μόλυνση΄που επιφέρει το ακουστικό συστατικό στα υψηλά επίπεδα του λόγου σήματος προς θόρυβο. Εκεί η απόδοση είναι σχεδόν το ίδιο καλή με αυτή του οπτικού-μόνο αναγνωριστή. Σε αυτά τα επίπεδα απόδοσης, μπορούν να εξαχθούν χρήσιμες εφαρμογές, ακόμα και γι αυτό το απλό λεξιλόγιο και το σημαντικότερο πλεονέκτημα της ολοκληρωματικής προσέγγισης, είναι το γεγονός ότι ο αναγνωριστής της ίδιας μορφής, μπορεί να χρησιμοποιηθεί σε όλα τα επίπεδα ακουστικού θορύβου.

Ένα άλλο πρόσφατο σχέδιο, χρησιμοποίησε επίσης τεχνικές PCA, αλλά τις ενσωμάτωσε σε ένα πιο περίπλοκο σύστημα, το οποίο χρησιμοποίησε ένα 'μοντέλο ενεργού περιγράμματος', σχετισμένο με σχήματα φιδιού και παραμορφομένων φορμών, για την έρευνα, εντοπισμό και αναπαράσταση του περιγράμματος της περιοχής των χειλιών σε κάθε πλαίσιο εικόνας.



Σχήμα 4: *Αναγνώριση λέξης ψηφίων με χρήση οπτικο-ακουστικού αναγνωριστή, με ποικίλα επίπεδα εξομειωμένου ακουστικού θορύβου.* ***(α)*** *HMMs με απόκλιση εξαρτώμενη από την κατάσταση,* ***(β)*** *HMMs με απόκλιση εξαρτώμενη από την κατάσταση και ακουστική ανίχνευση και μάσκα,* ***(γ)*** *HMMs με απόκλιση εξαρτώμενη από τις παραμέτρους του διανύσματος χαρακτηριστικών και ακουστική ανίχνευση και μάσκα. Η λεπτή οριζόντια γραμμή δείχνει οπτική-μόνο αναγνώριση, η παχιά συνεχής γραμμή δείχνει ακουστική-μόνο αναγνώριση και η διακεκομένη γραμμή οπτικο-ακουστική αναγνώριση.*

Σε μια άλλη εφαρμογή, συνδιάστηκαν τα HMMs με μοντελοποίηση χρησιμοποιώντας MLPs. Μια μορφή MLP χρησιμοποιήθηκε για τον υπολογισμό των PDFs ενός σετ από HMMs με εισόδους από τα συνδιασμένα διανύσματα οπτικο-ακουστικών

χαρακτηριστικών. Η αναγνώριση έγινε με τα HMM που περιγράφηκαν παραπάνω. Σημαντικό κέρδος σημειώθηκε στην ακουστική θορυβώδη ομιλία. Σε μια παραλλαγή, το διάνυσμα χαρακτηριστικών ενσωματώνει επίσης τις διαφορές μεταξύ εικόνων για να υπογραμμιστεί η σημασία των διαφόρων ειδών και επιπέδων των κινήσεων των χειλιών.

3.2 Σύνθεση εικόνας ομιλίας

Η εξέλιξη της σύνθεσης εικόνας ομιλίας, έχει πλήρως περιγραφεί σε άλλα άρθρα. Τα κυριότερα χαρακτηριστικά μπορούν να περιγραφούν ως εξής. Οι περισσότερες σύγρονες μέθοδοι, υιοθετούν γραφικά και πραγματοποιούν μοντέλα του προσώπου στις τρεις διαστάσεις. Έχουν γίνει επίσης και τρισδιάστατα, πιο λεπτομερή μοντέλα μερικών αρθρωτών όπως τα δόντια και η γλώσσα, μαζί με στρατηγικές για την χρησιμοποίησή τους. Στην πραγματικότητα, σε όλα αυτά τα μοντέλα, το κυριότερο πρόβλημα είναι ο έλεγχός τους με οικονομικό τρόπο, χρησιμοποιώντας μόνο ένα μικρό αριθμό παραμέτρων οι οποίες μπορούν να παραχθούν από ένα απλό (πχ φωνητικό ή ορθρογραφικό) χαρακτηριστικό της ομιλίας που θα συνθεθεί. Τα μοντέλα γίνονται κινούμενες εικόνες με δυο γενικούς τρόπους.

(α) Ο πρώτος, διαμορφώνει το πρόσωπο για ένα μικρό σετ από εξιδανικευμένες κινήσεις, που αντιστοιχούν σε συγκεκριμένους ήχους ομιλίας ή φωνήματα, μαζί με κάποιες μεταβατικές κινήσεις. Στη συνέχεια συνθέτονται οι ακολουθίες εικόνων από μία εκφορά λόγου βασισμένη σε φωνήματα, παρεμβάλλοντας πλαίσια μεταξύ μιας κατάλληλης ακολουθίας από διαμορφώσεις. Αυτή η μέθοδος μπορεί να τροποποιηθεί, έτσι ώστε να συμπεριλάβει και την συνάρθρωση. Τα μοντέλα που είναι έτσι σχεδιασμένα, χρησιμοποιούν μια θεωρία κινήσεων ομιλίας για την παραγωγή ενός εύχρηστου και ελέγξιμου πεδίου συναρθρωτικών αποκλίσεων. Το τελευταίο απαιτεί πρόσθετες παραμέτρους ελέγχου.

(β) Ο δεύτερος τρόπος, αφορά στη μοντελοποίηση της λεπτομερούς ανατομίας του προσώπου, δηλαδή του δέρματος των μυών και των οστών κ.α. Οι κινήσεις των μοντέλων, ελέγχονται εξακριβώνοντας τις κινήσεις των μυικών παραμέτρων με τη χρήση πχ του FACS. Αυτό, από τη μία μπορεί να προσφέρει πολύ ακριβείς και διακρίσιμες κινούμενες εικόνες του προσώπου, από την άλλη όμως έχει δυο μειονεκτήματα.

Πρώτον, δεν υπάρχει κάποια άμεση μέθοδος για την παραγωγή ηχητικών φωνημάτων, προσδιορίζοντας τις χρονικά μεταβαλλόμενες παραμέτρους των μυών με τη χρήση αλγορίθμων. Δεύτερον, σημαντικοί αρθρωτές όπως τα δόντια και η γλώσσα, είναι ορατοί μόνο μερικώς ή με διακοπές. Γι αυτό και είναι δύσκολο να παρατηρηθούν και να αναλυθούν οι κινήσεις τους, έτσι ώστε να εξαχθούν παράμετροι ελέγχου από αυτούς.

Σε μια προσπάθεια να υπερβούν μερικές από αυτές τις δυσκολίες, έχουν χρησιμοποιηθεί νευρωνικά δίκτυα, σε συστήματα σύνθεσης ομιλούντων προσώπων. Για παράδειγμα, ένα πρόσφατο σύστημα χρησιμοποίησε ένα MLP για την παραγωγή απεικονίσεων μεταξύ φωνημάτων και εικόνων επιλεγμένων από ένα προκαθορισμένο σετ πλαισίων προσώπου. Έχει επίσης χρησιμοποιηθεί και ένα TDNN για να επιλύσει το πρόβλημα της εκτίμησης των αρθρωτών από ακουστική ομιλία.

Ένας εναλλακτικός τρόπος αντιμετώπισης των δυσκολιών που περιγράφηκαν παραπάνω, χρησιμοποιεί ένα πρόσφατο πακέτο προγραμμάτων το οποίο δίνει μία εντελώς data driven προσέγγιση στη σύνθεση εικόνας ομιλίας. Εκεί, συνδιάζονται οι δύο στατιστικές τεχνικές της PCA και των HMM. Σε αυτή την προσέγγιση, φιάχνονται εικόνες από πραγματικά μοντέλα ομιλίας και περιλαμβάνουν όλους τους ορατούς αρθρωτές., καθώς επίσης και την υφή και σκίαση του δέρματος. Η PCA χρησιμοποιείται για την αναπαράσταση εικόνων προσώπου από ισοδύναμα διανύσματα παραμέτρων.

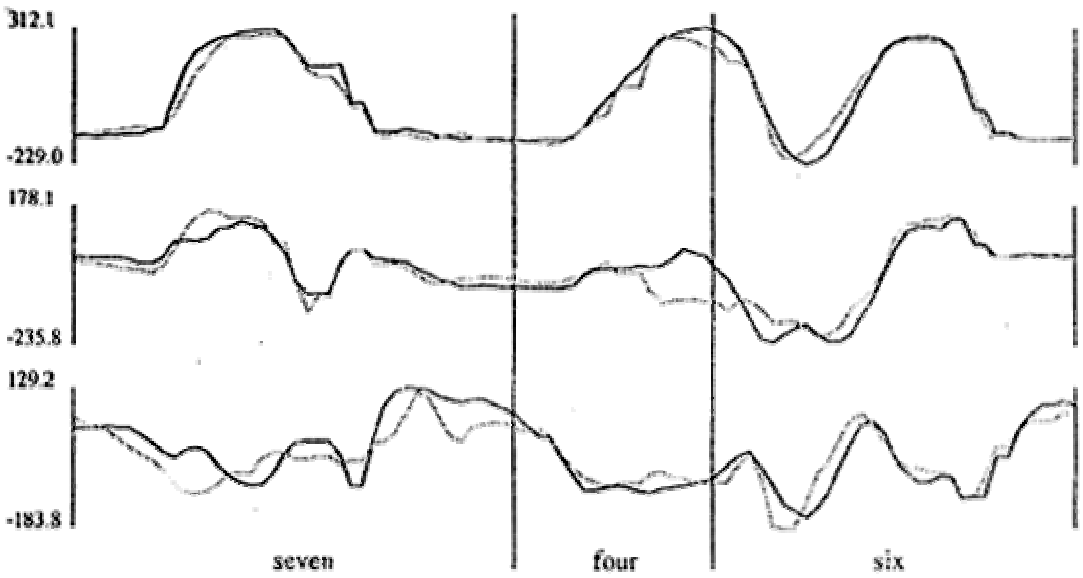
Οι υπολογίσιμες πραγματικές εκφορές πραγματικού λόγου, μαζί με την φυσική τους ποικιλία και φωνητική ευαισθησία, έχουν πραγματοποιηθεί από HMMs και είναι ευρέως διαδεδομένα στην αναγνώριση ομιλίας. Αν και αυτή είναι η πιο συνηθισμένη χρήση τους, ένα HMM μπορεί να παράγει μια ακολουθία από παραμέτρους χαρακτηριστικών τα οποία θα περιλαμβάνουν την ομιλία που νοντελοποιείται από το HMM. Έχει επιδειχθεί ένας συνθέτης εικόνας ομιλίας, βασισμένος σε HMM με είσοδο ακουστική ομιλία. Αυτός ο συνθέτης, χρησιμοποίησε ένα πλήρως ενωμένο HMM με δεκαέξη μόνο καταστάσεις, κάθε μια από τις οποίες αντιστοιχούσε σε ένα διανυσματικώς κβαντισμένο σχήμα χειλιού. Επίσης, τα χαρακτηριστικά της γλώσσας και των δοντιών εισήχθησαν περισσότερο τυχαία στο μοντέλο, δηλαδή δεν είχαν προβλεφθεί.

Χρησιμοποιώντας κατάλληλα συνδιασμένα σετ από HMMs, είναι δυνατή η παραγωγή γενικευμένων εκφορών λόγου. Η data driven προσέγγιση σημαίνει επίσης ότι η σύνθεση δεν εξαρτάται ούτε από μοντέλα προσώπου, ούτε από προεπιλεγμένα σετ από εικόνες.

Αν και ο υπό εξέταση πρωτότυπος συνθέτης παράγει μόνο εικόνες, είναι δυνατό να συμπληρωθούν τα HMM διανύσματα με ακουστικά δεδομένα και να συντεθούν ακουστικά και οπτικά σήματα παράλληλα.

Ένας πρωτότυπος συνθέτης εικόνας ομιλίας ικανός για σύνθεση ακολουθιών λέξεων ψηφίων, έχει κατασκευαστεί και εκτιμηθεί. Τα δεδομένα εκμάθησης αποκτήθηκαν με τη χρήση 200 εκφορών από τις RCG-10 λίστες τριπλετών όπως περιγράφηκε σε προηγούμενη παράγραφο. Όλα τα 17000 πλαίσια των 32x24 μονόχρωμων εικόνων της περιοχής του στόματος, εισήλθαν σε PCA και οι τιμές των πρώτων δεκαπέντε πρωτευουσών συντεταγμένων χρησιμοποιήθηκαν για την κωδικοποίηση κάθε πλαισίου. Τα αρχικά δεδομένα εικόνας τμηματοποιήθηκαν με το χέρι και όλες οι κωδικοποιημένες ακολουθίες πλαισίων έλαβαν φωνητικές ‘ταμπέλες’.

Για τη δημιουργία του πρωτότυπου συνθέτη εικόνας ομιλίας, τα HMMs κατασκευάστηκαν για κάθε τρίφωνο που μπορούσε να εμφανισθεί στο πλαίσιο ομιλούμενων ψηφίων. Μεταβάσεις μπορούσαν να γίνουν μόνο στην ίδια ή στην επόμενη κατάσταση. Οι αρχικές παράμετροι προήλθαν από τα κωδικοποιημένα δεδομένα εκμάθησης ως εξής. Για κάθε δείγμα από κάθε τρίφωνο, υπολογίστηκε ο ελάχιστος αριθμός καταστάσεων που απαιτείται για να μοντελοποιηθούν τα κυριότερα χαρακτηριστικά του παραδείγματος. Ο τελικός αριθμός των καταστάσεων σε κάθε τρίφωνο HMM, τέθηκε ίσος με τον αριθμό στο δείγμα αυτού του τριφώνου με το μεγαλύτερο αριθμό καταστάσεων. Η διαδικασία, κατέληξε σε HMMs με 3-9 καταστάσεις (3-4 καταστάσεις ήταν το πιο σύνηθες). Στα HMM υποτέθηκε μια κανονική κατανομή πολλών μεταβλητών, των τιμών των 15 πρωτευουσών συντεταγμένων σε κάθε κατάσταση. Οι αρχικές τιμές των μέσων τιμών και αποκλίσεων κάθε κατάστασης και οι αρχικές πιθανότητες μετάβασης μεταξύ διαδοχικών καταστάσεων, υπολογίστηκαν από το παράδειγμα στο σετ εκμάθησης το οποίο έδωσε τον μέγιστο αριθμό καταστάσεων. Η εκμάθηση, ή ο καθαρισμός των παραμέτρων, καθενός από τα τρίφωνα HMMs έγινε τότε εφαρμόζοντας την Baum-Welch διαδικασία επανεκτίμησης σε όλα τα παραδείγματα του ίδιου τρίφωνου στα δεδομένα εκμάθησης.



Σχήμα 5: *Ένα γραφικό παράδειγμα σύνθεσης εικόνας ομιλίας μιας τριπλέτας ψηφίων ('seven four six'). Οι βέλτιστα-ενωμένες χρονικά μεταβαλλόμενες καμπύλες των τιμών της πρώτης, τέταρτης και πέμπτης πρωτεύουσας συντεταγμένης έχουν γίνει για: καταγραμμένες εικόνες εκφορών λόγου (μαύρη γραμμή) και ένα παράδειγμα σύνθεσης βασισμένης σε HMM (γκρι γραμμή).*

Οι ακολουθίες ομιλίας με ψηφία, συντέθηκαν από τα φωνήματα μιας εισόδου ορθογραφίας. Τα HMMs για την κατάλληλη ακολουθία τρίφωνων βαφτίστηκαν ως πιθανοκρατικές, συγχρόνων πλαισίων μηχανές, με σκοπό να παράγουν μια ακολουθία από PCA-κωδικοποιημένες εξόδους. Κάθε συνδιασμός και αριθμός ψηφίων μπορούσε να συντεθεί ως συνεχής εκφορά λόγου. Παρόλα αυτά, τα HMMs παράγουν PCA-κωδικοποιημένες εξόδους ανεξάρτητα οπότε πρέπει να περιορισθούν στο να παράγουν αλλαγές που ποικίλουν ελάχιστα, ώστε να εξομοιώνουν τους ανατομικούς περιορισμούς του ανθρώπινου μηχανισμού παραγωγής ομιλίας. Αυτό έγινε δυνατό υπολογίζοντας ένα κατάλληλο τετραγωνικό μονοπάτι για κάθε τιμή του PCA κώδικα διαμέσου κάθε κατάστασης στα HMMs. Τα tracks τότε "λειάνθηκαν", με τη βοήθεια ενός αλγορίθμου 5 σημείων. Αν και τα tracks δεν είχαν την ίδια στατιστική με το HMM, η απόκλιση είναι μικρή στην πράξη. Το σχήμα 5 δείχνει τα tracks των πρωτευουσών συντεταγμένων για μια πρότυπη και μια συντεθημένη τριπλέτα. Οι επεξεργασμένες ακολουθίες εξόδου

συντελεστών πρωτευουσών συντεταγμένων από τα HMM, χρησιμοποιήθηκε για την ανακατασκευή μιας ακολουθίας 32x24 μονόχρωμων εικόνων οι οποίες εκτέθηκαν σε οθόνη γραφικών, με ρυθμό 1/50 δευτερολέπτα. Ένα παράδειγμα ανασυσταμένης εικόνας φαίνεται στο **σχήμα 2**. Αυτός είναι ένας γρήγορος αλγόριθμος. Έχει βρεθεί ότι η συνολική σύνθεση εικόνας τυχαία επιλεγμένων τριπλέτων ψηφίων χρησιμοποιώντας εκπαιδευμένα HMM, μπορεί να γίνει λίγες φορές σε πραγματικό χρόνο, ακόμα και με τον συγκεκριμένο μη βέλτιστο κώδικα. Πειράματα έχουν δείξει ότι δεν υπήρχε μεγάλη διαφορά στο ποσοστό αναγνώρισης απλών ψηφίων στη σύνθεση τριπλέτων ψηφίων (63,9%) και απλών τριπλέτων σε πρότυπες τριπλέτες, όταν αυτές ήταν εκτεθιμένες στα ίδια επίπεδα γκρίζου και στην ίδια χωρική ανάλυση (68,7%).

Η πρωτότυπη σύνθεση είναι σήμερα επεκταμένη σε ένα μεγαλύτερο λεξιλόγιο. Επιπρόσθετα, έχει εφαρμοστεί ένα μοντέλο ικανό να κωδικοποιεί υψηλής ανάλυσης έγχρωμες εικόνες το οποίο ανακτά τα πλεονεκτήματα της PCA κωδικοποίησης. Η αύξηση της χωρικής ανάλυσης των ακολουθιών εικόνας, θα πρέπει να βελτιώσει την ευκρίνεια των χαρακτηριστικών του προσώπου, καθώς επίσης και να κάνει τις εικόνες ακόμα πιο φυσικές. Η χρήση έγχρωμων εικόνων σε ένα συνθέτη εικόνας ομιλίας θα αυξήσει αρκετά την ρεαλιστικότητα και αποδοχή των συντιθέμενων εκφορών λόγου, ακόμα και αν διαφαίνεται ότι έχει μικρή προσφορά στην πιστότητα της εικόνας ομιλίας.

4.Σύνοψη

Η εξέλιξη αυτόματης αναγνώρισης και σύνθεσης εικόνας ομιλίας με τη χρήση υπολογιστών, έχει γίνει πολύ γρήγορη, ουσιαστικά από τα μέσα της δεκαετίας του ’70. Αυτό οφείλεται σε τρεις λόγους: **(α)** τη συνεχή βελτίωση στις μελέτες της ανάλυσης, σύνθεσης, αναγνώρισης και αντίληψης των ακουστικών σημάτων ομιλίας, **(β)** την χρήση ισχυρών νέων τεχνικών για την επεξεργασία δεδομένων, όπως τα νευρωνικά δίκτυα και στατιστικές τεχνικές όπως τα κρυφά μοντέλα Markov, και **(γ)** τη χρήση πολύ ισχυρών επεξεργαστών στους οποίους εφαρμόζονται αυτές οι τεχνικές, και ειδικότερα στη χρήση τους για τη μεταχείρηση του μεγάλου όγκου δεδομένων που εμπλέκεται στην επεξεργασία εικόνας και στα γραφικά. Η επιτυχία τεχνικών για καταγραφή, ανάλυση και μεταχείριση των ακουστικών δεδομένων ομιλίας, έχει γίνει ένα ερέθισμα για τους

ερευνητές εικόνας ομιλίας. Ειδικότερα, οι πιο πρόσφατες data driven τεχνικές, συμπεριλαμβανομένων πχ της PCA και των HMMs, δείχνουν ότι είναι δυνατό να χειρίζονται τα δεδομένα εικόνας ομιλίας με μικρότερη ανάγκη για *α priori* γνώση της δομής και των κανόνων που τη διέπουν. Για παράδειγμα, δεν είναι πια απαραίτητο να πρέπει να επιλεχθούν και να εξαχθούν τα 'σχετικά' ορατά χαρακτηριστικά του προσώπου όπως χρειαζόταν στις πρώτες προσεγγίσεις. Επίσης, τώρα είναι δυνατό να κατασκευαστούν ολοκληρωμένα συστήματα αναγνώρισης που συνδιάζουν τα οπτικά και ακουστικά χαρακτηριστικά της ομιλίας και να τα επεξεργαστούν παράλληλα χωρίς στην πορεία να γίνονται κατηγοριοποιήσεις και να χάνονται δεδομένα τα οποία μπορεί να χρησιμεύουν στα αργότερα στάδια. Η βελτίωση της ισχύος των υπολογιστών είναι πολύ σημαντική για το υπολογιστικό φόρτο και πολυπλοκότητα αυτών των τεχνικών. Σήμερα, μπορεί να διεκπεραιωθεί υψηλής ταχύτητας ανάλυση εικόνας χωρίς να καταφεύγουμε σε ειδικό υλισμικό και οι τεχνικές συμπίεσης επιτρέπουν την πολύ συμπαγή κωδικοποίηση και καταχώρηση εικόνων. Αντιστρόφως, ενώ στην αρχή της δεκαετίας του '80 η παραγωγή (κινούμενων) εικόνων περιγραμμάτων ομιλούντων προσώπων ήταν μόλις εφικτή σε συγκεκριμένα χρονικά πλαίσια, σήμερα η αναπαράσταση με κινούμενες εικόνες μπορεί να γίνει για όλο το πρόσωπο στα ίδια περίπου χρονικά πλαίσια. Για την ανάδειξη του πως συνδιάζονται οι τεχνικές διαχείρισης δεδομένων με την ισχύ των επεξεργαστών, αυτό το άρθρο έδειξε μια πτυχή κατα την οποία η σύνθεση εικόνας ομιλίας και η οπτικο-ακουστική αναγνώριση ομιλίας, μπορούν να χειριστούν μέσα από την εφαρμογή των ίδιων, δύο data driven μεθόδων. Γι αυτό και αντιπροσωπεύουν τις δυο πλευρές του ίδιου νομίσματος. Ένα πολύ συμπαγές PCA σχήμα κωδικοποίησης μπορεί να χρησιμοποιηθεί για την αναπαράσταση εικόνων ομιλούντων προσώπων. Τα HMMs μπορούν σε εκείνο το σημείο να χρησιμοποιηθούν για την αναπαράσταση ομιλούντων προσώπον. Από τη μία, αυτά μπορούν να χρησιμοποιηθούν για αναγνώριση, από την άλλη μπορούν να παράγουν εξόδους για σύνθεση σε μικρά χρονικά παράθυρα, εγγίζοντας τον πραγματικό χρόνο. Αυτές οι εξελίξεις, όπως πολλές από αυτές που συζητήθηκαν στο άρθρο, δεν θα ήταν δυνατές χωρίς ισχυρούς επεξεργαστές. Η συνεχής εξέλιξη του υλισμικού και των τεχνικών επεξεργασίας είναι τόσο σφοδρή, που η βασική πρόκληση είναι να κινηθούμε μπροστά ώστε να τα εκμεταλευθούμε πλήρως. Το άρθρο

αυτό έδειξε ότι αν το πεδίο συνεχίσει να εξελίσσεται με τον ίδιο γρήγορο ρυθμό, η πρόκληση θα επιτευχθεί.

Περίληψη

Η χρησιμοποίηση του προσώπου ενός ομιλητή συνεισφέρει αρκετά στην καλύτερη κατανόηση της ομιλίας. Σε περιπτώσεις όπως αναπηρία στην ακοή ή στο θάλαμο πλοήγησης αεροπλάνων, η ενίσχυση του ακουστικού σήματος με οπτικό, βοηθά αποτελεσματικά στην αναγνώριση της ομιλίας. Η οπτικοακουστική αναγνώριση και σύνθεση ομιλίας, είναι διτροπικά προβλήματα τα οποία επιλύονται με στατιστικές μεθόδους ή με χρήση νευρωνικών δικτύων. Η πρώτη βασική διεργασία σε τέτοια ζητήματα, είναι η εξαγωγή των ουσιωδών χαρακτηριστικών του προσώπου που μεταβάλλονται κατά την ομιλία. Η αναγνώριση εικόνας ομιλίας μπορεί να γίνει με MLP (πέρσεπτρον πολλών επιπέδων) τα οποία μπορούν να ταξινομήσουν δεδομένα ομιλίας έπειτα από εκπαίδευση. Τα νευρωνικά δίκτυα χρονικής καθυστέρησης (TDNN), χρησιμοποιούνται για πιο ικανοποιητική επίλυση του προβλήματος των χρονικών αποκλίσεων. Η στατιστική μέθοδος της ανάλυσης σε πρωτεύουσες συνταταγμένες, μπορεί να επιτύχει συμπίεση και κωδικοποίηση των δεδομένων εικόνας και η ταυτόχρονη χρήση τους με κρυφά μοντέλα Markov (HMM), οδηγεί σε ικανοποιητικά ποσοστά αναγνώρισης ομιλίας. Τα αποτελέσματα είναι βελτιωμένα αν χρησιμοποιηθούν ταυτόχρονα ακουστικά και οπτικά σήματα ομιλίας. Τα οπτικοακουστικά διανύσματα μπορεί να περιλαμβάνουν ειδικά χαρακτηριστικά όπως η γλώσσα και τα δόντια, εκτός από άλλα γενικότερα όπως τα σχήματα των χειλιών. Η σύνθεση εικόνας ομιλίας, είναι διαδικασία που περιλαμβάνει συνήθως μοντέλα της λεπτομερής ανατομίας του προσώπου. Και εδώ, η ταυτόχρονη χρήση της PCA για την αναπαράσταση των εικόνων προσώπου ως διανύσματα παραμέτρων και των HMM για σύνθεση εικόνας ομιλίας, έχει ικανοποιητικά αποτελέσματα. Το γενικό συμπέρασμα είναι ότι τελικά η σύνθεση εικόνας ομιλίας και η οπτικοακουστική αναγνώριση ομιλίας, αντιπροσωπεύουν τις δυο όψεις του ίδιου νομίσματος. Η γοργή ανάπτυξη ισχυρών επεξεργαστών και η ανάλογη εκμετάλλευσή τους, θα επιφέρει σημαντικές βελτιώσεις και στον τομέα της αναγνώρισης ομιλίας.

Λεξικό Αγγλικών όρων και συντμήσεων

Acoustic cues	Ακουστικά χαρακτηριστικά	
Articulatory movement	Κίνηση αρθρωτών	
Coarticulation	Συνάρθρωση	Αποκλίσεις των σχημάτων του στόματος για φώνημα, όταν υπάρχει σε διαφορετικές λέξεις
Data compression	Συμπίεση δεδομένων	
Digit triple	Τριπλέτα ψηφίων	Πχ 'six zero four'
Digit word	Λέξη ψηφίου	Πχ 'six'
Facial cues	Χαρακτηριστικά προσώπου	
Frames	Πλαίσια	Τα πλαίσια που αποτελείται μια εικόνα
Mapping	Απεικόνιση	
Musculature	Μυικό σύστημα	
Oral region	Περιοχή στόματος	
Quasi-cinematographically	Ψεύδο-κινηματογραφικά	
Training data	Δεδομένα εκπαίδευσης/εκμάθησης	
Triphone	Τρίφωνο	Φωνητικός φθόγγος μέσα στα πλαίσια ενός προηγούμενο και ενός επόμενου φθόγγου
Vector quantization	Διανυσματική κβάντιση	
Video speech synthesis	Σύνθεση εικόνας ομιλίας	
Visual speech	Εικόνα ομιλίας	Η εικόνα του προσώπου ενός ομιλητή
Vowel identification	Αναγνώριση φωνήεντος	
HMM	Κρυφά μοντέλα Markov	
MLP	Πέρσεπτρον πολλών επιπέδων	
PCA	Ανάλυση σε πρωτεύουσες συντεταγμένες	
TDNN	Νευρωνικά δίκτυα χρονικής καθυστέρησης	