ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ & ΔΙΟΙΚΗΣΗΣ

Εργαστήριο Ευφυών Συστημάτων & Ρομποτικής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΔΗΜΙΟΥΡΓΙΑ ΔΙΑΜΕΣΟΛΑΒΗΤΗ ΦΩΝΗΣ ΓΙΑ ΤΗΝ ΚΙΝΗΣΗ ΡΟΜΠΟΤΙΚΟΥ ΟΧΗΜΑΤΟΣ

ΒΛΑΣΣΗ ΑΝΤΩΝΙΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΤΣΟΥΡΒΕΛΟΥΔΗΣ ΝΙΚΟΛΑΟΣ

ΧΑΝΙΑ, ΟΚΤΩΒΡΙΟΣ 2004

Υπεβλήθη στο τμήμα Μηχανικών Παραγωγής και Διοίκησης για την μερική ικανοποίηση των απαιτήσεων για την απόκτηση του Διπλώματος Μηχανικού Παραγωγής και Διοίκησης από το Πολυτεχνείο Κρήτης.

Εξεταστική Επιτροπή

Επίκουρος Καθηγητής Τσουρβελούδης Νικόλαος (Επιβλέπων)
Αναπληρωτής Καθηγητής Κουϊκόγλου Βασίλης
Λέκτορας Νικολός Ιωάννης

Στους γονείς μου

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ

Με το πέρας της εργασίας αυτής θα ήθελα να ευχαριστήσω όλους όσους με βοήθησαν στην διεκπεραίωση της παρούσας διπλωματικής, κάτι που θα ήταν αδύνατο να επιτύχω χωρίς αυτούς.

Πρώτα απ' όλους, θα ήθελα να ευχαριστήσω τον επιβλέποντα Επίκουρο Καθηγητή κ. Νικόλαο Τσουρβελούδη, που μου εμπιστεύτηκε το θέμα. Η καθοδήγηση του στάθηκε πολύτιμη για τη ολοκλήρωση αυτής της εργασίας. Επιπλέον μου έδωσε την ευκαιρία να αποκτήσω σημαντικές εμπειρίες καθ' όλη τη διάρκεια της συνεργασίας μας.

Θερμές ευχαριστίες στους Σάββα Πιπερίδη, μέλος ΕΤΕΠ του Εργαστηρίου Ρομποτικής, και Ελευθέριο Δοϊτσίδη, υποψήφιο Διδάκτορα του τομέα Συστημάτων Παραγωγής, για την κατανόηση και βοήθεια που μου προσέφεραν κατά της διάρκεια του έργου μου. Εκτός εργαστηρίου, θα ήθελα να ευχαριστήσω προσωπικά τους Αθανάσιο Τσαλατσάνη και Αντώνη Σγούρο για την ανεκτίμητη βοήθεια τους.

Τέλος, ευχαριστώ τους φίλους μου για την ηθική συμπαράσταση και βοήθεια καθώς και την οικογένειά μου που στάθηκε δίπλα μου όλο αυτό τον καιρό, από τη στιγμή της εισαγωγής μου στο τμήμα μέχρι και την ολοκλήρωση των σπουδών μου.

ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 ΚΙΝΗΤΡΟ

Οι περισσότερες ερευνητικές μελέτες στα έντροχα ρομποτικά οχήματα είχαν επικεντρωθεί τις τελευταίες δεκαετίες στην εύρεση μεθόδων πλοήγησης με σκοπό την αποφυγή εμποδίων, είτε με χρήση αισθητήρων υπερήχων, είτε με χρήση κάμερας και λήψη εικόνων. Τα τελευταία όμως χρόνια έχει παρατηρηθεί έντονη ανάπτυξη της επεξεργασίας του λόγου καθώς και τεχνικών αναγνώρισης της ανθρώπινης φωνής. Έτσι, η πλοήγηση ρομποτικών οχημάτων μπορεί πλέον να επιτευχθεί και μέσω τέτοιων τεχνικών (αναγνώρισης – σύνθεσης λόγου).

Η φωνή είναι το φυσικό μέσο επικοινωνίας του ανθρώπου. Είναι λογικό λοιπόν να αναζητούμε παρόμοια επικοινωνία και με την μηχανή. Σε αυτή τη βάση στηρίχτηκαν διάφορες μελέτες και εφαρμογές αναγνώρισης φωνής με σκοπό την κίνηση ρομποτικών οχημάτων και όχι μόνο.

1.2 ΣΚΟΠΟΣ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η πλοήγηση ενός ρομποτικού οχήματος μέσω φωνητικών εντολών. Πιο συγκεκριμένα, στόχος είναι η ανάπτυξη προγράμματος διαμεσολαβητή φωνής που θα μετατρέπει τις φωνητικές εντολές του χρήστη σε εντολές κίνησης του οχήματος.

1.3 ΔΙΑΤΥΠΩΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Στην εργασία αυτή, η εφαρμογή του διαμεσολαβητή φωνής γίνεται στο ATRV-Mini. Το ρομποτικό αυτό όχημα είναι μοντέλο της RWI, ενώ αποτελεί το πρώτο έντροχο ρομπότ που αποκτά το τμήμα Μηχανικών Παραγωγής και Διοίκησης του Πολυτεχνείου Κρήτης. Από την άφιξη του μέχρι και σήμερα έχουν ήδη εφαρμοστεί στο ATRV-Mini διάφορες τεχνικές πλοήγησης, ανίχνευσης και αποφυγής εμποδίων. Η εργασία αυτή όμως αποτελεί την πρώτη απόπειρα εφαρμογής συστήματος αναγνώρισης φωνής.

Το πρόβλημα μπορεί ουσιαστικά να διατυπωθεί ως εξής: «Να χρησιμοποιηθεί κατάλληλο σύστημα αναγνώρισης φωνής μεμονωμένων λέξεων και να προσαρμοστεί κατάλληλα ώστε να αναπτυχθεί πρόγραμμα που να επιτρέπει στο ρομποτικό όχημα να κινείται μέσω δεσμευμένων λέξεων (φωνητικών εντολών) που θα δίνει ο εκάστοτε χρήστης. Το πρόγραμμα θα μπορεί να αναγνωρίζει όλες τις φωνές, ώστε να καθίσταται ευέλικτο ως προς την χρήση του».

1.4 ΠΕΡΙΓΡΑΦΗ ΔΟΜΙΚΩΝ ΣΤΟΙΧΕΙΩΝ

Για την εφαρμογή της παρούσας διπλωματικής εργασίας απαιτείται:

- Ένα μικρόφωνο που να δέχεται ως είσοδο φωνητικό σήμα συχνότητας 16 KHz. Στην περίπτωση μας το μικρόφωνο που χρησιμοποιήθηκε είναι ένα απλό Multimedia Microphone της εταιρίας Digitus Accessories.
- Ένας Η/Υ με κάρτα ήχου στην οποία θα συνδεθεί το προαναφερθέν μικρόφωνο. Στην εργασία αυτή χρησιμοποιήθηκε ένας φορητός υπολογιστής Pentium Centrino στα 1700 MHz με κάρτα ήχου και με λογισμικό Windows 2000 ή Windows XP.
- Πρόγραμμα αναγνώρισης φωνής (μεμονωμένων λέξεων) που θα τρέχει στον Η/Υ και θα είναι σε θέση να αναγνωρίζει δεσμευμένες λέξεις ή φράσεις. Εδώ χρησιμοποιήθηκε ένα πρόγραμμα αναγνώρισης φωνής που δέχεται ελληνικές λέξεις προς αναγνώριση.
- Ρομποτικό όχημα πάνω στο οποίο θα προσαρμοστεί και εφαρμοστεί το σύστημα διαμεσολαβητή φωνής. Στην περίπτωση μας έγινε χρήση του ρομποτικού οχήματος ATRV-Mini που προαναφέραμε και για το οποίο θα γίνει εκτενέστερη αναφορά στο επόμενο κεφάλαιο.

1.5 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ

Η εργασία είναι οργανωμένη με τον ακόλουθο τρόπο:

Στο Κεφάλαιο 2 παρουσιάζεται το έντροχο ρομποτικό όχημα που χρησιμοποιήθηκε με τα τεχνικά χαρακτηριστικά και το λογισμικό του.

Στο Κεφάλαιο 3 αναφέρονται γενικά στοιχεία για την αναγνώριση φωνής, ιστορικά στοιχεία της εξέλιξης της, αλλά και αντίστοιχες εφαρμογές της σε ρομποτικούς μηχανισμούς.

Στο Κεφάλαιο 4 παρατίθεται το απαραίτητο θεωρητικό υπόβαθρο και περιγράφονται τα Κρυφά Μοντέλα Markov που αποτελούν την βάση του συστήματος αναγνώρισης που αναλύεται στο επόμενο κεφάλαιο.

Στο Κεφάλαιο 5 γίνεται αρχικά η περιγραφή του λογισμικού αναγνώρισης σε θεωρητικό επίπεδο. Στη συνέχεια αναφέρονται στοιχεία για την ψηφιοποίηση και το μικρόφωνο, απαραίτητα για την αναλυτική περιγραφή των προγραμμάτων αναγνώρισης και ηχογράφησης που ακολουθεί. Αναπτύσσεται η δομή του αλγορίθμου κίνησης, καθώς και ο τρόπος σύνδεσης – επικοινωνίας των λογισμικών αναγνώρισης και κίνησης.

Το Κεφάλαιο 6 αφιερώνεται αποκλειστικά στο θέμα της κίνησης του οχήματος, παρουσιάζονται δηλαδή κάποια γενικότερα στοιχεία για την κίνηση του οχήματος καθώς και η ανάπτυξη του αλγόριθμου κίνησης.

Στο Κεφάλαιο 7 περιγράφεται η εφαρμογή του προγράμματος αναγνώρισης στο ρομποτικό όχημα σε συνδυασμό με το πρόγραμμα κίνησης σε πραγματικό χρόνο. Τέλος, παρουσιάζονται συμπεράσματα και μέσα από αυτά προκύπτουν θέματα για περαιτέρω έρευνα.

ΚΕΦΑΛΑΙΟ 2 : ΠΕΡΙΓΡΑΦΗ ATRV-Mini

2.1 ΕΙΣΑΓΩΓΗ

Το όχημα που χρησιμοποιήθηκε είναι ένα από τα δύο έντροχα οχήματα ATRV-Mini (Damon και Fidias) της Real World Interface (RWI) που αποκτήθηκαν από το Πολυτεχνείο Κρήτης το Σεπτέμβρη του 2000 (Σχήμα 2.1). Πρόκειται για έντροχα ρομπότ εσωτερικών και εξωτερικών χώρων με πολλές δυνατότητες. Τα ATRV-Mini του Εργαστηρίου Ευφυών Συστημάτων και Ρομποτικής έχουν εξοπλιστεί με ασύρματη και ενσύρματη επικοινωνία. Στην παρούσα εργασία θα χρησιμοποιηθεί η ενσύρματη για μεγαλύτερη ταχύτητα, παρ' όλο που αυτό καθιστά αδύνατη την πλοήγηση σε ανοιχτό χώρο.

Το ATRV-Mini διαθέτει σύστημα κίνησης που αποτελείται από δύο σερβοκινητήρες συνεχούς ρεύματος και τέσσερις τροχούς διαφορικής κίνησης. Ο κάθε ηλεκτροκινητήρας είναι συνδεδεμένος με δύο τροχούς. Διαθέτει, επίσης, 24 αισθητήρες ήχου (sonars) οι οποίοι έχουν ως σκοπό τον εντοπισμό εμποδίων που βρίσκονται σε καθορισμένο εύρος απόστασης γύρω από το ρομπότ, παγκόσμιο σύστημα συντεταγμένων (GPS), πυξίδα, κάμερα και 2 προφυλακτήρες (bumpers).

Σχήμα 2.1: Το ATRV-Mini της Real World Interface.

Ο ενσωματωμένος στο ATRV-Mini υπολογιστής αποτελείται από έναν επεξεργαστή INTEL PENTIUM III 500MHz, 64 MB SDRAM, 6 GB IDE HD, 100Mbps Ethernet και λειτουργεί σε περιβάλλον Linux. Η διαχείριση και ο έλεγχος των συστημάτων του ATRV-Mini πραγματοποιείται από το περιβάλλον MOBILITY της RWI.

Τέλος, το ATRV-Mini διαθέτει οθόνη υγρών κρυστάλλων στην οποία εμφανίζονται βασικές λειτουργίες όπως ενδείξεις στάθμης μπαταρίας, έλεγχος sonar και μοτέρ. Οι λειτουργίες αυτές ελέγχονται με το rFlex της RWI [1].

2.2 ΤΕΧΝΙΚΕΣ ΠΡΟΔΙΑΓΡΑΦΕΣ

Στον Πίνακα 2.1 περιγράφονται οι τεχνικές προδιαγραφές του οχήματος. Εκτός από τον εξοπλισμό, ο οποίος προσφέρεται από την εταιρία RWI, το όχημα παρέχει και κάποιες τροποποιήσεις κατά παραγγελία έτσι ώστε να καλύπτει τις ερευνητικές ανάγκες του εργαστηρίου Ευφυών Συστημάτων και Ρομποτικής του Πολυτεχνείου Κρήτης. Στον εξοπλισμό αυτό περιλαμβάνονται κάμερα, GPS και επιπλέον αισθητήρες.

Μήκος	62.2 cm
Πλάτος	53.3 cm
Ύψος	45 cm
Βάρος	38.6 kg
Σώμα	Αλουμίνιο
Ταχύτητα	0-1.5 $m/\sec$
Ωφέλιμο Φορτίο	9 kg
Χρόνος Λειτουργίας	3-6 hr (εξαρτάται από το έδαφος)
Κίνηση	4-wheel, PWM
Τρόπος Κατεύθυνσης	Skid steering
Γωνία Στροφής	0 (στρίβει στο κέντρο)
Μπαταρίες	Δύο 12 V, 12 amp/hr
Κινητήρες	Δύο 0.1 HP, 24 V DC servo motors
Υπολογιστής	Pentium III EBX
I/O Ports	Ethernet, RS-232, Joystick
Αισθητήρες	24 Sonar
Κάμερα	Sony EVI D30
GPS	

Πίνακας 2.1: Τεχνικά Χαρακτηριστικά του οχήματος ATRV-Mini [1]

2.3 ΠΕΡΙΓΡΑΦΗ ΒΑΣΙΚΟΥ ΕΞΟΠΛΙΣΜΟΥ ΤΟΥ ATRV-MINI

2.3.1 Αισθητήρες Υπερήχων

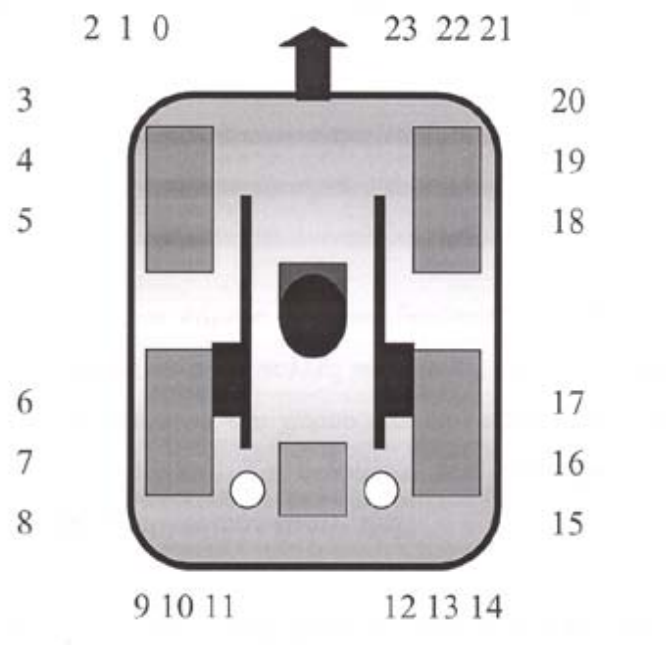
Η λέξη sonar προέρχεται από τα αρχικά των λέξεων “SOund NAvigation and Ranging”, σε ελεύθερη μετάφραση “Ηχητικός προσδιορισμός πορείας και απόστασης”. Αναπτύχθηκε ως μέσο εντοπισμού εχθρικών υποβρυχίων στη διάρκεια του δευτέρου παγκοσμίου πολέμου. Αποτελείται από ένα πομπό, έναν μετατροπέα και έναν δέκτη.

Ένας ηλεκτρικός παλμός που παράγεται από τον πομπό μετατρέπεται σε ηχητικό κύμα και μεταδίδεται στο νερό ή στον αέρα. Όταν το κύμα συναντήσει ένα αντικείμενο, αντηχείται. Η αντήχηση αυτή επιστρέφει στον μετατροπέα και μετασχηματίζεται πάλι σε ηλεκτρικό παλμό ο οποίος αναλύεται από τον δέκτη.

Δεδομένου ότι τόσο η ταχύτητα του ήχου στο μέσο (νερό, αέρα) όσο και ο χρόνος αποστολής και λήψης του ήχου είναι γνωστά, είναι δυνατή η μέτρηση της απόστασης του αντικειμένου. Η ίδια διαδικασία επαναλαμβάνεται πολλές φορές το δευτερόλεπτο. Τα χαρακτηριστικά ενός αποδοτικού συστήματος sonar είναι ένας ισχυρός πομπός μεγάλης εμβέλειας, ένας αξιόπιστος μετατροπέας και ένας ευαίσθητος δέκτης.

Υπάρχουν τρία είδη sonar: Τα ενεργητικά (active sonar), τα παθητικά (passive sonar) και τα ακουστικά (acoustic sonar). Τα ενεργητικά sonars χρησιμοποιούν την διάταξη πομπού – μετατροπέα - δέκτη για να στέλνουν ηχητικούς παλμούς και να λαμβάνουν την αντήχηση των αντικειμένων. Τα παθητικά sonars διακρίνουν τον χαρακτηριστικό ήχο που προκαλούν αντικείμενα όπως τα υποβρύχια, οι φάλαινες κλπ από τον ήχο του περιβάλλοντος, δεδομένου ότι η συχνότητα του ήχου που εκπέμπεται από κάθε τέτοιο αντικείμενο είναι διαφορετική και συγκεκριμένη. Η ταυτόχρονη χρήση από δύο αντικείμενα ενεργητικών sonar δημιουργεί την διάταξη των ακουστικών sonars, η οποία χρησιμοποιείται για την ενημέρωση της θέσης μεταξύ των αντικειμένων. Παρόμοια διάταξη με τα ακουστικά sonar χρησιμοποιούν τα δελφίνια και οι νυχτερίδες για την επικοινωνία τους.

Το έντροχο ρομπότ ATRV-Mini, χρησιμοποιεί 24 ενεργητικά sonars (Σχήμα 2.2).



Σχήμα 2.2: Διάταξη αισθητήρων υπερήχων στο ATRV-Mini [2]

2.3.2 Παγκόσμιο Σύστημα Συντεταγμένων

Το ΠΣΣ (Global Positioning System, GPS), είναι η πιο σημαντική ανακάλυψη στην πλοήγηση μετά την πυξίδα. Ένα σύστημα ΠΣΣ αποτελείται από τρία τμήματα: το τμήμα ελέγχου (Control Segment), το δορυφορικό τμήμα (Space Segment) και την μονάδα χρήσης (User Segment). Το τμήμα ελέγχου αποτελείται από ένα δίκτυο επίγειων σταθμών και αποτελεί το νευρικό σύστημα του συστήματος ΠΣΣ, παρέχοντας συνεχώς πληροφορίες σε κάθε δορυφόρο. Το δορυφορικό τμήμα αποτελείται από ένα πλήθος δορυφόρων, καθένας από τους οποίους εκτελεί

περιστροφή της γης συνήθως κάθε 12 ώρες. Κάθε δορυφόρος φέρει τέσσερα ρολόγια και στέλνει συνεχώς ραδιοσήματα (radio signals) τα οποία χρησιμοποιούν οι μονάδες χρήσης για να υπολογίζουν τη θέση τους. Η μονάδα χρήσης προσδιορίζει τη θέση της από τα ραδιοσήματα του δορυφορικού σταθμού. Αποτελείται από έναν δέκτη σημάτων, ένα ρολόι, μνήμη και έναν επεξεργαστή που εκτελεί υπολογισμούς.

Η μονάδα χρήσης προσδιορίζει τη θέση της μετρώντας τον χρόνο που χρειάζονται τα ραδιοσήματα τεσσάρων δορυφόρων για να ταξιδέψουν σε αυτήν. Απαιτούνται τρεις δορυφόροι ώστε η θέση της μονάδας χρήσης να υπολογισθεί τρισδιάστατα. Ο τέταρτος δορυφόρος χρησιμοποιείται για να επιτευχθεί μεγαλύτερη ακρίβεια. Το σφάλμα ενός αξιόπιστου συστήματος ΠΣΣ είναι μικρότερο των 100 μέτρων. Οι παράγοντες που το επηρεάζουν είναι οι καιρικές συνθήκες με σφάλμα θέσης 5 με 10 μέτρα, τα τυχόν σφάλματα από τον χειριστή και οι σκόπιμες στρατιωτικές παρεμβολές, με σφάλμα θέσης μεγαλύτερο από 100 και λιγότερο από 300 μέτρα. Τα σφάλματα των στρατιωτικών παρεμβολών μπορούν να εξαλειφθούν από ένα εξελιγμένο σύστημα ΠΣΣ, το διαφορικό παγκόσμιο σύστημα συντεταγμένων (Differential Global Positioning System, DGPS). Το ΔΠΣΣ περιέχει έναν ακόμη ραδιοδέκτη συνδεδεμένο ηλεκτρονικά με τον ραδιοδέκτη του συστήματος ΠΣΣ μειώνοντας το σφάλμα κάτω από τα 100 μέτρα. Το σύστημα ΠΣΣ του ATRV Mini είναι ΔΠΣΣ. [3]

2.3.3 Πυξίδα

Αν και το ΠΣΣ είναι ικανό να εντοπίζει το στίγμα του, δεν μπορεί να δώσει ουσιαστικές οδηγίες διεύθυνσης στο έδαφος. Αυτό συμβαίνει γιατί το ΠΣΣ δεν γνωρίζει τη διεύθυνση στόχευσης του. Το πρόβλημα αυτό λύνει η πυξίδα η οποία μαζί με το ΠΣΣ αποτελούν ένα ολοκληρωμένο σύστημα πλοήγησης. Με την πυξίδα είναι δυνατή η πληροφόρηση για τα τέσσερα σημεία του ορίζοντα. Το ATRV-Mini χρησιμοποιεί ηλεκτρονική πυξίδα η οποία παρέχει εκτός από την κλασική μαγνητική ένδειξη, υπηρεσίες όπως μέτρηση θερμοκρασίας (thermometer) και μέτρηση κλίσης (inclinometer).

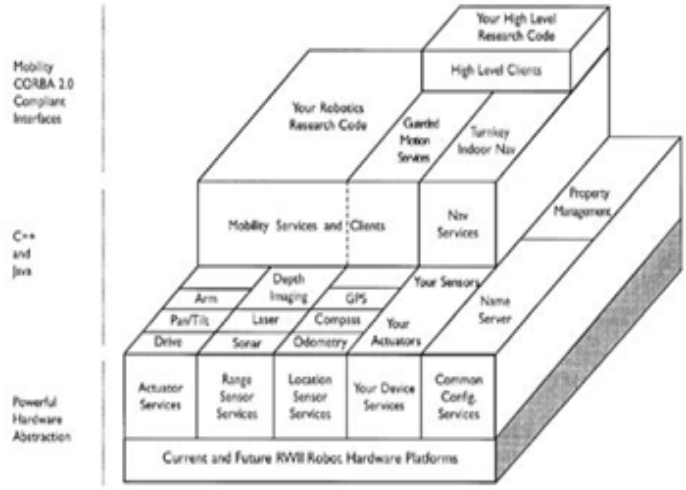
2.3.4 Κάμερα

Σημαντική συσκευή για την πλοήγηση έντροχου ρομπότ αποτελεί η κάμερα. Με τη βοήθεια της είναι δυνατή η σύλληψη του χώρου σε φωτογραφίες ή βίντεο. Οι φωτογραφίες και το βίντεο επεξεργάζονται και παρέχουν σημαντικά στοιχεία για την πορεία του έντροχου ρομπότ. Με την κάμερα είναι δυνατή η ανίχνευση τροχιάς και η αναγνώριση εμποδίων ή άλλων έντροχων ρομπότ. Η κάμερα που χρησιμοποιείται από το ATRV-Mini είναι η SONY EVI-D30. Λειτουργεί σε NTSC Color Standard, έχει 12X ZOOM και διαθέτει δύο μοτέρ που της επιτρέπουν κίνηση στον οριζόντιο άξονα με άνοιγμα 200^0 με μέγιστη ταχύτητα 800/sec και στον κατακόρυφο άξονα με άνοιγμα 50^0 με μέγιστη ταχύτητα 50^0/sec. Η SONY EVI-D30 έχει την ικανότητα αναγνώρισης και παρακολούθησης κίνησης. [3]

2.4 ΤΟ ΛΟΓΙΣΜΙΚΟ MOBILITY ΤΟΥ ATRV-MINI

Το MOBILITY είναι ένα αντικειμενοστρεφές εργαλείο που παρέχεται από την RWI για την δημιουργία προγραμμάτων ελέγχου για συστήματα ενός ή περισσοτέρων έντροχων ρομπότ. Αποτελείται από ένα σύνολο λογισμικών εργαλείων, το

αντικείμενο του μοντέλου του έντροχου ρομπότ, βασικές μονάδες (modules) ελέγχου του έντροχου ρομπότ (κίνηση, sonar, κάμερα, ΠΣΣ) και ένα αντικειμενοστρεφές περιβάλλον εργασίας για την απλοποίηση της ανάπτυξης κώδικα.



Σχήμα 2.3: Το περιβάλλον του Mobility της RWI [1]

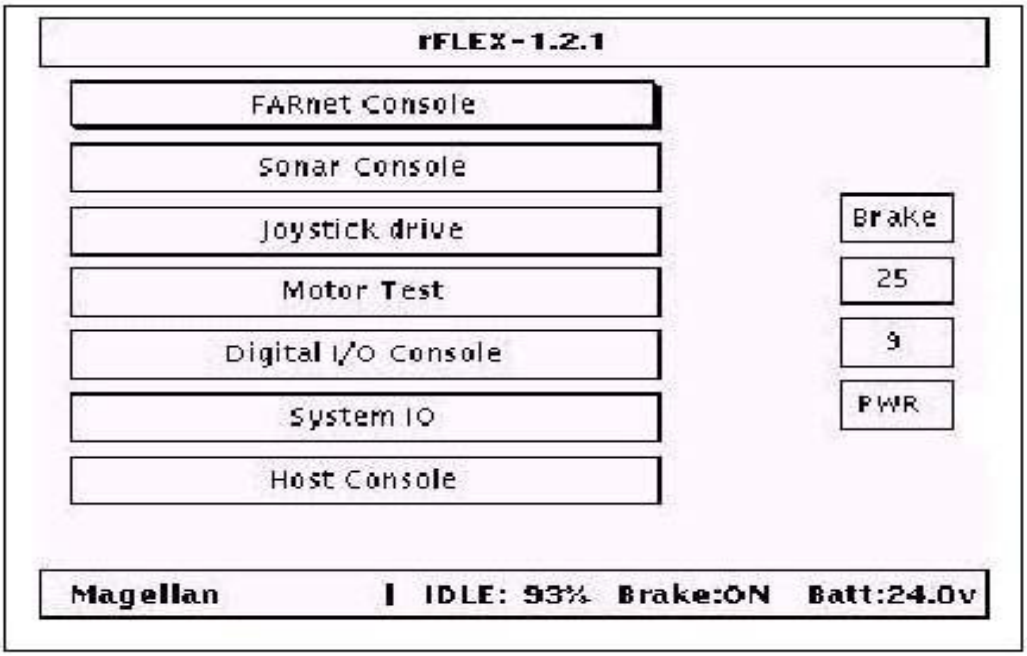
Όπως φαίνεται στο Σχήμα 2.3, το Mobility προσδιορίζει το αντικείμενο του έντροχου ρομπότ χρησιμοποιώντας το CORBA standard, το οποίο του δίνει τη δυνατότητα να υποστηρίζει πολλές γλώσσες προγραμματισμού σε διάφορες πλατφόρμες. Το περιβάλλον του Mobility επιτρέπει στο χρήστη να τροποποιήσει βασικά μέρη του συστήματος του ρομπότ και να προσθέσει νέα, ανάλογα με τις ανάγκες του. Το αντικείμενο του έντροχου ρομπότ αποτελείται από μια σειρά άλλων αντικειμένων. Καθένα από αυτά αντιπροσωπεύει μέρη του ρομπότ, όπως τους αισθητήρες και το μηχανισμό κίνησης. Τα αντικείμενα αυτά μπορούν να τροποποιηθούν ή να χρησιμοποιηθούν ως συναρτήσεις σε νέους αλγορίθμους. Το Mobility υποστηρίζει γλώσσες προγραμματισμού όπως η Java και η C++.

2.5 ΣΥΣΤΗΜΑ ΕΛΕΓΧΟΥ ΤΟΥ ATRV-MINI MINI, RFLEX

Το rFlex είναι το σύστημα ελέγχου του έντροχου ρομπότ και των περιφερειακών του που λειτουργεί χωρίς την χρήση υπολογιστικής μονάδας και βρίσκεται πάνω στο ρομπότ. Αποτελείται από ένα απλό αλληλεπιδραστικό περιβάλλον εργασίας με το οποίο πραγματοποιείται διαχείριση, διαμόρφωση και διάγνωση των περιφερειακών. Οι κύριες λειτουργίες του rFlex (Σχήμα 2.4) είναι:

- Εκκίνηση – Τερματισμός λειτουργίας του ρομπότ
- Ορισμός του τύπου του δικτύου στο οποίο μετέχει το ρομπότ
- Ενεργοποίηση και έλεγχος των αισθητήρων υπερήχων
- Ενεργοποίηση της οδήγησης με χειριστήριο
- Έλεγχος των μοτέρ
- Ενεργοποίηση ή απενεργοποίηση των φρένων
- Ένδειξη στάθμης μπαταρίας
- Μεταφορά στην οθόνη του λειτουργικού συστήματος
- Έλεγχος των θυρών επικοινωνίας

Οι πληροφορίες παρέχονται στον χρήστη μέσω μιας οθόνης υγρών κρυστάλλων που βρίσκεται στο πίσω μέρος της κορυφής του έντροχου ρομπότ. Η εναλλαγή μεταξύ των λειτουργιών πραγματοποιείται με ένα αέναο διακριτό ποτενσιόμετρο.



Σχήμα 2.4: Η κεντρική οθόνη λειτουργιών του rFlex [1]

ΚΕΦΑΛΑΙΟ 3 : ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΦΩΝΗΣ

3.1 ΓΕΝΙΚΑ ΠΕΡΙ ΑΝΑΓΝΩΡΙΣΗΣ ΦΩΝΗΣ

Οι τεχνολογίες αναγνώρισης φωνής επιτρέπουν σε υπολογιστές οι οποίοι είναι εφοδιασμένοι με μικρόφωνα να ερμηνεύσουν την ανθρώπινη φωνή, π.χ. για εγγραφή ή μέθοδο ελέγχου.

Οι εφαρμογές της αναγνώρισης φωνής έχουν να κάνουν είτε με μείωση του κόστους (αυτές δηλαδή οι οποίες αντικαθιστούν τους ανθρώπους, με αναγνωριστές φωνής), είτε με δημιουργία νέων δυνατοτήτων και πιο γρήγορων υπηρεσιών, όπως με πρόσβαση σε βάσεις δεδομένων και πληροφοριών που έχουν σχέση π.χ. με κλείσιμο αεροπορικών θέσεων, δελτία καιρού, χρηματιστήριο, φωνητική δακτυλογράφηση, υπηρεσίες τραπεζικών συναλλαγών μέσω φωνής και με πλήθος άλλες εφαρμογές. Επίσης έχουν ευρεία εφαρμογή τα τελευταία χρόνια στον τομέα της ρομποτικής, του αυτοματισμού και γενικότερα της αλληλεπίδρασης ανθρώπου – μηχανής, ενώ πολύ πρόσφατα χρησιμοποιήθηκε σε αυτοκινητοβιομηχανίες για πλοήγηση του οχήματος και ενεργοποίηση μηχανισμών μέσω φωνητικών εντολών.

Τι είναι όμως αναγνώριση φωνής; Η αναγνώριση φωνής (*speech recognition*) είναι η διαδικασία μετατροπής ενός ακουστικού σήματος, που λαμβάνεται μέσω μικροφώνου ή τηλεφωνικής γραμμής, σε μια ακολουθία λέξεων. Οι αναγνωρισμένες λέξεις μπορούν να είναι τα τελικά αποτελέσματα μιας εφαρμογής, όπως εντολές για έλεγχο ή εισαγωγή δεδομένων. Μπορούν επίσης να χρησιμοποιηθούν και ως είσοδο για μετέπειτα επεξεργασία, προκειμένου να επιτευχθεί η κατανόηση. Έχουν γίνει κατά καιρούς διάφορες προσπάθειες προσέγγισης στην αναγνώριση φωνής χρησιμοποιώντας ποικίλες τεχνολογίες και μοντέλα αναγνώρισης. Οι πιο επιτυχημένες προσεγγίσεις θεωρούνται αυτή των Νευρωνικών Δικτύων καθώς και αυτή που βασίζεται στην τεχνολογία της Στατιστικής Αναγνώρισης Προτύπων. [4]

Στην εργασία αυτή θα ασχοληθούμε με τη Στατιστική Αναγνώριση. Συνοπτικά, μπορούμε να πούμε ότι στην προσέγγιση αυτή το σύστημα κατασκευάζει ένα δίκτυο, που υλοποιεί τη γραμματική και σε κάθε επιτρεπόμενη πρόταση – λέξη αντιστοιχίζεται ένα σύνολο από κρυφά μοντέλα Markov (Hidden Markov Models, HMMs). Όταν πρόκειται να αναγνωριστούν νέα δεδομένα φωνής, το σύστημα υπολογίζει τις πιθανότητες τα δεδομένα αυτά να είχαν παραχθεί με βάση τα αποθηκευμένα HMMs. Το αποτέλεσμα της αναγνώρισης είναι η πρόταση με τη μεγαλύτερη πιθανότητα. Η δομή των HMMs, καθώς και οι αλγόριθμοι εκπαίδευσης που έχουν αναπτυχθεί για τον καθορισμό των παραμέτρων αναγνώρισης, παρέχουν υψηλές επιδόσεις σε εφαρμογές που λειτουργούν ανεξάρτητα από τον ομιλητή, εφαρμογές συνεχούς ομιλίας και μεγάλων λεξιλογίων. Περισσότερα θα αναλυθούν σε επόμενη παράγραφο.

3.1.1 Ιστορική Ανασκόπηση

Η αυτόματη μηχανική αναγνώριση φωνής είναι ο στόχος έρευνας για πάνω από τέσσερεις δεκαετίες και ενέπνευσε τέτοια θαύματα επιστημονικής φαντασίας όπως ο υπολογιστής HAL στου Stanley Kubrick’s το γνωστό έργο - A *space odyssey* και το

ρομπότ R2D2 στο κλασσικό έργο *Star Wars του* George Lucas. Όμως, παρά την αίγλη του να σχεδιάσει κανείς μία έξυπνη μηχανή που μπορεί να αναγνωρίσει τον προφορικό λόγο και να καταλάβει την έννοια του και παρά τις τεράστιες ερευνητικές προσπάθειες που έγιναν στην προσπάθεια να δημιουργηθεί μία τέτοια μηχανή, βρισκόμαστε ακόμη μακριά από το να επιτύχουμε τον επιθυμητό στόχο, δηλαδή μία μηχανή που θα μπορεί να καταλάβει μία συζήτηση επί οποιουδήποτε θέματος από όλους τους ομιλητές και σε όλα τα περιβάλλοντα.

Παρόλα αυτά, για να εκτιμήσουμε το μέγεθος της προόδου που έχει επιτευχθεί σε αυτή την περίοδο των τεσσάρων δεκαετιών, αξίζει τον κόπο να κάνουμε μία σύντομη ανασκόπηση από μερικά φωτεινούς σταθμούς της ερευνητικής πορείας.

Η μελέτη της αυτόματης αναγνώρισης και καταγραφής φωνής άρχισε το 1936 στα εργαστήρια ATT &T's Bell Labs. Συγχρόνως χρηματοδοτήθηκαν και εκτελεστήκαν και άλλες έρευνες από τα Πανεπιστήμια και την Κυβέρνηση των ΗΠΑ (κυρίως από το Στρατό και το DARPA (Υπηρεσία Προχωρημένων Ερευνητικών Προγραμμάτων Άμυνας). Αλλά μόνον στις αρχές της 10ετίας του 1980 η τεχνολογία έφτασε στην εμπορική αγορά.

Η πρώτη εταιρεία που παρουσίασε εμπορικό προϊόν ήταν η Covox το 1982. Η Covox έφερε τον ψηφιακό ήχο (μέσω το the Voice Master, Sound Master, και The Speech Thing) στο Commodore 64, Atari 400/800 και τελικά στο IBM PC στα μέσα της δεκαετίας του 1980. Μαζί με αυτήν την εισαγωγή ήχου στους υπολογιστές ήρθε και η αναγνώριση φωνής [9].

Πιο αναλυτικά, οι πρώτες απόπειρες για επινόηση συστημάτων αυτόματης αναγνώρισης φωνής από μηχανή έγιναν τη δεκαετία του 1950 όπου διάφοροι ερευνητές προσπάθησαν να διερευνήσουν τις βασικές ιδέες της ακουστικής-φωνητικής. Το 1952 στα Εργαστήρια Bell οι Davis, Biddullph και Balashek δημιούργησαν ένα σύστημα για μεμονωμένη ψηφιακή αναγνώριση ενός μόνο ομιλητή. Αυτό το σύστημα βασιζόταν αρκετά στην μέτρηση φασματικών αντηχήσεων μέσα στην περιοχή φωνηέντων του κάθε ψηφίου. Σε μία ανεξάρτητη προσπάθεια στο RCA Laboratories, το1956, οι Olson και Belar προσπάθησαν να αναγνωρίσουν δέκα διαφορετικές συλλαβές ενός ομιλητή ενσωματωμένες σε δέκα μονοσύλλαβες λέξεις. Το σύστημα πάλι βασιζόταν σε φασματικές μετρήσεις κυρίως μέσα στις περιοχές φωνηέντων. Το 1959 στο University College οι Fry και Denes προσπάθησαν να δημιουργήσουν έναν αναγνωριστή φωνημάτων που θα αναγνώριζε 4 φωνήεντα και 9 σύμφωνα. Χρησιμοποίησαν έναν αναλυτή φασμάτων και έναν ταυτιστή μοτίβων για να πάρουν την απόφαση αναγνώρισης. Μία νέα πλευρά αυτής της έρευνας ήταν η χρήση στατιστικών πληροφοριών για μία επιτρεπτή ακολουθία φωνημάτων στην Αγγλική (μία στοιχειώδης μορφή γλωσσικής σύνταξης) για τη βελτίωση της γενικής ακρίβειας φωνημάτων για λέξεις που αποτελούνται από δύο ή περισσότερα φωνήματα. Άλλη μία προσπάθεια που αξίζει να σημειώσουμε σε αυτή την περίοδο ήταν ο αναγνωριστής φωνηέντων των Forgie και Forgie ο οποίος κατασκευάστηκε το 1959 στο MIT Lincoln Laboratories και στον οποίο δέκα φωνήεντα ενσωματωμένα σε μορφή /b/- φωνήεν - /t/ αναγνωρίσθηκαν με ένα τρόπο ανεξάρτητου ομιλητή. Πάλι ένας αναλυτής φίλτρου χρησιμοποιήθηκε για να παρέχει φασματικές πληροφορίες και έγινε ένας υπολογισμός χρονικής μεταλλαγής των αντηχήσεων της φωνητικής οδού ώστε να αποφασισθεί ποιο ήταν το φωνήεν.

Στη δεκαετία του 1960 ήρθαν στην επιφάνεια και δημοσιεύθηκαν διάφορες βασικές ιδέες όσον αφορά την αναγνώριση φωνής. Όμως, η δεκαετία ξεκίνησε με πολλά Ιαπωνικά εργαστήρια να μπαίνουν στην έρευνα της αναγνώρισης και να κατασκευάζουν ειδικής σκοπιμότητας υλικό ως μέρος των συστημάτων τους. Ένα από τα πρώτα Ιαπωνικά συστήματα που περιγράφεται από τους Suzuki και Nakata

του ραδιοφωνικού ερευνητικού εργαστηρίου στο Τόκυο ήταν ένας αναγνωριστής φωνηέντων για υλικό. Ένας περίπλοκος αναλυτής φάσματος τραπέζης φίλτρων χρησιμοποιήθηκε μαζί με λογικό και συνέδεε τις εξόδους κάθε καναλιού του αναλυτή φάσματος με ένα κύκλωμα απόφασης φωνηέντων, και ένα σχέδιο πλειοψηφίας αποφάσεων χρησιμοποιήθηκε για να επιλέγει το φωνήεν που αρθρώνεται. Μία άλλη προσπάθεια υλικού στην Ιαπωνία ήταν η εργασία των Sakai και Doshita στο Πανεπιστήμιο του Κιότο το 1962, οι οποίοι κατασκεύασαν έναν αναγνωριστή φωνημάτων υλικού. Ένας διατμηματιστής φωνής για υλικό χρησιμοποιήθηκε με 0-διασταυρωτική ανάλυση διαφορετικών περιοχών της ομιλουμένης εισόδου για να παρέχει την έξοδο αναγνώρισης. Μία Τρίτη Ιαπωνική προσπάθεια ήταν ο ψηφιακός αναγνωριστής υλικού των Nagata και των συνεργατών του στα εργαστήρια NEC το 1963. Αυτή η προσπάθεια ήταν ίσως η πιο αξιοσημείωτη ως η αρχική απόπειρα αναγνώρισης φωνής στα NEC και οδήγησε σε ένα μακροχρόνιο και υψηλά αποδοτικό πρόγραμμα έρευνας.

Στη δεκαετία του 1960 ξεκίνησαν τρία βασικά ερευνητικά προγράμματα τα οποία είχαν μεγάλες επιπτώσεις στην έρευνα και εξέλιξη της αναγνώρισης φωνής τα τελευταία είκοσι χρόνια. Το πρώτο από αυτά τα προγράμματα ήταν οι προσπάθειες του Martin και των συνεργατών του στα εργαστήρια του RCA που άρχισε στο τέλος της δεκαετίας του 1960 για να βρουν ρεαλιστικές λύσεις για τα προβλήματα που σχετίζονται με την έλλειψη ομοιομορφίας σε κλίμακες χρόνου στην φωνή. Ο Martin ανέπτυξε ένα σετ στοιχειωδών μεθόδων για τη χρονική εξομάλυνση, που βασίζονται την ικανότητα να αναγνωρίζει κανείς την αρχή και το τέλος του λόγου και οι οποίες μείωσαν σημαντικά την ποικιλομορφία των σκορ αναγνώρισης. Ο Martin τελικά ανέπτυξε τη μέθοδο και ίδρυσε μία από τις πρώτε εταιρείες, την Threshold η οποία κατασκεύαζε και πουλούσε προϊόντα αναγνώρισης φωνής. Περίπου τον ίδιο καιρό, στη Σοβιετική Ένωση ο Vintsyuk πρότεινε τη χρήση μεθόδων δυναμικού προγραμματισμού για εξίσωση χρόνου όσον αφορά ένα ζεύγος από εκφράσεις φωνής. Αν και η ουσία των εννοιών της δυναμικής διαστροφής χρόνου καθώς και στοιχειώδεις παραλλαγές των αλγορίθμων για συνδεδεμένη αναγνώριση λέξεων ενσωματώθηκαν στην εργασία του Vintsyuk, ήταν τελείως άγνωστη στη Δύση και δεν ήλθε στο φως μέχρι και τα μέσα της δεκαετίας του 1980. Αυτό συνέβη πολύ μετά και αφού επίσημες μέθοδοι προτάθηκαν και χρησιμοποιήθηκαν από όλους.

Ένα τελικά αξιοσημείωτο επίτευγμα στη δεκαετία του 60 ήταν η πρωτοποριακή έρευνα του Reddy στον τομέα συνεχούς αναγνώρισης φωνής με δυναμική παρακολούθηση φωνημάτων. Η έρευνα του Reddy τελικά έδωσε την ώθηση για ένα μακροχρόνιο και πολύ επιτυχημένο ερευνητικό πρόγραμμα αναγνώρισης φωνής στο Carnegie Mellon University (στο οποίο μετακόμισε ο Reddy στα τέλη της δεκαετίας του 1960), το οποίο μέχρι σήμερα, παραμένει παγκόσμιος ηγέτης στα συστήματα της συνεχούς αναγνώρισης φωνής.

Στη δεκαετία του 1970 η έρευνα αναγνώρισης φωνής πέτυχε πολλούς σημαντικούς σταθμούς. Πρώτον, ο τομέας αναγνώρισης μεμονωμένων λέξεων ή ευδιάκριτων εκφράσεων έγινε μία βιώσιμη τεχνολογία που βασίζεται στις μελέτες του Velichko και Zagoruyko στη Ρωσία, του Sakoe και Chiba στην Ιαπωνία, και Itakura στις ΗΠΑ. Οι Ρωσικές μελέτες βοήθησαν να προοδεύσει η χρήση των ιδεών αναγνώρισης μοτίβων στην αναγνώριση φωνής. Η Ιαπωνική έρευνα έδειξε ότι θα μπορούσαν να εφαρμοσθούν επιτυχώς οι μέθοδοι δυναμικού προγραμματισμού και η έρευνα του Itakura έδειξε πως οι ιδέες γραμμικής προβλεπτικής κωδικοποίησης (LPC), οι οποίες είχαν ήδη επιτυχώς χρησιμοποιηθεί στην κωδικοποίηση φωνής χαμηλής αναλογίας θα μπορούσαν να επεκταθούν σε συστήματα αναγνώρισης φωνής με τη χρήση ενός κατάλληλου μέτρου απόστασης βασιζόμενο σε φασματικές παραμέτρους LPC.

Ένας άλλος σταθμός της δεκαετίας του 1970 ήταν η αρχή μίας μακροχρόνιας πολύ επιτυχημένης ομαδικής προσπάθειας για την αναγνώριση φωνής μεγάλου λεξιλογίου στην IBM όπου οι ερευνητές μελέτησαν τρεις διαφορετικές εργασίες για μία περίοδο σχεδόν δύο δεκαετιών, δηλαδή τη New Raleigh Language για απλές αναζητήσεις σε τράπεζες δεδομένων, τη γλώσσα κειμένου με πατέντα λέιζερ για τη μεταγραφή σχεδίων λέιζερ και την αλληλογραφία γραφείου, που ονομάζεται Tangora.

Τελικά, στα εργαστήρια AT&T Bell οι ερευνητές άρχισαν μία σειρά πειραμάτων με σκοπό να κατασκευάσουν συστήματα αναγνώρισης φωνής τα οποία θα ήταν πραγματικά ανεξάρτητα από ομιλητή. Για να επιτευχθεί αυτός ο στόχος χρησιμοποιήθηκε μία ευρεία κλίμακα περίπλοκων αλγορίθμων ώστε να καθορίζεται ο αριθμός διαφόρων μοτίβων που απαιτούνται για να αντιπροσωπεύουν όλες τις παραλλαγές διαφόρων λέξεων σε ένα ευρύ φάσμα πληθυσμού. Αυτή η έρευνα βελτιώθηκε σε μία δεκαετία έτσι ώστε οι τεχνικές για τη δημιουργία μοτίβων ανεξάρτητα ομιλητή είναι τώρα και κατανοητές και ευρέως χρησιμοποιούμενες.

Ακριβώς όπως η αναγνώριση μεμονωμένων λέξεων ήταν το επίκεντρο της έρευνας στη δεκαετία του 1970, το πρόβλημα αναγνώρισης συνδεδεμένων λέξεων ήταν το επίκεντρο της έρευνας στη δεκαετία του 1980. Εδώ ο στόχος ήταν να δημιουργηθεί ένα δυνατό σύστημα ικανό να αναγνωρίσει μία γρήγορα ομιλουμένη σειρά λέξεων που θα βασιζόταν στο ταίριασμα ενός μοτίβου αλληλουχίας ατομικών λέξεων. Μία μεγάλη ποικιλία αλγορίθμων αναγνώρισης συνδεδεμένων λέξεων διαμορφώθηκε και χρησιμοποιήθηκε, που περιλάμβανε την διπλού επιπέδου μέθοδο δυναμικού προγραμματισμού του Sakoe στην Nippon Electric Corporation (NEC) τη μία διαδρομή μέθοδο των Bridle και Brown της Joint Speech Research Unit (JSRU), τη μέθοδο δημιουργίας επιπέδων των Myers και Rabiner στο Bell Labs και τη μέθοδο πλαισίου διαμόρφωσης σύγχρονων επιπέδων του Lee και Rabiner στα Bell Labs. Κάθε μία από αυτές τις βέλτιστες μεθόδους είχε τα δικά της πλεονεκτήματα εφαρμογής, τα οποία διερευνήθηκαν για μία ευρεία κλίμακα εργασιών.

Η έρευνα στον τομέα της φωνής στη δεκαετία του 1980 χαρακτηρίστηκε από μία στροφή προς την τεχνολογία από μεθόδους βασισμένες σε φόρμες ταύτισης σε μεθόδους στατιστικής μοντελοποίησης – ιδιαίτερα η μέθοδος του κρυφού μοντέλου Markov. Αν η μεθοδολογία του κρυφού μοντέλου Markov (HMM) ήταν γνωστή και κατανοητή σε μερικά εργαστήρια (κυρίως στην IBM, Institute of Defense Analyses (IDA) και Dragon Systems) μόνο μετά την ευρεία δημοσίευση των μεθόδων και της θεωρίας των HMM στα μέσα της δεκαετίας του 1980 αυτή η τεχνική εφαρμόσθηκε ευρέως σε σχεδόν κάθε εργαστήριο έρευνας αναγνώρισης φωνής από τον κόσμο.

Μία άλλη «νέα» τεχνολογία η οποία επανήλθε στα τέλη της δεκαετίας του 1980 ήταν η ιδέα της εφαρμογής νευρωνικών δικτύων σε προβλήματα αναγνώρισης φωνής. Τα νευρικά δίκτυα πρωτοπαρουσιάστηκαν στη δεκαετία του 1950 αλλά αρχικά δεν αποδείχθηκαν χρήσιμα γιατί είχαν πολλά πρακτικά προβλήματα. Όμως στη δεκαετία του 1980 προβήκαμε σε μία βαθύτερη κατανόηση των δυνάμεων και των περιορισμών της τεχνολογίας, καθώς επίσης και των σχέσεων της τεχνολογίας με τις κλασσικές μεθόδους διαχωρισμού των σημάτων κατά κατηγορίες. Επίσης προτάθηκαν πολλοί νέοι τρόποι εφαρμογής συστημάτων [8].

Τελικά, η δεκαετία του 1980 ήταν μία δεκαετία στην οποία δόθηκε μεγάλη ώθηση στα συστήματα αναγνώρισης συνεχούς ομιλίας και μεγάλου λεξιλογίου από το Defense Advanced Research Project Agency (DARPA) των ΗΠΑ, το οποίο χρηματοδότησε ένα μεγάλο ερευνητικό πρόγραμμα με σκοπό να επιτύχει υψηλή λεκτική ακρίβεια για το έργο διαχείρισης τράπεζας δεδομένων για αναγνώριση συνεχούς ομιλίας 1000 λέξεων. Μεγάλες ερευνητικές συνεισφορές ήταν: το αποτέλεσμα των προσπαθειών

στο CMU (Το πολύ γνωστό σύστημα SPHINX), BBN με το σύστημα BYBLOS, Lincoln Labs, SRI [10], MIT και AT&T Bell labs [11].

Μία άλλη εταιρεία που ιδρύθηκε το 1982 και της οποίας το τελικό προϊόν έγινε ο αδιαφιλονίκητος ηγέτης στην αγορά αναγνώρισης φωνής ήταν τα συστήματα Dragon. Η εταιρία Scansoft κατασκευάζει σήμερα το προϊόν Dragon Naturally Speaking [9].

Το πρόγραμμα του DARPA συνεχίστηκε στη δεκαετία του 1990, οπότε η έμφαση μετατοπίσθηκε προς το μέτωπο της φυσικής γλώσσας στον αναγνωριστή και η εργασία μετατοπίζεται προς την ανάκληση πληροφοριών αεροπορικών ταξιδιών. Συγχρόνως, η τεχνολογία αναγνώρισης φωνής χρησιμοποιείται όλο και περισσότερο σε τηλεφωνικά δίκτυα για να αυτοματοποιήσει και να βελτιώσει τις υπηρεσίες τηλεφωνητών [8].

3.1.2 Η αυτόματη αναγνώριση φωνής

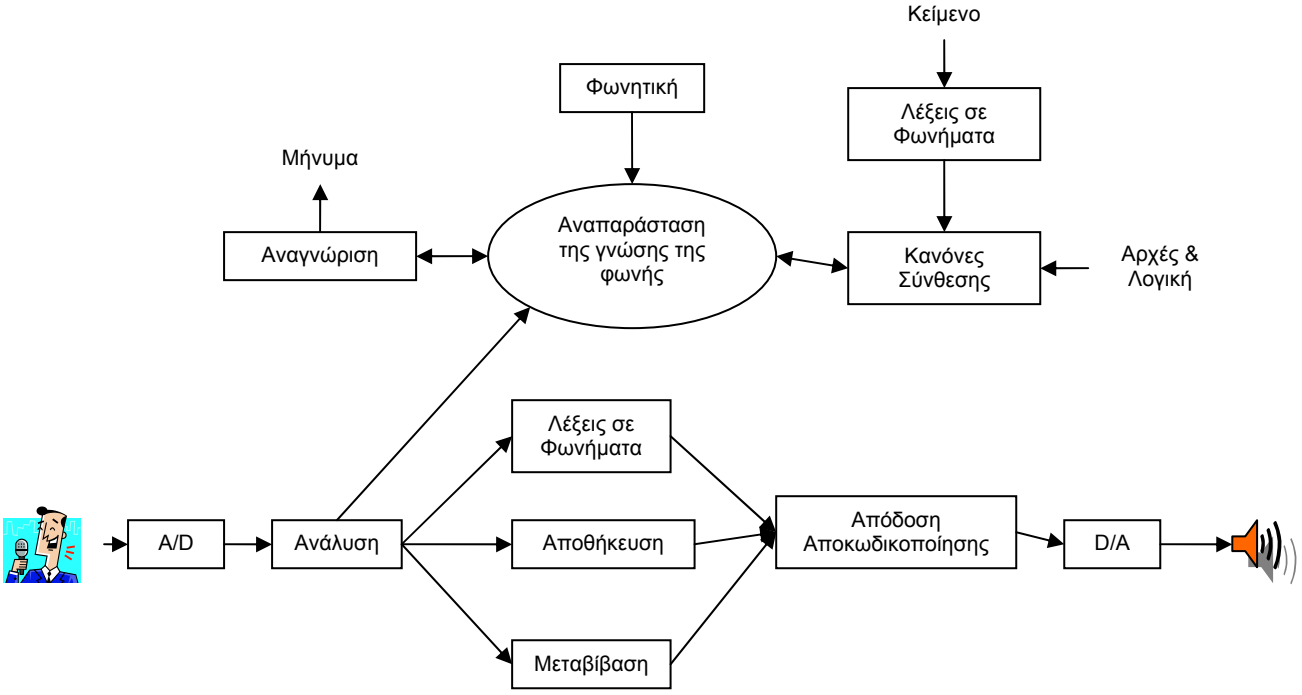
Είναι χρήσιμο να καταλάβουμε ότι η αυτόματη αναγνώριση φωνής είναι μέρος του ευρύτερου τομέα της *επεξεργασίας της φωνής.* Μπορούμε να διακρίνουμε τρεις κύριες περιοχές σε αυτό τον τομέα, αν και υπερκαλύπτονται αρκετά : *κωδικοποίηση, αναγνώριση και σύνθεση.* Το Σχήμα 3.1 παρουσιάζει τη σχέση μεταξύ αυτών των περιοχών. Η βελτίωση και η σύμπηξη της ομιλίας είναι χρήσιμες και για την αναγνώριση και για την κωδικοποίηση. [5]

Ενώ η φωνή μεταφέρει πληροφορίες για το μήνυμα, τον ομιλητή και το περιβάλλον ηχογράφησης, ο σκοπός της αναγνώρισης της φωνής είναι να απομονώσει το μήνυμα. Εξάλλου, στην αναγνώριση ή την επαλήθευση της ταυτότητας του ομιλητή, ο στόχος είναι να απομονώσουμε ειδικές για τον ομιλητή πληροφορίες. Όμως σε όλες αυτές τις περιπτώσεις όπου τα εισερχόμενα σήματα αναλύονται ή κωδικοποιούνται, η κύρια λειτουργία αφορά το διαχωρισμό των σχετικών πληροφοριών από τις μη σχετικές. Στο επίπεδο ανάλυσης σήματος, αυτό περιλαμβάνει τις διαδικασίες βελτίωσης της ομιλίας και αφαίρεσης θορύβου.

3.1.3 Κατηγορίες Συστημάτων Αναγνώρισης

Τα συστήματα αναγνώρισης φωνής μπορούν γενικά να χωριστούν σε κατηγορίες σχετικά με το εάν απαιτούν από το χρήστη να «εκπαιδεύσει» το σύστημα για να αναγνωρίσει τα δικά τους ιδιαίτερα μοτίβα ομιλίας ή όχι, εάν το σύστημα μπορεί να αναγνωρίσει συνεχή ομιλία ή απαιτεί από τους χρήστες να διασπάσουν την ομιλία τους σε μεμονωμένες λέξεις, και εάν το λεξιλόγιο που αναγνωρίζει το σύστημα είναι μικρό (στην τάξη των 10 ή το πολύ 100 λέξεων) ή μεγάλο (χιλιάδων λέξεων). [6]

Πιο συγκεκριμένα, τα συστήματα που απαιτούν σύντομο χρόνο εκπαίδευσης μπορούν (από το 2001 και μετά) να συλλάβουν συνεχή ομιλία με μεγάλο λεξιλόγιο σε φυσιολογικό ρυθμό με ακρίβεια περίπου 98% (κάνοντας λάθος 2 λέξεις στις 100) και διαφορετικά συστήματα που δεν απαιτούν εκπαίδευση μπορούν να αναγνωρίσουν ένα μικρό αριθμό λέξεων (για παράδειγμα, τα 10 ψηφία του δεκαδικού συστήματος) όπως ομιλούνται από τους περισσότερους αγγλόφωνους. Τέτοια συστήματα είναι δημοφιλή για την καθοδήγηση εισερχομένων τηλεφωνημάτων στους προορισμούς τους σε μεγάλους οργανισμούς. [7]



Σχήμα 3.1: Αλληλεπίδραση μεταξύ περιοχών του τομέα επεξεργασίας φωνής [5]

Τα συστήματα αναγνώρισης φωνής, ανάλογα με τις δυνατότητες τους, συχνά επίσης κατηγοριοποιούνται σε συστήματα απομονωμένων λέξεων/φράσεων, συστήματα συνδεδεμένων λέξεων και συστήματα συνεχούς ομιλίας. Τα συστήματα απομονωμένων λέξεων/φράσεων είναι οι πιο περιοριστικοί αναγνωριστές, αλλά μπορούν να λειτουργήσουν ικανοποιητικά σε μια μεγάλη ποικιλία εφαρμογών. Τα συστήματα συνδεδεμένων λέξεων είναι λιγότερο περιοριστικά και αρχίζουν να αποκτούν επίδοση κατάλληλη για μια σειρά από ενδιαφέρουσες εφαρμογές. Οι αναγνωριστές συνεχούς ομιλίας είναι ελάχιστα περιοριστικοί και απαιτητικοί από το χρήστη. Με το χρόνο η επίδοση τους βελτιώνεται και θα μπορούν να χρησιμοποιηθούν σε ιδιαίτερα απαιτητικές εφαρμογές. Είναι λοιπόν θέμα χρόνου να επιτευχθεί η αξιόπιστη επικοινωνία ανάμεσα σε ανθρώπους και μηχανές με στόχο την παροχή βέλτιστων υπηρεσιών στο χρήστη.

Τα συστήματα αναγνώρισης φωνής μπορούν να αναπτυχθούν με δεδομένα εκπαίδευσης που είναι είτε εξαρτημένα από ομιλητή (*speaker-dependent)* ή έχουν συλλεγεί ανεξάρτητα από αυτόν (*speaker-independent*). Η διαφορά έγκειται στο αν τα λεκτικά πρότυπα κατασκευάζονται με ανάλυση των δεδομένων φωνής των ιδίων των χρηστών ή με επεξεργασία δεδομένων που προέρχονται από ένα ανεξάρτητο και αντιπροσωπευτικό δείγμα ομιλητών. Η ποσότητα των δεδομένων εκπαίδευσης που απαιτούνται για συστήματα εξαρτημένα από ομιλητή είναι φυσικά κατά πολύ μικρότερη από αυτήν για κατασκευή συστήματος για ανεξάρτητους ομιλητές. [4]

3.1.4 Αρνητικοί παράγοντες εξέλιξης και τεχνικά προβλήματα

Ο τομέας της επεξεργασίας φωνής έχει γνωρίσει ραγδαία εξέλιξη τα τελευταία χρόνια χάρη στις νέες μεθόδους οι οποίες μειώνουν το υπολογιστικό κόστος υλοποίησης

των αλγορίθμων επεξεργασίας φωνής και στην ραγδαία εξελισσόμενη τεχνολογία υλικού και λογισμικού που παρέχουν νέες δυνατότητες και πιο γρήγορες υπηρεσίες. Οι παραπάνω παράγοντες έχουν καταστήσει εφικτή την έξοδο της αναγνώρισης φωνής από το πειραματικό στάδιο και την εμφάνιση της στην αγορά ως εμπορικό προϊόν. Εμπορικά συστήματα για την αναγνώριση της φωνής είναι διαθέσιμα σε πολλά καταστήματα από τη δεκαετία του 1990. Όμως είναι ενδιαφέρον να σημειώσουμε ότι παρά την εμφανή επιτυχία της τεχνολογίας, ελάχιστοι άνθρωποι χρησιμοποιούν τέτοια συστήματα αναγνώρισης φωνής, ακόμα και σήμερα.

Φαίνεται ότι οι περισσότεροι χρήστες υπολογιστών μπορούν να δημιουργήσουν και να εκδώσουν με ευκολία έγγραφα με το συνηθισμένο πληκτρολόγιο, παρά το γεγονός ότι οι περισσότεροι άνθρωποι μπορεί να μιλούν αρκετά γρηγορότερα από το να δακτυλογραφούν. Επιπροσθέτως μεγάλη χρήση των οργάνων ομιλίας έχει ως αποτέλεσμα μια φωνητική υπερφόρτιση.

Μερικά από τα κύρια τεχνικά προβλήματα στην αναγνώριση φωνής είναι: [8]

- Οι διαφορές μεταξύ των ομιλητών είναι συχνά μεγάλες και δύσκολο να τις υπολογίσουμε. Δεν είναι καθαρό ποια χαρακτηριστικά της ομιλίας είναι ανεξάρτητα του ομιλητή.
- Η ερμηνεία πολλών φωνημάτων, λέξεων και φράσεων είναι ευαίσθητη στο περιεχόμενο. Για παράδειγμα, τα φωνήματα είναι συχνά συντομότερα στις μεγάλες λέξεις από ότι στις μικρές. Οι λέξεις έχουν διαφορετικές σημασίες σε διαφορετικές προτάσεις, π.χ. «Philip lies» θα μπορούσε να ερμηνευθεί ή ως ο Philip είναι ψεύτης ή ως ο Philip είναι ξαπλωμένος σε ένα κρεβάτι.
- Ο τονισμός και η χροιά της φωνής μπορούν να αλλάξουν τελείως τη σωστή ερμηνεία μίας λέξης ή μίας πρότασης «Πήγαινε!» «Πήγαινε;» «Πήγαινε.» μπορούν καθαρά να αναγνωρισθούν από έναν άνθρωπο αλλά όχι τόσο καθαρά από έναν υπολογιστή.
- Οι λέξεις και οι προτάσεις μπορούν να έχουν διάφορες σωστές ερμηνείες, έτσι ώστε στην ουσία ο ομιλητής να αφήνει την εκλογή της σωστής ερμηνείας στον ακροατή.
- Η γραπτή γλώσσα μπορεί να χρειάζεται στίξη σύμφωνα με αυστηρούς κανόνες που δεν παρουσιάζονται τόσο έντονα στην ομιλία, και που είναι δύσκολο να τα συμπεράνεις χωρίς να ξέρεις τη σημασία τους (κόμματα, κατάληψη προτάσεων, εισαγωγικά).

Μία γενική λύση για πολλά από τα ανωτέρω προβλήματα απαιτεί ανθρώπινη γνώση και πείρα και επομένως θα απαιτούσε προηγμένες τεχνολογίες τεχνητής νοημοσύνης για να χρησιμοποιηθεί σε έναν υπολογιστή. Σήμερα, αρκετά Στατιστικά Μοντέλα Αναγνώρισης χρησιμοποιούνται συχνά για την αποσαφήνιση και βελτίωση όσον αφορά την ακρίβεια στην αναγνώριση φωνής.

Η κατανόηση της σημασίας των ομιλουμένων λέξεων θεωρείται από μερικούς ως χωριστός τομέας, ο τομέας της κατανόησης της φυσικής γλώσσας. Όμως υπάρχουν πολλά παραδείγματα προτάσεων που ακούγονται μεν το ίδιο αλλά μπορούν μόνο να αποσαφηνισθούν με αναφορά στο κείμενο. Ένα πολύ γνωστό μπλουζάκι που φορούσαν οι ερευνητές της εταιρείας υπολογιστών Apple δήλωνε: “I helped Apple wreck a nice beach”.

Επιπλέον, μια από τις πιο δύσκολες πλευρές της έρευνας, όσον αφορά την αναγνώριση φωνής από μηχανή, είναι η εμπλοκή πολλών επιστημών και η τάση πολλών ερευνητών να εφαρμόζουν μονολιθική προσέγγιση σε ατομικά προβλήματα.

Είναι σκόπιμο εδώ να παραθέσουμε τις επιστήμες που έχουν χρησιμοποιηθεί για ένα η περισσότερα προβλήματα αναγνώρισης φωνής [8]:

1. **Επεξεργασία Σημάτων –** Η διαδικασία να εξάγουμε σχετικές πληροφορίες από ένα σήμα φωνής με αποτελεσματικό σταθερό τρόπο. Περιλαμβανόμενος στην επεξεργασία σήματος είναι ο τύπος της φασματικής ανάλυσης που χρησιμοποιείται για να χαρακτηρίσουμε τις μεταβλητές ιδιότητες του σήματος φωνής καθώς και διαφόρους τύπους επεξεργασίας σημάτων (και μετεπεξεργασία) ώστε να κάνουμε το σήμα φωνής δυνατό στο περιβάλλον ηχογράφησης (βελτίωση σήματος).

2. **Φυσική (ακουστική) –** Η επιστήμη της κατανόησης της σχέσης μεταξύ του φυσικού σήματος φωνής και των φυσιολογικών μηχανισμών (τον ανθρώπινο φωνητικό μηχανισμό) που παρήγαγαν την φωνή και με τους οποίους η φωνή γίνεται αντιληπτή (τον ανθρώπινο ακουστικό μηχανισμό).

3. **Αναγνώριση μοτίβων –** Το σύνολο αλγορίθμων που χρησιμοποιείται για να συγκεντρώσουμε τα δεδομένα ώστε να δημιουργήσουμε με ένα η περισσότερα πρωτότυπα μοτίβα ενός συνόλου δεδομένων και να ταιριάσουμε (συγκρίνουμε) ένα ζεύγος μοτίβων με βάση τις μετρήσεις χαρακτηριστικών των μοτίβων.

4. **Θεωρία επικοινωνίας και πληροφοριών –** Οι μέθοδοι για τον υπολογισμό παραμέτρων των στατιστικών μοντέλων. Οι μέθοδοι για να εντοπίσουμε την παρουσία ειδικών μοτίβων ομιλίας, το σετ σύγχρονης κωδικοποίησης και αποκωδικοποίησης αλγορίθμων (συμπεριλαμβάνοντας το δυναμικό προγραμματισμό, μαζικούς αλγορίθμους, και αποκωδικοποίηση Viterbi) που χρησιμοποιείται για να ερευνήσουμε ένα μεγάλο αλλά ορισμένο πλέγμα για τον καλύτερο δρόμο στον οποίο αντιστοιχεί «καλύτερη» αναγνωρίσιμη ακολουθία λέξεων.

5. **Γλωσσολογία –** Οι σχέσεις μεταξύ ήχων (η φωνολογία), λέξεων μίας γλώσσα (η σύνταξη), σημασία των ομιλουμένων λέξεων (η σημαντική) και η αίσθηση που προέρχεται από την έννοια (η πραγματική). Σε αυτή την επιστήμη περιλαμβάνονται η μεθοδολογία της γραμματικής και η ανάλυση της γλώσσας.

6. **Φυσιολογία –** Η κατανόηση των υψηλότερης τάξης μηχανισμών μέσα στο ανθρώπινο κεντρικό νευρικό σύστημα που επιφέρουν την παραγωγή φωνής και αντίληψης στα ανθρώπινα όντα. Πολλές σύγχρονες τεχνικές προσπαθούν να ενσωματώσουν αυτό τον τύπο γνώσης μέσα στο πλαίσιο των τεχνητών νευρωνικών δικτύων (τα οποία εξαρτώνται πολύ από μερικές από τις ανωτέρω επιστήμες).

7. **Επιστήμη των υπολογιστών –** Η μελέτη των αποτελεσματικών αλγορίθμων που θα εφαρμόσουν και στο λογισμικό και στο υλικό, τις διάφορες μεθόδους που χρησιμοποιούνται σε ένα πρακτικό σύστημα αναγνώρισης φωνής.

8. **Ψυχολογία –** Η επιστήμη της κατανόησης των παραγόντων που θα καταστήσουν ικανή μία τεχνολογία να χρησιμοποιηθεί από ανθρώπινα όντα σε πρακτικές εργασίες.

3.2 ΕΦΑΡΜΟΓΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΦΩΝΗΣ ΣΤΗ ΡΟΜΠΟΤΙΚΗ

Όπως προαναφέρθηκε, η αναγνώριση φωνής έχει πολλές εφαρμογές, και ποικίλα συστήματα έχουν αναπτυχθεί για να καλύψουν κάθε νέα ανάγκη που δημιουργείται, ή ακόμα και να δημιουργήσουν νέες ανάγκες με τις δελεαστικές ευκολίες που παρέχουν κυρίως σε θέματα ταχύτητας και ευελιξίας.

Η εφαρμογή της αναγνώρισης φωνής στον τομέα της ρομποτικής έχει αρκετά χρόνια που έχει ξεκινήσει με σχετική επιτυχία καθώς ο έλεγχος του ρομπότ ξεφεύγει από τον κλασικό έλεγχο μέσω του πληκτρολόγιου, κάτι που το καθιστούσε καθόλου ευέλικτο.

Κρίνεται εδώ σκόπιμο να παραθέσουμε κάποια ενδεικτικά παραδείγματα εφαρμογής τεχνικών αναγνώρισης σε ρομποτικά οχήματα και όχι μόνο. Τα παραδείγματα έχουν επιλεγεί από την τελευταία κυρίως δεκαετία για να φανεί το εύρος δυνατοτήτων που πλέον παρέχει η εφαρμογή τέτοιων τεχνικών στα ρομπότ.

3.2.1 Το ρομπότ AIBO της SONY

Η Sony για αντικατάσταση του ρομπότ AIBO [ERS-110] εισήγαγε τον Οκτώβρη του 2000 το "AIBO 2^{ης} γενιάς" [ERS-210] (Σχήμα 3.2) με μεγαλύτερες δυνατότητες να εκφράσει συναισθήματα και πιο οικεία επικοινωνία με τους ανθρώπους.

Το νέο [ERS-210] είχε αρκετές νέες δυνατότητες και χαρακτηριστικά τα οποία είχαν ζητηθεί από πολλούς μέσα στα οποία ήταν και η αναγνώριση φωνής (αναγνώριση απλών λέξεων, του ονόματός του).

Από τα κύρια χαρακτηριστικά του "AIBO" [ERS-210] είναι η αυξημένη ικανότητα επικοινωνίας όταν χρησιμοποιείται με το "AIBO Life". Έτσι μπορεί να αναγνωρίζει και να αντιδρά στο άκουσμα του ονόματός του, και να αναγνωρίζει μέχρι περίπου 50 απλές λέξεις. Μπορεί επίσης να μιμηθεί τον τονισμό μιας φωνής που ακούει χρησιμοποιώντας τη δική του "γλώσσα AIBO".

Σχήμα 3.2: Το Ρομπότ Ψυχαγωγίας "AIBO", μοντέλο ERS-210 [12]

Επίσης, η SONY εισήγαγε τελευταία το ERS-7 (Σχήμα 3.3), το τελευταίο της μοντέλο, που είναι κοινωνικό, αυτόνομο, έξυπνο, συναισθηματικό και επικοινωνιακό περισσότερο από κάθε άλλο προηγούμενο μοντέλο. Καταλαβαίνει και ανταποκρίνεται στην ανθρώπινη φωνή, ενώ επικοινωνεί με τον κάτοχο του με ποικίλους τρόπους ένας από τους οποίους είναι και τα κινητά τηλέφωνα.

Σχήμα 3.3: Το νέο μοντέλο SONY AIBO ERS -7 [12]

3.2.2 Το ρομπότ PaPeRo της NEC

Το R100 είναι ένα πρωτότυπο ρομπότ που ανέπτυξε η NEC. Έχει δυνατότητες αναγνώρισης εικόνας και φωνής, καθώς και τεχνολογίες επικοινωνίας με το διαδίκτυο, και μηχατρονική, αναγνωρίζει πρόσωπα και αντιλαμβάνεται προφορικές εντολές, ενώ κινείται απαλά μέσα στο σπίτι, αποφεύγοντας εμπόδια όπως καρέκλες και τραπέζια.

Το νέο ρομπότ της NEC λέγεται PaPeRo (Σχήμα 3.3) και είναι μικρότερο, ελαφρύτερο και με περισσότερα χαρακτηριστικά από το αρχικό 'Personal Robot R100'. Η αρχιτεκτονική του PaPeRo του επιτρέπει να βρίσκεται και μόνο του εκτός εργαστηρίου, όπου μπορεί να βελτιώσει την ικανότητα για αναγνώριση και τη σχέση του με ανθρώπους.

Σχήμα 3.4: Το Προσωπικό Ρομπότ PaPeRo της NEC [13]

Το PaPeRo είναι το ακρωνύμιο των λέξεων "Partner-type Personal Robot". Είναι ένα πρωτότυπο "Προσωπικό Ρομπότ" που οι κατασκευαστές του ισχυρίζονται ότι επικοινωνεί με τους ανθρώπους ως ένα μέλος της οικογένειας. Από το 1997, που τα εργαστήρια της NEC ασχολούνται με την ανάπτυξη "Προσωπικών Ρομπότ" δημιούργησαν το πρώτο πρωτότυπο "R100" τον Ιούλιο του 1997 και το διάδοχό του "PaPeRo" τον Ιανουάριο του 2001.

Στο νέο PaPeRo 2003 αναπτύχθηκε μια πλατφόρμα λογισμικού από τη NEC αρχικά μόνο για την ιαπωνική αγορά. Από τα βασικά πλεονεκτήματα που έχει σε σχέση με τους προκατόχους τους είναι: η καλύτερη αναγνωρισιμότητα και επικοινωνία σε πραγματικό χώρο, το μικρότερο μέγεθος του σε συνδυασμό με την πλουσιότερη «γνώση» από το διαδίκτυο, την προσομοίωση ανθρώπινου χαρακτήρα, την

ευκολότερη ανάπτυξη μέσω μιας ειδικής γλώσσας και τέλος η άνετη αλληλεπίδραση που βασίζεται στην εξέλιξη της αμοιβαίας κατανόησης.

3.2.3 Το ρομπότ Robonaut της NASA

Τον Αύγουστο του 2004 παρουσιάστηκε το ρομπότ της NASA (Σχήμα 3.5) με χέρια και δάχτυλα που μοιάζουν ανθρώπινα ενώ γίνεται προσπάθεια να δημιουργηθούν για το ρομπότ αυτό πόδια ή ακόμα και ένα πόδι ή ρόδες.

Από την παρουσίασή του και μετά, πολλές βελτιώσεις έχουν γίνει και το ρομπότ μπορεί να κινείται σε ανώμαλες επιφάνειες, έχει άνεση στο σκαρφάλωμα ενώ μπορεί να "ζήσει" εκτός του σκάφους, και έτσι μπορεί να χρησιμοποιηθεί από ανθρώπους μέσα στο σκάφος οι οποίοι ελέγχοντας το ασύρματα (μέσω αναγνώρισης φωνής) μπορούν να κάνουν επιδιορθώσεις στο σκάφος ή άλλες εξωτερικές δουλειές.

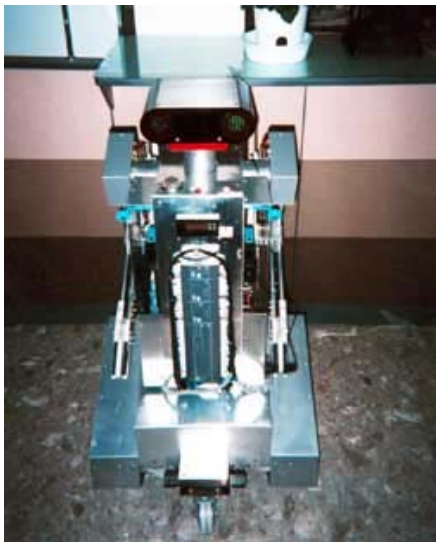
Σχήμα 3.5: Το ρομπότ Robonaut της NASA [14]

3.2.4 Η οικογένεια ρομπότ TMSUK της Thames

Όλα ξεκίνησαν το 1992 όταν η εταιρία Thames ξεκίνησε την πορεία της στην ρομποτική έρευνα και ανάπτυξη. Ξεκίνησε σχεδόν αθώα καθότι η Thames, ακολουθώντας την επιτυχία της ταινίας TRECON ξανάφτιαξε το εργοστάσιο. Το νέο κτίριο είχε χώρο υποδοχής αλλά δεν είχε άτομο στην υποδοχή. "Φτιάξε ένα!" είπαν.

Σχήμα 3.6: TMSUK-1 (1993) [15]

Έτσι, το 1993 η εταιρία Thames ολοκληρώνει την κατασκευή του πρώτου της ρομπότ, TMSUK-1 (Σχήμα 3.6). Αυτό ήταν το πρώτο ρομπότ που χρησιμοποιήθηκε στην υποδοχή. Ο πελάτης χαιρετάται από το ρομπότ μετά την είσοδό του στο κτίριο. Μετά καλείται να ακολουθήσει το ρομπότ στην αίθουσα συσκέψεων, στο γραφείο κλπ. ανάλογα με τις εντολές που δέχεται από τους εργοδότες του.



Σχήμα 3.7: TMSUK-2 (1996) [15]

Ένας πελάτης ήθελε ένα ρομπότ για το γραφείο του. Έτσι δημιουργήθηκε το TMSUK-2 (Σχήμα 3.7) το έτος 1996. Είναι το ίδιο με το αρχικό ρομπότ αλλά μιλάει περισσότερο. Χρησιμοποιεί φωνητική αναγνώριση και μπορεί να αναγνωρίσει 35 τρόπους ομιλίας. Αναγνωρίζει ιδιωματικές προφορές και παραπονιέται ότι πεινάει όταν τελειώνει η μπαταρία του.

Σχήμα 3.8: TMSUK-3 (1997) [15]

Το ρομπότ TMSUK-3 (Σχήμα 3.8), που κατασκευάστηκε το 1997, είναι το πρώτο που χρησιμοποίησε δίκτυο κινητής τηλεφωνίας για να δεχτεί εντολές από τον χειριστή του. Έτσι ο χειριστής μπορεί να βρίσκεται αρκετά χιλιόμετρα μακριά.

Τέλος, το μοντέλο TMSUK-04 (Σχήμα 3.9), που κατασκευάστηκε το 1999 και παρουσιάστηκε στο Λονδίνο το Φεβρουάριο του 2002, είναι παρόμοιο με το TMSUK 3. Χρησιμοποιεί σήματα PHS για να στείλει εντολές, αλλά έχει μεγαλύτερους βαθμούς ελευθερίας και μπορεί να χρησιμοποιηθεί από μία πιο σύνθετη μονάδα ελέγχου.

Σχήμα 3.9: TMSUK-04 (1999) [15]

ΚΕΦΑΛΑΙΟ 4: ΚΡΥΦΑ ΜΟΝΤΕΛΑ MARKOV

4.1 ΜΟΝΤΕΛΑ ΣΗΜΑΤΩΝ

Γενικά, οι διαδικασίες που συμβαίνουν γύρω μας, στον πραγματικό κόσμο, παράγουν εξόδους, οι οποίες μπορούν να χαρακτηρισθούν ως σήματα. Τα σήματα μπορούν να είναι διακριτά (π.χ. χαρακτήρες από μία ορισμένη αλφάβητο, ανύσματα από ένα βιβλίο κώδικα κλπ), η συνεχή (π.χ. δείγματα ομιλίας, μετρήσεις θερμοκρασίας, μουσική κλπ). Η πηγή του σήματος μπορεί να είναι στάσιμη (δηλαδή, οι στατιστικές του ιδιότητες να μην αλλάζουν με το χρόνο). Τα σήματα μπορεί να είναι καθαρά (δηλαδή να έρχονται αυστηρά από μία και μοναδική πηγή), ή μπορεί να διαφθείρονται από άλλες πηγές σημάτων (π.χ. θόρυβο), ή από τη μεταβίβαση παραμορφώσεων, δονήσεων κλπ.

Ένα πρόβλημα βασικού ενδιαφέροντος είναι να χαρακτηρίσουμε τέτοια σήματα χρησιμοποιώντας μοντέλα σημάτων. Υπάρχουν διάφοροι λόγοι για τους οποίους ενδιαφερόμαστε για την εφαρμογή μοντέλων σημάτων. Πρώτα από όλα, ένα μοντέλο σημάτων πρέπει να παρέχει τη βάση για τη θεωρητική περιγραφή ενός συστήματος επεξεργασίας σημάτων το οποίο μπορεί να χρησιμοποιηθεί για την επεξεργασία του σήματος έτσι ώστε να έχουμε το επιθυμητό αποτέλεσμα. Παραδείγματος χάριν εάν ενδιαφερόμαστε να βελτιώσουμε ένα φωνητικό σήμα το οποίο έχει διαφθαρεί από θόρυβο ή παραμόρφωση μεταβίβασης, μπορούμε να χρησιμοποιήσουμε ένα μοντέλο για να σχεδιάσουμε ένα σύστημα το οποίο θα απομακρύνει βέλτιστα το θόρυβο και θα εξαλείψει την παραμόρφωση. Ένας δεύτερος λόγος γιατί τα μοντέλα σημάτων είναι σημαντικά είναι ότι πιθανώς είναι ικανά να μας επιτρέψουν να μάθουμε πολλά για την πηγή του σήματος (δηλαδή την πραγματική διαδικασία που παρήγαγε το σήμα) χωρίς να είναι διαθέσιμη η πηγή. Αυτή η ιδιότητα είναι ιδιαίτερα σημαντική όταν το κόστος για τη λήψη του σήματος από την πραγματική πηγή είναι υψηλό. Σε αυτή την περίπτωση, με ένα καλό μοντέλο μπορούμε να προσομοιώσουμε την πηγή και να μάθουμε όσο το δυνατόν περισσότερα μέσω προσομοιώσεων. Τελικά, ο πιο σημαντικός λόγος γιατί τα μοντέλα είναι σημαντικά είναι ότι συχνά εφαρμόζονται εξαιρετικά καλά στην πράξη, και μας δίνουν τη δυνατότητα να πραγματοποιήσουμε σημαντικά πρακτικά συστήματα – για παράδειγμα, συστήματα πρόβλεψης, συστήματα αναγνώρισης, συστήματα ταυτότητας κλπ, με πολύ αποτελεσματικό τρόπο.

Υπάρχουν διάφορες πιθανές επιλογές για τον τύπο του μοντέλου σημάτων που θα χρησιμοποιηθεί για το χαρακτηρισμό των ιδιοτήτων ενός δεδομένου σήματος. Γενικά μπορεί κανείς να διχοτομήσει τα μοντέλα σημάτων στην τάξη των ντετερμινιστικών μοντέλων και στην τάξη των στατιστικών μοντέλων. Τα ντετερμινιστικά μοντέλα γενικά διερευνούν μερικές γνωστές ειδικές ιδιότητες του σήματος π.χ. ότι το σήμα είναι σε μορφή ημιτονοειδούς κύματος ή εκθετικό άθροισμα κλπ. Σε αυτές τις περιπτώσεις, η προδιαγραφή του μοντέλου είναι γενικά απλή. Το μόνο που απαιτείται είναι να καθορίσουμε τις τιμές των παραμέτρων του μοντέλου (π.χ. το εύρος, τη συχνότητα, τη φάση του ημιτονοειδούς κύματος, εύρη και αναλογίες εκθετικών κλπ). Η δεύτερη ευρεία τάξη μοντέλων σημάτων είναι ένα σετ από στατιστικά μοντέλα στα οποία προσπαθεί κανείς να χαρακτηρίσει μόνον τις στατιστικές ιδιότητες του σήματος. Παραδείγματα τέτοιων στατιστικών μοντέλων περιλαμβάνουν τις Γκαουσιανές διαδικασίες, τις διαδικασίες Poisson, τις διαδικασίες

Markov και τις κρυφές διαδικασίες Markov μεταξύ άλλων. Η βασική προϋπόθεση του στατιστικού μοντέλου είναι ότι μπορεί να χαρακτηριστεί ως παραμετρική τυχαία διαδικασία, και ότι οι παράμετροι της στοχαστικής διαδικασίας μπορούν να υπολογισθούν με ακριβή, καλά οριζόμενο τρόπο.

Για τις εφαρμογές ενδιαφέροντος, δηλαδή την επεξεργασία φωνής, και τα ντετερμινιστικά και τα στοχαστικά μοντέλα μπορεί να έχουν μεγάλη επιτυχία. Σε αυτή την εργασία θα ασχοληθούμε αυστηρά με έναν τύπο του στοχαστικού μοντέλου, δηλαδή το κρυφό μοντέλο Markov (HMM).

Ούτε η θεωρία των HMMs, ούτε οι εφαρμογές της είναι νέα πράγματα. Η βασική θεωρία δημοσιεύθηκε σε μία σειρά κλασσικών εργασιών του Baum και των συνεργατών του στο τέλος της δεκαετίας του 1960 και στις αρχές της δεκαετίας του 1970 και χρησιμοποιήθηκε για εφαρμογές στην επεξεργασία φωνής από τον Baker στο CMU και από τον Jelinek και τους συνεργάτες του στη δεκαετία του 1970 στην IBM. Όμως, ευρεία κατανόηση και εφαρμογή της θεωρίας των HMM στην επεξεργασία φωνής παρουσιάστηκε μόνον τα τελευταία χρόνια.[24]

4.2 ΓΕΝΙΚΗ ΕΠΙΣΚΟΠΗΣΗ ΤΩΝ HMM

Τα Κρυφά Μοντέλα Markov (Hidden Markov Models, HMMs) είναι η τεχνολογία που έχει επικρατήσει στα περισσότερα συστήματα αναγνώρισης φωνής. Ειδικά η δομή των μοντέλων αυτών και ο αλγόριθμος εκπαίδευσης που χρησιμοποιείται για να καθοριστούν οι παράμετροι αναγνώρισης κάνουν την προσέγγιση κατάλληλη για εφαρμογές ανεξαρτήτου ομιλητή, συνεχούς λόγου και μεγάλου λεξιλογίου.

Τα HMMs λειτουργούν βασικά ως ακουστικά μοντέλα τα οποία προσφέρουν μια απεικόνιση από δειγματοληφθέντα σήματα φωνής σε μια σειρά από φωνητικές μονάδες. Στην ιεραρχική δομή του συστήματος, τα HMMs χρησιμοποιούνται επίσης για την απεικόνιση των συγκεκριμένων σειρών φωνητικών μονάδων σε σειρές από λέξεις.
Ο όρος μοντέλο υποδηλώνει το γεγονός ότι επιθυμούμε να μοντελοποιήσουμε το σήμα φωνής. Ένα HMM είναι ένα στατιστικό μοντέλο της διαδικασίας παραγωγής φωνής, το οποίο σημαίνει ότι από μια δεδομένη σειρά από λέξεις μπορεί να παραχθεί το σήμα φωνής. Εκτός από μοντέλο παραγωγής, τα HMMs μπορούν να χρησιμοποιηθούν επίσης για τον υπολογισμό της πιθανότητας του ότι ένα δεδομένο σήμα φωνής παράχθηκε από μια συγκεκριμένη πρόταση. [16]

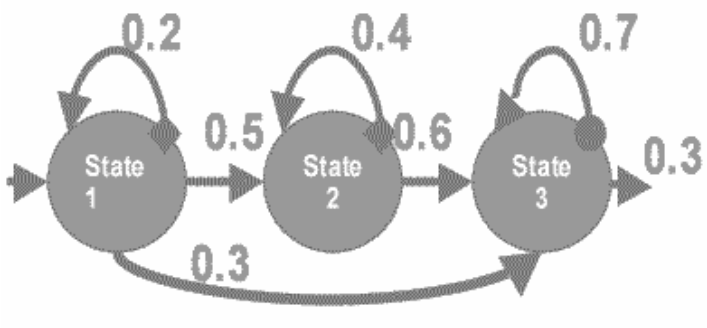
Στην αναγνώριση φωνής κατασκευάζεται ένα δίκτυο που υλοποιεί την γραμματική και για κάθε επιτρεπόμενη πρόταση αντιστοιχίζεται ένα σύνολο από μοντέλα HMMs. Όταν νέα δεδομένα φωνής πρόκειται να αναγνωριστούν, το σύστημα υπολογίζει τις πιθανότητες τα δεδομένα αυτά να είχαν παραχθεί με βάση καθένα από τα αποθηκευμένα HMMs. Το αποτέλεσμα της αναγνώρισης είναι η πρόταση με την μεγαλύτερη πιθανότητα.

Λόγω της χρονικής μεταβλητότητας του σήματος της φωνής, είναι σημαντικό για ένα σύστημα αναγνώρισης φωνής να έχει ένα μηχανισμό με τον οποίο θα μοντελοποιήσει το χρόνο. Στα HMMs, μια στάσιμη αλυσίδα Markov χρησιμοποιείται ως τέτοιος μηχανισμός. Το βασικό χαρακτηριστικό είναι ότι η πιθανότητα μετάβασης από μια κατάσταση σε μια άλλη εξαρτάται μόνο από τις δύο καταστάσεις και όχι από προηγούμενες μεταβάσεις.

Το πρόγραμμα αναγνώρισης φωνής χρησιμοποιεί μοντέλα Markov σε πολλά επίπεδα του συστήματος, συγκεκριμένα:

- Οι προτάσεις χωρίζονται σε λέξεις (δηλαδή η γραμματική).
- Οι λέξεις χωρίζονται σε φωνήματα (δηλαδή το λεξικό).
- Τα φωνήματα χωρίζονται σε καταστάσεις αρχής, μέσης και τέλους.

Κάθε μια από αυτές τις σειρές μοντελοποιούνται με μια αλυσίδα Markov. Στο χαμηλότερο επίπεδο επιτρέπεται η μετάβαση από μια κατάσταση στον εαυτό της. Το γεγονός αυτό επιτρέπει στα HMMs να μοντελοποιούν την διάρκεια ενός ήχου, το οποίο είναι πολύ σημαντικό λόγω της υψηλής μεταβλητότητας στο ρυθμό ομιλίας διαφορετικών ομιλητών. Στο Σχήμα 4.1 φαίνεται η τοπολογία ενός HMM το οποίο αποτελείται από τρεις καταστάσεις και στο οποίο φαίνονται η πιθανότητες μετάβασης από τη μια κατάσταση σε άλλη. Η πιθανότητα εμφάνισης ενός συμβόλου σε μια κατάσταση δε φαίνεται. [17]

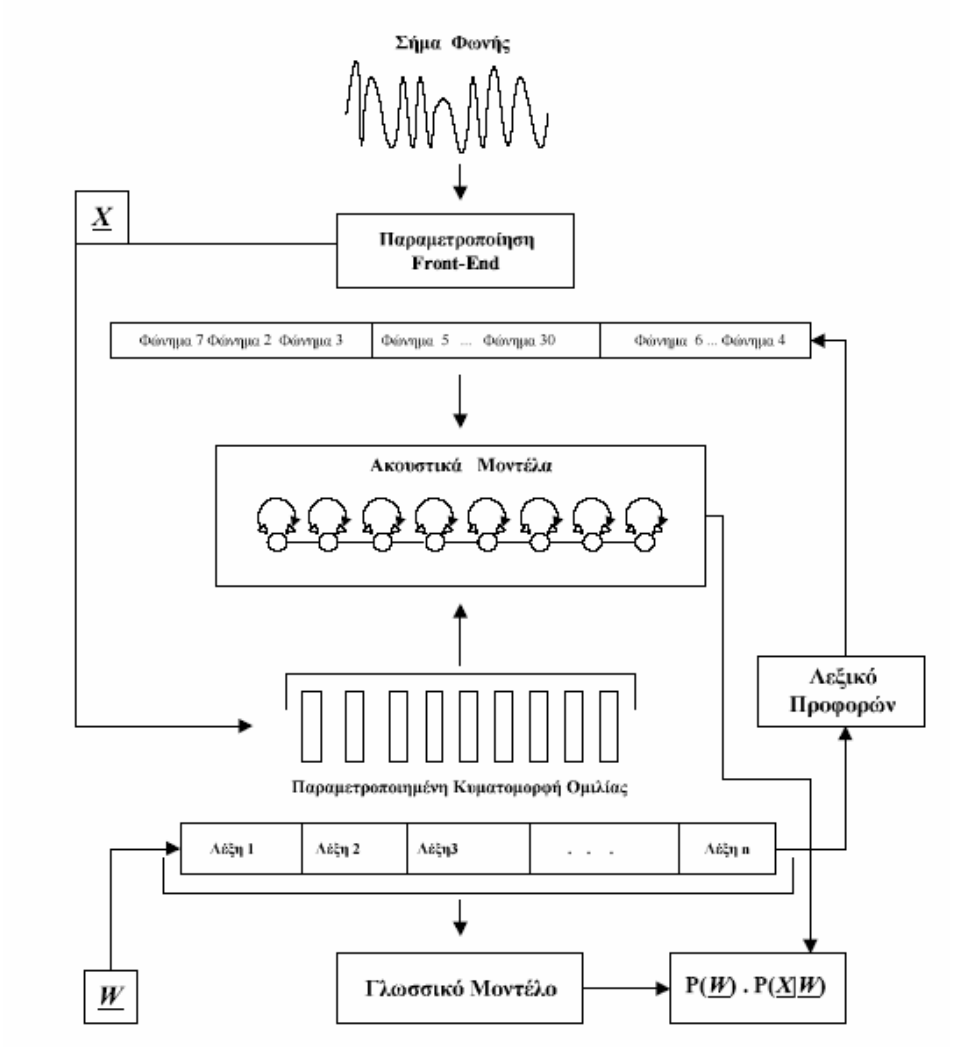


Σχήμα 4.1: Τοπολογία ενός HMM [5]

4.3 ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ

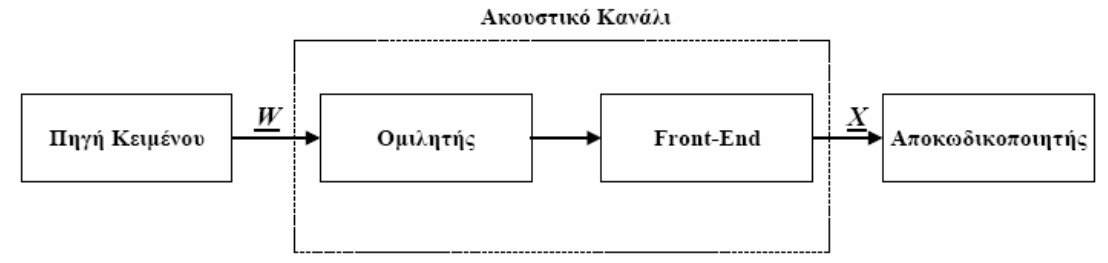
Ορίζοντας το πρόβλημα της αναγνώρισης (αποκωδικοποίησης), λέμε ότι ζητείται να καθοριστεί με βάση κάποιο κριτήριο ότι «εστάλη» (προφέρθηκε) η ακολουθία λέξεων $\underline{W}$ δεδομένης της ακολουθίας διανυσμάτων $\underline{X}$ στην είσοδο του αποκωδικοποιητή. Οι στατιστικές μέθοδοι αναγνώρισης προϋποθέτουν την ύπαρξη κάποιου αντίστοιχου στατιστικού μοντέλου για τον υπολογισμό της πιθανότητας $P(\underline{X}/\underline{W})$. Επιπλέον, ως κριτήριο αποκωδικοποίησης θεωρείται και η ελαχιστοποίηση της πιθανότητας σφάλματος. Η πιθανότητα σφάλματος ελαχιστοποιείται αν αποκωδικοποιήσουμε την ακολουθία εκείνη για την οποία μεγιστοποιείται η α-posteriori πιθανότητα δεδομένου ότι ο αποκωδικοποιητής «έλαβε» την ακολουθία $\underline{X}$.

Τα παραπάνω φαίνονται αναλυτικότερα στο Σχήμα 4.2, όπου περιγράφεται η διαδικασία υπολογισμού της πιθανότητας $P(\underline{X}/\underline{W})$ μιας ακολουθίας λέξεων $\underline{W}$ δεδομένης της ακολουθίας διανυσμάτων $\underline{X}$. Η αρχική πιθανότητα $P(\underline{W})$ καθορίζεται άμεσα από το γλωσσικό μοντέλο. Η πιθανότητα $P(\underline{X}/\underline{W})$ υπολογίζεται χρησιμοποιώντας ένα σύνθετο HMM που αναπαριστά την ακολουθία $\underline{W}$ και αποτελείται από απλά HMM φωνητικά μοντέλα, συνδεδεμένα σειριακά μεταξύ τους σύμφωνα με τις προφορές στο λεξικό προφορών. [17]



Σχήμα 4.2: Αναγνώριση Φωνής με Στατιστικές Μεθόδους [18]

Επίσης, το σύστημα του Σχήματος 4.3 μπορεί να εφαρμοστεί σε μια ευρύτατη ομάδα προβλημάτων που περιλαμβάνει αναγνώριση απομονωμένων λέξεων ή φράσεων, αναγνώριση συνδεδεμένων λέξεων, ακόμη και αναγνώριση συνεχούς ομιλίας. Παρά την αυξημένη πολυπλοκότητα τέτοιων μεθόδων, το βασικό μοντέλο αναγνώρισης προτύπων είναι η βάση σχεδόν όλων των μεθόδων που χρησιμοποιούνται σήμερα.



Σχήμα 4.3: Μοντέλο Αποκωδικοποίησης ενός Ψηφιακού Τηλεπικοινωνιακού Συστήματος [4]

Χρησιμοποιώντας αλυσίδες Markov, μία πρόταση μπορεί να αναλυθεί σε μια σειρά από φωνητικές καταστάσεις. Αν αυτή η ακολουθία είναι γνωστή τότε το πρόβλημα της αναγνώρισης έχει λυθεί. Όμως, όταν γίνεται η αναγνώριση, μόνο η κυματομορφή του σήματος φωνής είναι διαθέσιμη και η φωνητική ακολουθία είναι κρυφή. Αυτή την κρυφή ακολουθία προσπαθούμε να αναγνωρίσουμε. Χρειαζόμαστε έτσι μια σχέση μεταξύ της κυματομορφής και των καταστάσεων της αλυσίδας Markov. Αυτή η σχέση μοντελοποιείται θεωρώντας ότι κάθε κατάσταση παράγει μια ακουστική παρατήρηση ανεξάρτητη από τις παρατηρήσεις που παράγονται από άλλες καταστάσεις. Το μοντέλο που χρησιμοποιείται στην απεικόνιση καταστάσεων σε ήχους καθορίζει τον τύπο των HMMs, συνεχής ή διακριτής πυκνότητας. Σε όλες σχεδόν τις περιπτώσεις, τα πιο ακριβή συνεχής πυκνότητας HMMs προσφέρουν επαρκή ταχύτητα αναγνώρισης.

Τα συνεχής πυκνότητας HMMs μοντελοποιούν τη σχέση μεταξύ των ακουστικών παραθύρων και των καταστάσεων χρησιμοποιώντας μίγματα από Γκαουσιανές. Ενώ είναι περισσότερο ακριβή από τα διακριτής πυκνότητας HMMs, απαιτούν περισσότερο υπολογιστικό χρόνο. Από την άλλη, στα διακριτής πυκνότητας, μια διαδικασία η οποία καλείται κβαντισμός διανύσματος (vector quantization - VQ) χρησιμοποιείται για την απεικόνιση κάθε παραθύρου της κυματομορφής σε ένα σετ από σύμβολα. Τα διακριτά σύμβολα χρησιμοποιούνται στη συνέχεια ως είσοδοι στο HMM. Το πλεονέκτημα αυτής της προσέγγισης είναι ότι κάθε κατάσταση του HMM μοντελοποιεί μόνο ένα πεπερασμένο αριθμό από σύμβολα. Αυτό τυπικά έχει ως αποτέλεσμα ταχύτερο υπολογιστικό χρόνο. Το μειονέκτημα έγκειται στη διαδικασία κβαντοποίησης, η οποία μπορεί να προκαλέσει λάθη. [5]

Όπως περιγράψαμε νωρίτερα, το αποτέλεσμα της αναγνώρισης είναι εκείνη η πρόταση η οποία αντιστοιχεί στο HMM που θα μπορούσε να παράγει την παρατηρηθείσα ακουστική κυματομορφή με την μεγαλύτερη πιθανότητα. Το πρόβλημα είναι ότι είναι σχεδόν ανέφικτο να πιστοποιήσουμε όλες τις δυνατές προτάσεις. Για παράδειγμα, μια ακολουθία από οκτώ ψηφία έχει εκατό εκατομμύρια πιθανές προτάσεις. Για να μειώσουμε την πολυπλοκότητα, το σύστημα της αναγνώρισης χρησιμοποιεί την ιεραρχική δομή των HMMs, δηλαδή τα HMMs προτάσεων είναι αλληλουχία HMMs λέξεων, τα οποία με τη σειρά τους είναι αλληλουχία HMMs φωνημάτων. Κατά τη διάρκεια της αναγνώρισης, οι υπολογισμοί μπορούν να διαμοιραστούν ανάμεσα σε υποθετικές προτάσεις, δηλαδή για παράδειγμα ότι ο ίδιος υπολογισμός απαιτείται για τα εφτά πρώτα ψηφία των προτάσεων «1-2-3-4-5-6-7-8» και «1-2-3-4-5-6-7-7». Στις περισσότερες περιπτώσεις όμως, ακόμα και αν διαμοιράζουμε τους υπολογισμούς, δεν είναι αρκετό για εφαρμογές πραγματικού χρόνου και απαιτείται η διαγραφή των λιγότερο πιθανών υποθέσεων. Η διαδικασία αυτή βελτιώνει το χρόνο της αναγνώρισης αλλά μπορεί να προκαλέσει λάθη στην αναγνώριση. [16]

ΚΕΦΑΛΑΙΟ 5: ΛΟΓΙΣΜΙΚΟ ΑΝΑΓΝΩΡΙΣΗΣ – ΘΕΩΡΗΤΙΚΗ ΚΑΙ ΠΡΑΚΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

5.1 ΛΟΓΙΣΜΙΚΟ ΑΝΑΓΝΩΡΙΣΗΣ

Μεγάλη πρόοδος έχει γίνει στην τεχνολογία αναγνώρισης φωνής τα τελευταία χρόνια και η πρόοδος αυτή έχει γίνει περισσότερο εμφανής στην περιοχή της αναγνώρισης με μεγάλο λεξιλόγιο. Τα σημερινά εργαστηριακά αλλά και εμπορικά συστήματα είναι ικανά να μεταγράφουν σε κείμενο ένα συνεχή λόγο (με ευρύ λεξιλόγιο) από οποιονδήποτε ομιλητή με μέσο όρο λάθους 5-10%. Αρχικά τα συστήματα αναγνώρισης ήταν ικανά να αναγνωρίζουν μια υπαγόρευση κειμένων όπου ο ομιλητής άφηνε μία μικρή παύση ανάμεσα στις λέξεις. Σήμερα ο περιορισμός αυτός δεν υπάρχει πλέον και η έρευνα έχει εστιασθεί στην περαιτέρω αύξηση του ποσοστού ορθής αναγνώρισης.

Το σύστημα μας βέβαια παρ' όλο που ανήκει στην παραπάνω κατηγορία, αυτή δηλαδή των συστημάτων αναγνώρισης συνεχούς λόγου μεγάλου λεξιλογίου, περιορίζεται σε αναγνώριση μεμονωμένων λέξεων από ένα συγκεκριμένο λεξιλόγιο, κάτι που θα παρατεθεί σε επόμενη παράγραφο. Αυτό οφείλεται στην περιορισμένη χρήση που μας παρέχει η άδεια που έχουμε για χρήση του λογισμικού.

Στο σημείο αυτό είναι σκόπιμο να παρατεθούν στοιχεία που δείχνουν την δομή του συστήματος αναγνώρισης και το πως αυτό ενσωματώνει την τεχνολογία των κρυφών μοντέλων Markov που αναλύθηκαν στο Κεφάλαιο 4.

5.1.1 Αρχιτεκτονική συστήματος αναγνώρισης φωνής μεγάλου λεξιλογίου

Για να μετατραπεί η σχεδιαστική φιλοσοφία της αναγνώρισης φωνής με HMMs, σύντομη περιγραφή της οποίας έλαβε χώρα στο προηγούμενο κεφάλαιο, σε πρακτικό επίπεδο απαιτείται η επίλυση ενός αριθμού προβλημάτων: [18]

- **Παραμετροποίηση ενός front-end,** δηλαδή παραμετροποίηση του σήματος στο πρώτο στάδιο της αναγνώρισης ώστε να μπορεί να εξαχθεί από τις κυματομορφές ομιλίας όλη η απαραίτητη πληροφορία με ένα συμπαγή τρόπο, συμβατό, παράλληλα με τα μοντέλα που είναι βασισμένα σε HMMs.

- **Σχεδιασμός των ακουστικών μοντέλων** (ακριβής αναπαράσταση των κατανομών του κάθε ήχου από τα HMMs), ώστε να καλύπτουν καθένα από τα συμφραζόμενα στα οποία ο ήχος μπορεί να εμφανίζεται. Επιπλέον, οι παράμετροι των HMMs πρέπει να μπορούν να υπολογιστούν από δεδομένα εκπαίδευσης, παρόλο που δεν θα ήταν ποτέ δυνατό να συγκεντρωθούν ικανοποιητικές ποσότητες δεδομένων για να καλύψουν όλες τις δυνατές εκφράσεις.

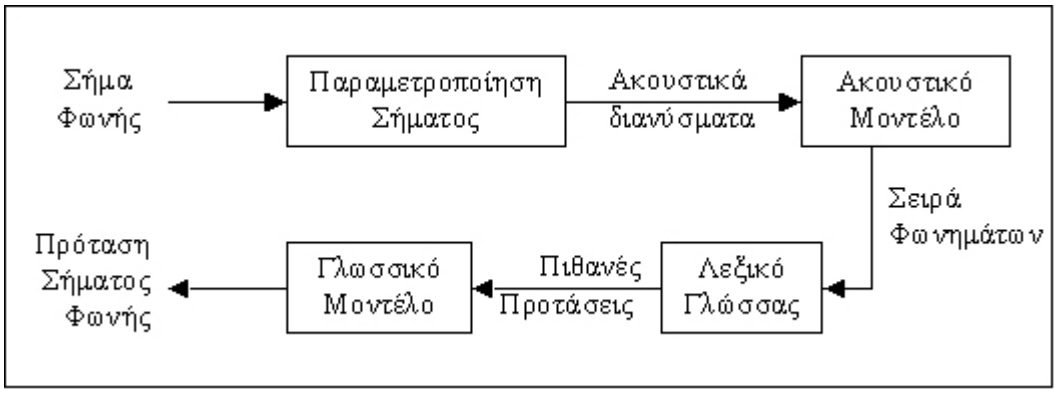
- **Σχεδιασμός του γλωσσικού μοντέλου**, ώστε να δίνει ακριβείς προβλέψεις βασιζόμενο στο προηγούμενο ιστορικό. Ωστόσο, καθώς η αραιή εμφάνιση

δεδομένων είναι ένα διαρκές πρόβλημα, το γλωσσικό μοντέλο πρέπει να είναι σε θέση να λειτουργεί ακόμα και για εκφράσεις για τις οποίες δεν εμφανίζονται καθόλου παραδείγματα στα δεδομένα της εκπαίδευσης.

- **Αποκωδικοποίηση**, ώστε οι εμφανιζόμενες εκφράσεις να ελέγχονται παράλληλα και σταδιακά να αποβάλλονται από το υπόλοιπο του ελέγχου όταν η πιθανότητα κάποιων εξ αυτών γίνεται πολύ μικρή. Ο σχεδιασμός αποκωδικοποιητών με ορθή απόδοση είναι ο κρισιμότερος παράγοντας για την υλοποίηση ενός συστήματος αναγνώρισης φωνής το οποίο μπορεί να λειτουργήσει γρήγορα και με ακρίβεια στα σημερινά υπολογιστικά συστήματα.

Παρακάτω θα ακολουθήσει αναλυτικότερη περιγραφή καθενός από τα τέσσερα αυτά θέματα.

Στο σημείο όμως αυτό κρίνεται απαραίτητη μια γραφική απεικόνιση (Σχήμα 5.1) του γενικού διαγράμματος ενός συστήματος αναγνώρισης συνεχούς λόγου μεγάλου λεξιλογίου με σκοπό να δειχθεί κατά κάποιο τρόπο η εξάρτηση των παραπάνω στοιχείων. Το προς αναγνώριση σήμα φωνής μετά από κατάλληλη επεξεργασία μετατρέπεται σε μία ακολουθία διανυσμάτων ακουστικών χαρακτηριστικών. Κάθε ένα από τα διανύσματα αυτά είναι μία περιεκτική κωδικοποίηση του φάσματος βραχέως χρόνου το οποίο υπολογίζεται σε χρονικό παράθυρο ανάλυσης μήκους 25 ms με ρυθμό περίπου 10 ms. Αυτό σημαίνει ότι υπάρχει μερική επικάλυψη των παραθύρων ανάλυσης. Για παράδειγμα ένα σήμα φωνής διάρκειας τριών δευτερολέπτων εκπροσωπείται με τριακόσια ακουστικά διανύσματα. Η αναγνώριση βασίζεται στις αρχές της στατιστικής αναγνώρισης προτύπων και περιλαμβάνει την εύρεση της ακολουθίας των λέξεων που είναι πιο πιθανόν να έχουν παράγει την ακολουθία των παρατηρούμενων διανυσμάτων χαρακτηριστικών.



Σχήμα 5.1: Γενική αρχιτεκτονική του συστήματος αναγνώρισης συνεχούς λόγου μεγάλου λεξιλογίου [19]

Η πιθανότητα μίας οποιασδήποτε πρότασης προσδιορίζεται αποσυνθέτοντας την πρόταση σε μία ακολουθία από λέξεις και στην συνέχεια αποσυνθέτοντας κάθε λέξη σε μία ακολουθία βασικών ήχων που ονομάζονται φωνήματα χρησιμοποιώντας ένα λεξικό. Στην αναγνώριση συνεχούς λόγου τα άγνωστα πρότυπα είναι τα φωνήματα και αντιστοιχούν σε στοιχειώδεις ήχους της γλώσσας (φθόγγους). Στην περίπτωση μας δεν υπάρχουν προτάσεις. Συνεπώς η αποσύνθεση του 1^{ου} σταδίου δεν είναι απαραίτητη.

Το φώνημα είναι η ελάχιστη λειτουργική μονάδα μίας γλώσσας η οποία δεν μπορεί

να αναλυθεί σε μικρότερα λειτουργικά στοιχεία και συνίσταται από ένα σύνολο διαφοροποιητικών χαρακτηριστικών. Αντικατάσταση ενός φωνήματος με ένα άλλο σε μία λέξη μεταβάλλει την σημασία της. Τα φωνήματα πραγματώνονται στον λόγο μέσω των φθόγγων. Το σύνολο των φωνημάτων που χρησιμοποιεί μία γλώσσα αποτελούν το φωνητικό αλφάβητο της γλώσσας. Το κάθε φώνημα παρίσταται με ένα ειδικό σύμβολο. Η Διεθνής Φωνητική Ένωση (International Phonetic Association) έχει δημιουργήσει το Διεθνές Φωνητικό Αλφάβητο (International Phonetic Alphabet), το οποίο περιλαμβάνει όλα τα φωνήματα όλων των γλωσσών του πλανήτη.

Στόχος του συστήματος αναγνώρισης φωνής είναι να αναγνωρίσει τα φωνήματα της κυματομορφής φωνής και να δώσει στην έξοδο τις λέξεις της πρότασης που αρθρώθηκαν.

Η πιθανή σειρά φωνημάτων υπολογίζεται από το ακουστικό μοντέλο. Το ακουστικό μοντέλο περιέχει για κάθε φώνημα ένα στατιστικό μοντέλο το οποίο καλείται κρυφό μοντέλο Markov (Hidden Markov Model, HMM). Όπως προαναφέρθηκε στο προηγούμενο κεφάλαιο, το κάθε HMM δέχεται ως είσοδο ένα ακουστικό διάνυσμα και επιστρέφει την πιθανότητα το ακουστικό διάνυσμα να είναι το φώνημα που περιγράφεται από το μοντέλο. Το φώνημα που αναγνωρίζεται από ένα ακουστικό διάνυσμα είναι αυτό όπου το HMM δίνει την μεγαλύτερη πιθανότητα. Το ακουστικό μοντέλο μετατρέπει την αλληλουχία των ακουστικών διανυσμάτων σε μια αλληλουχία φωνημάτων.

Τα φωνήματα συσχετίζονται με τα ορθογραφικά σύμβολα (γράμματα) της φυσικής γλώσσας που αναγνωρίζεται. Γίνεται εύρεση από το λεξικό της γλώσσας όλων των πιθανών λέξεων που εκπροσωπούνται από την αλληλουχία των φωνημάτων. Κατόπιν το γλωσσικό μοντέλο επιστρέφει για κάθε υποψήφια πρόταση που σχηματίστηκε την πιθανότητα να είναι σημασιολογικά σωστή. Στην έξοδο του συστήματος επιστρέφεται η πρόταση με την μεγαλύτερη πιθανότητα.

5.1.2 Παραμετροποίηση Σήματος (Front-End)

Το σήμα φωνής είναι ένα σήμα πολύπλοκο και πλούσιο επειδή παρέχει πληροφορίες σχετικές με το αποδιδόμενο μήνυμα αλλά και πληροφορίες σχετικές με τον ομιλητή (χροιά φωνής, συναισθηματική κατάσταση ομιλητή κ.α.). Όμως το σύστημα αναγνώρισης φωνής θεωρεί ότι το σήμα φωνής σε μικρά διαστήματα είναι σταθερό δηλ. τα φασματικά χαρακτηριστικά του μεταβάλλονται λίγο. Τα μικρά αυτά διαστήματα ονομάζονται παράθυρα λόγου (speech windows).

Το πρώτο στάδιο είναι η παραμετροποίηση του σήματος φωνής. Το σήμα χωρίζεται σε μικρότερα τμήματα φωνής και από το κάθε τμήμα εξάγεται ένα διάνυσμα με τα φασματικά χαρακτηριστικά. Το κενό μεταξύ των τμημάτων είναι συνήθως 10 msec και τα τμήματα συνήθως επικαλύπτονται για να έχουμε μεγαλύτερο παράθυρο ανάλυσης. Έτσι το μήκος του επικαλυπτόμενου παραθύρου συνήθως επιλέγεται 25 msec. Το κάθε τμήμα φωνής που εξάγουμε συνήθως το πολλαπλασιάζουμε με μία συνάρτηση παραθύρου[1] (π.χ. Hamming).

[1] Συναρτήσεις παραθύρου: υπάρχουν αρκετές συναρτήσεις παραθύρου. Οι πλέον διαδεδομένες εξ αυτών είναι η Τετραγωνική, η Hann, η Hamming, η Blackman και η Kaiser-Bessel. Οι συναρτήσεις αυτές χρησιμεύουν στον διαχωρισμό του φωνητικού σήματος σε μικρότερα τμήματα φωνής. Αυτό γίνεται αν πολλαπλασιάσουμε την εκάστοτε λέξη (ή φώνημα) με το παράθυρο λόγου, οπότε παίρνουμε το μικρότερο δυνατό τμήμα λόγου (frame).

Συχνά επίσης στο σήμα αυτό κάνουμε προ-έμφαση ώστε να ενισχυθούν οι υψηλές συχνότητες αντισταθμίζοντας έτσι την εξασθένηση αυτή που προκαλείται από τα χείλη [18].

5.1.3 Ακουστικό μοντέλο

Όπως αναφέρθηκε, το ακουστικό μοντέλο περιέχει για κάθε φώνημα της γλώσσας που αναγνωρίζεται ένα HMM. Κάθε HMM δέχεται ως είσοδο το ακουστικό διάνυσμα και επιστρέφει την πιθανότητα να εκπροσωπεί το ακουστικό διάνυσμα το φώνημα που μοντελοποιεί το συγκεκριμένο HMM. Για να είναι σε θέση το HMM να επιστρέφει αυτή την πιθανότητα στην εκμάθηση του HMM χρησιμοποιούνται πολλά και διαφορετικά δείγματα του φωνήματος.

Ένα HMM για αναγνώριση φωνημάτων έχει ένα αριθμό καταστάσεων (συνήθως 3) και μία τοπολογία από αριστερά προς τα δεξιά δηλαδή ως αρχική κατάσταση επιλέγεται η πρώτη από τις καταστάσεις του μοντέλου, ως τελική η τελευταία από τις καταστάσεις του μοντέλου ενώ όταν το HMM φεύγει από μία κατάσταση δεν επιστρέφει ποτέ σε αυτήν.

Η Ελληνική γλώσσα μπορεί να παρασταθεί με 32 φωνήματα, έτσι θεωρητικά χρειάζεται ένα ακουστικό μοντέλο με 32 HMM. Στην πράξη το κάθε φώνημα παρουσιάζει διαφορετικές μορφές ανάλογα με το περιβάλλον συνάρθρωσης, επομένως για να προκύψουν αξιόπιστα HMM πρέπει να γίνει εκμάθηση για κάθε διαφορετικό περιβάλλον. Η συνηθέστερη προσέγγιση είναι η υλοποίηση διαφορετικού μοντέλου HMM για κάθε φώνημα σε διαφορετικό περιβάλλον συνάρθρωσης (HMM τρι-φωνημάτων). Με 32 φωνήματα έχουμε 32*32*32=32.768 δυνατούς συνδυασμούς φωνημάτων και περιβάλλοντος άρθρωσης. Ορισμένοι από τους συνδυασμούς αυτούς δεν εμφανίζονται ποτέ και στην πράξη χρειάζεται να γίνει εκμάθηση σε περίπου 20.000 μοντέλα HMM τρι-φωνημάτων.

Στη σχεδίαση/εκμάθηση του ακουστικού μοντέλου λαμβάνονται υπόψη το μέγεθος των δεδομένων εκμάθησης (φωνήματα σε διαφορετικό περιβάλλον άρθρωσης και από διαφορετικούς ομιλητές), ο αριθμός των HMM και το μέγεθος των παραμέτρων των ακουστικών διανυσμάτων.

Στην αναγνώριση φωνής τα ακουστικά διανύσματα που έγιναν στο στάδιο της παραμετροποίησης του σήματος φωνής μετατρέπονται στην αλληλουχία φωνημάτων όπως υπολογίστηκε στατιστικά από τα HMMs. [18]

5.1.4 Γλωσσικό μοντέλο

Το γλωσσικό μοντέλο είναι ένας μηχανισμός για τον υπολογισμό της πιθανότητας να υπάρχει μία λέξη *W* σε μία εκφώνηση όπου προηγούνται n λέξεις. Σε μία γλώσσα δεν επιτρέπεται οποιαδήποτε σειρά λέξεων (συντακτικά, γραμματικά και σημασιολογικά) π.χ. «Η αρκούδα κοιμήθηκε» είναι δεκτή πρόταση ενώ η πρόταση «Η θάλασσα κοιμήθηκε» δεν είναι δεκτή πρόταση σημασιολογικά. Το γλωσσικό μοντέλο μοντελοποιεί το συντακτικό, την γραμματική και την σημασιολογία μιας γλώσσας. Στο προηγούμενο παράδειγμα, ένα σωστά εκπαιδευμένο γλωσσικό μοντέλο θα επιστρέψει μεγαλύτερη πιθανότητα στην πρόταση «Η αρκούδα κοιμήθηκε» από ότι στην πρόταση «Η θάλασσα κοιμήθηκε».

Στην περίοδο εκμάθησης του γλωσσικού μοντέλου, χρησιμοποιείται ένα πολύ μεγάλο

σώμα κειμένων (π.χ. 40.000.000 λέξεων) και βρίσκονται οι πιο συχνά εμφανιζόμενες λέξεις με ταυτόχρονο υπολογισμό των πιθανοτήτων εμφάνισής τους (μονογράμματα). Κατόπιν υπολογίζονται οι πιθανότητες εμφάνισης δύο λέξεων στη σειρά (διγράμματα) και οι πιθανότητες εμφάνισης τριών λέξεων στη σειρά (τριγράμματα). Θεωρητικά μπορούν να υπολογισθούν και οι πιθανότητες εμφάνισης τεσσάρων, πέντε κλπ λέξεων στη σειρά (τετραγράμματα, πενταγράμματα) αλλά τότε χρειάζεται πολύ μεγαλύτερο σώμα κειμένων το οποίο δεν είναι πρακτικά εύκολα υλοποιήσιμο.
Μεγάλη βαρύτητα έχει το είδος του σώματος κειμένων που χρησιμοποιείται στην εκμάθηση του γλωσσικού μοντέλου. Στην περίπτωση π.χ. που το σύστημα αναγνώρισης φωνής προορίζεται να χρησιμοποιηθεί από δημοσιογράφο που θέλει να αναγνωρίζει εκφωνημένα πολιτικά κείμενα με αξιοπιστία θα πρέπει το σώμα κειμένων που θα χρησιμοποιηθεί για την εκμάθηση του γλωσσικού μοντέλου να περιέχει πολιτικά κείμενα. [18]

5.1.5 Αποκωδικοποίηση

Σύμφωνα με τα παραπάνω, τελικά η αλληλουχία φωνημάτων που επιστρέφει το ακουστικό μοντέλο μετατρέπονται με τη βοήθεια του λεξικού σε πιθανές προτάσεις οι οποίες έχουν διαφορετική αλληλουχία λέξεων. Το σύστημα επιστρέφει στην έξοδο την πρόταση που έχει την μεγαλύτερη πιθανότητα ορθότητας όπως δόθηκε από το γλωσσικό μοντέλο.

Προκειμένου δηλαδή να επιτευχθεί αναγνώριση με τα τρία αυτά στοιχεία πρέπει να βρεθεί η ακολουθία των λέξεων $\underline{W}$ που μεγιστοποιεί την πιθανότητα ορθότητας.

Διάφορες μέθοδοι έχουν προταθεί για έρευνα και διάσχιση τέτοιων μοντέλων, βασιζόμενες κυρίως στην εφαρμογή των γνωστών αλγορίθμων depth-first και breadth-first. Στην σχεδίαση με depth-first η πιο πιθανή υπόθεση αναζητείται μέχρι να φτάσουμε στο τέλος της ομιλίας. Στην σχεδίαση με breadth-first όλες οι υποθέσεις ελέγχονται παράλληλα. Η αποκωδικοποίηση εκμεταλλεύεται την μέθοδο βελτιστοποίησης του Bellman γνωστή ως αποκωδικοποίηση Viterbi. Αυτή η μέθοδος είναι η *Viterbi Beam Search* και χρησιμοποιείται από το σύστημα αναγνώρισης φωνής για την αναζήτηση της πιο πιθανής υπόθεσης. Ο αλγόριθμος αυτός είναι αρκετά επαρκής και συνδυαζόμενος με άλλες τεχνικές προσδίδει στο ολοκληρωμένο σύστημα την ικανότητα να αναγνωρίζει σε πραγματικό χρόνο λεξικά με παραπάνω από 10.000 λέξεις, εκφωνημένες με εκατομμύρια δυνατούς τρόπους. Η διαδικασία διαγραφής των λιγότερο πιθανών υποθέσεων βελτιώνει το χρόνο της αναγνώρισης αλλά μπορεί να προκαλέσει λάθη στην αναγνώριση. [18]

Φυσικά όλα τα παραπάνω στην περίπτωση της εφαρμογής στην εργασία αυτή είναι πολύ πιο απλουστευμένα, καθώς οι λέξεις δίνονται από τον χρήστη μεμονωμένα. Παρόλα αυτά η παραπάνω ανάπτυξη κρίθηκε απαραίτητη για την καλύτερη κατανόηση του μοντέλου αναγνώρισης.

Αξίζει τέλος να σημειωθεί ότι το εν λόγω σύστημα αναγνώρισης είναι η πρώτη, και ίσως η μόνη επιτυχημένη προσπάθεια εφαρμογής τεχνικών αναγνώρισης φωνής με ελληνικό λεξιλόγιο.

5.2 ΜΙΚΡΟΦΩΝΟ ΚΑΙ ΨΗΦΙΟΠΟΙΗΣΗ

Για την επίτευξη της διαδικασίας αναγνώρισης απαιτείται όπως προαναφέραμε η χρήση ενός μικροφώνου. Παραθέτουμε κάποια βασικά χαρακτηριστικά της ψηφιοποίησης και των μικροφώνων για να είναι κατανοητή η περαιτέρω διαδικασία [20].

5.2.1. Μικρόφωνο

Σε ένα σύστημα αναγνώρισης φωνής, το μικρόφωνο αντικαθιστά το τύμπανο του αυτιού. Τα παρόντα κύματα πίεσης στον αέρα μεταβιβάζονται σε ένα ανάλογο ηλεκτρικό σήμα ίσως ποικίλλοντας την πυκνότητα της σκόνης άνθρακα σε ένα κουτί ή με κάποιο άλλο μέσο.

Τα καλύτερα μικρόφωνα έχουν ομοιόμορφη συχνότητα αντίδρασης, που σημαίνει ότι «ακούν» ήχους εξίσου καλά από μία ευρεία κλίμακα συχνοτήτων. Για την αναγνώριση φωνής, το μικρόφωνο θα πρέπει να έχει καλή αντίδραση συχνότητας σε ολόκληρη την κλίμακα της φυσιολογικής ανθρώπινης ακοής.

Ένα χαρακτηριστικό της υψηλής ποιότητας μικροφώνου είναι η ικανότητα να ματαιώνουν το θόρυβο. Αυτό συχνά γίνεται χρησιμοποιώντας δύο μικρόφωνα, το ένα με κατεύθυνση τον ήχο που μας ενδιαφέρει και το άλλο προσανατολισμένο στην αντίθετη κατεύθυνση. Ο περιβάλλων θόρυβος θα πρέπει να ακούγεται το ίδιο και στα δύο μικρόφωνα. Ο ήχος που μας ενδιαφέρει θα πρέπει να είναι δυνατότερος στο πρώτο μικρόφωνο. Αφαιρώντας τα δύο σήματα, ο θόρυβος ματαιώνεται αφήνοντας μόνον το σήμα με το μεγαλύτερο διαφορικό. Αυτό βελτιώνει το λόγο σήμα-θόρυβο για το μικρόφωνο.

Μία άλλη μέθοδος ματαίωσης θορύβου είναι να ψάξουμε για σταθερές συχνότητες θορύβου παρούσες στο περιβάλλον. Αυτές οι συχνότητες μπορεί να προέρχονται από φώτα νέον, ανεμιστήρες υπολογιστών, το μουρμουρητό 60 Hz του ηλεκτρικού συστήματος ή το αργό σφύριγμα των αγωγών εξαεριστήρων. Επειδή αυτοί οι ήχοι είναι σταθεροί, μπορούμε να βελτιώσουμε το λόγο σήμα-θόρυβο για την ομιλία σβήνοντας όλα τα σήματα σε αυτές τις συχνότητες σταθερού θορύβου. Αυτή η μέθοδος δεν είναι τόσο καλή όσο η ματαίωση θορύβου που βασίζεται στα 2 μικρόφωνα, γιατί χάνει μέρος του πραγματικού σήματος εκτός από τον θόρυβο. Παρόλα αυτά είναι κάποια βελτίωση.

5.2.2. Ψηφιοποίηση

Το ανάλογο σήμα από το μικρόφωνα μπορεί να καταγραφεί άμεσα σε μαγνητική ταινία, χρησιμοποιώντας ένα απλό μαγνητόφωνο. Αυτό το σήμα μπορεί κατόπι να αναπαραχθεί με εξαιρετική πιστότητα ξαναπαίζοντας την ταινία μέσω ενός ενισχυτή και ηχείων. Ενδιαφερόμαστε περισσότερο για τη ψηφιακή επεξεργασία του σήματος. Για αυτό το λόγο πρέπει να μετατρέψουμε τις εισερχόμενες τάσεις σε αριθμούς. Αυτό

γίνεται δειγματολογώντας το σήμα πολλές φορές ανά δευτερόλεπτο. Καταγράφουμε το επίπεδο του σήματος σε κάθε μία από αυτές τις φορές.

Η δειγματολόγηση είναι γνωστή σε οποιονδήποτε έχει παρακολουθήσει μία ταινία ή ένα πρόγραμμα τηλεόρασης. Στην τηλεόραση, το περιεχόμενο ζωγραφίζεται στην οθόνη με ταχύτητα περίπου 30 εικονιδίων ανά δευτερόλεπτο. Το ενδιαφέρον είναι ότι σε γενικές γραμμές η ανθρώπινη αντίληψη δε φαίνεται να ενοχλείται από αυτά τα 30 εικονίδια ανά δευτερόλεπτο. Αντίθετα βλέπουμε ομαλή κίνηση. Αλλά αυτό είναι ίσως λιγότερο εκπληκτικό όταν αντιληφθούμε ότι το ανθρώπινο νευρικό σύστημα, έτσι κι αλλιώς, παράγει εκρήξεις δεδομένων. Δε φαίνεται να λειτουργούμε με ανάλογα σήματα.

Υπάρχει ένα θεώρημα που ονομάζεται το θεώρημα του Nyquist. Βασικά δηλώνει ότι για να αναλύσουμε σωστά ένα σήμα σε κάποια συχνότητα πρέπει να το δειγματίσουμε σε τουλάχιστον διπλάσια συχνότητα. Εάν εφαρμοσθεί στην ομιλία, σημαίνει ότι ένα σήμα στα 1000 Hz πρέπει να δειγματολογηθεί σε περισσότερο από 2000 Hz, αλλιώς δε μπορεί να αναλυθεί στις σωστές του τιμές.

Υπάρχουν διάφορες μορφές αρχείου (τύποι) στα οποία μπορούν να σωθούν δεδομένα ομιλίας. Στα τελευταία χρόνια η πιο δημοφιλής μορφή είναι η Microsoft WAV. Υπάρχουν πολλές άλλες μορφές. Η διαφορά είναι στο πως πολλά δείγματα παρουσιάζονται ανά δευτερόλεπτο (samples/sec) και ως ακριβώς παρουσιάζεται το κάθε δείγμα και εάν είναι «στέρεο» (stereo mode, 2 κανάλια) ή «μόνο» (mono mode, 1 κανάλι) και εάν τα Κβάντα είναι λογαριθμικά διατεταγμένα ή με κάποιον άλλο τρόπο.

ΚΕΦΑΛΑΙΟ 6 : ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ – ΑΛΓΟΡΙΘΜΟΣ ΚΙΝΗΣΗΣ

6.1 ΚΙΝΗΣΗ ΤΟΥ ΡΟΜΠΟΤ

Η κίνηση του ρομπότ μπορεί να επιτευχθεί με τρεις τρόπους:

- Μέσω χειριστηρίου (joystick). Το χειριστήριο στην ουσία χρησιμεύει για την ευκολότερη μετακίνηση του κάθε ρομπότ, καθότι είναι αρκετά βαρύ για να το μετακινεί ένας μέσος χειριστής σε μεγάλες αποστάσεις χειροκίνητα. Στην περίπτωση αυτή οι εντολές κίνησης δίνονται από τον χειριστή εκείνη τη στιγμή.
- Μέσω του υπολογιστή του ρομπότ. Δεδομένης της οθόνης υγρών κρυστάλλων που παρέχει το ρομπότ, σε συνδυασμό με το σύστημα rFlex που περιγράψαμε στο 1^{ο} Κεφάλαιο και την υποδοχή για πληκτρολόγιο που παρέχεται στο εσωτερικό του κάθε ρομπότ, θα μπορούσε κάποιος να χρησιμοποιήσει τον Η/Υ του ρομπότ απευθείας για τον προγραμματισμό της κίνησης του.
- Μέσω υπολογιστή που είναι συνδεδεμένος σε δίκτυο (είτε σε τοπικό ίδιο με αυτό του ρομπότ, είτε στο διαδίκτυο) Στην περίπτωση αυτή, η κίνηση του ρομπότ πραγματοποιείται μέσω του αντίστοιχου λογισμικού (Mobility) του ρομπότ που αναλύθηκε στο Κεφάλαιο 2.

Στις περισσότερες των περιπτώσεων, η κίνηση του ρομπότ επιβάλλεται να πραγματοποιηθεί με τον τρίτο τρόπο, δηλαδή μέσω υπολογιστή λόγω της ευελιξίας, της ακρίβειας και της αυτοματοποίησης της κίνησης. Ευελιξία γιατί ο χρήστης με αλλαγή μερικών κάθε φορά παραμέτρων μπορεί να πραγματοποιήσει όλες τις δυνατές κινήσεις του ρομπότ πολύ εύκολα. Ακρίβεια γιατί η κίνηση μέσω του λογισμικού Mobility απαιτεί ακριβή προσδιορισμό όλων των παραμέτρων της κίνησης (γραμμική και γωνιακή ταχύτητα και χρόνο), ενώ με το χειριστήριο ο έλεγχος των παραμέτρων αυτών είναι σχεδόν ανέφικτος. Τέλος, αυτοματοποίηση με την έννοια των δυνατοτήτων που παρέχει στον χρήστη ένα λογισμικό περιβάλλον που μπορεί να λειτουργήσει μέσα από ποικίλα προγραμματιστικά περιβάλλοντα, με όλα τα θετικά που αυτό συνεπάγεται.

Αντίθετα, ο πρώτος και ο δεύτερος τρόπος δεν είναι καθόλου ευέλικτοι όπως είναι εύκολα κατανοητό, καθώς και στις δύο περιπτώσεις ο χρήστης πρέπει να έρχεται σε άμεση επαφή με το ρομπότ καθ' όλη τη διάρκεια της κίνησης του.

Στην εργασία αυτή θα ασχοληθούμε, λοιπόν, αποκλειστικά από τον έλεγχο του ρομπότ μέσω υπολογιστή που έχει δυνατότητα σύνδεσης μέσω δικτύου με τον υπολογιστή του ρομπότ.

Το περιβάλλον του Mobility που θα χρησιμοποιηθεί επιτρέπει στο χρήστη να τροποποιήσει βασικά μέρη του συστήματος του ρομπότ και να προσθέσει νέα, ανάλογα με τις ανάγκες του. Τα αντικείμενα που ορίζονται μέσα στο λογισμικό του ρομπότ μπορούν να τροποποιηθούν ή να χρησιμοποιηθούν ως συναρτήσεις σε νέους αλγορίθμους. Στην εργασία αυτή η ανάπτυξη των αλγορίθμων, τόσο για την

κίνηση του ρομπότ όσο και για την ανάπτυξη του διαμεσολαβητή φωνής (που αναπτύχθηκαν στο Κεφάλαιο 5) έγινε σε προγραμματιστικό περιβάλλον C++.

6.2 ΠΡΟΓΡΑΜΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ ΚΑΙ ΗΧΟΓΡΑΦΗΣΗΣ

Το πρόγραμμα αναγνώρισης έχει τελικά πάρει την μορφή ενός εκτελέσιμου αρχείου, το οποίο δέχεται ως είσοδο ένα αρχείο ήχου τύπου Microsoft WAV το οποίο έχει ηχογραφηθεί σε κατάσταση «μόνο» (mono mode) με 16KHz συχνότητα δειγματοληψίας και 16 δείγματα/δευτερόλεπτο (sample/sec) και το αποτέλεσμα της αναγνώρισης το τοποθετεί σε ένα αρχείο κειμένου τύπου ASCII (text file -.txt). Αλλά αυτό θα είναι το τελικό αποτέλεσμα που προϋποθέτει αρκετά βήματα, τα οποία και περιγράφονται σε αυτήν την παράγραφο. Πιο συγκεκριμένα, έχουμε δύο προγράμματα, αυτό που εκτελεί την αναγνώριση αυτή καθ' αυτή και αυτό που εκτελεί την ηχογράφηση πριν την αναγνώριση. Αναλύονται αμέσως παρακάτω.

6.2.1 Πρόγραμμα Αναγνώρισης

Για να λειτουργήσει η διαδικασία της αναγνώρισης απαιτείται η εγκατάσταση ενός «κλειδιού» καθότι όπως προαναφέραμε το πρόγραμμα είναι κλειδωμένο, και συνεπώς δεν μπορούμε να εκμεταλλευτούμε όλο το εύρος των δυνατοτήτων του. Το «κλειδί» αυτό είναι ουσιαστικά ένα USB (Universal Serial Bus) HASP Driver. Αυτό εγκαθίσταται στον υπολογιστή στον οποίο θα τρέξει το πρόγραμμα τρέχοντας το εκτελέσιμο αρχείο HDD32.exe.

Η λειτουργία του προγράμματος αναγνώρισης φωνής στηρίζεται στις εξής τρεις συναρτήσεις:

- "virec_init", η οποία κάνει την αρχικοποίηση του προγράμματος αναγνώρισης και χρειάζεται να τρέξει μόνο μία φορά.

- "virec_recogn", η οποία αναγνωρίζει το περιεχόμενο ενός Microsoft WAV αρχείου το οποίο έχει ηχογραφηθεί σε mono mode με 16KHz συχνότητα δειγματοληψίας και 16 bit/sample και το αποτέλεσμα της αναγνώρισης το τοποθετεί σε ένα αρχείο κειμένου.

- "virec_close" η οποία αποδεσμεύει την μνήμη που είχε παραχωρηθεί για το πρόγραμμα της αναγνώρισης.

Οι συναρτήσεις αυτές υπάρχουν μέσα σε ένα Dynamic Link Library (.dll) αρχείο το οποίο για να μπορέσει να μετατραπεί σε εκτελέσιμο πρέπει να περάσει από ένα μεταφραστή (compiler). Αυτό επιτεύχθηκε μέσω της γλώσσας προγραμματισμού C++ και συγκεκριμένα του λογισμικού της Borland C++ Builder 5. Εκεί δημιουργήθηκε μια «εφαρμογή» (application) που είχε ως αποτέλεσμα την εξαγωγή ενός εκτελέσιμου αρχείου (.exe), το οποίο μπορεί πλέον να κληθεί μέσα από οποιοδήποτε πρόγραμμα.

Το αρχείο αυτό (exe) που έχει δημιουργηθεί, όπως αναφέρθηκε παίρνει ως είσοδο το αρχείο ήχου (WAV) και δίνει ως έξοδο ένα αρχείο κειμένου (txt). Όσον αφορά στο περιεχόμενο του αρχείου ήχου πρέπει να είναι μια, και μόνο μία κάθε φορά, από τις ελληνικές λέξεις που δέχεται το πρόγραμμα (Πίνακας 6.1) ήδη ηχογραφημένη σε

κατάσταση «μόνο» με 16KHz συχνότητα δειγματοληψίας και 16 δείγματα/δευτερόλεπτο. Αντίστοιχα, το περιεχόμενο του αρχείου κειμένου που δίνει το πρόγραμμα θα είναι η αντίστοιχη ελληνική λέξη ακριβώς όπως αυτή αναπαρίσταται στον Πίνακα 6.1. Θεωρητικά το πρόγραμμα αναγνωρίζει 22 διακριτά σύνολα χαρακτήρων (που είναι οι λέξεις του Πίνακα). Ουσιαστικά όμως, οι λέξεις της ελληνικής γλώσσας που αναγνωρίζει είναι 18. Ο αριθμός 22 προκύπτει καθότι τρεις αριθμοί (οι 7, 8 και 9) μπορούν να εκφραστούν στην ελληνική γλώσσα με δύο τρόπους, ενώ υπάρχει και η περίπτωση του κενού αρχείου ήχου που συμβολίζεται με </s> στον πίνακα.

Ελληνική Λέξη	**Λατινική Προφορά**
μπροστά	brosta
πίσω	piso
στροφή	strofi
ξεκίνα	ksecina
σταμάτα	stamata
δεξιά	deksia
αριστερά	aristera
ευθεία	efTia
μηδέν	miDen
ένα	ena
δύο	dio
τρία	tria
τέσσερα	tesera
πέντε	pende
έξι	eksi
επτά	epta
εφτά	efta
οκτώ	okto
οχτώ	oxto
εννέα	enea
εννιά	eJa
</s>	###

Πίνακας 6.1: Ελληνικό Λεξικό Λογισμικού Αναγνώρισης

6.2.2. Πρόγραμμα Ηχογράφησης

Ένα θέμα που προκύπτει είναι ότι ένα πλήρες πρόγραμμα αναγνώρισης σε πραγματικό χρόνο (real–time) απαιτεί την επί τόπου ηχογράφηση του αρχείου φωνής, ενώ το συγκεκριμένο πρόγραμμα αναγνώρισης απαιτεί την προϋπαρξη του συγκεκριμένου αρχείου στον ίδιο χώρο του σκληρού δίσκου που βρίσκεται το πρόγραμμα. Αυτό μπορεί να αντιμετωπιστεί με την δημιουργία ενός άλλου προγράμματος που να συνεργάζεται κατά κάποιον τρόπο με το πρόγραμμα αναγνώρισης φωνής και να ζητάει από τον χρήστη να μιλήσει, να ηχογραφεί, να αποθηκεύει το αρχείο φωνής και μετά να καλεί το πρόγραμμα αναγνώρισης.

Για τον σκοπό αυτό δημιουργήθηκε ένα πρόγραμμα βασισμένο: (α) στην Microsoft Foundation Class (Βιβλιοθήκη) (MFC) που παρέχει η Visual C++ 6.0, καθώς και (β) σε ένα πλήθος κλάσεων που έχουν δημιουργηθεί σε C++. Αναλυτικότερα έχουμε τα εξής:

(α) Η βιβλιοθήκη MFC της Microsoft Visual C++ περιέχει περίπου 200 MFC κλάσεις, ενώ παρέχει ένα περιβάλλον πάνω στο οποίο μπορεί κάποιος να κτίσει μια εφαρμογή των Windows. Ακόμα συμπυκνώνει τα περισσότερα από τα API (Application Programming Interface, Περιβάλλον Προγραμματιστικών Εφαρμογών) των Win32 σε ένα σύνολο καλά οργανωμένων κλάσεων. [21]

Τα βασικά χαρακτηριστικά της MFC βιβλιοθήκης είναι τα ακόλουθα:

- Ευκολία επαναχρησιμοποιήσιμου κώδικα.
- Εξαγωγή μικρότερων σε μέγεθος εκτελέσιμων αρχείων.
- Δυνατότητα καθοδήγησης για ταχύτερη ανάπτυξη του προγράμματος, ειδικά για νέους χρήστες του αντικειμενοστραφούς προγραμματισμού
- Τα προγράμματα που χρησιμοποιούν την βιβλιοθήκη αυτή πρέπει να γράφονται αναγκαστικά σε C++ και να χρησιμοποιούν κλάσεις. Απαιτεί λοιπόν μια στοιχειώδη γνώση των κλάσεων και των αρχών του αντικειμενοστραφούς προγραμματισμού.

(β) Πρόκειται για ένα απλό πρόγραμμα ηχογράφησης το οποίο έχει την δυνατότητα να ηχογραφήσει και να αναπαράγει ήχους μέσω της κάρτας ήχου του υπολογιστή. Οι κλάσεις (classes της C++) που χρησιμοποιήθηκαν είναι οι ακόλουθες:

- Η CWave, που είναι η κύρια κλάση για την φόρτωση και την αποθήκευση του αρχείου ήχου. Έτσι καθορίζεται η συχνότητα ηχογράφησης, η «μόνο» ή «στέρεο» κατάσταση (αντίστοιχα mode 1 και 2) καθώς και ο αριθμός των δειγμάτων/δευτερόλεπτο που θέλουμε.
- Η CWaveDevice ασχολείται με τις συσκευές ήχου (κάρτες ήχου) για λογαριασμό της κλάσης CWave.
- Η CWaveInterface απαριθμεί όλες τις διαθέσιμες συσκευές ήχου καθώς και το όνομα της κάθε μιας.
- Η CWaveBufferis είναι η κλάση που συμπυκνώνει την προσωρινή μνήμη (buffer) για την κλάση CWave και τις κλάσεις CWaveIn/CWaveOut.
- Η CWaveIn είναι η βασική κλάση για τη διαδικασία της ηχογράφηση (αρχή ηχογράφηση, παύση, συνέχιση, λήξη ηχογράφησης) και για τη δημιουργία μιας νέας CWave από την ηχογράφηση.
- Η CWaveOut επιτρέπει την αναπαραγωγή της κλάσης CWave με ποικίλους τρόπους.

Ο συνδυασμός των παραπάνω με εφαρμογή στο προγραμματιστικό περιβάλλον της Visual C++ 6.0 έδωσε ως αποτέλεσμα ένα εκτελέσιμο MFC αρχείο (.exe), το οποίο επίσης καλείται από οποιοδήποτε πρόγραμμα. Αυτό, σε συνδυασμό με το πρόγραμμα αναγνώρισης φωνής που αναλύθηκε παραπάνω, κάνουν εφικτή την δυνατότητα επεξεργασίας και αναγνώρισης σε σχεδόν πραγματικό χρόνο.

6.3 ΑΛΓΟΡΙΘΜΟΣ ΚΙΝΗΣΗΣ

Είδαμε στην προηγούμενη παράγραφο ότι το πρόγραμμα αναγνώρισης φωνής δίνει ως έξοδο ένα αρχείο κειμένου (txt). Αυτό το αρχείο ουσιαστικά περιέχει μια από τις

λέξεις που αναγνωρίζει το πρόγραμμα. Η μορφή όμως αυτή στην οποία βρίσκεται η μεταβλητή μας στο αρχείο που δίνει το πρόγραμμα αναγνώρισης δεν είναι καθόλου ευέλικτη, καθώς είναι ένα string, δηλαδή ένα διακριτό σύνολο από ελληνικούς χαρακτήρες, το οποίο δεν είναι αναγνωρίσιμο από όλα τα λογισμικά και όλους τους μεταφραστές (compilers).

Για τον σκοπό αυτό, απαιτείται η άμεση κωδικοποίηση του σε μορφή αριθμού. Δηλαδή σε κάθε εντολή να αντιστοιχεί και ένας ακέραιος αριθμός. Έτσι, μέσα από ένα προγραμματιστικό περιβάλλον της C++ κωδικοποιούμε τη «Λέξη» σε έναν αριθμό, τον οποίο αποστέλλουμε με κάποιο τρόπο στον Η/Υ του ρομπότ για να εκτελέσει την αντίστοιχη κίνηση.

Στις επόμενες δύο παραγράφους θα αναλύσουμε αφενός ποιος είναι αυτός ο τρόπος επικοινωνίας των δύο προγραμμάτων με στόχο την αποστολή των κωδικοποιημένων λέξεων, αφετέρου ποιες είναι οι κινήσεις που μπορεί να εκτελέσει το όχημα μέσα από το πρόγραμμα αυτό.

6.3.1 Επικοινωνία Μεταξύ Προγραμμάτων

Ο Η/Υ του ATRV-Mini λειτουργεί σε λειτουργικό περιβάλλον Linux και συγκεκριμένα έχει την διανομή Red Hat 6.2. Αντίστοιχα, ο υπολογιστής μέσω του οποίου έγινε η εφαρμογή λειτουργεί σε λειτουργικό σύστημα Windows 2000 ή Windows XP (η δομή των δύο αυτών λειτουργικών συστημάτων είναι λίγο πολύ η ίδια καθώς τα Win XP δομήθηκαν πάνω στην βάση των Win 2000).

Είναι λοιπόν λογικό να υπάρχει πρόβλημα συμβατότητας στην προσπάθεια επίτευξης επικοινωνίας των δύο αυτών συστημάτων. Επίσης γεννάται το ερώτημα πώς μπορεί να επικοινωνήσει «αυτόματα» ο ένας υπολογιστής με τον άλλο για να μεταφέρει τα απαραίτητα δεδομένα (δηλαδή τους κωδικοποιημένους αριθμούς) αφού το κάθε πρόγραμμα τρέχει και σε διαφορετικό υπολογιστή. Η απάντηση έρχεται σχετικά εύκολα αν θεωρήσει κανείς τον Η/Υ του ρομπότ ως εξυπηρετητή (server) και τον Η/Υ που τρέχει Windows ως πελάτη (client), αν θεωρήσουμε δηλαδή ότι έχουμε TCP/IP δίκτυο ενός πελάτη με έναν εξυπηρετητή (Single Client – Single Server). Το πρόβλημα αυτό θα μπορούσε να επιλυθεί μέσω TCP/IP sockets (η λέξη sockets θα μπορούσε να μεταφραστεί ως “υποδοχές”). Είναι όμως μια αρκετά πολύπλοκη διαδικασία που τελικά δεν χρησιμοποιήθηκε καθώς βρέθηκε μια πιο απλή λύση, αυτή του Λογισμικού Samba, που παρέχει δυνατότητα αποστολής αρχείων μεταξύ δύο διαδικασιών που συμβαίνουν σε διαφορετικούς υπολογιστές.

Η παραπάνω παράγραφος περιέχει για πολλούς αναγνώστες πολλές «άγνωστες» λέξεις και έννοιες. Για τον λόγο αυτό κρίθηκε σκόπιμη μια σύντομη αναφορά σε αρκετές από αυτές τις έννοιες για την καλύτερη κατανόηση [22,23]:

α) Χρήση υπολογιστών με την μέθοδο Client-Server: Η σχέση πελάτης/κεντρικός υπολογιστής δικτύου είναι μία υπολογιστική αρχιτεκτονική που περιλαμβάνει διαδικασίες πελάτη, ο οποίος απαιτεί εξυπηρέτηση από τις διαδικασίες του κεντρικού υπολογιστή. Η περίπτωση μας περιλαμβάνει έναν πελάτη και έναν κεντρικό υπολογιστή. Παρόλα αυτά δεν είναι καθόλου σπάνιες οι περιπτώσεις ενός κεντρικού υπολογιστή με πολλούς πελάτες.

Η μέθοδος αυτή είναι η λογική επέκταση του προγραμματισμού ανά μονάδες. Αυτός ο προγραμματισμός παίρνει ως βασική προϋπόθεση ότι ο διαχωρισμός ενός μεγάλου τμήματος λογισμικού στα αρθρώματα που τον αποτελούν (modules)

δημιουργεί τη δυνατότητα για ευκολότερη εξέλιξη και καλύτερη ικανότητα συντήρησης. Η υπολογιστική μέθοδος πελάτης-εξυπηρετητής προχωράει ένα βήμα μακρύτερα αναγνωρίζοντας ότι όλα αυτά τα αρθρώματα δεν χρειάζεται να εκτελεστούν μέσα στον ίδιο χώρο μνήμης. Με αυτή την αρχιτεκτονική η μονάδα που καλεί γίνεται «ο πελάτης» (αυτός που ζητά μία υπηρεσία), και η μονάδα η οποία καλείται γίνεται «ο εξυπηρετητής» (αυτός ο οποίος παρέχει την υπηρεσία).

Η λογική επέκταση αυτού είναι να έχουμε πελάτες και εξυπηρετητές που εργάζονται (τρέχουν) στις κατάλληλες πλατφόρμες υλικού και λογισμικού. Για παράδειγμα, εξυπηρετητές του συστήματος διοίκησης βάσης δεδομένων να τρέχουν σε πλατφόρμες ειδικά σχεδιασμένες και διαμορφωμένες για να εκτελούν αναζητήσεις, ή ακόμα εξυπηρετητές αρχείων να τρέχουν σε πλατφόρμες με ειδικά στοιχεία για τα αρχεία διαχείρισης.

β) Πρωτόκολλο TCP/IP: Το Πρωτόκολλο Ελέγχου Μεταβίβασης (TCP) και το Πρωτόκολλο Διαδικτύου (IP) είναι δύο τελείως διαφορετικά πρωτόκολλα δικτύου εάν μιλήσουμε τεχνικά, όμως συνήθως χρησιμοποιούνται τόσο πολύ μαζί ώστε ο όρος TCP/IP έχει γίνει μόνιμη ορολογία για να αναφερθούμε στο ένα ή και στα δύο πρωτόκολλα.

Το IP αντιστοιχεί στο επίπεδο του δικτύου (επίπεδο 3) στο μοντέλο OS1, ενώ το TCP αντιστοιχεί στο Επίπεδο Μεταφοράς (επίπεδο 4) στο OS1. Με άλλα λόγια ο όρος TCP/IP αναφέρεται στις επικοινωνίες δικτύου όπου η μεταφορά TCP χρησιμοποιείται για να παραδώσει δεδομένα σε όλα τα δίκτυα IP. Το μέσο άτομο στο διαδίκτυο δουλεύει κυρίως σε περιβάλλον TCP/IP. Οι χρήστες του διαδικτύου για παράδειγμα, χρησιμοποιούν το TCP/IP για να επικοινωνήσουν με τους εξυπηρετητές του διαδικτύου. Γενικά, το TCP/IP παρέχει επικοινωνία διασύνδεσης μεταξύ των διαφόρων μηχανημάτων ενός δικτύου ενώ χρησιμοποιείται πολύ και στο διαδίκτυο και σε μικρά τοπικά δίκτυα υπολογιστών.

Το Πρωτόκολλο Διαδικτύου – IP δημιουργήθηκε στη δεκαετία του 1970 για να υποστηρίξει τις πρώτες δικτυώσεις υπολογιστών με το λειτουργικό σύστημα Unix. Σήμερα το IP έχει γίνει μόνιμο για όλα τα λειτουργικά συστήματα δικτύων (network operating systems, NOS), ώστε να επικοινωνούν μεταξύ τους. Πολλά πρωτόκολλα υψηλού επιπέδου όπως το HTTP και το TCP βασίζονται στο IP. Σήμερα υπάρχουν στην παραγωγή δύο παραλλαγές του IP. Σχεδόν όλα τα δίκτυα χρησιμοποιούν την έκδοση 4 του IP (IPv4), αλλά ένας αυξανόμενος αριθμός εκπαιδευτικών και ερευνητικών δικτύων έχουν υιοθετήσει την επόμενη γενεά του IP την έκδοση 6 (IPv6).

γ) Samba: Έχει δοθεί μεγάλη έμφαση στη συνύπαρξη του Unix και του Windows. Ο Οργανισμός Usenix οργάνωσε ένα ετήσιο συνέδριο (LISA/NT 14-17 Ιουλίου ,1999) γύρω από αυτό το θέμα. Δυστυχώς, τα δύο συστήματα προέρχονται από πολύ διαφορετικές κουλτούρες και έχουν δυσκολία να συνεργάζονται χωρίς διαμεσολάβηση, και αυτό, βέβαια είναι δουλειά του Samba. Το Samba τρέχει σε πλατφόρμες Unix, αλλά μιλάει στους πελάτες των Windows ως ιθαγενής. Επιτρέπει σε ένα σύστημα Unix να εισχωρήσει στην «Περιοχή Δικτύου» των Windows χωρίς να προκαλέσει αναταραχή. Οι χρήστες των Windows μπορεί να έχουν εύκολη πρόσβαση στις υπηρεσίες καταχώρησης και εκτύπωσης χωρίς να ξέρουν ή να ενδιαφέρονται εάν αυτές οι υπηρεσίες προσφέρονται από υποδοχείς Unix.

Όλα αυτά τα διαχειρίζεται ένα σχήμα πρωτοκόλλου που τώρα είναι γνωστό ως «Κοινό Σύστημα Καταχώρησης Διαδικτύου», ή CIFS. Αυτό το όνομα χρησιμοποιήθηκε για πρώτη φορά από την Microsoft και προσφέρει ελπίδες για το μέλλον. Στην καρδιά του CIFS βρίσκεται η τελευταία ενσάρκωση του πρωτοκόλλου

SMB (Server Message Block), το οποίον έχει μία μακρά και ανιαρή ιστορία. Το Samba είναι μία ανοικτή πηγή εφαρμογής του CIFS.

Το Samba διατίθεται ελεύθερα, αντίθετα με άλλες εφαρμογές του SMB/CIFS και επιτρέπει την ιδιότητα αλληλολειτουργίας των εξυπηρετητών Linux/Unix και των πελατών που βασίζονται σε Windows.

Πρόκειται για ένα λογισμικό που μπορεί να λειτουργήσει σε άλλη πλατφόρμα εκτός από το Microsoft Windows, για παράδειγμα σε UNIX , Linux, IBM 390, OpenVMS και άλλα λειτουργικά συστήματα. Χρησιμοποιεί πρωτόκολλο TCP/IP που εγκαθίσταται στον υποδοχέα εξυπηρετητή. Όταν διαμορφωθεί σωστά επιτρέπει στον υποδοχέα και αλληλοεπιδρά με έναν πελάτη τηςMicrosoft Windows.

Τι κάνει όμως τελικά το Samba; Το Samba αποτελείται από δύο κύρια προγράμματα: τα smbd και nmbd. Η δουλειά τους είναι να εφαρμόσουν τις 4 βασικές σύγχρονες υπηρεσίες του CIFS οι οποίες είναι:

- Υπηρεσίες καταχώρησης και εκτύπωσης
- Πιστοποίηση και Εξουσιοδότηση
- Ανάλυση και Απόφαση ονόματος
- Υπηρεσία Ανακοινώσεων (αναζήτηση)

Οι υπηρεσίες καταχώρησης και εκτύπωσης είναι βέβαια ο ακρογωνιαίος λίθος των σουϊτών CIFS. Αυτές παρέχονται από smbd. Το SMBD Daemon επίσης χειρίζεται «διαμοιρασμό αρχείων» και την «πιστοποίηση χρηστών». Δηλαδή, μπορεί κανείς να προστατεύσει τις υπηρεσίες κοινής αρχειοθέτησης και εκτύπωσης απαιτώντας προσωπικούς κωδικούς. Στον τρόπο κοινής χρήσης, το απλούστερο αλλά και λιγότερο προτεινόμενο σχέδιο μπορεί να δοθεί ένας κωδικός σε ένα κοινό αρχείο ή εκτυπωτή (που απλά λέγεται «share») (κοινή χρήση). Αυτός ο κωδικός κατόπι δίνεται σε όλους στους οποίους επιτρέπεται να το χρησιμοποιούν. Με τον τρόπο πιστοποίησης ή αναγνώρισης του χρήστη, ο κάθε χρήστης έχει το δικό του όνομα χρήστη και κωδικό και το σύστημα διοίκησης μπορεί να χορηγήσει ή να αρνηθεί πρόσβαση σε ατομική βάση.

Το Σύστημα Περιοχής NT Windows παρέχει ένα περαιτέρω επίπεδο προηγμένης πιστοποίησης για το CIFS. Η βασική ιδέα είναι ότι ένας χρήστης χρειάζεται μόνο να μπει μία φορά στο σύστημα για να έχει πρόσβαση σε όλες τις εξουσιοδοτημένες υπηρεσίες του δικτύου. Το σύστημα Περιοχής NT χειρίζεται αυτή την περίπτωση με έναν εξυπηρετητή πιστοποίησης ταυτότητας που ονομάζεται Ελεγκτής Περιοχής. Μία Περιοχή NT (που δεν πρέπει να την συγχέουμε με το Σύστημα Ονόματος Περιοχής (DNS)) είναι βασικά μία ομάδα μηχανημάτων που μοιράζονται τον ίδιο ελεγκτή Περιοχής.

Έτσι, με τη χρήση του λογισμικού Samba, και θεωρώντας τον Η/Υ του ρομπότ ως τον εξυπηρετητή και το PC των Windows ως πελάτη, μπορούμε να πετύχουμε την πρόσβαση του ενός στο άλλο και την ανάγνωση των απαραίτητων για την εφαρμογή μας αρχείων κειμένου από το ήδη δηλωμένο “Shared” φάκελο των Windows.

Αυτό επιτεύχθηκε με την εγκατάσταση του λογισμικού Samba στο ρομπότ “Damon”, όπου πλέον είναι σε θέση να έχει πρόσβαση σε όλους τους φακέλους “shared” του εκάστοτε Η/Υ που συνδέεται με αυτόν, με την προϋπόθεση ότι ο χρήστης κάνει κάθε φορά τις απαιτούμενες ενέργειες μέσα από την γραμμή εντολών. Συγκεκριμένα, η διαδικασία που πρέπει αν ακολουθήσει ένας χρήστης ενός Η/Υ που τρέχει σε Windows για να πετύχει αυτή την πρόσβαση είναι η ακόλουθη:

1. Ανοίγει ένα command prompt στα windows.
2. Συνδέεται μέσω telnet με το ρομπότ.
3. Συνδέεται ως root.
4. Τρέχει την εντολή: mount -t smbfs "\\\\WIN_IP\Shared " /mnt/test/ -o username=name. Όπου: WIN_IP η διεύθυνση IP του υπολογιστή που τρέχει σε Windows και από τον οποίο είναι συνδεδεμένος, Shared ο φάκελος που έχει ιδιότητες shared φακέλου στον Η/Υ των Windows και name το όνομα του χρήστη ομοίως στον Η/Υ των Windows.
5. Τέλος. Τώρα στον φάκελο /mnt/test στον Η/Υ του ρομπότ μπορεί να δει κανείς όλα τα περιεχόμενα του Shared φακέλου που έχει δηλώσει και να τα διαβάσει μέσα από το Linux του ρομπότ.

6.3.2 Επεξήγηση και Ψευδοκώδικας του Αλγορίθμου Κίνησης

Ο Αλγόριθμος Κίνησης που αναπτύχθηκε σε C++ μπορεί να διαχωριστεί σε δύο κατηγορίες κινήσεων:

α) Όταν ο χρήστης δίνει μια από τις εντολές: “Μπροστά”, “Πίσω”, “Δεξιά”, ”Αριστερά“, ”Σταμάτα“. Στην περίπτωση αυτή το πρόγραμμα σε συνεργασία με το λογισμικό Mobility καλεί τις απαραίτητες συναρτήσεις για την κίνηση του ρομπότ αντίστοιχα μπροστά, πίσω, στροφή δεξιά, στροφή αριστερά και τερματισμός κίνησης.

β) Όταν ο χρήστης δίνει δύο αριθμούς που αντιστοιχούν στις συντεταγμένες x και y του χώρου του ρομπότ. Οι αριθμοί αυτοί μπορεί να είναι από το 1 έως και το 9 και έχουν ως φυσική μονάδα το μέτρο. Δίνονται πάντα ως ζεύγη από τον χρήστη ακόμα και αν η μετάβαση σε μια διάσταση είναι μηδενική.

Δεδομένου ότι το πρόγραμμα αναγνώρισης φωνής δέχεται μόνο μία λέξη κάθε φορά προς αναγνώριση, η επίτευξη αναγνώρισης συνδυασμού λέξεων επαφίεται αποκλειστικά στον αλγόριθμο που αναπτύχθηκε. Έτσι, καταφέραμε να αναγνωρίζουμε τους εξής συνδυασμούς λέξεων στηριζόμενοι στη λογική ότι μπορούμε να επαναλάβουμε την διαδικασία της αναγνώρισης όσες φορές επιθυμούμε:

- Συνδυασμό δύο οποιοδήποτε αριθμών από το 0 έως το 9. Έτσι, ο πρώτος αριθμός που θα ειπωθεί καταχωρείται στην μεταβλητή x, ενώ ο δεύτερος στην μεταβλητή y. Οι μεταβλητές x και y εκφράζουν τις συντεταγμένες του επιπέδου κίνησης (x,y) στις οποίες θέλουμε να πάει το όχημα.

- Συνδυασμό των λέξεων «μπροστά», «πίσω», «δεξιά» και «αριστερά» ως πρώτη λέξη με την λέξη «σταμάτα» ως δεύτερη. Με τον τρόπο αυτό το όχημα εκτελεί μόνο μια εκ των τεσσάρων αυτών κινήσεων και μετά σταματάει.

- Συνδυασμό μίας εκ των λέξεων «μπροστά» και «πίσω» ως πρώτη λέξη με μία εκ των «δεξιά» και «αριστερά» ως δεύτερη. Έτσι, το ρομπότ εκτελεί συνδυασμό στροφής, είτε προς τα δεξιά είτε προς τα αριστερά, και γραμμικής κίνησης, προς τα μπροστά ή προς τα πίσω.

Έτσι, στον Πίνακα 6.2 παρουσιάζονται όλοι οι επιτρεπτοί συνδυασμοί εκφώνησης λέξεων στο πρόγραμμα αναγνώρισης φωνής με την σειρά που αναγράφονται, καθώς και τι αποτέλεσμα έχει ο κάθε συνδυασμός όσον αφορά στην κίνηση του ρομπότ.

Επιτρεπτός Συνδυασμός Ηχογράφησης Λέξεων	**Περιγραφή Κίνησης Οχήματος**
Μπροστά Σταμάτα	Κινείται προς τα μπροστά σε ευθεία γραμμή για περίπου 8 sec και διανύει διάστημα περίπου 1.75 m
Πίσω Σταμάτα	Κινείται προς τα πίσω σε ευθεία γραμμή για περίπου 8 sec και διανύει διάστημα περίπου 1.75 m
Δεξιά Σταμάτα	Εκτελεί στροφή προς τα δεξιά περίπου 90^0 για περίπου 8 sec.
Αριστερά Σταμάτα	Εκτελεί στροφή προς τα αριστερά περίπου 90^0 για περίπου 8 sec.
Μπροστά Δεξιά	Εκτελεί συνδυασμό κίνησης δεξιάς στροφής και ευθύγραμμης εμπρόσθιας για περίπου 67 sec.
Πίσω Δεξιά	Εκτελεί συνδυασμό κίνησης δεξιάς στροφής και ευθύγραμμης οπίσθιας για περίπου 7 sec.
Μπροστά Αριστερά	Εκτελεί συνδυασμό κίνησης αριστερής στροφής και ευθύγραμμης εμπρόσθιας για περίπου 7 sec.
Πίσω Αριστερά	Εκτελεί συνδυασμό κίνησης αριστερής στροφής και ευθύγραμμης οπίσθιας για περίπου 7 sec.
Όλοι οι συνδυασμοί από το 0 έως το 9 ανά δύο	Εκτελεί όλες τις απαιτούμενες κινήσεις για να βρεθεί τελικά στο σημείο (x,y) που ορίζεται μονοσήμαντα από τους δύο αριθμούς.

Πίνακας 6.2: Περιγραφή επιτρεπτών ηχογραφήσεων και κινήσεις που αυτές αντιστοιχούν

Πρέπει να τονιστεί στο σημείο αυτό ότι οι τιμές του Πίνακα 6.2 έχουν προκύψει από πειραματικές μετρήσεις στο χώρο του εργαστηρίου ρομποτικής και συνεπώς τυχόν αποκλίσεις από τις τιμές αυτές οφείλονται σε παράγοντες όπως η ολισθηρότητα του εδάφους, η τριβή, καθώς και η εναλλαγή υλικών στο έδαφος.

Σύμφωνα με την λογική που αναλύθηκε αμέσως παραπάνω, ακολουθεί ο ψευδοκώδικας μόνο του αλγορίθμου κίνησης, δεδομένου ότι έχει τελειώσει η διαδικασία της αναγνώρισης και έχει εξαχθεί το αντίστοιχο αρχείο:

- Διάβασε αρχείο κειμένου txt ➔ σώσε το περιεχόμενο ως “λέξη”

- ΑΝ “λέξη” = κάποιος αριθμός από 0 έως 9
  ΤΟΤΕ δώσε στην μεταβλητή “command” την τιμή 0
  ΑΝ συντεταγμένη **x** ελεύθερη
  x=“λέξη”
  Διαφορετικά ΑΝ συντεταγμένη x δεσμευμένη
  y=“λέξη”
  Επανέλαβε την διαδικασία από την αρχή μέχρι να δεσμευτούν **x** και **y**
  Τέλος ΑΝ

- ΑΝ "λέξη" = κάποια λέξη από τις "μπροστά", "πίσω", "δεξιά" ή "αριστερά" ΚΑΙ αν αυτή είναι η **πρώτη** "λέξη" που εκφωνείται, τότε:

  - ΑΝ "λέξη" = "μπροστά"
    ΤΟΤΕ δώσε στην μεταβλητή "command" την τιμή 1
    Τέλος ΑΝ

  - ΑΝ "λέξη" = "πίσω"
    ΤΟΤΕ δώσε στην μεταβλητή "command" την τιμή 2
    Τέλος ΑΝ

  - ΑΝ "λέξη" = "δεξιά"
    ΤΟΤΕ δώσε στην μεταβλητή "command" την τιμή 3
    Τέλος ΑΝ

  - ΑΝ "λέξη" = "αριστερά"
    ΤΟΤΕ δώσε στην μεταβλητή "command" την τιμή 4
    Τέλος ΑΝ

  Τέλος ΑΝ

- ΑΝ "λέξη" = κάποια λέξη από τις "δεξιά" "αριστερά", ή "σταμάτα" ΚΑΙ αν αυτή είναι η **δεύτερη** "λέξη" που εκφωνείται, τότε:

  Βρες ποια λέξη είναι με αλλεπάλληλους ελέγχους
  Αντιστοίχισε μια μεταβλητή "command" στον συγκεκριμένο συνδυασμό λέξεων
  ΤΟΤΕ δώσε στην μεταβλητή "command" την αντίστοιχη τιμή
  Κάλεσε πρόγραμμα για αντίστοιχη κίνηση (με δεδομένη γραμμική ή γωνιακή ταχύτητα)
  Δώσε ως είσοδο το "command"

  Δηλαδή:

  - ΑΝ ο συνδυασμός λέξεων είναι "μπροστά σταμάτα":
    ΤΟΤΕ δώσε στην μεταβλητή "command" την τιμή 1
    Κάλεσε πρόγραμμα για κίνηση μόνο μπροστά (με δεδομένη γραμμική ταχύτητα)
    Δώσε ως είσοδο το "command"
    Τέλος ΑΝ

  - ΑΝ ο συνδυασμός λέξεων είναι "πίσω σταμάτα":
    ΤΟΤΕ δώσε στην μεταβλητή "command" την τιμή 2
    Κάλεσε πρόγραμμα για κίνηση μόνο πίσω (με δεδομένη γραμμική ταχύτητα)
    Δώσε ως είσοδο το "command"
    Τέλος ΑΝ

  - ΑΝ ο συνδυασμός λέξεων είναι "δεξιά σταμάτα":
    ΤΟΤΕ δώσε στην μεταβλητή "command" την τιμή 3
    Κάλεσε πρόγραμμα για κίνηση μόνο στροφής δεξιάς (με δεδομένη γωνιακή ταχύτητα)
    Δώσε ως είσοδο το "command"
    Τέλος ΑΝ

- ΑΝ ο συνδυασμός λέξεων είναι “αριστερά σταμάτα”:
  ΤΟΤΕ δώσε στην μεταβλητή “command” την τιμή 4
  Κάλεσε πρόγραμμα για κίνηση μόνο στροφής αριστερής (με δεδομένη γωνιακή ταχύτητα)
  Δώσε ως είσοδο το “command”
  Τέλος ΑΝ

- ΑΝ ο συνδυασμός λέξεων είναι “μπροστά δεξιά”:
  ΤΟΤΕ δώσε στην μεταβλητή “command” την τιμή 5
  Κάλεσε πρόγραμμα για κίνηση στροφής δεξιάς με κίνηση προς τα μπροστά (με δεδομένη γραμμική και γωνιακή ταχύτητα)
  Δώσε ως είσοδο το “command”
  Τέλος ΑΝ

- ΑΝ ο συνδυασμός λέξεων είναι “πίσω δεξιά”:
  ΤΟΤΕ δώσε στην μεταβλητή “command” την τιμή 6
  Κάλεσε πρόγραμμα για κίνηση στροφής δεξιάς με κίνηση προς τα πίσω (με δεδομένη γραμμική και γωνιακή ταχύτητα)
  Δώσε ως είσοδο το “command”
  Τέλος ΑΝ

- ΑΝ ο συνδυασμός λέξεων είναι “μπροστά αριστερά”:
  ΤΟΤΕ δώσε στην μεταβλητή “command” την τιμή 7
  Κάλεσε πρόγραμμα για κίνηση στροφής αριστερής με κίνηση προς τα μπροστά (με δεδομένη γραμμική και γωνιακή ταχύτητα)
  Δώσε ως είσοδο το “command”
  Τέλος ΑΝ

- ΑΝ ο συνδυασμός λέξεων είναι “πίσω αριστερά”:
  ΤΟΤΕ δώσε στην μεταβλητή “command” την τιμή 8
  Κάλεσε πρόγραμμα για κίνηση στροφής αριστερής με κίνηση προς τα πίσω (με δεδομένη γραμμική και γωνιακή ταχύτητα)
  Δώσε ως είσοδο το “command”
  Τέλος ΑΝ

Τέλος ΑΝ

- ΑΝ x και y δεσμευμένες
  Κάλεσε πρόγραμμα κίνησης του ρομπότ μέσω συντεταγμένων
  Δώσε ως είσοδο τις συντεταγμένες
  Τέλος ΑΝ

- ΑΝ “λέξη” = ΑΓΝΩΣΤΗ
  Τότε επανέλαβε την διαδικασία από την αρχή
  Τέλος ΑΝ

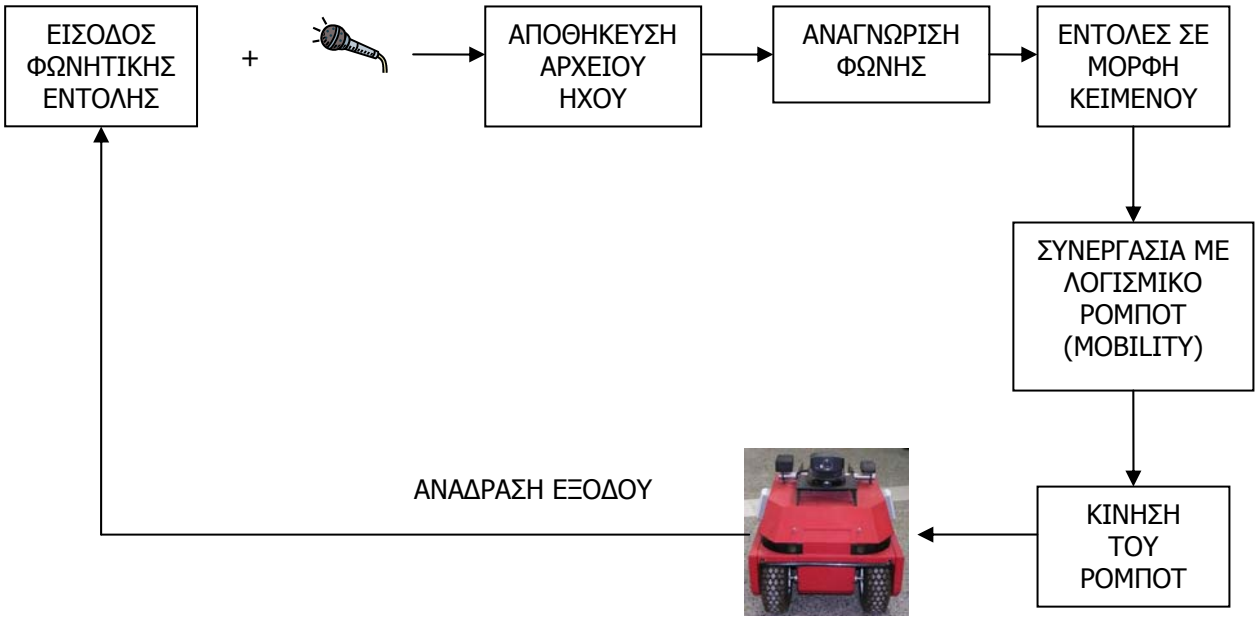
- Τέλος Προγράμματος.

ΚΕΦΑΛΑΙΟ 7 : ΕΦΑΡΜΟΓΗ ΠΡΟΓΡΑΜΜΑΤΩΝ ΑΝΑΓΝΩΡΙΣΗΣ ΚΑΙ ΚΙΝΗΣΗΣ

7.1 ΔΟΜΙΚΑ ΤΜΗΜΑΤΑ ΣΥΣΤΗΜΑΤΟΣ

Έχοντας ήδη μελετήσει κανείς τα προηγούμενα κεφάλαια μπορεί να καταλήξει εύκολα στο συμπέρασμα ότι η διαδικασία της αναγνώρισης δεν είναι μια απλή διαδικασία, αλλά μια σύνθετη που για να επιτευχθεί πρέπει να αναλυθεί σε περισσότερες απλές, οι οποίες λαβαίνουν χώρα σειριακά. Έτσι, μπορούμε να καταλήξουμε σε κάποια δομικά τμήματα που απαρτίζουν όλη την εφαρμογή και στα οποία μπορεί να χωριστεί το σύστημα, και είναι τα ακόλουθα:

α) *Το Πρόγραμμα Ηχογράφησης και Αποθήκευσης του Αρχείου (Βλ. Κεφάλαιο 5)*
β) *Το Πρόγραμμα Αναγνώρισης Φωνής (Βλ. Κεφάλαιο 5)*
γ) *Το Πρόγραμμα Κίνησης (Βλ. Κεφάλαιο 6)*
δ) *Μικρόφωνο (Βλ. Κεφάλαιο 5)*
ε) *Ρομπότ (Βλ. Κεφάλαιο 2)*
στ) *Χειριστής υπολογιστή (άνθρωπος)*

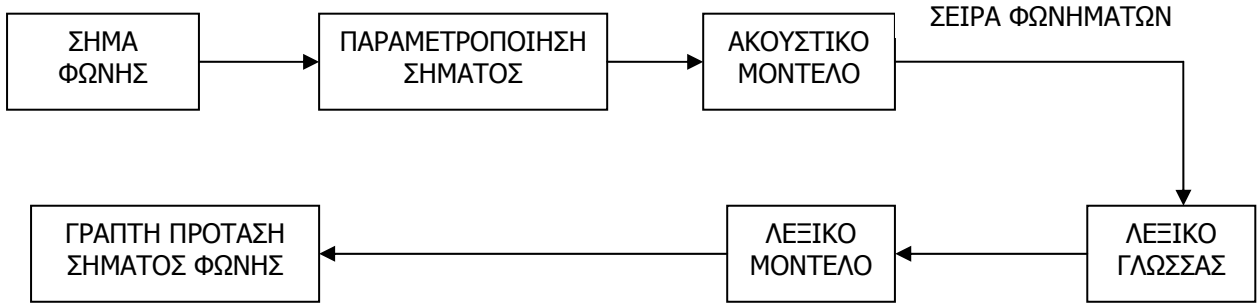


Σχήμα 7.1: Δομικό Διάγραμμα Αλληλεπίδρασης Ανθρώπου-Ρομπότ στην Αναγνώριση Φωνής

Τα τρία πρώτα είναι τα προγράμματα που χρησιμοποιήθηκαν για την επίτευξη του στόχου της εργασίας (Κεφάλαιο 1), ενώ τα τρία τελευταία αποτελούν τα απαραίτητα φυσικά εξαρτήματα (hardware) και η απαραίτητη ανθρώπινη παρουσία στην

διαδικασία της αναγνώρισης. Η αλληλεπίδραση αυτών των δομικών στοιχείων φαίνεται στο Σχήμα 7.1 που απεικονίζεται το Δομικό Διάγραμμα της εργασίας στο σύνολο της.

Το σύστημα αναγνώρισης φωνής, όπως άλλωστε είδαμε και στο Κεφάλαιο 3 μπορεί να αναλυθεί σε ένα άλλο δομικό διάγραμμα με ακόμα περισσότερα δομικά μέρη, η αλληλεπίδραση των οποίων, αν και θα πρέπει πλέον να είναι γνωστή στον αναγνώστη, φαίνεται στο Σχήμα 7.2.



Σχήμα 7.2: Δομικό Διάγραμμα Διαδικασίας Αναγνώρισης Φωνής

7.2 ΠΕΡΙΓΡΑΦΗ ΕΦΑΡΜΟΓΗΣ ΣΕ ΠΡΑΓΜΑΤΙΚΟ ΧΡΟΝΟ

Στο σημείο αυτό θα παρουσιάσουμε συνοπτικά την εφαρμογή της αναγνώρισης με βήματα, όπως αυτή πραγματοποιείται σε πραγματικό χρόνο. Ο σκοπός αυτής της παραγράφου είναι να εξοικειωθεί ο αναγνώστης με την σειρά εκτέλεσης των εργασιών και το γραφικό περιβάλλον.

Βήμα 1: Εκκίνηση, από τον χρήστη, του προγράμματος κίνησης στο ρομπότ. Το πρόγραμμα ανοίγει το αρχείο για ανάγνωση, που θα πάρει από το πρόγραμμα αναγνώρισης. Όσο το αρχείο είναι κενό μπαίνει σε αναμονή έως ότου αυτό να πάρει στο εσωτερικό του έστω και μια εκ των αναμενόμενων τιμών.

Βήμα 2: Εκκίνηση, από τον χρήστη, του προγράμματος της C++ στον Η/Υ που τρέχει σε Windows.

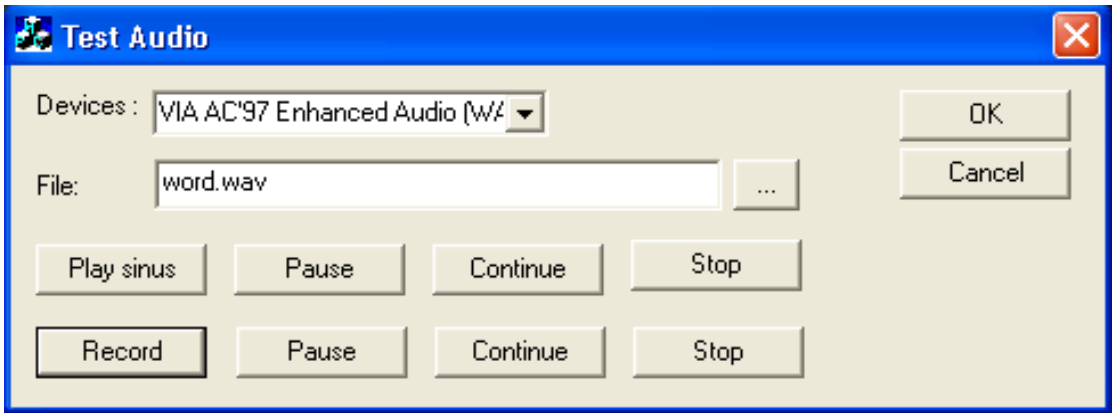
Βήμα 3: Πραγματοποιείται αυτόματα εκκίνηση, μέσα από το προαναφερθέν πρόγραμμα, του προγράμματος ηχογράφησης. Τότε ανοίγει το παράθυρο του Σχήματος 7.3, από το οποίο ο χρήστης μπορεί να επιλέξει να ηχογραφήσει (Record), να παύσει για λίγο την ηχογράφηση (Pause), να την επανεκκινήσει (Continue) και να την διακόψει (Stop). Ακόμα έχει δυνατότητα επιλογής κάρτας ήχου (Devices), σε περίπτωση ύπαρξης περισσοτέρων από μία. Τέλος, μπορεί να ανοίξει ένα αρχείο (File) για να το παίξει (Play sinus). Όταν τελειώσει με όλες τις ενέργειες που επιθυμεί πατάει το OK για να εξέλθει ή το Cancel αν θέλει να ακυρώσει την διαδικασία.

Βήμα 4: Αυτόματη εκκίνηση του προγράμματος της αναγνώρισης μέσα από το πρόγραμμα του Βήματος 2. Στο σημείο αυτό ανοίγει το παράθυρο του Σχήματος 7.4, από το οποίο ο χρήστης μπορεί να επιλέξει το όνομα του αρχείου εισόδου (.wav) καθώς και το όνομα του αρχείου εξόδου (.txt). Και τα δύο αυτά αρχεία θα βρίσκονται στον ίδιο φάκελο με αυτόν του εκτελέσιμου προγράμματος της αναγνώρισης. Στο

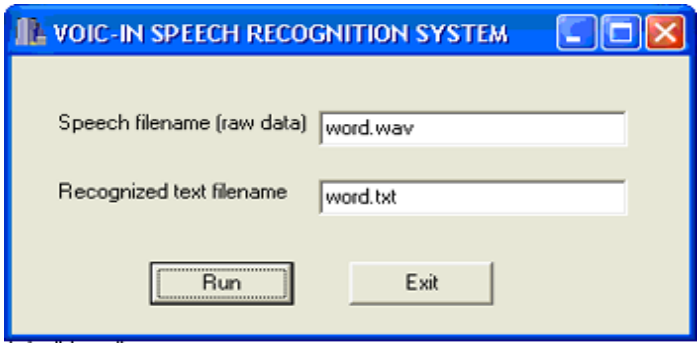
σημείο αυτό ο χρήστης πρέπει να σιγουρευτεί ότι έχει ενσωματώσει το USB HASP Drive στον Η/Υ που τρέχει το πρόγραμμα. Διαφορετικά, ο μεταφραστής θα βγάλει μήνυμα λάθους. Πατώντας “Run” τρέχει το πρόγραμμα, ενώ με το κουμπί “Exit” εξέρχεται ο χρήστης από το προγραμματιστικό περιβάλλον της αναγνώρισης.

Βήμα 5: Στο σημείο αυτό, ελέγχει το πρόγραμμα του Βήματος 2 αν το αρχείο εξόδου έχει αποθηκεύσει έγκυρα δεδομένα και αναλόγως είτε εκκινεί την διαδικασία ηχογράφησης και αναγνώρισης από το Βήμα 3, είτε κωδικοποιεί τα δεδομένα του αρχείου και τερματίζει πηγαίνοντας στο Βήμα 6.

Βήμα 6: Αποτελεί το τελευταίο βήμα του αλγορίθμου. Πραγματοποιείται ενημέρωση του Βήματος 1 και έτσι το πρόγραμμα του ρομπότ παίρνει τα απαραίτητα δεδομένα και τίθεται στην αντίστοιχη κίνηση.

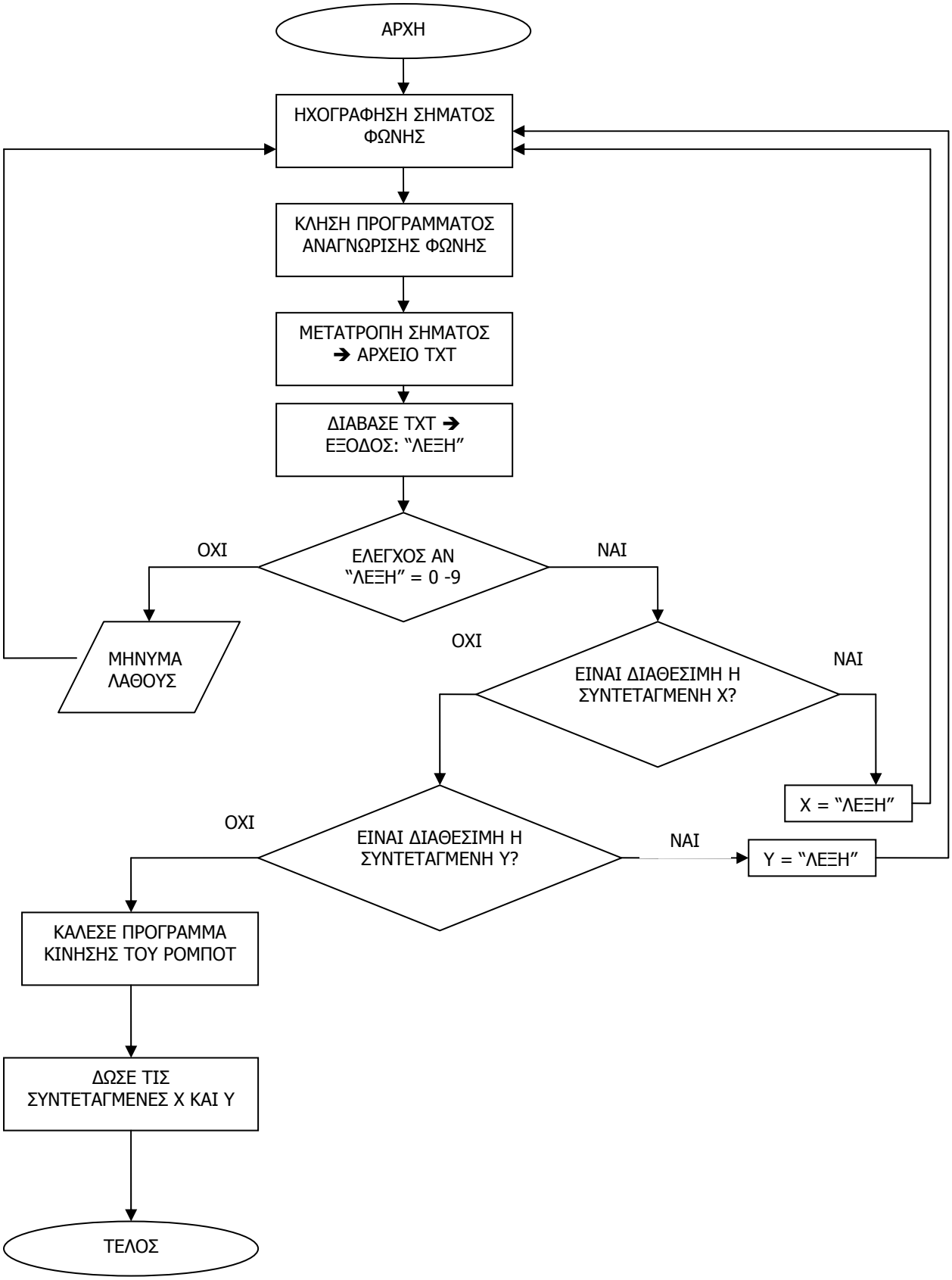


Σχήμα 7.3: Γραφικό Περιβάλλον Προγράμματος Ηχογράφησης

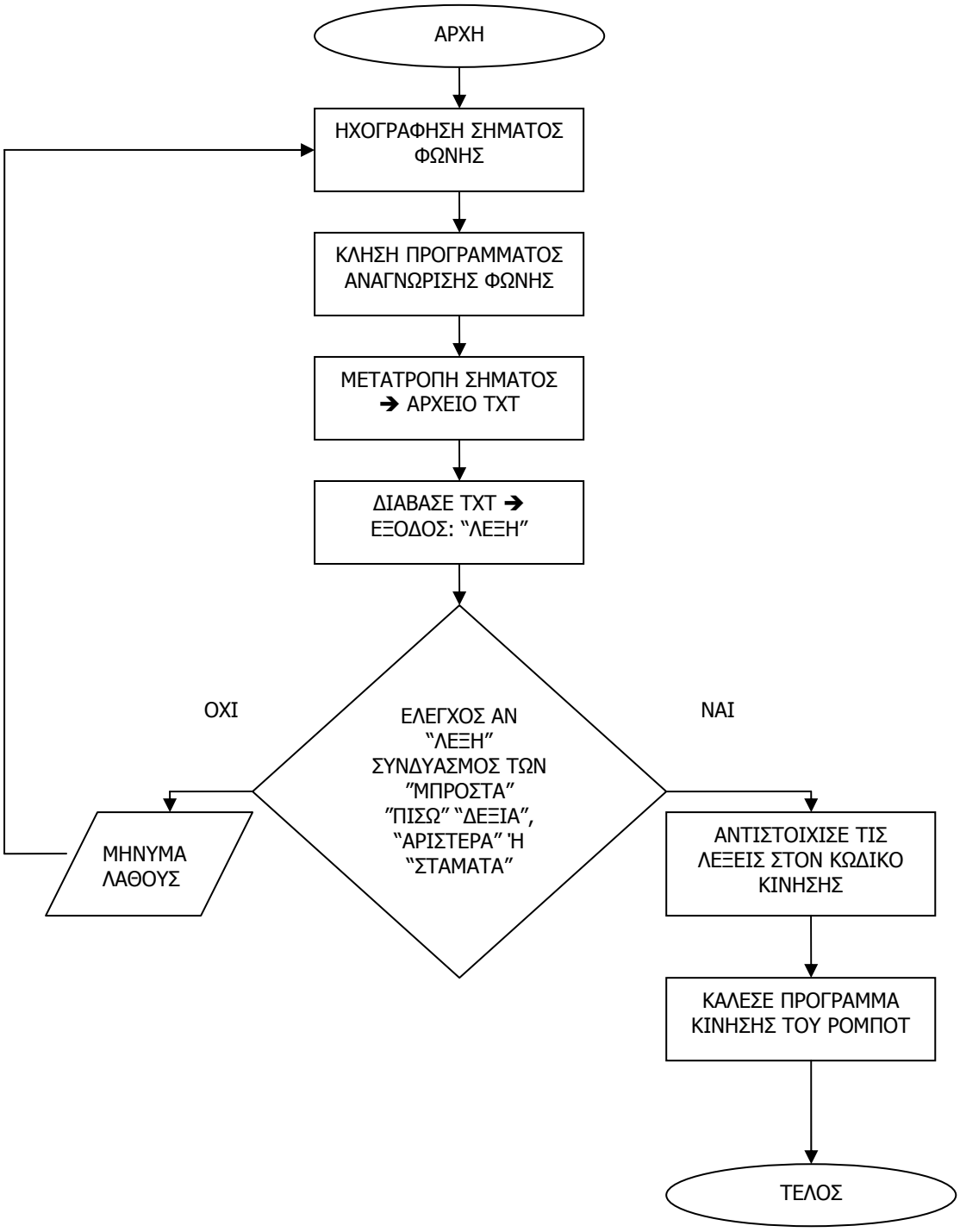


Σχήμα 7.4: Γραφικό Περιβάλλον Προγράμματος Αναγνώρισης Φωνής

Παρατίθενται ακόμα και δύο διαγράμματα ροής της παραπάνω διαδικασίας. Το διάγραμμα του Σχήματος 7.5 αναφέρεται στην περίπτωση που ο χρήστης θέλει να κινήσει το όχημα σε συγκεκριμένο σημείο του χώρου, οπότε και εκφωνεί συντεταγμένες x και y. Αντίστοιχα, το διάγραμμα του Σχήματος 7.6 περιγράφει την διαδικασία της αναγνώρισης στην περιπτώσεις κίνησης «μπροστά», «πίσω», «δεξιά», «αριστερά», «σταματήματος» και συνδυασμού αυτών.



Σχήμα 7.5: Διάγραμμα Ροής Συνολικής Διαδικασίας για Κίνηση μέσω Συντεταγμένων



Σχήμα 7.6: Διάγραμμα Ροής Συνολικής Διαδικασίας για Κίνηση μέσω Δεσμευμένων Εντολών

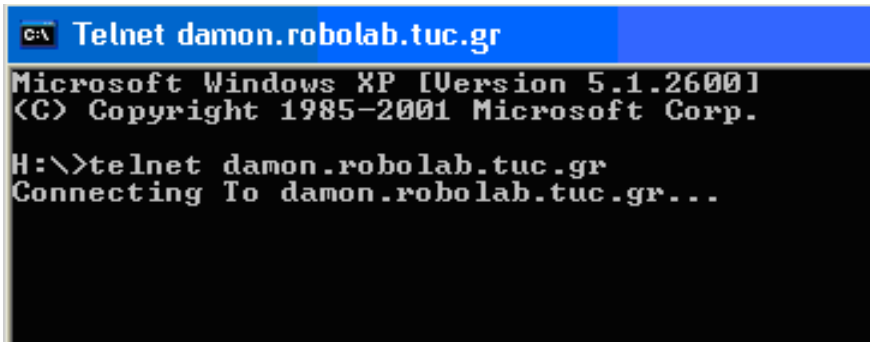
7.3 ΠΑΡΑΔΕΙΓΜΑ ΕΦΑΡΜΟΓΗΣ

Στο σημείο αυτό παρατίθεται ένα παράδειγμα εφαρμογής του προγράμματος αναγνώρισης φωνής στο ρομπότ ATRV-Mini σε πραγματικό χρόνο και κάτω από τις συνθήκες εσωτερικού χώρου (συγκεκριμένα του εργαστηρίου Ρομποτικής). Το παράδειγμα θα παρουσιαστεί με την μορφή βημάτων για καλύτερη κατανόηση.

Είναι σκόπιμο να υπενθυμίσουμε εδώ στον αναγνώστη ότι μπορεί να εκφωνεί και ηχογραφεί μία λέξη την φορά. Οπότε, δεδομένου ότι όλες οι επιτρεπτές κινήσεις απαιτούν την ηχογράφηση και αναγνώριση δύο λέξεων, θα του ζητηθεί δύο φορές από το πρόγραμμα αναγνώρισης η εκφώνηση μιας λέξης.

Το παράδειγμα που θα παρουσιάσουμε θα έχει ως αποτέλεσμα την στροφή του ATRV Mini προς τα δεξιά κατά περίπου 90^0. Τυχόν απώλειες στην τιμή αυτή οφείλονται σε ποικίλους παράγοντες (όπως η τριβή μεταξύ του ρομπότ και του εδάφους). Η εφαρμογή έγινε από τον χρήστη administrator ενός H/Y Laptop Centrino στα 1700 MHz που τρέχει σε Windows XP συνδεδεμένο στο δίκτυο του Robolab με IP address 147.27.9.54.

ΒΗΜΑ 1: Σύνδεση του Laptop με τον H/Y του ATRV-Mini Damon μέσω telnet με την εντολή: telnet damon.robolab.tuc.gr (Σχήμα7.7) δίνοντας τα απαραίτητα στοιχεία για την σύνδεση. Αλλαγή σε root user δίνοντας επίσης τα απαραίτητα στοιχεία.



Σχήμα 7.7: Σύνδεση με το ρομπότ μέσω telnet

ΒΗΜΑ 2: Στο ίδιο παράθυρο telnet εκκινούμε τον name server και το base του Damon.

ΒΗΜΑ 3: Σε νέο παράθυρο telnet συνδεόμαστε πάλι ως root και για να φορτώσουμε το πρόγραμμα Samba στον φάκελο shared που θέλουμε γράφουμε: mount -t smbfs "\\\\147.27.9.54\Shared Documents " /mnt/test/ -o username=administrator.

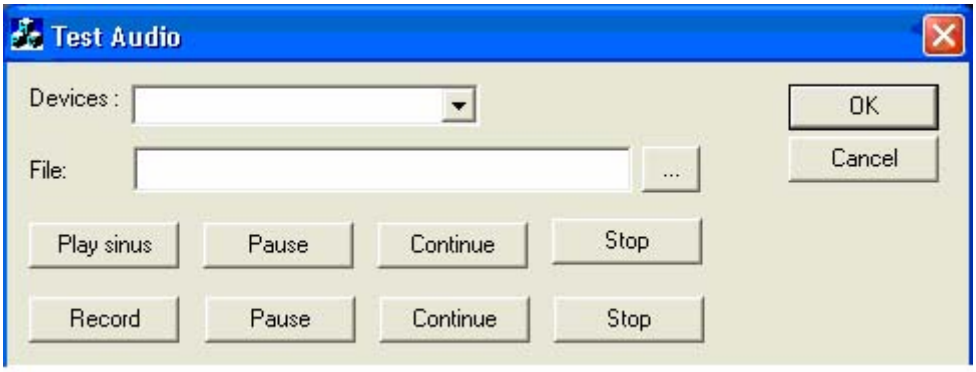
ΒΗΜΑ 4: Στο δεύτερο παράθυρο που έχουμε ανοίξει εκκινούμε το πρόγραμμα της κίνησης του ρομπότ με την εντολή: speechrobot –robot ATRVMini.

ΒΗΜΑ 5: Εκκινούμε μέσα από τα Windows και τον compiler της Visual C++ το πρόγραμμα αναγνώρισης φωνής που έχουμε δομήσει. Το παράθυρο του Σχήματος 7.8 εμφανίζεται. Στο παράθυρο αυτό πατάμε εκτελούμε την εξής διαδικασία:

- Πάτημα κουμπιού Record
- Εισαγωγή στο μικρόφωνο της λέξης «Δεξιά»

- Πάτημα του κουμπιού Stop
- Πάτημα του κουμπιού OK

Τώρα η λέξη «Δεξιά» έχει αποθηκευτεί ως word.wav στον ίδιο φάκελο με τα λοιπά προγράμματα.



Σχήμα 7.8: Πρόγραμμα ηχογράφησης

ΒΗΜΑ 6: Εμφανίζεται αυτομάτως το πλαίσιο του Σχήματος 7.4, δηλαδή το πρόγραμμα αναγνώρισης φωνής, που δείχνει ότι θα πάρει ως είσοδο το αρχείο word.wav και ότι θα δώσει ως έξοδο το word.txt. Στο σημείο αυτό πατάμε διαδοχικά τα κουμπιά Run και Exit.

ΒΗΜΑ 7: Το βήμα αυτό είναι επανάληψη του βήματος 5. Δηλαδή εμφανίζεται το ίδιο πλαίσιο (Σχήμα 7.8) και εκτελούμε τα ίδια βήματα, μόνο που τώρα ηχογραφούμε την λέξη «Σταμάτα».

ΒΗΜΑ 8: Το βήμα αυτό είναι επανάληψη του βήματος 6. Δηλαδή εμφανίζεται το πλαίσιο του Σχήματος 7.4 και ακολουθούμε την ίδια διαδικασία.

ΒΗΜΑ 9: Το ρομπότ παίρνει τα δεδομένα από τα προγράμματα που έτρεξαν και εκτελεί την κίνηση που φαίνεται στο Σχήμα 7.9, όπου φαίνεται η αρχική θέση του ρομπότ, μια τυχαία ενδιάμεση θέση και η τελική του θέση.

Παρατηρούμε ότι το όχημα έχει στρέψει κατά περίπου 90^0 με μία απόκλιση της τάξης των 5^0-10^0, η οποία ουσιαστικά οφείλεται στην εναλλαγή του υλικού του δαπέδου, όπως εύκολα παρατηρεί κανείς στο Σχήμα 7.9.

Σχήμα 7.9: Τρία στιγμιότυπα κατά την διάρκεια της δεξιάς στροφής του ATRV Mini.

7.4 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΤΑΣΕΙΣ

Στην εργασία αυτή ασχοληθήκαμε με την ανάπτυξη ενός μοντέλου αναγνώρισης φωνής και την εφαρμογή αυτού σε ρομποτικό όχημα. Το σύστημα αναγνώρισης φωνής που χρησιμοποιήθηκε ανήκει στην κατηγορία συστημάτων αναγνώρισης συνεχούς ομιλίας και μεγάλου λεξιλογίου. Στη συγκεκριμένη εφαρμογή παρόλα αυτά έγινε χρήση μερικών μόνο δυνατοτήτων του προαναφερθέντος προγράμματος.

7.4.1 Συμπεράσματα

Το σημαντικότερο πλεονέκτημα της μεθόδου που εφαρμόσθηκε είναι η αξιοπιστία του προγράμματος αναγνώρισης, κάτι που παρατηρήθηκε από την ιδιαίτερα μικρή απόκλιση μεταξύ των επιθυμητών και των τελικών πραγματικών αποτελεσμάτων. .Βέβαια, αυτό ήταν αναμενόμενο, αν λάβει κανείς υπόψη, ότι τέτοια συστήματα χαρακτηρίζονται από την ελαχιστοποίηση του σφάλματος αναγνώρισης.

Αξιοσημείωτη είναι η έλλειψη συστημάτων αναγνώρισης φωνής στην αγορά. Μελέτες δείχνουν ότι ενώ στα πρώτα στάδια ανάπτυξης του κλάδου αυτού τα συστήματα αναγνώρισης είχαν κατακλείσει την αγορά, τα τελευταία χρόνια μόνον ελάχιστα εμπορικά συστήματα είναι αυτά που αξίζουν.

Το σύστημα αναγνώρισης που χρησιμοποιήθηκε είναι το πρώτο σύστημα με ελληνικό λεξικό και που υποστηρίζει χιλιάδες λέξεις της ελληνικής γλώσσας. Γενικότερα η πλειοψηφία των εμπορικών συστημάτων υποστηρίζει φωνήματα-λέξεις της Αγγλικής γλώσσας.

Το βασικότερο μειονέκτημα της υλοποίησης του προγράμματος είναι η αδυναμία αναγνώρισης περισσότερων των μία λέξεων, κάθε φορά που τρέχει το πρόγραμμα. Έτσι, αναγκαστήκαμε να καταφύγουμε σε ευρετικές μεθόδους για την επίτευξη συνδυασμού λέξεων.

Τέλος, αξίζει να σημειωθεί η αδυναμία ανατροφοδότησης της τελικής θέσης του ρομπότ, με αποτέλεσμα να μη μπορεί να υπολογισθούν επακριβώς τυχόν σφάλματα στην κίνηση του οχήματος.

7.4.2 Μελλοντικές Προτάσεις

Μία πιθανή εξέλιξη της παρούσας εργασίας είναι η επιλογή ενός άλλου προγράμματος αναγνώρισης ή η προσαρμογή αυτού έτσι ώστε να επιτρέπει την αναγνώριση περισσότερων λέξεων και συνεχόμενου λόγου. Με τον τρόπο αυτό θα μπορεί να προσαρμόζεται στις εκάστοτε ανάγκες του κάθε χρήστη.

Παράλληλα, θα ήταν ενδιαφέρουσα η προοπτική πραγματοποίησης του συστήματος αναγνώρισης σε πραγματικό χρόνο με την κίνηση του ρομπότ, χωρίς δηλαδή να απαιτείται καμία παρέμβαση από το χρήστη-ομιλητή. Η επίτευξη αυτού καθιστά το σύστημα περισσότερο ευέλικτο , δεδομένου ότι ο χρήστης-ομιλητής μπορεί να βρίσκεται μακριά από τον υπολογιστή κατά τη διάρκεια της αναγνώρισης και κίνησης του οχήματος. Έτσι διευκολύνεται η χρήση του συστήματος σε εξωτερικούς χώρους.

Με περισσότερη έμφαση στον τομέα της κίνησης και λιγότερη σε αυτόν της αναγνώρισης φωνής, θα μπορούσε να εξελιχθεί το σύστημα κίνησης και αναγνώρισης με σκοπό την αποφυγή εμποδίων, καθώς και τη δημιουργία ενός γραφικού περιβάλλοντος διεπαφής με δυνατότητα αναπαράστασης της κίνησης του οχήματος σε πραγματικό χρόνο.

Η τελειοποίηση του συγκεκριμένου συστήματος θα ήταν η αλληλεπίδραση μεταξύ του ρομπότ και του χρήστη (π.χ. σε μορφή διαλόγου). Με τον τρόπο αυτό επιτυγχάνεται ο μέγιστος βαθμός ευελιξίας και ελαχιστοποίηση τυχόν σφαλμάτων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] IS Robotics, Inc., *ATRV-Mini All-Terrain Mobile Robot User's Guide*, Real World Interface Division, 2000.

[2] Μ. Διαμαντάκης, "Ανάπτυξη αναδρομικού αλγόριθμου πλοήγησης με τη βοήθεια συσκευών υπερήχου και χαρτογράφησης δωματίου με το έντροχο ρομπότ ATRV Mini", Διπλωματική εργασία, Τμήμα Μηχανικών Παραγωγής και Διοίκησης, Πολυτεχνείο Κρήτης, 2004.

[3] Α. Τσαλατσάνης, "Οπτικό σύστημα του ATRV-Mini, λειτουργικότητα και εφαρμογές", Διπλωματική εργασία, Τμήμα Μηχανικών Παραγωγής και Διοίκησης, Πολυτεχνείο Κρήτης, 2001.

[4] Β. Δουμπιώτης, "Αναγνώρισης Φωνής με τεχνικές κανονικοποίησης", Διπλωματική Εργασία, Τμήμα Ηλεκτρονικών και Μηχανικών Υπολογιστών, Πολυτεχνείο Κρήτης, 1998.

[5] E. Keller, *Fundamentals of Speech Synthesis and Speech Recognition*, University of Lausanne, Switzerland, 1994.

[6] Fact Index, http://www.fact-index.com.

[7] Brainy Encyclopedia Home, http://www.brainyencyclopedia.com.

[8] L. Rabiner & B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

[9] ScanSoft Inc., επίσημη ιστοσελίδα προϊόντος Dragon Naturally Speaking, http://www.dragon-medical-transcription.com.

[10] M. Weintraub et al., "Linguistic Constraints in Hidden Markov Models Based Speech Recognition", Proc. ICASSP 89, Glasgow, Scotland, May 1989.

[11] C. H. Lee, L. R. Rabiner, R. Pieraccini and J. G. Wilpon, "Acoustic Modelling for Large Vocabulary Speech Recognition", *Computer Speech and Language 4*, 1990.

[12] Sony Corporation, επίσημη ιστοσελίδα του AIBO ρομπότ, http://www.aibo.com.

[13] NEC Corporation, NEC Personal Robot Center, http://www.incx.nec.co.jp/robot.

[14] NASA Robonaut, επίσημη ιστοσελίδα του ρομπότ Robonaut της NASA http://robonaut.jsc.nasa.gov.

[15] Tmsuk Co. LTD, Κατασκευαστής των ρομπότ TMSUK, http://www.tmsuk.co.jp.

[16] Ν. Τσουράκης, "Λογοτυπογράφος: Το εργαλείο υπαγόρευσης του συστήματος αναγνώρισης φωνής της ελληνικής γλώσσας", Διπλωματική εργασία, Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών, Πολυτεχνείο Κρήτης, 2001.

[17] Β. Διγαλάκης, "Εισαγωγή στην Επεξεργασία Φωνής", Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών, Πολυτεχνείο Κρήτης.

[18] S. Young, *Large Vocabulary Continuous Speech Recognition: a Review*, Cambridge University Engineering Department, 1996.

[19] Ηλεκτρονικός Λογογράφος, http://www.logografos.gr.

[20] D. Colton, *Automatic Speech Recognition Tutorial*, BYU Hawaii, 2003.

[21] Microsoft Corp., MSDN Library, http://msdn.microsoft.com.

[22] Ιστοσελίδα τεχνικών υπολογιστικών όρων, http://about.com.

[23] Samba, επίσημη ιστοσελίδα λογισμικού, www.samba.org.

[24] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Proceedings of the IEEE, Vol. 2, February 1989.