ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ Α.Τ.Ε.Ι. ΠΕΛΟΠΟΝΝΗΣΟΥ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε.

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«Οι Εξελίξεις στην αυτόματη αναγνώριση ομιλίας »

Φοιτήτρια : Μπαλαδήμα Κωνσταντίνα
Α.Μ: 2011034

Επιβλέπων καθηγητής: Νόκας Γεώργιος

ΣΠΑΡΤΗ, ΝΟΕΜΒΡΙΟΣ 2017

Πίνακας περιεχομένων

Λογοκλοπή

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

"Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάση επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων.
Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δε μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας."

Όνομα και Επώνυμο Συγγραφέα (Με Κεφαλαία):

……

Υπογραφή (Ολογράφως, χωρίς μονογραφή):

……

Ημερομηνία (Ημέρα – Μήνας – Έτος):

……

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα της Πτυχιακής μου, Καθηγητή του Τμήματος Μηχανικών Πληροφορικής Τ.Ε. του Α.Τ.Ε.Ι Πελοποννήσου κ. Νόκα Γεώργιο για την συνεχή καθοδήγηση και υποστήριξη σε όλη την διάρκεια της δύσκολης αυτής προσπάθειας εκπόνησης της Πτυχιακής μου εργασίας. Παρά τα εμπόδια και τις αντιξοότητες που συναντήσαμε σε ολόκληρη την πορεία, ακόμα και όταν πίστεψα πως αυτή η προσπάθεια ήταν μάταιη, δεν σταμάτησε στιγμή να με ενισχύει με κάθε δυνατό τρόπο και να πιστεύει πως θα τα καταφέρω. Είμαι απόλυτα σίγουρη πως χωρίς την δική του υπομονή και επιμονή τίποτα δεν θα είχε πραγματοποιηθεί.

Επίσης, θα ήθελα να δώσω ένα μεγάλο ευχαριστώ στις φίλες μου, οι οποίες υπήρξαν ο διαρκής υποστηρικτής και εμψυχωτής στην προσπάθειά μου να ολοκληρώσω την εργασία μου.

Τέλος, θα ήθελα να εκφράσω την μεγαλύτερή μου ευγνωμοσύνη στην οικογένειά μου, μιας και ότι έχω πετύχει και ότι έχω ολοκληρώσει στην ζωή μου το οφείλω αποκλειστικά σε εκείνη.

Κεφάλαιο 1 – Εισαγωγή

1.1 Γενικά

Ο ρόλος της τεχνολογίας αδιαμφισβήτητα κατέχει κεντρική θέση στην εξέλιξη της κοινωνίας. Η τεχνολογία κινεί και καθοδηγεί την αλλαγή, την μετάβαση από το ένα στάδιο στο άλλο. Οι ιστορικές περίοδοι διακρίνονται βάσει της τεχνολογίας που υπήρχε διαθέσιμη, κι επίσης, διαφοροποιούσε τον ένα πολιτισμό από τον άλλο. Ιστορικά, δηλαδή, αποτέλεσε χαρακτηριστικό συγχρονικού και διαχρονικού διαχωρισμού. Τα τεχνολογικά μέσα αντικατοπτρίζουν τις ανάγκες του πολιτισμικού και κοινωνικού πλαισίου εντός του οποίου αναπτύσσονται, καθώς αποτελούν τη λύση σε κάποιο υφιστάμενο πρόβλημα. Επίσης, αντανακλούν χαρακτηριστικά της εκάστοτε κοινωνίας, όπως η ιεραρχία, οι σχέσεις μεταξύ των ανθρώπων, και τα ήθη της εποχής.

Τέλος, τα τεχνολογικά επιτεύγματα αποτελούν εφαρμογές της κατακτηθείσας γνώσης όσον αφορά τις επιστήμες.

Παρατηρούμε, λοιπόν, ότι όσο εξαφανίζονται οι εκάστοτε πολιτισμοί κι οδηγούμαστε σε μια παγκοσμιοποιημένη κοινωνία, η τεχνολογία αρχίζει να είναι κοινή για ολοένα και μεγαλύτερο μέρος της ανθρωπότητας. Σήμερα, λοιπόν, το σύνολο του λεγόμενου «ανεπτυγμένου κόσμου» χρησιμοποιεί τα ίδια τεχνολογικά μέσα. Η ύπαρξη του διαδικτύου επιτρέπει την επικοινωνία και τη διάδοση ιδεών, ειδήσεων και εξελίξεων σε παγκόσμιο επίπεδο σε μηδενικό χρόνο, γεγονός που μας καθιστά ένα «παγκόσμιο χωριό». Λόγω του ότι η τεχνολογία αντανακλά και τον τρόπο ζωής της κοινωνίας που την ανέπτυξε, τη νοοτροπία και τις αξίες της, η εξάπλωσή της συνεπάγεται και καθιέρωση και των στοιχείων αυτών. Έτσι, λοιπόν, σήμερα η τεχνολογία του δυτικού πολιτισμού έχει εξαπλωθεί παγκόσμια, από κοινού με κάποιες από τις δυτικές αξίες.

Υπάρχει συλλογική απαίτηση για αποτελεσματικότητα, κι ανταποδοτικότητα, δεδομένου του ανταγωνιστικού περιβάλλοντος. Υπάρχει ανάγκη για άμεση μετάδοση τεράστιου όγκου πληροφορίας σε παγκόσμια κλίμακα, για συνδεσιμότητα και επικοινωνία, κι αλληλεπίδραση. Ο σύγχρονος άνθρωπος είναι αποδέκτης μιας συνεχούς ροής πληροφορίας, επικοινωνίας κι ενημέρωσης, η οποία γίνεται προσβάσιμη μέσω ποικίλων συσκευών που γίνονται ολοένα και πιο μικρές, ευέλικτες, πολυλειτουργικές κι εν τέλει, απαραίτητες. Με τη βοήθειά τους ενημερώνεται, συμμετέχει, μαθαίνει, κι εργάζεται. Καλείται να είναι διαρκώς συνδεδεμένος, και να

διαχειρίζεται εξαιρετικά πολυάριθμα ερεθίσματα κάθε στιγμή, κατάσταση πολύ διαφορετική από αυτήν του παρελθόντος.[1][2]

Η αλλαγή όμως δεν είναι μόνο ποσοτική, αλλά και ποιοτική. Με τη θεαματική ανάπτυξη στις τεχνολογίες επικοινωνίας, έχει σημειωθεί και η αντίστοιχη ποιοτική αλλαγή στην ίδια τη φύση της επικοινωνίας. Από την αρχαιότητα μέχρι και πολύ πρόσφατα στην ιστορία της ανθρωπότητας η μετάδοση πληροφορίας συντελούνταν μέσω γραπτού κειμένου και στατικής εικόνας. Με τον ερχομό του διαδικτύου τα κείμενα και οι εικόνες άρχισαν να μεταδίδονται με το πάτημα ενός κουμπιού, χωρίς κανένα γεωγραφ`ικό περιορισμό ή κίνδυνο φυσικής φθοράς. Η βελτίωση της ταχύτητας του διαδικτύου έφερε την ουσιαστική επανάσταση στις τηλεπικοινωνίες και σηματοδότησε τη αλλαγή στην αντίληψή μας για την επικοινωνία. Το ταχύτερο διαδίκτυο επέτρεψε στις επικοινωνιακές διαδικασίες να εκτελούνται σε πραγματικό χρόνο, άρα τις κατέστησε πιο φυσικές κι αυθόρμητες. Μπορούμε λοιπόν να σχολιάζουμε και να επηρεάζουμε τα γεγονότα την ώρα που συμβαίνουν, από την άνεση του χώρου μας κι ανέξοδα. Είναι επίσης στη δική μας ευχέρεια να επικοινωνήσουμε ή να μεταδώσουμε μια πληροφορία διατηρώντας την ανωνυμία μας.

Με την αναδυόμενη τεχνολογία επεξεργασίας φυσικής γλώσσας αποκτούμε αυξανόμενες δυνατότητες που αφορούν γραμματικό έλεγχο, σημασιολογική αναγνώριση, και μετάφραση, κάτι που μας δίνει την ευχέρεια να παράγουμε προϊόντα πολύ ανώτερα από αυτά που θα μπορούσαμε βάσει των δικών μας γνώσεων. Έχει αλλάξει όμως και η ίδια η χρήση της γλώσσας. Εμφανίστηκαν νέες συντομεύσεις οι οποίες αντανακλούν την αρχιτεκτονική του πληκτρολογίου, η τάση για μικρότερες λέξεις, η χρήση ξένων λέξεων μέσα στα κείμενα, κι αρχικών γραμμάτων αντί ονομάτων.

Νέοι όροι που προέρχονται από τον κόσμο της τεχνολογίας χρησιμοποιούνται μέσα στον καθημερινό και επαγγελματικό λόγο, και μια γενικότερη στροφή στη θέαση της ζωής και των ανθρώπων ως τεχνολογικά δημιουργήματα, υπό την έννοια ότι όλοι και όλα αντιμετωπίζονται ως εν δυνάμει μηχανές που λειτουργούν με συγκεκριμένους τρόπους υπό συγκεκριμένες συνθήκες, και η αξία τους υπολογίζεται με όρους ωφέλειας.

Η έκταση των κειμένων έχει μειωθεί σημαντικά, απαιτείται μόνο το νόημα χωρίς προλόγους και περίτεχνες φράσεις, το λεξιλόγιο έχει απλουστευθεί και συντομευθεί. Αυτό συμβαίνει διότι οι συσκευές γίνονται όλο και μικρότερες και είναι εξαιρετικά άβολη η συγγραφή, επομένως περικόπτεται ό, τι δεν είναι απαραίτητο. Σε αυτό το

πρόβλημα έρχεται να απαντήσει η τεχνολογία αναγνώρισης φωνής, καθώς θα καταστήσει ευκολότερη την επικοινωνία θα μειώσει την πληκτρολόγηση η οποία εξακολουθεί να δημιουργεί προβλήματα λόγω του μικρού μεγέθους των συσκευών. Οι χρήστες θα μπορούν να κάνουν αναζήτηση στο κινητό τους ή στο διαδίκτυο, να συζητούν και να γράφουν χρησιμοποιώντας λογισμικά που αναγνωρίζουν την φωνητική εντολή, κάτι που θα διευκολύνει τη χρήση των συσκευών εν γένει, αλλά και σε κάποιες περιπτώσεις ακόμη περισσότερο. Τέτοιες είναι το on-line chatting και η χρήση του κινητού κατά τη βάδιση ή την οδήγηση (όπου η πληκτρολόγηση είναι από δύσκολη έως επικίνδυνη).

Οι αλλαγές στη χρήση της γλώσσας και τη φύση της επικοινωνίας είναι πλέον ευρέως αποδεκτές και επιθυμητές όχι μόνο στην ανεπίσημη επικοινωνία, αλλά και στο επαγγελματικό πλαίσιο. Η παλαιότερη τάση για χρήση εντυπωσιακών και μακρών εκφράσεων, που αποδείκνυε τις γλωσσικές ικανότητες του ομιλητή, έχει αντικατασταθεί από την ανταλλαγή σύντομων φράσεων επικεντρωμένων στην ουσία, χάριν συντομίας και αποτελεσματικότητας.

Συνακολούθως, παρατηρείται η τάση να παραγκωνίζεται η πρόσωπο με πρόσωπο επικοινωνία ως μη αξιόπιστη, και να προτιμάται η ανταλλαγή ηλεκτρονικών κειμένων όταν πρόκειται για επαγγελματικές συναλλαγές ή και για πληροφορίες που χρειάζεται να είναι απόλυτα ξεκάθαρες σε όλα τα συμβαλλόμενα μέρη. Αυτή η υποτίμηση της κατά πρόσωπο επικοινωνίας σε συνδυασμό με τη χρήση της τεχνολογίας ως μέσο επικοινωνίας, διασκέδασης κι εργασίας, έχει δημιουργήσει μια νέα κατάσταση διαβίωσης, όπου έχιε μειωθεί αισθητά το επίπεδο διασύνδεσης των ανθρώπων μεταξύ τους κι έχει αυξηθεί η εξάρτηση από την τεχνολογία. Υπάρχουν δύο τρόποι θεώρησης αυτής της μεταβολής. Από τη μια κάποιοι τη θεωρούν θετική, καθώς έχουν αυξηθεί αλματωδώς οι δυνατότητες, η εμβέλεια δράσης και δημιουργίας, η πρόσβαση στην πληροφορία για οτιδήποτε συμβαίνει οπουδήποτε αναφορικά με όλους τους τομείς.

Υπό αυτό το πρίσμα, ο χρήστης μπορεί πολύ απλά να καταφέρει πράγματα που παλαιότερα ήταν αδιανόητα. Για παράδειγμα, μπορεί να πλοηγηθεί στους δρόμους μιας πόλης, να δει αεροφωτογραφίες, να συνομιλήσει με κάποιον σε μια γλώσσα που δε γνωρίζει, να δει οπτικοποιημένες έννοιες τις φυσικής, να παρακολουθήσει μαθήματα από μεγάλα πανεπιστήμια του εξωτερικού. Όλα αυτά συμβαίνουν από την άνεση του γραφείου κι ανέξοδα, ενώ παλαιότερα προϋπέθεταν δαπανηρά ταξίδια και πολυετείς σπουδές. Από την άλλη, υπάρχει και η αντίθετη άποψη, που βλέπει αυτές τις εξελίξεις

ως αρνητικές, καθώς στερούν τη χαρά και την ουσία της ανθρώπινης συνύπαρξης, κι οδηγούν στην απομάκρυνση και την αποξένωση μεταξύ των ανθρώπων. Η προαναφερθείσα «τεχνολογική θεώρηση» παρατηρείται και στις διατομικές σχέσεις: είναι μονοδιάστατες, βραχύβιες, επιφανειακές, χρησιμοθηρικές και προσανατολισμένες στο κέρδος και την ανταποδοτικότητα. Υπάρχει φόβος έκθεσης του πραγματικού εαυτού αυτό που προβάλλουμε είναι ένα ηλεκτρονικό alter ego.

Η οθόνη μιας συσκευής προσφέρει ασφάλεια, καθώς επιτρέπει στο χρήστη να διαμορφώσει και να προβάλλει την ταυτότητα που επιθυμεί και να αποκρύψει άλλα στοιχεία. Τα κοινωνικά μας δίκτυα είναι γεμάτα με επαφές που δεν θα συναναστραφούμε ποτέ στην πραγματική ζωή.

Τόσο η θετική όσο και η αρνητική θεώρηση εμπεριέχουν μεγάλες δόσεις αλήθειας. Οι αλλαγές στον τρόπο ζωής παγκοσμίως είναι σαρωτικές, κι αξίζει τον κόπο να τις παρατηρήσουμε και να τις περιγράψουμε. Η βιομηχανία της τεχνολογίας με τα επιτεύγματά της στον τομέα της τεχνητής νοημοσύνης αλλάζει τα δεδομένα στις καθημερινές μας εμπειρίες. Κανένας τομέας δεν μπορεί να μείνει ανεπηρέαστος. Η προηγμένη αναγνώριση προτύπων, στα πλαίσια της μηχανικής μάθησης, επιτρέπει στις μηχανές να αναλύσουν τεράστιους όγκους δεδομένων, πράγμα που δίνει άλλη δυναμική και ώθηση στην ανθρώπινη σκέψη.

Στα πλαίσια αυτών των εξελίξεων αναδύεται ο φόβος σχετικά με τον περιορισμό της αυτόνομης σκέψης των ανθρώπων, της ιδιωτικότητάς τους και του σεβασμού στις πολιτισμικές ιδιαιτερότητές τους. Έχει διαχρονικά παρατηρηθεί ότι υπό το πρόσχημα της ασφάλειας, για παράδειγμα, έχουν αναπτυχθεί τεχνικές για την παρακολούθηση της ηλεκτρονικής συμπεριφοράς των ανθρώπων, για την εξαγωγή συμπερασμάτων σχετικά με τις προτιμήσεις τους κι τις ασχολίες τους. Τα δεδομένα αυτά χρησιμοποιούνται για την κατηγοριοποίησή τους σε ομάδες. Πολλοί είναι αυτοί που ισχυρίζονται ότι αυτό θα είναι και το τέλος της ιδιωτική ζωής και της ελευθερίας. Οποιαδήποτε ηλεκτρονική πληροφορία που αφορά τους χρήστες αποθηκεύεται και χρησιμεύει για την κατάταξη σε κατηγορίες, με σκοπό την εκμετάλλευση από διάφορους φορείς ελέγχου. Αυτό θα μπορούσε να χαρακτηρισθεί κι ως μια απειλή για τα προσωπικά δεδομένα. Οι τελευταίες κυβερνήσεις της Αμερικής και άλλων μεγάλων χωρών αγοράζουν αυτά τα δεδομένα από τις εταιρείες που τα συλλέγουν για τον έλεγχο των πολιτών. Εδώ τίθεται και το ζήτημα της ηθικής, και πιο συγκεκριμένα, η ανάγκη αποσαφήνισης του τι σημαίνει προσωπικό δεδομένο, ποιος πρέπει να το διαχειρίζεται και με τι σκοπό. Είναι

άγνωστο με ποια κριτήρια γίνονται αυτές οι κατηγοριοποιήσεις, όταν πρόκειται για υποθέσεις δικαιοσύνης και παράβασης του νόμου. Ίσως οι δημιουργοί των αλγορίθμων να έθεσαν τα κριτήρια με βάση τις δικές τους στερεοτυπικές αντιλήψεις σχετικά με την εθνικότητα, το φύλο, την εκπαίδευση και την οικονομική κατάσταση.

Προκύπτει λοιπόν ότι το ζήτημα της συλλογής δεδομένων από τη συμπεριφορά των χρηστών στο διαδίκτυο είναι ένα ζήτημα που αφορά το σύνολο των ανθρώπων, καθώς όλοι υπόκεινται σε παρακολούθηση και κατηγοριοποίηση. Είναι επίσης αναγκαίο να θεσπιστεί ένα ικανό θεσμικό πλαίσιο που θα διέπει αυτές τις διαδικασίες, υπό την εποπτεία και συνδρομή κοινωνικών επιστημόνων, προκειμένου να υπάρξει η μέγιστη δυνατή δεοντολογία και δικαιοσύνη.

Φυσικά σε ό, τι αφορά τον επιχειρηματικό κόσμο οι εξελίξεις στην τεχνητή νοημοσύνη και τη ρομποτική είναι πραγματικά θαυμαστές, καθώς έχει επιτευχθεί αυτοματοποίηση στις διαδικασίες ελέγχου και παραγωγής, στην ανάλυση οικονομικών στοιχείων. Η κυρίαρχη τάση αφορά την προσπάθεια αυτοματισμού διαδικασιών και υπολογιστικής αναπαράστασης της ανθρώπινης γνώσης. Στόχος είναι η μεγαλύτερη δυνατή τυποποίηση και εκτέλεση έργων από αλγορίθμους για την αποφυγή των ανθρώπινων λαθών που σχετίζονται με κόπωση, φτωχή κρίση, λανθασμένη αντίληψη και υποκειμενικότητα. Η τεχνητή νοημοσύνη διευρύνει την ανθρώπινη, αλλά θα πρέπει να σιγουρευτούμε ότι λειτουργεί πάντα προς όφελος του κοινωνικού συνόλου.

Κεφάλαιο 2 – Νοημοσύνη και λόγος

2.1 Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη στοχεύει στην εκτέλεση περιορισμένων και συγκεκριμένων στόχων. Οι αλγόριθμοι, δηλαδή, σχεδιάζονται και προορίζονται για έργα όπως η αναγνώριση προσώπου, η αναζήτηση αποτελεσμάτων, η αναγνώριση φωνής, κλπ. Ο μακροπρόθεσμος στόχος είναι η δημιουργία πιο συστημάτων που θα μπορούν να εκτελούν πολλαπλά έργα σειριακά ή ταυτόχρονα, ώστε να προσομοιάσουν σε έναν ανθρώπινο νου, δηλαδή σε μια πιο γενικευμένη και σφαιρική μορφή τεχνητής νοημοσύνης. Τα μέχρι στιγμής δεδομένα δείχνουν ότι όταν πρόκειται για ένα πολύ συγκεκριμένο έργο με πεπερασμένο αριθμό βημάτων, ο υπολογιστής αποδεικνύεται αποτελεσματικότερος από τον άνθρωπο, ενώ όταν το έργο απαιτεί σφαιρική αντιμετώπιση, τότε ο άνθρωπος αποδίδει πολύ καλύτερα.

Η τεχνητή νοημοσύνη ξεκίνησε ως μια προσπάθεια να κατανοήσουμε τη φύση του ανθρώπινου νου μέσω της αναπαράστασης και της κατασκευής νοημόνων συστημάτων ικανών για εκτέλεση συγκεκριμένων έργων. Μέχρι σήμερα έχουν παραχθεί εντυπωσιακά αποτελέσματα, και η πορεία στο μέλλον πρόκειται να είναι θεαματική. Οι εφαρμογές της τεχνητής νοημοσύνης ήδη βρίσκουν εφαρμογή σε όλες τις πλευρές της καθημερινής κι επαγγελματικής ζωής, και πρόκειται να διευρυνθούν περαιτέρω στο μέλλον. Η κατασκευή συστημάτων ικανών για ορισμένες μορφές σκέψης συγκεντρώνει το ενδιαφέρον πολλών επιστημών, όπως η πληροφορική, η ψυχολογία, τα μαθηματικά, η φιλοσοφία και η γλωσσολογία. Είναι ένας διεπιστημονικός κλάδος όπου όλες αυτές οι επιστήμες συμβάλλουν με τα πειραματικά τους δεδομένα στην ανάπτυξη συστημάτων με νοημοσύνη που λειτουργούν επικουρικά στην ανθρώπινη νοημοσύνη. Έχει τις ρίζες της στην μελέτη της φύσης, των χαρακτηριστικών και του φυσικού υποστρώματος της νοημοσύνης, πεδίο που υφίσταται από την αρχαιότητα. Αρχικά φιλόσοφοι, κι ύστερα επιστήμονες προσπάθησαν να κατανοήσουν πώς ο όραση, η μάθηση, η μνημονική ανάκληση και ο συλλογισμός εκτελούνται. Όταν αναπτύχθηκαν και οι πρώτοι υπολογιστές το 1950, έθεσαν αυτά τα ερωτήματα σε μια πιο πρακτική και συστηματική βάση. Σήμερα η τεχνητή νοημοσύνη περιλαμβάνει πολλές υποενότητες κι εφαρμογές, έτσι τα όριά της έχουν γίνει δυσδιάκριτα.

Έχουν δοθεί αρκετοί περιγραφικοί ορισμοί, οι οποίοι εμπίπτουν σε τέσσερις κατηγορίες: συστήματα που σκέπτονται σαν άνθρωποι, συστήματα που δρουν σαν άνθρωποι, συστήματα που σκέπτονται λογικά, και συστήματα που δρουν λογικά. Οι

παραπάνω κατηγορίες ορισμών προκύπτουν από φιλοσοφικές διακρίσεις μεταξύ του τι συνιστά σκέψη και τι δράση, πού βρίσκεται η διαχωριστική γραμμή μεταξύ τους και πώς αποτυπώνεται στα υπολογιστικά συστήματα.

Στα πλαίσια της θέασης της τεχνητής νοημοσύνης ως προσομοίωσης με την ανθρώπινη δράση, οι απαιτήσεις είναι από το σύστημα είναι να μπορεί να επεξεργάζεται φυσική γλώσσα, να αναπαριστά γνώση μέσω της αποθήκευσης, να χρησιμοποιεί την υποθηκευμένη γνώση για να απαντά ερωτήσεις μέσω συλλογισμού, και να είναι ικανό για μηχανική μάθηση, δηλαδή να αξιοποιεί δεδομένα για να τροποποιεί τη συμπεριφορά του. Οι κυριότερες εφαρμογές αυτής της προσέγγισης είναι η μηχανική όραση και η ρομποτική κίνηση.

Η επόμενη ομάδα ορισμών βλέπει την τεχνητή νοημοσύνη ως προσομοίωση της ανθρώπινης σκέψης, ως μοντελοποίησης των γνωστικών δεξιοτήτων του ανθρώπου. Αυτό φυσικά προϋποθέτει μια πολύ καλή ιδέα και γνώση για το πώς λειτουργεί ο ανθρώπινος νους, προκειμένου να μπορεί να μοντελοποιηθεί στη συνέχεια. Κι εδώ χρησιμοποιούνται τα ερευνητικά δεδομένα από το χώρο της πειραματικής ψυχολογίας. Αφού λοιπόν σχηματιστεί μια θεωρία βασιζόμενη σε δεδομένα, για το πώς συντελείται μια λειτουργία, τότε μεταφράζεται σε υπολογιστικό πρόγραμμα. Αφού κατασκευαστεί το πρόγραμμα, χρησιμοποιείται με πραγματικά δεδομένα προκειμένου να εξεταστεί η συμπεριφορά του. Εάν η σχέση εισερχομένου κι εξερχομένου προσομοιάζει σε αποτελεσματικότητα και χρονικότητα τον ανθρώπινο νου, τότε συμπεραίνουμε ότι το πρόγραμμα είναι επιτυχημένο ως προς το σκοπό του.

Η τρίτη προσέγγιση βλέπει την τεχνητή νοημοσύνη ως εφαρμογή των νόμων της ανθρώπινης σκέψης. Βασίζεται στην ανάπτυξη της φορμαλιστικής λογικής, στα πλαίσια της οποίας τα βήματα επίλυσης ενός προβλήματος, δηλαδή η μετάβαση από μια αρχική κατάσταση σε μια κατάσταση-στόχο οργανώνεται σε έναν πεπερασμένο αριθμό βημάτων, η πορεία είναι γραμμική και μονοκατευθυντική, και προβλέπονται συγκεκριμένες προϋποθέσεις για την μετάβαση από το ένα βήμα στο επόμενο. Φυσικά, αυτή η προσέγγιση προϋποθέτει την ύπαρξη ακριβών παραδοχών και διατυπώσεων για τα αντικείμενα και τα γεγονότα του φυσικού κόσμου, καθώς και για τις μεταξύ τους σχέσεις.

Δεδομένης της ύπαρξης αυτών των παραδοχών διατυπωμένων ως νόμων, και επαρκούς μνήμης, το σύστημα αποδίδει μια λύση στο πρόβλημα που του δίνεται. Γι’ αυτό η συγκεκριμένη θεώρηση είναι πιο χρήσιμη στην επίλυση μαθηματικών

προβλημάτων. Όσον αφορά τη χρησιμότητα σε καταστάσεις που απομακρύνονται από το φορμαλισμό των μαθηματικών, η περιγραφή του πραγματικού κόσμου γίνεται ολοένα και δυσκολότερη, λόγω του περίπλοκου χαρακτήρα του. Ένα άλλο μεγάλο ζήτημα είναι ότι στον πραγματικό κόσμο πολύ σπάνια λειτουργούμε σε συνθήκες απόλυτης βεβαιότητας, αλλά έχουμε μόνο ενδείξεις και εικασίες βάσει των οποίων λαμβάνουμε αποφάσεις, δοκιμάζουμε, και τροποποιούμε αναλόγως. Υπολογιστικά είναι αρκετά μεγάλη πρόκληση η αναπαράσταση της αβεβαιότητας και του χειρισμού της στη λήψη απόφασης, κι επιπλέον χρειάζεται μεγάλη υπολογιστική ισχύς. Τέλος δίνεται μεγάλη έμφαση στο σωστό συμπερασμό.

Η τέταρτη προσέγγιση βλέπει την τεχνητή νοημοσύνη ως την ικανότητα του συστήματος να δρα λογικά, δηλαδή να ολοκληρώνει έργα φτάνοντας σε μια επιθυμητή κατάσταση, δεδομένης των υπαρχόντων γνώσεων και αντιλήψεων. Εδώ λοιπόν αναφερόμαστε στα συστήματα ως πράκτορες., και η τεχνητή νοημοσύνη αποσκοπεί στην κατασκευή λογικών πρακτόρων. Είναι επίσης απαραίτητο να υπάρχουν συγκεκριμένοι νόμοι της σκέψης. Όπως προαναφέρθηκε δίνεται έμφαση στο σωστό συμπερασμό, το οποίο είναι μέρος και του να είναι κάποιος πράκτορας λογικός, καθώς η επιτυχής εκτέλεση ακολουθίας πράξεων μας δείχνει σωστή εκτίμηση της κατάστασης και των αξιοποίηση της υπάρχουσας γνώσης. Όμως μπορεί να υπάρξει απόκλιση μεταξύ του συμπερασμού και της λογικής πράξης, καθώς το πρώτο δεν συνεπάγεται πάντα το δεύτερο, και το αντίστροφο. Δηλαδή υπάρχουν καταστάσεις όπου τα συμπεράσματα είναι σωστά αλλά όχι η εκτέλεση πράξης, αλλά και άλλες καταστάσεις όπου δεν έχει προηγηθεί συμπερασμός, αλλά εκτελείται πράξη. Και στα πλαίσια αυτής της θεώρησης απαιτείται πολύ καλή γνώση βασισμένη σε ψυχολογικά ερευνητικά δεδομένα όσον αφορά τις ανθρώπινες γνωστικές λειτουργίες, προκειμένου αυτές να αναπαρασταθούν υπολογιστικά. Οι γνώσεις για τη λειτουργία του νου αλλά και του κόσμου διατυπώνονται σε προτάσεις φυσικής γλώσσας, και στη συνέχεια σε κώδικα. Η προσέγγιση αυτή ενέχει το πλεονέκτημα της εναρμόνισης με τη νοημοσύνη του ανθρώπου.

Παρατηρώντας αυτές τις διαφορετικές προσεγγίσεις, αντιλαμβανόμαστε και τη βασική διαφοροποίηση μεταξύ σκέψης και συμπεριφοράς, η οποία διακρίνει τους επιστήμονες που ασχολούνται με την τεχνητή νοημοσύνη. Αξίζει να γίνει μια μικρή αναφορά στη συμβολή των διαφορετικών πεδίων στην ανάπτυξή της. Οι φιλόσοφοι έθεσαν τις βάσεις για την κατανόηση του νου ως μηχανή η οποία αποθηκεύει γνώση με

κωδικοποιημένο τρόπο και την αξιοποιεί για να καταλήξει στις σωστές πράξεις. Οι μαθηματικοί παρείχαν τα εργαλεία για τη διαχείριση της βεβαιότητας και της ασάφειας, και τη γλώσσα για παραγωγή των αλγορίθμων. Οι ψυχολόγοι ενδυνάμωσαν την ιδέα ότι ο νους επεξεργάζεται πληροφορίες, και οι μηχανικοί υπολογιστών υλοποίησαν όλα τα παραπάνω. Καταλαβαίνουμε λοιπόν ότι το κομμάτι της τεχνητής νοημοσύνης που κατανοεί και επεξεργάζεται την ακοή ανταποκρίνεται ελαστικά, μπορεί να κατανοήσει καταστάσεις από τα συμφραζόμενα και μέσα από όμοιες καταστάσεις ταιριάζει τις ομοιότητες και τις διαφορές.

2.2 Επεξεργασία φυσικής γλώσσας

Η φυσική γλώσσα αναγνωρίζεται και ερμηνεύεται υπολογιστικά. Ο τομέας της επεξεργασίας της φυσικής γλώσσας αφορά πολλές επιμέρους εργασίες οι οποίες αποτελούν και διακριτά πεδία έρευνας, αλλά συνδυάζονται σε ενιαία προγράμματα. Σήμερα η έρευνα προσανατολίζεται στην επίλυση προβλημάτων όπως η διαχείριση παραπάνω από μίας πρότασης, η τμηματοποίηση και η ερμηνεία των προτάσεων χωριστά, η αποκρυπτογράφηση των ασυνήθιστων ή και άγνωστων λέξεων, η εργασία με περίπλοκο συντακτικό, η σημασιολογική ανάλυση, και οι αμφισημία.[2] Τα συστήματα που έχουν ανταποκριθεί με σχετική επιτυχία στα παραπάνω έργα έχουν αναπτυχθεί στα πλαίσια κάποιων σκοπών μεγαλύτερης κλίμακας. Ένα από αυτά είναι η μηχανική μετάφραση προτάσεων και κειμένων, πέρα από λέξεις μεμονωμένες.

Ακόμη και σήμερα δεν είμαστε σε θέση να εμπιστευτούμε τα συστήματα μηχανικής μετάφρασης, αλλά θα πρέπει να τα χρησιμοποιούμε επικουρικά και να γνωρίζουμε και τις δύο γλώσσες με τις οποίες εργαζόμαστε. Δηλαδή είτε γίνεται αντιγραφή κι επικόλληση συγκεκριμένων φράσεων από κείμενο, είτε ολόκληρου του κειμένου αλλά ακολουθεί κι επιμέλεια από άνθρωπο που γνωρίζει και τις δύο γλώσσες. Η μηχανική μετάφραση είναι δύσκολη καθώς απαιτεί βαθιά κατανόηση του κειμένου. Για να γίνει επιτυχημένα, θα πρέπει να προηγηθεί ανάγνωση του κειμένου κι αντίληψη της κατάστασης στην οποία αναφέρεται. Έπειτα, πρέπει να παραχθεί ένα αντίστοιχο κείμενο στην γλώσσα-στόχο που να περιγράφει την ίδια ή μια παρόμοια κατάσταση.

Μια άλλη εφαρμογή της επεξεργασίας φυσικής γλώσσας αφορά την ανάκτηση πληροφοριών. Εδώ, το έργο αφορά την επιλογή αρχείων ανάλογα με την αναζήτηση. Για το σκοπό αυτό, κάθε αρχείο φέρει κάποια αναγνωριστικά, όπως ετικέτες, λέξεις-κλειδιά. Συνήθως είναι διαχωρισμένα σε μέρη, ανάλογα με το θέμα τους, έτσι ώστε να

αποδίδονται με την κατάλληλη αναζήτηση. Η αναζήτηση είναι μια σειρά λέξεων που πληκτρολογεί ο χρήστης. Στα πρώτα συστήματα ανάκτησης χρησιμοποιούταν ένας συνδυασμός των λέξεων-κλειδιών με τη λογική Bool, όμως λόγω της περιορισμένης ευελιξίας που παρουσίαζε, εφαρμόζεται το μοντέλο του χώρου διανυσμάτων, όπου κάθε λίστα λέξεων επεξεργάζεται σαν ένα διάνυσμα σε χώρο n διαστάσεων, όπου n είναι ο αριθμός των ξεχωριστών μερών στη συλλογή αρχείων.

Τέλος, η επεξεργασία φυσικής γλώσσας χρησιμεύει στην κατηγοριοποίηση κειμένων. Έχει αποδεχτεί αποτελεσματική στο έργο της κατηγοριοποίησης κειμένου σε προκαθορισμένες θεματικές κατηγορίες. Για παράδειγμα, ο χρήστης μπορεί να αναζητήσει όλα τα κείμενα που σχετίζονται με ένα θέμα, όπως έναν τομέα της βιομηχανίας ή μια γεωγραφική περιοχή. Το ζήτημα όμως είναι ότι ο χρήστης δεν γνωρίζει ποιος είναι ο συνδυασμός λέξεων που όταν πληκτρολογηθεί στην αναζήτηση θα αποδώσει όλα τα σχετικά, και μόνο αυτά, κείμενα.

2.3 Αναγνώριση ομιλίας

Το όραμα της τεχνολογίας, ο υπολογιστής που να κατανοεί την ανθρώπινη ομιλία, είναι ήδη μια πραγματικότητα. Λόγω της πρόσφατης ανάπτυξης στην τεχνολογία αναγνώρισης ομιλίας, έχουν αρχίσει αλλά είναι σίγουρο πως θα υπάρξουν πολλές περισσότερες εφαρμογές καθοδηγούμενες από ομιλία διαθέσιμες στο προσεχές μέλλον. Μάλιστα, υπάρχουν ήδη μερικά λογισμικά στην αγορά, όπως τα πακέτα φωνητικών εντολών και υπαγόρευσης για υπολογιστή, καθώς επίσης και φωνητικά συστήματα ελέγχου και οδήγησης.

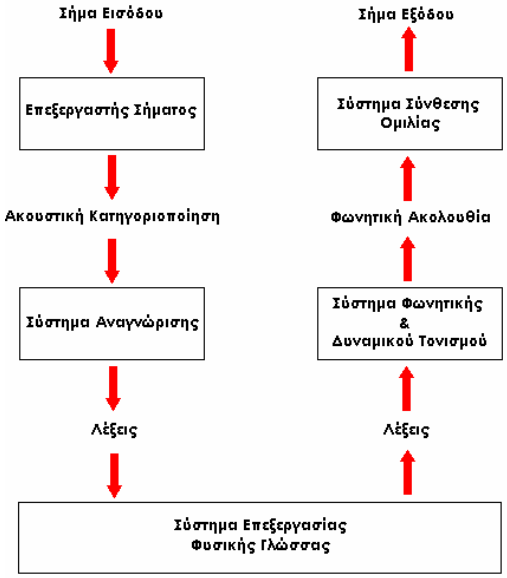
Η ομιλία είναι ο κυρίαρχος και ο πιο διαδεδομένος τρόπος ανθρώπινης επικοινωνίας. Αν και το μεγαλύτερο μέρος της διδασκαλίας και της εκμάθησης γίνεται στη γραπτή μορφή, η ομιλία είναι ο πλέον χρησιμοποιούμενος τρόπος καθημερινής επικοινωνίας. Επομένως, λογικά, η ομιλία θα είναι επίσης ο πιο διαδεδομένος και εύχρηστος τρόπος επικοινωνίας μεταξύ ανθρώπου και μηχανής. Αλλά δυστυχώς αυτό δεν είναι όσο εύκολο όσο ακούγεται. Παλιότερα τα περισσότερα συστήματα αναγνώρισης ομιλίας ήταν πολύ φτωχά τόσο στην ακρίβεια όσο και στην ταχύτητα. Επομένως ήταν αδύνατο να χρησιμοποιηθούν.

Ένα σύστημα αναγνώρισης ομιλίας είναι ένα σύστημα που αντιγράφει την ομιλία σε κείμενο. Μπορεί να θεωρηθεί σαν μια γραφομηχανή που ενεργοποιείται με την φωνή, όπου ένα πρόγραμμα μεταφέρει την ομιλία σε μια έξοδο του υπολογιστή,

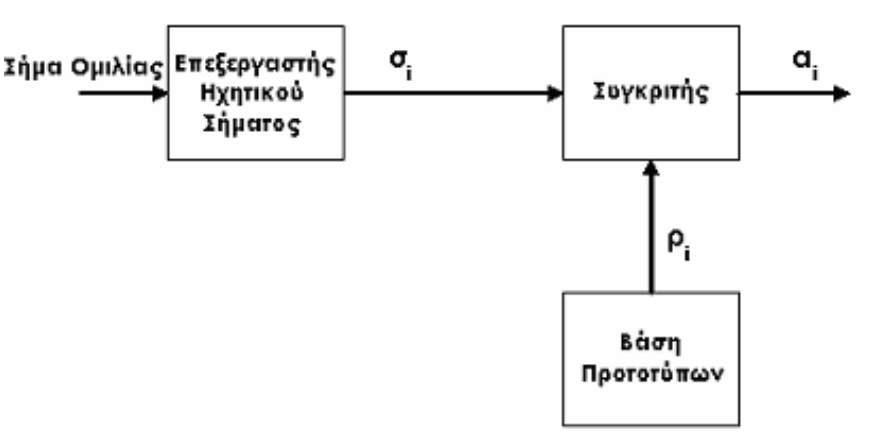
π.χ. την οθόνη. Τα συστήματα αναγνώρισης ομιλίας μπορούν να είναι δύο τύπων. Πρώτον, συστήματα αναγνώρισης μιας απομονωμένης λέξης.Πρόκειται για συστήματα που αναγνωρίζουν μονάχα μια λέξη τη φορά.

Δεύτερον, συστήματα αναγνώρισης συνεχούς ομιλίας, δηλαδή η αναγνώριση γίνεται ταυτόχρονα με την ομιλία.

Οι ίδιες βασικές τεχνικές για την ανάλυση, τη σημασιολογική ερμηνεία, και τη βασισμένη στα συμφραζόμενα ερμηνεία μπορούν να χρησιμοποιηθούν για την προφορική ή γραπτή γλώσσα, υπάρχουν όμως μερικές σημαντικές διαφορές που προκαλούν επιπτώσεις στο σχεδιασμό συστημάτων. Παραδείγματος χάριν στην περίπτωση της προφορικής γλώσσας το σύστημα πρέπει να εξετάσει την αβεβαιότητα, επειδή όπως είναι γνωστό η ομιλία είναι πολλές φορές διφορούμενη, δηλαδή με μια αλλαγή στον τονισμό ή στα σημεία στίξης να αλλάζει το νόημα. Επίσης οι προφορικές γλώσσες είναι δομικά διαφορετικές από τις γραπτές γλώσσες και μερικές φορές το σύστημα έχει μόνο μια εικασία για αυτό που ειπώθηκε. Στην πραγματικότητα, μερικές φορές ένα αντίγραφο της τέλεια κατανοητής ομιλίας δεν είναι κατανοητό όταν διαβάζεται.Η προφορική γλώσσα εμφανίζεται πιο επαυξημένη και περιέχει τις ιδιαίτερες πληροφορίες που δεν γίνονται αντιληπτές στη γραπτή μορφή. Έχει επίσης πολλές διορθώσεις, αφού ο ομιλητής διορθώνει τα λάθη που κάνει κατά την ομιλία του.Επιπλέον ο προφορικός διάλογος έχει μια συχνή αλληλεπίδραση της αναγνώρισης και της επιβεβαίωσης που διατηρεί τη συνομιλία, η οποία δεν εμφανίζεται σε γραπτές μορφές.



Γενικά μπορούμε να πούμε ότι ένα σύστημα αναγνώρισης ομιλίας αποτελείται από δυο βασικά τμήματα, την επεξεργασία του ηχητικού σήματος και την γλωσσολογική επεξεργασία. Όπως φαίνεται στην εικόνα οι ήχοι που παράγονται από τον ομιλητή μετατρέπονται στην ψηφιακή μορφή από έναν αναλογικό σε ψηφιακό μετατροπέα. Αυτό το σήμα υποβάλλεται σε επεξεργασία έπειτα για να εξαγάγει τα διάφορα χαρακτηριστικά γνωρίσματα, όπως η ένταση του ήχου σε διάφορες συχνότητες και η αλλαγή στην ένταση κατά τη διάρκεια του χρόνου. Μπορούμε να σκεφτούμε αυτή την διαδικασία σαν μια συσκευή που σε τακτά και αρκετά μικρά χρονικά διαστήματα παίρνει δείγματα σi του σήματος εισόδου. Τα σήματα αυτά συγκρίνονται με μια βάση πρωτοτύπων, όπου διαθέτει γνωστά σήματα ρi. Από τη σύγκρισή τους προκύπτει το πλέον κοντινό στοιχείο και δίνεται ως έξοδος το ακουστικό του σύμβολο αi



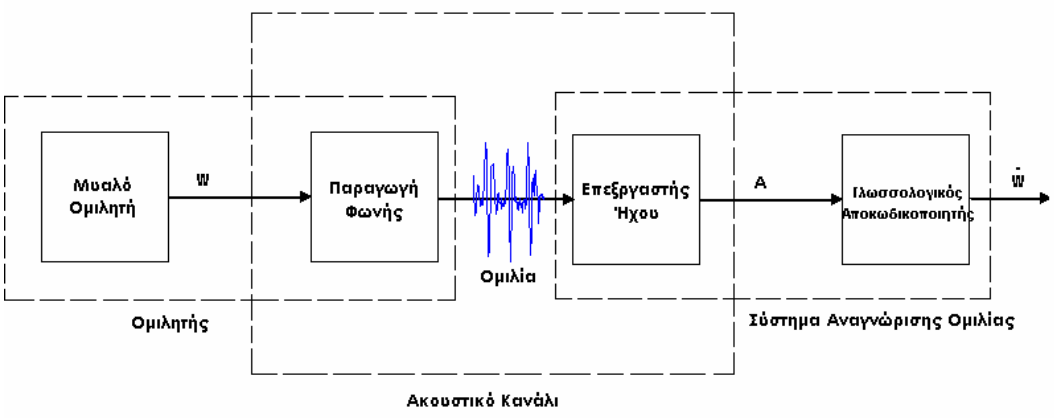
Αυτά τα χαρακτηριστικά σύμβολα αi χρησιμεύουν ως η είσοδος στο σύστημα αναγνώρισης ομιλίας που γενικά χρησιμοποιεί τεχνικές Κρυμμένων Μοντέλων Markov (HMM) για να προσδιοριστεί η πλέον πιθανή ακολουθία λέξεων που θα μπορούσε να έχει παραγάγει την έξοδο. Συγκεκριμένα δίνεται μια ακολουθία συμβόλων και πρέπει να προσδιορίσει την πλέον πιθανή ακολουθία λέξεων που θα μπορούσε να έχει παραγάγει εκείνη την είσοδο. Ένα σημαντικό ζήτημα που εξετάζεται είναι σε ποιο επίπεδο πρέπει να είναι βασισμένο το HMM. Μια επιλογή είναι να υπάρξει ένα δίκτυο HMM που να καθορίζει την πιθανή ακολουθία των συμβόλων κωδικοποίησης που θα μπορούσαν να αναγνωρίσουν τη λέξη. Για τις εφαρμογές περιορισμένου λεξιλογίου αυτό είναι μια βιώσιμη τεχνική. Είναι δυνατό να ληφθούν αρκετά παραδείγματα εκπαίδευσης για κάθε λέξη για να καθορίσει το δίκτυο. Το αποτέλεσμα είναι ένα ισχυρό σύστημα αναγνώρισης. Αλλά για μια εφαρμογή μεγάλου λεξιλογίου είναι δύσκολο να βρεθούν αρκετά στοιχεία εκπαίδευσης, γι’ αυτό σε τέτοιες περιπτώσεις

χρησιμοποιούνται τεχνικές υποθέσεων. Ένα άλλο πρόβλημα παρουσιάζεται στις εφαρμογές συνεχούς ομιλίας.

Εδώ μια λέξη θα αναγνωριζόταν διαφορετικά ανάλογα με τις περιβάλλουσες λέξεις. Λόγω αυτών των δυσκολιών τα συστήματα αναγνώρισης ομιλίας μεγάλου λεξιλογίου είναι χαρακτηριστικά βασισμένα σε μικρότερες μονάδες. Το φώνημα θα φαινόταν να είναι μια φυσική μονάδα. Σε ένα βασισμένο σε φωνήματα σύστημα, θα υπήρχε ένα δίκτυο HMM που να καθορίζει την πιθανή ακολουθία που πραγματοποιεί κάθε φώνημα. Σημειώστε ότι το φώνημα /t/ θα φανεί πολύ διαφορετικό ανάλογα με το πλαίσιό του.

Όλη η διαδικασία επικοινωνίας, από τον ομιλητή έως τον υπολογιστή φαίνεται στην εικόνα που βρίσκεται παρακάτω. Ο ομιλητής φαίνεται να αποτελείται από δύο τμήματα: το μυαλό, που καθορίζει τις λέξεις W που θα προφερθούν, από όργανο ομιλίας του. Ομοίως το σύστημα αναγνώρισης αποτελείται από δύο μέρη: τον επεξεργαστή ομιλίας και τον γλωσσολογικό αποκωδικοποιητή. Το δεύτερο περιλαμβάνει τα ακουστικά και γλωσσολογικά μοντέλα.

Επομένως ένα σύστημα αναγνώρισης ομιλίας δίνει την πλέον πιθανή ακολουθία λέξεων για να χρησιμεύσει ως η είσοδος στο σύστημα επεξεργασίας φυσικής γλώσσας. Όταν το σύστημα φυσικής γλώσσας θέλει να παραγάγει μια έκφραση, περνά μια πρόταση σε μια ενότητα που μεταφράζει τις λέξεις μέσα σε μια φθογγική ακολουθία και καθορίζει ένα περίγραμμα, και περνά έπειτα αυτές τις πληροφορίες στο σύστημα σύνθεσης, το οποίο παράγει την προφορική έξοδο.



Γίνεται αντιληπτό από τα παραπάνω ότι ένα σύστημα αναγνώρισης ομιλίας (speech recognizer) για να μπορέσει να αναγνωρίσει σωστά χρειάζεται εκτός από τα

ακουστικά μοντέλα και κάποια γλωσσολογική γνώση π.χ. κάποιες ακολουθίες λέξεων έχουν μεγαλύτερη πιθανότητα να εμφανιστούν ανάλογα με το θέμα (domain). Σε ένα σύστημα που αναγνωρίζει ομιλία σχετική με τον καιρό είναι πιο πιθανό να εμφανίζονται λέξεις όπως βροχή, συννεφιά κ.λ.π. και όχι πολιτική, δικηγόρος κ.τ.λ. Αυτή η γλωσσολογική γνώση περιγράφεται από ένα γλωσσολογικό μοντέλο (language model).

Σε διαλογικά συστήματα (dialogue systems) , όπου το σύστημα έχει την πρωτοβουλία, ο χρήστης ακολουθεί τις ερωτήσεις του συστήματος χωρίς να μπορεί να παίρνει πρωτοβουλίες. Αντίθετα σε mixed initiative συστήματα όπου το σύστημα και ο χρήστης μοιράζονται την πρωτοβουλία είναι βασικό να ξέρουμε σε κάθε σημείο του διαλόγου αυτό που είναι πιο πιθανό να πει ο χρήστης, ώστε να υποβοηθάτε ο speech recognizer από το κατάλληλο γλωσσολογικό μοντέλο. Δηλαδή ανάλογα με την πρόβλεψη γι' αυτό που σκοπεύει να κάνει ο χρήστης (dialogue act) δίνουμε μεγαλύτερη πιθανότητα σε κάποιες ακολουθίες λέξεων διευκολύνοντας τον speech recognizer.

Όπως φαίνεται από τα παραπάνω, ένα σύστημα αναγνώρισης ομιλίας χρειάζεται ένα καλό γλωσσολογικό μοντέλο για να μπορεί να αποφασίζει όταν το σύστημα αναγνώρισης δεν είναι σε θέση να αναγνωρίσει την είσοδο που δέχθηκε. Η βελτίωση του γλωσσολογικού μοντέλου επιτυγχάνεται μέσα από την εκπαίδευση του συστήματός μας, δηλαδή με καταχώρηση δεδομένων από ενδεικτικές πηγές που επιλέγει ο μηχανικός, τέτοιες ώστε το σύστημα να αποκτήσει αρκετή γνώση.

Κεφάλαιο 3 - Επεξεργασία του λόγου

3.1 Ψυχοφυσιολογία της επεξεργασίας λόγου

Το ηχητικό κύμα της ομιλίας περιλαμβάνει γλωσσικές πληροφορίες, όπως τα χαρακτηριστικά της φωνής του ομιλητή και το συναίσθημά του. Αυτή η ανταλλαγή πληροφοριών μέσω του λόγου ξεκάθαρα παίζει ένα πολύ σημαντικό ρόλο στις ζωές μας.Οι ακουστικές και γλωσσικές δομές του λόγου έχει επιβεβαιωθεί ότι είναι συνδεδεμένες με την ευφυΐα, και ότι είναι άμεσα αποτελέσματα της κουλτούρας και της κοινωνικής μας ανάπτυξης. Το ηχητικό κύμα του λόγου μεταφέρει επίσης το νόημα που επιθυμεί να μεταδώσει ο ομιλητής, πληροφορίες για το άτομό του. Όλα αυτά συνοψίζονται σε σημασιολογικό και πραγματολογικό περιεχόμενο. Αδιαμφισβήτητα, η ικανότητα απόκτησης και παραγωγής γλώσσας, και κατασκευής και χρήσης των γλωσσικών εργαλείων, είναι τα δυο κύρια χαρακτηριστικά που διαχωρίζουν τους ανθρώπους από τα υπόλοιπα ζώα.

Επιπλέον, η γλώσσα και η πολιτισμική ανάπτυξη είναι αλληλένδετες. Ακόμη κι αν ο γραπτός λόγος είναι αποτελεσματικός για την ανταλλαγή γνώσης και διαρκεί περισσότερο από τον προφορικό, η ποσότητα των πληροφοριών που ανταλλάσσεται μέσω του προφορικού λόγου είναι αρκετά μεγαλύτερη. Με πιο απλοποιημένους όρους, τα βιβλία, τα περιοδικά, μεταφέρουν πολλή πληροφορία, αλλά αυτή η μετάδοση είναι μονοκατευθυντική, άρα ελλειμματική. Η παραγωγή του ανθρώπινου λόγου ξεκινά με την αρχική εννοιολογική κατάκτηση της ιδέας ότι π ομιλητής θέλει να μεταφέρει κάτι στον ακροατή. Στη συνέχεια, ο ομιλητής μετατρέπει αυτή την ιδέα σε γλωσσικές δομές μέσω της επιλογής των κατάλληλων λέξεων ή φράσεων οι οποίες το αναπαριστούν, και στη συνέχεια τις βάζει στην κατάλληλη σειρά ανάλογα με τους γραμματικούς και συντακτικούς κανόνες της γλώσσας του, οι οποίοι μπορεί να τηρηθούν απόλυτα ή χαλαρά, ανάλογα το πλαίσιο της επικοινωνίας, την επισημότητα, και τη σχέση μεταξύ των ομιλητών. Αυτή τη διεργασία την υποστηρίζει ο ανθρώπινος εγκέφαλος, ο οποίος παράγει κινητικά σήματα στους κατάλληλους νευρώνες ώστε να κινηθούν οι αντίστοιχοι μύες των φωνητικών οργάνων. Αυτή η διαδικασία μπορεί να διακριθεί σε δύο υπό-ομάδες. Τη φυσιολογική επεξεργασία που περιλαμβάνει νευρώνες και μυς, και τη φυσική διαδικασία μέσα από την οποία το ηχητικό κύμα του λόγου παράγεται και μεταδίδεται.

Τα χαρακτηριστικά του λόγου ως φυσικό φαινόμενο είναι ουσιαστικά ότι συνεχή, αν και η γλώσσα ως επικοινωνιακό εργαλείο αποτελείται από διακριτές κωδικοποιημένες μονάδες. Μια πρόταση δομείται με τη χρήση βασικών λεξικών μονάδων, με την κάθε λέξη να αποτελείται από συλλαβές και κάθε συλλαβή να αποτελείται από φωνήματα, τα οποία με τη σειρά τους μπορούν να ταξινομηθούν σαν φωνήεντα και σύμφωνα. Αν και η συλλαβή καθαυτή δεν μπορεί να οριστεί επακριβώς, γνωρίζουμε ότι σχηματίζεται από το συνδυασμό ενός φωνήεντος με ένα ή περισσότερα σύμφωνα. Ο αριθμός των φωνηέντων και των συμφώνων ποικίλλει, ανάλογα με τη μέθοδο ταξινόμησης και τη γλώσσα υπό εξέταση. Εν πολλοίς, τα Αγγλικά έχουν 12 φωνήεντα και 24 σύμφωνα, ενώ τα Ιαπωνικά έχουν 5 φωνήεντα και 20 σύμφωνα. Ο αριθμός των φωνημάτων σε μια γλώσσα σπάνια φτάνει τα 50. Από τη στιγμή που υπάρχουν κανόνες συνδυασμού για το σχηματισμό των φωνημάτων σε συλλαβές, ο αριθμός των συλλαβών σε κάθε γλώσσα περιλαμβάνει μόνο ένα μικρό μέρος όλων των πιθανών φωνημικών συνδυασμών. Αντίθετα με το φώνημα, το οποίο είναι η μικρότερη μονάδα λόγου από γλωσσολογικής ή φωνημικής πλευράς, η φυσική μονάδα του παραγόμενου λόγου είναι το αλλόφωνο. Αν και ο αριθμός των λέξεων σε κάθε γλώσσα είναι πολύ μεγάλος και αυξάνεται συνεχώς, ο συνολικός αριθμός τους είναι πολύ μικρότερος από όλες τις συλλαβές και τους πιθανούς φωνημικούς συνδυασμούς. Έχει υποστηριχθεί ότι ο αριθμός των πιο συχνά χρησιμοποιούμενων λέξεων είναι μεταξύ 2.000 και 3.000 και ότι ο αριθμός των λέξεων που χρησιμοποιείται από το κάθε άτομο είναι 5000 με 10.000 λέξεις. Ο τόνος και ο τονισμός παίζουν πολύ σημαντικό ρόλο στην ανάδειξη των πιο σημαντικών λέξεων, στο σχηματισμό ερώτησης, και στην αναγνώριση συναισθήματος του ομιλητή.

3.2 Ομιλία και Ακοή

Ο λόγος προφέρεται για το σκοπό της επικοινωνίας, επομένως θα πρέπει να υπάρχει δέκτης ικανός για την αποκωδικοποίησή του. Άρα, σε αυτή τη διαδικασία παίζει πρωταγωνιστικό ρόλο η ακοή. Το ηχητικό κύμα του λόγου που παράγεται από τα φωνητικά όργανα και μεταδίδεται μέσω του αέρα στα αυτιά των ακροατών, ενεργοποιεί τα ακουστικά όργανα προκειμένου αυτά να παράγουν νευρικούς παλμούς τα οποία μεταφέρονται στον εγκέφαλο του ακροατή μέσω του ακουστικού νευρικού συστήματος. Αυτό επιτρέπει τη γλωσσολογική πληροφορία την οποία ο ο ομιλητής αποσκοπεί να μεταδώσει να είναι κατανοητή από τον ακροατή. Το ίδιο ηχητικό κύμα

μεταδίδεται φυσικά στα αυτιά και του ομιλητή επίσης, επιτρέποντάς του να ελέγχει συνεχώς τα φωνητικά του όργανα μέσω της ακοής της ίδιας του της ομιλίας ως ανατροφοδότηση.

Η σημαντική διαφορά αυτού του μηχανισμού ανατροφοδότησης είναι ξεκάθαρα εμφανής σε ανθρώπους που η ακοή τους έχει υποστεί βλάβη για πάνω από ένα ή δύο χρόνια. Είναι επίσης προφανές ότι είναι δύσκολο να μιλήσει κάποιος χωρίς να ακούει τη φωνή του με κάποια μικρή χρονοκαθυστέρηση. Η εσωτερική σύνδεση μεταξύ παραγωγής λόγου και ακοής ονομάζεται αλυσίδα λόγου. Με όρους παραγωγής, η αλυσίδα λόγου αποτελείται από γλωσσολογικά, φυσιολογικά, και ακουστικά στάδια, η σειρά των οπίων αντιστρέφεται κατά το άκουσμα. Η ανθρώπινη ακοή συντελείται από έναν μηχανισμό που παρουσιάζει ιδιαίτερα υψηλού επιπέδου οργάνωση, που μέχρι στιγμής δεν έχει καταστεί δυνατό να αναπαρασταθεί πλήρως με μέσα τεχνητής νοημοσύνης.

Ένα πλεονέκτημα της ανθρώπινης ακοής είναι η επιλεκτική ακοή η οποία επιτρέπει στον ακροατή να ακούσει μόνο μία φωνή ακόμη κι αν άλλοι άνθρωποι μιλούν ταυτόχρονα κι ακόμη κι αν η φωνή αυτού του ατόμου ακούγεται να μιλά ακαθόριστα, με έντονη διάλεκτο ή ιδιαίτερο τόνο. Από την άλλη, η ανθρώπινη ακοή έχει έναν μηχανισμό που παρουσιάζει πολύ μικρή ισχύ. Ένα παράδειγμα από το εγγενές μειονέκτημά του είναι ότι δεν είναι ικανό το αυτί να ξεχωρίσει δύο τόνους που είναι παρόμοιοι σε συχνότητα ή που έχουν ένα μικρό κενό χρόνου μεταξύ τους. Η υψηλού επιπέδου ακοή μας υποστηρίζεται από τον περίπλοκο μηχανισμό που διαθέτουμε για την αντίληψη της γλώσσας και την κατανόησή της, ο οποίος διαμεσολαβείται από τον εγκέφαλο που χρησιμοποιεί πληροφορίες με διαφορετικό περιεχόμενο για την εκτέλεση των διάφορων νοητικών διεργασιών. Οι εγγενείς σχέσεις μεταξύ αυτών των μηχανισμών, επομένως, επιτρέπουν στους ανθρώπους να επικοινωνούν αποτελεσματικά μεταξύ τους.

3.3 Ακουστικά χαρακτηριστικά του λόγου

Ο λόγος είναι ένα ηχητικό κύμα ημιτονοειδούς μορφής, κι ως τέτοιο διαθέτει ορισμένα χαρακτηριστικά, όπως συχνότητα, εύρος, ταχύτητα και κατεύθυνση. Η συχνότητα αφορά τον αριθμό παλμών μέσα σε μια επαναλαμβανόμενη μονάδα χρόνου, κι ουσιαστικά αφορά την οριζόντια μορφή του ηχητικού κύματος. Το εύρος αφορά την αλλαγή του μέσα σε μια μεμονωμένη χρονική περίοδο. Η ταχύτητα είναι η απόσταση

που διανύει το κύμα ανά μια μονάδα χρόνου καθώς μεταδίδεται μέσω ενός ελαστικού μέσου. Τέλος η κατεύθυνση αφορά την πληροφορία που εξάγουμε από τη σχετική θέση ενός στόχου σε σχέση με ένα σημείο έναρξης, το οποίο μαθηματικά δίνεται από ένα διάνυσμα.[3]

Η χρήση του όρου «ήχος» περιορίζεται στα επιστημονικά πεδία της ψυχολογίας και της φυσιολογίας, οι οποίες ενδιαφέρονται για την πρόσληψή του από τον εγκέφαλο.Το πεδίο της ψυχοακουστικής είναι ένα νέο επιστημονικό πεδίο που προέκυψε από τον συνδυασμό των δύο παραπάνω, κι ασχολείται ακριβώς με την αντίληψη του ήχου, κι εξετάζει την επίδραση του ηχητικού κύματος στον εγκέφαλο και το νου. Γνωρίζουμε ότι η ανθρώπινη αντίληψη είναι περιορισμένη σε κάποιες συχνότητες, από 20Hz ως 20.000Hz, όριο το οποίο μειώνεται με την αύξηση της ηλικίας. Υπάρχουν έξι στοιχεία τα οποία διαχωρίζουν μεταξύ τους οι ερευνητές του χώρου όσον αφορά το πώς προσλαμβάνει ο άνθρωπος τον ήχο: η ένταση, η διάρκεια, ο χωρικός εντοπισμός, η ηχητική υφή, η χροιά, και η ταλάντωση. Η ταλάντωση αφορά την αντιληπτική ιδιότητα του ήχου η οποία επιτρέπει τη σειροθέτηση όσον αφορά τη συχνότητα. Είναι η ποιότητα εκείνη που καθιστά δυνατό τον χαρακτηρισμό ορισμένων ήχων ως υψηλότερων ή χαμηλότερων, και συναντάται μόνο σε ήχους με συχνότητα τέτοια που τους κάνει να ξεχωρίζουν από το θόρυβο, δηλαδή πιο καθαρή και σταθερή. Η χροιά είναι το χαρακτηριστικό που μοιάζει να δίνει χρώμα στον ήχο. Η ηχητική υφή περιγράφει το πώς τα αρμονικά συστατικά συνδυάζονται σε μια ενιαία σύνθεση κι αποδίδει την αίσθηση της πυκνότητας του ήχου. Τέλος, ο χωρικός εντοπισμός αφορά την ικανότητα του ακροατή να εντοπίζει την πηγή του ήχου μέσα στο χώρο και να προσανατολίζεται μόνο με βάση ακουστικές πηγές. Η πειραματική ενασχόληση των ψυχολόγων και των φυσιολόγων με όλα αυτά τα αντιληπτικά στοιχεία του ήχου οδήγησε σε μια πολύ καλή ποσοτικοποίησή τους και στην ανάπτυξη μεθόδων καταμέτρησης. Έτσι, πολλά από αυτά πλέον ονομάζονται «στατιστικά μέρη» του λόγου, καθώς αλλαγές στο ένα στοιχείο επιφέρει αλλαγές και σε κάποιο άλλο όσον αφορά την αντίληψη. Ο απώτερος σκοπός είναι να κατασκευαστούν συστήματα τεχνητής νοημοσύνης που να εκτελούν κάποια από τα παραπάνω για την επίτευξη των έργων που επιτελεί και ο άνθρωπος.

3.4 Ψηφιοποίηση

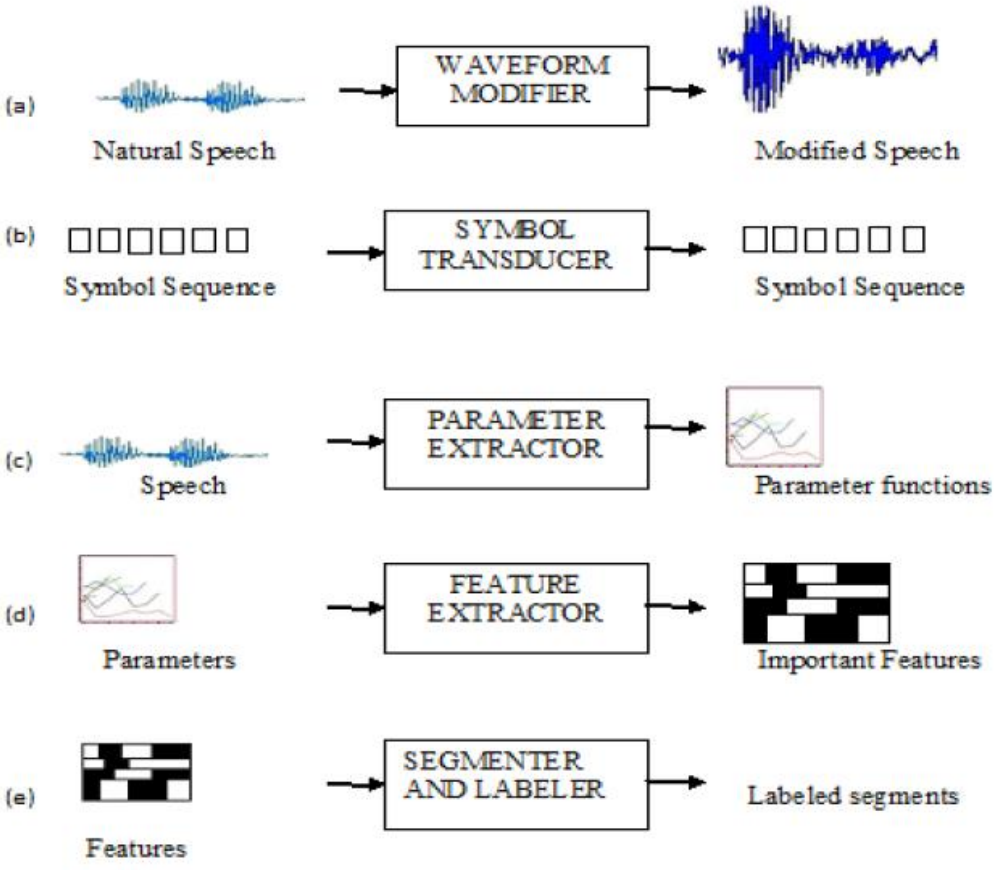
Το ηχητικό σήμα του λόγου μπορεί να γίνει ένα προσβάσιμο αντικείμενο μέσω της μετατροπής του σε ηλεκτρικό σήμα χρησιμοποιώντας το μικρόφωνο. Το ηλεκτρικό σήμα συνήθως μετασχηματίζεται από αναλογικό σε ψηφιακό πριν από κάθε άλλη γλωσσική επεξεργασία, για δύο λόγους. Ο πρώτος λόγος είναι ότι οι ψηφιακές τεχνικές διευκολύνουν την υψηλού επιπέδου επεξεργασία του σήματος που δε θα μπορούσε να επιτευχθεί με τις αναλογικές τεχνολογίες. Ο δεύτερος λόγος είναι ότι η ψηφιακή επεξεργασία είναι πολύ πιο αξιόπιστη και μπορεί να επιτευχθεί με τη χρήση ενός μόνο κυκλώματος.Η γρήγορη ανάπτυξη των υπολογιστών και των ολοκληρωμένων κυκλωμάτων σε συνδυασμό με την ανάπτυξη των ψηφιακών επικοινωνιών και δικτύων έχουν δώσει ώθηση στην εφαρμογή των τεχνικών ψηφιακής επεξεργασίας του ήχου.Η μετατροπή από αναλογικό σε ψηφιακό σήμα, διαδικασία γνωστή ως ψηφιοποίηση, αποτελείται από τη διεργασίες δειγματοληψίας, κβαντισμού και κωδικοποίησης. Η δειγματοληψία αφορά την απεικόνιση ενός συνεχόμενα μεταβαλλόμενου σήματος ως μια περιοδική ακολουθία τιμών. Ο κβαντισμός αφορά την αναπαράσταση μιας τιμής από την κυματομορφή με ένα πεπερασμένο σύνολο τιμών. Η κωδικοποίηση είναι η απόδοση ενός συγκεκριμένου αριθμό σε κάθε τιμή. Για ένα τέτοιο έργο, η διττή κωδικοποίηση που χρησιμοποιεί διττή αριθμητική αναπαράσταση είναι αυτή που χρησιμοποιείται και συχνότερα. Αυτές οι διαδικασίες, επομένως, διευκολύνουν ένα συνεχές αναλογικό σήμα να μετατραπεί σε μια ακολουθία κωδικών που έχουν επιλεγεί από ένα πεπερασμένο σύνολο (Furui, 2000).[11]

Κεφάλαιο 4 – Στοιχεία φωνής

4.1 Βασικά δομικά στοιχεία

Για την διαδικασία της αναγνώρισης ομιλίας υπάρχουν ορισμένα δομικά στοιχεία τα οποία εκτελούν μετασχηματισμούς. Μια κυματομορφή, ένα σύμβολο (α), λαμβάνει ένα σήμα ομιλίας εισόδου και παράγει ένα τροποποιημένο σήμα. Η τροποποίηση μπορεί να είναι ένα μέρος από μεγάλες τιμές του σήματος ή αποτέλεσμα φιλτραρίσματος του φάσματος συχνοτήτων που μεταβάλλει το σχήμα του σήματος ή ενισχύει την ομιλία και μειώνει θόρυβο που είναι παρών. Ενας μετατροπέας, σύμβολο (β), μπορεί να λάβει σε ένα διακριτό σημείο ακολουθία συμβόλων και αποδίδει μια τροποποιημένη αλληλουχία στην έξοδό του. Αν η είσοδος ήταν μια ακολουθία λέξεων σε μια γλώσσα και η έξοδος ήταν μια ισοδύναμη ακολουθία λέξη σε μια άλλη γλώσσα, αυτό το αισθητήριο θα ήταν μια συσκευή μετάφρασης. Παράμετρος απαγωγέας, (αναπαρίσταται στο σύμβολο γ) ο οποίος λαμβάνει ένα σήμα ομιλίας εισόδου και αποδίδει παραμέτρους του κύματος ομιλίας. Συχνά αποκαλείται προ-επεξεργασία.

Ένα χαρακτηριστικό απαγωγέα, σχήμα 1 (δ), μπορεί να λάβει τις παραμέτρους και να παράγει ένα πιο αφηρημένο σύνολο των σημαντικών πληροφοριών που φέρουν χαρακτηριστικά, όπως τον προσδιορισμό του τι τμήματα της ομιλίας εκφράζει, αν ο ήχος είναι δυνατά, κι αν τα φωνήεντα είναι αρκετά συγχρονισμένα. Επίσης, το σημείο (ε) μπορεί να δεχθεί το σύνολο των χαρακτηριστικών και να παράγει μια γραμμική σειρά από φωνήματα ή άλλων τομέων για αναγνώριση. Η αναγνωριστική μονάδα στο σχήμα (στ) λαμβάνει ακολουθία συμβόλων εισόδου το οποίο μπορεί να συγκριθεί με τις αναμενόμενες αλληλουχίες αναφοράς για διάφορες μονάδες για να καθορίσει τι γλωσσικές μονάδες φαίνεται να είναι στην είσοδο. Το αναγνωριστικό πιο κοινή μονάδα είναι μια » λέξη αντιστοίχισης», η οποία βρίσκει την πλησιέστερη αντιστοίχιση λέξεων, με βάση την οποία λέξης αποθηκεύονται εγχόρδων προφορά είναι πιο όπως η συμβολοσειρά εισόδου. Με αυτά τα δομικά στοιχεία, έχουμε τις βασικές προϋποθέσεις για τη συζήτηση των κύριων πηγών γνώσεις που απαιτούνται για τη μηχανή κατανόηση της ομιλίας. Όλα τα παραπάνω απεικονίζονται στην φωτογραφία που ακολουθεί.



4.2 Τύποι του λόγου

4.2.1 Μεμονωμένες λέξεις

Μεμονωμένες λέξεις που εκφέρονται ανεξάρτητα μεταξύ τους. Τα συστήματα αυτά επιδέχονται την άρνηση «Ακούστε / Not » ώστε να μπουν σε κατάσταση αναμονής, όπου απαιτείται από τον ομιλητή να προφέρει εκφράσεις. Μεμονωμένες Διατυπώσεις θα μπορούσε να είναι ένα καλύτερο όνομα για αυτή την κατηγορία(O' Shoughnessy, 2003) . [10]

4.2.2 Συνδεδεμένες λέξεις

Συνδεδεμένη αναγνώριση λέξεων είναι το σύστημα όπου οι λέξεις χωρίζονται από παύσεις. Συνδεδεμένη αναγνώριση λέξεων είναι μια κατηγορία ακολουθιών λέξεων, όπου το σύνολο των ακολουθιών μπορεί να προέρχεται από μικρές έως μέτριες σε μέγεθος λέξεις, καθώς και αλφαριθμητικούς χαρακτήρες. Όπως και στην αναγνώριση

μεμονωμένων λέξεων, αυτό το σύνολο επίσης έχει μια ιδιότητα που η βασική μονάδα αναγνώρισης είναι η λέξη (Arora, 2012).

4.2.3 Συνεχής ομιλία

Η συνεχής αναγνώριση ομιλίας αφορά την ομιλία όπου οι λέξεις συνδέονται μεταξύ τους αντί να χωρίζονται από παύσεις. Ως αποτέλεσμα το σύνολο των πληροφοριών είναι αρκετά μεγαλύτερο, κι έτσι απαιτείται μεγαλύτερη απόδοση και διαδικασία συνεχούς αναγνώρισης ομιλίας. Τα στοιχεία αναγνώρισης με δυνατότητες συνεχούς ομιλίας είναι μερικά από τα πιο δύσκολα να δημιουργηθούν επειδή χρησιμοποιούν ειδικές μεθόδους για να καθορίσουν τα όρια των εκφράσεων (Arora, 2012).

4.3 Ταξινόμηση συστημάτων φωνητικής αναγνώρισης

Τα συστήματα αναγνώρισης ομιλίας ταξινομούνται ως διακριτά ή συνεχή συστήματα που εξαρτώνται από τον εισερχόμενο ήχο ή είναι ανεξάρτητα. Τα διακριτά συστήματα διατηρούν ένα ξεχωριστό ακουστικό μοντέλο για κάθε λέξη, ο συνδυασμός των λέξεων ή φράσεων που αναφέρονται ως απομονωμένες αναγνώριση λέξη ομιλίας (ISR). Η συνεχής αναγνώριση ομιλίας (ΕΚΕ) στα συστήματα ανταποκρίνεται σε ένα αίτημα του χρήστη που προφέρει λέξεις, φράσεις ή προτάσεις που βρίσκονται σε μια σειρά από συγκεκριμένη σειρά. Ένα σύστημα ήχων απαιτεί την εγγραφή του χρήστη, όπως ένα παράδειγμα λέξης, φράσης ή μια φράση πριν από την αναγνώρισή της από το σύστημα, δηλαδή ο χρήστης εκπαιδεύει το σύστημα. Ένα σύστημα εισερχομένων ανεξάρτητο, δεν απαιτεί καμία εγγραφή πριν από τη χρήση του συστήματος. Έχει αναπτυχθεί για να λειτουργήσει για οποιοδήποτε εισερχόμενο ενός συγκεκριμένου τύπου. Τα συστήματα ήχων που εξαρτώνται είναι πιο εύκολο να κατασκευάσει κανείς και να είναι πιο ακριβή από ό, τι τα συστήματα ανεξάρτητα από τους ήχους. Ως εκ τούτου, το επίκεντρο του ενδιαφέροντος στον τομέα των συστημάτων αναγνώρισης φωνής είναι κατά κύριο λόγο ήχοι που βασίζονται σε μεμονωμένες λέξεις και συστήματα που χρησιμοποιούνται με βάση περιορισμένο λεξιλόγιο. Έτσι, ξεπερνούν τους περιορισμούς στην κατάσταση της τεχνολογίας που απαιτείται μεγαλύτερη έμφαση στην αλληλεπίδραση ανθρώπου-προς-υπολογιστή. Η πρόκληση είναι ο προδιορισμός του κατά πόσο η βελτιωμένη τεχνολογία αναγνώρισης ομιλίας θα

μπορούσε να χρησιμοποιηθεί για να υποστηρίξει την ενίσχυση της ανθρώπινης αλληλεπίδρασης με τις μηχανές. Ένα σημαντικό στοιχείο για τη δημιουργία του συστήματος αναγνώρισης ομιλίας είναι το μέγεθος του λεξιλογίου. Το λεξιλόγιο επηρεάζει την πολυπλοκότητα και την ακρίβεια του συστήματος. Το μέγεθος του λεξιλογίου μπορεί να είναι μικρό, μεσαίο ή μεγάλο. Ένα άλλο σημαντικό προσδιοριστικό για τον προσδιορισμό της πολυπλοκότητας του συστήματος της αναγνώρισης ομιλίας είναι ο τύπος του λόγου που χρησιμοποιεί το σύστημα αναγνώρισης: διακριτή ή συνεχής. Σε ένα διακριτό σύστημα ομιλία του, ο χρήστης πρέπει να βάλει μία παύση ανάμεσα σε κάθε λέξη που καθιστά το έργο της αναγνώρισης ομιλίας πολύ πιο εύκολο. Η συνεχής ομιλία είναι πιο δύσκολη λόγω των πολλών εισερχομένων. Κατ 'αρχάς, είναι δύσκολο να βρούμε το όριο έναρξης και λήξης των λέξεων. Ένα άλλο πρόβλημα είναι ότι η παραγωγή των φωνημάτων επηρεάζει την παραγωγή κάθε φθόγγου. Επίσης ο ρυθμός ομιλίας επηρεάζει την αναγνώριση της συνεχούς ομιλίας (Annusuya, 2014) .

4.4 Αναγνώριση προτύπων

Η αναγνώριση προτύπων είναι η επιστήμη που ασχολείται με την αναγνώριση κανονικοτήτων σε θορυβώδη και πολύπλοκα περιβάλλοντα με τρόπο αυτόματο. Είναι μια προσέγγιση που παρουσιάζει τεράστιες απαιτήσεις σε ευρύτητα, καθώς υπάρχει αυξανόμενη ανάγκη για αναγνώριση προτύπων σε πολλούς τομείς. Πρωταρχικής σημασίας είναι η Θεωρία Ταξινόμησης (Classification Theory), που δεν είναι άλλη από την ταξινόμηση αντικειμένων σε ξεχωριστές κατηγορίες ή κλάσεις.[3] Είναι πάντα το πρώτο βήμα για την αναγνώριση προτύπων. Βασίζεται στη στατιστική θεωρία αποφάσεων και παρέχει τη μαθηματική τεκμηρίωση και τις διαδικασίες για την αναπαράσταση των κατηγοριοποιημένων στοιχείων με τη χρήση διανυσμάτων(O' Shoughnessy, 2003).

4.5 Μηχανική αντίληψη

Πρόκειται για είδος προγραμμάτων που μπορούν να ταξινομήσουν πρότυπα, και χρησιμεύουν στην αναγνώριση ομιλίας και ομιλητή, στην αναγνώριση χειρογράφων, στην αναγνώριση χαρακτήρων, και στην αναγνώριση δακτυλικών αποτυπωμάτων. Οι εφαρμογές της μηχανικής αντίληψης απαντώνται στην αναγνώριση ακολουθιών DNA, στα τεστ μαστογραφίας, στη διάγνωση μέσω ηλεκτροεγκεφαλογραφήματος, στη

διάγνωση μέσω ηλεκτρο-καρδιογραφημάτων, στην αναγνώριση με σάρωση της ίριδας, και στην τοπογραφική τηλεσκόπηση (Shabtai, 2010). [4][5]

4.6 Συστήματα Αναγνώρισης Προτύπων

Η παρούσα παράγραφος περιγράφει σε μεγαλύτερη λεπτομέρεια την ακολουθία ενεργειών που εκτελεί ένα σύστημα αναγνώρισης προτύπων οποιασδήποτε χρησιμότητας. Αρχικά υπάρχει ο αισθητήρας, ο οποίος είναι το μέσο συλλογής του υλικού, κάμερα ή μικρόφωνο. Υπάρχει εξάρτηση από το εύρος ζώνης, τις παραμορφώσεις του σήματος και τη διακριτική ικανότητα. Στη συνέχεια καταμερισμός (ή μερισμός) και ομαδοποίηση, να διαχωριστούν δηλαδή σε ομάδες ανάλογα με τα χαρακτηριστικά τους και να μην υπάρχει επικάλυψη. Τα χαρακτηριστικά, στη συνέχεια, τυγχάνουν μεγαλύτερης ανάλυσης κι εξάγονται τα πιο χρήσιμα και προβλεπτικά χαρακτηριστικά ως προς κλιμάκωση, στροφή και ανάκλαση, και λαμβάνουν ένα διάνυσμα. Το επόμενο βήμα αφορά την ταξινόμηση, βάσει του διανύσματος, ώστε, τέλος, να γίνει η μετα-επεξεργασία, κατά την οποία γίνεται χρήση οποιασδήποτε άλλης πληροφορίας για τη βελτίωση της απόδοσης (O' Shoughnessy, 2003).[4][5]

4.7 Στάδια σχεδίασης

Προκειμένου να δημιουργηθεί ένα σύστημα αυτόματης αναγνώρισης ομιλίας, εκτελούνται ορισμένα βήματα. Αρχικά γίνεται η συλλογή των δεδομένων, τα οποία θα αποτελέσουν το υλικό εκπαίδευσης. Συλλέγονται αρκετά δεδομένα για την επαρκή εκπαίδευση, δοκιμή και αξιολόγηση του συστήματος. Στη συνέχεια επιλέγονται τα χαρακτηριστικά βάσει των οποίων επιθυμούμε να γίνει η εκπαίδευση. Τα επιλέγουμε βάσει επίδρασης θορύβου, απλότητας, κ.α. Έπειτα επιλέγουμε το μοντέλο, το οποίο είναι μια συνάρτηση πυκνότητας πιθανότητας. Το επόμενο βήμα είναι η εκπαίδευση του μοντέλου, δηλαδή εκπαιδεύεται ο ταξινομητής να κατηγοριοποιεί τα δεδομένα με μια μεθοδολογία που επιλέγουμε εμείς. Τέλος, γίνεται αξιολόγηση των αποτελεσμάτων μέσω μέτρησης του σφάλματος ταξινόμησης και υπολογιστικής πολυπλοκότητας (Ruthven, 1995).

4.8 Εκμάθηση και προσαρμοστικότητα

Η εκμάθηση του συστήματος επιτυγχάνεται με τη χρήση διάφορων μεθοδολογιών. Η εκμάθηση με επιτήρηση, διαχειρίζεται τα ονόματα των κατηγοριών ως ετικέτες και τις ερμηνεύει ως ενδεικτικά της κλάσης στην οποία ανήκουν. Η εκμάθηση χωρίς επιτήρηση κάνει χρήση συνόλων δεδομένων που δεν έχουν ταξινομηθεί προκειμένου να σχηματίζει ομαδοποιήσεις των προτύπων εισόδου. Τέλος, η προσαρμοστικότητα είναι ένα βασικό χαρακτηριστικό, καθώς υποτίθεται ότι τα χαρακτηριστικά των προτύπων αλλάζουν διαχρονικά και ο ταξινομητής θα πρέπει να είναι σε θέση να εκτελεί το έργο του χωρίς να απαιτείται εκ νέου επιτήρηση. [5]

Αποδίδεται σχηματικά η διαδικασία αναγνώρισης προτύπων, ως μια διαδικασία αναγωγής της πληροφορίας, αποτύπωσης της πληροφορίας, ή χαρακτηρισμού της πληροφορίας. Το μαθηματικό υπόβαθρο είναι πλούσιο. Υπάρχει βάση στην Γραμμική Άλγεβρα, και πιο συγκεκριμένα στις ορίζουσες και στην αντιστροφή πινάκων, στις ιδιοτιμές και στα διανύσματα, και στην παραγώγιση πινάκων. Επίσης, υπάρχει εκτενές υπόβαθρο και στη Θεωρία Πιθανοτήτων, και πιο συγκεκριμένα στη στατιστική συσχέτιση, στο θεώρημα Bayes, και στην κανονική κατανομή (Duin, 2007).

Κεφάλαιο 5 – Συστήματα αναγνώρισης φωνής

5.1 Σύστημα Αναγνώρισης φωνής

Για να αναγνωρίσουμε ένα άτομο από τη φωνή του, χρησιμοποιούμε τα λεγόμενα “φωνητικά αποτυπώματα”. Η ανθρώπινη φωνή δεν είναι ικανή να εκφέρει έναν τόνο κάθε φορά. Αντιθέτως παράγει συνεχόμενους βασικούς τόνους που δημιουργούν τη χροιά που ακούμε. Κάποιοι από τους βασικούς τόνους είναι τυχαίοι και κάποιοι είναι πολλαπλάσιοι των βασικών και ονομάζονται αρμονικές. Από όλα τα χαρακτηριστικά της ανθρώπινης φωνής τα σημαντικότερα είναι η συχνότητα και η ένταση. Η συχνότητα είναι η ταχύτητα με την οποία πάλλεται ο αέρας όταν μιλάμε, ενώ η ένταση είναι η δύναμη με την οποία εξέρχεται ο αέρας από το στόμα. Η μοναδικότητα της φωνής κάθε ανθρώπου οφείλεται τόσο στη φυσιολογία του αναπνευστικού συστήματός του, όσο και στο περιβάλλον μέσα στο οποίο έμαθε να μιλά. Ο συνδυασμός όλων των παραπάνω χαρακτηριστικών δημιουργεί την ανθρώπινη φωνή. Κάθε άνθρωπος όταν μιλά, δημιουργεί ένα μοναδικό διάγραμμα φωνής, που μπορεί να παρουσιαστεί σαν μια γραφική παράσταση που παρουσιάζει τη συχνότητα, την ένταση καθώς και τους τόνους που χρησιμοποιούνται για να διαμορφώσουν τη φωνή. Η διαδικασία από εκεί και πέρα που χρησιμοποιείται στηρίζεται στα μαθησιακά μοντέλα. Το σύστημα που χρησιμοποιούμε για να κάνει την αναγνώριση, “εκπαιδεύεται” και “μαθαίνει” συγκεκριμένες λέξεις από συγκεκριμένο άτομο, φτιάχνοντας μια βάση φωνητικών δεδομένων και συγκρίνοντας, όταν χρειαστεί τη φωνή του. Αναλόγως με τον εξοπλισμό που διαθέτουμε, το ποσοστό επιτυχημένης αναγνώρισης μπορεί να είναι αρκετά υψηλό. Όμως υπάρχουν κάποιοι περιορισμοί και κάποιοι κανόνες που πρέπει να ακολουθούνται. Για παράδειγμα, οι λέξεις και οι φράσεις που θα χρησιμοποιηθούν πρέπει να είναι ακριβώς οι ίδιες. Επίσης ο εξοπλισμός και οι συνθήκες ηχογράφησης πρέπει να είναι ίδιες. Άλλοι παράγοντες που μπορούν να αλλοιώσουν τη διαδικασία αναγνώρισης είναι η κακή υγεία του ατόμου και εξωτερικές παρεμβολές.[1]

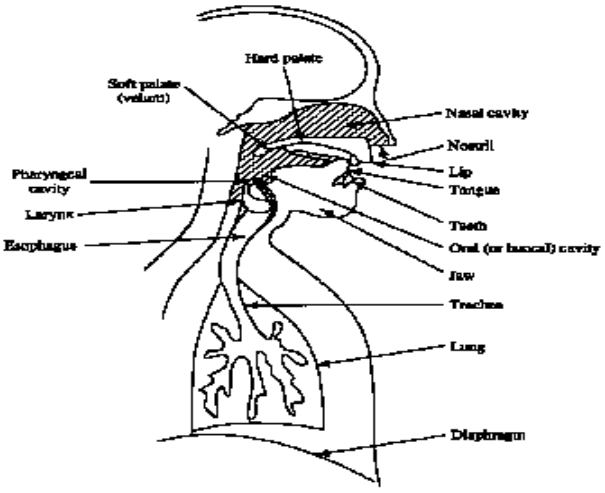
5.2 Βασικές Αρχές Αναγνώρισης Φωνής

Η αναγνώριση φωνής είναι ένα υβριδικό βιομετρικό σύστημα που συνδυάζει φυσικά χαρακτηριστικά με συμπεριφορές, αυξάνοντας με αυτόν τον τρόπο την ποιότητα

του βιομετρικού προτύπου. Ξεκινώντας από τα φυσικά χαρακτηριστικά, να πούμε ότι η ανθρώπινη φωνή εξαρτάται από πάρα πολλούς παράγοντες. Το μήκος των φωνητικών χορδών για παράδειγμα, είναι ένας από αυτούς. Το σχήμα του στόματος, των ρινικών κοιλοτήτων, και του λάρυγγα έχει επίσης ιδιαίτερη σημασία. Όλα αυτά αλληλεπιδρούν μεταξύ τους και διαμορφώνουν ένα σύνολο ιδιοτήτων (χροιά, ύψος κ.λπ.) το οποίο χαρακτηρίζει με μοναδικό τρόπο κάθε ανθρώπινη φωνή. Στο σύνολο των φυσικών ιδιοτήτων έρχονται να προστεθούν και μερικές ιδιότητες με βάση συμπεριφορές, όπως ο ρυθμός ομιλίας ή ο τονισμός, οι οποίες διασφαλίζουν ακόμα περισσότερο τη μοναδικότητα του τελικού δείγματος.

5.2.1 Επαλήθευση ομιλητών

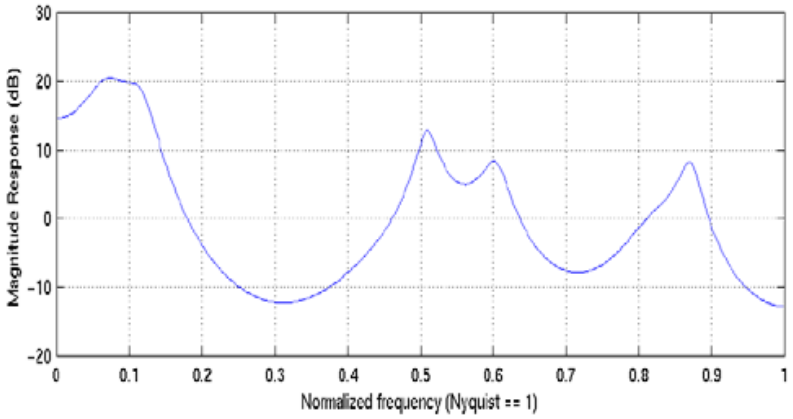
Τα συγκεκριμένα χαρακτηριστικά της ομιλίας του ομιλητή οφείλονται στις διαφορές των φυσιολογικών και συμπεριφορικών πτυχών του συστήματος λεκτικής παραγωγής των ανθρώπων. Η κύρια φυσιολογική πτυχή του ανθρώπινου συστήματος λεκτικής παραγωγής είναι η μορφή φωνητικών κομματιών. Το φωνητικό κομμάτι θεωρείται γενικά ως όργανο λεκτικής παραγωγής επάνω από τις φωνητικές πτυχές, το οποίο αποτελείται από τα εξής: (1) λαρυγγικός φάρυγγας (κάτω από την επιγλωττίδα), (2) προφορικός φάρυγγας (πίσω από τη γλώσσα, μεταξύ της επιγλωττίδας και της μαλθακής ύπερων,) (3) στοματική κοιλότητα (μπροστά από τη μαλθακή ύπερων και οριακά από τα χείλια, τη γλώσσα και τον ουρανίσκο), (4) ρινικός φάρυγγας (επάνω από τη μαλθακή ύπερων, πίσω από τη ρινική κοιλότητα),και (5) ρινική κοιλότητα (επάνω από τον ουρανίσκο και επεκτείνεται από το φάρυγγα στα ρουθούνια). Η σκιασμένη περιοχή στο παρακάτω σχήμα απεικονίζει το φωνητικό κομμάτι.

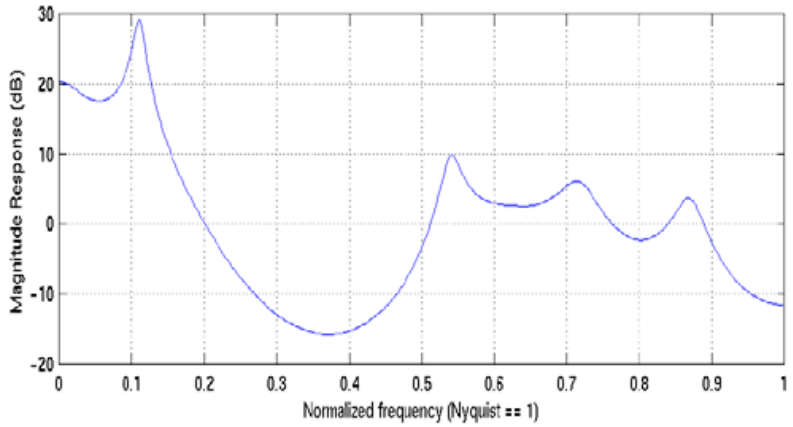


Το φωνητικό κομμάτι τροποποιεί το φασματικό περιεχόμενο ενός ακουστικού κύματος καθώς περνά διάμεσων του παράγοντας έτσι την ομιλία. Ως εκ τούτου, είναι κοινό στα συστήματα επαλήθευσης ομιλητών να χρησιμοποιηθούν τα χαρακτηριστικά γνωρίσματα που προέρχονται μόνο από το φωνητικό κομμάτι. Προκειμένου να χαρακτηριστούν τα χαρακτηριστικά γνωρίσματα του φωνητικού κομματιού, ο ανθρώπινος μηχανισμός λεκτικής παραγωγής αντιπροσωπεύεται ως discrete-time σύστημα της μορφής που απεικονίζεται στο σχήμα που ακολουθεί.

Το ακουστικό κύμα παράγεται όταν η ροή αέρος από τους πνεύμονες μεταφέρεται από την τραχεία μέσω των φωνητικών πτυχών. Αυτή η πηγή διέγερσης μπορεί να χαρακτηριστεί ως τη φώνηση, το ψιθύρισμα, το προστριβόμενο σύμφωνο, τη συμπίεση, τη δόνηση ή έναν συνδυασμό αυτών. Η διέγερση Phonated εμφανίζεται όταν διαμορφώνεται η ροή αέρος από τις φωνητικές πτυχές. Η ψιθυριστή διέγερση παράγεται από τη ροή αέρος που ορμά κατευθείαν ένα μικρό τριγωνικό άνοιγμα μεταξύ του αρυταινοειδούς χόνδρου στο πίσω τμήμα των σχεδόν κλειστών φωνητικών πτυχών. Η διέγερση Frication παράγεται από τις συστολές στο φωνητικό κομμάτι. Αποτελέσματα διέγερσης συμπίεσης από την απελευθέρωση ενός εντελώς κλειστού και φωνητικού κομματιού. Η διέγερση δόνησης προκαλείται από τον αέρα που ωθείται μέσω μιας περάτωσης εκτός από τις φωνητικές πτυχές, ειδικά στη γλώσσα. Η ομιλία που παράγεται από τη phonated διέγερση λέγετε εκφρασμένη, όταν παράγεται από τη phonated διέγερση συν το προστριβόμενο σύμφωνο λέγεται μικτή εκφρασμένη, και όταν παράγεται από άλλους τύπους διεγέρσεων λέγεται άναρθρη . Είναι δυνατό να αντιπροσωπευθεί το φωνητικό κομμάτι σε μια παραμετρική μορφή ως λειτουργία μεταφοράς H(z). Προκειμένου να υπολογιστούν οι παράμετροι του H(z) από το λεκτικό κυματοειδές, είναι απαραίτητο να υποτεθεί κάποια μορφή για το H(z). Ιδανικά, η λειτουργία μεταφοράς πρέπει να περιέχει τους πόλους καθώς επίσης και τα μηδενικά. Επιπλέον, εάν μόνο οι εκφρασμένες περιοχές της ομιλίας χρησιμοποιούνται τότεένα πρότυπο πόλων για το H(z) είναι ικανοποιητικό. Επιπλέον, η γραμμική ανάλυση πρόβλεψης μπορεί να χρησιμοποιηθεί για να υπολογίσει αποτελεσματικά τις παραμέτρους ενός προτύπου πόλων. Τέλος, μπορεί επίσης να διαπιστωθεί ότι το πρότυπο πόλων είναι το ελάχιστο μέρος της φάσης του αληθινού προτύπου και έχει φάσματα μεγεθών, το οποίο περιέχει τον όγκο των εξαρτώμενων πληροφοριών του ομιλητή.

Η παραπάνω αναφορά υπογραμμίζει επίσης την εξαρτώμενη φύση των προτύπων φωνητικών κομματιών. Δεδομένου ότι το πρότυπο προέρχεται από την παρατηρηθείσα ομιλία, εξαρτάται από την ομιλία. Το παρακάτω σχήμα επεξηγεί τις διαφορές στα πρότυπα για δύο ομιλητές που λένε το ίδιο φωνήεν.





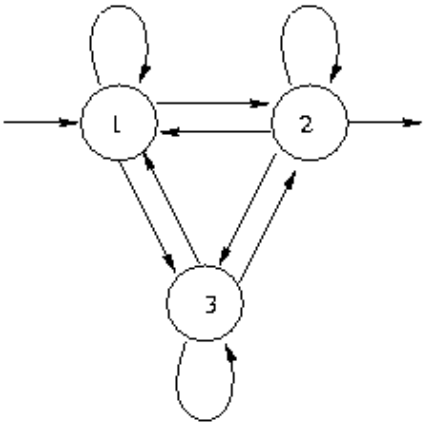
5.3 Επιλογή των χαρακτηριστικών γνωρισμάτων

Τα χαρακτηριστικά γνωρίσματα LPC ήταν πολύ δημοφιλή στα συστήματα αναγνώρισης ομιλίας και επαλήθευσης ομιλητών. Εντούτοις, η σύγκριση δύο διανυσμάτων χαρακτηριστικών γνωρισμάτων LPC απαιτεί τη χρήση των υπολογιστικά ακριβών μέτρων ομοιότητας όπως η απόσταση Itakura - Saito και ως εκ τούτου τα χαρακτηριστικά γνωρίσματα LPC είναι ακατάλληλα για τη χρήση σε πραγματικό χρόνο συστήματα. Ο Furui πρότεινε τη χρήση του φάσματος που ορίστηκε ως ο αντίστροφος μετασχηματισμός Φουριέ του λογαρίθμου του φάσματος μεγέθους, στις εφαρμογές αναγνώρισης ομιλίας. Η χρήση του φάσματος επιτρέπει την ομοιότητα μεταξύ δύο φασματικών διανυσμάτων χαρακτηριστικών γνωρισμάτων που υπολογίζονται ως απλή ευκλείδεια απόσταση. Επιπλέον, Ο Atal έχει καταδείξει ότι το φάσμα που προέρχεται από τα χαρακτηριστικά γνωρίσματα LPC οδηγεί στην καλύτερη απόδοση από την όψη FAR και FRR, για ένα σύστημα επαλήθευσης ομιλητών. Συνεπώς, έχουμε αποφασίσει να χρησιμοποιήσουμε το παραγόμενο από το LPC φάσμα για το σύστημα επαλήθευσης των ομιλητών μας.

5.3.1 Ομιλητής που διαμορφώνει

Χρησιμοποιώντας την φασματική ανάλυση όπως περιγράφεται στο προηγούμενο τμήμα, μια έκφραση μπορεί να αντιπροσωπευθεί ως ακολουθία διανυσμάτων χαρακτηριστικών γνωρισμάτων. Οι εκφράσεις, προφορικές από το ίδιο πρόσωπο αλλά σε διαφορετικούς χρόνους οδηγούν σε παρόμοια ακόμα διαφορετική ακολουθία διανυσμάτων χαρακτηριστικών γνωρισμάτων. Ο σκοπός της διαμόρφωσης φωνής είναι να χτιστεί ένα πρότυπο που συλλαμβάνει αυτές τις παραλλαγές στο αποσπασματικό σύνολο χαρακτηριστικών γνωρισμάτων. Υπάρχουν δύο τύποι προτύπων που έχουν χρησιμοποιηθεί εκτενώς στα συστήματα επαλήθευσης ομιλητών και λεκτικής αναγνώρισης: πιθανολογικά πρότυπα και πρότυπα προτύπων. Το πιθανολογικό πρότυπο μεταχειρίζεται τη διαδικασία λεκτικής παραγωγής ως παραμετρική τυχαία διαδικασία και υποθέτει ότι οι παράμετροι της ελλοχεύουσας πιθανολογικής διαδικασίας μπορούν να υπολογιστούν κατά τρόπο ακριβή, καλά καθορισμένο. Το πρότυπο προτύπων προσπαθεί να διαμορφώσει τη διαδικασία λεκτικής παραγωγής κατά τρόπο μη- παραμετρικό με τη διατήρηση διάφορων ακολουθιών διανυσμάτων, χαρακτηριστικών γνωρισμάτων που προέρχονται από τις πολλαπλάσιες εκφράσεις της ίδιας λέξης από το ίδιο πρόσωπο. Τα μοντέλα προτύπων εξουσίασαν την πρόωρη εργασία στην επαλήθευση ομιλητών και τη λεκτική αναγνώριση επειδή το μοντέλο προτύπων είναι διαισθητικά λογικότερο. Εντούτοις, η πρόσφατη

εργασία στα πιθανολογικά πρότυπα έχει καταδείξει ότι αυτά τα πρότυπα είναι πιο εύκαμπτα και ως εκ τούτου επιτρέπουν καλύτερα να διαμορφώσουν τις διαδικασίες λεκτικής παραγωγής. Ένα πολύ δημοφιλές πιθανολογικό πρότυπο για τη διαμόρφωση της διαδικασίας λεκτικής παραγωγής είναι το Hidden Markov Model (HMM). Τα HMMs είναι επεκτάσεις στα συμβατικά Markov πρότυπα, όπου οι παρατηρήσεις είναι μια πιθανολογική λειτουργία του state, δηλ., το πρότυπο είναι μια διπλά ενσωματωμένη πιθανολογική διαδικασία όπου η ελλοχεύουσα πιθανολογική διαδικασία δεν είναι άμεσα αισθητή (είναι κρυμμένη). Το HMM μπορεί μόνο να αντιμετωπισθεί μέσω ενός άλλου συνόλου πιθανολογικών διαδικασιών που παράγουν την ακολουθία παρατηρήσεων. Κατά συνέπεια, το HMM είναι μια μηχανή πεπερασμένων καταστάσεων .Ένα πλήρως συνδεδεμένο three-state HMM απεικονίζεται παρακάτω.



5.3.2 Ταίριασμα Σχεδιών

Η διαδικασία ταιριάσματος σχεδίων περιλαμβάνει τη σύγκριση ενός δεδομένου συνόλου διανυσμάτων χαρακτηριστικών γνωρισμάτων εισαγωγής ενάντια στο πρότυπο ομιλητών για την απαιτημένη ταυτότητα και τον υπολογισμό ενός ταιριαστού αποτελέσματος. Για τα Hidden Markov πρότυπα που συζητούνται παραπάνω, το ταιριαστό αποτέλεσμα είναι η πιθανότητα ότι ένα δεδομένο σύνολο διανυσμάτων χαρακτηριστικών γνωρισμάτων παρήχθη από το πρότυπο.

5.4 Μέθοδοι Παραμετροποίησης της ομιλίας

Η διαδικασία παραμετροποίησης της ομιλίας έχει σκοπό την περιγραφή των σημαντικών πληροφοριών που εμπεριέχονται στο σήμα ομιλίας με ένα διάνυσμα παραμέτρων το οποίο ονομάζεται παραμετρικό διάνυσμα ομιλίας. Γενικά, αυτές οι παράμετροι πρέπει να παρέχουν επαρκή αναπαράσταση του συνόλου των πληροφοριών

που είναι σχετικές με την ομιλία ενώ ταυτόχρονα να παραμένουν ανεπηρέαστες από διάφορες πηγές ανεπιθύμητης μεταβλητότητας. Για παράδειγμα, τέτοιες είναι οι πηγές παρεμβολών από το περιβάλλον, γραμμικές και μη γραμμικές παραμορφώσεις που εισάγονται από το κανάλι μετάδοσης και το μικρόφωνο.

Ειδικότερα, στις εφαρμογές αναγνώρισης ομιλητή, οι παράμετροι ομιλίας πρέπει να αναπαριστούν τα ιδιαίτερα χαρακτηριστικά της συγκεκριμένης φωνής με επαρκή ακρίβεια. Για να επιτευχθεί αυτό, ένα δεδομένο σύνολο παραμέτρων ομιλίας πρέπει να χαρακτηρίζεται από έναν αριθμό ποιοτικών χαρακτηριστικών, όπως ευαισθησία στην ατομικότητα της ανατομίας του φωνητικού συστήματος του χρήστη (γλωττίδα, φωνητικό κανάλι), ικανότητα να λαμβάνεται υπόψη το ύφος ομιλίας του ατόμου. Οι παράμετροι ομιλίας που αφορούν στο σχήμα του φωνητικού καναλιού είναι λιγότερο επιρρεπείς στην κατάσταση της υγείας του ατόμου, όταν συγκριθούν με τη θεμελιώδη συχνότητα ομιλίας, η οποία σχετίζεται με τις κινήσεις της γλωττίδας. Επιπλέον, όπως απέδειξε ο Doddington (1974), η θεμελιώδης συχνότητα είναι πιο επιρρεπής σε προσπάθειες μίμησης και δεν ενδείκνυται η χρήση της σε εφαρμογές οι οποίες απαιτούν μικρή πιθανότητα μη εξουσιοδοτημένης πρόσβασης, δηλαδή υψηλής ασφάλειας.

Η πρόοδος στην κατανόηση των ιδιοτήτων του ακουστικού συστήματος του ανθρώπου, από το οποίο προέρχονται οι σύγχρονες μέθοδοι παραμετροποίησης του σήματος ομιλίας, οδήγησαν σε μια ποικιλία παραμέτρων ομιλίας.

5.5 Παράμετροι ομιλίας που μοντελοποιούν την μη γραμμική αίσθηση της ακοής

Ακολουθώντας την πρόοδο που έχει επέλθει το δεύτερο μισό του εικοστού αιώνα στο πεδίο της ψυχοακουστικής και συγκεκριμένα στο ακουστικό σύστημα του ανθρώπου [Zwicker (1961), Patterson & Moore (1986), Glasberg & Moore (1990), Moore & Glasberg (1996)], η διαδικασία της παραμετροποίησης του σήματος ομιλίας έχει εμπλουτιστεί με νέες, βιολογικά εμπνευσμένες, μεθόδους παραμετροποίησης του σήματος ομιλίας. Σ' αυτές τις μεθόδους περιλαμβάνεται μια ποικιλία τεχνικών παραμετροποίησης του σήματος ομιλίας οι οποίες προσομοιάζουν από πολλές απόψεις την ακουστική αίσθηση του ανθρώπου.

Βασιζόμενοι στην προσέγγιση του κρίσιμου εύρους ζώνης από τον Zwicker και στην προσέγγιση της μη γραμμικής αίσθησης του ύψους της φωνής από τον Koenig (1949), οι Davis & Mermelstein (1980) πρότειναν τις φημισμένες πλέον σήμερα παραμέτρους MFCC. Στην παρουσίαση των Davis & Mermelstein (1980), αποδεικνυόταν ότι οι παράμετροι MFCC υπερτερούν των παραμέτρων LPC, LPCC καθώς και άλλων παραμέτρων όσον αφορά στην

αναγνώριση ομιλίας. Αποδείχθηκε επίσης [Davis & Mermelstein (1980), Reynolds (1994), Chen et al. (1997)] ότι σε θορυβώδεις συνθήκες περιβάλλοντος διατηρούν την υψηλότερη ευρωστία τους σε σύγκριση με άλλες παραμέτρους όπως οι LPCC, PLP, κλπ.

Οι παράμετροι MFCC, όπως υπολογίζονται στην εργασία των Davis & Mermelstein (1980), περιλαμβάνουν προσεγγίσεις των κρίσιμων ζωνών οι οποίες δεν συνάδουν απόλυτα με τη σημερινή αντίληψη για το θέμα. Ωστόσο, οι παράμετροι MFCC είχαν μεγάλη επίπτωση στη μεθοδολογία παραμετροποίησης του σήματος ομιλίας και κατ' επέκταση στις εφαρμογές αναγνώρισης ομιλίας.

5.5.1 Υπολογισμός των παραμέτρων MFCC

Μετά την εισαγωγή των παραμέτρων MFCC στην εργασία των Davis & Mermelstein (1980), έχει προταθεί ένας μεγάλος αριθμός τροποποιήσεων και βελτιώσεων της αρχικής ιδέας. Αυτές διαφέρουν κυρίως στον αριθμό, το σχήμα, το εύρος ζώνης και τον τρόπο που είναι διατεταγμένα αυτά τα φίλτρα καθώς και στον τρόπο με τον οποίο στρεβλώνεται (warped) το φάσμα ισχύος. Επιπλέον των προαναφερόμενων μεταβλητών, διαφορετικά είναι δυνατόν να είναι επίσης το εύρος συχνοτήτων ενδιαφέροντος καθώς και ο αριθμός των συντελεστών MFCC που χρησιμοποιούνται στην ταξινόμηση.

Ένας από τους κύριους λόγους για μια τέτοια ποικιλία υλοποιήσεων είναι η προσπάθεια για παρακολούθηση της προόδου που έχει συντελεσθεί στο πεδίο της ψυχοακουστικής (psychoacoustics) το τελευταίο χρονικό διάστημα. Για παράδειγμα, υπάρχουν ποικίλες προσεγγίσεις της μη γραμμικής αίσθησης του ύψους της φωνής από το ανθρώπινο σύστημα ακοής. Μια αρχική προσέγγιση, η οποία αναφέρεται ως κλίμακα Koenig (Koenig, 1949), είναι γραμμική κάτω από 1000 Hz και λογαριθμική πάνω από 1000 Hz. Παρέχει μια εύκολη υπολογιστικά αναπαράσταση της κλίμακας mel, η οποία δεν είναι βέβαια πολύ ακριβής και αποκλίνει σημαντικά από την αρχική κλίμακα τόσο για συχνότητες μικρότερες όσο και μεγαλύτερες των 1000 Hz.

5.6 Άλλες τεχνικές παραμετροποίησης

5.6.1 MFCC FB-40

Εξ' αιτίας της ικανοποιητικής απόδοσής τους στην αναγνώριση ομιλίας καθώς και λόγω του ότι αποτελεί την εξ' ορισμού επιλογή παραμετροποίησης για το σύστημα αναγνώρισης ομιλίας Sphinx-III,επιλέχθηκε ως η μία εκ των δύο μεθόδων παραμετροποίησης του σήματος ομιλίας βάσει του μετασχηματισμού Fourier.

5.6.2 PLP(PLP-FB19)

Οι παράμετροι PLP (Hermansky, 1990) βασίζονται σε μια τράπεζα 18 φίλτρων κατανεμημένα σύμφωνα με την κλίμακα Bark για την κάλυψη της περιοχής συχνοτήτων [0, 5000] Hz.

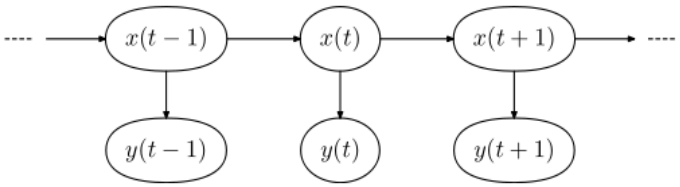
Εδώ, αυτή η τράπεζα φίλτρων προσαρμόστηκε στο επιθυμητό εύρος συχνοτήτων αφαιρώντας το πρώτο φίλτρο (χαμηλότερη συχνότητα) και όλα τα φίλτρα που η κεντρική τους συχνότητα βρίσκεται πέραν της συχνότητας 6855 Hz. Αυτή η τροποποίηση οδήγησε σε μια τράπεζα 19 φίλτρων τα οποία καλύπτουν το εύρος συχνοτήτων [100, 6400] Hz, η οποία είναι η καλύτερη δυνατή υλοποίηση στην επιθυμητή περιοχή συχνοτήτων.

Κεφάλαιο 6 – Αλγόριθμοι αναγνώρισης

6.1 Εισαγωγή στα Κρυφά Μαρκοβιανά Μοντέλα

Τα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models - HMM) είναι ένα πανίσχυρο στατιστικό εργαλείο για την μοντελοποίηση των παραγωγικών ακολουθιών που μπορούν να χαρακτηριστούν από μία ελλοχεύουσα διαδικασία που παράγει μία αισθητή ακολουθία[16]. Τα Κρυφά Μαρκοβιανά Μοντέλα βρίσκουν εφαρμογή σε πολλά πεδία ενδιαφέροντος στην επεξεργασία σήματος και ειδικά στην επεξεργασία ομιλίας. Πιο συγκεκριμένα έχουν επιτυχή εφαρμογή σε χαμηλού επιπέδου διεργασίες της Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing-NLP), όπως στην επισήμανση Μέρος-Του-Λόγου, στην κατάτμηση φράσης και στην εξαγωγή στοχευμένων πληροφοριών από έγγραφα. Η Μαρκοβιανή θεωρία πήρε το όνομά της από τον Andrei Markov στις αρχές του 20ου αιώνα, αλλά η θεωρία των Κρυφών Μαρκοβιανών Μοντέλων στην πραγματικότητα αναπτύχθηκε από τον Baum και τους συνεργάτες του στην δεκαετία του ΄60.[6]

Το παρακάτω διάγραμμα δείχνει τη γενική αρχιτεκτονική ενός στιγμιότυπου HMM. Κάθε οβάλ σχήμα αντιπροσωπεύει μια τυχαία μεταβλητή που μπορεί να υιοθετήσει οποιαδήποτε από μια σειρά αξιών. Η τυχαία μεταβλητή x(t) είναι η κρυμμένη κατάσταση τη χρονική στιγμή t(με x(t) ∈ {x1, x2, x3}). Η τυχαία μεταβλητή y(t) είναι η παρατήρηση τη χρονική στιγμή t(με y(t) ∈ {y1, y2, y3, y4}). Τα βέλη στο διάγραμμα (συχνά ονομάζεται διάγραμμα trellis) δηλώνουν όρους εξάρτησης.



Από το διάγραμμα, είναι σαφές ότι η υπό όρους πιθανοτική κατανομή της κρυφής μεταβλητής x(t) τη χρονική στιγμή t, λαμβάνοντας υπόψη τις τιμές της κρυφής μεταβλητής x ανά πάσα στιγμή, εξαρτάται μόνο από την αξία της κρυφής μεταβλητής x(t-1): οι τιμές τη χρονική στιγμή t2 και πριν δεν έχουν καμία επιρροή.

Αυτό ονομάζεται ιδιότητα Markov . Ομοίως, η αξία της παρατηρούμενης μεταβλητής y(t) εξαρτάται μόνο από την αξία των κρυμμένων μεταβλητών x(t) (τόσο σε

χρόνο t).

Στο βασικό τύπο του Κρυφού Μαρκοβιανού Μοντέλου που εξετάζεται εδώ, ο χώρος καταστάσεων των κρυμμένων μεταβλητών είναι διακριτός, ενώ και ίδιες οι παρατηρήσεις μπορεί να είναι και αυτές διακριτές ή συνεχείς (κατά κανόνα από τη διανομή Gaussian). Οι παράμετροι του Κρυφού Μαρκοβιανού Μοντέλου είναι δύο τύπων, πιθανότητες μετάβασης και πιθανότητες εκπομπής (επίσης γνωστές ως πιθανότητες εξόδου). Οι πιθανότητες μετάβασης ελέγχουν τον τρόπο με τον οποίο η κρυμμένη κατάσταση τη χρονική στιγμή t επιλέγεται δοσμένης της κρυφής κατάστασης τη χρονική στιγμή t-1.

Ο χώρος των κρυφών καταστάσεων θεωρείται ότι αποτελείται από ένα σύνολο από N πιθανές τιμές, κατά το πρότυπο της κατηγορηματικής διανομής. Αυτό σημαίνει ότι για κάθε μία από τις N πιθανές καταστάσεις που μπορεί να πάρει μια κρυφή μεταβλητή τη χρονική στιγμή t, υπάρχει μια πιθανότητα μετάβασης από αυτήν την κατάσταση σε καθεμία από τα N πιθανές καταστάσεις της κρυφής μεταβλητής τη χρονική στιγμή t+1, για ένα σύνολο N^2 πιθανοτήτων μετάβασης. (Σημειώστε, ωστόσο, ότι το σύνολο των πιθανοτήτων μετάβασης για τη μετάβαση από την κάθε δεδομένη κατάσταση πρέπει να έχει άθροισμα 1, που σημαίνει ότι μία πιθανότητα μετάβασης μπορεί να προσδιοριστεί αφού γίνουν γνωστά τα άλλα, αφήνοντας ένα σύνολο από N (N-1) παραμέτρων μετάβασης).

Επιπλέον, για κάθε ένα από τα N πιθανές καταστάσεις, υπάρχει μια σειρά από πιθανότητες εκπομπής που διέπουν τη διανομή της παρατηρούμενης μεταβλητής σε μια συγκεκριμένη χρονική στιγμή εξαιτίας της κατάστασης των κρυμμένων μεταβλητών εκείνη τη στιγμή. Το μέγεθος αυτού του συνόλου εξαρτάται από τη φύση των παρατηρούμενων μεταβλητών.

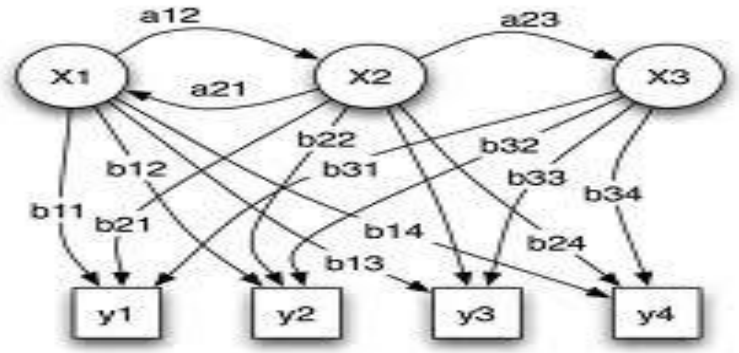
6.2 Hidden Markov Models

Ένα Hidden Markov Model (HMM) συνιστά έναν τύπο στοχαστικής μοντελοποίησης, κατάλληλο για μη στάσιμες στοχαστικές ακολουθίες, των οποίων οι στατιστικές ιδιότητες διέπονται από τυχαίες μεταβάσεις μεταξύ k διαφορετικών στάσιμων διεργασιών. Με άλλα λόγια τα HMM χρησιμοποιούνται για την μοντελοποίηση διεργασιών που είναι κατά τμήματα στάσιμες. Μία διεργασία θεωρείται στάσιμη όταν οι στατιστικές της ιδιότητες δεν μεταβάλλονται καθώς εξελίσσεται ο χρόνος.

6.3 Hidden Markov Models – Αναγνώριση και Εκπαίδευση.

Στο στάδιο της αναγνώρισης, υποθέτουμε ότι έχουμε στην διάθεση μας περισσότερα του ενός HMM, κάθε ένα εκ των οποίων περιγράφεται από ένα διαφορετικό σύνολο παραμέτρων. Προφανώς, κάθε HMM μοντελοποιεί μία διαφορετική, στάσιμη κατά τμήματα διεργασία. Για παράδειγμα, ένα HMM μπορεί να μοντελοποιεί μία διεργασία δύο πηγών εκπομπής παρατηρήσεων.

Κατά την φάση της αναγνώρισης, ο στόχος είναι ο ακόλουθος: με δεδομένη μία ακολουθία παρατηρήσεων και ένα πλήθος M από HMM (κάθε ένα εκ των οποίων μοντελοποιεί διαφορετική διεργασία), να αποφασιστεί πιο από τα HMM είναι περισσότερο πιθανό να εκπέμψει την συγκεκριμένη ακολουθία παρατηρήσεων. Δύο μέθοδοι που δίνουν λύση σε αυτό το πρόβλημα είναι η μέθοδος Baum-Welch (γνωστή και ως μέθοδος οποιασδήποτε διαδρομής – any path method) και η μέθοδος Viterbi (ή μέθοδος καλύτερης διαδρομής – best path method). Οι δύο αυτές μέθοδοι υπολογίζουν, για κάθε HMM, ένα αποτέλεσμα (score) που βασίζετε σε πιθανότητες. Το HMM που γεννά το μέγιστο σκορ θεωρείτε ως το πιο πιθανό να έχει εκπέμψει τη συγκεκριμένη ακολουθία παρατηρήσεων. Η φάση αναγνώρισης προϋποθέτει ότι όλες οι παράμετροι που προσδιορίζουν τα HMM έχουν προηγουμένως εκτιμηθεί και είναι επομένως γνωστές.Στην φάση της εκπαίδευσης γίνετε εκτίμηση των παραμέτρων του εκάστοτε HMM. Προς την κατεύθυνση αυτή, χρησιμοποιείται μία ακολουθία παρατηρήσεων ικανού μήκους (ή και περισσότερες), που έχει γεννηθεί από την αντίστοιχη στοχαστική διεργασία, προκειμένου να γίνει εκτίμηση των άγνωστων παραμέτρων (πχ. Χρησιμοποιώντας τεχνικές που βασίζονται στην λογική της μέγιστης πιθανοφάνειας για την εκτίμηση των παραμέτρων). Δύο δημοφιλείς επαναληπτικές τεχνικές εκπαίδευσης είναι η μέθοδος Baum-Welch και η μέθοδος Viterbi.



6.4 Νευρωνικά δίκτυα.

Οι μέθοδοι νευρωνικών δικτύων κυρίως αναφέρονται στην ταξινόμηση μοντέλων μέσα από ένα σύνολο δεδομένων. Το αδύνατο σημείο των νευρωνικών δικτύων είναι η μη συμμετοχή της παραμέτρου του χρόνου. Εναλλακτικά, ο χρόνος προσεγγίζεται ως εξωτερικός μηχανισμός. Τα Νευρωνικά Δίκτυα Χρονικής Καθυστέρησης είναι ένα από τα πιο αποτελεσματικά μοντέλα. Κάνουν χρήση πολυεπίπεδων νευρωνικών δικτύων. Τα Νευρωνικά Δίκτυα Χρονικής Καθυστέρησης μετατρέπουν το χρονικό πρόβλημα σε χωρικό. Τα συστήματα που βασίζονται σε τέτοιου είδους δίκτυα παρουσιάζουν υψηλή πολυπλοκότητα και απαιτούν μεγάλη επεξεργασία δεδομένων. Ακόμη και στις περιπτώσεις όπου ο παράγοντας του χρόνου έχει εισαχθεί ως μηχανισμός στο σύστημα, το υπολογιστικό κόστος, η πολυπλοκότητα της εκπαίδευσης αλλά και η δυσκολία ερμηνείας των αποτελεσμάτων είναι τα συχνότερα προβλήματα για την χρήσης τους.

Κεφάλαιο 7 – Αναγνώριση Ομιλίας

7.1 Συστήματα Αναγνώρισης Ομιλίας

Αναγνώριση ομιλίας καλείται το σύνολο των διαδικασιών που διαδοχικά εφαρμόζονται στο ακουστικό σήμα και έχουν ως αποτέλεσμα την παραγωγή ακολουθίας διακριτών συμβολών, η οποία περιγράφει το φωνητικό περιεχόμενο του ακουστικού σήματος. Η πλήρης περιγραφή του ακουστικού σήματος με λέξεις και προτάσεις που υπακούουν στους γραμματικούς, συντακτικούς και εννοιολογικούς κανόνες, είναι αποτέλεσμα επιπρόσθετης επεξεργασίας των παραγόμενων φωνητικών συμβολών. Το πρόβλημα που γενικά αντιμετωπίζεται στα συστήματα αναγνώρισης ομιλίας είναι:

Έχοντας τις παρατηρήσεις (διακριτές ή συνεχείς) μιας φωνητικής διαδικασίας, να δημιουργηθεί το μοντέλο του σήματος όμιλος το οποίο να εξηγεί και να χαρακτηρίζει το σήμα έτσι ώστε να υπάρχει η δυνατότητα με χρήση του μοντέλου αυτού να προσδιοριστεί η ταυτότητα άγνωστων ακολουθιών παρομοίων παρατηρήσεων.’’

Γενικά ένα σύστημα αναγνώρισης ομιλίας αποτελείται από την βαθμίδα δημιουργίας των διανυσμάτων των παρατηρήσεων (βαθμίδα εξαγωγής παραμέτρων), την βαθμίδα καθορισμού του μοντέλου που περιγράφει αυτές τις παραμέτρους (βαθμίδα εκπαίδευσης ή βαθμίδα δημιουργίας της μνήμης του συστήματος)και τέλος την βαθμίδα αναγνωρίσεις που προσδιορίζει την ταυτότητα ακολουθιών άγνωστων παρατηρήσεων χρησιμοποιώντας την μνήμη του συστήματος.[12]

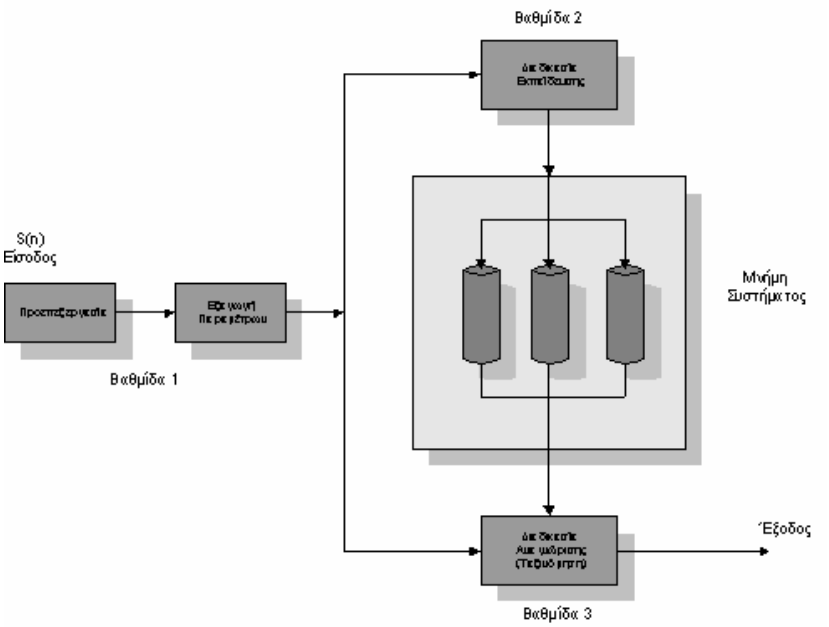
Τα συστήματα αναγνώρισης ομιλίας ταξινομούνται σε κατηγόριες ανάλογα με τις τεχνικές που υλοποιούνται στις επιμέρους βαθμίδες τους Η ταξινόμηση αυτή γίνεται με βάση:

τη μέθοδο που χρησιμοποιείται στη βαθμίδα εξαγωγής παραμέτρων.

- τη βασική μονάδα αναγνώρισης / κωδικοποίησης που χρησιμοποιείται από το σύστημα.
- το τρόπο ταξινόμησης των άγνωστων πρότυπων (βαθμίδα αναγνώρισης).
- το τρόπο δημιουργίας της μνήμης του συστήματος (βαθμίδα εκπαίδευσης).

Μία μεγάλη ποικιλία τεχνικών και μεθόδων χρησιμοποιούνται για την αναγνώριση ομιλίας. Συνήθως ξεκινάμε με την ψηφιοποίηση και δειγματοληψία της ομιλίας. Ακολουθεί η επεξεργασία του, όπου περιλαμβάνονται: φασματική ανάλυση, LPC ανάλυση, μετασχηματι-σμός cepstral, μετασχηματισμός Fourier κλπ. Το επόμενο στάδιο είναι η αναγνώριση φωνημάτων, ομάδων φωνημάτων ή λέξεων. Αυτό μπορεί να επιτευχθεί με κάποιο από τους εξής τρόπους: DTW (Dynamic Time Warping), HMM

(Hidden Markov modelling), NNs (Νευρωνικά Δίκτυα). Ο σκοπός τους είναι η αναγνώριση προτύπων ομιλίας. Τα ανώτερα συστήματα χρησιμοποιούν και στατιστικές μεθόδους για το σκοπό αυτό, όπως και γραμματικές γνώσεις. Άλλα, δουλεύουν με χρήση παραμέτρων της ανθρώπινης ομιλίας (pitch, έμφαση, ένταση, ρυθμός κ.τ.λ.) για την επεξεργασία του σήματος ομιλίας. Μερικά άλλα προσπαθούν να «καταλάβουν» την ομιλία, δηλ. προσπαθούν να μετατρέψουν τις λέξεις σε μια αναπαράσταση αυτών που ο ομιλητής εννοεί.[7]



7.2 Κατηγορίες συστημάτων

Τα συστήματα ομιλίας μπορούν να κατηγοριοποιηθούν με διάφορους τρόπους και βάση διαφορετικών παραμέτρων. Ας δούμε πώς:

α) Είδος ομιλίας

Ανάλογα με το είδος ομιλίας διακρίνονται τρεις κατηγόριες συστημάτων:

✓ Συστήματα αναγνώρισης διακριτής ομιλίας (μεμονωμένα προφερόμενων λέξεων).Χρονικά είναι τα πρώτα συστήματα αναγνώρισης ομιλίας που εμφανίσθηκαν και παρουσιάζουν την υψηλότερη αξιοπιστία. Ο ομιλητής είναι υποχρεωμένος να προφέρει την κάθε λέξη μεμονωμένα.

✓ Συστήματα αναγνώρισης διασυνδεδεμένων λέξεων. Αποτελούν εξέλιξη των συστημάτων αναγνώρισης διακριτής ομιλίας και αποτελούν την πρώτη προσπάθεια αναγνώρισης προτάσεων στις οποίες δεν χωρίζονται οι λέξεις με διαστήματα σιγής. Παρουσιάζουν ικανοποιητική αξιοπιστία, ιδιαίτερα μετά την εμφάνιση ικανών αλγόριθμων

Δυναμικού Προγραμματισμού. Χρονικά είναι τα πρώτα συστήματα αναγνώρισης ομιλίας προφέρει ακολουθία λέξεων, αλλά τις περισσότερες φορές περιορίζεται από τον αριθμό και το είδος των λέξεων που μπορεί να χρησιμοποιήσει (ψηφία, γράμματα αλφάβητου κα).

✓ Συστήματα αναγνώρισης συνεχούς ομιλίας. Χρονικά είναι τα τελευταία συστήματα αναγνώρισης ομιλίας. Ο ομιλητής αποδεσμεύεται από κάθε είδους περιορισμό. Η πλήρης αποδέσμευση του ομιλητή δημιουργεί πολλά προβλήματα στην διαδικασία αναγνώρισης, όπως μεγαλύτερη επίδραση του φαινομένου συνάρθρωσης, δυσκολία στον εντοπισμό των ορίων των λέξεων στη συνεχή φωνητική ακολουθία, κ.α. Η αξιοπιστία των συστημάτων αναγνώρισης συνεχούς ομιλίας παραμένει σε χαμηλά επίπεδα σε σύγκριση με την αξιοπιστία των συστημάτων αναγνώρισης διακριτής ομιλίας ή διασυνδεδεμένων λέξεων.

β) Μοντέλο ομιλητή

Η παράμετρος αυτή προσδιορίζει τον χρήστη του συστήματος αναγνώρισης ομιλίας.Σύστημα ανεξάρτητο από το χρήστη σημαίνει ότι μπορεί να αναγνωρίσει πρότυπα ομιλίας από έναν μεγάλο αριθμό χρηστών

Σύστημα εξαρτώμενο από το χρήστη είναι αυτό που απαιτεί χωριστή εκπαίδευση για κάθε χρήστη του. Τα συστήματα αυτά είναι συνήθως πολύ πιο ακριβή, αλλά η επαναλαμβανόμενη διαδικασία εκπαίδευσης μπορεί να δημιουργεί προβλήματα. Διακρίνεται σε σύστημα κάθε χρήστη ή σύστημα κατηγορίας χρήστη (άνδρα ή γυναίκα, άγγλο ή αμερικάνο)

Το σύστημα προσαρμοζόμενο στο χρήστη, είναι μια προσπάθεια συμβιβασμού των δύο παραπάνω. Είναι κατασκευασμένο με τέτοιο τρόπο, ώστε να προσαρμόζει την λειτουργία του σε κάθε νέο χρήστη που συναντά, βάση γενικών μοντέλων περιγραφής των χαρακτηριστικών του χρήστη.

γ) Μέγεθος λεξιλογίου

Το μέγεθος του λεξιλογίου επηρεάζει την πολυπλοκότητα, τις υπολογιστικές απαιτήσεις και την ακρίβεια του συστήματος.

Διακρίνουμε συστήματα:

- μικρού λεξιλογίου με μερικές λέξεις ως μερικές δεκάδες λέξεις (μέχρι 100 λέξεις).
- μεσαίου λεξιλογίου (από 100 μέχρι 1000 λέξεις).
- μεγάλου λεξιλογίου (περισσότερες από 1000 λέξεις).

δ) Μονάδα αναγνώρισης

Ανάλογα με την μονάδα αναγνώρισης διακρίνουμε συστήματα:

- αναγνώρισης λέξεων.

- αναγνώρισης τμημάτων λέξεων.
- αναγνώρισης φωνημάτων.

ε) Τεχνική αναγνώρισης

Τέλος, ανάλογα με την τεχνική αναγνώρισης διακρίνονται τέσσερις κατηγόριες:

I. συστήματα σύγκρισης πρότυπων (Template Matching), στα οποία η ταξινομήσει γίνεται με άμεση σύγκριση των χαρακτηριστικών των φωνητικών μονάδων.
II. πιθανοτικά συστήματα αναγνώρισης, που χρησιμοποιούν πιθανοτικά μοντέλα για την περιγραφή της ακουστικής διεργασίας (κρυμμένα μοντέλα Markov - HMM).
III. συντακτικά συστήματα, που ταυτοποιούν πρότυπα αποτελούμενα από μικρότερα αρχέγονα πρότυπα.
IV. συστήματα δικτύων που αποτελούνται από ένα σύνολο διασυνδεδεμένων κόμβων σε διαφορετικά επίπεδα, όπως τα νευρωνικά δίκτυα.

7.3 Βασική φωνητική μονάδα αναγνώρισης

Η επιλογή των φωνητικών μονάδων για τις οποίες η βαθμίδα εκμάθησης δημιουργεί πρότυπα διανύσματα αναφοράς και στις οποίες η βαθμίδα αναγνώρισης ταυτοποιεί πρωταρχικά το πρότυπο διάνυσμα του αγνώστου ακουστικού σήματος, καθορίζεται από την ζητούμενη απο το σύστημα αξιοπιστία αναγνώρισης , τους περιορισμούς στο μήκος μνήμης και από το μέγεθος του λεξιλογίου. Οι κυριότερες φωνητικές μονάδες που έχουν χρησιμοποιηθεί από συστήματα αναγνώρισης ανάλυσης ή σύνθεσης ομιλίας είναι οι ακόλουθες:

- Φωνήματα: Αποτελούν την μικρότερη σε χρονική διάρκεια φωνητική μονάδα αναγνώρισης. Χρησιμοποιούνται κύρια σε συστήματα αναγνώρισης μεγάλου λεξιλογίου. Παρουσιάζουν μικρή αξιοπιστία αναγνλωρισης σε σχέση με συστήματα που χρησιμοποιούν πρότυπα αναφοράς μεγαλύτερης χρονικής διάρκειας, αλλά δεν απαιτούν την χρήση χρονοβόρων αλγορίθμων χρονικής ευθυγράμμισης. Τέλος αναφέρεται ότι το σύστημα είναι ανεξάρτητο του μεγέθους του λεξικού.
- Συλλαβές: Είναι περισσότερο αξιόπιστες φωνητικές μονάδες από τα φωνήματα, επειδή περιγράφουν και τα μεταβατικά φαινόμενα μεταξύ των φθόγγων. Ο αριθμός τους είναι σχετικά μεγάλος .
- Λέξεις: Τα πρότυπα αναφοράς περιγράφουν τη χρονική μεταβολή των παραμέτρων του μοντέλου ομιλίας κατά ην διάρκεια προφοράς των λέξεων. Χρησιμοποιούνται σε συστήματα αναγνώρισης μεμονωμένα προφερόμενων λέξεων (διακριτής ομιλίας)

μικρού έως μέσου λεξιλογίου, παρουσιάζοντας και την υψηλότερη αξιοπιστία αναγνώρισης. Απαιτείται η χρήση χρονοβόρων διαδικασιών χρονικής ευθυγράμμισης για να είναι δυνατή η σύγκριση και η ταξινόμηση του συνόλου των προτύπων διανυσμάτων της άγνωστης λέξης.

7.4 Βαθμίδα αναγνώρισης

Στις παραγράφους που ακολουθούν επιχειρείται μια ομαδοποίηση των τεχνικών που χρησιμοποιούνται στην βαθμίδα αναγνώρισης, ανεξάρτητα από την φωνητική μονάδα που χρησιμοποιείται. Οι τεχνικές που χρησιμοποιούνται στην βαθμίδα αναγνώρισης διακρίνονται ανάλογα με το μοντέλο σύγκρισης που χρησιμοποιείται για να ταξινομηθούν τα άγνωστα πρότυπα διανύσματα σε:

(α) Μεθόδους Διανυσματικής Σύγκρισης (Template Matching)

(β) Πιθανοτικές ή Στοχαστικές (Bayes Classifies, Hidden Markov Models - HMM)

(γ) Συντακτικές (Syntactic Pattern Recognition)

(δ) Νευρωνικών Δικτύων (Neural Networks)

7.5 Αναγνώριση Ομιλίας Διανυσματικής Σύγκρισης (Template Matching)

Τα συστήματα διανυσματικής σύγκρισης είναι τα πλέον διαδεδομένα συστήματα ταξινόμησης προτύπων. Η ταυτοποίηση πραγματοποιείται με την αναζήτηση της κατηγορίας, τα χαρακτηριστικά της οποίας ταιριάζουν περισσότερο με τα χαρακτηριστικά του αγνώστου προτύπου διανύσματος.

Το μέτρο διαφοροποίησης των προτύπων υπολογίζεται ποσοτικά με την συνάρτηση απόστασης ή ομοιότητας που επιλέγεται έτσι ώστε να αποδίδει μετην μεγαλύτερη κατα το δυνατό αξιοπιστία τις διαφορές του φωνητικού περιεχομένου της ομιλίας.

7.6 Πιθανοτικά ή Στοχαστικά Συστήματα Αναγνώρισης Ομιλίας

Στην δεκαετία του ’70 γνώρισαν μεγάλη διάδοση και ανάπτυξη συστήματα αναγνώρισης ομιλίας που χρησιμοποίησαν πιθανοτικά μοντέλα στην βαθμίδα αναγνώρισης. Τα πλεονεκτήματα που οδήγησαν στη χρήση πιθανοτικών μοντέλων στην βαθμίδα αναγνώρισης είναι η βελτίωση της αξιοπιστίας των συστημάτων και η δυνατότητα που παρέχεται για ταυτόχρονη ενσωμάτωση πληροφορίας του τρόπου σύνταξης των προτάσεων ή για την δομή των προτάσεων που χρησιμοποιούνται σε

συγκεκριμένες εφαρμογές. Στην δεκαετία του ’80 υπήξε μια αλματώδης διάδοση των πιθανότικών συστημάτων λόγω της απλότητας και της ακρίβειας με την οποία μπορεί να μοντελοποιηθεί ο μηχανοσμός παραγωγής ομιλίας σαν μια διαδικασία εκτίμησης γεγονότων ενός κρυμμενου μοντέλου Markov. Ο Jelinek το 1975 έθεσε τις θεμελιώδεις αρχές εφαρμογής των πιθανοτικών μοντέλων σε συστήματα αναγνώρισης ομιλίας και διέκρινε τις τρεις βαθμίδες από τις οποίες αποτελούνται.

Τον γλωσσικό αποκωδικοποιητή, ο οποίος υπολογίζει την δεσμευμένη πιθανότητα να εμφανιστεί η πρόταση W όταν παρατηρηθεί η ακολυθία ακουστικών συμβόλων A. Το πλέον αξιόπιστο πιθανοτικό μοντέλο του γλωσσικού αποκωδικοποιητή είναι κρυμμένο μοντέλο Markov (HMM). Το μοντέλο θεωρεί ότι τα πρότυπα διανύσματα προέρχονται από μια διαδικασία Markov μετάβασης σε στοιχειώδεις φωνητικές μονάδες (φωνήματα, συλλαβές, λέξεις) που απαρτίζουν τη λέξη ή την πρόταση, οι οποίες δεν μπορούν να παρατηρηθούν άμεσα και γι’ αυτό το λόγο το μοντέλο ονομάζεται κρυμμένο μοντέλο Markov.

7.7 Συστήματα Συντακτικής Αναγνώρισης

Τα συστήματα συντακτικής αναγνώρισης προτύπων δημιουργήθηκαν και αναπτύχθηκαν σχετικά πρόσφατα και χρησιμοποιούνται κύρια για να ταυτοποιούν πρότυπα τα οποία περιγράφονται από μεγάλων διαστάσεων διανύσματα (όπως κινούμενα αντικείμενα, εικόνα, ομιλία) ή πρότυπα των οποίων τα χαρακτηριστικά δεν είναι η παραμετρική τους περιγραφή, αλλά ο τρόπος δόμησής τους από μικρότερα πρότυπα τα οποία ονομάζονται αρχέγονα. Χαρακτηριστικά παραδείγματα τέτοιων προτύπων είναι τα μεγάλα δακτυλικα αποτυπώματα, τα χαρακτηριστικά προσώπων, τα μέλη σώματος σε ακτινογραφίες Χ, τα χρωμοσώματα, η συνεχής ομιλία.

Τα συστήματα συντακτικής αναγνώρισης δεν επεμβαίνουν στον τρόπο με τον οποίο αναγνωρίζονται τα αρχέγονα πρότυπα. Η συντακτική αναγνώριση αποτελεί μια διαδικασία αναζήτησης των δομικών χαρακτηριστικών των αρχέγονων προτύπων της κατηγορίας που μοιάζει περισσότερο στην αντίστοιχη δομή αγνώστου προτύπου.

Η μνήμη του συστήματος περιέχει ένα σύνολο κανόνων που περιγράφουν την αναμενόμενη δομή των κατηγοριών. Το σύνολο των κανόνων ονομάζεται γραμματική της κατηγορίας.Η δομή των αρχέγονων προτύπων μπορεί να περιγραφεί από γραμματικές δέντρου, περιορισμένης σύνταξης γραμματικές, πλέγματος, διαγραμμάτων. Αν η γραμματική των κατηγοριών περιέχει και πληροφορίες για την πιθανοτική συμπεριφορά των κανόνων τότε η γραμματική ονομάζεται πιθανοτική ή στοχαστική, διαφορετικά

ονομάζεται δομική.

Συντακτική ανάλυση (parsing) ονομάζεται η διαδικάσια με την οποία καθορίζεται αν το άγνωστο πρότυπο μπορεί να αναπαραχθεί από την γραμματική κάποιας κατηγορίας. Οι κυριότερες εφαρμογές των συντακτικών συστημάτων έχουν γίνει σε συστήματα αναγνώρισης συνεχούς ομιλίας όπως είναι το HAPPY, DRAGON, το σύστημα των Bahl, Jelinek και Mercer, το LITHAN.

7.8 Συστήματα Αναγνώρισης Νευρονικών Δικτύων

Τα συστήματα αναγνώρισης δικτύων αποτελούνται από ένα σύνολο διασυνδεδεμένων κόμβων οι οποίοι βρίσκονται σε διαφορετικά επίπεδα. Ανάλογα με το επίπεδο στο οποίο ανήκουν οι κόμβοι διακρίνονται σε τρεις κατηγορίες:

- Στους κόμβους εισόδου στους οποίους εφαρμόζεται το άγνωστο πρότυπο διάνυσμα,
- Στους κρυφούς κόμβους των ενδιάμεσων επιπέδων, στους οποίους πραγματοποιείται η κύρια επεξεργασία χρησιμοποιώντας την μνήμη του συστήματος, και
- Στους κόμβους εξόδου οι οποίοι αναδεικνύουν την φωνητική κατηγορία που ανήκει το άγνωστο πρότυπο διάνυσμα που εφαρμόζεται στην είσοδο.

Το χαρακτηριστικό των συστημάτων αναγνώρισης δικτύων είναι ότι η πληροφορία διαχωρισμού των διαφορετικών φωνηκών κατηγοριών είναι ενσωματωμένη στον τρόπο σύνδεσής τους και στον τρόπο λειτουργίας των κόμβων. Όταν η επεξεργασία στους κόμβους περιέχει μη γραμμικούς μετασχηματισμούς, τότε το δίκτυο ονομάζεται νευρωνικό, διότι ο τρόπος λειτουργίας του δικτύου μοιάζει με τον τρόπο διάδοσης και επεξεργασίας της πληροφορίας στο νευρικό σύστημα των ζώων. Τα νευρωνικά δίκτυα είναι από τα πλέον διαδεδομένα δίκτυα επεηεργασίας και αναγνώρισης σημάτων σε επερυνητικό επίπεδο αφού παρουσιάζουν υψηλή αξιοπιστία αναγνώρισης και υπάρχει δυνατότητα υλοποίησής τους σε παράλληλης οργάνωσης επεξεργαστές.

Τα πρώτα μοντέλα νευρωνικών δικτύων εμφανίστηκαν την δεκαετία του 1940. Οι πρώτες μαθηματικές προσεγγίσεις των McCulloch και Hebb, Rosenblatt, Windrow, Posch δίνουν τα πρώτα μοντέλα τα οποία έχουν περιορισμένες δυνατότητες. Νεώτερες εργασίες των Hopfield (Hopfield et al, 1986), Rumelhart και McClelland, Sejnowski, Feldman (Biermann et al, 1970), Glossbery (Carpenter et al, 1986) που κύρια βασίζονται σε επιτυχείς μοντελοποιήσεις της λειτουργίας των νευρικών κυττάρων δημιούργησαν ένα πλήθος δικτύων που αποδίδουν υψηλή αξιοπιστία αναγνώρισης φθόγγων σε συστήματα αναγνώρισης ομιλίας.[13]

7.8.1 Διαδικασία εκπαίδευσης

Η διαδικασία εκπαίδευσης είναι το σύνολο των επεξεργασιών με τις οποίες ένα σύστημα αναγνώρισης ομιλίας χρησιμοποιώντας ένα καθορισμένο σύνολο επαναλήψεων των φωνητικών κατηγοριών του λεξιλογίου (που ονομάζεται σύνολο εκπαίδευσης) δημιουργεί τα πρότυπα διανύσματα αναφοράς. Το σύνολο των προτύπων αναφοράς αποτελέι τη μνήμη του συστήματος αναγνώρισης. Η επιλογή του αλγορίθμου της διαδικασίας εκμάθησης γίνεται με βάση την μέθοδο ταυτοποίησης που χρησιμοποιείται στην βαθμίδα αναγνώρισης. Βασική αρχή των αλγορίθμων εκμάθησης είαι ο προσδιορισμός της μνήμης του συστήματος με ταυτόχρονη ελαχιστοποίηση του λάθους αναγνώρισης ή της ενδοδιασποράς των αποστάσεων των προτύπων διανυσμάτων αναφοράς από το σύνολο εκμάθησης.

7.9 Πιθανοτικά Συστήματα Αναγνώρισης Ομιλίας

Ακολουθώντας την διάκριση των πιθανοτικών μοντέλων όπως ορίστηκε κατά την περιγραφή της δομής των πιθανοτικών συστημάτων ανανώρισης ομιλίας, μπορούμε αντίστοιχα να διακρίνουμε τρία τμήματα μνήμης:

- Την μνήμη του ακουστικού επεξεργαστή: Δημιουργείται με τους αλγορίθμους που υλοποιούν την διαδικασία εκπαίδευσης των δομικών συστημάτων.
- Την μνήμη του γλωσσικού αποκωδικοποιητή: Η επιλογή του αριθμού των κρυφών καταστάσεων του μοντέλου και του αριθμού των παρατηρήσεων εξόδου του κρυμμένου μοντέλου Markov καθορίζει την περιγραφική ακρίβεια των διαδικάσιών του, την τάξη και το πλήθος των επαναλήψεων που απαιτούνται για έναν ικανοποιητικό προσδιορισμό των παραμέτρων του. Έχουν προταθεί πολλά κριτήρια με τα οποία μπορεί να δημιουργηθεί η μνήμη του γλωσσικού αποκωδικοποιητή.
- Μνήμη γλωσσικού μοντέλου: Για τα συστήματα που χρησιμοποιούν πιθανοτικά μοντέλα λέξεων απαιτείται ένα ιδιαίτερα μεγάλο μέγεθος μνήμης αποθήκευσης των δεσμευμένων πιθανοτήτων του μοντέλου, ιδιαίτερα στα συστήματα αναγνώρισης ομιλίας μεγάλου λεξιλογίου. Μεγάλη είναι η απαίτηση για το κείμενο εκμαθησης, το οποίο πολλές φορές για μια ικανοποιητική προσέγγιση των παραμέτρων του μοντέλου φτάνει στην τάξη των 106 λέξεων (Lee, 1988).

Πρέπει να σημειωθεί ότι η αξιοπιστία του γλωσσικού μοντέλου εξαρτάται σημαντικά και από την εφαρμογή στην οποία θα χρησιμοποιηθεί. Συστήματα αναγνώρισης ομιλίας φυσικής γλώσσας απαιτούν εκμάθηση με κείμενα διαφορετικού περιεχομένου και

μεγάλου μεγέθους. Χαρακτηριστικό παράδειγμα αποτελεί το σύστημα Tangora (Lee, 1988) το οποίο είναι μια γραφομηχανή κατευθυνόμενη από ομιλία, που χρησιμοποιεί κατά την διαδικασία προσδιορισμού των παραμέτρων του γλωσσικού μοντέλου κείμενο 250 εκατομμυρίων λέξεων.

7.10 Συντακτικά Συστήματα Αναγνώρισης Ομιλίας

Η μνήμη των συντακτικών συστημάτων αναγνώρισης ομιλίας μπορεί να διαιρεθεί σε δύο τμήματα:

- Την μνήμη των αρχέγονων προτύπων: Ο αριθμός και η παραμετρική τους περιγραφή προσδιορίζονται συνήθως εμπειρικά. Ειδικοί αλγόριθμοι όπως η ευρετική μέθοδος, ο αλυσωτός κώδικας, η μέθοδος των πολυγώνων χρησιμοποιούνται στα μεγάλων διαστάσεων πρότυπα διανύσματα για να προσδιορίσουν τα αρχέγονα πρότυπα (Pavlidis, 1976). Μεγαλύτερης κλίμακας εφαρμογή βρίσκουν οι αλγόριθμοι δημιουργίας και μνήμης των δομικών συστημάτων όπως ο k-means, isodata για τον προσδιορισμό της παραμετρικής περιγραφής των αρχέγονων προτύπων, όταν αυτά παρουσιάζουν υψηλή διασπορά.
- Την μνήμη της γραμματικής σύνταξης των λέξεων: Από τους αλγορίθμους δημιουργίας δομικών γραμματικών μπορούμε να αναφέρουμε τις τενικές K-tails, την στρατηγική που προτάθηκε από τον Solomonoff, τον αλγόριθμο των Crespi και Reghizzi για τις γραμματικές κανόνων του Gips. Για τις γραμματικές δέντρου χρησιμοποιούνται οι ίδιοι αλγόριθμοι εκμάθησης μιας και μπορούν να αναπαραχθούν από τις γραμματικές κανόνων.

7.11 Συστήματα Αναγνώρισης Δικτύων

Αν και τα συστήματα αναγνώρισης νευρωνικών δικτύων είναι γνωστά από τη δεκαετία του ’40, εντούτοις η έλλειψη ικανοποιητικών αλγορίθμων εκμάθησης καθυστέρησε την υλοποίησή τους σε εφαρμογές. Η δημιοργία αλγορίθμων που χρησιμοποιούν σαν κριτήριο προσδιορισμού της μνήμης του δικτύου την ελαχιστοποίηση του λάθους πρόβλεψης της εξόδου των, επέτρεψαν την εφαρμογή των μοντέλων νευρωνικών δικτύων και στα συστήματα αναγνώρισης ομιλίας.

Ο αλγόριθμος προσδιορισμού των συντελεστών βαρύτητας των κόμβων βασίζεται στην αναζήτηση της περιοχής των περσσοτέρων ενεργοποιήσεων με την τοποθέτηση του προτύπου εκμάθησης στην είσοδο του δικτύου. Ο αναδρομικός αλγόριθμος που

επαναπροσδιορίζει τους συντελεστές βαρύτητας των κόμβων που ανήκουν στην περιοχή υψηλών ενεργοποιήσεων αποτελέι μια απλή προσομοίωση των μηχανισμών μάθησης του ανώτερου εγκεφαλικού φλοιού.

Σε όλους τους αλγορίθμους εκπαίδευσης η επιλογή των αρχικών συντελεστών του δικτύου δεν φαίνεται να παίζει σημαντικό ρόλο στην αξιοπιστία του συστήματος αναγνώρισης αλλά επιρεάζει σημαντικά την ταύτητα σύγκλισης του αλγορίθμου εκμάθησης.

7.11.1 Διαδικασία αναγνώρισης συστημάτων διανυσματικής σύγκρισης

Η διαδικασία αναγνώρισης των δομικών συστημάτων αναγνώρισης ομιλίας είναι απλή και στηρίζεται στην ταξινόμηση των αγνώστων προτύπων διανυσμάτων σε ένα από τα πρότυπα διανύσματα αναφοράς που αοτελούν το αλφάβητο του συστήματος. Η ταξινόμηση γίνεται με χρήση του κανόνα των K-κοντινότερων γειτόνων ή του κανόνα του κοντινότερου γείτονα που αποτελεί ειδική περίπτωση του προηγούμενου κανόνα (Dermatas, 1991).[18]

Τα ακόλουθα βήματα περιγ΄ραφουν μια ολοκληρωμένη διαδικασία αναγνώρισης, ενός δομικού συστήματος που χρησιμοποιεί το φώνημα σαν βασική φωνητική μονάδα:

ΒΗΜΑ 1: Παραγωγή του προτύπου παραμετρικού διανύσματος για κάθε πλαίσιο του αγνώστου ακουστικού σήματος.

ΒΗΜΑ 2: Χρησιμοποιώντας την κατάλληλη συνάρτηση απόστασης και το αλφάβητο του συστήματος ταξινομούνται τα άγνωστα πρότυπα διανύσματα. Η έξοδος της βαθμίδας αυτής δεν είναι μονοσήμαντη εξαιτίας ης μεγάλης αμφιβολίας που υπάρχει, ιδιαίτερα στην περιοχή συνένωσης των φωνημάτων. Έτσι δημιουργείται ένα πλέγμα λύσεων T x L, όπου T είναι ο συνολικός αριθμός των εναλλακτικών λύσεων κάθε ταξινόμησης.

ΒΗΜΑ 3: Η αποκωδικοποίηση του T x L πλέγματος πολλαπλών λύσεων γίνεται με χρήση αλγορίθμων δυναμικού προγραμματισμού και προκύπτει η τελική ακολουθία φωνητικών συμβόλων που αποτελεί την φωνητική περιγραφή του ακουστικού σήματος.

ΒΗΜΑ 4: Στην συνέχεια, με χρήση πάλι του αλγορίθμου δυναμικού προγραμματισμού και ενός πίνακα αντικαταστάσεων (confusion matrix) γίνεται η διαδικασία προσπέλασης στο λεξικό για να προσδιοριστούν οι λέξεις που αντιστοιχούν στην φωνητική ακολουθία. Η διαδικασία προσπέλασης στο λεξικό δεν δίνει μονοσήμαντη έξοδο αλλά πλέγμα W x L’ λέξεων, μήκους W και βάθους L’ λέξεων (Murveit et al, 1986).

ΒΗΜΑ 5: Το πλέγμα λέξεων αποκωδικοποιείται με χρήση γλωσσικού μοντέλου. Μαρκοβιανές διαδικασίες πρώτης, δεύτερης και τρίτης τάξης χρησιμοποιούνται στην υλοποίηση του γλωσσικού μοντέλου (Paeseler et al,1989; Derouault et al, 1986; Katz,

1987; Kuhn et al, 1990; Tomita , 1986; Ney et al, 1994). Με το βήμα αυτό ολοκληρώνεται η διαδικασία αναγνώρισης ενός δομικού συστήματος.

Τα βήματα που περιγράφηκαν αποτελούν την διαδικασία αναγνώρισης των δομικών συστημάτων αλλά και όλων των συστημάτων αναγνώρισης που έχουν δομημένη αρχιτεκτονική (modular). Στα ολοκληρωμένα συστήματα αναγνώρισης δεν είναι δυνατή η διάκριση των διαφόρων τμημάτων που αποτελούν την βαθμίδα αναγνώρισης.

7.11.2 Κατανόηση ομιλίας

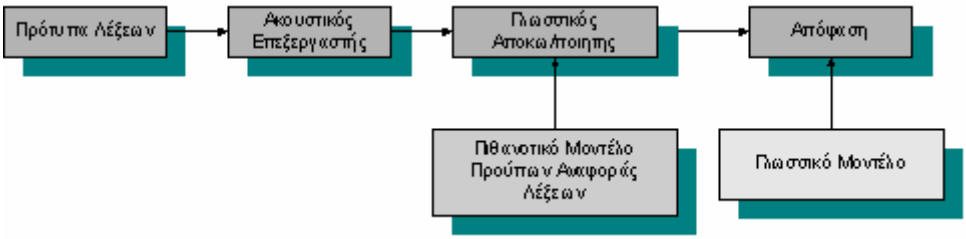
Γενικά, δεν μπορούμε να αποκτήσουμε ενδείξεις σχετικές με τη σύνθετη δομή των συντακτικών και πραγματολογικών σχέσεων της ομιλίας από ακουστικά δεδομένα μόνο. Παρά το γεγονός ότι η ακουστική δομή περιέχει όλη την πληροφορία που εκπέμπεται από τον ομιλητή στον ακροατή, μπορεί με ανάλυση να αξιολογηθεί μόνο σαν φορέας πληροφορίας, αφού οι διαδικασίες κατανόησης της ομιλίας και αναγνώρισης των ομιλητών εξαρτώνται, από φυσιολογικής πλευράς από παράγοντες άρθρωσης και από πνευματικής πλευράς από διαδικασίες μάθησης, που και τα δυο διέπονται από αρχές της κυβερνητικής. Έτσι η ακουστική μελέτη της διαδικασίας της ομιλίας περιορίζεται στην αναπαράσταση του φυσικού φορέα υπό μορφή ταλαντώσεων και παλμών μεταδιδόμενων μπρος και πίσω μέσω του αέρα. Η φυσιολογική κωδικοποίηση και αποκωδικοποίηση που γίνεται στα περιφερειακά και στα κεντρικά νευρικά συστήματα πρέπει να ληφθεί υπόψιν για τη μελέτη της αντίληψης, αλλά αυτό ανάγεται στον τομέα της ψυχοακουστικής (Winckel, 1966).

Οι διάφορες ομάδες ήχων ομιλίας απαιτούν για την παραγωγή τους διαφορετικές μορφές άρθρωσης με αποτέλεσμα να έχουν διαφορετικά ακουστικά χαρακτηριστικά. Κάθε ήχος χαρακτηρίζεται από μερικές συγκεντρώσεις ενέργειας (συντονισμούς) σε κάποιες συγκεκριμένες θέσεις συχνοτήτων. Επιπρόσθετα, τα εκρηκτικά, τα τυρβώδη και τα μη τυρβώδη φωνήματα έχουν ενέργεια σε πολλές ζώνες συχνοτήτων, λόγω των θορυβώδους ήχου πηγών τους. Κατά την διαδοχική παραγωγή παραγωγή ήχων ομιλίας υπάρχει μια αξιοσημείωτη ακουστική μετάβαση από τον ένα ήχο στον επόμενο, καθώς τα μέσα άρθρωσης «ίπτανται» μεταξύ των θέσεων που απαιτούνται για κάθε ένα από τους ήχους. Η μορφή και η θέση αυτής της μετάβασης καθορίζεται από τους συντονισμούς των δυο γειτονικών ήχων και είναι περισσότερο παρατηρίσιμη στην F_2, η οποία θεωρείται ότι μεταφέρει την περισσότερη πληροφορία σχετικά με τον τόπο άρθρωσης. Παρότι υπάρχουν μεταβάσεις άρθρωσης, και επομένως και ακουστικές, μεταξύ όλων των γειτονικών ήχων ομιλίας, ο πραγματικός τόπος άρθρωσης ενός συγκεκριμένου ήχου

μπορεί να αλλάξει συναρτήσει των ήχων που προηγούνται ή έπονται αυτού. Το φαινόμενο αυτό που ονομάζεται συνάρθρωση, μπορεί να είναι μονοκατευθυντήριο ή δικατευθυντήριο και οι ακουστικές του συνέπειες αφορούν το ότι τα πρότυπα συντονισμού (και ιδιαίτερα η F_2) των ξεχωριστών ήχων (ή ήχου) μετατίθενται προς τα πάνω ή προς τα κάτω μέσα στη ζώνη συχνοτήτων. Η μετάθεση του προτύπου συντονισμού του ενός ή και των δυο γειτονικών ήχων προκαλεί μεταβολές στη μορφή και στη θέση της μετάβασης του συντονισμού μεταξύ των ήχων.[20]

Ο άνθρωπος επεξεργάζεται τη φυσική γλώσσα για να κατανοήσει τα όσα ακούει από ένα ομιλητή ή διαβάζει σε ένα κείμενο. Η διαδικασία αυτή συνεπάγεται μια σειρά διαδοχικών λειτουργιών ή μετατροπών του ακουστικού ή οπτικού ερεθίσματος (δηλαδή της ομιλίας ή του κειμένου) που φτάνει στα αντίστοιχα αισθητήρια όργανα, τα αυτιά και τα μάτια, έτσι ώστε να καταλήξει στο νοητικό μήνυμα.

Ο πρωταρχικός σκοπός της ομιλίας είναι να μεταδίδει πληροφορία από το κεντρικό νευρικό σύστημα ενός ανθρώπου σε αυτό ενός άλλου. Ο μηχανισμος ομιλίας με τον οποίο μεταδίδεται αυτή η πληροφορία είναι κυρίως μια ολοκλήρωση κάποιων τμημάτων των συστημάτων για την επεξεργασία της τροφής και για την αναπνοή. Η ευελιξία της ομιλίας οφείλεται σε μεγάλο μέρος στο γεγονός ότι η οδός της τροφής και η οδός της αναπνοής τέμνονται. Έτσι, ένα ιδιαίτερο σύνθετο σύνολο ήχων διαφόρων τύπων μπορεί να παραχθεί από τον μηχανισμό παραγωγής ομιλίας.



7.12 Τύποι του σήματος ομιλίας

7.12.1 Διακριτή Ομιλία (Discrete Speech)

Τα συστήματα αναγνώρισης διακριτής ομιλίας απαιτούν από το χρήστη να προφέρει τις λέξεις με μικρές ενδιάμεσες παύσεις. Αυτό διευκολύνει το σύστημα για δύο λόγους: πρώτον εντοπίζεται εύκολα η αρχή και το τέλος της κάθε λέξης, και δεύτερον

αναγκάζει το χρήστη να προφέρει τις λέξεις πιο καθαρά .[10]

7.12.2 Συνεχής ομιλία (Continuous Speech)

Η συνεχής ομιλία είναι πολύ πιο δύσκολή για το σύστημα. Οι λέξεις ακολουθούν η μια την άλλη χωρίς διαστήματα σιγής και δύσκολα διακρίνεται το τέλος της μιας λέξης από την αρχή της επόμενης. Ακόμα ο χρήστης μπορεί να προφέρει τις ίδιες λέξεις διαφορετικά στην συνεχή ομιλία.

7.12.3 Ομιλία με θόρυβο (Speech with Background Noise)

Τα συστήματα αναγνώρισης συχνά αντιμετωπίζουν ποικιλία ήχων πέραν του σήματος ομιλίας που μας ενδιαφέρει. Αυτό είναι ιδιαίτερα δύσκολο όταν ο θόρυβος αποτελείται από άλλες φωνές.

7.12.4 Ομιλία με διαταραχή (Speech with Distortion)

Μερικές φορές το σήμα έχει διαταραχή. Αυτό μπορεί να προκληθεί από κακό μικρόφωνο,

αντήχηση του δωματίου κλπ. Αν η διαταραχή είναι συνεχής , το σύστημα μπορεί να εκπαιδευτεί ώστε να την αντιμετωπίζει ικανοποιητικά. Αν όχι , υπάρχει σοβαρό πρόβλημα στο σύστημα.

7.12.5 Ομιλία με περιττό θόρυβο (Speech with Superfluous Noises)

Όταν μιλάμε, συχνά κάνουμε θορύβους ανεπιθύμητους στο σήμα ομιλίας. Ήχοι όπως «εεεε....», θόρυβοι γλώσσας και χειλιών, καθάρισμα του λαιμού και άλλα παρόμοια πρέπει να αναγνωρίζονται σαν άχρηστη πληροφορία και να απομακρύνονται.

Ένας χώρος στον οποίο γίνεται εκτεταμένη χρήση των νευρωνικών δικτύων είναι η αναγνώριση ομιλίας. Σε αυτό το κεφάλαιο θα κάνουμε μια αναφορά στην εφαρμογή αυτών στο πρόβλημα της αναγνώρισης ομιλίας και θα αναφέρουμε κάποιες αρχιτεκτονικές που εφαρμόζονται σε αυτό το πεδίο.

Στο φως τριών δεκαετιών προσπάθειας με στόχο την δημιουργία συστημάτων ικανών να αναγνωρίσουν ανθρώπινη φυσική ομιλία, μπορεί κάποιος να αναρωτηθεί αν το πρόβλημα έχει επιλυθεί, και αν όχι, τι απαιτείται προκειμένου να εκπληρωθεί αυτή η πρόκληση. Για να είμαστε ακριβείς στο πεδίο αυτό έχουν γίνει σημαντικά βήματα, και πάρα πολλά σημαντικά μαθήματα έχουμε πάρει από τις διάφορες αποτυχίες και επιτυχίες των ερευνητών.

Παρόλα αυτά, τα συστήματα που έχουν αναπτυχθεί ως τώρα δεν είναι ικανά και δεν παρέχουν την απαιτούμενη αξιοπιστία με την οποία οι άνθρωποι μπορούν να επικοινωνούν μεταξύ τους χρησιμοποιώντας την ομιλία. Αυτό λοιπόν το κενό έχει οδηγήσει στη συνεχή έρευνα για την δημιουργία νέων μοντέλων και τεχνικών που θα μας φέρουν πιο κοντά στην δημιουργία μηχανών που θα μπορούν να εκτελέσουν τα παραπάνω όσο το δυνατόν πλησιέστερα στον τρόπο με τον οποίο τις εκτελούν οι άνθρωποι. Προβλήματα τα οποία κάνουν πιο δύσκολη την δημιουργία αυτών των μηχανών είναι, η διαφορετική προφορά της ίδιας λέξης, διάφοροι ιδιωματισμοί κ.α.

Τα νευρωνικά δίκτυα είναι η πιο πρόσφατη ανακάλυψη στην έρευνα νέων

μοντέλων αναγνώρισης ομιλίας. Παρακάτω θα προσπαθήσουμε να περιγράψουμε σύντομα και κατανοητά αρχιτεκτονικές νευρωνικών δικτύων οι οποίες αποτελούν σχετικά πρόσφατη δου-λειά. Αξίζει να σημειωθεί ότι ως τώρα έχουν γίνει πειράματα και εφαρμογές σε μέρη του προβλήματος και όχι στο σύνολο του. Αν και μια πάρα πολλά υποσχόμενη δουλειά έχει αρχίσει προς αυτή την κατεύθυνση, κανένα ολοκληρωμένο - ευρέως λεξιλογίου- σύστημα αναγνώρισης ομιλίας βασισμένο στα νευρωνικά δίκτυα και μόνο, δεν έχει πραγματοποιηθεί ως τώρα. Όμως τι είναι αυτό που γεμίζει τις 'μπαταρίες' των επιστημόνων για την συνέχιση της έρευνας προς αυτή την κατεύθυνση (των νευρωνικών δικτύων), και ποια είναι η υπόσχεση πίσω από τα "νευρωνικά δίκτυα", "συνδετικότητα" ή "παράλληλα καταμερισμένη επεξεργασία ", όπως αυτά τα μοντέλα συχνά έχουν ονομαστεί. Μια επιμέρους λίστα κάποιων "ελκυστικών" χαρακτηριστικών που παρέχουν τα νευρωνικά δίκτυα είναι η ακόλουθη.[23]

Μαζικός παραλληλισμός (Massive parallelism): Ένα νευρωνικό δίκτυο αποτελείται από πολλές μικρότερες υπολογιστικές μονάδες. Ο υπολογισμός γίνεται παράλληλα και με έναν καταμερισμό μεταξύ πολλών και συνδεδεμένων μεταξύ τους στοιχείων. Σημαντικά οφέλη που προκύπτουν σαν αποτέλεσμα των προηγουμένων είναι η ταχύτητα, η απλότητα (όσον αφορά την εφαρμογή τους ως hardware), και η ανοχή σε τυχόν λάθη.

Ικανοποιητική δημιουργία (Constraint satisfaction): Η επεξεργασία στα νευρωνικά δίκτυα δεν πραγματοποιείται σειριακά και δεν εξαρτάται από την απόδοση καθεμιάς ξεχωριστά υπολογιστικής μονάδας (νευρώνας), αλλά από το σύνολο.

Εκπαίδευση (Learning): Οι μαζικές παράλληλες αρχιτεκτονικές δεν μπορούν να προγραμματιστούν εύκολα, και τα μοντέλα στα οποία η επεξεργασία γίνεται με παράλληλο διαχωρισμό εξαρτώνται από κάποιο αυτόματο αλγόριθμο εκμάθησης. Ένας μεγάλος αριθμός τέτοιων αλγορίθμων υπάρχουν αυτή τη στιγμή, συγκαταλεγομένων και των τεχνικών εκπαίδευσης για πολυεπίπεδο διαχωρισμό και κατάταξη προτύπων (multilayer perceptron (back propagation)), μηχανές Boltzmann, κβαντισμό διανυσμάτων εκμάθησης και σύνδεσμος δικτύων. Αυτοί οι αλγόριθμοι εκμάθησης διαχειρίζονται τοπικά υπολογιστικά στοιχεία προκειμένου να βελτιστοποιήσουν στο σύνολο τους κάποιους ευρύτερους στόχους.

Στοχαστικά μοντέλα, αβεβαιότητα, μεταβλητότητα, ασάφεια (stochastic modeling, uncertainty, variability, fuzziness): Τα μοντέλα που χρησιμοποιούν τη μέθοδο των νευρωνικών δικτύων αντιμετωπίζουν τη μεταβλητότητα και τον θόρυβο βρίσκοντας κάθε φορά την κατάλληλη πιθανοτική γενίκευση. Τα νευρωνικά δίκτυα δεν απαιτούν καμιά συγκεκριμένη στατιστική βοήθεια, και για αυτό δεν χρειάζεται να γίνει καμιά

παραμετρική υπόθεση.

Μη γραμμικά μοντέλα (Nonlinear modeling): Τα νευρωνικά δίκτυα είναι μη γραμμικά μοντέλα που μπορούν να αντικαταστήσουν μη γραμμικούς ταξινομητές και συναρτήσεις ταξινόμησης. Επίσης μπορούν να διαχειριστούν εργασίες μεταξύ διαφορετικών μοντέλων, ακόμα και μεταξύ πολύπλοκων σχέσεων. Αυτό μπορεί να οδηγήσει σε καλύτερες επιδόσεις σε σχέση με την αντίστοιχη ενός γραμμικού μοντέλου σε διάφορες τμηματοποιήσεις, mapping, και interpolar tasks.

Ανακάλυψη κρυμμένης γνώσης (Discovery of "hidden" knowledge): Τα νευρωνικά δίκτυα παράγουν κρυμμένη γνώση, αφαιρούν, και γενικεύουν κατά την διαδικασία της λύσης ενός πιο πολύπλοκου προβλήματος. Στην εφαρμογή ενός πολυεπίπεδου perceptron, αυτή η κρυμμένη λογική είναι συχνά κωδικοποιημένη στα συνδετικά βάρη που ονομάζονται "κρυφές μονάδες"(Rumelhart and McClelland, 1986) [26]. Αν η πληροφορία μπορεί να εξαχθεί ή να κωδικοποιηθεί ικανοποιητικά μέσα σε αυτό το δίκτυο, μπορεί αυτή να δώσει μηχανισμούς για να ελαττώσει το κενό μεταξύ των προσεγγίσεων που βασίζονται σε πληροφορίες και στα στοχαστικά μοντέλα.

Ομοιομορφία (Uniformity): Ο υπολογισμός στα νευρωνικά δίκτυα γίνεται με απλά κρυφά υπολογιστικά στοιχεία τα οποία επικοινωνούν μεταξύ τους. Τα υπολογιστικά βήματα που ακολουθούνται από μια συγκεκριμένη μονάδα (συνήθως πολλαπλασιασμοί και προσθέσεις) είναι γενικά ανεξάρτητα από το πρόβλημα που το δίκτυο προσπαθεί να επιλύσει. Αυτό είναι κάτι πολύ χρήσιμο κατά την σχεδιάσει hardware μιας και οι μονάδες είναι απλές (φτηνές) και οι ίδιες (μονάδες) μπορούν να χρησιμοποιηθούν για μια ποικιλία προβλημάτων. Η ομοιο-μορφία είναι επίσης χρήσιμη σαν μέσο προκειμένου να επιτύχουμε την επιθυμητή σύγκληση, π.χ. τον δυναμικό συνδυασμό μεταξύ διαφορετικών σημάτων ή πληροφοριών εισόδου (για παράδειγμα στην ομιλία, σε φωνητικές και οπτικές σειρές ή συντακτικές, νοηματικές, και σε σχέση με την προφορά σειρές κ.α.).

Ταχύτητα εκπαίδευσης σε σχέση με την αναγνώριση (Speed-learning vs. Recognition): Εξαιτίας του μαζικού παράλληλου υπολογισμού που λαμβάνει χώρα στα νευρωνικά δίκτυα δίνεται η δυνατότητα σε αυτά να τρέχουν πιο ικανοποιητικά και με καλύτερα αποτελέσματα. Από την άλλη όμως κάποια νευρωνικά δίκτυα απαιτούν να γίνει μια κάποια καθόλου ευκαταφρόνητη εκπαίδευση.

Αντιγραφή των λειτουργιών του ανθρώπινου εγκεφάλου για την πραγματοποίηση των υπολογισμών. Με τα νευρωνικά δίκτυα γίνεται μια προσπάθεια να εξομοιωθεί ο τρόπος που γίνεται κάποιος υπολογισμός από το νευρικό μας σύστημα. Αν και υπάρχουν κάποιες ομοιότητες των υπαρχόντων νευρωνικών δικτύων με το νευρικό μας σύστημα, αυτή η αναλογία δεν πάει πολύ μακριά. Φτάνει μέχρι τα πρώτα πολύ απλά στάδια

λειτουργίας του εγκεφάλου. Οι γνώσεις μας αυτή την στιγμή για τον τρόπο που γίνονται οι υπολογισμοί στον εγκέφαλο καλύπτουν μόνο κάποια πεδία της όλης διαδικασίας και όχι όλο το σύνολο αυτής. Επίσης η υπάρχουσα τεχνολογία δεν είναι ακόμα σε τέτοια επίπεδα έτσι ώστε να επιτρέπει τέτοιες συγκρίσεις. Πρέπει να πούμε ότι είναι απαραίτητο να αντιγράψουμε πλήρως τον εγκέφαλο προκειμένου να δημιουργήσουμε χρήσιμα συστήματα αναγνώρισης ομιλίας σε Η/Υ. Ο ανθρώπινος εγκέφαλος είναι μια ζωντανή απόδειξη του γεγονότος ότι έξυπνη επικοινωνία με ομιλία είναι δυνατή. Με τη διόραση και την διαίσθηση αλλά και με διάφορες πληροφορίες που μπορούμε να συλλέξουμε από τις διάφορες υπολογιστικές μεθόδους που χρησιμοποιεί ο εγκέφαλος μπορούμε να πάρουμε χρήσιμες ιδέες και νέα μοντέλα για πρακτική σχεδίαση συστημάτων.

Κεφάλαιο 8 – Εφαρμογές

8.1 Αναγνώριση ομιλίας σε εφαρμογές

Τα τελευταία χρόνια υπάρχει πληθώρα εφαρμογών που χρησιμοποιεί την αναγνώριση Ομιλίας με την πιο κοινή στο κόσμο να είναι ο ψηφιακός οδηγός της Apple το “Siri”. Το καλό με την φωνητική αναγνώριση είναι ότι μπορείς να χρησιμοποιήσεις ηλεκτρονικές συσκευές, όπως υπολογιστές ή κινητά τηλέφωνα, απλά με την ομιλία. Σημαντικό θέμα στην αναγνώριση ομιλίας είναι η απόδοση του συστήματος που χρησιμοποιούμε. Η απόδοση συνήθως αξιολογείται για την ακρίβεια και την ταχύτητα. Η ακρίβεια συνήθως υπολογίζεται με τον Λόγο Σφάλματος ανά Λέξη (WER) , ενώ η ταχύτητα μετριέται με την ταχύτητα ως προς τον πραγματικό χρόνο.

Ωστόσο, η αναγνώριση ομιλίας είναι ένα σύνθετο πρόβλημα. Οι φωνές διαφέρουν στην έμφαση, προφορά, άρθρωση, τραχύτητα, έρρινα, ένταση και την ταχύτητα. Επίσης η ομιλία διαστρεβλώνεται από θορύβους χώρου και ηχώ. Η ακρίβεια της αναγνώρισης διαφοροποιείται ανάλογα με τα εξής:

- Μέγεθος λεξιλογίου
- Εξάρτηση ομιλητή /Ανεξαρτησία
- Απομονωμένες, ασυνεχής, ή συνεχής ομιλία
- Περιορισμούς ειδικούς ή γλώσσας
- Διάβασμα/Αυθόρμητη ομιλία
- Αντίξοες συνθήκες

Η μέτρηση της προόδου της απόδοσης της αναγνώρισης ομιλίας είναι δύσκολη και αμφιλεγόμενη. Μερικά καθήκοντα αναγνώρισης είναι πιο δύσκολα από ότι άλλα. Σε κάποια καθήκοντα ο Λόγος Σφάλματος ανά Λέξη είναι 1%, σε άλλα μπορεί και υψηλό μέχρι και 50%. Μερικές φορές φαίνεται ακόμα ότι οι επιδόσεις δεν εξελίσσονται αλλά πάνε προς τα πίσω, καθώς ερευνητές αναλαμβάνουν πιο δύσκολα καθήκοντα που έχουν μεγαλύτερα ποσοστά λάθους. Από την πλευρά της έρευνας και επειδή η πρόοδος είναι αργή και είναι δύσκολο να μετρηθεί, υπάρχει η αντίληψη ότι η απόδοση έχει σταθεροποιηθεί και ότι η χρηματοδότηση έχει στερέψει ή έχει επιλέξει άλλες προτεραιότητες. Αυτές οι αντιλήψεις δεν είναι νέες. Το 1969, ο John Pierce έγραψε μία ανοιχτή επιστολή, που προκάλεσε το ύψος της χρηματοδότησης να στεγνώσει για αρκετά χρόνια. Το 1993 υπήρξε μία ισχυρή αίσθηση ότι η απόδοση είχε σταθεροποιηθεί και υπήρχαν διάφορα ερευνητικά εργαστήρια αφιερωμένα στο θέμα. Ωστόσο, στην

δεκαετία του 1990, η χρηματοδότηση συνέχισε να βελτιώνεται αργά, αλλά σταθερά. Πλέον συναντάμε εφαρμογές με ενσωματωμένη αναγνώριση ομιλίας σε διάφορα σημεία όπως :

- Αυτοματισμούς σπιτιών
- Αυτόματη μετάφραση
- Ενσωματωμένα συστήματα σε αυτοκίνητα
- Ελικόπτερα
- Ηλεκτρονική υγεία
- Ρομποτικά συστήματα
- Τηλεματική
- Τηλέφωνα

8.2 Τρόπος λειτουργίας

Ένας τρόπος για την διαδικασία αναγνώρισης ομιλίας είναι να ληφθεί η κυματομορφή της, να χωριστεί σε κενά σημεία ανάμεσα από τις λέξεις και να γίνει προσπάθεια αναγνώρισης σε κάθε λέξη. Στην συνέχεια επιλέγονται οι λέξεις που ταιριάζουν περισσότερο από μία βάση λέξεων της βιβλιοθήκης που κάνει την αναγνώριση. Εφόσον μαζευτούν όλοι οι πιθανοί συνδυασμοί λέξεων για όλη την φράση συγκρίνονται όλες με τον ήχο της ομιλίας και επιλέγεται η λέξη που ταιριάζει περισσότερο. Ο τρόπος λειτουργίας της αναγνώρισης βασίζεται κυρίως στον αλγόριθμο που θα χρησιμοποιηθεί. Συνήθως θέλουμε να κάνουμε αναγνώριση ομιλίας σε κάποιο σύστημα με όχι αρκετή επεξεργαστική ισχύ, αλλά δεν θέλουμε να έχουμε αλλοίωση αποτελέσματος. Για αυτή την περίπτωση υπάρχουν ειδικές βιβλιοθήκες που με την χρήση του μοντέλου πελάτη – εξυπηρετητή συνδέονται σε έναν εξυπηρετητή που κάνει την αναγνώριση ομιλίας και επιστρέφει το αποτέλεσμα. Τα πιο γνωστά συστήματα που κάνουν μεγάλη επεξεργασία δεδομένων και χρησιμοποιούν αναγνώριση ομιλίας, βασίζονται σε αυτόν τον τρόπο ή σε τέτοιες βιβλιοθήκες.

8.3 Λογισμικά Αναγνώρισης φωνής

Χωρίς να φτάνει το πλήθος εταιρειών και προϊόντων των συστημάτων δακτυλικών αποτυπωμάτων, η αναγνώριση φωνής έχει μια αρκετά ευπρεπή εκπροσώπηση στην αγορά. Απ' ότι φαίνεται, οι περισσότερες εταιρείες του χώρου ρίχνουν το βάρος τους στην κατασκευή συστημάτων κατάλληλα για τηλεφωνικά δίκτυα ή για το Internet. Πολλές από

αυτές μάλιστα, όπως η ITT, εντάσσουν τα προϊόντα τους στα γενικότερα πλαίσια προγραμμάτων υπηρεσιών, τα οποία και χρεώνουν ανάλογα με παραμέτρους όπως η συχνότητα των αναγνωρίσεων σε ορισμένη μονάδα χρόνου, ο αριθμός των σταθμών εργασίας κ.λπ. Έτσι, ο λιγότερο επικερδής τομέας της ασφάλειας των προσωπικών υπολογιστών φαίνεται να μπαίνει σε δεύτερη μοίρα, τουλάχιστον προσωρινά. Παρ' όλα αυτά, και εκεί δε θα μπορούσαν να λείψουν κάποια πολύ αξιόλογα προϊόντα, όπως μπορούμε να δούμε παρακάτω.

8.3.1 Αναγνώριση φωνής στο Internet

Αν θέλετε να διαπιστώσετε την αποτελεσματικότητα της αναγνώρισης φωνής και παράλληλα έχετε μικρόφωνο και σύνδεση στο Internet, υπάρχουν τουλάχιστον δύο τρόποι για να το πετύχετε. Ο πρώτος είναι να κατεβάσετε το πρόγραμμα DemoKey από τις σελίδες των προϊόντων SpeakerKey (η απευθείας διεύθυνση είναι (http://www.speakerkey.com/speakerkey/docs/down.htm). Το πρόγραμμα αυτό έχει δύο φάσεις. Στην πρώτη, πρέπει να απαγγείλετε δώδεκα ζευγάρια αριθμών που σας προτείνει το πρόγραμμα ώστε να δημιουργήσει ένα φωνητικό προφίλ σας. Στη δεύτερη, τη φάση της αναγνώρισης, το πρόγραμμα προσπαθεί να σας αναγνωρίσει μετά την απαγγελία δύο ζευγαριών αριθμών.[24]

Ο δεύτερος τρόπος είναι να δοκιμάσετε τη διεύθυνση (http://iris1.let.kun.nl/TSpublic/cave). Στη συγκεκριμένη σελίδα (η οποία αποτελεί στην ουσία ένα on-line demo της μεθόδου) έχουν πρόσβαση μόνο εξουσιοδοτημένοι χρήστες μέσω ενός συστήματος αναγνώρισης φωνής. Αν θέλετε να πειραματιστείτε λοιπόν, δεν έχετε παρά να εγγραφείτε στη λίστα των εξουσιοδοτημένων χρηστών αφήνοντας το φωνητικό σας αποτύπωμα. Την επόμενη φορά που θα επισκεφθείτε την ιστοσελίδα θα πρέπει να απαγγείλετε τη φράση-κλειδί ώστε το σύστημα να σας αναγνωρίσει και να σας επιτρέψει την πρόσβαση.

8.3.2 Εταιρείες και Προϊόντα

8.3.2.1 VoicEntry II

Αν έχετε αμφιβολίες σχετικά με το πόσο οικονομικό μπορεί να είναι ένα βιομετρικό σύστημα για τον υπολογιστή σας, το λογισμικό VoicEntry II της T-Netix (http://www.t-netix.com) είναι η καλύτερη απόδειξη γι' αυτό. Πρόκειται για ένα ολοκληρωμένο πρόγραμμα προστασίας του συστήματός σας, με λειτουργίες όπως διαχείριση χρηστών,

έλεγχο πρόσβασης στα Windows, και κρυπτογράφηση μεμονωμένων αρχείων ή φακέλων. Επιπλέον, στο πρόγραμμα περιέχεται ένας screen-saver ο οποίος απενεργοποιείται μόνο μετά από προφορική εντολή του εξουσιοδοτημένου χρήστη. Το πρόγραμμα δίνει επίσης τη δυνατότητα στο χρήστη να εφαρμόσει βιομετρικό έλεγχο σε οποιοδήποτε άλλο πρόγραμμα προστασίας οθόνης θέλει εκείνος. Η τιμή του πακέτου είναι μόλις 50 δολάρια, δηλαδή γύρω στα 44 €, τη στιγμή που ένα σύστημα δακτυλικών αποτυπωμάτων αντίστοιχων δυνατοτήτων φτάνει τα 132 €.

8.3.2.2 SpeakerKey

Η ITT Industries (http://www.ittind.com) είναι μία από τις πρώτες εταιρείες πουασχολήθηκαν με την αναγνώριση φωνής. Χρησιμοποιώντας την πολύχρονη πείρα της κατασκεύασε τη σειρά προγραμμάτων SpeakerKey, μία από τις πιο ολοκληρωμένες λύσεις για περιπτώσεις όπου η αναγνώριση ταυτότητας μέσω τηλεφωνικού ή άλλου είδους δικτύου είναι απαραίτητη. Το πακέτο περιλαμβάνει τρεις εφαρμογές, κάθε μία από τις οποίες ειδικεύεται και σε έναν τύπο δικτύου: Η εφαρμογή PhoneKey είναι κατάλληλη για περιπτώσεις επιβεβαίωσης ταυτότητας μέσω τηλεφώνου, η εφαρμογή NetKey για τοπικά δίκτυα υπολογιστών, ενώ η εφαρμογή WebKey έχει ως πεδίο δράσης τον Παγκόσμιο Ιστό. Αναλυτικότερες πληροφορίες μπορείτε να πάρετε στη διεύθυνση (http://www.speakerkey.com).

8.3.2.3 CMU Sphinx

Η Sphinx είναι η πιο γνωστή βιβλιοθήκη για φωνητική αναγνώριση, είναι ανοιχτού κώδικα και αναπτύσσετε από το πανεπιστήμιο του Carnegie Mellon εδώ και 32 χρόνια, αλλά από το 2000 και από την κοινότητα του ανοιχτού κώδικα. Το 2000 η ομάδα Sphinx του πανεπιστημίου του Carnegie Mellon αφοσίωσε στην κοινότητα του ανοιχτού κώδικα διάφορες συνιστώσες για φωνητική αναγνώριση. Η πιο σημαντική δυνατότητα της συγκεκριμένη βιβλιοθήκης είναι η δυνατότητα να προσαρμοστεί σε μία συγκεκριμένη διάλεκτο με σχετικά εύκολο τρόπο με την χρήση διαφόρων εργαλείων που έχουν αναπτυχθεί. Βέβαια μπορεί να προσαρμοστεί και σε ολόκληρες γλώσσες και όχι μόνο σε διαλέκτους συγκεκριμένης γλώσσας. Άλλη σημαντική δυνατότητα είναι ότι μπορεί να επιλεχθεί ένα λεξικό δεδομένων και η αναγνώριση ομιλίας θα δουλεύει μόνο για τις συγκεκριμένες λέξεις.[25]

8.3.2.4 Microsoft Speech API

Το Speech API της Microsoft παρέχει μία διεπαφή υψηλού επιπέδου που εφαρμόζει όλες τις χαμηλού επιπέδου λεπτομέρειες που απαιτούνται για τον έλεγχο και την διαχείριση των εργασιών σε πραγματικό χρόνο των διαφόρων μηχανών ομιλίας . [26]

8.3.2.5 iSpeech

Το API της iSpeech για φορητές συσκευές είναι μία δωρεάν και αξιόλογη λύση, βέβαια προσφέρει και web έκδοση για το API που είναι επι-πληρωμή βασιζόμενο στις λέξεις που μεταφράζονται. Το συγκεκριμένο API ενώ υποστηρίζει διάφορες γλώσσες, δεν υποστηρίζει δυστυχώς την ελληνική και επίσης δεν είναι και τόσο εύκολη η αναγνώριση ομιλίας όπως στο Speech Recognizer . [27]

8.3.2.6 Speech Recognizer

Η συγκεκριμένη βιβλιοθήκη είναι μονόδρομος για όποιον θέλει να κάνει φωνητική αναγνώριση στο λειτουργικό Android,αν και χρειάζεται να είναι συνδεδεμένο στο Internet για να δουλέψει η αναγνώριση, η αποτελεσματικότητά του είναι αρκετά καλή μιας και βασίζεται στην μηχανή αναγνώρισης του Google Now . Επίσης η βιβλιοθήκη είναι πολύ εύκολη στην χρήση, πολύ εύκολη στην παραμετροποίηση, αλλά υποστηρίζει και την ελληνική γλώσσα [28]

8.3.2.7 Pocketsphinx

Η συγκεκριμένη βιβλιοθήκη είναι ανεπτυγμένη από το πανεπιστήμιο του Carnegie Mellon, που έχει αναπτύξει το CMU Sphinx . Η συγκεκριμένη βιβλιοθήκη είναι γραμμένη σε JavaScript και είναι μία βιβλιοθήκη που προορίζεται να χρησιμοποιείται για επεξεργασία αναγνώριση ομιλίας στο διαδίκτυο, από ιστοσελίδες. Στην ουσία είναι ένα μικρό κομμάτι της CMU Sphinx, αν και συγκριτικά με την CMU Sphinx το αποτέλεσμα της Pocketsphinx δεν είναι αρκετά καλό, το αποτέλεσμα παραμένει αρκετά καλό.

8.3.2.8 Annyang, annyang – node

Ο Tal Ater δημιούργησε την Javascript βιβλιοθήκη annyang , για αναγνώριση ομιλίας σε ιστοσελίδες και στο διαδίκτυο που δουλεύει εξ ολοκλήρου στην πλευρά του

χρήστη όπως η Pocketsphinx . Η συγκεκριμένη βιβλιοθήκη είναι δωρεάν για χρήση και με MIT άδεια για επεξεργασία. Υπάρχει βέβαια και η Javascript βιβλιοθήκη annyang-node που βασίζεται στην βιβλιοθήκη του Tal Ater , αλλά είναι σχεδιασμένη να μπορεί να κάνει την φωνητική αναγνώριση στην πλευρά του εξυπηρετητή της ιστοσελίδας ώστε να μην επιβαρύνει το σύστημα του χρήστη.

8.3.2.9 Speech API

Το Speech API αν και beta έκδοση παρέχει την δυνατότητα να χρησιμοποιηθεί σε ιστοσελίδες είτε με την χρήση Javascript , είτε με την χρήση Flash , είτε με την χρήση PHP. Μάλιστα η συγκεκριμένη βιβλιοθήκη μπορεί να χρησιμοποιηθεί και από άλλες γλώσσες :

- Javascript
- Flash
- PHP
- Python
- Ruby
- Java

8.3.2.10 HTML5 Speech Recognition

Η συγκεκριμένη βιβλιοθήκη είναι η πρόταση της Google που έχει ενσωματώσει μάλιστα και στον translate.google.com ώστε να μπορεί να γίνει αυτόματα η μετάφραση ομιλίας. Η Google είχε καταθέσει ολόκληρη σχεδίαση στην Κοινοπραξία Παγκόσμιου Ιστού W3C για το πώς θα μπορούσε να είναι η αναγνώριση ομιλίας στην HTML5.

8.4 Άλλα προϊόντα

Επίσης, άξια αναφοράς προϊόντα είναι το VoiceGuardian της Keyware (http://www.keywareusa.com), καθώς και τα πακέτα της VeriVoice (http://www.verivoice.com), τα οποία μάλιστα καλύπτουν ένα μεγαλύτερο εύρος εφαρμογών, συμπεριλαμβανομένης της υποστήριξης έξυπνων καρτών (smart cards).

8.4.1 MLS Αναγνώριση Φωνής[PRO-083]

Μια πρωτοποριακή και καινοτομική εφαρμογή της MLS που αξιοποιεί τη δύναμη

της φωνής σας! Το Πρόγραμμα Φωνητικών Εντολών της MLS δίνει στον υπολογιστή σας τη δυνατότητα να αναγνωρίζει ελληνικά και σε εσάς τη δυνατότητα να χειρίζεστε τις εφαρμογές του MS Office, αλλά και όλες τις εφαρμογές που συνοδεύουν τα Windows, απλά και μόνο με τη φωνή σας. Μεταξύ των εντυπωσιακών δυνατοτήτων τις οποίες σας προσφέρει η πρώτη εφαρμογή Αναγνώρισης Φωνής για τα ελληνικά δεδομένα είναι και οι ακόλουθες: πλοήγηση στο διαδίκτυο, τροποποίηση και διόρθωση κειμένων στο Word, το Σημειωματάριο, το Wordpad και το Outlook Express, επιλογή της αγαπημένης σας μουσικής, επικόλληση κειμένων, δημιουργία φωνητικών πλήκτρων συντόμευσης, χειρισμός της προβολής των παρουσιάσεων του Powerpoint, κ.α .

8.4.2 Julius

Το Julius είναι ένα υψηλής απόδοσης λογισμικό αποκωδικοποίησης της φωνής με δυνατότητα αναγνώρισης ενός μεγάλου λεξιλογίου (συνεχόμενου λόγου). Βασισμένο σε λέξεις των τριών γραμμάτων και στο πλαίσιο HMM, μπορεί να αποκωδικοποιήσει μια λέξη των 60k σχεδόν σε πραγματικό χρόνο, στους περισσότερους υπάρχοντες υπολογιστές. Έχουν ενσωματωθεί οι περισσότερες τεχνικές αναζήτησης. Έχει επίσης δημιουργηθεί προσεκτικά ώστε να είναι ανεξάρτητο από τις μοντελοποιημένες δομές, και οι διάφοροι τύποι HMM όπως τρισύλλαβες λέξεις, υποστηρίζονται με οποιοδήποτε συνδυασμό, προτάσεις και φωνήματα. Το Julius δούλευε άριστα στα λειτουργικά συστήματα Linux και Unix, αλλά δουλεύει με επιτυχία και σε Windows. Το Julius είναι λογισμικό ανοικτού κώδικα και διανέμεται με άδεια BSD.Το Julius αναπτύχθηκε από το 1997 για την ιαπωνική έρευνα LVCSR ως μέρος μίας δέσμης λογισμικών, και οι εργασίες συνεχίστηκαν στο Continuous Speech Recognition Consortium (CSRC), στην Ιαπωνία, από το 2000 έως το 2003. Στις τελευταίες εκδόσεις ενσωματώθηκε στο Julius ένας αναλυτής αναγνώρισης φωνής που βασίζεται στη γραμματική, με το όνομα Julian. Το Julian είναι μία τροποποιημένη έκδοση του Julius που χρησιμοποιεί τη γραμματική ως πρότυπο γλώσσας. Μπορεί να χρησιμοποιηθεί για να δημιουργηθεί ενός είδους συστήματος φωνητικών εντολών με τη βοήθεια ενός μικρού λεξιλογίου.

8.4.3 VoxForge

Το VoxForge δημιουργήθηκε για να συλλέξει μεταγραφή ομιλίας (transcription) για τη χρήση της από εργαλεία αναγνώρισης ομιλίας του ελεύθερου και ανοικτού κώδικα λογισμικού (Open Source) για Linux / Unix, Windows και Mac. Όλα τα αρχεία ήχου που έχουν υποβληθεί υπό την άδεια GPL θα συγκεντρωθούν σε ακουστικά μοντέλα για χρήση από λογισμικό ανοιχτού κώδικα για αναγνώριση ομιλίας, όπως το Sphinx, το ISIP, το

Julius και το HTK (σημείωση: το HTK έχει περιορισμούς διανομής). Το VoxForge άρχισε πρόσφατα να χρησιμοποιεί το LibriVox ως πηγή ηχητικών δεδομένων[36]

8.4.4 HTK

Το HTK (Hidden Markov Model Toolkit) είναι ένα λογισμικό εργαλείο για τον χειρισμό HMMs. Προορίζεται κυρίως για την αναγνώριση φωνής, αλλά έχει χρησιμοποιηθεί σε πολλές άλλες εφαρμογές αναγνώρισης προτύπων που απασχολούν HMMs.

8.4.5 CSLU Toolkit

Το CSLU Toolkit είναι μία βιβλιοθήκη λογισμικού που περιλαμβάνει μια ολοκληρωμένη σουίτα εργαλείων που επιτρέπουν την εξερεύνηση, τη μάθηση και την έρευνα μεταξύ της αλληλεπίδρασης της ομιλίας και του ανθρώπου με τον υπολογιστή. Τα εργαλεία περιλαμβάνουν:

- Ήχο.
- Προβολή.
- Αναγνώριση φωνής.
- Γεννήτρια φωνής.
- Εφέ.

8.4.6 Dragon NaturallySpeaking

Το Dragon NaturallySpeaking είναι ένα λογισμικό αναγνώρισης φωνής που αναπτύχθηκε από τη Dragon Systems, και πωλείται από τη Nuance Communications για προσωπικούς υπολογιστές με Windows λογισμικό. Ήταν ένα από τα πρώτα προγράμματα που έκαναν πρακτική την αναγνώριση φωνής σε προσωπικούς υπολογιστές. Το NaturallySpeaking χρησιμοποιεί μια απλή οπτική διεπαφή. Οι λέξεις που υπαγορεύονται εμφανίζονται σε ένα πλωτό tooltip καθώς εκφωνούνται, και όταν ο ομιλητής κάνει παύση, το πρόγραμμα μεταφέρει τις λέξεις στο ενεργό παράθυρο στη τοποθεσία του κέρσορα. Όπως και άλλα λογισμικά αναγνώρισης φωνής, το NaturallySpeaking έχει τρεις βασικούς τομείς λειτουργικότητας: την υπαγόρευση, όπου ο προφορικός λόγος μεταφέρεται σε γραπτό κείμενο, τις εντολές ελέγχου, όπου ο προφορικός λόγος αναγνωρίζεται ως εντολές και τελικά το text-to-speech, όπου το γραπτό

κείμενο μετατρέπεται σε σύνθεση ροής ήχου. Οι αρχικές εκδόσεις αυτού του λογισμικού έπρεπε να εκπαιδευτούν για περίπου 10 λεπτά ώστε να αναγνωρίζουν τη φωνή του ομιλητή, ωστόσο στην έκδοση 9 αυτή η απαίτηση εγκαταλείφθηκε. Τα προφίλ της φωνής μπορούν να προσεγγιστούν από διαφορετικούς ηλεκτρονικούς υπολογιστές σε ένα δικτυωμένο περιβάλλον, ωστόσο και σε όλους τους υπολογιστές το hardware για τον ήχο και οι ρυθμίσεις του προγράμματος πρέπει να είναι πανομοιότυπες. Η Nuance ισχυρίζεται ότι χρησιμοποιώντας το NaturallySpeaking, για την εγγραφή μιας έκθεσης 900 λέξεων θα χρειαζόταν 6 λεπτά, ενώ η εγγραφή μιας έκθεσης ίδιου μεγέθους με ταχύτητα πληκτρολόγησης 40 λέξεων το λεπτό, θα έπαιρνε 22 λεπτά. Η Nuance κυκλοφόρησε το Dragon NaturallySpeaking 10.1, που υποστηρίζεται σε περιβάλλον Windows Vista 64-bit, στο τέλος του Μαρτίου του 2009[37]

8.4.7 IBM ViaVoice

Το IBM ViaVoice είναι μια σειρά από προϊόντα λογισμικού αναγνώρισης συνεχόμενης φωνής σχεδιασμένο από την IBM. Η τρέχουσα έκδοση έχει σχεδιαστεί κυρίως για χρήση σε φορητές συσκευές.

8.5 Εφαρμογές

Μερικές από τις πιο γνωστές εφαρμογές που έχει ενσωματωθεί η αναγνώριση ομιλίας είναι οι :

- Αυτοματισμούς σπιτιών
- Αυτόματη μετάφραση
- Ενσωματωμένα συστήματα σε αυτοκίνητα
- Ελικόπτερα
- Ηλεκτρονική υγεία
- Ρομποτικά συστήματα
- Τηλέφωνα

8.5.1 Αυτοματισμούς σπιτιών

Με την εξέλιξη της τεχνολογίας αλλά και της αναγνώρισης ομιλίας δεν θα μπορούσε αυτή η δυνατότητα να λείπει από έναν αυτοματισμού σπιτιού, όταν υπάρχει οικονομικό IC , από την SensoryInc.com , που έχει την δυνατότητα να κάνει αναγνώριση ομιλίας.[29]

8.5.2 Αυτόματη μετάφραση

Από το 2011 η Google έχει προσθέσει στην διαδικτυακή της πλατφόρμα μετάφρασης, την δυνατότητα μετάφρασης κατευθείαν από το μικρόφωνο, χωρίς να χρειάζεται να πληκτρολογήσεις κάποιο κείμενο. [30]

8.5.3 Ενσωματωμένα συστήματα σε αυτοκίνητα

Με την αύξηση των δυνατοτήτων των ενσωματωμένων συστημάτων σε αυτοκίνητα, η φωνητική αναγνώριση είναι μία ελκυστική πρόταση ώστε να προσφέρει στους οδηγούς μία ασφαλής, εύκολη στην χρήση διεπαφή στα αυτοκίνητά τους. Μάλιστα για την βελτίωση της απόδοσης, επειδή είναι προφανές ότι υπάρχει πολύς θόρυβος σε ένα αυτοκίνητο έχουν επιλεχθεί διάφορες τεχνικές πολλών μικροφώνων για να γίνεται βελτιστοποίηση της απόδοσης. [31]

8.5.4 Ελικόπτερα

Στις 22 Ιουνίου του 2007 μία νέα τεχνολογία που επιτρέπει στους πιλότους ελικοπτέρων από φωνητικές εντολές δοκιμάστηκε με επιτυχία στην αεροπορία του Ηνωμένου Βασιλείου. Σχεδιάστηκε για την αντιμετώπιση του προβλήματος των πιλότων που ξοδεύουν πάρα πολύ χρόνο ψάχνοντας μέσα στο πιλοτήριο τι να επιλέξουν ,ένα πρόβλημα που επιδεινώθηκε από την έλευση των πολύπλοκων οθονών πολλαπλών λειτουργιών, η απευθείας εισαγωγής φωνής του QinetiQ συστήματος ενσωματώνει την τεχνολογία αναγνώρισης ομιλίας για να διευκολύνει τον άμεσο έλεγχο ηλεκτρονικού εξοπλισμού αεροσκαφών με την χρήση μικροφώνων.[32]

8.5.5 Ηλεκτρονική υγεία

Με την εισαγωγή της αναγνώρισης ομιλίας στην ηλεκτρονική υγεία προσφέρει στους Ιατρούς την δυνατότητα να έχουν μία εύκολη πλοήγηση σε πολύπλοκους Ηλεκτρονικούς Φακέλους Υγείας σε πολύ λίγο χρόνο. Τους προσφέρει επίσης την δυνατότητα να υπαγορεύουν, επεξεργάζονται και να υπογράφουν Ιατρικές συνταγές μόνο με την ομιλία.[33]

8.5.6 Τηλέφωνα(Siri/ Google Now/ Cortana)

Τον Οκτώβριο του 2011 η Siri ανακοινώνει τον έξυπνο προσωπικό οδηγό για κινητά τηλέφωνα το Siri. Αν και στην αρχή η εταιρία ήθελε να αναπτύξει εφαρμογές για όλα τα λειτουργικά συστήματα για κινητά τηλέφωνα, μετά την αγορά της εταιρίας από την Apple (28-04-10) αυτό ακυρώθηκε. Πλέον το Siri είναι ο πιο διαδεδομένος προσωπικός οδηγός μιας και είναι ενσωματωμένος σε όλα τα κινητά και tablet που βγάζει πλέον η Apple. Βλέποντας η Google την τεράστια απήχηση που είχε το Siri της Apple αποφάσισε να αναπτύξει τον δικό της έξυπνο προσωπικό οδηγό που τον ονόμασε Google Now, το Google Now εμφανίστηκε αρχικά τον Ιούλιο του 2012(ένα χρόνο μετά το Siri) στην Jelly Bean έκδοση του Android και 1 χρόνο μετά(29-04-13) η Google ανακοίνωσε και την έκδοση για iOS . Αξίζει να αναφερθεί και η προσπάθεια της Microsoft με τον Cortana που ανακοινώθηκε τον Απρίλιο του 2014 και εκτιμά ότι θα μπορεί να ανταγωνιστεί τους αντιπάλους της.

8.5.7 Λοιπές συσκευές

Τον Φεβρουάριο του 2013 η Google ανακοινώνει τα Google Glass , ειδικά σχεδιασμένα γυαλιά με μία μικρή οθόνη στο λίγο πάνω από το δεξί μάτι και υποδοχή για φακούς.[35] Τα συγκεκριμένα γυαλιά είχαν έναν ενσωματωμένο υπολογιστή και λειτουργούσε εξολοκλήρου με την χρήση φωνητικής αναγνώρισης. Τον Ιανουάριο του 2015 η Google ανακοίνωσε το τέλους του συγκεκριμένου project. Οι τεχνικές προδιαγραφές του ήταν :

- Android 4.0.4 ή πιο πρόσφατο
- Οθόνη ανάλυσης 640x360
- Κάμερα 5-megapixel, με δυνατότητα καταγραφής βίντεο 720p
- Wi-Fi 802.11b/g
- Bluetooth
- Χώρος αποθήκευσης 16GB (12 GB διαθέσιμα)
- Μνήμη RAM 682MB
- Γυροσκόπιο 3 αξόνων
- Επιταχυνσιόμετρο 3 αξόνων
- Μαγνητόμετρο 3 αξόνων (πυξίδα)
- Αισθητήρας του φωτός του περιβάλλοντος και αισθητήρας εγγύτητας
- Αισθητήρας αγωγιμότητας των οστών

Συμπεράσματα

Το κύριο συμπέρασμα της παρούσας πτυχιακής είναι ότι η φωνή δεν είναι τίποτα περισσότερο από ένα σύνολο κυματιδίων και από αυτά ένα συγκεκριμένο σύνολο μπορεί να αναγνωριστεί μέσα από διάφορες τεχνικές και με συγκεκριμένες παραμέτρους και να χρησιμοποιηθεί με επιτυχία στις εφαρμογές. Πρέπει να λάβουμε υπόψη μας ότι τα κύματα αυτά μπορούν προσαρμόζονται ανάλογα με το περιεχόμενο τους στην εφαρμογή στην οποία γίνεται η αναγνώριση. Υπάρχει η δυνατότητα επιλογής από ένα πολύ μεγάλο αριθμό βάσεων αναπαράστασης ενός σήματος φωνής η οποία καθορίζει την συμπεριφορά της ανάλυσης με χαρακτηριστικά χρόνου ή συχνότητας.

Βιβλιογραφία

[1] Δημήτρης Καρτσακλής :Computer για όλους,¨Βιομετρικά συστήματα αναγνώρισης¨, 1/9/1999, Τεύχος 182.

[2] Κώστας Νάκος, Νίκος Βλασσόπουλος, Γιώργος Ροπόδης : Computer για όλους ,¨Αναγνώριση φωνής¨,Τεύχος 194 1/10/2000.

[3] E. D. Brill, "A simple rule-based part-of-speech tagger," in Proceedings of the third Conference on Applied Natural Language Processing (ANLP'92), Trento, Italy, 1992.

[4] E. D. Brill, "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging," Computational Linguistics, vol. 21, no. 4, pp. 543--566, 1995.

[5] E. D. Brill, "Unsupervised Learning of Disambiguation Rules for Part of Speech Tagging," in Proceedings of the 3rd Workshop on Natural Language Processing Using Very Large Corpora, Massachusetts, USA, 1995.

[6] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," IEEE ASSP Magazine, pp. 4-15, January 1986.

[7] Phil Blunsom (August 19, 2004). Hidden Markov Models.

[8] Weismann, J. & Saloman, R. (1999). «Gesture recognition for virtual reality applications using data glove and neural networks». In Proceedings of the IEEE Conference on Neural Networks, Washington, USA.

[9] Camurri, A. & Volpe, G. (2004). Gesture-based Communication in Human- Computer Interaction. LNAI 2915, Genova, Italy, Springer Verlag.

[10] Barnewell Thomas P III., Kambiz NAyebi, and Craig H. Richardon, “Speech Coding A Computer Laboratory Textbook”, *John Wiley &Sons, .Inc., 1996.*

[11] Γουμενίδης Θεόδωρος, ’’Κωδικοποίηση Φωνής - Κωδικοποιητές Κυματομορφής ”, Φεβρουάριος 2004.

[12] Jeremy Bradbury, ’’Linear Predictive Coding”, December 2000.

[13] Jason Woodard, “Speech Coding”, http://www.ecs.soton.ac.uk/~ipw/index.html

[14] Niranjan Dhanakoti, “Speech Signal Processing”, project report 2002.

[15] Bryan Douglas, “Voice Encoding Methods for Digital Wireless Communications Systems”, Fall 1997.

[16] Eddie L. T. Choy, “Waveform Interpolation Speech Coder at 4 kb/s”, August 1998

[17] Susanna Varho, “New Linear Predictive Methods for Digital Speech Processing”, 2001

[18] Nadim Batri, “Robust Spectral Parameter Coding in Speech Processing”, May 1998

[19] Alan McCree and Jan Carlos De Martin, “A 1.7 Kb/s Melp Coder with

mproved Analysis and Quantization", DSPS R&D. Texas Instruments, Dallas, Texas

[20] Andreas Spanias, "Speech Coding: A Tutorial Review".

[21] Andreas Spanias, "Multimedia Signal Processing Lecture Notes"

[22] Schussler Mare, "Design and Simulation of a Speech Coder for Mobile Communication Systems", Master's Thesis, 1994

[23] Antti Kiviluoto, "Speech Coding Standards"

[24] H. W. Glen Shires, «Web Speech API Specification,» Google Inc, 19 10 2012.

[25] C. M. University, «CMUSphinx Wiki,»: http://cmusphinx.sourceforge.net/wiki/.

[26] Microsoft, «Speech API Overview,»

[27] iSpeech, «Make your Apps Humans,»

[28] Google, «Speech Recognizer,».Available: http://developer.android.com/reference/android/speech/SpeechRecognizer.html

[29] G. V. Ph.D, «Future Technology Developments and Domotics,».Available: ttp://airconditioningpros.co.uk/vanderheiden-miami-pdf-home-assistance/.

[30] Google, «Google Translate, Now With Voice Input,» Google, . Available: http://googlesystem.blogspot.gr/2011/04/google-translate-now-with-voice-input.html

[31] H. Iwamida, «Voice Recognition Technology for Car-Mounted Devices,». Available: http://www.fujitsu-ten.com/business/technicaljournal/pdf/12-2E.pdf.

[32] «QinetiQ speech recognition technology allows voice control of aircraft systems,» . Available: http://www.qinetiq.com/media/news/releases/Pages/qinetiq-voice-control-technology-for-aircraft.aspx.

[33] http://www.nuance.com/for-healthcare/by-solutions/speech-recognition/index.htm.

[34] Sun Microsystem, .Available: http://java.sun.com/developer/Books/jdbc/ch07.pdf.

[35] Google,«AsyncTask,». Available: http://developer.android.com/reference/android/os/AsyncTask.html.

[36] J. Krumm, "Ubiquitous Computing", σελ. 276, Boca Raton ,2009

[37] Wikipedia. "Dragon

[38] NaturallySpeaking",http://en.wikipedia.org/wiki/Dragon NaturallySpeaking, 2009