ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Περίληψη

Κεφ. 6 Η ΡΟΗ ΤΟΥ ΛΟΓΟΥ

από το βιβλίο SPEAKERS MODELING

Το Κεφάλαιο 6 του βιβλίου SPEAKERS MODELING διαπραγματεύεται την Ροή του Λόγου. Γίνεται μια περιγραφή της αναπαράστασης ομιλίας μέσω των διαφόρων συστημάτων σύνθεσης ομιλίας από Ηλ. Υπολογιστές και μαθηματικά μοντέλα, όπως το Η.Μ.Μ.(*Hidden Markov Model).*

Η ανάπτυξη της Τεχνολογίας εστιάζεται στον καθορισμό τριών τύπων ομιλίας ήτοι:

α. Διακριτών Λέξεων,
β. Συνδεδεμένης Λέξεως,
γ. Συνεχούς Ομιλίας.

Η επιμέρους ανάπτυξη καθενός τύπου ξεχωριστά, αναλύει τα ειδικά χαρακτηριστικά του δίνοντας έμφαση στις αναγενόμενες αντιδράσεις που αφορούν την αντιμετώπιση των φαινομένων της **συνάρθρωσης(*coarticulation*) και της ενδιάμεσης λέξης(*cross-word*).**

Ο Έλεγχος ύπαρξης των λεξιλογίων μεσαίου, μεγάλου και μικρού μεγέθους, η σωστή ανάπτυξη φωνημάτων κατά την ροή του Λόγου, είναι μερικές από τις μεθόδους στις οποίες βασίζεται η σύγχρονη έρευνα. Ο αντικειμενικός σκοπός είναι η δυνατή αναγνώριση μέσω του λειτουργικού Η/Υ, του λόγου. Για τον σκοπό αυτό αναπτύχθηκαν δύο μέθοδοι :

α. μέθοδος εσωτ. Λέξεων ,
β. μέθοδος Υπολέξεων & χρήσεως Τριφώνων.

Επεκτείνοντας τις ικανότητες κατηγοριοποίησης των δικτύων συνδετικότητας , δημιουργείται η απαίτηση κατασκευής των νευρωνικών δικτύων ως μέσου ικανού να παρέξει υψηλής ποιότητας αναγνώριση συνεχούς ομιλίας. Η πλέον διαδεδομένη μέθοδος εφαρμογής είναι η ανάπτυξη των Υβριδικών συστημάτων όπως το MS-TDNN.

Τέλος, η συνεχής ομιλία με τα λόγια χαρακτηριστικά της αποβαίνει ένα αρκετά χρήσιμο εργαλείο στην μοντελοποίηση της ομιλίας. Η **χρήση παύσεων(ομιλία *staccato*)** αποτελεί ένα προσόν αλλά και περιορισμό μαζί. Τα αναπτυσσόμενα εμπορικά πακέτα διαθέτουν συνεχή ομιλία της οποίας η συνεχής ακρίβεια στην έκφραση , εξαρτάται από την χρήση των συμφώνων στο σύνολο της ομιλίας.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΤΙΤΛΟΣ

Σελ.

ΕΙΣΑΓΩΓΗ 6

6.1 ΕΙΣΟΔΟΣ ΣΥΣΤΗΜΑΤΟΣ ΔΙΑΚΡΙΤΩΝ ΛΕΞΕΩΝ 8

6.2 ΕΙΣΟΔΟΣ ΣΥΣΤΗΜΑΤΟΣ ΣΥΝΔΕΔΕΜΕΝΩΝ ΛΕΞΕΩΝ 9

6.3 ΣΥΝΕΧΗΣ ΟΜΙΛΙΑ 9

6.3.1 Όρια λέξεων **10**

6.3.2 Φαινόμενο συνάρθρωσης διαπλεκ. Λέξεων **10**

6.3.3 Συστήματα **11**

6.3.3.1 Συστήματα βασισμένα σε λέξεις **11**

6.3.3.2 Συστήματα Υπολέξεων **12**

6.3.3.3 Ακουστικά-Φωνητικά συστήματα **15**

6.4 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ 15

6.5 ΕΠΙΛΕΓΟΝΤΑΣ ΤΗΝ ΡΟΗ ΤΟΥ ΛΟΓΟΥ 19

6.5.1 Σύγκριση συνεχούς ομιλίας με τύπο, Αυθόρμητη ομιλία **19**

6.5.2 Συνεχής ομιλία συγκριτικά με πληροφόρηση **20**

6.6 ΟΜΙΛΙΑ ΔΙΑΚΡΙΤΩΝ ΛΕΞΕΩΝ ΚΑΙ ΣΥΝΔΕΔΕΜΕΝΩΝ ΛΕΞΕΩΝ 21

6.6.1 Ακρίβεια **21**

6.6.2 Ευκολία χρήσης **22**

6.6.3 Εκτίμηση Αποδόσεως **23**

6.7 ΣΥΝΕΧΗΣ ΟΜΙΛΙΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ 23

6.7.1 Ακρίβεια στα Εμπορικά πακέτα **24**

6.7.2 Εύχρηστη χρήση εμπορικών πακέτων **25**

6.7.3 Εκτίμηση απόδοσης εμπορικών πακέτων **25**

ΣΥΜΠΛΗΡΩΜΑ ΕΡΓΑΣΙΑΣ:
΄΄THE HIDDEN MARKOV MODEL΄΄ 26- 52

ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

Αριθμός		*Σελίδα*
6.0	Φασματογράφημα ΄΄do re mi"με παύσεις	7
6.1	Φασματογράφημα "do re mi" άνευ παύσεων	8
6.2	Δίκτυο νευρώνων κρυφού ελέγχου από τον LEVIN	17
6.3	Φωνοτοπικός χάρτης του Kohonen	17

ΓΛΩΣΣΑΡΙ

Λέξη.

Acoustic Representation: Η αναπαράσταση της ομιλίας. Αναφέρεται στην μηχανική ανάλυση & παρουσίαση του σχηματισμού της ομιλίας από εξωτερικούς του ανθρώπουμηχανισμούς .

Coarticulation patterns: Μορφές συνάρθρωσης. Διακρίνονται σε HARD και SOFT.

Discrete Word: Τύπος ομιλίας. Η επικοινωνία με τον τύπο αυτό γίνεται/ επιτυγχάνεται με ευρεία χρήση διαλειμμάτων ομιλίας, οπότε και παρατηρείται η απρόσκοπτη ομιλία και η ανάλυση της από τον επεξεργαστή(processor).

Fenones: Είδος μονάδος υπολέξης το οποίο ορίζεται για την καλύτερη ανάλυση των υποψήφιων (candidate)λέξεων. Το μέγεθος τους είναι μικρότερο(subphonetic) από το του φωνήματος και χρησιμοποιούνται από το τμήμα έρευνας της IBM.

Fuzzy: Ορισμός που δίδεται στην λεπτομερή μορφοποίηση των ορίων κάθε λέξης, κατά την προφορά της και τους ήχους που χρησιμοποιούνται για τον σκοπό αυτό. Είναι δείκτης ασάφειας μεταξύ ΄΄γειτονικών΄΄ λέξεων.

IntrawordContext Dependent Units: Μονάδα ενδιαμέσων λέξεων. Παρέχει πληροφορίες σχετικές με το περιεχόμενο ενδιάμεσα της λέξεως, καθώς και την πολυπλοκότητα και τον ρυθμό ανάπτυξης αυτής κατά την γενική εφαρμογή των λεξιλογίων.

Multi-Stage Searching: Τύπος έρευνας που επιδρά στην ελαχιστοποίηση της πολυπλοκότητας των συστημάτων

Obtrusiveness: Φορτικότητα ή επίδραση, κατά την εφαρμογή ενός μεγέθους. Εκφράζει την επιρροή του εκάστοτε μεγέθους στην αντίστοιχη κάθε φορά διαδικασία.(ΠΑΥΣΗ κατά την ΟΜΙΛΙΑ)

Performance scoring : Διαδικασία αξιολόγησης. Το αποτέλεσμα περιγράφεται σε πίνακα με στοιχεία ~Σωστές αναγνωρίσεις, ~Διαγραφές, ~Εισαγωγές, ~Υποκαταστάσεις

Semiphones: Όμοια με το ΄΄fenone΄΄. Η χρήση του περιορίζεται στον συνδυασμό μεταξύ ΄΄triphones΄΄, ΄΄phonemes΄΄, ΄΄classical diphones΄΄.

Staccato Effect: Φαινόμενο κατά το οποίο απαιτείται η χρήση διακοπών κατάλληλων σε διάρκεια μεταξύ επιλεγμένων τόνων και λέξεων

Test frames: Σχηματισμοί ελέγχου. Χρησιμοποιούνται σε κάθε αλγόριθμο ως μέρος της εισόδου με μορφή που τους επιτρέπει, να συγκρίνουν το περιεχόμενο τους με την υποψήφια για ένταξη λέξη. Αποτελεί μέρος της διαδικασίας που συντελείται ανά επίπεδο στην TWO-LEVEL DYNAMIC PROGRAMMING MATCHING.

Triphones:

Word Candidate: Υποψήφια για χρήση λέξη. Κατά την αναγνώριση ομιλίας συντελείται η επιλογή των λέξεων που θα αναγνωρισθούν και εκτιμηθούν στην τελική σύνταξη του λεξιλογίου.

Word-Modeling Technologies: Τεχνικές πρόπλασης λέξεων. Κατά την εφαρμογή τους εξετάζονται οι συνθήκες αναγνώρισης και χρήσης των λέξεων καθώς επίσης και η οριοθέτηση μεταξύ των λέξεων.

Word-pair grammar: Τεχνική δομής των λέξεων και φράσεων. Πρόκειται για μια προσέγγιση που γίνεται από τα εμπορικά συστήματα αναγνώρισης, προκειμένου να κατηγοριοποιήσουν τις αποδεχόμενες σε κάθε εφαρμογή ακολουθίες λέξεων(word-sequences)

Κεφάλαιο 6
από το βιβλίο ***SPEAKERS MODELING***

Η ΡΟΗ ΤΟΥ ΛΟΓΟΥ

Εισαγωγή

Οι εμπνευστές των συστημάτων αναγνώρισης ομιλίας συχνά υποδεικνύουν την αναγνώριση ομιλίας ως την φυσική σύνδεση μεταξύ Ηλ. Υπολογιστή και ανθρώπου. Η φυσικότητα επισημαίνει την ικανότητα της επικοινωνίας μέσω του συστήματος αναγνώρισης χρησιμοποιώντας έναν συνήθη και γνωστό τρόπο ομιλίας. Οι δύο (2) όψεις της φυσικότητας , αυτή του μεγάλου λεξιλογίου και αυτή της ευμετάβλητης γραμματικής, έχουν εξετασθεί στα κεφάλαια 3&4 . Ένα άλλο σημαντικό περιεχόμενο είναι η ροή του λόγου αυτή καθεαυτή. Ο όρος ΄΄ροή του λόγου΄΄ δεν αποτελεί συνήθη τεχνική ορολογία στην βιομηχανία των συστημάτων αναγνώρισης ομιλίας.

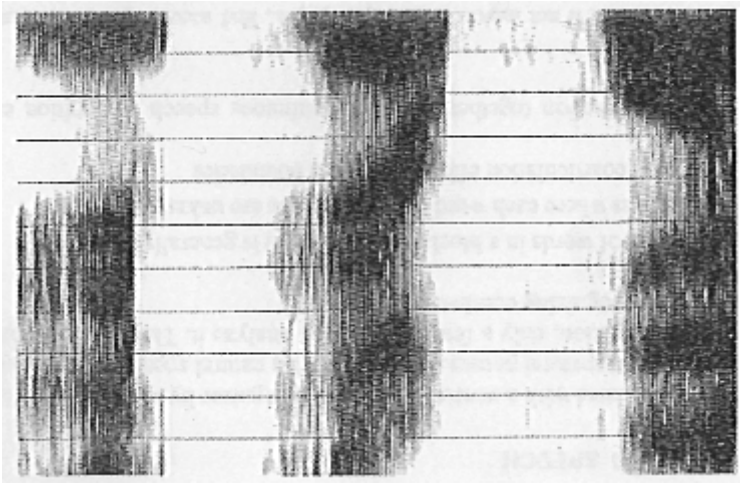
Η χρήση του όρου αυτού στο κείμενο αυτού του βιβλίου είναι απαραίτητη , επειδή δεν υπάρχει γενικά αποδεκτός όρος ο οποίος να χαρακτηρίζει το σύνολο των επιλογών ώστε να περιγραφεί στο παρόν Κεφάλαιο.

Η ροή του λόγου αναφέρεται στο πως ο χρήστης ενός συστήματος αναγνώρισης ομιλίας πρέπει να ομιλεί :

> **Is ¨ the ¨ speaker ¨ required ¨to ¨ pause¨between ¨ words? Or Can the speech be uttered in a natural fashion?**

Φαίνεται παράξενο το ότι τέτοιου είδους ερωτήσεις είναι αναγκαίες να υποβληθούν . Από την στιγμή κατά την οποία ομιλούμε εκφράζοντας σειρές

ανεξάρτητων λέξεων , είναι εύκολο να φανταστούμε ότι η ακουστική αναπαράσταση του λόγου θα περιέχει επίσης φυσικά διαλείμματα μεταξύ ανεξάρτητων λέξεων , όμοια σε μορφή αυτή της φασματογραφικής οθόνης του τρίπτυχου ΄΄ do re mi ΄΄ όπως εμφανίζεται στο σχήμα 6.0. Ατυχώς δεν είναι αυτή η ουσία της υπόθεσης. Όσον αφορά το άκουσμα , οι λέξεις που εκφράζουμε κατά την ροή του λόγου, συγχρόνως, όπως φαίνεται στο φασματικό διάγραμμα του σχήματος 6.1 και όπως θα δούμε στο κεφάλαιο 2, η διαδικασία της αναγνώρισης ολοκληρώνεται μέσω των επιδράσεων από την συναρθρούμενη ομιλία και άλλων διαστρεβλώσεων.



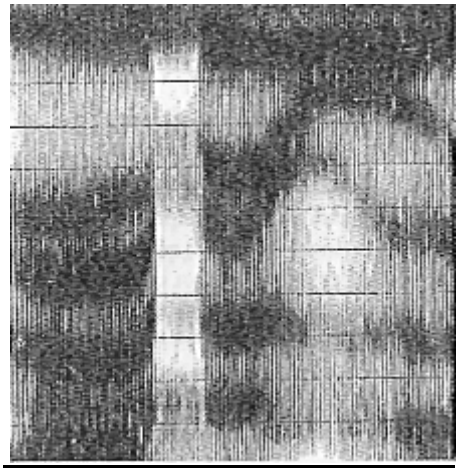
Εικόνα Σφάλμα! Δεν υπάρχει κείμενο καθορισμένου στυλ στο έγγραφο.**-1 φασματογράφημα ΄΄do re mi ΄΄ με παύσεις**

Τα κείμενα που βασίζονται σε αυτού του είδους τα παραδείγματα κυριαρχούν ως προς τις επιλογές ροής ομιλίας εντός των εμπορικών πακέτων συστημάτων αναγνώρισης ομιλίας. Η εστίαση της τεχνολογίας ξεκινά με τον καθορισμό των τριών (3) τύπων ροής ομιλίας οι οποίοι εφαρμόζονται σε εμπορικά συστήματα :

ή Διακριτών λέξεων ή Διακριτού συλλαβισμού
ή Συνδεδεμένης λέξεως
ή Συνεχούςομιλίας

Η μέγιστη προσοχή αποδίδεται στον τύπο της συνεχούς ομιλίας, εμπεριέχοντας συζητήσεις επί των προβλημάτων τοποθέτησης διαχωριστικών των λέξεων και υπερπήδησης των φαινομένων συνάρθρωσης διαπλεκόμενων λέξεων.

Το κύριο σημείο της εφαρμογής καθορίζεται με την απόδοση όσων αναφέρονται στον ΄΄μύθο ΄΄ περί συνεχούς ομιλίας. Έτσι δίδεται η ευκαιρία να περιγραφεί η ακρίβεια και η εύκολη χρήση των οδηγιών συνδυασμένης με διακριτό, συνδεδεμένο ή συνεχή λόγο. Οι τομείς διακριτού και συνεχούς λόγου ολοκληρώνονται με τεχνικές εκτίμησης της απόδοσης των συστημάτων αναγνώρισης ομιλίας.



Εικόνα Σφάλμα! Δεν υπάρχει κείμενο καθορισμένου στυλ στο έγγραφο.**-2 Φασματόγραφημα ΄΄do re mi ΄΄ άνευ παύσεων**

6.1 ΕΙΣΟΔΟΣ ΣΥΣΤΗΜΑΤΟΣ ΔΙΑΚΡΙΤΩΝ ΛΕΞΕΩΝ

Με σκοπό την επίτευξη επικοινωνίας μέσω συστήματος αναγνώρισης ομιλίας διακριτών λέξεων, ένας χρήστης απαιτείται να κάνει χρήση παύσεων μεταξύ των λέξεων. Οι παύσεις μεταξύ των λέξεων εξυπηρετούν δύο(2) σκοπούς :

α. Να παρεμποδίσει την συνάρθρωση στην ομιλία λόγω δυσλεξίας, η οποία διαστρεβλώνει την ακουστική εικόνα της λέξεως προς αναγνώριση
β. Να επιτρέψει στον επεξεργαστή χρόνου να ολοκληρώσει τις αναλύσεις του

Η ανάπτυξη των ταχύτερων επεξεργαστών Ηλ. Υπολογιστών έχει καταστήσει εφικτή την μείωση του μεγέθους των παύσεων. Παραδοσιακά, το μήκος αυτού του είδους των παύσεων ήταν 1/4 του δευτερολέπτου ή και περισσότερο, αλλά ορισμένοι κατασκευαστές έχουν καταφέρει να μειώσουν την τιμή αυτή στο 1/10 του δευτερολέπτου. Οποιαδήποτε και να είναι η διάρκεια της παύσεως, η αναγνώριση διακριτής λέξεως απαιτεί μια καθαρή αρχή και λήξη για οποιαδήποτε λέξη ή φράση.

Η εξέλιξη της αναγνώρισης όπως παρουσιάστηκε σε προηγούμενα κεφάλαια, ειδικότερα στο κεφάλαιο 3 όπου παρουσιάστηκαν οι μέθοδοι μοντελοποίησης λέξης, περιγράφουν τις βασικές αρχές της αναγνώρισης διακριτής λέξης. Η αναγνώριση με την μέθοδο διακριτής λέξης χρησιμοποιείται τόσο με περιγράμματα όσο και με **μη διακρινόμενα (hidden) μοντέλα *Markov*(HMM΄s).** Όταν τα HMM΄S χρησιμοποιούνται σε συστήματα αναγνώρισης ομιλίας με την μέθοδο της διακριτής λέξεως τα αρχικά και τα τελικά εδάφια κάθε μοντέλου αναφοράς HMM δυνατόν να περιέχει άλλα μοντέλα σιγής και/ή θορύβου περιβάλλοντος. Το περιεχόμενο αυτών των εδαφίων διευκολύνει η αναγνώριση των ορίων λέξεων.

Η είσοδος της μεθόδου διακριτής λέξεως δύναται να χρησιμοποιηθεί σε

λεξιλόγιο οποιουδήποτε μεγέθους, με όλες τις φόρμες των μοντέλων ομιλίας, και μέσα σε οποιοδήποτε τύπο περιβάλλοντος ομιλίας. Αυτό απαιτείται δια ζώσης σε όλες τις εφαρμογές υπαγόρευσης.

6.2 ΕΙΣΟΔΟΣ ΣΥΣΤΗΜΑΤΟΣ ΣΥΝΔΕΔΕΜΕΝΗΣ ΛΕΞΕΩΣ

Ο όρος ομιλία συνδεδεμένης λέξεως (*connected-word speech*) χρησιμοποιείται για την αναφορά σε δύο διαφορετικούς τύπους ροής ομιλίας. Η μία από τις έννοιες που καλύπτει αυτός ο τύπος της ομιλίας, είναι συνώνυμη με τον συνεχή λόγο . Αυτή η κατηγορία συνδεδεμένης ομιλίας θα εξετασθεί στην παράγραφο 6.3 .

Η δευτερεύουσα έννοια αναφέρεται σε μία ροή του λόγου που απαιτεί από τον ομιλητή να εισάγει μια στιγμιαία διακοπή μεταξύ των λέξεων. Μέχρι πρόσφατα, αυτές οι διακοπές είχαν διάρκεια κατά μέγιστο πενήντα milliseconds , αποδίδοντας έτσι στον λόγο συνδεδεμένης λέξεως το λεγόμενο *staccato effect.* Μερικοί ερευνητές έχουν επιτύχει να περιορίσουν το διάστημα παύσεως αλλά , όπως και στην αναγνώριση διακριτής λέξεως, η ομιλία συνδεδεμένης λέξεως στηρίζεται στις παύσεις ώστε να περιορίσει την συνάρθρωση διαπλεκόμενων λέξεων. Λόγω αυτού, ο τρόπος ομιλίας με εφαρμογή του *staccato effect* παραμένει απαραίτητος.

6.3 ΣΥΝΕΧΗΣ ΛΟΓΟΣ

Ένας ομιλητής επικοινωνεί με τον ακροατή συνεχούς λόγου μέσω ομιλίας σε φυσική ροή χωρίς αφύσικες παύσεις. Αν και ο τρόπος φυσικής ομιλίας είναι το επιδιωκόμενο αποτέλεσμα της αναγνώρισης ομιλίας , εν τούτοις μερικές συσκευές μόνο έχουν την δυνατότητα ανάλυσής του. Οι κύριες προκλήσεις οι οποίες αποσκοπούν στην αναγνώριση συνεχούς λόγου είναι:

¨Ο αριθμός των λέξεων ανά σύνολο εισαγωγής ομιλίας είναι γενικά άγνωστος

¨Οι θέσεις κατά τις οποίες μια λέξη ξεκινά και τελειώνει είναι άγνωστες

¨ Φαινόμενα διαπλεκόμενης ομιλίας επιδρούν στην οριοθέτηση των λέξεων

Αυτές οι τρείς (3) επισημάνσεις ενεργούν μαζί κάνοντας την αναγνώριση συνεχούς λόγου εξαιρετικά δύσκολη.

> «Είναι συχνά δύσκολο , αν όχι ακατόρθωτο, να καθορισθούν επακριβώς (Π.χ. να βρεθούν αυτόματα και με ακρίβεια) τα όρια των λέξεων λόγω της διαπλοκής του ήχου της ομιλίας. Τοιουτοτρόπως, για παράδειγμα , το όριο μεταξύ του ψηφίου 3 και του ψηφίου 8 στην [ψηφιακή ακολουθία 238] είναι απροσδιόριστο διότι ο ήχος τέλους IiI [iy] στο 3 διαπλέκεται με τον αρχικό ήχο IeI IeyI στο 8,

όταν πρόκειται για την έκφραση της Αγγλικής γλώσσας» (Lawrence Rabiner& Biing-Hwang Juang, Senior Researchers, AT&T Bell Labs Fundamental of speech Recognition 1993 σελ. 392)

Οι δύο πρώτες επισημάνσεις που αναφέρονται προηγούμενα μπορεί να συγχωνευθούν στο βασικό πρόβλημα προσδιορισμού της θέσης των ορίων λέξεων, στην ρήμη του λόγου, όπως απεικονίζεται αυτή η διεργασία στο σχήμα 6.1. Το τελευταίο πρόβλημα αναφέρεται στην διαπίστωση όταν γειτονικά σύμφωνα-φωνήεντα επιδρούν το ένα στο άλλο ακόμα και αν εμφανίζονται σε διαφορετικές λέξεις.

6.3.1 Όρια Λέξεων

Αντίθετα από την εισαγωγή διακριτής λέξεως και συνδεδεμένης λέξεως τα κυκλώματα αναγνώρισης συνεχούς λόγου δεν μπορούν να περιγράψουν ξεκάθαρα λέξεις που να είναι κατάλληλες για ανάλυση . Τουναντίον, ο κύριος σκοπός της αναγνώρισης ομιλίας έχει διαρθρωθεί ώστε να περιέχει την εκτίμηση μεταξύ αντιφατικών υποθέσεων σε συνδυασμό με τον αριθμό των λέξεων σε λεγόμενα και τις θέσεις τους. Σε κάθε ψηφιοποιημένο σκελετό(κεφ. 2, παρ. 2.2.2) το σύστημα αναγνώρισης θα πρέπει να καθορίζει κατά πόσο έχει φθάσει στο όριο μιας λέξης . Στο σχήμα 6.1 , για παράδειγμα, το σύστημα αναγνώρισης απαιτείται να αποφασίσει κατά πόσο η πρώτη λέξη είναι η νότα «do» ή η λέξη «door» ή και ακόμα η λέξη «during». Δεν υπάρχει τίποτα στο σήμα από μόνο του, το οποίο να υποδεικνύει την σωστή επιλογή και υπάρχει σημαντικός αριθμός παραδειγμάτων από περιγράμματα ομιλίας στα οποία προκαλείται σύγχυση, όπως αυτό παρατηρείται σε ανθρώπους ακροατές. Η φράση «Grey tape» για Παράδειγμα, είναι περίπου ίδια σε άκουσμα με την φράση «Great ape» η φράση «An ice chest» θα μπορούσε σχεδόν να ακουστεί ως «A nice chest» καθώς και το ακουστικό περίγραμμα της φράσης «How to wreck a nice beach» με αυτό της φράσης «How to recognize speech».

Εάν όλες οι λέξεις σε ένα λεξιλόγιο εκτιμώνται ως ενδεχόμενες πιθανές λέξεις (*word candidate)*ουσιαστικά σε κάθε τμήμα της εισαγωγής, τότε ο αριθμός των υπολογισμών που απαιτείται για να ολοκληρωθεί η αναγνώριση συνεχούς ομιλίας θα μπορούσε εύκολα να καταβάλλει ένα αξιόπιστο Η/Υ , ακόμα και εάν πρόκειται για μικρού μεγέθους λεξιλόγια. Συνεπώς, τα συστήματα αναγνώρισης διαθέτουν πλήθος τεχνικών μείωσης του αριθμού των υποψήφιων λέξεων καθώς επίσης τεχνικές καθορισμού θέσεως ορίων λέξεως. Μερικά από αυτά έχουν ήδη περιγραφεί κατά την ανάλυση της δομής εφαρμογής(ιδέ Κεφ. 4). 'Αλλα είδη θα περιγραφούν στο τμήμα 6.3.3 .

6.3.2 Φαινόμενο Συνάρθρωσης Διαπλεκόμενων λέξεων

Υπάρχουν δύο (2) τύποι φαινόμενου Διαπλεκόμενης ομιλίας(συνήθως καλούνται φαινόμενα δυσλεξίας από διαπλεκόμενη ομιλία)τα οποία οι ερευνητές πρέπει να τα διατηρούν ως ήπια και σκληρά. Τα ήπια φαινόμενα είναι παρόμοια με τα αντίστοιχα που περιγράφηκαν στο κεφάλαιο 2(τμήμα 2.1.3). Αυτά

είναι ήσσονος σημασίας, αναμενόμενες μεταβολές που συχνά επιδρούν με θόρυβο (αλλαγή του φωνητικού ΄΄t΄΄ της λέξης ΄΄wait΄΄ σε ΄΄d΄΄ , όταν αυτή χρησιμοποιείται στην πρόταση ΄΄wait a minute΄΄) συριγμού ή πλαταγίσματος των χειλιών (φωνητικό ΄΄k΄΄ στην λέξη ΄΄cool΄΄).

Τα σκληρά φαινόμενα συνάρθρωσης ομιλίας είναι περισσότερο ακραία. Γενικά επικεντρώνουν την ενεργητικότητά τους στην διαγραφή των μεμονωμένων φωνημάτων(όπως το ΄΄t΄΄ στο τέλος της λέξης ΄΄what ΄΄ στην πρόταση ΄΄what time is it?΄΄), μεταβολή φωνημάτων (όπως η αποκοπή του φωνητικού ΄΄d΄΄ στην λέξη ΄΄did΄΄ ακολουθούμενης από την λέξη ΄΄you΄΄ όπου και εκεί συνίσταται αποκοπή του ΄΄y΄΄ και μετατροπή τους σε ΄΄jh΄΄ στην φράση ΄΄Did you know?΄΄) ή ένας συνδυασμός διαγραφής και μεταβολής (όμοια με την ΄΄what do you want ?΄΄ στην οποία επέρχεται η αλλαγή της σε ΄΄wah ju wan?΄΄).

Τα φαινόμενα συνάρθρωσης ομιλίας εμφανίζονται αυξημένα στις λειτουργικές; λέξεις όπως ΄΄the΄΄, ΄΄and΄΄, ΄΄do΄΄, και ΄΄to΄΄. Οι λειτουργικές λέξεις είναι τυπικά μονοσύλλαβες ομιλούμενες χωρίς έμφαση . Επηρεάζονται τόσο έντονα στην δυσλεξία από την συνάρθρωση ώστε γίνονται μη αναγνωρίσιμες Τα πιο κοινά παραδείγματα είναι οι λέξεις ΄΄wanna΄΄ και ΄΄gonna΄΄ οι οποίες απορροφούν την λειτουργική λέξη ΄΄to΄΄ , αν και παρόλα αυτά όλες οι λειτουργικές λέξεις είναι τρωτές στην διαγραφή /διαστρέβλωση της ηχητικής απόδοσης. Περισσότερες πληροφορίες στα ήπια και σκληρά φαινόμενα διαπλοκότητας ομιλίας δυνατόν να βρεθούν στα βιβλία των Giachin (1990) , Rabiner.

6.3.3 Συστήματα

Τα συστήματα συνεχούς ομιλίας (C*ontinuous speech*) διαφέρουν στην χρήση των ορίων λέξεων και της συνάρθρωσης ομιλίας. Οι μέθοδοι που χρησιμοποιούν γενικά αντανακλούν το μέγεθος των λεξιλογίων τους και των μονάδων ανάλυσης που χρησιμοποιείται για αναγνώριση (ιδέ Κεφ. 3). Οι προσεγγίσεις του πεδίου αυτού γενικά ομαδοποιούνται ως εξής:

¨ **Συστήματα που βασίζονται στις λέξεις (μικρό & μεγάλο μέγεθος λεξιλογίου)**

- **Συστήματα Υπολέξεων/ τριφωνία(μεγάλο λεξιλόγιο)**
- **Συστήματα Ακουστικά Φωνητικά**

Ο Jelinek το 1976 και οι Rabiner & Juang το 1993 παρέχουν μια τεχνική αναφορά και ο Kaisse το 1985 γνωστοποιεί μια τεχνική γλωσσική προοπτική στο αντικείμενο που αποσκοπεί στην πρόκληση αναγνώρισης ταχέως συνεχούς ομιλίας.

6.3.3.1 Συστήματα βασισμένα σε λέξεις (μικρό & μεγάλο λεξιλόγιο) Μια αποτελεσματική και ενεργός τεχνική που εφαρμόζεται για τον καθορισμό των ορίων λέξεων σε μικρά και μεσαία μεγέθη συστημάτων, είναι η χρήση μιας δομικής τεχνικής όπως η γραμματική(ιδέ κεφ.4). Αυτού του είδους η προσέγγιση χρησιμοποιείται από τα περισσότερα εμπορικά πακέτα αναγνώρισης. Μια τεχνική(*grammar*) πεπερασμένης κατάστασης ή μια τεχνική ζεύγους λέξεων θα εξειδικεύσει τις επιτρεπτές ακολουθίες λέξεων σε μια εφαρμογή. Παραπλήσιες

λέξεις μπορούν στην συνέχεια να επιδοθούν ως μονάδες ώστε **να περιέχουν ήπια και σκληρά φαινόμενα συνάρθρωσης.** Καθώς εξελίσσεται η διαδικασία εισαγωγής, μια γραμματική ταυτοποιεί το σημείο έναρξης της επόμενης μη ταυτοποιημένης λέξεως.
Αρκετές έρευνες και εκτιμήσεις των καταστάσεων χρησιμοποιούνται για την οργάνωση των υπολογιστικών απαιτήσεων του συνεχούς λόγου έτσι ώστε ο λόγος να δύναται να αναγνωρισθεί μέσω του λειτουργικού προγράμματος ενός PC, οπωσδήποτε εντός ενός στενά περιορισμένου διαστήματος χρόνου , με συνεκτίμηση των παρακάτω:

¨Ενarμόνιση δύο επιπέδων δυναμικού προγραμματισμού´´
¨ Αλγόριθμο κατασκευής επιπέδου´´
¨Αλγόριθμο ενός επιπέδου ´´

Η **εναρμόνιση δύο επιπέδων δυναμικού προγραμματισμού** και ο **Αλγόριθμος κατασκευής επιπέδου** μειώνουν τις απαιτήσεις σε λογισμικό Η/Υ με τον καταμερισμό της ερευνητικής διαδικασίας σε δύο(2) βήματα, τα οποία καλούνται επίπεδα. Σε κάθε αλγόριθμο , το πρώτο επίπεδο αποτελείται από λειτουργίες επιλογής τυχαίων πλαισίων εισαγωγής ώστε να χρησιμοποιηθούν ως πλαίσια ελέγχου(*test frames*). Τα δημιουργούμενα πλαίσια ελέγχου κάθε φορά συγκρίνονται με την υποψήφια λέξη. Στο δεύτερο επίπεδο , οι υποψήφιοι με την μεγαλύτερη βαθμολογία από το πρώτο επίπεδο συνενώνονται ώστε να αποτελέσουν ένα συνεχές *string*-αλληλουχία λέξεων. **Ο αλγόριθμος ενός επιπέδου** αποδίδει το πρώτο και δεύτερο βήμα μαζί και στην συνέχεια οπισθοδρομεί για να βρεί τις καλύτερες βαθμολογίες.
Οι Rabiner &Juang το 1993 (κεφ.7) παρέχουν λεπτομερείς περιγραφές των μεθόδων αντιμετώπισης των προβλημάτων οριοθέτησης λέξεων, Οι αυθεντικο΄θ τρόποι τυποποιήσεως των αλγόριθμων έρευνας που περιγράφονται σε αυτό το τμήμα, περιλαμβάνονται στον Sakoe (1979) και εστιάζεται στην εναρμόνιση του προγραμματισμού δύο επιπέδων, επίσης στους Myer & Rabiner (1981) όσον αφορά τον αλγόριθμο δομής επιπέδου και στον Ney (1984) για τον αλγόριθμο μιας διέλευσης.

6.3.3.2 Συστήματα Υπολέξεων /τριφωνίας(μεγάλο μέγεθος λεξιλογίου) Τα προβλήματα που εμφανίζονται κατά την αναγνώριση συνεχούς λόγου συνθέτονται όταν ο αριθμός των υποψήφιων λέξεων που πρέπει να εκτιμηθούν είναι μεγάλος. Η έλλειψη των εμπορικών συστημάτων αναγνώρισης , μεγάλου λεξιλογίου είναι απόδειξη ότι το πρόβλημα δεν έχει επιλυθεί ικανοποιητικά. Όπως και στα συστήματα που βασίζονται σε λέξεις, οι τεχνικές πεπερασμένης κατάστασης (ιδέ κεφ. 4 ,τμήμα 4.2) δυνατόν να χρησιμοποιηθούν για να περιορίσουν τις επιλογές λέξεων. Ατυχώς οι διευκρινιστικές τεχνικές , όπως αυτές της σταθερής κατάστασης και ζεύγους λέξεων, υπόκεινται σε δύο(2) σημαντικούς περιορισμούς όταν χρησιμοποιούνται σε συστήματα μεγάλων λεξιλογίων, όπως παρακάτω:

α. Είναι παρα πολύ περιοριστικοί σε μερικά συστήματα μεγάλων λεξιλογίων, ειδικά σε συντακτικό ελεύθερης φόρμας.

Β. Εάν το ενεργό λεξιλόγιο φθάνει και δέκα χιλιάδες, η τεχνική (γραμματική) λίγο έχει να κάνει ώστε να μειώσει την πρόκληση αναγνώρισης συνεχούς λόγου.
Άλλες μέθοδοι απαιτούν οι χρήστες να έχουν την δυνατότητα της σύνταξης.

Τα συστήματα με μεγάλο λεξιλόγιο βασίζονται στις μονάδες υπολέξεων περισσότερο παρά στις ολόκληρες λέξεις. Από την στιγμή που ο συνολικός αριθμός των μονάδων υπολέξεων είναι αρκετά μικρότερος από τον αριθμό των λέξεων σε ένα σύστημα , είναι εμφανές ότι τα προβλήματα που αντιμετωπίζει η αναγνώριση συνεχούς λόγου , αυξάνονται. Με την συλλογιστική αυτή, μερικοί ερευνητές πρόσθεσαν μοντέλα τριφωνίας στην διαστρέβλωση από συνάρθρωση ομιλίας στα συστήματα που οι ίδιοι είχαν αναπτύξει και συμπεριέλαβαν ειδικά μοντέλα για λειτουργικές λέξεις.

Αυτοί κατέληξαν ότι:

> « Κατά την χρήση μονάδων ομιλίας υψηλής πιστότητας (ανεξάρτητες του περιβάλλοντος) ,συμπεριλαμβανομένων τόσο των μονάδων εξωτερικά των λέξεων όσο και εσωτερικά, η πολυπλοκότητα της συνολικής εφαρμογής αυξάνεται δραματική ανάλογα με τον αριθμό των βασικών μονάδων...Μια εφαρμογή ολικής έρευνας... είναι συνολικά μη πρακτική, εάν όχι απίθανη(Roberto Pieraccini, Chin-Hull Lee, Egidio Giachin& Lawrence Rabiner, Researchers , AT&T Bell Labs, Μείωση της πολυπλοκότητας σε ένα σύστημα αναγνώρισης ομιλίας μεγάλου λεξιλογίου, 1991, σελ.729)

Αντίθετα με τα μοντέλα τριφωνίας εξωτερικών λέξεων ο αριθμός των μοντέλων εσωτερικών λέξεων είναι τεράστιος:

> Υπάρχουν 2381 τριφωνίες σε λέξεις στον στόχο μας των 997 λέξεων . Αλλά υπάρχουν 7057 τριφωνίες οι οποίες όταν βρίσκονται ανάμεσα στις τριφωνίες λέξεων συνίστανται εξ ίσου (Kai-Fu Lee, Hsiao-Wuen Hon, Mei-Yuh Hwang&Sanjoy Mahajan, Carnegie Mellon University[Lee is currently at Apple Computer Corporation] , ΄΄Recent progress and future outlook of the SPHINX speech recognition system΄΄ 1990,σελ.60)

Ο λόγος είναι ότι η ακουστική πληροφορία για την μια πλευρά της διαστρέβλωσης τριφωνίας είναι άγνωστη. Μια τέτοιου είδους τριφωνία για το τέλος των λέξεων όπως ΄΄plain΄΄ και ΄΄stain΄΄ θα περιέχει ακουστική πληροφορία κατά την μετάβαση από το φωνήεν στο τελικό *n* στην μία κατάσταση του HMM. Θα περιέχει επίσης ακουστικά δεδομένα για το *n* στην δεύτερή του κατάσταση, αλλά από την άλλη δεν θα διαθέτει πλέον κανένα οδηγό για την επιλογή ακουστικού υλικού στην Τρίτη κατάσταση από την στιγμή που ένας μεγάλος πίνακας λέξεων θα ακολουθεί το ΄΄΄plain΄΄ ή το ΄΄stain΄΄.

Μερικοί ερευνητές έχουν προτείνει μεθόδους για την μείωση του αριθμού των μονάδων υπολέξεων. Μερικές προτάσεις υιοθετούν μια από τις επόμενες κύριες στρατηγικές:

¨ Χρήση άλλων μονάδων πλήν τριφωνιών

¨ Χρήση συνόλων από μονάδες υπολέξεων

Ενας αριθμός μονάδων υπολέξεων, όπως τα *semiphones* και τα *fenones* χρησιμοποιούνται κατά κόρον . Αυτές οι μονάδες είναι μεγαλύτερες , μικρότερες ή απλά διαφορετικές από τα τρίφωνα. Τα ***semiphone*s** (ανακαλύφθηκαν και προτάθηκαν από τον Doug Paul του εργαστηρίου Lincoln στο M.I.T.) ,για παράδειγμα συνδυάζουν τρίφωνα εξαρτώμενα περιεχόμενου , φωνήματα ανεξάρτητα περιεχομένου και κλασικά δίφωνα(τα μοντέλα των κλασικών διφώνων ξεκινούν στο κέντρο του ενός φωνήματος και τελειώνουν στο κέντρο του επόμενου).Τα *fenones* είναι υποφωνητικά(μικρότερα του φωνήματος) ως μονάδες ,υποστηρίχθηκαν δε από την **IBM και χρησιμοποιήθηκαν στην έρευνα της πάνω στην συνεχή ομιλία, τόσο καλά όσο και το σύστημα διακριτών λέξεων TANGORA. Όπως και στα τρίφωνα , όλες αυτές οι μονάδες υπολέξεων αναπαριστώνται με την χρήση HMM'ς.**

Η μοντελοποίηση γενικευμένου τριφώνου (αναπτύχθηκε από την CMU) τόσο για περιβάλλον ενδιάμεσων όσο και για περιβάλλον μεταξύ λέξεων είναι η πιο δημοφιλής τεχνική ομαδοποίησης. **Ένα γενικευμένο τρίφωνο δημιουργείται από ομαδοποίηση και κατηγοριοποίηση των ακουστικών δεδομένων από περιβάλλον για ένα απλό φώνημα. Τα ακουστικά περιεχόμενα *s, th, f* φια παράδειγμα, πιθανόν να ευρίσκονται μεταξύ των συνθλιβέντων σε γενικευμένη τριφωνία για το φωνήεν *aa* (όπως και στην λέξη *΄΄sock΄)* . Αν και τα αποτελεσματικά τρίφωνα είναι περισσότερο δημιουργικά, οι ερευνητές της CMU έχουν ανακαλύψει ότι η ακρίβεια αναγνώρισης δεν κλιμακώνεται ακόμα καο αν πρόκειται για πολύ καλά προετοιμασμένο μοντέλο(ιδέ κεφ. 3, τμήμα 3.1.5)**

Η πολυεπίπεδη έρευνα , όπως ο αλγόριθμος νιοστών όρων(ιδέ κεφ.3, τμήμα 3.2.2 και την σημείωση στο 3.2.2), μπορεί να μειώσει την αύξηση όσον αφορά την πολυπλοκότητα για τα συστήματα με μοντέλα τριφώνου διασταυρούμενων λέξεων(*cross-word)* κατά τρόπο ώστε:

¨**Επίπεδο 1**: Η βέλτιστη ακολουθία λέξεων επιτυγχάνεται χρησιμοποιώντας μοντέλα ενδιάμεσων διασταυρούμενων λέξεων που διαθέτουν τριφωνία μόνο

¨**Επίπεδο 2**: Τα αποτελέσματα του επιπέδου 1 , επανεκτιμώνται χρησιμοποιώντας μοντέλα τριφωνίας διασταυρούμενων λέξεων μόνο

Η αποτελεσματική έρευνα και η εκτίμηση όλων των μοντέλων και τεχνικών στηρίζονται σε καλά μοντέλα. Παράλληλα με την ανάπτυξη όλων των μοντέλων τριφωνίας, είναι δύσκολο να καθορισθεί ένα αρκετά μεγάλο δείγμα για κάθε φαινόμενο συναρθρούμενης διασταυρούμενης λέξεως ώστε να δομηθεί το καλό μοντέλο. Ευτυχώς, το γεγονός των λιγότερων προσεγγίσεων με ευαίσθητα δεδομένα μπορούν να χρησιμοποιηθούν στην διαχείριση μερικών φαινομένων συνάρθρωσης διασταυρούμενων λέξεων . Τα ήπια φαινόμενα αυτού του τύπου είναι όμοια με διαδρομές συναρθρούμενης - διασταυρούμενης λέξεως δίνοντας έτσι την δυνατότητα να χρησιμοποιηθούν μοντέλα εν ισχύ για να τα αναπαραστήσουν . Τα περισσότερα φαινόμενα συνάρθρωσης ομιλίας μπορούν να προβλεφθούν από τα φωνητικά περιβάλλοντα τους. Μερικοί ερευνητές έχουν

κατασκευάσει μοντέλα τριφώνου διασταυρούμενης λέξεως με χρήση φωνητικών κανόνων που λειτουργούν για αυτά.

Ο Paul (1991) περιγράφει τα ημίφωνα(*semiphones*) του εργαστηρίου Lincoln. Κοιτώντας στον Bahl(1988) , υπάρχουν πληροφορίες για *fenones.* Ο Kai-Fu Lee (1990b) επεξηγεί τα γενικευμένα τρίφωνα. Στον Swartz(1992) δίδονται η περιγραφή και εξηγήσεις για το πως οι Bolt Beranek &Newman (BBN) χρησιμοποιούν το παράδειγμα N-best στο σύστημα BIBLOS για να διαχειριστούν την συνάρθρωση διασταυρούμενης λέξεως.

6.3.3.3 Ακουστικά-Φωνητικά (μεγάλο λεξιλόγιο) Συστήματα.Τα ακουστικά - φωνητικά συστήματα ή μοντέλα είναι γενικά δομημένα σε μοντέλα ανεξάρτητα περιβάλλοντος λειτουργίας(ιδέ κεφ. 3, τμήμα 3.1.4). Από την στιγμή που ο αριθμός των φωνημάτων σε μια λέξη είναι γενικά μικρός και σταθερός, τα ακουστικά-φωνητικά συστήματα δεν παρέχουν από την δημιουργία τους αναπαραστάσεις μονάδων και οι οποίες συνδέονται με συστήματα τριφώνου.

Το κύριο πρόβλημα που αντιμετωπίζουν τα ακουστικά-φωνητικά συστήματα είναι η απουσία ομαδοποιημένης πληροφορίας. Μερικοί αναλυτές διευθύνουν το πρόβλημα με την κατασκευή ειδικών μοντέλων υπολέξεων για την διαχείριση σιωπής και συνάρθρωσης ομιλίας, διασταυρούμενης λέξεως. Όπως και οι κατασκευαστές συστημάτων τριφώνων ρισκάρουν να δημιουργήσουν πιθανή έκρηξη των συστημάτων.

Η εταιρεία **Speech Systems Inc.(SSI) κατανόησε το πρόβλημα που δημιουργούν φαινόμενα συνάρθρωσης ομιλίας διασταυρούμενης λέξεως συνδιάζοντας μια γραμματική σταθερού σημείου με την επεξεργασία φωνητικών δύο επιπέδων**. Στο πρώτο επίπεδο , ένα δένδρο αποφάσεων χρησιμοποιείται για τον σχηματισμό τμημάτων που βασίζονται στην ευρεία κατηγοροποίηση φωνητικών. Στο δεύτερο επίπεδο, η αλυσίδα τμημάτων κωδικοποιείται μέσω ενός άλλου δένδρου αποφάσεων και χρησιμοποιείται ως εισαγωγή στα φωνητικά HMM'S.

Άλλοι ερευνητές εισήγαγαν πληροφορίες στατιστικής σχετικά με τα μοντέλα φωνημάτων αυτοπροσώπως στο περιβάλλον τους.
Οι μονάδες φωνημάτων δεν επεκτείνονται αδρά στα γειτονικά φωνήματα. Όπως και να έχει πάντως, εξωτερικεύοντας η εξάρτηση περιβάλλοντος, είναι δυνατόν να μοντελοποιηθεί εύκολα από τον συνδυασμό περιβαλλοντικώς πληροφοριών στον ακουστικό δρομέα (Herman Ney, University of Technology, AACHEN GERMANY & ANDREAS NOU, FOUNDEX, ASPECT GMbH. «Acoustic-phonetic modeling in the SPICOS system» 1994, p.312)
Τα ομοίου τύπου συστήματα δυνατόν να περιγράφουν ως συστήματα ημιανεξάρτητα του περιβάλλοντος. Ο Meisell (1991) παρέχει επιπρόσθετες πληροφορίες στα συστήματα SSI. Ο Fissore (1989) περιγράφει ένα σύστημα με ειδικά μοντέλα για διαπλοκή ομιλίας διασταυρωμένης λέξεως.

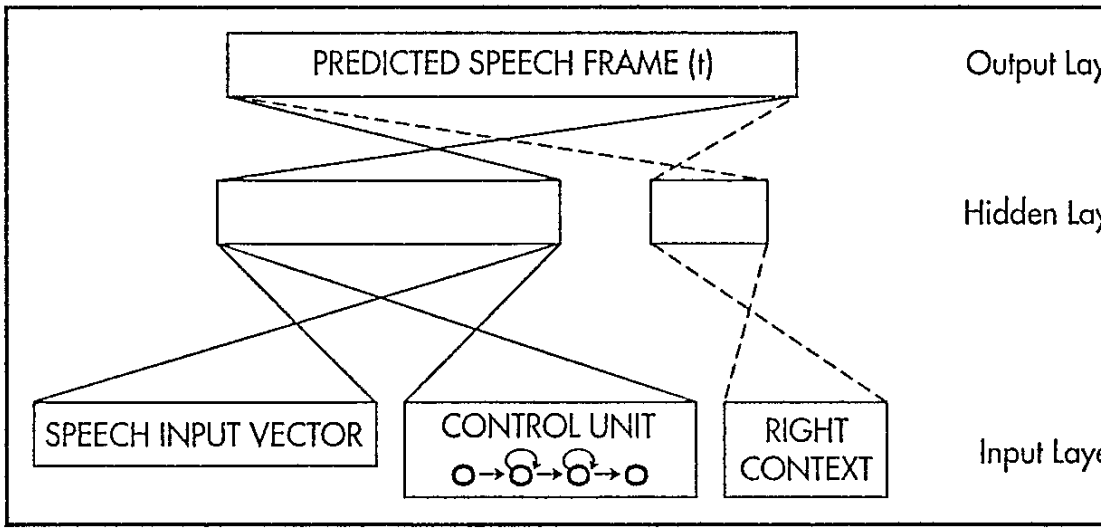
6.4 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Εκτείνοντας τις ικανότητες κατηγοριοποίησης των δικτύων συνδετικότητας στην αναγνώριση συνεχούς ομιλίας είναι μια ενδιαφέρουσα κατεύθυνση έρευνας στην αναγνώριση ομιλίας (PATRICK HAFFNER, CNET LANNION A TSS/RCP και Michael FRANZINE & ALEX WAIBEL, CARNEGIE MELLON UNIVERSITY, «Integrating time alignment and neural networks for high performance continuous speech recognition 1991 p.105). Στην απαίτηση κατασκευής και σχεδιασμού νευρωνικών δικτύων ικανών να παρέχουν υψηλής ποιότητας αναγνώριση συνεχούς ομιλίας, ένας κατασκευαστής / σχεδιαστής θα πρέπει να διαθέτει την δυνατότητα να αναπαράγει την άψογη κατηγοριοποίηση, δυνατότητα των νευρωνικών δικτύων όταν θα πρέπει να αντιμετωπιστεί η εσωτερική αδυναμία των στην διαχείριση παραλληλισμού χρόνου (Κεφ. 2, τμήμα 2.5)

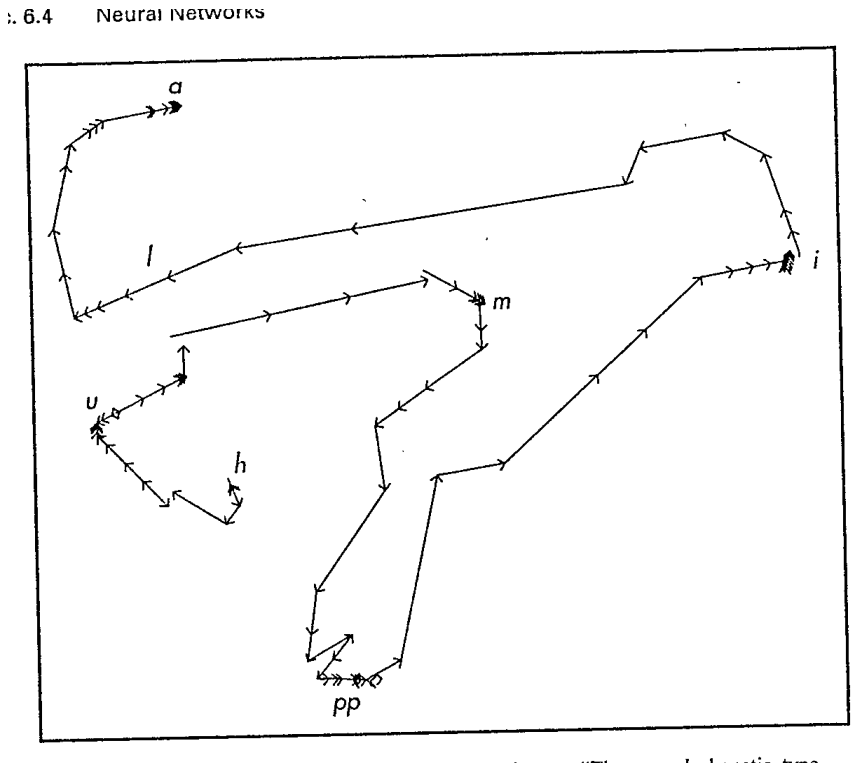
Η πλέον δημοφιλής προσέγγιση της αναγνώρισης συνεχούς ομιλίας είναι η σχεδίαση υβριδικών συστημάτων. Επιπρόσθετα Δε είναι και η πλέον επιτυχημένη επειδή καταγράφει τις δυνατότητες των νευρωνικών δικτύων, σε σχέση με τον ανώτερο χρόνο παραλληλισμού των συμβαντικών τεχνικών, όπως είναι η τεχνική HMM και χρόνος δυναμικού warping (ιδε' κεφ.2 τμήμα 2.3.1).

Επιλύοντας ένα πρόβλημα πραγματικού χρόνου σχεδόν πάντοτε απαιτείται η δεξιότητα ενός ετερογενούς συστήματος… Αυτό είναι τουλάχιστο μερικό διότι η δομή των δύσκολων προβλημάτων είναι τυπικά ετερογενής… Η αυτόματη αναγνώριση ομιλίας (ASR) δεν αποτελεί εξαίρεση (Herve' Bourlard, Faculete Polytechnique de Mons, Belgium and International Computer Science Institute of the University of California at Berkley & Nelson Morgan, International Computer Science Institute, Connectionist Speech Recognition: A Hybrid Approach, 1994 pp 3-4).

Μερικά υβρίδια χρησιμοποιούν το νευρωνικό δίκτυο ως προεπεξεργαστή (post-processor) για ένα σύστημα βασισμένο στην τεχνική HMM, αλλά ένας αυξούμενος αριθμός χρησιμοποιεί ένα δίκτυο ως το κύριο συστατικό αναγνώρισης η ως το πρώτο επίπεδο από μια διαδικασία δύο βημάτων. Το πολυεπίπεδο του CMO TDNN (MS-TOWN) είναι ένα παράδειγμα ενός τέτοιου είδους βασισμένο σε υβρίδια δίκτυο. Το MS-TOWN είναι μια υβριδική έκδοση των συστημάτων TDNN στην μετακίνηση του από την αναγνώριση διάκριτης λέξεως στην συνεχή ομιλία. Οι ρυθμισθείσεις έξοδοι των μοντέλων TDNN μεταδίδονται σε πίνακα χρόνου δυναμικού warping για στιγμιαίο παραλληλισμό. Κατά την διάρκεια που το MS TOWN αποδίδει πολύ καλά στους αρχικούς ελέγχους με βάσεις δεδομένων συνεχούς ομιλίας περιέχουσε έως και τέσσερις χιλιάδες λέξεις, ο έλεγχος έχει επεκταθεί και σε μεγαλύτερα λεξιλόγια.



Εικόνα Σφάλμα! Δεν υπάρχει κείμενο καθορισμένου στυλ στο έγγραφο.**-3**
Δίκτυο Νευρώνων κρυφού ελέγχου απο τον LEVIN



εικόνα 2.2 1 Φωνοτοπικός Χάρτης του KOHONEN (som)

Αλλά υβριδικά συστήματα χρησιμοποιούν ένα πολύ επίπεδο ποσοστιαία δίκτυο (Multi-Layer Perception) για να παρέχουν πρόβλεψη σήματος σε ένα HMM. Η λειτουργία του νευρωνικού δικτύου είναι περισσότερο για να αντιπαραβάλλει ακουστικές παραστάσεις (patterns) παρά για να τις πιστοποιεί (classify). Το Συνδεδεμένο προβλέψιμο νευρωνικό δίκτυο για το σύστημα (MU είναι ένα παράδειγμα που αναφέρεται σε όσα ελέχθησαν ανωτέρω. Περιέχει ένα δίκτυο επανεκκίνησης (feed forward) το οποίο χρησιμοποιεί δεδομένα από προηγούμενα κομμάτια συνεχούς ομιλίας ώστε να παρέχει πρόβλεψη μελλοντικών τιμών σε επεξεργασμένους τομείς. Όταν το επόμενο τμήμα του λόγο έχει ληφθεί, το LPNN βαθμολογεί και διορθώνει την δική του απόδοση. Η έξοδος του LPNN αποστέλλεται σε ένα αλγόριθμο dynamic type warping για time normalization.

Το **Νευρωνικό Δίκτυο Κρυμμένου Ελέγχου (HCNN) του LEVIN, όπως αυτό φαίνεται στο σχήμα 6.2, είναι ένα άλλο παράδειγμα υβριδικού συστήματος** πρόβλεψης. Σε αντίθεση με τα περισσότερα λοιπά υβριδικά, το HCNN λαμβάνει είσοδο από τρείς ξεχωριστές πηγές:

- **Προεπεξεργασμένες ψηφιοποιημένες μορφές δειγμάτων από ανύσματα ομιλίας**
- **Μια ανεξάρτητη μονάδα ελέγχου**
- **Μονάδες contextual**

Η μονάδα ελέγχου χειρίζεται στιγμιαία απόκλιση χωρίς να μεταβάλλει τις υπόλοιπες παραμέτρους δικτύου και η δομή του είναι παράλληλη με αυτή του HMM. Το δίκτυο στο σχήμα 6.2 είναι RIGHT-CONTEXT DEPENDENT επειδή οι contextual μονάδες παρέχουν ακουστική πληροφορία σχετικά με τα φωνήματα στο δεξιό μέρος του τρέχοντος τμήματος. Αυτό το δίκτυο θα μπορούσε επίσης να σχεδιασθεί να παρέχει και το αριστερό κλάδο εξίσου.

Ένα από τα πλέον ασυνήθη υβριδικά συστήματα είναι η **Νευρωνική φωνητική γραφομηχανή του KOHONEN (Kohonen's Neural Phonetic type-writer). Συνδυάζει δύο αρχιτεκτονικές δικτύων:**

Α) Την αρχιτεκτονική Εκμάθηση ανύσματος ποιοτικού ελέγχου (LVQ) ή Learning Vector Quantization.

Β)Την αρχιτεκτονική αυτοδόμησης χάρτη (SOM).

Το περιεχόμενο της αυτό-οργάνωσης γεννά φωνοτοπικούς χάρτες, όπως αυτός για την φινλανδική λέξη humppila στο σχήμα 6.3. Η νευρωνική φωνητική γραφομηχανή έχει μια ομιλιτική γραμματική (linguistic) η οποία μεταβάλλει την έξοδο του δικτύου σε μια ορθογραφική αποκρυπτογράφηση. Εχει την δυνατότητα να χειριστεί λεξιλόγιο χιλίων λέξεων σε χρόνο που προσεγγίζει τον πραγματικό με αναφερθείσα ακρίβεια 98%

Το βιβλίο BOULARD & MORGAN (1994) και το έγγραφο από Morgan & Boulard (1995) παρέχουν τεχνικές περιγραφές των

νευρωνικών δικτύων που χρησιμοποιούνται στην αναγνώριση συνεχούς ομιλίας. Οι HALFFNER & WAIBEL (1992) περιγράφων το MS-TDNN. Ο Tebelskis (1992) περιγράφει το LPNN, ο LEVIN (1990) περιγράφει το HCNN, και ο KOHONNEN (1998) περιγράφει την νευρωνική φωνητική γραφομηχανή.

6.5 ΕΠΙΛΕΓΟΝΤΑΣ ΤΗΝ ΡΟΗ ΛΟΓΟΥ

Η ροή του λόγου για μια εφαρμογή αναγνώρισης είναι συχνά η πλέον ενδεικνυόμενη κατάσταση εξέλιξης του συστήματος. Πιο συχνά, οι ερευνητές απορρίπτουν όλα πλην των συστημάτων συνεχούς ομιλίας. Αυτό είναι ατυχές για αρκετούς λόγους όπως:

- **Η συνεχής ομιλία δεν είναι αναγκαία για μερικούς τύπους εφαρμογών και σχεδίασης εφαρμογών.**

- **Η ροή του λόγου ενδέχεται να είναι λιγότερο ενδιαφέρουσα ως παράγων από ότι άλλες καταστάσεις όπως το μέγεθος λεξιλογίου ή η ευκαμψία.**

- **Για μερικές εφαρμογές, η αναγνώριση διακριτής λέξεως δυνατόν να αποτελεί την μόνη επιλογή.**

Τα περισσότερα όργανα των συστημάτων διοίκησης & ελέγχου είναι εκ του φυσικού καταξιωμένα απέναντι σε μικρές εκρήξεις λόγου. Τόσο τα διακριτά όσο και τα συνεχή συστήματα είναι κατάλληλα για τέτοιου είδους συστήματα. Για να παραμείνουν μόνο τα συστήματα συνεχούς ομιλίας θα πρέπει να μειωθούν πάρα πολύ οι τιμές προζήτησής τους σε συνδυασμό με την ανάπτυξη της καταλληλότητας τους, σε τέτοιο βαθμό όσο άλλες καταστάσεις που επηρεάζουν την προσφορά των προϊόντων αναγνώρισης διακριτής λέξεως. Αυτές περιέχουν εξαιρετικά υψηλής αξιοπιστίας και απόδοσης εργαλεία εφαρμογών εναλλαγής προϊόντων όπως το σύστημα VOICE MED της Kurzwell AI's

6.5.1 ΣΥΓΚΡΙΣΗ ΣΥΝΕΧΟΥΣ ΟΜΙΛΙΑΣ ΑΠΕΝΑΝΤΙ ΣΤΟΝ ΕΛΕΥΘΕΡΟ ΤΥΠΟ SPONTANEOUS SPEECH

Η συνεχής ομιλία συχνά είναι μπλεγμένη με την ομιλία ελεύθερου τύπου. Ο ελεύθερος τύπος, spontaneous speech αναφέρεται στην επικοινωνία ανθρώπων διασκεδάζοντας ο ένας τον άλλον, επιλέγοντας οποιοδήποτε τίτλο, τρόπο ομιλίας και το λεξιλόγιο που επιθυμούν.

Η αναγνώριση συνεχούς ομιλίας μέσω Η/Υ είναι πολύ περισσότερο περιοριστική. Παρέχει ομιλία άνευ παύσεων μεταξύ των λέξεων, αλλα τονίζει περιεχόμενα στον χαρακτηρισμό της συζήτησης τρόπου ομιλίας, και / η άλλες μορφές επικοινωνίας. Οι αποδεχόμενες επιλογές καθορίζονται από καταστάσεις άλλων συστημάτων:

- **Το μέγεθος και η φύση του λεξιλογίου (Κεφ. 3)**
- **Η φύση της γραμματικής (Κεφ. 4)**
- **Η φύση της δομής της εφαρμογής (Κεφ. 8 & 9)**

Ουσιαστικά όλα τα εμπορικά συστήματα αναγνώρισης ομιλίας θέτουν σοβαρές επιφυλάξεις στο λεξιλόγιο και την δομή, τα οποία είναι προσβάσιμα για τους ομιλητές. Μερικά συστήματα αναγνώρισης συνεχούς ομιλίας εξακολουθούν να προσφέρουν περισσότερες από εκατό (100) λέξεις. **Αν και η αύξηση της υπολογιστικής ικανότητας των PC ενεργοποιεί εφαρμογές λεξιλογίου που φθάνουν έως και 10.000 λέξεις, τα περισσότερα συστήματα συνεχούς ομιλίας εξειδικεύουν ένα συντελεστή μέγιστης διακλάδωσης. Τα περισσότερα εμπορικά συστήματα συνεχούς ομιλίας στηρίζονται σε αυστηρές γραμματικές σε σταθερού επιπέδου γραμματικές ή ζεύγους λέξεων. Ένα σύστημα δομημένης φαρμακευτικής ορολογίας χρησιμοποιώντας μια από τις προαναφερθείσες γραμματικές θα μπορούσε για παράδειγμα, να περιορίσει την είσοδο σχετικά με την πίεση στο αίμα ασθενούς με την παρακάτω μορφή:**
Σε εξάρτηση με το προϊόν της αναγνώρισης, αυτές οι προτάσεις δυνατόν είναι να εισαχθούν σε μια διακριτή ή συνεχή ροή του λόγου. Αν και οι προτάσεις στο σχήμα 6.4 είναι αιτιολογημένες εξ αρχής, δεν επιτρέπουν αποκλίσεις. Ένας φυσικός δεν είναι δυνατόν να πει:

Η πίεση αίματος ενός ασθενούς είναι ΑΡΙΘΜΟΣ στα ΑΡΙΘΜΟΣ σήμερα.
Αυτό είναι ολοκληρωτικά μέσα στα φυσικά όρια του ασθενούς. Οτιδήποτε αποκλίνει από τις μορφές που φαίνονται στο σχήμα είναι μη αποδεκτό.

Η μοντελοποίηση N-γραμμάτων (ΚΕΦ.4, τμήμα 4.3.1) επιτρέπει μια αρκετά κοντινή εκτίμηση στο ελεύθερο τύπο ομιλίας. Ουσιαστικά όλα τα εμπορικά συστήματα ορολογίας που χρησιμοποιούν μοντέλα N-γραμμών απαιτούν είσοδο διακριτής λέξης επειδή ο συνδυασμός της συνεχούς ομιλίας με ένα μεγάλο λεξιλόγιο και είσοδο ελευθέρου τύπου έχει απορία της απάντησης τεραστίων υπολογιστικών (τμήμα 6.7.1 και κεφ.10, τμήμα 10.1) Το SPEECHMAGIC, σύστημα ομιλίας της PHILIPS, το πρώτο εμπορικό προϊόν ορολογίας συνεχούς ομιλίας που χρησιμοποιεί N-γραμμών μοντέλο, έχει διαδικασίας αναγνώρισης μετά αφότου ένας χρήστης έχει ολοκληρώσει την ορολογία.

6.5.2 ΣΥΝΕΧΗΣ ΟΜΙΛΙΑ ΣΥΓΚΡΙΤΙΚΑ ΜΕ ΤΗΝ ΠΛΗΡΟΦΟΡΗΣΗ

Η επιθυμία για την συνεχή ομιλία, είναι συνυφασμένη με την σύγχυση μεταξύ του ελεύθερου τύπου ομιλίας και της συνεχούς ομιλίας, έχει αναδύσει μια δεύτερη μη κατανοητή σχετικά με την απαίτηση της ανθρώπινης επικοινωνίας. Όταν κάποιος κατανοεί την ομιλούσα γλώσσα του άλλου προσώπου, αυτόν / η εφαρμόζει πλέον των ακουστικών και στατιστικών γνώσεων επι του σκοπού. Μια ευρέως φάσματος γνώση και πληροφορία επιτυγχάνεται περιέχοντας:

- **Το ακροαστικό ρεύμα λόγου**

- **Η δομή προτάσεων**
  Έννοια των λέξεων, περιέχοντας συνώνυμα και αντωνυμίες

- **Το αντικείμενο**

- **Γνώση περί του ομιλητή**

Τα Υφιστάμενα εμπορικά συστήματα αναγνώρισης έχουν πληροφορίες μόνο στο προς ακρόαση ρεύμα ομιλίας συνδυαζόμενο με στοιχεία δομής προτάσεων προσαρμοσμένο σε αυτά Ακόμα και ο συνδυασμός συνεχούς ομιλίας, μεγάλου λεξιλογίου και δομής ελεύθερου τύπου δεν καθιστά εφικτό από τα συστήματα να κατορθώσουν την αξιοπιστία του ανθρώπινου λόγου.
Αυτό απαιτεί κατ' ελάχιστο μια συστοιχία για εργαλεία ανάλυσης ομιλούσας γλώσσας συγκρινόμενα με εκείνου που αναφέρθηκαν ανωτέρω. Όπως και να είναι πάντως, είναι σημαντικό να θυμόμαστε ότι ακόμα και μετά την χρήση αυτών και άλλων εργαλείων, οι άνθρωποι ακόμα δεν κατανοούν ο ένας τον άλλον. Χωρίς τους περιορισμούς της ανθρώπινης κατανόησης, μπορεί να γίνει πιο εύκολα (ειρωνικά) κατανοητό ότι η τεχνολογία ομιλίας έχει την ικανότητα να έχει πρόσβαση σε όλα το αναλυτικά εργαλεία της λίστας όπου θα πρέπει να εκτελέσει χωρίς σφάλμα.

6.6 ΟΜΙΛΙΑ ΔΙΑΚΡΙΤΩΝ ΛΕΞΕΩΝ ΚΑΤ ΟΜΙΛΙΑ ΣΥΝΔΕΔΕΜΕΝΩΝ ΛΕΞΕΩΝ

Οι συσκευές αναγνώρισης για τον διακριτό και συνδεδεμένο λόγο, αναμένουν ξεκάθαρα όρια λέξεων και δεν ανέχονται την διαπλεκόμενη διάρθρωση. Η συνδεδεμένη ομιλία είναι γενικά χρήσιμη στην επιλογή μεταξύ συστημάτων αναγνώρισης διακριτής λέξεως. Συχνά περιορίζεται σε κάποια ορισμένα λεξιλόγια και υποσύνολα τους, συνήθως ψηφία (0 έως 9). Όταν χρησιμοποιείται κατ' αυτό τον τρόπο, ο συνδεδεμένος λόγος διευκολύνει την είσοδο του καθορισμένου λεξιλογίου σε εφαρμογές επαναληπτικής εισόδου δεδομένων. Η ομιλουμένη έναρξη μακρών σειρών ψηφίων, για παράδειγμα, αποβαίνει ταχύτερη και λιγότερο βαρετή. Ωστόσο, αυτά θα πρέπει να αρθρώνονται ξεκάθαρα.

6.6.1. ΑΚΡΙΒΕΙΑ. Η χρήση των παύσεων (ή ομιλία staccato) είναι ένα προσόν τόσο όσο και ένα περιορισμός. Προσδίδει ακρίβεια αναγνώρισης μέσω της αναγνώρισης των ορίων για κάθε λέξη.

Στον ίδιο χρόνο, η ανάπαυλα επιφέρει ταυτότητα στα συστήματα διακριτής λέξεως και συνδεδεμένης λέξεως σε σφάλμα εισαγωγής (κεφ.2, τμήμα 2.6.2) τα οποία επιδρούν από ξαφνικά, απροσδόκητο θόρυβο, όπως το κουδούνισμα ενός κοντινού τηλεφώνου ή δυνατού background λόγου. Αυτά τα προβλήματα δεν είναι δυνατό να εξαλειφθούν αλλά μπορούν να μειωθούν μέσω της ανάπτυξης αλγορίθμων αναγνώρισης πιο σοφιστικέ συνδυασμών με:

- **Υψηλής ποιότητας κατευθυντικά μικρόφωνα**

- **Καλή μοντελοποίηση του περιβάλλοντος ομιλίας (Κεφ.7)**
- **Καλή μοντελοποίηση του ομιλητή**

- **Υψηλά επίπεδα αποδοχής**

- **Εύχρηστες τεχνικές διόρθωσης σφαλμάτων.**

Μια αποτελεσματική μέθοδος για τη διόρθωση σφαλμάτων είναι να περιέχεται μια ή περισσότερες λέξεις, όπως το «Oops», «backup» ή «stratch that» και το οπίο επιδρά στο σύστημα ώστε να εκτελέσει «backup». Αυτές οι λέξεις οδηγούν το σύστημα να μετακινήσει την πιο πρόσφατη απόφαση αναγνώρισης.

Οι στρατηγικές διόρθωσης σφάλματος συχνά συνδυάζονται με τεχνικές επιβεβαίωσης οι οποίες επιτρέπουν στους χρήστες να ξανακοιτάξουν το πιο πρόσφατο σύνολο των αναγνωρισθέντων λέξεων. Πολλά προϊόντα αναγνώρισης επιτρέπουν την σχεδίαση της επιβεβαίωσης χρησιμοποιώντας ψηφιοποιημένο ή συνθετοποιημένο έξοδο λόγου των τμημάτων που έχουν αναγνωρισθεί.

Από την στιγμή που η παύση μεταξύ των λέξεων δεν αποτελεί έναν φυσικό τρόπο ομιλίας, οι εφαρμογές διακριτής λέξεως και συνδεδεμένης λέξεως και στις οποίες έχουν πρόσβαση χρήστες μιας φοράς, απαιτείται να ελέγχει την ροή λόγου αυτών των χρηστών. Αυτή είναι μια περίπτωση που αντιμετωπίζει τις περισσότερες τηλεφωνικές εφαρμογές. Η είσοδος ψηφιακών ακολούθων μέσω της τηλεφωνίας μπορεί να συμβαδίσει με την χρήση των προειδοποιητικών ήχων (beeps) μεταξύ των ψηφίων προκειμένου να εμποδίσει τους χρήστες να ομιλούν πολύ γρήγορα. Το κλειδί (κεφ.4, τμήμα 4.5) δυνατόν να χρησιμοποιείται για να τοποθετήσει το συγκεκριμένο λεξιλόγιο προσαρμοσμένο στον υπερφίαλο λόγο.

6.6.2 <u>ΕΥΚΟΛΙΑ ΧΡΗΣΗΣ</u>. Η φορτικότητα της παύσης μεταξύ εξαρτάται από την φύση της εφαρμογή και από τον χρήστη. Σε πολλές εφαρμογές δεν αποτελεί οργανική δυσκολία λόγου (impediment). Όταν οι ανταποκρίσεις του χρήστη είναι περιορισμένες σε απλές λέξεις, για παράδειγμα, οι παύσεις περνούν απαρατήρητες. Οι περισσότερες εφαρμογές οργάνων διοίκησης και ελέγχου καθώς και εφαρμογές εισόδου δεδομένων μπορούν να δομηθούν έτσι ώστε να αποτελούνται κυρίως από μικρές εκφράσεις (Κεφ. 8 και 9).

Εάν μια εφαρμογή ερευνά τις αποκρίσεις λέξεων, αλλα είναι πιθανότατα ότι οι χρήστες θα εμπεδώσουν το επιθυμητό λεξιλόγιο στην συνεχής ομιλία, μπορεί να χρησιμοποιηθεί η επισήμανση των λέξεων. Σε τέτοιες περιπτώσεις η επισήμανση λέξεων είναι ένας τρόπος γενίκευσης ενός συστήματος αναγνώρισης μεμονωμένων λέξεων στον οποίο ο χρήστης δεν υποχρεούται να συλλαβίσει τις λέξεις στην απομόνωση; Αυτό οδηγεί σε περισσότερο φιλικά προς τον χρήστη συστήματα.

Στην συνέχεια η χρήση του εντοπισμού λέξεων λειτουργεί καλύτερα όταν το σύνολο των λέξεων - στόχων είναι μικρό και το επίπεδο της σύγχυσης είναι χαμηλό. Πρόσφατες έρευνες στο SRI προτείνουν αυτοί οι περιορισμοί να μην είναι αναγκαίοι. Αποδείχθηκε η ύπαρξη σημαντικών αυξήσεων στην ακρίβεια του εντοπισμού λέξεων όταν αυτές μοντελοποιούσα ένα μεγάλο αριθμό μη σημαντικών λέξεων (σημαντικός = κλειδί), τόσο καλά όσο ο εντοπισμός λέξεων κλειδιών. Αυτές οι ανακαλύψεις είναι δυνατόν να υποστηρίζουν την αύξηση του μεγέθους και της ευκαμψίας των εφαρμογών εντοπισμού λέξεων στο μέλλον και αυτές γίνουν μεγαλύτερες.

Γενικά είναι η παύση αυτή η οποία συνδυαζόμενη με την επιθυμητή αργία των συστημάτων διακριτών λέξεων ώστε να δημιουργείται η πηγή της μέγιστης ενόχλησης για τους χρήστες. Ομιλητές με εμπειρία στην ευγλωττία, είναι εκείνοι που χρησιμοποιούν την υπαγόρευση αναγνώρισης ομιλίας. Ο εκνευρισμός προφέρεται επιλεκτικά από τους ομιλητές καθόσον έχουν ευχέρεια στην ταχεία απαγόρευση μιας αναφοράς που έχει διορθωθεί και μεταγραφεί από κάποιον άλλον. Αντιλαμβάνονται την χρήση του λόγου διακριτών-λέξεων ως σύνθημα που να κάνει την δουλειά των πιο δύσκολη. Εάν η υποκίνηση του ομιλητή για να χρησιμοποιήσει ένα σύστημα υπαγόρευσης είναι μεγάλη, τότε οι απαιτούμενες παύσεις αποβαίνουν κατι από λιγότερο από εμπόδιο.

6.6.3 ΕΚΤΙΜΗΣΗ ΑΠΟΔΟΣΕΩΣ. Μια σημαντική μέτρηση εκτιμήσεως είναι η επιθυμία του πλήθους των χρηστών να μετατρέψουν τις απαιτήσεις της ομιλίας διακριτών λέξεων. Το επίπεδο αποδοχής εξαρτάται από:

- **Τύπο χρηστών που εμπλέκεται κατά περίπτωση**
- **Φύση της εφαρμογής**
- **Σχεδίαση της εφαρμογής**
- **Προετοιμασία / συμμετοχή των χρηστών στην σχεδίαση των εφαρμογών**

Μερικές ομάδες χρηστών, σημαντικότερη οι φυσικοί και υψηλόβαθμοι διοικητικοί, προβάλλουν λιγότερη υπομονή στην ομιλία διακριτών λέξεων από ότι οι άλλοι ομιλητές. Από την στιγμή που η ανυπομονησία δεν είναι εξαπλωμένη μέσα στις ομάδες αυτές, η αποδοχή της ομιλίας διακριτών λέξεων πρέπει να ισοσταθμίσθει απέναντι σε άλλους σκοπούς χρήσης συστημάτων αναγνώρισης ομιλίας για τέτοιους πληθυσμούς.

> Η χρήση ομιλίας διακριτής λέξεως για υψηλή πίεση, εφαρμογές υψηλών ταχυτήτων, όπως η διακίνηση αξιών (stock trading), είναι δυνατόν να δημιουργήσει αντίδραση εάν οι ομιλητές έχουν την πεποίθηση ότι η ροή του διακριτού λόγου αλληλοπαρεμβάλλεται με την ικανότητα του στην απόδοση άλλων λειτουργιών. Αυτή η αντίληψη είναι δυνατόν να αντανακλά μια απόκριση φτωχής σχεδίασης εφαρμογών, αντίσταση αλλαγής, ενόχληση σε οποιοδήποτε σύστημα ομιλίας ή νες τεχνολογίας, ή αυτόνομη και ανεπαρκή εκπαίδευση.

Όλα αυτά τα θέματα πρέπει να εξετασθούν.
Η βαθμολογία της απόδοσης απαιτεί την πινακοποίηση των παρακάτω στοιχείων:

- **Σωστές αναγνωρίσεις**
- **Διαγραφές**
- **Εισαγωγές**
- **Αντικαταστάσεις**

Μια περισσότερο λεπτομερής ανάλυση μπορεί να περιέχει σύγκριση διαφορετικών τύπων σφαλμάτων και προβλήματα που συνδέονται με συγκεκριμένα σημεία λεξιλογίου ή ομάδες λέξεων. Στο Κεφ. 8 (τμήμα 8.5) γίνεται η ανάλυση των προϊόντων και η εκτίμηση εφαρμογών με περισσότερες λεπτομέρειες.

6.7 ΣΥΝΕΧΗΣ ΟΜΙΛΙΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ

Οι άνθρωποι πιστεύουν ότι εάν μπορεί να έχεις 50.000 λέξεις, γιατί να μην έχεις και συνεχή ομιλία; Είναι μια άποψη θέσεως πολύ δύσκολη.

Η πρόκληση που τέθηκε από τον BUNTING μπορεί να εκφυλισθεί από μία μόνο λέξη;

Jeet;

Εάν δεν είναι ακόμα ξεκάθαρο τη σημαίνει αυτό, φανταστείτε ότι είναι μεσημέρι και έχετε ήδη πεινάσει. Κατόπιν φανταστείτε κάποιον που εμφανίζεται πριν από εσάς και να ρωτάς jeet; Αυτή η ακραία έκφραση παράφραση του «Did you eat?» δεν είναι ασυνήθης όπως φαντάζεται κάποιος. Θα μπορούσατε να έχετε

ανταποκριθεί με την φράση:

No Skweet!

Ή με την φράση

Ahminuheet later

Άλλο παράδειγμα συστημένων εκφράσεων περιέχονται τα «wanna», «gonna», «doncha» έχουν γίνει πλέον συνήθεια. Δεν είναι πλέον παρατηρήσιμα από τους ανθρώπους για τι χρησιμοποιούμε την περίληψη καταστάσεων ή συζήτηση για να μειώσουμε περισσότερες συγχύσεως αυτού του είδους. Όταν αυτά τα εργαλεία αποτύχουν, παρατηρούνται σύγχυση και μη κατανόηση ως αποτέλεσμα.

Το humor που εκφράζει το τραγούδι «Maizry Doats», για παράδειγμα είναι φανερά εκπεφρασμένο στους (forcing listeners) ακροατές που προσπαθούν να κατανοήσουν στον περιπεπλεγμένο λόγο χωρίς να έχουν επιτύχει την ουσία.

Τα εμπορικά προϊόντα αναγνώρισης ομιλίας κινούνται στα πλαίσια του «Maizry Doats» :

> Διαθέτουν πολύ λίγο από το συγκεκριμένο περιβάλλον για να βοηθήσουν στην ερμηνεία αυτών που ένας ομιλητής λέει και καμμία πληροφόρηση ώστε να τα καταστήσει ικανά για να δομήσουν μια ερμηνεία. Όπως φαίνεται στο τραγούδι, η χρήση της ροής του λόγου διακριτών λέξεων βοηθά στην αντίληψη αλλά αντίθετα προς την ακουστική του «Maizry Doats» τα συστήματα αναγνώρισης ομιλίας δεν είναι κατάλληλα για να εκφράσουν οποιοδήποτε humor το οποίο μπορεί να αποτελεί μέρος της καταστάσεως των.

6.7.1 <u>ΑΚΡΙΒΕΙΑ ΣΤΑ ΕΜΠΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ.</u> Τα υφιστάμενα εμπορικά συστήματα βασίζονται κυρίως σε εξωτερικές μεθόδους επαύξησης της ακρίβειας και παρέχουν διόρθωση σφαλμάτων. Τα συστήματα που βασίζονται στην είσοδο μικρών εντολών ή δεδομένων χρησιμοποιούν μερικές από τις ίδιες τεχνικές με αυτές των διακριτών και συνδεδεμένων λέξεων συστημάτων.

Όταν τα λεγόμενα είναι μακροσκελή και μπλεγμένα, η επιλογή της διόρθωσης σφάλματος επιδρά σημαντικά στην εύχρηστη εκμετάλλευση του συστηματικού και στην συνολική αποδοτικότητα. Η δυσκολία του να επιτευχθεί υψηλής ποιότητας αναγνώριση συνεχούς ομιλίας για μεγάλα πακέτα λεξιλογίου είναι ο κύριος λόγος γιατί τα πλειστά των συστημάτων αυτών δεν έχουν δημιουργηθεί από τα εργαστήρια ερευνών.

Τα εργαλεία του συστήματος υπαγόρευσης εταιρείας PHILIPS, το SPEECH MAGIC αποφεύγουν μερικά από αυτού του είδους τα προβλήματα αποδίδοντας την αναγνώριση μετά την ολοκλήρωση της ορολογίας. Αυτό απλά σημαίνει ότι ο χρόνος που απαιτείται για την διόρθωση των αναγνωρισμένων λαθών και την αναγνώριση του χαμένου λεξιλογίου αποδίδεται αργότερα, και ίσως, από κάποιο πρόσωπο διαφορετικό αυτού που υπαγόρευσε το υλικό.

Η ακρίβεια των συστημάτων συνεχούς ομιλίας μπορεί να εξασθενήσει από μη επικοινωνιακή ομιλία, όπως το «UH» , αυτό-διόρθωση, παρεμβολές. Όταν τέτοιου είδους σφάλματος και υπερφίαλος λόγος αναμένεται να είναι συχνοί, δυνατόν να χρησιμοποιηθεί η λειτουργία επισήμανσης λέξεων. Σε

άλλες περιπτώσεις η οικειότητα του χρήστη σε μια εφαρμογή ή προχωρημένης δομής υποβολή, μπορεί να μειώσει αισθητά τον μη-επικοινωνιακό λόγο. Ο Wayne Wand του CMU έχει ανακαλύψει ότι η δημιουργία μοντέλων μερικών από αυτά τα φαινόμενα ομιλίας επαυξάνει την ακρίβεια στην ελεύθερη απόδοση στα συστήματα έρευνας συνεχούς ομιλίας. Περισσότερες πληροφορίες, παρέχονται στην εργασία του Ward στην περιοχή αυτή και στο Κεφ.7 (Τμήμα 7.1.4)

6.7.2 ΕΥΧΡΗΣΤΗ ΧΡΗΣΗ ΤΩΝ ΕΜΠΟΡΙΚΩΝ ΠΑΚΕΤΩΝ Όπως και με τον διακριτό & συνδεδεμένο λόγο, η ευκαμψία της γραμματικής ενός συστήματος και η ποιότητα των στρατηγικών διόρθωσης σφαλμάτων μπορεί να επιφέρει άμεση επίδραση στην εύκολη χρήση των συστημάτων. Οι περισσότεροι χρήστες δεν θα ευχαριστηθούν με το να επαναλάβουν ολόκληρες σειρές ομιλίας ειδικότερα αν ο χρόνος είναι καθοριστικός παράγων για την επίτευξη του σκοπού.

Μερικά συστήματα υπαγόρευσης παρέχουν ορατή επαλήθευση των αναγνωρισθέντων τμημάτων σε μια οθόνη Η / Υ και επιτρέπουν τον συνδυασμό πληκτρολογίου, ποντικιού και ομιλίας προκειμένου διορθωθούν τα σφάλματα. Αυτά τα εργαλεία είναι σχεδιασμένα για να αντικαταστήσουν τους μεταφραστές. Αντίθετα με τους ανθρώπινους μεταφραστές, τέτοιου είδους συστήματος παρουσιάζουν δυσκολία στην απόκριση κοινών εντολών όπως:

- **Delete that paragraph (Διαγραφή παραγράφου)**
- **Change the first sentence to read (Αλλαγή κειμένου πρώτης πρότασης)**
- **Get the address from the file (Εύρεση Διεύθυνσης αρχείου)**

Οι χρήστες των συστημάτων υπαγόρευσης οι οποίοι ωστόσο έχουν οικειότητα με τις συνηθισμένες συσκευές υπαγόρευσης

6.7. 3 ΕΚΤΙΜΗΣΗ ΑΠΟΔΟΣΕΩΣ ΣΤΑ ΕΜΠΟΡΙΚΑ ΠΑΚΕΤΑ Είναι πολύ πιο δύσκολο να εκτιμηθεί η απόδοση ενός συστήματος αναγνώρισης συνεχούς ομιλίας από το να διατιμηθεί κάποιο σύστημα διακριτών λέξεων. Ένα απλό μετρημα των συνολικών σφαλμάτων, ακόμα και αν είναι ομαδοποιηημένα ανά τύπο σφάλματος, παρέχει μια λανθασμένη εικόνα . Εάν μια εσφαλμένη θέση τέλους λέξεως συμβεί νωρίς σε μία έκφραση μπορεί να παράγει ένα φαινόμενο ΄΄domino effect΄΄ από λανθασμένες αναγνωρίσεις. Ο καλύτερος οδηγός για την παραγωγή και την εφαρμογή της διατίμησης είναι οι ανάγκες και οι προτεραιότητες της εφαρμογής.

> Η επιλογή των πλέον κατάλληλων μεθόδων μέτρησης αποσκοπεί στον καθορισμό των σχετικών εφαρμογών και , ειδικότερα του τρόπου της επιβεβαίωσης και διόρθωσης από τον ομιλητή.(David Pallett, National Institute of Standards and Technology, ΄΄Performanve assessment of autonatic speech recognizers΄΄, 1985, p.380)

Τα θέματα που σχετίζονται με την εκτίμηση αποδόσεως παρουσιάζονται με

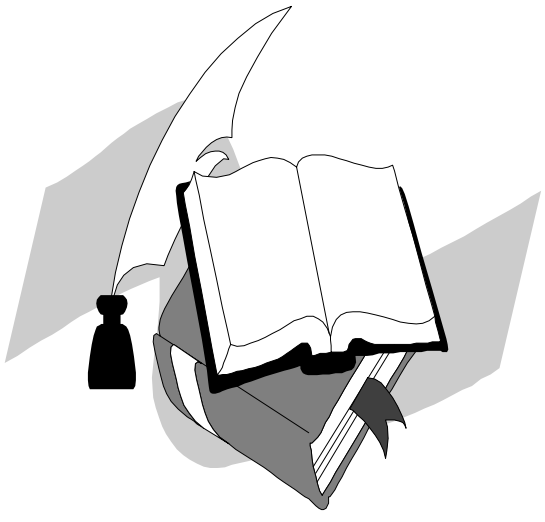
περισσότερες λεπτομέρειες στο κεφάλαιο 8 .

♦ ♥ ♠ ♣ ♦ ♥ ♠

Σ Υ Μ Π Λ Η Ρ Ω Μ Α Ε Ρ Γ Α Σ Ι Α Σ

ΘΕΜΑ : ΄΄THE HIDDEN MARKOV MODEL΄΄

Αναφορές :΄΄ DISCRETE-TIME PROCESSING OF SPEECH SIGNALS΄΄, J.R.Deller, J.G. Proakis, J.H.L. Hansen (Chapt. 12)



ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΤΙΤΛΟΣ

Σελ. Ε**ΙΣΑΓΩΓΗ**
28

2. **ΑΝΑΠΤΥΞΗ ΘΕΩΡΙΑΣ**
30

2.1 ΕΙΣΑΓΩΓΗ
2.2 ΤΥΠΟΙ ΄΄MOORE΄΄ & ΄΄MEALY΄΄
ΓΙΑ ΤΟ ΗΜΜ
2.3 ΤΑ ΔΥΟ ΠΡΟΒΛΗΜΑΤΑ ΗΜΜ
2.4 Η ΔΙΑΚΡΙΤΗ ΠΑΡΑΤΗΡΗΣΗ ΤΟΥ ΗΜΜ
Τυπικές Διαδικασίες
2.5 ΑΝΑΓΝΩΡΙΣΗ ΜΕ ΧΡΗΣΗ
ΔΙΑΚΡΙΤΗΣ ΠΑΡΑΤΗΡΗΣΗΣ ΗΜΜ
2.6 ΕΦΑΡΜΟΓΗ/ΕΚΠΑΙΔΕΥΣΗ ΤΗΣ ΔΙΑΚΡΙΤΗΣ
ΠΑΡΑΤΗΡΗΣΗΣ ΗΜΜ
2.7 ΣΥΝΕΧΗΣ ΠΑΡΑΤΗΡΗΣΗ ΗΜΜ
2.8 ΠΙΘΑΝΟΤΗΤΕΣ ΔΙΑΡΚΕΙΑΣ ΚΑΤΑΣΤΑΣΗΣ
ΣΤΗΝ ΔΙΑΚΡΙΤΗ ΠΑΡΑΤΗΡΗΣΗ ΗΜΜ
2.9 ΚΛΙΜΑΚΩΣΗ ΤΟΥ F-B ΑΛΓΟΡΙΘΜΟΥ
2.10 ΕΦΑΡΜΟΓΕΣ ΜΕ ΑΚΟΛΟΥΘΙΕΣ
ΠΟΛΛΑΠΛΗΣ ΠΑΡΑΤΗΡΗΣΗΣ
2.11 ΚΡΙΤΗΡΙΑ ΕΝΑΛ. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ
ΣΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΗΜΜ

3. **ΠΡΑΚΤΙΚΕΣ ΕΚΔΟΣΕΙΣ**
3.1 ΑΚΟΥΣΤΙΚΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ
3.2 ΜΕΓΕΘΟΣ & ΔΟΜΗ ΤΟΥ ΜΟΝΤΕΛΟΥ
3.3 ΑΝΑΠΤΥΞΗ ΜΕ ΑΝΑΚΡΙΒΗ ΔΕΔΟΜΕΝΑ
3.4 ΑΚΟΥΣΤΙΚΕΣ ΜΟΝΑΔΕΣ ΜΟΝΤΕΛΟΠΟΙΗΜΕΝΕΣ
ΑΠΟ ΗΜΜ

ΣΥΜΠΛΗΡΩΜΑ ΕΡΓΑΣΙΑΣ
ΘΕΜΑ : ΄΄THE HIDDEN MARKOV MODEL΄΄

Εισαγωγή

Στις δεκαετίες 1970 και 1980 , οι ερευνητές ομιλίας ξεκίνησαν να επιστρέφουν στην στοχαστική διαδικασία των μοντέλων ομιλίας με σκοπό να διευθυνσιοδοτηθεί το πρόβλημα της μεταβλητότητας, ειδικότερα στα συστήματα μεγάλης κλίμακας.

Ο όρος ΄΄ΣΤΟΧΑΣΤΙΚΉ ΔΙΑΔΙΚΑΣΊΑ΄΄ χρησιμοποιείται για να επιδείξει ότι τα μοντέλα στα οποία εφαρμόζεται η προσέγγιση αυτή χαρακτηρίζουν από την φύση τους , ένα μέρος της μεταβλητότητας του λόγου. Αυτό διαπιστώνεται και κατά την απευθείας χρήση των δεδομένων της ομιλίας σε μαζικές προσεγγίσεις του μοντέλου χωρίς την εφαρμογή της πιθανολογικής μορφής του. Ο όρος ΄΄ΔΟΜΙΚΕΣ ΜΕΘΟΔΟΙ΄΄ χρησιμοποιείται επίσης στην περιγραφή των στοχαστικών προσεγγίσεων, από την στιγμή Δε που καθεμία από τις μεθόδους βασίζεται πάνω σε μοντέλο με πολύ σημαντική δομή.

Οι δύο διαφορετικοί τύποι των στοχαστικών μεθόδων έχουν ερευνηθεί. Ο πρώτος είναι το μοντέλο HMM **(Hidden Markov Model)** , όπου είναι εφικτή η χρήση του σε συμβατικές ακολουθιακές υπολογιστικές μηχανές. Από την συνεχή έρευνα, το HMM έχει ανακηρυχθεί η βάση για πολλά επιτυχή λεξιλόγια και εμπορικά συστήματα αναγνώρισης ομιλίας. Η ιστορική εξέλιξη του μοντέλου HMM βασίζεται στην επεξεργασία ομιλίας και βαθμιαία έγινε ευρύτερα γνωστό στο πεδίο ομιλίας. Ο δεύτερος τρόπος ήταν το τεχνητό νευρωνικό δίκτυο (**Artificial Neural Network**) , το οποίο εξελίχθηκε ως μέρος μιας ευρύτερου φάσματος έρευνας με θέμα τι ς εναλλακτικές αρχιτεκτονικές παράλληλου προγραμματισμού των βιολογικών νευρικών συστημάτων.

Η είσοδος του HMM στο πεδίο αναγνώρισης ομιλίας ήταν αποτέλεσμα της ανεξάρτητης έρευνας που έκανε ο BAKER στο Carnegie- Mellon University. Ταυτόχρονα υπήρχε η εξέταση του μοντέλου από τον Poritz στο Institute for Defense Analysis. Μαζί με τον Levinson, ο Poritz ανακάλυψε ότι η κινητικότητα γύρω από την ανάπτυξη της μορφής HMM ξεκινούσε από την δεκαετία του ΄50 , όπου οι αναλυτές μελετούσαν το πρόβλημα του χαρακτηρισμού τυχαίων διαδικασιών για τις οποίες υπήρχαν μη περατωθείσες παρατηρήσεις. Η προσέγγιση τους γινόταν μέσω της διπλά στοχαστικής διαδικασίας **(Doubly Stochastic Process**), κατά την οποία τα παρατηρούμενα δεδομένα θεωρούντο ως το ΄΄πέρασμα΄΄ της πραγματικής (**Hidden**) διαδικασίας μέσα από αισθητήρα (**Sensor**) ο οποίος παρήγαγε την επόμενη διαδικασία(**Observed**) .

Και οι δύο διαδικασίες χαρακτηρίζονταν μόνο απο αυτή που ήταν δυνατό να παρατηρηθεί . Το αποτέλεσμα ήταν ο αλγόριθμος εκτίμησης - μεγιστοποίησης(ΕΜ). Αργότερα οι Baum και οι συνεργάτες του ανέπτυξαν μια ειδική περίπτωση HMM κατά την οποία αναπτύσετο μια ειδική περίπτωση του ΕΜ αλγόριθμου , ο **F-B algorithm .**

2. ΑΝΑΠΤΥΞΗ ΘΕΩΡΙΑΣ

2. 1 Εισαγωγή

Ένα HMM είναι στην ουσία ένας αυτοματισμός ορισμένων καταστάσεων(**Finite State Automation)** , μια αφηρημένη μηχανή η οποία χρησιμοποιείται για την μοντελοποίηση της έκφρασης λόγου . Η έκφραση μπορεί να έχει το μέγεθος λέξεως, μονάδος υποδιαίρεσης λέξεως , πρότασης, φωνήματος ανάλογα με το μέγεθος του λεξιλογίου για το οποίο χρησιμοποιείται το HMM.

Χωρίς να χάνεται η γενικότητα κατά πολύ , θα θεωρήσουμε ως μονάδα εξέτασης την λέξη , ώστε να κατανοήσουμε την λειτουργία του HMM. Ήδη είμαστε εξοικειωμένοι με την ιδέα να μειώσουμε την έκφραση λόγου σε μία σειρά(ακολουθία) παραμέτρων . Δεχόμαστε ότι για τον έλεγχο που θα ενεργήσουμε ισχύει :

Test features : t(1), t(2), t(3), ...,t(i) ,...,t(I) (1)

Στην εξέταση του HMM τα ανωτέρω ονομάζονται παρατηρήσεις ή παρατηρούμενα(**observable**), από την στιγμή που αυτές οι τιμές αναπαριστούν την πληροφορία από την εισερχόμενη έκφραση ομιλίας Έτσι λοιπόν μπορούμε να θεωρήσουμε την έκφραση των παρατηρήσεων , μία άλλη μορφή της (1) :

obs = y(1), y(2), y(3), ..., y(t), ..., y(T) (2)

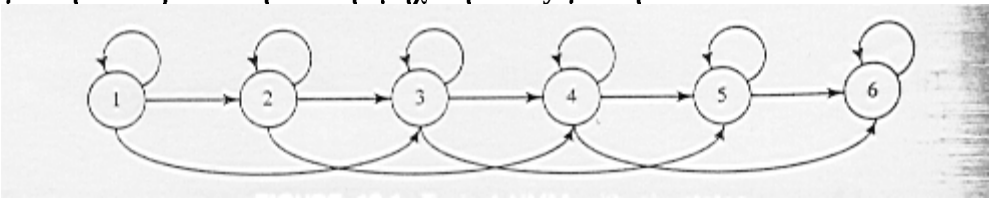
2.2 Τύποι ΄΄MOORE΄΄ και ΄΄Mealy΄΄ για το HMM

Το HMM, το οποίο πάντοτε εκφράζεται με ειδική λέξη (ή άλλου είδους έκφραση) είναι μια μηχανή ορισμένων καταστάσεων (**Finite State Machine**) που διαθέτει την ικανότητα να δημιουργεί ακολουθίες ή σειρές παρατηρήσεων. Σχετίζεται περισσότερο με την δημιουργία των σειρών παρατήρησης που προέρχονται από την παρατήρηση της ειδικής λέξης του HMM. Κατά την διάρκεια της εκπαιδευτικής φάσης το HMM ΄΄διδάσκει΄΄ την στατιστική ανανέωση των παρατηρήσεων σύμφωνα με την ειδική για το καθένα λέξη.

Ενόσω διαρκεί η φάση αναγνώρισης υποτίθεται ότι το κάθε ένα από τα υφιστάμενα HMM παράγει την ακολουθία παρατήρησης. Γεννάται το ερώτημα κατά πόσο είναι εφικτό το συγκεκριμένο κάθε φορά HMM να παράγει αυτή την εισερχόμενη ακολουθία; Η λέξη που υπάρχει σχέση του HMM και φαίνεται σύμφωνα με την μέγιστη πιθανοφάνεια (**Highest Likelihood**) θεωρείται ως η σημαντική λέξη . Στο σημείο αυτό πρέπει να σημειωθεί ότι ο αποκλειστικός

σκοπός του HMM δεν είναι η δημιουργία ακολουθιών παρατήρησης , αλλά υποτίθεται αυτό γίνεται για τις ανάγκες της απόδειξης.

Το διάγραμμα αναπαράστασης ενός HMM φαίνεται στο σχήμα 2.1 . Οι καταστάσεις φαίνονται με τους ακέραιους. Η δομή ή η τυπολογία καθορίζεται με τις μεταπτώσεις μεταξύ των καταστάσεων. Η δημιουργία της ακολουθίας παρατηρήσεων γίνεται απο τις μεταπηδήσεις μεταξύ καταστάσεων εκπέμποντας μια παρατήρηση ανά πήδημα μεταξύ των καταστάσεων. Δύο τύποι HMM, που διαφέρουν ελάχιστα μεταξύ τους χρησιμοποιούνται ανάλογα με το είδος της επεξεργασίας. Ο ένας τύπος αναφέρεται στην ακουστική επεξεργασία και ο άλλος στην επεξεργασία γλώσσας. Η διαφορά τους έγκειται στην δημιουργία των παρατηρήσεων το μεν ένα δημιουργεί παρατήρηση όταν αφικνείται σε κάθε κατάσταση , το Δε άλλο κατά την διάρκεια της μετάπτωσης. Έτσι έχουμε την μηχανή Moore για την $1^{η}$ περίπτωση και την μηχανή Mealy για την $2^{η}$.



εικόνα 2.1 1Τυπική Μορφή HMM 6 Καταστάσεων

Θα εξετάσουμε τον τύπο του Moore στην παράγραφο αυτή. Οι μορφές των μεταπτώσεων είναι ιδανικά ίδιες και για τις δύο περιπτώσεις. Σε κάθε χρόνο παρατήρησης μια κατάσταση μετάπτωσης εκφράζει το μοντέλο. Οι πιθανοφάνειες των μεταπτώσεων αυτών εμφανίζονται ως τόξα μεταξύ των καταστάσεων. Θεωρώντας την πιθανότητα μετάπτωσης καταστάσεων αji από την κατάσταση i στην κατάσταση j, κατασκευάζουμε τον πίνακα πιθανοτήτων μετάπτωσης όπου S είναι το σύνολο των καταστάσεων στο μοντέλο.

$$a(i|j) \stackrel{\text{def}}{=} P(\underline{x}(t) = i|\underline{x}(t-1) = j),$$

Τα μεγέθη α(j/i) δεν εξαρτώνται από τον χρόνο. Κάθε στήλη εκφράζει μοναδικότητα καθόσον μια μετάπτωση συμβαίνει μια φορά τον χρόνο . Οι πιθανότητες ως μεγέθη εκφράζονται ως εξής:

Υπό την προϋπόθεση :

$$a(i|j_1) = P(\underline{x}(t) = i|\underline{x}(t-1) = j_1, \underline{x}(t-2) = j_2, \underline{x}(t-3) = j_3, \ldots),$$

δημιουργείται μια τυχαία επεξεργασία γνωστή ως $1^{ης}$ τάξεως Markov διαδικασία. Όταν στην διαδικασία αυτή οι μεταβλητές λαμβάνουν διακριτές τιμές τότε μιλάμε για αλυσίδα Markov. Η αλυσίδα Markov , εφόσον οι τιμές δεν εξαρτώνται από τις τιμές του χρόνου t, åßíáé ïïïãåíÞò.

Áí êáé ç äéáöïñÜ åóôéÜæåôáé ìüíï óôïõò ôýðïõò, åí ôïýôïéò ïé ìåôáðôþóåéò ÷ñçóéìïðïéïýíôáé ðåñéóóüôåñï áðü ôéò êáôáóôÜóåéò , êáé ôïýôï äéüôé ïé ðñþôåò äéáèÝôïõí áíáãíþñéóç ðïõ âïçèÜ áôçí åõÝëéêôç ÷ñÞóç ôïõò. Ç uij εκφράζει την μετατόπιση μεταξύ I, j . Έτσι για την τυχαία διαδικασία u ισχύει :

$$P(u(t)=u\ i/j\ )=P(x(t)=i/x(t-1)=j)=a(i/j)$$

Για καλύτερη εξέταση των ανωτέρω και στο πως επιδρούν στον τύπο Moore δημιουργήθηκε το άνυσμα κατάστασης πιθανότητας

$$\pi(t) \stackrel{\text{def}}{=} \begin{bmatrix} P(\underline{x}(t) = 1) \\ P(\underline{x}(t) = 2) \\ \vdots \\ P(\underline{x}(t) = S) \end{bmatrix}.$$

$$\pi(t) = \mathbf{A}^{t-1}\pi(1).$$

$$\pi(t) = \mathbf{A}\pi(t-1).$$

Ρίχνοντας μια ματιά στις παρατηρήσεις συμπεραίνεται ότι η ακολουθία των παρατηρήσεων εξιδανικεύεται ως μια στοχαστική διαδικασία

διακριτού χρόνου : $f_{\underline{x}|\underline{s}}(\xi|i) \stackrel{\text{def}}{=} f_{\underline{x}(t)|\underline{s}(t)}(\xi|i)$ for arbitrary t. Στον τύπο του Moore , με την έναρξη της κατάστασης δημιουργείται και η αντίστοιχη παράτήρηση.

2.3 Τα Δύο προβλήματα HMM

Οι δύο περιπτώσεις κατά τις οποίες εστιάζεται η χρήση του HMM είναι :

Α. Δοθέντος μιας σειράς παρατηρήσεων για συγκεκριμένη λέξη με ποιο τρόπο θα χρησιμοποιήσουμε το HMM ώστε να αναπαριστά την λέξη; Αυτό αναπαριστά το πρόβλημα Εφαρμογής HMM

Β. Δοθέντος του προβλήματος εκπαίδευσης στο HMM, τίθεται το ερώτημα πως ευρίσκεται η πιθανοφάνεια η οποία παράγεται από ακολουθία παρατηρήσεων εισερχόμενης ομιλίας; Αυτό αναπαριστά το πρόβλημα της αναγνώρισης.

2.4 Η Διακριτή Παρατήρηση του HMM- Τυπικές διαδικασίες

Η διακριτή παρατήρηση του μοντέλου HMM περιορίζεται στην παραγωγή ενός ορισμένου συνόλου διακριτών παρατηρήσεων . Με την χρήση των μεθόδων κβάντισης(VQ) ανυσμάτων, όλα τα ανύσματα παρατήρησης συγκεντρώνονται σε επιτρεπτό όριο. Στο σημείο αυτό εισάγεται η έννοια του κώδικα βιβλίου (codebook). Το codebook ÷ñçóéìïðïéåßôáé êáé ãéá ôçí êáÜíôéóç ôùí áíôóìÜôùí ðáñáôçñÞóåùò. Áí õðÜñ÷ïõí K áñéèìüò ðéèáíþí áíôóìÜôùí ðáñáôÞñçóçò ìå ôï âéâëßï êþäéêá , ôüôå åßíáé åýêïëï íá áíôéóôïé÷ßóïõìå óå êÜèå áêÝñáéï ìéá ðáñáôÞñçóç, Üò õðïèÝóïõìå k, 1<k<K.

Είναι εύκολο να σημειωθεί ότι για την αληθινή επεξεργασία υφίσταται άνυσμα κατάλληλο για το οποίο ενεργοποιείται η διαδικασία κβάντισης παρατηρήσεων μέχρις ότου οι μεταβλητές του ανύσματος είναι μόνο ακέραιοι. Γενικότερα, η κβαντισμένη περιοχή δυνατόν να αποτελέσει μια ευδιάκριτη έκφραση συμβόλου από ένα αλφάβητο k συμβόλων .Αυτό το τελευταίο προσδίδει πολλές φορές έμφαση στον χαρακτηρισμό διακριτό σύμβολο HMM. Κατά την εξέταση που γίνεται σε αυτό το εδάφιο είναι εύκολο να κατανοηθεί ότι για κάθε παρατήρηση αντιστοιχεί ένας ακέραιος.

Κατά την διακριτή παρατήρηση του HMM, προκύπτει ότι η κβαντισμένη παρατήρηση-συνάρτηση πυκνότητας πιθανότητας για δεδομένη κατάσταση i λαμβάνει k κρούσεις σε πραγματικό χρόνο. Σ την περίπτωση αυτή καλόν είναι να γνωρίζουμε την κατανομή πιθανότητας στα k σύμβολα για τα οποία υφίσταται ξεχωριστή κατάσταση για το καθένα με βάρος:

$$b(k|i) \stackrel{\text{def}}{=} P(\underline{y}(t) = k|\underline{x}(t) = i).$$

Από τον ορισμό του ανύσματος πιθανότητας καταλήγουμε στην έκφραση της εξίσωσης κατάστασης (state equation) και εξίσωσης παρατήρησης (observation equation)

$$\mathbf{p}(t) \stackrel{\text{def}}{=} \begin{bmatrix} P(\underline{y}(t)=1) \\ P(\underline{y}(t)=2) \\ \vdots \\ P(\underline{y}(t)=K) \end{bmatrix}. \quad \mathbf{p}(t) = \mathbf{B}\boldsymbol{\pi}(t), \quad \mathbf{p}(t) = \mathbf{B}\mathbf{A}^{t-1}\boldsymbol{\pi}(1).$$

Τελικά καταλήγουμε ότι η μαθηματική επεξήγηση του HMM είναι διαμορφωμένη κατά κάποιο τρόπο σύμφωνα με την γενική κατάσταση αλλά και να αποδίδει τις διακριτές παρατηρήσεις:

$$\mathcal{M} = [S, \boldsymbol{\pi}(1), \mathbf{A}, \mathbf{B}, [\mathbf{y}_k, 1 \le k \le K]],$$

2.5 Αναγνώριση με χρήση Διακριτής παρατήρησης HMM

Μεταξύ των δύο προβλημάτων του HMM , το πρόβλημα της αναγνώρισης είναι πιο εύκολο , για αυτό και θα εξετασθεί πρώτο . Υποθέτουμε ότι ένα πλήρως εκπαιδευμένο HMM αντιπροσωπεύει ένα λεξιλόγιο το οποίο αφορά λέξεις που έχουν την δυνατότητα να αναπαριστούν την ακολουθία παρατήρησης. Σε αυτό το σημείο θέλουμε να ελέγξουμε την πιθανότητα πιθανοφάνειας κάθε μοντέλου σχετικά με την ακολουθία παρατήρησης.

Ας ξεκινήσουμε ορίζοντας μερικές εκφράσεις που θα φανούν αργότερα χρήσιμες , κατά την επεξήγηση των μεγεθών . Ειδικότερα :

partial sequence: $y_{t_1}^{t_2} \stackrel{\text{def}}{=} [y(t_1), y(t_1+1), y(t_1+2), \ldots, y(t_2)].$

Forward
partial sequence $y_1^t \stackrel{\text{def}}{=} [y(1), y(2), \ldots, y(t)]$

Back ward
partial sequence $y_{t+1}^T \stackrel{\text{def}}{=} [y(t+1), y(t+2), \ldots, y(T)].$

ο όρος backward σημαίνει ότι ξεκινά από την τελευταία παρατήρηση. Η ερώτηση κλειδί που συνάγεται πλέον είναι :

΄΄τι εννοούμε με τον όρο πιθανοφάνεια για ένα HMM ;΄΄

Σε ισχύ ευρίσκονται δύο γενικές μετρήσεις της πιθανοφάνειας, οι οποίες χρησιμοποιούνται στο πρόβλημα της αναγνώρισης. Η κάθε μία οδηγεί ξεχωριστά σε συγκεκριμένο αλγόριθμο, γι αυτό και θα εξετασθούν ξεχωριστά.

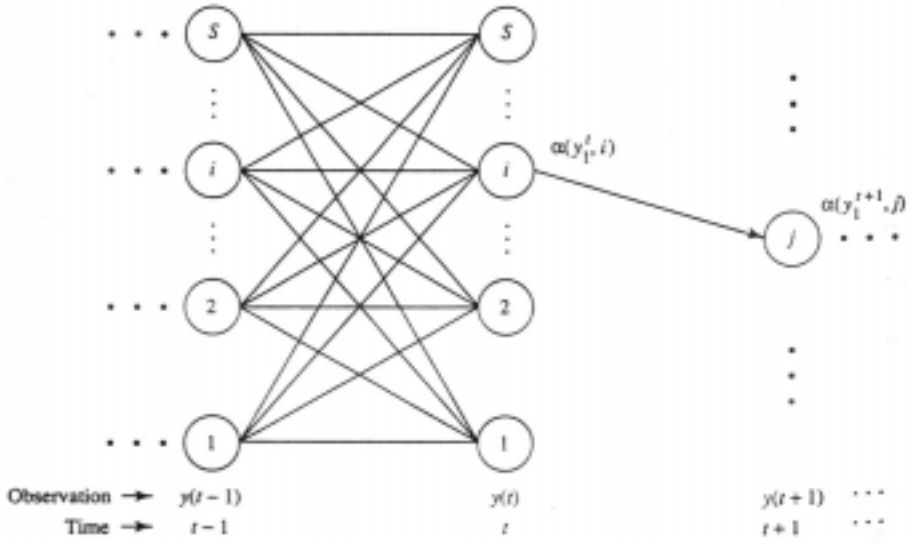
Μέθοδος ΄΄Any Path΄΄

Προσέγγιση 1(F-B). Η ονομασία της μεθόδου αυτής προήλθε από το ότι ο αλγόριθμος χρησιμοποιεί πιθανότητα η οποία βασίζεται σε παρατηρήσεις που έχουν παραχθεί μέσω ακολουθίας οποιασδήποτε κατάστασης διαμέσου του μοντέλου. Επειδή κατά τον υπολογισμό της πιθανότητας πιθανοφάνειας δημιουργείται συσσώρευση ογκωδών υπολογισμών μέσω του αθροίσματος **P(y(m)=Σj p(y,j/m)),** παίρνουμε τον αλγόριθμο **(F-B) Baum.** Ορίζεται μια ακολουθία προχώρησης και μία ακολουθία οπισθοδρόμησης. Από τις δύο ακολουθίες χρησιμοποιείται περισσότερο η πρώτη.

Από την ανάπτυξη των υπολογισμών προκύπτει ο υπολογισμός σε πίνακα μιας εξίσωσης από την οποία εξάγονται οι τιμές για την ακολουθία προχώρησης.

$$\alpha(y_1^{t+1}, j) = \alpha(y_1^t, i) P(\underline{x}(t+1) = j \,|\, \underline{x}(t) = i) \times P(\underline{y}(t+1) = y(t+1) \,|\, \underline{x}(t+1) = j) = \alpha(y_1^t, i) a(j|i) b(y(t+1)|j).$$

$$\alpha(y_1^{t+1}, j) = \sum_{i=1}^{S} \alpha(y_1^t, i) a(j|i) b(y(t+1)|j).$$



εικόνα 2.2 2Κατάσταση Πίνακα για την Εκφραση της Εμπρόσθιας Επανάληψης

Όμοια χρησιμοποιώντας την ανωτέρω διαδικασία είναι εφικτός ο υπολογισμός της ακολουθίας οπισθοδρόμησης:

$$\beta(y_{t+1}^{T}|i) = \sum_{j=1}^{S} \beta(y_{t+2}^{T}|j)a(j|i)b(y(t+1)|j).$$

Το επόμενο βήμα είναι ο υπολογισμός της πιθανότητας πιθανοφάνειας, όπου παρατηρείται η ευχέρεια της επίλυσης αν χρησιμοποιηθεί ο αλγόριθμος Backward.:

$$\alpha(y_1^1, j) = P(\underline{x}(1) = j)b(y(1)|j)$$

Τα βήματα του αλγόριθμου φαίνονται παρακάτω :

Initialization: Initialize $\alpha(y_1^1, j)$ for $j = 1, \ldots, S$ using (12.35).

Recursion: For $t = 2, \ldots, T$
For $j = 1, \ldots, S$
Update $\alpha(y_1^t, j)$ using (12.34).
Next j
Next t

Termination: Compute likelihood $P(y|\mathcal{M})$ using (12.40).

Απο την προσεκτική ανάλυση του αλγόριθμου αυτού είναι εμφανές ότι απαιτούνται O(S^2 T) βήματα . Αν υποτεθεί S=5 & T=100 , τότε απαιτούνται περίπου

$100 \times 5^2 = 2500$ βήματα και τα οποία μεταφράζονται σε εξήντα εννέα (69) τάξεις μεγεθών για να ολοκληρωθεί τελικά ο υπολογισμός . Τέλος είναι σημαντικό να σημειωθεί το πρόβλημα πράξεων το οποίο υφίσταται , επειδή κατά την κατασκευή του αλγόριθμου εκτελούνται αρκετοί των πολλαπλασιασμών μεταξύ των πιθανοτήτων. Αυτό επιδρά αρνητικά στην λειτουργία του όλου υπολογισμού του αλγόριθμου.

Με την κλιμακωτή εφαρμογή του αλγόριθμου εξαλείφονται τα ανωτέρω προβλήματα.

Προσέγγιση 2 (προσέγγιση χώρου καταστάσεων) Η μέθοδος προτάθηκε από τους DELLER και SNIDER για να μειωθεί το πρόβλημα πολυπλοκότητας που εμφανίζεται κατά την εξέταση της μεθόδου προσέγγισης 1.Τυπικά η πολυπλοκότητα είναι της τάξεως

O(3ST)=O(S^3/2 T)

Στην μέθοδο 2 επιχειρείται η ανάλυση του HMM σαν σύστημα χώρου κατάστασης. Εξετάζεται η εξίσωση κατάστασης και η εξίσωση παρατήρησης:

$$\boldsymbol{\pi}(t) = \mathbf{A}\boldsymbol{\pi}(t-1) + \boldsymbol{\delta}(t)\mathbf{u}(t)$$

$$\mathbf{p}(t) = \mathbf{B}\boldsymbol{\pi}(t).$$

Ο πίνακας που αντιστοιχεί στην εξίσωση καταστάσεως είναι διαγώνιος. Για τις ανάγκες του υπολογισμού δίνεται ο μετασχηματισμός **B=PSP^(-1).** Με βάση τα ανωτέρω προκύπτει , ότι εφόσον ο Α είναι διαγώνιος ο αριθμός των πράξεων που απαιτούνται είναι O(ST).Σε ορισμένες περιπτώσεις ο μέσος όρος κόστους ανά μοντέλο είναι **O[(1-K)ST].**

Η επιλογή αυτής της μεθόδου διαθέτει υπολογιστικά αλλά αριθμητικά δεδομένα που πλεονεκτούν ιδιαίτερα κατά την ανοικοδόμηση του τύπου , σε αντίθεση με την μέθοδο 1.

$$-\log P(y|m) = -\sum_{t=1} \log P(\underline{y}(t) = y(t)|m).$$

Ένας προσεκτικός έλεγχος των υπολογισμών που λαμβάνονται υπόψη και στις δύο ανωτέρω μεθόδους, δίνει το αποτέλεσμα ότι πρόκειται για περίπου ίδιους . Η δεύτερη μέθοδος επικεντρώνεται σε αναδιοργάνωση των υπολογισμών πίνακα.

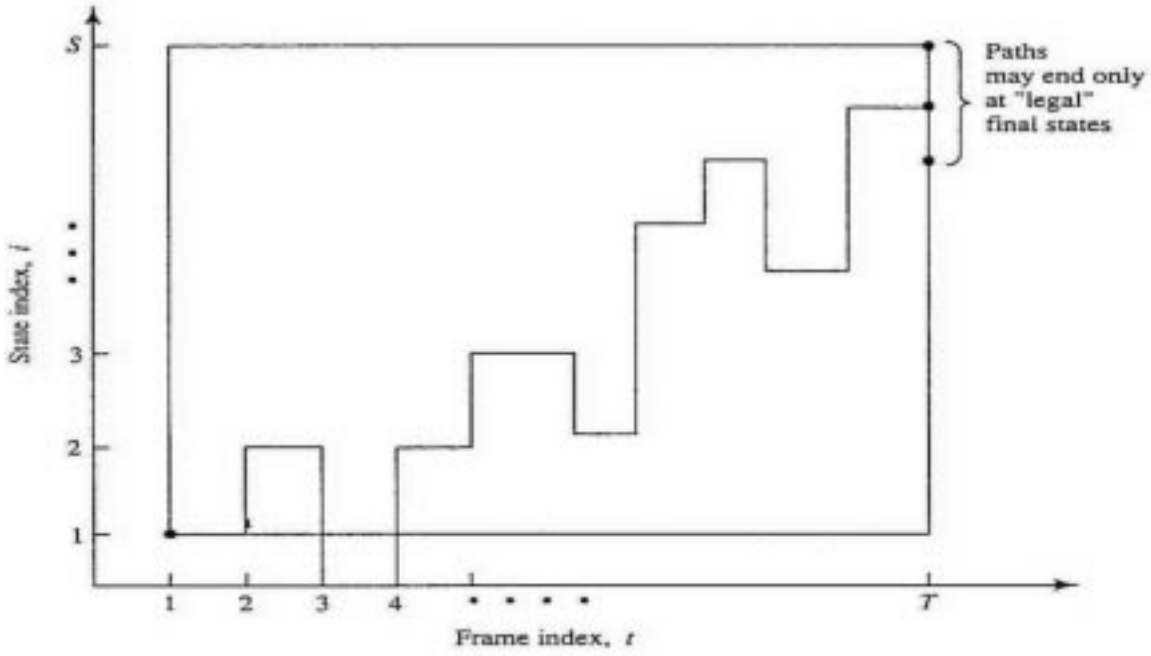
]Μέθοδος ΄΄Best Path΄΄

Προσέγγιση Viterbi. Κατά τις προηγούμενες προσεγγίσεις τύπου Any Path η υπολογισθείσα πιθανοφάνεια του μοντέλου HMM βασιζόταν στην πιθανότητα που δημιουργούσε το μοντέλο αυτό κάθεαυτό με χρήση ακολουθιών μήκους T . Μια άλλη μέθοδος υπολογισμού πιθανοφάνειας, η οποία είναι μικρότερη σε μέγεθος, βασίζεται στην πιθανότητα το μοντέλο HMM να δημιουργεί αυτόκλητα την βέλτιστη ακολουθία παρατηρήσεων χρησιμοποιώντας τις βέλτιστες καταστάσεις.

Παρατηρούμε το μέγεθος:

$$s^* \stackrel{\text{def}}{=} \underset{s}{\operatorname{argmax}}\, P(y, s|m)$$

Όπως διαφαίνεται το πρόβλημα εστιάζεται κυρίως σε πρόβλημα κανονικοποίησης ακολουθίας.Τα επι μέρους βήματα του προβλήματος παρουσιάζονται από το σχήμα όπως παρακάτω:



εικόνα 2.2 3 Εσχάρα Ερευνας HMM ως πρόβλημα DP

Μία εσχάρα ορίζεται με άξονες χρόνου κατάστασης. Κατά την ανάλυση των εξαγομένων από την εσχάρα, τίθενται δύο απλοί περιορισμοί:

Α Τα ακολουθιακά σημεία εσχάρας κατά μήκος οποιουδήποτε μονοπατιού να έχουν την μορφή (t,i) (t+1, j), 1<ι, j<S .Αυτό εκφράζει ότι κάθε μονοπάτι εξελίσσεται ανά μονάδα στον χρόνο

Β Τα τελικά σημεία εσχάρας κάθε μονοπατιού να είναι της μορφής (T, If) üðïõ If åßíáé ç ôåëéêÞ êáôÜóôáóç ôïõ ìïíôÝëïõ.

Óôï äéÜãñáììá öáßíåôáé ôï μονοπάτι ελάχιστου κόστους. Παρόλα αυτά, το κόστος πράξεων είναι μεγάλο , το οποίο ξεπερνάται με χρήση αρνητικών λογαρίθμων όπως το

$$D = \sum_{t=1}^{T} d'\left[(t, i_t)\middle|(t-1, i_{t-1})\right],$$

Στον ανωτέρω υπολογισμό του D καθορίζονται τρείς (3) τύποι κόστους :

$$d'_N(t, i) = b(y(t)|i) \stackrel{\text{def}}{=} P(\underline{y} = y(t)|\underline{x}(t) = i).$$
$$d'_T[(t, i)|(t-1, j)] = a(i|j) \stackrel{\text{def}}{=} P(\underline{x}(t) = i|\underline{x}(t-1) = j),$$
$$d'[(t, i)|(t-1, j)] = d'_T[(t, i)|(t-1, j)]d'_N(t, i)$$
$$= a(i|j)b(y(t)|i)$$

Με την χρήση των αλγόριθμων καθίσταται ευκολότερος ο υπολογισμός του κόστους για το ελάχιστο ικανό μονοπάτι , αλλά και το αντίστοιχο για το μέγιστο D. Ειδικότερα:

$$D_{\min}(t, i_t) \stackrel{\text{def}}{=} \text{distance from } (0,0) \text{ to } (t, i_t) \text{ over the best path.}$$
$$D_{\min}(t, i_t) = \min_{i_{t-1}}\left\{D_{\min}(t-1, i_{t-1}) + d'\left[(t, i_t)\middle|(t-1, i_{t-1})\right]\right\}$$

Περατώνοντας την ανάλυση η κατάληξη οδηγεί στην έκφραση του D ως:

$$D^* = \min_{\text{legal } i_T}\left\{D_{\min}(T, i_T)\right\}.$$

Η ελαχιστοποίηση της σχέσης για τα If καταλήγει στο συμπέρασμα ότι μεταξύ ΄λων των τελικών καταστάσεων παραμένει ως η επικράτεστερη . Στην τελικά διαμορφωμένη έκφραση των τελικών καταστάσεων είναι:

$$i_T^* = \underset{\text{legal } i_T}{\operatorname{argmin}}\left\{D_{\min}(T, i_T)\right\}.$$
$$D^* = -\log P(y, s^*|\mathcal{M})$$

Από την τελευταία σχέση είναι δυνατή η έκφραση της πιθανοφάνειας:

$$\Psi(t, i_t) = \underset{i_{t-1}}{\operatorname{argmin}}\left\{D_{\min}(t-1, i_{t-1}) + \left[-\log a(i_t|i_{t-1})\right] + \left[-\log b(y(t)|i_t)\right]\right\}$$
$$= \operatorname{argmin}\left\{D_{\min}(t-1, i_{t-1}) + \left[-\log a(i_t|i_{t-1})\right]\right\}.$$

Ο αλγόριθμος Vitterbi, ο οποίος φαίνεται στο σχήμα

Initialization: "Origin" of all paths is node (0, 0).

For $i = 1, 2, \ldots, S$

$D_{\min}(1, i) = a(i|0)b(y(1)|i) = P(\underline{x}(1) = i)b(y(1)|i)$

$\Psi(1, i) = 0$

Next i

Recursion: For $t = 2, 3, \ldots, T$

For $i_t = 1, 2, \ldots, S$

Compute $D_{\min}(t, i_t)$ according to (12.67).

Record $\Psi(i_t, j)$ according to (12.71).

Next i_t

Next t

Termination: Distance of optimal path $(0, 0) \rightarrow (T, i_T^*)$ is given in (12.68).

Best state sequence, $\mathcal{I}^*$, is found as follows:

Find i_T^* as in (12.69).

For $t = T - 1, T - 2, \ldots, 0$

$i_t^* = \Psi(t, i_{t+1}^*)$

Next t

Με μια σύντομη εκ νέου ματιά και στους δύο τύπους αλγορίθμων εξάγεται το συμπέρασμα ότι ο VITTERBI απαιτεί (S-1) T λιγότερες προσθέσεις για κάθε Hmm. Τόσο ο αλγόριθμος VITTERBI όσο και ο F-B μπορούν να δημιουργήσουν πιθανοφάνειες για κάθε δρόμο λύσεις οι οποίοι δεν εμφανίσθηκαν ποτέ στα δεδομένα κύλισης (training data). Αυτή η εξέλιξη αναπαριστά μια αποτελεσματική ευθυγράμμιση της κατανομής πιθανότητας που συνδυάζεται με ένα Hmm.

Αυτό-κανονικοποίηση

Μια από τις πλέον χρήσιμες ιδιότητες είναι ότι το Hmm διαθέτει σύμφωνη κανονικοποίηση που παρέχει. Το κύριο σημείο αυτής της ιδιότητας φαίνεται στο προηγούμενο σχήμα 12.4. Ο αριθμός των καταστάσεων στο κάθε μοντέλο είναι πάντοτε μικρότερος από το μήκος T της μεταβλητής. Αυτό σημαίνει ότι για κάθε δρόμο λύσης του HMM το μέγιστο μήκος του είναι T.

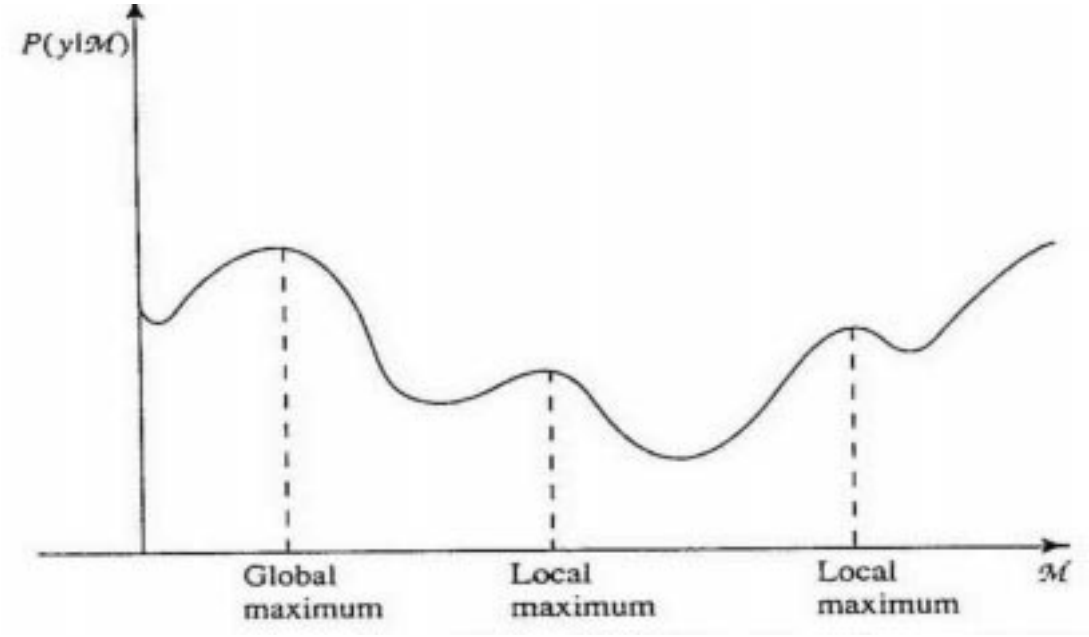
Μετά από Την λειτουργία αυτή καθίσταται εφικτό να ορισθεί η self-normalizing ιδιότητα ως συνέπεια της διπλά στοχαστικής δομής του Hmm η οποία στην ουσία επιτρέπει τη συνεχή δημιουργία σειρών αναφοράς με ακριβώς το ίδιο μήκος όπως η ακολουθία ελέγχου.

Εγκάρσια-Ερευνα

Η εγκάρσια έρευνα στο HMM λαμβάνει από την εξέταση της εσχάρας, μέσω της

$$D_{\min}(t, i_t) \leq D_{\min}(t, i_t^*) + \delta(t).$$

Η εγκάρσια έρευνα συμβαίνει τόσο για τον F-B αλγόριθμο όσο και για τον Viterbi.



εικόνα 2.2 4 Ορισμός Πιθανοφάνειας HMM

2.6 Εφαρμογή/Εκπαίδευση της Διακριτής Παρατήρησης HMM

Επανεκτίμηση του F-B:

Το αντικείμενο της επανεκτίμησης κατά την εφαρμογή ενός Hmm εστιάζεται στην σωστή αναπαράσταση συγκεκριμένων λέξεων. Εδικότερα χρησιμοποιούνται σειρές της μορφής y=y=(y(1).....(yt)) γίνει δυνατή η εύρεση του κατάλληλου μοντέλου της μορφής (12.21).

Εξάλλου στην περίπτωση του F-B αλγόριθμου τηρείται η διαδικασία αναλυτικής εκτίμησης του μοντέλου m το οποίο αντιστοιχεί σε τοπικό μέγιστο της πιθανοφάνειας P(y/m). Αυτή ακριβώς είναι και η μέθοδος εύρεσης των παραμέτρων σε Hmm.

Ειδικότερα ορίζονται τυχαίες μεταβλητές Yj(t) στις οποίες το μοντέλο λειτουργεί για συγκεκριμένο χρόνο t. Κατά την διαδικασία οι παρακάτω τέσσερις ποσότητες, των οποίων η ύπαρξη αντανακλάται σε εμφάνιση των κυρίων σημείων:

$$\xi(i,j;\bullet) = P(\underline{u}(\bullet) = u_{ji}|y,m) = \sum_{t=1}^{T-1} \xi(i,j;t),$$

$$\gamma(i;\bullet) = P(\underline{u}(\bullet) \in u_{\bullet i}|y,m) = \sum_{t=1}^{T-1} \gamma(i;t),$$

$$\nu(j;\bullet) = P(\underline{u}(\bullet) \in u_{j\bullet}|y,m) = \sum_{t=1}^{T} \nu(j;t),$$

$$\delta(j,k;\bullet) = P(\underline{y}_j(\bullet) = k|y,m) = \sum_{t=1}^{T} \delta(j,k;t) = \sum_{\substack{t=1 \\ y(t)=k}}^{T} \nu(j;t).$$

όπου

$n(u_{j/i})$ = αριθμός μεταπτώσεων για τυχαίες ακολουθίες παρατήρησης μήκους T

$n(u_{\bullet/i})$ = αριθμός μεταπτώσεων για το σύνολο $u_{\bullet/i}$ για τυχαία ακολουθία παρατήρησης μήκους T

$n(u_{j/\bullet})$= αριθμός στιγμών παρατήρησης K και κατάπτωσης j που υπάρχουν σε συνδυασμό μεταξύ τους κατά την τυχαία ακολουθία παρατήρησης μήκους T

Στη συνέχεια είναι εύκολο να δειχθεί ότι :

$$\xi(i,j;\bullet) = \mathcal{E}[\underline{n}(u_{ji})|y,m]$$

$$\gamma(i;\bullet) = \mathcal{E}[\underline{n}(u_{\bullet i})|y,m]$$

$$\nu(j;\bullet) = \mathcal{E}[\underline{n}(u_{j\bullet})|y,m]$$

$$\delta(j,k;\bullet) = \mathcal{E}[\underline{n}(\underline{y}_j(t) = k)|y,m].$$

Μετά τους παραπάνω ορισμούς είναι δυνατή η έκφραση των εκτιμήσεων αλλά και των παραμέτρων του μοντέλου προς εξέταση :

$$\mathcal{E}[\underline{n}(u_{j1i})|y, m] = [1 \times P(u_{j1i}|y, m)] + [0 \times P(\text{not } u_{j1i}|y, m)] = \xi(i, j; \bullet).$$

$$\bar{a}(j|i) = \frac{\mathcal{E}[\underline{n}(u_{j1i})|y, m]}{\mathcal{E}[\underline{n}(u_{\cdot 1i})|y, m]} = \frac{\xi(i, j; \bullet)}{\gamma(i; \bullet)},$$

$$\bar{b}(k|j) = \frac{\mathcal{E}[\underline{n}(y_j(t) = k)|y, m]}{\mathcal{E}[\underline{n}(u_{j1\cdot})|y, m]} = \frac{\delta(j, k; \bullet)}{\nu(j; \bullet)},$$

$$P(\underline{x}(1) = i) = \gamma(i; 1).$$

$$\bar{a}(j|i) = \frac{\sum_{t=1}^{T-1} \alpha(y_1^t, i)a(j|i)b(y(t+1)|j)\beta(y_{t+2}^T|j)}{\sum_{t=1}^{T-1} \alpha(y_1^t, i)\beta(y_{t+1}^T|i)},$$

$$\bar{b}(k|j) = \frac{\sum_{\substack{t=1 \\ y(t)=k}}^{T} \alpha(y_1^t, j)\beta(y_t^T|j)}{\sum_{t=1}^{T} \alpha(y_1^t, j)\beta(y_{t+1}^T|j)},$$

$$P(\underline{x}(1) = i) = \frac{\alpha(y_1^1, i)\beta(y_2^T|i)}{P(y|m)}$$

Δεδομένης της διαδικασίας που περιγράφηκε το μοντέλο m εκφράζεται σχετικά με το y :

m=argmax P(y/m)

οπωσδήποτε αυτή η συνάρτηση\ έκφραση του m διαθέτει πολλά μέγιστα. Το πλέον κατάλληλο μοντέλο m δηλώνεται ως σφαιρικό μέγιστο.

Τελικά καταλήγουμε στο συμπέρασμα ότι η διαδικασία F-B, όπως και η διαδικασία Viterbi συμπεριφέρονται και καλούνται **«επανεκτίμηση μέσω αναγνώρισης»** για λόγους που έχουν περιγραφεί ανωτέρω.

Επανεκτίμηση Viterbi

Αν και ο F-B αλγόριθμος είναι ο ενδεδειγμένος τρόπος για την εφαρμογή της διακριτής παρατήρησης HMM, ωστόσο μι απλούστερη και εξίσου αποτελεσματική οδός είναι αυτή του αλγόριθμου Viterbi. Από επανεκτίμηση του αλγορίθμου Viterbi προκύπτουν οι εξισώσεις :

$$\bar{a}(j|i) = \frac{n(u_{j1i})}{n(u_{\cdot 1i})},$$

$$\bar{b}(k|j) = \frac{n(\underline{y}_j(t) = k)}{n(u_{j1\cdot})}.$$

Ο αλγόριθμος Viterbi παρέχει έναν καλύτερο χαρακτηρισμό των λύσεων του μοντέλου που παρέχεται από τον F-B αλγόριθμο.

2.7 Συνεχής Παρατήρηση HMM

Στη γενικότερη περίπτωση των συνεχών παρατηρήσεων και ανυσματικά εκτιμηθέντων, η τυπική κατανομή του HMM περιέχει μια πολυμεταβαλλόμενη συνάρτηση πυκνότητας πιθανότητας.

Αναγνώριση

Στο πρόβλημα της αναγνώρισης, για κάθε εισερχόμενη παρατήρηση ορίζεται η πιθανοφάνεια της δημιουργούμενης παρατήρησης:

$$b(\mathbf{y}(t)|j) \stackrel{\text{def}}{=} f_{\Sigma|\Xi}(\mathbf{y}(t)|j).$$

Μετά από αυτό είναι αποδεκτές όλες οι μέθοδοι προσέγγισης που περιγράφηκαν και μπορούν να εφαρμοσθούν.

Ειδικότερα, επισημαίνεται ότι με την χρήση του αρνητικού λογάριθμου της b(y(t)) παράγεται μια μέγιστη πιθανοφάνεια –απόσταση της y(t) απο την αξία ττηςς μέσης τιμήςς, εφόσον υφίσταται η προΰπόθεση η συνάρτηση πυκνότητας πιθανότητας (pdf)έχει ορισθεί ως GAUSSIAN.

Εφαρμογή/Εκπαίδευση

Διαδικασία F-B: Το πλέον χρησιμοποιημένο μοντέλο είναι αυτό της GAUSSIAN ΠΥΚΝΌΤΗΤΑΣ ΔΕΙΓΜΑΤΟΣ, η οποία έχει την μορφή :

$$f_{\Sigma|\Xi}(\xi|i) = \sum_{m=1}^{M} c_{im}\, \mathcal{N}(\xi; \mu_{im}, \mathbf{C}_{im})$$

όπου **Cim**= συνιστώσα μείγματος για τα μ-οστά συστατικά στην κατάσταση i. Εκδηλώνει την χρήση της πολυμεταβλητής pdf με μέση τιμή μ και τον πίνακα τιμών Cim.

Αδό ôçν åîÝôáός ôùí óõíèçêþí êáé ôùí ðáñáìÝôñùí , ðñïêýðôåé üôé ïé óõíéóôþóåò δείγματος εχουν την έκφραση :

$$\bar{c}_{il} = \frac{v(i;\bullet,l)}{\sum_{m=1}^{M} v(i;\bullet,m)},$$

$$\bar{\mu}_{il} = \frac{\sum_{t=1}^{T} v(i;t,l)\mathbf{y}(t)}{v(i;\bullet,l)},$$

$$\overline{C}_{il} = \frac{\sum_{t=1}^{T} \nu(i;t,l)[\mathbf{y}(t) - \boldsymbol{\mu}_{il}][\mathbf{y}(t) - \boldsymbol{\mu}_{il}]^T}{\nu(i;\bullet,l)}$$

όπου **Cil** =λόγος των αναμενόμενων χρονικών στιγμών του μονοπατιού στην κατάσταση i που χρησιμοποιούν το συστατικό l τάξης προς τον αριθμό των μονοπατιών που μεταστασθμεύουν στην κατάσταση .

Όπως και στην περίπτωση της διακριτής παρατήρησης, έτσι και στο μοντέλο που επεξεργαζόμαστε οι ανωτέρω σχέσεις οδηγούν στην αναπαράσταση πλήρως του τοπικού μέγιστου της πιθανοφάνειας P(Y/m).

Η εύρεση του μέγιστου αυτού γίνεται με την βοήθεια του μεγέθους μil και Cil .Αυτό σημαίνει ότι κάθε στιγμή πρέπει να υφίσταται η δυνατότητα χρήσης των δεδομένων εκπαίδευσης/εφαρμογής για την επανεκτίμηση της ακολουθίας

Οπωσδήποτε όμως οι πυκνότητες αυτές απεικονίζουν μόνο ένα μέρος της περιγραφής του μοντέλου . Απαιτείται επιπλέον η εύρεση των εκτιμήσεων για τις πιθανότητες μετάδοσης κατάστασης και αρχικής κατάστασης. Ο πλήρης αλγόριθμος F-B έχει περιγραφεί προηγούμενα.

Διαδικασία Viterbi

Κατά την εφαρμογή μιας διαδικασίας Viterbi τα ανύσματα μέσης τιμής και οι πίνακες συμμεταβολής για τις πυκνότητες παρατηρήσεων επανεκτιμώνται με απλή αναλογική.

Αν η παρατήρηση **y(x)** παράγεται κατά την κατάσταση i , τότε ισχύει **y(x) → i.** Έστω ότι έχουν οριστεί **N(i) ανύσματα** παρατήρησης. Τότε :

$$\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{\substack{t=1 \\ \mathbf{y}(t) \sim i}}^{T} \mathbf{y}(t)$$

$$\mathbf{C}_i = \frac{1}{N_i} \sum_{\substack{t=1 \\ \mathbf{y}(t) \sim i}}^{T} [\mathbf{y}(t) - \boldsymbol{\mu}_i][\mathbf{y}(t) - \boldsymbol{\mu}_i]^T.$$

Οπωσδήποτε τα συστατικά δείγματος όταν **M>1**,εμφανίζουν σε μια κατάσταση ελέγχοντας τα ανύσματα παρατήρησης. Αυτό δύναται να πραγματοποιηθεί μέσω clustering, **K-means** αλγόριθμο για παράδειγμα με **K=M.** Για **Nil** ανύσματα παρατήρησης έχω ότι :

$$c_{il} = \frac{N_{il}}{N_i}.$$

Μερικά αποτελέσματα από το πείραμα ψηφιακής αναγνώρισης φαίνονται στο σχ.12.8

2.8 Πιθανότητες Διάρκειας Κατάστασης στην Διακριτή Παρατήρηση HMM

Ένα από τα πλεονεκτήματα που εμφανίζονται κατά την χρήση του HMM είναι ο αποκλειστικός χαρακτηρισμός της ακουστικής δομής της συμφωνίας που πρόκειται να μοντελοποιηθεί.

Πέραν των ανωτέρω, οι καταστάσεις ενός μοντέλου δύνανται να αναπαραστήσουν, σε πρώτη προσέγγιση, ανεξάρτητα ακουστικά φαινόμενα όπως ένας ήχος φωνήεντος σε μια λέξη ή μια μετάδοση φωνήματος. Επιπρόσθετα, σε αρκετές περιπτώσεις ο αριθμός καταστάσεων είναι τέτοιος ώστε να μπορεί να επεξηγήσει αυτού του είδους τα φαινόμενα.

Για παράδειγμα, η αναπαράσταση ενός φωνήματος από ένα HMM απαιτεί τρεις καταστάσεις - μία για να εκτελεσθεί η μετάδοση για κάθε ένα από τα άκρα του φωνήματος ήτοι σύνολο δύο και τέλος μια σταθερή κατάσταση αναφοράς. Είναι προφανές ότι το HMM αυτοργανώνεται προκειμένου να αποτελέσειένα αναλυτικό κριτήριο.

Παρά το ότι τα ακουστικά φαινόμενα δεν διαθέτουν διάρκεια εκθετικά κατανεμημένα, εν τούτοις το HMM διαθέτει εκθετικές κατανομές πιθανότητες. Αυτό εκφράζεται με το μέγεθος d στην :

$$P(\underline{d}_i = d) = [a(i|i)]^{d-1}[1 - a(i|i)].$$

Ο αντικειμενικός σκοπός είναι η εύρεση μιας διαδικασίας υπολογισμού της κάθε παραμέτρου ξεχωριστά του μοντέλου εκείνου για το οποίο μεγιστοποιείται η πιθανοφάνεια P(y/m). Η διαδικασία αυτή οπωσδήποτε θα περιέχει έναν F-B αλγόριθμο για την φάση της αναγνώρισης.

Από την ανάλυση της διαδικασίας προκύπτουν οι παρακάτω εκφράσεις :
Επανεκτίμηση πιθανοτήτων μετάδοσης :

$$\bar{a}(j|i) = \frac{\sum_{t=1}^{T} \alpha(y_1^t, i)a(j|i)\beta'(y_{t+1}^T|j)}{\sum_{j=1}^{S}\sum_{t=1}^{T} \alpha(y_1^t, i)a(j|i)\beta'(y_{t+1}^T|j)},$$

Αναλογία αναμενόμενου αριθμού στιγμών συμβ.κ :

$$\bar{b}(k|i) = \frac{\sum_{\substack{t=1 \\ y(t)=k}}^{T} \left[\sum_{s<t} \alpha'(y_1^s, i)\beta'(y_{s+1}^T|i) - \sum_{s<t} \alpha(y_1^s, i)\beta(y_{s+1}^T|i)\right]}{\sum_{k=1}^{K}\sum_{\substack{t=1 \\ y(t)=k}}^{T} \left[\sum_{s<t} \alpha'(y_1^s, i)\beta'(y_{s+1}^T|i) - \sum_{s<t} \alpha(y_1^s, i)\beta(y_{s+1}^T|i)\right]},$$

αναλογία αναμενόμενων αριθμών με διάρκεια d :

$$P(\underline{d}_i = d) = \frac{\sum_{t=1}^{T} \alpha^r(y_1^t, i) P(\underline{d}_i = d) \beta(y_{t+d+1}^T | i) \prod_{s=t+1}^{t+d} b(y(s)|i)}{\sum_{d=1}^{D} \sum_{t=1}^{T} \alpha^r(y_1^t, i) P(\underline{d}_i = d) \beta(y_{t+d+1}^T | i) \prod_{s=t+1}^{t+d} b(y(s)|i)},$$

Εκφραση της P(x(1)=i/y^t)

Παρότι έχουμε καταλήξει στο συμπέρασμα ότι οι πυκνωτές διάρκειας αποδεικνύουν την απόδοση του HMM, εν τούτοις η αιτιολόγηση έχει το δικό της κόστος. Το κόστος αυτό αναλύει στην παράμετρο D και τα μεγέθη που την αποτελούν. Στην περίπτωση αυτή μόνο οι παράμετροι της συνάρτησης πυκνότητας πιθανότητας εξετάζονται ως προς την καταλληλότητά τους.

Τελικά επισημαίνεται ότι η F-B προσέγγιση σε μοντέλα πυκνότητας διάρκειας επέχει εναλλακτικής λύσεως. Κατά την φάση αναγνώρισης, μια προσέγγιση κατά Viterbi όλου του μοντέλου είναι χρονοβόρα.

2.9 Κλιμάκωση του F-B Αλγόριθμου

Το μείζον πρόβλημα κατά την εφαρμογή του αλγόριθμου F-B είναι ο μεγάλος αριθμός των πολ/σμών και άλλων πράξεων.

Λαμβάνοντας την

$$\bar{a}(j|i) = \frac{\sum_{t=1}^{T-1} \alpha(y_1^t, i) a(j|i) b(y(t+1)|j) \beta(y_{t+2}^T | j)}{\sum_{t=1}^{T-1} \alpha(y_1^t, i) \beta(y_{t+1}^T | i)},$$

θα δείξουμε τα βήματα της διαδικασίας επανεκτίμησης απευθείας σε όρους των ακολουθιών F-B

$$\hat{\alpha}(y_1^1, i) \stackrel{\text{def}}{=} \frac{\alpha(y_1^1, i)}{\sum_{j=1}^{S} \alpha(y_1^1, j)}.$$

$$c(1) \stackrel{\text{def}}{=} \frac{1}{\sum_{j=1}^{S} \alpha(y_1^1, j)},$$

$$\hat{\alpha}(y_1^1, i) = c(1)\alpha(y_1^1, i).$$

$$\tilde{\alpha}(y_1^2, i) = \sum_{j=1}^{S} \hat{\alpha}(y_1^1, j) a(i|j) b(y(t)|i).$$

$$\tilde{\alpha}(y_1^2, i) = c(1)\alpha(y_1^2, i).$$

$$\hat{\alpha}(y_1^2, i) \stackrel{\text{def}}{=} \frac{\tilde{\alpha}(y_1^2, i)}{\sum_{j=1}^{S} \tilde{\alpha}(y_1^1, j)} = c(2)\tilde{\alpha}(y_1^2, i) = c(2)c(1)\alpha(y_1^2, i).$$

$$\tilde{\alpha}(y_1^t, i) = \sum_{j=1}^{S} \hat{\alpha}(y_1^{t-1}, j)a(i|j)b(y(t)|i)$$

$$\hat{\alpha}(y_1^t, i) = c(t)\tilde{\alpha}(y_1^t, i) = \left(\prod_{\tau=1}^{t} c(\tau)\right)\alpha(y_1^t, i),$$

$$c(t) = \frac{1}{\sum_{i=1}^{S} \tilde{\alpha}(y_1^t, i)}.$$

$$\hat{\alpha}(y_1^t, i) = \frac{\sum_{j=1}^{S} \hat{\alpha}(y_1^{t-1}, j)a(i|j)b(y(t)|i)}{\sum_{k=1}^{S}\sum_{j=1}^{S} \hat{\alpha}(y_1^{t-1}, j)a(k|j)b(y(t)|k)}.$$

$$\hat{\alpha}(y_1^t, i) = \frac{\sum_{j=1}^{S} \left(\prod_{\tau=1}^{t-1} c(\tau)\right)\alpha(y_1^{t-1}, j)a(i|j)b(y(t)|i)}{\sum_{k=1}^{S}\sum_{j=1}^{S} \left(\prod_{\tau=1}^{t-1} c(\tau)\right)\alpha(y_1^{t-1}, j)a(k|j)b(y(t)|k)}$$

$$= \frac{\alpha(y_1^t, i)}{\sum_{k=1}^{S} \alpha(y_1^t, k)}.$$

Φαίνεται ότι η ανωτέρω διαδικασία κλιμακοποιεί εκάστη α(y,i) επί του αθροίσματος όλων των καταστάσεων στο χρόνο .

Τελικά αφού έχουμε καθορίσει τη ίδια διαδικασία και για τον όρο β( ) υπολογίζουμε την πιθανότητα του μοντέλου P(y/m) ως

$$P(y|m) = \sum_{i=1}^{S} \alpha(y_1^T, i).$$

Τελικά καταλήγουμε στην

$$P(y|m) = \left(\prod_{\tau=1}^{T} c(\tau)\right)^{-1}.$$

και με χρήση του

$$-\log P(y|m) = \sum_{\tau=1}^{T} \log c(\tau),$$

2.10 Εφαρμογή με ακολουθίες Πολλαπλής Παρατήρησης

Με σκοπό να παρέξουν μια πιο ολοκληρωμένη παρουσίαση των στατιστικών μεταβλητών ώστε να μοιάζουν με παρούσες εκφράσεις , υφίσταται μια καλή μέθοδος για να εφαρμοσθεί ένα δοθέν HMM με πολλαπλές συμφωνίες

εκπαίδευσης. Καλόν είναι να σημειωθεί ότι οι εξισώσεις επανεκτίμησης του Baum-Welsh είναι ευθειακές (straight forward).

Αρχικά θεωρούμε ότι εργαζόμαστε σε μοντέλο με κανονική αντιστοίχηση και χωρίς πολλές υποθέσεις όσον αφορά την ύπαρξη των μεταβλητών.

Οι αρχικές πιθανότητες κατάστασης είναι της μορφής

Το μήκος στο οποίο γίνεται η παρατήρηση είναι T1.
Με την χρήση των σχέσεων :

$$\xi(i,j;\bullet) = P(\underline{u}(\bullet) = u_{ji}|y,m) = \sum_{t=1}^{T-1} \xi(i,j;t),$$

$$\gamma(i;\bullet) = P(\underline{u}(\bullet) \in u_{\cdot|i}|y,m) = \sum_{t=1}^{T-1} \gamma(i;t),$$

$$\nu(j;\bullet) = P(\underline{u}(\bullet) \in u_{j|\cdot}|y,m) = \sum_{t=1}^{T} \nu(j;t),$$

$$\delta(j,k;\bullet) = P(\underline{y}_j(\bullet) = k|y,m) = \sum_{t=1}^{T} \delta(j,k;t) = \sum_{\substack{t=1 \\ y(t)=k}}^{T} \nu(j;t).$$

μπορούμε να εκφράσουμε τα ορίσματα :

$$\bar{a}(j|i) = \frac{\sum_{l=1}^{L} \frac{1}{P(y^{(l)}|m)} \sum_{t=1}^{T_l-1} \alpha^{(l)}(y_1^t, i)\, a(j|i)\, b(y(t+1)|j)\, \beta^{(l)}(y_{t+2}^{T_l}|j)}{\sum_{l=1}^{L} \frac{1}{P(y^{(l)}|m)} \sum_{t=1}^{T_l-1} \alpha^{(l)}(y_1^t, i)\, \beta^{(l)}(y_{t+1}^{T_l}|i)},$$

$$\bar{b}(k|j) = \frac{\sum_{l=1}^{L} \frac{1}{P(y^{(l)}|m)} \sum_{\substack{t=1 \\ y(t)=k}}^{T_l} \alpha^{(l)}(y_1^t, j)\, \beta^{(l)}(y_{t+1}^{T_l}|j)}{\sum_{l=1}^{L} \frac{1}{P(y^{(l)}|m)} \sum_{t=1}^{T_l-1} \alpha^{(l)}(y_1^t, j)\, \beta^{(l)}(y_{t+1}^{T_l}|j)},$$

και τελικά να καταλήξουμε στην έκφραση :

$$\bar{b}(k|j) = \frac{\sum_{l=1}^{L} \sum_{\substack{t=1 \\ y(t)=k}}^{T_l} \hat{\alpha}^{(l)}(y_1^t, j)\, \hat{\beta}^{(l)}(y_{t+1}^{T_l}|j)}{\sum_{l=1}^{L} \frac{1}{P(y^{(l)}|m)} \sum_{t=1}^{T_l} \hat{\alpha}^{(l)}(y_1^t, j)\, \hat{\beta}^{(l)}(y_{t+1}^{T_l}|j)},$$

με την βοήθεια της σχέσεως:

$$\bar{b}(k|j) = \frac{\delta(j,k;\bullet)}{v(j;\bullet)}.$$

Τελειώνοντας , επισημαίνεται ότι η χρήση της **Viterbi** διαδικασίας για την επανεκτίμηση με πολλαπλές παρατηρήσεις είναι πολύ ευθεία.

2.11 Κριτήρια εναλλακτικής βελτιστοποίησης στην εφαρμογή του HMM

Μέχρι τώρα χρησιμοποιήθηκε η προσέγγιση μέγιστης πιθανοφάνειας για τον σχεδιασμό του **HMM**. Όπως διεφάνει η όλη φιλοσοφία βασίζεται στην μεγιστοποίηση του μεγέθους της πιθανότητας P(y/m) δημιουργίας ακολούθων παρατηρήσεων.

Παρά τα θετικά αποτελέσματα της μεθόδου στην πράξη εν τούτοις υπάρχουν δύο θεμελιώδη προβλήματα:

α) Το σήμα δύναται να μην προστίθεται στους περιορισμούς του HMM ή το μοντέλο να μην αποδίδεται με ακρίβεια.

β) Η προσέγγιση μέγιστης πιθανοφάνειας δεν περιέχει τίποτα από αρνητική εφαρμογή.

Στην εναλλακτική προσέγγιση , η επιτυχία συνίσταται στην εύρεση των HMM παραμέτρων που αυτές ελαχιστοποιούν την επικαλούμενη πληροφορία διάκρισης ή αλλιώς την ενδιάμεση εντροπία μεταξύ στατιστικών στοιχείων του σήματος και του μοντέλου. Η τεχνική απεικόνισης καλείται MDI (ελάχιστη πληροφορία διάκρισης)

Η DI είναι μια σχέση μέτρησης πιθανολογικής απόστασης

$$J_{DI} = \int_{-\infty}^{\infty} f_{\underline{v}|\varepsilon}(\mathbf{v}|1) \log \frac{f_{\underline{v}|\varepsilon}(\mathbf{v}|1)}{f_{\underline{v}|\varepsilon}(\mathbf{v}|2)} d\mathbf{x}$$

Υποθέτοντας την ύπαρξη R διαφορετικών HMM αναπαρίστανται από m1, m2, m3,.....,mR. Επιπλέον η παρατήρησή τους γίνεται με L ακολουθίες παρατήρησης μήκους T1, T2, T3,.....TL αντίστοιχα. Με μια μικρή κανονικοποίηση ο μέσος όρος αμοιβαίας πληροφορίας μεταξύ των τυχαίων ποσοτήτων y και m ε,ιναι :

$$\overline{M}(\underline{y}, \underline{m}) = \sum_{l=1}^{L} \sum_{r=1}^{R} P(\underline{y} = \mathbf{y}^{(l)}, \underline{m} = m_r) \log \frac{P(\underline{y} = \mathbf{y}^{(l)}, \underline{m} = m_r)}{P(\underline{y} = \mathbf{y}^{(l)}) P(\underline{m} = m_r)}.$$

Το πρόβλημα αυτό επικεντρώνεται στην επίλυση αυστηρά μη γραμμικών συνόλων εξισώσεων με διάφορους περιορισμούς. Δεν υπάρχει γνωστή κλειστού

τύπου επίλυση για την μεγιστοποίηση αυτή αλλά και μια συνήθης μέθοδος βελτιστοποίησης όπως αυτή του αλγόριθμου απότομης κατάδοσης πρέπει να εφαρμοσθεί. Για τον λόγο αυτό η μέθοδος είναι μέτρια υπολογιστικά και παρέχει εξίσου μέτρια αποτελέσματα.

3. Πρακτικές Εκδόσεις

Έχοντας ολοκληρώσει την παρουσία των πλέον σημαντικών από πλευράς θεωρίας σημείων του HMM και των αλγορίθμων του, ας δούμε τώρα τι διατίθεται στην πράξη.

3.1 Ακουστικές Παρατηρήσεις

Μέχρι στιγμής έχει γίνει μια περιληπτική θεώρηση των ανυσμάτων παρατήρησης (y(1), y(2), y(3),.....y(t))
χωρίς να δίνεται ιδιαίτερη σημασία τι περιέχουν αυτά τα ανύσματα. Πολλοί χαρακτηρισμοί τους έχουν δοθεί αλλά οι πλέον κοινοί είναι οι παράμετροι LP και cepstral.

Σε μια τυπική εφαρμογή η ομιλία δειγματοληπτείται σε **8Khz** και αναλύεται σε πλαίσια (frames) των **256 σημείων** με επικάλυψη των **156 σημείων.** Οι απολήξεις πλαισίων ανάλυσης είναι **m=100.200,.....M** στην γρήγορη ανάλυση. Αυτοί οι χρόνοι μεταβάλλονται σε χρόνους παρατήρησης **t=1,2,.....,T.** Για μια συνηθισμένη έκφραση λέξεων απαιτείται **M=8000 και T=80 παρατηρήσεις.**

Σε κάθε πλαίσιο υπολογίζονται **8-10 LP παράμετροι**, οι οποίες στην συνέχεια μετατρέπονται σε **12 συνιστώσες cepstral.** Σε τελική ανάλυση σε κάθε πλαίσιο περιλαμβάνονται μια μέτρηση ενέργειας μικρού χρόνου και μια διαφορική μέτρηση ενέργειας σε σύνολο 26 εξελίξεων ανά παρατήρηση.

3.2 Μέγεθος και Δομή του μοντέλου.

Μέχρι τώρα είναι δεκτό από την ροή της εξέτασης ότι το HMM χρησιμοποιείται συχνά σε υψηλότερου επιπέδου συσκευές αναγνώρισης ομιλίας που προβάλλουν γλωσσολογικούς περιορισμούς. Στην παρούσα εξέταση θα περιοριστεί το ενδιαφέρον στο ακουστικό επίπεδο για τα HMM.

Ένα άλλο στοιχείο είναι η δομή και το μέγεθος του μοντέλου. Με τον όρο δομή εννοούνται όλα τα στοιχεία τα οποία δίνουν τη μορφή των επιτρεπτών καταστάσεων μετάδοσης , ενώ με τον όρο μέγεθος τον αριθμό των καταστάσεων που περιέχονται στο μοντέλο.

Όπως έχει αποδειχθεί πειραματικά οι καταστάσεις απεικονίζουν ιδανικευμένα ακουστικά φαινόμενα. Έτσι λοιπόν ο αριθμός των καταστάσεων είναι σύμφωνος με τον αριθμό των ακουστικών φαινομένων. Επί παραδείγματι, εάν οι λέξεις έχουν μοντελοποιηθεί με διακριτές παρατηρήσεις τότε 5-10 καταστάσεις είναι αρκετές για να εκφράσουν τους ανεξάρτητους ήχους της λέξεως.

Ως απλός υπολογισμός , το μέσο μήκος των εκφράσεων είναι το κατάλληλο μέσο για τον καθορισμό του αριθμού των απαραίτητων καταστάσεων. Μια πιο ακριβής περιγραφή του αξιώματος αυτού αποτελεί το **fenone** που θα περιγραφεί στην συνέχεια. Το fenone είναι μια ακουστική μονάδα η οποία επιτυχώς είναι μήκους ενός πλαισίου. Στο σχήμα 0-4 φαίνονται τα **phone και fenone models.**

Η πλέον γενική μορφή του HMM είναι η γνωστή ως εργοδική η οποία επιτρέπει τη ύπαρξη απεριόριστων καταστάσεων μετάπτωσης. Κανένα από τα στοιχεία του πίνακα μετάπτωσης A δεν περιορίζεται η τιμή του στο μηδέν. Ένα παράδειγμα δομής αυτής της μορφής φαίνεται στο σχήμα 12.12, με ένα μοντέλο έξι καταστάσεων όπου είναι εμφανές ότι η μορφή αυτή δεν συνάδει με το μοντέλο ακολουθιακής κατάταξης ώστε να είναι σύμφωνο με το σήμα. Το εργοδικό μοντέλο όταν χρησιμοποιείται με την ομιλία διαθέτει την ευκολία εκείνη με την οποία αναπαριστάται με την ακολουθιακή δομή.

Το πιο ιδανικό μοντέλο στην αναγνώριση ομιλίας είναι το μοντέλο Bakis έξι καταστάσεων στο σχήμα 12.13. Η επιλογή ενός τέτοιου περιορισμένου μοντέλου δεν απαιτεί καμιά αλλαγή των διαδικασιών εφαρμογής. Στην περίπτωση F-B οι καταστάσεις που έχουν τιμή D παραμένουν μηδέν σε όλο το μήκος.

Η χρήση codebook υποβοηθεί την εξέλιξη των μετρήσεων. Με το σχήμα 12.12 φαίνεται η αντιμετώπιση των διαστροφών ομιλίας. Ειδικότερα επισημαίνεται η χρήση του VQ codebook όταν υφίσταται η διακριτή παρατήρηση. Συνήθως οι αναλυτές συσκευές ομιλίας χρησιμοποιούν codebooks μεγέθους 32-256.

Σε εφαρμοσμένη περίπτωση αλλά μη συνήθης, ένας ομιλητής απαιτείται να περιορίσει τους ήχους που παράγει, εξαιτίας της ανικανότητας ομιλίας πιθανώς. Στην περίπτωση αυτή απαιτείται ένα πιο απλό codebook ακόμα και με ένα σχετικά πολύπλοκο λεξιλόγιο.

3.3 <u>Ανάπτυξη με ανακριβή δεδομένα</u>

Από την μέχρι τώρα εξέταση καθίσταται σαφές ότι απαιτούνται πολλά δεδομένα για την ακριβή ανάπτυξη ενός μοντέλου HMM. Για παράδειγμα αν θεωρήσουμε την πιθανότητα b(j/i)=δ.
Οι υπόλοιπες πιθανότητες θα ορισθούν εξαιτίας ενός μοντέλου m και θα ακολουθιακών στιγμών είναι υπό μορφή αθροίσματος.

Μια πιο τυπική προσέγγιση ώστε να εφαρμόζονται ανακριβή δεδομένα είναι αυτή που καλείται ως διαγραφείσα μεταβολή (deleted interpolation) η οποία προτάθηκε από τον Bahl το 1983. Για την κατανόηση της μεθόδου αυτής , είναι απαραίτητο να εξετασθεί πρώτα η σχέση των συνδεδεμένων καταστάσεων..

Δύο καταστάσεις είναι συνδεδεμένες, όταν μοιράζονται κοινές κατανομές πιθανότητας παρατηρήσεων.

Αυτό είναι εύκολο να κατανοηθεί από το σχήμα στο οποίο ακολουθεί και εμφανίζεται η έκφραση του HMM 3 καταστάσεων υπό μορφή συνδεδεμένων καταστάσεων.
Το πλεονέκτημα της εφαρμογής των συνδεδεμένων καταστάσεων είναι ότι για το ίδιο ποσό δεδομένων μπορούν να αφαιρεθούν περισσότερες παράμετροι δίνοντας έτσι την δυνατότητα ύπαρξης μικρότερης μεταβολής.

Επιστρέφοντας στην αρχική κατεύθυνση της επεξήγησης της μεθόδου Deleted interpolation , αυτό θα συμβεί προκειμένου να απαντηθεί το ερώτημα :

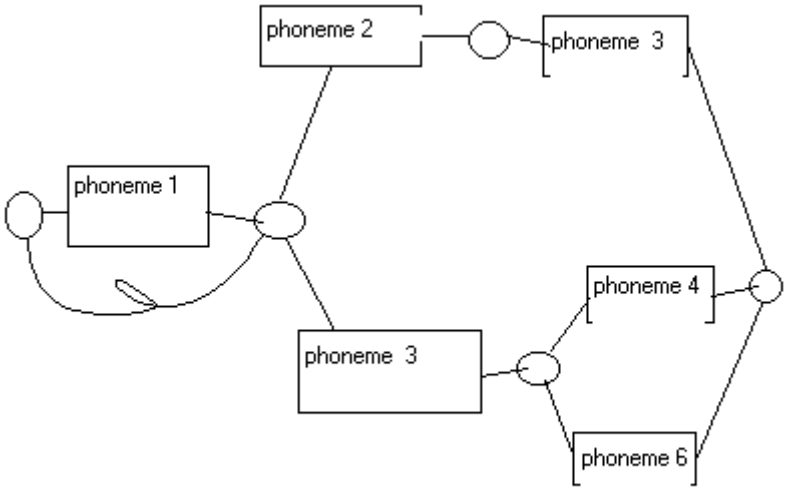
Ποιο από τα δύο μοντέλα HMM είναι προτιμητέο;

αυτό με τη συνήθη μορφή ή αυτό με την των 3 καταστάσεων

Η μέθοδος **Deleted interpolation** είναι μια μέθοδος που χρησιμοποιεί ένα υβριδικό μοντέλο στο οποίο περιέχονται αυτόματα οι κατάλληλες αναλογίες από κάθε μοντέλο. Η μορφή του μοντέλου που προκύπτει από την εφαρμογή της μεθόδου έχει

= πίνακας καταστάσεων μετάπτωσης για μοντέλο
= πίνακας καταστάσεων παρατήρησης και
=

Στην ουσία η μέθοδος αυτή συνήθως χρησιμοποιείται στην περιγραφή για λίγο πιο πολύπλοκες διαδικασίες από αυτή που περιγράφηκε. Υ



Εικόνα Σφάλμα! Δεν υπάρχει κείμενο καθορισμένου στυλ στο έγγραφο.**-4**
Συγκρότηση & λειτουργία phoneme

3.4 Ακουστικές Μονάδες μοντελοποιημένες από HMM

Εδώ στο τμήμα αυτό θα εξετασθεί η μοντελοποίηση HMM ακουστικών συσκευών με γνώμονα την καλύτερη δυνατή βελτιστοποίηση του σήματος.

Ξεκινάμε με το **phone.** Η πιο απλή μορφή χρήσης του phone είναι η λεγόμενη **context-independent** για την οποία ένα απλό μοντέλο κατασκευάζεται για κάθε phone. Το phone χρησιμοποιείται σε πολλά ακουστικά συστήματα. Έτσι υφίστανται και λειτουργούν triphones μοντέλα. Για τα **triphones** μοντέλα και προκειμένου να εξασφαλισθεί η επεξεργασία ακριβών δεδομένων, έχουμε την μέθοδο αντικειμένου τριφώνου.

Τα context independent phones χρησιμοποιούνται ως λέξεις μικρής συνάρθρωσης (μονοσύλλαβες, δισύλλαβες). **Οι επικαλούμενες λέξεις λειτουργικές απαιτούν 4% του λεξιλογίου στην DARPA 30% των ομιλουμένων λέξεων και ευθύνονται για το 50% των σφαλμάτων.**

Άλλα είδη phones είναι τα δίφωνα (ζεύγη phones) στα οποία ορίζονται το στατικό μοντέλο και το μοντέλο μεταπτώσεων.

Ανεβαίνοντας την ιεραρχία των γλωσσολογικών εργαλείων, αναλύεται το φώνημα μια ολοκληρωμένη μονάδα που δύναται να έχει πολλαπλές ακουστικές εφαρμογές. Από την θέα δομής μοντέλων , το φώνημα δύναται να αποτελέσει ένα δίκτυο δομημένο σε μοντέλα phones. Στο σχήμα 12.16 εμφανίζονται οι ανωτέρω μορφές .

Πάνω από το φώνημα στην ακουστική ιεραρχία συγκαταλέγονται η συλλαβή και τα μοντέλα με συλλαβή κατά το ήμισυ.

Στο κάτω μέρος της ακουστικής ιεραρχίας βρίσκεται μια ασυνήθιστη ακουστική μονάδα το fenone, η οποία επινοήθηκε και ονομάσθηκε από τους αναλυτές της IBM. Το fenone εφαρμόσθηκε με επιτυχία από το 1980. Εστιάζεται στο πρόβλημα σύμφωνα με το οποίο ο κυματικός τύπος της ομιλίας για κάθε λέξη μειώνεται σε ακολουθία παρατήρησης ανυσματικά κβαντισμένης. Παρέχεται Δε η δυνατότητα να δημιουργηθεί μια ακολουθία παρατηρήσεων για οποιαδήποτε έκφραση της ίδιας λέξης. Μπορούμε να μοντελοποιήσουμε την απόκλιση αντικαθιστώντας κάθε παρατήρηση με αυθεντική ακολουθία ανάπτυξης μέσω ενός ικρού HMM ικανού να δημιουργήσει το περιβάλλον μεταβλητότητας του αυθεντικού μοντέλου παρατήρησης.

Πειραματικά τα fenones έχουν μειώσει την αναλογία σφάλματος σε 28% . Στο σύστημα TANGORA χρησιμοποιούνται ευρέως.

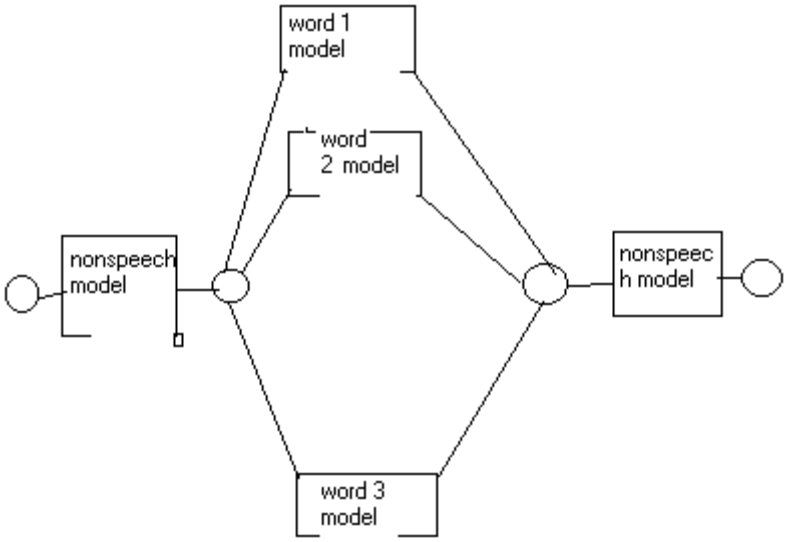
3.5 Διερεύνηση Συστημάτων Αναγνώρισης πουβασίζονται σε HMM

Στην παράγραφο αυτή γίνεται προσπάθεια να παρουσιασθούν οι δυνατότητες και οι στρατηγικές στις οποίες χρησιμοποιούνται τα HMM καθώς και η απόδοσή τους σε κάθε περίπτωση.

Τα συστήματα ομιλίας διακρίνονται σε δύο είδη:

α IWR = αναγνώριση μέσω μεμονωμένης λέξεως

β CSR = αναγνώριση συνεχούς ομιλίας



Εικόνα Σφάλμα! Δεν υπάρχει κείμενο καθορισμένου στυλ στο έγγραφο.**-5 Απεικόνιση HMM μη ομιλίας στο αρχικό και τελικό στάδιο**

3.5.1 IWR άνευ συντακτικού

Η απλούστερη περίπτωση των προθέσεων αναγνώρισης είναι η αναγνώριση της διακριτής λέξεως που ομιλείται ξεχωριστά. Στα πλαίσια της πρόθεσης αυτής είναι και η προσέγγιση κατά HMM , με δύο κύρια σημεία κατά την εφαρμογή του :

α Ένα απλό HMM χρησιμεύει για την αναπαράσταση μιας λέξεως

β Μονάδες υποδιαίρεσης της λέξης μπορούν να μοντελοποιηθούν από το HMM

Υφίστανται τρείς(3) περιπτώσεις εφαρμογής της αναγνώρισης μέσω HMM:

Α Η πρώτη περίπτωση περιγράφεται από την αυτόματη τοποθέτηση των σημείων πέρατος της κάθε λέξης και τα διαχωρίζουν από την ομιλία τόσο στην εφαρμογή όσο και στην αναγνώριση όσον αφορά τις προθέσεις τους . Στην περίπτωση αυτή το μοντέλο αναπαριστά μόνο ομιλία και ως εκ

τούτου οι ακολουθίες παρατήρησης θα πρέπει να αναγνώρίζονται πολύ προσεκτικά ώστε να αναπαριστούν μόνο ομιλία .

β Η δεύτερη περίπτωση περιέχει στο αρχικό στάδιο μη ομιλία και στα δύο άκρα των δειγμάτων ώστε να λαμβάνεται υπόψη σε κάθε ανεξάρτητο μοντέλο. Κατά την διάρκεια της αναγνώρισης , οι ακολουθίες παρατήρησης δυνατόν εναλλακτικά να περιέχουν δύο (2) άκρα μη ομιλίας.

Γ Η τρίτη περίπτωση είναι η ανάπτυξη ξεχωριστών μοντέλων για την ομιλία στα δύο (2) άκρα του μοντέλου και τα οποία αναπαριστούν μόνο λέξεις. Εδώ οι εισερχόμενες ακολουθίες παρατήρησης για αναγνώριση δυνατόν να έχουν μη ομιλία και στα δύο άκρα τα οποία κυρίως λειτουργούν για το μοντέλο.

Μια αξιόλογη εξέλιξη της διαδικασίας που περιγράφηκε ανωτέρω είναι οι περιορισμοί ανάμεσα στην ομιλία και στην μη ομιλία δυνατόν να μην είναι γνωστή σε δεδομένα ανάπτυξης.

Κατά την αναγνώριση φάσης , το μεγάλο ΗΜΜ ΔΊΚΤΥΟ μπορεί να ερευνηθεί κατά τον ίδιο τρόπο όπως και με κάθε άλλο ΗΜΜ. Μια κάθετη έρευνα είναι το πλέον χρήσιμο εργαλείο σε τέτοια περίπτωση . Κατά΄την ολοκλήρωση της έρευνας, η αναγνωρισθείσα λέξη ανακαλύπτεται μέσω οπισθοδρόμησης διαμέσου του δρόμου μέγιστης πιθανοφάνειας.

4 ΣΥΜΠΕΡΑΣΜΑΤΑ

α Η συνεχής ομιλία αποτελεί την πιο ολοκληρωμένη έκφραση μεγάλων λεξιλογίων . Επομένως η χρήση γλωσσολογικών εργαλείων είναι απαραίτητη.

Β Τα συστήματα που αποτελούν εργαλεία για την απεικόνιση του CSR χρησιμοποιούν την περιοχή έρευνας των ΗΜΜ ΛΌΓΩ ΤΗΣ ΙΔΙΌΤΗΤΑΣ ΤΩΝ ΝΑ ΑΥΤΟ ΟΡΓΑΝΩΝΟΝΤΑΙ ανάλογα με τις απαιτήσεις.

Γ Το ΗΜΜ κατέχει μείζονα ρόλο τόσο στην ακουστική όσο και στην γλωσσολογική επεξεργασία εξαιτίας του γεγονότος ότι αυτό λειτουργεί κατά τον ίδιο τρόπο με της μηχανής ΄΄FINITE STATE AUTOMATON΄΄, όπου η δημιουργία της λέξης γίνεται με βάση την ισχύουσα γραμματική .