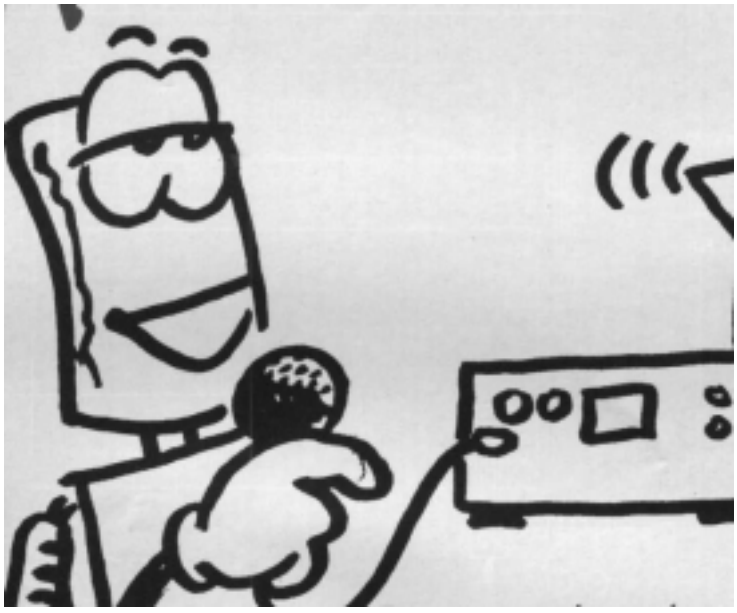
ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
Μεταπτυχιακό Πρόγραμμα Σπουδών

Νεκτάριος Μαργαρίτης (Α.Μ 97517)

Περιβάλλον ομιλίας–Ανθεκτική ανάλυση και αναπαράσταση ομιλίας



Εργασία στο μάθημα: **Επικοινωνία με ομιλία**
Διδάσκων :κ. **Γεώργιος Κουρουπέτογλου**

Αθήνα 1999

ΠΑΡΑΤΗΡΗΣΗ :

*Στο πρώτο κεφάλαιο της εργασίας μας ,**πραγματοποιήσαμε αυστηρή-επακριβή μετάφραση για όλο το κεφάλαιο**

*Στα ακόλουθα δυο κεφάλαια ,**κάναμε αυστηρή μετάφραση** , και **μεταφράσαμε το μεγαλύτερο μέρος της ύλης**

ΠΕΡΙΛΗΨΗ :

Ξεκινώντας ,θα ασχοληθούμε με τις *διάφορες μορφές θορύβου .Θόρυβος* ,είναι ήχοι που μεταδίδονται σ'ένα σύστημα αναγνώρισης και δεν είναι μέρος του σήματος εισόδου που φέρει τη πληροφορία . Τέσσερις βασικοί τύποι επηρεάζουν τα συστήματα αναγνώρισης ομιλίας :

- *θόρυβος υποβάθρου*
- *θόρυβος καναλιού*
- *θόρυβος από την ομιλία*
- *οι Χωρίς Μήνυμα φωνήσεις*

(ενώ. το *αντίξοο περιβάλλον* έχει τουλάχιστον ένα απ'τα γνωρίσματα : *πολύ θορυβο ,αγνωστες ιδιότητες θορύβου)*

Έτσι ,ο θόρυβος υποβάθρου παράγεται στο μέρος όπου η ομιλία γεννιέται , και εισάγεται μαζί με την ομιλία μας στη συσκευή εισόδου(εμφανίζεται σε κάθε συχνότητα).Ο θόρυβος καναλιού παράγεται από τα συστήματα (συσκευές) εισόδου που μεταφέρουν την ομιλία στο κυρίως σύστημα αναγνώρισης .Επίσης θα θεωρήσουμε ότι υ*πάρχουν δυο βασικά κανάλια ,τα μικρόφωνα και τα τηλέφωνα* . Έτσι η παραμόρφωση οφείλεται : στην απόκριση των συσκευών εισόδου και στο προσθετικό ηλεκτρικό θόρυβο του καναλιού. Τέλος ,για τον Χωρίς Μήνυμα (Non Communication) θόρυβο ομιλίας ,αυτός μπορεί να είναι ο ήχος από τα χείλια στην αρχή μίας πρότασης ,κομπιάσματα, διορθώσεις λαθών στη μέση των λέξεων , κομμένες λέξεις, μη γραμματικές δομές ,επιφωνήματα όπως το «ωχ» κ.α ,ενώ η ομιλία Lombard περιγράφει τις αλλαγές που προκαλούνται στην ομιλία, όταν ο ομιλητής προσπαθεί να ακουστεί πάνω από τον θόρυβο

Σκοπός λοιπόν σε αυτό το κεφαλαίο είναι να επέμβουμε άμεσα στο θόρυβο ,με τη κατάλληλη επιλογή και χρήση μικροφωνου.Τα μικρόφωνα ποικίλουν σε ποιότητα και τύπο .Θα ασχοληθούμε κυρίως με τα ΜΟΝΟΚΑΤΕΥΘΥΝΤΙΚΑ(αποκρίνονται σε ήχους από μια συγκεκριμένη κατεύθυνση),τα *πολυκατευθυντικά* ,τα *αφαίρεσης θορύβου* και *τα κοντινής ομιλίας* (κατασκευάζονται ώστε να αποκρίνονται στους ήχους που γεννιούνται κοντά ή απέναντι της ηχητικής πηγής)μικρόφωνα.

Ακόμα πιο ουσιαστικό για την ελάττωση θορύβου, είναι το επόμενο κεφάλαιο , αφού εστιάζει στην ανθεκτική ανάλυση ομιλίας .Μελετάμε την *PLP ανάλυση*(που ενσωματώνει 3 γνωρίσματα - ψυχοακουστικες έννοιες ,δηλ επαληθευμένες-ιδιότητες του ανθρώπινου οργανικού συστήματος), ενώ τεράστιο ενδιαφέρον έχουν τα τρία *Ακουστικά μοντέλα*(που ξεχωρίζουν το σήμα εισόδου από τον θορυβο ,με τον ίδιο τρόπο που το κάνουν και οι άνθρωποι) που αναπτύσσονται. Ολοκληρώνουμε το κεφαλαίο ,με *την ανθεκτική εκτίμηση σφάλματος* και τα μοντέλα *ARMA*

Στο τελευταίο κεφάλαιο αναδεικνύουμε τη σημασία *της φασματικής δυναμικής ,του φασματικού δυναμικου φιλτραρισματος ,των τεχνικών κανονικοποίησς ,και της μοντελοποίησς στο πεδίο αυτοσυσχέτησης για ανθεκτικό ASR* . Κλείνουμε το κεφάλαιο ,με την εργασία που έχει γίνει πάνω στη *φασματική λειανση cepstral* και στα μέτρα ομοιότητας .

ΠΕΡΙΕΧΟΜΕΝΑ :

ΚΕΦΑΛΑΙΟ 8 : Η ΧΡΗΣΗ ΜΙΑΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ ΑΝΘΕΚΤΙΚΗΣ ΟΜΙΛΙΑΣ (Jean-Claude Junqua ,Jean-Paul Haton)

*** ΜΕΤΑΦΡΑΣΗ ΑΓΓΛΙΚΩΝ ΟΡΩΝ**
***ΠΙΝΑΚΑΣ ΣΥΝΤΜΗΣΕΩΝ**

ΚΕΦΑΛΑΙΟ 7 :ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΟΜΙΛΙΑΣ

ΕΙΣΑΓΩΓΗ

Ένα σύστημα αναγνώρισης ομιλίας αναμένεται να λειτουργεί κατάλληλα στο περιβάλλον στο οποίο στόχος είναι να εφαρμοστεί . Αυτό το περιβάλλον μπορεί να ποικίλει , από ένα ήσυχο, προσωπικό γραφείο έως ένα εργοστάσιο, μια προκυμαία εκφόρτωσης ,ή ένα αυτοκίνητο .Όσο αυξάνονται οι εφαρμογές αναγνώρισης ομιλίας σε αριθμό και ποικιλία, τόσο χρειάζονται να υπερνικήσουν τις προκλήσεις από το αυξανόμενο σύνολο των αντίξοων περιβάλλοντων ομιλίας.

Δεν μπορούμε να κατασκευάσουμε συστήματα αναγνώρισης ομιλίας τα οποία λειτουργούν μόνο μέσα σε ένα ήσυχο , ελεγχόμενο περιβάλλον εργαστηρίου. Σε πραγματικές συνθήκες θα υπάρχουν πολλές φωνές και θόρυβος υπόβαθρου (όπως ηχοι από το δρόμο , μουσική ή το βουητό ενός κλιματιστικού) για να τους ανταγωνιστούμε. (Pat Russo, President, AT & T Business Communivations Systems, ομιλία του Keynote στο Advanced Speech Applications and Technology συνέδριο 1994).

Η ικανότητα ενός συστήματος αναγνώρισης να λειτουργεί με ακρίβεια κατω από αντίξοες συνθήκες ονομάζεται Ανθεκτικότητα .Η ανθεκτικότητα είναι κρίσιμη για την επιτυχία μιας εφαρμογής ,διότι εάν οι χρήστες πρέπει να `αγωνιζονται` για να κάνουν ένα αναγνωριστή ομιλίας να λειτουργήσει , δεν θα το χρησιμοποιήσουν.

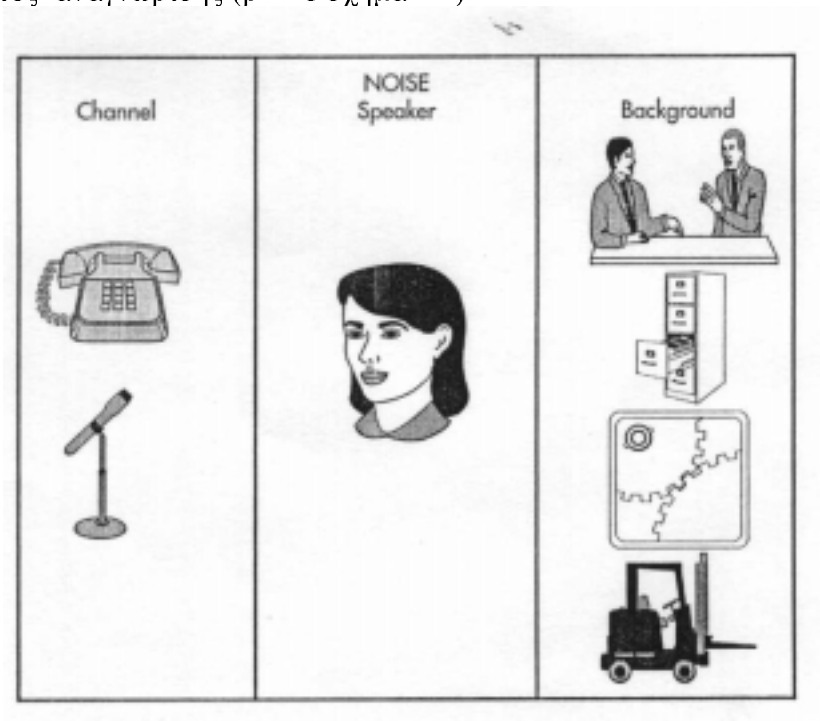
Η τεχνολογία εστιάζει ξεκινώντας με μια κατάταξη των τύπων του θορύβου που επηρεάζουν την λειτουργία ενός συστήματος αναγνώρισης, και τους Τρόπους μέσα από τους οποίους το περιβάλλον ομιλίας συγκρούεται με τη διαδικασία αναγνώρισης ομιλίας, συμπεριλαμβάνοντας την απόκριση του ομιλητή στο θόρυβο υπόβαθρου .Εξετάζει τεχνολογίες που χρησιμοποιούνται για να περιορίσουν ή να ελαχιστοποιήσουν τις αρνητικές επιδράσεις του ήχου (ονομάζονται τεχνικές περιορισμού θορύβου) και η έρευνα κατευθύνεται στον εμπλουτισμό του σήματος ομιλίας , από μόνο του (καλείται εμπλουτισμός ομιλίας).Η ακρίβεια της εφαρμογής ,επανεξετάζει την (αρνητική) επίδραση διαφόρων τύπων θορύβου στην ακρίβεια και περιγράφει πώς ένας κατασκευαστής της εφαρμογής μπορεί να ελαχιστοποιήσει τις αρνητικές τους επιδράσεις.

7.1 ΤΙ ΕΙΝΑΙ ΘΟΡΥΒΟΣ ?

Ήχοι οι οποίοι μεταδίδονται σε ένα σύστημα αναγνώρισης και δεν είναι μέρος του σήματος εισόδου που φέρει τη πληροφορία ,είναι θόρυβος . Υπάρχουν τέσσερις βασικοί τύποι θορύβου που επηρεάζουν τα συστήματα αναγνώρισης ομιλίας :

- Ο Θόρυβος Υπόβαθρου
- Ο Θόρυβος Καναλιού
- Ο Θόρυβος απ'την ομιλία ,όπως ο ήχος (πλατάγιασμα) των χειλιών
- Οι Χωρίς μήνυμα Φωνήσεις ,όπως το `Ωχ`

Ο καθένας παίζει καποιο ρόλο στη πρόκληση της ακρίβειας ενός συστήματος αναγνώρισης (βλέπε σχήμα 7.1)



Σχήμα 7.1 Τύποι θορύβου

Ο θόρυβος υποβάθρου ,αναφέρεται στο θόρυβο που παράγεται στο μέρος όπου η ομιλία γεννιέται και εισάγεται προς το σύστημα αναγνώρισης .Ο θόρυβος καναλιού, παράγεται από τα συστήματα (συσκευές) εισόδου που μεταφέρουν την ομιλία στο κυρίως σύστημα αναγνώρισης . Ο θόρυβος ομιλίας και οι φωνήσεις που δεν φέρουν μήνυμα, αναπαριστούν μερικές από τις συμβολές των ομιλητών προς το αντίξοο περιβάλλον αναγνώρισης . Αλλοι παράγοντες αναφέρθηκαν στο κεφάλαιο 5.

Είσοδο ομιλίας που περιέχει θόρυβο υποβάθρου και/ή θόρυβο καναλιού καλείται *αλλοιωμενος ή ενθόρυβος λόγος*. Αντίθετα λόγος που περιέχει λίγο ή καθόλου θόρυβο από υπόβαθρο ονομάζεται καθαρός λόγος. Ο τελευταίος, ακόμα, ίσως αναφέρεται και στο σύστημα ομιλίας που περιέχει θόρυβο από το υπόβαθρο και το κανάλι, όπου όμως είναι προσαρμοσμένος στα μοντέλα (αναφοράς) του συστήματος αναγνώρισης.

Ένα παρόμοιο σύνολο ορισμών υπάρχει για την έννοια *του αντίξοου περιβάλλοντος* το οποίο έχει ένα ή περισσότερα από τα ακόλουθα γνωρίσματα :

- Πολύ θόρυβο
- Άγνωστες ιδιότητες θορύβου
- Θόρυβο που δεν μοντελοποιείται στα σχέδια (αναφοράς) του συστήματος αναγνώρισης.

Ο τρίτος τύπος αναφέρεται στην ικανότητα των συστημάτων αναγνώρισης να ξεπερνούν την αρνητική επίδραση μερικών τύπων θορύβου, συνεκτιμώντας τους κατά το σχεδιασμό του μοντέλου.

Τα χαρακτηριστικά του περιβάλλοντος ομιλίας , όπου έχουν την μέγιστη αρνητική επίδραση στην συσκευή αναγνώρισης είναι :

- Ο φύση του θορύβου υπόβαθρου (π.χ η ομιλία περιβάλλοντος)
- Η Ποικιλομορφία στο θόρυβο
- Η ηχώ του θορύβου
- Ο τύπος του καναλιού ομιλίας που χρησιμοποιείται (τηλέφωνο ή μικρόφωνο)
- Η ποιότητα του καναλιού ομιλίας

Ακόμα , χαρακτηριστικά του περιβάλλοντος που πρέπει να προσθέσουμε , είναι η απόκριση του ομιλητή στον θόρυβο (καλείται *αποτέλεσμα LOMBARD*). Ολοι αυτοί οι παράγοντες πρέπει να υπολογιστούν για την σχεδίαση ενός συστήματος ή εφαρμογής αναγνώρισης.

7.1.1 Ο ΛΟΓΟΣ SNR (signal to noise ratio)

Ενας από τους τρόπους να εκτιμήσουμε τη ενδεχόμενη αρνητική επίδραση του θορύβου στην ακρίβεια της αναγνώρισης είναι να υπολογίσουμε τον *λόγο (κλάσμα) σήμα προς θόρυβο* (SNR) στην είσοδο , μετρώντας την ισχύ της ομιλίας (που συνιστά το σήμα εισόδου) προς την ισχύ του θορύβου. Μάλιστα, εξαιτίας της ιδιότητας μερικών μικροφώνων να απορροφούν και το υπόβαθρο, ο SNR για ένα σύστημα ομιλίας , μετριέται το ίδιο καλά τόσο μέσω της απόκρισης του μικροφώνου, όσο και στο περιβάλλον ομιλίας .Γενικά μετράμε τον SNR σε dB.Ετσι SNR πολλων dB αντιπροσωπεύει ισχυρότερα σήματα ομιλίας ως προς τον (περιβάλλων) θόρυβο υπόβαθρου. Π.χ ένα ησυχο εργαστήριο ή ιδιωτικό γραφείο , ίσως έχει SNR της τάξης των 45 dB. Ένα εργοστάσιο ή ο ηλεκτρικός σιδηρόδρομος έχουν SNR μικρότερο των 5 dB

Ο SNR έχει ισχυρή επίδραση στην ακρίβεια του συστήματος ομιλίας . Αποτελέσματα που εκτίθονται για έρευνα , και υπαίθριες εφαρμογές , δειχνουν ξεκάθαρα ότι η ακρίβεια αναγνώρισης ελαττώνεται με μειωσεις του SNR . Δυστυχως όμως τα περισσότερα περιβάλλοντα στον κόσμο έχουν σχετικά φτωχούς SNR

Ενώ , εργαστηριακά μπορεί να επιτευχθεί κλάσμα SNR που να αγγίζει τα 90 db (παρόλο που 50 dB είναι πιο τυπική τιμή) , μετρήσεις σε ομιλία σε φυσικό περιβάλλον , δείχνουν κατά μέσο όρο, λόγο SNR με τιμή μόνο 4,8 db . Ξεκάθαρα λοιπόν ,πολύ από την παρούσα (εργαστηριακά βασισμένη) γνώση μας , πάνω στην αντίληψη της ομιλίας ,ίσως να μην εφαρμόζεται στις

επικοινωνίες για τις τυπικές περιπτώσεις (Thomas Carell , Northwestern University , «Acoustical cues to auditory object formation in sentences" 1992,manuscript p.2)

Κατά συνέπεια η ανάλυση του SNR και των άλλων χαρακτηριστικών θορύβου , είναι θεμελιώδη στοιχεία για τον σχεδιασμό των παραπάνω εφαρμογών.

7.1.2 ΘΟΡΥΒΟΣ ΥΠΟΒΑΘΡΟΥ

Όλος ο χώρος της Ακουστικής έχει αγνοηθεί στη βιομηχανία αναγνώρισης ομιλίας .Οι άνθρωποι ,μόλις που έχουν ασχοληθεί με το σήμα που παρουσιάζεται στο κυρίως σύστημα αναγνώρισης . Αλλα παίρνεις όλα τα είδη θορύβου και άσχετα σήματα που μειώνουν σημαντικά την ικανότητα του κυρίως συστήματος αναγνώρισης να λειτουργήσει.(Bill Porter ,General Manager,AT&T Intelligent Acoustics Systems,personal communication,1995).

Ο θόρυβος υπόβαθρου είναι μέρος του ηχητικού περιβάλλοντος και εισάγεται μαζί με την ομιλία μας στην συσκευή εισόδου. Συμπεριλαμβάνει άλλες φωνές, μηχανικό θόρυβο και ήχους που προέρχονται από την ανθρώπινη δραστηριότητα, π.χ άνοιγμα συρταριών και περπάτημα. Επειδή τις περισσότερες φορές ο θόρυβος υποβάθρου επικάθεται (υπερτίθεται) πάνω στην ομιλία που δίνουμε στην είσοδο, αυτό συχνά καλείται *προσθετικός θόρυβος ή θόρυβος του περιβάλλοντος*. Εμφανίζεται σε κάθε συχνότητα ή φάσμα συχνοτήτων συμπεριλαμβανομένων και των σημαντικών για την ανθρώπινη ομιλία.

Ο θόρυβος υποβάθρου συχνά περιγράφεται χρησιμοποιώντας:

- τον λόγο S.N.R.
- την ποικιλομορφία
- την φασματική ισχύ.

Η ποικιλομορφία αναφέρεται στα συστατικά του θορύβου. Τα περιβάλλοντα που χαρακτηρίζονται από διακεκομμένο θόρυβο (π.χ. τηλεφωνικό κουδούνισμα) είναι πιο προκλητικά για τα συστήματα αναγνώρισης απ' ότι ένας σταθερός θόρυβος από μηχανουργείο. Αυτό αληθεύει ακόμη και αν ο συμπεριλαμβανόμενος θόρυβος είναι εξαιρετικά ισχυρός , επειδή οι συνιστώσες του θορύβου είναι πιο εύκολα αναγνωρίσιμες, και φιλτράρονται από την είσοδο ή ανιχνεύονται και συμπεριλαμβάνονται στα μοντέλα αναφοράς ενός συστήματος αναγνώρισης. Οι υψηλά μεταβαλλόμενοι και διακεκκομένοι θόρυβοι διαφεύγουν της σύλληψης ,και κατά συνέπεια είναι πιο δύσκολο για το σύστημα αναγνώρισης να τους αναγνωρίσει και να τους περιορίσει.

Το φάσμα ισχύος περιγράφει την περιοχή συχνοτήτων όπου κάθε ήχος παρουσιάζει την μεγαλύτερή του ένταση . Ο περιορισμός του θορύβου και οι μέθοδοι εμπλουτισμού της ομιλίας είναι δυσκολότερες όταν το φάσμα ισχύος του θορύβου συμπίπτει με το φάσμα ισχύος της ομιλίας. Ευτυχώς τα ακουστικά, πρότυπα των περισσότερων θορύβων υποβάθρου, ακόμα και του διακεκκομένου θορύβου, μεταβάλλονται πιο αργά απ' ότι συμβαίνει με την ομιλία. Αυτή η διαφορά επιτυγχάνει την διάκριση του καθαρού θορύβου υποβάθρου από την ομιλία ,που θα αναλυθεί. Όμως όταν ο θόρυβος υποβάθρου εμπεριέχει ομιλία , εντούτοις αυτός ο τρόπος διάκρισης δεν είναι

κατάλληλος .Κατά συνέπεια, διακεκκομενος θόρυβος υπόβαθρου συμπεριλαμβανομένων και της τηλεόρασης και του ραδιοφώνου αποτελούν το πιο δύσκολο περιβάλλον για τα συστήματα αναγνώρισης ομιλίας .

Δεν είναι όμως όλοι οι θόρυβοι υποβάθρου προσθετικοί. Κάθε δωμάτιο ή άλλα εσωτερικού χώρου περιβάλλοντα ομιλίας παράγουν αντήχηση. Η αντήχηση στα δωμάτια αποτελεί μια μοναδική σχεδιαστική πρόκληση για συστήματα αναγνώρισης ομιλίας σε καταναλωτικά προϊόντα, αυτοκίνητα και στα ηχεία. Η ακουστική των εσωτερικών χώρων τα κάνει να συμπεριφέρονται όπως μια αίθουσα με αντήχηση (βλέπε κεφ. 2, ενότητα 2.1.2.3). Αυτά αλλάζουν το φάσμα ισχύος με το να ενισχύουν μερικές συχνότητες και να καταπνίγουν άλλες. Αυτό καλείται *παραμόρφωση σήματος*. Διάφοροι παράγοντες επηρεάζουν τον χρόνο αντήχησης συμπεριλαμβανομένου του μεγέθους και του σχήματος του περιβάλλοντος ομιλίας, των υλικών κατασκευής, της επίπλωσης, του τύπου του μικροφώνου που χρησιμοποιείται και την απόσταση του ομιλητή από το μικρόφωνο.

Μια δεύτερη επίδραση του χρόνου αντήχησης είναι ότι παρατείνει χαρακτηριστικούς ήχους στους οποίους και προκαλεί επαναλαμβανόμενη είσοδο στο μικρόφωνο. Ακούμε αυτό το φαινόμενο σαν ηχώ.

Όταν παράγεις ένα σήμα ομιλίας θέλεις το φάσμα του σήματος να είναι δυνατό στιγμιαία και να σβήνει γρήγορα. Η αντήχηση επιβραδύνει τη (διάχυση) μετάδοση της ισχύος του σήματος στο δωμάτιο και κάνει δύσκολο για το σύστημα αναγνώρισης να καταλάβει τί λαμβάνει.

Το πιο επιθυμητό περιβάλλον ομιλίας είναι αυτό που δεν έχει καθόλου αντήχηση (καλείται χωρίς ηχώ). Αφού όμως περιβαλλον χωρίς ηχώ είναι τουλάχιστον ανύπαρκτο ειδικά για τα ελευθέρων χεριών-συστήματα αναγνώρισης ομιλίας, οι σχεδιαστές εφαρμογών πρέπει πάντα να προσανατολίζονται στην επίλυση προβλημάτων ανάδρασης.

7.1.3. ΘΟΡΥΒΟΣ ΚΑΝΑΛΙΟΥ

Ο θόρυβος καναλιού αναφέρεται στην επίδραση που εμφανίζεται πάνω στο σήμα (ομιλίας) εισόδου, από το τμήμα (συσκευή) εισόδου της ομιλίας (καλείται *το κανάλι ομιλίας*). Η λειτουργία του καναλιού ομιλίας είναι να μετατρέπει τα ηχητικά κύματα σε αναλογικά ηλεκτρικά κύματα (οπότε συμπεριφέρεται σαν μετατροπέας μορφών ενέργειας) και να εκπέμπει αυτά τα σήματα. Υπάρχουν δύο βασικά κανάλια που χρησιμοποιούνται για αναγνώριση λόγου:

α) τα μικρόφωνα,

β) τα τηλέφωνα.

Αυτά όταν μετατρέπουν το σήμα εισόδου παράγουν τμήματα παραμόρφωσης πάνω στο σήμα. Η παραμόρφωση είναι ένα υποπροϊόν της ακουστικής απόκρισης και ικανότητας της συσκευής εισόδου. Παράγεται πριν την επεξεργασία ομιλίας από το σύστημα αναγνώρισης. Επίσης το κανάλι συμβάλει στο σήμα τον προσθετικό του θόρυβο, συνήθως ηλεκτρικό θόρυβο.

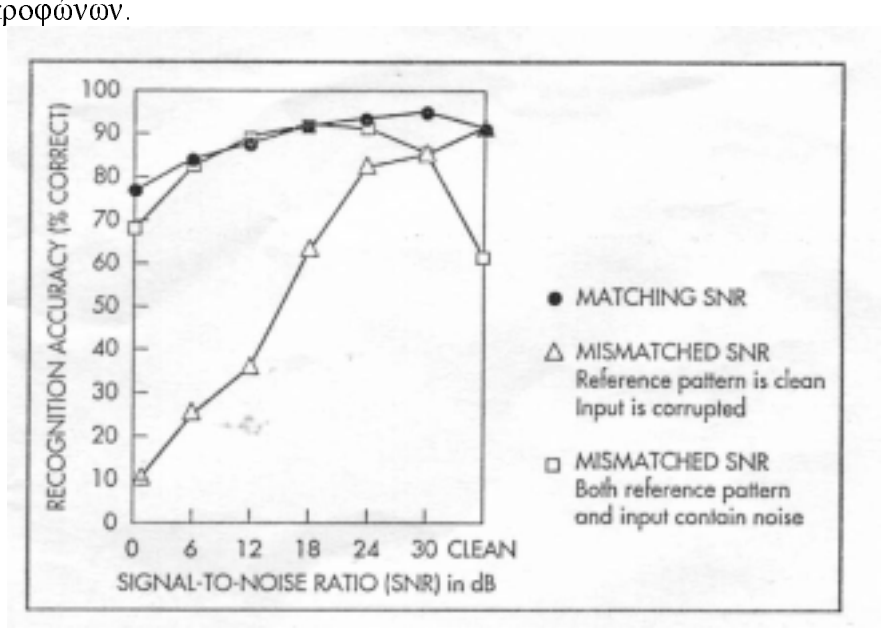
Η αρνητική επιδραση του καναλιού στην ακρίβεια αναγνώρισης είναι μοναδική. Για ανεξάρτητα συστήματα ομιλίας, λαμβάνεται υπόψη ακόμη και η επίδραση της ποικιλομορφίας του ομιλητή (βλέπε κεφ. 5). Οι σχεδιαστές των ανεξάρτητων συστημάτων ομιλίας μαζεύουν προσεκτικά δείγματα ομιλίας από τηλεφωνικά δίκτυα ,τα οποία θα χρησιμοποιηθούν για τις

εφαρμογές τους. Προκειμένου να προσαρμόσουν διαφορετικές συνθήκες θορύβου καναλιού, αναπτύσσουν διαφορετικά είδη μοντέλων για τα κινητά τηλέφωνα και για τα καλωδιακού δικτύου .

7.1.3. 1 ΜΙΚΡΟΦΩΝΑ

Κάθε μάρκα και μοντέλο μικροφώνου παράγει ένα μοναδικό σχηματισμο από παραμόρφωση και προσθετικό ηλεκτρικό θόρυβο. Επειδή είναι εξαιρετικά δύσκολο να απομακρύνουμε τα χαρακτηριστικά θορύβου του μικροφώνου πάνω από το σήμα, τα συμπεριλαμβάνουμε γενικά στα μοντέλα αναφοράς της εφαρμογής ή του συστήματος. Όπως φαίνεται στο σχήμα 7.2 όταν το μικρόφωνο χρησιμοποιείται σε ένα αναπτυγμένο σύστημα, είναι διαφορετικό περίπτωση από εκείνο που χρησιμοποιήθηκε για την ανάπτυξη του μοντέλου , η επίδραση στην ακρίβεια αναγνώρισης είναι ισχυρή και "δηλητηριώδης".

Rabiner & Juang (1993 κεφ. 5). περιγράφουν μια περίπτωση σε ένα σύστημα μεγάλου λεξιλογίου όπου παρατήρησαν μια ελάττωση της ακρίβειας από ποσοστό 85%, σε κάτω του 19%, σαν αποτέλεσμα της αλλαγής μικροφώνων.



Σχημα 7.2 Αποτελέσματα των κακοταιριασμένων μικροφώνων

Τα μικρόφωνα επίσης ποικίλουν στην ποιότητα και στον τύπο. Τα *πολυκατευθυντικά* μικρόφωνα έχουν συγκριτικά ομοιόμορφη απόκριση ,σε αντίθεση με τα *κατευθυντικά* μικρόφωνα (επίσης καλούνται διαφορικά μικρόφωνα) που σχεδιάζονται να αποκρίνονται σε ήχους από μία συγκεκριμένη κατεύθυνση (καλούνται *μονοκατευθυντικά* μικρόφωνα) ή από δύο συγκεκριμένες κατευθύνσεις (καλούνται *δικατευθυντικά* μικρόφωνα). Ακριβώς επειδή μπορούν να απομονώσουν ηχητικές πηγές βασιζόμενα στην τοποθεσία τους ,τα κατευθυντικά μικρόφωνα είναι καλύτερα από τα πολυκατευθυντικα μικρόφωνα στην αναγνώριση ομιλίας. Τα κατευθυντικά μικρόφωνα διαφέρουν στο εύρος της ευαισθησίας τους (καλείται μέγεθος της δέσμης). Μερικά είναι ευαίσθητα σε ήχους που πηγαζουν από ένα συγκριτικά ανοιχτό χώρο,ενώ άλλα ειναι δέσμης στενής (σαν μολύβι).

Τα μικρόφωνα αφαίρεσης θορύβου προτιμούνται σε περιβάλλοντα υψηλού θορύβου , γιατί ανθίστανται σε ήχους προερχόμενους από μακρινές πηγές. Γενικά κατασκευάζονται από δικατευθυντικά μικρόφωνα ή από ένα ζευγάρι μονοκατευθυντικων μικροφώνων όπου το ένα κοιτάζει προς την πηγή ομιλίας και το άλλο ως προς το περιβάλλον. Έτσι η φασματική είσοδος από το δεύτερο μικρόφωνο(του περιβάλλοντος) απομακρύνονται πάνω από το σήμα (ομιλίας) εισόδου, πριν την επεξεργασία. Αυτή η μέθοδος βελτιώνει τον SNR ιδιαίτερα.

Τα *κοντινής ομιλίας* μικρόφωνα σχεδιάζονται για να τοποθετούνται κοντά ή απέναντι της ηχητικής πηγής (όπως το στόμα του ομιλητή) και τυπικά βρίσκονται στα μικρόφωνα των headset. Κατασκευάζονται ώστε να αποκρίνονται αρχικά στους ήχους που γεννιούνται από τις πραπάνω πηγές. Το σχέδιο και και η τοποθέτηση των παραπάνω μικροφώνων ελαττώνει τις αρνητικές συνέπειες της ανάδρασης και βελτιώνει τον SNR. Όταν τα μικρόφωνα αυτά δεν μπορούν να χρησιμοποιηθούν, π.χ. σε κιόσκια και άλλου μακρινού πεδίου εφαρμογές ,τα στενής ζώνης κατευθυντικά μικρόφωνα ή

σειρά κατευθυντικών μικροφώνων μπορούν να χρησιμοποιηθούν ώστε να καλύψουν τον απαιτούμενο χώρο ομιλίας . Ένα προχωρημένο πείραμα παρουσιάστηκε από την IBM και είναι ιδιαίτερα χρήσιμο για να καταλάβουμε την σχέση ανάμεσα στον τύπο μικροφώνου και στον θόρυβο. Οι ερευνητές σύγκριναν 5 τύπους μικροφώνων

α) τα μικρόφωνα χειρός -αφαίρεσης θορύβου

β) τα στηριζόμενα σε μόνιμη βάση (σταντ) μικρόφωνα

γ) τα (τοποθετημένα σε πλαίσιο) μπροστά από το στόμα -κοντινής ομιλίας

δ) μονοκατευθυντικά με κλιπ (μανταλάκι)

ε) πολυκατευθυντικά με κλιπ (μανταλάκι)

σε 5 διαφορετικές συνθήκες όσο αφορά τον θόρυβο υποβάθρου :

- καφετέρια κατά το μεσημέρι (ομιλίες, άλλοι θόρυβοι),
- εργαστήρια υπολογιστών (μηχανικοί ήχοι, ήχοι προερχόμενοι από τους παρευρισκόμενους),
- γραφείο γραμματείας (διακεκομμένος θόρυβο σε ένα γενικά ήσυχο περιβάλλον),
- φωτοτυπικό εργαστήριο (δυνατοί και επαναλαμβανόμενοι μηχανικοί θόρυβοι),
- ήσυχο γραφείο(snr benchmark για ήσυχα περιβάλλοντα).

Σε όλες τις περιπτώσεις υπήρχε μια ισχυρή σύνδεση ανάμεσα στον SNR και στην ακρίβεια αναγνώρισης. Τα κατευθυντικά μικρόφωνα ήταν λιγότερα ευαίσθητα στους θορύβους του υπόβαθρου σε σχέση με όλους τους άλλους τύπους. Τα μικρόφωνα αφαίρεσης θορύβου είναι ιδιαίτερα ανθεκτικά σε ήχους προερχόμενους από πηγές στο υπόβαθρο και λειτουργούσαν καλά σε περιβάλλοντα υψηλού θορύβου (καφετέρια και φωτοτυπικό εργαστήριο). Επίσης τα κοντινής ομιλίας headset μικρόφωνα λειτούργησαν πολύ καλά .Οι Rabiner & Juang ερευνητές της IBM βρήκαν ότι δεν είναι προτρέψιμο να εξασκείσαι με ένα μικρόφωνο και να χρησιμοποιήσεις άλλο στη πράξη. Για περισσότερες πληροφορίες για τις έρευνες της IBM προτρέξτε Das,et al.(1993). Ακόμα βλέπε Rabiner & Juang (1993, Κεφ. 5).

7.1.3.2 ΤΟ ΤΗΛΕΦΩΝΙΚΟ ΚΑΝΑΛΙ

Οι εφαρμογές αναγνώρισης ομιλίας που σχεδιάστηκαν για χρήση πάνω στα τηλέφωνα πρέπει να προσαρμόζουν τον θόρυβο που γεννιέται από το

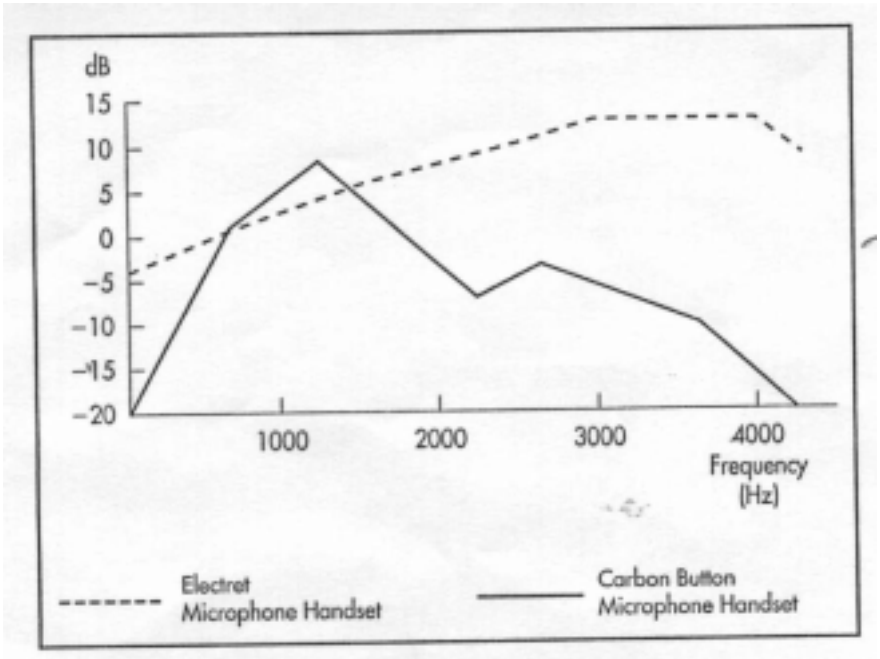
μικρόφωνο του ακουστικού. Αν το σύστημα αναγνώρισης είναι στο τηλεφωνικό δίκτυο απ' ότι στο ακουστικό, το σύστημα πρέπει επιπλέον να κατευθύνει το θόρυβο στο τηλεφωνικό δίκτυο.

Τρεις τύποι μικροφώνων χρησιμοποιούνται συνήθως σε τηλεφωνικούς εξοπλισμούς.

- Πυκνωτικά μικρόφωνα
- Τα μικρό φωνα άνθρακος ,για τηλεφωνικό ακουστικό
- Πυκνωτικά μικρόφωνα ,για τηλεφωνικό ακουστικό

Αυτές οι κλάσεις των μικροφώνων ποικίλουν στην αντήχηση που λαμβάνεται από το περιβάλλον και στην παραμόρφωση που προκαλούν .Τα μικρόφωνα είναι ιδιαίτερα δύσκολα τόσο για τα συστήματα αναγνώρισης όσο και για τους ανθρώπους .Κυμαίνοντας την απόσταση και την θέση του ομιλητή σε σχέση με το μικρόφωνο, ελαττώνεται η ισχύς του σήματος ομιλίας στο μικρόφωνο, συνεισφέροντας στον μικρό SNR. Η είσοδος από μικρόφωνο είναι επίσης ευάλλωτη στην αντήχηση. Οι ομιλητές κάνουν την αναγνώριση ομιλίας ακόμη πιο δύσκολη όταν απομακρύνονται ή γυρνάνε από το μικρόφωνο καθώς μιλάνε, ή όταν αλλάζουν την ένταση της ομιλίας καθώς το χρησιμοποιούν.

Τα πυκνωτικά μικρόφωνα και τα μικρόφωνα άνθρακος ,είναι ακόμα πιο ακατάλληλη για συστήματα αναγνώρισης. Και οι δύο κατηγορίες χρησιμοποιούνται ως κοντινής ομιλίας συσκευές σε τηλεφωνικά ακουστικά, Όπως η εικόνα 7.3 δείχνει δεν έχουν όλα την ίδια απόκριση συχνότητας.



Σχήμα 7.3 Σύγκριση Πυκνωτικού μικροφώνου και μικροφώνου άνθρακος

Τα πυκνωτικά μικρόφωνα κυριάρχησαν στην Αμερική ,ενώ τα τηλέφωνα στις περισσότερες άλλες χώρες χρησιμοποιούσαν τα μικρόφωνα άνθρακος . Τα πυκνωτικά μικρόφωνα παρέχουν περισσότερη ισορροπία στην καμπύλη απόκρισης απ' ότι τα μικρόφωνα άνθρακος . Τα τελευταία ,επίσης επιδεικνύουν φτωχή απόκριση για το εύρος συχνοτήτων μεταξύ 1500Hz-2000Hz, καθώς και για τις υψηλές συχνότητες , οι οποίες βοηθούν στον

προσδιορισμό των άφωνων τυρβώδων συμφώνων(β,θ,σ,ζ,φ,χ) (κεφ.2 ενότητα2.1). Όπως και με τα τυπικά μικρόφωνα , τα πυκνωτικά μικρόφωνα και τα μικρόφωνα άνθρακος ,ποικίλουν εξαιρετικά σε ποιότητα και στους αλγόριθμους που χρησιμοποιούνται για τη μετατροπή των σημάτων εισόδου .Ακόμα περισσότερο κάθε μοντέλο συνεισφέρει την δικιά του παραμόρφωση και τον δικό του προσθετικό θόρυβο.

Το τηλεφωνικό δίκτυο φέρει και άλλη παραμόρφωση στο σήμα. Επιπροσθέτως του προσθετικού θορύβου, βάζοντας και άλλες λειτουργίες του δικτύου, εξασθενίζουν συχνότητες σήματος κάτω τον 100Hz και πάνω των 3100Hz. Τόσο οι άνθρωποι όσο και οι συσκευές αναγνώρισης ομιλίας, πρέπει να εστιάζουν τις επεξεργασίες τους σε συχνότητες μεταξύ 100-3100Hz. . Αυτή η ζώνη καλείται *τηλεφωνικό εύρος ζώνης*.

Η φύση και ο τύπος του προσθετικού θορύβου και η παραμόρφωση , ποικίλουν από δίκτυο σε δίκτυο .Είναι μεγάλα υποπροϊόντα του τύπου , της ποιότητας και της συντήρησης του εξοπλισμού.

Ο θόρυβος δικτύου μπορεί να είναι συστηματικός ή τυχαίος. *Ο Συστηματικός θόρυβος* είναι ένα μόνιμο χαρακτηριστικό του τηλεφωνικού δικτύου. Κάθε διακόπτης προκαλεί τα δικά του χαρακτηριστικά θορύβου στις καμπύλες. Παρόλο που ποικίλει από δίκτυο σε δίκτυο και απο τηλεφωνικό δίκτυο σε τηλεφωνικό δίκτυο, ο συστηματικός θόρυβος του δικτύου είναι σταθερός και προβλέψιμος. Αληθεύει ιδιαίτερα για τα κλασσικά τηλέφωνα.

Γνωρίζοντας το κανάλι μπορούμε να προβλέψουμε τι θα συμβεί. Ο παλμός πάντα θα παραμορφωθεί με ένα συγκεκριμένο τρόπο. Οι συχνότητες θα έχουν πάντα μια συγκεκριμένη ελάχιστη καθυστέρηση φάσης (James Martin ,Chairman,James Maritn Associates,Telecommunications and the compyter,1990, σελ 640).

Ο *τυχαίος θόρυβος* είναι απρόβλεπτος. Τα συστατικά του στοιχεία συμπεριλαμβάνουν (το σφύριγμα από) τον ηλεκτρικό λευκό θόρυβο, τον ατμοσφαιρικό θόρυβο, και τριξίματα .Αυτοί μπορούν να παράγουν ξαφνική στιγμιαία μεταβολή στον SNR ή ακόμα περισσότερο να προκαλέσουν την αχρηστία του καναλιού. Οι τυχαίες πηγές παραμόρφωσης περιλαμβάνουν πτώση ή απότομη άνοδο στο πλάτος του σήματος και αλλαγές στην φάση. Τα χαρακτηριστικά θορύβου σε ένα καλωδιακό τηλεφωνικό δίκτυο είναι προκλητικά ,αλλά ο συστηματικός θόρυβος η παραμόρφωση αυτών των δικτύων (μερικά χαρακτηριστικά τυχαίου θορύβου) μπορούν να συμπεριληφθούν σε μοντέλα.. Τυπικά μία απλή τηλεφωνική επικοινωνία χρησιμοποιεί περισσότερα από ένα δίκτυα. κάνοντας αναγκαία την μοντελοποίηση των χαρακτηριστικών πολλών δικτύων.

Η αναγνώριση ομιλίας σε δίκτυα κινητής τηλεφωνίας αντιπροσωπεύει μία ακόμα μεγαλύτερη πρόκληση .Η διακύμανση στην εκπομπή είναι μεγαλύτερη από εκείνη των καλωδιακών δικτύων και είναι συνδιασμός της ισχύος εκπομπής του τηλεφώνου, της κυψελίδος, των υψηλών επιπέδων θορύβου υποβάθρου ,και της αντήχησης . Ο χειρισμός του θορύβου του καναλιού και του υποβάθρου απαίτησε επιπλέον συλλογή δεδομένων. Το project της Texas Instrument : Voice Across America εξέλεξε μεγάλο αριθμό δειγμάτων αντιπροσωπεύοντας τόσο το τηλεφωνικό δίκτυο όσο και τα χαρακτηριστικά των handset παντού στις Η.Π.Α. Αυτές οι επιδράσεις του δικτύου περιορίζονται όταν η συσκευή αναγνώρισης λόγου ενσωματώνεται πάνω στο κινητό τηλέφωνο, απ' ότι όταν στο δίκτυο. Οι Foster &Schalk (1993) και Rabiner & Juang (1993, Κεφ. 5) προσφέρουν επιπλέον πληροφορίες σχετικά

με το τηλεφωνικό κανάλι.

7.1.4 ΧΩΡΙΣ ΜΗΝΥΜΑ(Non Communication)ΘΟΡΥΒΟΣ ΟΜΙΛΙΑΣ

Κατά την διάρκεια της ομιλίας παράγονται ήχοι οι οποίοι πρέπει να επεξεργαστούν από το σύστημα αναγνώρισης σαν `χωρίς μήνυμα` είσοδο .Τέτοιοι θόρυβοι από ομιλία είναι «το καθάρισμα του λαιμού ,οι ήχοι από τα χείλια και την γλώσσα (πλατάγιασμα)». Μερικοί είναι προβλέψιμοι (επισημαίνεται π.χ ήχος από το χείλια στην αρχή μιας κουβέντας) και μπορούν να αναγνωριστούν από το σύστημα ή να μοντελοποιηθούν σαν ειδικές λέξεις ή σαν «σκουπίδια». Τέτοιες `χωρίς μήνυμα ` αρθρώσεις χαρακτηρίζουν τον αυθόρμητο λόγο.

Ο αυθόρμητος λόγος περιέχει διορθώσεις στη μέση της κουβέντας και αλλαγές ρημάτων, μη γραμματικές δομές και κομμένες λέξεις. Είναι πολύ δύσκολο να παραχθούν κανόνες οι οποίοι προσφέρουν καλή κάλυψη της αλληλουχίας των λέξεων που οι άνθρωποι χρησιμοποιούν όταν μιλάνε με αυθόρμητο λόγο. (Wayne Ward & Sheryl Young ,Carnegie Mellon Univercity,"Ευελικτη χρήση σημασιολογικών περιορισμών στην αναγνώριση ομιλίας -Flexible use of semantic constraints in speech recognition,"1993,σελ. 49)

Οι ομιλητές κομπιάζουν , τραυλίζουν, διορθώνουν λάθη στην μέση των λέξεων (π.χ είναι κιτρ/ εεε όχι μπλε) και γεμίζουν τις παύσεις με `χωρίς μήνυμα `ήχους όπως το «ωχ» π.χ (βλέπε κεφ. 6 τομέας 6.7.1) .Αυτά έχουν ξεκινήσει να μοντελοποιούνται αλλά ακόμα είναι λίγο κατανοητά. Ακουστικά βασιζόμενα συστήματα αναγνώρισης ομιλίας όπως εμπορικά συστήματα αναγνώρισης των ημερών μας, θα είναι ικανά να χειριστούν μερικά από αυτά τα φαινόμενα, αλλά η αληθινή ανθεκτικότητα θα επιτευχθεί μόνο με την ανάπτυξη εμπορικών συστημάτων που θα καταλαβαίνουν την ομιλούμενη γλώσσα (βλέπε κεφ. 4 ,ενότητα 4.4) Τα συστήματα αναγνώρισης της ομιλούμενης γλώσσας, χρησιμοποιούν πληροφορίες για τους σκοπούς του ομιλητή, καθομιλούμενες δομές, και άλλες γνωστικές –γλωσσολογικές πηγές ,ώστε να επεξεργαστούν το λόγο. Ο Wayne Ward του CMU ερευνά ενεργά την `χωρίς μήνυμα` συμπεριφορά ομιλίας σε συνεχή λόγο. Για περισσότερες πληροφορίες βλέπε Ward (1991) και Wayne Ward & Sheryl Young (1993).

Το αποτέλεσμα LOMBARD

Το αποτέλεσμα LOMBARD (επίσης καλείται και λόγος LOMBARD) περιγράφει τις αλλαγές όπου προκαλούνται στην ομιλία όταν ένας ομιλητής προσπαθεί να κάνει τον εαυτό του να ακουστεί πάνω από τον θόρυβο. Τα χαρακτηριστικά του λόγου LOMBARD πρωτοπεριγράφτηκαν από έναν Γάλλο φυσιολόγο τον E. LOMBARD το 1911 και εκδόθηκαν σε ένα περιοδικό για ειδικούς στα μάτια, αυτιά, μύτη και διαταραχές στο λαιμό. Στις αρχές του 1950 το ενδιαφέρον των ερευνητών πάνω στον άνθρωπο με άνθρωπο επικοινωνία , άρχισε να στρέφεται και να εξετάζει τις επιδράσεις του αυξημένου φωνητικού έργου (όπως π.χ το να φωνάζει κανείς), μια συνέπεια της επίδρασης LOMBARD στη διανοητικότητα του λόγου. Ενα μέρος αυτής της εργασίας αποτέλεσε την βάση για την κατανόηση της

επίδρασης του αποτελέσματος LOMBARD ,πάνω στην ακρίβεια των συσκευών αναγνώρισης ομιλίας. Για περισσότερες πληροφορίες πάνω στην επίδραση LOMBARD βλέπε την γνήσια δημοσίευση του Lombard (1911) στο Ann. Maladies Oreille,Larynx, Nez, Pharynx (in french), Gardner (1996), Pickett(1956) ,Rostolland & Parant (1973).

7.2 Η ΑΠΟΚΡΙΣΗ ΤΩΝ ΟΜΙΛΗΤΩΝ ΣΤΟΝ ΘΟΡΥΒΟ ΥΠΟΒΑΘΡΟΥ

Ακόμα και αν οι επιδράσεις από το κανάλι , το μικρόφωνο και τον θόρυβο ,απομακρυνθούν, οι επιδράσεις από το στρες και από το αποτέλεσμα LOMBARD θα συμβάλλουν μοναδικά σε ελάττωση της ικανότητας αναγνώρισης (John H.L Hansen, `Εργαστηριο Επεξεργασίας Ανθεκτικού Λόγου, Πανεπιστήμιο Duke, προσωπική επικοινωνία, 1994)

Όπως συμβαινει και με τα συστήματα αναγνώρισης ,οι ομιλητές δεν μένουν ανεπηρέαστοι από στρες και από το θόρυβο υπόβαθρου. Το στρες παράγει συναισθηματικές και φυσικές αποκρίσεις οι οποίες επιδρούν στο τρόπο που το άτομο ομιλεί. Ο θόρυβος συμβάλλει στην προσπάθεια του ατόμου να επικοινωνήσει ακόμα και μικρές αυξήσεις στον θόρυβο μικρότερες των 10Db , παράγουν μεταβολές στον τρόπο που οι άνθρωποι ομιλούν. Αυτές οι αλλαγές μεγενθύνονται όταν ο θόρυβος υπόβαθρου γίνεται ισχυρότερος .Μερικά από τα χαρακτηριστηκά ομιλίας και ακουστικής που έχουν να κάνουν με την προσπάθεια του ομιλητή να υπερκεράσει τον θόρυβο συμπεριλαμβάνουν :

- Αυξανόμενο φωνητικό έργο
- Μεγαλύτερη διάρκεια των λέξεων εξαιτίας του αυξανόμενου μήκους των φωνηέντων
- Αλλαγές στις θέσεις των formant(συχνότητες φωνοσυντονισμού της φωνητικής οδού) με φωνηέντα
- Αυξημένα πλάτη στα των formant(συχνότητες φωνοσυντονισμού της φωνητικής οδού)
- Παράλειψη μερικών τελικών συμφώνων

Αυτή η συμπεριφορά *αποκαλείται λόγος LOMBARD ή το αποτέλεσμα LOMBARD*. Το αποτέλεσμα LOMBARD κάνει την ομιλία ευκολότερη στο να ακουστεί και να κατανοηθεί όταν αυτός που ακούει είναι ένας άλλος άνθρωπος .Σε αντίθεση έχει μια ισχυρά αρνητική επίδραση στην ακρίβεια αναγνώρισης και μπορεί να προκαλέσει αποκλίσεις στην ακρίβεια, της τάξης του 25%

Τα ήδη υπάρχοντα εμπορικά συστήματα αναγνώρισης ομιλίας δεν είναι ικανά για αυτόματη προσαρμογή στις ακουστικές επιδράσεις του λόγου LOMBARD.

Ενας λόγος ,είναι η ύπαρξη διαφωνιών όσο αφορά την χρήση και το περιεχόμενο των ακουστικών μεταβολών που παρουσιάσαμε παραπάνω. Μερικοί ερευνητές έχουν ανακαλύψει εσωτερική – και πολύ-ομιλητική μεταβλητότητα στα χαρακτηριστικά του λόγου LOMBARD

Για μερικές παραμέτρους που δοκιμάσαμε , υπάρχουν μοναδικές μεταβολές ανάμεσα στην αντρική και στην γυναικεία ομιλία. Αυτή η μεταβλητότητα μερικών φαινομένων μπορεί να είναι πολύ εξαρτήσημη από το περιεχόμενο

(Jean-Claude Junqua & Yolande Anglande, C.R.I.N./I.N.R.I.A. Vandoeuvre les Nancy,Γαλλια, «Ακουστικές και διανοητικές σπουδές του λόγου LOMBARD»,1990,σελ.844)

Ωστόσο οι περισσότεροι ερευνητές έχουν προσμετρήσει περισσότερο κοινό περιεχόμενο στα λειτουργικά ακουστικά πρότυπα του λόγου LOMBARD. Διαφορές ανάμεσα στα αποτελέσματα αυτών των δύο ερευνητικών ομάδων ίσως να οφείλονται τόσο στις συγκεκριμένες παραμέτρους που αυτοί εξέτασαν, όσο και στο μέγεθος και την απόκλιση των πληθυσμών που χρησιμοποιήσανε για έρευνα

Επιπρόσθετες μεταβολές στον λόγο , εξαιτίας του στρεσαρίσματος, έχει βρεθεί να αυξάνουν τις επιδρασεις του θορύβου. Αυτές είναι λιγότερο μελετημένες και κατανοητές.

Για περισσότερες πληροφορίες σχετικά με τον λόγο LOMBARD απευθυνθείτε στις δημοσιεύσεις των John H.L Hansen(1993),Hansen & Bria (1990 &1992) ,Stanton,et al.(1989), and Jungua & Anglade(1990). Στις δημοσιεύσεις Hansen ,Hansen & Bria περιγράφετε επισης η στρεσαρισμενη ομιλία.

7.3 ΑΚΟΥΣΤΙΚΑ ΜΟΝΤΕΛΑ

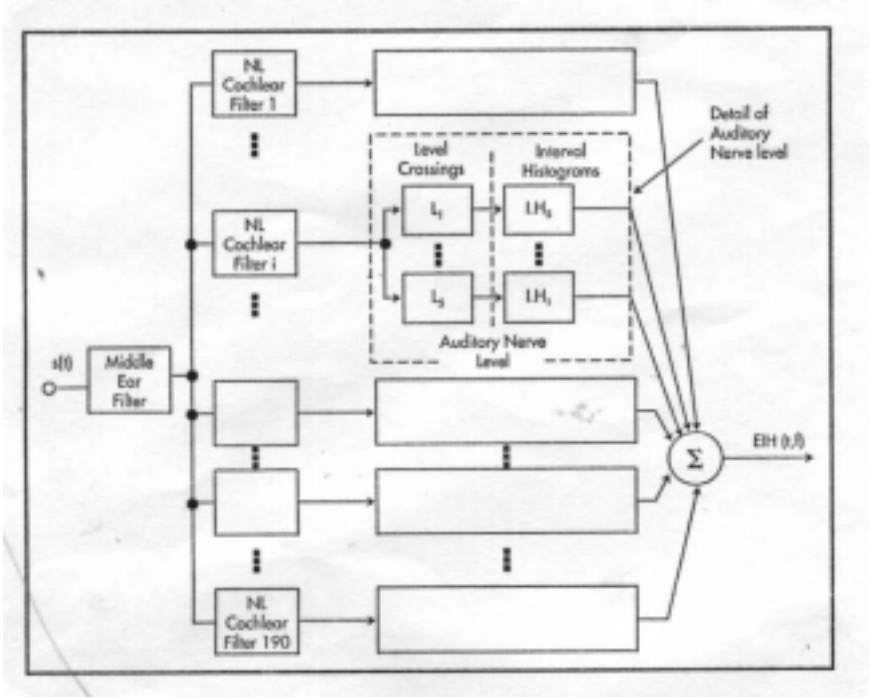
Ο σκοπός της ακουστικής μοντελοποίησης είναι να επιφέρει στην συσκευή αναγνώρισης την ικανότητα να φιλτράρει τον θόρυβο από το σήμα εισόδου με τον ίδιο τρόπο που και ο άνθρωπος κάνει. Η ακουστική μοντελοποίηση προσφέρει την υπόσχεση να παράγουμε πιο ικανοποιητικά , ανθεκτικά και ακριβή συστήματα αναγνώρισης ,για ομιλία η οποία παράγεται σε θορυβώδη περιβάλλοντα. Αυτή τη στιγμή δεν υπάρχουν εμπορικά συστήματα αναγνώρισης ομιλίας τα οποία να βασίζονται πάνω στην ακουστική μοντελοποίηση.

Μερικά χαρακτηριστικά της ακουστικής μοντελοποίησης, όπως η κλίμακα mel (βλεπε κεφ. 2 τομεας 2,1,4) έχουν ηδη γίνει στανταρ στοιχεία σε ερευνητικά και εμπορικά συστήματα αναγνώρισης .Η χρήση αυτών και άλλων χαρακτηριστικών έχει προωθηθεί σε ένα ευρή φάσμα ερευνητικών συστημάτων. Ερευνητές στο DUKE UNIVERSITY έχουν ανακαλύψει ένα δυαδικού καναλιού, αλγόριθμο φιλτραρίσματος - μείωσης θορύβου, βασισμένο πάνω στις παραμέτρους της κλίμακας mel. Αυτός αποδείχτηκε ότι λειτουργεί καλά σε έλεγχους πάνω σε ομιλία που παράγεται πάνω σε λευκό θόρυβο και μέσα σε θορυβώδη πιλοτήρια αεροπλάνων

Ερευνητικά συστήματα που χρησιμοποιούν άλλες ιδιότητες της ακουστικής επεξεργασίας των θηλαστικών, αναπτύσσονται στα εργαστήρια . Οι Cheng και O`Shaughnessy των INRS τηλεποικοινωνιών του Καναδά έχουν αναπτύξει επιπρόσθετους αλγόριθμους μείωσης θορύβου βασισμένους πάνω στην lateral inhibition συμπεριφορά του ανθρώπινου ακουστικού συστήματος . *lateral inhibition είναι* η ικανότητα να διακρίνεις ήχους από δυο διαφορετικές (διακριτές) ηχητικές πηγές .Οι αλγόριθμοι σχεδιάστηκαν για να χειρίζονται διαφορετικούς τύπους προθετικού θορύβου όταν το SNR είναι κάτω των 5 dB και σε αντίθεση με τις περισσότερο συνηθισμένες προσεγγίσεις δεν χρειάζονται ειδικούς (άνευ ομιλίας) ανιχνευτές.

ΤΟ σύστημα του Chitza : (Ensemble Interval Histogram)Συντονισμένων Παύσεων Ιστόγραμμα (ΕΙΗ) , που παρουσιάζεται στην εικόνα 7,4 ,εστιάζει

τα σημαντικότερα στοιχεία του ακουστικού νεύρου.



Σχήμα 7.4 Το μοντελο **ΕΙΗ**

Το ΕΙΗ είναι από τα πιο πολύπλοκα συστήματα συστήματα μοντελοποίησης. Παρουσιάζει κωδικοποίηση ομιλίας χρησιμοποιώντας μοντέλα των περισσοτέρων περιφερειακών συστημάτων ακουστικής . Το μέσο ακουστικό φίλτρο απομακρύνει συχνότητες κάτω των 1000Hz (υψιπερατο φίλτρο) καθόσο τα << φίλτρα κοχλία >> σχεδιάστηκαν ώστε να μιμούνται την συμπεριφορά των δομών που παρουσιάζονται ακόμα εσωτερικότερα στο αυτί . Η λειτουργία των ζωνοδιαβατών φίλτρων και των ιστογραμμάτων παύσεως, αναπαριστά αναπαριστά τη συνδυασμένη λειτουργία των ξεχωριστών ινών του ακουστικού νεύρου .Σε πρόσφατα πειράματα ο ΕΙΗ κωδικοποιητής διέκρινε δίφωνα (βλεπε κεφ. 6 τομεας 6,3,3,2)περισσοτερο όπως ο άνθρωπος ,παρά από ότι τα συμβατικά συστήματα κωδικοποίσης.

Πιο αναλυτικά αυτό το μοντέλο βασίζεται στο συγχρονισμό του σήματος ομιλίας .Ένα σύνολο 85 φίλτρων κοχλία , ισαπεχόντων στη λογαριθμηκή κλίμακα ,από 200 Hz εως 3200 Hz, χρησιμοποιούνται για να μοντελοποιήσουν τη βασιλική μεμβράνη . Ακολούθως, στην εξοδο του κάθε φίλτρου, μοντελοποιείται ο μηχανισμός πυροδότησης των νευρικών ινών .Το επόμενο βήμα παράγει ένα ψευδοφάσμα ,που αποκτείται μέσω (κατάλληλης) άθροισης των ιστογραμμάτων όλων των φίλτρων.(Αφού)τα ιστογράμματα κατασκευάζονται μέσω εντοπισμού των περιοχών που πυροδοτούνται ταυτόχρονα ,στη διάταξη των ινών που προσομοιώνουν τα παραπάνω φίλτρα. Ο Ghitza το1987 αντικατέστησε τα παραπάνω φίλτρα κοχλία με ένα σύνολο φίλτρων Hamming .Τα αποτελέσματα έδειξαν βελτιωμένη απόδοση , δίνοντας έμφαση στο ρόλο που παίζει η σύγχρονη μέτρηση ,για εξαγωγή κατάλληλης πληροφορίας ομιλίας ,ακόμα και σε περιβάλλοντα με θόρυβο . Η εξαγωγή της κύριας περιοδικής πληροφορίας είναι ενας τρόπος για επίτευξη ανθεκτικότητας έναντι του θορύβου .Τελικά , ο Ghitza το1988,με τη

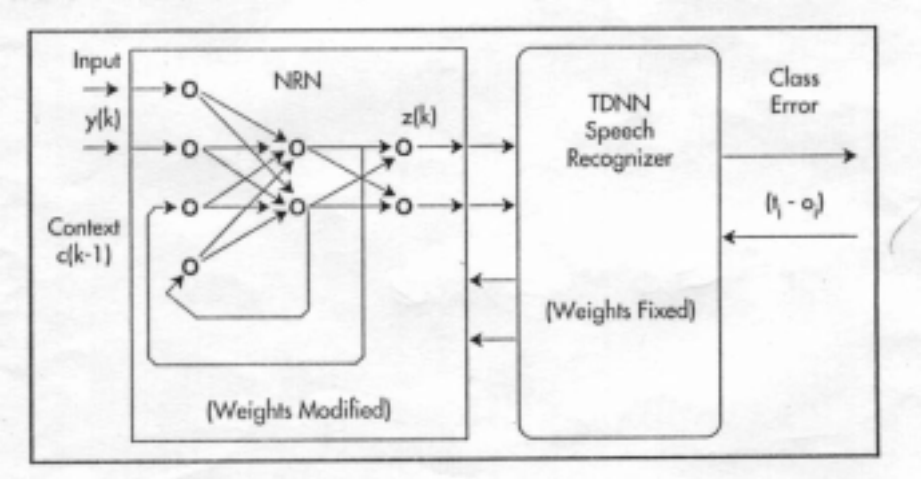
προσθήκη μηχανισμού ανάδρασης ,που χρησιμοποιεί πληροφορία που λαμβάνεται από ΕΙΗ φάσμα ,επέφερε παραπέρα βελτίωση στην ακρίβεια της αναγνώρισης ομιλίας .Ετσι, το 1995 ,επιβεβαιώθηκε η ανθεκτικότητα του ΕΙΗ φάσματος έναντι της σύνηθης mel-cepstral ανάλυσης.

Τα συστήματα που περιγράφτηκαν σ΄αυτήν την ενότητα συνενώνουν γνωστές συμπεριφορές του περιφερειακού ακουστικού συστήματος (κυρίως του κοχλία και του ακουστικού νεύρου) σε συστήματα αναγνώρισης ομιλίας .Οι παράγωγοι αυτών των συστημάτων παραδέχονται ότι γνωστές συμπεριφορές συνιστούν έναν μικρό τομέα από την ανθρώπινη ακουστική ικανότητα. Πρόσφατες ανακαλύψεις από ερευνητές πάνω στην αντίληψη ομιλίας ,έχουν αναδείξει πόσο πολύπλοκη είναι η ακουστική επεξεργασία για τον άνθρωπο. Ένα γνώρισμα αυτής της αντίληψης ,η ακουστική αντικειμενική μορφοποίηση, (βλεπε κεφ. 2 , τομεα 2,1,4), εξηγεί πως ο άνθρωπος καταφέρνει να διακρίνει την ομιλία από τον θόρυβο υπόβαθρου, όταν ακούει μια φωνή. Η ακουστική αντικειμενική μορφοποίηση κατέφερε , εν μέρει, να ακούσει μερικές συχνότητες μέσα σε ένα μεγάλο εύρος ζώνης συχνοτήτων , σαν αυτές να ήταν μια συχνότητα ,επειδή αυτές μεταβάλλονται μαζι (καλείται *συνδιαμόρφωση)* πράγμα που επιτεύχθηκε μέσω της ικανότητας του ανθρώπινου συστήματος επεξεργασίας. Αφού λοιπόν είναι πλήρως κατανοητές ,διαδικασίες όπως η ακουστική αντικειμενική μορφοποίηση θα φέρουν μεγάλες βελτιώσεις στην ικανότητα των συστημάτων αναγνωρισης ομιλίας να λειτουργούν στον θόρυβο.

Αναλυτικές περιγραφές της ακουστικής μοντελοποίησης που περιγράφτηκαν σ'αυτόν τον τομέα μπορούν να βρεθούν στους Chitza(1994),Cheng & O'Shaughnessy(1991),kai Nandkuman & Hansen(1992).Οι Carrell(to appear) , Carrell&Opie(1992),και Darwin(1981) περιγράφουν την έρευνα του πάνω στην ακουστική αντικειμενική μορφοποίηση και διαμόρφωση.

7,4 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Η πλειονότητα των νευρωτικών δικτύων που σχεδιάστηκε να χειρίζεται τον θόρυβο, επίτυγχε αυτό το σκοπό ,μέσω αντιστοίχησης της αλλοιωμένης από θορυβο εισόδου, σε καθαρό λόγο .Η αναγνώριση ομιλίας , παρουσιάζεται στην έξοδο αυτών των νευρωτικών δικτύων , μέσω χωριστών συσκευών αναγνώρισης ομιλίας. Τα περισσότερα από αυτά τα νευρωνικά δίκτυα λειτουργουν με είσοδο διακριτών λέξεων. Το *δίκτυο απομάκρυνσης θορύβου* (NRN) αποτελει ένα τέτοιο παράδειγμα (βλεπε εικονα 7.5). Είναι ένα αναδρομικό δίκτυο, εκγυμναζόμενο χρησιμοποιώντας οπισθοδρομική διάδοση. Το NRN αντιστοιχεί την αλλοιωμένη από θόρυβο ομιλία σε καθαρό λόγο και στέλνει την έξοδο του προς ένα TDNN δίκτυο (βλεπε κεφ. 2 παρ. 2.5)



Σχήμα 7.5 Το *δίκτυο απομάκρυνσης θορύβου* (NRN)

Ερευνητές στα Interpreting τηλεφωνικά ερευνητικά εργαστήρια του Εξελιγμένου Ερευνητικού Ινστιτούτου Τηλεπικοινωνιών (Advanced telecomunnications Research Institute-ATR) της Ιαπωνιας και ο καθηγητης Helge Sorensen του πανεπιστημίου του Aalborg στην Γερμανία έχει αναπτύξει έμπροσθεν τροφοδοτούντα δίκτυα ελάττωσης θορύβου . Όπως το NRN, αυτά τα δικτυα μετατρέπουν θορυβώδη είσοδο σε καθαρή ομιλία και είναι σχεδιασμένα να λειτουργούν συνδεδεμένα με τα υπάρχοντα συστήματα αναγνώρισης. Το δίκτυο του Sorensen χρησιμοποιεί ένα ακουστικό μοντέλο για να προεπεξεργαστεί είσοδο. Όταν ο Sorensen δοκιμασε αυτη την σχεδιαση σαν την εμπρόσθια κατάληξη για ένα στανταρ σύστημα αναγνώρισης η ακρίβεια αναγνώρισης βελτιώθηκε εως και 65%

Το *επιλεκτικά εκγυμναζόμενο νευρονικό δίκτυο* (STNN) είναι αποτέλεσμα μιας Γαλλοαμερικανικής συνεργασίας. Σε αντίθεση με τα δίκτυα που προαναφέρθηκαν παρουσιάζει ικανότητα αναγνώρισης σε επίπεδο λέξης μέσω συλλαβισμού (κεφ.3 & 3.5.3)σε καθαρό, σε αλλοιωμένο από θόρυβο και σε Lombard λόγο. Είναι επιλεκτικά εκπαιδευόμενο να τοποθετεί ένα φωνητικό σημείο αναφοράς μέσα σε μία διακριτής λέξης εισοδο και να παρέχει μία τελεια διακριτή ανάλυση. Το STNN είναι ένα έμπροσθεν τροφοδότησης και οπισθοδρομικής διάδοσης δίκτυο.

Οι ερευνητές στην Wright Air Force βάση , έχουν αναπτύξει ένα υβριδικό νευρονικό δίκτυο και ένα σύστημα μοντελοποίησης κοχλία για τον εμπλουτισμό της ομιλίας. Αυτό μιμείται τη δικαναλική (καλείται binaural-που σημαίνει : χρησιμοποιώ και τα δύο αυτιά) ανάλυση ομιλίας η οποία έγινε από το ανθρώπινο ακουστικό συστημα επεξεργασίας .

Ο σκοπός είναι να βελτιώσουμε την αναγνώριση του αλλοιωμένου λόγου πάνω στη μονοφωνική (monaural) (μονού καναλιού) περίπτωση μέσω μοντελοποίησης βασικών γνωστών απόψεων του δικαναλικού ακουστικού συστήματος (Martin DeSimio & Timothy Anderson,Wright Patterson AFB, "Αναγνώριση φωνήματος με μοντελα δικαναλικου κοχλία και αναπαράσταση στερέωσης ,»1993, σελ. 1-521).

Τα πρώτα δύο στάδια του συστήματος κάνουν δικαναλικό φιλτράρισμα, διαφορές, άλλες ακουστικής μοντελοποίησης αναλύσεις σαν μέρος της αναλογικής προς ψηφιακής μετατροπή επεξεργασία. Το τρίτο στάδιο

χρησιμοποιεί την ακουστική μοντελοποίηση για να ανασυνδιάσει τα σήματα και μετά τροφοδοτεί τα αποτελέσματα σε ένα αυτοδιοργανώμενο χάρτη (Self Organizing Mup - SOM). Ο SOM παράγει αναγνώριση του ομιλιτή ανεξάρτητα του φαινομένου. Αρχικές δοκιμές του μοντέλου αυτού μείωσαν σημαντικά τα λάθη του σήματος και εισόδου αλλά αύξησαν τα λάθη υποκατάστασης .

Οι Moon & Hwang (1993) περιγράφουν το NRN . Ο Sorensen (1991) και Ohkura & Sugiyama (1991) περιγράφουν τα έμπροσθεν τροφοδότησης , ελάττωσης θορύβου δύκτια τους , και ο Anglade ,et al. (1993) περιγράφει το STNN. Οι DeSimio & Anderson (1993) παρέχουν μια λεπτομερή περιγραφή του Wright Patterson μοντελου κοχλια .

7.5 ΕΚΤΙΜΗΣΗ (της εφαρμογής)

Με σκοπό να χειριστούμε τον θόρυβο υποβάθρου είναι χρήσιμο να καταλάβουμε τους τύπους και τα χαρακτηριστικά του θορύβου που περιέχεται. Αυτό πραγματοποιείται μέσω εκτίμησης της εφαρμογής . Αυτή η εκτίμηση πρέπει να γίνει σε συνθήκες όλων των περιβάλλοντων στα οποία θα τοποθετηθεί η εφαρμογή .Η εκτίμηση αποτελείται από μια ανάλυση του θορύβου στο αναμενόμενο περιβάλλον ομιλίας. Οι βασικοί παράμετροι αυτής της εκτίμησης είναι οι:

- η η φύση του θορύβου υποβάθρου (ομιλία, θόρυβοι κλπ)
- μεταβλητότητα στον θόρυβο υποβάθρου και καναλιού
- η ένταση του θορύβου του καναλιού και του ηχου του υποβάθρου
- τύπος του καναλιού ομιλίας που χρησιμοποιείται(τηλέφωνο-μικρ/νο)
- η ποιότητα του καναλιόυ ομιλίας

Τα περιβάλλοντα με μεγάλο βαθμο θορύβου και μικρό SNR (βλεπε ενοτητα 7.1) θα επηρεάσουν δυσμενώς την ακρίβεια αναγνώρισης .Αυτές οι συνθήκες , εντούτοις, ίσως να μην αναπαριστούν τα πιο προκλητικά περιβάλλοντα ομιλίας. Τα περιβάλλοντα που περιέχουν μεγάλο βαθμό ηχώς ή μεταβλητού θορύβου ειδικότερα κτυπήματα από τηλέφωνο και δυνατές φωνές, θα παρουσιάσουν υψηλότερες τάξεις λάθους από περιβάλλοντα ομιλίας που χαρακτηρίζονται από μηχανικούς θορύβους (μηχανών). Οι ενοχλητικοί θόρυβοι είναι πολύ δύσκολο να δειγματοληπτηθούν και είναι πολύ πιθανό να παράγουν τα ηχητικα κακοταιριάσματα που υποβιβάζουν την ακρίβεια (εικ. 7.2). Ήσυχα περιβαλλοντα που έχουν περιστασιακές εξάρσεις θορύβου μπορούν να προκαλέσουν περισσότερα προβλήματα από περιβαλλοντα με δυνατό , επίμονο θόρυβο γιατί το σύστημα ίσως να μην έχει την κατάλληλη πληροφορία να απορρίψει τον θόρυβο. Ο θόρυβος υποβάθρου είναι η πιο δύσκολη περίπτωση θορύβου για επεξεργασία. Αυτός ανταγωνίζεται απευθείας με το αρχικό σήμα λόγου χρησιμοποιώντας την ίδια ζώνη συχνοτήτων και συγκρίσιμες πύλες μεταβολής συχνότητας.

Όποτε είναι δυνατό, μορφές θορύβου, πηγές, κατευθύνσεις, ηχώ και απόσταση από τον θόρυβο υποβάθρου μιας εφαρμογής , πρέπει να προσδιορίζονται. Αυτή η πληροφορία παρέχει ένα μεσο εκτίμησης της καταλληλότητας της εφαρμογής για την αναγνώριση ομιλιας .

Τα προϊόντα αναγνώρισης θα πρέπει να είναι εντελώς δοκιμασμένα μέσα στο περιβάλλον της εφαρμογής. Αν ένας εξαρτώμενος αναγνωριστής ομιλίας πρόκειται να χρησιμοποιηθεί, η δοκιμή πρέπει να βασιστεί πάνω σε μοντέλα παρόμοια με το αναμενόμενο περιβάλλον θορύβου. Τυπικά συνθήκες με

λιγότερο θόρυβο (ήσυχα γραφεία) χρειάζονται λιγότερα δείγματα εκμάθησης, ειδικά αν το σύστημα αναγνώρισης πρόκειται να κατασκευαστεί με υψηλής πιστότητας κατευθυντικά μικρόφωνα. Περιβάλλοντα με θόρυβο , όπως πατώματα εργοστασίων, χρειάζονται περισσότερα tokens και ακόμη θα πρέπει να συμπεριληφθεί και δειγματοληψία θορύβου υποβάθρου. Η αναποτελεσματικότητα ενός ήσυχου περιβάλλοντος μπορεί να εμπλουτιστεί παρέχοντας εξάσκηση μέσασε θόρυβο. Αν υπάρχει μία πιθανότητα να προκληθεί θόρυβος ομιλίας, λίγη εκπαίδευση μέσα σε ένα θορυβώδες περιβάλλον (π.χ. καφετέρια) θα μπορούσε να βοηθήσει στην ακρίβεια του συστήματος αναγνώρισης .Η παρουσία υψηλή ηχούς , ίσως πρέπει να διευθετηθεί μέσω χρήσης μικροφώνων κοντινής ομιλίας (τέτοιων όπως των μικροφώνων πανω στα headset), στενής δέσμης κατευθυντικών μικροφώνων ή μικροφωνικών σειρών (διατάξεων).

7.6 ΤΕΧΝΙΚΕΣ ΣΧΕΔΙΑΣΜΟΥ «ΑΝΘΕΚΤΙΚΩΝ» ΣΥΣΤΗΜΑΤΩΝ ΟΜΙΛΙΑΣ

Η "*Ανθεκτικότητα*" , περιγράφει την ικανότητα του συστήματος αναγνώρισης ομιλίας ή της εφαρμογής , να λειτουργήσει σε ποικίλα περιβάλλοντα. Η σημασία της ανθεκτικότητας αυξάνεται, σαν ένα υποπροϊόν της επιθυμίας να χρησιμοποιήσουμε αναγνώριση ομιλίας σε μία μεγαλύτερη ποικιλία περιβάλλοντων.

Παρόλο που πολύ έρευνα έχει αφιερωθει στο να προωθησει συστήματα μεμονομενων λέξεων για εκδόσεις συνεχούς αναγνώρισης λόγου, έχει πραγματοποιηθεί περιορισμένη πρόοδος προς τη κατεύθυνση του θορύβου και της επίδρασης Lombard (John H.L Hansen,Εργαστηριο Επεξεργασιας Ανθεκτικού Λόγου, Πανεπιστημιο Duke, προσωπικές επικοινωνίες ,1994).

Πολλές προσεγγίσεις εστιάζουν στον προσθετικό θόρυβο και βασίζονται στη γνώση σχετικά με τα φασματικά πρότυπα καθώς και πρότυπα έντασης αυτού του θορύβου. Γενικά η πληροφορία σχετικά με τον προσθετικό θόρυβο λαμβανεται μέσω *των τεχνικών ανίχνευσης θορύβου* οι οποίες εφαρμόζονται αυτόματα από το σύστημα αναγνώρισης ομιλίας κατά την διάρκεια περιόδων μη ομιλίας.

Η προσπάθεια να περιοριστούν οι αρνητικές συνέπειες του θορύβου από υπόβαθρο, κανάλι, ή ομιλία ,κατευθύνονται στην αύξηση της νοήμων επεξεργασίας πάνω στην ομιλία. Όταν αυτή περιλαμβάνει ταυτοποίηση, μείωση ή απομάκρυνση του θορύβου από το σήμα καλείται μείωση θορύβου. Οι πιο επιτυχείς και συνήθες τεχνικές είναι:

- εκγύμναση
- προεπεξεργασία
- αφαίρεση θορύβου

Όταν εστιάζουμε πάνω στην βελτίωση της ποιότητας του σήματος ομιλίας , η επεξεργασία καλείται εμπλουτισμός ομιλίας ,παρόλο που ο όρος αυτός μερικές φορές χρησιμοποιείται για να αναφερθεί στην μείωση / περιορισμό του θορύβου.

Αυτές οι τεχνικές, συνδιαζόμενενες με την κατάλληλη επιλογή μικροφώνου και τοποθέτησης , μπορούν να εμπλουτίσουν σε μεγάλο βαθμό την ακρίβεια και την σκληρότητα των περισσότερων εφαρμογών. Οι έρευνες πάνω στην μείωση του θορύβου και στον εμπλουτισμό της ομιλίας

συνεχίζονται, παρόλο που ο εμπλουτισμός της ομιλίας εφαρμόζεται μόνο στα εργαστήρια. Η τεχνολογία των μικροφώνων που συναντάμε στο εμπόριο βελτιώνεται σταθερά και βοηθάει στην προώθηση της αναγνώρισης ομιλίας σε περισσότερο σκληρά περιβάλλοντα ομιλίας. Καμία απ' αυτές τις προσεγγίσεις ,ειδικά όταν εφαρμόζεται σε συστήματα απλών μικροφώνων είναι συγκρίσιμη στην ικανότητα μείωσης θορύβου και εμπλουτισμού ομιλίας των ανθρώπινων όντων.

Έχουμε εκατομμύρια χρόνων εξέλιξης ,για να μας βοηθήσει να αντιμετωπίσουμε τον ήχο και την ηχώ. Οι υπολογιστές δεν έχουν αυτέ τις ικανότητες. Τόσο πολλά πράγματα που ακούγονται ξεκάθαρα σε εμάς ακούγονται διαστρεβλωμένα σε ένα υπολογιστή. Οι υπολογιστές πρέπει να ξανακαλύψουμε τον τρόπο με τον οποίο οι άνθρωποι επεξεργάζονται τον ήχο (Bill Portar,General Manager,AT&T Εξυπνα Ακουστικά Συστήματα , προσωπική επικοινωνία ,1995).

Ώσπου οι υπολογιστές να μάθουν να χειρίζονται το ίδιο καλά τον θόρυβο όσο ο άνθρωπος, η ευθύνη της διευθέτησης του θορύβου και των θεμάτων δημιουργίας θορύβου θα εξακολουθεί να εξαρτάται από τον σχεδιαστή της εφαρμογής.

7.6.1 ΕΚΓΥΜΝΑΣΗ

Αν ο αναμενόμενος θόρυβος υποβάθρου είναι γνωστός και αν υπάρχει ένας μικρός αριθμός διακριτών συνθηκών θορύβου , είναι δυνατό να "εκπαιδεύσουμε" το μοντέλο στον αναμενόμενο θόρυβο το ίδιο καλά όπως και σε συνθήκες ησυχίας. Αυτή η διαδικασία (καλείται *multi-style* εκμάθηση) επιτρέπει στον θόρυβο να αποδιδεται σαν μέρος του κανονικού σήματος. Είναι συνήθης τεχνική για ανάπτυξη εξαρτόμενης από τον ομιλητή εφαρμογής και χρησιμοποιείται το ίδιο καλά για ανάπτυξη μοντέλων ανεξαρτήτων από τον ομιλητή για τηλεφωνικά συστήματα. Η multi style εκμάθηση δουλεύει πολύ καλά σε περιβάλλοντα που περιέχουν θόρυβο. Όταν όμως παρουσιάζονται μεταβλητές ή απρόβλεπτες συνθήκες θορύβου, η δειγματοληψία εισαγωγής στοιχείων μπορεί να γίνει μακριά και δύσκολη.

Μια πιο αποτελεσματική προσέγγιση για συνθήκες μεταβλητού θορύβου ,περιλαμβάνει ανάπτυξη ανοσίας στο θόρυβο σαν μέρος της δημιουργίας των μοντέλων.

Εστιάζουμε στην ανάπτυξη πιθανοθεωρητικών μέτρων που είναι ανθεκτικά στις επιδράσεις του θορύβου παρά στην προσπάθεια να απομακρύνουμε τον θόρυβο από τον αποκλιμακούμενο λόγο (Beth Carlson &Mark Clements ,Georgia Institute of Technology, «Speech recognition in noice using a projection –based likelihood measure for mixture density HMM's," 1992, σελ. 237).

Μερικοί ερευνητές έχουν βρει ότι προσθέτοντας θόρυβο περιβάλλοντος σε μοντέλα καθαρού λόγου θα αυξήσουμε την ανθεκτικότητα. Άλλοι ερευνητές τροποποιούν τους αλγόριθμους που δημιουργούν αυτά τα μοντέλα.

Consult Ohkyra ,et al.(1991)για αξιολόγηση της ανθεκτικότητας των δύο αλγορίθμων εκμάθησης. Carlson & Clements (1992) 1992 γράφει άλλες ανθεκτικές μετρήσεις. Das, et al.(1993) περιγράφει μία μελέτη της IBM συνιστώντας την περίληψη του θορύβου στα μοντέλα (7.1 3.1).

7.6.2 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ

Σκοπός μας είναι να καταλάβουμε και να περιορίσουμε την μεταβλητότητα στο σήμα ομιλίας εξαιτίας των αλλαγών του περιβάλλοντος και κατά αυτόν τον τρόπο ,τελικα, να αποφύγουμε την ανάγκη για παραπέρα εκμάθηση του συστήματος αναγνώριση σε διαφορετικά περιβάλλοντα (Hynec Hermansky of US West and Nelson Morgan & Hans-Gunter Hirsch of International computer Scince Institute, "Αναγνώριση ομιλίας σε προσθετικό και μεταβαλλόμενο θόρυβο βασιζόμενο σthRASTA φασματική επεξεργασία," 1993, σελ.83)

Η τεχνική της προεπεξεργασίας απομακρύνει τον προσθετικό θόρυβο από το σήμα πριν αυτός ψηφιοποιηθεί. Η συνήθης χρησιμοποιούμενη προσέγγιση είναι το φιλτράρισμα του σήματος. Η λειτουργία των φίλτρων είναι να περιορίζουν τις συχνότητες που μπορούν να περάσουν μέσα απ' αυτά. Μόνο τα *χαμηλοπερατά φίλτρα* αφήνουν να περάσουν συχνότητες μικρότερες από μία σχεδιαζόμενη συχνότητα (ή φάσμα συχνοτήτων*), υψηπερατά φίλτρα* περιοριζουν τις συχνότητες τις πάνω από μία ορισμένη συχνότητα (ή φάσμα συχνοτήτων) και τα *ζωνοπερατά φίλτρα* που περιορίζουν όλες εκτός από ένα ορισμένο εύρος συχνοτήτων. Τα υψηπερατά και καλά σχεδιασμένα ζωνοπερατά φίλτρα έχουν αποδειχθεί τα πιο επιτυχή για συστήματα αναγνώρισης ομιλίας. Μερικά φίλτρα βασίζονται πάνω στην γραμμική (Hz) κλίμακα αλλά ένας αυξανόμενος αριθμός σχεδιάζονται χρησιμοποιώντας την κλίμακα mel (κεφ2. Ενοτητα 2.1.4) ή μία αλγοριθμική κλίμακα.

Hermansky, et al.,(1993)Hansen &Applebaum (1993), and Hansen & Arslan(1995) περιγράφοyn συγκεκριμένες προεπεξεργαστικές υλοποιήσεις .

7.6.3 ΑΚΥΡΩΣΗ ΘΟΡΥΒΟΥ

Η ακύρωση θορύβου απομακρύνει ή εξασθενιζει τις συχνότητες που συνοδεύονται από θόρυβο. Αυτό μπορεί να επιτευχθεί με ποκίλους τρόπους. Οι πιο συνήθεις μέθοδοι είναι η φασματική *αφαιρεση και συγκαλυψη*. Όπως και το όνομα μαρτυρά η μέθοδος αυτή αφαιρεί τις συχνότητες του θορύβου από το σήμα. Όταν ο θόρυβος επικαλύπτει ένα μεγάλο εύρος συχνοτήτων συμπεριλαμβανομένων και των σημαντικών συχνοτήτων ομιλίας, η τεχνική συγκάληψης , η οποία ρυθμίζει την ισχύ αυτών των συχνοτήτων ώστε να προσεγγίζει περισσότερο το σήμα ομιλίας, έχει αποδεικτεί περισσότερο αποτελεσματική. Η φασματική αφαίρεση και συγκάλυψη χρησιμοποιούν την συλλεγόμενη πληροφορία από την ανίχνευση θορύβου ,για να υπολογίσουν το φάσμα ισχύος του προσθετικού θορύβου και το αφαιρούν από το σήμα μας. Συχνά, φέρουν εις πέρας την μέθοδο αυτή, χρησιμοποιώντας δύο ή περισσότερους μικροφωνικούς αισθητήρες (κάψες) , οι οποίοι τοποθετούνται είτε στο ίδιο μικρόφωνο είτε παρατάσσονται σε μια μικροφωνική σειρα (διάταξη).

Μερικοί αλγόριθμοι αφαίρεσης θορύβου σχεδιάστηκαν για τηλεφωνικά συστήματα απομακρύνοντας τον θόρυβο από *τον πλευρικό τόνο* του ακουστικού. Ο πλευρικος τόνος είναι μέρος της κυκλωματικής ανάδρασης φωνης στο τηλεφώνο, που επιτρέπει στον ομιλιτή να ακούει την φωνή του. Με την ελάττωση του θορύβου *στον πλευρικό τόνο,* αυτοί οι αλγόριθμοι

ελαττώνουν την επίδραση Lombard. Τα μικρόφωνα αφαίρεσης θορύβου είναι κατάλληλα για περιπτώσεις υψηλού θορύβου επειδή ελαττώνουν τον θόρυβο από το σημείο εισόδου. (τεχνικές περιγραφές της μεθόδου αφαίρεσης θορύβου δίνονται από Klatt (1976), Rabiner & Juang (1993, κεφ. 5)).

7.6.4 ΜΙΚΡΟΦΩΝΑ

Τα μικρόφωνα που επιλέγονται για χρήση πρέπει να ικανοποιούν τις ανάγκες του συστήματος αναγνώρισης, του περιβάλλοντος, των εφαρμογών, του χρήστη. Τα μικρόφωνα που χρησιμοποιούνται γενικά για αναγνώριση λόγου ,είναι υψηλής ποιοτητας *κατευθυντικα, αφαίρεσης θορύβου και κοντινής ομιλίας* μικρόφωνα. Τα κατευθυντικά μικρόφωνα ελαττώνουν την επικίνδυνη επίδραση του θορύβου που προέρχεται από το υπόβαθρο, γιατί είναι πιο ευαίσθητα στον ήχο που έρχεται από συγκεκριμένη κατεύθυνση, όπως του ομιλιτή. Η αφαίρεση θορύβου ελαττώνει τις συχνότητες του θορύβου από το θόρυβο υπόβαθρου μέσα στο σήμα. (περισσότερες πληροφορίες στη παραγραφο 7.1.3.1)

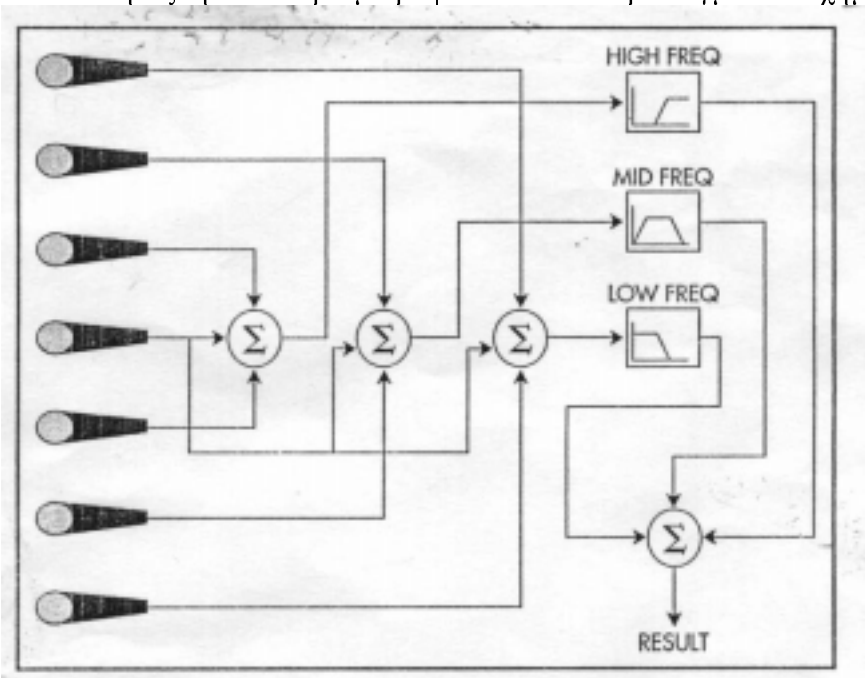
Η ποιότητα του μικροφώνου πρέπει να αυξάνει με την αύξηση της απόστασης από το στόμα του ομιλιτή, ειδικά για περιβάλλον που περιέχει υπόβαθρο έντονης ομιλίας. Γι' αυτό τον λόγο οι παραγωγοί συστημάτων αναγνώρισης προτιμούν τα μικρόφωνα κοντινής ομιλίας. Αυτά τα μικρόφωνα σχεδιάζονται ώστε να διεγείρονται από τις πολύ κοντινές ηχητικές πηγές και είναι κατάλληλα όταν οι ομιλητές τοποθετούν το μικρόφωνο πολύ κοντά ή απέναντι στο στόμα τους. Τέτοια μικρόφωνα μπορούν να χρησιμοποιηθούν στα τηλέφωνα και γενικότερα στα μικρόφωνα που είναι κρεμαστά μπροστά από το στόμα. του ομιλητή.

Η κατάλληλη τοποθέτηση των μικρόφωνων αυτών είναι καθοριστική για την ακριβή αναγνώριση ήχου. Η τοποθέτηση του μικροφώνου απέναντι από το στόμα(αντί της άμεσης τοποθέτησης μπροστά στο στόμα) λαμβάνει το κατάλληλο φάσμα χωρίς τον υπερβολικό θόρυβο που δημιουργείται από την εκφορά των τελικών συμφώνων (π.χ. π, τ, κ). Επίσης ελαττώνουμε την επίδραση του θορύβου που δημιουργείται από `ομιλία χωρίς μήνυμα` και είναι λιγότερο πιθανό να κατακρατήσει υγρασία το μικρόφωνο Ακόμη και με την κατάλληλη τοποθέτηση ,τα περισσότερα μικρόφωνα κοντινής ομιλίας χρειάζονται προκάλυμμα για να προστατέψουν από την παραμόρφωση που γεννιέται από την προφορά τελικών συμφώνων αλλά και να μειώσουν την κατακράτηση υγρασίας . Δυστυχώς, η επιτυχής τοποθέτηση των μικροφώνων είναι δύσκολη. Ακόμα και αν οι ομιλητές γνωρίζουν να τοποθετούν τα μικρόφωνα σωστά, κακές συνήθειες ίσως εμφανιστούν. Η τοποθέτηση μικροφώνων σε headset είναι καλύτερα ρυθμίσιμη σε σχέση με τα συνηθισμένα, γιατί γίνεται η πιο εύκολη η μετακίνηση του από τον χρήστη. Οι χρήστες τείνουν να τοποθετούν το μικρόφωνο μπροστά στο κέντρο του στόματος ή κάτω και μακριά απ' αυτό. Παρόμοια προβλήματα εμφανίζονται με τα boom μικρόφωνα.

Τα μικρόφωνα χειρός αποδίδουν καλύτερα όταν υπάρχει μικρή ομιλία στο υπόβαθρο. Τα στηριζόμενα σε μόνιμη βάση (στάνη) μικρόφωνα, ή αυτά που ενσωματώνονται στα ρούχα ή στον εξοπλισμό μας, πρέπει να δοκιμαστούν προσεκτικά πριν χρησιμοποιουθούν. Και αυτό γιατί αν αλλάξει η απόσταση, το ρούχο, ο μηχανικός ήχος, ή ελαττωθεί η ισχύς του σήματος ομιλίας (με

απομάκρυνση του ομιλητή από το μικρόφωνο) θα πρέπει να ξαναρυθμιστούν.

Τα μικρόφωνα κοντινής ομιλίας είναι ακατάλληλα για εφαρμογές μακρινού πεδίου (π.χ. κιόσκια) και μπορεί να απορριφθούν από τους χρήστες ως άβολα και φορτικά. Αυτά τα μικρόφωνα μπορούν να αντικατασταθούν από κατευθυντικά μικρόφωνα αφαίρεσης θορύβου μεγάλης κατευθυντικότητας ή από σειρά μικροφώνων. Ένα παράδειγμα επιτυχημένης



Σχήμα 7.6 Σχηματική αναπαράσταση της ιδέας της AT& T., που απορρέει από τη γραμμική παράταξη σειράς μικροφώνων , στο κιόσκι *Alice's Interactive Wonderland Kiosk* .

εφαρμογής σειράς μικροφώνων ,σε θορυβώδες περιβάλλον ομιλίας με αντήχηση , είναι το κιόσκι *Alice's Interactive Wonderland Kiosk* στο Epcot center στη Disney World. Η γραμμική παράταξη σειράς μικροφώνων ,που σχεδιάστηκε από την AT&T και φαίνεται στο σχήμα.7.6,στερεωθηκε στο ταβανι του κιοσκι , πανω από το μικρό καναπέ οπου κάθονταν οι ομιλητές .Οι ομιλητές επικοινωνούν με κινούμενα σχέδια που τους κάνουν ερωτήσεις .Οι απαντήσεις τους αναγνωρίζονταν χρησιμοποιώντας ένα αναγνωριστή εντοπισμού λέξεων με εικόνα (βλεπε κεφ. 4 ,ενοτητα 4.5) σε έναν υπολογιστή .Η παράταξη σειράς μικροφώνων δημιουργεί ένα τρισδιάστατο χώρο μέγιστης ακουστικής ευαισθησίας (μερικές φορές αναφέρεται και σαν sweet spot) γύρω από το καναπέ . Το σχήμα του είναι παρόμοιο σαν χάρτινης κούπας με πλατύ πυθμένα. Η οργανωμένη κατά φάσεις (χρονικές περιόδους) είσοδος στα μικρόφωνα ,που παρουσιάζονται στο σχήμα 7.6, ρυθμίζει την είσοδο ώστε να ερμηνευθεί τις διαφορές στον χρόνο, με τις οποιες η ομιλία φτάνει σε κάθε ζευγάρι μικροφώνων. Η είσοδος περνά μέσα από διάφορα φίλτρα (φαίνονται στο δεξιό μέρος της παρένθεσης) πριν να σταλεί στον προεπεξεργαστή. Η γραμμική παράταξη μικροφώνων που φαίνεται στην εικόνα 7.6 σχεδιάστηκε προσεκτικά για την ακουστική που απαιτείται για τα κιόσκια. Έχει κάνει το σύστημα απρόσβλητο απο ήχους που πηγάζουν από το εξωτερικό μέρος του sweet spot, όπως τους ήχους από την γειτονική δραστηριότητα και συνεισφέρει στην υψηλή ακρίβεια (καλύτερη από 95% του συστήματος αναγνώρισης).

Μερικά συστήματα αναγνώρισης μπορούν να χρησιμοποιήσουν. Σε αυτά

τα μικρόφωνα ο χρήστης πρέπει να πιέζει ένα κομβίο ή ένα πετάλι ποδιού κατά την διάρκεια της ομιλίας.Τα «πιεσε-να-μιλησεις» μικρόφωνα είναι χαρακτηριστικό πολλών μικροφώνων που κρατάμε στο χέρι, αλλά μπορεί να χρησιμοποιηθεί και με όλους τους άλλους σχηματισμούς μικροφώνων, συμπεριλαμβανομένων και των headset. Εμπλουτίζει την ακρίβεια σε περιβάλλοντα υψηλού θορύβου με το να διεγείρει το σύστημα αναγνώρισης με την είσοδο της ομιλίας. Συστήματα που χρησιμοποιούν «πιεσε-να-μιλησεις» μικρόφωνα , ίσως επισης να κατασκευαστούν ώστε να επιτρέπουν στους χρήστες να σταματάνε στο μέσο της ομιλιας χωρίς να σηματοδοτούν το τέλος της ομιλίας.Τα «πίεσε-να-μιλήσεις» κομβια , είναι λιγότερα χρησιμα για "ελευθερων χεριων" εφαρμογες ,εκτός και αν μπορεί να χρησιμοποιηθεί ένα πετάλι ποδιού . Αν η μέθοδος «πιεσε-να-μιλησεις» επιλεγει, οι χρήστες πρέπει να εκπαιδευτούν στον καταλληλο συγχρονισμό του κομβίου και της ταχύτητας ομιλίας. Ειδικότερα πρέπει να μάθουν να πατάνε το κομβιο αμέσως πριν ξεκινήσουν να μιλάνε, παρά καθώς ξεκινάνε να μιλάνε. Επίσης να αφήσουν το κομβιο αφού η ομιλία είναι εντελώς ολοκληρωμένη, παρά καθώς αυτοί τελειώνουν.

Μία τεχνική λογισμικού παρόμοια μ' αυτή του «πιεσε-να-μιλησεις» ,για ευμετάβλητα περιβάλλοντα ή για περιβάλλοντα υψηλού θορύβου,είναι μέσω *συνθηματικών λέξεων* όπως «Υπολογιστής» ή «Είσοδος». Αυτές οι λέξεις σηματοδοτούν τον εισερχόμενο λόγο και επιτρέπουν στο σύστημα να σταματάει να άκουει κατά την διάρκεια περιόδων μη ομιλίας.

Ο έλεγχος της ενίσχυσης του μικροφώνου είναι κάτι που πρέπει να ληφθεί υπόψη. Σε μερικά συστήματα ο έλεγχος αυτός είναι αυτόματος. Άλλα συστήματα επιτρέπουν στον χρήστη την ρύθμιση αυτή. Ο αυτόματος έλεγχος ενίσχυσης συνήθως προτιμείται περισσότερο. Ένας κίνδυνος της δια του χειρός ενίσχυσης είναι ότι ίσως αποτελεί περισσότερο το ανακλαστικό της αντίδρασης του χρήστη σε ισχυρό θόρυβο υποβάθρου ή λάθος αναγνώρισης απ' ότι μία απαίτηση του συστήματος αναγνώρισης. Μία ρύθμιση ενίσχυσης που είναι πολύ μεγάλη μπορεί να δημιουργήσει παραμόρφωση. Ένα αυτόματο σύστημα προϋποθέτει υπευθυνότητα για τον υπολογισμό της κατάλληλης ενίσχυσης. Αυτό μπορεί να βασιστεί πάνω σε δεδομένα εγγραφής , την αρχική είσοδο , ή ένα μέσο όρο πάνω σε όλους τους ομιλητές. Η επιλογή της κατάλληλης μεθόδου ενίσχυσης εξαρτάται από τις απαιτήσεις της εφαρμογής. Τα περιλαμβανόμενα, υψηλού θορύβου περιβάλλοντα και εφαρμογές σε συνθήκες που απαιτούν απαλής ομιλίας είσοδο, είναι ευκολότερα για συστήματα αυτόματου ελέγχου ενίσχυσης απ' ότι εφαρμογές σε μεταβαλλόμενα περιβάλλοντα .Σε όλες τις περιπτώσεις η μεθοδολογία ελέγχου ενίσχυσης , θα πρέπει να ελέγχεται σαν μέρος της αξιολόγησης του προϊόντος .

Η ποιότητα των μικροφώνων και η τεχνολογία ακύρωσης θορύβου μεσω μικροφώνου έχει βελτιωθεί σταθερά .Κατασκευαστές μικροφώνων όπως η GENTEX ,παράγουν directional μικρόφωνα σχεδιασμένα για ελαφρά headset ή για τοποθέτηση στις οθόνες των υπολογιστών. Επισης αυτά τα μικρόφωνα , σχεδιάζονται να είναι λιγότερο ευαίσθητα στην υγρασία και στην ηλεκτρομαγνητική παρεμβολή. Η διαθεσιμότητα της υψηλότερης ποιότητας μικροφώνων ,αυξάνεται καθώς οι τιμές πέφτουν και ενδιαφέρον για ακουστική είσοδο σε υπολογιστή μεγαλώνει .Π.χ μικρόφωνα κατάλληλα για αναγνωριση ομιλιας προσφέρονται σαν βασικός εξοπλισμός στους προσωπικούς υπολογιστές .

Η ανεπιθύμητη παραμόρφωση που παρουσιάζεται στα ακουστικά έχει παράγει *λογισμικό ακύρωσης ηχούς* , προκείμενου να μειώσει , την παρόμοια με «βαρέλι» , ποιότητα των ακουστικών και στις μακράς απόστασης τηλεπικοινωνίες . Ακόμα μπορεί να χρησιμοποιηθεί ώστε να παρέχει ένα καθαρότερο σήμα σε ένα σύστημα αναγνώρισης ομιλιας

7,7 ΧΕΙΡΙΖΟΜΕΝΟΙ ΤΟΝ ΘΟΡΥΒΟ ΚΑΝΑΛΙΟΥ

Ένα από τα θέματα που αντιμετωπίζουν οι κατασκευαστές συστημάτων και εφαρμογών αναγνώρισης ομιλιας είναι η φύση και η ποιότητα της συσκευής εισόδου που χρησιμοποιείται για αναγνωριση.Οι συχνότητες αλλά και η ποσότητα του θορύβου που γεννιέται από το κανάλι εισόδου , τα χαρακτηριστικά συχνοτικης απόκρισης του και η παρουσία των περισσοτέρων του ενός καναλιού επιδρούν στην ποιότητα της αναγνώρισης.

7,7,1 Η ΠΟΙΟΤΗΤΑ ΤΟΥ ΜΙΚΡΟΦΩΝΟΥ

Τα μικρόφωνα ποικίλουν πολύ τόσο στην ποιότητα όσο και στην λειτουργία .Όλες οι εφαρμογές χρειάζονται καλής ποιότητας κατευθυντικά μικρόφωνα , αλλά ακόμα και ένα υψηλής ποιότητας μικρόφωνο μπορεί να μην είναι καλά προσαρμοσμένο σε ένα συγκεκριμένο σύστημα αναγνώρισης. Προκείμενου λοιπόν να ελαχιστοποιήσουν τις ασυμβατότητες , μερικοί παραγωγοί παρέχουν ή συνιστούν συγκεκριμένους τύπους μικροφώνου , για τα προϊόντα τους. Εάν το μικρόφωνο είναι του κατασκευαστή ή ένα άλλο χρησιμοποιείται για μια εφαρμογή, δεν συνίσταται να εκπαιδευτείτε σε κάποιο μικρόφωνο και να χρησιμοποιείτε διαφορετικό κατά την ανάπτυξη της εφαρμογής . Κάτι τέτοιο θα επιδράσει εναντίον της ακρίβειας ακόμα και σε καλές συνθήκες ομιλιας (βλέπε παράγραφο 7,1,3 και εικόνα 7,2)

7.7.2 ΤΗΛΕΦΩΝΑ

Η κατασκευή μίας τηλεφωνικής εφαρμογής είναι ιδιαίτερα απαιτητική (Βλέπε επίσης κεφ. 8).Οι διακυμάνσεις στη ποιότητα του τηλεφώνου ,τα χαρακτηριστικά της γραμμής μεταφοράς ,και ο τύπος εκπομπής συνδυάζονται με το τηλεφωνικό σφύριγμα και μπορούν να κάνουν λέξεις όπως το six τουλαχιστον ακαταλαβίστικες . Στην θέα αυτής της αποκρουστικής κατάστασης του τηλεφωνικού καναλιού , δεν είναι προς έκπληξη το γεγονός της ύπαρξης κατασκευαστών, που συγκεντρώνουν περισσότερα των 10.000 tokens προκειμένου να σχεδιάσουν ψηφιακά μοντέλα για συνεχή ομιλία.

Στις ΗΠΑ η ποιότητα των τηλεφωνικών δικτύων είναι συγκριτικά ομοιόμορφη. Οι διαφορές είναι στο μεγαλύτερο μέρος, εξαιτίας των χαρακτηριστικών που διακρίνουν τη χρήση καλωδιακού ή κινητού τηλεφώνου καθώς και δορυφορικής σύνδεσης. Αν δεν είναι η συσκευή αναγνώρισης ενσωματωμένη στο τηλέφωνο, η ποιότητα του τηλεφώνου αποτελεί ένα δεύτερο επίπεδο της παραπάνω μεταβλητότητας Ένα τηλέφωνο κόστους μικρότερο των 10 δολαρίων θα έχει απόκριση πολύ διαφορετική από ένα άλλο που κοστίζει πάνω από 100 δολάρια. Και αυτή η διαφορά οφείλεται

στη ποιότητα των στοιχείων και των αλγορίθμων που χαρακτηρίζουν τα δύο τηλέφωνα.

Εφαρμογές που σχεδιάστηκαν για χρήση εκτός της περιοχής των ΗΠΑ πρέπει να γίνονται με προσοχή. Η μεγαλύτερη πηγή δυσκολίας, προέρχεται απ' την ιδιοσυγκρασία στη συμπεριφορά του εθνικού συστήματος τηλεφωνικού δικτύου. Έτσι, μερικά απ΄αυτά τα δίκτυα, χαρακτηρίζονται από υψηλά επίπεδα θορύβου και φτωχή ποιότητα εκπομπής, Όμως ακόμα τα χαρακτηριστικά του δικτύου ενδέχεται να ποικίλουν δραματικά μέσα και κατά μήκος της ίδιας της χώρας. Τα περισσότερα τηλέφωνα έξω απ' τις ΗΠΑ, χρησιμοποιούν μικρόφωνα άνθρακα, όπου προσθέτουν ένα υπολογίσιμο μέρος παραμόρφωσης σήματος και προσθετικού θορύβου.

7. 8 ΧΕΙΡΙΖΟΜΕΝΟ ΤΟΝ ΘΟΡΥΒΟ ΑΠΟ ΟΜΙΛΙΑ ΧΩΡΙΣ ΜΗΝΥΜΑ

Το ενδιαφέρον για τις επιδράσεις του θορύβου από ομιλία `χωρίς μήνυμα` έχει επεξεργαστεί μαζί με την εμφάνιση συστημάτων υπαγόρευσης ελεύθερης μορφής . Θα αναπτυχθεί σαν ελεύθερης μορφής , συστήματα συνεχούς ομιλίας εμφανίζονται και χρησιμοποιούνται από διαφορετικούς πληθυσμούς χρηστών. Είναι ένα θέμα που αντιμετωπίζει συστήματα όλων των τύπων και λειτουργιών , ειδικότερα εκείνων που χρησιμοποιούν συνεχή ομιλία.

Τα λάθη αναγνώρισης που προκύπτουν απ΄' τη ταυτοποίηση α-λογης επικοινωνίας ως έγκυρης απόκρισης ,μπορούν να βγάλουν το σύστημα αναγνώρισης εκτός συγχρονισμού με το πρόγραμμα της εφαρμογής .

Θόρυβος από ομιλία `χωρίς μήνυμα` , όπως το "ωχ" ή οι ήχοι από τα χείλια, είναι δυσκολότερο να σβηστεί απ'ότι ο θόρυβος καναλιού ή υποβάθρου. Ένα μέρος αυτού, ειδικά η διάθεση αυτοδιόρθωσης είναι λίγο κατανοητός όπως π.χ.

"Ο αριθμός είναι 55-12 ΟΧ!! 555-2134"

Άλλη, `χωρίς μήνυμα ` συμπεριφορά μπορεί να διευθετηθεί στις μέρες μας. Κατάλληλη τοποθέτηση του μικροφώνου και κυρίως μικρόφωνα κοντινής ομιλίας μπορούν να ελαττώσουν τις επιδράσεις ενός μέρους του θορύβου ομιλίας και ειδικότερα τις συμπιέσεις του αέρα που συνοδεύουν βασικούς ήχους ομιλίας. Καλοσχεδιασμένες εφαρμογές , καλά μοντέλα αναφοράς και ο εγκλιματισμός του χρήστη στην εφαρμογή, όλα μαζί βοηθούν να μειωθούν τα λάθη αναγνώρισης που συνδέονται με την είσοδο ομιλίας `χωρίς μήνυμα . Μοντέλα αναφοράς, μπορούν να δημιουργηθούν για συχνά συναντώμενους τύπους ομιλίας `χωρίς μήνυμα ` , όπως οι αρχικά παραγόμενοι ήχοι, απ' τα χείλια (σε μία κουβέντα) και το γέμισμα των κενών ομιλίας (π.χ. με το εεε) . Δημιουργία μοντέλων γεμίσματος των κενών ομιλίας , εμπλουτίζει την "αυθεντικότητα" των συστημάτων συνεχούς ομιλίας, ιδιαίτερα.(Βλέπε κεφ.6 ενότητα 6.3 και 6.7).Αντίγραφα και άλλοι μηχανισμοί διόρθωσης λάθους μπορούν να βοηθήσουν τους χρήστες να βγάλουν το σύστημα από λάθος δρόμο. Οπτική ανατροφοδότηση από Video ή ακουστική επιβεβαίωση της εισόδου, μπορεί να βοηθήσει τους χρήστες να διορθώσουν τα λάθη, επίσης καλά. Η σχεδίαση συστημάτων ενημέρωσης μέσω λόγου(βλέπε κεφ. 2, παράγραφος 2.8) συνεισφέρει σπουδαία στην ανθεντικότητα της εφαρμογής , μέσω ελάττωσης της πιθανότητας ότι το

interface ομιλίας θα χάσει το συγχρονισμό με το λοιπό software της εφαρμογής.

7.9 ΧΕΙΡΙΖΟΜΕΝΟ ΤΟΝ ΛΟΓΟ LOMBARD

Είναι επιτακτική ανάγκη για τους ερευνητές ομιλίας να εστιάσουν πάνω σε βελτιωμένα σχήματα μοντελοποίηση ομιλίας, προκειμένου να διευθετήσουν καλύτερα τη ποικιλία των δυναμικών κινήσεων των αρθρώσεων όπου προκαλούνται όταν οι ομιλητές είναι στρεσαρισμένοι (συμπεριλαμβανομένου και του αποτελέσματος Lombard) (John H.L. Hansen, Roburt Speech Processing Labortory, Duke Universtity, Personal Communication, 1993)

Λίγη προσοχή έχει δοθεί στην μείωση της αρνητικής επίδρασης της ομιλίας Lombard στην ακρίβεια αναγνώρισης. Η πιο συχνά εφαρμοζόμενη τεχνική για την ελάττωση της ομιλίας Lombard είναι η "πολυμορφική" εκπαίδευση, όπου παράγει μοντέλα εξαρτώμενα του ομιλητή, που περιλαμβάνουν συγχρόνως Lombard και μη Lombard ομιλία.. Αυτή η προσέγγιση δουλεύει άριστα εάν το λεξιλόγιο είναι σχετικά μικρό, οι συνθήκες θορύβου είναι ομοιόμορφες και οι ομιλητές συνεργάσιμοι. Αφού όμως τα χαρακτηριστικά της ομιλίας Lombard , έχουν βρεθεί να ποικίλουν με τις συνθήκες θορύβου, οι εναλλάξιμες ή άγνωστες συνθήκες θορύβου κάνουν την "πολυμορφική" εκπαίδευση δύσκολη να επιτευγχθεί . Εάν δε, μεταβλητές συνθήκες stress συνοδεύουν το θόρυβο (Βλέπε κεφ. 6 Παράγραφος 6.7.2) μπερδεύουν την εκπαίδευση ακόμα περισσότερο.

Αντί της εκπαίδευσης σε συνθήκες πολλαπλή ομιλίας, είναι πιο επιθυμητό να αναπτύξουμε αλγόριθμους αναγνώρισης που χρησιμοποιούν μόνο φυσιολογικό λόγο για την εκπαίδευση και να ερμηνεύουν ανεπιφύλακτα την διακύμανση της ομιλίας, εξ'αιτίας του φόρτου εργασίας και του stress του ομιλητή (Bill Sstanton, U.S. Air Foca Academy, and Leah Jamieson & George Allen, Purdue Univercity, "Αυθεντική αναγνώριση ηχηρού και Lombard λόγου μέσα σε περιβάλλον μαχητικού πιλοτηρίου" 1989, σελ. 675)

Η "πολυμορφική" εκπαίδευση , είναι τεχνική εμπλουτισμού ομιλίας (Βλέπε κεφ. 7 εισαγωγή, και ενότητα 7.1.2. και 7.4) που εφαρμόζεται κατά τη διάρκεια προεπεξεργασίας (Βλέπε κεφ. 2, ενότητα 2.2 και τις υποενότητές της). Αυτή κάνει την είσοδο να μοιάζει περισσότερο στα εξαρτώμενα από τον ομιλητή μοντέλα, για κανονική ομιλία αποθηκευμένη στο σύστημα . Ελεγχοι , χρησιμοποιώντας αυτή τη προσέγγιση , βελτίωσαν την ακρίβεια αναγνώρισης για ομιλία Lombard έως 42% . Μία άλλη προσέγγιση ,σχεδιασμένη για ανεξάρτητα του ομιλητή συστήματα , απομακρύνει ακουστικά χαρακτηριστικά περισσότερο επηρεασμένα απ' το αποτέλεσμα Lombard , για μοντέλα ανεξάρτητα από τον ομιλητή .Το STNN που σχεδιάστηκε στη Γαλλία ,αποτελεί μια άλλη προσέγγιση για χειρισμό του λόγου Lombard (Βλέπε ενότητα 7.4)

Μέχρι στιγμής αυτές οι τεχνικές έχουν δοκιμαστεί σε μικρές ομάδες ομιλητών κάτω από ελεγχόμενες συνθήκες . Εκτιμήσεις της αποτελεσματικότητάς τους στο χειρισμό της ομιλίας Lombard χρειάζονται υπολογίσιμα περισσότερες δοκιμές σε διαφορετικούς πληθυσμούς ομιλητών , ομιλούντων κάτω από ποικίλες συνθήκες θορύβου Ένα βήμα προς αυτή τη κατεύθυνσης είναι το σύστημα ΙΚΑΡΟΣ του Πανεπιστημίου DUKE. Ο ΙΚΑΡΟΣ εκμεταλλεύεται την ισχύ της τεχνολογίας επεξεργασίας ψηφιακού

σήματος της IBM (Βλέπε κεφ. 2 ενότητα 2.2. και τις υποενότητές της) προκειμένου να επιτύγχει εκτεταμένη προεπεξεργασία του σήματος εισόδου σε πραγματικό χρόνο .Η προεπεξεργασία αποτελείται από εξαναγκασμένο επαναληπτικό εμπλουτισμό ομιλίας και αλγόριθμους αντιστάθμισης του στρες, όπου οι ερευνητές John Hausen και Douglas Cairus του Πανεπιστημίου Duke βρήκαν να είναι αποτελεσματικοί στην βελτίωση της ακρίβειας για την ομιλία Lombard και τον στρεσαρισμένο λόγο. Η χρήση εμπορικής τεχνολογίας πραγματικού χρόνου ωθεί την έρευνα της ομιλίας Lombard και του στρεσαρισμένου λόγου πιο κοντά προς τις συνθήκες που θα αντιμετωπιστούν σε πραγματικές εφαρμογές.

Επιπρόσθετες τεχνικές πληροφορίες πάνω σε αυτό το θέμα μπορούν να βρεθούν στο Hansen (1993) Hansen & Apple baum (1990), Hansen & Bria (1990 και 1992) Hanson & Apple baum (1990) Jungua & Anglade (1990) και Stanton etc. (1989) Η εργασία του Hansen περιλαμβάνει τη σχεδίαση δυναμικών αλγορίθμων προς χειρισμό της ομιλίας Lombard Ο Hansen και άλλοι ερευνητές έχουν επίσης εξετάσει τρόπους χειρισμού ενός συνδυασμού ομιλίας Lombard, στρες και θορύβου υποβάθρου.

7.10 ΑΝΑΓΝΩΡΙΣΗ ΟΜΙΛΙΑΣ ΣΕ ΣΚΛΗΡΟ ΠΕΡΙΒΑΛΛΟΝ

Η αναγνώριση ομιλίας σε οχήματα ή για καταναλωτικά προϊόντα και υπηρεσίες ,είναι δύσκολη Αυτά τα αντίξοα περιβάλλοντα θα συζητηθούν στις ακόλουθες ενότητες.

ΑΝΑΓΝΩΡΙΣΗ ΟΜΙΛΙΑΣ ΜΕΣΑ ΣΕ ΟΧΗΜΑΤΑ

Οι κατασκευαστές αυτοκινήτων και συστημάτων αναγνώρισης ομιλίας έχουν εργαστεί σε τηλεφωνικά συστήματα επιλογής (για αυτοκίνητα) και εφαρμογές ελέγχου περιβάλλοντος (έλεγχος μεσώ φωνής , μη κρίσιμων , λειτουργιών όπως άνοιγμα παραθύρων και ανάμα του καλοριφέρ) από το πρώτο μισό της δεκαετίας του 1980 , προς πρόβλεψη της νομοθεσίας (των states) που απαιτεί την λειτουργία τηλεφώνων και άλλων συστημάτων χωρίς την βοήθεια των χεριών . Τα συστήματα έλεγχου περιβάλλοντος , δεν έχουν βγει ακόμα από τα εργαστήρια .

Το πρώτο εμπορικό κυψελοειδές τηλεφωνικό σύστημα επιλογής (για αυτοκίνητα) κατασκευάστηκε το 1986 από την Voice Control Systems . Η τεχνολογική γνώση των καταναλωτών δεν ξεκίνησε να αναπτύσσεται πριν την δεκαετία του 1990, με την εκτεταμένη χρήση των τηλεφώνων στα αυτοκίνητα και την εμφάνιση των συστημάτων αναγνώρισης ομιλίας σ΄αλλα περιβάλλοντα.

7.10.1 ΟΧΗΜΑΤΑ

Το όχημα , είναι περιβάλλον υψηλής επιθυμίας για εφαρμογή αναγνώρισης ομιλίας ,αφού τα μάτια του οδηγου θα πρεπει να παραμενουν στον δρομο κατά το περισσοτερο δυνατο και στην ιδανικη περιπτωση και τα δυο χερια θα περπει να κρατανε το τιμονι . ‘Ομως , τα οχηματα επισης αντιπροσωπευουν , ένα από τα πιο αντιξοα , προς αναγνωριση ομιλιας,

περιβαλοντα.Αναλογως του αυτοκίνητου, όπως και του δρόμου , της κυκλοφορίας και των καιρικών συνθηκών , το επίπεδο θορύβου υποβάθρου μπορεί να είναι εξαιρετικά υψηλό και ενδέχεται να περιέχει όλη την γκάμα των τύπων θορύβου του περιβάλλοντος :

- Ο σχετικά υπάρχων θόρυβος της μηχανής μπορεί να συνδυαστεί, με τον διακεκομμένο θορυβο από τις μηχανές άλλων αυτοκίνητων και με τους ήχους της δραστηριότητας εντός του αυτοκίνητου.
- Το SNR του αέρα, της μηχανής, και του θορύβου από τα λάστιχα, ποικίλει ανάλογα με τις εξωτερικές συνθήκες όπως το άνοιγμα του παραθύρου και την ταχύτητα
- Το φάσμα ισχύος του θορύβου υποβάθρου, κυμαίνεται από τις πολύ χαμηλές συχνότητες (θόρυβος μηχανής) , ακολουθεί το φάσμα ομιλιας , έως τις υψηλές συχνότητες (αέρας και ρόδες).
- Η λειτουργία του ραδιόφωνου προσθέτει μουσική , ομιλία υποβάθρου και μια τουλάχιστον απέραντη ποικιλία ήχων , συμπεριλαμβανόμενου και του σφυρίγματος του λευκού θορύβου.
- Η διαμόρφωση του εσωτερικού χώρου του οχήματος ,κάνει το περιβάλλον επιρροπο προς υψηλά επίπεδα αντήχησης.

Η φωνή του ομιλητή , είναι επισης αρκετά μεταβλητή .Ευμετάβλητα επίπεδα θορύβου υποβάθρου , ερμηνεύουν ένα μέρος της διακύμανσης της ομιλίας σαν λόγο Lombard .Το στρες , τα ισχυρά συναισθήματα , η θέση του κεφαλιού, και η κούραση, είναι πρόσθετα στοιχεια.Το εύρος της ποικιλίας μπορεί να είναι ακραίο ,για εμπορικά φορτηγά ή αντιπροσώπους πωλήσεων που ίσως οδηγούν 8 ή περισσότερες ώρες μέσα σε μια μέρα.

Ο προσεχτικός σχεδιασμός των ανθρώπινων παραγόντων είναι ένα προκλητικό αλλά και κρίσιμο στοιχείο της ανάπτυξης της εφαρμογής για αυτό το περιβάλλον.Τα συστήματα επιλογής μέσω φωνής πάνω στα κινητά τηλέφωνα , έχουν επιτύγχει , όταν οι χρηστές κρατούν το τηλεφωνικό σετ χειρός .Τα συστήματα επιλογής μεσώ φωνής , χωρίς την βοήθεια χεριών και τα συστήματα ελέγχου περιβάλλοντος για έλεγχο μεσω φωνής , παραθύρων ραδιοφώνων και κλιματισμού – θέρμανσης , αντιπροσωπεύουν ακόμα μεγαλύτερες προκλήσεις.Τα κατευθυντικά μικρόφωνα που υπάρχουν μέσα στο ακουστικό του τηλεφώνου, οι μικροφωνικές διατάξεις και τα headsets έχουν δοκιμαστεί . Οι οδηγοί φορτηγών και αυτοκινήτων απορρίπτουν από κοινού τα headsets, αφήνοντας τα στηριζόμενα στο πρόσωπο μικρόφωνα σαν την μόνη βιώσιμη επιλογή .Μερικοί κατασκευαστές έχουν χρησιμοποιήσει ένα πιεσε-να-μιλησεις κομβία για να σηματοδοτήσουν την εισερχόμενη ομιλία(βλέπε ενότητα 7.7.1). Μερικά συστήματα χρησιμοποιούν σειρά μικροφώνων και πολύπλοκες τεχνικές φιλτραρισματος για να αντισταθμίσουν την ευμετάβλητη τοποθέτηση του κεφαλιού , και το θορυβο του περιβάλλοντος .Κανένα από αυτά τα συστήματα δεν ήταν ικανό να παρακάμψει την συγκέντρωση παραγόντων θορύβου, του ανθρώπινου παράγοντα, του παράγοντα ομιλητή ,και τα εμπόδια από την αποδοχή του χρηστή, που αντίκρισε η αναγνωριση ομιλιας , για εργασίες διαφορετικές της επιλογής μεσώ φωνής στο τηλέφωνο αυτοκινήτου.

7.10.2 ΚΑΤΑΝΑΛΩΤΙΚΑ ΠΡΟΪΟΝΤΑ ΚΑΙ ΥΠΗΡΕΣΙΕΣ

Τα χαρακτηριστικά που κάνουν την χωρις-χρηση-χεριων αναγνωριση ομιλιας ,δύσκολη στα οχηματα , επεκτείνεται σε πολλούς άλλους τύπους εφαρμογων.Αναμεσα σε αυτές είναι καταναλωτικά προϊόντα και υπηρεσίες .

Οι υπηρεσίες προς καταναλωτές στα κιόσκια, αντικρίζουν προκλήσεις που είναι συγκρίσιμες προς αυτές που διευθετήθηκαν από την AT&T στηνAlice's Interactive Wonderland(ενότητα 7.6.4).Μερικές εφαρμογές μπορούν να χρησιμοποιήσουν μια γραμμική διάταξη μικρόφωνων .Τηλεφωνικά ακουστικά και της στενής δέσμης μικρόφωνα , αρμόζουν περισσοτερο σε άλλες << εν πορεία >> υπηρεσίες , όπως οι αυτόματες μηχανές ταμειου.Παρομοιως, μια υπηρεσία << για περαστικούς οδηγούς>> σε ένα εστιατόριο (fast food) ή σε μια τράπεζα , χρειάζεται ένα κατευθυντικό μικρόφωνο στενής δέσμης , ικανής να λαμβάνει τον λόγο του καταναλωτή , χωρίς επιπλέον να εισάγει τον θορυβο της μηχανής και του σωλήνα απορρόφησης , Αυτή η τεχνολογία είναι επισης υπό σχεδιασμό στα << ευφυή ακουστικά συστήματα>> της AT&T

Όσοι αναπτύσσουν φορητά καταναλωτικά προϊόντα, πρεπει επισης να σχεδιάσουν ,λαμβάνοντας υπόψη την ποικιλομορφία του θορύβου υποβάθρου , την αντοχή στο χρόνο και την αντίσταση στην ζημιά , το μικρό μέγεθος και τα πολύ χαμηλά γενικά έξοδα .

Τα καταναλωτικά προϊόντα και οι υπηρεσίες , υπόσχονται να είναι μια από τις πιο επικερδείς βιομηχανίες για αναγνωριση ομιλιας .Επιτυχία σ'αυτή την <<αρένα>> θα διασφαλίσει ευρεία αποδοχή της αναγνώρισης ομιλιας. Προκείμενου όμως να επιτύχουν , τα προβλήματα που παρουσιάζονται από το περιβάλλον ομιλιας θα πρεπει να επιλυθούν .

7.11 ΣΧΕΔΙΑΖΟΝΤΑΣ ΓΙΑ ΠΟΛΛΑΠΛΑ ΠΕΡΙΒΑΛΛΟΝΤΑ

Εάν μια εφαρμογή πρόκειται να χρησιμοποιηθεί σε περισσότερα από ένα και μόνο περιβάλλοντα, ή εάν ο ομιλητής αναμένεται να είναι εν κινήσει ,τα προϊόντα θα πρεπει να δοκιμαστούν σε κάθε πιθανό περιβάλλον .

Tokens που συλλέγονται από ένα περιβάλλον δεν μπορεί να είναι αναμενόμενο να λειτουργήσουν καλά σε περιβάλλοντα διαφορετικών χαρακτηριστικών θορύβου (ή καναλιού). Τα υπάρχοντα μοντέλα θα πρεπει να δοκιμαστούν και ενημερωθούν με δεδομένα από νέα περιβάλλοντα. Αυτή η προσέγγιση είναι δεδομένη διαδικασία λειτουργίας για σχεδιαστές συστημάτων αναγνώρισης – ανεξάρτητων από τον ομιλητή. Από τους παραγωγούς, εκείνοι που αναπτύσσουν εφαρμογές για κινητά τηλέφωνα , π,χ , μαζεύουν ξεχωριστά σετ από tokens για καθένα από αυτά τα περιβάλλοντα .

Η << ευκίνητη προσαρμογή >> (βλέπε κεφ. 5 , ενότητα 5.4) μπορεί να βοηθήσει στην μεταφορά μιας εφαρμογής από το ένα περιβάλλον στο άλλο , αλλά δεν θα έπρεπε να χρησιμοποιηθεί σαν τακτική εάν τα χαρακτηριστικά του νέου περιβάλλοντος είναι γνωστά. Μοντελοποιώντας το νέο περιβάλλον πριν η ανάπτυξη παράγει μεγαλύτερη ακρίβεια πιο γρήγορα .

Τα καταναλωτικά προϊόντα , αναπαριστούν μια ειδική περιπτωση των συστημάτων πολλαπλού περιβάλλοντος . Είναι πιθανόν να χρησιμοποιηθούν σε μεγαλύτερου εύρους συνθήκες, με άγνωστα χαρακτηριστικά θορύβου . Η χρήση του λόγου σε τέτοια προϊόντα είναι νέα , και πολλά μέρη της πρόκλησης είναι στην διαδικασία του να ανακαλυφθούν.

ΚΕΦΑΛΑΙΟ 7 :ΠΡΟΣ ΤΗΝ ΚΑΤΕΥΘΥΝΣΗ ΤΗΣ ΑΝΑΛΥΣΗΣ ΤΗΣ ΑΝΘΕΚΤΙΚΗΣ ΟΜΙΛΙΑΣ

7.1.ΠΡΟΚΑΤΑΡΚΤΙΚΑ

Όταν ένα σήμα ομιλίας αλλοιώνεται από θόρυβο, ή παραμορφώνεται από κανάλια επικοινωνίας όπως τηλεφωνικές γραμμές ή μικρόφωνο, ή απλά δέχεται επιδράσεις από τα συναισθήματα ή το στρές μας, ένας αριθμός μεθόδων που μπορούν να εφαρμοστούν προκειμένου να αποφέρουν μία περισσότερο ανθεκτική αναπαράσταση ομιλίας, από ότι μια παραδοσιακή διαδικασία απόκτησης σήματος, ακολουθούμενη από ένα LPC ή μια ανάλυση τράπεζας φίλτρων.

Στο στάδιο της απόκτησης σήματος τα κατευθυντικά και ακύρωσης - θορύβου μικρόφωνα μπορούν να χρησιμοποιηθούν για να ελαττώσουν την επίδραση του θορύβου από το υπόβαθρο. Σειρές (διατάξεις) μικροφώνων και τεχνικές ακύρωσης θορύβου, μπορούν επίσης να συνεισφέρουν στην αύξηση του SNR. Μετά την απόκτηση της ομιλίας , τεχνικές εμπλουτισμού ομιλίας, μέσω φιλτραρίσματος και μείωσης θορύβου, μπορούν να βελτιώσουν τη ποιότητα του σήματος της ομιλίας. Τεχνικές απόκτησης ανθεντικού σήματος και αλγόριθμοι ακύρωσης θορύβου, συζητούνται στην ενότητα 7.2, ενώ οι υπόλοιπες ενότητες του κεφαλαίου εστιάζονται στην ανάλυση της ανθεκτικής ομιλίας. Το παραμετρικό φιλτράρισμα, ο εμπλουτισμός ομιλίας, και μέθοδοι μείωσης θορύβου καλύπτονται στα κεφάλαια 8 και 9.

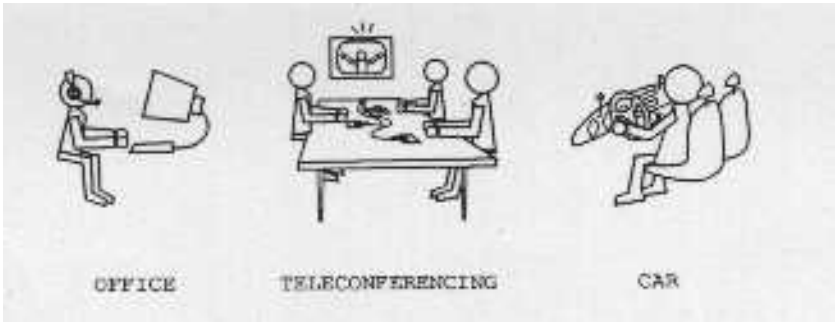
Όταν το σήμα έχει αποκτηθεί και προεπεξεργαστεί οι αλγόριθμοι ανάλυσης ομιλίας μετατρέπουν το ωηφιοποιημένο σήμα ομιλίας στις φασματικές συνιστώσες του. Προκειμένου να διεκπεραιώσουμε το πρόβλημα της ανάλυσης της ανθεκτικής ομιλίας έχουν αναπτυχθεί ακουστικά μοντέλα που αναπαράγουν υπολογιστικώς τα φυσιολογικά όργανα ή ψυχοακουστικά χαρακτηριστικά .

Στην υποενότητα 7.3.1. θα παρουσιαστούν τρία διαφορετικά ακουστικά μοντέλα .. που έχουν αναδειχθεί να είναι πιο επιτυχή από τις παραδοσιακές μεθόδους ανάλυσης περιβάλλον αντίξοων συνθηκών. Το πρόβλημα της εξαγωγής της ομιλίας από το θόρυβο μπορεί επίσης να θεωρηθεί σαν ένα πρόβλημα παραμετρικού υπολογισμού. Κατά συνέπεια στην υποενότητα 7.3.2 δίνεται μία κριτική της χρήσης στατιστικής και μοντελοποίησης ARMA προκειμένου να βελτιωθεί η παραδοσιακή ανάλυση γραμμικής πρόβλεψης.

7.2. ΑΠΟΚΤΗΣΗ ΣΗΜΑΤΟΣ

Προκειμένου να μετατρέψουμε ένα σήμα από την αναλογική του αναπαράσταση στη ψηφιακή ένας αριθμός τμημάτων επεξεργασίας είναι αναγκαίος. Το σήμα χρειάζεται να ληφθεί από ένα ή περισσότερα μικρόφωνα

,να φιλτραριστεί, να κβαντιστεί , και να κωδικοποιηθεί σε μία ψηφιακή αναπαράσταση. Η εικόνα 7.1 δείχνει διαφορετικές καταστάσεις όπου ένα ή περισσότερα μικρόφωνα συνήθως χρησιμοποιούνται. Καθένα από αυτά τα βήματα μπορεί να παράγει θόρυβο και παραμόρφωση όπου ελαττώνουν την πληροφορία που μεταδίδεται στο μπροστά μέρος ενός αυτόματου συστήματος αναγνώρισης ομιλίας.



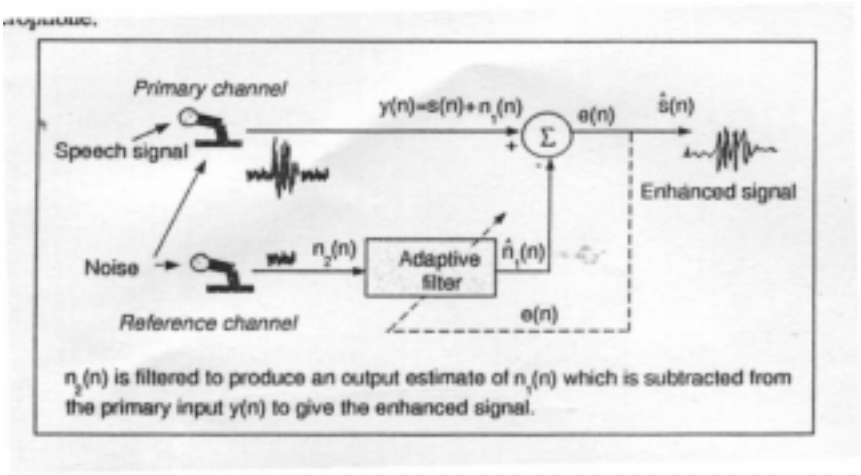
Σχήμα7.1 Παράδειγμα διαφορετικών συνθηκών απόκτησης σήματος

Προκειμένου να βελτιώσει το SNR, ένας αριθμός ερευνητών χρησιμοποίησε πολλά μικρόφωνα και ανέπτυξε αλγόριθμους επεξεργασίας διαφόρων πηγών εισόδου. Ο αλγόριθμος προσαρμοστικής ακύρωσης θορύβου χρειάζεται μία ανεξάρτητη επιπρόσθετη αναφορά για το θόρυβο (Βλέπε εικόνα 7.2). Δύο σήματα εισόδου που παρέχονται από δύο μικρόφωνα επεξεργάζονται , και οι μόνιμες ή μη μόνιμες παρεμβολές μπορούν να ακυρωθούν. Τα χαρακτηριστικά του θορύβου υπολογίζονται στο μικρόφωνο αναφοράς. Προκειμένου να πραγματοποιήσουμε το προσαρμοστικό φιλτράρισμα, η μέθοδος ελαχίστων τετραγώνων (LMS) γενικά χρησιμοποιείται. Περισσότερες λεπτομέρειες σχετικά με αυτή τη τεχνική μπορούν να βρεθούν στον Windrow et al.,1975,και μερικές πρόσφατες μελέτες και κριτικές(π.χ[Powell et al.,1987 Nakadai and Suganara,1990]) και κριτικές (e.g.[Van Comperadle 1992]. Προκειμένου να είναι αποτελεσματική η προσαρμοστική ακύρωση θορύβου απαιτεί να έχει ένα παρόμοιο σήμα θορύβου στο αλλοιωμένο σήμα και στο σήμα αναφοράς (καθόλου απώλειες). Σε μερικές εφαρμογές αυτό μπορεί να είναι ένας περιορισμός γιατί αν τα δύο μικρόφωνα είναι πολύ μακριά, είναι δύσκολο να υπολογίσεις το θόρυβο, και αν είναι πολύ κοντά ,το σήμα ομιλίας θα ληφθεί και από το μικρόφωνο αναφοράς .

Μία άλλη προσέγγιση στην ακύρωση θορύβου ονομάζεται ενεργή ακύρωση θορύβου. Γενικά αυτή η τεχνική έναν κύριο σένσορα τοποθετημένο στο σημείο όπου ο θόρυβος πρέπει να ακυρωθεί και ένα πιο δευτερεύων ή περισσότερο δευτερεύοντες σένσορες, τοποθετημένους σε άλλες θέσεις, για να παρέχουν πληροφορίες αξιοποιούμενες από τον αλγόριθμο για την ακύρωση του θορύβου στο κύριο σένσορα. Εντούτοις ενεργή ακύρωση θορύβου μπορεί επίσης να εφαρμοστεί σε συστήματα μονού σένσορα. Για παράδειγμα στον Openheim et al.,1992, η δευτερεύων πηγή θορύβου μοντελοποιείται από μια στοχαστική διαδικασία.Στον Zangri,1993, ένας αλγόριθμος ενεργούς ακύρωσης θορύβου δύο μικροφώνων ήταν επιτυχής στην εξασθένιση του θορύβου μέσα στη καμπίνα ενός ελικοφόρου

αεροσκάφους. Επίσης έχει αναφερθεί καλύτερη απόδοση απ' ότι για τη μέθοδο μονού σένσορα , στον Openheim et al.,1992 .

Άλλες δημοφιλείς προσεγγίσεις βασίστηκαν στους Frost[Frost III,1972;Farrell et al.,1992;Slyh and Moses,1993] και Griffiths-Jim [Griffiths and Jim ,1982; Van Compernolle et al.,1990]αλγόριθμοι χρησιμοποιούν προσαρμοστικές μεθόδους βασιζόμενες στην ελαχιστοποίηση της μέσης τετραγωνικής ενέργειας. Αυτοί οι αλγόριθμοι θεωρούν ότι το επιθυμητό σήμα είναι ανεξάρτητο από όλες τις πηγές παρεμβολής. Παρέχουν καλή βελτίωση όταν ο θόρυβος είναι προθετικός αλλά δεν έχουν καλή απόδοση σε περιβάλλοντα αντήχησης.. Μία πολλών μικροφώνων λύση βασιζόμενη στην επεξεργασία ετεροσυσχέτισης παρουσιάστηκε να αυξάνει την ακρίβεια αναγνώρισης. Αυτή η μέθοδος παρακινήθηκε από μία ανάλογη προς την ετεροσυσχέτιση επεξεργασία στο ανθρώπινο δι-ακουστικό σύστημα. Η μορφοποίηση δέσμης έχει μελετηθεί σε μεγάλο βαθμό σε υποβρύχια επεξεργασία sonar. Αυτή υπόσχεται προεπεξεργαστική προσέγγιση για ASR ομιλίας με θόρυβο και αντήχηση. Αυτή η προσέγγιση μπορεί να συνδυαστεί με άλλες συμπληρωματικές μεθόδους ώστε να επιτύχει ανθεκτική αναγνώριση ομιλίας (π.χ.[Stern et al., 1992]).



Σχήμα 7.2 Η αρχή της προσαρμοστικής ακύρωσης θορύβου

7.3 ΑΝΘΕΚΤΙΚΗ ΑΝΑΛΥΣΗ ΟΜΙΛΙΑΣ

7.3.1 ΠΑΝΩ ΣΤΗ ΧΡΗΣΗ ΑΚΟΥΣΤΙΚΩΝ ΜΟΝΤΕΛΩΝ, ΓΙΑ ΚΑΛΥΤΕΡΗ ΑΝΑΛΥΣΗ ΟΜΙΛΙΑΣ

7.3.1.1. ΠΡΟΚΑΤΑΡΚΤΙΚΑ

Ένα ακουστικό μοντέλο μπορεί αν αναπτυχθεί χρησιμοποιώντας δύο διαφορετικές προσεγγίσεις :

1) Μία φυσιολογική ή δομική προσέγγιση
2) Μία συναρτησιακή προσέγγιση .

Με τη πρώτη προσέγγιση επιχειρείται μία εξήγηση για όλα τα παρατηρούμενα φαινόμενα από το μοντέλο.. Στη συναρτησιακή προσέγγιση, σημαντικές απόψεις μοντελοποιούνται για να περιληφθούν σαν ένα μαύρο κουτί , σε ένα περισσότερο γενικό σύστημα. Στα συστήματα που παρουσιάζονται σε αυτή την ενότητα η δεύτερη προσέγγιση επιλέχτηκε επειδή ο σκοπός ήταν να συμπεριληφθεί το ακουστικό μοντέλο σαν ένα μπροστά μέρος ένα σύστημα αναγνώρισης ομιλίας. Οι ιδιότητες των αποκρίσεων της νευρικής ίνας ή οι πολύ καλά γνωστές ψυχοακουστικές έννοιες, όπου παρουσιάζονταν σπουδαίες, επιλέχθησαν και συμπεριλήφθηκαν σε ένα υπολογιστικό μοντέλο , υπεύθυνο για την ανάλυση ομιλίας. Οι γνώσεις μας πάνω στην φυσιολογία και στην ψυχοακουστική δε μας επιτρέπει άμεσα όλους τους μηχανισμούς της ακουστικής περιφέρειας. Συνεπώς μερικά βασικά φαινόμενα μοντελοποιήθηκαν γενικά και η αυτόματη αναγνώριση ομιλίας βελτιστοποιείται σαν ένα σύνολο.

Οι ακόλουθες ενότητες παρουσιάζουν τρία ακουστικά μοντέλα :

1) Την αντιληπτικά βασισμένη ,γραμμικής πρόβλεψης ,ανάλυση (PLP) βασισμένη σε ψυχοακουστικές καλά επαληθευόμενες έννοιες
2) Το E.I.H. ιστόγραμμα : ένα φυσιολογικό μοντέλο ενσωματώνοντας μία προσημείωση του μηχανισμού του ακουστικού νεύρου και μία σύγχρονη μέτρηση. Επίσης παρουσιάζεται μία τροποποίηση αυτού του μοντέλου που ονομάζεται ανάλυση χρονικά σύγχρονης γραμμικής πρόβλεψης
3) Ένα υβριδικό ακουστικό μοντέλο όπου φυσιολογικές και ψυχοακουστικές διαπιστώσεις έχουν προσημειωθεί .

7.3.1.2. Η ΑΝΤΙΛΗΠΤΙΚΑ–ΒΑΣΙΣΜΕΝΗ ΓΡΑΜΜΙΚΗΣ ΠΡΟΒΛΕΨΗΣ ΜΕΘΟΔΟΣ ΑΝΑΛΎΣΗΣ (PLP)

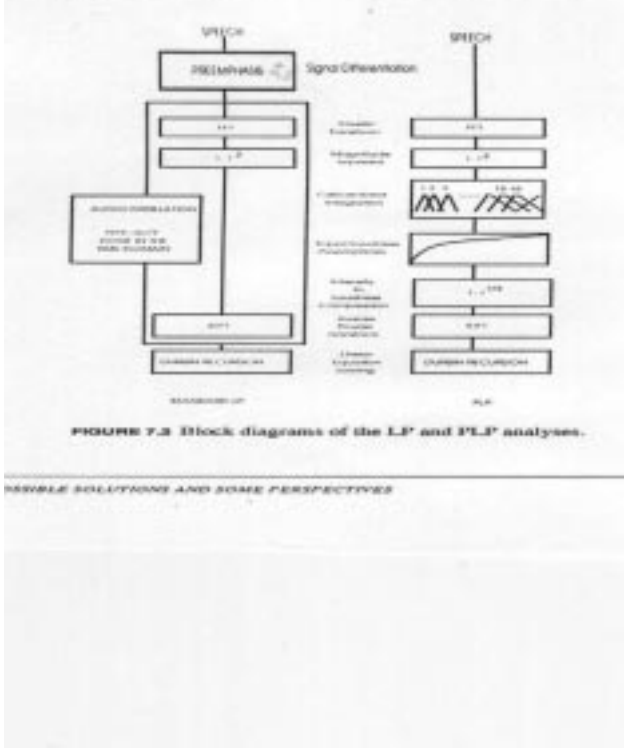
7.3.1.2.1. ΕΙΣΑΓΩΓΗ

Η PLP είναι μία σχετικά νέα προσέγγιση όπου μοντελοποιεί ,ένα δια των αισθήσεων αντιληπτό ακουστικό φάσμα , μέσω μίας όλων των πόλων συνάντησης, χρησιμοποιώντας την τεχνική αυτοσυσχέτισης LP. Όπως φαίνεται στην εικόνα 7.3. η PLP διαφέρει από την συνήθη LP στα εξής:

1) Στην ενσωμάτωση της κρίσιμης μπάντας του φάσματος ισχύος της ομιλίας (όπου παράγεται το ακουστικό φάσμα)
2) Προέμφαση ίσης έντασης ήχου
3) Τριτή ρίζα της έντασης - συμπίεσης του ήχου.

Αυτές οι τρείς επεξεργασίες προσημειώνουν μερικές ιδιότητες του ανθρώπινου οργανικού συστήματος. Συνοδεύονται από ένα ολοπολικό μοντέλο που παρέχει μία συμπαγή αναπαράσταση του ακουστικού φάσματος με όλους τους πόλους του. Επίσης αυτό το ολοπολικό μοντέλο έχει το πλεονέκτημα της ενυπάρχουσας έμφασης του στις φασματικές κορυφές .Η συμβατότητα των παραμέτρων της PLP ανάλυσης με τη συνήθη LP ανάλυση

αποδεικνύεται χρήσιμη για τις πρακτικές εφαρμογές. Στις ακόλουθες ενότητες ,συζητούνται τα διαφορετικά βήματα αυτής της μεθόδου.



Σχήμα 7.3 Block διαγράμματα της LP και PLP ανάλυσης

7.3.1.2.2. ΠΑΡΑΤΗΡΩΝΤΑΣ ΤΟ ΑΚΟΥΣΤΙΚΟ ΦΑΣΜΑ

Προκειμένου να παρατηρήσουμε το ακουστικό φάσμα δεκαεπτά έξοδοι από κρίσιμα ζωνοπερατά φίλτρα χρησιμοποιούνται. Οι κεντρικές τους ενότητες είναι ίσα τοποθετημένες στο πεδίο bark και ορίζουν από :

$$z = 6\log\left(\frac{f}{600} + \sqrt{\left(\frac{f}{600}\right)^2 + 1}\right)$$

όπου f είναι η συχνότητα σε Hertz και το z καλύπτει το εύρος $0 \le f \le 5\text{kHz}$ $(0 \le z \le 16.9\,\text{bark})$. Η κεντρική συχνότητα του κ-οστου ζωνοπερατού φίλτρου είναι ίση με z_K=0,9994κ. Το κρίσιμο ζωνοπερατό φίλτρο προσομοιώνεται αθροίζοντας το φάσμα ισχύος P(ω) όπου αποκτείται με τον FFT αλγόριθμο μέσα από ένα παράθυρο Hamming των είκοσι second ομιλίας. Σε αυτό το άθροισμα η υιοθετούμενη συνάρτηση βάρους είναι η :

$$C_{\text{K}}(\omega)= \begin{cases} 10^{1.0(z-z_K+0.5)} & \text{για Z} \le z_K - 0.5 \\ 1 & \text{για } z_K - 0.5 < \text{Z} < z_K + 0.5 \\ 10^{-2,5(z-z_K-0.5)} & \text{για Z} \ge z_K + 0.5 \end{cases}$$

Επειδή η κεντρική συχνότητα εξόδου του μηδέν δεν είναι καλά ορισμένη (αλλά ποτέ δε χρησιμοποιείται) η τιμή μηδέν (0) του φίλτρου εξόδου τίθεται ίση με τη τιμή ένα (1) του φίλτρου εξόδου. Τα φίλτρα είναι ασύμμετρα. Η ίσης ένταση καμπύλη προσεγγίζεται από την :

$$\text{E}(\omega)=1.151 \sqrt{\frac{(\omega^2+144*10^4)\omega^2}{(\omega^2+16*10^4)(\omega^2+961*10^4)}}$$

,όπου ω είναι η γωνιακή συχνότητα. Αυτή η καμπύλη προσεγγίζει την ακουστική απόκρουση πάνω στα μέσης συχνότητας επίπεδα έντασης (+10dB/oct από 0 έως 0.4kHz ,οριζόντια από 0.4κHz έως 1.2 κHz ,+6dB/oct από 1.2 εως 3.1kHz ,και οριζόντια από 3.1 μέχρι 5kHz) Η F_K δίνεται από τον τύπο : $F_{\kappa} = \text{E}(\omega_{\text{K}}) \int_0^{11} c_K(\omega) P(\omega) d\omega$

και είναι η ίσης έντασης ήχου, σταθμισμένη έξοδος , του K-οστού κρίσιμου ζωνοπερατού φίλτρου. Τελικά η τιμή της έντασης λαμβάνεται από το μετασχηματισμό :

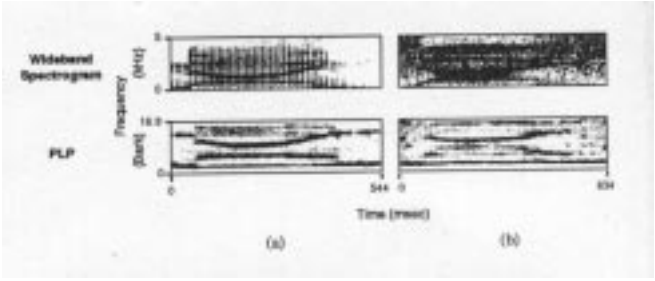
$$Q(\omega_K) = F_k^{\frac{1}{3}}$$

Η έξοδος από όλες αυτές τις επεξεργασίες ,είναι μία διακριτή αναπαράσταση του ακουστικού φάσματος (δεκαοκτώ (18) τιμές) όπου παρέχεται στην ολοπολική συνάρτηση μοντελοποίησης.

7.3.1.2.3. ΠΡΟΣΕΓΓΙΣΗ ΤΟΥ ΑΚΟΥΣΤΙΚΟΥ ΦΑΣΜΑΤΟΣ ΑΠΟ ΕΝΑ ΟΛΟΠΟΛΙΚΟ ΜΟΝΤΕΛΟ

Κάθε μη αρνητική συνάρτηση μπορεί να προσεγγιστεί από το φάσμα ενός όλων των πόλων μοντέλο χρησιμοποιώντας την μέθοδο αυτοσυσχέτισης LP. Αυτό επιτυγχάνεται μέσω αντιστοίχισης του ακουστικού φάσματος στο πεδίο αυτοσυσχέτισης χρησιμοποιώντας τον αντίστροφο διακριτικό μετασχηματισμό Fourier. Ακολούθως η επίλυση των γραμμικών Yule-Walker σχέσεων[Makhoul ,1975] παρέχει ένα σύνολο συντελεστών για το ολοπολικό φίλτρο. Ο ρόλος του όλων των πόλων μοντέλου είναι βασικός για την ελάττωση της διαστατικότητας του ακουστικού φάσματος και την αύξηση της ακουστικής ανάλυσης. Η ακουστική ανάλυση αυξάνεται επειδή οι θέσεις των κορυφών στο όλων των πόλων μοντέλο δεν περιορίζονται στις κρίσιμες κεντρικές συχνότητες. Παραπέρα ένα χαμηλής τάξης ολοπολικό μοντέλο επιτρέπει την απομάκρυνση των ασήμαντων φασματικών κορυφών. Η εικόνα 7.4 παρουσιάζει για τη περίπτωση της λέξης ”nine” το σύνηθες ευρείας-μπάντας φασματόγραμμα και το ψεύδο φασματόγραμμα που απορρέει από μία δωδέκατης τάξης PLP ανάλυση (στατικών συντελεστών) λαμβάνοντας

βάρη από την RPS. Όπως περιγράφεται στην επόμενη ενότητα ο συνδυαμός της PLP ανάλυσης και της RPS κατανομής αποδεικνύεται να δίνει καλή απόδοση στην ASR.

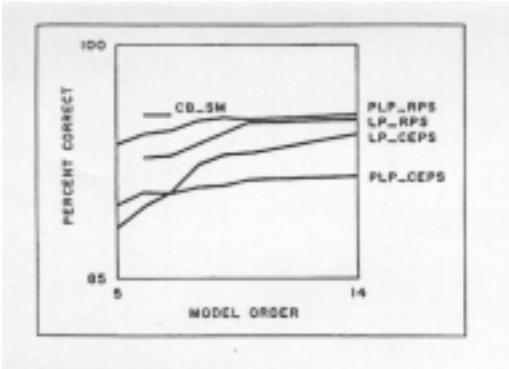


Σχήμα 7.4 Ευρείας μπάντας φασματαγράμματα και PLP ψευδο-φασματογράμματα ,για τη λέξη ”nine” στη περίπτωση καθαρού (α) και ενθόρυβου λόγου Lombard

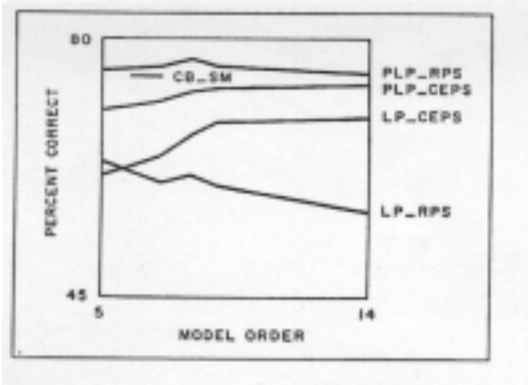
7.3.1.2.4 ΕΦΑΡΜΟΓΕΣ ΤΗΣ PLP ΑΝΑΛΥΣΗΣ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΟΜΙΛΙΑΣ

Η PLP αναπαράσταση εξαιτίας της χαμηλής της διαστατικότητας και της βολικής της μορφής έχει ενδιαφέρον για πρακτικά συστήματα αναγνώρισης ομιλίας. Το λαμβανόμενο φάσμα από τη PLP είναι περισσότερο εξομαλυνμένο από το φάσμα της LP τόσο σε διαστάσεις συχνότητας όσο και χρόνου. Αυτό το χαρακτηριστικό του PLP φάσματος είναι ένα επιθυμητό χαρακτηριστικό για το μπροστά μέρος ενός συστήματος αναγνώρισης ομιλίας. Ένας αριθμός από μελέτες αποδεικνύουν ότι τα χαμηλότερης τάξης μοντέλα της PLP ανάλυσης (πέντε (5) έως οκτώ (8)) δίνουν την καλύτερη ακρίβεια αναγνώρισης. Στις εικόνες 7.5. και 7.6 γίνεται σύγκριση για το παραπάνω σκοπό της LP και της PLP συνδυασμένης με του Ευκλείδιου και του RPS μέτρου απόστασης. Μπορούμε να παρατηρήσουμε ότι τα καλύτερα αποτελέσματα λαμβάνονται με το PLP-βάσης μπροστά-μέρος συνδυασμένο με το RPS μέτρο απόστασης.

Επίσης στη ταυτοποίηση του ομιλητή , τα χαρακτηριστικά της PLP βρέθηκαν να είναι σταθερά τα καλύτερα σε ένα αριθμό πειραμάτων.



Σχήμα 7.5 Σύγκριση μπροστά μέρους ASR σε αναγνώριση εξαρτώμενη του ομιλητή



Σχήμα 7.6 Σύγκριση μπροστά μέρους ASR σε αναγνώριση ανεξάρτητη του ομιλητή

Υπολογιστικώς , οι σύνηθες LP και η PLP είναι περίπου ίσης αποδοτικότητας .Ωστόσο ο μικρός αριθμός των συντελεστών της PLP ανάλυσης επιτρέπουν υπολογιστική και αποθηκευτική οικονομία.

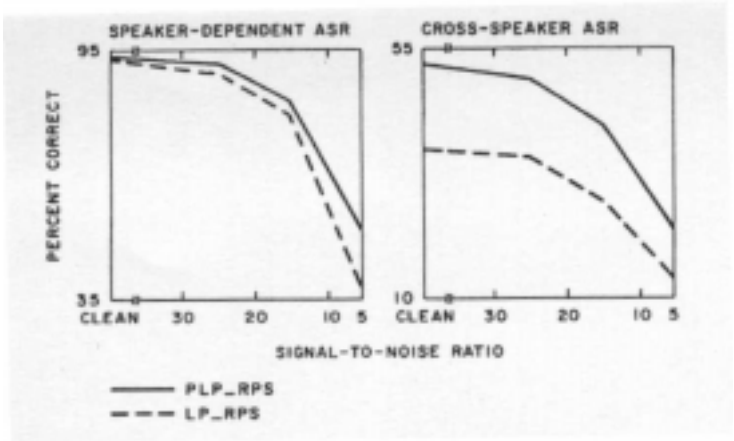
Ένα πιθανό μειονέκτημα της PLP ανάλυσης είναι ότι κάνει μία κρίσιμη ολοκλήρωση μετά το μετασχηματισμό Fourier για να λάβει το ακουστικό φάσμα . Αυτό οδηγεί σε μία σταθερή φασματική ανάλυση. Χρησιμοποιώντας μία πραγματική τράπεζα φίλτρων, είναι πιθανό να παρατηρηθεί μία καλή φασματική ανάλυση στις χαμηλές συχνότητες και μία καλή προσωρινή ανάλυση στις υψηλές συχνότητες (όπως στο ανθρώπινο αυτί)

7.3.1.2.5. ΕΦΑΡΜΟΓΗ ΤΟΥ PLP-ΒΑΣΙΣΜΕΝΟΥ ΜΠΡΟΣΤΑ-ΜΕΡΟΥΣ- ΣΥΣΤΗΜΑΤΟΣ ΑΝΑΓΝΩΡΙΣΗΣ ΟΜΙΛΙΑΣ , ΣΕ ΑΝΑΓΝΩΡΙΣΗ ΘΟΡΥΒΩΔΟΥΣ ΟΜΙΛΙΑΣ

Σε ένα αριθμό πειραμάτων τα PLP και LP- βασισμένα μπροστά -μέρη συστημάτων αναγνώρισης ,συγκρίθηκαν ,όταν προσθετικός θόρυβος προστέθηκε στο σήμα ομιλίας. Η εικόνα 7.7 δείχνει μία σύγκριση των LP_RPS και PLP_RPS , σε αναγνώριση εξαρτώμενη από τον ομιλητή και έτερου-ομιλητή αναγνώριση (στη τελευταία τα πρότυπα του ενός ομιλητή χρησιμοποιήθηκαν σαν αναφορές και τα πρότυπα του άλλου ομιλητή συγκρίθηκαν προς αυτές). Η PLP έδωσε τα καλύτερα αποτελέσματα ειδικά στην έτερου-ομιλητή αναγνώριση. Η σχετική ανθεκτικότητα της PLP ανάλυσης στο θόρυβο φαίνεται να είναι εξαιτίας του φιλτραρίσματος σε κρίσιμες φασματικές ζώνες. Όταν αποτιμήθηκε η απόδοση της PLP χρησιμοποιώντας ομιλία Lombard ,η καλύτερη απόδοση αποκτήθηκε για μοντέλο χαμηλής τάξης (πέντε (5)) ανάλυσης. Περισσότερα πειράματα μπορούν να βρεθούν στον Junqua 1989.

Πρόσφατα ,η επεξεργασία RASTA (Βλέπε κεφάλαιο 8.2.3) έχει συνδυαστεί με την PLP ανάλυση για να αντιμετωπίσει παραμορφώσεις οφειλόμενες στο κανάλι. Αυτός ο συνδυασμός βρέθηκε να δίνει μεγάλη βελτίωση συγκρινόμενος προς παραδοσιακές μεθόδους ανάλυσης ,όταν υπάρχουν διαφορές ανάμεσα στις συνθήκες εκπαίδευσης και τις πραγματικές, που επιφέρουν σε μια προσθετική παραμόρφωση στο πεδίο του λογαριθμικού

φάσματος. Εντούτοις αυτό το κέρδος στην απόδοση ήταν κατά κύριο λόγο οφειλόμενο στην επεξεργασία RASTA και όχι στην PLP ανάλυση.



Σχήμα 7.7 Σύγκριση της PLP_RPS και LP_RPS σε θόρυβο(Λευκό-Gaussian προσθετικό θόρυβο)

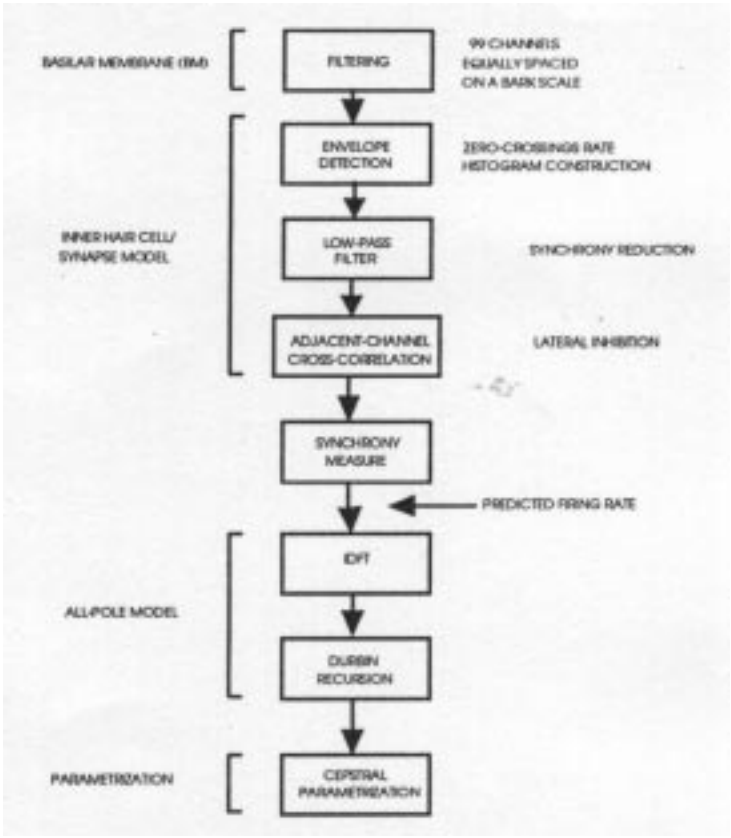
7.3.1.3.1 ΤΟ Ε.Ι.Η ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

(Το περιεχόμενο αυτής της ενότητας συγχωνεύτηκε στην αντίστοιχη ενότητα 7.3 του προηγούμενου κεφ: Speaking Environment)

7.3.1.3.2. ΤΟ ΣΥΧΡΟΝΟΥ ΧΡΟΝΟΥ ΓΡΑΜΜΙΚΗΣ ΠΡΟΒΛΕΨΗΣ ΑΚΟΥΣΤΙΚΟ ΜΟΝΤΕΛΟ

Μία άλλη ακουστική ανάλυση βασιζόμενη σε κατασκευή ιστογράμματος παρουσιάστηκε το 1989 από τον Junqua. Αυτή η ανάλυση ονομάζεται «σύχγρονου χρόνου γραμμικής πρόβλεψης» (SLP) ανάλυση, μοντελοποιεί μερικές μηχανικές επιδράσεις της βασιλικής μεμβράνης (BM), και αντιστοιχεί μηχανικές δονήσεις σε μία νευρωνική αναπαράσταση. Τα φαινόμενα που σχετίζονται με το πέρασμα το αέρα διαμέσω του ακουστικού καναλιού του μέσου αυτιού δεν έχουν θεωρηθεί καθόλου. Επίσης ο αυτόματος έλεγχος ενίσχυσης, ο οποίος συχνά θεωρείται να λαμβάνει χώρα στο μέσο αυτί, επίσης δεν παρουσιάζεται.

Κοιτάζοντας το σχήμα 7.9 διαπιστώνουμε ότι ο τομέας φιλτραρίσματος αποτελείται από ενενήντα εννέα φίλτρα ,ίσα τοποθετημένα , στην bark κλίμακα (από μηδέν 0 μέχρι πέντε 5 kHz.) Αυτά τα φίλτρα είναι συμμετρικά ,με κλίσεις των10 dB/bark. Οι κλίσεις έχουν ρυθμιστεί μέσω βελτιστοποιήσης των αποτελεσμάτων αναγνώρισης , σε μερικά προκαταρκτικά πειράματα (ειδικότερα τα ασύμμετρα φίλτρα οδήγησαν σε χαμηλότερη ακρίβεια αναγνώρισης). Η κύρια λειτουργία αυτού του τομέα όπου προσομειώνει το φιλτράρισμα της βασιλικής μεμβράνης, είναι να διαχωρίζει πολύπλοκους συνδυασμούς ήχων σε περιοχές υψηλού SNR.



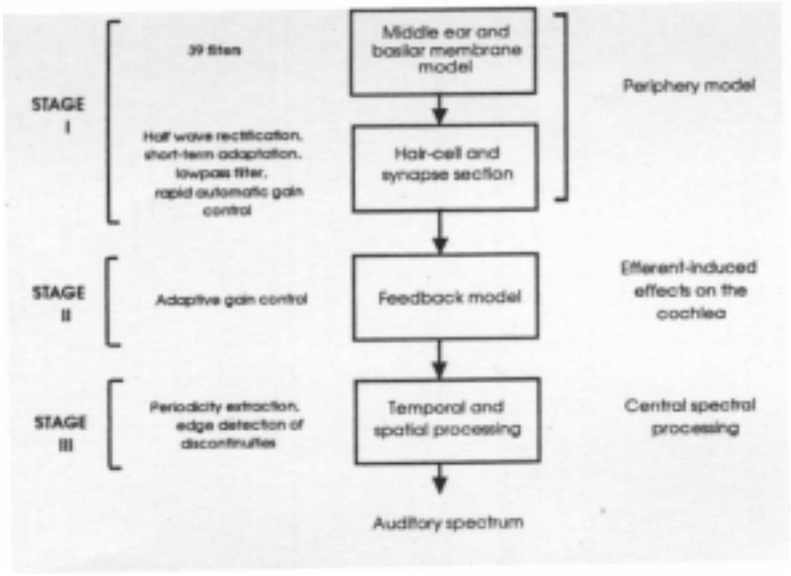
Σχήμα 7.9 Block διάγραμμα της SLP ανάλυσης (μετά τον Junqua,1989)

Ο δεύτερος τομέας του μοντέλου ,είναι ένα τριών βημάτων εσωτερικού κελιού/σύναψης μοντέλου .Καταρχήν, ένας ανιχνευτής περιβάλλουσας, προσδιορίζει, για κάθε κανάλι, την περιβάλλουσα του φάσματος. Κατά δεύτερον, ένα χαμηλοπερατό φίλτο με μία πολύ βαθμιαία ελάττωση (3 dB στα 2 kHz ,9 dB στα 4 kHz και 13dB στα 6 kHz) ελαττώνει τον συγχρονισμό των υψηλής συχνότητας διεγέρσεων. Ο λόγος παρουσίας αυτού του φίλτρου είναι ότι ο συγχρονισμός στις υψηλές συχνότητες δεν είναι φανερός.

Το χαμηλοπερατό φιλτράρισμα προκαλεί μία πτώση των ιστογραμμάτων στις υψηλές συχνότητες ,οδηγώντας σε ένα πιο ομαλό φάσμα. Τρίτον , ένας γειτονικού καναλιού μηχανισμός ετεροσυσχέτισης , εμπλουτίζει τις κορυφές (ειδικότερα παρουσία θορύβου) και προσωμειώνει μερικές από τις ιδιότητες του μηχανισμού lateral inhibition (που εξηγήσαμε αναλυτικά σε προηγούμενο κεφάλαιο). Το αποτέλεσμα των γειτονικών-καναλιών μηχανισμού ετεροσυσχέτισης ,είναι να επεξεργάζεται την χρονική λειτουργία του μοντέλου, κατά τέτοιο τρόπο ώστε να αποφέρεται μία αναπαράσταση της φασματικής περιβάλλουσας του σήματος , η οποία είναι σχετικά σταθερή πάνω σε μία μεγάλη ποικιλία SNR [Deng et al.,1988].

Ο τρίτος τομέας υλοποιεί ένα μέτρο συγχρονισμού (μέσω άθροισης όλων των καναλιών)όπου εξακριβώνει τις περιοχές στην προσωμειούμενη διάταξη ίνας, όπου τις πυροδοτεί συγχρονισμένα. Το εύρος αυτής της περιοχής χρησιμοποιείται σαν ένας εκτιμητής της σχετικής φασματικής έντασης[Ghitza,1987].

Ο τελευταίος τομέας είναι ένα ολοπολικό μοντέλο ακολουθούμενο από μία cepstral παραμετροποίηση.



Σχήμα 7.10 Ενα σύνθετο ακουστικό μοντέλο που αναπαράγει τα αποτελέσματα της μεταφοράς ώσης στον κοχλία και στο κεντρικό ακουστικό σύστημα

7.3.1.4. ΕΝΑ ΑΚΟΥΣΤΙΚΟ ΜΟΝΤΕΛΟ ΜΕ ΕΛΕΓΧΟ ΑΝΑΔΡΑΣΗΣ ΚΑΙ ΚΕΝΤΡΙΚΗ ΑΚΟΥΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ

7.3.1.4.1. ΠΕΡΙΛΗΨΗ

Ενα ακουστικό μοντέλο που μόλις πρόσφατα εισήχθη, είναι αυτό του σχήματος 7.10 Το πρώτο στάδιο αυτού του ακουστικού μοντέλου είναι παρόμοιο με τα δύο πρώτα στάδια που μόλις μελετήσαμε. Ο λόγος της παρουσίας του ελέγχου μέσω ανάδρασης , είναι προκειμένου να αποκτήσουμε μία περισσότερο ανθεκτική αναπαράσταση και να ελαττώσουμε τις επιδράσεις του θορύβου .Οποτεδήποτε ανιχνεύεται θόρυβος, αυτό το μοντέλο αυξάνει το κατώφλι των καμπύλων συντονισμού των ακουστικών νευρικών ινών και κατόπιν ελαττώνει την κίνηση της βασιλικής μεμβράνης (στάθμη πυροδότησης). Το κατώφλι των ακουστικών νευρικών ινών ορίζεται σαν η ελάχιστη αξιόπιστη απειροστή αύξηση , στην μέσης τάξης εκφόρτιση πάνω από την αυθόρμητη κίνηση. Ο έλεγχος ανάδρασης μπορεί επίσης να οριστεί σαν ο πλευρικά ζευγαρομένος προσαρμοστικός έλεγχος ενίσχυσης ,όπου προσομειώνει τα αποτελέσματα της μεταφοράς ώθησης προς την μυική σύναψη στον κοχλία. Υπάρχουν σημαντικές αποδείξεις ότι ηλεκτρικές διεγέρσεις στην δίοδο μεταφορά ώθησης του ελαιοκοχλία μπορούν να επενεργήσουν στη μηχανική του κοχλία και να εμποδίσουν την θετικά-οδηγούμενη δραστηριότητα των ακουστικών νευρικών ινών. Στον Gao et al.,1992, το 1992 ,συλλαβές παρουσίας θορύβου παρουσιάζουν την μείωση θορύβου που προκλήθηκε από αυτόν τον μηχανισμό ελέγχου μέσω ανάδρασης.

7.3.2. ΑΝΘΕΚΤΙΚΗ ΕΚΤΙΜΗΣΗ ΦΑΣΜΑΤΟΣ ΚΑΙ ΜΟΝΤΕΛΑ ARMA

7.3.2.1 ΒΕΛΤΙΩΜΕΝΗ AR ΜΟΝΤΕΛΟΠΟΙΗΣΗ

Τα AR θεωρούν ότι η ομιλία μπορεί να μοντελοποιηθεί είτε με λευκό θόρυβο ,είτε με ένα τρένο (σειράς) παλμών που διεγείρουν ένα γραμμικό σύστημα με μία ολοπολκή συνάρτηση μεταφοράς. Εντούτοις, όταν το σήμα εισόδου ξεφεύγει από αυτές τις υποθέσεις (π.χ. όταν υπάρχουν μηδενικά στη συνάρτηση μεταφοράς της φωνητικής περιοχής όπως τους έρρινους ήχους) η μοντελοποίηση ARMA είναι μία άλλη δυνατή λύση για βελτίωση της φασματικής εκτίμησης. Τα μοντέλα ARMA χρειάζονται μερικές ακόμα υποθέσεις πάνω στο μοντέλο θορύβου ,όπως στο μοντέλο του προσθετικού φάσματος ισχύος. Τα ARMA ή μηδενικού πόλου μοντέλα[Steiglitz ,1977;Atal and Schroeder,1978;Cadzow,1982;Morikawa and Fujisaki,1982; Fujisaki and ljungqvist,1987;Wang et al., 1990], υπερτερούν των AR στον προσδιορισμό της φασματικής δομής .Είναι καλά γνωστό ότι παρέχουν ένα καλύτερο υπολογισμό των ένρινων ήχων και της ομιλίας που είναι αλλοιωμένη από το θόρυβο. Εάν λευκός θόρυβος επιπροστίθεται σε ένα σήμα που παράγεται από μία καθεαυτού AR επεξεργασία, έχει καταδειχθεί ότι φασματικά μηδενικά εισάγονται[Kay,1979]. Μία ARMA διαδικασία όπου μετατρέπει τη πηγή u_i στο σήμα ομιλίας s_i έχει ως εξής :

$$s_i = -\sum_{k=1}^{p} a_k s_{i-k} + \sum_{j=1}^{q} b_j u_{i-j}$$

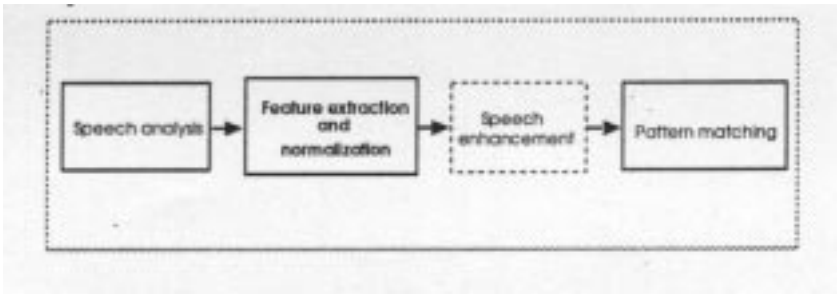
, όπου οι συντελεστές $a_k \, kai \, b_j$ αναφέρονται σαν AR και κινούμενου μέσου (MA) παράμετροι, αντιστοίχως. Για q=0 το μοντέλο ARMA γίνεται ένα AR. Η κύρια δυσκολία στον υπολογισμό ενός ARMA πηγάζει από τη μη γραμμικότητα του προβλήματος (τα u_{i-j} και b_j είναι άγνωστα). Αυτό οδηγεί σε μία αυξημένη πολυπλοκότητα συγκρινόμενη προς την AR μοντελοποίηση . Ενας αριθμός λύσεων έχει προταθεί προκειμένου να επιλυθεί ο υπολογισμός των παραμέτρων του ARMA.

Η ανθεκτική στατιστική εκτίμηση και η ARMA μοντελοποίηση αποτελούν υποσχόμενες τεχνικές για πιο ακριβείς αναλύσεις ομιλίας αλλοιώμενες από θόρυβο ή για μερικές κατηγορίες ήχων που μοντελοποιούνται φτωχά με την LP ανάλυση. Εντούτοις όπως και με τα ακουστικά μοντέλα, αυτές οι τεχνικές είναι ακόμα υπολογιστικώς «ακριβές».Πέραν τούτου, η μη γραμμικές τεχνικές συχνά έχουν προβλήματα σταθεροποίησης.

ΚΕΦΑΛΑΙΟ 8 :Η ΧΡΗΣΗ ΜΙΑΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ ΑΝΘΕΚΤΙΚΗΣ ΟΜΙΛΙΑΣ

8.1 ΕΙΣΑΓΩΓΗ

Γενικά στην αναλυτική επεξεργασία των χαρακτηριστικών καμία υπόθεση δε γίνεται για τα χαρακτηριστικά του θορύβου .Αυτό μπορεί να θεωρηθεί σαν ένα πλεονέκτημα ή και σαν ένα μειονέκτημα. Εάν αναπαραστήσουμε τα διαφορετικά βήματα που απαντώνται στην ASR όπως τα βλέπουμε στην εικόνα 8.1, οι τεχνικές που περιγράφονται σε αυτό το κεφάλαιο αντιστοιχούν στο κουτί με τίτλο : «εξαγωγή χαρακτηριστικών και κανονικοποίηση» . Εντούτοις , όσο οι αλγόριθμοι της κανονικοποίησης της εξαγωγής χαρακτηριστικών τείνουν να γίνουν όλο και περισσότερο πολύπλοκοι , δεν υπάρχει μία καθαρή διαχωριστική γραμμή ανάμεσα στο βήμα της εξαγωγής χαρακτηριστικών και κανονικοποίησης , και στα άλλα βήματα ανάγνωρισης.



Σχήμα 8.1 Το βήμα « εξαγωγή χαρακτηριστηκών και η κανονικοποίηση» στην ASR

8.2. ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

8.2.1. ΧΡΟΝΙΚΕΣ ΠΑΡΑΓΩΓΟΙ ΤΗΣ ΟΜΙΛΙΑΣ

8.2.1.1. ΟΙ ΜΕΤΑΒΑΣΕΙΣ ΣΤΗΝ ΟΜΙΛΙΑΣ ΚΑΙ Η ΑΝΤΙΛΗΨΗ ΤΟΥ ΛΟΓΟΥ

Είναι αρκετά γνωστό ότι οι μεταβάσεις παίζουν ένα σημαντικό ρόλο στην αντίληψη της ομιλίας. Η σημασία της δυναμικής φασματικής πληροφορίας και τα φαινόμενα ενάρθρωσης που περιλαμβάνονται στη παραγωγή συμφώνων και φωνηέντων σε συνεχή ομιλία, συνοδεύονται από μία καλή βιβλιογραφία(π.χ[LehisteandPeterson,1961;Lindblom,1963;Ohman,1966;Stevens and House,1963; Stevsns et al.,1966]). Έτσι μία περίοδος περίπου των

δέκα msec , που συμπεριλαμβάνει τη μέγιστη φασματική μετάβαση , βρέθηκε να φέρει κρίσιμη φωνητική πληροφορία για ταυτοποίηση συλλαβής .Πέρα όμως των μετρήσεων σε διάρκεια ,αναφέρεται ότι οι φασματικές μεταβάσεις ανάμεσα σε σύμφωνα και φωνήεντα , περιέχουν τη περισσότερο σημαντική φωνητική πληροφορία για ταυτοποίηση συλλαβής. Το μερίδιο του φωνήεντος πάνω στο σήμα ομιλίας , που τοποθετείται μετά από ουσιαστικές παύσεις , βρέθηκε να είναι αμελητέο για την αντίληψη της συλλαβής. Αυτά τα πειράματα επιβεβαίωσαν τη σημασία των φασματικών δυναμικών χαρακτηριστικών και οδήγησαν στη σαφή μοντελοποίηση των φασματικών δυναμικών χαρακτηριστικών στα ASR συστήματα.

8.2.1.2. ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΗΣ ΔΥΝΑΜΙΚΗΣ ΤΗΣ ΟΜΙΛΙΑΣ

Δύο κύριες τεχνικές για την αναπαράσταση της δυναμικής της ομιλίας ερευνήθηκαν :

1) Χρονικής-διαφοράς συνάρτηση ,μιας μικρού χρόνου παραμετρικής αναπαράστασης ομιλίας
2) Παλλινδρονούμενοι συντελεστές της παραμετρικής αναπαράστασης ομιλίας

Διάφορες μελέτες δείχνουν ότι οι παλινδρονούμενοι συντελεστές είναι ομαλότεροι και λειτουργούν καλύτερα από ότι η απλή η διαφορά ανάμεσα στα δύο πλαίσια. Μια απόσταση που αναπαριστά στατικά ,και της πρώτης παραγώγου ,τα χαρακτηριστικά , μπορεί να οριστεί από την:

$$d(k,l)=\sum_{j=1}^{M}(W(p_{k,j}-p_{i,j}))^2+\sum_{j=1}^{R}((1-W)(r_{k,j}-r_{i,j}))^2$$

όπου κ είναι το κ-ιοστό πλαίσιο ελέγχου , l είναι το l-ιοστό πλαίσιο αναφοράς ,R είναι ο αριθμός των συντελεστών της πρώτης παραγώγου , W είναι ένας παράγοντας μεταξύ 0 και 1, $p_{i,j}$ είναι το j-οστός σταθμισμένος cepstral συντελεστής για το i-οστό πλαίσιο και $r_{i,j}$ είναι ο j-οστός συντελεστής της πρώτου παραγώγου για το i-οστό πλαίσιο. Αν το $r_{i,j}$ αναπαριστά ένα παλινδρονούμενο συντελεστή , αυτός μπορεί να υπολογιστεί ως ακολούθως :

$$r_{i,j}=\frac{\sum_{q=-Q}^{Q} q\,p_{i+q,j}}{\sum_{q=-Q}^{Q} q^2}.$$

όπου Q είναι μία σταθερά έτσι ώστε 2Q+1 είναι ο αριθμός των πλαισίων του παράθυρου που χρησιμοποιήθηκε για να υπολογιστούν οι συντελεστές

παλινδρόμησης . Ορθογώνια πολυώνυμα [Draper and Smith ,1981]μπορούν να χρησιμοποιηθούν για να υπολογιστούν τα υψηλότερης ταξης παλινδρόμενα χαρακτηριστικά. Όπως αποδεικνύεται στους Hanson and Applebaum,1993 , ο ρόλος των χρονικών παραγώγων είναι να δώσουν έμφαση στα ίχνη των formants (συχνοτήτων φωνοσυντονισμού της φωνητικής οδούς) και να ελαττώσουν τις συχνότητες χαμηλής διαμόρφωσης.

Η πρώτη χρονική παράγωγος των cepstral συντελεστών χρησιμοποιείται πολύ στην ASR (Βλέπε υποενότητα 8.2.1.3) . Στον υπολογισμό των χρονικών παραγώγων συνήθως, το χρονικό μήκος του παραθύρου έχει εμπειρικά τιμή μεταξύ σαράντα και εκατό msec [Furui,1990].

Οι μεγαλύτερη τάξης χρονικές παράγωγοι έχουν επίσης μελετηθεί και εφαρμοστεί για την εξακρίβωση του ομιλητή και για αναγνώριση ομιλίας .Οι περισσότερες από τις μελέτες , συγκεντρώνονται στη χρήση της δεύτερης παραγώγου(π.χFurui and Rosenberg,1980; Furui,1981;Furui,1986b; Hanson and Applebaum ,1990a; Hanson and Applebaum ,1990b;Ney,1990]).Ακόμα και αν υπάρχουν διάφοροι τρόποι να υπολογιστεί η δεύτερη χρονική παράγωγος, οι περισσότερες υπάρχουσες προσεγγίσεις υπολογίζουν τη δεύτερη παράγωγο σαν χρονική παράγωγο της πρώτης χρονικής παραγώγου.

8.2.1.3. ASR ΜΕ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΟΝΙΚΗΣ ΠΑΡΑΓΩΓΟΥ

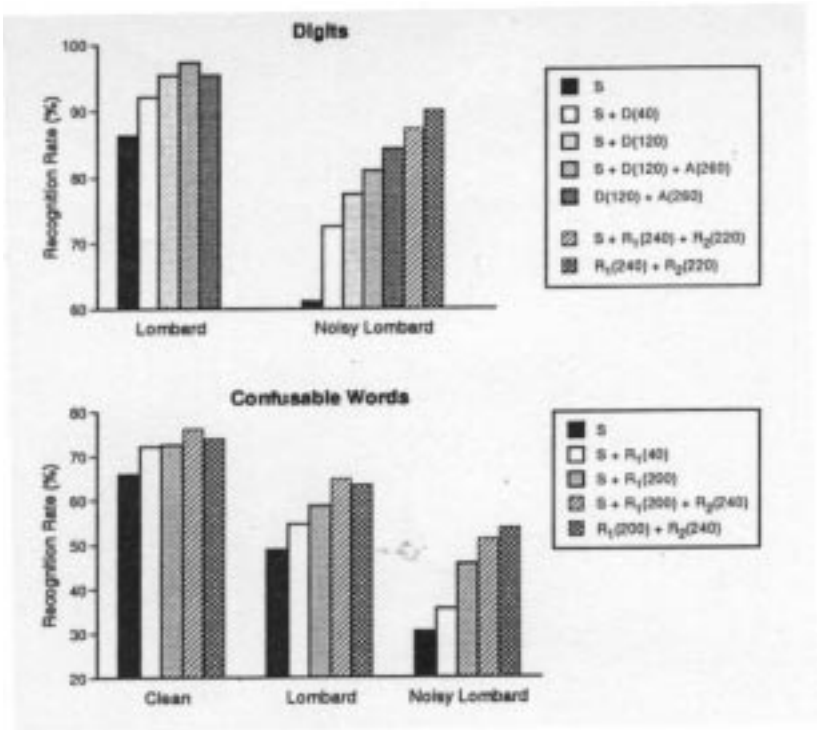
Τα συστήματα αυτόματης αναγνώρισης ομιλίας που χρησιμοποιούν στατικά και δυναμικά φασματικά χαρακτηριστικά , φάνηκε να είναι αποτελεσματικά για ανεξάρτητη του ομιλητή αναγνώριση και ταυτοποίηση(π.χ.[Furui,1981;SoongandRosenberg,1986;Furui,1986c;Junqua, 1987;Kitamura and Hayahara, 1988;Hanson and Applebaum ,1990]). Η πρώτη παράγωγος έχει βρεθεί να είναι αποτελεσματική για την αυτόματη αναγνώριση του καθαρού λόγου αλλά επίσης και για Lombart ομιλία και για ενθόρυβη Lombart ομιλία [Hanson and Applebaum,1990b].Πιο γενικά οποτεδήποτε υπάρχει κακό ταίριασμα ανάμεσα στα δεδομένα εκμάθησης και στα δεδομένα δοκιμής τα δυναμικά χαρακτηριστικά βοηθούν το σύστημα αναγνώρισης. Τα δυναμικά φασματικά χαρακτηριστικά επηρεάζονται λιγότερο από το θόρυβο από ότι τα στατικά φασματικά χαρακτηριστικά ομιλία [Hanson and Applebaum,1990b;Openshaw et al.,1992] Στην εξαρτώμενη αλλά και στην ανεξάρτητη απο τον ομιλητή αναγνώριση της ενθορύβου ομιλίας, τα δυναμικά και μέσου όρου φασματικά χαρακτηριστικά ,που υπολογίζονται από ένα δυδιάστατο mel cepstrum βρέθηκε να είναι ανώτερα του μονοδιάστατου mel cepstra, που αναπαριστά το στατικό χαρακτηριστικό[Kitamura and Hayahara,1988;Kitamura et al.,1990]. Άλλα πειράματα έδειξαν ότι το δυναμικό cepstrum είναι ανθεκτικό έναντι του προσθετικού λευκού και ρόζ θορύβου, ακόμα και αν ο θόρυβος επιδρά στη διαμόρφωση πλάτους. Σε αυτά τα πειράματα ο συνδυασμός δυναμικού cepstrum και του συνηθισμένου πρώτης παραγώγου cepstrum , έδωσαν καλύτερη απόδοση από ότι το συνηθισμένο cepstrum και το πρώτης παραγώγου cepstrum ,συνδιασμένα[Aikawa and Saito,1994].

Η βελτίωση που αποφέρεται στην ακρίβεια αναγνώρισης μέσω της δεύτερης παραγώγου είναι λιγότερο φανερή.Ο Furui , 1986b, βρήκε ότι η δεύτερη παράγωγος δεν επέφερε καμία βελτίωση , ενώ στους Ney ,το 1990,και το 1991 ο Dubois , η χρήση της δεύτερης παραγώγου έφερε βελτιωμένη απόδοση .Για την αναγνώριση θορύβου, Lombard , και θορυβούδους Lombard ομιλίας ,οι Hanson and Applebaum ,1990b ανέφεραν ένα κέρδος στην ακρίβεια ομιλίας όταν χρησιμοποίησαν τη δεύτερη παράγωγο. Ακόμα παραπέρα επίσης βρήκαν ότι στη περίπτωση συνθηκών κακού ταιριάσματος ανάμεσα στα δεδομένα εκμάθησης και δοκιμής, η απομάκρυνση του στατικού χαρακτηριστικού και η διατήρηση της πρώτης και δεύτερης παραγώγου ήταν προνομιούχος. Αυτό επίσης επιβεβαιώθηκε με πειράματα τηλεφωνικής ομιλίας όπου ένα επιπρόσθετο κακό ταίριασμα εισήχθη τεχνητά.

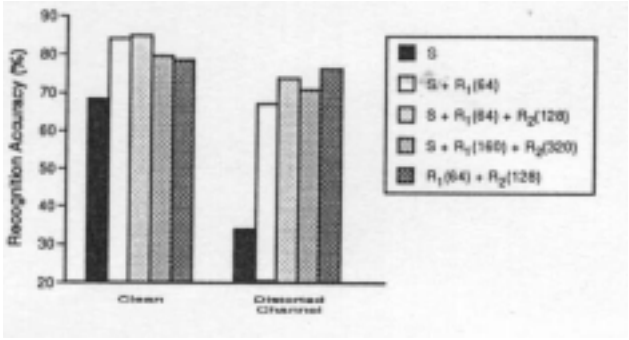
Εκτός της πολύ μεγάλης βελτίωσης που αποκτήθηκε με τη πρώτη παράγωγο, της ελαφράς βελτίωσης που δόθηκε από τη δεύτερη Παράγωγο, και της καλύτερης απόδοσης που αποκτήθηκε μέσω της υλοποίησης παλινδρόμησης, μπορούν να εξαχθούν και τα ακόλουθα συμπεράσματα:

1) Μεγάλα παράθυρα παραγώγου βοηθάνε στη περίπτωση μεμονωμένων λέξεων. Εντούτοις αυτά ελαττώνουν την ακριβεια αναγνώρισης στη περίπτωση συνεχούς ομιλίας. Επιπρόσθετα πειράματα και άλλα μεγέθη παραθύρου επιβεβαιώνουν αυτή τη παρατήρηση[Hanson et al., 1995];

2) Τα χαρακτηριστικά της παραγώγου βοηθούν περισσότερο στις περιπτώσεις με το μέγιστο κακό ταίριασμα. Για τις μεμονωμένες λέξεις και για τη συνεχή ομιλία ο συνδυασμός της πρώτης και δεύτερης παραγώγου χωρίς το στατικό χαρακτηριστικό (R1 + R2) αντισταθμίζουν καλά το κακό ταίριασμα ανάμεσα σε συνθήκες εκμάθησης και δοκιμής. Εντούτοις , το R1+R2 από μόνο του ελαττώνει την ακρίβεια αναγνώρισης για τη καθαρή ομιλία.



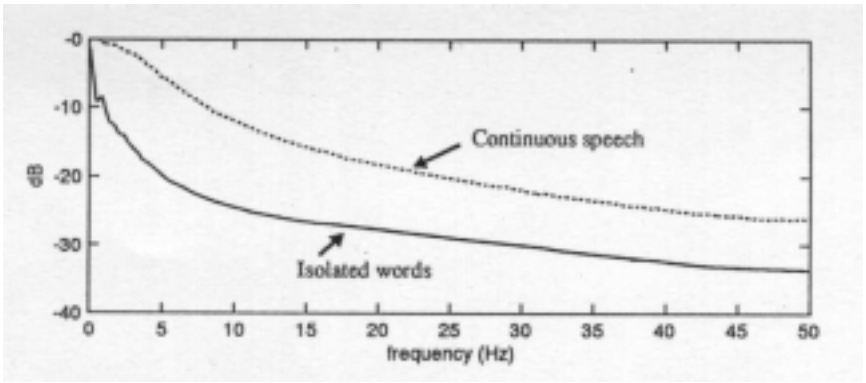
Σχήμα 8.2 Βλέπουμε τη βελτίωση στην αναγνώριση,για ποικίλες συνθήκες,χρησιμοποιώντας PLP-βάσης χρονικές παραγώγους



Σχήμα 8.3 Ακρίβεια σε αναγνώριση γράμματος ,λαμβανόμενη με διάφορα σύνολα ,MFCC-βάσης ,χαρακτηριστικών

Στη μελέτη του Nadeu and Juang,1994, ερευνήθηκε το αντάλλαγμα ανάμεσα στη διακύμανση του λάθους εκτίμησης και της ανάλυσης χρόνου. Όπως βλέπουμε στο σχήμα 8.4 βρέθηκε ότι το μεγάλου όρου φάσμα ισχύος της χρονικής αλληλουχίας των φασματικών παραμέτρων ,έχει ένα μεγαλύτερο εύρος ζώνης για συνεχή ομιλία από ότι για μεμονωμένες λέξεις. Οσο τα μικρά παράθυρα παλινδρόμησης έχουν λιγότερη ανάλυση συχνότητας (μεγαλύτερο εύρος ζώνης) από ότι τα μεγάλα παράθυρα παλινδρόμησης, αυτά είναι πιο κατάλληλα να παίρνουν τη πληροφορία σε συνεχή λόγο. Κατά συνέπεια τα μεγάλα παράθυρα είναι πιο κατάλληλα για μεμονωμένες λέξεις.

Στον Lee et al., 1990 , αναφέρθηκε ότι η προσθήκη της δεύτερης παραγώγου στη παραμετρική αναπαράσταση ομιλίας δεν είναι πάντοτε ευεργετική για τον εκάστοτε ομιλητή ακόμα και αν η ολική απόδοση βελτιώνεται. Μία από τις πιθανές αιτίες ίσως είναι ότι η δεύτερης τάξης cepstral ανάλυση παρέχει παρατηρήσεις με πολύ θόρυβο. Ένας πιθανός λόγος είναι ότι η χρήση χρονικών παραγώγων σαν επιπρόσθετων παραμέτρων στην παραμετρική αναπαράσταση ομιλίας ίσως να μην είναι η καταλληλότερη.



Σχήμα 8.4 Μέσο μεγάλου-όρου φάσμα ισχύος για μεμονομένες λέξεις και συνεχή ομιλία ,σαν συνάρτηση της συχνότητας.

8.2.2 AR ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΣΤΟ ΠΕΔΙΟ ΑΥΤΟΣΥΣΧΕΤΗΣΗΣ

Βασιζόμενοι στο γεγονός ότι στην ακολουθία αυτοσυσχέτισης επιδρά λιγότερο ο θόρυβος από ότι στο καθεαυτού σήμα νέες τεχνικές μοντελοποίησης έχουν αναπτυχθεί στο πεδίο αυτόσυσχέτισης. Η Short-time Modified Coherence (SMC) ,αναπαράσταση είναι μία ολοπολική μοντελοποίηση της ακολουθίας αυτοσυσχέτισης ακολουθούμενης από ένα φασματικό διαμορφωτή. Η μίας-μεριάς αυτοσυσχετιζόμενη γραμμικής πρόβλεψης κωδικοποίηση (OSALPC) είναι μία AR μοντελοποίηση του αιτιούδους μέρους της ακολουθίας αυτοσυσχέτισης .

Η μίας-πλευράς ακολουθία αυτοσυσχέτισης από μία ακουλουθία αυτοσυσχέτισης R (m) μπορεί να οριστεί[Herando and Nadeu,1991] ως εξής :

$$R^{+}(m) = \begin{cases} R(m) & , m>0 \\ R(0) & ,m=0 \\ 0 & ,m<0 \end{cases}$$

με $R^{+}(m) + R^{+}(m) = R(m)$, $-\infty \le m \le +\infty$.

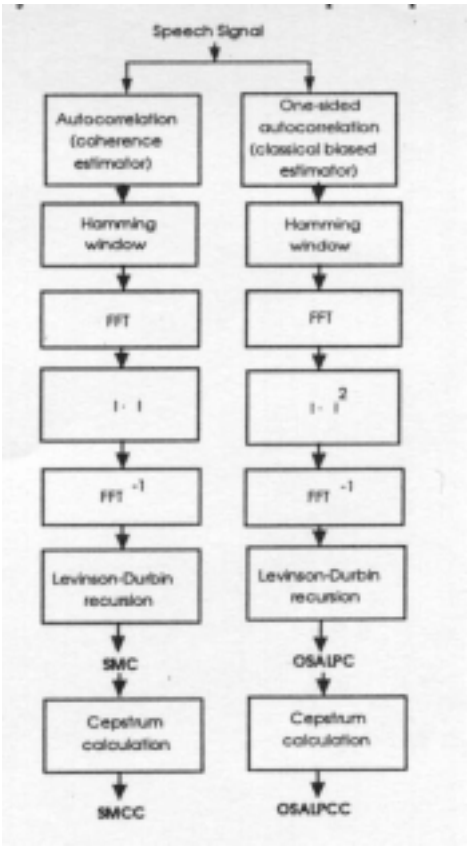
Η SMC αναπαράσταση [Mansour and Juang,1988b]ελαττώνει τη μεταβλητότητα ανάμεσα σε γειτονικά πλαίσια μέσω ελάττωσης της αλληλεπίδρασης ανάμεσα στο μήκος του πλαισίου ομιλίας και στη περίοδο υψηλού τόνου όπως συγκρίθηκαν στην LPC ανάλυση. Βέβαια αυτό έχει σαν αποτέλεσμα μία αύξηση στην υπολογιστική πολυπλοκότητα. Επίσης ο συνδυασμός της SMC παραμετροποίησης με υψηπερατό φιλτράρισμα του ενθόρυβου σήματος ομιλίας[Lecomte et al.,1989],αναφέρθηκε να επιφέρει μία καλύτερη βελτίωση στο θόρυβο από ότι ένα ανθεκτικό LP-βάσης μπροστά-μέρους συστήματος αναγνώρισης.

Αλγοριθμικά η διαφορά ανάμεσα στην OSALPC και στην SMC εντοπίζεται στη μέθοδο προσδιορισμού της πρώτης ακολουθίας αυτοσυσχέτισης και στον φασματικό διαμορφωτή που συμπεριλαμβάνεται στην SMC . Σε μία δεύτερη υλοποίηση της OSALPC χρησιμοποιήθηκε ο ίδιος εκτιμητής αυτοσυσχέτησης με εκείνον της SMC και αυτό επέφερε μία ελαφρά βελτίωση[Hernando και Nadeu,1994]. Όταν οι δύο μέθοδοι συγκρίθηκαν, αναφέρθηκε ότι για χαμηλό SMR , η OSALPC απέδωσε καλύτερα της LPC και της SMC[Hernando και Nadeu,1991]. Παραπέρα ,πιο πρόσφατες μελέτες πάνω στη OSALPC [Hernando και Nadeu,1994]έδειξαν ότι :

- Η χρήση ενός μη συμμετρικού lifter(φασματικού λειαντή)είναι επιθυμητή,
- Η μεγάλης τάξης πρόβλεψη και η cepstral αναπαράσταση βασιζόμενη στη OSALPC και συμπεριλαμβάνοντας δυναμικά χαρακτηριστικά δίνει καλή απόδοση
- Ένα μέτρο προέκτασης cepstral δε βελτίωσε περισσότερο την ακρίβεια αναγνώρισης ,παρουσία θορύβου από αυτοκίνητο.

Η φασματική μοντελοποίηση στο πεδίο αυτοσυσχέτισης έχει βρεθεί να βελτιώνει την αναγνώριση ενθόρυβου ομιλία χωρία καμία γνώση πάνω στα

χαρακτηριστικά του θορύβου. Η SMC και OSALPC είναι αρκετά απλά τεχνικές όπου παρέχουν βελτιωμένες αναπαραστάσεις ομιλίας σε σύγκριση με την LP ανάλυση.



8.5 Block διαγράμματα της SMC καιOSALPC

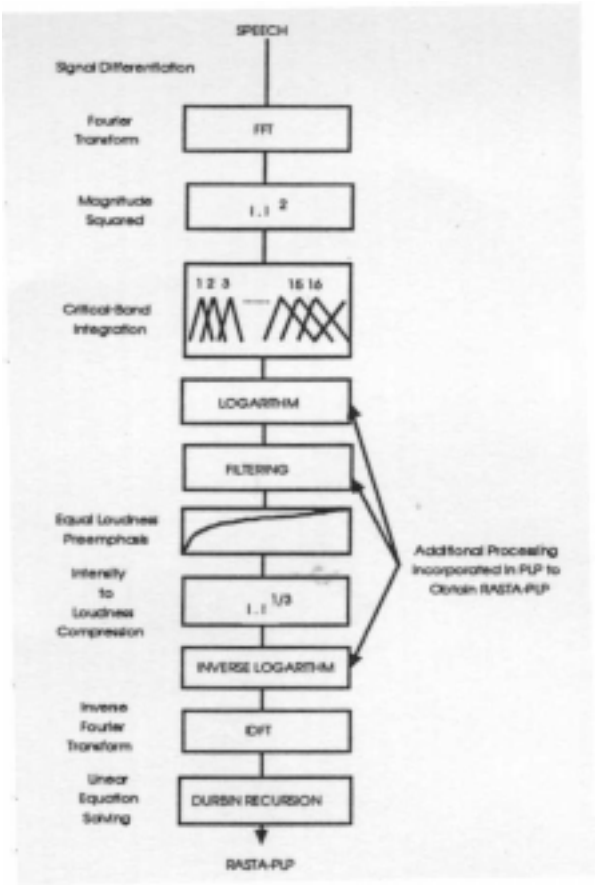
8.2.3 ΕΠΕΞΕΡΓΑΣΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

8.2.3.2. ΦΙΛΤΡΑΡΙΣΜΑ ΤΡΟΧΙΩΝ ΧΡΟΝΟΥ

Στην ανεξάρτητη του ομιλητή ASR θέλουμε να πάρουμε τη φωνητική πληροφορία από το σήμα ομιλίας και να απορρίψουμε όλες τις άλλες άσχετες πηγές πληροφορίας. Αυτή η φωνητική πληροφορία περιέχεται σε μία περιορισμένη ζώνη συχνοτήτων. Για να αντιμετωπίσουμε επιτυχώς τη μεταβλητότητα και τις διαφορετικές συνθήκες εξάσκησης και δοκιμών ,είναι βασικό να απομονώσουμε και να εξάγουμε τη κατάλληλη πληροφορία από το σήμα ομιλίας . Κατά συνέπεια τεχνικές φιλτραρίσματος χρησιμοποιούνται προκειμένου να εξάγουν τη κατάλληλη ανεξάρτητη από το χρήστη πληροφορία και να αποσιωπήσουν την επίδραση άλλων παραγόντων όπως την επίδραση του καναλιού. Τα φίλτρα μπορούν να υλοποιηθούν σε διάφορους τύπους και σε διάφορα πεδία π.χ. μερικές μέθοδοι αρχικά αναπτύχθηκαν για να βελτιώσουν την ανεξαρτησία ως προς τον ομιλητή ,να δώσουν την επίδραση του θορύβου υπόβαθρου , για να εξασθενήσουν την

επίδραση των διακυμάνσεων του ακουστικού καναλιού και για να αντιμετωπίσουν τη θορυβώδη ομιλία Lombard[Hanson and Applebaum,1993].

Το 1991 ο Hermansky et al., πρότεινε τη χρήση ενός ζωνοπερατού φίλτρου όπου συνδύαζε ένας πρώτης τάξης παλινδρονούμενο φίλτρο με ένα πρώτης τάξης πόλο. Αυτή η μέθοδος αναφέρεται σαν η συγγενής φασματική τεχνική RelAtive SpecTrAl (RASTA), και συνδιαζόμενη με τη PLP ανάλυση ,δίνει την RASTA-PLP ανάλυση. Το σχήμα 8.6 δείχνει τις τροποποιήσεις που πραγματοποιήθηκαν στη PLP ανάλυση για να αποκτηθεί η RASTA - PLP ανάλυση.

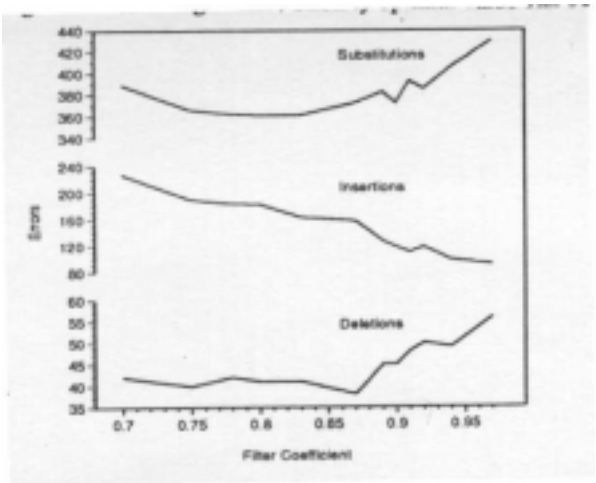


Σχήμα 8.6 Block διάγραμμα της RASTA-PLP ανάλυσης

Η αλλαγή του πραγματικού πόλου του φίλτρου αντιστοιχεί στην αλλαγή σταθεράς χρόνου. Η εικόνα 8.7 ανακεφαλαιώνει την επίδραση αυτής της παραμέτρου πάνω στο βαθμό αναγνώρισης για ένα συνεχούς συλλαβιστού ονόματος συστήματος αναγνώρισης, για τηλεφωνική ομιλία. Όταν η τιμή του πραγματικού πόλου μειώνεται η σταθερά χρόνου ελαττώνεται και οι πιο χαμηλής διαμόρφωσης συχνότητες φιλτράρονται. Για την εικόνα 8.7 που αφορά το σύστημα αναγνώρισης που μόλις αναφέραμε, μία σχετικά κατάλληλη τιμή είναι η τιμή 0,90.

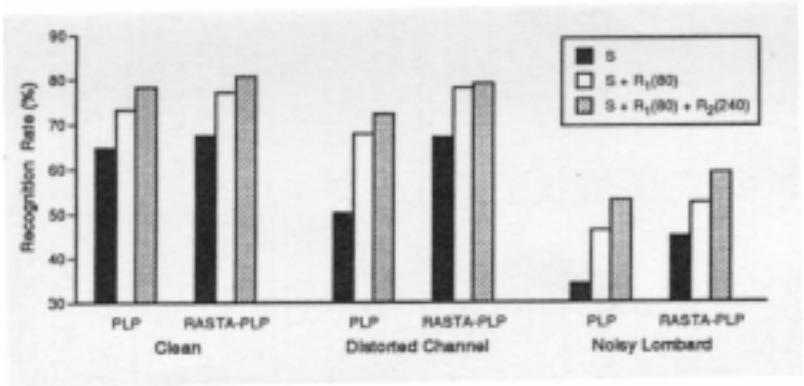
Οι τροχιές χρόνου (σταθεράς) φιλτραρίσματος βρέθηκε να εξουδετερώνουν σημαντικά την επίδραση του θορύβου συνέλιξης που προέρχεται από το κανάλι επικοινωνίας [Hermansky et al.,1991, Hermansky

et al,1992, Hermansky και Morgan,1994]. Ένα από τα μειονεκτήματα του RASTA φιλτραρίσματος (και πιο γενικά του υψυπερατού και του ζωνοπερατού φιλτραρίσματος) είναι ότι κατά τις παύσεις στο λόγο, το φιλτράρισμα περιλαμβάνει ένα βασικό χρόνο απόκρισης ,στον οποίο τα αργώς μεταβαλλόμενα στοιχεία δεν έχουν κατασταλεί εντελώς. Αυτό το φαινόμενο μπορεί να οδηγήσει σε ελάττωση της ακρίβειας αναγνώρισης όταν οι συνθήκες εκμάθησης και δοκιμής είναι παρόμοιες,. Προκειμένου να ξεπεραστεί αυτό το πρόβλημα εναλλακτικές μέθοδοι έχουν προταθεί .

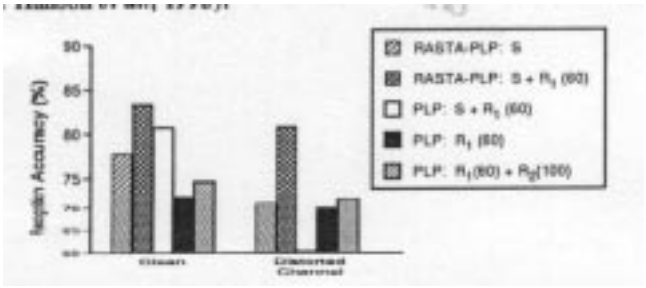


Σχήμα 8.7 Η ακρίβεια αναγνώρισης σαν συνάρτηση της σταθεράς χρόνου ,που δίνεται από τη τιμή του πραγματικού πόλου του RASTA φίλτρου

Η RASTA επεξεργασία έχει συγκριθεί με τις χρονικές παραγώγους του πεδίου cepstral[Hanson and Applebaum ,1993, Hermansky και Morgan,1994,Hanson et al ,1995,Junqua et al.,1995a]. Οι χρονικές παράγωγοι μπορούν επίσης να θεωρηθούν σαν γραμμικά φίλτρα χρονικών πλασματικών παραμέτρων. Οι εικόνες 8..8 και 8.9 παρουσιάζουν μία σύγκριση αυτών των τεχνικών. Τα πειράματα και τα σχόλια αυτών των εικόνων είναι παρόμοια με εκείνα που περιγράφηκαν για τις εικόνες 8.2 και 8.3.



Σχήμα 8.8 Σύγκριση του βαθμού αναγνώρισης για PLP και RASTA-PLP αναλύσεις ,συγχεόμενες μεμονομένες λέξεις και διαφορετικα σύνολα χαρακτηριστικών



Σχήμα 8.9 Σύγκριση της ακρίβειας αναγνώρισης για διάφορα RASTA-PLP και PLP-βασης σύνολα χαρακτηριστικών σε συνεχή αναγνώριση συλλαβιστού ονόματος

Μπορούμε να διαπιστώσουμε ότι, παρόμοια με τις χρονικές παραγώγους , η επεξεργασία RASTA είναι πιο χρήσιμη όταν υπάρχει ένα σημαντικό κακοταίριασμα ανάμεσα στις συνθήκες εκμάθησης και δοκιμής και όταν αυτό το κακό ταίριασμα μπορεί να μοντελοποιηθεί με γραμμικά φίλτρα . Τέτοια είδη παραμόρφωσης συμπεριλαμβάνουν τις διακυμάνσεις του μικροφώνου και πιο γενικά τις επιδράσεις από το κανάλι Η επεξεργασία RASTA είναι χρήσιμη τόσο για την ανάγνωριση μεμονωμένων λέξεων όσο και για την αναγνώριση συνεχούς λόγου. Από τα αποτελέσματα 8.8. και 8.9 ότι η επεξεργασία RASTA δεν είναι απλά το αποτέλεσμα της πρώτης παραγώγου ,αλλά κάτι παραπάνω.

Εξαρτώμενοι από την έκφραση αναγνώρισης η αποτελεσματικότητα της RASTA επεξεργασίας μπορεί να ποικίλει. Όπως συζητήθηκε και στην υποενότητα 8.2.1.3. τα μεγάλου χρόνου παράθυρα για τον υπολογισμό χρονικών παραγώγων[Applebaum and Hanson,1991], και για το RASTA φιλτράρισμα[Van Hamme et al.,19994, Hermansky και Morgan,1994] ,δεν είναι ικανοποιητικά ,για επεξεργασία συνεχούς λόγου βασισμένη σε ανεξάρτητες του κειμένου μονάδες λέξεων.

Προκειμένου να αντιμετωπισθεί ο προσθετικός θόρυβος, προτάθηκε η Γραμμικη-Λογαριθμικη(Γραμ-Λογ) RASTA[Hermansky και Morgan,1992, Hermansky et al.,1993].Παρόμοια με τον Hirsch et al., ο οποίος χρησιμοποίησε υψιπερατό φιλτράρισμα σε ένα φασματικό πεδίο ισχύος για να καταπνίξει τον προσθετικό θόρυβο, η αρχική RASTA επεξεργασία τροποποιήθηκε ώστε να πραγματοποιεί φιλτράρισμα στο πεδίο φασματικής ισχύος όταν ο θόρυβος είναι προσθετικός και στο λογαριθμικό φασματικό πεδίο κατά τη παρουσία θορύβου συνέλιξης. Προκειμένου να επιτευχθεί αυτό ο λογαριθμικός μετασχηματισμός στην επεξεργασία RASTA αντικαταστήθηκε από τον μετασχηματισμό $y=\ln(1+Jx)$ όπου J είναι μία θετική σταθερά εξαρτώμενη από το σήμα. Εάν το $J<<1$, ο μετασχηματισμός μπορεί να προσεγγηθεί από ένα γραμμικό μετασχηματισμό. Εάν το $J>>1$, .. μπορεί να προσεγγηθεί από ένα λογαριθμικό μετασχηματισμό. Στον Hermansky et al.,1993, η σταθερά J προσαρμοζόταν αυτόματα με μετρήσεις της ενέργειας του μέσου θορύβου. Η επεξεργασία Γραμ-Λογ RASTA αποδείχτηκε να είναι αποτελεσματική και για τον προσθετικό και για το θόρυβο συνέλιξης[Hermansky και Morgan,1994].

8.2.4 ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

8.2.4.1 ΠΡΟΣΑΡΜΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Η προσαρμογή χαρακτηριστικών ,είναι ένα άλλο τμήμα της τεχνικής που μπορεί να βοηθήσει το σύστημα αναγνώρισης να αντεπεξέλθει τα αποτελέσματα του θορύβου, την αντίδραση Lombard και πιο γενικά τις περιβαλλοντολογικές αλλαγές. Το 1991 στους Barbier and Chollet, ένα πολυστρωματικό perceptron χρησιμοποιήθηκε για να αποδείξει την αντιστοιχία ανάμεσα στις ενθόρυβες και χωρίς θόρυβο cepstral παραμέτρους. Το πολυστρωματικό perceptron εκγυμνάστηκε να παράγει την έξοδο το πιο κοντινό κατά προσέγγιση χωρίς θόρυβο cepstral διάνυσμα όταν παρουσιάζονταν επιτυχώς , αντίστοιχα cepstral διανύσματα με θόρυβο. Μία DTW χρησιμοποιήθηκε για να βρεί την αντιστοιχία ανάμεσα στα cepstral διανύσματα χωρίς θόρυβο και στα cepstral διανύσματα με θόρυβο. Στο στάδιο της αναγνώρισης ο cepstral μετασχηματισμός εφαρμοζόταν στις cepstral παραμέτρους των λέξεων που θα αναγνωριζόντουσαν. Η βελτίωση στη ακρίβεια αναγνώρισης πραγματοποιήθηκε ακόμα και για ομιλητή διαφορετικό από εκείνον για τον οποίο είχε γίνει η προσαρμογή. Μία άλλη υλοποίηση της ίδιας ιδέας επίσης αναφέρθηκε το 1991 στους Chollet και Mokbel(βλεπε επίσης 9.3.5.1). Σε αυτή την περίπτωση, ένας γραμμικός μετασχηματισμός χρησιμοποιήθηκε ανάμεσα στις αρθρώσεις με και χωρίς θόρυβο , και στο στάδιο αναγνώρισης ο μετασχηματισμός εφαρμόστηκε στις καθαρές φόρμες αναφοράς. Η υπόθεση που έγινε από αυτό το είδος της προσαρμοστικής μεθόδου , είναι ότι προσαρμοστικές μέθοδοι για τα περιβάλλοντα εκμάθησης και δοκιμής , είναι και για τα δύο διαθέσιμα. Το 1994 οι Lee και Wang , επέκτειναν τη γραμμική cepstral προσαρμογή στη προσαρμογή των παραμέτρων της πρώτης παραγώγου. Πειράματα σε μεμονωμένα ψηφία έδειξαν την αποτελεσματικότητα της μεθόδου για την αντιστάθμιση θορύβου.

8.2.4.2 Cepstral ΑΝΤΙΣΤΑΘΜΙΣΗ ΚΑΙ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

Καθόσον η cepstral παραμετροποίηση είναι πιο δημοφιλής αναπαράσταση ομιλίας χρησιμοποιούμενη στην αναγνώριση ομιλίας ένας αριθμός μεθόδων κανονικοποίησης έχει αναπτυχθεί στο cepstral πεδίο. Αυτές μπορούν να εφαρμοστούν απευθείας μετά την εξαγωγή των χαρακτηριστικών. Αυτές μία από τις πιο απλές αλλά αποτελεσματικές μεθόδους είναι η τεχνική (CMN-Cepstral Mean Normalization)cepstral Μέση Κανονικοποίηση . Μία άμεση υλοποίηση αυτής της τεχνικής υπολογίζει τον μέσο, κάθε cepstral τιμής ,πάνω στην άρθρωση, και μετά αφαιρεί αυτό τον μέσο από την τιμή σε κάθε πλαίσιο[Atal,1974]. Η υπόθεση που γίνεται από αυτή την τεχνική[Furui,1981,Schwartz et al.,1993,Mokbel et al.,1994], που δεν χρησιμοποιεί καμία γνώση για το περιβάλλον, είναι ότι ο μέσος όρος του cepstrum πάνω στις παύσεις ομιλίας αναπαριστά τη παραμόρφωση από το

κανάλι. Έτσι χρειάζονται πολλοί όροι για τον υπολογισμό της παραμόρφωσης από το κανάλι γεγονός ακατάλληλο για πραγματικού χρόνου εφαρμογές. Ωστόσο υπολογισμός μέσης τιμής με λίγους όρους έχει προταθεί(π.χ[Rosenberg et al.,1994]).Υλοποιήσεις λίγων όρων ,υποθέτουν ότι η παραμόρφωση από το κανάλι μεταβάλλεται αργά ,συγκρινόμενη με το σήμα ομιλίας.

Το να καταπνίξουμε τη παραμόρφωση του καναλιού μπορεί να θεωρηθεί σα μία φασματική κανονικοποίηση ή σα μία διαδικασία παραμετρικού φιλτραρίσματος. Κατά συνέπεια η CMN συχνά συγκρίνεται με ζωνοπερατό ή υψιπερατό φιλτράρισμα τροχιών χρόνου (τροχιές χρόνου βλέπε 8.2.3.2.) (π.χ[Schwartz et al.,1993,Mokbel et al.,1994,Junqua et al.,1995a]). Σε μερικές περιπτώσεις η CMN φαίνεται να είναι ανώτερη από ένα υψιπερατό ή ζωνοπερατό φιλτράρισμα(π.χ[Van Hamme et al.,1994]). Εντούτοις όταν χρειάζεται ταυτόχρονη αντιστάθμιση του προσθετικού θορύβου και των επιδράσεων του καναλιού η αποτελεσματικότητα της CMN είναι κατά τι περιορισμένη.

Normalization Methods	Meaning	Characteristics
CMN	Cepstral Mean Normalization	No knowledge of the testing environment; computationally very simple; not well suited for real-time; no joint compensation.
CDCN	Codebook-Dependent Cepstral Normalization	No knowledge of the testing environment; computationally expensive; adapt to the environment.
SDCN	SNR-Dependent Cepstral Normalization	Needs knowledge of the testing environment; computationally simple; no adaptation to the environment.
ISDCN	Interpolated SNR-Dependent Cepstral Normalization	No knowledge of the testing environment; little computation overhead as compared to SDCN; adapt to the environment.
FCDCN	Fixed Codebook-Dependent Cepstral Normalization	Needs knowledge of the testing environment; computational complexity is low; no adaptation to the environment.
BSDCN	Blind SNR-Dependent Cepstral Normalization	No knowledge of the testing environment; computationally simple; adapt to the environment.
MFCDCN	Multiple Fixed Codebook-Dependent Cepstral Normalization	No knowledge of the testing environment; computationally simple; adaptation to the environment by environment selection.
PDCN	Phone-Dependent Cepstral Normalization	Needs knowledge of the testing environment; computationally simple; no adaptation to the environment.
SPDCN	SNR-Dependent Phone-Dependent Cepstral Normalization	Needs knowledge of the testing environment; computationally simple; no adaptation to the environment.
IMFCDCN	Interpolated Multiple Fixed Codebook-Dependent Cepstral Normalization	No knowledge of the testing environment; slight computation overhead as compared to MFCDCN; environment selection possibly followed by interpolation.
IPDCN	Interpolated Phone-Dependent Cepstral Normalization	No knowledge of the testing environment; slight computation overhead as compared to IPDCN; environment selection possibly followed by interpolation.

Πίνακας 8.1 Σύγκριση της CMN καιδιάφορων cepstral κανονικοποιήσεων που προτείνονται από τη CMU

Το 1987 και το 1988 ο Chen, προκειμένου να αντισταθμίσει τη ποικιλομορφία των cepstral συντελεστών όταν ο λόγος παράγεται με διαφορετικά στυλ ομιλίας, ένα διορθωτικό μέσου όρου διάνυσμα προστέθηκε στις cepstral παραμέτρους. Για να αποκτηθούν οι αντισταθμητικές παράμετροι ,οι cepstral μέσοι και οι διακυμάνσεις υπολογίστηκαν για τα διάφορα στυλ ομιλίας. Ο Chen έδειξε ότι με αυτή την απλή αντισταθμητική μέθοδο ουσιαστική μείωση στη τάξη των λαθών μπορεί να αποκτηθεί. Αυτή η τεχνική φάνηκε να είναι χρήσιμη για την αντιμετώπιση της αντίδρασης Lombard (βλέπε υποενότητα 9.3.5.1)

Τέλος, γενικότερα, στο πίνακα 8.1 συνοψίζουμε έναν αριθμό μεθόδων cepstral κανονικοποίησης

8.2.5. ΕΚΤΙΜΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΣΕ ΘΟΡΥΒΟ

Προκειμένου να βελτιώσουν την απόδοση του συστήματος αναγνώρισης ομιλίας σε θόρυβο, πολλοί ερευνητές πρότειναν έναν στατιστικό εκτιμητή προκειμένου να λάβουν τα χαρακτηριστικά της ομιλίας.

Οι Porter και Boll,1984,καιVan Compernolle,1989, χρησιμοποίησαν έναν ελάχιστου μέσου τετραγωνικού σφάλματος (MMSE) αλγόριθμο που τον εφάρμοσαν αντίστοιχα στις ενέργειες εξόδου μιας τράπεζας φίλτρων και στους συντελεστές του διακριτού μετασχηματισμού DFT. Ο εκτιμητής MMSE χρησιμοποιεί στατιστική θορύβου και ομιλίας και την ενθόρυβη φασματική τιμή για να παρέχει την αναμενόμενη τιμή της εκτιμούμενης καταστατικής συνάρτησης κατανομής [Porter και Boll,1984]. Στη πραγματικότητα η ιδέα είναι να εφαρμοστεί μία φασματική ανάλυση (αποσύνθεση). Οι Erell και Weintraub,1990, και Erell και Weintraub,1993 όρισαν ένα καταλληλότερο κριτήριο όπου η εκτίμηση επιζητά να ελαχιστοποιήσει την παραμόρφωση όπως μετρήθηκε από την ASR έμμετρη απόσταση.

Ένας αλγόριθμος μεταγεννέστερης εκτίμησης (MAP) για εμπλουτισμό ομιλίας που οδήγει σε ένα ανθεκτικό σύνολο χαρακτηριστικών για αναγνώριση ομιλίας σε θόρυβο, προτάθηκε το 1998 από το Hansen ,και το1998 από τους Hansen και Clements Ο εκτιμητής MAP μεγιστοποιεί τη συνάντηση πυκνότητας πιθανότητας των αγνώστων παραμέτρων, δοθέντων των ενθόρυβο παρατηρήσεων. Αυτός χρησιμοποιήθηκε το 1978 από τον Lim, προκειμένου να υπολογίσει τις παραμέτρους ομιλίας με θόρυβο. Προκειμένου να βελτιωθεί η εκτίμηση των παραμέτρων ένας αριθμός φασματικών περιορισμός πάνω σε όλων των πόλων μοντέλο ομιλίας και στα χαρακτηριστικά της φωνητικής περιοχής εισήχθη μεταξύ των υπολογιστικών τμημάτων του MAP.[Hansen ,1988].

8.2.6 ΑΛΛΕΣ ΤΕΧΝΙΚΕΣ ΠΟΥ ΠΑΡΕΧΟΥΝ ΒΕΛΤΙΩΜΕΝΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Με τη χρήση μοντέλων κάλυψης ο Usagawa et al., το 1994 πρότεινε μία νέα μέθοδο εξαγωγής παραμέτρων ομιλίας για περιβάλλοντα με θόρυβο.

Αυτή η μέθοδος δεν εξαρτάται από το επίπεδο θορύβου και μπορεί να αντιμετωπίσει μη σταθερό θόρυβο. Οι παράμετροι του μοντέλου κάλυψης προέρχονται από ψυχοακουστικά πειράματα. Καθώς τα υψηλού-επιπέδου ομιλίας στοιχεία ,καλύπτουν τα χαμηλού-επιπέδου στοιχεία , αυτή η μέθοδος παράγει μία πιο ανθεκτική αναπαράσταση ομιλίας. Αυτό το μοντέλο κάλυψης εισάγεται ανάμεσα στον υπολογισμό του φάσματος ισχύος και στον υπολογισμό των συντελεστών της LPC. Όταν αξιολογείται σαν ένα μπρος μέρος ενός αυτόματου συστήματος αναγνώρισης μεμονωμένων λέξεων ,με ή χωρίς συσκευή ακύρωσης θορύβου, αυτή η μέθοδος φαίνεται να βελτιώνει την ακρίβεια αναγνώρισης[Usagawa et al., το 1994]

Ένα άλλο σύνολο από χαρακτήρες, προτάθηκε το 1994 από τον Assaleh et al.,βασιζόμενο στο μοντέλο διαμόρφωσης (MM) . Αυτό το μοντέλο αποσυνθέτει την ομιλία σε στοιχεία διαμορφωμένα κατά πλάτος και συχνότητα. Βρέθηκε ότι η MM-βάσης προσέγγιση είναι περισσότερο ανθεκτική στη γήρανση από ότι η cepstral-βάσης προσέγγιση. Πιο συγκεκριμένα η MM-βάσης μέθοδο έδωσε παρόμοια αποτελέσματα για δύο διαφορετικές βάσεις δεδομένων που καταγράφησαν με τέσσερις μήνες διαφορά, ενώ μία cepstrum-βάσης προσέγγιση επέφερε μία (15%) ελάττωση στην ακρίβεια αναγνώρισης.

8.3 ΘΟΡΥΒΟΥ - ΑΝΘΕΚΤΙΚΗ ΠΑΡΑΜΟΡΦΩΣΗ ΚΑΙ ΜΕΤΡΗΣΕΙΣ ΟΜΟΙΟΤΗΤΑΣ

8.3.1 ΦΑΣΜΑΤΙΚΗ ΛΕΙΑΝΣΗ (cepstral lifters)

Προκειμένου να σχεδιάσουμε το καλύτερο δυνατό σύστημα για ASR είναι σημαντικό να θεωρήσουμε την ανάλυση και το μέτρο απόστασης σε μία συνδυασμένη μέθοδο , από ότι να τα θεωρήσουμε σαν διαφορετικά θέματα. Βλέπουμε αυτή τη τάση σε θέματα πρόσφατων μελετών, ιδιαίτερα σε μελέτες πάνω στην αναγνώριση ομιλίας που χρησιμοποιούν cepstral παραμέτρους. Οι συντελεστές cepstral έχουν ευρέως μελετηθεί με ένα σταθμούμενο (με βάρη) cepstral μέτρο απόστασης , για μεγαλύτερη βελτίωση του αποτελέσματος της αναγνώρισης(π.χ[Yegnanarayana and Reddy ,1979,Paliwal,1982,Tohkura,1985,Hanson et al.,1985, Tohkura,1987,Junqua et al.,1993]).

Σταθμίζοντας (με βάρη) τους cepstral συντελεστές , επηρεάζουμε μοναδικά την τελική παραμετρική αναπαράσταση ομιλίας , που χρησιμοποιείται στο ταίριασμα προτύπου. Οι συντελεστές βαρύτητας μπορούν να προέρθουν από μία στατιστική προσέγγιση [Tohkura,1985, Tohkura,1987] ή από μία μελέτη της ευαισθησίας τους στα φασματικά χαρακτηριστικά. Ολα τα σταθμισμένα (με βάρη) σχήματα έχουν αυξανόμενα βάρη πάνω στους πρώτους cepstral συντελεστές.

Η χαμηλής τάξης όροι της cepstral αναπαράστασης περιγράφουν τα ομαλά χαρακτηριστικά του φάσματος ,που αντιστοιχούν κυρίως στην απόκριση των φωνητικών οργάνων, από ότι στην τέλεια φασματική δομή. Αυτή η τέλεια φασματική δομή , σχηματίζει τεχνουργήματα που επενεργούν

δυσμενώς στο αποτέλεσμα του ταιριάσματος φασματικών προτύπων, στο σταθμισμένο cepstral μέτρο απόστασης [Hermansky ,1987]. Τέτοια τεχνουργήματα πρέπει να ελαχιστοποιηθούν με τη βοήθεια κάποιων μέσων. Ενας απλός τρόπος να επιτευχθεί αυτό είναι να κοπούν οι άπειρες σειρές των σταθμισμένων cepstral συντελεστών , σε ένα μικρό νούμερο.

Διάφοροι ερευνητές πρότειναν διάφορους lifter(φασματικούς λειαντές) . Με την PLP ανάλυση το βαθμιαίο απώτατο άκρο του lifter βρέθηκε ότι δεν χρειάζεται[Junqua et al.,1993]. Επειδή το PLP φάσμα είναι σχετικά ομαλό ,η αποκοπή των cepstral συντελεστών για το PLP φάσμα , δεν επιδρά σημαντικά στο σχήμα του .Ακόμα βρέθηκε ότι το RPS μέτρο απόστασης είναι πολύ ευαίσθητο στα φασματικά μέγιστα και όχι αρκετά ευαίσθητο στις φασματικές κλίσεις. Κατά συνέπεια, προτάθηκε ένα νέο μέτρο απόστασης όπου σταθμίζει κάθε συντελεστή cepstral , μέσω του αύξοντα αριθμού του,υψωμένου σε μία δύναμη [Junqua et al.,1993]. Αυτός ο γενικός εκθετικός φασματικός λειαντής, ορίζεται από τον τύπο $E_n = n^s$, $s \geq 0$, και συνιστά μία συνεχή σειρά ανάμεσα στον cepstral του Ευκλείδη (S=0) και τα RPS (S=1) μέτρα απόστασης.

Ὅπως είδαμε στην ενότητα 8.2.1.2 , ο δυναμικός cepstrum που προτάθηκε από τον Aikawa et al.,το 1992, υλοποιήθηκε σαν έναν cepstral lifter. Πιο πρόσφατα, ένα προσαρμοστικό cepstral στάθμισμα (με βάρη), που ενισχύει τα στοιχεία στενού εύρους ζώνης, και εξασθενίζει τα στοιχεία ανοιχτού εύρους ζώνης, προτάθηκε το 1994 από τους Assaleh and Mammone. Σε μία πλαίσιο προς πλαίσιο βάση , η προσαρμογή ελαττώνει τη σημασία των άσχετων διακυμάνσεων. Στο γενικό πλαίσιο της ταυτοποίησης του ομιλητή, η προσαρμοστική cepstral στάθμιση (με βάρη) βρέθηκε να είναι ιδανική για ζωνοπερατή φασματική λείανση.

Παρόλο το γεγονός ότι αυτή οι cepstral lifters έχει βρεθεί να είναι ευεργητικοί για ASR σε χωρίς θόρυβο ομιλία, ένας αριθμός αυτών των cepstral lifter έχει επίσης βρεθεί να βελτιώνει την απόδοση του ASR παρουσία θορύβου. Οσο αφορά τον γενικό εκθετικό lifter, τελικά αυτός βρέθηκε να είναι ο καλύτερος στην RPS σταθμούμενη με την PLP ανάλυση, σε ένα μέσο SNR (δεκαπέντε με είκοσι dB) (με βέλτιστο γύρω στο S=0,6)[Junqua,1989]. Εντούτοις, όταν το SNR ελαττώθηκε ,το βέλτιστο μετακινήθηκε προς τον RPS lifter (S=1).

8.3.2. ΜΕΤΡΑ ΑΝΘΕΚΤΙΚΗΣ ΠΑΡΑΜΟΡΦΩΣΗΣ

8.3.2.1. ΕΙΣΑΓΩΓΗ

Ενα μέτρο ανθεκτικής παραμόρφωσης θα έπρεπε να δίνει περισσότερο βάρος στις παραμορφώσεις ανάμεσα στα κομμάτια του φάσματος όπου είναι τα λιγότερα επηρρεάσμενα από το θόρυβο. Δίνοντας αυτό τον ορισμό ένας αριθμός από τεχωικές θα μπορούσε να ταξινομηθεί σε αυτή τη κατηγορία. Ωστόσο στα επόμενα κεφάλαια θα θεωρήσουμε μόνο τα μέσα παραμόρφωσης όπου χειρίζονται κατάλληλα τα διανύσματα των χαρακτηριστικών ,ώστε να αποφέρουν μία περισσότερο ανθεκτική αναπαράσταση ομιλίας στο θόρυβο,

πριν να χρησιμοποιηθούν στο στάδιο του ταιριάσματος προτύπων. Τα περισσότερα από τα ανθεκτικής παραμόρφωσης μέτρα εκμεταλλεύονται το γεγονός ότι, παρουσία θορύβου οι κορυφές του φάσματος εξομαλύνονται (ειδικότερα οι υψηλότερες) και ότι η σημαντική πληροφορία συγκεντρώνεται εκεί. Ετσι αυτά τα μέτρα απόστασης είναι ευαίσθητα στις κορυφές και «βαρύνουν» τα μέρη με τα πιο υψηλά SNR ,από ότι τα μέρη με τους χαμηλούς SNR.

Είναι σημαντικό να σχετίσουμε ένα μέτρο μαθηματικής απόστασης σε μερικά είδη σημαντικών -δια των αισθήσεων αντιληπτών - φαινομένων. Π.χ. οι Flanagan,1955a, Flanagan,1955b,και Flanagan,1972,Gray and Markel, 1976, Shikano,1981, πρότειναν μέτρα ,όπου βασίστηκαν στο γεγονός ότι η ανθρώπινη ακοή έχει μεγαλύτερη ευαισθησία στις κορυφές , από ότι στις κοιλάδες ,του μικρού σε χρόνο ,φάσματος.

Ενα μέτρο φασματικού ταιριάσματος πρέπει να είναι ακριβές στα περιβάλλοντα με θόρυβο. Είναι επίσης σπουδαίο να αυξήσει την ομοιότητά του προς το ανθρώπινο αισθητήριο. Ετσι οι ακόλουθες δύο επιπρόσθετες συνθήκες θα πρέπει να προστεθούν στον ορισμό του Gray and Markel για το μέτρο απόστασης (Βλέπε 3.2.1) :

- τα ανθρώπινα ακουστικά χαρακτηριστικά θα έπρεπε να επηρεάζουν τη σχεδίαση του μέτρου απόστασης ,
- περιβαλλοντολογικά χαρακτηριστικά θα έπρεπε να ληφθούν υπόψη.

Στις επόμενες ενότητες φασματικά μέτρα που λαμβάνουν υπόψη τους τα ανθρώπινα ακουστικά χαρακτηριστικά, παρουσιάζονται, καθώς και συζητάται η εφαρμοσιμότητά τους στο ASR παρουσία θορύβου. Ενα ιδιαίτερο χαρακτηριστικό αυτών των μέτρων είναι ότι δεν χρειάζονται καθόλου γνώση των χαρακτηριστικών θορύβου.([Shikano και Itakyra,1992]).

8.3.2.2 ΣΥΧΝΟΤΙΚΑ-ΣΤΑΘΜΙΣΜΕΝΑ ΚΑΙ ΣΥΧΝΟΤΙΚΑ ΠΑΡΑΜΟΡΦΩΜΕΝΑ ΜΕΤΡΑ ΦΑΣΜΑΤΙΚΟΥ ΤΑΙΡΙΑΣΜΑΤΟΣ .

Τα ανθρώπινα ακουστικά χαρακτηριστικά δεν είναι τα ίδια για όλο το φάσμα ακουστικών συχνοτήτων : υπάρχει μεγαλύτερη ευαισθησία στις απλές συχνότητες. Σύμφωνα με αυτή την ιδιότητα, εμελετήθησαν, τα μέτρα φασματικού ταιριάσματος, που βασίζονται στα ακουστικά χαρακτηριστικά. ε Οι Matsumoto και Imai, το 1986 ,μετά από σύγκριση διαφόρων μέτρων φασματικού ταιριάσματος σε θόρυβο, ανέφεραν ότι τα φασματικά σταθμούμενα μέτρα βελτίωσαν την ανθεκτικότητα του ASR. Αυτό επιβεβαιώθηκε από τους Soong και Sondhi το 1987 ,όπου ένα καινούριο συχνοτικά-σταθμισμένο μέτρο ,(παρουσιαζόμενο καλύτερο από το μέτρο παραμόρφωσης των Itakura –Saito, σε χαμηλό SNR) προτάθηκε. Αυτό το μέτρο παραμόρφωσης ορίζεται από τη σχέση :

$$d_{WI} = \log \int_{\pi}^{\pi} F(\omega) \frac{|B(\omega)|^2 d\omega}{|A(\omega)|^2} \frac{}{2\pi}.$$

, όπου ω είναι η γωνιακή συχνότητα , B(ω) και A(ω) είναι αντίστοιχα, το φάσμα αναφοράς και δοκιμής, και F(ω) είναι η συνάρτηση βάρους ,οριζόμενη στους Soong και Sondhi το 1987, σαν ένα «εύρος ζώνης-διαπλατισμένο» φάσμα δοκιμής. Η απόδοση αυτού του μέτρου παραμόρφωσης φάνηκε να είναι η ίδια με εκείνου που δεν είχαν χρησιμοποιηθεί βάρη για μεγάλα SNR αλλά πολύ καλύτερη για την περίπτωση των μεσαίων και χαμηλών SNR. Αυτό οφείλεται σε δύο σημαντικά χαρακτηριστικά του μέτρου :

1) Μία ανομοιόμορφη φασματική στάθμιση
2) Μία προσαρμοστική ρύθμιση του παράγοντα βάρους (στάθμισης).

Τα συχνοτικά-παραμορφωμένα μέτρα παραμόρφωσης ερευνήθηκαν το 1985 από τον Nocerino et al.,και το 1988 από τον Noda.Ο Nocerino et al., πρότεινε μία ικανοποιητική διαδικασία παραμόρφωσης της κλίμακας συχνοτήτων προς μία κρίσιμη bark κλίμακα. Ο Noda πρότεινε ένα συχνοτικά παραμορφωμένο μέτρο απόστασης ,που επέκτεινε τη κλίμακα συχνοτήτων σύμφωνα με το τοπικό SNR σε κάθε συχνότητα του φάσματος. Αυτό το μέτρο απόστασης παρείχε μοναδικά κέρδη σε σχέση με τα συμβατικά μέτρα απόστασης για επιβεβαίωση ομιλητών σε θόρυβο.

8.3.2.3. ΦΑΣΜΑΤΙΚΗΣ ΚΛΙΣΗΣ ΚΑΙ ΚΑΘΥΣΤΕΡΗΣΗΣ ΟΜΑΔΟΣ ΜΕΤΡΑ ΑΠΟΣΤΑΣΗΣ

Το φασματικής κλίσης μέτρο παραμόρφωσης ([Nocerino,1985]) δίνει έμφαση στις θέσεις των φασματικών κορυφών ,όπου είναι αντιληπτικά σημαντικές. Καθόσο αυτό το μέτρο ήταν αρχικά βασισμένο στις διαφορές ανάμεσα στις φασματικές κλίσεις κρίσιμων φασματικών περιοχών , οι Hanson και Wakita το 1986 εφάρμοσαν αυτό το μέτρο σε ένα ολοπολικό μοντέλο φάσματος με την RPS μέτρηση απόστασης. Η ακόλουθη φόρμουλα (των [Hanson και Wakita , 1986]) ορίζει ένα φασματικής κλίσης μέτρο απόστασης των δύο ολοπολικού μοντέλου φάσματος $1/A_T(\omega)$ (δοκιμαστικό)και $1/A_T(\omega)$ (αναφοράς) :

$$d_{SS} = \frac{1}{\pi} \int_{0}^{\pi} \{ \frac{\partial}{\partial \omega} \log \left| \frac{1}{A_T(\omega)} \right|^2 - \frac{\partial}{\partial \omega} \log \left| \frac{1}{A_R(\omega)} \right|^2 \}^2 d\omega.$$

Όπως ήδη αναφέρθηκε το RPS μέτρο απόστασης αποδείχτηκε να λειτουργεί σταθερά καλύτερα από ότι το σύνηθες Ευκλείδιο cepstral μέτρο, ειδικότερα για την αναγνώριση ενθόρυβης ομιλίας[Hanson and Wakita,1986, Hanson and Wakita,1987].

Η εξομαλυμένη ομαδικής καθυστέρησης cepstral απόσταση, εφαρμοζόμενη για στάθμιση των συντελεστών cepstral, προσαυξάνει τις κορυφές μέσω διακρισιμότητας του φάσματος φάσης. Καθόσο αυτό το μέτρο απόστασης έχει φανεί να αποδίδει καλά το θόρυβο, ένα από τα μειονεκτήματά του είναι ότι κάθε φασματική κορυφή με το ίδιο εύρος ζώνης συνεισφέρει ίσα στην απόσταση ,ανεξαρτήτως της ισχύος της και της συχνότητας. Προκειμένου να αντιμετωπισθεί αυτό το πρόβλημα προτάθηκε η σταθμισμένη απόσταση καθυστέρησης ομάδος (WGD), η οποία βασίστηκε στη διαφορά καθυστέρησης ομάδος σταθμιζόμενη από το φάσμα ισχύος. Στην ανεξάρτητη του ομιλητή αναγνώριση και στον προσθετικό λευκό-Gaussian ή χαμηλοπερατά φιλτραρισμένο θόρυβο, η WGD υπερτερούσε ,στους χαμηλούς και μεγάλους SNR, του RPS μέτρου απόστασης .Πάνω στη βάση αυτού του μέτρου απόστασης, μία συχνοτικά σταθμισμένη συνεχής πυκνότητα HMM εισήχθη[Matsumoto,1992].

8.3.2.4 ΜΕΤΡΑ cepstral ΠΡΟΒΟΛΗΣ

Οπως αναφέρθηκε στους Mansour and Juang,1988a ,δεν υπάρχει προφανής λόγος να διατηρηθούν τα συμμετρικά χαρακτηριστικά του μέτρου απόστασης ,εάν κάποιος γνωρίζει με βεβαιότητα ότι τα σήματα και δοκιμής αναφοράς έχουν διαφορετικούς βαθμούς αλλοίωσης από θόρυβο. Βασιζόμενοι σε αυτό και σε μία μελέτη των αποτελεσμάτων του προσθετικού λευκού θορύβου στα cepstral διανύσματα, οι συγγραφείς πρότειναν μία οικογένεια μέτρων παραμόρφωσης ,περισσότερο ανθεκτικών στο θόρυβο[Mansour and Juang,1988a, Mansour and Juang,1989b]. Αυτοί έδειξαν ότι ο προσθετικός λευκός θόρυβος ελαττώνει το μέτρο των cepstral διανυσμάτων και ότι η γωνία απόκλισης ανάμεσα σε δύο τέτοια δυανύσματα είναι ελάχιστα ευαίσθητη στην αλλοίωση από προσθετικό λευκό θόρυβο. Αυτοί επίσης πρότειναν μία βελτιστοποίησης-πλαισίου προσαρμοστική εξίσωση ,η οποία οδήγησε στο ακόλουθο γενικό μέτρο παραμόρφωσης :

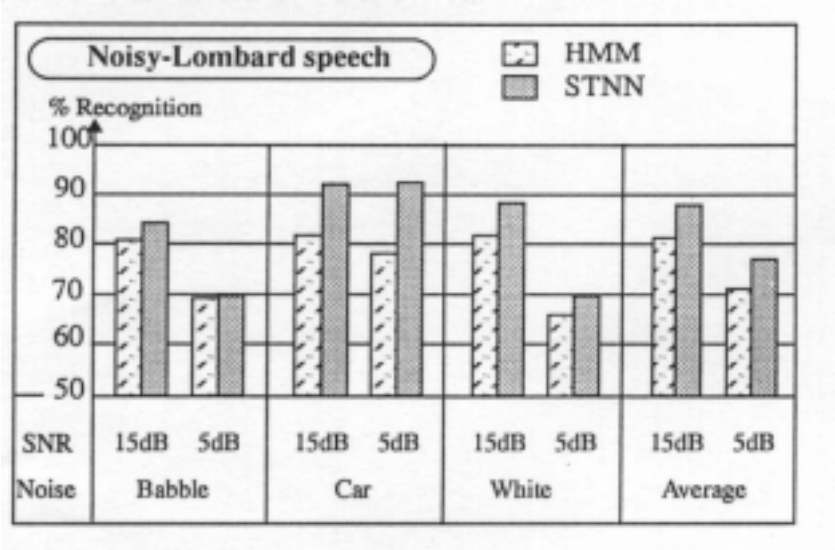
$$d = \left| c_l \right|^a (1 - \cos\beta)$$

όπου β η γωνία ανάμεσα στα c_t, c_r και c_t, c_r είναι αντίστοιχα ,τα δοκιμαστικά και αναφοράς cepstral διανύσματα . Αυτό το γενικής παραμόρφωσης μέτρο ονομάστηκε cepstral προβολής μέτρο, και έδωσε τα καλύτερα αποτελέσματα για α=1.Για διάφορα ολοπολικά μοντέλα και εξαρτώμενη του ομιλητή ενθόρυβη αναγνώριση ομιλίας, το παραπάνω μέτρο με α=1 βρέθηκε να βελτιώνει μοναδικά την ακρίβεια αναγνώρισης ,σε διάφορα πειράματα που συμπεριλάμβαναν διαφορετικούς cepstral lifters

8.3.3 ΔΙΑΚΡΙΝΤΙΚΑ ΜΕΤΡΑ ΟΜΟΙΟΤΗΤΑΣ

Προκειμένου να επιτύγχουμε διάκρση ανάμεσα στα διανύσματα χαρακτηριστικών, ένας αριθμός τεχνικών έχει αναπτυχθεί. Ανάμεσα στις τεχνικές που είναι διαθέσιμες να διακρίνουν ανάμεσα σε κατανομές των παραμετρικών διανυσμάτων, οι ταξινομητές νευρωνικών δικτύων προσφέρουν σημαντικά πλεονεκτήματα (Βλέπε υποενότητα 3.3.4.). Αυτοί επιτρέπουν τη δημιουργία πολύπλοκων αντιστοιχιών ,χωρίς την υπόθεση ενός συγκεκριμένου τύπου κατανομής για τις συναρτήσεις πυκνότητας πιθανότητας. Ακόμα παραπέρα ,αυτοί χρησιμοποιούν διορθωτική εκγύμνανση και για μία απλή τροπολογία δικτύου απαιτούν συχνά λιγότερα δεδομένα εκγύμνανσης από ότι το συμβατικό HMM. Για ενθόρυβη αναγνώριση ομιλίας, η MLP ταξινομητές έχουν βρεθεί να είναι οι καλύτεροι σε πιο συμβατικές μεθόδους ,σε έναν αριθμό λειτουργιών διάκρισης.

Προκειμένου να κάνουμε διάκριση ανάμεσα σε συγχεόμενα γράμματα από θόρυβο, το επιλεκτικά εκγυμναζόμενο νευρωνικό δίκτυο (STNN) βρέθηκε να υπερτερεί του HMM συστήματος αναγνώρισης[Anglede et al.,1993]. Το σχήμα 8.10 δείχνει μία σύγκριση ανάμεσα στο STNN και στο συνεχής πυκνότητας HMM για ενθόρυβη Lombard ομιλία με διαφορετικά SNR, για συγχεόμενα σύνολα από τα γράμματα {M,N}. Τέλος τα μέτρα ομοιότητας μπορούν επίσης να χρησιμοποιηθούν προκειμένου να πάρουμε μία ανθεκτική αναπαράσταση ομιλίας ,σε πολύ-ομιλητικές διαφορές(π.χ.[Hoshimi et al.,1992]).



Σχήμα 8.10 Συγκριτική αξιολόγηση σε θόρυβο ανάμεσα στο STNN και στο συνεχής πυκνότητας HMM ,για τη διάκριση των δύο γραμμάτων {M,N}σε διαφορετικάSNR.

ΜΕΤΑΦΡΑΣΗ ΑΓΓΛΙΚΩΝ ΟΡΩΝ (κατά τη σειρά παρουσιασής τους)

Robustness	ανθεκτικότητα
Speech enhancement	εμπλουτισμός ομιλίας
Background noise	θόρυβος υποβάθρου
Channel noise	θόρυβος καναλιού
Corrupted speech	αλλοιωμένος λόγος
Clean speech	καθαρός λόγος
Adverse environment	αντιξοο περιβάλλον
Lombard effect	αποτέλεσμα(επίδραση) Lombard
Additive noise	προσθετικός θόρυβος
Power spectrum	φάσμα ισχύος
Signal distortion	παραμόρφωση σήματος
Directional mic.	Κατευθυντικό μικρ.
omnidirectional mic.	πολυκατευθυντικό μικρ.
Differential mic.	Διαφορικό μικρ.
Close-talking mic.	Κοντινής ομιλίας μικρ.
Electret speakerphone	πυκνωτικό μικρ.
carbon speakerphone	μικρόφωνο άνθρακος
noise canceling mic.	Μικρόφωνο αφαίρεσης θορύβου
screen	προκάλυμα μικρόφωνου
screen	φιλτράρω
systematic noise	συστηματικός θόρυβος
random noise	τυχαίος θόρυβος
non-communication speech noise	χωρίς μήνυμα θόρυβος ομιλίας
formants	συχνότητες φωνοσυντονισμού φωνητικής οδού
auditory model	ακουστικό μοντέλο
Comodulation	συνδιαμόρφωση
Noise Removing Network	δίκτυο απομάκρυνσης θορύβου
Backpropagation	οπισθοδρομική διάδοση
Selectively Trained Neural Network	επιλεκτ. εκγυμναζόμενο νευρ.δίκτυο
Assessment	εκτίμηση
Noise tracking techniques	τεχνικές ανίχνευσης θορύβου
Training	εκγύμνανση,εκμάθηνση
Subtraction and masking	αφαίρεση και συγκάλυψη
Linear array of mic.	Γραμμική παράταξη μικροφώνων
Humidity	υγρασία

Signal acquisition	απόκτηση σήματος
Robust speech analysis	εύρωστη (ανθεκτική)ανάλυση ομιλίας
Adaptive noise cancellation	προσαρμοστική ακύρωση θορύβου
Least mean square method	μέθοδος ελαχ. τετραγώνων
Perceptually based	αντιληπτικά βασισμένη
linear prediction analysis	γραμμικής πρόβλεψης ανάλυση
time Synchronous	σύγχρονων χρόνων
All-pole	ολοπολίκο
Feedback control	έλεγχο ανάδρασης

Robust spectral estimation	ανθεκτική εκτίμηση φάσματος
Speech transitions	μεταβάσεις στην ομιλία
Speech perception	αντίληψη του λόγου
Speech dynamics	δυναμική της ομιλίας
Regression	παλινδρόμηση
Temporal derivatives	χρονικές παράγωγοι
Autocorrelation domain	πεδίο αυτοσυσχέτισης
Discriminant analysis	διακριτική ανάλυση
Time trajectories	τροχιές χρόνου
RelAtive SpecTrAl technique	συγγενής φασματική τεχνική
Feature adaptation	προσαρμογή χαρακτηριστικών
Multi-layer-perceptron	πολυστρωματικό perceptron
Cepstral compensation	cepstral αντιστάθμιση
Normalization	κανονικοποίηση
Maximum A Posteriory	μεταγεννέστερα
Similarity measures	μέτρα ομοιότητας
Minimum mean square error	ελαχ. μέσο τετραγωνικό σφάλμα
Cepstral lifters	φασματικοί λειαντές
cepstral projection measures	μέτρα cepstral προβολής

Επίσης ,μερικές ακόμα σημαντικές λέξεις είναι οι :

Adverse	αντίξοο
Dissipation	μετάδοση ήχου(διάγχυση)
Compound	μίγμα ,συνδιασμός
Embed	ενσωματώνω
Phase	οργανώνω κατά φάσεις(χρονικές περιόδους)
Reccurent	αναδρομικό
Semantic	σημασιολογικός
Smack(lip)	πλατάγιασμα
Spontaneous	αυθόρμητος
Vocalization	φώνησης
Envelope	περιβάλλουσα
Increment	απειροστή αύξηση
Optimize	βελτιστοποιώ
Convolution	συνέλιξη
Trailing edge	απότατο άκρο
Fricatives	τυρβώδη σύμφωνα

ΠΙΝΑΚΑΣ ΣΥΝΤΜΗΣΕΩΝ

EIH	Ensemble Interval Histogram
NRN	Noise Removing Network
STNN	Selectively Trained Neural Network
LMS	Least mean square method
PLP	Perceptually based linear prediction analysis
SLP	time Synchronous linear prediction analysis
SMC	Short time modified coherence
OSALPC	Autocorrelation linear prediction coding
LDA	Log cepstral domain
IMELDA	Integrate mel scale represantation using LDA
RASTA	RelAtive SpecTrAl technique
MMSE	Minimum mean square error
MAP	Maximum A Posteriory

ΒΙΒΛΙΟΓΡΑΦΙΑ

*Κεφάλαιο THE SPEAKING ENVIRONMENT του βιβλίου "USING SPEECH RECOGNITION" **(Jodith A.Markowitz**)

*Κεφάλαιο TOWARDS ROBUST SPEECH ANALYSIS του βιβλίου «ROBUSTNESS IN AUTOMATIC SPEECH RECOGNITION" **(Jean-Claude Junqua ,Jean-Paul Haton)**

*Καφάλαιο ON THE USE OF A ROBUST SPEECH REPRESENTATION ANALYSIS του βιβλίου «ROBUSTNESS IN AUTOMATIC SPEECH RECOGNITION" **(Jean-Claude Junqua ,Jean-Paul Haton)**