



**ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

Μεταπτυχιακό Πρόγραμμα Σπουδών Ηλεκτρονικού Αυτοματισμού

**Παναγιωτάκης Σπυρίδων (Α.Μ. 97522)**  
**Παστρομάς Σωτήριος (Α.Μ. 97527)**

**ΤΑ ΠΕΔΙΑ ΕΦΑΡΜΟΓΗΣ ΤΗΣ ΑΥΤΟΜΑΤΗΣ**  
**ΑΝΑΓΝΩΡΙΣΗΣ ΟΜΙΛΙΑΣ**

Εργασία στο μάθημα: Επικοινωνία με Ομιλία  
Διδάσκων: Γεώργιος Κουρουπέτρογλου

Αθήνα 1999

**ΤΑ  
ΠΕΔΙΑ ΕΦΑΡΜΟΓΗΣ  
ΤΗΣ  
ΑΥΤΟΜΑΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ  
ΟΜΙΛΙΑΣ  
(ASR: Automatic Speech Recognition)**

## Περιεχόμενα

Περιεχόμενα .....	3
Περίληψη .....	6
Εισαγωγή .....	7
1. Το μοντέλο Star Trek .....	7
2. Οι ανθρώπινοι παράγοντες (Human factors).....	9
2.1 Κατανόηση του έργου.....	10
2.2 Κατανόηση των χρηστών.....	12
2.2.1 Ποιος είναι ο χρήστης;.....	12
2.2.2 Τι πραγματικά θέλουν ή χρειάζονται οι χρήστες;.....	13
2.3 Επαναλαμβανόμενος σχεδιασμός κι έλεγχος.....	14
2.4 Οι νέοι ορίζοντες.....	15
3. Εφαρμογές.....	16
3.1 Εκτέλεση εντολών κι ελέγχου.....	17
3.1.1 Ζητήματα σχεδιασμού για την εκτέλεση εντολών κι ελέγχου .....	18
1. Μέγεθος κι επιλογή λεξιλογίου.....	18
2. Μεταγλώττιση.....	19
3. Γλωσσολογική δομή.....	19
4. Μοντέλο ομιλητή.....	20
5. Ροή ομιλίας.....	20
3.1.2 Το περιβάλλον ομιλίας της εκτέλεσης εντολών κι ελέγχου .....	20
3.1.3 Οι ανθρώπινοι παράγοντες της εκτέλεσης εντολών κι ελέγχου .....	21
3.1.4 Παραδείγματα της εκτέλεσης εντολών κι ελέγχου.....	22
1. Κλήση μακρινής απόστασης.AT&T.....	22
2. Έλεγχος φωνής για άτομα με αναπηρία.Montefiore Hospital .....	23
3. Έλεγχος των τηλεοπτικών κάμερων κλειστού κυκλώματος.NASA Jet Propulsion Laboratory.....	23
3.2 Εισαγωγή Δεδομένων.....	24
3.2.1 Ζητήματα σχεδιασμού για την εισαγωγή δεδομένων....	26
1. Μέγεθος κι επιλογή λεξιλογίου.....	26
2. Μεταγλώττιση.....	26
3. Γλωσσολογική δομή.....	27
4. Μοντέλο ομιλητή.....	27
5. Ροή ομιλίας.....	27
3.2.2 Το περιβάλλον ομιλίας της εισαγωγής δεδομένων.....	27
3.2.3 Οι ανθρώπινοι παράγοντες της εισαγωγής δεδομένων...	28
3.2.4 Παραδείγματα της εισαγωγής δεδομένων.....	29
1. Συντήρηση παγίδων ατμού. Eastman Kodak Company.....	29
2. Διαγράμματα υγείας δίπλα στα κρεβάτια. KorTeam International.....	30
3. Απόθεμα εργατικών αποζημιώσεων. Navistar International.....	31
3.3 Πρόσβαση Δεδομένων.....	31

3.3.1	Ζητήματα σχεδιασμού της πρόσβασης δεδομένων.....	33
	1. Μέγεθος κι επιλογή λεξιλογίου.....	33
	2. Μεταγλώττιση.....	33
	3. Γλωσσολογική δομή.....	34
	4. Μοντέλο ομιλητή.....	34
	5. Ροή ομιλίας.....	34
3.3.2	Το περιβάλλον ομιλίας της πρόσβασης δεδομένων.....	35
3.3.3	Οι ανθρώπινοι παράγοντες της πρόσβασης δεδομένων.....	35
3.3.4	Παραδείγματα της πρόσβασης δεδομένων.....	35
	1. Τραπεζικές συναλλαγές μέσω τηλεφώνου National Westminster Bank P.I.C.London, England.....	35
	2. Φωνητική υλοποίηση κλήσεων. Sprint & Texas Instruments.....	36
3.4	Υπαγόρευση.....	37
3.4.1	Σχεδιαστικά θέματα υπαγόρευσης.....	38
	1. Μέγεθος λεξιλογίου και επιλογή.....	38
	2. Μεταγλώττιση.....	39
	3. Δομή της γλώσσας.....	39
	4. Μοντέλο ομιλητή.....	39
	5. Ροή ομιλίας.....	40
3.4.2	Περιβάλλον ομιλίας της υπαγόρευσης.....	40
3.4.3	Ανθρώπινοι παράγοντες υπαγόρευσης.....	41
3.4.4	Παραδείγματα υπαγόρευσης.....	41
	1. DragonDictate. Dragon συστήματα.....	42
	3. Η σειρά εργαλείων λειτουργίας υπαγόρευσης. Philips συστήματα υπαγόρευσης.....	42
3.5	Χρησιμοποιώντας το τηλέφωνο.....	43
3.5.1	Σχεδιαστικά θέματα χρήσης τηλεφώνου.....	45
	1. Μέγεθος λεξιλογίου και επιλογή.....	45
	2. Μεταγλώττιση.....	46
	3. Γλωσσική δομή.....	46
	4. Μοντέλο ομιλητή.....	46
	5. Ροή ομιλίας.....	47
3.5.2	Περιβάλλον ομιλίας χρησιμοποιώντας το τηλέφωνο....	47
3.5.3	Ανθρώπινοι παράγοντες στη χρήση τηλεφώνου.....	48
3.5.4	Παράδειγμα κλήσης τηλεφώνου.....	50
	Διαχείριση κλήσης. Wildfire Επικοινωνίες.....	50
4.	Λειτουργίες επιτρεπόμενης ομιλίας (speech-enabled).....	51
5.	Εκτίμηση.....	51
	1. Επιλογή ενός κατάλληλου αλγόριθμου ή προϊόντος για μια συγκεκριμένη λειτουργία.....	52
	2. Εκτίμηση της απόδοσης μιας εφαρμογής πριν την ανάπτυξη.....	52
	3. Καθορισμός βελτιώσεων σε υπάρχουσες λειτουργίες αναγνώρισης ομιλίας.....	53
	4. Καθορισμός της αξίας της επιπλέον ομιλίας σ' ένα υπάρχον σύστημα.....	53
	5.1 Επίσημη και μη επίσημη εκτίμηση.....	53
	5.2 Εκτίμηση προϊόντος.....	54

5.2.1	Αναφορικές βάσεις δεδομένων – Συλλογές.....	55
5.2.2	Πρωτόκολλα δοκιμών.....	57
5.2.3	Τόποι λαθών.....	57
5.2.4	Βαθμολογία.....	58
5.2.5	Πέρα από τη βαθμολογία.....	59
5.3	Δοκιμασμός εφαρμογής.....	59
5.3.1	Τι πρέπει να δοκιμαστεί;.....	60
5.3.2	Βαθμολογία στη δοκιμασία εφαρμογής.....	63
5.3.3	Πέρα από τη βαθμολογία της δοκιμής εφαρμογής.....	63
	Πίνακας αγγλικών όρων .....	65
	Πίνακας συντμήσεων .....	67
	Βιβλιογραφία-αναφορές .....	68
	Παράρτημα .....	69

## Περίληψη

Η αναγνώριση ομιλίας μπορεί να προσδώσει σε μία εφαρμογή ή ένα προϊόν μεγαλύτερη χρησιμότητα, ευχρηστία και φυσικότητα. Προσφέρει δυνατότητα χειρισμού απαλλαγμένη από τη χρήση των χεριών, αμεσότερο έλεγχο του προσωπικού περιβάλλοντος στα άτομα με ειδικές οπτικές και κινητικές ανάγκες, ταυτόχρονο έλεγχο των διαφόρων εξαρτημάτων ενός εξοπλισμού, μειωμένο κόστος και μεγαλύτερη αποδοτικότητα.

Από τα τέλη της δεκαετίας του 1980 κι έπειτα, έχει παρατηρηθεί εντυπωσιακή αύξηση στο πλήθος και τις μορφές των εφαρμογών αναγνώρισης ομιλίας που αναπτύχθηκαν, με τις πιο πρόσφατες εμπορικές επιτυχίες να έχουν συντελέσει στη χρήση των σχετικών προϊόντων από ένα ευρύ φάσμα ανθρώπων. Στις περισσότερες περιπτώσεις, η τεχνολογία αυτή χρησιμοποιείται ως μέσο για την υλοποίηση μίας ή περισσότερων από τις ακόλουθες τέσσερις βασικές λειτουργίες:

- Της εκτέλεσης εντολών κι ελέγχου, η οποία αναφέρεται στον προφορικό έλεγχο οποιουδήποτε τύπου εξοπλισμού κι εφαρμόζεται από το χειρισμό βαρέων εξορυκτικών μηχανημάτων μέχρι την “πλοήγηση” στα Microsoft Windows,
- Της εισαγωγής δεδομένων, η οποία χρησιμοποιεί την ομιλία ως μέσο για τον εφοδιασμό με δεδομένα των βάσεων δεδομένων, των συστημάτων ποιοτικού λογιστικού ελέγχου και άλλων λογισμικών,
- Της πρόσβασης δεδομένων, η οποία χρησιμοποιείται για την αναζήτηση κάποιας συγκεκριμένης πληροφορίας σε μια βάση δεδομένων ή ένα αρχείο και
- Της υπαγόρευσης, η οποία μετατρέπει την ομιλία σε κείμενο για γράμματα, αναφορές και άλλες καταγραφές.

Οι περισσότερες εφαρμογές συνήθως επιτελούν συνδυασμό μίας ή περισσότερων από τις παραπάνω λειτουργίες. Κάθε μία από τις λειτουργίες αυτές περιγράφεται ως προς:

- Τις απαιτήσεις σχεδιασμού
- Το περιβάλλον ομιλίας
- Τους ανθρώπινους παράγοντες
- Τα παραδείγματα εφαρμογής

Μια από τις μεγαλύτερες προκλήσεις που η αναγνώριση ομιλίας καλείται ν’ αντιμετωπίσει είναι οι ανεδαφικές προσδοκίες με τις οποίες συνήθως την προσεγγίζουν τόσο οι χρήστες, όσο οι εφαρμογείς (developers) και οι οποίες οφείλονται, κυρίως, στη μειωμένη κατανόηση τόσο του στόχου που πρέπει να επιτευχθεί, όσο των περιορισμών που υπεισέρχονται. Για το λόγο αυτό, ο σωστός χειρισμός των ανθρώπινων παραγόντων προσαυξάνει την πιθανότητα επιτυχίας ενός συστήματος αναγνώρισης.

Η μέτρηση της απόδοσης ενός συστήματος αναγνώρισης γίνεται με την εκτίμηση, μια διεργασία πολύ σημαντική, η οποία καθορίζει την επιτυχημένη ή μη ανάπτυξή του. Επιδίδεται στην αξιολόγηση:

- Αλγοριθμικών μεθόδων που αφορούν δοκιμές καλής λειτουργίας προϊόντων ή εφαρμογών αναγνώρισης,
- Προϊόντων αναγνώρισης ομιλίας που χρησιμοποιούνται για συγκεκριμένη εφαρμογή,
- Σχεδιασμού εφαρμογών αναγνώρισης.

## **Εισαγωγή**

*“Ποια είναι η σπουδαιότητα της αναγνώρισης ομιλίας; Η πράξη έχει δείξει ότι είναι καθ’ όλα εξαρτώμενη από την εφαρμογή που χρησιμοποιεί ο καθένας.”* (Larry Dooling, President and CEO of Verbex Voice Systems, Personal Communication, Advanced Speech Applications and Technologies ’95, April 5, 1995).

Οι περισσότερες εφαρμογές της αναγνώρισης ομιλίας έχουν σχεδιαστεί έτσι ώστε να επιτελούν μία ή περισσότερες από τις ακόλουθες λειτουργίες:

- Εισαγωγή δεδομένων (Data Entry)
- Εκτέλεση εντολών κι ελέγχου (Command and Control)
- Πρόσβαση πληροφοριών (Information Access)
- Υπαγόρευση (Dictation)

Κάθε μία από τις λειτουργίες αυτές χαρακτηρίζεται από ένα μοναδικό σύνολο:

- Απαιτήσεων ομιλίας
- Απαιτήσεων από τον ανθρώπινο παράγοντα
- Τυπικών παραδειγμάτων

Ανάμεσα στους πιο σημαντικούς παράγοντες, αν και συνήθως είναι αυτοί που αγνοούνται συχνότερα, συμπεριλαμβάνονται οι ανθρώπινοι παράγοντες (Human Factors) μιας εφαρμογής. Όπως συμβαίνει με όλες τις τεχνολογίες ανθρωπίνων διεπαφών, η επιτυχία μιας εφαρμογής αναγνώρισης ομιλίας είναι άρρηκτα συνδεδεμένη με την αποδοχή που θα τύχει από τους ανθρώπους που πρόκειται να τη χρησιμοποιήσουν. Η αναγνώριση ομιλίας, όμως, έχει επιπρόσθετα ν’ αντιμετωπίσει μια μεγαλύτερη πρόκληση, καθώς τόσο οι χρήστες, όσο οι εφαρμογείς (developers) συχνά την προσεγγίζουν με ανεδαφικές προσδοκίες. Για τους λόγους αυτούς, η εργασία μας ξεκινά με την εξέταση του μοντέλου Star Trek, ενός συνόλου δηλαδή από ανεδαφικές προσδοκίες σχετικά με την αναγνώριση φωνής που μπορούν να προσβάλλουν την επιτυχία κάθε σχετικής εφαρμογής, για ν’ ακολουθήσει μια επισκόπηση των λοιπών ζητημάτων που άπτονται του ανθρώπινου παράγοντα. Στη συνέχεια παρατίθεται μια λεπτομερής ανάλυση των τεσσάρων κύριων λειτουργιών που αναφέρθησαν και το κεφάλαιο κλείνει με μια αποτίμηση των σχετικών προϊόντων κι εφαρμογών.

### **1. Το μοντέλο Star Trek**

Το μοντέλο Star Trek στην αναγνώριση ομιλίας πηγάζει από τα φουτουριστικά συστήματα επικοινωνίας που χρησιμοποιούνται στις ταινίες, τα βιβλία και τα τηλεοπτικά προγράμματα Star Trek. Οι εν λόγω προηγμένες συσκευές τεχνικής νοημοσύνης παρουσιάζονται ικανές για προφορική (verbal) αλληλεπίδραση, συγκρίσιμη με αυτή των ανθρώπινων όντων.

Τα προβλήματα ξεκινούν από τη στιγμή που οι εφαρμογείς και οι χρήστες ξεχνούν πως τα συστήματα αναγνώρισης ομιλίας του Star Trek δεν αποτελούν παρά εικόνες του μέλλοντος. Οι χρήστες ενθουσιάζονται με το ενδεχόμενο και μόνο της εισαγωγής τους στον κόσμο της επιστημονικής φαντασίας, ενθουσιασμός που όμως μετατρέπεται σε αγανάκτηση και θυμό όταν τα συστήματα αναγνώρισης ομιλίας του σήμερα αποτυγχάνουν ν’ αποδειχθούν τόσο αξιοθαύμαστα όσο οι φανταστικές επικοινωνιακές συσκευές του 23<sup>ου</sup> αιώνα. Η διευθέτηση των

ουτοπικών προσδοκιών αυτού του είδους αποτελεί ζήτημα ζωτικής σημασίας κατά το σχεδιασμό κάθε σχετικής εφαρμογής.

*“Προτού οι χρήστες μας δοκιμάσουν μια εφαρμογή αναγνώρισης ομιλίας, προδιαθέτουμε τις προσδοκίες τους, ούτως ώστε να καταλάβουν ποιες είναι οι δυνατότητες της τεχνολογίας σήμερα και τι μπορούν να περιμένουν. Πολλοί χρήστες φαντάζονται την εισαγωγή και την εξαγωγή φωνής (voice input and output) όπως συμβαίνει στο Star Trek και αναμένουν να μπορούν να επιτυγχάνουν φυσική προφορική υπαγόρευση χωρίς λάθη. Είναι πραγματικά επικίνδυνο να τους οδηγεί κανείς σε τέτοια μονοπάτια.”* (Janet Ho, Vice President of Engineering, Kor Team International and Gabriel Groner, President, Insight Solutions, “Speech Recognition in a hospital bedside charting system”, 1994, p.47).

Ο πρωταρχικός λόγος που οι χρήστες στρέφονται στο μοντέλο Star Trek είναι ότι αυτό είναι το μοναδικό μοντέλο αναγνώρισης ομιλίας που γνωρίζουν. Η τεχνολογία της αναγνώρισης ομιλίας στη δεκαετία του '90 παραμένει απόμακρη, χωρίς να διαθέτει κάποιο τυποποιημένο υπόδειγμα αλληλεπίδρασης.

*“Η αγορά θα περιμένει από το υπόδειγμα αυτό να τη φέρει στο σημείο όπου καθώς θα βαδίζει κάποιος προς το PC ή το workstation του και υπάρχει εκεί ένα μικρόφωνο, να μπορεί να πει: «Γνωρίζω πως να το χειριστώ αυτό. Γνωρίζω πως λειτουργεί με αυτή τη νέα εφαρμογή, ακόμη κι αν ποτέ πριν στο παρελθόν δεν έχω χρησιμοποιήσει αυτή την εφαρμογή.”* (Michael Krajner, Manager of the Speech Group, BBN, Industry Leaders Panel, Advanced Speech Applications and Technologies Conference '94).

Ο σχεδιασμός ενός τέτοιου προτύπου (standard) αποτελεί σκοπό της βιομηχανίας, αν και κάποιες από τις ουσιώδεις αρχές του περιορίζονται από τον υπεισερχόμενο ανθρώπινο παράγοντα. Έως ότου ένα τυποποιημένο υπόδειγμα παρουσιαστεί και γίνει ευρέως αποδεκτό το μοντέλο Star Trek θα υφίσταται. Οι δε εφαρμογείς, που προσαρτούν την αναγνώριση ομιλίας στα συστήματά τους ως μέσο εντυπωσιασμού, επιδιώκοντας να προσδώσουν κάτι από την αύρα του Star Trek, ενδεχομένως θα συνεχίσουν να παράγουν ελλιπώς σχεδιασμένα και άσχημα ολοκληρωμένα συστήματα, όπως αποτυπώνεται και στο ακόλουθο παράδειγμα:

Μια εταιρία X που διέθετε ένα εξελιγμένο σύστημα αυτοματισμού γραφείου, αποφάσισε να προσθέσει την αναγνώριση ομιλίας ως δευτερεύον interface στο προϊόν της, προσδοκώντας με τον τρόπο αυτό να ενισχύσει την high-tech φύση του. Επέλεξαν για το σκοπό αυτό την τεχνολογία μιας έγκριτης εταιρίας, της οποίας το εξαρτώμενο από το χρήστη σύστημα αναγνώρισης ομιλίας είχε εφαρμοστεί επιτυχώς σε ευρεία κλίμακα εφαρμογών και περιβαλλόντων ομιλίας. Η ομιλία θα εφαρμόζονταν σε μια συσκευή εισόδου, η οποία μπορούσε είτε να κρατάτε κοντά στο στόμα, είτε να στηρίζεται στο PC του χρήστη, μόνο που αμέλησαν να καθορίσουν αυστηρά τον τρόπο με τον οποίο θα έπρεπε να χρησιμοποιηθεί η ομιλία. Ένας ρεπόρτερ που δοκίμασε το σύστημα εξύμνησε το προϊόν αλλά βρήκε την αναγνώριση ομιλίας προβληματική, κρίνοντάς την ως αναξιόπιστη και ανώριμη τεχνολογία. Η εταιρία συμφώνησε και, χωρίς να λάβει καμιά ευθύνη για την αποτυχία του interface ομιλίας να λειτουργήσει επαρκώς, αισθάνθηκε ανακουφισμένη όταν ο ρεπόρτερ ανέφερε ότι η ομιλία είχε αναπτυχθεί από κάποια άλλη εταιρία.

Αν οι ερευνητές της εταιρίας X είχαν κατανοήσει τη φύση της αναγνώρισης ομιλίας και είχαν ελέγξει τους ανθρώπινους παράγοντες, ώστε να προκαταλάβουν τους εναλλακτικούς τρόπους με τους οποίους θα μπορούσε να χρησιμοποιηθεί το interface ομιλίας, θα είχαν αναγνωρίσει την πηγή του προβλήματος. Οι χρήστες αυτού του εξαρτώμενου από τον ομιλητή συστήματος θα έπρεπε να εκφωνούν τις εντολές κρατώντας το μικρόφωνο το πολύ μία ίντσα μακριά από το στόμα τους, ενώ αυτοί χρησιμοποίησαν το σύστημα με το μικρόφωνο να



βρίσκεται στο PC, σχεδόν δύο πόδια μακριά. Στις περισσότερες περιπτώσεις, αποστάσεις μικρότερες των δύο ποδιών ελάχιστα επηρεάζουν την ικανότητα των ανθρώπων (ή των συσκευών επικοινωνίας του Star Trek) να κατανοούν την ομιλία. Δυστυχώς, όμως, η τεχνολογία των μικροφώνων και της αναγνώρισης ομιλίας της δεκαετίας μας αντιλαμβάνεται αξιοσημείωτες διαφορές στο σήμα φωνής και στον προσλαμβανόμενο θόρυβο, καθώς το μικρόφωνο απομακρύνεται από το στόμα του ομιλητή. Αυτές οι αποκλίσεις από το αρχικό σήμα ομιλίας επιδρούν καταλυτικά στην ακρίβεια ενός συστήματος αναγνώρισης ομιλίας.

Ένα ακόμη σφάλμα που διαπράχθηκε από την εταιρία X ήταν το πνεύμα του «Δεν εφευρέθηκε εδώ». Η εταιρία X αγνόησε ότι οι περισσότεροι άνθρωποι δε διαχωρίζουν στην αποτίμησή τους για ένα προϊόν τις συνιστώσες που το απαρτίζουν και τις εταιρίες που έχουν κατασκευάσει κάθε μία από αυτές. Σε περίπτωση χαμηλής απόδοσης οποιασδήποτε από τις συνιστώσες και ιδιαίτερα του user-interface, υποβαθμίζεται η αποδοχή ολόκληρου του προϊόντος συνολικά.

Όμως, η αναγνώριση ομιλίας οφείλει να προσφέρει πλεονεκτήματα πέραν της απλής καινοτομίας:

*“Υπάρχουν πλεονεκτήματα και μειονεκτήματα όσον αφορά την εφαρμογή της τεχνολογίας της ομιλίας στις ανθρώπινες διεπαφές (interfaces). Είναι πολύ σημαντικό να γνωρίζεις τους λόγους για τους οποίους χρησιμοποιείς την αναγνώριση ομιλίας, όπως και να μπορείς να τους εκφράσεις κατηγορηματικά.”* (Bruce Balentine, Human Factors Specialist, Scott Instruments [Currently Voice Control Systems, Inc.], Goodlistener Cookbook, 1992, p.47).

Οι παράμετροι αυτοί απαιτούν κατανόηση της τεχνολογίας, του έργου που πρόκειται να επιτελέσει η εφαρμογή, και των ιδιαιτεροτήτων των ανθρώπων που θα τη χρησιμοποιήσουν.

## **2. Οι ανθρώπινοι παράγοντες (Human factors)**

Η αναγνώριση ομιλίας μπορεί να προσδώσει σε μία εφαρμογή ή ένα προϊόν μεγαλύτερη χρησιμότητα, ευχρηστία και φυσικότητα. Προσφέρει δυνατότητα χειρισμού απαλλαγμένη από τη χρήση των χεριών, αμεσότερο έλεγχο του προσωπικού περιβάλλοντος στα άτομα με ειδικές οπτικές και κινητικές ανάγκες, ταυτόχρονο έλεγχο των διαφόρων εξαρτημάτων ενός εξοπλισμού, μειωμένο κόστος και μεγαλύτερη αποδοτικότητα. Με δεδομένα τα πλεονεκτήματα αυτά, καθώς και άλλα εξίσου σημαντικά

*“είναι λογικό ν’ αναρωτιόμαστε για ποιο λόγο η ομιλία δεν είναι περισσότερο διαδεδομένη στην πράξη... Μια προφανής απάντηση είναι ότι η τεχνολογία δεν έχει ακόμη προαχθεί στο βαθμό που χρειάζεται προκειμένου να προσφέρει την απαιτούμενη απόδοση για μαζικές εφαρμογές σε ένα ανταγωνιστικό κόστος. Ένα άλλο πιθανό ενδεχόμενο, όμως, είναι ότι οι υπάρχουσες δυνατότητες της αναγνώρισης ομιλίας δεν έχουν ακόμη αναπτυχθεί επαρκώς, εξαιτίας της μειωμένης κατανόησης των απαιτήσεων που θα πρέπει να πληρεί ένα επιτυχημένο interface ομιλίας.”* (Bob Dampier and Gragam Leedham, “Human Factors”, 1992, p.261).

Η ανεπαρκής κατανόηση πηγάζει από πολλές αιτίες, μεταξύ των οποίων:

- Η ελλιπής εξοικείωση των εφαρμογέων (developers) με τη σχετική τεχνολογία.
- Η ελλιπής εξοικείωση των χρηστών με τη σχετική τεχνολογία.
- Οι ταχύτερες αλλαγές που συμβαίνουν στην εμπορική τεχνολογία της αναγνώρισης ομιλίας.
- Η χρήση της αναγνώρισης ομιλίας σε νέους τύπους συστημάτων, όπως είναι τα καταναλωτικά προϊόντα.

Απ' όλα τα παραπάνω είναι προφανές ότι η ανάπτυξη τέτοιων εφαρμογών αποτελεί πρόκληση. Οι σχετικά νέοι με την αναγνώριση ομιλίας εμπορικοί αντιπρόσωποι μπορούν να ωφεληθούν τα μέγιστα από τα μαθήματα που παίρνουν από τους άλλους, έχοντας κατά νου ότι:

*“Κάθε φορά που εισάγεται μια νέα τεχνολογία στους ανθρώπους, είναι αδύνατο να προβλεφθούν οι αντιδράσεις. Το μόνο που μπορεί να προβλεφθεί είναι ότι σίγουρα θα κάνουν κάτι απρόβλεπτο.”* (Judith Tschirgi, Director Services and Speech Technology, AT & T Network Systems, Personal Communication, 1994).

Η πρόκληση εντοπίζεται στο εξής: να εστιάσει κανείς το ενδιαφέρον του πρωταρχικά στα μηχανικά ζητήματα ή μήπως ν' ασχοληθεί με τους ανθρώπινους παράγοντες σαν να είναι ασυσχέτιστοι με τα υπόλοιπα στάδια του σχεδιασμού;

*“Όμως, ακόμη και οι πιο απλές αποφάσεις που σχετίζονται με τους ανθρώπινους παράγοντες συνήθως εμπεριέχουν συμβιβασμούς. Επίσης, ακόμη και τα πιο κοινά σφάλματα λειτουργούν αθροιστικά. Από τη στιγμή που οι χρήστες συναντούν πρόσθετα προβλήματα κι εκνευρίζονται, εγκαταλείπουν γρήγορα την οποιαδήποτε χρήση ενός συστήματος, εκτός κι αν αναγκάζονται να το χρησιμοποιήσουν.”* (Martha Lindeman, President, Users First, “How to design “caller friendly” applications”, 1993, p.16).

Ο σωστός χειρισμός των ανθρώπινων παραγόντων προσαυξάνει την πιθανότητα επιτυχίας ενός συστήματος, μειώνοντας ταυτόχρονα το κόστος από την τροποποίηση ενός συστήματος που έχει ήδη αναπτυχθεί.

Υπάρχουν διάφοροι τρόποι για να επιτευχθεί ο σχεδιασμός των ανθρώπινων παραγόντων. Όλοι, όμως, περιλαμβάνουν την κατανόηση:

- Της δύναμης, των περιορισμών και της φύσης της τεχνολογίας αναγνώρισης ομιλίας.
- Των στόχων και της δομής του έργου (task).
- Των ανθρώπων που πρόκειται να χρησιμοποιήσουν το σύστημα.

Το μεγαλύτερο μέρος του βιβλίου αυτού καταπιάνεται με το πρώτο ζήτημα. Τα δύο επόμενα είναι στενά συνδεδεμένα με το πρώτο. Βιβλία ολόκληρα έχουν γραφτεί σχετικά με το ζήτημα των ανθρώπινων παραγόντων. Οι ακόλουθες παράγραφοι δεν επιχειρούν να εξαντλήσουν ολόκληρο το πεδίο, προσφέρουν όμως κάποιες βασικές κατευθυντήριες γραμμές και θέτουν αρκετά θέματα προς προβληματισμό.

## **2.1 Κατανόηση του έργου**

Οι αντικειμενικοί στόχοι ενός έργου εφαρμόζονται εντός των συστημάτων που δημιουργούνται για την υλοποίησή του και ανακλώνται στον τρόπο με τον οποίο δομείται το σύστημα, την οργάνωση των υπο-έργων, τις συνιστώσες εισόδου/εξόδου και την επικοινωνία με τους ανθρώπους που χειρίζονται το σύστημα. Οι διάφορες όψεις της αρχικής σύλληψης του συστήματος μεταφράζονται στην πορεία σε σχεδιαστικές προδιαγραφές και κατασκευαστικές λεπτομέρειες.

Οι στόχοι ενός έργου διαμορφώνουν τα βασικά χαρακτηριστικά του ανθρώπινου interface, μεταξύ των οποίων συγκαταλέγονται:

- Η πληροφορία που θα πρέπει να λάβει το σύστημα από το χρήστη
- Η πληροφορία που θα δοθεί στο χρήστη

- Οι επιλογές που θα προσφέρονται στους χρήστες στα διάφορα σημεία της αλληλεπίδρασης, καθώς και οι προτεραιότητές τους
- Η κατάλληλη τοποθέτηση της συσκευής εισόδου
- Οι μορφές του χειρισμού σφαλμάτων
- Οι συνιστώσες του Interface

Αυτές οι όψεις ενός έργου, μαζί με κάποιες άλλες, φιλτράρονται κι εκλεπτίζονται στην πορεία, με βάση τα στοιχεία που αφορούν τους ανθρώπους που πρόκειται να χειριστούν το σύστημα. Οι ενδεχόμενες αλλαγές, με τη σειρά τους, υφίστανται αυστηρό έλεγχο, προκειμένου να πιστοποιηθεί κατά πόσο προάγουν την άμεση και αποδοτική ικανοποίηση των στόχων του έργου.

*“Αρχικά χρησιμοποιήσαμε μια «λαμπερή» και «καθαρή» φωνή για την εκφώνηση των ερωταποκρίσεων, επειδή θεωρήσαμε ότι κάτι τέτοιο θα ήταν καλαίσθητο και περισσότερο ανθρώπινο. Όμως οι χρήστες αντέδρασαν με φλυαρίες. Ως εκ τούτου αποφασίσαμε να διατυπώσουμε τις ερωταποκρίσεις μας με λιγότερο φιλικό τρόπο, περισσότερο εστιασμένο όμως, στην ενθάρρυνση των χρηστών να προφέρουν τις ζητούμενες λέξεις.”* (Robert Perdue, Supervisor, AT & T Bell Laboratories, Personal communication, 1994).

Οι δύο κύριες επιδιώξεις της κατανόησης ενός έργου είναι να καθοριστεί κατά πόσο η αναγνώριση ομιλίας είναι κατάλληλη για το συγκεκριμένο έργο και αν ναι με ποια μορφή θα πρέπει να εφαρμοστεί. Σε γενικές γραμμές, εκείνες οι περιπτώσεις που απαιτούν αποκλειστική προσήλωση του ατόμου στην ακουστική είσοδο (auditory input) δεν αποτελούν ιδανικές υποψήφιας για εφαρμογή της αναγνώρισης ομιλίας. Παράδειγμα μιας τέτοιας περίπτωσης αποτελούν οι επικοινωνίες για τον έλεγχο της εναέριας κυκλοφορίας. Ο έλεγχος της εναέριας κυκλοφορίας είναι μια αρκετά ψυχοφθόρα εργασία, που απαιτεί την προσεκτική παρακολούθηση πολλαπλών ταυτόχρονων συνδιαλέξεων μέσω ενός ενθόρυβου καναλιού. Ως εκ τούτου, η εκπαίδευση των ελεγκτών εναέριας κυκλοφορίας αποτελεί ιδανικό υποψήφιο για την αναγνώριση ομιλίας, αν και παραδόξως, η φύση του έργου δεν προσφέρεται για κάτι τέτοιο. Αν και έχουν αναπτυχθεί πολυάριθμα συστήματα αναγνώρισης ομιλίας για την εκπαίδευση των ελεγκτών, εν τούτοις το μόνο σύστημα που σχεδιάστηκε για εφαρμογή σε πραγματικές επικοινωνίες ελέγχου της εναέριας κυκλοφορίας ήταν αυτό της BBN. Η προσέγγιση που επιχειρήθηκε από τους επιστήμονές της, η επικέντρωση δηλαδή στα κύρια σημεία μιας συζήτησης (gisting), είναι η μοναδική κατάλληλη για το συγκεκριμένο έργο, καθώς η φιλοσοφία στην οποία στηρίζεται επιδιώκει τη μικρότερη δυνατή παρέμβαση (non-intrusive). Στόχος του συστήματος αυτού είναι η αναγνώριση και η εξαγωγή πληροφοριών από τις επικοινωνίες ελέγχου της εναέριας κυκλοφορίας χωρίς να προκαλείται επέμβαση στους εν λόγω διάλογους. Οι συνδιαλεγόμενοι κατ’ ουσία αγνοούν πλήρως την ύπαρξη του συστήματος. Το πλεονέκτημα από μια τέτοια προσέγγιση είναι ότι η απαιτούμενη πληροφορία συγκεντρώνεται άμεσα, διευκολύνοντας έτσι την εργασία του ελεγκτή εναέριας κυκλοφορίας, αφού δεν απαιτείται πλέον η ανάκληση (recall) και καταγραφή των συνδιαλέξεων.

Σε μία άλλη περίπτωση, η UME Corp ανέπτυξε μια χρηματιστηριακή εφαρμογή, η οποία επιτρέπει σε έναν χρηματιστή που συνομιλεί με έναν πελάτη του, να επικοινωνεί ταυτόχρονα και με ένα σύστημα αναγνώρισης ομιλίας. Φυσιολογικά, η αναγνώριση ομιλίας θα έπρεπε να παρεμβαίνει στην επικοινωνία χρηματιστή-πελάτη, εν τούτοις το εν λόγω σύστημα σχεδιάστηκε έτσι ώστε να επιτρέπεται στον πελάτη ν’ ακούει τη στιγμή που ο χρηματιστής εκτελεί τις εντολές του.

## **2.2 Κατανόηση των χρηστών**

*‘Εγώ θα εστιαζόμουν κυρίως στην παρακολούθηση των αντιδράσεων των χρηστών που χρησιμοποιούν το προϊόν σας. Μερικές φορές είναι πολύ διαφωτιστικό και άλλες πάλι πολύ οδυνηρό να παρακολουθείς ανθρώπους που δυσκολεύονται. Οι περισσότεροι αποφεύγουν τα εγχειρίδια χρήσης και δε χρησιμοποιούν το help των προγραμμάτων. Ούτε καν προσπαθούν να καταλάβουν ποια είναι τα αντικείμενα του menu. Οι περισσότεροι επιζητούν να είναι ικανοί να χειρίζονται τέλεια κάθε λογισμικό, αμέσως μετά την εγκατάστασή του.’* (Ray Kurzweil, Founder and Chairman of Kurzweil Applied Intelligence Inc., Leader’s Panel on “Money-making opportunities in speech recognition”, Advanced Speech Applications and Technologies Conference ’95).

### **2.2.1 Ποιος είναι ο χρήστης;**

Οι χρήστες διαφέρουν μεταξύ τους σε πολλά ζητήματα, όπως είναι τα προσωπικά χαρακτηριστικά, οι συνυφασμένες με το έργο ιδιότητες και οι εμπειρίες με το σύστημα. Τα προσωπικά χαρακτηριστικά, όπως οι αντιληπτικές, κινητικές και γλωσσικές ικανότητες συντελούν στην εξειδίκευση του interface, ανεξάρτητα από την κυρίως φύση του έργου. Οι άνθρωποι με κινητική παραλυσία στο άνω μέρος του σώματος, για παράδειγμα, απαιτούν συστήματα απαλλαγμένα από τη χρήση των χεριών, με την εκάστοτε υλοποίηση να διαφέρει ανάλογα με την έκταση και τη φύση της παραλυσίας. Άλλες πάλι κατηγορίες χρηστών χρειάζονται διγλωσσικά (bilingual) ή πολυγλωσσικά (multi-lingual) interface ομιλίας.

Κάποια συνυφασμένα με το έργο χαρακτηριστικά πηγάζουν από τα προϊόντα που χρησιμοποιούνται για την επιτέλεση του έργου. Έτσι, η ακουστική και η οπτική οξύτητα, όπως και η επιδεξιότητα των χεριών, πολύ συχνά περιορίζονται από τον εξοπλισμό προσωπικής προστασίας. Ομοίως, τα προστατευτικά καλύμματα του κεφαλιού συχνά κουκουλώνουν ή παραμορφώνουν την ομιλία. Τα χαρακτηριστικά της ομιλίας των χρηστών επηρεάζονται επίσης από τον περιβαλλοντικό θόρυβο, το στρες και τις σωματικές δραστηριότητες που αποτελούν μέρος του έργου, όπως π.χ. η ανύψωση.

Όταν η αναγνώριση ομιλίας εφαρμόζεται για ένα συγκεκριμένο σκοπό, η διαμόρφωση του τελικού σχεδιασμού μπορεί να γίνει με δύο τρόπους. Και οι δύο, πάντως, εξαρτώνται από τη φύση των χρηστών. Σε περίπτωση που οι χρήστες έχουν προσωπική εμπειρία με το έργο, τότε το λεξιλόγιο και η γλώσσα που χρησιμοποιούν κατά την εκτέλεσή του πρέπει να ενσωματωθούν στο σύστημα. Διαφορετικά, σε περίπτωση μη πεπειραμένων χρηστών οι οποίοι ούτε προτίθενται να γίνουν γνώστες, δεν είναι και ιδιαίτερα έξυπνο να οργανωθεί η εφαρμογή γύρω από τη γλώσσα και το λεξιλόγιο των ειδικών.

Χαρακτηριστικό παράδειγμα της δεύτερης κατηγορίας χρηστών είναι εκείνοι που επιλέγουν να χρησιμοποιήσουν ένα οδηγούμενο από φωνή σύστημα τηλεχειρισμού (voice-driven remote control system) για τον προγραμματισμό ενός VCR. Οι περισσότεροι άνθρωποι αυτής της κατηγορίας αγνοούν τα περί προγραμματισμού των VCR κι επιλέγουν την εφαρμογή ομιλίας ως μέσο που θα τους αποτρέψει από τη διαδικασία εκμάθησής του. Για τους απλούς αυτούς ανθρώπους, η γλώσσα και το λεξιλόγιο που απαιτείται για μια επιτυχή εφαρμογή θα πρέπει ν’ αντανακλά τα γλωσσικά πρότυπα και τις προσδοκίες τους. Ομοίως, όταν ένα καινούργιο έργο δημιουργείται ή προκύπτει από κάποιο άλλο παλιότερο, ίσως χρειαστεί ν’ αναγνωριστούν, να εξειδικευτούν και να ελεγχθούν κάποια νέα πρότυπα συμπεριφοράς. Στην κατηγορία αυτή υπάγεται η αντικατάσταση της χειροκίνητης λειτουργίας ενός εξοπλισμού από φωνητική εντολοδότηση, όπως και η περιπλάνηση στα GUI’s (Graphical User Interfaces) μέσω φωνής αντί του γνωστού mouse. Τα GUI αυτά ενδέχεται να υιοθετούν ομιλητικά πρότυπα για τη λειτουργία τους της μορφής «αρχείο, άνοιξε», αν και

κάποιοι άνθρωποι ίσως να προτιμούν κάποιους περισσότερο γραμματικούς τύπους, όπως «άνοιξε ένα αρχείο» ή «άνοιξε το αρχείο X».

Ο σχεδιασμός των παραμέτρων βοήθειας (help modules), των διεργασιών διόρθωσης σφαλμάτων, των ερωταποκρίσεων και των εντολών εξαρτάται σε μεγάλο βαθμό από τον προβλεπόμενο βαθμό εξοικείωσης των χρηστών με το σύστημα ή με άλλα παρόμοια συστήματα. Σε άλλες περιπτώσεις ίσως πάλι χρειαστεί τα συστήματα να ενσωματώνουν δύο ή περισσότερα σύνολα τεχνικών αλληλεπίδρασης, προκειμένου να μπορούν ν' αντεπεξέλθουν σε τέτοιες διαφορές γνώσης κι εμπειρίας μεταξύ των χρηστών. Εκείνοι οι άνθρωποι που χρησιμοποιούν ένα σύστημα επανειλημμένως ενδεχομένως θ' απαιτήσουν την ενσωμάτωση μιας τεχνικής δυναμικού χρήστη (power user). Η μέθοδος της απρόσκλητης διακοπής (bargain) αποτελεί μια τεχνική αυτής της μορφής και απαντάται πολύ συχνά στις τηλεφωνικές εφαρμογές. Η τεχνική αυτή επιτρέπει στους χρήστες ν' αποκρίνονται προτού ολοκληρωθεί η διατύπωση κάποιας ερώτησης. Εν αντιθέσει, ο σχεδιασμός για εκείνους τους χρήστες που εντελώς αναπάντεχα συνέβηκε να χειριστούν το σύστημα, ή για εκείνους που πολύ σπάνια θα το χρησιμοποιούν, απαιτείται να ελαχιστοποιεί τη σύγχυση και την απορία που ενδεχομένως θα νιώθουν. Ένας τρόπος για να επιτευχθεί αυτό είναι η χρήση κανονικών, μακρύτερων ερωταποκρίσεων που περιέχουν λεπτομερείς εξηγήσεις.

Η απουσία εξοικείωσης με τις εφαρμογές των υπολογιστών ενδέχεται να κάνει τη χρήση ομιλίας περισσότερο ελκυστική, αν και δεν αποκλείεται μια εφαρμογή με ομιλία να φαντάζει ακόμη πιο παράξενη. Για παράδειγμα, κάποιες κατηγορίες χρηστών αισθάνονται άβολα με τις νέες τεχνολογίες και δεν είναι διατιθέμενοι να συνομιλούν με μια μηχανή. Η αντίδραση αυτή είναι συνυφασμένη με αυτό που η Judith Tschirgi της AT & T αποκαλεί «τεχνοφοβία» (technophobia). Κάποιοι άλλοι, πάλι, ενδέχεται να εκνευρίζονται αρκετά εύκολα με μια μηχανή που μοιάζει να προσπαθεί να τους πείσει να κάνουν κάτι που δεν επιθυμούν. Όταν η τεχνοφοβία αγνοηθεί, οι χρήστες αρνούνται να υπακούσουν στις επιταγές της εφαρμογής.

*“Στη δική μας περίπτωση, νομίζω πως οι χρήστες φοβόντουσαν το σύστημα. Έμοιαζαν να το μεταχειρίζονται διστακτικά, αν και υποτίθεται ότι εξυπηρετούσε τη χαλάρωση και την εξοικείωση μαζί του.”* (Steve Rogers, Manager of Systems and Technology, Natwest Markets Ltd., Quoted by Teresa Yraslorza in Merrill Lynch, “Special Supplement:Voice recognition”, 1993, p.10).

### **2.2.2 Τι πραγματικά θέλουν ή χρειάζονται οι χρήστες;**

Οι άνθρωποι που καταφεύγουν σε ένα interface αναγνώρισης ομιλίας συνήθως έχουν έναν συγκεκριμένο στόχο που επιδιώκουν να ικανοποιήσουν. Οι στόχοι τους κυμαίνονται από τη λήψη πληροφορίας για κάποιον τραπεζικό λογαριασμό, την πραγματοποίηση μιας τηλεφωνικής κλήσεως, την ολοκλήρωση μιας διαχειριστικής επιθεώρησης, έως την υπαγόρευση μιας αναφοράς. Σε κάθε περίπτωση η ομιλία έχει δευτερεύουσα σημασία και απλά υποβοηθά στην πραγματοποίηση του στόχου.

Ο μόνος τρόπος για ν' αντιμετωπιστεί η ομιλία περισσότερο ευνοϊκά είναι να πειστούν οι χρήστες ότι μέσω αυτής μπορούν να επιτελούν τις εργασίες τους ευκολότερα, καλύτερα και γρηγορότερα. Η ικανοποίηση των χρηστών συνήθως προηγείται της ικανοποίησης των απλών απαιτήσεων ενός έργου. Συγκεκριμένα, υπάρχουν περιβαλλοντικές, προσωπικές και κοινωνικές παράμετροι που προσβάλλουν ή προσαυξάνουν την αποδοχή μιας εφαρμογής από τους χρήστες, όπως διαφαίνεται και στα ακόλουθα παραδείγματα:

- Οι επιθεωρητές της γραμμής παραγωγής ενός εργοστασίου οφείλουν να ολοκληρώνουν τις επιθεωρήσεις τους εντός του χρόνου που καθορίζεται από την

κίνηση της μονάδας παραγωγής. Χρησιμοποιώντας την αναγνώριση ομιλίας το άγχος για την αναγνώριση και καταγραφή των ελαττωματικών προϊόντων μέσα σε τόσο περιορισμένα χρονικά πλαίσια δύναται να μειωθεί σημαντικά.

- Οι άνθρωποι που επιχειρούν να χειριστούν ένα σύστημα αναγνώρισης ομιλίας ενδέχεται να εκνευριστούν αν τύχουν υποδοχής από ένα μηχάνημα και όχι από κάποιον άνθρωπο.
- Οι άνθρωποι που συνομιλούν με ένα σύστημα αναγνώρισης ομιλίας σε ένα ATM, ενδεχομένως θα αισθανθούν μεγάλη αμηχανία με την ιδέα ότι κάποιος τους βλέπουν να μιλούν σε ένα μηχάνημα.

Οι χρήστες έχουν τις δικές τους προτιμήσεις όσον αφορά το σχεδιασμό του τελικού εξοπλισμού, όπως και συγκεκριμένες ανάγκες. Ενδεχομένως θ' αρνηθούν τη χρησιμοποίηση ακουστικών, όπως ενδέχεται να επιθυμούν οθόνες με πιστές παρουσιάσεις των φορμών που καλούνται να συμπληρώσουν. Η ανίχνευση και ικανοποίηση των απαιτήσεων αυτών προσαυξάνει σημαντικά την αποδοχή μιας εφαρμογής.

Σε ορισμένες περιπτώσεις, οι απαιτήσεις των χρηστών δεν αποκλείεται να συγκρούονται με τους στόχους της εφαρμογής. Για παράδειγμα, κάποιος γιατροί, που τους έχει ανατεθεί το έργο της ηχογράφησης, παραπονιούνται, καθώς είναι υποχρεωμένοι να συντάσσουν μόνοι τους τις δικές τους υπαγορεύσεις αντί να τις αναλαμβάνει κάποιος άλλος. Η δυσφορία τους αυτή, όμως, αντιτίθεται στις ανάγκες των νοσοκομείων, τα οποία εφαρμόζουν τα συστήματα αναγνώρισης ομιλίας ως μέσο για τη μείωση του διαχειριστικού κόστους και την επιτάχυνση της συναλλαγής με τα ασφαλιστικά ταμεία.

Σε άλλες περιπτώσεις πάλι, οι απαιτήσεις ή οι περιορισμοί της τεχνολογίας της αναγνώρισης ομιλίας ενδέχεται να παρεμβαίνουν στην εκπλήρωση των επιδιώξεων των χρηστών.

*“Στην ερώτηση για τον αριθμό της πιστωτικής κάρτας των χρηστών, παρατηρήσαμε ότι το σύστημα δούλεψε καλύτερα με μικρές ακολουθίες ψηφίων, γι’ αυτό και ζητήσαμε από τους χρήστες να εισάγουν τους αριθμούς τους σε ομάδες των τεσσάρων ψηφίων. Πολλοί άνθρωποι, όμως, δεν γνώριζαν τη σημασία της λέξης «ψηφίο». Όταν, λοιπόν, την αντικαταστήσαμε με τη λέξη «αριθμός» οι άνθρωποι μπερδεύονταν περισσότερο, καθώς ο «αριθμός» μπορεί να υπονοεί τόσο ένα μονό ψηφίο, όσο και ολόκληρη την ακολουθία των ψηφίων. Έτσι, οι χρήστες ζητούσαν να μάθουν αν έπρεπε να εισάγουν ολόκληρη την ακολουθία των ψηφίων τέσσερις φορές.”* (Robert Perdue, Supervisor, AT & T Bell Laboratories, Personal communication, 1994).

Παρ’ όλ’ αυτά, στις περισσότερες περιπτώσεις οι απαιτήσεις και οι ανάγκες των χρηστών μπορούν να ικανοποιηθούν μέσα από προσεκτικό σχεδιασμό κι έλεγχο.

### **2.3 Επαναλαμβανόμενος σχεδιασμός κι έλεγχος (Iterative Design and Testing)**

*“Ο προσεκτικός σχεδιασμός κι ο επαναλαμβανόμενος έλεγχος της αλληλεπίδρασης χρήστη-συστήματος είναι απαραίτητα στοιχεία προκειμένου να γίνει αντιληπτή η πραγματική δυναμική της χρήσης ομιλίας σε μία μορφή interface που θα χαρακτηρίζεται ως περισσότερο φυσική κι ενστικτώδης.”* (Gerhard Deffner of Texas Instruments, Linda Jo Black of Sprint and

Mark Ollmann of Texas Instruments, “Wizard of Oz market testing of spoken commands interface style”, 1994, p.124).

Η χρήση του επαναλαμβανόμενου σχεδιασμού κι ελέγχου παρέχει στους επιστήμονες τη δυνατότητα ν’ αναλύουν μια εφαρμογή στα συστατικά της μέρη κι ακολούθως ν’ αξιολογούν τις τεχνολογικές απαιτήσεις και τις απαιτήσεις σε ανθρώπινους παράγοντες για κάθε μία από αυτές. Κάποια στοιχεία του επαναλαμβανόμενου σχεδιασμού μπορούν να υλοποιηθούν μέσω πιλοτικών συστημάτων ή πρωτοτύπων. Τα πρωτότυπα χρησιμεύουν, για παράδειγμα, στην παρατήρηση του τρόπου με τον οποίο αλληλεπιδρούν οι άνθρωποι με το σύστημα.

Μια άλλη τεχνική που εφαρμόζεται αρκετά συχνά είναι ο έλεγχος μέσω του Μάγου του Οζ (Wizard of Oz testing). Εμπνευσμένοι από τις ευφυείς ψευδαισθήσεις που δημιουργούσε ο χαρακτήρας του τίτλου στο βιβλίο και την ταινία «Ο μάγος του Οζ», οι ειδικοί περί των ανθρώπινων παραγόντων χρησιμοποιούν την προσέγγιση αυτή ως μέσο για τη μελέτη της συμπεριφοράς των χρηστών, όταν αυτοί αντιπαρατίθενται με διάφορους τύπους και μορφές interfaces. Οι έλεγχοι μέσω του Μάγου του Οζ προσομοιάζουν την αναμενόμενη αλληλεπίδραση μεταξύ ενός ανθρώπου κι ενός συστήματος αναγνώρισης που επιτελεί ένα συγκεκριμένο έργο. Με τον τρόπο αυτό διευκολύνεται η συλλογή δεδομένων όσον αφορά το λεξιλόγιο, τα πρότυπα ομιλίας (speech patterns) και γενικά το σύνολο των αντιδράσεων των χρηστών για μια σχεδιαζόμενη εφαρμογή ή συνιστώσα. Επίσης, ελαχιστοποιούν τον απαιτούμενο χρόνο για την ανάπτυξη εναλλακτικών σχεδιασμών, αποτελώντας έτσι πολύτιμο συμπλήρωμα του επαναλαμβανόμενου σχεδιασμού κι ελέγχου.

## **2.4 Οι νέοι ορίζοντες**

Οι πρόσφατες εμπορικές επιτυχίες έχουν συντελέσει στη χρησιμοποίηση της αναγνώρισης ομιλίας σε ευρεία κλίμακα εφαρμογών, οι οποίες χρησιμοποιούνται από ένα ευρύ φάσμα ανθρώπων. Οι διευρυμένοι αυτοί νέοι ορίζοντες έχουν θέσει ποικίλα ερωτήματα συνυφασμένα με τους ανθρώπινους παράγοντες. Μερικά από αυτά μπορούν να ικανοποιηθούν με τις τρέχουσες παραδοχές, ενώ για κάποια άλλα θ’ απαιτηθούν νέες λύσεις.

Μέχρι πρόσφατα, για παράδειγμα, οι τηλεφωνικές εφαρμογές δεν μπορούσαν να επεξεργαστούν την απάντηση ενός χρήστη προτού ολοκληρώνονταν η ερώτηση που αντιστοιχούσε στην απάντηση αυτή. Οι απαντήσεις, πάλι, όφειλαν να περιορίζονται στις λέξεις που καθορίζονταν από την ερώτηση. Οι δύο αυτοί περιορισμοί παρεμπόδιζαν την επιτυχία των τηλεφωνικών συστημάτων ομιλίας. Η ανάπτυξη, όμως, του συστήματος απρόσκλητης διακοπής (barge-in) και του εντοπισμού λέξεων (word spotting) μετέτρεψε την τηλεφωνία σε μία από τις πιο δημοφιλείς πλατφόρμες εφαρμογής της αναγνώρισης ομιλίας.

Ο σχεδιασμός των ανθρώπινων παραγόντων που θ’ απαιτηθεί για τα συστήματα του μέλλοντος δεν είναι ακόμη γνωστός φυσικά. Κάποιες ανακαλύψεις, χωρίς αμφιβολία, θ’ απομυθοποιήσουν παραδοχές που έχουν μεταφερθεί στην ομιλία από άλλες τεχνικές, όπως τα συστήματα IVR (Interactive Voice-Response) που λειτουργούν μέσω επαφής, ενώ κάποιες άλλες θα είναι καρποί μιας προσεκτικής ανίχνευσης των νέων οριζόντων. Οι Boitet και Loken-Kim (1993) έχουν αναφερθεί στους ανθρώπινους παράγοντες που θα διέπουν μια μελλοντική εφαρμογή: τη μετάφραση ομιλίας προς ομιλία σε πραγματικό χρόνο (speech-to-speech translation).

### 3. Εφαρμογές

Από τα τέλη της δεκαετίας του 1980 κι έπειτα, έχει παρατηρηθεί εντυπωσιακή αύξηση στο πλήθος και τις μορφές των εφαρμογών αναγνώρισης ομιλίας που αναπτύχθηκαν. Στις περισσότερες περιπτώσεις, η τεχνολογία αυτή χρησιμοποιείται ως μέσο για την υλοποίηση μίας ή περισσότερων από τις λειτουργίες που περιγράφονται στον ακόλουθο πίνακα:

<b>Όνομα Λειτουργίας</b>	<b>Περιγραφή Λειτουργίας</b>	<b>Παραδείγματα</b>
Εκτέλεση εντολών κι ελέγχου	Προφορικός έλεγχος εξοπλισμού ή προγραμμάτων λογισμικού	Προγραμματισμός ενός VCR, Έλεγχος μιας αναπηρικής πολυθρόνας, Πλοήγηση (navigation) στα Windows
Εισαγωγή Δεδομένων	Φωνητική εισαγωγή δεδομένων σε προγράμματα λογισμικού	Υλοποίηση ενός λογισμικού ελέγχου, Συμπλήρωση μιας φόρμας, Καταγραφή μιας παραγγελίας
Πρόσβαση Δεδομένων/Ανάκτηση Πληροφοριών	Εφαρμογή επερωτήσεων και άλλων αιτήσεων για την αναζήτηση κάποιας συγκεκριμένης πληροφορίας σε μια βάση δεδομένων ή ένα αρχείο	Τραπεζικές συναλλαγές μέσω τηλεφώνου, Εξεύρεση δρομολογήσεων, Υλοποίηση τηλεφωνικών κλήσεων
Υπαγόρευση	Καταγραφή αλληλογραφίας, αναφορών και άλλων κειμένων	Υπαγόρευση ενός γράμματος, Υπαγόρευση μιας ιατρικής αναφοράς

Η εκτέλεση εντολών κι ελέγχου αναφέρεται στον προφορικό έλεγχο οποιουδήποτε τύπου εξοπλισμού κι εφαρμόζεται από το χειρισμό βαρέων εξορυκτικών μηχανημάτων μέχρι την “πλοήγηση” στα Microsoft Windows. Η εισαγωγή δεδομένων χρησιμοποιεί την ομιλία ως μέσο για τον εφοδιασμό με δεδομένα των βάσεων δεδομένων, των συστημάτων ποιοτικού λογιστικού ελέγχου και άλλων λογισμικών. Μέχρι τα τέλη της δεκαετίας του '80 οι περισσότερες εφαρμογές της εισαγωγής δεδομένων αναπτύσσονταν για στρατιωτικούς και κατασκευαστικούς σκοπούς κυρίως. Από τότε, η χρήση τους εξαπλώθηκε δραματικά και περιλαμβάνει πλέον τη συμπλήρωση φορμών, την τήρηση διαγραμμάτων υγείας ασθενών στα νοσοκομεία και τον προγραμματισμό ραντεβού. Η πρόσβαση δεδομένων αποτελεί την πιο πρόσφατη λειτουργία αναγνώρισης ομιλίας και χρησιμοποιείται συνήθως σε τηλεφωνικές εφαρμογές. Τα συστήματα υπαγόρευσης μετατρέπουν την ομιλία σε κείμενο για γράμματα, αναφορές και άλλες καταγραφές.

Κάθε μία από τις λειτουργίες αυτές περιγράφεται ακολούθως ως προς:

- Τις απαιτήσεις σχεδιασμού
- Το περιβάλλον ομιλίας
- Τους ανθρώπινους παράγοντες
- Τα παραδείγματα εφαρμογής

Τα παραδείγματα επιλέχθηκαν για την παρουσίαση θεμάτων που θα συζητηθούν στις επόμενες παραγράφους και δε θα πρέπει να θεωρηθούν ως υποδείξεις συγκεκριμένων προϊόντων.



Είναι σημαντικό να σημειωθεί ότι οι περισσότερες εφαρμογές χρησιμοποιούν περισσότερες από μία από τις λειτουργίες του παραπάνω πίνακα. Οι εφαρμογές που στηρίζονται στα Windows, για παράδειγμα, αν και είναι συστήματα εκτέλεσης εντολών κι ελέγχου κυρίως, σχεδιασμένα για το χειρισμό menus και άλλων αντικειμένων στο περιβάλλον των Windows, εν τούτοις οι περισσότερες από αυτές έχουν επίσης τη δυνατότητα εισαγωγής δεδομένων, ενώ μία τουλάχιστον έχει και δυνατότητες υπαγόρευσης. Ομοίως, οι εφαρμογές υπαγόρευσης συνήθως ενσωματώνουν μια μικρή λίστα όρων σχετικών με την εκτέλεση εντολών κι ελέγχου για την πλοήγηση του συστήματος. Καθώς η εξοικείωση των καταναλωτών με την αναγνώριση ομιλίας μεγαλώνει και οι εμπορικές εφαρμογές πληθαίνουν, ο συνδυασμός των λειτουργιών σε μία εφαρμογή θα θεωρείται δεδομένος. Για τους σχετικούς επιστήμονες αυτό θ' απαιτήσει την εξισορρόπηση πολλαπλών ζητημάτων ομιλίας, περιβαλλόντων και ανθρωπίνων παραγόντων.

### **3.1 Εκτέλεση εντολών κι ελέγχου**

Η εκτέλεση εντολών κι ελέγχου από έναν εξοπλισμό περιλαμβάνει το χειρισμό υπολογιστών, βαρέων μηχανών και άλλων μηχανημάτων μέσω φωνής. Οι εφαρμογές πολεμικής διαχείρισης για στρατιωτικά τανκ, πλοία και πολεμικά αεροπλάνα ήταν από τις πρώτες ερευνητικές περιπτώσεις εκτέλεσης εντολών κι ελέγχου που αναπτύχθηκαν, ενώ η πρώτη εμπορική εφαρμογή αναφέρονταν στον έλεγχο κατασκευαστικού εξοπλισμού. Τελευταία, οι εφαρμογές της εκτέλεσης εντολών κι ελέγχου έχουν επεκταθεί και σε άλλες βιομηχανίες με πιο γνωστά παραδείγματα αυτά που αναφέρονται στην υλοποίηση τηλεφωνικών κλήσεων μακρινών αποστάσεων στις τηλεπικοινωνίες και στον έλεγχο λογισμικού, όπως των Windows και των εφαρμογών που τρέχουν πάνω από αυτά. Σε γενικές γραμμές, οι εφαρμογές στις άλλες βιομηχανίες περιλαμβάνουν έλεγχο σε:

- |                              |  |
|------------------------------|--|
| • Αρχιτεκτονική              | Εφαρμογές σε “έξυπνα” κτίρια   |
| • Αυτοκίνηση                 | Παράθυρα, θερμαντές και άλλα βοηθητικά συστήματα, τηλέφωνα αυτοκινήτων             |
| • Καταναλωτικά προϊόντα      | Παιχνίδια, PDA's (Personal Digital Assistants), VCR's                              |
| • Σχεδιασμό                  | Οθόνες και menus συστημάτων CAD (Computer-Aided Design)                            |
| • Κατασκευές                 | Βαρύς εξοπλισμός, έλεγχος κατασκευαστικών διεργασιών                               |
| • Περίθαλψη                  | Ραδιομικροσκόπια σε χειρουργικά κρεβάτια   |
| • Στρατιωτικές χρήσεις       | Εμφανίσεις χαρτών στις καμπίνες πλοήγησης (cockpits) των αεροπλάνων                |
| • Εξόρυξη                    | Βαρύς εξορυκτικός εξοπλισμός σε επικίνδυνους χώρους                                |
| • Αυτοματισμούς γραφείου     | Software για τη συμπλήρωση φορμών  |
| • Άτομα με ειδικές ανάγκες   | Αναπηρικές πολυθρόνες, κρεβάτια και άλλος εξοπλισμός                               |
| • Εξερεύνηση του διαστήματος | Οθόνες τηλεόρασης για ρομποτικούς επισκευαστικούς βραχίονες στα διαστημικά οχήματα |
| • Τηλεπικοινωνίες            | Εκτέλεση τηλεφωνικών κλήσεων   |

Η φωνητική εκτέλεση εντολών κι ελέγχου συνήθως χρησιμοποιείται για το χειρισμό συστημάτων υποστήριξης, έτσι ώστε τα χέρια του χρήστη να παραμένουν ελεύθερα για το χειρισμό του πρωτεύοντος εξοπλισμού χωρίς διακοπή. Αποτέλεσμα της δυνατότητας αυτής αποτελεί η διατήρηση της οπτικής και χειρονακτικής δεξιότητας επικεντρωμένης στο κυρίως έργο, καθώς και η βελτίωση της παραγωγικότητας, της ακρίβειας και της ασφάλειας. Για παράδειγμα, κατά τη λειτουργία ενός συστήματος σχεδιασμού υποβοηθούμενου από υπολογιστή (CAD) ο σχεδιαστής θα εξακολουθεί να χρησιμοποιεί το εργαλείο CAD την ίδια στιγμή που θα δίνει εντολές για την πλοήγηση της οθόνης και του menu με τη φωνή, οι στρατιωτικοί πιλότοι θα μπορούν ν' ανακτούν πρόσβαση σε οπτικές επιδείξεις, όπως η ανάγνωση χαρτών, χωρίς να μετακινούν τα χέρια τους από τον έλεγχο του σκάφους, ενώ οι οδηγοί αυτοκινήτων θα μπορούν να καλούν αριθμούς τηλεφώνων κρατώντας το τιμόνι και κοιτάζοντας το δρόμο.

Τα interfaces ομιλίας στα συστήματα υπολογιστικού λογισμικού, όπως τα Windows ή τα εκτελεστικά συστήματα πληροφοριών (EIS: Executive Information Systems), συνήθως σχεδιάζονται για να διευκολύνουν τη γρήγορη εξοικείωση με τις καινούργιες υπολογιστικές εφαρμογές, ώστε ν' αυξάνεται η παραγωγικότητα. Όταν πάλι εφαρμόζονται για το χειρισμό εξοπλισμού σε ατομικό προσωπικό περιβάλλον, η ομιλία επιτρέπει σε ανθρώπους με σοβαρές φυσικές αδυναμίες να περιορίσουν την εξάρτησή τους από τους άλλους.

### **3.1.1 Ζητήματα σχεδιασμού για την εκτέλεση εντολών κι ελέγχου**

#### **1. Μέγεθος κι επιλογή λεξιλογίου**

Οι απαιτήσεις σε λεξιλόγιο και η δομική πολυπλοκότητα των συστημάτων εκτέλεσης εντολών κι ελέγχου προσαυξάνουν την περιπλοκή του εξοπλισμού που πρόκειται να ελεγχθεί. Σε πολλές περιπτώσεις, το μέγεθος ενός λεξιλογίου εκτείνεται από 10 έως 50 λέξεις. Αυτό το περιορισμένο λεξιλόγιο αντανακλά το γεγονός ότι συνήθως είναι πεπερασμένο το πλήθος των λειτουργιών που μπορεί να εκτελέσει μια απλή μηχανή. Η τεχνολογία των φωνητικών VCR, για παράδειγμα, απαιτεί λεξιλόγιο με λιγότερες από 70 λέξεις για τον προγραμματισμό ενός VCR. Στο αντίθετο άκρο εντοπίζονται οι καμπίνες πλοήγησης των στρατιωτικών αεροσκαφών με τον πολύπλοκο έλεγχο πτήσης, τη διαχείριση μαχών και τα εξελιγμένα ηλεκτρονικά, οι οποίες έχουν γίνει τόσο πολυσύνθετες ώστε οι ερευνητές που ασχολούνται με τις εφαρμογές της ομιλίας σε αυτές έχουν υπολογίσει το εύρος του απαιτούμενου λεξιλογίου στις 1300 με 5000 λέξεις. Ο σχεδιασμός λεξιλογίου για τα interfaces των εφαρμογών λογισμικού κυμαίνεται ανάμεσα στα δύο αυτά άκρα, με τα πολυ-λειτουργικά (multi-tasking) συστήματα και τις πολυσύνθετες (multi-features) εφαρμογές ν' απαιτούν τα περισσότερο εκτεταμένα λεξιλόγια.

Μία από τις δυσκολότερες και πιο κουραστικές όψεις στο σχεδιασμό μιας εφαρμογής ομιλίας είναι η επιλογή του λεξιλογίου. Μια μερική λύση είναι αυτή που έχει αναπτυχθεί για την εκτέλεση εντολών κι ελέγχου στα υπολογιστικά προγράμματα: Προϊόντα, όπως το Verbex' Listen για Windows, περιλαμβάνουν εργαλεία ικανά να εξετάζουν τον κώδικα ενός τμήματος του software και να εξάγουν από αυτόν το λεξιλόγιο που απαιτείται για τον έλεγχο των οθονών και των menus, ενώ κάποια άλλα ενσωματώνουν εργαλεία που επιτρέπουν στους τελικούς χρήστες να δημιουργούν τα δικά τους εξατομικευμένα λεξιλόγια και τις προσωπικές τους λίστες.

Πέρα, όμως, από τις περιπτώσεις του γραφικού λογισμικού (graphically-oriented software), η ανάπτυξη λεξιλογίου για εφαρμογές της εκτέλεσης εντολών κι ελέγχου μπορεί να είναι πονηρή. Στις περισσότερες εφαρμογές ο έλεγχος του εξοπλισμού εκτελείται χειροκίνητα

μέσω μεγάλων μη προφορικών διεργασιών και ο σχεδιαστής εφαρμογών ομιλίας καλείται να εντοπίσει εκείνες τις απλές προφορικές εκφράσεις που σχετίζονται με τις χειροκίνητες λειτουργίες. Στις περιπτώσεις όπου η εργασία είναι ελάχιστα κατανοητή από το χρήστη, όπως για παράδειγμα στον προγραμματισμό ενός VCR, ο δομικός σχεδιασμός του λεξιλογίου και της γλώσσας απαιτείται να είναι εστιασμένος στους ανθρώπινους παράγοντες.

## **2. Μεταγλώττιση**

Εφ' όσον οι προφορικές εντολές οφείλουν να επικοινωνούν μ' ένα τμήμα του εξοπλισμού, η μετάφραση σπανίως αναφέρεται σε κείμενο. Συνήθως πρόκειται για κάποιον ψηφιακό κώδικα ή κάποιο ηλεκτρονικό σήμα. Σε πρώτη φάση προσδιορίζεται το σύνολο των σημάτων εντολών που αναγνωρίζονται από τον εξοπλισμό, ώστε ακολούθως ο αναγνωριστής (recogniser) να μετατρέπει την ηχητική είσοδο σε μια μορφή η οποία, κατ' ελάχιστο, θα μπορεί να μετασηματιστεί γρήγορα κι αποδοτικά στο σωστό πρότυπο σχήμα.

## **3. Γλωσσολογική δομή**

Η δομή μιας εφαρμογής εκτέλεσης εντολών κι ελέγχου εξαρτάται από τον εξοπλισμό και στις περισσότερες περιπτώσεις είναι υψηλά δομημένη και συμπληρωμένη με διεργασίες διόρθωσης των λαθών. Οι γραμματικές πεπερασμένων καταστάσεων (finite-state grammars) και κάποια άλλα ισοδύναμα συστήματα παρέχουν τη δυνατότητα ελέγχου της αλληλουχίας των λειτουργιών προκειμένου να εξασφαλίζεται ότι ανταποκρίνονται στις ανάγκες μιας μηχανής.

Ο εντοπισμός των λέξεων κλειδιά (keyword spotting) είναι η μέθοδος που χρησιμοποιείται συνήθως στις τηλεφωνικές εφαρμογές. Εν αντιθέσει με τις περισσότερες εφαρμογές εκτέλεσης εντολών κι ελέγχου, τα τηλεφωνικά συστήματα, σε γενικές γραμμές, δεν έχουν πεπειραμένους χρήστες. Συνήθως χρησιμοποιούνται από μεγάλους πληθυσμούς αρχαρίων χρηστών, των οποίων οι αποκρίσεις σ' ένα σύστημα αναγνώρισης ομιλίας ενδέχεται να είναι απρόβλεπτες. Αν και από την οπτική γωνία των χρηστών εμφανίζεται σαν να μην υπάρχει καμία δομή, εν τούτοις ο εντοπισμός των λέξεων-κλειδιά επιτρέπει στους χρήστες να μιλούν αυθόρμητα, υπό την προϋπόθεση να χρησιμοποιούν κάποιες από τις αναμενόμενες λέξεις.

Τα χαρακτηριστικά απόδοσης ενός εξοπλισμού ενδέχεται να παρουσιάζουν ιδιαίτερες επικοινωνιακές προκλήσεις. Μία από τις δυσκολότερες είναι η έκφραση της κίνησης. Όταν ο εξοπλισμός είναι ένα PC ή κάποια άλλη συσκευή που επιτρέπει πολλαπλούς και ταυτόχρονους τρόπους επικοινωνίας, η ομιλία ενδεχομένως θα συνδυάζεται και με άλλες συσκευές προκειμένου να προάγεται ένα πακέτο ελέγχου ολόκληρου του εξοπλισμού. Αυτή η πολυτροπική προσέγγιση (multimedia approach) επιτρέπει το χειρισμό κάθε εισόδου με τρόπο που να ταιριάζει στις δυνατότητες της κάθε μιας. Στις περιπτώσεις όπου η ομιλία είναι ο μοναδικός τρόπος εκδήλωσης των εντολών, όλες οι λειτουργίες της εκτέλεσης εντολών κι ελέγχου θα πρέπει να ικανοποιούνται με τρόπο απλό και φυσικό.

Κάποιοι εφαρμογείς έχουν εκμεταλλευτεί την ασάφεια της γλώσσας προκειμένου να υπερπηδήσουν το πρόβλημα αυτό. Το 1987 η RCA σχεδίασε ένα πρωτότυπο σύστημα το οποίο χρησιμοποιούσε τη φωνή για τον έλεγχο των κάμερων που παρακολουθούσαν τους ρομποτικούς επισκευαστικούς βραχίονες στα διαστημόπλοια της NASA. Το σύστημα αυτό ενσωμάτωνε κάποιους ασαφείς αλλά αποτελεσματικούς χωρικούς τροποποιητές (modifiers), όπως "λιγάκι" και "όλο αριστερά", για τον έλεγχο της έκτασης και του είδους των κινήσεων της κάμερας. Η μετάφραση μιας έκφρασης όπως το "λιγάκι" ήταν ένα καλά καθορισμένο

ποσοστό μετακίνησης της κάμερας, μόνο που η γλώσσα ήταν περισσότερο φυσική για τους αστροναύτες από το να διατυπώνουν μια ακριβή αριθμητική έκφραση.

#### **4. Μοντέλο ομιλητή**

Στις περισσότερες εφαρμογές της εκτέλεσης εντολών κι ελέγχου ο πληθυσμός των χρηστών αποτελείται είτε από μεμονομένα άτομα, είτε από μικρά γκρουπ ατόμων. Ως εκ τούτου, η εξαρτώμενη από τον ομιλητή αναγνώριση χρησιμοποιείται πολύ συχνά προκειμένου να αυξηθεί η ακρίβεια και να υποβοηθηθεί ο εγκλιματισμός των χρηστών στο λεξιλόγιο και τη γραμματική του συστήματος.

Σε περιπτώσεις εφαρμογών, πάντως, που σχεδιάζονται για ευκολία στη χρήση, όπως είναι τα interfaces των υπολογιστών, η εξάρτηση από τον ομιλητή ενδέχεται να μη λειτουργεί πάντα παραγωγικά. Στα συστήματα αυτά δεν αποκλείεται να ενσωματώνονται μεγαλύτερα λεξιλόγια από αυτά που η εφαρμογή μπορεί να εκπαιδευτεί, με αποτέλεσμα οι απαιτήσεις της εγγραφής (enrollment) να μην εξυπηρετούν την ευκολότερη, ταχύτερη και απλούστερη χρήση των υπολογιστών. Ο εντοπισμός αυτών των γλωσσολογικών χαρακτηριστικών και περιορισμών των ανθρωπίνων παραγόντων στα interfaces των υπολογιστών έχει αναγκάσει κάποιους πωλητές (vendors) να επιλέξουν μοντελοποίηση είτε προσαρμοζόμενη στον ομιλητή, είτε εντελώς ανεξάρτητη από αυτόν. Οι τηλεφωνικές εφαρμογές συνήθως χρησιμοποιούν τεχνολογία ανεξάρτητη του ομιλητή, γεγονός που μέχρι το 1993 συνεπάγονταν ότι τα λεξιλόγια τους όφειλαν να παραμένουν μικρά. Με την εισαγωγή, όμως, της υπολεκτικής μοντελοποίησης (subword modeling) οι εφαρμογές αυτής της μορφής έχουν απαλλαγεί από τον περιορισμό αυτόν, αν και η ακρίβεια της αναγνώρισης για λέξεις που παράγονται με χρήση της υπολεκτικής μοντελοποίησης είναι λιγότερο ικανοποιητική σε σχέση με αυτή για τις λέξεις που παράγονται από μέγιστο αριθμό δειγμάτων. Ως εκ τούτου, το λεξιλόγιο πρέπει να επιλέγεται προσεκτικά, προκειμένου να αποφεύγονται οι λέξεις που μπορούν να προκαλέσουν σύγχυση.

#### **5. Ροή ομιλίας**

Τα περισσότερα συστήματα εκτέλεσης εντολών κι ελέγχου χαρακτηρίζονται από σύντομες ριπές (bursts) ομιλίας, οι οποίες αναπαριστούν την κάθε εντολή. Αυτό επιτρέπει τη χρήση της αναγνώρισης διακριτών λέξεων ή την αναγνώριση συνεχούς ομιλίας. Επίσης, η ισχύς του θορύβου, καθώς και άλλοι παράγοντες, διαδραματίζουν σημαντικό ρόλο στην επιλογή της ροής ομιλίας.

Στις τηλεφωνικές εφαρμογές οι χρήστες εισάγουν συνεχή ομιλία, αν και η αναγνώριση αναφέρεται είτε σε μεμονωμένες λέξεις, είτε σε σύντομες φράσεις.

#### **3.1.2 Το περιβάλλον ομιλίας της εκτέλεσης εντολών κι ελέγχου**

Αρκετές εφαρμογές της εκτέλεσης εντολών κι ελέγχου χαρακτηρίζονται από υψηλό θόρυβο. Τα χαρακτηριστικά του θορύβου ποικίλουν ανάλογα με την εφαρμογή:

- Ένα σύστημα κλήσεων για τηλέφωνα αυτοκινήτων οφείλει να ανταπεξέρχεται ενός καναλιού με ενθόρυβη είσοδο, του θορύβου του αυτοκινήτου, της αντίχησης, του ραδιοφώνου, του θορύβου του δρόμου, καθώς και του φαινομένου Lombard λόγω του ενθόρυβου περιβάλλοντος.

- Ένα παιχνίδι με αναγνώριση ομιλίας είναι πολύ πιθανό να “βομβαρδίζεται” από τις φωνές και τους ήχους των παιδιών που παίζουν τριγύρω. Μπορεί επίσης να συνεισφέρει τη δική του φωνή και τους μηχανικούς θορύβους στο περιβάλλον, ενώ ο ομιλητής ενδέχεται να είναι απρόσεκτος ή και να αγνοεί την ακριβή θέση του μικροφώνου.
- Ένα τμήμα του εξοπλισμού σ’ ένα εργοστάσιο θα περιτριγυρίζεται από τη βοή του υπόλοιπου εξοπλισμού, ενώ δεν αποκλείεται το περιβάλλον να χαρακτηρίζεται από διακοπόμενους κρότους και θορυβώδεις συζητήσεις ανάμεσα στους εργάτες.

Παρ’ όλ’ αυτά, όμως, τα φτωχά περιβάλλοντα ομιλίας για τις περισσότερες εφαρμογές της εκτέλεσης εντολών κι ελέγχου, η ακρίβεια δεν παύει ν’ αποτελεί πρωταρχικό στόχο. Τα προϊόντα απαιτείται να ελέγχονται επισταμένως προκειμένου να εξασφαλίζεται ότι λειτουργούν σωστά εντός των προκαθορισμένων περιβαλλόντων, καθώς και να ενισχύονται με επαρκείς διεργασίες διόρθωσης των σφαλμάτων για τις περιπτώσεις που αποτυγχάνουν να υπερκαλύψουν το θόρυβο.

### **3.1.3 Οι ανθρώπινοι παράγοντες της εκτέλεσης εντολών κι ελέγχου**

Πολλοί άνθρωποι που χρησιμοποιούν εφαρμογές της εκτέλεσης εντολών κι ελέγχου αντιμετωπίζουν την αναγνώριση ομιλίας απλά ως μια συνιστώσα της μηχανής που χειρίζονται. Γι’ αυτούς, ο ρόλος της ομιλίας περιορίζεται στη μετατροπή των προφορικών εντολών απ’ ευθείας σε πράξη. Συνήθως χαρακτηρίζονται από μικρή ανεκτικότητα στα λάθη και σχεδόν καθόλου υπομονή στην αργή ανταπόκριση ή την αναποτελεσματικότητα.

Τα συστήματα εκτέλεσης εντολών κι ελέγχου οφείλουν να είναι απλά κι εύκολα στη χρήση, με λεξιλόγιο που να εξυπηρετεί την απομνημόνευση και τη χρησιμοποίησή του. Τα συστήματα ελέγχου του περιβάλλοντος για ανθρώπους με ειδικές ανάγκες απαιτούν προσαρμογή στις ανάγκες καθενός ατόμου ξεχωριστά. Έτσι, γι’ ανθρώπους με σοβαρά προβλήματα ομιλίας ενδεχομένως θ’ απαιτηθεί προσεκτική επιλογή του λεξιλογίου, προκειμένου να επαυξηθεί η διαφοροποίηση των λέξεων, στα πλαίσια πάντα των δυνατοτήτων του ατόμου. Δεν αποκλείεται πάλι ν’ αποδειχθεί αναγκαία η ενσωμάτωση περισσότερο εκτεταμένων εγγραφών (enrollment) αν και οι ηχογραφήσεις (sessions) ίσως χρειαστεί να παραμείνουν σύντομες. Τα συστήματα, επίσης, ενδέχεται να χρειαστεί να προσαρμοστούν σε μια πιο αργή παραγωγή λέξεων ή και στην παρουσία παύσεων μεταξύ των λέξεων, γεγονός που μπορεί να σημαίνει πως ο καθορισμός των ορίων κάθε λέξης πρέπει να περιλαμβάνει μεγαλύτερα περιθώρια.

Η εκτέλεση προφορικών εντολών κι ελέγχου δεν αποκλείεται να μετατραπεί σε αντικείμενο διαφωνίας σε περίπτωση θεωρηθεί αφορμή ομαδικών απολύσεων: Αν ένα και μόνο άτομο μπορεί να ελέγξει περισσότερες μηχανές από πριν μέσω της ομιλίας του, ίσως να μην κριθεί αναγκαίο ν’ απασχολείται περισσότερο προσωπικό για το ίδιο έργο. Η υλοποίηση, όμως, τέτοιων μειώσεων στο προσωπικό εξ’ αιτίας του αυτοματισμού οφείλει να γίνεται με σύνεση, λαμβάνοντας πάντα υπ’ όψιν τους ανθρώπινους παράγοντες, όπως την ασφάλεια και το πλήθος των εργασιών που μπορεί να επιτελέσει μόνος του ένας απλός εργάτης.

### 3.1.4 Παραδείγματα της εκτέλεσης εντολών κι ελέγχου

Τα ακόλουθα παραδείγματα παρουσιάζουν τρεις διαφορετικούς τύπους εφαρμογής της εκτέλεσης εντολών κι ελέγχου. Τα συστήματα της AT&T και της NASA απεικονίζουν τη σπουδαιότητα που έχει ο προσεκτικός έλεγχος ακόμη και στις πιο απλές εφαρμογές με τις φαινομενικά ξεκάθαρες οδηγίες.

Πρόσθετα παραδείγματα συστημάτων εκτέλεσης εντολών κι ελέγχου είναι αυτά που αφορούν τα interfaces υπολογιστικών λογισμικών, τους ανθρώπους με ειδικές ανάγκες, τη διαχείριση τηλεφωνικών κλήσεων και τα καταναλωτικά προϊόντα.

#### 1. Κλήση μακρινής απόστασης. AT&T

Ένας από τους λόγους χρήσης της αναγνώρισης ομιλίας στις παραδοσιακές, υποβοηθούμενες από το λειτουργό (operator-assisted), υπηρεσίες είναι η ανάγκη να προσφερθεί μια υψηλού επιπέδου εξυπηρέτηση 24 ώρες καθημερινά, η οποία θα είναι αποδοτική τόσο για τις περιόδους υψηλής, όσο και για τις περιόδους χαμηλής χρήσης. Προκειμένου να εκτιμήσουν την επίδραση των ανθρώπινων παραγόντων στην εκτέλεση ενός τέτοιου συστήματος κλήσεων μεγάλων αποστάσεων, τα εργαστήρια Bell της AT&T προσδιόρισαν τέσσερις συνθήκες κλήσεως και μία συνθήκη διαφυγής. Για να ελαχιστοποιηθούν τα σφάλματα, στις συνθήκες κλήσεως εκχωρήθηκαν οι ακόλουθες, ακουστικά διαφορετικές, λέξεις ή φράσεις της απάντησης του χρήστη:

##### ΣΥΝΘΗΚΗ ΚΛΗΣΗΣ

χρέωση (collect)  
χρέωση κάρτας (calling card)  
χρέωση τρίτου ατόμου (third-party billing)  
άνθρωπος-προς-άνθρωπο (person-to-person)  
λειτουργός (operator)

##### ΑΠΑΝΤΗΣΗ ΧΡΗΣΤΗ

χρέωση (collect)  
χρέωση κάρτας (calling card)  
τρίτο νούμερο (third number)  
άνθρωπος (person)  
λειτουργός (operator)

Η λίστα των επιλογών ξεκινούσε με το απλό ερώτημα: “Έχετε καλέσει την AT&T. Παρακαλώ πείτε...”. Ο αρχικός σχεδιασμός χρησιμοποιούσε την ανεξάρτητη από τον ομιλητή αναγνώριση διακριτών λέξεων και υποβλήθηκε σε έλεγχο σχεδόν 75000 κλήσεων από τον οποίο προέκυψε ότι:

“ *Είναι εξαιρετικά δύσκολο, αν όχι αδύνατο, να αναγκάσεις τους πραγματικούς συνδρομητές μιας τέτοιας υπηρεσίας να εκφωνούν μόνο τις επιτρεπτές λέξεις εισόδου.*” (Jay Wilpon, Lawrence Rabiner, Chin-Hui Lee & E.R. Goldman, Researchers, AT&T Bell Laboratories, “ Automatic recognition of keywords in unconstrained speech using hidden Markov models”, 1990, p. 1870).

Η πράξη απέδειξε ότι μόνο το 65% του συνόλου των χρηστών ακολούθησε το συγκεκριμένο μοντέλο, ενώ περίπου το 20% των χρηστών προσέθεταν επιπλέον ομιλία (extraneous speech) στις αποκρίσεις τους (π.χ. “θέλω να κάνω μια συλλογική κλήση (collect call) παρακαλώ”). Το δίλημμα που καταγράφηκε στ’ αποτελέσματα των μελετών των εργαστηρίων Bell, ήταν αν άξιζε τον κόπο να δημιουργηθεί μια πραγματικά πολύπλοκη γραμματική για ένα τόσο απλό έργο ή αν θα έπρεπε να εγκαταλειφθούν οι εφαρμογές που απευθύνονται σε περιστασιακούς χρήστες. Αναγνωρίζοντας ότι η χρήση επιπλέον ομιλίας θα ήταν κοινό και αναπόφευκτο πρόβλημα σε κάθε αντίστοιχη τηλεφωνική εφαρμογή αποφασίστηκε το σύστημα να υλοποιηθεί εκ νέου, χρησιμοποιώντας αυτή τη φορά τον

εντοπισμό κάποιων λέξεων-κλειδιά. Το τελικό σύστημα, η Επεξεργασία Κλήσεων μέσω Αναγνώρισης Φωνής ( VRCP: Voice Recognition Call Processing), περιελάμβανε μοντέλα για τις λέξεις-κλειδιά, τη σιωπή (silence), τους θορύβους μετάδοσης και την επιπλέον ομιλία. Με βάση τις απαντήσεις που δόθηκαν στο δοκιμαστικό έλεγχο κατασκευάστηκαν εξ' αρχής κάποια μοντέλα που δε χρησιμοποιούσαν τις λέξεις-κλειδιά, εξασφαλίζοντας έτσι ότι η πιο συχνά επαναλαμβανόμενη επιπλέον ομιλία αναπαριστάνονταν επαρκώς. Τελευταία, τα μοντέλα της επιπλέον ομιλίας επαναδημιουργήθηκαν χρησιμοποιώντας μια περισσότερο λεπτομερή, και υψηλότερης τάξης, ακουστική ανάλυση των γενικών προτύπων ομιλίας, όπως αυτά καταγράφονται στις βάσεις δεδομένων για την ομιλία.

Ένα άλλο τηλεπικοινωνιακό σύστημα εκτέλεσης εντολών κι ελέγχου είναι αυτό που περιγράφεται από τον Lenning (1989).

## **2. Έλεγχος φωνής για άτομα με αναπηρία. Montefiore Hospital**

Το VARRS (Voice Activated Robotic Retrofit System) είναι ένα πρωτότυπο σύστημα που δίνει τη δυνατότητα σ' ένα κλινήρες (bed-bound) άτομο να ελέγχει τη θέση του κρεβατιού του με τη φωνή. Το σύστημα σχεδιάστηκε για να χρησιμοποιείται από ηλικιωμένα άτομα με προβλήματα όρασης ή παραλυσία, που είναι αδύνατο να χρησιμοποιούν χειροκίνητους ελέγχους.

*“Η αδυναμία αυτών των ασθενών να εκτελέσουν απλές ρυθμίσεις της θέσης του κεφαλιού ή του ποδιού ενός κρεβατιού τους δημιουργεί την αίσθηση μειωμένης ανεξαρτησίας, επιφορτίζοντας με μη αναγκαία βάρη το νοσηλευτικό και το βοηθητικό προσωπικό”.* (Marc Liang, MD & Krishna Narayanan, MD, Montefiore Hospital, “The application of voice recognition to robotic positioning of a hospital bed”, 1989, p. 31).

Το VARRS διαθέτει ένα μικρό λεξιλόγιο που περιλαμβάνει: ένα password, δύο λέξεις για τα τμήματα του κρεβατιού (κεφάλι και πόδια), πέντε κατευθυντήριους όρους (πάνω (up), κάτω (down), ανασήκωσε (tilt), πίσω (backward) και μπροστά (forward)) και τρεις προγραμματισμένες θέσεις για το κρεβάτι (θέση τηλεόρασης, φαγητού και διαβάσματος). Οι μελέτες των ανθρώπινων παραγόντων επέβαλλαν τη χρήση της εξαρτημένης από τον ομιλητή εκπαίδευσης του συστήματος, ταυτόχρονα με την εφαρμογή ενός password, προκειμένου ν' αποφεύγεται η απόκριση του συστήματος σε άλλες φωνές. Με τον τρόπο αυτό, η κλίση αποκρίνεται μόνο στη φωνή του ασθενή, εκτελώντας μονάχα εκείνες τις εντολές των οποίων προηγείται άμεσα το ειδικό password. Υπάρχει, επίσης, μια εντολή ασφαλείας (fail-safe command), η οποία διακόπτει την κίνηση του κρεβατιού εντός ενός χιλιοστού του δευτερολέπτου αφ' ότου εκφωνηθεί. Η εντολή ασφαλείας προστέθηκε ως μέτρο υποστήριξης (backup) στην περίπτωση εσφαλμένης αναγνώρισης.

Από τη στιγμή που σχεδιάστηκε το σύστημα VARRS έχουν αναπτυχθεί κι άλλοι έλεγχοι για τα νοσοκομειακά κρεβάτια, όπως είναι το Hill-Rom's Enhancemate.

## **3. Έλεγχος των τηλεοπτικών κάμερων κλειστού κυκλώματος. NASA Jet Propulsion Laboratory**

Ο έλεγχος μέσω φωνής του εξοπλισμού των διαστημικών οχημάτων παρέχει στους αστροναύτες τη δυνατότητα να επιτελούν ταυτόχρονα πολλαπλές και πολύπλοκες

λειτουργίες.Μια απ' αυτές είναι ο έλεγχος των τηλεοπτικών κάμερων κλειστού κυκλώματος που καταγράφουν τη λειτουργία των ρομποτικών επισκευαστικών βραχιόνων στα διαστημόπλοια.Στην περίπτωση αυτή, ένας και μόνο αστροναύτης πρέπει να χειρίζεται ταυτόχρονα και τους επισκευαστικούς βραχιόνες και τις κάμερες.Το Jet Propulsion Laboratory μαζί με το Lockheed σχεδίασαν ένα φωνητικό σύστημα, το επονομαζόμενο VCS, για την εκτέλεση εντολών κι ελέγχου από τις κάμερες κλειστού κυκλώματος.Το σύστημα υποβλήθηκε σε δοκιμαστικό έλεγχο στο διαστημόπλοιο Orbiter Discovery τον Οκτώβριο του 1990.

Οι τριάντα-τρεις (33) από τις σαράντα-μία (41) εντολές που επιλέγησαν για το σύστημα ομιλίας χρησιμοποιούνταν για τις λειτουργίες ελέγχου των διακοπών (switches) στις χειροκίνητα ελεγχόμενες κάμερες.Οι αστροναύτες μπορούσαν να επιλέγουν και να λειτουργούν τις κάμερες, να ενεργοποιούν την εγγραφή εικόνας από μία κάμερα και να φυλάσσουν τις κάμερες.Οι υπόλοιπες εντολές αναφέρονταν στην ταυτοποίηση των χρηστών, καθώς και τη χωρική ταχύτητα και μετατόπιση των κάμερων και είχαν τη μορφή: “περισσότερο (more)”, “όχι τόσο πολύ (too much)”, “σιγά-σιγά (easy)”.Το λεξιλόγιο σχετικά με την ταχύτητα και τη μετατόπιση προσέθετε ευελιξία στον έλεγχο των χωρικών μετακινήσεων, παρέχοντας ταυτόχρονα στους αστροναύτες ένα μέσο για την επανόρθωση των σφαλμάτων αναγνώρισης.Η εξαρτημένη από το χρήστη αναγνώριση εφαρμόστηκε αφού προηγουμένως δοκιμάστηκε προσεκτικά κατά τη διάρκεια επαναλαμβανόμενων ελέγχων πριν από την πτήση.Οι δοκιμές πραγματοποιήθηκαν σε περιβάλλον που προσομοίαζε επακριβώς τις συνθήκες θορύβου στο διαστημόπλοιο, έτσι ώστε τα μοντέλα ομιλητών να περιέχουν τον κατάλληλο περιβαλλοντικό θόρυβο (background noise).

*“Κατά τη διάρκεια των δοκιμών, πρωταρχικός στόχος ήταν η εξοικείωση του πληρώματος με την τεχνολογία και τη λειτουργία του VIS, καθώς και η εξαγωγή ενός ικανοποιητικού συνόλου ιχνών (templates) που θα μπορούσε να χρησιμοποιηθεί στην πτήση.”* (George Solarar, Project manager, NASA Lyndon Space Center & Marc Sommers, Project Engineer, Lockheed Engineering & Sciences Company, “Space Shuttle voice recognition flight experiment”, 1991, p.14).

Τις δοκιμές ακολούθησαν τρεις ξεχωριστές προσομοιώσεις πτήσεων, σχεδιασμένες για τη διόρθωση των σφαλμάτων αναγνώρισης και την αναβάθμιση των προβληματικών ιχνών.Πριν από τον έλεγχο στο διαστημόπλοιο η αναγνώριση είχε επιτευχθεί σχεδόν στο 100% για καθένα από τους αστροναύτες.Η αναγνώριση κατά τη διάρκεια της πτήσης κατέδειξε την ανάγκη ενσωμάτωσης στο VIS μιας μεθόδου για τον αυτόματο έλεγχο της απολαβής (gain), προκειμένου ν'αντισταθμίζονται οι μεταβολές στην ένταση της εισόδου.

Ένα παρόμοιο σύστημα που αναπτύχθηκε στο GEIRCA's Advanced Technology Laboratory περιγράφεται στο Hackenberg (1987).

### **3.2 Εισαγωγή Δεδομένων (Data Entry)**

Η εισαγωγή δεδομένων μέσω φωνής επιτελεί την άμεση πληροφορική είσοδο δεδομένων στο υπολογιστικό σύστημα που θα τα επεξεργαστεί.Πριν τα τέλη της δεκαετίας του '80, οι εφαρμογές της εισαγωγής δεδομένων ήταν το κύριο αντικείμενο των εμπορικών συστημάτων αναγνώρισης ομιλίας.Οι πρώτες εφαρμογές της περιστρέφονταν γύρω από τις ανάγκες του στρατού, των εργοστασίων και των αποθηκών.Αυτό που κέντριζε το ενδιαφέρον αυτών των πελατών στην αναγνώριση ομιλίας ήταν η δυνατότητα για είσοδο δεδομένων χωρίς τη χρήση χειριών.



Μέχρι τα τέλη του '80 η χρήση της αναγνώρισης ομιλίας ως μέσον για την εισαγωγή δεδομένων επεκτάθηκε και σε άλλες βιομηχανίες και λειτουργίες:

- Κατασκευές Εισαγωγή δεδομένων σχετικών με την τοποθεσία
- Οδοντιατρική Καταγραφή των εξετάσεων ρουτίνας
- Εκπαίδευση Υποβοηθούμενη από υπολογιστή διδασκαλία με χρήση πολυμέσων
- Μηχανικός σχεδιασμός Εισαγωγή των απαιτήσεων σε κατασκευαστικά υλικά
- Οικονομία Καταγραφή ασφαλιστικών συμβολαίων
- Δημόσια διοίκηση Συλλογή μετρήσεων, προστασία δασών από δασοφύλακες
- Υγεία Προγραμματισμός των ραντεβού των ασθενών, Διαγράμματα υγείας των ασθενών, Καταγραφή συνταγών από τους φαρμακοποιούς, Ανάλυση δειγμάτων και ακτινογραφιών από τους εργαστηριακούς τεχνικούς
- Ξενοδοχεία/Πανσιόν Καταγραφή της ώρας άφιξης και της παραμονής από τους υπαλλήλους
- Ασφάλειες Επιθεώρηση των κατεστραμμένων ιδιοκτησιών
- Συστήματα γραμματείας Συμπλήρωση αιτήσεων
- Διανομή πακέτων Ταξινόμηση πακέτων
- Εμπόριο Καταγραφή πληροφοριών για τα επιστρεφόμενα προϊόντα
- Τηλεφωνικές επισκευές Έλεγχος των τηλεφωνικών γραμμών από το προσωπικό ανάβασης στους τηλεφωνικούς στύλους

Η εισαγωγή δεδομένων μέσω φωνής συχνά αναφέρεται ως “λειτουργία για απασχολημένα χέρια (hands-busy) ή απασχολημένα μάτια (eyes-busy)”, επειδή επιτρέπει στους εργαζομένους να επιτελούν χειρωνακτικές εργασίες ταυτόχρονα με την προφορική εισαγωγή δεδομένων.

*“Από τη στιγμή που παίρνει πολύτιμο χρόνο σ’έναν χειρουργό να πάρει τα μάτια του από το μικροσκόπιο και να συγκεντρωθεί εκ νέου για να καταγράψει δεδομένα σ’ένα χαρτί, η χρήση της αναγνώρισης ομιλίας αυξάνει την αποδοτικότητα αξιοσημείωτα, ενώ περιορίζει τα συμπτώματα πονοκεφάλων και υπερκόπωσης των ματιών στους χειρουργούς.”* (Lick Cook, Staff Writer, Managing Automation, “Voice recognition: Time for another look?”, 1990)

Η αναγνώριση ομιλίας είναι ο μοναδικός τρόπος εισόδου δεδομένων που επιτρέπει στους χρήστες να διατηρούν τα μάτια, τα χέρια και τη σκέψη τους συγκεντρωμένα στο έργο που επιτελούν την ίδια στιγμή που εισάγουν δεδομένα.

Στις κατασκευαστικές και αποθηκευτικές εφαρμογές η ομιλία μπορεί να συνδυαστεί με άλλες τεχνολογίες για την περισυλλογή συμπληρωματικής πληροφορίας που διαφεύγει. Η απελευθερωμένη από τη χρήση των χειροκίνητων συσκευών που απαιτούνται για την είσοδο δεδομένων άλλης μορφής. Για παράδειγμα, το bar code είναι το πιο αποδοτικό μέσο για την περισυλλογή της κωδικοποιημένης πληροφορίας που σχετίζεται με κάποιο προϊόν. Όμως, δεν έχει τη δυνατότητα καταγραφής καταστατικής πληροφορίας, όπως είναι η κατάσταση των υλικών κατά την άφιξή τους στο σημείο παραλαβής όπου στάλθηκαν συσκευασμένα, ή του τύπου της πλατφόρμας που χρησιμοποιήθηκε για τη μεταφορά τους. Ο συνδυασμός της

αναγνώρισης ομιλίας με το barcode επιτρέπει τη συνάθροιση όλων αυτών των πληροφοριών, διευκολύνοντας ταυτόχρονα την παραγωγή barcodes για τα αντικείμενα που δεν έχουν.

### **3.2.1 Ζητήματα σχεδιασμού για την εισαγωγή δεδομένων**

#### **1. Μέγεθος κι επιλογή του λεξιλογίου**

Οι παραδοσιακές εφαρμογές της εισαγωγής δεδομένων χαρακτηρίζονταν από λεξιλόγια μικρότερα των πενήντα λέξεων, μετά την πρόσφατη ανάπτυξη, όμως, των λεξιλογικών δυνατοτήτων της αναγνώρισης ομιλίας, οι εφαρμογείς έχουν συνειδητοποιήσει πως οι εφαρμογές που συνοδεύονται από περιορισμένα λεξιλόγια δεν επιτελούν παρά ένα μικρό τμήμα μιας δυναμικής εφαρμογής εισαγωγής δεδομένων.

Στις περισσότερες περιπτώσεις, οι απαιτήσεις σε λεξιλόγιο περιλαμβάνουν κάποιους τυποποιημένους όρους (π.χ ψηφία, ναι, όχι, ακύρωση), μαζί με τις λέξεις που συνοδεύουν κάθε εφαρμογή. Συχνά, το μεγαλύτερο ποσοστό ενός λεξιλογίου είναι χαρακτηριστικό της κάθε εφαρμογής, γεγονός που σημαίνει ότι τα περισσότερα προϊόντα με ενσωματωμένα λεξικά δεν επαρκούν για να ικανοποιήσουν τις λεξιλογικές ανάγκες κάθε εφαρμογής. Αυτός είναι ο κύριος λόγος διαφωνίας στην επιλογή ενός προϊόντος με εργαλεία ανάπτυξης λεξιλογίου.

#### **2. Μεταγλώττιση**

Οι εφαρμογές εισαγωγής δεδομένων συχνά χαρακτηρίζονται από ειδικές μεταφραστικές απαιτήσεις. Στις περισσότερες περιπτώσεις η επιθυμητή έξοδος μιας αναγνώρισης δεν περιορίζεται στην τυποποιημένη γραπτή αναπαράσταση των λέξεων που έχουν εισαχθεί. Μια βάση δεδομένων με τις διάφορες ατέλειες, για παράδειγμα, μπορεί ν'αναπαριστά τη φράση "χαραγμένος πίνακας" με τον κώδικα F435. Όπως θα δούμε, το σύστημα συντήρησης για την παγίδευση ατμού της Eastman Kodak (Eastman Kodak's steam trap maintenance system) αποτελεί παράδειγμα των προβλημάτων που μπορούν να προκύψουν όταν ένα σύστημα ομιλίας πρέπει να επικοινωνήσει ευθέως με το βασικό λογισμικό.

Δεν αποκλείεται μεμονωμένες λέξεις ή φράσεις να υποκαθιστούν πολύπλοκες λειτουργίες του πληκτρολογίου, διαδοχικά click του ποντικιού να συνενώνονται σε μία λέξη ή φράση, ή ακόμα και σε μια προκαθορισμένη ορθογραφία (spelling). Αποδοχές αυτής της μορφής δεν αποκλείεται ν'αποτελούν τις κατευθυντήριες γραμμές για τη δημιουργία ενός φωνητικού συστήματος. Μια Αμερικάνικη εταιρία κατασκευής φορτηγών, για παράδειγμα, αδυνατούσε να πραγματοποιήσει ακριβή ποιοτικό έλεγχο των ατελειών της παραγωγής της, επειδή κάποια κρίσιμα δεδομένα καταγράφονταν με διάφορες ορθογραφίες, αν και αντιστοιχούσαν στα ίδια ελαττώματα. Έτσι, το ίδιο ελάττωμα στα μπροστινά φρένα (front-brakes) που παρατηρούταν σε μια σειρά από φορτηγά καταγράφονταν ως "μπροστινά φρένα", "μπρ.φρνα", "φρένα, εμπρός", και "μπρ.φρν", προκαλώντας την στατιστική καταγραφή τεσσάρων διαφορετικών ελαττωμάτων. Η αναγνώριση ομιλίας, οι οθόνες επαφής (touch screens) και η χρήση ποντικιού εξάλειψαν το πρόβλημα αυτό, διατηρώντας μια αντιστοιχία ένα-προς-ένα μεταξύ της εισόδου και της μετάφρασης. Από τις τεχνολογίες αυτές η αναγνώριση ομιλίας είναι η μόνη που παραμένει ευέλικτη καθώς μεγαλώνει το λεξιλόγιο.

### **3. Γλωσσολογική δομή**

Οι περισσότερες εφαρμογές της εισαγωγής δεδομένων είναι υψηλά δομημένες και αναπαριστώνται καλύτερα μέσω των γραμματικών περιορισμένων καταστάσεων και άλλων ισοδυνάμων. Η χρήση μιας δομημένης γλώσσας έχει δύο πλεονεκτήματα, πολύ σημαντικά για τις περισσότερες εφαρμογές της εισαγωγής δεδομένων: επαυξημένη ακρίβεια (enhanced accuracy) και ταχύτερη επεξεργασία. Τα πλεονεκτήματα αυτά πηγάζουν από τη μειωμένη περιπλοκή (perplexity). Η ακρίβεια και η επιτάχυνση αυξάνουν με τη μείωση του αριθμού των λέξεων στο ενεργό λεξιλόγιο, ενώ η ακρίβεια βελτιώνεται, επίσης, καθώς οι λέξεις του ενεργού λεξιλογίου γίνονται περισσότερο διακριτές μεταξύ τους.

### **4. Μοντέλο ομιλητή**

Για τις εφαρμογές της εισαγωγής δεδομένων με περιορισμένο λεξιλόγιο, υπάρχει μια αντιστάθμιση στο μοντέλο του ομιλητή μεταξύ ακρίβειας και ευκολίας στη χρήση. Η εξαρτώμενη από τον ομιλητή μοντελοποίηση, η οποία επικεντρώνει την αναγνώριση στα πρότυπα ομιλίας του κάθε χρήστη, επαυξάνει την ακρίβεια, επιτρέποντας την εγγραφή (enrollment) ακόμη και στις περιπτώσεις που διαχωρίζεται σε πολλαπλά τμήματα. Τα ανεξάρτητα από τον ομιλητή μοντέλα είναι εύκολα στη χρήση αλλά λιγότερο ακριβή, καθώς αφ' ενός δεν επικεντρώνονται στον ομιλητή και αφ' ετέρου ενδέχεται να μην περιέχουν επαρκή αναπαράσταση του περιβάλλοντος ομιλίας.

Καθώς τα λεξιλόγια των εφαρμογών αυτών μεγαλώνουν, γίνεται δυσκολότερη η προσέλευση χρηστών στα εκτεταμένα προγράμματα εγγραφής που απαιτούνται. Τα ανεξάρτητα από τον ομιλητή συστήματα, καθώς και τα προσαρμοσμένα στον ομιλητή συστήματα με τη σταδιακή προσαρμογή (on-the-fly adaptation), μπορούν να χρησιμοποιηθούν στις περιπτώσεις που οι έλεγχοι υπό πραγματικές συνθήκες έχουν δείξει ότι αποδίδουν ικανοποιητικά εντός του περιβάλλοντος ομιλίας.

### **5. Ροή ομιλίας**

Όπως με τα συστήματα της εκτέλεσης εντολών κι ελέγχου, έτσι και οι εφαρμογές της εισαγωγής δεδομένων χαρακτηρίζονται από σύντομες ριπές ομιλίας. Η επιλογή μεταξύ αναγνώρισης συνεχούς ομιλίας και αναγνώρισης διακριτών λέξεων εξαρτάται κυρίως από παράγοντες όπως ο περιβαλλοντικός θόρυβος. Σε γενικές γραμμές, πάντως, τα συστήματα συνεχούς ομιλίας παρουσιάζουν μεγαλύτερη ανοχή στο θόρυβο από τα συστήματα διακριτών λέξεων, με την ποιότητα να ποικίλλει ανάλογα με το προϊόν, το λεξιλόγιο, τους χρήστες και το περιβάλλον.

#### **3.2.2 Το περιβάλλον ομιλίας της εισαγωγής δεδομένων**

Τα περισσότερα από τα περιβάλλοντα εφαρμογής της εισαγωγής δεδομένων είναι ιδιαίτερα επιβαρυνμένα. Χαρακτηρίζονται από φυσικές ιδιαιτερότητες (όπως σκόνη, υψηλές θερμοκρασίες, δονήσεις και ατυχήματα) ικανές να καταστρέψουν τον υπολογιστικό εξοπλισμό. Ένας κατασκευαστής, του οποίου οι υπάλληλοι αρνούσαν να χρησιμοποιήσουν τα πλαστικά καλύμματα των ηλεκτρολογίων, αναγκάστηκε ν' αντικαθιστά τον εξοπλισμό κάθε δύο ή τρεις μήνες.

Οι συνθήκες αυτές συνάγουν στη χρήση ανθεκτικού, εργοστασιακού εξοπλισμού ή στη μετάδοση μέσω ραδιοσυχνοτήτων. Η μετάδοση μέσω ραδιοσυχνοτήτων επιτρέπει στο υπολογιστικό hardware να φυλάσσεται σε ασφαλέστερη και καθαρότερη περιοχή. Επίσης, επιτρέπει μεγαλύτερη κινητικότητα στους χρήστες, οι οποίοι ίσως χρειαστεί να κινηθούν γύρω από αντικείμενα κατά τη διάρκεια επιθεωρήσεων ή αναλύσεων. Η ραδιομετάδοση, όμως, ενδέχεται να παρουσιάσει παρεμβολές από τις άλλες συχνότητες μετάδοσης του περιβάλλοντος.

Κάποια προϊόντα σχεδιάζονται ειδικά για ιδιαίτερος ενθόρυβα περιβάλλοντα ομιλίας, ενσωματώνοντας ειδικούς αλγόριθμους χειρισμού των θορύβων. Αν και η ικανότητα των ανεξάρτητων από τον ομιλητή μοντέλων στο χειρισμό των θορύβων αυξάνει συνεχώς, εντούτοις τα περισσότερα προϊόντα που προορίζονται για ιδιαίτερος ενθόρυβα και πειστικά περιβάλλοντα είναι εξαρτημένα από τον ομιλητή. Η εξαρτημένη από τον ομιλητή τεχνολογία είναι περισσότερο ακριβής είναι περισσότερο ακριβής στους υψηλούς θορύβους, επειδή διαθέτει τη δυνατότητα μοντελοποίησης του ομιλητή, του περιβαλλοντικού θορύβου και του φαινομένου Lombard με τρόπο ιδιαίτερα πειστικό.

Όλα τα προϊόντα πρέπει να ελέγχονται εντός του προκαθορισμένου περιβάλλοντος εφαρμογής τους, προκειμένου να πιστοποιείται η ικανότητά τους να αντεπεξέρχονται στις αναμενόμενες συνθήκες θορύβου. Αυτό είναι ιδιαίτερα κρίσιμο για τα περιβάλλοντα υψηλού θορύβου και τα κινητά συστήματα. Οι εφαρμογές που απαιτούν κινητά ή φορητά συστήματα, συνήθως προορίζονται για χρήση μέσα σε πλήθος διαφορετικών περιβαλλόντων ομιλίας, συμπεριλαμβανομένων των εξωτερικών χώρων.

### **3.2.3 Οι Ανθρώπινοι παράγοντες της εισαγωγής δεδομένων**

Οι άνθρωποι που χρησιμοποιούν παραδοσιακές εφαρμογές της εισαγωγής δεδομένων στις εργοστασιακές γραμμές παραγωγής, στις αποθήκες και τα λιμάνια σπάνια είναι πεπερασμένοι στη χρήση υπολογιστών. Ενδέχεται, μάλιστα, να αισθάνονται άβολα με τους υπολογιστές. Η αποδοχή, όμως, μπορεί να αυξηθεί με τη διαμόρφωση ενός interface που θα χαρακτηρίζεται από οικίες συνιστώσες. Αν, για παράδειγμα, ένα έργο απαιτεί τη συμπλήρωση μιας οικείας φόρμας, η εξοικείωση του χρήστη θα επαυξηθεί αν η οθόνη σχεδιαστεί έτσι ώστε ν'αναπαριστά τη φόρμα. Αυτός ο τρόπος αναπαράστασης δίνει έμφαση στο στόχο μιας εισαγωγής δεδομένων. Ομοίως, μια εφαρμογή επιθεώρησης του ποιοτικού ελέγχου ή της γραμμής παραγωγής μπορεί να δομηθεί με τρόπο που να θυμίζει κάποιες γνωστές φόρμες, έστω κι αν τα ζητούμενα της φόρμας έχουν αντικατασταθεί από ερωτήσεις προς το χρήστη.

Ο προορισμός της χρήσης μιας εφαρμογής και του ρόλου της αναγνώρισης ομιλίας ως προς το στόχο του έργου συντελεί στο σχηματισμό θετικής εικόνας πριν από την υλοποίηση του συστήματος. Σε περίπτωση που η προεργασία αυτή γίνεται από τους επιβλέποντες ή άλλα μέλη μιας ομάδας, η αποδοχή του συστήματος από τους χρήστες θα είναι μεγαλύτερη. Η συμμετοχή κάποιων από τους χρήστες στο σχεδιασμό μιας εφαρμογής συντελεί στον καλύτερο χειρισμό των ζητημάτων που άπτονται των ανθρώπινων παραγόντων, της αποδοχής από τους χρήστες και της ολοκλήρωσης του έργου. Οι χρήστες, για παράδειγμα, ενδέχεται να προτιμούν την τοποθέτηση του μικροφώνου σε κάποιο συγκεκριμένο σημείο ή να έχουν ήδη διαμορφώσει κάποιο πρότυπο λειτουργίας του έργου.

Πολλές εφαρμογές της εισαγωγής δεδομένων απαιτούν οι επερωτήσεις προς τους χρήστες να γίνονται μέσω σύνθεσης ομιλίας ή ψηφιακά ηχογραφημένης ομιλίας. Η λειτουργία των επερωτήσεων αναφέρεται στην αίτηση συγκεκριμένων δεδομένων και την επαλήθευση μιας αναγνωρισμένης εισόδου. Η ποιότητα και ο τόνος (pitch) των επερωτήσεων απαιτείται να διαφέρει εμφανώς από τον περιβαλλοντικό θόρυβο, όπως και να είναι αποδεκτός από τους

χρήστες. Η έκταση και ο ρυθμός ομιλίας τους πρέπει ν'ακολουθεί τις απαιτήσεις του έργου, καθώς και την εμπειρία των χρηστών με το σύστημα. Οι γραμμές παραγωγής ίσως απαιτούν γρήγορο ρυθμό, ενώ τα διαγράμματα υγείας δίπλα στα κρεβάτια όχι. Σε περίπτωση που συνυπάρχουν αρχάριοι κι έμπειροι χρήστες ενδέχεται ν'απαιτηθεί η δημιουργία δύο ή περισσότερων interfaces, τα οποία θ'αντανακλούν τα προσόντα των διαφόρων χρηστών. Το σύστημα KorTeam για τα διαγράμματα υγείας δίπλα στα κρεβάτια αποτελεί παράδειγμα ενός συστήματος με πολλαπλά interfaces.

Οι εφαρμογές που χρησιμοποιούν οθόνες για την έξοδο των πληροφοριών πρέπει να διαθέτουν ευκρινείς οπτικές επερωτήσεις, προσαρμοσμένες στη φωτεινότητα, την απόσταση μεταξύ χρήστη και οθόνης και τους υπόλοιπους οπτικούς παράγοντες του περιβάλλοντος ομιλίας. Οι οθόνες που δομούνται με τη δομή φορμών πρέπει να παρουσιάζουν την παρούσα θέση ευκρινώς, καθώς και να διαθέτουν εργαλεία διευκόλυνσης της πλοήγησης για τη διόρθωση των σφαλμάτων ή την επανεξέταση.

Στις εργοστασιακές και αποθηκευτικές εφαρμογές, οι χρήστες συνήθως βρίσκονται σε κίνηση. Ως εκ τούτου τα συστήματα αναγνώρισης δεν πρέπει να προσδένονται σε κάποιο εξάρτημα του εξοπλισμού, εκτός και αν αυτό μεταφέρεται από το χρήστη. Ενδέχεται, επίσης, ν'απαιτηθεί η ενσωμάτωση μικροφώνου πάνω στον προστατευτικό εξοπλισμό του προσώπου ή του κεφαλιού.

### **3.2.4 Παραδείγματα της εισαγωγής δεδομένων**

Τα παρακάτω παραδείγματα απεικονίζουν την ποικιλομορφία των εφαρμογών της εισαγωγής δεδομένων. Πρόσθετα παραδείγματα εφαρμογής αποτελούν τα συστήματα που έχουν αναπτυχθεί για εργοστασιακές και οικονομικές χρήσεις.

#### **1. Συντήρηση παγίδων ατμού (stream trap maintenance). Eastman Kodak Company**

Οι παγίδες ατμού είναι εξαρτήματα των συστημάτων θέρμανσης ατμού. Υποβοηθούν την απομάκρυνση του υδροποιημένου νερού από το σύστημα θέρμανσης, παγιδεύοντας τον ατμό ώστε να μη χάνεται ενέργεια. Όταν μια παγίδα ατμού χαλάσει το νερό υδροποιείται πολύ γρήγορα με αποτέλεσμα να μειώνεται η απόδοση του συστήματος θέρμανσης. Στα εργοστάσια, οι παγίδες ατμού πρέπει να παραμένουν συνεχώς σε λειτουργία, διαφορετικά παρατηρούνται απώλειες στην παραγωγή.

Η εταιρία Eastman Kodak εφάρμοσε την αναγνώριση ομιλίας στη διεργασία συντήρησης των παγίδων ατμού:

*“Οι μηχανικοί εργάζονταν σε ομάδες. Συνήθως ένας μηχανικός πραγματοποιούσε τους ελέγχους ενώ οι υπόλοιποι κατέγραφαν τ'αποτελέσματα και βοηθούσαν μετακινώντας τον εξοπλισμό. Στις περισσότερες περιπτώσεις χρειαζόνταν ανεμόσκαλες (ladders) ή ανελκυστήρες για να προσεγγιστούν οι παγίδες.*

*Ο κύριος λόγος εισαγωγής των φωνητικών υπολογιστών είναι η παροχή στους μηχανικούς της δυνατότητας να εργάζονται ανεξάρτητα χωρίς χαρτί και μολύβι. Η φωνητική είσοδος των δεδομένων επιτρέπει στους μηχανικούς να διατηρούν τα χέρια τους διαθέσιμα για το χειρισμό του εξοπλισμού, την ίδια στιγμή που εισάγουν δεδομένα. Ένα άλλο κορυφαίο πλεονέκτημα είναι η εξάλειψη της υπαλληλικής νοοτροπίας, τονώνοντας την εργασία.” (Lonnie Lood & Joh Lewis, Eastman Kodak Company, “Integrated Voice data System for Central Stem trap maintenance”, 1991, p.94).*

Εν συντομία, επιτεύχθηκε καλύτερη αξιοποίηση του προσωπικού, εισαγωγή δεδομένων απαλλαγμένη από τη χρήση των χεριών και ταχύτερος εκσυγχρονισμός των βάσεων δεδομένων συντήρησης.

Για την υλοποίηση των στόχων αυτών χρησιμοποιήθηκε ένα φορητό σύστημα αναγνώρισης ομιλίας, το οποίο πέρασε γρήγορα από το σχεδιασμό στην πράξη, όπου και αποδείχθηκε η χρησιμότητά του. Επιπλέον επέτρεψε στους εργαζόμενους να διατηρούν ελευθερία κινήσεων, αφού οι επερωτήσεις που τους απηύθυνε μέσω ψηφιακής φωνής εξάλειψαν την ανάγκη καταφυγής σε συμβουλευτικές οθόνες κατά τη διάρκεια της εισαγωγής δεδομένων. Ως εκ τούτου η εισαγωγή δεν απαλλάχθηκε μόνο από τη χρήση των χεριών αλλά και από τη χρήση των ματιών.

Η Kodak καταμέτρησε πλήθος προβλημάτων, όπως είναι φυσιολογικό σε τέτοιες εργοστασιακές εφαρμογές. Ο περιβαλλοντικός θόρυβος μεταβαλλόταν από ελάχιστη σε μέγιστη ένταση, καταδεικνύοντας διάφορες τροποποιήσεις, του μικροφώνου μετά την υλοποίηση του συστήματος. Τα καλώδια που συνέδεαν τα ακουστικά με τη μονάδα ομιλίας αποδείχθηκαν πολύ εύθραυστα για το συγκεκριμένο εργοστασιακό περιβάλλον, αναγκάζοντας τους μηχανικούς συντήρησης να χρησιμοποιούν προστατευτικά καλύμματα για τα καλώδια. Επιπλέον οι μηχανισμοί συντήρησης δεν είχαν ιδιαίτερη εξοικείωση με τους υπολογιστές, με αποτέλεσμα να απαιτηθούν πολλαπλές τροποποιήσεις στο interface του χρήστη για να διευκολυνθεί η εξοικείωσή τους. Τέλος, η βάση δεδομένων που προορίζονταν για την αποθήκευση των δεδομένων συντήρησης δεν είχε σχεδιαστεί για την άμεση είσοδο δεδομένων από PC. Προκειμένου να επιλυθεί το πρόβλημα αυτό οι σχεδιαστές υλοποίησαν ένα “ρομπότ” λογισμικού, το οποίο διάβαζε τα αρχεία ομιλίας από το PC παράγοντας προσομοιωμένα δεδομένα πληκτρολογίου για τη βάση.

## **2. Διαγράμματα υγείας δίπλα στα κρεβάτια (Bedside charting).** **KorTeam International**

Στα νοσοκομειακά περιβάλλοντα παρατηρείται μια διαρκώς αυξανόμενη ανάγκη βελτίωσης των μεθόδων καταγραφής της θεραπείας των ασθενών και επιτάχυνσης επεξεργασίας των καταγραφών αυτών, με αποτέλεσμα να έχει μετατραπεί σε μόδα η εισαγωγή αυτοματισμών στην υγεία συνολικά. Χρησιμοποιώντας τεχνικές επισκόπησης και έρευνας της αγοράς, η KorTeam International κατάληξε πως ένας τρόπος να ικανοποιηθούν κάποιες από τις απαιτήσεις για βελτίωση της καταγραφής ήταν η σχεδίαση ενός υπολογιστικού συστήματος παρακολούθησης των διαγραμμάτων υγείας. Οι ειδικοί επιστήμονες που ρωτήθηκαν ενέκριναν το όλο εγχείρημα και συνέστησαν ότι ένα τέτοιο σύστημα θα έπρεπε να είναι απαλλαγμένο από τη χρήση χεριών. Εξέφρασαν την προσδοκία τους να περιοριστούν οι πρόσθετες επιβαρύνσεις κατά τις εξετάσεις και τις εξυπηρετήσεις των ασθενών που απαιτούν την αποκλειστική χρήση χεριών και ματιών. Η μόνη ενδεδειγμένη λύση ήταν η αναγνώριση ομιλίας.

Πριν από την επιλογή του τελικού προϊόντος, η KorTeam προχώρησε στη δημιουργία μιας λίστας απαιτήσεων σχετικά με:

- Τις απαιτήσεις του έργου των διαγραμμάτων υγείας
- Τα χαρακτηριστικά των χρηστών
- Τις συνθήκες κάτω από τις οποίες οι γιατροί και οι νοσοκόμοι θα λειτουργούσαν το σύστημα.

Επίσης ερευνήθηκαν οι απαιτήσεις που θα έπρεπε να πληρεί το μικρόφωνο, λαμβάνοντας υπ’ όψιν τις προτιμήσεις του προσωπικού, την απόκριση και τη

διακριτικότητα. Το interface του χρήστη που αναπτύχθηκε περιείχε ένα μενού κι έτρεχε πάνω στο περιβάλλον των Microsoft Windows, παρουσιάζοντας τη μορφή της φόρμας που πρέπει να συμπληρωθεί. Η χρήση εικόνων της πραγματικής φόρμας απαίτησε την ενσωμάτωση ειδικών εργαλείων για την προσθήκη, διαγραφή και αλλαγή των φορμών, καθώς οι φόρμες χρειάζεται να τροποποιούνται περιοδικά από τους χρήστες. Το interface σχεδιάστηκε με βάση τις ανάγκες τεσσάρων κατηγοριών χρηστών, οι οποίες ορίστηκαν ως προς την εξοικείωση με τους υπολογιστές και την εμπειρία με τα επαγγέλματα υγείας. Ο προσεκτικός σχεδιασμός του πλάνου εφαρμογής επιμήκυνε το χρόνο μεταξύ της σύλληψης της ιδέας και την υλοποίησή της ως το σύστημα Chartwell για τα διαγράμματα υγείας δίπλα στα κρεβάτια. Με τον τρόπο, όμως, αυτόν είχε ήδη επαυξηθεί η εξοικείωση και η αποδοχή των χρηστών ως προς το σύστημα, αυξάνοντας, έτσι, την πιθανότητα να σταφθεί το εγχείρημα με επιτυχία.

### **3. Απόθεμα εργατικών αποζημιώσεων (Worker's Compensation reserving). Navistar International**

Η διατήρηση αποθέματος είναι ένας τρόπος εξοικονόμησης των κεφαλαίων που θ' απαιτηθούν για τη επαρκή κάλυψη των αναγκών μιας εταιρίας όσον αφορά τις διεκδικήσεις εργατικών αποζημιώσεων. Συνεπάγεται έναν πολύπλοκο υπολογισμό βασισμένο στους παράγοντες που υπεισέρχονται σε κάθε διεκδίκηση. Μετά από δεκαοκτώ χρόνια, ο ειδικός της Navistar περί του αποθεματικού για τις εργατικές αποζημιώσεις επρόκειτο να συνταξιοδοτηθεί, χωρίς να υπάρχει κάποιος ικανός να τον αντικαταστήσει.

Η λύση που προτάθηκε ήταν η ανάπτυξη ενός συστήματος τεχνητής νοημοσύνης (ενός έμπειρου συστήματος δηλαδή), το οποίο θα κωδικοποιούσε τη λογική λήψης αποφάσεων που χρησιμοποιούν οι ειδικοί περί των αποθεματικών. Η αναγνώριση ομιλίας συμπεριλήφθηκε στον όλο σχεδιασμό επειδή οι υπάλληλοι που θα χειρίζονταν το σύστημα δεν ήταν εξοικειωμένοι με τη χρήση υπολογιστών και η ομιλία τους έμοιαζε περισσότερο φυσική και οικεία σε σχέση με τους άλλους τρόπους εισόδου δεδομένων. Προκειμένου να διευκολυνθεί η χρήση του έμπειρου συστήματος και η συλλογή των δεδομένων το λεξιλόγιο και το interface του χρήστη έτυχαν ιδιαίτερου σχεδιασμού κι ενσωματώθηκαν στα στάνταρτ του συστήματος.

Ένα πρωτότυπο του φωνητικού συμβούλου περί του αποθέματος εργατικών αποζημιώσεων τέθηκε υπό τον έλεγχο των ειδικών και των χρηστών του συστήματος. Η εξοικείωση με το πρωτότυπο διευκόλυνε την προσαρμογή των χρηστών και τους επέτρεψε να προτείνουν βελτιώσεις, όπως την εισαγωγή ενός απλού συστήματος φωνητικού password. Επίσης, επαλήθευσαν την ακρίβεια του συστήματος, την καταλληλότητα του λεξιλογίου και τη σπουδαιότητα του interface ομιλίας:

*“Δε νομίζω ότι θα μπορούσα να το χρησιμοποιήσω χωρίς την ομιλία.” (Ray Skarcyuski, Manager, Worker's Compensation Department, Navistar International, Personal communication, 1988).*

### **3.3 Πρόσβαση Δεδομένων (Data Access)**

Η πρόσβαση δεδομένων (γνωστή και ως ανάκτηση πληροφοριών (information retrieval) ή πρόσβαση πληροφοριών (information access)) είναι μια σχετικά πρόσφατη λειτουργία της αναγνώρισης ομιλίας. Στόχος της είναι η ανάκτηση πληροφοριών από μια on-line πηγή, συνήθως μια βάση δεδομένων. Ο ρόλος της ομιλίας εντοπίζεται στην ενημέρωση μιας εφαρμογής ως προς τα δεδομένα που πρόκειται να μεταφερθούν. Οι πρώτες εφαρμογές

αναφέρονταν σε συστήματα τηλεφωνικής εξυπηρέτησης, τα οποία επέκτειναν τις υπάρχουσες ψηφιακές εφαρμογές στους χρήστες με τηλέφωνα περιστροφικού καντράν.

Το πλήθος των εφαρμογών της πρόσβασης δεδομένων έχει πλέον αυξηθεί και περιλαμβάνει:

- |                                 |   |
|---------------------------------|---|
| • Αυτοκίνηση                    | Εργαλεία πλοήγησης των αστικών περιοχών   |
| • Τραπεζικές συναλλαγές         | Πρόσβαση των ιδιοκτητών στους προσωπικούς τους τραπεζικούς λογαριασμούς                                 |
| • Καταναλωτικά προϊόντα         | Χρήση των προσωπικών ψηφιακών βοηθών (PDA's)  |
| • Εξυπηρέτηση πελατών           | Δυνατότητα τηλεφωνικών παραγγελιών  |
| • Επισκευές εξοπλισμών          | Πρόσβαση του προσωπικού στα επισκευαστικά δεδομένα των εγχειριδίων χρήσης (manual)                      |
| • Οικονομία                     | Πρόσβαση των χρηματιστών στα δεδομένα της χρηματιστηριακής αγοράς                                       |
| • Δικηγορία                     | Έρευνα υποθέσεων μέσω on-line νομικών βάσεων δεδομένων  |
| • Εφαρμογή των νόμων            | Πρόσβαση των τροχονόμων στη βάση δεδομένων με τις άδειες οδήγησης                                       |
| • Συστήματα γραμματείας         | Ενοποίηση υπολογιστών-τηλεφώνων   |
| • Μη κερδοσκοπικούς Οργανισμούς | Παροχή ανακοινώσεων και πληροφοριών στα μέλη  |
| • Λιανικό εμπόριο               | Παροχή μισθοδοτικών πληροφοριών στους καταστηματάρχες   |
| • Τηλεπικοινωνίες               | Πρόσβαση σε υπηρεσίες καταλόγου, Υλοποίηση κλήσεων μέσω αναφοράς του ονόματος του επιθυμητού συνομιλητή |
| • Τουρισμός                     | Πρόσβαση των τουριστών σε βάσεις δεδομένων με ταξιδιωτικές οδηγίες.                                     |

Μια δημοφιλής εφαρμογή της πρόσβασης δεδομένων είναι η υλοποίηση τηλεφωνικών κλήσεων μετά από απλή αναφορά του ονόματος του ατόμου που επιθυμούμε να καλέσουμε. Εν αντιθέσει με την υλοποίηση κλήσεων της εκτέλεσης εντολών κι ελέγχου, η οποία απλά μετατρέπει τα ψηφία ενός αριθμού τηλεφώνου σε τηλεπικοινωνιακά σήματα, η υλοποίηση κλήσεων μέσω ονόματος (dial-by-name) αναμένει την εισαγωγή από τους χρήστες κάποιου προκαθορισμένου ονόματος ή άλλου αναγνωριστή (π.χ. "κάλεσε το γραφείο). Η διεργασία που λαμβάνει χώρα είναι:

1. Η ομιλία αναγνωρίζεται από ένα σύστημα αναγνώρισης ομιλίας
2. Το αναγνωρισμένο πρότυπο αναζητείται σ'έναν πίνακα δεδομένων
3. Ανασύρεται ο αριθμός τηλεφώνου που αντιστοιχεί στο αναγνωρισμένο πρότυπο
4. Υλοποιείται η κλήση στον αριθμό αυτόν

Οι περισσότερες εφαρμογές πρόσβασης δεδομένων χαρακτηρίζονται από συγκεκριμένους στόχους: Καταγραφή της κατάστασης ενός τραπεζικού λογαριασμού, εμφάνιση κάποιας συγκεκριμένης πληροφορίας, υλοποίηση μιας κλήσεως. Οι πιο πρόσφατες εφαρμογές, όπως το AudioNav της Amerigon και το LawTalk της West Publishing για την WESTLAW, ικανοποιούν πολυπλοκότερες αλληλεπιδράσεις και χρησιμοποιούν μεγαλύτερα λεξιλόγια.



Η αναγνώριση ομιλίας είναι ένας τρόπος επέκτασης των υπάρχουσων ψηφιακών τηλεφωνικών υπηρεσιών (touch-tone telephone services) πρόσβασης δεδομένων σε πελάτες με τηλεφωνικές συσκευές περιστρεφόμενου καντράν. Αυτό είναι ιδιαίτερα χρήσιμο στις αγροτικές και τις υπερπόντιες περιοχές των Ηνωμένων Πολιτειών. Όπως οι ψηφιακές υπηρεσίες, έτσι και τα συστήματα αναγνώρισης ομιλίας μπορούν να προάγουν υπηρεσίες χωρίς διακοπή, 24 ώρες ημερησίως.

### 3.3.1 Ζητήματα σχεδιασμού της πρόσβασης δεδομένων

#### 1. Μέγεθος κι επιλογή λεξιλογίου

Οι περισσότερες από τις υπάρχουσες εφαρμογές πρόσβασης δεδομένων χαρακτηρίζονται από μικρά λεξιλόγια, αποτελούμενα από λίγες, φανερά διακριτές λέξεις κλειδιά. Οι λεξιλογικές απαιτήσεις των εφαρμογών της πρόσβασης δεδομένων αυξάνονται με την ενσωμάτωση περισσότερο εκτεταμένων προσβάσεων στις πληροφορίες μεγάλων βάσεων δεδομένων και την επέκτασή τους σε ανάκτηση πληροφοριών πλήρους κειμένου (full-text retrieval). Ακόμη και μια φαινομενικά απλή επερώτηση βάσεων δεδομένων, όπως η:

*“Λίστα όλων των μπουλονιών (bolts) για τα οποία η αποθηκευμένη και απογεγραμμένη ποσότητα είναι μικρότερη από 500”*

απαιτεί λεξιλόγιο που θα καλύπτει τα ονόματα όλων των διαθέσιμων πεδίων της βάσης, καθώς και το περιεχόμενό τους. Οι εφαρμογές αυτού του είδους είναι δύσκολο ν’ αναπτυχθούν, αν και είναι ιδιαίτερα χρήσιμες για οργανισμούς κι εταιρίες των οποίων οι υπάλληλοι χρειάζονται την πρόσβαση σε τέτοια δεδομένα, αλλά δεν επιθυμούν την εκμάθηση κατασκευής επερωτήσεων μέσω της SQL. Η χρήση εισόδων ομιλίας με μεγάλο λεξιλόγιο, για τη διατύπωση φυσικών επερωτήσεων, επιτρέπει σε επαγγελματίες χωρίς εμπειρία στη χρήση της SQL ή της Δυαδικής διατύπωσης επερωτήσεων (Boolean query formation) να ανακτούν την αναγκαία πληροφορία όποτε τη χρειάζονται.

#### 2. Μεταγλώττιση

Ένας από τους πρωταρχικούς στόχους πολλών συστημάτων πρόσβασης δεδομένων είναι η αντικατάσταση της πληκτρολόγησης πολύπλοκων και ασαφών προτύπων με έναν περισσότερο φυσικό τρόπο εισόδου. Για παράδειγμα, ο ακόλουθος κώδικας σε SQL αντιστοιχεί στην προφορική επερώτηση:

*“Λίστα όλων των μπουλονιών με αποθηκευμένη ποσότητα μικρότερη από 500”:*

```
SELECT TABLE1.ITEM_NUMBER, TABLE2.OH_BALANCE
FROM MATERIALS_TABLE TABLE1, WAREHOUSE_TABLE TABLE2
WHERE TABLE1.ITEM_NUMBER=TABLE2.ITEM_NUMBER AND
TABLE2.OH_BALANCE < 500 AND TABLE1.ITEM_TYPE='BOLT'
```

Επίσης οι δυο ακόλουθες αιτήσεις αντιστοιχούν στην τριμηνιαία ιστορία της εταιρίας Coca-Cola, όπως αυτές γίνονται σ’ ένα ευρείας χρήσεως πληροφοριακό σύστημα ενός εμπορικού ορόφου:

KEYBOARD	SPEECH
ko83q	Ιστορία της Coca Cola

Σε κάθε περίπτωση, η ομιλία αποτρέπει τους χρήστες από την ανάγκη εκμάθησης ενός ιδιαίτερα πολύπλοκου συστήματος πρόσβασης δεδομένων.

### **3. Γλωσσολογική δομή**

Τόσο η πραγματοποίηση κλήσεων μέσω φωνής, όσο και οι υπόλοιπες εφαρμογές της πρόσβασης δεδομένων που προορίζονται για χρήση από μικρές ομάδες ατόμων, υλοποιούνται μέσω των γραμματικών περιορισμένων καταστάσεων ή άλλων ισοδυνάμων. Οι τηλεφωνικές εφαρμογές, που χαρακτηρίζονται από μεγάλο πλήθος περιστασιακών χρηστών στηρίζονται στον εντοπισμό κάποιων λέξεων-κλειδιά. Σε αρκετές περιπτώσεις οι εφαρμογές αυτές υιοθετούν τα πρότυπα ερωτήσεων-αποκρίσεων που χρησιμοποιούνται στα συστήματα IVR. Δηλαδή, ο χρήστης αποκρίνεται με μια συγκεκριμένη λέξη ή φράση στη λίστα των επιλογών που του παρουσιάζεται.

Για την ανάκτηση πληροφοριών πλήρους κειμένου (full-text information retrieval) απαιτείται η χρήση ενός μοντέλου στατιστικής γλώσσας. Το ίδιο σύστημα Law Talk για την WESTLAW χρησιμοποιεί το μοντέλο δύο γραμματικών Dragon Dictate συνδυασμένο με το WIN, το σύστημα επερωτήσεων φυσικής γλώσσας της WEST. Το WIN αναλύει την αναγνωρισμένη είσοδο και τη μετατρέπει σε SQL επερωτήσεις.

### **4. Μοντέλο Ομιλητή**

Η επιλογή του μοντέλου ομιλητή εξαρτάται από τη φύση και το μέγεθος του πληθυσμού των χρηστών. Τα συστήματα που προορίζονται για εκτεταμένη χρήση ενσωματώνουν τεχνολογία ανεξάρτητη του ομιλητή, ενώ τα φωνητικά συστήματα υλοποίησης κλήσεων που εφαρμόζονται σε μια απλή τηλεφωνική συσκευή, ή στα τηλέφωνα ενός νοικοκυριού είναι εξαρτημένα από τον ομιλητή. Η χρήση της εξαρτημένης από τον ομιλητή τεχνολογίας επαυξάνει την ακρίβεια ενός συστήματος, ενώ συνάμα επιτρέπει στους χρήστες να προσαρμόζουν το σύστημα στις απαιτήσεις τους, δημιουργώντας τις προσωπικές τους εγγραφές για ταχεία υλοποίηση κλήσεων μέσω ονόματος. Το σύστημα Law Talk για την WESTLAW, που χαρακτηρίζεται από μεγάλο λεξιλόγιο, είναι προσαρμοζόμενο στον ομιλητή, καθώς προορίζεται για επαναλαμβανόμενη χρήση από τους χρήστες. Η προσαρμογή στον ομιλητή επαυξάνει την ακρίβεια, προσαρμόζοντας την αναγνώριση στη φωνή του κάθε χρήστη. Με τον τρόπο αυτό διευκολύνεται η διάκριση ανάμεσα σ'ένα μεγάλο πλήθος υποψήφιων λέξεων.

### **5. Ροή ομιλίας**

Η πιο κατάλληλη ροή ομιλίας για την πρόσβαση δεδομένων είναι η συνεχής ομιλία σε συνδυασμό με τον εντοπισμό λέξεων. Η αναγνώριση συνεχούς ομιλίας ή διακριτών λέξεων μπορεί να εφαρμοστεί μόνο σε περίπτωση που οι χρήστες δεν αναμένεται να χρησιμοποιούν επιπλέον ομιλία.

### **3.3.2 Το περιβάλλον ομιλίας της πρόσβασης δεδομένων**

Το πιο γνωστό περιβάλλον ομιλίας είναι η δημόσια τηλεφωνία. Η χρήση του τηλεφώνου ως πλατφόρμα. Για την ανάπτυξη εφαρμογών αναγνώρισης ομιλίας αναπτύσσεται σε ακόλουθη παράγραφο.

Οι υπολογιστικές εφαρμογές (desktop applications) συνήθως αναφέρονται σε περιβάλλον γραφείου. Τα γραφεία χαρακτηρίζονται από διακεκομμένο θόρυβο (όπως ένα τηλέφωνο που χτυπάει) και περιβάλλουσα ομιλία (background speech). Εν αντιθέσει, τα καταναλωτικά προϊόντα δεν έχουν συγκεκριμένα περιβάλλοντα ομιλίας, γι' αυτό και κατά το σχεδιασμό τους λαμβάνεται πάντα υπ' όψιν η χειρότερη δυνατή περίπτωση, δηλαδή δυνατός, άτακτος θόρυβος και περιβάλλουσα ομιλία. Το AudioNav της Amerigon έχει σχεδιαστεί για ένα από τα πιο επιβαρημένα περιβάλλοντα ομιλίας, την αυτοκίνηση.

### **3.3.3 Οι ανθρώπινοι παράγοντες της πρόσβασης δεδομένων**

Από τη στιγμή που αρκετές εφαρμογές της πρόσβασης δεδομένων δημιουργούνται για χρήση στην τηλεφωνία, οι ανθρώπινοι παράγοντες που υπεισέρχονται αντιστοιχούν σ' αυτούς των άλλων τηλεφωνικών εφαρμογών. Το ζήτημα των ανθρωπίνων παραγόντων για τις εφαρμογές στα καταναλωτικά προϊόντα δεν έχει ακόμη αναλυθεί πλήρως. Τα συστήματα πλοήγησης αυτοκινήτου (Car navigation systems) και τα λοιπά συστήματα παροχής οδηγιών (directions) οφείλουν να συμμορφώνονται στα καθιερωμένα πρότυπα συνομιλιών και τις τελετουργίες κάθε έργου, συμπεριλαμβανομένου του τρόπου με τον οποίο δίνονται οι οδηγίες, του χειρισμού των ανθρωπίνων σφαλμάτων και της παροχής των οδηγιών με τρόπο που να γίνονται εύκολα κατανοητές.

### **3.3.4 Παραδείγματα της πρόσβασης δεδομένων**

Στη συνέχεια παρατίθενται κάποια χαρακτηριστικά παραδείγματα συστημάτων πρόσβασης δεδομένων. Η εφαρμογή της Sprint/Texas Instruments απεικονίζει τη χρήση του Μάγου του Οζ για τον έλεγχο της διαδικασίας επαναλαμβανόμενου σχεδιασμού. Άλλα συστήματα πρόσβασης δεδομένων είναι τα προϊόντα που έχουν σχεδιαστεί για εργοστασιακές, νομικές και καταναλωτικές χρήσεις.

### **1. Τραπεζικές συναλλαγές μέσω τηλεφώνου. National Westminster Bank P.I.C. London, England**

Η ευκολία και η απλότητα στη χρήση του τηλεφώνου έχει γεννήσει μια διαρκώς αναπτυσσόμενη βιομηχανία εξυπηρέτησης τραπεζικών συναλλαγών από το σπίτι. Η ευρεία χρήση των ψηφιακών τηλεφωνικών εφαρμογών ενίσχυσε την ανάπτυξη των τηλεφωνικών συναλλαγών, αν και οι ψηφιακές διευκολύνσεις εκτός της Βόρειας Αμερικής δεν έχουν αναπτυχθεί ιδιαίτερα. Για τον λόγο αυτό, έξω από τις Η.Π.Α οι τηλεφωνικές τραπεζικές συναλλαγές υλοποιούνται μέσω της αναγνώρισης ομιλίας.

Η Actionline της National Westminster Bank αποτελεί χαρακτηριστικό παράδειγμα αυτών των εφαρμογών. Πρωταρχικό της στόχο αποτελεί η διευκόλυνση των καταναλωτών,

ενώ λειτουργικά συνδυάζει την εκτέλεση εντολών κι ελέγχου για τη διακίνηση κεφαλαίων με την πρόσβαση στα δεδομένα τραπεζικών λογαριασμών.

Το σύστημα απαιτεί συγκεκριμένη δικτυακή και hardware διαμόρφωση αποτελούμενη από:

- Πολλαπλά συστήματα ομιλίας στηριγμένα σε PC
- Μια host βάση δεδομένο εφοδιασμένη με την πληροφορία των τραπεζικών λογαριασμών
- Μια τηλεφωνική γραμμή υποστήριξης για την υποβοήθηση των χρηστών (helpdesk)

Κάθε σύστημα ομιλίας έχει τη δυνατότητα ταυτόχρονου χειρισμού μέχρι δεκαέξι κλήσεων. Ο host περιέχει την απαιτούμενη πληροφορία των λογαριασμών και μπορεί να ικανοποιεί τις εντολές μεταφοράς κεφαλαίων των χρηστών. Η γραμμή υποβοήθησης (helpdesk) απασχολεί ανθρώπινο δυναμικό για την υποστήριξη του συστήματος. Ολόκληρος ο εξοπλισμός συνδέεται σ' ένα ενδοτραπεζικό ψηφιακό δίκτυο.

Οι ανθρώπινοι παράγοντες υπεισέρχονται σε ζητήματα που κυμαίνονται από τον προσδιορισμό των τύπων των πελατών που θα ενδιαφέρονταν για τη χρήση ενός τέτοιου συστήματος, μέχρι το σχεδιασμό του interface και πιο συγκεκριμένα την επιλογή της φωνής που θα χρησιμοποιηθεί στις επερωτήσεις, τη σύνταξη και το μήκος των επερωτήσεων, καθώς και τον τρόπο με τον οποίο η μηχανή θα διακινεί τα κακά νέα για ένα λογαριασμό.

*“Οφείλαμε να εκπαιδύσουμε μια μηχανή να πληροφορεί τον πελάτη, με αποδεκτό τρόπο, ότι αυτός ή αυτή υπερέβη το λογαριασμό του/της ή ακόμη χειρότερα ότι αυτός ή αυτή εισήγαγε μη έγκυρα δεδομένα...Για παράδειγμα, ενώ για έναν άνθρωπο είναι απολύτως αποδεκτό να λέει “συγγνώμη” σε κάποιον άλλο, για τις μηχανές δεν είναι και τόσο αποδεκτό να αντιμετωπίζουν παρόμοιες καταστάσεις με τον ίδιο τρόπο.” (Geoff Hammond, Information Technology Manager, Telephone Banking Project, National Westminster Bank, “Deploying Speech Technology in telephone banking services”, 1990, p.31).*

Ο πιλοτικός έλεγχος έγινε μεμονωμένα με τους τραπεζικούς πελάτες να επιβεβαιώνουν απλά τα στοιχεία κάποιων αναμενόμενων χρηστών του συστήματος. Ακολούθησε η ανάπτυξη του τελικού συστήματος. Από την αρχική υλοποίηση του συστήματος κι έπειτα, η Actionline έχει καταγράψει την προσδοκώμενη διευκόλυνση των πελατών, ενώ έχει περιορίσει και τα έξοδά της, αντικαθιστώντας το χειρισμό των χάρτινων αποδείξεων (paper documents) με τη φθηνότερη ηλεκτρονική επεξεργασία.

## **2. Φωνητική υλοποίηση κλήσεων. Sprint & Texas Instruments**

Το FONCARD της Sprint Voice σχεδιάστηκε αρχικά για τον περιορισμό του φαινομένου εξαπάτησης της καλούμενης κάρτας (calling card fraud), αλλά στην πορεία μετεξελίχθηκε σε μια ελκυστική υπηρεσία για τους συνδρομητές.

*“Το αρχικό ενδιαφέρον στην τεχνολογία αυτή εστιαζόταν στη δημιουργία μιας συσκευής ασφάλειας η οποία θα συντελούσε στην εξάλειψη της εξαπάτησης. Αν και η επαλήθευση των στοιχείων είναι ιδιαίτερα σημαντική για μας, αυτό που έχουμε συνειδητοποιήσει είναι ότι η ασφάλεια μάλλον δε θ' αποτελέσει το κύριο αντικείμενο της τεχνολογίας αυτής. Είναι εξαιρετικά απλή στη χρήση της.” (Tommye Wealand, Director, Strategic Planning Research, Sprint, Personal communication, 1994).*

“Ένα τέτοιο σύστημα χρειαζόταν να συνδυάζει την επαλήθευση του ομιλητή, την αναγνώριση ομιλίας και την αυξημένη λειτουργικότητα ως προς την είσοδο ομιλίας. Αρκετά χρόνια εργασίας αφιερώθηκαν στον καθορισμό και την ανάπτυξη ενός τέτοιου προτύπου.

Η διεργασία σχεδιασμού στηρίχθηκε στην επαναλαμβανόμενη ανάπτυξη και τον έλεγχο. Βασική συνιστώσα της όλης διεργασίας αποτέλεσε ο έλεγχος μέσω του Μάγου του Οζ. Διάφορες παράμετροι του interface, όπως η εφαρμογή του password εισόδου, ελέγχθηκαν ανεξάρτητα με τη συμβολή των χρηστών των δημόσιων τηλεφωνικών γραμμών. Ο έλεγχος αυτός, για παράδειγμα, επέδειξε ότι οι χρήστες αρέσκονταν στην υλοποίηση κλήσεων μέσω ονομάτων (π.χ. “κάλεσε το λογιστή μου”) και ότι η απλότητα χρήσης επαυξανόταν όταν τα passwords αποτελούνταν από οικεία αριθμητικά ή λεκτικά πρότυπα. Περαιτέρω έλεγχοι επέτρεψαν στους εφαρμογείς να περιορίσουν την αλληλεπίδραση συνενώνοντας την επαλήθευση του ομιλητή με την ανεξάρτητη του ομιλητή αναγνώριση ψηφίων, βελτιώνοντας τις επερωτήσεις κι ενσωματώνοντας τόσο την είσοδο ομιλίας όσο και τις ψηφιακές εισαγωγές (touch-time input).

*“Τόσο από την παραδοσιακή σκοπιά των ανθρώπινων παραγόντων, όσο από την οπτική γωνία της έρευνας αγοράς, φαντάζει περισσότερο χρονοβόρο και κοπιαστικό. Όμως, με δεδομένο ότι η ευκολία στη χρήση έχει γίνει σημαντικός παράγοντας εμπορικής επιτυχίας, ήταν επιτακτική η ανάγκη βελτίωσης της ερευνητικής μας μεθοδολογίας. Οι εξονυχιστικοί έλεγχοι αυτού του είδους συνέβαλαν δραματικά στη θετική ανταπόκριση των πελατών στο φωνητικό FONCARD.”* (Gerhard Deffner of Texas Instruments, Linda Jo Black of Sprint & Mark Ollmann of Texas Instruments, “Wizard of Oz Market testing of spoken command interface style”, 1994, p.128).

### **3.4 Υπαγόρευση**

Ένας από τους βασικούς σκοπούς της αναγνώρισης ομιλίας είναι να δημιουργήσει έναν ανθρώπινο ακουστικό δακτυλογράφο. Τεχνολογικά επιτεύγματα στα τέλη του 1980 έφεραν αυτό το στόχο πιο κοντά στη πραγματικότητα με την εμφάνιση εμπορικών, μεγάλων σε λεξιλόγιο προϊόντων υπαγόρευσης. Οι κατασκευαστές αυτών των προϊόντων έχουν καθορίσει δύο βασικές τάξεις χρηστών:

- Επιχειρήσεις, υγιεινή, και νομικούς επαγγελματίες με εκτεταμένες ανάγκες στη παραγωγή αρχείων
- Άνθρωποι με ανικανότητες οι οποίοι έχουν επαγγελματικές και/ή προσωπικές ανάγκες υπαγόρευσης.

Η χρήση της αναγνώρισης ομιλίας στη παραγωγή αναφοράς επιταχύνει την κίνηση της πληροφορίας σε οργανισμούς και συστήματα υπολογιστών που το χρειάζονται. Αυτό είναι ιδιαίτερα σημαντικό στην υγιεινή, όπου, όσο πιο γρήγορα υποβάλλεται η τεκμηρίωση στις ασφαλιστικές εταιρίες, στα προγράμματα ιατρικής βοήθειας, στην ιατρική φροντίδα, τόσο πιο γρήγορα γίνεται η πληρωμή.

Στους νομικούς επαγγελματίες, τα συστήματα δημιουργίας εγγράφων βασισμένα στον υπολογιστή (καλείται *σλλογή τεκμηρίων*) υπάρχουν εδώ και πολλά χρόνια. Αυτά τα συστήματα έχουν καθιερωθεί ως μια εύκολη και έξυπνη μέθοδος παραγωγής επίσημων δοκιμίων, όπως διαθήκες και ιδιαίτερων τύπων συμβόλαια. Η δομημένη υπαγόρευση αναφορών είναι μια επέκταση αυτής της δυνατότητας. Η χρήση των μεγάλης κλίμακας λέξεων *σκανδάλης* που χρησιμεύουν στο να τοποθετούνται μεγάλα τμήματα κειμένου μέσα σ’ ένα

έγγραφο είναι πολύ πιο γρήγορη μέθοδος από το να γράφεται το κείμενο λέξη προς λέξη. Αυτό το κέρδος μπορεί να επεκταθεί σε γενικές λειτουργίες επιχειρήσεων και σε μερικές βιομηχανίες οι οποίες χρειάζεται να υποβάλλουν περιοδικές αναφορές σε διάφορες αντιπροσωπείες της κυβέρνησης.

Επαγγελματίες με επαναληπτικό σύνδρομο άγχους (RSI-repetitive stress syndrome), δυσλειτουργίες κίνησης ή οπτικές βλάβες, βλέπουν την αναγνώριση ομιλίας σαν ένα μέσο απελευθέρωσης από την εξάρτησή τους από άλλους ανθρώπους. Η υπαγόρευση χωρίς χέρια επιτρέπει σ' αυτούς τους ανθρώπους να δημιουργήσουν δουλειά και προσωπικά κείμενα. Τα συστήματα υπαγόρευσης χρησιμοποιούνται σε μικρότερη έκταση ως βοήθεια προς τους ανθρώπους με τα παραπάνω προβλήματα, έτσι ώστε να μπορούν να παρακολουθήσουν τις διαδικασίες ενός δικαστηρίου ή άλλες βασικές προφορικές διαδικασίες. Η Janet Baker (1994) παρουσιάζει επιπρόσθετες πληροφορίες πάνω στην υπαγόρευση.

### **3.4.1 Σχεδιαστικά θέματα υπαγόρευσης**

#### **1. Μέγεθος λεξιλογίου και επιλογή**

Το λεξιλόγιο μιας διαδικασίας υπαγόρευσης μπορεί να διαχωριστεί στους εξής τέσσερις τύπους:

- Έμφυτο λεξιλόγιο
- Λεξιλόγιο του χρήστη
- Λεξικό
- Ενεργό λεξιλόγιο

Το *έμφυτο λεξιλόγιο* είναι αυτό το οποίο φορτώνεται μαζί με τη λειτουργία και περιέχει όλες τις λέξεις μέσα σ' αυτή. Το *λεξιλόγιο του χρήστη* μπορεί να είναι διαφορετικό από το έμφυτο μιας λειτουργίας ή μπορεί να είναι ένα στοιχείο αυτού. Μερικά προϊόντα επιτρέπουν στους χρήστες να δημιουργήσουν νέες λέξεις οι οποίες γίνονται κομμάτι του έμφυτου λεξιλογίου. Αυτές οι λέξεις είναι προσβάσιμες και σε άλλους χρήστες χωρίς περαιτέρω διαδικασίες. Άλλα προϊόντα ωστόσο, αποθηκεύουν λέξεις καθορισμένες απ' το χρήστη μέσα σ' ένα ιδιωτικό αρχείο αυτού. Οι άλλοι ομιλητές οι οποίοι θέλουν να χρησιμοποιήσουν αυτές τις λέξεις μπορούν, είτε να τις ορίσουν για τα δικά τους αρχεία ομιλίας είτε να τις εισάγουν. Το *λεξικό* είναι μια εφεδρική πηγή για τους σχεδιαστές και/ή τους λειτουργικούς χρήστες. Δεν φορτώνεται μαζί με μια λειτουργία, αλλά είναι προσπελάσιμο όταν ο σχεδιαστής και/ή ο χρήστης επιθυμεί να προσθέσει μια νέα λέξη στο έμφυτο λεξιλόγιο. Τα πιο πολλά προϊόντα έχουν τη δυνατότητα να αναφέρουν αν μια νεο-ορισμένη είσοδος υπάρχει ήδη στο εφεδρικό λεξικό. Όταν μια λέξη προστίθεται στη λειτουργία από το εφεδρικό λεξικό, είναι σαν μια ανεξάρτητη του ομιλητή βασική φόρμα λεξιλογίου. Το *ενεργό λεξιλόγιο* είναι ένα χαρακτηριστικό της τρέχουσας λειτουργίας του συστήματος. Είναι η λίστα των λέξεων που μπορούν να εμφανιστούν στον επόμενο χρόνο και είναι καθορισμένη από το μοντέλο γλώσσας του συστήματος.

Αποτελεσματική υπαγόρευση απαιτεί ένα πολύ μεγάλο επεκτάσιμο λεξιλόγιο. Θεωρητικά, όσο μεγαλύτερο είναι το έμφυτο λεξιλόγιο της λειτουργίας τόσο λιγότερο αποτελεσματικό είναι το σύστημα στο να αντιμετωπίσει μια άγνωστη λέξη. Στη πραγματικότητα, είναι η καταλληλότητα του λεξιλογίου η οποία είναι κρίσιμη. Ακόμη κι αν το μέγιστο έμφυτο λεξιλόγιο είναι είκοσι ή εξήντα χιλιάδες λέξεις, πρέπει να περιέχει τις λέξεις και τις φράσεις οι οποίες χρειάζονται για τη λειτουργία. Αυτό είναι ένα σημαντικό

χαρακτηριστικό της λειτουργίας κι αυτό επειδή ο λάθος υπολογισμός αγνώστων λέξεων αποτελεί το πιο συχνό σφάλμα που παράγουν τα συστήματα υπαγόρευσης. Οι μέθοδοι επέκτασης του έμφυτου λεξιλογίου ποικίλλουν με κάθε προϊόν, αλλά γενικά απαιτείται είσοδο ομιλίας (γραπτή ή προφορική). Ένας από τους σκοπούς της έρευνας στο σχεδιασμό μεγάλων λεξιλογίων είναι η αυτοματοποίηση της διαδικασίας πρόσθεσης νέων λέξεων, αλλά δεν υπάρχουν προς το παρόν εμπορικά συστήματα τα οποία να μπορούν αυτόματα να παράγουν νέο λεξιλόγιο μόνο από προφορική είσοδο. Μέχρι τη πραγματοποίηση αυτών των συστημάτων, η ευθύνη της ταυτοποίησης και μοντελοποίησης των λέξεων και γλώσσας στο πεδίο ορισμού της λειτουργίας, θα είναι αποκλειστική του σχεδιαστή του προϊόντος ή του σχεδιαστή της λειτουργίας. Οι χρήστες θα μπορούν να παίζουν περιορισμένο ρόλο σ' αυτό το χώρο.

## **2. Μεταγλώττιση**

Η μεταγλώττιση της εισαγόμενης ομιλίας σχεδόν πάντα μετατρέπεται σε κείμενο. Τα περισσότερα συστήματα υπαγόρευσης επιτρέπουν τη δημιουργία λέξεων σκανδάλης οι οποίες συμπεριφέρονται σαν σειρές εντολών. Μια λέξη σκανδάλης μπορεί να προκαλέσει τη παραγωγή μεγάλων τμημάτων μορφοποιημένου κειμένου και είναι κομμάτι σχεδίασης των περισσότερων διαδικασιών παραγωγής αναφοράς.

Τα περισσότερα συστήματα επίσης περιέχουν μια μικρή ομάδα λέξεων εντολών-και-ελέγχου οι οποίες χρησιμοποιούνται στην κίνηση μέσα στο σύστημα ή σε έγγραφο και πραγματοποιούν συνήθεις λειτουργίες, όπως πρόσθεση νέων λέξεων. Ο χαρακτήρας, αριθμός και οι μεταγλωττίσεις αυτών των λέξεων εξαρτώνται από το προϊόν.

## **3. Δομή της γλώσσας**

Τα συστήματα υπαγόρευσης απαιτούν το μεγαλύτερο βαθμό ευκαμψίας στη δομή της γλώσσας. Χωρίς αυτή, ελεύθερη μορφή εισόδου δεν είναι δυνατή. Γι' αυτό, πεπερασμένης κατάστασης γραμματικές και άλλες περιορισμένες μέθοδοι σχεδιασμού είναι ακατάλληλες. Η πιο επιτυχημένη προσέγγιση είναι η στατιστική μοντελοποίηση γλώσσας. Οι πεπερασμένης κατάστασης γραμματικές και άλλα παρόμοια σχήματα, ωστόσο, απασχολούν τη *δομημένη υπαγόρευση*. Αυτή, η οποία καλείται και παραγωγή αναφοράς, είναι υπερπλήρης σε πολλαπλές επιλογές και αντικείμενα τελειοποιημένης πρότασης. Από πολλές πλευρές η δομημένη υπαγόρευση είναι πιο τυπική σε δεδομένα εισόδου απ' ότι η υπαγόρευση. Ένα καλό παράδειγμα της δομημένης υπαγόρευσης είναι

Μερικοί σχεδιαστές έχουν εκμεταλλευτεί το γεγονός ότι τα μοντέλα γλώσσας είναι στατιστικά στη φύση. Αυτοί έχουν συμπεριλάβει την ικανότητα αλλαγής μοντέλων εκθέτοντάς τα σε μεγάλα κείμενα τα οποία περιέχουν μοτίβα γλώσσας που είναι τυπικά των εγγράφων τα οποία το σύστημα θα παράγει.

## **4. Μοντέλο ομιλητή**

Το ιδανικό μοντέλο ομιλητή για ένα σύστημα υπαγόρευσης είναι το ανεξάρτητο του ομιλητή. Αυτό απαιτεί ένα επίπεδο ακρίβειας η οποία δεν είναι ακόμη δυνατή στα ανεξάρτητα του ομιλητή μοντέλα, για συστήματα με πολύ μεγάλο λεξιλόγιο. Εμπορικά προϊόντα τα οποία προσφέρουν μοντέλα ανεξάρτητα του ομιλητή, όπως η σειρά εργαλείων συστήματος

υπαγόρευσης του Philips, καθορίζουν αυτά τα μοντέλα ως δευτερεύουσες εκδόσεις και τα διαχωρίζουν ως θηλυκά και αρσενικά (ή υψηλής και χαμηλής φωνής) μοντέλα.

Τα εμπορικά συστήματα υπαγόρευσης κάνουν χρήση της *προσαρμογής του ομιλητή*. Η προσαρμογή του ομιλητή εναρμονίζει τις ανεξάρτητες του ομιλητή βασικές φόρμες μέσα στο σύστημα με τις ξεχωριστές ακουστικές μορφές του κάθε χρήστη. Αυτές οι τροποποιήσεις μπορεί να επιτευχθούν με διάφορους τρόπους, όμως οι πιο κοινοί είναι η *γρήγορη εγγραφή* και η προσαρμογή του ομιλητή. Εξαρτώμενο απ' το σύστημα, λέξεις καθορισμένες από συγκεκριμένους χρήστες μπορούν να παραμείνουν εξαρτημένες του ομιλητή και να αποθηκευτούν στο αρχείο του ομιλητή που τις δημιούργησε, ή μπορούν να ξαναφορμαριστούν σαν βασικές μορφές και να αποτελέσουν κομμάτι του γενικού παρόντος λεξιλογίου της λειτουργίας.

Άνθρωποι με βλάβες οι οποίες επιδρούν στην ομιλία, όπως η ALS, η προοδευτική μυασθένεια, ή μυϊκή δυστροφία, είναι πιθανόν να έχουν δυσκολίες με τα συστήματα υπαγόρευσης. Οι μορφές ομιλίας τους μπορεί να είναι ασύμβατες και οι διακοπές μεταξύ των λέξεων να είναι μεγάλες. Η on-the-fly προσαρμογή του ομιλητή μπορεί να βοηθήσει, αλλά η διάκριση λέξης από τον ομιλητή μπορεί να μην είναι αρκετή για τη σωστή λειτουργία του συστήματος υπαγόρευσης.

## **5. Ροή ομιλίας**

Η συνεχόμενη ομιλία είναι η πιο επιθυμητή ροή ομιλίας για υπαγόρευση και αποτελεί το στόχο προς τον οποίο οι πωλήσεις των συστημάτων υπαγόρευσης κινούνται. Το 1994, τα Συστήματα Υπαγόρευσης Philips εισήγαγαν το πρώτο εμπορικό συνεχόμενης ομιλίας σύστημα υπαγόρευσης. Είναι μια σειρά εργαλείων για τους κατασκευαστές λειτουργίας.

Τα συστήματα αναγνώρισης διακριτής λέξης αναγκάζουν το χρήστη να εισάγει ένα στιγμιαίο διάλειμμα μεταξύ των λέξεων. Αυτά τα διαλείμματα εξολοθρεύουν δύο προβλήματα τα οποία συνεισφέρουν στη δυσκολία αναγνώρισης συνεχόμενης ομιλίας : επιδράσεις περιορισμού και εξακρίβωση ορίων λέξης. Το αποτέλεσμα είναι κατά πολύ ταχύτερο και ακριβές με λιγότερη υπολογιστική ισχύς. Η μετάπτωση από την αναγνώριση διακριτής λέξης στη συνεχόμενη ομιλία θα είναι ένα παραπροϊόν της υλοποίησης πιο ισχυρών ολοκληρωμένων κυκλωμάτων στους ηλεκτρονικούς υπολογιστές.

### **3.4.2 Περιβάλλον ομιλίας της υπαγόρευσης**

Πολλά συστήματα υπαγόρευσης χρησιμοποιούνται μέσα σε σχετικά ήσυχα γραφεία και εργαστήρια. Τα επίπεδα θορύβου του υποβάθρου σ' αυτά τα περιβάλλοντα είναι χαμηλά αλλά αποτελούνται πρωταρχικά από ομιλία υποβάθρου και ασυνεχή θόρυβο.

Συστήματα υπαγόρευσης έχουν χρησιμοποιηθεί με αξιοσημείωτη επιτυχία σε πιο δύσκολους χώρους, κυρίως σε ειδικής ανάγκης δωμάτια νοσοκομείου. Τα ειδικής ανάγκης δωμάτια προσφέρουν μια καταπληκτική πρόκληση σε κάθε σύστημα αναγνώρισης, επειδή μπορεί να ακούγονται δυνατές φωνές ή θόρυβοι και το άγχος των ομιλητών να είναι μεγάλο. Αυτά τα δωμάτια επίσης, είναι χαρακτηριστικά των χώρων ομιλίας για τα μελλοντικά συστήματα υπαγόρευσης αφού έτσι αρχίζουν και προσαρμόζονται σαν καταναλωτικά προϊόντα.



### **3.4.3 Ανθρώπινοι παράγοντες υπαγόρευσης**

Μια βασική πηγή δυσκολίας προέρχεται από τη λανθασμένη προσδοκία ότι η υπαγόρευση μαζί με την αναγνώριση ομιλίας θα είναι συγκρίσιμες στη χρήση μιας ανθρώπινης εγγραφής. Τα συστήματα υπαγόρευσης επί του παρόντος, αντιπροσωπεύουν τη πιο προχωρημένη μορφή τεχνολογίας εμπορικού συστήματος αναγνώρισης, αλλά δεν δουλεύουν όπως οι άνθρωποι ακροατές ή οι πεπειραμένοι ηχογράφοι. Μια λειτουργία θα αποτύχει αν επιτραπεί στους χρήστες η διατήρηση μη αληθινών προσδοκιών για τις ικανότητές τους. Κατά συνέπεια, το ταίριασμα του εργαλείου και του σχεδιασμού της λειτουργίας με τον αριθμό των χρηστών είναι κρίσιμο για την επιτυχία της λειτουργίας.

Ομιλητές οι οποίοι είναι συνηθισμένοι να υπαγορεύουν γρήγορα μια ανακοίνωση που έχει διορθωθεί και εγγραφεί από κάποιον άλλο, θεωρούν την αναγνώριση ομιλίας μια αγγαρεία που κάνει τη δουλειά τους πιο δύσκολη. Φθείρονται όταν αναγκάζονται να :

- σπαταλούν χρόνο για γρήγορη εγγραφή
- σταματούν ενδιάμεσα από τις λέξεις όταν μιλάνε
- επιβεβαιώνουν την ακρίβεια των αναγνωριζόμενων κειμένων
- πραγματοποιούν εισαγωγικές λειτουργίες οι οποίες δίνονται κανονικά από τους εγγραφείς.

Συστήματα τα οποία μπορούν να κατασκευαστούν ώστε να επιτρέπουν σε ένα δεύτερο πρόσωπο να εισαχθεί στην υπαγόρευση μπορεί να είναι πιο ικανοποιητικά, εάν όντως δεν ανατρέπουν τα οφέλη από τη μείωση διαχειριστικού κόστους και επιταχύνουν πληροφορία σε αποδέκτες που τη χρειάζονται.

Πολλοί χρήστες θεωρούν τη διακριτή λέξη είσοδο ενοχλητική και αφύσικη, ακόμη κι αν η διακοπή είναι στιγμιαία. Για αυτούς τους χρήστες η διακοπή είναι ένα εμπόδιο στη παραλαβή. Άλλοι, συμπεριλαμβανομένων και πολλών ανθρώπων με δυσλειτουργίες οι οποίοι χρησιμοποιούν αυτά τα συστήματα, θεωρούν τη διακοπή μεταξύ των λέξεων ως μια ελάχιστη επιβάρυνση σε σύγκριση με τα οφέλη που αποκομίζουν από τη χρήση ενός συστήματος αναγνώρισης ομιλίας.

Οι λειτουργίες υπαγόρευσης στις οποίες συμμετέχουν περισσότεροι από ένας, όπως στη δημιουργία των ανακοινώσεων υψηλού κινδύνου δωματίων, το φόρτωμα και ξεφόρτωμα των μοντέλων χρήστη μπορεί να φαίνεται ότι απαιτεί αρκετό χρόνο. Εάν ένα σύστημα υπαγόρευσης είναι επιθυμητό για άλλους λόγους, ο κατασκευαστής μπορεί να φτιάξει μια λειτουργία η οποία να επιτρέπει συνεισφορές από διάφορους ανθρώπους να παραχθούν ανεξάρτητα και να συναρμολογηθούν αργότερα.

Τα συστήματα υπαγόρευσης πρέπει να παρέχουν εργαλεία φιλικά στο χρήστη για να επικυρώνουν την αναγνώριση, να διορθώνουν λάθη, να προσθέτουν λέξεις και να εξερευνούν το κείμενο. Τα πιο πολλά εμπορικά προϊόντα τώρα, προσφέρουν αυτά τα εργαλεία και τα έχουν κάνει ελεύθερα προς χρήση. Η ανάγκη για χωρίς έλεγχο υπαγόρευση είναι βασική για τη χρησιμότητα των συστημάτων στους ανθρώπους με δυσλειτουργίες. Εταιρίες οι οποίες βλέπουν τους ανθρώπους με δυσλειτουργίες ως ένα σημαντικό κομμάτι της αγοράς έχουν δημιουργήσει χωρίς έλεγχο προϊόντα και πιο πολλά συστήματα προσθέτουν εργαλεία και επιλογές.

### **3.4.4 Παραδείγματα υπαγόρευσης**

Τα δύο συστήματα που περιγράφονται παρακάτω αντιπροσωπεύουν διαφορετική σχεδίαση και στρατηγικές υλοποίησης. Κάποιες όψεις των ήδη υπαρχόντων συστημάτων, όπως η *DragonDictate's* διακριτής λέξης αναγνώριση, τελικά θα αντικατασταθούν από πιο εξελιγμένες μορφές όπως η συνεχής αναγνώριση ομιλίας. Οι πιο πολλές μορφές που περιγράφονται εδώ αντιπροσωπεύουν σχεδιαστικές αποφάσεις βασισμένες πάνω σε σκοπούς υλοποίησης.

#### **1. DragonDictate. Dragon συστήματα**

Η *DragonDictate* σχεδιάστηκε για άμεση παραγωγή δοκιμίων υπαγόρευσης. Προσαρμογή, σύνταξη κειμένου και ενημέρωση λεξιλογίου γίνονται την ίδια στιγμή από το πρόσωπο που κάνει την υπαγόρευση. Η προσαρμογή περιλαμβάνει προσαρμογή ακουστικών μοντέλων, μοντέλων γλώσσας και παρασκευαστικού θορύβου.

Αν και το σύστημα χρησιμοποιείται από κατασκευαστές που φτιάχνουν ειδικές λειτουργίες υπαγόρευσης, η βασική μορφή του *DragonDictate* είναι ένα ετοιμοπαράδοτο από ομιλία σε κείμενο σύστημα υπαγόρευσης το οποίο μπορεί να χρησιμοποιηθεί αμέσως αφού εγκατασταθεί. Η προσαρμογή του ομιλητή γίνεται στον αέρα. Κάθε διακριτής λέξης γεγονός αναγνώρισης παράγει μια ταξινομημένη λίστα λέξεων επιτυγχάνοντας τον υψηλότερο βαθμό αναγνώρισης. Εάν η υποψήφια με το μεγαλύτερο βαθμό δεν είναι η σωστή λέξη, ο ομιλητής μπορεί να διαλέξει μια άλλη από τη λίστα λέγοντας το όνομά της. Εάν η λέξη αυτή δεν βρίσκεται στη ταξινομημένη λίστα, τότε πιθανότατα δεν υπάρχει στο *DragonDictate* υπάρχον λεξιλόγιο που αποτελείται από εξήντα χιλιάδες λέξεις. Για να προσθέσει μια λέξη στο υπάρχον λεξιλόγιο, ο ομιλητής μπορεί προφορικά να βάλει σε ενέργεια μια αποθηκευμένη στο λεξιλόγιο λειτουργία. Η νέα εισαγωγή μπορεί να τυπωθεί με το πληκτρολόγιο του υπολογιστή ή να συλλαβιστεί λεκτικά χρησιμοποιώντας το στρατιωτικό αλφάβητο σε ελεύθερη μορφή. Αν το σύστημα κλειδώνει τη λέξη στο αποθηκευμένο λεξιλόγιό του, θα ξαναβρεί το βασικό μοντέλο λεξιλογίου και θα την προσθέσει στο υπάρχον λεξιλόγιο λειτουργίας αντί να προσθέσει το ανεξάρτητο του ομιλητή μοντέλο του χρήστη για τη λέξη. Άλλες άμεσες εισαγωγικές ενέργειες και εξερεύνηση κειμένων είναι διαθέσιμες μέσω της χρήσης του μικρού εντολής και ελέγχου λεξιλογίου.

Τα *DragonDictate* γλωσσικά μοντέλα απορρέουν από στατιστικές αναλύσεις γενικών επιχειρήσεων κειμένων. Πελατειακοί οργανισμοί και σχεδιαστές λειτουργιών μπορούν να προσαρμόσουν το γλωσσικό μοντέλο εφαρμόζοντας τα βελτιωτικά εργαλεία του συστήματος σε μεγάλες ποσότητες αντιπροσωπευτικών κειμένων των εγγράφων που το σύστημα θα παράγει. Το σύστημα είναι επίσης ικανό να *εξατομικεύει* το γλωσσικό μοντέλο μέσω μιας στον αέρα προσαρμογής του τύπου ομιλίας του χρήστη.

#### **2. Η σειρά εργαλείων λειτουργίας υπαγόρευσης. Philips συστήματα υπαγόρευσης**

Το αντικείμενο της σειράς εργαλείων είναι να καθιστούν δυνατή στους σχεδιαστές τη δημιουργία μιας ποικιλίας λειτουργιών διαδικασίας ομιλίας. Αυτό εξηγεί και τη δομημένη κατασκευή της σειράς εργαλείων. Οι διαδικασίες υπαγόρευσης μπορούν να περιλάβουν ψηφιακή εγγραφή και παίξιμο ηχογραφημένου μέρους μόνο, ή μπορούν να συμπεριλάβουν αναγνώριση ομιλίας. Η δομημένη προσαρμογή είναι σίγουρα παρούσα στην κατασκευή

σχεδίου λειτουργίας της αναγνώρισης ομιλίας : προσαρμογή του ομιλητή, υπαγόρευση, αναγνώριση και εργαλεία σύνταξης κειμένου είναι ξεχωριστές υπομονάδες που διευκολύνουν την ενσωμάτωση του τριμερούς λογισμικού.

Αν και το σύστημα παρέχει μοντέλα ανεξάρτητα του ομιλητή, οι χρήστες ενθαρρύνονται να χρησιμοποιούν τη γρήγορη διαδικασία εγγραφής πριν χρησιμοποιήσουν την αναγνώριση. Η υπαγόρευση γίνεται είσοδος δια μέσου ενός συνεχόμενης ομιλίας συστήματος υπαγόρευσης. Η *SpeechNote* υπομονάδα αποθηκεύει την είσοδο σαν μια υψηλής πιστότητας ψηφιακή εγγραφή. Καμιά άλλη σύνταξη ή άλλη αλληλεπίδραση μαζί με το σύστημα δε χρειάζεται. Αργότερα, όταν η *SpeechMagic* υπομονάδα αναγνώρισης ομιλίας συμπεριληφθεί, χρησιμοποιεί συνεχόμενης ομιλίας αναγνώριση για να μετατρέψει την αποθηκευμένη υπαγόρευση σε κείμενο. Κατά τη διάρκεια της διαδικασίας αναγνώρισης, η *SpeechMagic* διατηρεί έναν σύνδεσμο μεταξύ του αναγνωρισμένου κειμένου και της αυθεντικής ψηφιακής εγγραφής. Αυτή η σύνδεση αργότερα διευκολύνει τη σύνταξη του αναγνωρισμένου κειμένου βάζοντας τον κέρσορα του υπολογιστή να κινείται μέσω του ανοιχτού (on-screen) κειμένου ταυτόχρονα με το παίξιμο (playback) της ηχογραφημένης υπαγόρευσης. Η σύνδεση διατηρείται και μετά τη σύνταξη, βάζοντας έτσι σε λειτουργία τη διαδικασία καθορισμού λέξεων μέσα στο έγγραφο οι οποίες δεν περιλαμβάνονται στο λεξιλόγιό της. Το πρόσωπο το οποίο συντάσσει το έγγραφο μαρκάρει τις νέες λέξεις τις οποίες η λειτουργία χρειάζεται να προσθέσει στο λεξιλόγιό της.

Η σειρά εργαλείων σχεδιάστηκε για σχεδιαστές λειτουργιών και δεν προτίθεται να γίνει ένα σύστημα που να τελειώνει το χρήστη (end-user). Οι σχεδιαστές χρησιμοποιούν τη *SpeechPro* υπομονάδα για να χτίσουν γλωσσικά μοντέλα χρησιμοποιώντας μεγάλες ποσότητες κειμένου οι οποίες αντιπροσωπεύουν έγγραφα που η λειτουργία θα δημιουργήσει. Κάθε γλωσσικό μοντέλο δομείται από στατιστικές αναλύσεις γλωσσικών μορφών που απαντώνται σ' αυτά τα έγγραφα. Λεξιλόγια εφαρμογών περιέχουν ένα μέγιστο εικοσιπέντε χιλιάδων λέξεων παρμένες από τα κείμενα εκπαίδευσης και από το αποθηκευμένο λεξικό των τρακοσίων χιλιάδων λέξεων.

### **3.5 Χρησιμοποιώντας το τηλέφωνο**

Η δυνατότητα χρήσης αναγνώρισης ομιλίας μέσω τηλεφώνου έχει αποτελέσει το σκοπό της βιομηχανίας σαράντα χρόνια τώρα, αλλά δεν ήταν εφαρμόσιμη μέχρι τα μέσα της δεκαετίας του '90. Η εμπορική επιτυχία δεν ήρθε μέχρι τις αρχές του 1990. Από τότε, η δημοτικότητα της αναγνώρισης ομιλίας μέσω τηλεφώνου συνεχίζεται να αυξάνει.

Η δημοτικότητα και επιτυχία των εφαρμογών στο τηλέφωνο προέρχεται από αρκετούς παράγοντες. Μεταξύ αυτών είναι η σπουδαιότητα, η εξοικείωση και η παντοδυναμία του τηλεφώνου τόσο, όσο η ύπαρξη ενός καλά καθορισμένου παραδείγματος συνομιλίας : η άμεση μορφή απόκρισης (prompt-and-response pattern). Το παράδειγμα καθιερώθηκε για τις ψηφιακές (touch-tone) αλληλεπιδρούσας φωνής απόκρισης (Interactive Voice Response ή IVR) εφαρμογές και έχει εισαχθεί στις εφαρμογές αναγνώρισης ομιλίας. Το prompt-and-response παράδειγμα στέφθηκε εν μέρει με επιτυχία, επειδή είναι υψηλά ελέγξιμο. Το επίπεδο ελέγχου είναι υπεύθυνο για την εύκολη ικανοποίηση του αντικειμενικού στόχου του καλούντος, γρήγορα και αποτελεσματικά.

Η υλοποίηση της αναγνώρισης ομιλίας μέσω τηλεφώνου αξιώνει τρεις βασικές προδιαγραφές :

- *Συστήματα επιπέδου δικτύου*

Η τεχνολογία ενσωματώνεται μέσα σ' ένα δημόσιο τηλεφωνικό δίκτυο ή ανήκει σ' έναν έξυπνο περιφερειακό υπολογιστή, όπως ένας

μικροϋπολογιστής συνδεδεμένος σε ένα δημόσιο τηλεφωνικό δίκτυο.

- *Συστήματα προσωπικών υπολογιστών* Λογισμικό και/ή τυποποιημένο σταθερό πρόγραμμα (firmware) για προσωπικούς υπολογιστές οι οποίοι μπορούν να χρησιμοποιηθούν με ιδιωτικά τηλεφωνικά δίκτυα. Πολλά απ' αυτά τα συστήματα υπάρχουν σε πλακέτες για προσωπικούς υπολογιστές.
- *Συστήματα επιπέδου ολοκληρωμένων κυκλωμάτων* Τυποποιημένο σταθερό πρόγραμμα σε ολοκληρωμένα κυκλώματα ή σειρά ολοκληρωμένων κυκλωμάτων.

Εταιρίες τηλεπικοινωνιών όπως η AT&T, η Ameritech και η McCaw Cellular, χρησιμοποίησαν πρώτες την αναγνώριση ομιλίας στα τηλεφωνικά τους δίκτυα για διαχειριστικές υπηρεσίες και φωνητική πληκτρολόγηση. Από το 1994, η επιπέδου δικτύου αναγνώριση ομιλίας αναπτύχθηκε ως μια υπηρεσία με πληρωμή σε επιχειρήσεις και μόνιμους πελάτες.

Την ίδια στιγμή, οι εφαρμογές βασισμένες στον PC (Personal Computer) κέρδιζαν μια παρουσία στις επιχειρήσεις. Μια από τις πρώτες εφαρμογές αναπτύχθηκε από τον InterVoice, έναν IVR παροχέα. Το InterVoice σύστημα φωνητικής πληκτρολόγησης βγήκε στην αγορά της τραπεζικής βιομηχανίας σαν μια μέθοδος επέκτασης των τραπεζικών μέσω τηλεφώνου υπηρεσιών σε πελάτες με περιστροφική κλήση τηλεφώνων. Σύντομα κι άλλες IVR εταιρίες άρχισαν να εγκρίνουν την τεχνολογία αναγνώρισης ομιλίας, συνενώνοντάς τη με τα δικά τους προϊόντα και ενσωματώνοντας αυτή στα μηχανήματά τους. Από το 1992, πιο πολλές βιομηχανίες άρχισαν να συνενώνουν την αναγνώριση ομιλίας μέσα στα ιδιωτικά τηλεφωνικά τους δίκτυα.

Η χρήση ολοκληρωμένων κυκλωμάτων ή σειρών ολοκληρωμένων κυκλωμάτων επίσης, άρχισε στις αρχές του 1990 και αποτελεί μέρος της γενικής ροπής του υλικού των ηλεκτρονικών υπολογιστών προς τη σμίκρυνση. Τα συστήματα επιπέδου ολοκληρωμένων κυκλωμάτων επιτρέπουν σ' αυτή τη τεχνολογία να ενσωματωθεί μέσα σε μικρές τηλεφωνικές χειροσυσκευές, οι οποίες αρχικά χρησιμοποιούνται για φωνητική πληκτρολόγηση.

Ένας από τους πιο συχνά εγκωμιαζόμενους λόγους όπου η αναγνώριση ομιλίας χρησιμοποιείται στις επιχειρήσεις, είναι η επέκταση υπηρεσιών στους πελάτες που έχουν τηλέφωνο με περιστροφική κλήση. Αν και η τεχνολογία τονικής επαφής (touch-tone) κυριαρχεί στις Ηνωμένες Πολιτείες της Αμερικής, σπάνια χρησιμοποιείται εκτός Βορείου Αμερικής. Τυπικά, οι δυνατότητες υπηρεσιών στο πελάτη οι οποίες επεκτείνονται μοιάζουν με αυτές που ήδη προσφέρονται στους πελάτες με τη τεχνολογία τονικής επαφής. Υπάρχουν αρκετές καταστάσεις οι οποίες κάνουν την αναγνώριση ομιλίας για αυτό το σκοπό, ιδιαίτερα ωφέλιμη:

- Υπηρεσίες επί εικοσιτετράωρου (και Σαββατοκύριακο) βάσεως
- Υπηρεσίες σε πολύ μακρινές αγορές
- Υπηρεσίες υποστήριξης σε ώρες αιχμής

Οι πελάτες με τηλέφωνα περιστροφικής κλήσης δεν χρειάζεται να περιμένουν σε ώρες εργασίας για να αποκτήσουν πρόσβαση σε έναν οργανισμό, μόνο και μόνο για να μάθουν ότι περιμένουν σε μια σειρά για υπηρεσίες. Πολλές κλήσεις μπορούν να χειριστούν αυτόματα, ακόμη κι αν ο πελάτης δεν έχει τηλέφωνο τονικής επαφής. Επίπεδα υπηρεσιών αυτού του

είδους παρουσιάζουν ένα ανταγωνιστικό πλεονέκτημα για επιχειρήσεις, οι οποίες έχουν επιλέξει τη χρήση αναγνώρισης ομιλίας.

Ένας δεύτερος επιχειρησιακός λόγος για τη χρήση αναγνώρισης ομιλίας στο τηλέφωνο, είναι η μείωση του κόστους προσωπικού για υπάλληλους υποδοχής και υπηρεσίες απάντησης. Μπορεί να χρησιμοποιηθεί για παράδειγμα, από επιχειρήσεις που ενδιαφέρονται για τη προσφορά υπηρεσιών σε πελάτες επί εικοσιτετραώρου βάσεως χωρίς να αυξήσουν το κόστος προσωπικού για απογευματινές και διημέρου υπηρεσίες. Ένα ασυνήθιστο παράδειγμα είναι η υλοποίηση ενός συστήματος ομιλίας σε K-Mart, το οποίο επιτρέπει τη πρόσβαση υπαλλήλων σε πληροφορία αποθηκευμένη σε εταιρικές βάσεις δεδομένων. Τα κέρδη κόστους αυξάνονται αμέσως.

Η προσωπική χρήση της αναγνώρισης ομιλίας (α.ο.) επικεντρώνεται σε μεγάλο βαθμό στη φωνητική πληκτρολόγηση. Παρέχει ασφάλεια στα οχήματα επειδή επιτρέπει στους οδηγούς να έχουν τα μάτια τους στο δρόμο και τα χέρια τους στο τιμόνι, ενώ πληκτρολογούν στο τηλέφωνο του αυτοκινήτου. Η χρήση της α.ο. στην κλήση τηλεφώνου στα οχήματα είναι επιτακτική σε μερικές πολιτείες, σημειώνουμε την Kalifornia και το Michigan, όπου έχουν προβλεφθεί νόμοι που υποχρεώνουν τους οδηγούς να κρατούν τα χέρια τους στο τιμόνι καθώς τηλεφωνούν.

### **3.5.1 Σχεδιαστικά θέματα χρήσης τηλεφώνου**

#### **1. Μέγεθος λεξιλογίου και επιλογή**

Μέχρι η AT&T να ανακοινώσει τη συγχώνευση του μοντέλου υπολέξεων στο Conversant σύστημα το 1993, οι εμπορικές τηλεφωνικές λειτουργίες ήταν περιορισμένες σε πολύ μικρά λεξιλόγια. Πολλά απ' αυτά είχαν λιγότερες από είκοσι λέξεις. Η χρήση των ψηφίων 'ναι' και 'όχι' και μια μικρή επιλογή λέξεων ελέγχου είναι ακόμη κατάλληλες για μερικές διεργασίες, αλλά η εισαγωγή του μοντέλου υπολέξεων προσφέρει στους σχεδιαστές μεγαλύτερες δυνατότητες λεξιλογίου.

Η μοντελοποίηση υπολέξεων έχει απελευθερώσει την ανάπτυξη λεξιλογίου από την εντατική δειγματοληψία, και ο χρόνος ανάπτυξης λεξιλογίου έχει μειωθεί από μήνες σε λεπτά. Από το 1994, η μοντελοποίηση υπολέξεων άρχισε να εμφανίζεται σαν σχεδιαστικό χαρακτηριστικό και άλλων προϊόντων.

Η μοντελοποίηση υπολέξεων επίσης, απελευθερώνει την α.ο. από το μοντέλο λειτουργίας με τονική επαφή. Οι δυνατότητες απάντησης που παρουσιάζονται στους καλούντες μπορούν τώρα να είναι αρκετά πιο ουσιαστικές. Είναι πιο εύκολο για τους καλούντες να θυμούνται να πουν 'ρουχισμός' για να αναφερθούν στο τμήμα ρουχισμού, παρά να πουν το '3', κάνοντας τη συνομιλία πιο απλή, πιο γρήγορη, πιο κατανοητή και πιο ακριβής.

Από το 1994, το σύστημα Conversant μπορούσε να προγραμματιστεί ώστε να χρησιμοποιεί ένα ενεργό λεξιλόγιο δύο χιλιάδων λέξεων. Ένα τόσο μεγάλο λεξιλόγιο μπορεί να είναι αρκετό για τις ανάγκες των περισσότερων εισόδου και απάντησης τηλεφωνικών λειτουργιών, αλλά χρησιμοποιεί τη λειτουργία νέων τύπων λειτουργιών. Ακόμη και στην υπάρχουσα μορφή, ένα μεγαλύτερο ενεργό λεξιλόγιο είναι δυνατό να συμπεριλάβει μεγαλύτερο αριθμό συνωνύμων και σταθερών λέξεων. Επεκτείνει επίσης τον αριθμό επιλογών που παρουσιάζονται σε μια λίστα. Έρευνα στο US WEST ανακάλυψε ότι ο αριθμός επιλογών που παρουσιάζονται στον καλούντα είναι μεγάλος, εάν αυτός μπορεί να πατήσει barge-in (παρεμβαίνω) όταν η επιθυμούμενη επιλογή έχει σημειωθεί. Μεγαλύτερες λίστες παρουσιάζονται από τη μοντελοποίηση υπολέξεων επειδή οι επιλογές είναι πιο ουσιαστικές.

Η χρήση της μοντελοποίησης υπολέξεων έχει δύο όψεις. Αυξάνει ναι μεν τη φυσικότητα και την ελαστικότητα μιας λειτουργίας, αλλά από την άλλη η ακρίβεια

αναγνώρισης είναι μικρότερη για λεξιλόγιο δημιουργημένο με μοντελοποίηση υπολέξεων σε σύγκριση με το λεξιλόγιο που δημιουργείται με δειγματοληψία. Ένας σχεδιαστής κατά συνέπεια, πρέπει να είναι προσεκτικός ώστε να αυξάνει τις ακουστικές διαφορές στο ενεργό λεξιλόγιο.

Το Αγγλικό αλφάβητο αποτελεί μια σημαντική σειρά λεξιλογίου για πολλές τηλεφωνικές λειτουργίες, αλλά ακόμη χρησιμοποιείται σπάνια. Εάν τυχόν χρειαστεί ορθογραφία σε μια διεργασία, ακριβής αναγνώριση μπορεί να επιτυγχανθεί μέσω προσεκτικής σχεδίασης. Η υλοποίηση στρατιωτικού αλφαβήτου είναι γενικά ακατάλληλη για τηλεφωνικές λειτουργίες, επειδή αρκετές απ' αυτές περιμένουν χρήστες ενός χρόνου (one-time) και επειδή τελικά, όλα τα τηλεφωνικά συστήματα σχεδιάζονται για να χρησιμοποιηθούν από μια εξαιρετικά απαιτητική ομάδα χρηστών: τους καταναλωτές.

## **2. Μεταγλώττιση**

Η ενεργοποιημένη φωνητική πληκτρολόγηση μετατρέπεται σε touch-tone ή παλμικού Τόνου σήματα. Η μεταγλώττιση όλων των άλλων εισόδων εξαρτάται αποκλειστικά από τη φύση και τις επικοινωνιακές απαιτήσεις της λειτουργίας.

## **3. Γλωσσική δομή**

Το πιο ευρέως χρησιμοποιούμενο γλωσσικό σχήμα είναι η τετελεσμένης κατάστασης γραμματική, ή μια συγκρινόμενη, υψηλά δομημένη τεχνική. Οι σχεδιαστές ειδικευμένοι στις τηλεφωνικές λειτουργίες συχνά ονομάζουν αυτή τη δομημένη προσέγγιση *scripting*. Όπως άλλοι τύποι τετελεσμένης κατάστασης γραμματικής ανάπτυξης, ο *scripting* ενδιαφέρεται για τη συχνότητα γεγονότων που μπορούν να συμβούν σε μια λειτουργία, δίνει όμως πολύ μεγαλύτερη έμφαση στη σχεδίαση καλών εντολών.

Άλλη μια τεχνική η οποία γίνεται γνωστή είναι ο εντοπισμός (ή στόχευση) λέξεων. Ο εντοπισμός δεν εξουδετερώνει την ανάγκη του *scripting* της διεργασίας, αλλά αφαιρεί τις αρχικές γλωσσικές απαιτήσεις που εναποτίθενται στον καλούντα. Είναι μια από τις τεχνολογίες που έχει συνεισφέρει στην επιτυχία της α.ο. μέσω τηλεφώνου, και είναι εξαιρετικά χρήσιμη για διαφορετικούς πλυθησμούς ομιλητών που θα εκτελέσουν μια λειτουργία μόνο μια φορά ή σπάνια. Ο εντοπισμός λέξεων αυξάνει τη πιθανότητα ότι αυτοί θα επιτύχουν το σκοπό τους, ακόμη κι αν δεν περιορίσουν τα λεγόμενά τους σε λέξεις ή φράσεις που απαιτούνται από τις εντολές. Αν εξαιρεθεί η αυξανόμενη ευρωστία, ο εντοπισμός λέξεων δεν είναι μια αλάνθαστη και προσεκτική σχεδίαση διόρθωσης λάθους, οπότε διαδικασίες υποχώρησης (fallback) χρειάζονται για να εμποδίσουν αναγνώριση λαθών από λάθος μονοπάτι του καλούντος.

## **4. Μοντέλο ομιλητή**

Μερικές λειτουργίες (όπως η φωνητική πληκτρολόγηση) μπορούν να είναι εξαρτημένες από τον ομιλητή, αλλά οι περισσότερες φωνητικές λειτουργίες απαιτούν ανεξάρτητη του ομιλητή αναγνώριση επειδή έχουν σχεδιαστεί για μεγάλο πληθυσμό χρηστών.

Ένα ανεξάρτητο του ομιλητή μοντέλο πρέπει πάντα να δοκιμάζεται πάνω σ' ένα δείγμα ομιλητών, οι οποίοι θα χρησιμοποιήσουν μια διεργασία για να επιβεβαιώσουν ότι τα πρότυπά τους ομιλίας συμπεριλαμβάνονται στο μοντέλο. Τέτοια δοκιμασία είναι εξαιρετικά σπουδαία όταν η μοντελοποίηση υπολέξεων έχει χρησιμοποιηθεί, επειδή οι υπολέξεις μπορεί

να μην εναρμονίζονται με τα συγκεκριμένα χαρακτηριστικά του ομιλητή του πληθυσμού χρηστών.

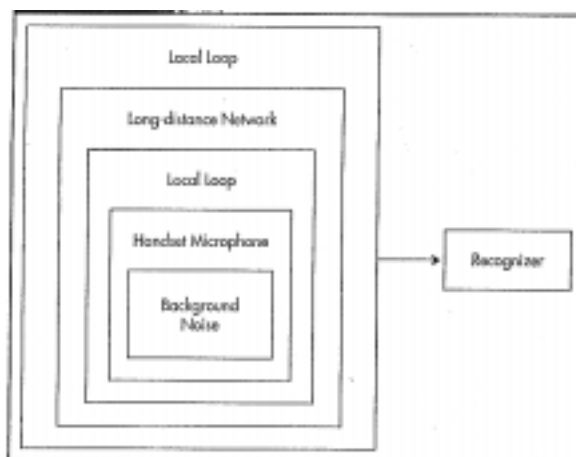
## 5. Ροή ομιλίας

Εάν το λεξιλόγιο είναι καλά καθορισμένο αλλά η συμπεριφορά του πληθυσμού των ομιλητών απρόβλεπτη, ο εντοπισμός λέξεων είναι μια σωστή εκλογή. Διεργασίες που χρησιμοποιούν εντοπισμό λέξεων περιμένουν στην είσοδο συνεχόμενη ομιλία, αλλά συγχρόνως την ψάχνουν για συγκεκριμένα κλειδιά-λέξεις.

Διακριτής λέξης και συνεχόμενης ομιλίας αναγνώριση (όχι χρησιμοποιώντας εντοπισμό λέξεων) είναι επίσης ευρέως χρησιμοποιούμενες για προσανατολισμένων ψηφίων (digit-oriented) είσοδο. Τα συστήματα διακριτής λέξης διασχίζουν (pace) την είσοδο εισάγοντας έναν τόνο ή μπίπ μεταξύ των λέξεων, για να αναδείξουν το κενό μεταξύ αυτών. Σιγά-σιγά αυτό εκτοπίζεται από την συνεχόμενη ομιλία είσοδο, το οποίο είναι γρηγορότερο και πιο φυσικό.

### 3.5.2 Περιβάλλον ομιλίας χρησιμοποιώντας το τηλέφωνο

Τα συστήματα αναγνώρισης ομιλίας για το τηλεφωνικό κανάλι πρέπει να είναι πολύ εύρωστα. Το εύρος του καναλιού είναι μικρό και το κανάλι το ίδιο είναι θορυβώδες. Η επόμενη εικόνα εμφανίζει τους τύπους θορύβου τους οποίους ένα σύστημα α.ο. πρέπει να αντιμάχεται. Η ομιλία βγαίνει σ' ένα τηλέφωνο πλημμυρισμένο από εσωτερικό θόρυβο, ο



οποίος ποικίλλει από την ιδεατή ησυχία (virtual silence) έως το βρυχηθμό (roar). Ένα σύστημα αναγνώρισης έχει λίγη προετοιμασία για τον εσωτερικό θόρυβο που πλημμυρίζει τη φωνή του καλούντος. Η ομιλία και ο εσωτερικός θόρυβος μορφοτροπούνται από ένα μικρόφωνο με άγνωστες ιδιότητες και ποιότητα. Το μικρόφωνο συνεισφέρει το δικό του θόρυβο τόσο πολύ όσο και οι παραπονημένες μορφές του χαρακτηριστικού σήματος. Το μορφοτροποποιημένο, ορισμένο σήμα μεταφέρεται μέσω ενός τοπικού τηλεφωνικού βρόχου στο κεντρικό γραφείο της τηλεφωνικής εταιρίας. Ο βρόχος είναι κομμάτι του τοπικού τηλεφωνικού δικτύου. Εισάγει πιο πολύ θόρυβο μέσα στο σήμα. Η κρούση του θορύβου στο τοπικό βρόχο ποικίλλει από δίκτυο σε δίκτυο και επηρεάζεται από τις τεχνικές κωδικοποίησης που χρησιμοποιούνται, την ηλικία των συστατικών μερών του δικτύου, τη συντήρηση και

άλλους παράγοντες. Εάν η κλήση πραγματοποιείται από μια εξωτερική περιοχή κλήσης, πρώτα διασχίζει το τοπικό βρόχο, έπειτα μπαίνει στο μεγάλης απόστασης δίκτυο (ή δίκτυα) όπου αντιμετωπίζει άλλη μια σειρά χαρακτηριστικών θορύβου και παραμόρφωσης, πριν εισαχθεί στο δεύτερο τοπικό τηλεφωνικό βρόχο.

Η κατάσταση είναι χειρότερη όταν η μετάδοση γίνεται μέσω ενός κυτταρικού τηλεφωνικού δικτύου. Η ποιότητα μετάδοσης είναι γενικά φτωχότερη απ' ό,τι όταν γίνεται με τις γειωμένες γραμμές τηλεφώνου. Επιπλέον, το σύστημα αναγνώρισης πρέπει να αντιμάχεται σήματα ποικίλης ισχύος το οποίο μπορεί να προκαλέσει εξασθένηση της φωνής.

Η καλύτερη αναγνώριση παράγεται από τα συστήματα επιπέδου ολοκληρωμένων κυκλωμάτων, τα οποία ενσωματώνονται στις τηλεφωνικές χειροσυσκευές. Αυτά τα συστήματα ρυθμίζονται στα μικροφωνικά χαρακτηριστικά της χειροσυσκευής κι έτσι αποφεύγεται ο επιπρόσθετος δικτυακός θόρυβος, επειδή η αναγνώριση γίνεται από το ίδιο το τηλέφωνο. Αφού τα επίπεδα ολοκληρωμένων κυκλωμάτων κυτταρικά τηλεφωνικά συστήματα χρησιμοποιούνται συχνά στα αυτοκίνητα, πρέπει να είναι ικανά να χειριστούν τις προκλήσεις του περιβάλλοντος του αυτοκινήτου, ειδικά αν είναι τοποθετημένα στη θέση τους και χρησιμοποιούνται σαν μικρόφωνα.

### **3.5.3 Ανθρώπινοι παράγοντες στη χρήση τηλεφώνου**

Όπως μαζί με το IVR χρησιμοποιείται τονικής επαφής είσοδο, η καταλληλότητα της διαδικασίας αυτόματης κλήσης πρέπει να συγκριθεί με τις ανάγκες των καλούντων. Όταν το μονοπάτι της κλήσης είναι περίπλοκο, όταν οι καλούντες συχνά απαιτούν πιο πολλή βοήθεια απ' ό,τι ένα αυτοματοποιημένο σύστημα μπορεί να προσφέρει, ή όταν οι υπάλληλοι παροχής υπηρεσιών προσφέρουν βαρετά πληροφορίες στους καλούντες, η αυτοματοποίηση δημιουργεί μια αίσθηση φτωχών υπηρεσιών πελάτη.

Εάν ο πληθυσμός ομιλητών αποτελείται βασικά από εκλεπτυσμένους τηλεφωνικούς χρήστες, όπως είναι οι επαγγελματίες επιχειρήσεων, αυτοί θα είναι εξοικειωμένοι με IVR διεργασίες οι οποίες απαιτούν τονικής επαφής είσοδο. Αυτό το παράδειγμα κάνει το σχεδιασμό των ανθρώπινων παραγόντων πιο απλό, επειδή παρέχει ένα βασικό εντολής-και-απάντησης μοντέλο το οποίο καθοδηγεί τον καλούντα μέσω μιας συγκρινόμενης διεργασίας αναγνώρισης ομιλίας. Αντίστροφα, κάθε παραβίαση του εδραιωμένου παραδείγματος, συμπεριλαμβανομένων και ερωτήσεων για απαντήσεις λέξεων από αριθμούς, μπορεί να μπερδέψει τον καλούντα. Η σχετικά μικρή ακρίβεια της α.ο. σε σύγκριση με την τονικής επαφής είσοδο, μπορεί να ενοχλήσει τους καλούντες οι οποίοι δεν θα λάβουν υπόψη τους αυτά τα πράγματα σαν περιβαντολογικό θόρυβο.

Οι μη τεχνολογικά καταρτισμένοι καλούντες μπορεί να μην έχουν το IVR παράδειγμα για να καθοδηγηθούν και μπορεί να μπερδευτούν, να τρομοκρατηθούν από τη τεχνολογία ή να θυμώσουν επειδή πρέπει να μιλήσουν σε ηλεκτρονικό υπολογιστή. Αυτοί οι χρήστες είναι πιο πιθανό να παράγουν άσχετη ομιλία και να απαντούν αργά στις εντολές. Η έλλειψη ενός γνωστού παραδείγματος μπορεί να επιφέρει ένα μεγάλο ποσοστό άσχετης ομιλίας. Η χρήση του εντοπισμού λέξεων όμως, διευκολύνει την αναγνώριση χωρίς να παρεμβαίνει στις φυσικές μορφές ομιλίας αυτών των χρηστών.

Κάθε τύπος χρήστη μπορεί να εκφράσει έκπληξη όταν εμφανιστεί μπροστά του ένα σύστημα α.ο. Αναφωνήσεις όπως 'Ει, μιλάω σ' έναν υπολογιστή!' δεν είναι ασυνήθιστες και παράγουν λάθη αναγνώρισης.

Το τονικής επαφής παράδειγμα παρουσιάζει ένα εμπόδιο όταν προτρέπει τους σχεδιαστές να μοντελοποιήσουν λειτουργίες α.ο. σ' αυτό. Αυτή η στρατηγική επιφέρει καλές και κακές σχεδιαστικές αποφάσεις. Ένα παράδειγμα καλής εκλογής είναι το βασικό σχήμα των απλών εντολών. Εμπειρία στο σχεδιασμό συστήματος τονικής επαφής έχει αποδείξει ότι



η συμπεριφορά του καλούντος επηρεάζεται πάρα πολύ από το τρόπο που μια εντολή ή ανακοίνωση γίνεται λέξη. Η μορφή

Για τις πωλήσεις, λέγε 'ένα'  
Για τη τεχνική υποστήριξη, λέγε 'δύο'

για παράδειγμα, παράγει συνέχεια πιο ακριβείς απαντήσεις από

Λέγε 'ένα' για τις πωλήσεις  
Λέγε 'δύο' για τη τεχνική υποστήριξη.

Μαζί με τις τονικής επαφής λειτουργίες

- Η γλώσσα μιας εντολής πρέπει να είναι καθαρή
- Οι επιλογές απάντησης πρέπει να είναι απλές και καθαρές
- Οι προεντολές γενικά, πρέπει να παρέχουν επιπλέον διευκρινίσεις
- Η ταχύτητα του συστήματος απάντησης στην είσοδο πρέπει να είναι αποδεκτή
- Οι καλούντες πρέπει να νιώθουν ότι επιτυγχάνουν τους σκοπούς τους
- Πρέπει να υπάρχει μια μέθοδος υποχώρησης σε περίπτωση αποτυχίας.

Η χρήση της εμπειρίας τονικής επαφής οδηγεί στη πράξη στο περιορισμό του αριθμού των επιλογών σε τρεις ή τέσσερις. Αυτή η υπόθεση έχει μεταφερθεί στις λειτουργίες α.ο. , αλλά πρόσφατα πειράματα της U S WEST προτείνουν ότι όταν οι χρήστες μπορούν να απαντήσουν με τοπικά σχετιζόμενες λέξεις, όπως 'πωλήσεις' ή 'τεχνολογική υποστήριξη', η λίστα μπορεί να είναι αρκετά μεγαλύτερη. Πιο εκτεταμένη έρευνα χρειάζεται για να διευκρινίσει καταστάσεις όπου τα όρια των μενού είναι κατάλληλα. Η «κοινής λογικής» διαπίστωση ότι οι άνθρωποι θα απαντήσουν με λέξεις καθορισμένες ως εύλογες απαντήσεις σε μια εντολή, αποδείχτηκε λάθος από τα εργαστήρια της AT&T Bell σε έρευνα που οδήγησε στη δική τους χρησιμότητα του εντοπισμού λέξεων. Η DSMI (Disease State Management Inc.) αντιμετώπισε ένα παρόμοιο φαινόμενο όταν δοκίμασε πως οι καλούντες απαντούσαν στην εντολή 'πες τον αριθμό πέντε'.

*'Είπαν 'κωδικός πέντε', συλλάβισαν πέντε και χτύπησαν το νούμερο 5 στο τονικής επαφής μονοπάτι-κλειδί'' (Leon Zak, Director of Systems, DSMI, personal communication, 1995).*

Με τις διευρυμένες επιλογές λεξιλογίου που εξασφαλίζει η μοντελοποίηση υπολέξεων, είναι πιθανό ότι αρκετές νέες γνώσεις θα αποκτηθούν μέσω των λαθών και της εμπειρίας.

Καθώς οι λειτουργίες α.ο. αναπτύσσονται σε λεξιλόγιο και ελαστικότητα, ταυτόχρονα αποχωρίζονται από το καθιερωμένο εντολής-και-απάντησης παράδειγμα. Η πρόκληση των ανθρώπινων παραγόντων είναι να καθορίσουν ένα νέο παράδειγμα συνομιλίας το οποίο θα γίνει αποδεκτό από ανθρώπους που χρησιμοποιούν τέτοια συστήματα.

### 3.5.4 Παράδειγμα κλήσης τηλεφώνου (Διαχείριση κλήσης, Wildfire Επικοινωνίες)

Η Wildfire Ηλεκτρονική Βοήθεια συντονίζει τις ενέργειες διαχείρισης κλήσης οι οποίες γίνονται από τους πιο πολλούς επαγγελματίες επιχειρήσεων. Αυτή λοιπόν, συνενώνει ένα μεγάλο εύρος ποικίλων διαχειριστικών διεργασιών κλήσης, συμπεριλαμβανομένων των:

- Διάθεση (placing) (χρησιμοποιώντας ενεργοποιημένη φωνητική πληκτρολόγηση) και επιστροφή κλήσεων
- Παραλαβή και ανακοίνωση εισαγόμενων κλήσεων στο χρήστη
- Παραλαβή μηνυμάτων
- Εγγραφή φωνητικού ταχυδρομείου μηνυμάτων (playing voicemail messages)
- Προγραμματισμός συνεχόμενων (follow-up) κλήσεων
- Προγραμματισμός υπενθυμίσεων στο χρήστη
- Προώθηση κλήσεων σε σελιδοποιούς (pagers) ή άλλα νούμερα
- Ενεργοποίηση χρηστών για τον ορισμό επιπέδων προτεραιότητας σε επαφές για μελλοντικό χειρισμό κλήσεων.

Η ποιότητα της Wildfire δεν οφείλεται σε πολύ εξελιγμένη τεχνολογία αναγνώρισης ομιλίας. Χρησιμοποιεί διακριτής λέξης αναγνώριση και εντοπισμό λέξεων, ανεξάρτητα και εξαρτημένα του ομιλητή συστήματα ομιλητή και το λεξιλόγιό της είναι μέτριο. Είναι πιο πολύ το αποτέλεσμα μιας προσεκτικής ενδοένωσης ξεχωριστών λειτουργιών και μιας υπερφυσικής ανθρώπινης διασύνδεσης. Αντικαθιστά πολύπλοκες τονικής επαφής σειρές κλειδιών με απλές λεκτικές εντολές, όπως 'κλήση γραφείου' για να γίνει μια κλήση ή 'θα το πάρω' σε απάντηση ότι έχει λεχθεί, ότι ένα συγκεκριμένο πρόσωπο καλεί. Εάν μια εισαγόμενη κλήση φτάσει ενώ μια άλλη βρίσκεται σε διαδικασία, η ανακοίνωση της Wildfire του ονόματος του νέου καλούντος ακούγεται μόνο από το χρήστη που μπορεί να δώσει οδηγία στη Wildfire να πάρει το μήνυμα, ή να τοποθετήσει την τρέχουσα κλήση σε αναμονή ενώ η νέα κλήση προωθείται. Οι χρήστες αρχικοποιούν εντολές στη Wildfire καλώντας τες με το όνομά τους. Το σύστημα απαντάει 'εδώ είμαι' ή 'Τι μπορώ να κάνω για σένα;', μιλώντας με μια φιλική παιδική φωνή. Όταν δοθεί μια εντολή μπορεί να απαντήσει 'Εντάξει' ή 'Το έπιασα'. Ένας τυπικός διάλογος για προγραμματισμό υπενθύμισης μπορεί να είναι ως εξής:

<u>Χρήστης</u>	<u>Wildfire</u>
Θυμισέ μου	Τι μπορώ να κάνω για σένα;
<Θέμα. Ησυχία μετά>	Τι να σου υπενθυμίσω;
Αύριο	Πότε;
8.30	Τι ώρα;
Το πρωί	Πρωί ή βράδυ;
Ναι	Αύριο στις 8:30 π.μ. Είναι σωστό;
Γειά Wildfire	Θα στο υπενθυμίσω αυτή την ώρα
	Ευχαριστώ, γειά.

Η Wildfire σχεδιάστηκε να χρησιμοποιείται από εξερχόμενους και εσωτερικούς (on-site) υπαλλήλους εταιρίας. Ανήκει στο εσωτερικό τηλεφωνικό δίκτυο μιας εταιρίας και συνδέεται με το δημόσιο τηλεφωνικό δίκτυο με μια T-1 γραμμή, έτσι ώστε να μπορεί να έχει κανείς πρόσβαση μαζί της από οποιοδήποτε τηλέφωνο του δημόσιου δικτύου.

#### **4. Λειτουργίες επιτρεπόμενης ομιλίας (speech-enabled)**

Ένα σύστημα επιτρεπόμενης ομιλίας (καλείται επίσης ενήμερη ομιλία) δεσμεύει τις συναρτήσεις (functions) μιας εφαρμογής στην ομιλούμενη είσοδο. Η ομιλία είναι μέρος της εφαρμογής. Αυτή η σύνδεση επιφέρει γρήγορες απαντήσεις και μεγαλύτερη συνέπεια, ενώ μειώνει τη πιθανότητα ότι η διασύνδεση ομιλίας θα χάσει το συγχρονισμό με το υπόλοιπο σύστημα. Ο Worker's Compensation Reserving Advisor για παράδειγμα, ενσωμάτωσε τον έλεγχο της ομιλούμενης διασύνδεσης μέσα στους ειδικευμένους κανόνες του συστήματος. Ο επόμενος κανόνας ενεργοποιεί γραφικά και ομιλία το ένα πίσω από το άλλο.

IF

The INJURY-TYPE is FRACTURE and  
Injury SEVERITY is UNKNOWN

THEN

ACTIVATE speech input for FRACTURE-SEVERITY  
DISPLAY SPEECH-INPUT instructions-display-screen  
ACTIVATE menu-popup screen for FRACTURE-SEVERITY

Όταν έχει πραγματοποιηθεί σωστά ένα σύστημα επιτρεπόμενης ομιλίας, μπορεί να βελτιώσει τη φυσικότητα του συστήματος σχεδιάζοντας τη διεργασία να περιμένει είσοδο, η οποία προσαρμόζεται στις κανονικές μορφές ομιλίας. Τελικά, αυτό θα εισάγει τη δυνατότητα της αντικατάστασης μιας γραφικής διασύνδεσης, ακριβώς όπως η GUI απελευθέρωσε τους χρήστες από τη γραμμή εντολών.

*“Αυτή είναι μια σημαντική περιοχή όπου οι χρήστες θα αρχίσουν να επιτυγχάνουν παραγωγικότητα και χρήσιμα οφέλη. Αντί να παρέχει ένα διαφορετικό τρόπο εξερεύνησης διαλόγων, η ομιλία μπορεί να διαπεράσει εντελώς τη GUI διασύνδεση και να εκτελέσει την πρόθεση του χρήστη”* (Bruce Armstrong, Manager of the Novell Speech Technologies Group, WordPerfect, The Novell Applications Group, “Speech recognition application program interface committee”, 1994, p. 23).

Με λίγα λόγια, το τελικό επιτρεπόμενης ομιλίας σύστημα είναι ικανό να επικοινωνήσει με ανθρώπους με ένα φυσικό, περιορισμένου σκοπού τρόπο. Αυτός ο τύπος συστήματος πρέπει να περιμένει την ενοποίηση της ομιλίας με ένα νοήμων υπολογιστικό λογισμικό, αλλά τα συστήματα επιτρεπόμενης ομιλίας πιο απλής μορφής έχουν ήδη αναπτυχθεί.

#### **5. Εκτίμηση**

Η εκτίμηση αφορά τη μέτρηση της απόδοσης ενός συστήματος. Μπορεί να εκφραστεί με την εκτίμηση των:

- Αλγοριθμικών μεθόδων (χρησιμοποιούνται στα συστήματα έρευνας)
- Προϊόντων για χρήση σε μια συγκεκριμένη εφαρμογή

- Της εφαρμογής της ίδιας.

Η εκτίμηση μπορεί να απαιτεί μεγάλο χρονικό διάστημα. Πολύ συχνά θεωρείται δευτερεύουσα και ξεχωριστή από τη σχεδίαση εφαρμογής. Η αλήθεια είναι ότι η υλοποίηση ενός καλού επίσημου προγράμματος εκτίμησης μπορεί να είναι κρίσιμη για την επιτυχή ανάπτυξη ενός συστήματος. Μερικές από τις αρχικές λειτουργίες εκτίμησης είναι:

## **1. Επιλογή ενός κατάλληλου αλγόριθμου ή προϊόντος για μια συγκεκριμένη λειτουργία**

Η εκτίμηση ενός προϊόντος συχνά πραγματοποιείται έχοντας υπόψη ένα συγκεκριμένο σκοπό. Ο αριθμός των καλών εμπορικών προϊόντων αναγνώρισης ομιλίας αυξάνεται ραγδαία. Κάθε προϊόν σχεδιάζεται για συγκεκριμένους τύπους λειτουργιών και παρουσιάζει μοναδικές ικανότητες. Μέχρι τα δεδομένα της βιομηχανίας για λειτουργία προγραμματικών διασυνδέσεων (API) να καθοριστούν πλήρως, το κόστος της επιλογής ενός ακατάλληλου προϊόντος μπορεί να είναι μεγάλο.

Τα API δεδομένα διευκολύνουν την αλλαγή αναγνωριστών αφού η εφαρμογή έχει δημιουργηθεί, αλλά δεν μπορούν να εκμηδενίσουν την ανάγκη για εξήγηση των μελλοντικών διαφορών μεταξύ των προϊόντων. Ούτε τα API δεδομένα εγγυώνται ότι ένα δεδομένο προϊόν θα λειτουργήσει σωστά σ' ένα περιβάλλον μιας λειτουργίας, ή με τον προσδοκώμενο πληθυσμό χρηστών. Τα προϊόντα θα συνεχίσουν να διαφέρουν σ' αυτούς τους τομείς και θα χρειαστεί να δοκιμαστούν για μια λειτουργία.

*“Τυπικά πολλοί πωλητές παραθέτουν απόδοση της τεχνολογίας τους κατά 98-99% ακρίβεια, χωρίς ποτέ να ξεκαθαρίζουν τις συνθήκες κάτω από τις οποίες η δοκιμασία έγινε. Χωρίς να ξέρουν κάτι από τη μορφή του λεξιλογίου που χρησιμοποιήθηκε στο test, πόσοι ομιλητές και δείγματα ομιλίας χρησιμοποιήθηκαν, πως εξελίχθηκε η εκπαίδευση ή τι αναλογία σήματος προς θόρυβο ή μικρόφωνο που χρησιμοποιήθηκε, είναι επιστημονικά αδύνατο να καθορίσουν τι 99% σωστή αναγνώριση ακριβώς σημαίνει, ή πως συγκρίνεται με άλλες μετρήσεις απόδοσης”* (David Pisoni, Howard Nusbaum & Subrata Das, “Automatic measurement of speech recognition performance”, 1986, p. 4).

Εν τέλει, είναι επιβεβλημένη η εκτίμηση των προϊόντων να γίνεται στο γενικό πλαίσιο της εφαρμογής.

## **2. Εκτίμηση της απόδοσης μιας εφαρμογής πριν την ανάπτυξη**

*“Είναι πολύ σημαντικό να αντιληφθούμε πως μετράμε την απόδοση αναγνώρισης ομιλίας. Αυτή είναι μια τεχνολογία διασύνδεσης χρήστη και την πρώτη φορά δεν θα φτάσετε την τελειότητα, ακόμη κι αν ανήκετε σ' έναν οργανισμό όπως η AT&T με μεγάλη εμπειρία. Ένας λόγος είναι ότι κάθε φορά που παρουσιάζετε μια νέα τεχνολογία σε ανθρώπους, δεν μπορείτε να προβλέψετε την αντίδρασή τους”* (Judith Tschirgi, Direction of Services and Speech Technology, AT&T Network Systems, personal communication, 1994).

Είναι πιο επικερδές και αποτελεσματικό από άποψη χρόνου, να προσδιορίσει κάποιος και στη συνέχεια να διορθώσει τις αδυναμίες σχεδίασης στη διάρκεια της ανάπτυξης, παρά να τροποποιήσει ένα σύστημα όταν αυτό έχει βγει στην αγορά. Εάν η ομιλία συγχωνεύεται στο προϊόν, λάθος υπολογισμοί μπορεί να βλάψουν την άποψη του πελάτη για το προϊόν και την εταιρία. Αυτό είναι αλήθεια αν η τεχνολογία αναπτύχθηκε από τον κατασκευαστή του

προϊόντος ή από μια εξωτερική πηγή. Από τη σκοπιά πάντως του χρήστη, δεν υπάρχει ‘δεν ανακαλύφθηκε εδώ’ για ένα συστατικό στοιχείο ή γνώρισμα του προϊόντος.

Όταν ο σχεδιαστής λειτουργίας δεν είναι εξοικειωμένος με την α.ο., οι ευκαιρίες για λάθη είναι πολλές. Αυτό επαυξάνεται από το γεγονός ότι η α.ο. είναι μια αναδυόμενη τεχνολογία διασύνδεσης. Όπως όλες οι τεχνολογίες ανθρώπινης διασύνδεσης, έτσι και η α.ο. είναι ευαίσθητη σε ψεγάδια του ανθρώπινου σχεδιασμού. Όντας ακόμη μια αναδυόμενη τεχνολογία, χαρακτηρίζεται από τεχνολογικές αλλαγές και από την έλλειψη ενός εύκολα προσβάσιμου σώματος γνώσης, όσον αφορά την ανθρώπινη σχεδίαση.

Εάν το σχέδιο είναι ένα πρωτότυπο σχεδιασμένο να εκτιμά το κέρδος της συνένωσης α.ο. μέσα σε μια εφαρμογή, ελαττώματα στον ανθρώπινο σχεδιασμό και άλλες όψεις του σχεδιασμού εφαρμογής μπορούν να αδυνατίσουν την ακρίβεια αναγνώρισης. Αυτό είναι πιο πιθανό να εξηγηθεί ως στοιχείο ελαττωμάτων στην τεχνολογία α.ο., παρά ένα συγκεκριμένο παράδειγμα μιας φτωχής ανθρώπινης-υπολογιστικής σχεδίασης διασύνδεσης. Τέτοιες απροσεξίες μπορεί να αποβούν μοιραίες στην αποδοχή της αναγνώρισης ομιλίας από έναν οργανισμό, και ατυχέστερα από χρήστες που συναντούν το αποτυχημένο σύστημα.

### **3. Καθορισμός βελτιώσεων σε υπάρχουσες λειτουργίες αναγνώρισης ομιλίας**

Τυπικά, συστηματική εκτίμηση μπορεί να εφαρμοστεί σ’ ένα υπάρχον σύστημα ή εφαρμογή η οποία δεν αποδίδει τα αναμενόμενα. Χρησιμοποιείται για να αναγνωρίζει προβλήματα και να δίνει ιδέες για την αντιμετώπιση αυτών.

*“Χωρίς συστηματική μέτρηση της απόδοσης αναγνώρισης κάτω από ελεγχόμενες συνθήκες, είναι σχεδόν αδύνατο να καθοριστεί εάν οι αλλαγές σ’ έναν αλγόριθμο αναγνώρισης θα επιφέρουν αξιόλογες αλλαγές στην απόδοση αυτής”* (David Pisoni, Howard Nusbaum & Subrata Das, “Automatic measurement of speech recognition performance”, 1986, p. 4).

### **4. Καθορισμός της αξίας της επιπλέον ομιλίας σ’ ένα υπάρχον σύστημα**

Η αναγνώριση ομιλίας μπορεί να ενσωματωθεί σε υπάρχοντα συστήματα ως επέκταση της λειτουργικότητας ή ως αύξηση της παραγωγικότητας. Αυτό πιθανόν συνεπάγεται την επανασχεδίαση της λειτουργίας ή του συστήματος. Όπως με τη δημιουργία νέων εφαρμογών και προϊόντων, είναι σημαντικό να δοκιμαστούν πλάνα για την επανασχεδίαση πριν την υλοποίηση. Τα αποτελέσματα της ενσωμάτωσης ομιλίας πρέπει τότε να συγκριθούν με τις υπάρχουσες προδιαγραφές. Αυτές οι συγκρίσεις θα βοηθήσουν στο να καθοριστεί εάν θα συνεχιστεί η προσθήκη της ομιλίας, και αν ναι, πως θα επιτευχθεί αυτή.

#### **5.1 Επίσημη και μη εκτίμηση**

Η επίσημη εκτίμηση αφορά πρωτόκολλα σχεδιαστικών δοκιμασιών, δημιουργία βάσεων δεδομένων των δειγμάτων ομιλίας και επιλογή δοκιμαστικού εξοπλισμού. Τα δοκιμαστικά πρωτόκολλα για εκτίμηση εφαρμογών εξετάζουν το βασικό σχεδιασμό λειτουργίας, ειδικά ανθρώπινου σχεδιασμού γνωρίσματα, και περιλαμβάνουν εκτίμηση της απόδοσης εφαρμογής με βάση τα:

- Πληθυσμούς στοχευόμενου ομιλητή (target speaker populations)
- Περιβάλλον στόχου
- Λεξιλόγιο
- Σχεδίαση εφαρμογής και scripts
- Διασυνδέσεις εφαρμογής
- Ανθρώπινοι παράγοντες.

Ο επίσημος δοκιμασμός αυτού του τύπου οδήγησε τα AT&T Bell εργαστήρια στην ανακάλυψη ότι οι μιας φοράς (one-time) χρήστες συστημάτων βασισμένων σε τηλέφωνο, είναι πιθανό να προσθέσουν άσχετη ομιλία στις απαντήσεις τους. Αυτή η αποκάλυψη και η συνειδητοποίηση ότι η υπάρχουσα τεχνολογία α.ο. δεν θα μπορούσε να εφαρμοστεί σε τέτοιες εφαρμογές, οδήγησε στην επιλογή της τεχνολογίας εντοπισμού λέξεων.

Η μη επίσημη εκτίμηση δεν απαιτεί προκαταγραφή συστηματικών δοκιμασιών σχεδίασης ή εξοπλισμού. Ανεξάρτητα απ' αυτό, μπορεί ακόμη να είναι συστηματική, ειδικά αν χρησιμοποιείται ως μέθοδος εξερεύνησης των δυνατοτήτων ενός ή πιο πολλών προϊόντων.

Ένα μέτρο για την επιλογή επίσημης ή όχι προσέγγισης είναι το κόστος της αποτυχίας. Εάν αυτό είναι μικρό, τότε μια μη επίσημη ή ελάχιστα επίσημη εκτίμηση είναι λογική. Χαμηλού κόστους ενέργειες περιλαμβάνουν εκτίμηση των:

- Προσωπική χρήση των χαμηλού κόστους προϊόντων αναγνώρισης
- Απόκτηση βασικής εξοικείωσης της τεχνολογίας α.ο.
- Διατύπωση της βασικής έννοιας της εφαρμογής

Εάν το κόστος είναι υψηλό, μια λεπτομερής εκτίμηση πρέπει να αναπτυχθεί. Υψηλού κόστους ενέργειες περιλαμβάνουν εκτίμηση των:

- Ευρεία χρήση ενός συστήματος αναγνώρισης σ' έναν οργανισμό
- Ανάπτυξη ή βελτίωση ενός προϊόντος
- Υπηρεσία πελάτη.

Το πρωταρχικό ρίσκο της εφαρμογής μη επίσημης εκτίμησης σε υψηλού κόστους ενέργειες,

είναι η αποτυχία της εφαρμογής ή του προϊόντος. Οι πελάτες μπορούν να βάλουν ετικέτα σε ένα προϊόν ως φτωχής σχεδίασης ή άχρηστο, ο οργανισμός μπορεί να απορρίψει την τεχνολογία α.ο. ως ανώριμη, και έχοντας μάθει τίποτα από την εμπειρία ένας σχεδιαστής μπορεί να επιμείνει σε λάθος απόψεις για τη τεχνολογία.

## **5.2 Εκτίμηση προϊόντος**

Ένα απ' αυτά που έμαθαν από το APRA SUR τεχνολογικό έργο το 1970 ήταν, ότι είναι αδύνατο να συγκρίνεις συστήματα τα οποία διαφέρουν στη λειτουργία, στο λεξιλόγιο και στην κατασκευή. Πως θα μπορούσε το Hearsay-II, ένα σύστημα πίνακα σχεδιασμένο για ανάκτηση κειμένων να συγκριθεί με το HWIM, ένα σύστημα που συνδύαζε ένα φωνητικό πλέγμα με ένα επαυξημένης μετάδοσης δίκτυο σε μιας μικρού μπάτζετ λειτουργίας; Ήταν λοιπόν επόμενο ότι ο μόνος τρόπος για σωστή παρουσίαση της απόδοσης μιας ομάδας συστημάτων, ήταν να καταστεί βέβαιο ότι είχαν παρόμοια λεξιλόγια και παρουσίαζαν ίδια αποτελέσματα πάνω στην ίδια λειτουργία. Αυτό είναι το πνεύμα εκτίμησης προϊόντος περιορισμένης σε εφαρμογή.

Η εκτίμηση προϊόντος περιορισμένης σε εφαρμογή συγκρίνει δύο ή περισσότερα προϊόντα αναγνώρισης σε προεπιλεγμένα χαρακτηριστικά της εφαρμογής:

- Πληθυσμούς στόχου ομιλητή
- Περιβάλλον στόχου
- Λεξιλόγιο
- Πολυπλοκότητα (perplexity) εφαρμογής
- Διασυνδέσεις εφαρμογής.

Δεν υπάρχουν διεθνή δεδομένα για αξιολόγηση προϊόντων περιορισμένων σε εφαρμογή. Η επιλογή ενός προϊόντος α.ο. γίνεται κυρίως λόγω των χρήσεων στις οποίες η α.ο. θα τοποθετηθεί, και στις προτεραιότητες που είναι συνδεδεμένες μ' αυτές τις λειτουργίες. Η εγκυρότητα των αποτελεσμάτων εξαρτάται από τη ποιότητα των εργαλείων και των μεθόδων που χρησιμοποιούνται.

Υπάρχει ένα ενδιαφέρον στο σχεδιασμό πιο γενικών εκτίμησης προϊόντων εργαλεία. Αυτό υπήρξε ένα από τα αντικείμενα του ESPRIT Μεθοδολογίας Εκτίμησης Ομιλίας (Speech Assessment Methodology-SAM) τεχνολογικού έργου. Η SAM αποπειράθηκε να καθιερώσει μεθόδους αξιολόγησης απόδοσης προϊόντων τα οποία χρησιμοποιούν δεδομένο εξοπλισμό. Η Bellcore ακολούθησε μια άλλη προσέγγιση. Έχτισε μια οικογένεια ανεξάρτητων του ομιλητή αναφορικών (reference) βάσεων δεδομένων και μια μεθοδολογία δεδομένων δοκιμασιών για εκτίμηση απόδοσης τεχνολογίας α.ο. πάνω από τηλεφωνικά δίκτυα. Οι πωλητές κανόνισαν να έχουν τους Bellcore ελέγχους απόδοσης στα προϊόντα τους, παρέχοντας έτσι λεπτομερή αποτελέσματα. Οι αναφορικές βάσεις δεδομένων παραμένουν στην υπηρεσία τεχνικού ελέγχου της Bellcore.

### **5.2.1 Αναφορικές βάσεις δεδομένων – Συλλογές**

Μια αναφορική βάση δεδομένων ή συλλογή, είναι ένα βασικό εργαλείο για εκτίμηση προϊόντων περιορισμένων σε εφαρμογές. Περιέχει δείγματα ομιλίας τα οποία απεικονίζουν τον ομιλητή, το περιβάλλον, το λεξιλόγιο, την πολυπλοκότητα και άλλες συνθήκες της εφαρμογής. Η ποιότητα της αναφορικής βάσεως δεδομένων (β.δ.) είναι ένας καθοριστικός παράγοντας στον προσδιορισμό της εγκυρότητας των αποτελεσμάτων των εφαρμογών. Συνήθως, τα δεδομένα καταγράφονται χρησιμοποιώντας υψηλής ποιότητας εξοπλισμό για να ελαχιστοποιηθεί ο θόρυβος που εισάγεται από τον εξοπλισμό καταγραφής και τα μικρόφωνα, σε σύγκριση με αυτά που θα χρησιμοποιηθούν για τη εφαρμογή. Οι ομιλητές είναι από τον πληθυσμό των αναμενόμενων χρηστών εφαρμογής. Όλα τα ανθρώπινα λάθη, παραβιάσεις από το script και όροι όχι από το λεξιλόγιο του συστήματος σχολιάζονται ή απομακρύνονται από τα δεδομένα, για να αποφευχθούν λάθη πλαστής αναγνώρισης τα οποία θα παρουσιάσουν λάθος απόδοση προϊόντος.

Υπάρχουν σημερινές συλλογές καθομιλούμενης γλώσσας οι οποίες μπορούν να χρησιμοποιηθούν ως αναφορικές β.δ. Εάν αυτές απεικονίζουν ακριβώς τις συνθήκες της εφαρμογής, η ανάμειξή τους σ' ένα δοκιμαστικό πρόγραμμα είναι τεράστιας αξίας. Ένα πρόβλημα με πολλές, αλλά όχι με όλες απ' αυτές τις συλλογές είναι ότι τα δείγματά τους δεν περιέχουν αυθόρμητη ομιλία. Σε πολλές περιπτώσεις, τα δεδομένα έχουν διαβαστεί από προετοιμασμένο κείμενο ή εκμαιεύτηκαν ζητώντας επαναλήψεις συγκεκριμένων λέξεων από το σύστημα συλλογής. Η ανάπτυξη μιας μέσα στο σπίτι (in-house) συλλογής είναι πιο χρονοβόρα, αλλά αν είναι καλά κατασκευασμένη, μπορεί να προσφέρει πιο ακριβή εκτίμηση της ικανότητας ενός προϊόντος να χειρίζεται τις συνθήκες μιας δοσμένης εφαρμογής, σ' ένα συγκεκριμένο περιβάλλον.

### Συλλογές καθομιλούμενης γλώσσας.

Ο αριθμός των συλλογών καθομιλούμενης γλώσσας αυξάνεται ραγδαία. Ένας λόγος είναι ότι οι APRA και NIST τις χρησιμοποιούν για να πραγματοποιούν ετήσια τέστ δοκιμής σε μεγάλο λεξιλόγιο συστήματα αναγνώρισης και συστήματα κατανόησης της καθομιλούμενης γλώσσας. Για να καθιερώσει ένα επίπεδο λειτουργίας και ομοιομορφία λεξιλογίου τα οποία χρειάζονται για συγκρινόμενες δοκιμές, η APRA συχνά βρίσκεται, για την κατασκευή, σε επαφή με συγκεκριμένες αναφορικές β.δ. Αυτές οι συλλογές τότε χρησιμοποιούνται από τους εργολάβους της APRA για να χτίσουν, παρόμοια με τις δοκιμές συστήματα αναγνώρισης ομιλίας. Αυτή η χρήση των συλλογών έχει καλλιεργήσει τεχνολογικές προόδους στην έρευνα και τα εμπορικά συστήματα.

Μια από τις πρώτες συλλογές αυτού του είδους ήταν η Texas Instruments' TI Digits. Ολοκληρώθηκε το 1984 και χρηματοδοτήθηκε ολοκληρωτικά από την Texas Instruments. Η συλλογή περιείχε δείγματα συνδεδεμένων αριθμών λεγομένων από 326 γυναίκες, άντρες, κορίτσια κι αγόρια. Η TI Digits συλλογή έχει γίνει ένα σίγουρο εργαλείο για ανάπτυξη και δοκιμή συστημάτων συνδεδεμένων ψηφίων.

Η TIMIT συλλογή αναπτύχθηκε κάτω από ένα συμβόλαιο της DARPA για να καλύψει όλες τις μορφές ήχου και φωνητικές συχνότητες στα Αμερικανικά Αγγλικά. Η TIMIT ήταν μια κοινή προσπάθεια συνεργασίας από τις Texas Instruments (TI), το ινστιτούτο τεχνολογίας της Μασαχουσέτης (MIT) και από το διεθνές ινστιτούτο έρευνας του Stanford (SRI). Περιέχει δείγματα ομιλίας από 630 ανθρώπους αντιπροσωπεύοντας όλες τις περιοχές διαλέκτων των Ηνωμένων Πολιτειών. Η TIMIT έχει χρησιμοποιηθεί στην έρευνα ακουστικών ομιλίας και για την αρχική ανάπτυξη στατιστικών τεχνικών μοντελοποίησης. Έχει επίσης χρησιμοποιηθεί για την εκτίμηση συστημάτων αναγνώρισης βασισμένα στους φθόγγους. Η DAPRA Resource Management συλλογή ολοκληρώθηκε το 1988. Περιέχει 2.800 προτάσεις ομιλούμενες από 100 γνήσιους ομιλητές της Αμερικανικής Αγγλικής γλώσσας. Η Air Travel Information System (ATIS) λειτουργία έχει οδηγήσει στη δημιουργία ενός αριθμού συλλογών υπό τη χρηματοδότηση της DAPRA. Η ATIS σχεδιάστηκε για χρήση στην ανάπτυξη ομιλούμενης τεχνολογίας κατανόησης. Περιέχει περίπου 14.000 φράσεις για δύο ομάδες διαλόγων.

Μερικές απ' αυτές τις συλλογές έχουν γίνει διαθέσιμες από το Linguistic Data Consortium, έναν αφιλοκερδή οργανισμό που έγινε μέλος του Πανεπιστημίου της Pennsylvania.

Μερικές συλλογές της καθομιλούμενης γλώσσας που σχεδιάστηκαν για να κατευθύνουν συγκεκριμένους τύπους θεμάτων εφαρμογών, μπορεί να γίνουν αποδεκτές για δοκιμή προϊόντων. Το Logica Cambridge Ltd. κατασκεύασε μια συλλογή και μια μεθοδολογία εκτίμησης της απόδοσης προϊόντων σε σχέση με διαφορετικές παραμέτρους συγκεκριμένου ομιλητή. Η συλλογή περιέχει δείγματα παρμένα με διάφορους ομιλούμενους ρυθμούς, από ανθρώπους με διάφορα μεγέθη φωνητικού συστήματος, χρησιμοποιώντας ποικίλους βαθμούς φωνητικής προσπάθειας. Άλλη μια συλλογή, η NOISEX – 92, περιέχει δείγματα ομιλούμενων ψηφίων ενσωματωμένα σε διάφορους τύπους θορύβου. Η NOISEX – 92 αναπτύχθηκε ως κομμάτι του ESPRIT έργου 2589 (το λεγόμενο SAM), για να διεγείρει τους κατασκευαστές και των δύο, τεχνολογίας ομιλίας και εφαρμογών, να προσδιορίσουν μορφές θορύβου και συνθήκες οι οποίες υποβιβάζουν την απόδοση αναγνώρισης.

### 5.2.2 Πρωτόκολλα δοκιμών



Η μεθοδολογία δοκιμών χρειάζεται να εκφράζει τις προτεραιότητες μιας εφαρμογής και τις συνθήκες της για χρήση. Οι καθοδηγήσεις του πωλητή για την εγκατάσταση του εξοπλισμού, την εγγραφή, την επιλογή μικροφώνου και άλλων γνωρισμάτων του λειτουργικού εξοπλισμού πρέπει να ακολουθηθούν, ή αλλιώς ο εξοπλισμός δεν θα είναι ικανός να αποδώσει τα μέγιστα. Τα πρωτόκολλα δοκιμής σχεδιάστηκαν για να επιβεβαιώσουν ότι ο εξοπλισμός λειτουργεί σωστά κατά τη διάρκεια των δοκιμών.

Τα πρωτόκολλα δοκιμών περιλαμβάνουν προδιαγραφές για τη σχεδίαση και χρήση μιας αναφορικής βάσης ή βάσεων δεδομένων. Αυτό ασφαλίσει ότι μια συλλογή που σχεδιάστηκε εσωτερικά θα προσφέρει απαραίτητα στοιχεία δοκιμών. Εάν μια υπάρχουσα συλλογή πρόκειται να χρησιμοποιηθεί (οι προδιαγραφές πρέπει να καθοδηγήσουν την επιλογή της συλλογής), πρέπει να εξακριβωθεί εάν τα προϊόντα που θα δοκιμαστούν είναι παρόμοια με αυτά για τα οποία η συλλογή αρχικά σχεδιάστηκε, και να επιβεβαιωθεί ότι οι εφαρμογές είναι ίδιες μ' αυτές που χρησιμοποιήθηκαν στα αρχικά τέστ.

### **5.2.3 Τύποι λαθών**

Υπάρχουν τρεις τύποι λαθών οι οποίοι γενικά υπολογίζονται στο αποτέλεσμα της απόδοσης ενός προϊόντος αναγνώρισης:

- Διαγραφή Κόψιμο λέξεων. Για παράδειγμα, το σύστημα ακούει '124' όταν ο ομιλητής λέει '1234'.
- Αλλαγή Αντικατάσταση μιας ομιλούμενης λέξης με μια άλλη, συνήθως παρόμοια. Για παράδειγμα, το σύστημα ακούει '9' όταν ο ομιλητής λέει '5'.
- Εισαγωγή Πρόσθεση μιας λέξης. Για παράδειγμα, το σύστημα ακούει '12384' όταν ο ομιλητής λέει '1234'.

Σε μερικές περιπτώσεις, απλές διαγραφές αναφέρονται ως καθόλου ακοής (no-hear) απαντήσεις. Τέτοια λάθη μπορεί να συμβούν όταν το σύστημα ψάχνει για μια απλή (ή μονολεκτική) απάντηση, όπως 'ναι' ή 'όχι'. Ακόμη αυτό συμβαίνει όταν η απάντηση του χρήστη δεν ακούγεται από το σύστημα. Άλλοι τύποι λαθών που μπορεί να εισαχθούν στο αποτέλεσμα είναι:

- Λάθος αποδοχή Η αναγνώριση μιας λέξης που δεν βρίσκεται στο λεξιλόγιο ως μια λέξη η οποία βρίσκεται μέσα σ' αυτό (καλείται έξω απ' το λεξιλόγιο).
- Διαχωρισμός Μια πολυσυλλαβική λέξη αναγνωρίζεται ως δύο ή περισσότερες μικρές λέξεις. Για παράδειγμα, η λέξη 'pandemonium' αναγνωρίζεται ως 'point and mention'.
- Συγχώνευση Δύο λέξεις αναγνωρίζονται σαν μια. Για παράδειγμα, η 'lamp and' αναγνωρίζεται ως 'lampron'.

Τα δεδομένα λάθος αποδοχής είναι σημαντικά μεγάλα για εφαρμογές των οποίων οι χρήστες είναι πιθανό να χρησιμοποιούν ορολογία εκτός λεξιλογίου. Αυτά πολύ συχνά γίνονται από ανθρώπους που έχουν λίγη ή καθόλου προηγούμενη εμπειρία με τη εφαρμογή. Τα λάθη διαχωρισμού και συγχώνευσης αφορούν τα συστήματα αναγνώρισης συνεχόμενης

ομιλίας, και είναι πιο κατάλληλα για συστήματα μεγάλου λεξιλογίου χρησιμοποιώντας στατιστικά γλωσσικά μοντέλα.

Τα συστήματα εντοπισμού λέξεων διατυπώνουν εκ νέου τα λάθος αποδοχής και διαγραφής λάθη στα:

- Λάθος σύνθημα  
Το σύστημα λανθασμένα αναγνωρίζει ένα κομμάτι της εισόδου σαν ένα από τις λέξεις κλειδιά. Για παράδειγμα, το σύστημα ακούει 'Lending' όταν ο ομιλητής λέει 'Please, let me have some information'.
- Λάθος απόρριψη  
Αποτυχία εντοπισμού μιας έγκυρης λέξης κλειδί στην είσοδο.

#### **5.2.4 Βαθμολογία**

Δύο απλές προσεγγίσεις της βαθμολογίας των συστημάτων διακριτής λέξης είναι η *ορθότητα επί τοις εκατό* και η *ακρίβεια λέξης*. Η ορθότητα επί τοις εκατό μετράει τον ολικό αριθμό λέξεων στην είσοδο (ολικές λέξεις) και τον αριθμό των λέξεων που αναγνωρίστηκαν σωστά (σωστός αριθμός). Στη συνέχεια διαίρει το σωστό αριθμό με τον ολικό αριθμό των λέξεων. Η ακρίβεια λέξης είναι καλύτερη στο ότι εξετάζει την απόδοση χρησιμοποιώντας τρεις τύπους λαθών: αλλαγές, εισαγωγές και διαγραφές. Ο αριθμός λαθών κάθε τύπου συνοψίζεται. Το άθροισμα αυτών των λαθών αφαιρείται από τις ολικές λέξεις για να βρεθεί ο σωστός αριθμός. Τότε η ορθότητα επί τοις εκατό μπορεί να υπολογιστεί. Η ορθότητα και η ακρίβεια λέξης μπορούν εύκολα να υπολογιστούν, αλλά μπορεί να παρουσιάσουν λανθασμένα τον πραγματικό αριθμό λαθών. Από τη φύση της, η ορθότητα αγνοεί τα λάθη εισαγωγής και γι' αυτό είναι πιθανό να παράγει ένα χαμηλό ρυθμό λάθους. Τα συστήματα τα οποία παράγουν πολλά λάθη εισαγωγής θα εμφανιστούν λανθασμένα, να αποδίδουν καλύτερα απ' αυτά που παράγουν άλλα είδη λαθών. Επειδή πράγματι μετράει τα λάθη εισαγωγής, η προσέγγιση ακρίβειας λέξης μπορεί να φτάσει σ' έναν φαινομενικό αρνητικού αποτελέσματος σωστό αριθμό, ανταποκρινόμενο στους ρυθμούς λαθών λέξης μεγαλύτερες από 100%.

Η χρήση μιας δυναμικού προγραμματισμού (Dynamic Programming – DP) ευθυγράμμισης συμβολοσειράς μεθόδου είναι πιο πολύπλοκη, αλλά μπορεί να χρησιμοποιηθεί με διακριτής λέξης, συνδεδεμένης και συνεχόμενης ομιλίας είσοδο. Χρησιμοποιεί τεχνικές δυναμικού προγραμματισμού για να ευθυγραμμίσει και στη συνέχεια να συγκρίνει τη συμβολοσειρά που παρουσιάζουν οι αναγνωρισμένες λέξεις, με τη συμβολοσειρά που περιέχεται στη πραγματική είσοδο. Ποινές καθορίζονται για αλλαγής, εισαγωγής και διαγραφής λάθη. Τα αποτελέσματα καταγράφονται σε όρους επί τοις εκατό λαθών λέξης ή πρότασης. Η DP ευθυγράμμιση συμβολοσειράς αναπτύχθηκε σ' έναν αριθμό τοποθεσιών έρευνας. Αρχικά υλοποιήθηκε σε μεγάλης κλίμακας δοκιμές από τη NIST για της DAPRA τα δοκιμαστικά τέστ το 1987. Η DP ευθυγράμμιση λέξεων, τώρα μια από τις ευρέως χρησιμοποιούμενες στις μεθοδολογίες βαθμολογίας, έχει αυτοματοποιηθεί και συγχωνεύει μια ποικιλία στατιστικών τεχνικών μεγάλης σπουδαιότητας.

#### **5.2.5 Πέρα από τη βαθμολογία**

Υπάρχουν πολλά καλά προϊόντα αναγνώρισης ομιλίας ικανά για καλά αποτελέσματα, πάνω στις μετρήσεις που περιγράφηκαν στη προηγούμενη ενότητα. Είναι μια επικίνδυνη παραπλάνηση να λαμβάνεται ένας απλός αριθμός, ο οποίος εμφανίζεται να ενσωματώνει την απόδοση ενός προϊόντος. Ένας τέτοιος αριθμός μερικές φορές καλείται μια φιγούρα αξίας.

*“Η ευκολία του να είναι ικανό να αναφερθεί ένας απλός αριθμός είναι τόσο σπουδαία, όπου ένας αναπόφευχτα θα παραχθεί ανεξάρτητα από τις αδυναμίες. Δεχόμενοι αυτό, πρέπει να διαλέξουμε μια καλή φιγούρα κέρδους, όπου η λέξη καλή αναφέρεται σ’ ένα μέτρο που είναι εύκολο να αποκτηθεί και το οποίο απεικονίζει επακριβώς την βασική απόδοση την οποία εμφανίζεται να μετράει, και το οποίο παράγει μια ποικιλία που είναι λογικά κοντά σε αυτή που οι περισσότεροι χρήστες με συγκεκριμένες εφαρμογές θα θελήσουν”* (Melvyn Hunt, “Figures of merit for assessing connected-word recognisers,” 1990,p.329-330).

Αυτοί είναι πολύ δύσκολοι σκοποί να επιτευχθούν. Σημειώνουν την ανάγκη για χρήση των προτεραιοτήτων της εφαρμογής ώστε να καθοδηγηθεί η εξήγηση απόδοσης αποτελεσμάτων προϊόντων.

Λεπτομερής ανάλυση των μορφών λάθους μπορεί να παρέχει πληροφορίες μιας πιο σπουδαίας τιμής από μια απλή φιγούρα κέρδους. Μια ασυνήθιστη μέτρηση υψηλού λάθους μπορεί να υποδεικνύει ότι ένα προϊόν υπολειτουργεί, κατά μια έννοια. Ένας μεγάλος αριθμός διαγραφής και αλλαγής λαθών για έναν αναγνωριστή ψηφιακής λέξης, μπορεί να αποτελέσει σημάδι ότι ο βηματισμός της ομιλούμενης εισόδου είναι πολύ γρήγορος, ή ότι η σύνδεση σε περιφερειακές συσκευές είναι φτωχή. Μερικά λάθη υποδεικνύουν την ασυμβατότητα μεταξύ του αναγνωριστή και των προτεραιοτήτων της εφαρμογής. Ένας αναγνωριστής ο οποίος χρειάζεται μικρότερο βήμα εισόδου, είναι ακατάλληλος για λειτουργίες που εξαρτώνται από τη ταχύτητα ή για τις οποίες οι χρήστες είναι πιθανόν να μιλούν γρήγορα. Άλλα λάθη απεικονίζουν προβλήματα στο σχεδιασμό της εφαρμογής ή στην υλοποίηση. Ο συνυπολογισμός σειράς μπερδεμένων λέξεων σε μια εφαρμογή για παράδειγμα, μπορεί να παρουσιάσει λάθη αλλαγής σε πολλά προϊόντα.

### **5.3 Δοκιμασμός εφαρμογής**

Η δοκιμή εφαρμογής αφορά την αξιολόγηση της σχεδίασης της εφαρμογής, παρά την ικανότητα ενός προϊόντος αναγνώρισης να λειτουργεί μ’ αυτή τη σχεδίαση. Η εκτίμηση αρχίζει μόλις η εφαρμογή καταστεί αποδεκτή και συνεχίζει και μετά το ολοκληρωμένο σύστημα βγει στο εμπόριο. Μια καλή αρχική προσέγγιση είναι επαναληπτικός σχεδιασμός και δοκιμές χρησιμοποιώντας:

- Δοκιμή του οδηγού του Oz (Wizard of Oz testing)
- Ανάπτυξη πρωτοτύπων
- Πεδία δοκιμών δοκιμαστικών συστημάτων
- Πεδία δοκιμών του ολοκληρωμένου συστήματος

Η δοκιμή του οδηγού του Oz είναι χρήσιμη στην εκτίμηση εναλλακτικών σχεδιαστικών απόψεων. Βοηθάει να καθοριστεί εάν μια εφαρμογή θα γίνει αποδεκτή από τους χρήστες. Μπορεί ακόμη να βοηθήσει στον καθορισμό λεξιλογίου και των προτύπων ομιλίας που χρειάζονται για τη εφαρμογή. Η τελευταία χρήση του οδηγού με πρωτότυπα μπορεί να προσδιορίσει μέρη της εφαρμογής στα οποία οι χρήστες θα μπερδευτούν ή θα κάνουν λάθη.

Η επίδειξη και τα πρωτότυπα συστήματα προσφέρουν μια γρήγορη και σχετικά ανέξοδη μέθοδο εκτίμησης πολλών γνωρισμάτων υλοποίησης, συμπεριλαμβανομένων και των:

- Συγκεκριμένων σχεδιαστικών επιλογών οι οποίες δοκιμάζονται καλά με τη δοκιμή του οδηγού του Oz
- Διαφορετικών διακλαδισμένων παραγόντων ή επίπεδα πολυπλοκότητας
- Σχεδιαστικές ανταλλαγές (tradeoffs), όπως ταχύτητα με ακρίβεια
- Λάθη και τύποι λαθών συνδυασμένα με συγκεκριμένους τομείς της εφαρμογής
- Τεχνικές υποχώρησης όπως προεντολές ή διαδρομή σ' έναν άνθρωπο
- Η ολική σχεδίαση της εφαρμογής

Αυτές οι μέθοδοι εκτίμησης μπορεί να χρησιμοποιηθούν μαζί με τη δοκιμή του οδηγού του Oz ή ανεξάρτητα. Τα πρωτότυπα επίσης, χρησιμεύουν στη δοκιμή διαχειριζόμενων γνωρισμάτων ενός συστήματος αναγνώρισης, όπως τα κατώφλια απόρριψης. Τα περισσότερα συστήματα αναγνώρισης έχουν προσαρμοζόμενα κατώφλια απόρριψης.

### **5.3.1 Τι πρέπει να δοκιμαστεί:**

Μια από τις πρώτες δοκιμές που εφαρμόζεται σε μian εφαρμογή είναι εάν θα γίνει αποδεκτή στο πληθυσμό των αναμενόμενων χρηστών. Η εφαρμογή θα πρέπει να φαίνεται λογική στους χρήστες ειδάλλως δεν θα τη χρησιμοποιήσουν. Εάν η εφαρμογή πρόκειται να γίνει ένα προϊόν ή μια ευρέως χρησιμοποιούμενη υπηρεσία, μια χρήσιμη μέθοδος για τη συλλογή πληροφοριών για την αποδοχή του χρήστη, είναι μέσω μιας έρευνας αγοράς.

*‘Μπορείς να χρησιμοποιήσεις μια εστιασμένης ομάδας προσέγγιση για να ανακαλύψεις εάν οι άνθρωποι την αποδέχονται ή θα σκεφτούν κατά μια φτωχή έννοια για την εταιρία, ειδικά αν το σύστημα δεν λειτουργεί καλά’* (Robert Perdue, Supervisor, AT&T Bell Laboratories, personal communication, 1994).

Μιας και έχει καθοριστεί ότι μια εφαρμογή αξίζει τον κόπο, η σχεδίαση και ο δοκιμαστικός κύκλος αρχίζει και καλύπτει τις παρακάτω όψεις της εφαρμογής.

#### **1. Λεξιλόγιο**

Μερικές από τις όψεις εκτίμησης του λεξιλογίου είναι:

- Ολοκληρωτικό μέγεθος λεξιλογίου
- Ανάγκη για πρόσθεση λέξεων
- Συγγεόμενες λέξεις
- Πολυπλοκότητα
- Σωστές επιλογές λέξεων

Οι χρήστες πρέπει να καταλαβαίνουν τις επιλογές του λεξιλογίου.

*‘Σε μian εφαρμογή πιστωτικής κάρτας υπάρχουν οι επιλογές ‘χαμένη κάρτα’, ‘κλεμμένη κάρτα’ ή ‘συνεργάζομαι’. Οι άνθρωποι δεν ξέρουν τι σημαίνει το ‘συνεργάζομαι’*’ (Robert Perdue, Supervisor, AT&T Bell Laboratories, personal communication, 1994).

Αυτό περιλαμβάνει τη γλώσσα σε ακουστικές και οπτικές εντολές. Η λειτουργία των εντολών είναι να κατευθύνει το χρήστη στις κατάλληλες απαντήσεις.

## **2. Δομή**

Ένα αντικείμενο της εκτίμησης εφαρμογής είναι εάν οι χρήστες καταλαβαίνουν και μπορούν να ακολουθήσουν τη δομή της εφαρμογής. Σαν κομμάτι της δουλειάς της σε αρκετές εφαρμογές βασισμένες στο τηλέφωνο, η Ameritech υπολόγισε πόσο καλά οι χρήστες καταλάβαιναν τι απαιτούνταν απ' αυτούς.

*“Κρατήσαμε εγγραφές από το τι λέχθηκε στην αρχή της συνομιλίας κι από τις δύο πλευρές, έτσι ώστε ήμασταν ικανοί να παγιδεύσουμε την αρχή μιας κλήσης και να προσέξουμε τους ρυθμούς ακρίβειας. Ήταν ένα σωρό άνθρωποι που δεν ήξεραν τι σήμαινε το κουδούνισμα (bong). Συνέχισαν να πηγαίνουν στη μηχανή και να δίνουν τα νούμερα.”* (Eileen Schwab, Network services engineer, Ameritech, personal communication, 1994).

Ένας δεύτερος στόχος είναι να επιβεβαιωθεί αν το σύστημα οδηγεί τους χρήστες στην αποπεράτωση της λειτουργίας γρήγορα και αποτελεσματικά. Η εκτίμηση μπορεί να συνοψίσει σημεία όπου η σύγχυση του χρήστη και τα ανθρώπινα λάθη συμβαίνουν, και μπορεί να υπολογίσει το χρόνο που απαιτείται για μετακίνηση μεταξύ συγκεκριμένων σημείων της λειτουργίας. Αυτές οι σκέψεις πρέπει να προστεθούν στους υπολογισμούς αυτών των χρηστών που ολοκλήρωσαν το στόχο τους.

Άλλο ένα αντικείμενο δοκιμής εφαρμογής είναι ο καθορισμός εάν η δομή της εφαρμογής ταιριάζει με τις προτεραιότητες της λειτουργίας. Η προτεραιότητα πολλών εργοστασιακών εφαρμογών είναι η ταχύτητα. Ενέργειες στη γραμμή παραγωγής βηματίζονται και η σύνδεση είναι σημαντική. Αντίθετα, μια προτεραιότητα των τηλεφωνικών εφαρμογών είναι η ελαστικότητα. Αυτές οι εφαρμογές πρέπει να είναι ικανές να χειρίζονται την απρόβλεπτη ομιλία και συμπεριφορά των ανθρώπων που δεν είναι εξοικειωμένοι με την εφαρμογή.

## **3. Μοντέλα ομιλητή**

Η εκτίμηση της μοντελοποίησης του ομιλητή περιλαμβάνει ακρίβεια των μοντέλων και εγγραφή εφαρμογών (συστήματα εξαρτώμενα του ομιλητή). Η δοκιμή με πρωτότυπα εφαρμογών κατά τη διάρκεια της ανάπτυξης, μπορεί να καθορίσει την καταλληλότητα των μοντέλων ομιλητή στη εφαρμογή. Η συνέχιση αυτής της εκτίμησης μετά την αρχική εμπορευματοποίηση μιας εφαρμογής είναι χρήσιμη και για τα δύο συστήματα εξαρτημένα ή όχι του ομιλητή. Οι χρήστες των εξαρτημένων του ομιλητή συστημάτων τείνουν να μιλούν διαφορετικά όταν συνηθίσουν ένα σύστημα. Το αποτέλεσμα είναι ότι τα αναφορικά μοντέλα που παράγονται από την αρχική εγγραφή, θα παρουσιάσουν λιγότερο ακριβή. Συνεχόμενη επιμέλεια θα καθορίσει επίσης, την αποτελεσματικότητα της επανεκπαίδευσης ή τις προσαρμοστικές εφαρμογές του ομιλητή.

Όταν ένα σύστημα έχει εμπορευματοποιηθεί, μπορεί να καθοριστεί εάν οι άνθρωποι που έρχονται σε επαφή με ένα σύστημα ανεξάρτητο του ομιλητή, δεν είναι περιορισμένοι σ' αυτούς των οποίων η ομιλία είχε αρχικά ταιριάζει με τα μοντέλα του συστήματος. Μπορεί να υπάρχει μεγαλύτερη ποικιλία στη διάλεκτο, στην ταχύτητα της ομιλίας, στη φωνητική πίεση και άλλα χαρακτηριστικά του ομιλητή που είχαν προβλεφθεί. Αυτό μπορεί να απαιτεί τη σχεδίαση νέων μοντέλων ομιλητή. Η χρήση των εγγραφών από το χώρο βοηθά στον

καθορισμό των πηγών των προβλημάτων αναγνώρισης και θα καθορίσει τα είδη των προσαρμογών που χρειάζονται.

Εάν μια άλλη λειτουργία πραγματοποιείται ενώ ένα πρόσωπο μιλάει σ' ένα σύστημα αναγνώρισης, θα πρέπει να καθοριστεί εάν και σε ποιο βαθμό απόδοσης της λειτουργίας επηρεάζει η ομιλία. Λειτουργίες που απαιτούν φυσική άσκηση (physical exertion) θα αλλάξουν τα πρότυπα ομιλίας. Εάν μια προστατευτική μάσκα ή άλλος εξοπλισμός φορεθεί στο πρόσωπο, τα προϊόντα θα πρέπει να δοκιμαστούν στην ομιλία που παράγεται όταν οι ομιλητές φορούν αυτό τον εξοπλισμό. Όταν το μικρόφωνο του αναγνωριστή τοποθετηθεί μέσα στη μάσκα μπορεί να εξασθενίσει το θόρυβο του περιβάλλοντος, αποδίδοντας τη δοκιμή ή την εγγραφή βασισμένη στις συνθήκες εσωτερικού θορύβου λιγότερο χρήσιμη. Εάν η πραγματοποίηση της λειτουργίας παράγει θόρυβο, αυτό χρειάζεται επίσης να συμπεριληφθεί στην εκτίμηση.

#### **4. Ροή ομιλίας**

Η εκτίμηση της ροής ομιλίας περιλαμβάνει αξιολόγηση της ακρίβειας, της ευκολίας χρήσης και της ταχύτητας. Μερικές φορές τα παραπάνω μπορούν να ανταλλάγουν. Μεγαλύτερη ταχύτητα για παράδειγμα, μπορεί να επιφέρει μείωση της ολικής ακρίβειας. Η διακριτής λέξης είσοδο μπορεί να μην είναι η κατάλληλη για μια εφαρμογή η οποία απαιτεί μεγάλη ταχύτητα, αλλά τα συνεχόμενης ομιλίας συστήματα ακόμη απαιτούν καθαρή έκφραση. Εάν η ακρίβεια είναι πιο σημαντική από τη ταχύτητα, η διακριτής λέξης είσοδο μπορεί να διασχιστεί χρησιμοποιώντας ακουστικούς τόνους ή οπτικές εντολές.

#### **5. Ομιλούμενο και λειτουργικό περιβάλλον**

Μια εφαρμογή πρέπει να δοκιμαστεί για να καθοριστεί πόσο καλά έχει μοντελοποιήσει τα χαρακτηριστικά του περιβάλλοντος ομιλίας. Από τα χαρακτηριστικά του περιβάλλοντος, αυτά τα οποία επηρεάζουν το σήμα εισόδου είναι:

- Ο λόγος σήματος προς θόρυβο
- Το φάσμα ισχύος θορύβου
- Η μεταβλητότητα του θορύβου

Εάν ένα σύστημα πρόκειται να χρησιμοποιηθεί σε παραπάνω από ένα περιβάλλον ομιλίας, τα χαρακτηριστικά θορύβου όλων των περιβαλλόντων μπορεί να μην παρουσιάζονται σωστά.

Τα περισσότερα προϊόντα αναγνώρισης έχουν προσαρμοζόμενα κατώφλια απόρριψης. Τουλάχιστον ένα ταίριασμα για μια υποψήφια λέξη πρέπει να υπερβεί το κατώφλι απόρριψης για να αναφέρει το σύστημα μια αναγνώριση. Τα συχνά διαγραφής ή αντικατάστασης λάθη πραγματοποιούνται όταν η φωνή του ομιλητή είναι πολύ ήπια ή όταν η αναλογία σήματος προς θόρυβο είναι μεγάλος. Προσαρμογή του κατωφλίου απόρριψης θα μειώσει αυτά τα λάθη.

Ένα από τα κέρδη της χρήσης πρωτοτύπων στις δοκιμές, είναι ότι μπορούν να καθορίσουν εάν οι φυσικές προδιαγραφές του συστήματος αναγνώρισης είναι κατάλληλες για την πραγματοποίηση της λειτουργίας, ή αν πρέπει να γίνουν προσαρμογές στον εξοπλισμό.

#### **6. Ανθρώπινοι παράγοντες**

*“Δεν μπορείς να αξιώσεις ότι πραγματοποιώντας μια εφαρμογή οι άνθρωποι θα συνεργαστούν”* (Robert Perdue, Supervisor, AT&T Bell Laboratories, personal communication, 1994).

Από τη στιγμή που η αναγνώριση ομιλίας είναι μια τεχνολογία διασύνδεσης, η δοκιμασία ανθρώπινων παραγόντων είναι μια σταθερά που υποστηρίζει την εκτίμηση όλων των άλλων όψεων της απόδοσής της σε μια εφαρμογή. Καθοδηγεί το σχεδιασμό και την αξιολόγηση του λεξιλογίου, των εντολών, των προεντολών και άλλων υποσυστημάτων υποστήριξης. Εξετάζει τους τρόπους που οι άνθρωποι θα συνομιλήσουν με το σύστημα, πόσο καλά καταλαβαίνουν τι κάνει το σύστημα, και πόσο καλά το ίδιο τους οδηγεί στους στόχους τους. Ξεχωρίζει τις όψεις της εφαρμογής οι οποίες θεωρούνται ανεκτίμητες με αυτές που ενοχλούν.

### **5.3.2 Βαθμολογία στη δοκιμασία εφαρμογής**

Οι τεχνικές βαθμολόγησης που συχνά εφαρμόζονται σε εφαρμογές είναι παρόμοιες με αυτές που χρησιμοποιούνται στην εκτίμηση προϊόντων. Μια φιγούρα κέρδους συχνά αναφέρεται ως η βάση της ακρίβειας.

Αυτό αποτελεί μια παραπλάνηση επειδή όλοι οι υπολογισμοί μονού αριθμού, είτε είναι η ορθότητα επί τοις εκατό είτε είναι η ακρίβεια λέξης, και οι διαδικασίες βαθμολόγησης βασισμένες στη χρήση της DP ευθυγράμμισης συμβολοσειράς, αποτυγχάνουν να μετρήσουν για μη κατάλληλες απαντήσεις του χρήστη και για no-hear λάθη (ή λάθη λάθος απόρριψης στις εφαρμογές εντοπισμού λέξεων). Ο λόγος αποτυχίας της σύλληψης no-hear απαντήσεων είναι ότι όταν συμβαίνουν αυτά τα λάθη συνήθως διεγείρουν μια προεντολή ακολουθούμενη από μια νέα απάντηση. Η αποτυχία της γνώσης του συμβάντος αυτών των λαθών συνεισφέρει στην απόρριψη μιας εφαρμογής από χρήστες που ενοχλούνται όταν αναγκάζονται να επαναλαμβάνουν σωστές απαντήσεις.

Μια επαύξηση (enhancement) της βαθμολογίας είναι να μετράει τις φορές που ένα πρόσωπο πρέπει να επαναλάβει μια σωστή απάντηση. Στις εφαρμογές βασισμένες στο τηλέφωνο, αυτό είναι δυνατό να συμβεί τόσο συχνά όσο 20% της κάθε φοράς. Ένας ρυθμός απόρριψης λάθους αυτής της έντασης μπορεί να είναι ένας δείκτης ότι η φωνή εισόδου είναι μικρής έντασης ή ότι το κατώφλι απόρριψης είναι πολύ υψηλό. Άλλη μια βελτίωση της βαθμολογίας είναι να θέσει υψηλότερες ποινές (penalties) σε λάθη από επιλεγμένα σημαντικά κομμάτια λεξιλογίου. Αυτό θα επιτρέψει στα προβλήματα με αυτά τα θέματα-λέξεις να σημειωθούν και να διορθωθούν γρήγορα.

### **5.3.3 Πέρα από τη βαθμολογία της δοκιμής εφαρμογής**

Προχωρημένοι ερευνητές και σχεδιαστές συστημάτων αναγνωρίζουν το γεγονός ότι για κάθε εφαρμογή αναγνώρισης ομιλίας οι δεδομένες μέθοδοι συνόψισης (tabulation), δεν αποδίδουν με ακρίβεια την απόδοση λειτουργίας ενός συστήματος. Καμία από τις απλές προσεγγίσεις για παράδειγμα, δεν υπολογίζει την έμφυτη πολυπλοκότητα της εφαρμογής και την επίδρασή της στο χρήστη.

*“Εάν έχεις μια πολύπλοκη εφαρμογή με πολλούς βρόχους μπορεί να θέλεις να υπολογίσεις το ποσοστό των ανθρώπων που επιτυγχάνουν το στόχο τους. Μπορεί να θέλεις να κοιτάξεις σε κάθε σημείο να δεις πόσος χρόνος τους πήρε να φτάσουν εκεί, και να το συγκρίνεις*

με ένα μη-αυτοματοποιημένο σύστημα'' (Judith Tschirgi, Direction of Services and Speech Technology, AT&T Network Systems, personal communication, 1994).

Ανάλυση των ταινιών εγγραφής των πεδίων δοκιμών και των πρωτοτύπων μπορεί να διαχωρίσει τα πραγματικά λάθη αναγνώρισης από τα λάθη που παράγονται από την επίδραση της εφαρμογής στο χρήστη. Πολλές τηλεφωνικές εφαρμογές συμπεριλαμβάνουν time-out παράθυρα για να διεγείρουν μια προεντολή ή το τερματισμό μιας εφαρμογής, όταν δεν λαμβάνεται καμιά απάντηση από τον καλούντα. Ένα λάθος αντικατάστασης μπορεί να συμβεί όταν ο ομιλητής αρχίσει να μιλάει πριν το σύστημα αρχίσει να ακούει. Άλλα λάθη παράγονται από θόρυβο.

Αντικατάστασης και εισαγωγής λάθη (και λάθος ειδοποίηση στον εντοπισμό λέξεων) μπορεί να προκαλέσουν στην εφαρμογή να λειτουργήσει όχι σωστά, και το σύστημα θα γνωρίζει ότι έγινε ένα λάθος. Αυτά τα λάθη μπορούν να χρησιμοποιηθούν για να καθορίσουν σημεία στην εφαρμογή όπου μια επιβεβαίωση απάντησης μπορεί να εισαχθεί, ή όπου περαιτέρω ανάλυση του λεξιλογίου μπορεί να εκμηδενίσει μπερδεμένες λέξεις.

Μερικά προϊόντα παράγουν μια βαθμολογημένη λίστα υποψήφιων λέξεων για κάθε γεγονός αναγνώρισης. Όταν ο υψηλότερος υποψήφιος είναι η σωστή λέξη, βοηθάει να συγκριθεί το ταίριασμα με αυτό της επόμενης λέξης στη λίστα (ο κοντινότερος γείτονας). Η εγγύτητα αυτών των αποτελεσμάτων μπορεί να υποδείξει ενδεχόμενα προβλήματα από λέξεις που μπερδεύουν. Περισσότερη εκπαίδευση (για εξαρτώμενα του ομιλητή συστήματα) μπορεί να αποδείξει τα αποτελέσματα ή μπορεί να χρειάζεται να αλλάξει τις λέξεις. Εάν ένα λάθος αντικατάστασης συμβεί, η λίστα των εναλλακτικών λύσεων μπορεί να χρησιμοποιηθεί για να καθορίσει εάν η σωστή λέξη είναι η υψηλότερη στη λίστα. Εάν υπάρχει ένα πρότυπο λαθών αλλαγής, ένας πίνακας σύγχυσης (confusion matrix) μπορεί να παραχθεί για να απεικονίζει τα πρότυπα των ακουστικών συγχύσεων. Ένας τέτοιος πίνακας μπορεί να χρησιμοποιηθεί στη διόρθωση λαθών για σειρές λέξεων με μεγάλη πιθανότητα σύγχυσης, όπως είναι το Αγγλικό αλφάβητο, ή για να τροποποιήσει τη μορφή του λεξιλογίου.

Ένα από τα πιο σημαντικά γνωρίσματα της εκτίμησης είναι η ευχαρίστηση του χρήστη. Κάθε πεδίο δοκιμής θα πρέπει να περιέχει έναν τρόπο επαφής με ανθρώπους που δοκιμάζουν το σύστημα για σχόλια και ενέργειες. Η βαθμολογία υπερέρχει των συνοψίσεων οποιουδήποτε άλλου τύπου.



## Πίνακας Αγγλικών όρων

Assessment	εκτίμηση
Automatic Speech Recognition	Αυτόματη αναγνώριση ομιλίας
Auditory	ακουστικός
Background noise	περιβαλλοντικός θόρυβος
Background speech	περιβάλλουσα ομιλία
Backup	υποστήριξη
Barge-in	απρόσκλητη διακοπή
Bed-bound	κλινήρης
Bedside charting	Διαγράμματα υγείας δίπλα σε κρεβάτια
Bilingual	διγλωσσικός
Bolt	μπαλόνι
Boolean Query formulation	Δυαδική διατύπωση ερωτήσεων
Broker	χρηματιστής
Burst	ριπή
Calling card fraud	εξαπάτηση καλούμενης κάρτας
Claim	διεκδίκηση
Cockpit	καμπίνα πλοήγησης
Command and Control	Εκτέλεση εντολών και ελέγχου
Compensation	αποζημίωση
Computer-Aided Design	Σχεδιασμός υποβοηθούμενος από υπολογιστή
Corpus	συλλογή
Custom control	Προσαρμοστικός έλεγχος
Data entry	εισαγωγή δεδομένων
Deployment	Ανάπτυξη, εμπορευματοποίηση
Developer	εφαρμογέας
Dial-by-name	υλοποίηση κλήσεων μέσω ονόματος
Dictation	Υπαγόρευση
Document	καταγράφω
Drain	διοχετεύω, απομακρύνω
Enrollment	εγγραφή
Executive Information System	Εκτελεστικά Συστήματα Πληροφοριών
Extraneous speech	επιπλέον ομιλία
Fallback	Υποχωρώ, φυλάω
Finite-state grammars	γραμματικές πεπερασμένων καταστάσεων
Full text	πλήρες κείμενο
Gain	απολαβή
Gisting	επικέντρωση στα κύρια σημεία μιας συζήτησης
Handset	χειροσυσκευή
Headset	Ακουστικά με ενσωματωμένο μικρόφωνο
Help modules	παράμετροι βοήθειας
Human factors	Ανθρώπινοι παράγοντες
Information Access	πρόσβαση πληροφοριών
Information retrieval	ανάκτηση πληροφοριών
Interaction	Συνομιλία, συνδιάλεξη
Interface	Διασύνδεση, διεπαφή
Iterative Design	Επαναλαμβανόμενος σχεδιασμός
Magnitude	Ένταση, ποσότητα

Maintenance system	σύστημα συντήρησης
Manual	εγχειρίδιο χρήσης
Modifier	τροποποιητής
Multi-feature	πολυσύνθετος
Multi-tasking	πολυλειτουργικός
Navigation	πλοήγηση
on-the-fly adaptation	σταδιακή προσαρμογή
Operator	Λειτουργός
Paper documents	χάρτινες αποδείξεις
Perplexity	περιπλοκή
Pitch	τόνος
Power user	δυναμικός χρήστης
Recall	ανακαλώ
Recognizer	αναγνωριστής
Reference databases	Αναφορικές βάσεις δεδομένων
Reserving	φύλαξη αποθεματικού
Rotary	περιστροφικός
Scheduler	Καταγραφέας προγράμματος
Score	Βαθμολογία, αποτέλεσμα
Session	ηχογράφηση
Software	λογισμικό
Speaker adaptive	προσαρμοζόμενο στον ομιλητή
Speaker dependent	εξαρτημένος από τον ομιλητή
Speaker independent	ανεξάρτητος από τον ομιλητή
Speech pattern	πρότυπο ομιλίας
Speech-to-speech translation	μετάφραση ομιλίας προς ομιλία
Spelling	ορθογραφία
Standard	προδιαγραφή
Steam trap	παγίδα ατμού
Switch	διακόπτης
Synthesis	Σύνθεση
Tabulate	Συνοψίζω, ταξινομώ σε πίνακες
Task	έργο
Telephone banking	τηλεφωνικές τραπεζικές συναλλαγές
Tilt	ανασηκώνω
Touch screen	οθόνη επαφής
Touch-tone services	ψηφιακές υπηρεσίες
Tradeoff	Ανταλλαγή
Transducer	μορφοτροπέας
Translation	μεταγλώττιση
Verbal	προφορικός
Voice driven	οδηγούμενος από φωνή

### Πίνακας συντμήσεων

ASR	Automatic Speech Recognition
CAD	Computer-Aided Design
CTI	Computer Telephony Integration
DSP	Digital Signal Processing
EIS	Executive Information Systems
GUI	Graphical User Interface
IVR	Interactive Voice-Response
PDA	Personal Digital Assistants
RSI	Repetitive Strain Injuries
SQL	Standard Query Language
VARRS	Voice Activated Robotic Retrofit System
VRCP	Voice Recognition Call Processing

## Βιβλιογραφία – Αναφορές

1. “*USING SPEECH RECOGNITION*” Chapters 8, 4 Judith A. Markowitz,  
*Prentice Hall, Upper Saddle  
River, New Jersey*
2. “*PRODUCT DESIGN WITH SPEECH RECOGNITION*” *Article from Internet*
3. “*TASK INTEGRATION IN MULTIMODAL SPEECH RECOGNITION  
ENVIRONMENTS*”  
Michael A. Grasso,  
Tim Finin,  
*Article from Internet*

# ΠΑΡΑΡΤΗΜΑ

## Commercial Speech Recognition

*Μια προσπάθεια αξιολόγησης του ομώνυμου web site*

## Εισαγωγικό Σημείωμα

Η ακόλουθη εργασία αποτελεί μια πρώτη προσπάθεια περιγραφής και αξιολόγησης του web site που βρίσκεται στην ηλεκτρονική διεύθυνση: "<http://www.tiac.net/users/rwilcox/speech.html>". Η όλη προσπάθεια έγινε στα πλαίσια του μαθήματος "Επικοινωνία με ομιλία", υπό τη μορφή συμπληρωματικής εργασίας.

Βασικός στόχος της έρευνάς μας αυτής είναι η ανίχνευση, παρουσίαση, καταγραφή και αξιολόγηση των νεότερων εξελίξεων στο χώρο της αναγνώρισης ομιλίας, όπως αυτά παρουσιάζονται μέσα από την εν λόγω σελίδα. Για το λόγο αυτό προτιμήσαμε να διατηρήσουμε τη μορφή της σελίδας όπως έχει στο web προκειμένου να γίνει φανερή η προσήλωσή μας στα περιεχόμενά της. Η εμμονή μας αυτή, όμως σε καμία περίπτωση δε σημαίνει ότι το όλο εγχείρημα έγινε αβίαστα, αφού τα σχόλια που συνοδεύουν την εργασία προέκυψαν έπειτα από εντατική, κοπιαστική και ιδιαίτερος χρονοβόρα μελέτη των links που υπήρχαν στη σελίδα. Αναλυτικά η μεθοδολογία που ακολουθήσαμε έχει ως εξής:

- Μελέτη της σελίδας-οδηγού
- Επίσκεψη στα διαθέσιμα links κάθε κατηγορίας
- Έρευνα των στοιχείων που συνοδεύουν κάθε link
- Αξιολόγηση των περιεχομένων τους
- Καταγραφή των συμπερασμάτων μας πάνω στο format της σελίδας-οδηγού

Σε αρκετές περιπτώσεις πολλά από τα προτεινόμενα links ήταν ανενεργά, γεγονός που δεν επέτρεψε την είσοδό μας στα συγκεκριμένα sites. Στις περιπτώσεις αυτές προτιμήσαμε να αντιγράψουμε απλώς τα σχόλια του δημιουργού της σελίδας οδηγού, προκειμένου να αποφευχθεί το ενδεχόμενο της εσφαλμένης κρίσης ενός site που δεν καταφέραμε (ίσως και λόγω ενός απλού δικτυακού προβλήματος) να επισκεφθούμε. Σε όσες περιπτώσεις η ακριβής μετάφραση κάποιων επιστημονικών και τεχνικών όρων από τα Αγγλικά στα Ελληνικά δε μας ήταν γνωστή προτιμήσαμε να διατηρήσουμε τον πρωτότυπο ξενόγλωσσο όρο, προκειμένου να περιορίσουμε το ενδεχόμενο της σύγχυσης στον αναγνώστη.

Αν και η όλη προσπάθεια έγινε λίγο βιαστικά λόγω της πίεσης χρόνου που υπήρχε, εν τούτοις θέλουμε να πιστεύουμε ότι το τελικό αποτέλεσμα είναι λίαν αξιοπρόσεκτο και αξιόλογο, καθώς και αρκετά κατανοητό κι ευπαρουσίαστο. Ελπίζουμε μόνο η γνώμη αυτή να μην αντικατοπτρίζει μονάχα την προσωπική μας πεποίθηση, αλλά να συμπίπτει απόλυτα με την άποψη που θα κρυσταλλώσει, αρχικά ο επιβλεπων καθηγητής και ακολούθως κάθε μελλοντικός αναγνώστης της εργασίας μας.

# Commercial Speech Recognition

---

## Contents

---

- [What's New](#)
- [FAQ Sites](#)
- [USENET Groups](#)
- [Recognition Engines](#)
- [Resellers](#)
- [Computer-Telephony Integration](#)
- [Major Research Labs](#)
- [Text-to-Speech and Compression](#)
- [Applications of Speech Recognition](#)
- [Mailing Lists](#)
- [FTP Sites](#)
- [Trade Shows](#)
- [Books](#)

- [Periodicals](#)
- [Notes](#)

## What's New

- Το [VoiceAction](#) είναι ένα ActiveX σύστημα προσαρμοστικού ελέγχου (custom control) για εφαρμογές αναγνώρισης ομιλίας, το οποίο έχει ενσωματωμένο ένα προκατασκευασμένο kit κατάλληλο για τους προγραμματιστές που επιθυμούν να προσθέσουν στις εφαρμογές τους δυνατότητες αναγνώρισης ομιλίας. Διαθέτει σχεδιασμό τεχνικής νοημοσύνης προκειμένου να προσαρμόζεται στις λεξιλογικές και γλωσσολογικές απαιτήσεις των χρηστών, ενώ παράλληλα επιτρέπει τη μετατροπή της ομιλίας σε κείμενο.
- Η εταιρία [Pacific Search Consultants](#), μεταξύ των υπολοίπων υπηρεσιών της ειδικεύεται στο πεδίο της φωνητικής τεχνολογίας προσφέροντας προϊόντα αναγνώρισης/επεξεργασίας/σύνθεσης ομιλίας, αναλαμβάνοντας το γλωσσολογικό, σημασιολογικό, ακουστικό, φωνολογικό, γραμματικό, διαλογικό σχεδιασμό εφαρμογών κι αναπτύσσοντας φωνητικά interfaces για την επικοινωνία ανθρώπου-υπολογιστή.
- Σε περίπτωση που κάποιος ενδιαφέρεται για τηλεφωνικές συσκευές ομιλίας (telephony speech engines), οι ακόλουθες διευθύνσεις αποτελούν τις demo pages κάποιων κορυφαίων προμηθευτών του είδους: [SpeechWorks](#), [VCS](#), [Nuance](#), [Locus](#), and [Vocalis](#).
- Η εταιρία [Brooktrout](#) προσφέρει υψηλής απόδοσης και ποιότητας ηλεκτρονικά προϊόντα επικοινωνίας κατάλληλα για παροχές υπηρεσιών, πωλητές και VARs που ασχολούνται με την ηλεκτρονική ανταλλαγή πληροφοριών, ενώ ειδικεύεται και στην παραγωγή hardware για computer telephony. Στόχος της είναι η αποδοτική ενοποίηση των επικοινωνιών φωνής, δεδομένων και fax εντός των τηλεπικοινωνιακών και δικτυακών περιβάλλοντων και τελευταία έχει αναπτυχθεί ιδιαίτερα, ύστερα από την εξαγορά της Rhetorex.
- Η [Black Ice](#) είναι μια εταιρία η οποία διακινεί ένα SDK σύστημα σύνθεσης ομιλίας στο οποίο έχει ενσωματωθεί το σύστημα προσαρμοστικού ελέγχου Active X.
- Η [SyVox, Inc.](#) παρέχει προϊόντα αναγνώρισης ομιλίας για την διευκόλυνση των εργατών που μετακινούνται εντός περιορισμένου χώρου.
- Η [SoftVoice](#) προσφέρει μια μηχανή μετατροπής κειμένου σε ομιλία που λειτουργεί σε περιβάλλον Windows και η οποία είναι ήδη διαθέσιμη προς αγορά.
- Στη διεύθυνση <http://www.freespeech98.com/prod.cfm?sid=> μπορεί κανείς να βρει το FreeSpeech98, ένα πακέτο υπαγόρευσης συνεχόμενης ομιλίας της Philips Speech Processing που λειτουργεί σε περιβάλλον Windows για PC.
- Η [Norcom Electronics](#) έχει παράγει έναν αναλογικό recorder μίνι-κασσέτας, ο οποίος συνεργάζεται και λειτουργεί με το λογισμικό υπαγόρευσης Dragon.
- Η εταιρία [Entropic](#) αναπτύσσει πακέτα αναγνώρισης και σύνθεσης ομιλίας που εφαρμόζονται στην τηλεφωνία και τα οποία είναι ευρέως διαδεδομένα στα Πανεπιστημιακά εργαστήρια.
- Η [Registry Magic](#) είναι μια εταιρία ανάπτυξης λογισμικού και προϊόντων αναγνώρισης ομιλίας τα οποία επιτρέπουν στους χρήστες να συνδιαλέγονται φυσικά με τα υπολογιστικά και τα τηλεφωνικά συστήματα, περιορίζοντας, έτσι, τα λειτουργικά έξοδα των επιχειρήσεων και προσαξίζοντας την παραγωγικότητα. Μεταξύ των προϊόντων της συγκαταλέγονται:
  - Ο [Virtual Operator](#), ένα σύστημα αναγνώρισης ομιλίας για την εκτέλεση εντολών κι ελέγχου που προάγει την εξυπηρέτηση και τη δρομολόγηση τηλεφωνικών κλήσεων, επιτελώντας όλες τις λειτουργίες ενός πραγματικού τηλεφωνητή. Εξυπηρετεί μέχρι και 12 κλήσεις ταυτόχρονα, καταλαβαίνει τη φυσική ομιλία, ελαχιστοποιεί την ανάγκη για touch-tone μενού, ενώ είναι σχεδιασμένο για την αυτοματοποίηση και διεθνών τηλεφωνικών συστημάτων για χρήση σε χώρες που τα touch-tone μενού δεν είναι διαδεδομένα. Διατίθεται σε διάφορες γλώσσες.
  - Ο [Virtual Dialer](#), μια προσωπική τηλεφωνική ατζέντα για την πρόσβαση δεδομένων μέσω της αναγνώρισης ομιλίας που επιτρέπει στους χρήστες να υλοποιούν κλήσεις εκφωνώντας απλά, το όνομα και την τοποθεσία του επιθυμητού συνομιλητή. Το σύστημα είναι προσβάσιμο από το γραφείο, το αυτοκίνητο, το σπίτι, το ξενοδοχείο, το αεροδρόμιο ή οπουδήποτε αλλού.
  - Το [Magic Calendar](#) και το [Magic Calendar για MS Outlook](#), δύο προϊόντα εισαγωγής δεδομένων που λειτουργούν ως καταγραφείς του προσωπικού προγράμματος (schedulers), εκτελώντας φωνητικές εντολές για την καταχώρηση και την αποθήκευση ραντεβού, τηλεφωνικών αριθμών, συναντήσεων, διακοπών κτλ. Με τον τρόπο αυτό η διαχείριση του

καθημερινού προγράμματος απελευθερώνεται από τη χρήση των χεριών λαμβάνοντας, πλέον, χώρα μέσω της φυσικής ομιλίας προς το PC.

- Η [Brite Voice Systems](#) προσφέρει μια μεγάλη ποικιλία αλληλεπιδρώντων (interactive) πληροφοριακών συστημάτων, όπως IVR και CTI (Computer Telephony Integration), καθώς και προϊόντων κινητής τηλεφωνίας συμπεριλαμβανομένων των συστημάτων υλοποίησης κλήσεων μέσω φωνής για κινητές συσκευές. Η Brite ήταν η πρώτη εταιρία που υλοποίησε ένα εμπορικό προϊόν για την φωνητική υλοποίηση κλήσεων στην ασύρματη τηλεφωνία κι εξακολουθεί να παραμένει παγκοσμίως πρώτη στην εφαρμογή συστημάτων αυτής της μορφής. Ενδεικτικά των προϊόντων της αποτελούν:
  - Το **VoiceSelect**, ένα προϊόν πρόσβασης δεδομένων που παρέχει στους συνδρομητές τη δυνατότητα να πραγματοποιούν τα τηλεφωνήματά τους προφέροντας απλά τον αριθμό ή ένα όνομα στη συσκευή τους. Αυτή η απαλλαγμένη από τη χρήση των χεριών τηλεφωνία προσφέρει άμεσα στους χρήστες τα πιο συχνά χρησιμοποιούμενα τηλέφωνα τους, προάγοντας ταυτόχρονα ασφάλεια κατά την οδήγηση και ευκολία στην υλοποίηση κλήσεων κατά τις ώρες αιχμής. Η Brite έχει ενσωματώσει στο VoiceSelect εξαρτημένη κι ανεξάρτητη του ομιλητή αναγνώριση ομιλίας, ενώ παράλληλα έχει μεταφέρει τη διαδικασία της αναγνώρισης στην πλευρά του δικτυακού μεταγωγού (switch) αντί της τηλεφωνικής συσκευής. Με τον τρόπο αυτό οι συνδρομητές μπορούν να κάνουν χρήση των πλεονεκτημάτων του VoiceSelect χρησιμοποιώντας τις συσκευές που ήδη έχουν, αναβαθμίζοντάς τις.
  - Το **BriteConnect**, ένα βραβευμένο IVR προϊόν πρόσβασης δεδομένων, που παρέχει στους χρήστες άμεση κι αυτοματοποιημένη πρόσβαση στις βάσεις δεδομένων μιας εταιρίας 24 ώρες τη μέρα, εξοικονομώντας χρόνο και χρήμα. Μέσα από τις τεχνολογίες text-to-speech και fax-back που υποστηρίζει εξασφαλίζει ευελιξία και εναλλακτικούς τρόπους ενημέρωσης στους πελάτες.
- Η [Labtec](#) προσφέρει μια μεγάλη ποικιλία ηχείων, μικροφώνων κι ακουστικών για την εξυπηρέτηση των αναγκών της αναγνώρισης ομιλίας μέσω PC. Μερικά από τα προϊόντα της είναι:
  - Ακουστικά για PC: C-110, C-105, C-184
  - Μικρόφωνα για PC: AM-242, AM-240, AM-232, AM-222, AM-252 (New), LVA-7330 (New), LVA-7370 (New)
  - Ακουστικά μ' ενσωματωμένο μικρόφωνο (headsets) για PC: C-324, C-316, C-350, C-322, C-315, LVA-8450 (New), LVA-8550 (New)
- Το [Speech Technology Magazine](#) προσφέρει δωρεάν συνδρομή σε ορισμένες κατηγορίες αναγνώστρων και διατηρεί αρχείο με τα δημοσιευμένα άρθρα των περασμένων ετών. Τα άρθρα του καλύπτουν επαρκώς τις εξελίξεις στη βιομηχανία της αναγνώρισης ομιλίας τα τελευταία δύο χρόνια.
- Η [Telex](#) παράγει μικρόφωνα για υπολογιστές, headsets που εξαλείφουν το θόρυβο, ένα ειδικό μικρόφωνο για υπαγόρευση εξ' αποστάσεως, ηχεία, καθώς και άλλες συσκευές εισόδου φωνής για τις εφαρμογές της αναγνώρισης ομιλίας. Τα καινούργια προϊόντα της, το **Aria Desktop Dictation Microphone**, το **NexVoice.USB desktop microphone** και το **Verba.USB headworn microphone** αναμένεται να κυκλοφορήσουν μέσα στο πρώτο τέταρτο του 1999.
- Η [Human Language Systems](#) λειτουργεί από το 1993 και προτείνει τις κατάλληλες εφαρμογές της αναγνώρισης ομιλίας στις επιχειρήσεις. Ειδικεύεται στα αυτοματοποιημένα κέντρα κλήσεων, στα συστήματα αναγνώρισης συνεχούς ομιλίας στα ομιλώντα διαλογικά συστήματα ανάκτησης πληροφοριών, καθώς και στη σύνθεση ομιλίας ενώ παράλληλα οργανώνει συμβουλευτικά κι εκπαιδευτικά μαθήματα για τα προϊόντα της. Διαθέτει γραφεία σε Boulder, CO και Γαλλία.
- Η [SCCS, Inc.](#) προσφέρει προϊόντα και υπηρεσίες για την επιτυχή εφαρμογή της αναγνώρισης ομιλίας όπως μικρόφωνα, headsets και ασύρματα εξαρτήματα, μαζί με το απαραίτητο λογισμικό. Ενδεικτικά αναφέρουμε τον ψηφιακό καταγραφέα φωνής **Olympus D1000** που είναι εφοδιασμένος με το σύστημα ηχογράφησης ViaVoice και το ασύρματο μικρόφωνο **Andrea Digital Super Directional Array** που εγκαθίσταται πάνω στο PC, εστιάζει στη φωνή του χρήστη και ακυρώνει το θόρυβο από το περιβάλλον (αναμένεται να διατεθεί εντός του εξαμήνου). Τέλος, η SCCS προσφέρει ένα πλήθος δυναμικών λύσεων σχεδιασμένων για την ιατρική βιομηχανία. Για παράδειγμα, η **Radiology Database** αποθηκεύει και ανακτά με ηλεκτρονικό τρόπο τις διάφορες ιατρικές αναφορές, ενώ η **Medical Notes Database** αποθηκεύει και ανακτά τα ιατρικά διαγράμματα (charts).

## FAQ Sites

---



Στις ακόλουθες διευθύνσεις μπορεί κανείς να αναζητήσει πληροφορίες και υποστήριξη για διάφορα τεχνολογικά θέματα, συμπεριλαμβανομένης της αναγνώρισης ομιλίας:

- [Comp.speech.users FAQ](#): Δημιουργήθηκε τον Ιούνιο του 1998 και παρέχει πληροφορίες για ζητήματα που αφορούν τη χρήση της τεχνολογίας της αναγνώρισης ομιλίας, τόσο από τη σκοπιά του hardware, όσο και του software.
- [Comp.speech.research FAQ](#): Προσφέρει εκτεταμένες πληροφορίες για την τεχνολογία της ομιλίας καλύπτοντας θέματα όπως η σύνθεση ομιλίας, η κωδικοποίηση ομιλίας, η αναγνώριση ομιλίας, η επεξεργασία ηχητικών σημάτων κ.α. Διαθέτει, τέλος, κι ένα αρκετά ενημερωμένο ftp site.
- [Comp.speech.research FAQ - Vendor list](#): Παρέχει εκτεταμένες πληροφορίες για τις επιχειρήσεις και τους οργανισμούς που εμπορεύονται προϊόντα της αναγνώρισης ομιλίας.
- [Dragon Dictate and a2x FAQ](#): Παρέχει πληροφορίες για τη χρήση του συγκεκριμένου προγράμματος υπαγόρευσης της Dragon.
- [Neural Net FAQ](#): Παρέχει πληροφορίες σχετικά με την τεχνολογία και τις εφαρμογές των νευρωνικών δικτύων.
- [Typing Injuries FAQ](#): Πρόκειται για ένα εκπαιδευτικό site που παρέχει πληροφορίες σχετικές με τα RSI (Repetitive Strain Injuries), τους τρόπους θεραπείας και τα προϊόντα που ενδείκνυνται για την αντιμετώπισή τους.

## Usenet Groups

---

Στα ακόλουθα newsgroups μπορεί κανείς να έρθει σ' επαφή και ν' αλληλογραφήσει με άλλους χρήστες τεχνολογικών υπηρεσιών, προκειμένου να ενημερωθεί, ν' ανταλλάξει απόψεις και να ζητήσει συμβουλές για διάφορα επιστημονικά θέματα. Εκτός της αναγνώρισης ομιλίας καλύπτονται και αρκετά άλλα πεδία εφαρμογών:

- [comp.speech](#) - Είναι το αρχικό newsgroup, το οποίο διαχωρίζεται στα θέματα που περιγράφονται ακολούθως.
- [comp.speech.users](#) - Εφαρμογές και προϊόντα που χρησιμοποιούν ομιλία.
- [comp.speech.research](#) - Σύνθεση (synthesis) και αναγνώριση ομιλίας.
- [comp.ai.nat-lang](#) - Επεξεργασία της φυσικής (natural) γλώσσας.
- [comp.ai.neural-nets](#) - Νευρωνικά δίκτυα.
- [alt.dcom.telecom](#) - Τηλεπικοινωνίες (unmoderated).
- [comp.dcom.telecom](#) - Τηλεπικοινωνίες (moderated).
- [comp.dcom.telecom.tech](#) - Τεχνολογία των τηλεπικοινωνιών.
- [comp.dsp](#) - Ψηφιακή επεξεργασία σήματος.
- [sci.lang](#) - Γλώσσες (Language).

## Company Web Sites - Recognition Engines

- 
- Η εταιρία [Lernout & Hauspie](#) διαθέτει προϊόντα αναγνώρισης, σύνθεσης και συμπίεσης ομιλίας τα οποία εκτός της Αγγλικής, αναγνωρίζουν και πολλές άλλες γλώσσες.
  - Η [Dragon Systems](#) έχει πρόσφατα εγκαινιάσει το δικό της web site. Μεταξύ των προϊόντων της περιλαμβάνονται: το **Dragon Naturally Speaking Professional 3.0**, ένα ευρύτατα διαδεδομένο προϊόν σύνθεσης φυσικής ομιλίας και το **Dragon Dictation 3.0**, ίσως το πιο ονομαστό προϊόν υπαγόρευσης.
  - Η [Verbex Voice Systems](#) είναι μία από τις πρωτοπόρες εταιρίες στην αναγνώριση συνεχόμενης ομιλίας. Τα προϊόντα της έχουν κατά καιρούς χρησιμοποιηθεί από πλήθος οργανισμών μεταξύ των οποίων συμπεριλαμβάνονται οι: UPS, L'Eggs, Canada Post, και Medical Labs. Ένα από τα προϊόντα της διαθέτει λογισμικό OEM συμβατό με τη πλατφόρμα Listen for Windows.
  - Το [Microsoft Research Speech Technology Group](#) ασχολείται με τη σχεδίαση και την ανάπτυξη της μηχανής Whisper, ενώ έχει ήδη ολοκληρώσει και προσφέρει το SAPI SDK. Η [Command Corporation](#) αυτή την περίοδο ασχολείται με την κατασκευή του αναγνωριστή IN CUBE.
  - Η [Northern Telecom](#) ασχολείται με την ανάπτυξη μηχανών αναγνώρισης ομιλίας.
  - Το [STAR Lab](#) της SRI Corp απασχολεί αυτή την περίοδο 25 επιστήμονες για την ανάπτυξη και το σχεδιασμό του DECIPHER, ενός ευρυζωνικού αναγνωριστή συνεχόμενης ομιλίας, ενώ έχει ήδη προχωρήσει σε αρκετά ικανοποιητικό βαθμό την κατασκευή μιας καινούργιας μηχανής αναγνώρισης με το όνομα Corona (αν ενδιαφέρεσαι να πειραματιστείς με τον τρόπο εφαρμογής της Corona στην

εκτέλεση εντολών και ελέγχου του Unix, μπορείς να ακολουθήσεις τη διεύθυνση που προσφέρεται [εδώ](#)).

- Η εταιρία [SpeechWorks](#) (γνωστή μέχρι πρόσφατα ως ALTech) ασχολείται με την ανάπτυξη μηχανών αναγνώρισης ομιλίας για την υλοποίηση δακτυλογραφίσεων και την εξυπηρέτηση πελατών μέσω τηλεφώνου.
- Το [IBM VoiceType Dictation](#) αποτελεί ένα web site αφιερωμένο στις εφαρμογές της υπαγόρευσης για τα PC.
- Το [AT&T Advanced Speech Products Group](#) προσφέρει το WATSON, ένα προϊόν αναγνώρισης συμβατό με το σύστημα αναγνώρισης και σύνθεσης ομιλίας SAPI, καθώς και με τις τεχνολογίες εξακριβωσης στοιχείων ομιλητών.
- Η [ART](#) (Advanced Recognition Technologies, Inc.) ασχολείται με τη σχεδίαση και την ανάπτυξη εφαρμογών υπαγόρευσης και αναγνώρισης ομιλίας, οι οποίες ενδείκνυνται για χρήση στις διάφορες καταναλωτικές συσκευές, όπως τα κινητά τηλέφωνα, τα παιχνίδια και οι PDAs.
- Η [Voice Control Systems](#) έχει ήδη εξαγοράσει αρκετές άλλες εταιρίες αναγνώρισης ομιλίας, όπως η PureSpeech. Μεταξύ των υπηρεσιών που προσφέρει συμπεριλαμβάνεται η φωνητική υλοποίηση κλήσεων για κινητά τηλέφωνα, η φωνητική εξακριβωση στοιχείων, οι αυτόματοι τηλεφωνητές, καθώς και οι φωνητικές δακτυλογραφίσεις.
- Η εταιρεία [Nuance](#) προσφέρει έναν αναγνωριστή συνεχόμενης ομιλίας ιδανικό για εφαρμογές με μεγάλο λεξιλόγιο, όπως είναι η καταγραφή εμπορευμάτων.
- Αν και η [Fonix](#) δεν αποτελεί κατασκευάστρια εταιρία προϊόντων αναγνώρισης, τελευταία απέκτησε τα δικαιώματα του εξαιρετικού αναγνώστη AcuVoice της TTS, καθώς και τον PowerScribe, ένα ιατρικό σύστημα υπαγόρευσης της Articulate Systems.
- Η Καναδική εταιρεία [Locus Speech Coporation](#) έχει πρόσφατα ανακοινώσει την κατασκευή μιας ανεξάρτητης του ομιλητή τηλεφωνικής συσκευής ομιλίας με ευέλικτο λεξιλόγιο, η οποία αναγνωρίζει Αγγλικά και Γαλλικά.
- Η [A&G Graphics Interface](#) έχει αναπτύξει μερικά εργαλεία SDK κατάλληλα για τη δημιουργία εφαρμογών αναγνώρισης ομιλίας.
- Η [Vocalis](#) ασχολείται κυρίως με τις εφαρμογές της ομιλίας στην computer telephony στην Ευρώπη. Ξεκινώντας το 1993 με την εξαγορά του management της Logica's Speech και του Natural Language Group, αποτελεί πλέον μια δημόσια επιχείρηση με έδρα την Αγγλία.
- Η [Sensory Inc.](#) γνωστή μέχρι πρότινος ως Sensory Circuits, ασχολείται με τη διακίνηση ολοκληρωμένων κυκλωμάτων για συστήματα αναγνώρισης ομιλίας, ιδανικών για τηλεπικοινωνιακές συσκευές και καταναλωτικά ηλεκτρονικά. Τα πεδία δράσης της περιλαμβάνουν την αναγνώριση ομιλίας, τη σύνθεση ομιλίας, την ηχογράφιση και το playback φωνής, καθώς και τη σύνθεση μουσικής, ενώ διαθέτει μεγάλη εμπειρία στην εφαρμογή της αναγνώρισης ομιλίας για το σχεδιασμό προϊόντων.
- Η [Philips Speech Processing](#) κατασκευάζει προϊόντα αναγνώρισης ομιλίας τελευταίας τεχνολογίας, συμπεριλαμβανομένων των συστημάτων φυσικής συνομιλίας για τη τηλεφωνία, των καταναλωτικών προϊόντων και των συστημάτων IT (Internet Telephony).
- Το [Defense Group Inc. \(DGI\)](#) ασχολείται κυρίως με την εφαρμογή ειδικών μεθόδων περιορισμού του περιβαλλοντικού θορύβου στα συστήματα αναγνώρισης ομιλίας, και διατίθεται για συνεργασία με τους σχεδιαστές μηχανών αναγνώρισης ομιλίας.
- Η [Oki Semiconductor](#) παράγει κυρίως ολοκληρωμένα κυκλώματα κατάλληλα για ενσωματωμένες εφαρμογές αναγνώρισης.
- Η [Speech Solutions](#) ασχολείται με την κατασκευή προσαρμοστικών εφαρμογών αναγνώρισης ομιλίας και προϊόντων που ενσωματώνουν τους ελέγχους ActiveX.
- Η [BaBel Technologies SA](#) στο Βέλγιο διακινεί προϊόντα για την αναγνώριση/σύνθεση και κωδικοποίηση της ομιλίας, καθώς και την οπτική αναγνώριση χαρακτήρων. Μεταξύ αυτών περιλαμβάνονται:
  - Το **BaBel SDK version 2.0**, ένας υψηλής ποιότητας συνθέτης ομιλίας για Αμερικάνικα, Γαλλικά, Γερμανικά και Ολλανδικά (προσφέρεται η δυνατότητα για on-line demo).
  - Το **BaBelDial version 2.1**, ένα προϊόν πρόσβασης δεδομένων για τη φωνητική υλοποίηση κλήσεων που επιτρέπει την κλήση των επιθυμητών τηλεφωνικών αριθμών με απλή εκφώνηση κάποιων λέξεων-κλειδιά.
  - Το **BaBel IWR version 1.04**, το οποίο ενοποιεί την τεχνολογία των τεχνητών νευρωνικών δικτύων με αυτήν της βασισμένης στα υβριδικά Hidden Markov μοντέλα, ανεξάρτητης του ομιλητή, αναγνώρισης ομιλίας, προκειμένου να επιτρέψει μια εύρωστη αναγνώριση λέξεων-κλειδιά για μεταβλητού εύρους λεξιλόγια.
  - Κωδικοποιητές ομιλίας στα 16, 8, 5.3, 2.4 Kb/s.

- Η λίστα που παρέχεται στη διεύθυνση [Yahoo's Voice Recognition](#) ενδέχεται να περιλαμβάνει κάποιες νεοεισελθούσες στο WWW εταιρίες. Πρέπει να σημειωθεί ότι η παραπάνω λίστα περιλαμβάνει αρκετές εταιρείες διακίνησης προϊόντων αναγνώρισης, οι οποίες όμως δεν κατασκευάζουν δικά τους συστήματα.

## Company Web Sites - Resellers and Integrators

---

- Η [Lab Resources](#) διακινεί τα προϊόντα της Dragon στο Wisconsin και το Illinois.
- Η [Voice Recognition Systems, Inc.](#) διακινεί τα προϊόντα υπαγόρευσης της Dragon στο Maryland.
- Η [Alma Information Systems](#) που εδρεύει στο Houston του Τέξας αποτελεί ένα διακινητή των προϊόντων λογισμικού LawTalk και OfficeTalk της Kolvox.
- Η [Peter Cohen Associates](#) αποτελεί ένα μεγάλο διακινητή των προϊόντων υπαγόρευσης της Dragon.
- Η [Responsive Systems Ltd.](#) είναι ο Βρετανικός και ο Ευρωπαϊκός προμηθευτής του προϊόντος VOICE της Kurzweil καθώς και ένας από τους βασικούς διακινητές του προϊόντος υπαγόρευσης DragonDictate. Διαθέτει ένα αρκετά ενημερωμένο και κατατοπιστικό site.
- Στο Καναδά το προϊόν υπαγόρευσης DragonDictate διακινείται από την εταιρεία [Talk2Me Technology](#).
- Η εταιρεία [Austin Computer](#), εκτός από διακινητής του προϊόντος DragonDictate αποτελεί προμηθευτή προϊόντων για την αυτοματοποίηση των υπολογιστών που διευκολύνουν τα άτομα με ειδικές ανάγκες στο χειρισμό τους.
- Η [Freedom of Speech](#) αποτελεί έναν βασικό παροχέα της Assistive Computer Technology (ACT) σε ανθρώπους με ειδικές ανάγκες.
- Η [NuGen](#) στο Michigan και η [Speech Recognition Systems](#) αποτελούν δύο παραδοσιακούς διακινητές προϊόντων της Dragon.
- Η [AM Technology](#) αποτελεί έναν διακινητή των προϊόντων της Dragon στη περιοχή της Βοστώνης.
- Η [AllVoice Computing](#) έχει εξοπλίσει με προϊόντα αναγνώρισης της Dragon αλλά και άλλων εταιρειών περισσότερες από χίλιους οργανισμούς στην Αγγλία.
- Ο [1st Voice](#) είναι ένας πιστοποιημένος διακινητής των συστημάτων της Dragon στο Palo Alto, CA, ενώ παράλληλα αναλαμβάνει την εκπαίδευση ατόμων στο χειρισμό του DragonDictate για Windows και Macintosh.
- Η [Narratek](#) διακινεί προϊόντα φωνητικού ελέγχου και υπαγόρευσης τα οποία χρησιμοποιούνται στις ιατρικές ηχογραφήσεις.
- Η [Computerized Business Systems](#) εδρεύει στο Charlotte και διακινεί ολόκληρη τη σειρά των προϊόντων υπαγόρευσης της IBM VoiceType και MedSpeak Radiology για τα οποία προσφέρει αρκετά demos για download.
- Η [A.D.A. Solutions by WorkLink](#) διακινεί τα προϊόντα DragonDictate, Dragon NaturallySpeaking, αποτροπής της πάθησης RSI, Dragon PowerSecretary και άλλα.
- Η [ABN, Inc.](#) λειτουργεί σαν προμηθευτής συστημάτων και δικτυακών υπηρεσιών στη περιοχή του Northeast και ειδικεύεται στους αυτοματισμούς και τις επιχειρησιακές εφαρμογές. Διακινεί κυρίως τα προϊόντα της IBM, της Kurzeil, της Dragon και ακόμη και της Wildcard (μέχρι πρότεινος γνωστή ως Kolvox).
- Η [Ficomp Systems](#) λειτουργεί σαν διακινητής προϊόντων αναγνώρισης συνεχόμενης ομιλίας και έχει αναπτύξει ένα επιτυχημένο σύστημα καταγραφής των τιμών που λειτουργεί με φωνή, για ένα από τα μεγαλύτερα ανταλλακτήρια συναλλάγματος του Chicago.
- Ο [VoiceAutomated](#) είναι ένας προμηθευτής προϊόντων λογισμικού για εφαρμογές υπαγόρευσης, όπως είναι το Continuous Speech της Philips.
- Η [Voice Recognition Systems](#) (Lexington, KY) διακινεί και προσαρμόζει τα προϊόντα υπαγόρευσης της Dragon στις διάφορες εφαρμογές, ενώ παράλληλα προωθεί τη χρήση ενός πρόσθετου προϊόντος που επιτρέπει την αντικατάσταση των μικροφώνων από καταγραφείς κασσετών.
- Η [Andrea Electronics](#) προσφέρει μια ολοκληρωμένη σειρά μικροφώνων για την ακύρωση του θορύβου, τα οποία είναι ιδανικά για χρήση στις διάφορες εφαρμογές αναγνώρισης.
- Η [Idioma Ltd.](#) έχει αναπτύξει ένα δικό της σύστημα αναγνώρισης ομιλίας, κατάλληλο για την προσωπική εκπαίδευση των χρηστών σε θέματα προφοράς και γλωσσικών δεξιοτήτων. Προσφέρει εκπαιδευτές για φωνητική θεραπεία, βελτίωση της προφοράς και εκμάθησης ανάγνωσης για παιδιά.
- Η [Voice Activated Systems Technologies \(VAST\)](#) στην Καλιφόρνια, έχει σχεδιάσει μερικά ιατρικά εξαρτήματα για την Ψυχιατρική, την Ουρολογία, την Οφθαλμολογία και την Ενδοκρινολογία, τα οποία στηρίζονται στο Naturally Speaking της Dragon και προσφέρουν τη δυνατότητα υπαγόρευσης συνταγών, αναφορών κι εκθέσεων προόδου.

- Η [Word Systems Inc.](#) αποτελεί ένα διακινητή των προϊόντων επεξεργασίας ομιλίας της Philips, καθώς και των συστημάτων ιατρικών ηχογραφήσεων κι εφαρμογών υπαγόρευσης μέσω του τηλεφώνου.
- Η [Shure Brothers](#) προσφέρει μικρόφωνα υψηλής απόδοσης πιστοποιημένα από τη SoundBlaster.
- Η [New World Creations](#) διακινεί πάνω από 100 προϊόντα σχετικά με την αναγνώριση ομιλίας τη φωνητική εξακρίβωση στοιχείων, τη μετατροπή κειμένου σε ομιλία, καθώς και μικρόφωνα. Επίσης προσφέρει μια σύγκριση των διαφόρων συστημάτων υπαγόρευσης, αλλά και των υπολοίπων προϊόντων αναγνώρισης ομιλίας, οργανωμένη κατά τιμή κι εφαρμογή, καθώς και μια διαθέσιμη πηγή downloading για την αναβάθμιση κάποιων προϊόντων.
- Η [21st Century Eloquence](#) είναι ένας προμηθευτής προϊόντων ομιλίας που εδρεύει στη Φλόριντα. Στο ιδιαίτερος ελκυστικό web site της μπορεί κανείς να ενημερωθεί για τα τελευταία βιομηχανικά νέα και να βρει ένα πίνακα με τα προϊόντα που διακινεί, όπως μικρόφωνα, recorders και ειδικά πρόσθετα λεξιλόγια.
- Η [Synapse](#) είναι ένας προμηθευτής προϊόντων για άτομα με ειδικές ανάγκες, τα οποία στοχεύουν στην αύξηση της παραγωγικότητας και την υποσκέλιση των περιορισμών που αντιμετωπίζουν. Μεταξύ αυτών περιλαμβάνονται: λογισμικά αναγνώρισης ομιλίας για τον απαλλαγμένο από τη χρήση των χεριών έλεγχο των υπολογιστών, συστήματα φυσικής ομιλίας, συστήματα υπαγόρευσης, μεταφραστές, mouse κεφαλής, συστήματα εκμάθησης κ.α.

## Computer-Telephony Integration Vendors

- 
- Το [Dialogic](#) είναι ένας κορυφαίος τηλεφωνικός εξοπλισμός και φορέας της τεχνολογίας DSP.
  - Μεταξύ των [IBM's Computer-Aided Telephony Products](#) περιλαμβάνονται τα CallPath και DirectTalk
  - Η [Digital Sound Corporation](#) προσφέρει μια δικτυακή υπηρεσία φωνητικού ταχυδρομείου, τη δυνατότητα φωνητικής συμπλήρωσης αιτήσεων φορμών, καθώς και το εργαλείο γραφικής τηλεφωνίας Passport.
  - Η [TELE DATA Consultants](#), μια εταιρεία ανάπτυξης λογισμικού για τηλεφωνικά interfaces, προσφέρει μια εξαιρετική [σύννοψη](#) των εφαρμογών των IVR.
  - Η [ITI Software](#) διαθέτει βιβλιοθήκες γραμμένες σε C κώδικα προορισμένες για χρήση στην computer-telephony, με πάνω από 30,000 θύρες για χρήση.
  - Η [Katalina Technologies](#) εμπορεύεται το VoiceGuide, ένα εργαλείο γραφικής σχεδίασης για εφαρμογές IVR, fax-back και φωνητικού ταχυδρομείου.
  - Η [Digital Equipment Corporation](#) προσφέρει το DECVoice για τη διασύνδεση του VAX σ' ένα τηλεφωνικό σύστημα.
  - Η [Universal Access Inc.](#) προσφέρει τον WebFax Server προκειμένου να παρέχεται η δυνατότητα για fax-back των Web pages.
  - Η [Word Systems](#) προσφέρει λύσεις που εφαρμόζονται στα interfaces της αναγνώρισης ομιλίας και της τηλεφωνίας.
  - Η [T-NETIX](#) διακινεί προϊόντα επεξεργασίας κλήσεων, όπως το σύστημα φωνητικής εξακρίβωσης στοιχείων SpeakEZ.
  - Η [Artisoft](#) διαθέτει μια μεγάλη γκάμα προϊόντων για Computer Telephony, Fax και Networking, όπως γίνεται αντιληπτό από το Web site της εταιρίας. Κύρια προϊόντα της αποτελούν: το [Visual Voice Pro](#), μια συλλογή γραφικών εργαλείων και ActiveX συστημάτων προσαρμοστικού ελέγχου των 32 bit που επιτρέπει στους εφαρμογείς να δημιουργούν μια μεγάλη ποικιλία εφαρμογών για Computer Telephony και το [TeleVantage](#), ένα έξυπνο τηλεφωνικό σύστημα σχεδιασμένο αποκλειστικά για μικρού και μεσαίου μεγέθους επιχειρήσεις.

## Major Research Labs

---

Ακολουθως προσφέρονται οι διευθύνσεις μερικών κορυφαίων ερευνητικών εργαστηρίων στον τομέα της αναγνώρισης ομιλίας.

- [Center for Spoken Language Understanding \(CSLU\) Home Page](#)
- [Carnegie-Mellon Interactive Systems Labs](#)
- [MIT Spoken Language Systems Group](#)

- [IEEE Communications Society](#)
- [Cambridge University Speech Vision Robotics](#)
- [U. Illinois ASR Lab](#)
- Η λίστα [Index - Speech](#) περιέχει σχεδόν 50 ζεύξεις για ερευνητικά εργαστήρια σε όλο τον κόσμο, κάθε ένα από τα οποία διαθέτει ένα όμορφο και φανταχτερό γραφικό (συντηρείται από την UCSC).
- Τα [Haskins Laboratories](#) προσφέρουν σχεδόν 100 ζεύξεις για σελίδες του Web σχετικές με την ομιλία, μαζί με μια αρκετά ενημερωμένη λίστα των προγραμματισμένων συνεδρίων ανά τον κόσμο.
- Το [Univeristy of Edinburgh](#) προσφέρει μια σπουδαία λίστα των διεθνών ερευνητικών κέντρων εκτός της Αγγλίας.
- Το [U.S. Navy Intelligent Multimedia Center](#) εξερευνά τις διεπαφές ανθρώπου υπολογιστή.
- Το [Linguistic Data Consortium](#) περιγράφει τις διάφορες συνεδριάσεις και προόδους του.
- Το [The Natural Language Software Registry](#) κατέχει πληθώρα πληροφοριών για τα διάφορα projects φυσικής γλώσσας, όπως το COSIMA, έναν αναγνωριστή συνεχόμενης ομιλίας στα Γερμανικά.
- Το [The University of Munich's](#) Bavarian Archive ασχολείται με την παραγωγή προϊόντων αναγνωρισής στα Γερμανικά.
- Το [IFA](#) στην Ολλανδία προσφέρει μια λίστα των διεθνών ερευνητικών πηγών για θέματα φωνητικής και ομιλίας.
- [Comp.speech FAQ - Sources of raw voice data](#)
- Η λίστα [Corpora](#) περιέχει ένα σύνολο αρχειοθετημένων ανακοινώσεων.
- Η ακόλουθη διεύθυνση περιέχει ερευνητικά papers που αναφέρονται στο σχεδιασμό των interfaces συνομιλίας: [Matt Marx's Home Page](#).
- Το [ROC](#) είναι ένα ενδιαφέρον σχεδιαστικό εργαλείο γλωσσικών μοντέλων, δημιουργημένο από το φοιτητή Stephen Yang.
- Η [International Phonetic Association](#) παρέχει γενικές πληροφορίες σχετικά με τα φωνητικά σύμβολα, τις εγγραφές ήχων και τις φωνητικές γραμματισειρές.
- Ακολούθως παρέχεται ένα δείγμα των Γαλλικών site για την ομιλία, συμπεριλαμβανομένου του εργαστηρίου [Institute of Communications](#), ενός άρθρου στο [Recherche](#) webzine και του εργαστηρίου [CLIPS](#).
- Το [UCLA](#) διαθέτει ένα ερευνητικό εργαστήριο αναγνώρισης ομιλίας, όπου διερευνάται ο τρόπος λειτουργίας της ανθρώπινης φωνής και ακοής.

## Text to Speech and Compression

- 
- Η [Gus Communications](#) κατασκευάζει συνθέτες ομιλίας που λειτουργούν σε περιβάλλον Windows ειδικά για χρήστες με ειδικές ανάγκες.
  - Το [The Digital Cambridge Research Laboratory](#) έχει ήδη πραγματοποιήσει αρκετά σημαντική δουλειά στον τομέα της σύνθεσης ομιλίας.
  - Στη διεύθυνση [University of Edinburgh](#) υπάρχει διαθέσιμη μια λίστα με περίπου μια ντουζίνα TTS demos από το Web.
  - Το [Laureate](#) είναι ένα καινούργιο σύστημα TTS, κατασκευασμένο από τα BT Labs της Αγγλίας. Στη σελίδα τους στο web υπάρχει διαθέσιμο ένα interactive demo.
  - Η [Voxware](#) ειδικεύεται στη συμπίεση και τη μορφοποίηση (morphing) φωνής. Τελευταίο τους προϊόν είναι το ToolVox για το web, το οποίο επιτρέπει την εφαρμογή αρχείων συμπιεσμένης φωνής στα διάφορα web sites.
  - Μερικές ακόμη ονομαστές εταιρίες που ειδικεύονται στη σύνθεση ομιλίας είναι η [Willow Pond](#), που προσφέρει μερικά δείγματα των ηχογραφήσεων που πραγματοποιούν οι μηχανές της και η [First Byte](#), η οποία φημίζεται για τις πλατφόρμες σύνθεσης ομιλίας για PC που προωθεί.
  - Η [Elan](#) αποτελεί έναν Γαλλικό προμηθευτή συστημάτων TTS και υλοποιεί αξιόλογη έρευνα στους τομείς του hardware και του software.
  - Η [Eloquent Technology](#) προσφέρει ένα demo του λογισμικού μετατροπής κειμένων σε ομιλία που διακινεί, το οποίο τρέχει σε περιβάλλον Windows.
  - Η [Lucent Technologies Bell Labs](#) προσφέρει ένα καινούργιο on line demo του συστήματος μετατροπής κειμένων σε ομιλία που διακινεί, το οποίο λειτουργεί σε τρεις γλώσσες: Αγγλικά, Γερμανικά και Κινέζικα.
  - Η [AcuVoice](#) διακινεί μια μηχανή TTS, καθώς κι ένα σύστημα SDK που λειτουργεί επίσης σαν αναγνώστης ([reader](#)) κειμένων για Windows PC.

- Αν και η εταιρία [Lernout & Hauspie](#) έχει ήδη αναφερθεί σαν διακινητής συσκευών αναγνώρισης, εν τούτοις διακινεί και αρκετά προϊόντα συμπίεσης φωνής και μετατροπής κειμένων σε ομιλία, τα οποία είναι συμβατά με αρκετές γλώσσες, όπως το TruVoice της Centigram.

## Applications of Speech Recognition, IVR and Computer-Integrated Telephony

- Η [Wildfire Communications](#) στοχεύει στη δημιουργία ενός προϊόντος εκτέλεσης εντολών κι ελέγχου για τη γραμματειακή υποστήριξη μέσω τηλεφώνου, το οποίο θ' αναγνωρίζει και θα εκτελεί διάφορες φωνητικές εντολές.
- Ένα μέρος του site της [KAMT](#) είναι αφιερωμένο στη μετατροπή των ιατρικών καταγραφών από αναλογική σε ψηφιακή μορφή και συγκρίνει διάφορα προϊόντα υπαγόρευσης.
- Η [Metroplex Voice Technologies](#) εξήγγειλε πρόσφατα την κυκλοφορία του MathTalk, ενός προγράμματος μαθηματικών υπολογισμών που χρησιμοποιεί macros και τη συσκευή υπαγόρευσης της Dragon.
- Η εταιρία [Digital Dreams](#) παράγει συσκευές φωνητικής εισόδου κατάλληλες για εφαρμογές πολυμέσων, όπως είναι ο Macromedia Director.
- Η [Digital Acoustics](#) διακινεί μια ιδιαίτερος αξιόλογη εφαρμογή φωνητικής υλοποίησης κλήσεων, η οποία προσαρτάται σε μια τηλεφωνική γραμμή και ακολούθως μπορεί να χρησιμοποιηθεί από οποιαδήποτε τηλεφωνική συσκευή εντός του σπιτιού.
- Το web site της Sun Microsystems για τα διάφορα [SpeechActs](#) διερευνά τις δυνατές εφαρμογές της ομιλίας και έχει πρόσφατα εμπλουτιστεί με ορισμένες ηχητικές ηχογραφήσεις.
- Το [CDTNews](#) είναι ένα μηνιαίο newsletter αφιερωμένο στις παθήσεις RSIs (Repetitive Strain Injuries). Το τεύχος του Νοεμβρίου του '95 αναφέρεται στην αναγνώριση ομιλίας.
- Στη διεύθυνση αυτή περιγράφεται ο organizer και recorder χειρός [Voicelt](#), ο οποίος δεν έχει άμεση σχέση με την αναγνώριση ομιλίας αλλά διαθέτει μια αρκετά διασκεδαστική και προτότυπη σελίδα στο web.
- Η [CAPITAL](#) είναι ένας εκπαιδευτικός οργανισμός που μελετά τους δυνατούς συνδυασμούς φωνής και υπολογιστών για χρήση στην εκπαίδευση και ιδιαίτερα στη βελτίωση της προφοράς.
- Το [The MRC Group](#) έχει κατασκευάσει κι εμπορεύεται το PowerScribe για χρήση στη Ραδιολογία, με βάση το PowerScribe της [Articulate Systems](#). Το MRC Group απασχολεί σχεδόν 2200 δακτυλογράφους (transcriptionists) σε 60 γραφεία.
- Η [Applied Voice Recognition \(AVRI\)](#) διακινεί τον [SpeechCOMMANDER](#), ένα προϊόν συνεχόμενης υπαγόρευσης για χρήση σε περιβάλλον γραφείου και αύξηση της παραγωγικότητας, το οποίο αναβαθμίζει εντυπωσιακά τη συσκευή υπαγόρευσης της IBM μέσω του προγράμματος SOHO που εφαρμόζει.
- Η [Parity Software](#) αποτελεί μια από τις σημαντικότερες εταιρίες διακίνησης εργαλείων για computer-telephony. Στα προϊόντα της συμπεριλαμβάνονται: ένα περιβάλλον για scripting, εργαλεία με έλεγχο ActiveX, έξυπνες λύσεις για GUI, ενώ πρόσφατα έχει προωθήσει ένα προϊόν με ActiveX το οποίο συνεργάζεται με τις μηχανές αναγνώρισης ομιλίας τύπου SAPI.
- Η [Unisys](#) προσπαθεί να επιτύχει την κατανόηση της φυσικής γλώσσας για την εξαγωγή από ένα κείμενο του κεντρικού νοήματος προκειμένου να δημιουργήσει έναν υποστηρικτή συνδιαλέξεων μέσω ενός αναγνωριστή ομιλίας που μετατρέπει τη φωνή σε κείμενο.
- Η [Kewill-DECTEL](#) αποτελεί ένα Βρετανικό φορέα συστημάτων αναγνώρισης ομιλίας ενδεδειγμένων για εφαρμογή στους μεγάλους αποθηκευτικούς χώρους, όπου απαιτείται η εν κινήσει και απαλλαγμένη από τη χρήση των χειρών εισαγωγή δεδομένων, π.χ. για την εκτέλεση των παραγγελιών.
- Η [Speech Machines](#) αποτελεί έναν από τους πρωτοπόρους στην ενοποίηση της αναγνώριση ομιλίας με το Internet και στοχεύει στην εισαγωγή της αναγνώρισης ομιλίας στα επιχειρησιακά περιβάλλοντα. Πρωτεύουσα εφαρμογή της εταιρίας αποτελεί το [Cybertranscriber](#), μια υπηρεσία που επιτρέπει την υπαγόρευση φωνητικών μηνυμάτων μέσω κινητών ή σταθερών τηλεφώνων, υπολογιστών και recorders, την αυτόματη δακτυλογράφησή τους και ακολούθως την αποστολή των δακτυλογραφημένων κειμένων στους χρήστες μέσω email. Το Cybertranscriber συνδυάζεται εύκολα με τις υπάρχουσες third-party υπηρεσίες, όπως τα τηλεφωνικά κέντρα, τα συστήματα voice mail ή τα νοσοκομειακά συστήματα υπαγόρευσης.

## Mailing Lists

---

Ακολούθως παρέχονται οι διευθύνσεις ορισμένων επιλεγμένων mailing lists, στις οποίες μπορεί κανείς να εγγραφεί προκειμένου να συμμετάσχει σε αλληλογραφία και συζητήσεις για θέματα που σχετίζονται με την αναγνώριση ομιλίας.

- [Colibri](#): Πρόκειται για ένα ηλεκτρονικό newsletter, το οποίο απευθύνεται σ' εκείνους τους χρήστες που ενδιαφέρονται να ενημερωθούν σχετικά με τα πεδία της γλώσσας, της ομιλίας και της λογικής.
- Το [ECTL - Electronic Communal Temporal Lobe](#) είναι μια moderated mailing list που απευθύνεται στους ερευνητές που ασχολούνται με τις εφαρμογές των interfaces ομιλίας στους υπολογιστές. Προκειμένου να γίνει κανείς συνδρομητής αρκεί να στείλει το όνομα, το ινστιτούτο, το τμήμα, το τηλέφωνο εργασίας και την ηλεκτρονική του διεύθυνση.
- Για να εγγραφεί κανείς συνδρομητής στην [Prosody](#) mailing list αρκεί να στείλει στη συγκεκριμένη διεύθυνση ένα mail, το οποίο θα περιέχει το μήνυμα: "subscribe prosody Your\_Name".
- Η [foNETiks](#) είναι ένα μηνιαίο moderated newsletter που απευθύνεται στους επιστήμονες της ομιλίας και της φωνητικής. Για την εγγραφή απαιτείται η αποστολή mail με το μήνυμα: "join fonetiks your\_first\_name your\_second\_name".
- Η [PC-TELEPHONY](#) είναι μια mailing list που ασχολείται με την ενοποίηση υπολογιστών και τηλεφώνων. Για την εγγραφή απαιτείται η αποστολή mail με το μήνυμα: "SUBSCRIBE PC-TELEPHONY".
- Στόχος της [TOUCHTON](#) mailing list είναι η προαγωγή εξειδικευμένων συζητήσεων για θέματα που σχετίζονται με την αναγνώριση ομιλίας και τις 'touch-tone' υπηρεσίες. Για την εγγραφή απαιτείται η αποστολή mail με το μήνυμα: "SUBSCRIBE TOUCHTON Your\_Name". Σε περίπτωση που κάποιος επιθυμεί να του αποστέλλονται τα περιεχόμενα της λίστας υπό τη μορφή περιοδικής ανασκόπησης μπορεί να στείλει στην Touchton ένα mail με την εντολή SET TOUCHTON DIGEST, αφού προηγουμένως παραλάβει την επιβεβαίωση εγγραφής.
- Στην λίστα συζήτησης [CCMI-L](#) συμμετέχουν για την ώρα 411 Τηλεπικοινωνιακοί διευθυντές. Για την εγγραφή απαιτείται η αποστολή mail με το μήνυμα: "subscribe ccmi-l".
- Εδώ παρέχεται η διεύθυνση για να γίνει κανείς συνδρομητής στην [Telecom Digest](#).

## FTP Sites

---

Στις ακόλουθες διευθύνσεις μπορεί κανείς ν' αναζητήσει πληροφοριακό υλικό και προγράμματα που αφορούν την αναγνώριση ομιλίας και κάποιες άλλες περιοχές. Σε περίπτωση που απαιτείται βοήθεια για την αποσυμπίεση κάποιου αρχείου με συγκεκριμένο format, λύση μπορεί να δοθεί μέσω επίσκεψης στο comp.compression FAQ, που προσφέρεται [εδώ](#).

- [Comp.speech archives](#): περιέχει περισσότερα από 100 άρθρα ομαδοποιημένα σε ένα shar file, το οποίο συνοδεύεται από ένα αρχείο με τα περιεχόμενα.
- [Comp.dsp FAQ](#): Πρόκειται για ένα αρχείο με πληροφορίες για τα DSP's.
- [ECTL](#): περιέχει τα αρχεία της ομόθυμης mailing list που αναφέρθηκε προηγουμένως, μαζί με κάποιες περιγραφές προϊόντων ομιλίας και τεχνικές αναφορές.

## Trade Shows

---

Ακολούθως αναφέρονται τα sites κάποιων συνεδρίων που γίνονται (ή έχουν ήδη γίνει) στην Αμερική και τα οποία έχουν αντικείμενο την αναγνώριση ομιλίας και τις εφαρμογές της:

- [AVIOS'99](#) - Διοργανώνεται από την American Voice Input/Output Society το διάστημα από 24-26 Μαΐου 1999 στο San Jose των Η.Π.Α. Παράλληλα με την έκθεση και την παρουσίαση των νέων προϊόντων φωνητικής τεχνολογίας θα λάβει χώρα το ετήσιο Voice I/O Conference, στο οποίο θα ανακοινωθούν τα πιο πρόσφατα ερευνητικά αποτελέσματα για την ομιλία και το οποίο είναι ευρέως διαδεδομένο ως το κύριο μέσο επιμόρφωσης των εσχάτως μνημένων στη φωνητική τεχνολογία. - Για περισσότερες πληροφορίες: (408) 248-1353.
- [Computer Telephony Show](#) - Η Computer Telephony Expo Spring 99 διοργανώνεται στο Los Angeles Convention Center του Los Angeles από 1-5 Μαρτίου 1999. Στο χώρο της έκθεσης μπορεί κανείς να βρει οτιδήποτε χρειάζεται προκειμένου να κάνει τις πιο συνετές αγορές σε προϊόντα CT (Computer Telephony). Περισσότερες από 600 εταιρίες αναμένεται να εκθέσουν τα προϊόντα τους σχετικά με CT,

Call Centers, IP Telephony και Next Generation Networks. Παράλληλα με την έκθεση θα λάβει χώρα συνέδριο για την παρουσίαση εργασιών και δημοσιεύσεων που σχετίζονται με την αναγνώριση ομιλίας. Η Φθινοπωρινή έκθεση Fall Expo θα γίνει στο Orlando τον Οκτώβριο. -Για περισσότερες πληροφορίες: 212-691-8215.

- [SpeechTEK](#) - Η SpeechTEK 98 Conference & Exhibition διοργανώθηκε από την COMTEK International τον Οκτώβριο του 1998 στη Νέα Υόρκη. Μεταξύ των θεμάτων που κάλυψε συμπεριλαμβάνονταν οι καταναλωτικές εφαρμογές, η υγεία, τα τραπεζικά συστήματα, οι κατασκευές, οι εφαρμογές γραφείου, τα τηλεπικοινωνιακά δίκτυα, οι τηλεφωνικές εφαρμογές κ.α. -Για περισσότερες πληροφορίες: (800) 777-8774

## Books

---

Σ' αυτή την κατηγορία μπορεί να βρει κανείς τους τίτλους κάποιων επιλεγμένων βιβλίων που σχετίζονται με την αναγνώριση ομιλίας και τα οποία αφορούν τους χρήστες, τους σχεδιαστές προϊόντων, τους προγραμματιστές και τους επιστήμονες. Αναλυτικά έχουμε:

1. Τα βιβλία που αφορούν τους χρήστες των προϊόντων ομιλίας είναι:  
Users of Speech Recognition

- [Computer Speech Technology](#)
- [337 KILLER VOICE PROCESSING APPS](#)
- [The Essential Simply Speaking Gold](#)
- [IBM Speech Recognition: The Future Now! with Cdrom](#)
- [Using Speech Recognition](#)

2. Τα βιβλία που αφορούν τους σχεδιαστές προϊόντων ομιλίας είναι:  
Speech Interface Design

- [Designing Interactive Speech Systems: From First Ideas to User Testing](#)
- [Handbook of Standards and Resources for Spoken Language Systems](#)
- [Spoken Dialog with Computers](#)
- [Speech Technology for Telecommunications](#)
- [Voice Communications with Computers](#)

3. Τα βιβλία που αφορούν τους προγραμματιστές είναι:  
Implementing Speech Recognition in Your Application

- [Speech Recognition : Theory and C++ Implementation](#) Σημείωση: Η B&N δέχεται παραγγελίες για το βιβλίο αυτό στην τιμή των \$57 αντί της τιμής των \$125 μέσω του [Amazon](#).
- [Computer Telephony Integration](#)
- [PC Telephony: The Complete Guide to Designing, Building & Programming Systems Using Dialogic Systems](#)

4. Τα βιβλία που αφορούν τους επιστήμονες και τους ερευνητές είναι:  
The Latest Research and Methods in Speech Recognition

- [Statistical Methods for Speech Recognition](#)
- [Voice Recognition \(Klevans, Rodman\)](#)
- [Speech Coding: A Computer Laboratory Textbook](#)
- [Progress in Speech Synthesis](#)
- [Robustness in Automatic Speech Recognition: Fundamentals and Applications](#)
- [Fundamentals of Speech Recognition](#)

## Periodicals



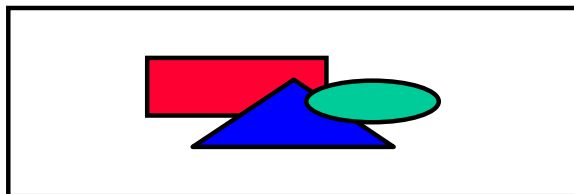
Ακολουθως παρέχονται οι διευθύνσεις και ορισμένες στοιχειώδεις πληροφορίες για κάποια επιλεγμένα περιοδικά που σχετίζονται με την αναγνώριση ομιλίας και τις εφαρμογές της:

- [ASRNews](#) - 12 τεύχη στην τιμή των \$325 -Για περισσότερες πληροφορίες: 781/861-6680.Σύνηθες μέγεθος: 14 σελίδες.
- Το [Computer Telephony](#) αποτελεί ένα μηνιαίο περιοδικό με θέματα που αφορούν την CT και τ' άλλα προϊόντα αναγνώρισης ομιλίας καθώς και συμβουλές για τις μεθόδους αγοράς, πώλησης και προώθησης των σχετικών εφαρμογών.Προσφέρονται 12 τεύχη στην τιμή των \$16.95 -Για περισσότερες πληροφορίες: 1-800-677-3435.Σύνηθες μέγεθος: 146 έγχρωμες σελίδες.
- Το [Speech Recognition Update](#) αποτελεί ένα καλογραμμένο, κατανοητό, επαγγελματικό περιοδικό για την αγορά, τα προϊόντα, τις επιχειρήσεις, τις εταιρίες και την τεχνολογία της αναγνώρισης ομιλίας.Προσφέρονται 12 τεύχη στην τιμή των \$375 -Για περισσότερες πληροφορίες: 818/708-0962.Σύνηθες μέγεθος: 24 σελίδες.
- Το [Speech Technology Magazine](#) προσφέρει δωρεάν συνδρομή σε ορισμένες κατηγορίες αναγνωστών και διατηρεί αρχείο με τα δημοσιευμένα άρθρα των περασμένων ετών.Τα άρθρα του καλύπτουν επαρκώς τις εξελίξεις στη βιομηχανία της αναγνώρισης ομιλίας τα τελευταία δύο χρόνια.Εκδίδεται κάθε τρεις μήνες και η ετήσια συνδρομή κοστολογείται στα \$12.95 -Για περισσότερες πληροφορίες: 203-834-1430.Μέγεθος συνηθισμένου τεύχους: 32 έγχρωμες σελίδες.
- [Teleconnect](#) - 12 τεύχη ετησίως στην τιμή των \$15 -Για περισσότερες πληροφορίες: 1-800-677-3435.Σύνηθες μέγεθος: 148 έγχρωμες σελίδες.
- [VoiceNews](#) - 12 τεύχη ετησίως στην τιμή των \$297 -Για περισσότερες πληροφορίες: 301/424-0114.Σύνηθες μέγεθος: 10 σελίδες.
- [Voice Processing Magazine](#) - Για περισσότερες πληροφορίες: 1-800-854-3112.

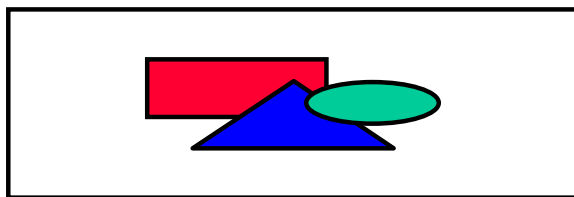
---

**Updated: February 9, 1999.**

**This site is maintained by Russ Wilcox as a service to those interested in the business side (as opposed to the technical underpinnings) of speech recognition. Content, suggestions, and corrections are heartily welcomed! Send [mail](#).**

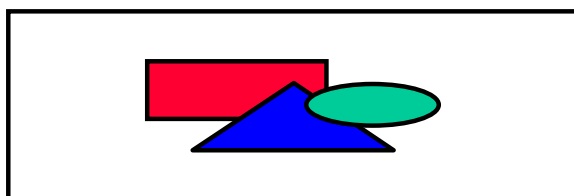


*As mentioned in the What's on the Web directory!*



*Listed as a Key Resource for the Speech topic.*

***DISCLAIMER: All selections and descriptions are my own.***



*Web Access Symbol*