ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μεταπτυχιακό Δίπλωμα Ειδίκευσης Ηλεκτρονικού Αυτοματισμού

Μαυροσκούφη Σταύρου (Α.Μ. 96515)

Μαυροσκούφη Βασιλικής (Α.Μ. 97537)

Ο Ρόλος της Φωνής στην Επικοινωνία Ανθρώπου Μηχανής

Εργασία στο μάθημα: Επικοινωνία με Ομιλία

Διδάσκων: Γεώργιος Κουρουπέτρογλου

Αθήνα 1999

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΙΝΑΚΑΣ ΣΥΝΤΜΗΣΕΩΝ

ARPA **Advanced Research Projects Agency**
Γραφείον Εργων Προχωρημένης Ερευνας

DMIs **Direct Manipulation Interfaces**
Διεπαφές Αμεσης Προσπέλασης

GUIs **Graphical User Interfaces**
Γραφικές Διεπαφές Χρήστη

ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

ΛΕΞΙΚΟ ΟΡΟΛΟΓΙΑΣ

carpal tunnel syndrome	σύνδρομο καρπικής σήραγγας	
coverage	κάλυψη	
desirability	επιθυμητότητα	
direct manipulation	άμεση προσπέλαση	
disfluencies	δυσχέρειες λόγου	
fine-grained	φυσικός	
intelligibility	κατανοητότητα	
guidelines	οδηγίες	
heuristics	ευριστικές μέθοδοι	
metaphor	αλληγορία	
modality	τρόπος	
multimodal	πολυτροπικός	
open ended	ανοικτό, δίχως τέλος	
perplexity	περιπλοκή	μέτρο της μέσης τιμής των πιθανοτήτων λέξεων κάθε κατάστασης της γραμματικής του μοντέλου γλώσσας ενός συστήματος.
recallability	ικανότητα ανάκλησης	
referent	αναφορά	
target system	σύστημα στόχος	
taxonomies	ταξινομίες	

Ο Ρόλος της Φωνής στην Επικοινωνία Ανθρώπου Μηχανής*

Philip R. Cohen και Sharon L. Oviatt

ΠΕΡΙΛΗΨΗ

Επικρατεί αισιοδοξία ότι στο εγγύς μέλλον θα σημειωθεί ραγδαία εξέλιξη στην αλληλεπίδραση ανθρώπου υπολογιστή με τη χρήση φωνής. Πρόσφατα κατασκευάσθηκαν πρωτότυπα συστημάτων που επιδεικνύουν αναγνώριση ομιλίας σε πραγματικό χρόνο ανεξάρτητη του ομιλητή και κατανόηση φυσικής γλώσσας για λεξιλόγια μετρίου μεγέθους (1000 με 2000 λέξεις). Ακόμη, διαφαίνονται στον ορίζοντα συστήματα αναγνώρισης ομιλίας για μεγαλύτερα λεξιλόγια. Ηδη οι κατασκευαστές υπολογιστών κατασκευάζουν υποσυστήματα αναγνώρισης ομιλίας στις νέες τους γραμμές παραγωγής. Ωστόσο, πριν η τεχνολογία αυτή χρησιμοποιηθεί ευρέως, απαιτείται να αποκτηθεί και να εφαρμοσθεί μία ουσιαστική γνωστική βάση σχετικά με την ανθρώπινη ομιλούμενη γλώσσα και τις επιδόσεις της, κατά τη διάρκεια αλληλεπίδρασης που βασίζεται σε υπολογιστή,. Αυτή η διατριβή κάνει ανασκόπηση των περιοχών εφαρμογών, στις οποίες η αλληλεπίδραση με ομιλία μπορεί να παίξει ένα σημαντικό ρόλο, προσδιορίζει πιθανά οφέλη που προκύπτουν από αλληλεπίδραση με μηχανές με ομιλία και προσπαθεί να συγκρίνει τη φωνή με εναλλακτικούς και συμπληρωματικούς τρόπους αλληλεπίδρασης ανθρώπου υπολογιστή. Η διατριβή επίσης αναφέρεται στις πληροφορίες που απαιτούνται για να κατασκευαστούν σταθερές εμπειρικές βάσεις για μελλοντική σχεδίαση διεπαφών ανθρώπου υπολογιστή. Τέλος, η διατριβή υποστηρίζει μια συστηματικότερη και επιστημονικότερη προσέγγιση στην κατανόηση της ανθρώπινης γλώσσας και απόδοσης με συστήματα αλληλεπίδρασης με φωνή.

* Η συγγραφή αυτής της διατριβής υποστηρίχθηκε εν μέρει από μία χορηγία του Εθνικού Ιδρύματος Επιστημών (No. IRI-9213472) στο SRI International.

ΕΙΣΑΓΩΓΗ

Από την αρχή της εποχής των υπολογιστών οι μελλοντολόγοι οραματίστηκαν τον διαλογικό υπολογιστή-μία μηχανή που θα μπορούσε να συμμετέχει σε συζητήσεις ομιλούμενης φυσικής γλώσσας. Για παράδειγμα, η περίφημη "δοκιμή" υπολογιστικής νοημοσύνης του Turing οραματίσθηκε έναν υπολογιστή που θα μπορούσε να φέρει εις πέρας μία συζήτηση στα Αγγλικά με τόση ευχέρεια που οι άνθρωποι δεν θα μπορούσαν να τη διακρίνουν από αυτή μεταξύ δύο ατόμων. Ωστόσο, παρ' όλη την εκτεταμένη έρευνα και τα πολλά αξιοσημείωτα επιστημονικά και τεχνολογικά επιτεύγματα, μέχρι πρόσφατα υπήρχαν ελάχιστοι διάλογοι ανθρώπου υπολογιστή και κανείς από αυτούς με ομιλία. Αυτή η κατάσταση έχει αρχίσει να μεταβάλλεται καθώς η σταθερή πρόοδος στην αναγνώριση ομιλίας και στις τεχνολογίες επεξεργασίας φυσικών γλωσσών, υποστηριζόμενη από σημαντικές προόδους στο υλικό υπολογιστών, κατέστησε δυνατή τη δημιουργία εργαστηριακών και πρωτότυπων συστημάτων, με τα οποία κάποιος μπορεί να συμμετέχει σε απλούς διαλόγους ερωτοαπαντήσεων. Παρ' όλο που αυτή η πρωταρχική δυνατότητα απέχει αρκετά από μία συζήτηση μεταξύ ανθρώπων, προξενεί σημαντικό ενδιαφέρον και γεννά αισιοδοξία για το μέλλον της αλληλεπίδρασης ανθρώπου υπολογιστή με χρήση φωνής.

Αυτή η διατριβή έχει ως σκοπό τον καθορισμό εφαρμογών, για τις οποίες η αλληλεπίδραση με ομιλία ενδέχεται να είναι ωφέλιμη, τη θέση της φωνής σε σχέση με εναλλακτικούς και συμπληρωματικούς τρόπους αλληλεπίδρασης ανθρώπου υπολογιστή και την εξέταση των εμποδίων που υπάρχουν για επιτυχημένη ανάπτυξη συστημάτων ομιλούμενης γλώσσας, λόγω της φύσης της αλληλεπίδρασης με ομιλούμενη γλώσσα.

Δύο γενικά είδη της τεχνολογίας εισόδου ομιλίας λαμβάνονται υπ' όψιν. Πρώτον, εξετάζεται ένας αριθμός υπαρχουσών εφαρμογών στις τεχνολογίες αναγνώρισης ομιλίας, για τις οποίες το σύστημα αναγνωρίζει τις λέξεις που λέγονται, αλλά δε χρειάζεται να κατανοεί το νόημα των λεγομένων. Κατά δεύτερο λόγο, επικεντρωνόμαστε σε εφαρμογές που απαιτούν πληρέστερη κατανόηση του προτιθέμενου νοήματος του ομιλητή, εξετάζοντας μελλοντικά συστήματα διαλόγου με ομιλία. Τέλος, συζητάμε πώς αυτή η κατανόηση ομιλίας θα παίξει ένα ρόλο σε μελλοντικές αλληλεπιδράσεις ανθρώπου υπολογιστή ειδικά σε όσες αφορούν στη συντονισμένη χρήση πολλαπλών τρόπων επικοινωνίας όπως τα γραφικα, η συγγραφή στο χέρι, και οι χειρονομίες. Υποστηρίζεται ότι η εξέλιξη παρεμποδίστηκε από την έλλειψη κατάλληλης επιστημονικής γνώσης σχετικά με αλληλεπιδράσεις του ανθρώπου με ομιλία ειδικά με υπολογιστές. Μία τέτοια γνωστική βάση είναι ουσιώδης στην ανάπτυξη καλά δομημένων οδηγιών διεπαφής με άνθρωπο που μπορούν να βοηθήσουν τους σχεδιαστές συστημάτων στην ανάπτυξη επιτυχημένων εφαρμογών που ενσωματώνουν αλληλεπίδραση με ομιλία. Δοθέντων των πρόσφατων τεχνολογικών επιτευγμάτων ο επιστημονικός χώρος είναι πλέον σε θέση να επεκτείνει συστηματικά αυτή τη γνωστική βάση.

Υπόβαθρο και Ορισμοί

Η αλληλεπίδραση ανθρώπου υπολογιστή με χρήση φωνής ενδέχεται να αφορά είσοδο ή έξοδο ομιλίας ίσως συνδυασμό αυτών των δύο ή συνδυασμό τους με άλλους τρόπους επικοινωνίας.

Ανάλυση Ομιλίας

Η διαδικασία της ανάλυσης ομιλίας συχνά χαρακτηρίζεται από 5 διαστάσεις:

- *Εξάρτηση από τον ομιλητή.* Οι αναγνωριστές ομιλίας περιγράφονται ως εξαρτημένοι/εκπαιδευόμενοι από τον ομιλητή, προσαρμοσμένοι στον ομιλητή και ανεξάρτητοι του ομιλητή. Για αναγνώριση εξαρτόμενη από τον ομιλητή συλλέγονται δείγματα ομιλίας του χρήστη και χρησιμοποιούνται ως μοντέλα για τα λεγόμενα του(της) που θα επακολουθήσουν. Για αναγνώριση προσαρμοσμένη στον ομιλητή διατίθενται αρχικά παραμετροποιημένα ακουστικά μοντέλα τα οποία μπορούν να ρυθμιστούν τελικά για ένα δεδομένο χρήστη μέσω της προφοράς ενός περιορισμένου συνόλου συγκεκριμένων λεγομένων. Τέλος, οι ανεξάρτητοι του ομιλητή αναγνωριστές σχεδιάζονται για να χειριστούν ομιλία οποιουδήποτε χρήστη, χωρίς εκπαίδευση, στο δεδομένο θέμα ομιλίας (βλέπε Flanagan, σε αυτό το τεύχος).

- *Συνέχεια ομιλίας.* Οσα λέγονται μπορούν να λεχθούν κατά ένα απομονωμένο τρόπο, με διακοπές μεταξύ των λέξεων ή ως συνεχής φυσική ομιλία.

- *Τύπος ομιλίας.* Προκειμένου να αναπτύξουν αρχικούς αλγορίθμους οι ερευνητές αρχικά χρησιμοποιούν ως δεδομένα αναγνωσμένη ομιλία, στην οποία οι ομιλητές αναγνώσκουν τυχαίες προτάσεις που προέρχονται από κάποιο σώμα κειμένου, όπως η Wall Street Journal. Σε συνέχεια αυτού του σταδίου ανάπτυξης αλγορίθμων η έρευνα αναγνώρισης ομιλίας προσπαθεί να χειριστεί αυθόρμητη ομιλία, στην οποία οι ομιλητές παράγουν νέα λεγόμενα στο επιλεγμένο θέμα ομιλίας.

- *Αλληλεπίδραση.* Ορισμένες δραστηριότητες αναγνώρισης ομιλίας, όπως η υπαγόρευση, μπορούν να χαρακτηριστούν ως μη αλληλεπιδραστικές, με την έννοια ότι ο ομιλητής δεν λαμβάνει ανατροφοδότηση από τον(τους) προτιθέμενο(ους) ακροατή(ές). Αλλα συστήματα σχεδιάζονται για επεξεργασία αλληλεπιδραστικής ομιλίας, στην οποία οι ομιλητές κατασκευάζουν τα λεγόμενα τους ως τμήμα μιας ανταλλαγής της σειράς ομιλίας με ένα σύστημα ή με άλλον ομιλητή.

- *Λεξιλόγιο και Γραμματική.* Ο χρήστης μπορεί να χρησιμοποιήσει λέξεις από ένα στενά περιορισμένο λεξιλόγιο και γραμματική ή από ευρύτερα λεξιλόγια και γραμματικές που προσεγγίζουν καλύτερα τα αντίστοιχα μιας φυσικής γλώσσας. Το λεξιλόγιο και η γραμματική του συστήματος μπορούν να επιλεγούν από τον σχεδιαστή του συστήματος ή αυτόν που αναπτύσσει την εφαρμογή είτε να προκύψουν από δεδομένα βασισμένα σε πραγματικούς χρήστες που μιλούν είτε σε ένα προσομοιωμένο σύστημα είτε σε ένα πρωτογενές πρωτότυπο συστήματος. Οι τρέχουσες τεχνολογίες αναγνώρισης ομιλίας απαιτούν μία εκτίμηση της πιθανότητας εμφάνισης κάθε λέξης στο πλαίσιο των άλλων λέξεων του λεξιλογίου. Επειδή αυτές οι πιθανότητες τυπικά προσεγγίζονται από την κατανομή των λέξεων σε ένα δεδομένο σώμα κειμένου, είναι δύσκολο αυτή τη στιγμή να επεκταθεί ένα λεξιλόγιο συστήματος παρ' όλο που η έρευνα προσανατολίζεται σε αναγνώριση ανεξάρτητη λεξιλογίου (Hon και Lee, 1991).

Οι πωλητές συχνά αναφέρουν ότι το υλικό τους για αναγνώριση ομιλίας προσφέρει πολύ υψηλή ακρίβεια αναγνώρισης, αλλά είναι μόνο στα πλαίσια ποσοτικής κατανόησης της διαδικασίας αναγνώρισης που κάποιος μπορεί να

συγκρίνει με σημασία την απόδοση των αναγνωριστών. Για να αξιολογήσουν την δυσκολία μιας δεδομένης διαδικασίας αναγνώρισης για ένα δεδομένο σύστημα, οι ερευνητές χρησιμοποιούν ένα μέτρο της *περιπλοκής (perplexity)* του μοντέλου γλώσσας αυτού του συστήματος, το οποίο μετρά, μιλώντας πρόχειρα, τη μέση τιμή των πιθανοτήτων λέξεων σε κάθε κατάσταση της γραμματικής (Bahl και al., 1983 Baker, 1975 Jelinek, 1976). Η ακρίβεια αναγνώρισης λέξης βρέθηκε, γενικά να είναι αντιστρόφως ανάλογη της περιπλοκής. Τα περισσότερα εμπορικά συστήματα προσφέρουν συστήματα αναγνώρισης ομιλίας που ισχυρίζονται ότι πετυχαίνουν ποσοστό ακρίβειας αναγνώρισης λέξης > 95% δεδομένης μίας περιπλοκής της τάξης του 10. Τουλάχιστον ένας πωλητής προσφέρει ένα σύστημα 1000 με 5000 λέξεων, ανεξάρτητο του ομιλητή, με περιπλοκές στο διάστημα 66 με 433 και ένα αντίστοιχο σφάλμα αναγνώρισης λέξης με ποσοστό του 3 με 15% για αναγνώριση μεμονωμένων λέξεων (Baker, 1991). Τα σημερινά εργαστηριακά συστήματα υποστηρίζουν αναγνώριση συνεχόμενης ομιλίας πραγματικού χρόνου, ανεξάρτητη του ομιλητή που προέρχεται από λεξιλόγιο περίπου 1500 λέξεων με περιπλοκή 50 έως 70, έχοντας ως αποτέλεσμα ρυθμούς σφάλματος αναγνώρισης λέξης μεταξύ 4 και 8% (Pallet και al., 1993). Τα πιο φιλόδοξα ανεξάρτητα του ομιλητή συστήματα αναγνωρίζουν αυτή τη στιγμή σε πραγματικό χρόνο, αναγνωσμένη ομιλία που προέρχεται από ένα λεξιλόγιο 5000 λέξεων του κειμένου της Wall Street Journal, με περιπλοκή 120, και έχουν ως αποτέλεσμα ρυθμό σφάλματος αναγνώρισης λέξης της τάξης του 5% (Pallet και al., 1993). Γίνονται προσπάθειες για ευρύτερα λεξιλόγια.

Το τελικό αποτέλεσμα της αναγνώρισης φωνής είναι η(οι) σειρά(ές) λέξεων υψηλότερης τάξης ή συχνά το πλέγμα των λέξεων που καλύπτει το σήμα. Για μικρά λεξιλόγια και στενά περιορισμένες γραμματικές ένας απλός διερμηνευτής μπορεί να απαντήσει στις ομιλούμενες λέξεις απ' ευθείας. Ωστόσο, για μεγαλύτερα λεξιλόγια και φυσικότερες γραμματικές πρέπει να εφαρμοσθεί στην έξοδο του αναγνωριστή η *κατανόηση φυσικής γλώσσας*, ώστε να ανακτηθεί το προτιθέμενο νόημα των λεγομένων.[1] Επειδή αυτή η διαδικασία κατανόησης της φυσικής γλώσσας είναι πολύπλοκη και δίχως τέλος, συχνά περιορίζεται από την εφαρμογή (π.χ. ανάκτηση πληροφοριών από μία βάση δεδομένων) και από το θέμα ομιλίας (π.χ. μία βάση δεδομένων σχετικά με αεροπορικές πτήσεις). Στο σημείο αυτό ο συνδυασμός αναγνώρισης ομιλίας και κατανόησης γλώσσας θα ορισθεί ως *κατανόηση ομιλίας* και τα συστήματα που χρησιμοποιούν τέτοια είσοδο θα ορισθούν ως *συστήματα ομιλούμενης γλώσσας*. Αυτή η διατριβή κάνει μια ανασκόπηση προηγούμενης εργασίας στις χρήσεις της αναγνώρισης ομιλίας, αλλά εστιάζει στις χρήσεις της ομιλούμενης γλώσσας.

Σύνθεση ομιλίας

Υπάρχουν τρεις τύποι τεχνολογίας σύνθεσης ομιλίας:

- *Ψηφιοποιημένη ομιλία.* Για την παραγωγή ενός λεγομένου η μηχανή συγκεντρώνει και παίζει ξανά προηγούμενα εγγεγραμμένα και συμπιεσμένα δείγματα ανθρώπινης ομιλίας. Παρ' όλο που μπορεί συχνά να ακουστεί μία αισθητή διακοπή μεταξύ δειγμάτων και ο συνολικός επιτονισμός ενδέχεται να είναι ανακριβής, μία τέτοια διαδικασία σύνθεσης μπορεί να προσφέρει ομιλία

[1] Ανέτρεξε στον Moore (σε αυτό το τεύχος) για μία συζήτηση σχετικά με το πώς αυτά τα στοιχεία μπορούν να ολοκληρωθούν.

υψηλής κατανοητότητας (intelligibility) που ηχεί ως ανθρώπινη. Αυτή η διαδικασία είναι ωστόσο περιορισμένη στην παραγωγή συνδυασμών των εγγεγραμμένων δειγμάτων.

- *Κείμενο σε ομιλία.* Η σύνθεση αυτή εμπλέκει μία αυτοματοποιημένη ανάλυση της δομής των λέξεων μέσα στο μορφολογικά τους συστατικά μέρη. Συνδυάζοντας τις προφορές αυτών των μονάδων υπολέξεων σύμφωνα με τους κανόνες γραμματικής και μορφοποίησης σε ήχο με μία ευρεία λίστα προφορών που αποτελούν εξαιρέσεις (για τα Αγγλικά), αυθαίρετο κείμενο μπορεί να αποδοθεί ως ομιλία. Επειδή αυτή η τεχνολογία μπορεί να χειριστεί κείμενο δίχως τέλος είναι βολική για εφαρμογές ευρείας κλίμακας, όπως είναι η ανάγνωση κειμένου δυνατά σε τυφλούς χρήστες ή η ανάγνωση ηλεκτρονικού ταχυδρομείου μέσω του τηλεφώνου. Η επιστήμη και η τεχνολογία κειμένου σε ομιλία καλύπτονται ευρέως σε άλλο σημείο αυτού του τεύχους (ανέτρεξε σε Allen και Carlson σε αυτό το τεύχος).

- *Εννοια σε ομιλία.* Με τα συστήματα κειμένου σε ομιλία το κείμενο που θα μετατραπεί προέρχεται από μία ανθρώπινη πηγή. Μελλοντικά συστήματα διαλόγου θα απαιτήσουν υπολογιστές να αποφασίζουν για τους ίδιους τι θα πουν και πως θα το πουν, ώστε να καταλήξουν σε μία σημαντική και αρμόζουσα στα συμφραζόμενα συμμετοχή σε διάλογο. Τέτοια συστήματα απαιτείται να καθορίσουν ποιες ενέργειες ομιλίας θα εκτελεστούν (π.χ. αίτηση, πρόταση), πώς θα αναφερθούμε σε οντότητες λεγομένων, τι θα πούμε σχετικά με αυτές, ποιες γραμματικές μορφές θα χρησιμοποιηθούν και ποιος επιτονισμός θα εφαρμοσθεί. Επιπλέον, το λεγόμενο θα έπρεπε να συμβάλλει στην εξέλιξη του διαλόγου έτσι το σύστημα θα έπρεπε να διατηρήσει μία εικόνα αυτών που έχουν ειπωθεί, ώστε να αναλυθούν και να γίνουν κατανοητά τα επακόλουθα λεγόμενα του χρήστη.

Οι περιοχές έρευνας της σύνθεσης ομιλίας και της δημιουργίας γλώσσας, έτυχαν σημαντικά λιγότερης προσοχής από την αναγνώριση και κατανόηση ομιλίας, αλλά θα θεωρηθούν σημαντικές, καθώς η δυνατότητα ανάπτυξης συστημάτων διαλογικής ομιλίας γίνεται πραγματοποιήσιμη.

Το υπόλοιπο αυτής της διατριβής εξερευνά τρέχουσες και μελλοντικές περιοχές εφαρμογών, στις οποίες η αλληλεπίδραση με ομιλία μπορεί να είναι ένας προτιμητέος τρόπος επικοινωνίας με υπολογιστές. Κατ' αρχάς καθορίζονται παράγοντες που μπορούν να επηρεάσουν την επιθυμητότητα (desirability) και την αποτελεσματικότητα της αλληλεπίδρασης με υπολογιστές που βασίζεται σε φωνή, ανεξάρτητα από το αν ομιλείται μία απλή γλώσσα εντολών ή μία ψευδοφυσική γλώσσα. Ακολουθεί συζήτηση για την αλληλεπίδραση με ομιλούμενη γλώσσα και σύγκρισή της με την αλληλεπίδραση που βασίζεται στο πληκτρολόγιο και το τρέχον κυρίαρχο παράδειγμα της γραφικής διεπαφής χρήστη. Αφού προσδιορισθούν οι συνθήκες που ευνοούν την αλληλεπίδραση με ομιλούμενη γλώσσα, αναγνωρίζονται κάποια κενά στην επιστημονική γνωστική βάση επικοινωνίας με ομιλία που παρουσιάζουν εμπόδια στην ανάπτυξη συστημάτων που βασίζονται σε ομιλούμενη γλώσσα. Παρατηρείται ότι τα μελλοντικά συστήματα θα είναι πολυτροπικά με τη φωνή να είναι ένας από τους διαθέσιμους τρόπους επικοινωνίας. Καταλήγουμε με προτάσεις για περαιτέρω έρευνα που απαιτείται να διεξαχθεί για την υποστήριξη της ανάπτυξης μονοτροπικών και πολυτροπικών συστημάτων που βασίζονται σε φωνή και υποστηρίζουμε ότι υπάρχει μία πιεστική ανάγκη για δημιουργία εμπειρικών οδηγιών διεπαφής του ανθρώπου, για όσους αναπτύσσουν συστήματα, πριν η

τεχνολογία που βασίζεται στη φωνή μπορέσει να πραγματοποιήσει τους ενδεχόμενους στόχους της.

ΠΟΤΕ Η ΑΛΛΗΛΕΠΙΔΡΑΣΗ ΜΕ ΥΠΟΛΟΓΙΣΤΕΣ ΜΕ ΟΜΙΛΙΑ ΕΙΝΑΙ ΧΡΗΣΙΜΗ;

Εως τώρα δεν υπάρχει θεωρία ή κατηγοριοποίηση σε δραστηριότητες και περιβάλλοντα που θα μπορούσαν να προβλέψουν(, όντας όλα ισότιμα,) πότε η φωνή θα ήταν ένας προτιμητέος τρόπος επικοινωνίας ανθρώπου υπολογιστή. Ωστόσο, καθορίσθηκε ένας αριθμός περιπτώσεων, στις οποίες η επικοινωνία με μηχανές με ομιλία μπορεί να είναι πλεονεκτική:

- όταν τα χέρια ή οι οφθαλμοί του χρήστη είναι απασχολημένοι,
- όταν μόνο ένα περιορισμένο πληκτρολόγιο και/ή μία οθόνη είναι διαθέσιμα,
- όταν ο χρήστης είναι ανίκανος,
- όταν η προφορά είναι το αντικειμενικό ζήτημα της χρήσης υπολογιστή, ή
- όταν προτιμάται η αλληλεπίδραση με φυσική γλώσσα.

Εξετάζονται εν συντομία οι παρόντες και οι μελλοντικοί ρόλοι της αλληλεπίδρασης με υπολογιστές με ομιλία γι' αυτά τα περιβάλλοντα. Επειδή η αλληλεπίδραση με ομιλούμενη φυσική γλώσσα είναι η δυσκολότερη στην εφαρμογή της, γίνεται εκτενής μελέτη του θέματος στην παράγραφο με τίτλο "Αλληλεπίδραση με Φυσική Γλώσσα".

Είσοδος Φωνής

Δραστηριότητες απασχολημένων χεριών/οφθαλμών

Η κλασική περίπτωση που ευνοεί την αλληλεπίδραση με μηχανές με ομιλία είναι αυτή στην οποία τα χέρια και/ή οι οφθαλμοί του χρήστη είναι απασχολημένα, εκτελώντας κάποια άλλη εργασία. Σε τέτοιες συνθήκες, χρησιμοποιώντας φωνή για επικοινωνία με τη μηχανή,. οι άνθρωποι είναι ελεύθεροι να δώσουν τη δέουσα προσοχή στη δραστηριότητά τους παρά να τη διακόψουν για να χρησιμοποιήσουν ένα πληκτρολόγιο. Μελέτες στο πεδίο αυτό προτείνουν πως, για παράδειγμα, οι πιλότοι των F-16 που μπορούν να επιτύχουν υψηλό συντελεστή αναγνώρισης ομιλίας μπορούν να φέρουν εις πέρας αποστολές όπως ο σχηματισμός πτήσης ή η πλοήγηση σε χαμηλό επίπεδο, ταχύτερα και ακριβέστερα, όταν χρησιμοποιούν έλεγχο με ομιλία πάνω σε διάφορα αεροπορικά υποσυστήματα σε σύγκριση με το πληκτρολόγιο και την εισαγωγή δεδομένων με πολυλειτουργικά πλήκτρα (Howard, 1987 Rosenhoover και al., 1987 Williamson, 1987). Παρόμοια αποτελέσματα βρέθηκαν για τους πιλότους ελικοπτέρων σε ενθόρυβα περιβάλλοντα κατά τη διάρκεια διαδικασιών ανίχνευσης και επικοινωνίας (Simpson και all 1985, Swider 1987).[2]

[2] Για περαιτέρω συζήτηση σχετικά με αναγνώριση ομιλίας σε στρατιωτικά περιβάλλοντα ανατρέξτε στον (Weinstein 1991, σε αυτό το τεύχος).

Αφθονούν επίσης οι εμπορικές εφαρμογές απασχολημένων χεριών–οφθαλμών. Για παράδειγμα, οι εγκαταστάτες συρμάτων που ζητούσαν τον σειριακό αριθμό ενός σύρματος και στη συνέχεια οδηγούνταν προφορικα από τον υπολογιστή για την εγκατάσταση του σύρματος, πέτυχαν αύξηση παραγωγικότητας κατά 20 με 30% με βελτιωση στην ακρίβεια και μικρότερο χρόνο εκπαίδευσης σε σχέση με την προγενέστερη μέθοδο καθορισμού και εγκατάστασης σύρματος (Marshall, 1992). Οι ταξινομητές πακέτων που ζητούσαν τα ονόματα των πόλεων, αντί να πληκτρολογούν κλειδιά για την ετικέτα προορισμού, πέτυχαν βελτίωση του χρόνου εισαγωγής κατά 37% κατά τη διάρκεια εργασιών με απασχολημένα χέρια-οφθαλμούς. (Visick και al., 1984). Ωστόσο, όταν το στοιχείο απασχολημένων χεριών-οφθαλμών στη διανομή πακέτων απομακρύνθηκε, η είσοδος με ομιλία δεν προσέφερε αισθητά πλεονεκτήματα ταχύτητας. Επιπρόσθετα, οι σχεδιαστές VLSI κυκλωμάτων μπορούσαν να ολοκληρώσουν κατά 24% περισσότερες εργασίες, όταν ήταν διαθέσιμες ομιλούμενες εντολές, από όταν χρησιμοποιούσαν μόνο διεπαφή πληκτρολογίου ή ποντικιού (ανέτρεξε στην παράγραφο με τίτλο "Αμεση Προσπέλαση") (Martin, 1989). Παρ' όλο που οι μελέτες στο συγκεκριμένο πεδίο σπάνια οδηγούν σε συμπέρασμα, πολλές μελέτες συστημάτων αναγνώρισης ομιλίας υψηλής ακρίβειας σε δραστηριότητες απασχολημένων χεριών-οφθαλμών βρήκαν ότι η είσοδος με ομιλία οδηγεί σε υψηλότερη παραγωγικότητα και ακρίβεια.

Η είσοδος με ομιλία όχι μόνο προσφέρει αύξηση της αποτελεσματικότητας για δεδομένη δραστηριότητα απασχολημένων χεριών-οφθαλμών, αλλά επίσης προσφέρει το ενδεχόμενο μεταβολής της φύσης αυτής της δραστηριότητας κατά ωφέλιμο τρόπο. Για παράδειγμα, αντί να πρέπει να θυμάται και να λέει ή να πληκτρολογεί τα γράμματα "YY" για να δηλώσει ένα αεροδρόμιο προορισμού, ένας χειριστής αποσκευών θα μπορούσε απλά να πει "Toronto", επομένως να χρησιμοποιήσει ένα όνομα ευκολομνημόνευτο (Martin 1989, Nye 1982). Παρόμοια πιθανά πλεονεκτήματα αναγνωρίζονται για τηλεφωνικές συσκευές που βασίζονται σε φωνή, στις οποίες κάποιος μπορεί να πει "κάλεσε τον Τομ" από το να πρέπει να θυμάται και να εισάγει έναν αριθμό τηλεφώνου (Rabiner και al. 1980). Αλλες δραστηριότητες με απασχολημένα χέρια-οφθαλμούς που ενδεχομένως να επωφεληθούν από την αλληλεπίδραση με φωνή, περιλαμβάνουν εισαγωγή δεδομένων και έλεγχο μηχανών σε εργοστάσια και εφαρμογές επιστημονικών χώρων (Martin, 1976), πρόσβαση σε πληροφορίες για στρατιωτικό έλεγχο και εντολές, διαχείριση πληροφορίας αστροναυτών κατά τη διάρκεια πρόσθετης πρόσβασης των οχημάτων στο διάστημα, υπαγόρευση ιατρικών διαγνώσεων (Baker, 1991) συντήρηση και επιδιόρθωση εξοπλισμού, έλεγχο εξοπλισμού αυτοκινήτου (π.χ. ραδιόφωνο, τηλέφωνα, κλιματιστικός έλεγχος) και βοηθημάτων πλοήγησης (Streeter και al. 1985).

Ενας βασικός παράγοντας που καθορίζει την επιτυχία των εφαρμογών εισόδου ομιλίας είναι η ακρίβεια αναγνώρισης ομιλίας. Για παράδειγμα, η βέλτιστη επίδοση που αναφέρθηκε κατά τη διάρκεια δοκιμαστικών πτήσεων F-16 επετεύχθη, όταν οι πιλότοι πέτυχαν ρυθμούς αναγνώρισης μεμονωμένων λέξεων μεγαλύτερους του 95%. Κάτω του 90% η απαιτούμενη προσπάθεια για τη διόρθωση σφαλμάτων αναγνώρισης θεωρήθηκε ότι υπερτερούσε των πλεονεκτημάτων για το χρήστη (Howard, 1987). Παρόμοια αποτελέσματα που φανερώνουν την εξάλειψη πλεονεκτημάτων, όταν ληφθεί υπ' όψιν η διόρθωση

σφαλμάτων, βρέθηκαν σε δραστηριότητες τόσο απλές όσο η εισαγωγή συνδεδεμένων ψηφίων (Hauptmann και Rudnicky, 1990).

Προκειμένου να επιτευχθεί ένα επαρκώς υψηλό επίπεδο ακρίβειας αναγνώρισης σε δοκιμαστικά πεδία, η είσοδος με ομιλία περιορίσθηκε σημαντικά για να επιτραπεί μόνο ένας μικρός αριθμός πιθανών λέξεων σε κάθε δεδομένη στιγμή. Ωστόσο, ακόμη και με αυτούς τους περιορισμούς η ακρίβεια συχνά μένει πίσω από τις εργαστηριακές δοκιμές, λόγω πολλών πολύπλοκων παραγόντων, όπως η φυσική και η συναισθηματική κατάσταση του χρήστη, ο ατμοσφαιρικός θόρυβος, ο μικροφωνικός εξοπλισμός, οι απαιτήσεις των πραγματικών δραστηριοτήτων, οι μέθοδοι του χρήστη και του συστήματος εκπαίδευσης και οι προσωπικές διαφορές που συναντώνται, όταν ένα πλήθος πραγματικών χρηστών δειγματοληπτείται. Ωστόσο, υπάρχει ο ισχυρισμός ότι οι περισσότερες αποτυχίες της τεχνολογίας ομιλίας ήταν αποτέλεσμα της μηχανικής και της διαχείρισης ανθρώπινων παραγόντων (Lea, 1992) παρά της χαμηλής ακρίβειας αναγνώρισης ανά χρήστη. Τα ζητήματα ανθρώπινων παραγόντων συζητούνται εκτενέστερα στη συνέχεια και από τον Kamm (σε αυτό το τεύχος).

Επιλογή περιορισμένου πληκτρολογίου/οθόνης

Οι επικρατέστερες τρέχουσες χρήσεις αναγνώρισης και σύνθεσης ομιλίας είναι οι εφαρμογές που βασίζονται στο τηλέφωνο. Οι αναλυτές ομιλίας χρησιμοποιούνται συνήθως στη βιομηχανία των τηλεπικοινωνιών για την υποστήριξη βοήθειας τηλεφωνικού καταλόγου αναφέροντας το επιθυμητό νούμερο στον καλούντα και ως εκ τούτου ελευθερώνοντας τον τηλεφωνητή για να χειρισθεί άλλη κλήση. Οι αναγνωριστές ομιλίας αναπτύχθηκαν για να αντικαταστήσουν ή να βελτιώσουν τηλεφωνικές υπηρεσίες (π.χ. συλλογή κλήσεων) χειριζόμενοι εκατοντάδες εκατομμύρια καλούντων κάθε έτος και έχοντας ως αποτέλεσμα κέρδη πολλών εκατομμυρίων δολλαρίων (Lennig 1989, Nakatsu και Wilpon σε αυτό το τεύχος). Οι αναγνωριστές ομιλίας για τηλεπικοινωνιακές εφαρμογές δέχονται ένα πολύ περιορισμένο λεξιλόγιο, ίσως ξεχωρίζοντας μόνο ορισμένες λέξεις κλειδιά στην είσοδο, αλλά απαιτείται να λειτουργούν με υψηλή πιστότητα για ένα ευρύ φάσμα του κοινού. Παρ' όλο που δεν θεωρούνται τόσο σοβαρές όσο οι αεροπορικές ή οι κατασκευαστικές εφαρμογές, οι τηλεπικοινωνιακές εφαρμογές θεωρούνται δύσκολες διότι οι καλούντες λαμβάνουν ελάχιστη ή καθόλου εκπαίδευση σχετικά με τη χρήση του συστήματος και ενδεχομένως να διαθέτουν εξοπλισμό χαμηλού επιπέδου, ενθόρυβες τηλεφωνικές γραμμές και απρόβλεπτα επίπεδα ατμοσφαιρικού θορύβου. Επιπλέον, είναι δύσκολο να προβλεφθεί και να δρομολογηθεί η συμπεριφορά του καλούντος (Basson, 1992 Kamm σε αυτό το τεύχος Spitz, 1991).[3]

Η σημαντική επιτυχία στην αυτοματοποίηση των απλούστερων τηλεφωνικών υπηρεσιών ανοίγει το δρόμο για περισσότερο φιλόδοξες εφαρμογές που βασίζονται στο τηλέφωνο, όπως η πρόσβαση σε πληροφορίες απομακρυσμένων βάσεων δεδομένων. Για παράδειγμα, ο καλών ενδέχεται να ζητήσει δρομολόγια τρένων και αεροπλάνων (ARPA 1993, Πρακτικά του Συνεδρίου Ομιλίας και Φυσικής Γλώσσας 1991, Peckham 1991), πληροφορίες τηλεφωνικού καταλόγου, ή ισοζύγια τραπεζικών λογαριασμών (Nakatsu, σε αυτό

[3] Για ένα θαυμάσιο απολογισμό των ανθρωπίνων παραγόντων και των τεχνικών δυσκολιών στις τηλεπικοινωνιακές εφαρμογές αναγνώρισης ομιλίας ανατρέξτε στους Karis και Dobroth (1991).

το τεύχος) και να λάβει την απάντηση ακουστικά. Αυτός ο γενικός χώρος της αλληλεπίδρασης ανθρώπου υπολογιστή είναι δυσκολότερο να υλοποιηθεί σε σχέση με τις απλές τηλεφωνικές υπηρεσίες, επειδή η κλίμακα της συμπεριφοράς του καλούντα είναι αρκετά ευρεία και επειδή απαιτούνται κατανόηση ομιλίας και συμμετοχή σε διάλογο παρά αναγνώριση λέξεων. Οταν απαιτείται να μεταβιβαστούν ακόμη και μέτριες ποσότητες δεδομένων, μία αλληλεπίδραση καθαρά με φωνή μπορεί να είναι δύσκολο να διεξαχθεί, παρ' όλο που η εμφάνιση των "εικονοτηλεφώνων" μπορεί κάλλιστα να βελτιώσει τέτοιες περιπτώσεις.

Η προκλητικότερη πιθανή εφαρμογή της τεχνολογίας ομιλούμενης γλώσσας που βασίζεται στο τηλέφωνο ίσως είναι η διερμηνεία της τηλεφωνίας (Kourematsu, 1992, Roe και al. 1991), στην οποία δύο καλούντες που μιλούν διαφορετικές γλώσσες μπορούν να συμμετέχουν σε ένα διάλογο με τη μεσολάβηση ενός συστήματος μετάφρασης ομιλούμενης γλώσσας (Kitano 1991, Yato και al. 1992). Τέτοια συστήματα σχεδιάζονται αυτή τη στιγμή για να ενσωματώνουν αναγνώριση ομιλίας, μετάφραση μηχανής και υποσυστήματα σύνθεσης ομιλίας και να διερμηνεύουν μία πρόταση κάθε στιγμή. Ενα πρόσφατο αρχικό πείραμα που οργανώθηκε από την ATR International (Ιαπωνία) με το Carnegie-Mellon University (ΗΠΑ) και τη Siemens A.G. (Γερμανία) ενέπλεξε διερμηνευόμενους μέσω μηχανής διαλόγους από Ιαπωνικά σε Αγγλικά και από Ιαπωνικά σε Γερμανικά (Pollack, 1993 Yato και al., 1992). Τα λεγόμενα μιας γλώσσας αναγνωρίσθηκαν και μεταφράσθηκαν από έναν τοπικό υπολογιστή, ο οποίος έστειλε μία μεταφρασμένη απόδοση του κειμένου στο ξένο σημείο, όπου πραγματοποιήθηκε μία σύνθεση από κείμενο σε ομιλία. Η AT&T έκανε επίδειξη ενός περιορισμένου συστήματος μετάφρασης ομιλίας από Αγγλικά σε Ιαπωνικά (Roe και al., 1991), παρ' όλη την ανυπαρξία αντίστοιχου συστήματος βασισμένου σε τηλέφωνο και η Nippon Electric Corporation έκανε επίδειξη παρόμοιου συστήματος από Ιαπωνικά σε Αγγλικά.

Πέραν της χρήσης τηλεφώνου ένας δεύτερος παράγοντας που συνδέεται με τον εξοπλισμό και που ευνοεί την αλληλεπίδραση που βασίζεται σε φωνή είναι το διαρκώς μειωμένο μέγεθος των φορητών υπολογιστών. Οι φορητοί υπολογιστές και οι συσκευές επικοινωνιών σύντομα θα είναι πολύ μικρές για να επιτρέψουν τη χρήση ενός πληκτρολογίου υποδηλώνοντας ότι οι τρόποι εισόδου για τέτοιες μηχανές θα είναι πιθανότατα η ψηφιοποιημένη πένα και η φωνή (Crane, 1991 Ivuattm 1992) με την οθόνη και τη φωνή να παρέχουν σύστημα εξόδου. Δεδομένου ότι αυτές οι συσκευές προτίθενται να αντικαταστήσουν και το τηλέφωνο οι χρήστες ήδη θα μιλούν *μέσω* αυτών. Μία φυσική εξέλιξη των συσκευών θα προσφέρει στο χρήστη τη δυνατότητα να μιλά επίσης *σε* αυτές.

Τέλος, ένα προκύπτον όφελος της τεχνολογίας φωνής είναι η αντικατάσταση των πολλών πλήκτρων ελέγχου στις ηλεκτρονικές συσκευές του καταναλωτή(π.χ, βιντεοκάμερα, τηλεφωνικοί δέκτες). Καθώς αυξάνει ο αριθμός των λειτουργιών που ελέγχονται από το χρήστη σε αυτές τις συσκευές, η διεπαφή με το χρήστη γίνεται υπερβολικά πολύπλοκη και μπορεί να οδηγήσει σε σύγχυση, όσον αφορά το πως θα εκτελεστούν ακόμη και απλές δραστηριότητες. Πρόσφατα ανακοινώθηκαν προϊόντα που επιτρέπουν στους χρήστες να προγραμματίσουν τις συσκευές τους, χρησιμοποιώντας απλές εντολές φωνής.

Ανικανότητα

Ενα βασικό ενδεχόμενο όφελος της τεχνολογίας φωνής θα είναι να βοηθήσει κουφούς χρήστες στην επικοινωνία με ανθρώπους που ακούνε χρησιμοποιώντας ένα τηλέφωνο (Bernstein, 1988). Ενα τέτοιο σύστημα θα αναγνώριζε την ομιλία του ατόμου που ακούει, θα την απέδιδε ως κείμενο και θα συνέθετε την απάντηση κειμένου του κουφού ατόμου (αν χρησιμοποιείται ένα τερματικό) ως ένα λεγόμενο. Αλλη χρήση της αναγνώρισης ομιλίας στην υπηρεσία των κουφών χρηστών θα ήταν η παραγωγή σε πραγματικό χρόνο υποτίτλων τηλεοπτικών προγραμμάτων ή ταινιών. Η αναγνώριση ομιλίας θα μπορούσε επίσης να χρησιμοποιηθεί από κινητικά ανίκανους χρήστες για να ελέγχουν ηλεκτρικά είδη του νοικοκυριού, αναπηρικές καρέκλες και μηχανικά πρόσθετα. Η σύνθεση από κείμενο σε ομιλία μπορεί να βοηθήσει χρήστες με δυσχέρειες στην ομιλία και την κίνηση, να βοηθήσει τυφλούς χρήστες με την αλληλεπίδραση με υπολογιστή και όταν συνδυαστεί με τεχνολογία αναγνώρισης οπτικού χαρακτήρα μπορεί να αναγνώσει τυπωμένο υλικό σε τυφλούς χρήστες. Τέλος, με δεδομένες τις επαρκείς δυνατότητες των συστημάτων αναγνώρισης ομιλίας η ομιλούμενη είσοδος μπορεί να γίνει μία προκαθορισμένη θεραπεία για επαναλαμβανόμενες κακώσεις άγχους, όπως το σύνδρομο καρπικής σήρραγας, το οποίο εκτιμάται ότι προσβάλει περίπου το 1,5% των εργαζομένων γραφείου σε ασχολίες που τυπικά εμπεριέχουν τη χρήση πληκτρολογίων (Tanaka και al., 1993) παρ' όλο που οι ίδιοι οι αναγνωριστές ομιλίας μπορούν να οδηγήσουν σε διαφορετικές βλάβες λόγω διαρκούς άγχους (Markinson, προσωπική επικοινωνία, 1993).[4]

Το αντικειμενικό ζήτημα είναι η προφορά

Η αναγνώριση ομιλίας θα γίνει συστατικό μέρος μελλοντικών βοηθημάτων βασισμένων σε υπολογιστή για την εκμάθηση ξένων γλωσσών και για τη διδασκαλία της ανάγνωσης (Bernstein και Rtischev, 1991 Bernstein και al., 1990 Mostow και al., 1993). Για τέτοια συστήματα η προφορά από τους ομιλητές κειμένων που προέρχονται από υπολογιστή θα αναλυόταν και θα δινόταν ως είσοδος σε ένα πρόγραμμα για τη διδασκαλία ξένων γλωσσών ή ανάγνωσης. Ενώ αυτές ενδέχεται να είναι ευκολότερες εφαρμογές αναγνώρισης ομιλίας σε σχέση με άλλες, επειδή οι λέξεις που αναφέρονται προέρχονται από τον υπολογιστή, το σύστημα αναγνώρισης θα ερχόταν πάλι αντιμέτωπο με κακές προφορές ή βραδείες προφορές απαιτώντας ένα βαθμό ευρωστίας που δεν λαμβάνεται συχνά υπ' όψιν σε άλλες εφαρμογές αναγνώρισης ομιλίας. Ουσιώδης έρευνα θα απαιτηθεί επίσης για την ανάπτυξη και τη δοκιμή νέου εκπαιδευτικού λογισμικού που θα επωφεληθεί της αναγνώρισης και σύνθεσης ομιλίας για τη διδασκαλία ανάγνωσης. Αυτή είναι ίσως μία από τις σημαντικότερες ενδεχόμενες εφαρμογές της τεχνολογίας ομιλίας, επειδή οι κοινωνιολογικές συνέπειες από την άνοδο των επιπέδων αλφαβητισμού σε ευρεία κλίμακα είναι τεράστιες.

[4] Το γενικό αντικείμενο της "βοηθητικής τεχνολογίας" καλύπτεται ευρέως από τον H. Levitt (σε αυτό το τεύχος) και για μία ανασκόπηση της αναγνώρισης ομιλίας για αποκατάσταση μπορείτε να ανατρέξετε στον Bernstein (1988).

Εξοδος Φωνής

Οπως με την είσοδο ομιλίας, οι παράγοντες που ευνοούν την έξοδο φωνής είναι μόνο άτυπα κατανοητοί. Ακριβώς όπως δραστηριότητες με υψηλό βαθμό οπτικής ή χειρωνακτικής δράσης μπορούν να είναι αποτελεσματικότερα πραγματοποιήσιμες χρησιμοποιώντας είσοδο με ομιλία, αντίστοιχα τέτοιες δραστηριότητες μπορούν επίσης να ευνοήσουν έξοδο ομιλούμενου συστήματος. Ενας χρήστης θα μπορούσε να επικεντρωθεί σε μία δραστηριότητα παρά να μετακινήσει το βλέμμα του(της) για να δει την οθόνη του συστήματος. Τυπικά περιβάλλοντα εφαρμογών περιλαμβάνουν το πέταγμα ενός σκάφους, στο οποίο ο πιλότος θα λάμβανε πληροφορίες σχετικά με την κατάσταση των υποσυστημάτων του σκάφους κατά τη διάρκεια κρίσιμων φάσεων της επιχείρησης (π.χ. προσγείωση, μανούβρες υψηλής ταχύτητας) και την οδήγηση ενός αυτοκινήτου στην οποία ο οδηγός θα λάμβανε πληροφορίες πλοήγησης κατά τη διάρκεια της οδήγησης. Αλλοι παράγοντες που θεωρούνται ότι ευνοούν την έξοδο φωνής περιλαμβάνουν απομακρυσμένη πρόσβαση σε υπηρεσίες πληροφοριών μέσω τηλεφώνου, έλλειψη ικανοτήτων ανάγνωσης, σκοτεινά περιβάλλοντα, και την ανάγκη για παγκατευθυντική παρουσίαση πληροφοριών, όπως στην έκδοση προειδοποιήσεων στις καμπίνες πιλότων, σε δωμάτια ελέγχου, εργοστάσια (Simpson και al., 1985 Thomas και al., 1984).

Υπάρχουν πολυάριθμες μελέτες σύνθεσης ομιλίας, αλλά δεν έχει προκύψει καθαρή εικόνα του πότε η επικοινωνία ανθρώπου υπολογιστή με τη χρήση εξόδου ομιλίας είναι αποδοτικότερη ή προτιμητέα. Ψυχολογική έρευνα εξέτασε την κατανοητότητα (intelligibility), φυσικότητα (naturalness), σαφήνεια (comprehensibility) και ικανότητα ανάκλησης (recallability) της συνθετοποιημένης ομιλίας (Luce και al., 1983 Nusbaum και Schwab, 1983 Simpson και al., 1985 Thomas και al., 1984).Η κατανοητότητα και η φυσικότητα είναι ορθογώνιες διαστάσεις με την έννοια ότι η συνθετική ομιλία που είναι παρούσα σε ένα περιβάλλον άλλων ανθρωπίνων φωνών μπορεί να είναι κατανοητή αλλά μη φυσική. Αντιστρόφως, η ανθρώπινη ομιλία σε ένα ενθόρυβο περιβάλλον μπορεί να είναι φυσική αλλά ακατανόητη (Simpson και al., 1985). Πολλοί παράγοντες επηρεάζουν την κατανοητότητα συνθετοποιημένης ομιλίας σε ένα πραγματικό περιβάλλον εφαρμογών όπως το κατώτατο όριο κατανόησης φωνημάτων, ο ρυθμός ομιλίας, το επίπεδο σήματος προς θόρυβο και η παρουσία άλλων ανταγωνιστικών φωνών καθώς επίσης και τα γλωσσολογικά και πραγματικά (pragmatic) συμφραζομένα (Simpson και Navarro 1984, Simpson και al. 1985).

Το κατά πόσο επιθυμούμε την εξόδο φωνής εξαρτάται από το περιβάλλον εφαρμογών. Οι πιλότοι προτιμούν να ακούν προειδοποιήσεις με χρήση συνθετικής παρά ψηφιοποιημένης ομιλίας καθώς η πρώτη είναι ευκολότερα διακρίσιμη από άλλες φωνές όπως η ραδιοφωνική κίνηση (Voorhees και al., 1983). Ωστόσο, σε προσομοιώσεις συστημάτων ελέγχου εναέριας κυκλοφορίας στις οποίες οι πιλότοι θα περίμεναν να αλληλεπιδράσουν με έναν άνθρωπο ψηφιοποιημένη ανθρώπινη φωνή προτιμήθηκε από συνθετοποιημένη φωνή υπολογιστή (Simpson και al., 1985). Οι χρήστες ενδέχεται να προτιμούν να λαμβάνουν πληροφορία οπτικά είτε σε μία ξεχωριστή οθόνη είτε σε μία heads-up οθόνη (Swider, 1987) φυλάσσοντας την έξοδο με ομιλία για κρίσιμα προειδοποιητικά μηνύματα (Simpson και al., 1985). Περαιτέρω έρευνα απαιτείται, ώστε να καθοριστούν αυτοί οι τύποι περιβαλλόντων επεξεργασίας

πληροφορίας, για τους οποίους η έξοδος με ομιλία είναι ωφέλιμη ή προτιμητέα. Επιπλέον, από το να επικεντρωνόμαστε στα οφέλη της εκφώνησης ενός λεγομένου σε σύγκριση με άλλους τρόπους παρουσίασης της ίδιας πληροφορίας, η μελλοντική έρευνα χρειάζεται να αξιολογήσει την επίδοση και τις προτιμήσεις του χρήστη ως μία συνάρτηση του *περιεχομένου* αυτού που μεταδίδεται, ειδικά αν ο υπολογιστής θα καθορίσει αυτό το περιεχόμενο (π.χ. η δημιουργία οδηγιών πλοήγησης για οδηγούς). Τέλος, έρευνα είναι απαραίτητη για την ανάπτυξη αλγορίθμων για τον καθορισμό του κατάλληλου επιτονισμού που θα χρησιμοποιηθεί κατά τη διάρκεια διαλόγου ανθρώπου υπολογιστή με ομιλία.

Περίληψη

Υπάρχουν πολυάριθμες υπάρχουσες εφαρμογές αλληλεπίδρασης ανθρώπου υπολογιστή που βασίζονται σε φωνή και νέες εφαρμογές αναπτύσονται ταχύτατα. Σε πολλές εφαρμογές, για τις οποίες η είσοδος του χρήστη μπορεί να περιοριστεί επαρκώς, για να επιτρέψει αναγνώριση υψηλής ακρίβειας, η είσοδος με φωνή βρέθηκε να οδηγεί σε γρηγορότερη εκτέλεση εργασιών και με λιγότερα σφάλματα από ό,τι η εισαγωγή με πληκτρολόγιο. Δυστυχώς, δεν υπάρχει ακόμη μέθοδος με αρχές που να προβλέπει πότε η είσοδος φωνής θα είναι ο αποδοτικότερος, αποτελεσματικότερος και προτιμότερος τρόπος επικοινωνίας. Παρομοίως καμία περιεκτική ανάλυση δεν έχει προσδιορίσει τις συνθήκες σχετικά με το πότε η φωνή θα είναι η προτιμότερη ή η αποδοτικότερη μορφή εξόδου υπολογιστή, παρ' όλο που ξανά οι δραστηριότητες με απασχολημένα χέρια/οφθαλμούς ενδέχεται να είναι μεταξύ των κύριων υποψηφίων για έξοδο φωνής.

Μία σημαντική περίσταση που ευνοεί την επικοινωνία ανθρώπου υπολογιστή μέσω φωνής είναι όταν ο χρήστης θέλει να αλληλεπιδράσει με τον υπολογιστή σε μία φυσική γλώσσα, όπως είναι η Αγγλική. Η επόμενη ενότητα μελετά μία τέτοια επικοινωνία με ομιλούμενη γλώσσα.

ΣΥΓΚΡΙΣΗ ΟΜΙΛΟΥΜΕΝΗΣ ΓΛΩΣΣΑΣ ΜΕ ΑΛΛΟΥΣ ΤΡΟΠΟΥΣ ΕΠΙΚΟΙΝΩΝΙΑΣ

Ενας χρήστης που θα μιλά σε έναν υπολογιστή ενδέχεται να περιμένει να είναι ικανός να μιλήσει σε μία φυσική γλώσσα, δηλαδή να χρησιμοποιήσει συνήθεις γλωσσολογικές κατασκευές, όπως ονόματα και ρηματικές φράσεις. Αντιστρόφως, αν επιλεγεί η αλληλεπίδραση με φυσική γλώσσα ως ένας τρόπος επικοινωνίας ανθρώπου υπολογιστή, οι χρήστες ενδέχεται να προτιμούν να μιλούν παρά να πληκτρολογούν. Σε κάθε περίπτωση οι χρήστες ενδέχεται να προσδοκούν να μπορούν να συμμετέχουν σε ένα διάλογο, στον οποίο τα λεγόμενα κάθε μέρους θέτουν το πλαίσιο για τη διερμηνεία των επακόλουθων λεγομένων. Αρχικά συζητείται η κατάσταση της ανάπτυξης των συστημάτων ομιλούμενης γλώσσας και στη συνέχεια συγκρίνεται η αλληλεπίδραση με ομιλούμενη γλώσσα με την αλληλεπίδραση με πληκτρολόγηση.

Πρωτότυπα συστημάτων ομιλούμενης γλώσσας

Βρίσκεται σε εξέλιξη έρευνα για την ανάπτυξη των συστημάτων ερωτο-απαντήσεων με ομιλούμενη γλώσσα-συστήματα που επιτρέπουν σε χρήστες να εκφράζουν τις ερωτήσεις τους ελεύθερα και τα οποία κατανοούν αυτές τις

ερωτήσεις και παρέχουν ακριβή απάντηση. Τα υποστηριζόμενα από το ARPA συστήματα πληροφοριών αεροπορικών ταξιδίων, το ATIS (1993) που αναπτύχθηκε από τους Bolt, Beranek και Newman (Kubala και al., 1992), Carnegie-Mellon University (Huang και al., 1993), το Τεχνολογικό Ινστιτούτο της Μασαχουσέτης (Zue και al., 1993), το SRI International (Appelt και Jackson, 1992) και άλλα ιδρύματα, επιτρέπουν σε αρχάριους χρήστες να αποκτήσουν πληροφορίες σε πραγματικό χρόνο από την βάση δεδομένων του Επίσημου Οδηγού Αεροπορικών Γραμμών μέσω της χρήσης ανεξαρτήτων του ομιλητή, συνεχώς διατυπωμένων ερωτήσεων στα Αγγλικά. Τα συστήματα αναγνωρίζουν τις λέξεις στα λεγόμενα του χρήστη, αναλύουν το νόημα των λεγομένων, συχνά παρά τα σφάλματα αναγνώρισης λέξης, ανακτούν πληροφορίες από (ένα υποσύνολο) της βάσης δεδομένων του Επίσημου Οδηγού Αεροπορικών Γραμμών και παράγουν ένα σύνολο απαντήσεων υπό μορφή πίνακα που απαντούν στην ερώτηση. Αυτά τα συστήματα απαντούν με τον σωστό πίνακα πτήσεων σε ποσοστό μεγαλύτερο του 70% των ερωτήσεων που είναι ανεξάρτητες των συμφραζομένων, όπως "Ποιες πτήσεις αναχωρούν από το San Francisco για τη Washington μετά τις 7:45 π.μ;". Πραγματοποιήθηκε ταχεία εξέλιξη στην ανάπτυξη αυτών των συστημάτων με μία 4-πλή μείωση στους ρυθμούς αναγνώρισης βαρύνοντων σφαλμάτων σε μία 20-μηνη περίοδο αναγνώρισης ομιλίας, μία 3.5-πλή μείωση σε μία 30-μηνη περίοδο για κατανόηση φυσικής γλώσσας και μία 2-πλή μείωση σε μία 20-μηνη περίοδο για το συνδυασμό τους ως ένα σύστημα κατανόησης ομιλούμενης γλώσσας. Αλλες κύριες προσπάθειες για την ανάπτυξη συστημάτων διαλόγου με ομιλία πρόκειται να γίνουν στην Ευρώπη (Mariani, 1992 Peckham, 1991) και την Ιαπωνία (Yato et al., 1992).

Το μεγαλύτερο μέρος της τεχνολογίας επεξεργασίας γλώσσας που χρησιμοποιείται για την κατανόηση ομιλούμενης γλώσσας βασίσθηκε σε τεχνικές συστημάτων φυσικής γλώσσας που βασίζονται σε πληκτρολόγιο.[5] Ωστόσο, η είσοδος με ομιλία παρουσιάζει ποιοτικά διαφορετικά προβλήματα για κατανόηση γλώσσας που δεν έχουν ανάλογο στην αλληλεπίδραση με πληκτρολόγιο.

Ομιλούμενη Γλώσσα προς Πληκτρολογημένη Γλώσσα

Μεθοδολογία Ερευνας

Στον απολογισμό των ευρημάτων για τη γλωσσική επικοινωνία που σχετίζονται με την αλληλεπίδραση ανθρώπου υπολογιστή με ομιλία, ορισμένα αποτελέσματα βασίζονται σε αναλύσεις αλληλεπίδρασης ανθρώπου με άνθρωπο, ορισμένα βασίζονται σε αλληλεπίδραση ανθρώπου προσομοιωμένου υπολογιστή και ορισμένα βασίζονται σε αλληλεπίδραση ανθρώπου υπολογιστή. Μελέτες επικοινωνίας ανθρώπου με άνθρωπο μπορούν να προσδιορίσουν τις επικοινωνιακές δυνατότητες που έχουν οι άνθρωποι στις αλληλεπιδράσεις τους με υπολογιστές και μπορούν να δείξουν τι θα μπορούσε να επιτευχθεί, αν οι υπολογιστές ήταν κατάλληλοι συζητητές. Ωστόσο, επειδή αυτό το επίπεδο ανταγωνιστικότητας στη συζήτηση θα είναι μη κατορθωτό για κάποιο χρονικό διάστημα, οι επιστήμονες ανέπτυξαν τεχνικές για προσομοίωση υπολογιστικών συστημάτων που αλληλεπιδρούν μέσω ομιλούμενης γλώσσας (Andry et al., 1990,

[5] Για μία συζήτηση της κατάστασης της έρευνας και της τεχνολογίας επεξεργασίας φυσικής γλώσσας, ανέτρεξε στον Bates (σε αυτό το τεύχος).

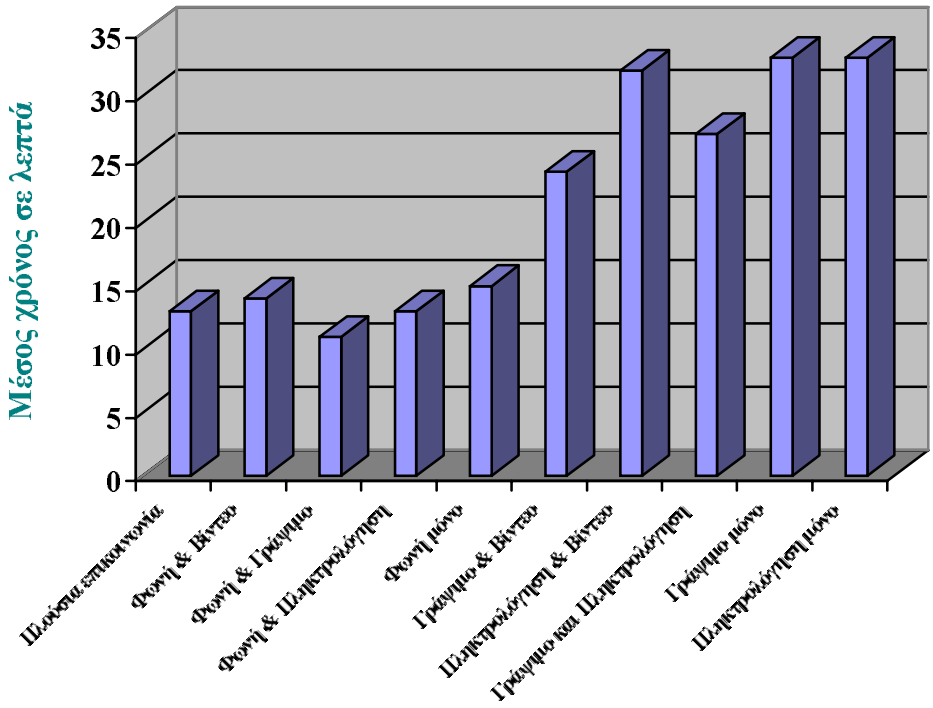
Fraser και Gilbert, 1991 Gould et al., 1983 Guymard και Siroux, 1988 Leiser, 1989 Oviatt et al., 1992, 1993a Pavan και Pelletti, 1990 Price, 1990) χρησιμοποιώντας ένα κρυμμένο ανθρώπινο βοηθό που ανταπαντά στην ομιλούμενη γλώσσα. Με αυτή τη μέθοδο, οι ερευνητές μπορούν να αναλύσουν τη γλώσσα,το διάλογο, την επίδοση και τις προτιμήσεις των ανθρώπων πριν την ανάπτυξη πλήρως λειτουργικών συστημάτων.

Σημαντικά μεθοδολογικά ζητήματα για τέτοιες προσομοιώσεις περιλαμβάνουν παροχή ορθής και γρήγορης απόκρισης και εκπαίδευση του βοηθού προσομοίωσης ώστε να λειτουργεί κατάλληλα. Οι άνθρωποι συμμετέχουν σε γρήγορη αλληλεπίδραση με ομιλία και έχουν απαιτήσεις για ταχύτητα στην αλληλεπίδρασή τους με υπολογιστές. Αργές αλληλεπιδράσεις μπορούν να έχουν ως αποτέλεσμα οι χρήστες να διακόπτουν το σύστημα με επαναλήψεις, ενώ αυτό επεξεργάζεται την προηγούμενη είσοδό τους (VanKatwijk et al., 1979) και εικάζεται ότι μπορούν επίσης να προκαλούν φαινόμενα χαρακτηριστικά της μη αλληλεπιδραστικής ομιλίας (Oviatt και Cohen, 1991a). Μία τεχνική που χρησιμοποιείται για την επιτάχυνση τέτοιων προσομοιώσεων εισόδου φωνής/εξόδου φωνής είναι η χρήση ενός φωνοκωδικοποιητή που μετατρέπει τη φυσικά ομιλούμενη απάντηση του βοηθού σε ένα λεγόμενο που ηχεί μηχανικά (Fraser και Gilbert, 1991 Guyomard και Siroux, 1988). Η ταχύτητα του "συστήματος" ως εκ τούτου καθοδηγείται από την γνώση και το χρόνο αντίδρασης του βοηθού καθώς επίσης και από τη δραστηριότητα που θα επακολουθήσει, αλλά όχι από την αναγνώριση ομιλίας, την κατανόηση γλώσσας και τη σύνθεση ομιλίας. Ωστόσο, επειδή οι άνθρωποι μιλούν διαφορετικά σε έναν υπολογιστή από ό,τι σε ένα άτομο (Fraser και Gilbert, 1991) ακόμη και για προτροπές για απλές απαντήσεις ναι/όχι (Basson, 1992 Basson et al., 1989) ο βοηθός δεν θα έπρεπε να παρέχει μία *τόσο ευφυή* απάντηση, καθώς αυτό μπορεί να αποκαλύψει ότι το "σύστημα" είναι μία προσομοίωση. Μία δεύτερη μέθοδος προσομοίωσης η οποία περιορίζει το βοηθό προσομοίωσης και επίσης υποστηρίζει μία γρήγορη απάντηση, είναι να παρέχονται στο βοηθό ορισμένα προκαθορισμένα πεδία και δομές στην οθόνη που μπορούν να επιλεγούν για απάντηση στο υποκείμενο (Andry et al., 1990 Dahlback et al., 1992 Leiser, 1989 Oviatt et al., 1992). Περαιτέρω έρευνα απαιτείται για την ανάπτυξη μεθοδολογιών προσομοίωσης που μπορούν να μοντελοποιήσουν επακριβώς συστήματα ομιλούμενης γλώσσας, έτσι ώστε πρότυπα αλληλεπίδρασης με τον προσομοιωτή να προβλέπουν πρότυπα αλληλεπίδρασης με το πραγματικό σύστημα ομιλούμενης γλώσσας.

Σύγκριση Τρόπων Επικοινωνίας που Βασίζονται σε Γλώσσα

Σε μία σειρά μελετών αλληλεπιδραστικής επικοινωνίας ανθρώπου με άνθρωπο ο Chapanis και οι συνάδελφοί του (Chapanis et al., 1972, 1977 Kelly και Chapanis, 1977 Michaelis et al., 1977 Ochsman και Chapanis, 1974) συνέκριναν την αποτελεσματικότητα της επικοινωνίας ανθρώπου με άνθρωπο όταν τα υποκείμενα χρησιμοποίησαν οποιουσδήποτε από 10 τρόπους επικοινωνίας (συμπεριλαμβανομένων των τρόπων πρόσωπο με πρόσωπο, φωνής μόνο, συνδεδεμένων τηλέτυπων, αλληλεπιδραστικού γραψίματος). Ο σημαντικότερος καθοριστικός παράγοντας της ταχύτητας επίλυσης προβλημάτων μίας ομάδας βρέθηκε να είναι η παρουσία ενός στοιχείου φωνής. Πιο συγκεκριμένα, μία ποικιλία δραστηριοτήτων πραγματοποιήθηκαν δύο με τρεις φορές ταχύτερα χρησιμοποιώντας έναν τρόπο φωνής σε σχέση με έναν μονίμου

αντιγράφου όπως φαίνεται στο Σχήμα 1. Κατά το ίδιο χρονικό διάστημα, η ταχύτητα οδήγησε σε μία 8-πλή αύξηση στον αριθμό μηνυμάτων και προτάσεων και μία 10-πλή αύξηση στο ρυθμό των λέξεων επικοινωνίας. Αυτά τα αποτελέσματα δηλώνουν το ουσιαστικό ενδεχόμενο να προκύψουν πλεονεκτήματα στην αποτελεσματικότητα από την χρήση επικοινωνίας με ομιλούμενη γλώσσα.



Σχήμα 1 *Η φωνή καθορίζει την αποτελεσματικότητα της δραστηριότητας (από Ochsman και Chapanis, 1974)*

Ερευνα από τους συγγραφείς επιβεβαίωσε αυτά τα αποτελέσματα αποτελεσματικότητας στους διαλόγους ανθρώπου με άνθρωπο κατά την εκτέλεση δραστηριοτήτων συναρμολόγησης εξοπλισμού (Cohen, 1984 Oviatt και Cohen, 1991b) βρίσκοντας ένα τριπλάσιο πλεονέκτημα ταχύτητας για αλληλεπιδραστική τηλεφωνική ομιλία σε σχέση με επικοινωνία με πληκτρολόγιο. Επιπλέον, η δομή τηλεφωνικών διαλόγων διέφερε από αυτή των διαλόγων με πληκτρολόγιο. Μεταξύ των διαφορών που σημειώθηκαν ήταν ότι οι διάλογοι με ομιλία παρουσίασαν περισσότερες υπαινικτικές φράσεις που σηματοδότησαν τη δομή του διαλόγου (όπως “επόμενο”, “εντάξει τώρα”) και οι ομιλητές αλληλεπίδρασαν με έναν πιο “φυσικό” τρόπο σε σχέση με τους χρήστες πληκτρολογίου. Πιο συγκεκριμένα, προκειμένου να εκτελεστεί μια υποδραστηριότητα, οι χρήστες συχνά έκαναν δύο αιτήσεις, μία για προσδιορισμό του αντικειμένου και μία για την ενέργεια, ενώ οι χρήστες πληκτρολογίου τυπικά ενοποιούσαν και τις δύο σε

ένα προστακτικό λεγόμενο. Παρόμοια ευρήματα μίας φυσικής προσέγγισης κατά τη διάρκεια αλληλεπίδρασης με ομιλία προς μία περισσότερο συντακτικά ολοκληρωμένη προσέγγιση για αλληλεπίδραση με πληκτρολόγιο βρέθηκαν σε μία μελέτη προσομοιωμένης αλληλεπίδρασης ανθρώπου υπολογιστή (Zoltan-Ford, 1991). Τέλος, η είσοδος με ομιλία ήταν πιο "έμμεση" από ό,τι η είσοδος με πληκτρολόγιο. Αυτό σημαίνει ότι, αντίθετα με την αλληλεπίδραση με πληκτρολόγιο, τα λεγόμενα δεν μετέφεραν κυριολεκτικά την πρόθεση του ομιλητή, ώστε ο ακροατής να εκτελέσει μία ενέργεια (Cohen, 1984). Μελλοντική έρευνα απαιτείται να γίνει στον τομέα στον οποίο τα αποτελέσματα γενικεύουν στην αλληλεπίδραση ανθρώπου υπολογιστή με ομιλία για συγκρίσιμες δραστηριότητες.

Ενα πλεονέκτημα της εισόδου φωνής είναι η εξάλειψη της πληκτρολόγησης που θα προσέφερε πιθανά αποθέματα παραγωγικότητας σε γραφείο (Baker, 1991 Jelinek, 1985). Σε μία μελέτη προσομοιωμένης "ακουστικής γραφομηχανής" οι Gould et al., (Gould, 1978, 1982 Gould et al., 1983) εξέτασαν πώς αρχάριοι και έμπειροι χρήστες υπαγόρευσης θα χρησιμοποιούσαν μία μηχανή που θα αναγνώριζε και θα πληκτρολογούσε την υπαγόρευση από το χρήστη ενός επαγγελματικού γράμματος σε σύγκριση με την υπαγόρευση και σύνταξη του γράμματος σε έναν άνθρωπο ή το γράψιμο και τη σύνταξη του γράμματος. Το σύστημα ακουστικής γραφομηχανής προσομοιώθηκε και τα υποκείμενα πληροφορήθηκαν ότι στην πραγματικότητα μιλούσαν σε ένα άτομο. Υπήρξε ισχυρισμός ότι οι χρήστες μίας ακουστικής γραφομηχανής ήταν τόσο ικανοποιημένοι με αυτό τον τρόπο επικοινωνίας όσο με τους άλλους και ότι η υπαγόρευση σε μία ακουστική γραφομηχανή θα μπορούσε ενδεχομένως να αποτελέσει ένα τόσο γρήγορο τρόπο σύνθεσης γράμματος όσο η πληκτρολόγηση.Υπάρχει, ωστόσο, αντίθετη απόδειξη, βάσει ενός αριθμού μελετών προσομοίωσης (Murray et al., 1991 Newell et al., 1990) ότι οι επεξεργαστές λέξεων ομιλίας μόνο είναι λιγότερο αποδοτικοί και προτιμητέοι από τις μεθόδους σύνθεσης που βασίζονται σε γράψιμο ή πληκτρολόγηση. Επιπλέον, μία συνδυασμένη μέθοδος χρήσης ομιλίας για είσοδο κειμένου και μίας οθόνης αφής για έλεγχο κέρσορα ήταν αποτελεσματικότερη από ό,τι η ομιλία μόνο, παρ' όλο που είναι λιγότερο αποτελεσματική σε σύγκριση με την σύνθεση και τη σύνταξη με χρήση πληκτρολογίων ή γραψίματος.

Καμία σειρά μελετών δεν εξέτασε με λεπτομέρεια τη γλωσσολογική και διαλεκτική δομή του υπαγορευμένου υλικού που θα μπορούσε να εξηγήσει γιατί η ομιλούμενη σύνθεση και σύνταξη είναι λιγότερο αποτελεσματικές από άλλους τρόπους. Σε μία μελέτη επικοινωνίας ανθρώπου με άνθρωπο βρέθηκε ότι οι άπειροι "διδάκτορες" που παρείχαν πληροφορίες σε έναν ακροατή άνθρωπο παρήγαγαν περισσότερες δομές λόγου από αυτές που θα απαιτούσε η σύνταξη, ώστε να κάνουν αποδεκτό το κείμενο, όπως επαναλήψεις, εκλεπτύνσεις και ανώφελες χρήσεις αναφορικών εκφράσεων σε σχέση με τους χρήστες αλληλεπιδραστικής ομιλίας ή αλληλεπιδραστικού πληκτρολογίου (Oviatt και Cohen, 1991a, 1991b). Ως εκ τούτου, η έλλειψη αλληλεπίδρασης με έναν ακροατή ενδέχεται να συντελέσει σε φτωχά διατυπωμένη είσοδο, δίδοντας ένα μεγαλύτερο βάρος στη φάση προσύνταξης, όπου η είσοδος ομιλίας είναι λιγότερο αποτελεσματική (Newell et al., 1990). Εν περιλήψει, παρ' όλο που οι συσκευές αυτόματης υπαγόρευσης αποτέλεσαν πόλο έλξης ως μία σημαντική ιδέα προϊόντος της τεχνολογίας ομιλίας, τα ενδεχόμενα πλεονεκτήματά τους παραμένουν υπό αμφισβήτηση.

Ο χώρος που μελετά τους τρόπους (modalities) δεν έχει ακόμη εξερευνηθεί συστηματικά. Δεν γνωρίζουμε επακριβώς πώς τα αποτελέσματα μελετών επικοινωνίας ανθρώπου με άνθρωπο μπορούν να προβλέψουν αποτελέσματα για μελέτες αλληλεπιδράσεων ανθρώπου προσομοίωσης ή ανθρώπου υπολογιστή. Επίσης, απαιτείται να διεξαχθούν περισσότερες μελέτες που θα συγκρίνουν τη δομή και το περιεχόμενο της ομιλούμενης γλώσσας ανθρώπου υπολογιστή με την πληκτρολογημένη γλώσσα ανθρώπου υπολογιστή, ώστε να γίνει κατανοητό πώς θα προσαρμοσθεί η τεχνολογία που αναπτύχθηκε για αλληλεπίδραση με πληκτρολόγιο σε συστήματα ομιλούμενης γλώσσας.

Κοινό στοιχείο σε πολλές επιτυχημένες εφαρμογές τεχνολογίας που βασίζεται σε φωνή είναι η έλλειψη κατάλληλης εναλλακτικής λύσης για τη φωνή δοθέντων των εργασιών και των περιβαλλόντων που σχετίζονται με τη χρήση υπολογιστή. Βασικά ερωτήματα παραμένουν, όπως ο προσδιορισμός εφαρμογών στις οποίες η φωνή θα προτιμηθεί, όταν άλλοι τρόποι επικοινωνίας είναι δυνατοί. Ορισμένες μελέτες αναφέρουν μία αναμφισβήτητη προτίμηση για ομιλία σε σύγκριση με άλλους τρόπους (Rudnicky, 1993), ωστόσο άλλες μελέτες κάνουν αναφορά σε αντίθετο συμπέρασμα (Murray et al., 1991 Newell et al/, 1990). Ως εκ τούτου, παρ' όλα τα προαναφερθέντα πιθανά πλεονεκτήματα της αλληλεπίδρασης ανθρώπου υπολογιστή με χρήση φωνής, δεν είναι προφανές γιατί οι άνθρωποι θα ήθελαν να μιλούν στον υπολογιστή τους κατά την εκτέλεση των καθημερινών εργασιών γραφείου. Προκειμένου να εξασφαλιστεί ένα πλαίσιο απάντησης στο ερώτημα αυτό η συζήτηση που ακολουθεί συγκρίνει την διεπαφή χρήστη άμεσης προσπέλασης που κυριαρχεί αυτή τη στιγμή με την πληκτρολογημένη ή ομιλούμενη φυσική γλώσσα.

Σύγκριση Αλληλεπίδρασης Με Φυσική Γλώσσα με Εναλλακτικούς Τρόπους

Υπάρχουν πολυάριθμοι εναλλακτικοί τρόποι αλληλεπίδρασης ανθρώπου υπολογιστή, όπως η χρήση πληκτρολογίων για διαβίβαση κειμένου, η ένδειξη και οι χειρονομίες με συσκευές όπως το ποντίκι, η ψηφιοποιημένη πένα, οι ιχνοσφαίρες και τα ψηφιοποιημένα γάντια. Είναι σημαντικό να κατανοηθεί ποιο ρόλο μπορεί να παίξει στην υποστήριξη αλληλεπίδρασης με άνθρωπο η ομιλία και πιο συγκεκριμένα η ομιλούμενη γλώσσα, ειδικά όταν διατίθενται αυτοί οι άλλοι τρόποι. Για να ξεκινήσουμε αυτή τη συζήτηση χρειάζεται να προσδιορίσουμε τις ιδιότητες των επιτυχημένων διεπαφών. Μία ιδανική διεπαφή θα έπρεπε να είναι:

Απαλλαγμένη σφαλμάτων. Η διεπαφή θα έπρεπε να αποτρέπει το χρήστη απο τη διατύπωση εσφαλμένων εντολών, θα έπρεπε να ελαχιστοποιεί παρερμηνείες της πρόθεσης του χρήστη και θα έπρεπε να προσφέρει απλές μεθόδους για διόρθωση σφαλμάτων.

Διαφανής. Η λειτουργικότητα του συστήματος εφαρμογής θα έπρεπε να είναι προφανής στο χρήστη.

Υψηλού επιπέδου. Ο χρήστης δεν θα έπρεπε να μάθει τις βαθύτερες δομές και γλώσσες του υπολογιστή, αλλά μάλλον θα έπρεπε να μπορεί να δηλώνει απλά τις επιθυμίες του(της) και να αφήνει το σύστημα να χειριστεί τις λεπτομέρειες.

Συνεπής. Στρατηγικές που δουλεύουν για την επίκληση μίας λειτουργίας υπολογιστή θα έπρεπε να εφαρμόζονται και για την επίκληση άλλων.

Εύκολη στην εκμάθηση. Ο χρήστης δεν θα έπρεπε να χρειάζεται τυπική εκπαίδευση, αλλά μάλλον μία σύντομη διαδικασία εξερεύνησης θα αρκούσε για την εκμάθηση του πώς να χρησιμοποιήσει ένα δεδομένο σύστημα.

Εκφραστική. Ο χρήστης θα έπρεπε να μπορεί να εκτελεί εύκολα οποιοδήποτε συνδυασμό δραστηριοτήτων που έχει στο μυαλό του μέσα στα όρια της προτιθέμενης λειτουργικότητας του συστήματος.

Χρησιμοποιώντας αυτό το σύνολο ιδιοτήτων ακολουθεί στη συνέχεια συζήτηση σχετικά με τη χρήση της άμεσης προσπέλασης και των τεχνολογιών φυσικής γλώσσας.

Αμεση Προσπέλαση (Direct Manipulation)

Το παράδειγμα της γραφικής διεπαφής με το χρήστη συνεπάγεται έναν τύπο αλληλεπίδρασης που προσφέρει στο χρήστη μενού, εικονίδια και συσκευές ένδειξης (π.χ., το "ποντίκι" [English et al., 1967]) για την επίκληση εντολών του υπολογιστή καθώς επίσης και πολλαπλά παράθυρα στα οποία εμφανίζεται η έξοδος. Αυτές οι γραφικές διεπαφές χρήστη (GUIs) που έγιναν δημοφιλείς από τον Apple Macintosh και τα Microsoft Windows, χρησιμοποιούν τεχνικές που πρωτοεμφανίστηκαν στο SRI International και στο Xerox's Palo Alto Research Center στα τέλη των δεκαετιών '60 και '70 (Englebart 1973, Kay και Goldberg 1977). Με τα GUIs οι χρήστες εκτελούν ενέργειες επιλέγοντας αντικείμενα και στη συνέχεια επιλέγοντας την επιθυμητή ενέργεια από ένα μενού, παρά πληκτρολογώντας εντολές.

Επιπρόσθετα, με πολλά GUIs ένας χρήστης μπορεί απευθείας να χειριστεί γραφικά αντικείμενα, ώστε να εκτελεστούν ενέργειες στα αντικείμενα που αντιπροσωπεύουν. Για παράδειγμα, ένας χρήστης μπορεί να αντιγράψει ένα αρχείο από ένα δίσκο σε έναν άλλο επιλέγοντας το εικονίδιό του με τη συσκευή ένδειξης και "σύροντάς" το από τη λίστα των αρχείων του πρώτου δίσκου στο δεύτερο. Αλλες ενέργειες άμεσης προσπέλασης περιλαμβάνουν τη χρήση μίας "μπάρας ολίσθησης" για την εμφάνιση διαφορετικών τμημάτων ενός αρχείου και το σύρσιμο του εικονιδίου ενός αρχείου στην κορυφή του εικονιδίου "σκουπίδια", προκειμένου να διαγραφεί. Εκτός από το ποντίκι υπάρχουν πολυάριθμες συσκευές ένδειξης όπως οι ιχνο-σφαίρες (trackballs) και τα χειριστήρια (joysticks) και ορισμένες συσκευές προσφέρουν πολλαπλές δυνατότητες όπως η χρήση πεννών για ένδειξη, χειρονομίες και γράψιμο. Τέλος, για να γενικεύσουμε σε μία διαφορετική διάσταση, οι χρήστες μπορούν τώρα να χειριστούν άμεσα εικονικούς κόσμους χρησιμοποιώντας γάντια που κινούνται από τον υπολογιστή και στολές σώματος (Fisher 1990, Kreuger 1977, Rheingold 1991) επιτρέποντας σε έξυπνες κινήσεις του σώματος να επηρεάσουν το εικονικό περιβάλλον.

Ισχυρά σημεία. Πολλοί συγγραφείς προσδιόρισαν πλεονεκτήματα καλά σχεδιασμένων διεπαφών άμεσης προσπέλασης που βασίζονται σε γραφικά (DMIs, Direct Manipulation Interfaces) (π.χ., Hutchins et al., 1986 Shneiderman, 1983) ισχυριζόμενοι ότι

- Τα DIMs που βασίζονται σε γνωστές αλληγορίες, είναι διαισθητικά και εύκολα στη χρήση.

- Τα GUIs μπορούν να έχουν μία συνεπή "όψη και αίσθηση" που καθιστά ικανούς τους χρήστες ενός προγράμματος να μάθουν γρήγορα κάποιο άλλο πρόγραμμα.

- Τα μενού καθιστούν τις διαθέσιμες επιλογές προφανείς, επομένως περιορίζουν τα λάθη του χρήστη κατά τη διατύπωση εντολών και τον προσδιορισμό των παραμέτρων τους.

- Τα GUIs μπορούν να προστατεύσουν το χρήστη από την υποχρέωση να μάθει τις βαθύτερες έννοιες και λεπτομέρειες του υπολογιστή.

Δεν είναι υπερβολή να λεχθεί ότι οι γραφικές διεπαφές χρήστη που υποστηρίζουν αλληλεπίδραση άμεσης προσπέλασης, γνώρισαν τόση επιτυχία που καμία σοβαρή εταιρεία υπολογιστών δεν θα προσπαθούσε να πουλήσει ένα μηχάνημα χωρίς GUI.

Αδυναμίες. Τα DIMs δεν επαρκούν για όλες τις ανάγκες. Μία καθαρή και εκφραστική αδυναμία είναι η έλλειψη διαθέσιμων μέσων για τον προσδιορισμό οντοτήτων. Επιτρέποντας απλώς στους χρήστες να επιλέξουν οντότητες που εμφανίστηκαν εκείνη τη στιγμή, τους παρέχει μικρή υποστήριξη για τον προσδιορισμό αντικειμένων που δεν βρίσκονται στην οθόνη (όπως ένα όνομα αρχείου σε μία λίστα 200 αρχείων), για τον καθορισμό χρονικών σχέσεων που δηλώνουν μελλοντικά ή παρελθοντικά γεγονότα, για τον προσδιορισμό και τη λειτουργία μεγάλων συνόλων οντοτήτων και για τη χρήση του γενικού πλαισίου της αλληλεπίδρασης. Κατά κύριο λόγο, όσοι ασχολούνται με την ανάπτυξη των GUIs εξασφάλισαν απλές ρουτίνες ταύτισης συμβολοσειρών που βρίσκουν αντικείμενα βασει της ακριβούς ή μερικής ταύτισης των ονομάτων τους. Αυτό που λείπει είναι ένας τρόπος που θα εξασφαλίζει στους χρήστες την περιγραφή οντοτήτων χρησιμοποιώντας κάποια μορφή γλωσσολογικής έκφρασης ώστε να δηλώνουν ή να διακρίνουν ένα προσωπικό αντικείμενο, ένα σύνολο αντικειμένων, μία χρονική περίοδο και ούτω καθεξής.[6] Το λιγότερο, μία γλώσσα περιγραφής θα έπρεπε να περιλαμβάνει κάποιο τρόπο για να βρεθούν οντότητες που διαθέτουν ένα δεδομένο σύνολο ιδιοτήτων, για να βρεθεί ποιες ιδιότητες έχουν ενδιαφέρον και ποιες όχι, για να βρεθεί πόσες οντότητες είναι επιθυμητές, για να τεθούν χρονικοί περιορισμοί σε ενέργειες που συνεπάγονται αυτές οι ιδιότητες και ούτω καθεξής. Επιπλέον, ένα χρήσιμο χαρακτηριστικό μίας γλώσσας περιγραφής είναι η ικανότητα της επαναχρησιμοποίησης αναφορών προηγούμενων περιγραφών. Ορισμένες από αυτές τις δυνατότητες συναντώνται σε τυπικές γλώσσες ερωτήσεων, ενώ όλες συναντώνται σε φυσικές γλώσσες.

Παρ' όλο που οι διεπαφές άμεσης προσπέλασης συχνά δεν είναι υψηλού επιπέδου προστατεύουν το χρήστη από λεπτομέρειες της εφαρμογής,. Για παράδειγμα, ένας κοινός τρόπος αίτησης για πληροφορίες από μία σχεσιακή βάση δεδομένων είναι η επιλογή ορισμένων πεδίων από τους πίνακες που κάποιος θέλει να δει. Προκειμένου να γίνει αυτό σωστά, ο χρήστης απαιτείται να μάθει τη δομή της βάσης δεδομένων-για παράδειγμα, να μάθει ότι τα δεδομένα απεικονίζονται σε έναν ή περισσότερους πίνακες που αποτελούνται από πολυάριθμα πεδία των οποίων οι σημασίες ενδέχεται να μην είναι προφανείς. Κατά αυτό τον τρόπο η βαθύτερη υλοποίηση υπό μορφή πίνακα έγινε η

[6] Φυσικά η εξάλειψη περιγραφών ήταν μία συνειδητή απόφασης σχεδίασης από τους δημιουργούς των GUIs.

αλληγορία της διεπαφής χρήστη. Μία εναλλακτική λύση είναι να αναπτυχθούν συστήματα και διεπαφές που μεταφράζουν τον τρόπο σκέψης του χρήστη σχετικά με το πρόβλημα και την εφαρμογή. Κάνοντας αυτό, ο χρήστης ενδεχομένως να μπορουσε να ανακτήσει ανεπιφύλακτα πληροφορία, χωρίς να απαιτείται να γνωρίζει τι κρατείται στη βάση πόσο μάλλον να μάθει τη δομή αυτής της βάσης. Αναλαμβάνοντας μία τέτοια αλληλεπίδραση υψηλού επιπέδου οι χρήστες ενδέχεται να μπορούν να συνδυάσουν την πρόσβαση σε πληροφορίες με άλλες εφαρμογές επεξεργασίας πληροφορίας, όπως είναι η εκτέλεση μιας προσομοίωσης, χωρίς αρχικά να πρέπει να σκεφτούν την ανάκτηση της βάσης και στη συνέχεια μεταβαίνοντας διανοητικά μεταξύ "εφαρμογών" να σκεφτούν την προσομοίωση.

Οταν είναι δυνατές πολυάριθμες εντολές, τα GUIs συνήθως παρουσιάζουν μία ιεραρχική δομή μενού. Καθώς ο αριθμός των εντολών αυξάνει, ο συνήθης χρήστης μπορεί να έχει δυσκολία να θυμάται σε ποιο μενού αυτές βρίσκονται. Ωστόσο, ο χρήστης που γνωρίζει που βρίσκεται η επιθυμητή ενέργεια σε μία ευρεία ιεραρχία ενεργειών, ακόμη χρειάζεται να διαπεράσει την ιεραρχία. Οι σχεδιαστές λογισμικού προσπάθησαν να ξεπεράσουν αυτό το πρόβλημα παρέχοντας διαφορετικά σύνολα μενού σε χρήστες διαφορετικών επιπέδων εμπειρίας, προεπιλέγοντας το πιο πρόσφατα χρησιμοποιούμενο στοιχείο ενός μενού και παρέχοντας άμεσες συνδέσεις σε κοινά χρησιμοποιούμενες εντολές μέσω των συνδυασμών ειδικών πλήκτρων. Ωστόσο, κάνοντας το τελευταίο, τα GUIs δανείζονται στοιχεία από τις διεπαφές που βασίζονται σε πληκτρολόγιο και τις γλώσσες εντολών.

Επειδή ο άμεση προσπέλαση δίνει έμφαση στην γρήγορη γραφική απόκριση σε ενέργειες (Shneiderman, 1983), ο χρόνος της ενέργειας του συστήματος στα DIMs είναι κυριολεκτικά ο χρόνος κατά τον οποίο έγινε επίκληση της ενέργειας. Παρ' όλο που ορισμένα συστήματα ενδέχεται να καθυστερήσουν ενέργειες έως συγκεκριμένους μελλοντικούς χρόνους, τα DIMs και τα GUIs προσφέρουν μικρή υποστήριξη σε χρήστες που θέλουν να εκτελέσουν ενέργειες σε έναν άγνωστο αλλά περιγράψιμο μελλοντικό χρόνο.

Τέλος, τα DIMs βασίζονται σημαντικά στα χέρια και στους οφθαλμούς του χρήστη. Δοθείσης της προηγούμενης συζήτησής μας ορισμένες δραστηριότητες θα εκτελούνταν καλύτερα με ομιλία. Μέχρις εδώ, ωστόσο, έχει διεξαχθεί ελάχιστη έρευνα που να συγκρίνει τα GUIs με την ομιλία. Πρώιμα εργαστηριακά αποτελέσματα ενός συστήματος σχεδίασης VLSI άμεσης προσπέλασης που υποστηρίζει αναγνώριση ομιλίας που εξαρτάται από τον ομιλητή, δείχνουν ότι οι χρήστες είναι τόσο ταχείς στην έκφραση εντολών μονής λέξης όσο κατά την επίκληση των ίδιων εντολών με κλικ του πλήκτρου του ποντικιού ή πληκτρολογώντας μία εντολή σύντμησης μονού γράμματος (Martin, 1989). Αυτό σημαίνει ότι δεν παρατηρήθηκε καμία απώλεια αποτελεσματικότητας που να οφείλεται στη χρήση ομιλίας για απλές δραστηριότητες, στις οποίες τα DIMs τυπικά υπερέχουν. Ας σημειωθεί ότι γενικά παρατηρείται ένα διπλάσιο έως τριπλάσιο πλεονέκτημα ταχύτητας, όταν η ομιλία συγκρίνεται με την πληκτρολόγηση πλήρων λέξεων (Chapanis et al., 1977 Oviatt και Cohen, 1991b). Σε μία πρόσφατη μελέτη αλληλεπίδρασης ανθρώπου υπολογιστή για την ανάκτηση πληροφορίας από μία μικρή βάση δεδομένων (240 είσοδοι) βρέθηκε ότι η ομιλία προτιμήθηκε με σημαντική διαφορά σε σχέση με τη χρήση ολίσθησης άμεσης προσπέλασης, παρ' όλο που ο συνολικός χρόνος ολοκλήρωσης της δραστηριότητας με φωνή ήταν μεγαλύτερος (Rudnicky, 1993).

Αυτή η μελέτη υποδεικνύει ότι για απλές δραστηριότητες, απαλλαγμένες του κινδύνου, η προτίμηση του χρήστη ενδέχεται να βασισθεί στο χρόνο εισόδου μάλλον παρά στους συνολικούς χρόνους ολοκλήρωσης της δραστηριότητας ή στη συνολική ακρίβεια πραγματοποίησης της δραστηριότητας.

Αλληλεπίδραση με Φυσική Γλώσσα

Ισχυρά σημεία. Η φυσική γλώσσα είναι η παραδειγματική περίπτωση ενός εκφραστικού τρόπου επικοινωνίας. Ενα βασικό πλεονέκτημα είναι η χρήση ψυχολογικά έντονων και μνημονικών περιγραφών. Τα Αγγλικά ή οποιαδήποτε άλλη φυσική γλώσσα παρέχουν ένα σύνολο ωραία επεξεργασμένων περιγραφικών εργαλείων, όπως η χρήση ονομαστικών φράσεων για προσδιορισμό αντικειμένων, ρηματικών φράσεων για προσδιορισμό γεγονότων, χρόνου και έγκλισης ρημάτων για περιγραφή χρονικών περιόδων. Από την ίδια τη φύση των προτάσεων αυτές οι δυνατότητες αναπτύσσονται ταυτόχρονα, καθώς οι προτάσεις πρέπει να έχουν κάποιο νόημα και περιγράφουν συχνότερα γεγονότα τοποθετημένα στο χρόνο.

Σε συνδυασμό με αυτή τη δυνατότητα της περιγραφής οντοτήτων, οι φυσικές γλώσσες προσφέρουν την ικανότητα αποφυγής εκτεταμένων επαναπεριγραφών μέσω της χρήσης αντωνυμιών και άλλων “αναφορικών” εκφράσεων. Τέτοιες εκφράσεις συνήθως προορίζονται να δηλώσουν τις ίδιες οντότητες που δηλώνουν και οι προγενέστερες και ο αποδέκτης οφείλει να συμπεράνει τη σύνδεση. Ετσι η χρήση αναφοράς παρέχει ένα οικονομικό πλεονέκτημα στον ομιλητή εις βάρος της υποχρέωσης του ακροατή να εξάγει συμπεράσματα.

Επιπλέον οι εντολές φυσικής γλώσσας μπορούν να προσφέρουν μία άμεση οδό για την επίκληση μίας ενέργειας ή την πραγματοποίηση επιλογών που θα ενσωματώνονταν βαθιά στο ιεραρχικό μενού ενεργειών ή θα απαιτούσαν πολλαπλές επιλογές μενού όπως η γραμματοσειρά, ο τύπος και το μέγεθος σε ένα πρόγραμμα επεξεργασίας λέξεων. Χρησιμοποιώντας τέτοιες εντολές ο χρήστης θα απέφευγε την υποχρέωση επιλογής πολυάριθμων εισόδων μενού για την απομόνωση της επιθυμητής ενέργειας. Επιπλέον, επειδή η επίκληση μίας ενέργειας ενδέχεται να συνεπάγεται μία περιγραφή των παραμέτρων της η ανάκτηση πληροφοριών σχετίζεται στενά με την επίκληση ενεργειών.

Ιδανικά, τα συστήματα φυσικής γλώσσας θα έπρεπε να απαιτούν μόνο ένα ελάχιστο εκπαίδευσης στον τομέα που καλύπτεται από το σύστημα-στόχο. Χρησιμοποιώντας φυσική γλώσσα οι άνθρωποι θα έπρεπε να μπορούσαν να αλληλεπιδρούν αμέσως με ένα σύστημα γνωστού περιεχομένου και λειτουργικότητας, χωρίς να πρέπει να μάθουν τις βαθύτερες δομές του υπολογιστή. Το σύστημα θα έπρεπε να διαθέτει επαρκές λεξιλόγιο καθώς επίσης γλωσσολογικές, σημασιολογικές και διαλογικές ικανότητες για να υποστηρίξει την επίλυση αλληλεπιδραστικών προβλημάτων από συνήθεις χρήστες, δηλαδή χρήστες που χρησιμοποιούν το σύστημα σπάνια. Για παράδειγμα, στο παρόν στάδιο ανάπτυξης πολλοί χρήστες μπορούν να επιλύσουν με επιτυχία προβλήματα λανθασμένης σχεδίασης με ένα από τα συστήματα ATIS (Advanced Research Projects Agency, 1993) μέσα σε λίγα λεπτά εισαγωγής στο σύστημα και στην κάλυψή του. Προκειμένου να αναπτυχθούν συστήματα με αυτό το επίπεδο ευρωστίας, το σύστημα πρέπει να είναι εκπαιδευμένο και δοκιμασμένο σε ένα

σημαντικό σύνολο δεδομένων που αντιπροσωπεύουν είσοδο από ένα ευρύ φάσμα χρηστών.[7] Αυτή τη στιγμή είναι άγνωστο το επίπεδο εκπαίδευσης που απαιτείται για να επιτευχθεί ένα δεδομένο επίπεδο απόδοσης με τη χρήση αυτών των συστημάτων.

Αδυναμίες. Γενικά, διαφαίνονται διάφορα μειονεκτήματα, όταν η φυσική γλώσσα ενσωματώνεται σε μία διεπαφή. Καθαρά συστήματα φυσικής γλώσσας πάσχουν από αδιαφανή γλωσσολογική και εννοιολογική κάλυψη-ο χρήστης γνωρίζει ότι το σύστημα δεν μπορεί να ερμηνεύσει κάθε λεγόμενο, αλλά δεν γνωρίζει ακριβώς τι μπορεί να ερμηνεύσει (Hendrix και Walter, 1987 Murray et al., 1991 Small και Weldon, 1983 Turner et al., 1984). Συχνά πρέπει να γίνουν πολλαπλές προσπάθειες, για να τεθεί μία ερώτηση ή μία εντολή, την οποία το σύστημα μπορεί να ερμηνεύσει σωστά. Ετσι, τέτοια συστήματα ενδέχεται να είναι επιρρεπή σε σφάλματα και,σύμφωνα με τον ισχυρισμό ορισμένων, (Shneiderman, 1980) να οδηγήσουν σε αποτυχία και απογοήτευση. Ενας τρόπος για να ξεπεραστούν αυτά τα προβλήματα προτάθηκε σε ένα σύστημα επεξεργασίας γλώσσας που βασίζεται σε μενού, στο οποίο οι χρήστες συνέθεταν ερωτήσεις σε μία ψευδοφυσική γλώσσα, επιλέγοντας φράσεις από ένα μενού (Tennant et al., 1983). Παρ' όλο που οι προκύπτουσες ερωτήσεις είναι εγγυημένα αναλυτέες, όταν υπάρχει ένας μεγάλος αριθμός επιλογών μενού που μπορούν να γίνουν, η διαδιακασία ερωτήσεων γίνεται δύσκαμπτη.

Πολλές προτάσεις φυσικής γλώσσας είναι ασαφείς και οι αναλυτές συχνά βρίσκουν περισσότερες ασάφειες από τους ανθρώπους. Ετσι, ένα σύστημα φυσικής γλώσσας συχνά ασχολείται με κάποια μορφή διευκρίνισης ή επιβεβαίωσης υποδιαλόγου, για να καθοριστεί εάν η διερμηνεία του είναι η προτιθέμενη. Η τρέχουσα έρευνα προσπαθεί να χειριστεί την ασάφεια της εισόδου φυσικής γλώσσας αναπτύσοντας αλγορίθμους περιγραφής πιθανοτήτων, για τους οποίους οι αναλύσεις θα κατατάσσονταν σύμφωνα με την πιθανότητα να συμβούν στον δεδομένο τομέα (ανέτρεξε στον Marcus, σε αυτό το τεύχος). Επίσης, η έρευνα αρχίζει να διερευνά το ενδεχόμενο χρήσης προσωδίας για την επιλογή μεταξύ ασαφών αναλύσεων (Bear και Price, 1990 Price et al., 1991). Μία τρίτη κατεύθυνση έρευνας συνεπάγεται την ελαχιστοποίηση των ασαφειών μέσω τεχνικών πολυτροπικών διεπαφών για την καθοδήγηση της γλώσσας του χρήστη (Cohen, 1991b Cohen et al., 1989 Oviatt et al., 1993).

Αλλο μειονέκτημα της αλληλεπίδρασης με φυσική γλώσσα είναι ότι οι αλγόριθμοι που ξεδιαλύνουν τι αναφέρεται δεν παρέχουν πάντα τη σωστή απάντηση, εν μέρει επειδή τα συστήματα διαθέτουν υποανεπτυγμένες γνωστικές βάσεις και έν μέρει επειδή το σύστημα έχει μικρή πρόσβαση στην κατάσταση ομιλίας, παρ' όλο που τα προγενέστερα λεγόμενα και οι γραφικές παρουσιάσεις του συστήματος δημιούργησαν αυτή την κατάσταση ομιλίας. Για να περιπλακούν τα πράγματα, τα συστήματα αυτή τη στιγμή έχουν δυσκολία στο να ακολουθήσουν τις μεταβολές των συμφραζομένων σύμφωνα με το διάλογο. Αυτοί οι περιορισμοί γνώσης λέξεων και συμφραζομένων υποτιμούν την αναζήτηση αναφορών και παρέχουν έναν άλλο λόγο, για τον οποίο τα συστήματα

[7] Η προσπάθεια ATIS απαίτησε τη συλλογή και το σχολιασμό των λεγομένων περισσοτέρων των 10.000 χρηστών, ορισμένα από τα οποία χρησιμοποιήθηκαν για την ανάπτυξη του συστήματος και τα υπόλοιπα για δοκιμή κατά τη διάρκεια συγκριτικών αξιολογήσεων που διεξήχθησαν από το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας.

φυσικών γλωσσών συνήθως σχεδιάζονται για την επιβεβαίωση των διερμηνειών τους.

Δεν είναι ξεκάθαρο πού η αλληλεπίδραση με πληκτρολογημένη φυσική γλώσσα θα είναι ένας τρόπος επιλογής. Μελέτες που συγκρίνουν την απάντηση ερωτήσεων σε βάση δεδομένων με πληκτρολογημένη φυσική γλώσσα με ερώτηση σε βάση δεδομένων χρησιμοποιώντας μία ψεύτικη γλώσσα ερωτήσεων (π.χ., SQL) (Chamberlin και Boyce, 1974) έχουν δεδομένα διφορούμενα αποτελέσματα, με ορισμένες μελέτες να συμπεραίνουν ότι η αλληλεπίδραση με φυσική γλώσσα προσφέρει γρηγορότερη και συμπαγέστερη διατύπωση ερωτήσεων (Jarke et al., 1985) ενώ άλλες συμπεραίνουν ότι η μέθοδος με ερωτήσεις σε βάση δεδομένων με χρήση SQL είναι ακριβέστερη και ευκολότερη στην εκμάθηση (Jarke et al., 1985 Shneidrman, 1980a). Ωστόσο, αυτές οι μελέτες παρουσιάζουν ατέλειες από τη χρήση προτοτύπων συστημάτων φυσικής γλώσσας παρά εμπορικών συστημάτων. Οταν ένα ποιοτικό σύστημα ανάκτησης από βάση δεδομένων με φυσική γλώσσα(INTELLECT Harris, 1977) μελετήθηκε στο χώρο, οι χρήστες ανέφεραν οφέλη αποδοτικότητας και μία καθαρή προτίμηση για αλληλεπίδραση φυσικής γλώσσας σε σύγκριση με μία προηγούμενη μέθοδο αλληλεπίδρασης με βάση δεδομένων με γλώσσα ερωτήσεων (Capindale και Crawford, 1990). Αλλη δυσκολία σε πολλές εργαστηριακές μελέτες είναι η έλλειψη κατάλληλων ελέγχων στην εκπαίδευση των υποκειμένων. Σε μία μελέτη που συγκρίνει τη χρησιμότητα της χρήσης φυσικής γλώσσας προς μία γλώσσα ερωτήσεων για πρόσβαση σε βάση δεδομένων (Shneiderman, 1980b) δε δόθηκε στους χρήστες στην κατάσταση φυσικής γλώσσας στην πραγματικότητα καμία εκπαίδευση στο περιεχόμενο της βάσης δεδομένων, με τη λογική ότι τα συστήματα φυσικής γλώσσας δεν θα απαιτούσαν εκπαίδευση ενώ οι χρήστες SQL εκπαιδεύτηκαν στα ονόματα αρχείων και πεδίων αυτής της βάσης δεδομένων. Δεν αποτελεί έκπληξη, υπό αυτές τις συνθήκες, ότι οι χρήστες φυσικών γλωσσών έκαναν πιο "προχωρημένα" σφάλματα με την έννοια της αίτησης πληροφορίας που δεν υπήρχε στη βάση.

Περίληψη: Περιπτώσεις που ευνοούν Αλληλεπίδραση με Μηχανές με Ομιλούμενη Γλώσσα

Θεωρητικά η άμεση προσπέλαση θα έπρεπε να ήταν ωφέλιμη όταν τα αντικείμενα προς χειρισμό είναι στην οθόνη, η ταυτότητά τους είναι γνωστή και δεν υπάρχουν πολλά αντικείμενα προς επιλογή. Επιπρόσθετα, οι γραφικές διεπαφές χρήστη περιορίζουν τις επιλογές των χρηστών, εμποδίζοντάς τους να κάνουν λάθη κατά τη διατύπωση εντολών. Η αλληλεπίδραση με υπολογιστές με φυσική γλώσσα προσφέρει πιθανά πλεονεκτήματα, όταν οι χρήστες χρειάζεται να προσδιορίσουν αντικείμενα, ενέργειες και γεγονότα από σύνολα πολύ ευρεία, για να εμφανιστούν και/ή να εξεταστούν ξεχωριστά και, όταν οι χρήστες χρειάζεται να επικαλεστούν ενέργειες σε μελλοντικούς χρόνους που πρέπει να περιγραφούν. Επιπλέον, η φυσική γλώσσα επιτρέπει σε χρήστες να σκεφθούν τα προβλήματά τους και να εκφράσουν τους στόχους τους με τους δικούς τους όρους παρά με αυτούς του υπολογιστή. Ωστόσο, επιτρέποντας στους χρήστες να το κάνουν αυτό τα συστήματα χρειάζεται να διαθέτουν επαρκείς συλλογιστικές και διερμηνευτικές ικανότητες για την επίλυση προβλημάτων κατά τη μετάφραση μεταξύ του μοντέλου αντίληψης του χρήστη και της εφαρμογής του συστήματος.

Συνδυάζοντας τα εμπειρικά αποτελέσματα περιπτώσεων που ευνοούν την αλληλεπίδραση που βασίζεται σε φωνή με την προαναφερθείσα ανάλυση των

αλληλεπιδράσεων, για τις οποίες η φυσική γλώσσα ενδέχεται να είναι καταλληλότερη, φαίνεται ότι εφαρμογές που απαιτούν ταχεία είσοδο από το χρήστη για πολύπλοκες περιγραφές θα ευνοήσουν την επικοινωνία με ομιλούμενη φυσική γλώσσα. Επιπλέον, αυτή η προτίμηση είναι πιθανόν να είναι ισχυρότερη, όταν είναι δυνατή μία ελάχιστη εκπαίδευση πάνω στις βαθύτερες δομές του υπολογιστή. Παραδείγματα αυτής της περιοχής εφαρμογών είναι η άσκηση ερωτήσεων σε μία βάση δεδομένων ή η δημιουργία κανόνων ενέργειας (π.χ., "Αν αργοπορήσω σε μία συνάντηση ειδοποιήστε τους συμμετέχοντες σε αυτήν"). Λόγω του πρόσφατου των χρησιμοποιήσιμων συστημάτων ομιλούμενης γλώσσας υπάρχουν ελάχιστες μελέτες που συγκρίνουν την αλληλεπίδραση με ομιλούμενη γλώσσα με τον άμεσο χειρισμό για την πραγματοποίηση πραγματικών δραστηριοτήτων.

Μέχρις αυτού του σημείου αντιπαραβάλλαμε την αλληλεπίδραση με ομιλία με άλλους τρόπους. Αξίζει να σημειωθεί ότι αυτοί οι τρόποι έχουν συμπληρωματικά πλεονεκτήματα και μειονεκτήματα, τα οποία επηρεάζουν την ανάπτυξη πολυτροπικών διεπαφών που αντισταθμίζουν τις αδυναμίες μίας τεχνολογίας διεπαφής σε σχέση με τα ισχυρά σημεία κάποιας άλλης (Cohen, 1991 Cohen et al., 1989). (Ανατρέξτε στην παράγραφο με τίτλο "Πολύτροπα Συστήματα.")

ΑΝΘΡΩΠΙΝΟΙ ΠΑΡΑΓΟΝΤΕΣ ΕΜΠΟΔΙΑ ΣΕ ΣΥΣΤΗΜΑΤΑ ΟΜΙΛΟΥΜΕΝΗΣ ΓΛΩΣΣΑΣ

Παρ' όλο που υπάρχουν πολυάριθμες τεχνικές προκλήσεις στην κατασκευή συστημάτων ομιλούμενης γλώσσας, πολλές από τις οποίες εμπεριέχονται με λεπτομέρεις σε αυτό το τεύχος, απαιτείται ιδιαίτερα η γνώση της διεπαφής και των ανθρωπίνων παραγόντων για αυτά τα συστήματα. Στη συνέχεια παρατίθενται απαιτούμενες πληροφορίες σχετικά με αυθόρμητη ομιλία, ομιλούμενη φυσική γλώσσα και αλληλεπίδραση με ομιλία.

Αυθόρμητη Ομιλία

Οταν ένα λεγόμενο *εκφωνείται αυθόρμητα* μπορεί κάλλιστα να περιέχει λάθος αρχές, δισταγμούς, πλήρεις παύσεις, μεταβάσεις, αποσπάσματα και άλλους τύπους τεχνικά "μη γραμματικών" λεγομένων. Αυτά τα φαινόμενα αναστατώνουν τους αναγνωριστές ομιλίας και τους αναλυτές φυσικής γλώσσας και πρέπει να ανιχνευθούν και να διορθωθούν πριν οι τεχνικές που βασίζονται στην παρούσα τεχνολογία αναπτυχθούν σταθερά. Η τρέχουσα έρευνα ξεκίνησε την διερεύνηση τεχνικών για την ανίχνευση και το χειρισμό δυσχερειών λόγου στην αλληλεπίδραση ανθρώπου υπολογιστή με ομιλία(Bear et al., 1992 Hindle, 1983 Nakatani και Hirschberg, 1993) και έχουν αναπτυχθεί τεχνικές εύρωστης επεξεργασίας που καθιστούν τις ρουτίνες ανάλυσης γλώσσας ικανές να ανακτήσουν το νόημα ενός λεγομένου παρά τα σφάλματα αναγνώρισης (Dowding et al., 1993 Huang et al., 1993 Jackson et al., 1991 Stallard και Bobrow, 1992).

Ο προσδιορισμός διαφορετικών τύπων ομιλούμενης γλώσσας ανθρώπου με άνθρωπο και ανθρώπου υπολογιστή αποκάλυψε ότι ο ανθρώπινος ρυθμός αυθόρμητων δυσχερειών λόγου και αυτοδιορθώσεων είναι ουσιαστικά χαμηλότερος όταν μιλά σε ένα σύστημα παρά σε άλλο άνθρωπο (Oviatt, 1993). Μία δυνατή σχέση πρόβλεψης, επίσης, αποδείχθηκε μεταξύ του ρυθμού

ομιλούμενων δυσχερειών λόγου και του μήκους των λεγομένων (Oviatt, 1993). Από το να πρέπει να επιλύσει δυσχέρειες λόγου, η έρευνα διεπαφής αποκάλυψε ότι οι τεχνικές που βασίζονται σε φόρμα μπορούν να μειώσουν σε ποσοστό μέχρι του 70% όλες τις δυσχέρειες λόγου που συμβαίνουν κατά τη διάρκεια αλληλεπίδρασης ανθρώπου υπολογιστή (Oviatt, 1993). Εν συντομία, η έρευνα προτείνει ότι ορισμένοι δύσκολοι τύποι εισόδου, όπως οι δυσχέρειες λόγου, μπορούν να αποφευχθούν, γενικά, μέσω στρατηγικής σχεδίασης της διεπαφής.

Φυσική Γλώσσα

Γενικά, επειδή η επικοινωνία ανθρώπου υπολογιστή με ομιλούμενη γλώσσα συνεπάγεται ότι το σύστημα κατανοεί μία φυσική γλώσσα αλλά όχι τη συνολική γλώσσα, οι χρήστες θα χρησιμοποιούν κατασκευές εκτός της κάλυψης του συστήματος. Ωστόσο, προσδοκάται ότι, δοθέντων επαρκών δεδομένων, στα οποία θα βασισθεί η ανάπτυξη γραμματικών και φορμών, η πιθανότητα ένας συνεργάσιμος χρήστης να δημιουργήσει λεγόμενα εκτός της κάλυψης του συστήματος είναι μικρή. Ακόμη δεν είναι αυτή τη στιγμή γνωστό:

- πώς να επιλέξει κανείς σχετικά "κλειστούς" τομείς, των οποίων το λεξιλόγιο και οι γλωσσολογικές κατασκευές μπορούν να αποκτηθούν μέσω επαναληπτικής εκπαίδευσης και δοκιμής μίας ευρείας συλλογής της εισόδου του χρήστη.

- πόσο καλά οι χρήστες μπορούν να διακρίνουν τις επικοινωνιακές δυνατότητες του συστήματος.

- πόσο καλά οι χρήστες μπορούν να παραμείνουν μέσα στα όρια αυτών των δυνατοτήτων.

- ποιο επίπεδο επίδοσης δραστηριότητας μπορούν να επιτύχουν οι χρήστες.

- ποιο επίπεδο παρερμηνείας οι χρήστες θα ανεχθούν και ποιο επίπεδο απαιτείται γι' αυτούς, ώστε να λύσουν τα προβλήματα αποτελεσματικά, και

- πόση εκπαίδευση είναι αποδεκτή.

Τα συστήματα δεν είναι έμπειρα στο χειρισμό προβλημάτων γλωσσολογικής κάλυψης παρά μόνο στο να αποκρίνονται ότι δεδομένες λέξεις δεν περιέχονται στο λεξιλόγιο ή ότι το λεγόμενο δεν κατανοήθηκε. Η ίδια η αναγνώριση ότι μία λέξη είναι εκτός λεξιλογίου είναι ένα δύσκολο ζήτημα (Cole et al., 1992). Αν οι χρήστες μπορούν να διακρίνουν το λεξιλόγιο του συστήματος, μπορούμε να είμαστε αισιόδοξοι ότι μπορούν να προσαρμοσθούν σε αυτό το λεξιλόγιο. Πράγματι, η έρευνα επικοινωνίας ανθρώπου με άνθρωπο έδειξε ότι χρήστες που επικοινωνούν πληκτρολογώντας, μπορούν να επιλύσουν προβλήματα τόσο αποτελεσματικά με ένα περιορισμένο λεξιλόγιο καθορισμένης εργασίας (500 με 1000 λέξεις) όσο και με ένα απεριόριστο λεξιλόγιο (Kelly και Chapanis, 1977 Michaelis et al., 1977). Η προσαρμογή του χρήστη στους περιορισμούς του λεξιλογίου βρέθηκε επίσης για προσομοιωμένη αλληλεπίδραση ανθρώπου υπολογιστή (Zoltan-Ford, 1983, 1991), παρ' όλο που αυτά τα αποτελέσματα χρειάζεται να επαληθευτούν όσον αφορά στην αλληλεπίδραση ανθρώπου υπολογιστή με ομιλία.

Για αλληλεπιδραστικές εφαρμογές, ο χρήστης ενδέχεται να ξεκινήσει να *μιμηθεί* ή να *μοντελοποιήσει* τη γλώσσα που παρατηρεί το σύστημα και

παρουσιάζεται η ευκαιρία στο σύστημα να παίξει έναν ενεργό ρόλο στη *διαμόρφωση* και *καθοδήγηση* της γλώσσας του χρήστη, ώστε να ταιριάζει στην κάλυψη στενότερα. Πολυάριθμες μελέτες ανθρώπινης επικοινωνίας έδειξαν ότι οι άνθρωποι θα προσαρμόσουν στους τύπους ομιλίας των συνομιλητών τους συμπεριλαμβανομένων της φωνητικής έντασης (Welkowitz et al., 1972), της διαλέκτου (Giles et al., 1987) και του ρυθμού (Street et al., 1983). Οι εξηγήσεις γι' αυτή τη σύγκλιση τύπων διαλόγου εμπεριέχουν κοινωνικούς παράγοντες όπως είναι η επιθυμία αποδοκιμασίας (Giles et al., 1987) και ψυχογλωσσολογικούς παράγοντες που σχετίζονται με περιορισμούς μνήμης (Levelt και Kelter, 1982). Παρόμοια αποτελέσματα βρέθηκαν σε μία μελέτη επικοινωνίας με πληκτρολογιο και με ομιλία με ένα προσομοιωμένο σύστημα φυσικής γλώσσας (Zoltan-Ford, 1983, 1984) που έδειξε ότι οι άνθρωποι θα μοντελοποιήσουν το λεξιλόγιο και το μήκος των απαντήσεων του συστήματος. Για παράδειγμα, αν οι απαντήσεις του συστήματος είναι λακωνικές, το πιθανότερο είναι να ισχύει το ίδιο για την είσοδο του χρήστη. Σε μία μελέτη προσομοίωσης αλληλεπιδράσεων με βάση δεδομένων με πληκτρολογημένη φυσική γλώσσα τα υποκείμενα μοντελοποίησαν απλές συντακτικές δομές και λεξικολογικά στοιχεία που παρατήρησαν στις παραφράσεις της εισόδου τους από το σύστημα (Leiser, 1989). Ωστόσο, δεν είναι γνωστό, αν η μοντελοποίηση συντακτικών δομών συμβαίνει στην αλληλεπίδραση ανθρώπου υπολογιστή με ομιλία. Αν οι χρήστες συστημάτων *ομιλούμενης* γλώσσας μάθουν να προσαρμόζουν τις γραμματικές δομές που παρατηρούν, τότε είναι δυνατές νέες μορφές εκπαίδευσης του χρήστη με τους σχεδιαστές του συστήματος να εμμένουν στην αρχή ότι οποιαδήποτε μηνύματα παρέχονται σε ένα χρήστη πρέπει να είναι αναλύσιμα από τον αναλυτή του συστήματος. Ενας τρόπος να εγγυηθεί κανείς αυτή τη συμπεριφορά συστήματος θα ήταν να απαιτηθεί το σύστημα να δημιουργεί τα λεγόμενά του, παρά απλώς να εξιστορεί τυποποιημένο κείμενο, χρησιμοποιώντας μία δικατευθυντήρια γραμματική. Οποιαδήποτε λεγόμενα το σύστημα θα δημιουργούσε, χρησιμοποιώντας αυτή τη γραμματική, θα ήταν τότε εγγυημένο ότι είναι αναλύσιμα.

Ενας αριθμός μελετών διερεύνησαν μεθόδους διαμόρφωσης της γλώσσας του χρήστη, ώστε να καλύπτεται από το σύστημα. Για εφαρμογές τηλεπικοινωνιών, η διατύπωση των προτροπών του συστήματος για πληροφορία που δίδεται μέσω του τηλεφώνου, επηρεάζει δραματικά το ρυθμό της συμμόρφωσης του καλούντος στις αναμενόμενες λέξεις και φράσεις (Basson, 1992 Rubin-Spitz και Yashchin, 1989 Spitz, 1991). Για συστήματα με ανατροφοδότηση που βασίζεται σε οθόνη, η ανθρώπινη ομιλούμενη γλώσσα μπορεί να διοχετευθεί αποτελεσματικά μέσω της χρήσης μιας φόρμας που ο χρήστης γεμίζει με ομιλία (Oviatt et al., 1993). Οι αλληλεπιδράσεις που βασίζονται σε φόρμα ελλατώνουν τη συντακτική ασάφεια της ομιλίας του χρήστη κατά 65%, μετρημένη ως ο αριθμός των αναλύσεων ανά λεγόμενο, επομένως οδηγώντας στη γλώσσα του χρήστη που είναι απλούστερη στην επεξεργασία. Κατά το ίδιο χρονικό διάστημα, όσον αφορά στις υπηρεσίες συναλλαγών που αναλύονται σε αυτή τη μελέτη, οι χρήστες βρέθηκαν να προτιμούν αλληλεπίδραση με ομιλία και γραπτή αλληλεπίδραση που βασίζονται σε φόρμες, από τις μη περιορισμένες κατά ένα παράγοντα του 2 προς 1. Ως εκ τούτου, η ανθρώπινη γλώσσα όχι μόνο μπορεί να διοχετευθεί αλλά φαίνεται να υπάρχουν περιπτώσεις όπου προτιμάται η καθοδήγηση και η έννοια της ολοκλήρωσης που παρέχονται από μία φόρμα.

Αλληλεπίδραση και Διάλογος

Οταν δίνεται η ευκαιρία να αλληλεπιδράσουν με συστήματα μέσω ομιλούμενης φυσικής γλώσσας, οι χρήστες θα προσπαθήσουν να συμμετέχουν σε διαλόγους, αναμένοντας προηγούμενα λεγόμενα και απαντήσεις να θέσουν ένα γενικό πλαίσιο για επόμενα λεγόμενα και ο συνομιλητής τους να κάνει χρήση αυτού του γενικού πλαισίου για να καθορίσει τις αναφορές των αντωνυμιών. Παρ' όλο που οι αντωνυμίες και άλλες κατασκευές που είναι εξαρτημένες από τα συμφραζόμενα, μερικές φορές συναντώνται σπανιότερα σε διαλόγους με μηχανές σε σχέση με τους διαλόγους ανθρώπου με άνθρωπο (Kennedy et al., 1988), η εξάρτηση από το κείμενο αποτελεί ένα θεμέλιο λίθο της αλληλεπίδρασης ανθρώπου υπολογιστή. Για παράδειγμα, λεγόμενα εξαρτημένα από τα συμφραζόμενα περιλαμβάνουν το 44% του σώματος κειμένου του ATIS που συλλέχθηκε για την ολότητα ομιλούμενης γλώσσας του ARPA (MADCOW Working Group, 1992). Γενικά μία λύση στο πρόβλημα της κατανόησης λεγομένων που είναι εξαρτημένα από τα σύμφραζόμενα θα είναι δύσκολο να δοθεί καθώς ενδέχεται να απαιτήσει το σύστημα να αναπτύξει ένα αυθαίρετο σύνολο γνώσης του κόσμου(Charniak, 1973 Winograd, 1972). Ωστόσο, εκτιμάται ότι μία απλή στρατηγική για καθορισμό αναφορών που χρησιμοποιείται σε επεξεργασία κειμένου και μία που χρησιμοποιεί μόνο τη συντακτική δομή προηγούμενων λεγομένων ενδεχομένως να επαρκούν για τον προσδιορισμό των σωστών αναφορών για αντωνυμίες σε περισσότερο του 90% των περιπτώσεων (Hobbs, 1978). Κατά πόσο αυτές οι τεχνικές θα δουλέψουν εξίσου καλά για διάλογο ανθρώπου υπολογιστή με ομιλία είναι άγνωστο. Ενας τρόπος να μετριαστεί η έμφυτη δυσκολία του καθορισμού αναφορών, όταν χρησιμοποιείται ένα πολυτροπικό σύστημα, ενδέχεται να είναι ο συνδυασμός ομιλούμενων αντωνυμιών και φράσεων οριστικών ονομάτων με ενέργειες ένδειξης (Cohen, 1991 Cohen et al., 1989).

Τα παρόντα συστήματα ομιλούμενης γλώσσας υποστήριξαν διαλόγους στους οποίους ο χρήστης θέτει πολλαπλές ερωτήσεις, ορισμένες από τις οποίες απαιτούν περαιτέρω επεξεργασία των απαντήσεων σε προηγούμενες ερωτήσεις (ARPA, 1993), ή διαλόγους στους οποίους ο χρήστης παρακινείται για πληροφορία (Andry 1992, Peckham 1991). Είναι πιθανόν να απαιτηθεί από χρήστες πιο διαφορετική συμπεριφορά διαλόγου, όπως η δυνατότητα να συμμετέχουν σε συμβουλευτικούς, διευκρινιστικούς και επιβεβαιωτικούς διαλόγους (Codd, 1974 Litman και Allen, 1987). Σε σχέση με τους διαλόγους επιβεβαίωσης η επικοινωνία με ομιλία είναι στενά αλληλεπιδραστική και οι ομιλητές αναμένουν γρήγορη επιβεβαίωση κατανόησης μέσω καναλιών οπισθοδρόμησης (π.χ., "uh huh") και άλλων σημάτων. Μελέτες έδειξαν ότι καθυστερήσεις επικοινωνίας τόσο μικρές όσο 0,25 δευτερόλεπτα μπορούν να διασπάσουν πρότυπα συζήτησης (Krauss και Bricker, 1967) οδηγώντας τους ομιλητές σε επεξεργασία και επαναδιατύπωση των λεγομένων τους. (Krauss και Weinheimer, 1966 Oviatt και Cohen, 1991a) και ότι οι τηλεφωνικές επικοινωνίες είναι ιδιαίτερα ευαίσθητες στις καθυστερήσεις. Η ανάγκη για έγκαιρες επιβεβαιώσεις θα αποτελέσει πρόκληση για πολλές εφαρμογές επεξεργασίας ομιλούμενης γλώσσας, ειδικά γι' αυτές που εμπλέκουν την τηλεφωνία.

Για την υποστήριξη μιας ευρύτερης κλίμακας διαλογικής συμπεριφοράς διερευνούνται μαθηματικά και υπολογιστικά γενικότερα μοντέλα διαλόγου που συμπεριλαμβάνουν μοντέλα διαλόγου και γραμματικές διαλόγου που βασίζονται

σε σχέδιο. Τα μοντέλα που βασίζονται σε σχέδιο, στηρίζονται στην παρατήρηση ότι τα λεγόμενα δεν είναι απλά σειρές λέξεων αλλά είναι η παρατηρητέα επίδοση των επικοινωνιακών ενεργειών, ή ενεργειών ομιλίας (Searle, 1969), όπως η αίτηση, πληροφορία, προειδοποίηση, πρόταση και επιβεβαίωση. Επιπλέον, οι άνθρωποι δεν εκτελούν μόνο ενέργειες τυχαία, αλλά σχεδιάζουν τις ενέργειές τους για να επιτύχουν ποικίλους στόχους και στην περίπτωση των επικοινωνιακών ενεργειών αυτοί οι στόχοι περιλαμβάνουν αλλαγές στις πνευματικές καταστάσεις των ακροατών. Για παράδειγμα, οι αιτήσεις των ομιλητών σχεδιάζονται να μεταβάλλουν τις προθέσεις των παραληπτών. Θεωρίες επικοινωνιακής ενέργειας και διαλόγου που βασίζονται σε σχέδιο (Allen και Perrault, 1980 Appelt, 1985 Cohen και Levesque, 1990 Cohen και Perrault, 1979 Perrault και Allen, 1980 Sidner και Israel, 1981) υποθέτουν ότι οι ενέργειες ομιλίας του ομιλητή είναι μέρος ενός σχεδίου και η δουλειά του ακροατή είναι να αποκαλύψει και να αποκριθεί κατάλληλα στο βαθύτερο σχέδιο παρά απλώς στα λεγόμενα. Για παράδειγμα, σε απάντηση μίας ερώτησης του πελάτη "Που είναι οι μπριζόλες που διαφημίζετε;" μία απάντηση του κρεοπώλη "Πόσες θέλετε;" είναι κατάλληλη διότι ο κρεοπώλης ανακάλυψε ότι το σχέδιο του πελάτη να πάρει ο ίδιος τις μπριζόλες πρόκειται να αποτύχει. Οντας συνεργάσιμος προσπαθεί να εκτελέσει ένα σχέδιο για να επιτύχει τον απώτερο στόχο του πελάτη να αποκτήσει τις μπριζόλες (Cohen, 1978). Η τρέχουσα έρευνα στο μοντέλο αυτό προσπαθεί να ενσωματώσει πολυπλοκότερα φαινόμενα διαλόγων, όπως οι διευκρινίσεις (Litman και Allen, 1987, 1990 Yamaoka και Iida, 1991) και να μοντελοποιήσει το διάλογο περισσότερο ως μία *κοινή* επιχείρηση, κάτι που οι συμμετέχοντες επιχειρούν μαζί (Clark και Wilkes-Gibbs, 1986 Cohen και Levesque, 1991 Grosz και Sidner, 1990).

Η προσέγγιση διαλογικής γραμματικής μοντελοποιεί το διάλογο απλά ως ένα δίκτυο μετάβασης σε πεπερασμένες καταστάσεις (Dahlback και Johnsson, 1992 Polanyi και Scha, 1984 Winograd και Flores, 1986), στο οποίο οι μεταβάσεις κατάστασης συμβαίνουν στη βάση του τύπου της επικοινωνιακής ενέργειας που συνέβει (π.χ., μία αίτηση). Τέτοια αυτόματα θα μπορούσαν να χρησιμοποιηθούν για την πρόβλεψη των "καταστάσεων" του επόμενου διαλόγου που είναι πιθανές και έτσι θα βοηθούσε τους αναγνωριστές ομιλίας μεταβάλλοντας τις πιθανότητες ποικίλων λεξικολογικών, συντακτικών, σημασιολογικών και πραγματικών πληροφοριών (Andry, 1992 Young et al., 1989). Ωστόσο, ένας αριθμός μειονεκτημάτων του μοντέλου είναι προφανής (Cohen, 1993 Levinson 1981). Κατά πρώτο λόγο, απαιτεί την εκτέλεση των επικοινωνιακών ενεργειών από τον ομιλητή κατά τον προσδιορισμό ενός λεγομένου, κάτι που είναι το ίδιο ένα δύσκολο πρόβλημα, για το οποίο προηγούμενες λύσεις απαίτησαν αναγνώριση σχεδίου (Allen και Perrault, 1980 Kautz, 1990 Perrault και Allen, 1980). Κατά δεύτερο λόγο, το μοντέλο υποθέτει ότι μόνο μία κατάσταση προκύπτει από μία μετάβαση. Ωστόσο τα λεγόμενα είναι πολυλειτουργικά. Ενα λεγόμενο, για παράδειγμα, ενδέχεται να είναι ταυτόχρονα μία απόρριψη και ένας ισχυρισμός. Το υποσύστημα διαλόγου γραμματικής θα χρειαζόταν επομένως να βρίσκεται σε πολλαπλές καταστάσεις ταυτόχρονα, μία ιδιότητα που τυπικά δεν είναι επιτρεπτή. Τέλος και το κυριότερο, το μοντέλο δεν αναφέρει πώς τα συστήματα θα μπορούσαν να επιλέξουν μεταξύ των επόμενων κινήσεων, δηλαδή, μεταξύ των καταστάσεων που είναι δυνατές εκείνη τη στιγμή προκειμένου να παίξει το ρόλο του ως συνεργάσιμος γνώστης. Κάτι ανάλογο της σχεδίασης είναι τότε επίσης πιθανό να ζητηθεί.

Η έρευνα διαλόγου είναι αυτή τη στιγμή ο ασθενέστερος σύνδεσμος στο πρόγραμμα έρευνας για την ανάπτυξη συστημάτων ομιλούμενης γλώσσας. Πρώτον και κύριον, η τεχνολογία διαλόγου έχει την ανάγκη μίας μεθοδολογίας καθορισμού, στην οποία ένας θεωρητικός θα μπορούσε να δηλώσει τυπικά τι θα έπρεπε να κάνει ένα σύστημα διαλόγου (π.χ., τι θα μετρούσε ως αποδεκτή συμπεριφορά διαλόγου). Οπως και σε άλλους κλάδους της επιστήμης των υπολογιστών, τέτοιοι καθορισμοί ενδέχεται τότε να οδηγήσουν σε μεθόδους για μαθηματική και εμπειρική αξιολόγηση του κατά πόσο ένα δεδομένο σύστημα διαθέτει τις προδιαγραφές. Ωστόσο, προκειμένου να συμβεί αυτό θα απαιτηθούν νέες θεωρητικές προσεγγίσεις. Κατά δεύτερο λόγο, απαιτείται να πραγματοποιηθούν περισσότερα πειράματα εφαρμογών που κυμαίνονται μεταξύ των απλούστερων μοντέλων διαλόγου που βασίζονται σε δήλωση έως τις περιεκτικότερες προσεγγίσεις που βασίζονται σε σχέδιο. Ερευνα που στοχεύει στην ανάπτυξη υπολογιστικά βολικών αλγορίθμων αναγνώρισης σχεδίου είναι αναγκαία σε κρίσιμο βαθμό.

ΠΟΛΥΤΡΟΠΙΚΑ ΣΥΣΤΗΜΑΤΑ

Υπάρχει μικρή αμφιβολία σχετικά με το ότι η φωνή θα φιγουράρει σε εξέχουσα θέση στον πίνακα πιθανών τεχνολογιών διεπαφής που είναι διαθέσιμες σε όσους ασχολούνται με την ανάπτυξη. Ωστόσο, με εξαίρεση τις συμβατικές εφαρμογές που βασίζονται στο τηλέφωνο οι διεπαφές ανθρώπου υπολογιστή που ενσωματώνουν φωνή θα είναι πιθανώς πολυτροπικές με την έννοια του συνδυασμού φωνής με τη χρήση μιας συσκευής ένδειξης, χειρονομιών, γραψίματος κ.λπ. για ανατροφοδότηση μέσω οθόνης (Cohen et al., 1989 Hauptmann and McAvinney, 1993 Oviatt, 1992 Wahlster, 1991). Πολλά συστήματα εφαρμογών απαιτούν πολυτροπική επικοινωνία, όπως οι αλληλεπιδράσεις που έμφυτα βασίζονται σε χάρτη. Τέτοια συστήματα μπορούν να περιπλέκουν συντονισμένη ομιλία, χειρονομία, ένδειξη ή γράψιμο στο χάρτη κατά την είσοδο και σύνθεση ομιλίας συντονισμένη με γραφικά για έξοδο. Από την προηγούμενη συζήτηση είναι προφανές ότι κάθε τεχνολογία διεπαφής έχει ισχυρά σημεία και αδυναμίες και ενδέχεται να είναι στρατηγικό να επιχειρηθεί να αναπτυχθούν διεπαφές που επωφελούνται από τα ισχυρά σημεία κάποιας για να ξεπεράσουν τις αδυναμίες άλλης (Cohen, 1991). Αυτό σημαίνει ότι οι χρήστες θα ήταν ικανοί να μιλούν, όταν το επιθυμούν, συμπληρώνοντας την ομιλία με άλλους τρόπους, όποτε αυτό απαιτείται.

Υπάρχουν πολλά πλεονεκτήματα στις πολυτροπικές διεπαφές:

Αποφυγή σφαλμάτων και εύρωστη απόδοση. Οι πολυτροπικές διεπαφές μπορούν να προσφέρουν το ενδεχόμενο αποφυγής σφαλμάτων που διαφορετικά θα γίνονταν σε μία μονότροπη διεπαφή. Για παράδειγμα, εκτιμάται ότι 86% των σφαλμάτων ανθρώπινης επίδοσης για κρίσιμες δραστηριότητες που συνέβησαν κατά τη διάρκεια μίας μελέτης διερμηνευμένης τηλεφωνίας θα μπορούσαν να είχαν αποφευχθεί ανοίγοντας ένα κανάλι γραψίματος που βασίζεται σε οθόνη (Oviatt, υπό έκδοση). Η πολυτροπική αναγνώριση επίσης προσφέρει τη δυνατότητα βελτίωσης της αναγνώρισης σε δυσμενείς συνθήκες. Για παράδειγμα, ταυτόχρονη χρήση αναγνωριστών ομιλίας που διαβάζουν τα χείλη ενδέχεται να αυξήσει το ρυθμό αναγνώρισης σε περιβάλλοντα υψηλού θορύβου (Garcia et al., 1992 Petajan et al., 1988) που διαφορετικά θα κατέστρεφαν αναγνωριστές ακουστικής ομιλίας. Εναλλακτικά, σε τέτοια περιβάλλοντα οι

χρήστες πολυτροπικών διεπαφών θα άλλαζαν απλά μεταξύ των τρόπων, για παράδειγμα, χρησιμοποιώντας το γράψιμο.

Διόρθωση σφαλμάτων. Οι πολυτροπικές διεπαφές προσφέρουν περισσότερες επιλογές για διόρθωση σφαλμάτων, σε σχέση με όσα συμβαίνουν. Τα σφάλματα αναγνώρισης εμφανίζουν πρόβλημα στους χρήστες, εν μέρει επειδή η πηγή τους δεν είναι προφανής. Οι χρήστες συχνά ανταπαντούν σε σφάλματα αναγνώρισης ομιλίας με υπεράρθρωση. Ομως, εφόσον οι αναγνωριστές δεν είναι τυπικά εκπαιδευμένοι σε ομιλία υπεράρθρωσης, αυτή η στρατηγική διόρθωσης οδηγεί σε μικρότερη πιθανότητα επιτυχημένης αναγνώρισης γι' αυτό το περιεχόμενο (Shriberg et al., 1992). Προβλήματα αναγνώρισης ενδέχεται κατ' αυτό τον τρόπο να επαναληφθούν πολυάριθμες φορές για το ίδιο περιεχόμενο οδηγώντας σε μία "κίνηση υποβιβασμού" που απογοητεύτει τους χρήστες και, μπορεί να προκαλέσει την εγκατάλειψη της εφαρμογής (Oviatt, 1992). Παρέχοντας την επιλογή της χρήσης άλλου τρόπου, όπως είναι το γράψιμο, ένας χρήστης μπορεί απλά να μεταβαίνει μεταξύ τρόπων, για να διορθώσει ένα σφάλμα του πρώτου τρόπου.

Παραλλαγή κατάστασης και χρήστη. Οι ποικίλες περιστάσεις, στις οποίες θα χρησιμοποιούνται οι φορητοί υπολογιστές, είναι δυνατόν να μεταβάλλουν τις προτιμήσεις των ανθρώπων για ένα τρόπο επικοινωνίας ή άλλο. Για παράδειγμα, ο χρήστης ενδέχεται κατά καιρούς να συναντά ενθόρυβα περιβάλλοντα ή να επιθυμεί μυστικότητα και επομένως θα προτιμούσε να μη μιλά. Επίσης, οι άνθρωποι ενδέχεται να προτιμούν να μιλούν για το περιεχόμενο κάποιας εργασίας, αλλά όχι για άλλες. Τέλος, διαφορετικοί τύποι χρηστών ενδέχεται συστηματικά να προτιμούν να χρησιμοποιούν έναν τρόπο παρά κάποιον άλλο. Σε όλες αυτές τις περιπτώσεις ένα πολυτροπικό σύστημα προσφέρει την ζητούμενη προσαρμοστικότητα

Καθώς ερευνούμε την πολυτροπική διεπαφή για πιθανές λύσεις σε προβλήματα που προκύπτουν από εφαρμογές ομιλίας μόνο, πολλά εμπόδια υλοποίησης χρειάζεται να ξεπερασθούν προκειμένου να ολοκληρωθούν και να συγχρονιστούν οι τρόποι. Για παράδειγμα, τα πολυτροπικά συστήματα θα μπορούσαν να παρουσιάσουν πληροφορία γραφικά ή με πολλαπλά συντονισμένους τρόπους (Feiner και McKeown, 1991 Wahlster, 1991) και να επιτρέπουν στους χρήστες να αναφέρονται γλωσσολογικά σε οντότητες που έχουν εισαχθεί γραφικά (Cohen, 1991 Wahlster, 1991). Απαιτείται η ανάπτυξη τεχνικών για το συγχρονισμό εισόδου από ταυτόχρονη ροή δεδομένων, έτσι ώστε, για παράδειγμα, είσοδοι χειρονομίας να μπορούν να βοηθήσουν στην επίλυση ασαφειών στην επεξεργασία ομιλίας και αντιστρόφως. Απαιτείται έρευνα στις πολυτροπικές διεπαφές, για να εξεταστούν, όχι μόνο οι τεχνικές για εδραίωση μίας παραγωγικής σύνθεσης μεταξύ τρόπων, αλλά επίσης και η επίδραση που συγκεκριμένες τεχνολογίες ολοκλήρωσης θα έχουν στην αλληλεπίδραση ανθρώπου υπολογιστή. Απαιτείται να επιχειρηθεί περισσότερο εμπειρική έρευνα στην ανθρώπινη χρήση πολυτροπικών συστημάτων, καθώς γνωρίζουμε ελάχιστα σχετικά με το πώς οι χρήστες χρησιμοποιούν πολλαπλούς τρόπους για επικοινωνία με άλλους ανθρώπους, πόσο μάλλον με υπολογιστές, ή σχετικά με την υποστήριξη τέτοιας επικοινωνίας αποτελεσματικότερα.

ΕΠΙΣΤΗΜΟΝΙΚΗ ΕΡΕΥΝΑ ΣΤΟΥΣ ΤΡΟΠΟΥΣ ΕΠΙΚΟΙΝΩΝΙΑΣ

Η παρούσα έρευνα και το κλίμα ανάπτυξης για την τεχνολογία που βασίζεται στην ομιλία είναι ενεργητικότερα σε σχέση με τα αντίστοιχα κατά το χρονικό διάστημα της αναφοράς του 1984 του Εθνικού Συμβουλίου Ερευνας (Εθνικό Συμβούλιο Ερευνας, 1984), σχετικά με αναγνώριση ομιλίας σε απαιτητικά περιβάλλοντα. Σημαντικά κεφάλαια για έρευνα και ανάπτυξη διατίθενται πλέον για την κατασκευή συστημάτων κατανόησης ομιλίας και έχουν αναπτυχθεί τα πρώτα ανεξάρτητα από τον ομιλητή, συνεχή, πραγματικού χρόνου συστήματα ομιλούμενης γλώσσας. Ωστόσο, ορισμένα από τα ίδια προβλήματα που προσδιορίσθηκαν τότε, εξακολουθούν να ισχύουν και τώρα. Πιο συγκεκριμένα, ελάχιστες απαντήσεις είναι διαθέσιμες σχετικά με το πώς οι άνθρωποι θα αλληλεπιδράσουν με συστήματα χρησιμοποιώντας φωνή καθώς και σχετικά με το πόσο καλά θα εκτελέσουν τις εργασίες στα τελικά περιβάλλοντα σε αντιδιαστολή προς το εργαστήριο. Μικρή έρευνα έχει γίνει για την εξάρτηση της επικοινωνίας από τον τρόπο που χρησιμοποείται, ή τον τύπο των εργασιών, εν μέρει επειδή δεν υπήρχαν ταξινομίες με αρχές ή περιεκτική έρευνα που να απευθύνεται σε αυτούς τους παράγοντες. Πιο συγκεκριμένα, η χρήση πολλαπλών τρόπων επικοινωνίας για την υποστήριξη αλληλεπίδρασης ανθρώπου υπολογιστή μόλις τώρα εφαρμόζεται.

Ευτυχώς ο χώρος είναι αυτή τη στιγμή σε θέση να καλύψει τα κενά της γνωστικής του βάσης σχετικά με την επικοινωνία ανθρώπου μηχανής. με ομιλία Χρησιμοποιώντας υπάρχοντα συστήματα που κατανοούν πραγματικού χρόνου, συνεχή λεγόμενα που επιτρέπουν στους χρήστες να επιλύσουν πραγματικά προβλήματα, ένας αριθμός μελετών ζωτικής σημασίας μπορούν τώρα να αναληφθούν κατά ένα συστηματικότερο τρόπο. Παραδείγματα περιλαμβάνουν:

- μακροχρόνιες μελέτες της γλωσσολογικής και της σχετικής με την επίλυση προβλημάτων συμπεριφοράς του χρήστη που θα διερευνούσαν πώς οι χρήστες προσαρμόζονται σε ένα δεδομένο σύστημα

- μελέτες της κατανόησης από το χρήστη των ορίων του συστήματος και της απόδοσης των χρηστών κατά την παρατήρηση των ορίων του συστήματος

- μελέτες διαφορετικών τεχνικών για την αποκάλυψη της κάλυψης ενός συστήματος και για την καθοδήγηση της εισόδου του χρήστη

- μελέτες που συγκρίνουν την αποτελεσματικότητα της τεχνολογίας ομιλούμενης γλώσσας με εναλλακτικές τεχνολογίες, όπως η χρήση συστημάτων φυσικής γλώσσας που βασίζονται σε πληκτρολόγιο, γλώσσες ερωτήσεων ή διεπαφές άμεσης προσπέλασης και

- μελέτες που αναλύουν τη γλώσσα, την απόδοση των εργασιών και τις προτιμήσεις του χρήστη για την χρησιμοποίηση διαφορετικών τρόπων σε προσωπικό επίπεδο και μέσα στα όρια μίας ολοκληρωμένης πολυτροπικής διεπαφής.

Οι πληροφορίες που αποκτώνται από τέτοιες μελέτες θα ήταν μία ανεκτίμητη προσθήκη στη γνωστική βάση του πώς η επεξεργασία ομιλούμενης γλώσσας μπορεί να εμπλακεί σε μία χρησιμοποιήσιμη διεπαφή ανθρώπου υπολογιστή. Αμείωτες προσπάθειες απαιτείται να γίνουν για την ανάπτυξη

καταλληλότερων μεθόδων προσομοίωσης ομιλούμενης γλώσσας, την κατανόηση του πώς θα κατασκευασθούν περιορισμένα αλλά εύρωστα συστήματα διαλόγου που βασίζονται σε μία ποικιλία τρόπων επικοινωνίας και τη μελέτη της φύσης του διαλόγου.

Μία ζωτική και υποτιμημένη συμβολή στην επιτυχημένη ανάπτυξη της τεχνολογίας φωνής για αλληλεπίδραση ανθρώπου υπολογιστή θα προέλθει από την ανάπτυξη ενός έγκυρου εμπειρικά και με αρχές συνόλου οδηγιών ανθρώπου διεπαφής για διεπαφές που ενσωματώνουν την ομιλία (cf. Lea, 1992). Οδηγοί GUIs παρέχουν τυπικές ευριστικές μεθόδους και προτάσεις για την κατασκευή "χρησιμοποιήσιμων" διεπαφών, παρ' όλο που συχνά αυτές οι προτάσεις δεν βασίζονται σε επιστημονικά τεκμηριωμένα γεγονότα και αρχές. Παρ' όλη την προφανή εσπιτυχία αυτών των οδηγών για GUIs, δεν είναι καθόλου σαφές ότι ένα απλό σύνολο ευριστικών θα δουλέψει για την τεχνολογία ομιλούμενης γλώσσας, επειδή η ανθρώπινη γλώσσα είναι περισσότερο μεταβλητή και δημιουργική σε σχέση με την συμπεριφορά που επιτρέπεται από τα GUIs. Απαντήσεις σε ορισμένες από τις ερωτήσεις που τέθηκαν πρωτύτερα θα ήταν πολύτιμες στον καθορισμό τυπικών εμπειρικών θεμελίων για την ανάπτυξη αποτελεσματικών οδηγιών μίας νέας γενιάς διεπαφών προσανατολισμένων στη γλώσσα.

Τέλος, ένα τέτοιο σύνολο οδηγιών που ενσωματώνει τα αποτελέσματα επιστημονικής θεωρίας και πειραμάτων θα ήταν ικανό να προβλέψει, δοθέντων μιας καθορισμένης επικοινωνιακής κατάστασης, δραστηριότητας, πληθυσμού των χρηστών, και ενός συνόλου των στοιχείων των τρόπων, με τι θα μοιάζει η αλληλεπίδραση χρήστη υπολογιστή για μία πολυτροπική διεπαφή συγκεκριμένης διαμόρφωσης. Τέτοιες προβλέψεις θα πληροφορούσαν εκ των προτέρων όσους ασχολούνται με την ανάπτυξη σχετικά με ενδεχόμενα προβλήματα και θα οδηγούσαν σε μία περισσότερο εύρωστη, χρησιμοποιήσιμη και ικανοποιητική διεπαφή ανθρώπου υπολογιστή. Δοθέντων της πολυπλοκότητας κατά την διαδικασία του σχεδιασμού και των σημαντικών εξόδων που απαιτούνται για τη δημιουργία εφαρμογών ομιλούμενης γλώσσας, αν οι σχεδιαστές αφεθούν στη διαίσθησή τους οι εφαρμογές θα δοκιμάζονται. Ως εκ τούτου, για επιστημονικούς, τεχνολογικούς και οικονομικούς λόγους απαιτείται μία συντονισμένη προσπάθεια για την ανάπτυξη μίας επιστημονικότερης κατανόησης των τρόπων επικοινωνίας και του πώς αυτοί μπορούν να ολοκληρωθούν με την υποστήριξη επιτυχούς αλληλεπίδρασης ανθρώπου υπολογιστή.

ΕΥΧΑΡΙΣΤΙΕΣ

Πολλές ευχαριστίες στους Jared Bernstein, Clay Coler, Carol Simpson, Ray Perrault, Robert Markinson, Raja Rajasekharan και John Vester για τις πολύτιμες συζητήσεις και τις πηγές υλικού.

ΠΕΡΙΛΗΨΗ

Η εξέλιξη στον τομέα της αλληλεπίδρασης ανθρώπου υπολογιστή με τη χρήση φωνής οδήγησε στην κατασκευή πρωτότυπων συστημάτων αναγνώρισης ομιλίας. Ωστόσο πριν η τεχνολογία αυτή χρησιμοποιηθεί ευρέως απαιτείται μια ουσιαστική γνωστική βάση πάνω στην ανθρώπινη ομιλούμενη γλώσσα και τις επιδόσεις της κατά τη διάρκεια αλληλεπίδρασης με υπολογιστή. Προς την κατεύθυνση αυτή συγκλίνει και η παρούσα διατριβή.

Τρεις είναι οι βασικές θεματικές ενότητες που πραγματεύεται η διατριβή. Η πρώτη αφορά στο πότε η αλληλεπίδραση με ομιλούμενη γλώσσα είναι ωφέλιμη, η δεύτερη στη σύγκριση της φωνής με εναλλακτικούς τρόπους αλληλεπίδρασης ανθρώπου υπολογιστή και η τρίτη στην εξέταση εμποδίων που υπάρχουν για επιτυχημένη ανάπτυξη συστημάτων ομιλούμενης γλώσσας.

Στην πρώτη ενότητα αναφέρονται περιπτώσεις που ευνοούν την αλληλεπίδραση με ομιλία, όπως όταν τα χέρια ή οι οφθαλμοί του χρήστη είναι απασχολημένοι, όταν διαθέτουμε περιορισμένο πληκτρολόγιο και/ή οθόνη, όταν ο χρήστης είναι ανίκανος (disabled), όταν η προφορά είναι το αντικειμενικό ζήτημα της χρήσης υπολγιστή (για δραστηριότητες εκμάθησης ξένων γλωσσών και διδασκαλίας της ανάγνωσης).Οι παραπάνω περιπτώσεις αφορούν στην είσοδο με ομιλία και ενδεχομένως και στην έξοδο.Ωστόσο δεν υπάρχει ακόμη μέθοδος που να προβλέπει πότε η είσοδος ή η έξοδος φωνής θα είναι ο αποτελεσματικότερος και προτιμότερος τρόπος επικοινωνίας.

Στη δεύτερη ενότητα γίνεται σύγκριση της ομιλούμενης γλώσσας με άλλους τρόπους επικοινωνίας, όπως την πληκτρολογημένη γλώσσα.και το τρέχον κυρίαρχο παράδειγμα της γραφικής διεπαφής χρήστη Για κάθε τρόπο αναφέρονται τα ισχυρά σημεία και οι αδυναμίες του καθώς και περιπτώσεις στις οποίες ευνοείται ο ένας ή ο άλλος τρόπος.

Στην τρίτη ενότητα τα εμπόδια που παρουσιάζονται αφορούν σε ανθρώπινους παράγοντες, όπως η αυθόρμητη ομιλία και οι δυσχέρειες λόγου που αυτή συνεπάγεται, η φυσική γλώσσα και οι τεράστιες δυνατότητές της που αναγκάζουν το χρήστη να περιοριστεί σε ένα μοντέλο γλώσσας κατανοητό από τη μηχανή και η προσδοκία του χρήστη για μια αλληλεπίδραση μέσω διαλόγου δεδομένου του προβλήματος που έχουν οι μηχανές να κατανοήσουν λέγομενα που εξαρτώνται από τα συμφραζόμενα.

Στο τέλος της διατριβής γίνεται η διαπίστωση ότι τα μελλοντικά συστήματα θα είναι πολυτροπικά με τη φωνή να είναι ένας από τους διαθέσιμους τρόπους επικοινωνίας. Αναφέρονται τα πλεονεκτήματα των πολυτροπικών διεπαφών, όπως η αποφυγή σφαλμάτων και η εύρωστη απόδοση, η διόρθωση σφαλμάτων και η δυνατότητα επιλογής. Η διατριβή κλείνει με μια αναφορά σε μελέτες που μπορούν να γίνουν στο συγκεκριμένο χώρο.

ΑΝΑΦΟΡΕΣ

Advanced Research Projects Agency. ARPA Spoken Language Systems Technology Workshop. Massachusetts Institute of Technology, Cambridge, Mass. 1993.

Allen, J. F. και C. F. Perrault. Analyzing Intention in dialogues. Artificial Intelligence, 15(3):143-178, 1980/

Andry, F. Static and dynamic predictions: A method to improve speech understanding in cooperative dialogues. In Proceedings of the International Conference on Spoken Language Processing, Banff, Alberta, Canada, Oct. University of Alberta, 1992.

Andry, F., E. Bilange, F. Charperntier, K. Choukri, M. Ponamale, and S. Soudoplatoff. Computerized simulation tools for the design of an oral dialogue system. In Selected Pulications, 1988-1990, SUNDIAL Project (Esprit P2218). Commission of the European Communities, 1990.

Appelt, K. Planning English Sentences. Cambridge University Press, Cambridge, U.K., 1985

Appelt, D. E., and E. Jackson. SRI International February 1992 ATIS benchmark test results. In Fifth DARPA Workshop on Speech and Natural Language, San Mateo, Calif. Morgan Kaufmann Publishers, Inc., 1992.

Bahl, L., F. Jelinek, and R. L. Mercer. A maximum likelihood approach to continuous speech recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5(2):179-190, March 1983.

Baker, J. F. Stochastic modeling for automatic speech understanding. In D. R. Reddy, ed., Speech Recognition, pp. 521-541. Academic Press, New York, 1975.

Baker, J. M. Large vocabulary speaker-adaptive continuous speech recognition research overview at Dragon systems. In Proceedings of Eurospeech'91: 2nd European Conference on Speech Communication and Technology, pp. 29-32, Genova, Italy, 1991.

Basson, S. Prompting the user in ASR applications. In Proceedings of COST232 Workshop-European Cooperative in Science and Technology, November 1992.

Basson, S., O. Christie, S. Levas, and J. Spitz. Evaluating speech recognition potential in automating directory assistance call completion. In AVIOS Proceedings. American Voice I/O Society, 1989.

Bear, J., J. Dowding, and E. Shriberg. Detection and correction of repairs in human-computer dialog. In D. Walker, ed. Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics, Newark, Delaware, June 1992.

Bear, J., and P. Price. Prosody, syntax and parsing. In Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics, pp. 17-22, Pittsburgh, Pa., 1990.

Bernstein, J. Applications of speech recognition technology in rehabilitation. In J. E. Harkins and B. M. Virvan, eds., Speech to Text: Today and Tomorrow. GRI Monograph Series, B., No. 2. Gallaudet University Research Institute, Washington, D. C., 1988.

Bernstein, J., M. Cohen, H. Murveit, D. Rtischev, and M. Weintraub. Automatic evaluation and training in English pronunciation. In Proceedings of the 1990 International Conference on Spoken Language Processing, pp. 1185-1188, The Acoustical Society of Japan, Kobe, Japan, 1990.

Bernstein, J., and D. Rtischev. A voice interactive language instruction system. In Proceedings of Eurospeech'91, pp. 981-984, Genova, Italy. IEEE, 1991.

Capindale, R. A., and R. G. Crawford. Using a natural language interface with casual users. International Journal of Man-Machine Studies, 32:341-362, 1990.

Chamberlin, D. D., and R. G. Boyce. Sequel: A structured English query language. In Proceedings of the 1974 ACM SIGMOD Workshop on Data Description, Access and Control, May 1974.

Chapanis, A., R. B. Ochsman, R. N. Parrish, and G. D. Weeks. Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem solving. Human Factors, 14:487-509, 1972.

Chapanis, A., R. N. Parrish, R. B. Ochsman, and G. D. Weeks. Studies in interactive communication: II. The effects of four communication modes on the linguistic performance of teams during cooperative problem solving. Human Factors, 19(2):101-125, April 1977.

Charniak, E., Jack and Janet in search of a theory of knowledge. In Advance Papers of the Third Meeting of the International Joint Conference on Artificial Intelligence, Los Altos, Calif. William Kaufmann, Inc., 1973.

Clark, H. H., and D. Wilkes-Gibbs. Referring as a collaborative process. Cognition, 22:1-39, 1986.

Codd, E. F. Seven steps to rendezvous with the casual user. In Proceedings IFIP TC-2 Working Conference on Data Base Management Systems, pp. 179-200. North-Holland Publishing Co., Amsterdam, 1974.

Cohen, P. R. On Knowing What to Say: Planning Speech Acts. PhD thesis, University of Toronto, Toronto, Canada. Technical Report No. 118, Department of Computer Science, 1978.

Cohen, P. R. The pragmatics of referring and the modality of communication. Computational Linguistics, 10(2): 97-146, April-June 1984.

Cohen, P. R. The role of natural language in a multimodal interface. In the 2nd FRIEND21 International Symposium on Next Generation Human Interface Technologies, Tokyo, Japan, November 1991. Institute for Personalized Information Environment.

Cohen, P. R. Models of dialogue. In M. Nagao, ed., Cognitive Processing for vision and Voice: Proceedings of the Fourth NEC Research Symposium. SIAM, 1993.

Cohen, P. R., and H. J. Levesque. Confirmations and joint action. In Proceedings of the 12th International Joint Conference on Artificial Intelligence, pp. 951-957, Sydney, Australia, Morgan Kaufmann Publishers, Inc. 1991.

Cohen, P. R., and C. R. Perrault. Elements of a plan-based theory of speech acts. Cognitive Science, 3(3): 177-212, 1979.

Cohen, P. R., M. Dalrymple, D. B. Moran, F. C. N. Pereira. J. W. Sullivan, R.a A. Gargan, J. L. Schlossberg, and S. W. Tyler. Synergistic use of direct manipulation and natural language. In Human Factors in Computing Systems: CHI'89 Conference Proceedings, pp. 227-234, New York, Addison Wesley Publishing Co. 1989.

Cole, R., L. Hirschman, L. Atlas, M. Beckman, A. Bierman, M. Bush, J. cohen, O. Garcia, B. Hanson, H. Hermansky, S. Levinson, K. Mckeown, N. Morgan, D. Novick, M. Ostendorf., S. Oviatt, P. Price, H. Silverman, J. Spitz, A. Waibel, C. Weinstein, S. Zahorain, and V. Zue. NSF Workshop on Spoken Language

Understanding. Technical Report CS/E 92-014, Oregon Graduate Institute, September 1992.

Crane, H. D. Writing and talking to computers. Business Intelligence Program Report D91-1557, SRI International, Menlo Park, Calif., July 1991.

Dahlback, N., and A. Jonsson. An empirically bhased computationally tractable dialogue model. In Proceedings of the 14th Annual Conference of the Cognitive Science society (COGSCI-92), Bloomington, Ind., July 1992.

Dahlback, N., A. Jonsson, and L. Ahrenberg. Wizard of Oz studies-why and how. In L. Ahrenberg, N. Dahlback, and A. Jonsson, eds., Proceedings from the Workshop on Empirical Models and Methodology for Natural Language Dialogue Systems, Trento, Italy, April. Association for computational Linguistics, 1992.

Dowing, J., J. M. Gawron, D. Appelt, J. Bear, L. Cherny, R. Moore, and D. Moran. Gemini: A natural language system for spoken-language understanding. In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, pp. 54-61, Columbus, Ohio, June 1993.

Englebart, D. Design considerations for knowledge workshop terminals. In National Computer Conference, pp. 221-227, 1973.

English, w. K., D. C. Englebart, and M. A. Berman. Display-selection techniques for text manipulation. IEEE Transactions on Human Factors in Electronics, HFE-8(1)L 5-15, March 1967.

Feiner, S. K., and K. R. McKeown. COMET: Generating coordinated multimedia explanations. In Human Factors in Computing Systems (CHI'91), pp. 449-450, New York, April. ACM Press, 1991.

Fisher, S. Virtual environments, personal simulation, and telepresence. Multimedia Review: The Journal of Multimedia Computing, 1(2), 1990.

Fraser, N. M., and G. N. Gilbert. Simulating speech systems. Computer Speech and Language, 5(1): 81-99, 1991.

Garcia, O. N. A. J. goldschen, and E. D. Petajan. Feature Extraction for Optical Speech Recognition or Automatic Lipreading. Technical Report, Institute for Information Science and Technology, Department of Electrical Engineering and Computer Science. The George Washington University, Washington, D. C., November 1992.

Giles, H., A. Mulac, J. J Bradac, and P. Johnson. Speech accommodation theory: The fist decade and beyond. In M. L. McLaughlin, ed., Communication Yearbook 10, pp. 13-48. Sage Publishers, Beverly Hills, California, 1987.

Gould, J. D. How experts dictate. Journal of Experimental Psychology: Human Perception and Performance, 4(4): 648-661, 1978.

Gould, J. D. Writing and speaking letters and messages. International Journal of Man-Machine Studies, 16(1): 147-171, 1982.

Gould, J.K., J. Conti, and T. Hovanyecz. Composing letters with a simulated listening typewriter. Communications of the ACM, 26(4): 295-308, April 1983.

Grosz, B., and C. Sidner. Plans for discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, eds., Intentions in Communication, pp. 417-444. MIT Press, Cambridge, Mass., 1990.

Guymard, M., and J. Siroux. Experimentation in the specification of an oral dialogue. In H. Niemann, M. Lang, and G. Sagerer, eds., Recent Advances in Speech

Understanding and Dialogue Systems. NATO ASI Series, vol. 46. Springer Verlag, Berlin, 1988.

Harris, R. User oriented data base query with the robot natural language query system. International Journal of Man-Machine Studies, 9:696-713, 1977.

Hauptmann, A. G., and P. McAvinney. Gestures with speech for direct manipulation. International Journal of Man-Machine Studies, 38:231-249, 1993.

Hauptmann, A. G., and A. I. Rudnicky. A comparison of speech and typed input. In Proceedings of the Speech and Natural Language Workshop, pp. 219-224, San Mateo, Calif., June. Morgan Kaufmann, Publishers, Inc., 1990.

Hendrix, G. G., and B. A. Walter. the intelligent assistant. Byte, pp. 251-258, December 1987.

Hindle, D. Deterministic parsing of syntactic non-fluencies. In Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics, pp. 123-128, Cambridge, Mass., June 1983.

Hobbs, J. R. Resolving pronoun reference. Lingua, 44, 1978.

Hon, H.-W., and K.-F. Lee. Recent progress in robust vocaburay-independent speech recognition. In Proceedings of the Speech and Natural Language Workshop, pp. 258-263, San Mateo, Calif., October. Morgan Kaufmann Publishers, Inc., 1991.

Howard, J. A. Flight resting of the AFTI/F-16 voice interactive avionics system. In Proceedings of Military Speech Tech 1987, pp. 76-82, Arlington, Va., Media Dimensions, 1987.

Huang, X., F. Alleva, M.-Y. Hwang. and R. Rosenfeld. An overview of the SPHINX-II speech recognition system. In Proceedings of the ARPA Workshop on Human Language Technology, San Mateo, Calif. Morgan Kaufmann Publishers, Inc., 1993.

Hutchins, E. L., J. D. Hollan, and D. A. Norman. Direct manipulation interfaces. In D. A. Norman and S. W. Draper, eds. User Centered System Design, pp. 87-124. Lawrence Erlbaum Publishers, Hillsdale, N. J., 1986.

Jackson, E., D. Appelt, J. Bear, R. Moore, and A. Podlozny. A template matcher for robust NL interpretation. In Proceedings of the 4th DARPA Workshop on Speech and Natural Language, pp. 190-194, San Mateo, Calif., February. Morgan Kaufmann Publishers, Inc., 1991.

Jarke, M., J. A. Turner, E. A. Stohr, Y. Vassiliou, N. H. White, and K. Michielsen. A field evaluation of natural language for data retrieval. IEEE Transactions on Software Engineering, SE-11(1): 97-113, 1985.

Jelinek, F. Continuous speech recognition by statistical methods. Proceedings of the IEEE, 64: 532-536, April 1976.

Jelinek, F. The development of an experimental discrete dictation recognizer. Proceedings of the IEEE, 73(11) : 1616-1624, November 1985.

Karis, D., and K. M. Dobroth. Automating services with speech recognition over the public switched telephone network: Human factors considerations. IEEE Journal of Selected Areas in Communications, 9(4): 574-585, 1991.

Kautz, H. A circumscriptive theory of plan recognition. In P. R. Cohen, J. Morgan, and M. E. Pollack, eds., Intentions in Communication. MIT Press, Cambridge, Mass., 1990.

Kay, A., and A. Goldberg. Personal dynamic media. IEEE Computer, 10(1): 31-42, 1977.

Kelly, M. J. and A. Chapanis. Limited vocabulary natural language dialogue. International Journal of Man-Machine Studies, 9:479-501, 1977.

Kennedy, A., A. Wilkes, L. Elder, and W. S. Murray. Dialogue with machines. Cognition, 30(1): 37-72, 1988.

Kitano, H. A dm-dialog. IEEE Computer, 24(6): 36-50, June 1991.

Krauss, R. M., and P. D. Bricker. Effects of transmission delay and access delay on the efficiency of verbal communication. Journal of the Acoustical Society of America, 41(2): 286-292, 1967.

Krauss, R. M., and S. Weinheirmer. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. Journal of Personality and Social Psychology, 4:343-346, 1966.

Kreuger, M. Responsive environments. In Proceedings of the National Computer Conference, 1977.

Kubala, F., C. barry, M. Bates, R. Bobrow, P. Fng, R. Ingria, J. Makhoul, L. Nguyen, R. Schwartz, and D. Stallard. BBN BYBLOS and HARC February 1992 ATIS benchmark results. In fifth DARPA Workshop on speech and Natural Language, San Mateo, Calif. Morgan Kaufmann Publishers, Inc., 1992.

Kurematsu, A. Future perspective of automatic telephone interpretation. Transactions of IEICE, E75(1): 14-19, January 1992.

Lea, W. A. Practical lessons from configuring voice I/O systems. In Proceedings of Speech Tech/Voice Systems Worldwide, New York. Media Dimensions, Inc., 1992.

Leiser, R. G. Exploiting convergence to improve natural language understanding. Interacting with Computers, 1(3):284-298, December 1989.

Lenning, M. Using speech recognition in the telephone network to automate collect and third-number-billed calls. In Proceedings of Speech Tech'89, pp. 124-125. Arlington, Va. Media Dimensions, Inc., 1989.

Levelt, W.J. M., and S. Kelter. Surface form and memory in question-answering. Cognitive Psychology, 14(1): 78-106, 1982.

Levinson, S. Some pre-observations on the modelling of dialogue. Discourse Processes, 4(1), 1981.

Litman, D. J., and J. F. Allen. Discourse processing and commonsense plans. In P. R. Cohen, J. Morgan, and M. E. Pollack, eds., Intentions in Communication, pp. 365-388. MIT Press, Cambridge, Mass., 1990.

Luce, P. A., T. C. Feustel, and D. B. Pisoni. Capacity demands in short-term memory for synthetic and natural speech. Human Factors, 25(1): 17-32, 1983.

MADCOW Working Group. Multi-site data collection for a spoken language corpus. In Proceedings of the Speech and Natural Language Workshop, pp. 7-14, San Mateo, Calif., February. Morgan Kaufmann Publishers, Inc., 1992.

Mariani, J. Spoken language processing in the framework of human-machine communication at LIMSI. In Proceedings of Speech and Natural Language Workshop, pp. 55-60, San Mateo, Calif. Morgan Kaufmann Publishers, Inc., 1992.

Marshall, J. P. A manufacturing application of voice recognition for assembly of aircraft wire harnesses. In Proceedings of Speech Tech/Voice Systems Worldwide, New York. Media Dimensions, Inc., 1992.

Martin, G. L. The utility of speech input in user-computer interfaces. International Journal of Man-Machine Studies, 30(4): 355-375, 1989.

Martin, T. B. Practical applications of voice input to machines. Proceedings of the IEEE, 64(4): 487-501, April 1976.

Michaelis, P. R., A. Chapanis, G. D. Weeks, and M. J. Kelly. Word usage in interactive dialogue with restricted and unrestricted vocabularies. IEEE Transactions on Professional Communication, PC-20(4), December 1977.

Mostow, J., A. G. Hauptmann, L. L. chase, and S. Roth. Towards a reading coach that listens: Automated detection of oral reading errors. In Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI93), Menlo Park, Calif., AI Press/The MIT Press, 1993.

Murray, I. R., J. L. Arnott, A. F. Newell, G. Cruickshank, K. E. P. Carter, and R. Dye. Experiments with a Full-Speed Speech-Driven Word Processor. Technical Report CS 91.09, Mathematics and Computer Science Department, University of Dundee, Dundee, Scotland, April 1991.

Nakatani, C., and J. Hirschberg. A speech-first model for repair detection and correction. In Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, pp. 46-53. Columbus, Ohio, June 1993.

National Research Council. Automatic Speech Recognition in Severe Environments. National Academy Press, Washington, D. C., 1984.

Newell, A. F., J. L. Arnott, K. Carter, and G. Gruickshank. Listening typewriter simulation studies. International Journal of Man-Machine Studies, 33(1): 1-19, 1990.

Nusbaum, H. C., and E. C. Schwab. The effects of training on intelligibility of synthetic speech: II. The learning curve for synthetic speech. In Proceedings of the 105th meeting of the Acoustical Society of America, Cincinnati, Ohio, May 1983.

Nye, J. M. Human factors analysis of speech recognition systems. In Speech Technology I, pp. 50-57, 1982.

Ochsman, R. B., and A. Chapanis. The effects of 10 communication modes on the behaviour of teams during co-operative problem-solving. International Journal of Man-Machine Studies, 6(5): 579-620, September 1974.

Oviatt, S. L. Pen/voice: Complementary multimodal communication. In Proceedings of Speech Tech'92, pp. 238-241, New York, February 1992.

Oviatt, S. L. Predicting spoken disfluencies during human-computer interaction. In K. Shirai, ed., Proceedings of the International Symposium on Spoken Dialogue: New Directions in Human-Machine communication, Tokyo, Japan, November 1993.

Oviatt, S. L. Toward multimodal support for interpreted telephone dialogues. In M> M. Taylor, F. Neel, and D. G. Bouwhuis, eds., Structure of Mutimodal Dialogue. Elsevier Science Publishers B. V., Amsterdam, Netherlands, in press.

Oviatt, S. L. and P. R. Cohen. discourse structure and performance efficiency in interactive and noninteractive spoken modalities. Computer Speech and Language, 5(4):297-326. 1991a

Oviatt, S. L., and P. R. Cohen. The contributing influence of speech and interaction on human discourse patterns. In J. W. Sullivan and S. W. Tyler, eds., Intelligent User Interfaces, pp. 69-83. ACM PRESS Frontier Series. Addison-Wesley Publishing Co., New York, 1991b.

Oviatt, S. L. P. R. Cohen, M. W. Fong, and M. P. Frank. A rapid semi-automatic simulation technique for investigating interactive speech and handwriting. In J. Ohala, ed., Proceedings of the 1992 International Conference on Spoken Language Processing, pp. 1351-1354, University of Alberta, October, 1992.

Oviatt, S. L., P. R. Cohen, M. Wang, and J. Gaston. A simulation-based research strategy for designing complex NL systems. In ARPA Human Language Technology Workshop, Princeton, N. J., March 1993.

Pallet, D. S., J. G, Fiscus, W. M. Fisher, and J. S. Garofolo. Benchmark tests for the DARPA spoken language program. In Proceedings of the ARPA Workshop on Human Language Technology, San Mateo, Calif., Morgan Kaufmann Publishers, Inc., 1993.

Pavan, S., and B. Pelletti. An experimental approach to the design of an oral cooperative dialogue. In Selected Publications, 1988-1990, SUNDIAL Project (Esprit P2218). Commission of the European Communities, 1990.

Peckham, J. Speech understanding an dialogue over the telephone: An overview of the ESPRIT SUNDIAL project. In Proceedings of the Speech and Natural language Workshop, pp. 14-28, San Mateo, Calif., February. Morgan Kaufmann Publishers, Inc., 1991.

Perrault, C. R., and J. F. Allen. A plan-based analysis of indirect speech acts. American Journal of Computational Linguistics, 6(3): 167-182, 1980.

Petajan, E., B. Bradford, D. Bodoff, and N. M. Brooke. An improved automatic lipreading system to enhance speech recognition. In Proceedings of Human Factors in computing Systems (CHI'88), pp. 19-25, New York. Association for Computing Machinery Press, 1988.

Polyani, R., and R. Scha. A syntactic approach to discourse semantics. In Proceedings of the 10th International Conference on Computational Linguistics, pp. 413-419, Stanford, Calif., 1984.

Pollack, A. Computer translator phones try to compensate for Babel. New York Times, January 29, 1993.

Price, P. J. Evaluation of spoken language systems: The ATIS domain. In Proceedings of the 3rd DARPA Workshop on Speech and Natural Language, pp. 91-95, San Mateo, Calif. Morgan Kaufmann Publishers, Inc., 1990.

Price, P., M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. The use of prosody in syntactic disambiguation. In Proceedings of the Speech and Natural Language Workshop, pp. 372-377, San Mateo, Calif., October. Morgan Kaufmann Publishers, Inc., 1991.

Proceedings of the Speech and Natural Language Workshop, San Mateo, Calif., October, 1991, Morgan Kaufmann Publishers, Inc.

Rabiner, L. R., J. g. Wilpon, and A. E. Rosenberg. A voice-controlled, repertory-dialer system. Bell System Technical Journal, 59(7): 1153-1163, September 1980.

Rheingold, H. Virtual Reality. Summit Books, 1991.

Roe, d. B., F. Pereira, R. W. Sproat, and M. D. Riley. Toward a spoken language translator for restricted-domain context-free languages. In Proceedings of Eurospeech'91: 2nd European Conference on Speech Communication and Technology, pp. 1063-1066, Genova, Italy. European Speech communication Association, 1991.

Rosenhoover, F. A., J. S. Eckel, F. A. Gorg, and S. W. Rabeler. AFTI/F-16 voice interactive avionics evaluation. In Proceedings of the National Aerospace and Electronics Conference (NAECON'87). IEEE, 1987.

Rubin-spitz. J., and D. Yashchin. Effects of dialogue design on customer responses in automated operator services. In Proceedings of Speech Tech'89, 1989.

Rudinicky, A. I. Mode preference in a simple data-retrieval task. In ARPA Human Language Technology Workshop, Princeton, N. J., March 1993.

Searle, J. R. Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge, 1969.

Shneiderman, B. Natural vs. precise concise languages for human operation of computers: Research issues and experimental approaches. In Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics, pp. 139-141, Philadelphia, Pa., June 1980a.

Shneiderman, B. Software Psychology: Human Factors in Computer and Information systems. Winthrop Publishers, Inc., Cambridge, Mass., 1980b.

Shneiderman, B. Direct manipulation: A step beyond programming languages. IEEE computer, 16(8): 57-69, 1983.

Shriberg, E., E. Wade, and P. Price. Human-machine problem-solving using spoken language systems (SLS): Factors affecting performance and user satisfaction. In Proceedings of Speech and Natural Language Workshop, pp. 49-54, San Mateo, Calif. Morgan Kaufmann Publishers, Inc., 1992.

Sidner, C., and D. Israel. Recognizing intended meaning and speaker's plans. In Proceedings of the Seventh International Joint Conference on Artificial Intelligence, pp. 203-208, Vancouver, B. C., 1981.

Simpson, c. A., and T. N. Navarro. Intelligibility of computer generated speech as a function of multiple factors. In Proceedings of the National Aerospace and Electronics Conference (NAECON), pp. 932-940, New York, May. IEEE, 1984.

Simpson, C. A., C. R. Coler, and E. M. Huff. Human factors of voice I/O for aircraft cockpit controls and displays. In Proceedings of the Workshop on Standardization for Speech I/O Technology, pp. 159-166, Gaithersburg, Md., March. national Bureau of Standards, 1982.

Simpson, C. A., M. E. McCauley, E. F. Roland, J. C. Ruth, and B. H. Williges. System design for speech recognition and generation. Human Factors, 27(2): 115-141, 1985

Small, D., and L. Weldon. An experimental comparison of natural and structured query languages. Human Factors, 25: 253-263, 1983.

Spitz, J. collection and analysis of data from real users: Implications for speech recognition/understanding systems. In Proceedings of the 4th DARPA Workshop on Speech and Natural Language, Asilomar, Calif., February. Defense Advanced Research Projects Agency, 1991.

Stallard, D., and R. Bobrow. Fragment processing in the DELPHI system. In Proceedings of the Speech and Natural Language Workshop, pp. 305-310, San Mateo, Calif., February. Morgan Kaufmann Publishers, Inc., 1992.

Street, R. L., Jr., R. M. Brady, and W. B. Putman. The influence of speech rate stereotypes and rate similarity on listeners' evaluations of speakers. Journal of Language and Social Psychology, 2(1): 37-56, 1983.

Streeter, L. A., D. Vitello, and S. A. Wonsiewicz. How to tell people where to go: comparing navigational aids. International Journal of Man-Machine Studies, 22:549-562, 1985.

Swider, R. F. Operational evaluation of voice command/response in an Army helicopter. In Proceedings of Military Speech Tech 1987, pp. 143-146, Arlington, Va. Media Dimensions, 1987.

Tanaka, S., D. K. wild, P. J. Seligman, W. E. Halperin, V. Behrens, and V. Putz-Anderson. Prevalence and Work-Relatedness of Self-Reported Carpal tunnel Syndrome Among U.S. Workers-Analysis of the Occupational Health Supplement Data of the 1988 national Health Interview Survey. National Institute of Occupational Safety and Health, and Centers for Disease Control and Prevention (Cincinnati). in submission.

Tennant, H. R., K. M. Ross, R. M. Saenz. C. W. Thompson, and J. R. Miller. Menu-based natural language understanding. In Proceedings of the 21st Annual Meeting of the Association for computational Linguistics, pp. 151-158, Cambridge, Mass., June 1983.

Thomas, J. Cl., M. B. Rosson, and M. Chodorow. Human factors and synthetic speech. In B. Shackel, ed., Proceedings of INTERACT'84, Amsterdam. Elsevier Science Publishers B. V. (North Holland), 1984.

Turner, J. A., M. Jarke, E. A. Stohr, Y. Vassiliou, and N. White. Using restricted natural language for data retrieval: A plan for field evaluation. In Y. Vassiliou, ed., Human Factors and Interactive computer systems, pp. 163-190. Ablex Publishing Corp., Norwood, N. J., 1984.

VanKatwijk, A. f., F. L. VanNes, H. C. Bunt, H. F. Muller, and F. F. Leopold. Naïve subjects interacting with a conversing information system. IPO Annual Progress Report, 14:105-112, 1979.

Visick, K., P. Johnson, and J. Long. The use of simple speech recognisers in industrial applications. In Proceeding of INTERACT '84 First IFIP conference on Human-Computer Interaction, London, U.K., 1984.

Voorhees, J. W., N. M. Bucher, E. M. Huff, C. A. Simpson, and D. H. Williams. voice interactive electronic warning system (views). In Proceedings of the IEEE/AIAA 5th Digital Avionics Systems conference, pp. 3.5.1-3.5.8, New York. IEEE, 1983.

Wahlster, W. User and discourse models for multimodal communication. In J. W. Sullivan and S. W. Tyler, eds., Intelligent User Interfaces, pp. 45-68. ACM Press Frontier Series. Addison Wesley Publishing Co., New York. 1991.

Weinstein, C. Opportunities for advanced speech processing in military computer-based systems Proceedings of the IEEE, 79(11): 1626-1641, November 1991.

Welkowitz, J., s. Feldstein, M. Finkelstein, and L. Aylesworth. Changes in vocal intensity as a function of interspeaker influence. Perceptual and Motor Skills, 10:715-718, 1972.

Williamson, J. T. Flight test results of the AFTI/F-16 voice interactive avionics program. In Proceedings of the American Voice I/O Society (AVIOS) 87 voice I/O Systems Applications Conference, pp. 335-345, Alexandria, Va., 1987.

Winograd, T. Understanding Natural Language. Academic Press, New York, 1972.

Winograd, T., and F. Flores. Understanding Computers and Cognition: A New Foundation for Design. Ablex Publishing co., Norwood, N. J., 1986.

Yamaoka, t., and H. Iida. Dialogue interpretation model and its application to next utterance prediction for spoken language processing. In Proceedings of

Eurospeech '91: 2nd European Conference on Speech communication and Technology. pp. 849-852, Genova, Italy. European Speech Communication Association, 1991.

Yato, F., T. Takezawa, S. Sagayama, J. Takami, H. Singer, N. Uratani, T. Morimoto, and A. Kurematus. International Joint Experiment Toward Interpreting Telephony (in Japanese). Technical Report, The Institute of Electronics, Information, and Communication Engineers, 1992.

Young, S. R., A. G. Hauptmann, W. H. Ward, E. T. Smith, and P. Werner. High level knowledge sources in usable speech recognition systems. Communications of the ACM, 32(2), February 1989.

Zoltan-Ford, E. Language Shaping and Modeling in Natural Language Interactions with Computers. PhD thesis, Psychology Department, Johns Hopkins University, Baltimore, Md., 1983.

Zoltan-Ford, E. Reducing variablility in natural-language interactions with computers. In M. J. Alluisi, S. de Groot, and E. A. Alluisi, eds., Proceedings of the Human Factors Society-28th Annual Meeting, vol. 2, pp. 768-772, San Antonio, Tex., 1984.

Zoltan-Ford, E. How to get people to say and type what computes can understand. International Journal of Man-Machine Studies, 34:527-547, 1991.

Zue, V., J. Glass, D. Goddeau, D. Goodine, L. Hirschman, M. Phillips, J. Polifroni, and S. Seneff. The MIT ATIS system: February 1992 progress report. In Fifth DARPA Workshop on Speech and Natural Language, San Mateo, Calif. Morgan Kaufmann Publishers, Inc., 1992.