



Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής

Ενότητα 2: Αρχιτεκτονική Συστημάτων
Φωνητικών Διεπαφών Χρήστη

Γεώργιος Κουρουπέτρογλου

koupe@di.uoa.gr



Περιεχόμενα ενότητας

Βασικές αρχιτεκτονικές συστημάτων φωνητικών διεπαφών χρήστη

Τι είναι φωνή;

Ο ήχος που παράγεται:

- από τους ανθρώπους και τα άλλα σπονδυλωτά
- χρησιμοποιώντας τους πνεύμονες και τις φωνητικές χορδές στον λάρυγγα.

Η φωνή δεν παράγεται πάντα ως ομιλία:

κλάμα, γουργούρισμα, άναρθρες φωνές μωρών, γέλιο, τραγούδι, κραυγή, μούγκρισμα, νιαούρισμα ενηλίκων

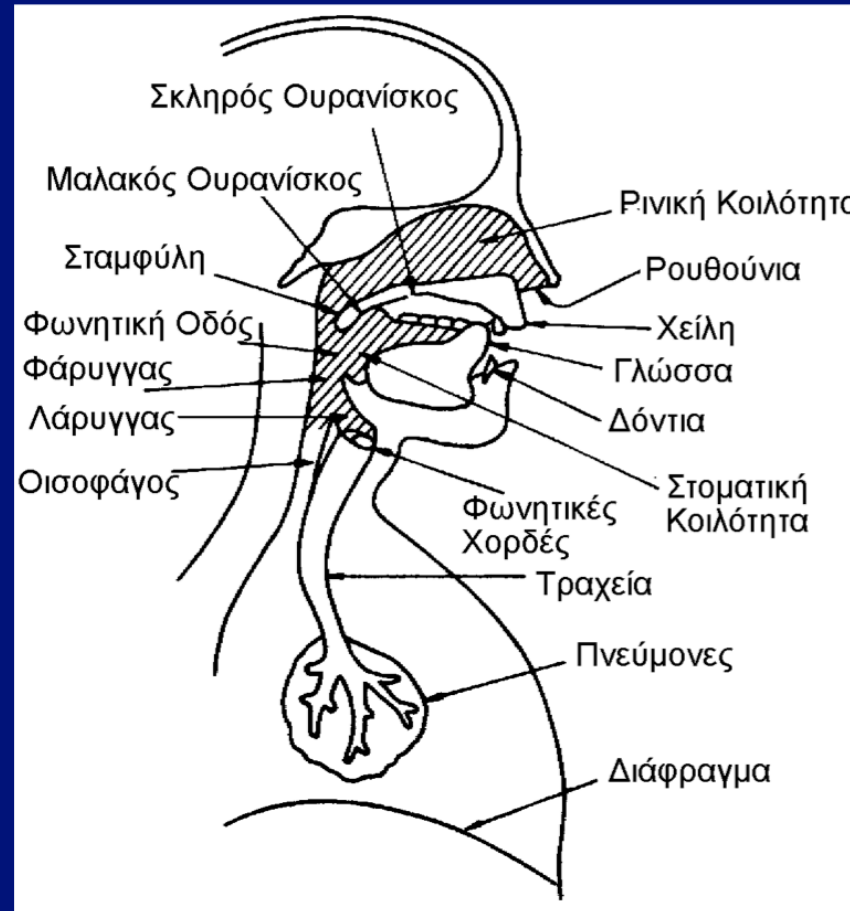
«Ο τραγουδιστής ελέγχει καλά τη φωνή του».

«Έχασε τη φωνή του».

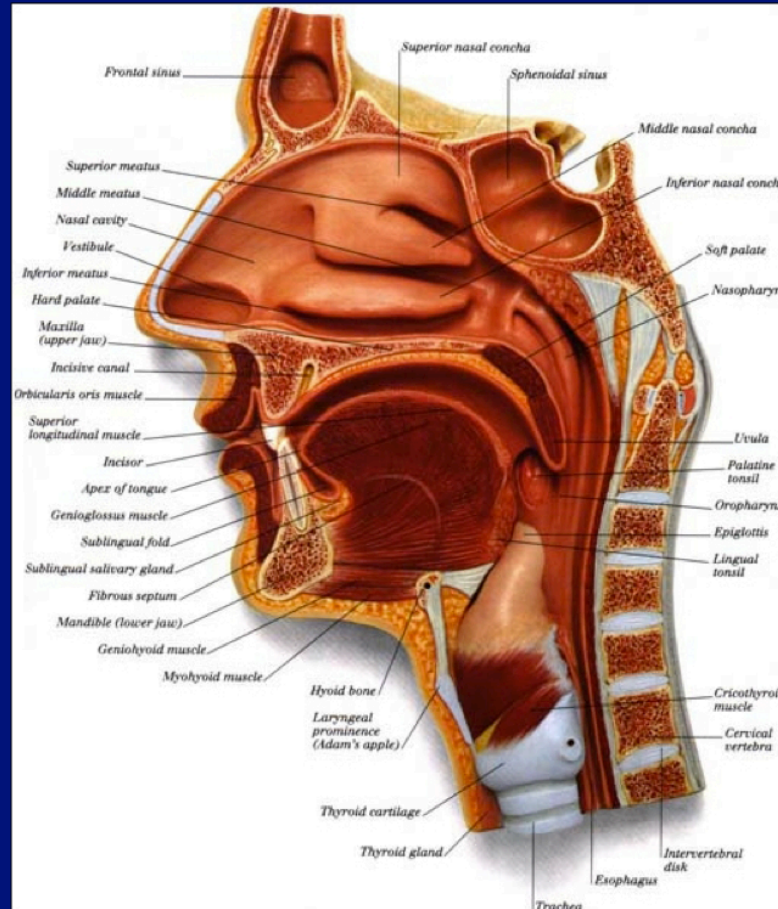
«Μια στριγκλίζουσα φωνή ακούστηκε πίσω μας».



Φωνητικό σύστημα ανθρώπου (1 από 2)



Φωνητικό σύστημα ανθρώπου (2 από 2)



Τι είναι ομιλία; (1 από 2)

- Ομιλία είναι ένας συγκεκριμένος τύπος φωνής.
- Ομιλία είναι προφορική επικοινωνία.
- Ομιλία είναι η διαδικασία της κατασκευής καθορισμένων φωνητικών ήχων που σχηματίζουν λέξεις με σκοπό να εκφράσουν σκέψεις και ιδέες.

«Η ομιλία του ήταν παραμορφωμένη».

«Διαπίστωσα μια ελαφρά διαφορετική χροιά στην ομιλία του».

«Ήσαν εντελώς άνετα μαζί, χωρίς να μιλούν».

Τι είναι ομιλία; (2 από 2)

- Φώνημα είναι η βασική μονάδα της ομιλίας ή το μικρότερο τμήμα ομιλίας που η αντικατάστασή του από ένα άλλο μπορεί να έχει σαν αποτέλεσμα αλλαγή στο νόημα:
 - “b” στην: “bike” ή “th” στην: “father”
 - “i” στις: λείπει, λύπη, λίπη, λοιπόν,

Γλωσσολογία, Φωνητική & Φωνολογία (1 από 2)

Η μελέτη του τρόπου που οι άνθρωποι παράγουν ήχους ομιλίας και ο τρόπος με τον οποίο χρησιμοποιούνται αυτοί στη γλώσσα αποτελεί μια χωριστή επιστημονική περιοχή που έχει ένα αρκετά ανεπτυγμένο θεωρητικό υπόβαθρο και αποτελεί μέρος της Γλωσσολογίας.

Χωρίζεται σε δυο κλάδους:

Γλωσσολογία, Φωνητική & Φωνολογία (2 από 2)

- **Φωνητική** (phonetics) ασχολείται με την παραγωγή, μετάδοση, αντίληψη και κατηγοριοποίηση των ήχων της ομιλίας
 - **αρθρωτική φωνητική** (articulatory phonetics) εξετάζει τη φύση των ήχων της ομιλίας βάσει της άρθρωσής τους,
 - **ακουστική φωνητική** (acoustic phonetics) στηρίζεται στην ανάλυση των ήχων της ομιλίας που παράγονται κατά την εκφώνηση
 - **αντιληπτική φωνητική** (auditory phonetics) εξετάζει τη διαδικασία της αντίληψης των ήχων της ομιλίας
- **Φωνολογία ή Φωνητική Γλωσσολογία** (phonology or linguistic phonetics): με αντικείμενο τη λειτουργία των ήχων ομιλίας στις διάφορες γλώσσες ή στα γλωσσικά συστήματα

Φωνήματα της Νέας Ελληνικής

Φωνήεντα: /i/ /u/
/ε/ /ο/
/α/

Σύμφωνα: /p/ /t/ /k/
/b/ /d/ /g/
/f/ /e/ /χ/
/v/ /D/ /γ/
/s/
/z/
/m/ /n/
/l/
/r/

Το διεθνές φωνητικό αλφάβητο

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill				r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	ʼ as in:
Ʉ Dental	ɗ Dental/alveolar	ɓ' Bilabial
ɠ (Post)alveolar	ɗ' Palatal	ɗ' Dental/alveolar
ɠ' Palatoalveolar	ɟ Velar	ɟ' Velar
ʙ Alveolar lateral	ɠ Uvular	ɠ' Alveolar fricative

SUPRASEGMENTALS

Primary stress: ˈ founaˈtʃən
 Secondary stress: ˌ
 Long: eː
 Half-long: eˑ
 Extra-short: e̥
 Syllable break: i.ækt
 Minor (foot) group: ˌ
 Major (intonation) group: ː
 Linking (absence of a break): ˌ

TONES & WORD ACCENTS

LEVEL: ˈ (high), ˌ (low)
 ACCENTS: ˥ (Rising), ˦ (Falling), ˧ (High rising), ˨ (Low rising), ˩ (Rising-falling), etc.
 CONTOUR: ˨˩˥ (Global rise), ˨˩˨˥ (Downstep), ˨˩˨˨˥ (Upstep), ˨˩˨˩˥ (Global fall)

VOWELS

Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ɱ Voiceless labial-velar fricative
 ɰ Voiced labial-velar approximant
 ɥ Voiced labial-palatal approximant
 ɸ Voiceless epiglottal fricative
 ʕ Voiced epiglottal fricative
 ʡ Epiglottal plosive

ɤ Alveolo-palatal fricatives
 ɺ Alveolar lateral flap
 ɥ Simultaneous ɸ and X
 Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary: k͡p t͡s

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. ɰ̥.

◌̥ Voiceless	◌̤ Breathily voiced	◌̦ Dental	◌̧ Apical
◌̨ Voiced	◌̩ Creaky voiced	◌̪ Laminar	◌̫ Nasalized
◌̬ Aspirated	◌̭ Linguolabial	◌̮ Labialized	◌̯ Palatalized
◌̰ More rounded	◌̱ Less rounded	◌̲ Velarized	◌̳ Pharyngealized
◌̴ Retraacted	◌̵ Centralized	◌̶ Mid-centralized	◌̷ Syllabic
◌̸ Non-syllabic	◌̹ Rhoticity		



Αλλόφωνα

Ως αλλόφωνο ενός φωνήματος ορίζεται ένα άλλο φώνημα το οποίο όταν αντικαταστήσει το πρώτο σε μια λέξη έχει ως αποτέλεσμα το να μην αλλάζει η σημασία ή η ταυτότητα της λέξης, παρά μόνο να δίνει μια σαφή διαφορά στο άκουσμα της προφοράς.

Διαφορετικής αιτιολογίας, όπως:

- **συνάρθρωση** (coarticulation) π.χ το φώνημα /κ/ στις περιπτώσεις: “κάπα” και “δεν κοιτάω”
- ιστορικές ή τοπικές εξελίξεις: /ι/ στις λέξεις “φίδι”, “βιάζομαι” και “πιάνω” ή του /λ/ στις λέξεις “λάμα” και “ήλιος”

Αλλόφωνα των συμφώνων (1 από 3)

Σύμφωνο	Αλλόφωνα σε σύμβολα IPA	Παραδείγματα	
β	v	βήτα	
γ	Δ j	γάμα γέρος	
δ	D	δέλτα	
ζ	z	ζήτα	
θ	T	θήτα	
κ	k c g o	κάπα κερνώ εκβολή δεν κοιτάω	
λ	l '	λάμα ήλιος	
μ	m M	μη μια	

Αλλόφωνα των συμφώνων (2 από 3)

Σύμφωνο	Αλλόφωνα σε σύμβολα IPA	Παραδείγματα	
ν	n ψ N m	ναός χτένια δεν κάνει δεν πάω	
ξ	ks	ξανθιά	
π	p b	τρύπα την πηγή	
ρ	r	ράφι	
σ	s z	σανός σγουρός	
ς	s	νέος	
τ	t	τάφος	
φ	f	φάρος	

Αλλόφωνα των συμφώνων (3 από 3)

Σύμφωνο	Αλλόφωνα σε σύμβολα IPA	Παραδείγματα
χ	Χ C	χάρος χέρι
ψ	ps	ψάρι
μπ	b mb	μπάλα εμπαίζω
ντ	d md	ντύνω εντολή
γκ	g f	γκρεμός γκίνια

Τεχνολογία επικοινωνίας με ομιλία (1 από 3)

Διαπροσωπική επικοινωνία (σε πραγματικό χρόνο)

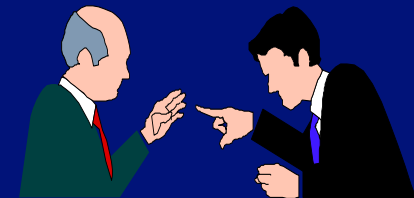
- Εκ του πλησίον ομιλία

Πρόβλημα παραγωγής ομιλίας 1.2%

Πρόβλημα αποδοχής ομιλίας 2.7%

Βοηθήματα παραγωγής ομιλίας

Βοηθήματα επικοινωνίας για κωφούς

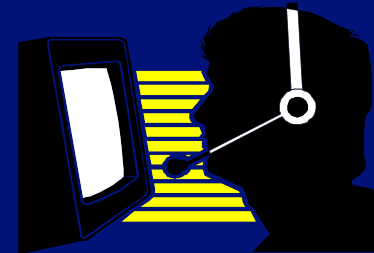


Τεχνολογία επικοινωνίας με ομιλία (2 από 3)

- Σε απόσταση ομιλία

Δίκτυο (π.χ. ψηφιακής τηλεφωνίας, κινητή τηλεφωνία, internet, ISDN)

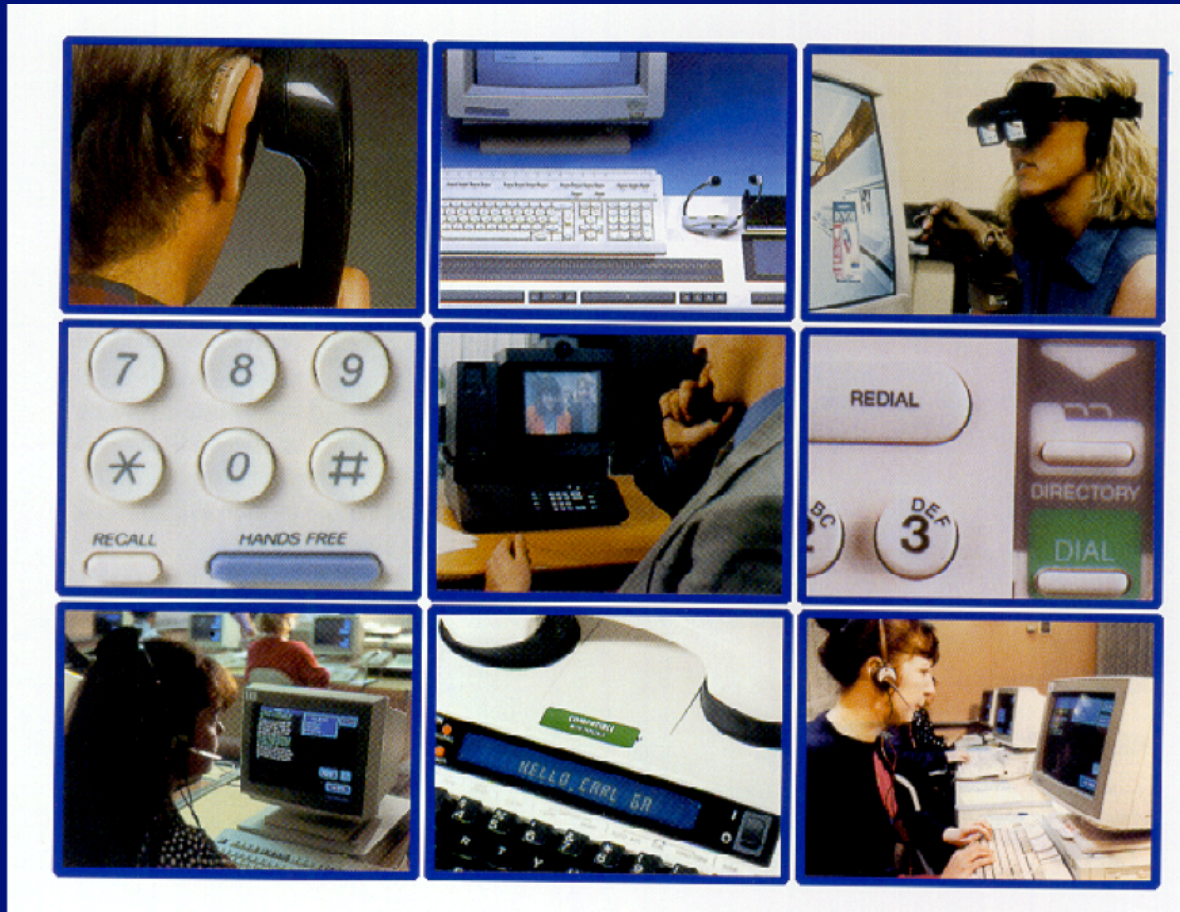
- Ψηφιοποίηση ομιλίας
- Συμπύεση ομιλίας
- Μετάδοση ομιλίας
- Αναπαραγωγή ομιλίας



Τεχνολογία επικοινωνίας με ομιλία (3 από 3)

Ποσοστά ΑμεΑ στο γενικό πληθυσμό στην ΕΕ	
Κινητικότητα	
Χρήστες αμαξιδίου	0.4
Δεν μπορούν να βαδίσουν χωρίς βοήθημα	5.0
Δεν μπορούν να χρησιμοποιήσουν τα δάκτυλα ή τον καρπό	0.3
Δεν μπορούν να χρησιμοποιήσουν έναν βραχίονα	0.1
Μειωμένη ισχύς χεριών	2.8
Μειωμένος συντονισμός χεριών	1.4
Ομιλίας (χωρίς ομιλία)	0.3
Διανοητικής	
Γλωσσική αναπηρία	0.7
Δυσλεξία	3.1
Νοητική υστέρηση	3.8
Ακοής	
Κώφωση	0.1
Βαρηκοΐα	10.0
Όρασης	
Τύφλωση	0.1
Χαμηλή όραση	1.5

Βοηθήματα διαπροσωπικής επικοινωνίας



Τα Συστήματα Εναλλακτικής και Επαυξητικής Επικοινωνίας

Αφορούν:

- Άτομα που παρουσιάζουν:
 - αλαλία, αναρθρία,
 - δυσνόητη ομιλία, δυσαρθρία,
 - προβλήματα κατανόησης και χρήσης του λόγου
- Ομάδες ατόμων με:
 - κινητικά προβλήματα, νοητικά προβλήματα,
 - αφασία, αυτισμό, σύνδρομο Rett,
 - αναπτυξιακά προβλήματα λόγου

Άτομα χωρίς όραση

- Πρόβλημα στην πρόσβαση στις έντυπες και ηλεκτρονικές πληροφορίες αλλά και στις υπηρεσίες του διαδικτύου
- Διάλογος με ομιλία:
 - Φωνητικές εντολές – Φωνητική πλοήγηση
 - Μετατροπή εγγράφων σε ομιλία

Παραδείγματα Περιπτωσιακής Απώλειας Λειτουργιών (1 από 2)

- Λειτουργία χωρίς όραση: απασχολημένα μάτια (π.χ. οδήγηση), σκοτάδι
- Λειτουργία με χαμηλή όραση: μικρή οθόνη, περιβάλλον με καπνό, ξεχασμένα γυαλιά
- Λειτουργία χωρίς ακοή: πολύ θορυβώδες περιβάλλον, απασχολημένη ακοή, επιβαλλόμενη ησυχία (βιβλιοθήκες, συναντήσεις)

Παραδείγματα Περιπτωσιακής Απώλειας Λειτουργιών (2 από 2)

- Λειτουργία με χαμηλή ακοή: θορυβώδες περιβάλλον
- Λειτουργία χωρίς ανάγνωση: άλλη γλώσσα, επισκέπτες/ τουρίστες, ξεχασμένα γυαλιά
- Λειτουργία χωρίς περιορισμένη νοητική ικανότητα: πανικόβλητος, επηρεασμένος από αλκοόλ
- Λειτουργία με μειωμένη ικανότητα χεριών: ζωηρό όχημα, διαστημικό όχημα

Τεχνολογία επικοινωνίας με ομιλία

- Διαπροσωπική επικοινωνία (όχι σε πραγματικό χρόνο)
 - Εκ του πλησίον -> Ανταλλαγή μηνυμάτων μέσω φωνής
 - Σε απόσταση -> Αλληλογραφία μέσω φωνής
- Επικοινωνία ανθρώπου-μηχανής



ομιλία (αναγνώριση) ->
<- ομιλία (σύνθεση)



κοινό τηλεφωνικό δίκτυο

αναγνώριση



ΟΜΙΛΙΑ

σύνθεση



Μιλώντας και ακούγοντας

- Η ομιλία και η ακοή είναι θεμελιώδεις για τις περισσότερες διαπροσωπικές αλληλεπιδράσεις.

“Close the door, please!”

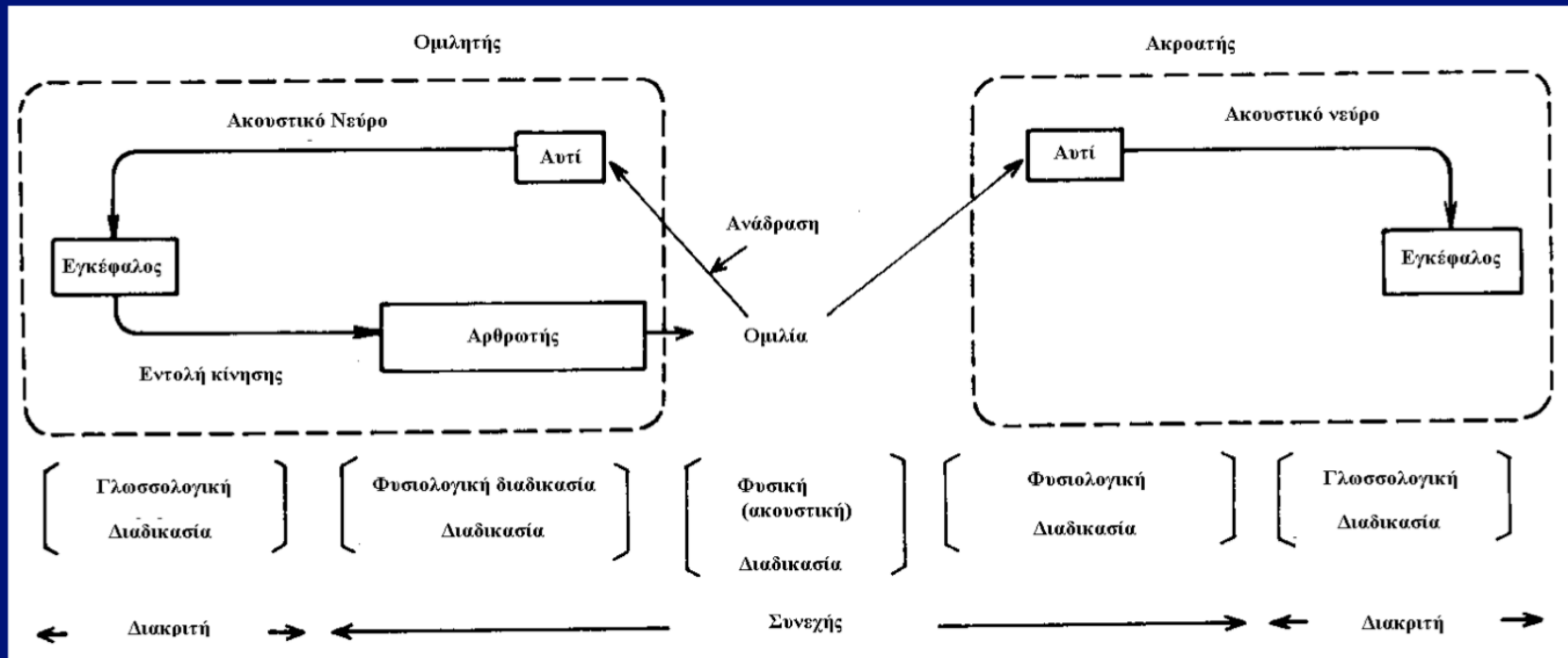
“For which city and state would you like the weather?”

“It is time to go to bed, Rona.”

“But I don’t want to go to bed now.”

“You have to go to school early tomorrow, therefore it is better you go to bed now”.

Διαδικασία Επικοινωνίας με Ομιλία



Προσωδία

- Ένταση ή ο δυναμικός τονισμός (intensity or stress),
- Μουσικότητα ή μουσικός τονισμός (pitch)
- Χρονισμός (timing)

Τονικές γλώσσες π.χ. κινεζική: 4 διαφορετικοί τόνοι (tones)

Χρήση προσωδίας:

- διάθεση του ομιλητή
- έμφαση - ρητορική
- τονισμός πολυσυλλαβικών λέξεων

Επιτονισμός (intonation) είναι η κύμανση της φωνής κατά την εκφώνηση μεγαλύτερων από τη λέξη τμημάτων του λόγου, είναι δηλαδή ο τόνος της πρότασης (έκφραση απορίας ή ερώτηση)

Άλλα χαρακτηριστικά (1 από 4)

- Διαφορετικές **προφορές** (accents) στην ίδια γλώσσα
- **Διάλεκτος** (dialect): ξεκάθαρες παραλλαγές της ίδιας γλώσσας, οι οποίες ομιλούνται από μία μεγάλη ομάδα ανθρώπων, συνήθως σε μία ορισμένη γεωγραφική περιοχή (συχνά χρήση εναλλακτικών λέξεων, γραμματικών αλλαγών)

- **Πλεονασμός** (redundancy)

*"Εφτά νομα σ' ένα δωμα
που να ξαπλω να κλείσεις μα»*

- **Το φαινόμενο του πάρτυ**

Two voices heard from the same location (mono) and then from two locations (stereo).

- **Οπτικές πληροφορίες – χειρονομίες - χειλεανάγνωση**



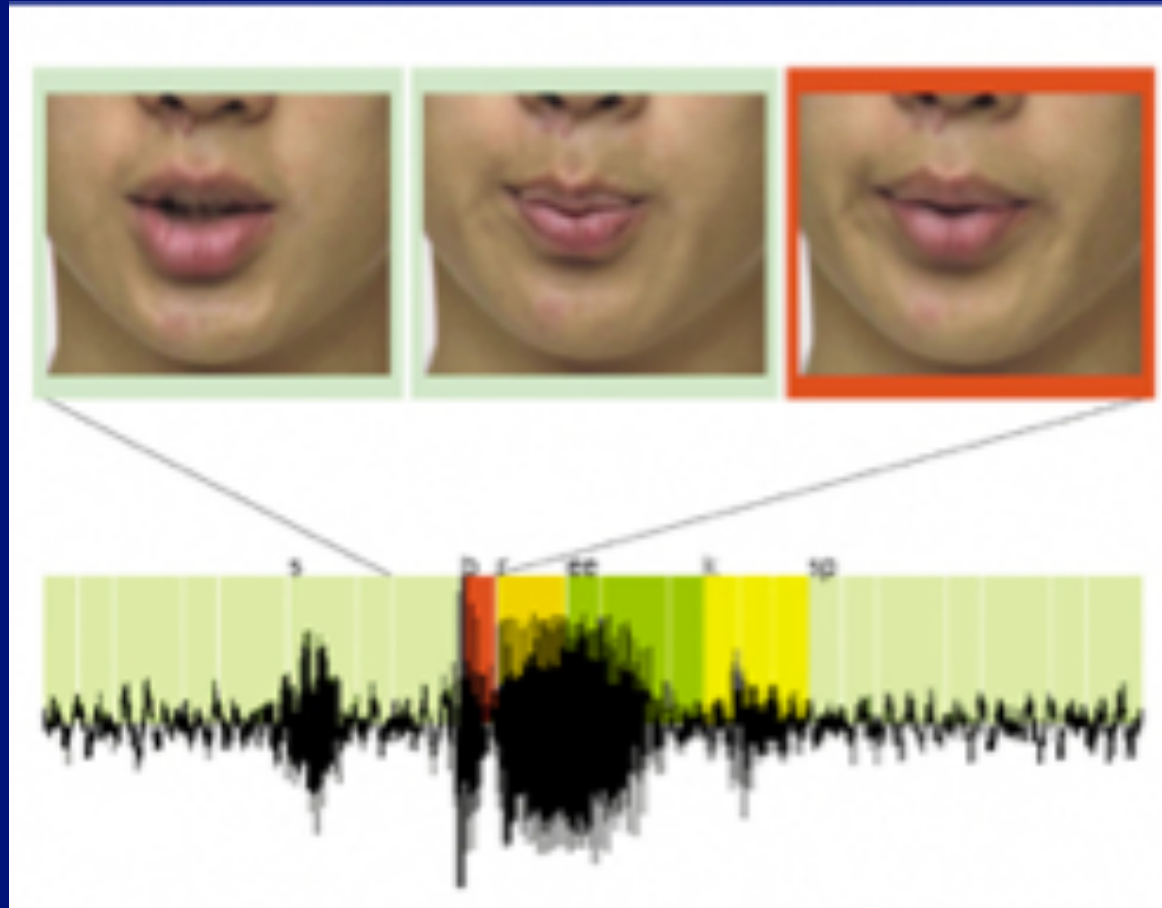
Άλλα χαρακτηριστικά (2 από 4)



Άλλα χαρακτηριστικά (3 από 4)



Άλλα χαρακτηριστικά (4 από 4)



Πληροφορίες που μεταφέρει το σήμα της ομιλίας

- Ποια φυσική γλώσσα
- Γλωσσολογική – το νόημα
- Ο τόνος ή το ύφος του ομιλητή
- Ποιος μιλά – ταυτότητα ομιλητή
- Η συναισθηματική κατάσταση του ομιλητή
- Παθολογία (π.χ. αλλαγή ομιλίας ύστερα από θυρεοειδεκτομή ή λαρυγγίτιδα)

Επεξεργασία Ομιλίας

- Ψηφιοποίηση ομιλίας
- Αναγνώριση ομιλίας
- Σύνθεση ομιλίας
- Αναγνώριση ομιλητή
- Κωδικοποίηση ομιλίας

Επεξεργασία Φωνής

«Επεξεργασία φωνής είναι ένας τρόπος χρήσης της τεχνολογίας ομιλίας ή των τεχνολογιών επεξεργασίας ομιλίας για την επίτευξη της διάδοσης και ανάκλησης των πληροφοριών.»

Η επεξεργασία φωνής αποτελείται από μια ομάδα τεχνολογιών που συνδυάζονται με διάφορους τρόπους για την επίτευξη συγκεκριμένων λειτουργιών. Συχνά αλληλεπιδρά με άλλα περιβάλλοντα, όπως τράπεζες δεδομένων, τηλεφωνικά συστήματα και μηχανές τηλεμοιοτυπίας σαν μέρος ενός μεγαλύτερου τηλεπικοινωνιακού συστήματος διαχείρισης πληροφοριών.

Λειτουργίες Επεξεργασίας Φωνής

- Αυτόματη αναγγελία
- Ταχυδρομείο ομιλίας (voice mailing)
- Εκτέλεση ή διεξαγωγή εντολής
- Παροχή πληροφοριών
- Πολλαπλές λειτουργίες

Λειτουργία των Φωνητικών οργάνων κατά την παραγωγή ομιλίας (1 από 3)



Λειτουργία των Φωνητικών οργάνων κατά την παραγωγή ομιλίας (2 από 3)



Λειτουργία των Φωνητικών οργάνων κατά την παραγωγή ομιλίας (3 από 3)

Κινούμενα σχέδια που δείχνουν πως τα ηχητικά κύματα μεταφέρονται από το εξωτερικό αυτί στα υγρά του εσωτερικού αυτιού και τη βασική μεμβράνη (όργανο του Corti).



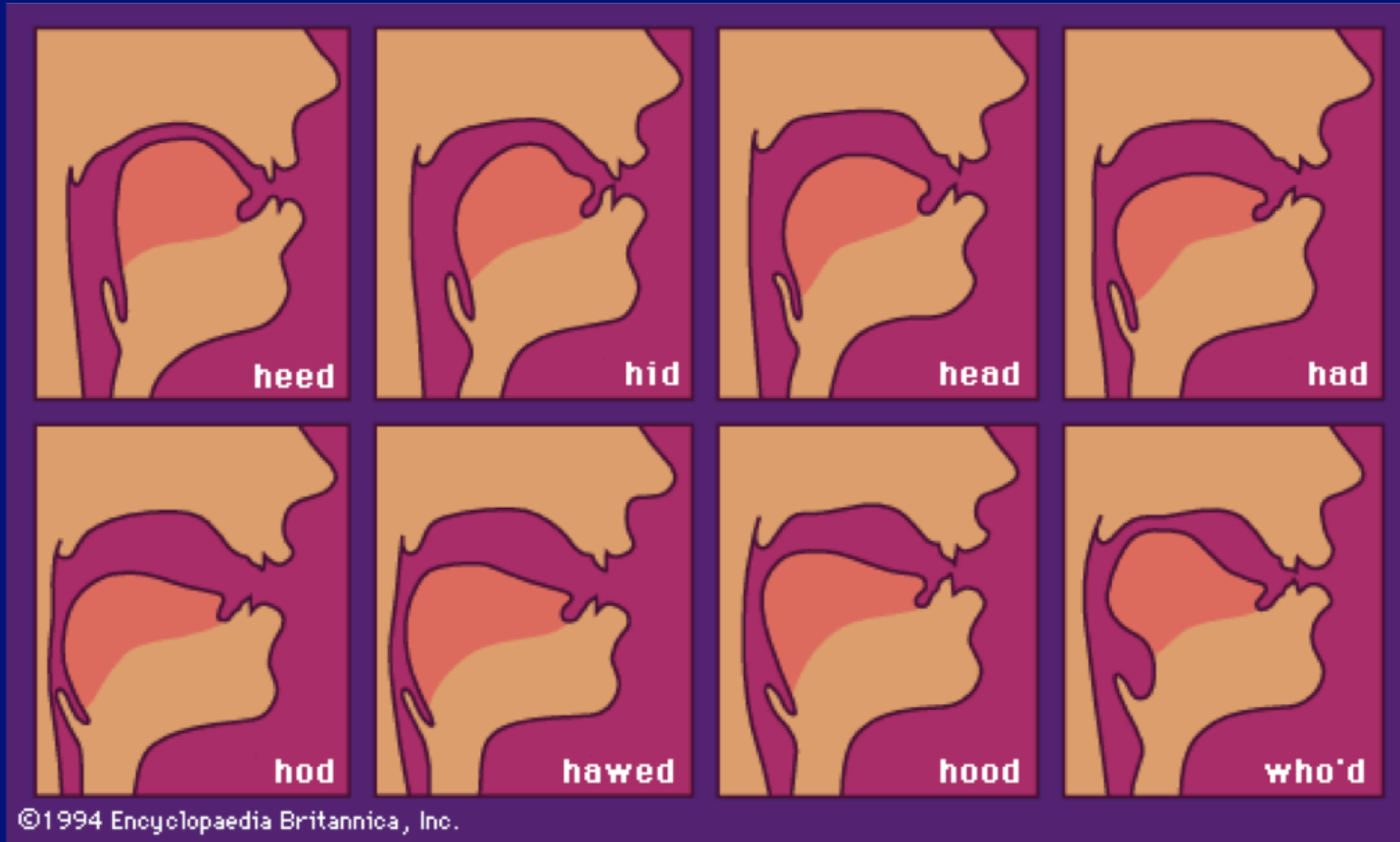
Κίνηση αρθρωτών κατά την παραγωγή ήχων ομιλίας

<http://www.uiowa.edu/~acadtech/phonetics/english/frameset.html>

Φωνολογία στην Αμερικανική Αγγλική



Μέσες θέσεις αρθρωτών κατά την παραγωγή φωνηέντων

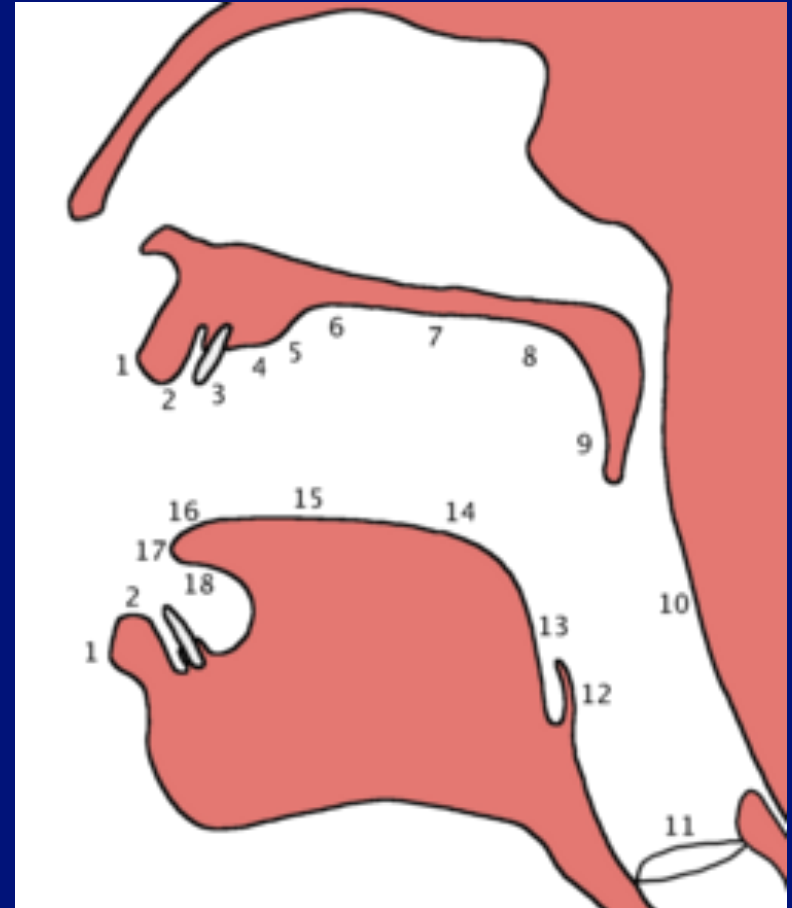


Οι θέσεις της άρθρωσης (1 από 3)

Είναι τα σημεία επαφής, στα οποία δημιουργείται μια **εμπόδιση** (φράξιμο) στη φωνητική οδό μεταξύ ενός **ενεργού** (μετακινούμενου) **αρθρωτή** (τυπικά κάποιο μέρος της γλώσσας) και ενός **παθητικού** (στατικού) **αρθρωτή** (τυπικά κάποιο σημείο της οροφής του στόματος).

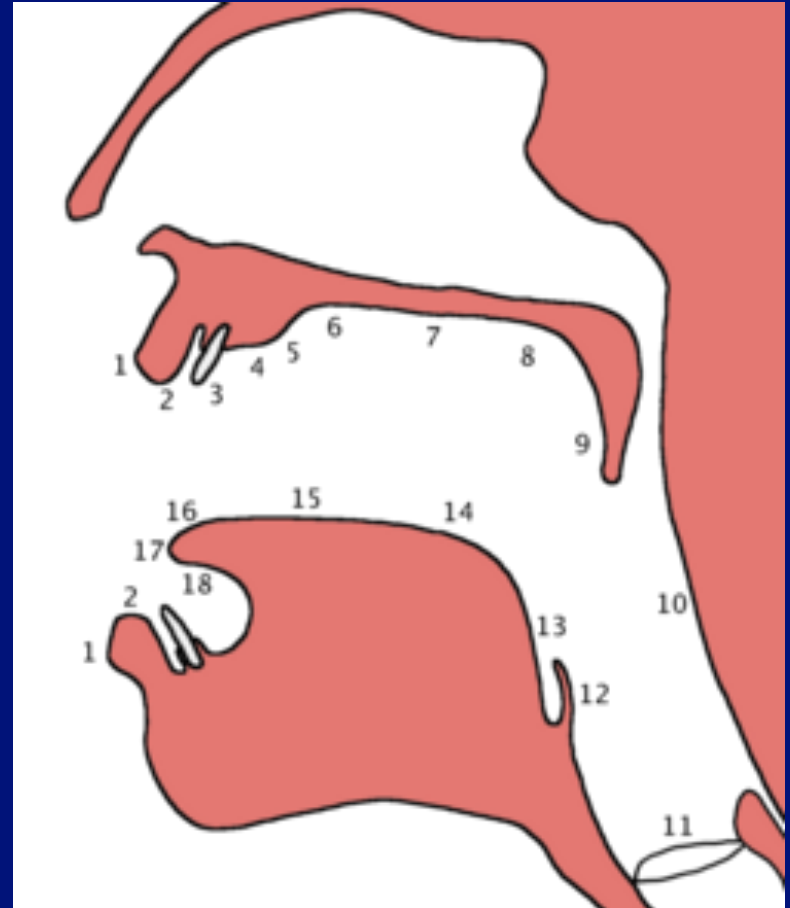
Οι θέσεις της άρθρωσης (2 από 3)

1. εξω-χειλικό (Exo-labial),
2. ενδο-χειλικό (Endo-labial),
3. Οδοντικά (Dental),
4. Φατνιακό (Alveolar),
5. μετα-φατνιακό (Post-alveolar),
6. προ-φατνιακό (Pre-palatal),
7. Ουρανικό (Palatal),
8. Υπερωικό (Velar),
9. Σταφυλικό (Uvular)

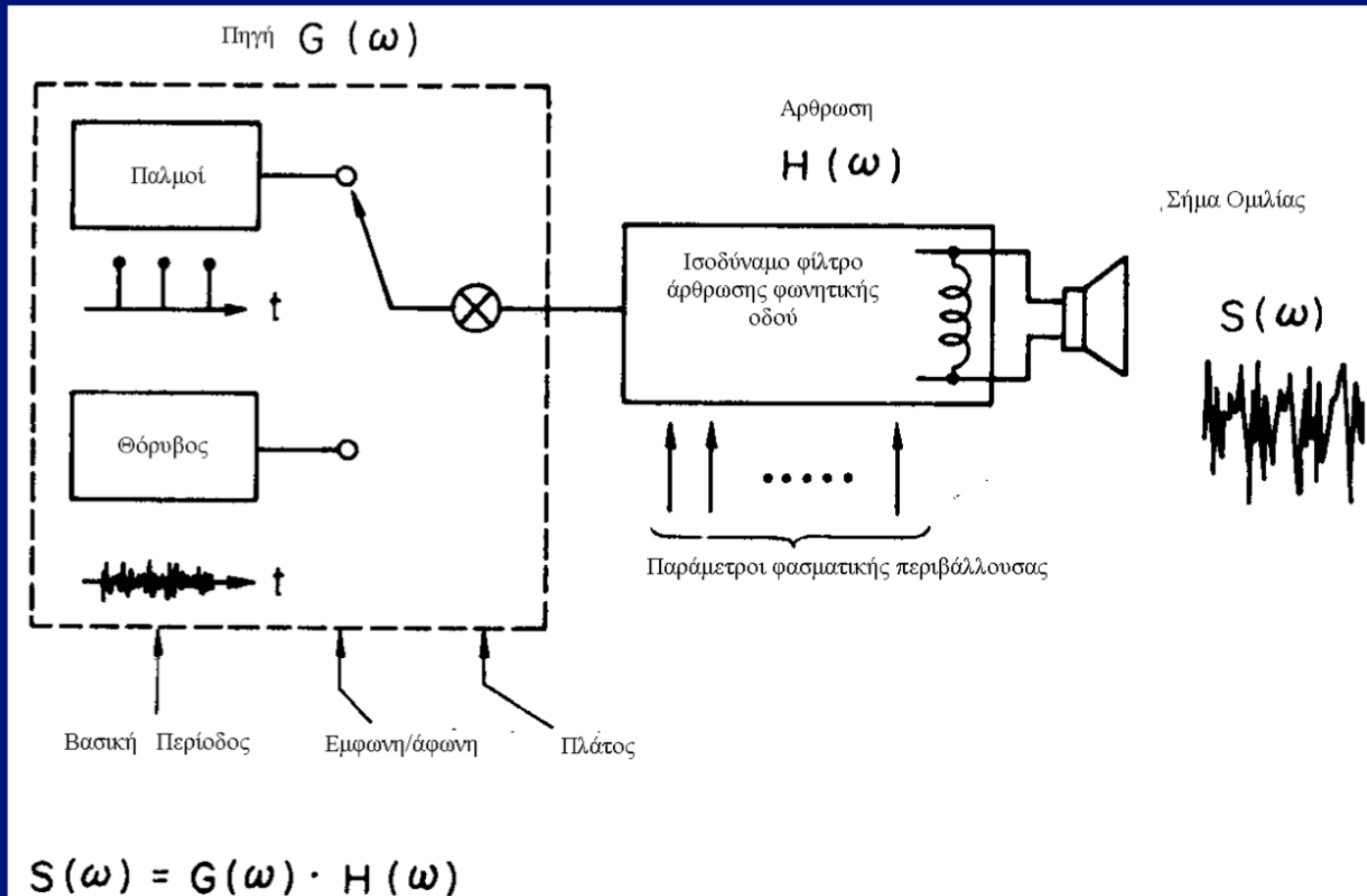


Οι θέσεις της άρθρωσης (3 από 3)

- 10. Φαρυγγικό (Pharyngeal),
- 11. Γλωττιδικό (Glottal),
- 12. Επιγλωττικό (Epiglottal),
- 13. Ριζικό (Radical),
- 14. μετα-ραχιαίο (Postero-dorsal),
- 15. Προ-ραχιαίο (Antero-dorsal),
- 16. Στρωματικό (Laminal),
- 17. Κορυφαίο (Apical),
- 18. Υποκορυφαίο (Sub-apical)



Σύνθεση Ομιλίας: Μοντέλο γραμμικά διαχωρίσιμου ισοδύναμου κυκλώματος του μηχανισμού παραγωγής ομιλίας



Βασική αρχιτεκτονική συστήματος μετατροπής κειμένου σε ομιλία

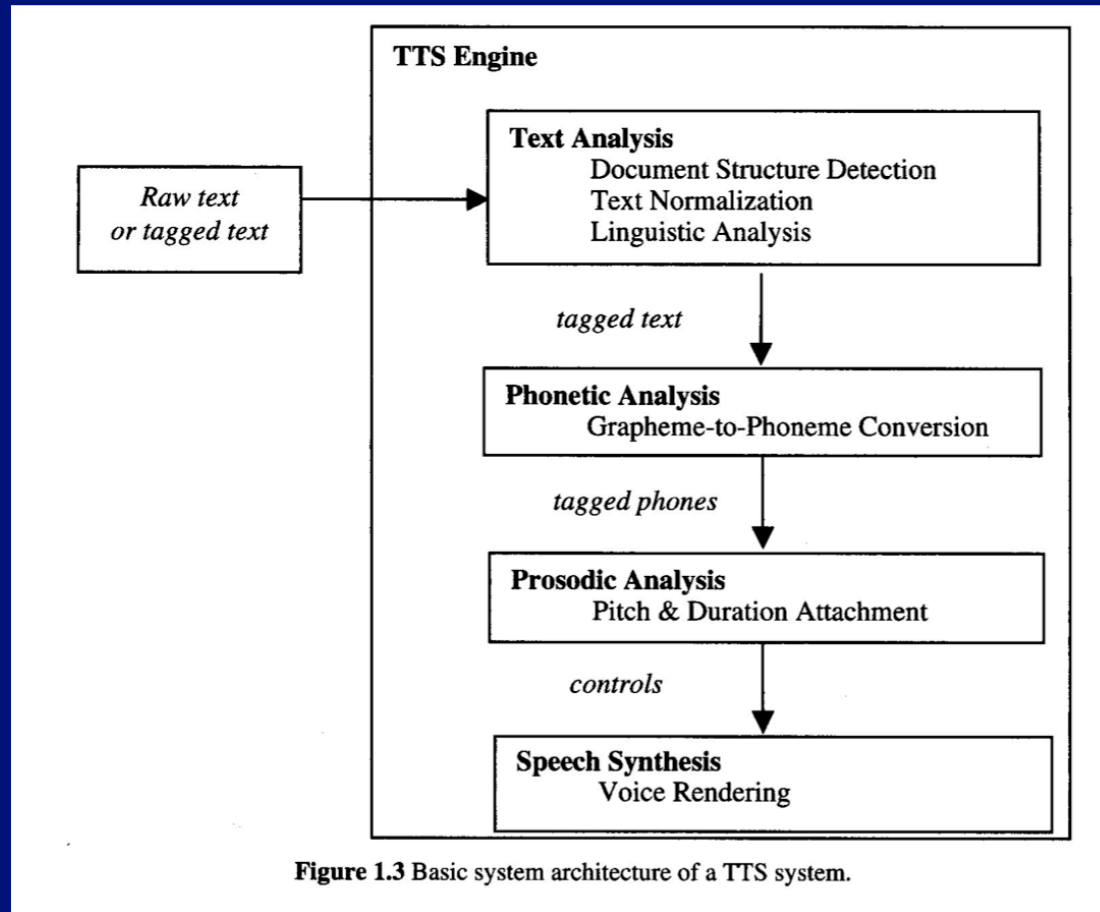
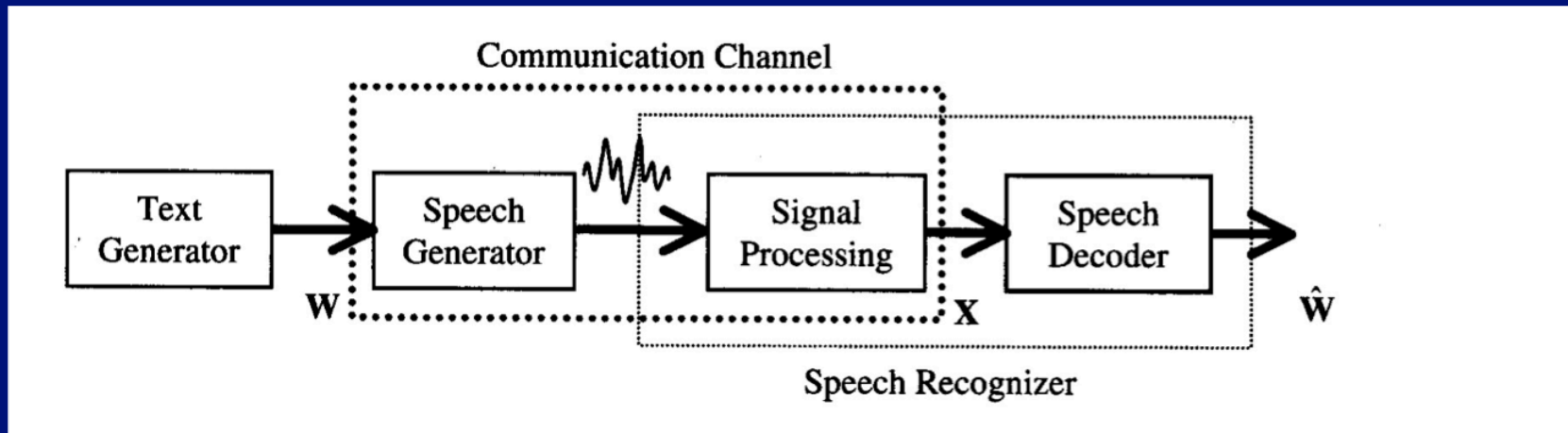


Figure 1.3 Basic system architecture of a TTS system.

Μοντέλο πηγής-καναλιού για ένα σύστημα αναγνώρισης ομιλίας



Βασική αρχιτεκτονική συστήματος αναγνώρισης ομιλίας (1 από 3)

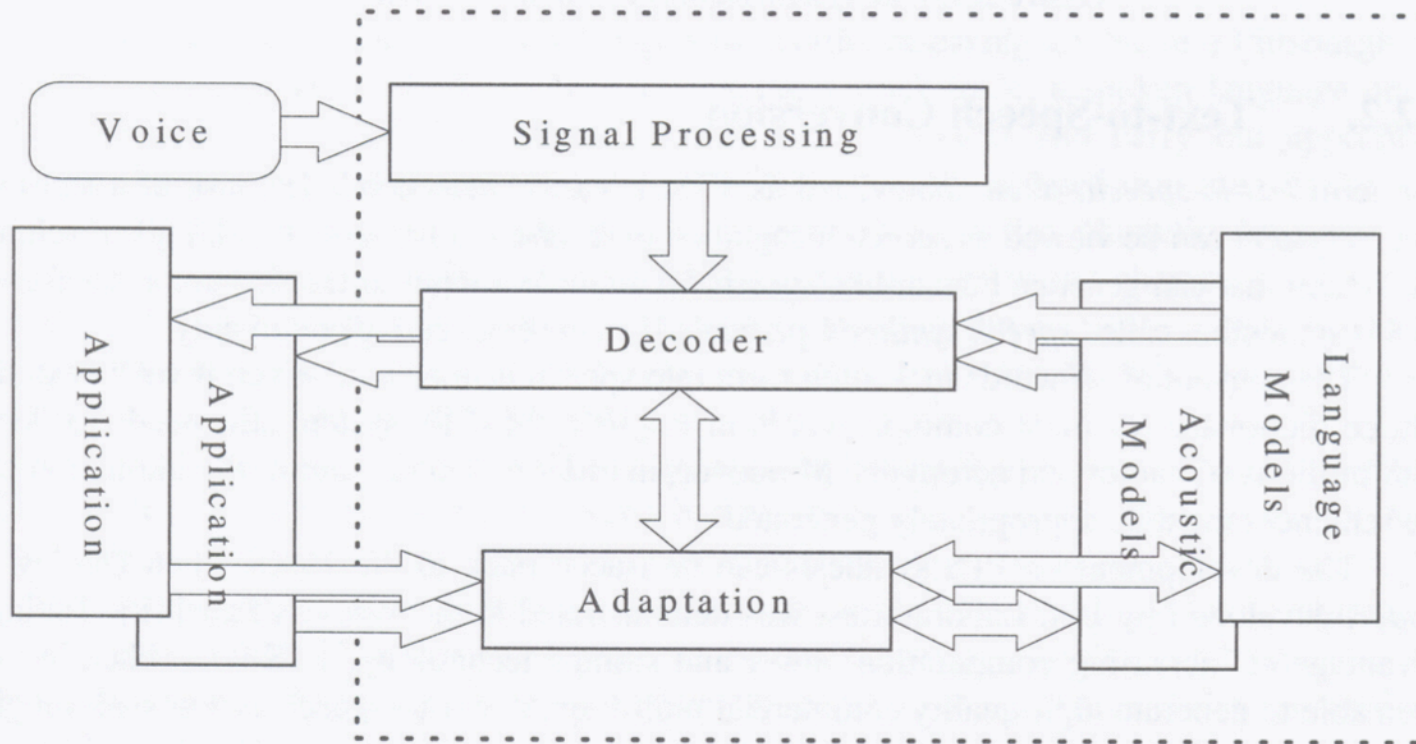


Figure 1.2 Basic system architecture of a speech recognition system [12].

Βασική αρχιτεκτονική συστήματος αναγνώρισης ομιλίας (2 από 3)

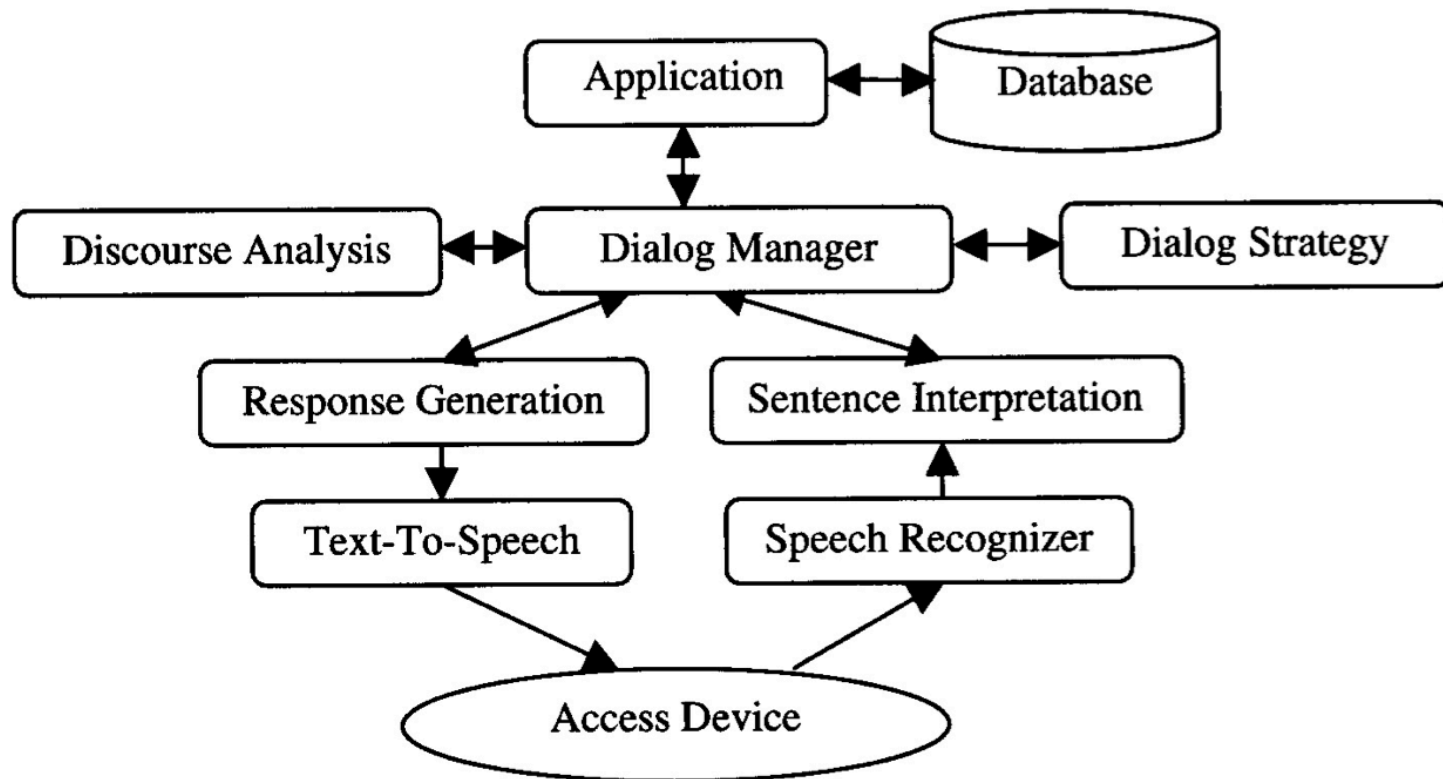


Figure 1.4 Basic system architecture of a spoken language understanding system.

Βασική αρχιτεκτονική συστήματος αναγνώρισης ομιλίας (3 από 3)

Table 1.1 Word error rate comparisons between human and machines on similar tasks.

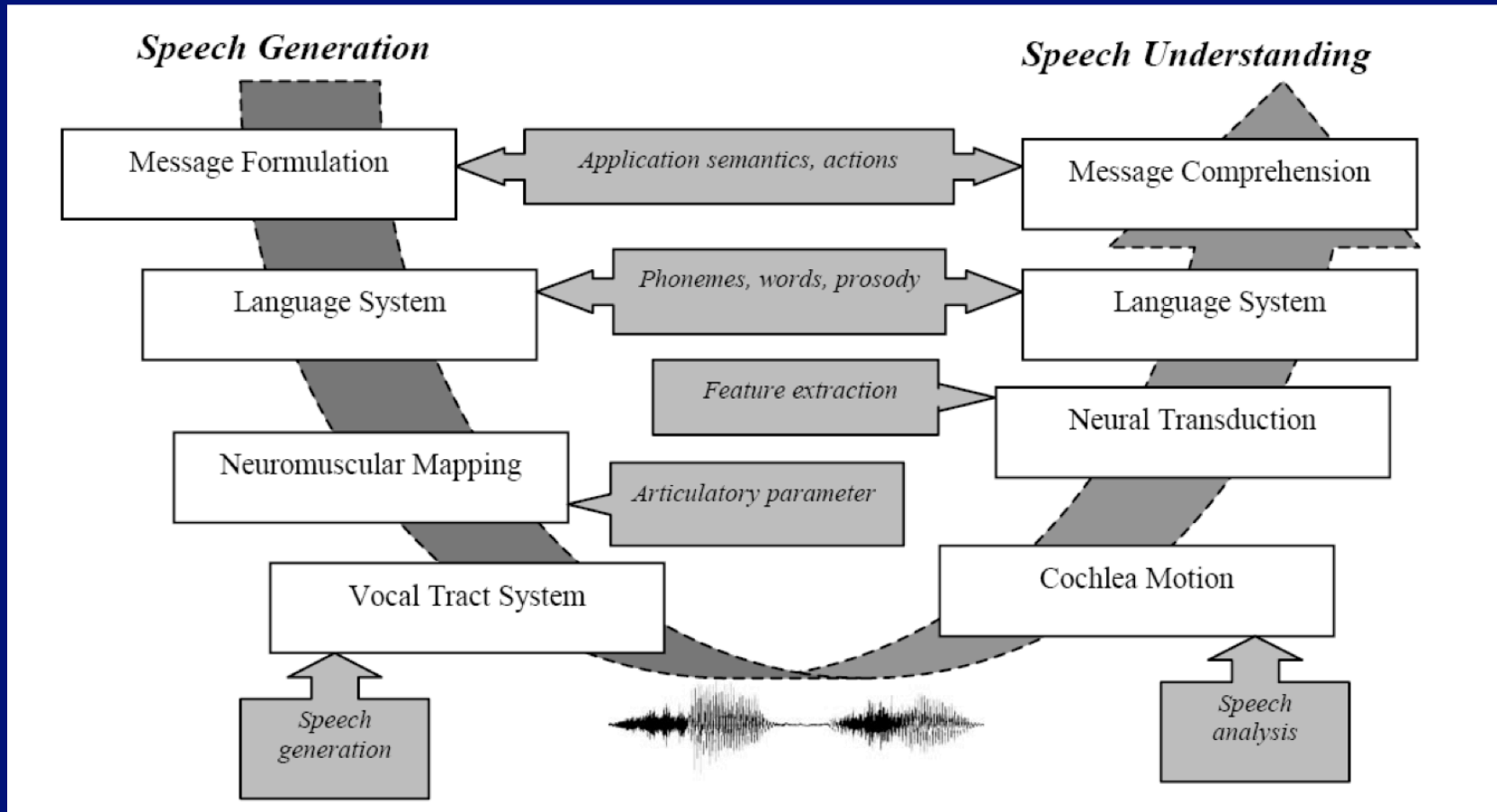
Tasks	Vocabulary	Humans	Machines
Connected digits	10	0.009%	0.72%
Alphabet letters	26	1%	5%
Spontaneous telephone speech	2000	3.8%	36.7%
WSJ with clean speech	5000	0.9%	4.5%
WSJ with noisy speech (10-db SNR)	5000	1.1%	8.6%
Clean speech based on trigram sentences	20,000	7.6%	4.4%

WSJ: Wall Street Journal



Δημιουργία και κατανόηση ομιλίας

(τα γκρι κουτιά υποδεικνύουν τα αντίστοιχα υποσυστήματα ενός συστήματος επεξεργασίας προφορικής γλώσσας)



Σύγκριση Γραφικών και Φωνητικών Διεπαφών (1 από 2)

Characteristic	GUI	VUI
Fonts	Choice of multiple sizes, colours, and types	Choice of multiple voices
Pictures	Present the picture to the user	Describe the picture with words (caption)
Backgrounds	Choice of multiple colours and patterns	Choice of background music or sounds
Menus	Large number of choices	Usually 7 ± 2 choices
Forms	Users may enter values in any order	Users enter values usually in predefined order
Prompting the user	Large menu of information and links	Limited number of prompts and options
User's response	Click a field and enter a value	Speak the menu choice or field value

Σύγκριση Γραφικών και Φωνητικών Διεπαφών (2 από 2)

Characteristic	GUI	VUI
Navigation	Click the link or use the keyboard to enter URL	Speak the option choice
Human memory	Screen acts as an extension to human memory	Caller is forced to use own mental memory
Dialog style	Mostly user-directed	Mostly application-directed
Input control	Apply the type constraints	Apply the grammar rules
Accuracy of input	Input is always recognized	Input may not be recognised correctly
Event handling	Display the event message	Present help as a prompt
Global commands	Available as a menu in a separate frame	Announced at the beginning of a session
Dates	Absolute dates are used (28.7.2003)	Relative dates are used ("next Monday")

Πλεονεκτήματα συνομιλίας με έναν υπολογιστή (1 από 2)

Οι χρήστες μπορούν να μιλήσουν και να ακούσουν έναν υπολογιστή:

- παρά τις όποιες φυσικές αναπηρίες όπως η τύφλωση ή χαμηλή φυσική δεξιότητα
- ξεπερνώντας τους περιορισμούς μικρών πληκτρολογίων και οθονών
- στην έλλειψη πληκτρολογίου σε μια συσκευή
- ενώ έχουν απασχολημένα τα χέρια και τα μάτια τους σε άλλες δραστηριότητες
- οποιαδήποτε στιγμή της ημέρας
- με άμεση σύνδεση χωρίς να τοποθετούνται σε αναμονή

Πλεονεκτήματα συνομιλίας με έναν υπολογιστή (2 από 2)

- χρησιμοποιώντας γλώσσες που δεν είναι δυνατόν να πληκτρολογηθούν
- για να διοχετεύσουν τα συναισθήματά τους
- για να έχουν πρόσβαση σε χρονικά μεταβαλλόμενα δεδομένα
- για να έχουν πρόσβαση σε δεδομένα εξαρτώμενα από τη θέση
- για να έχουν πρόσβαση σε δημόσιες και ιδιωτικές πληροφορίες
- για να ελέγξουν υπολογιστικές διεργασίες και δραστηριότητες

Παρά τις φυσικές αναπηρίες όπως η τύφλωση ή χαμηλή φυσική δεξιότητα

- Η ομιλία επιτρέπει στους χρήστες με ανικανότητες να έχουν πρόσβαση σε υπολογιστές.
- Οι χρήστες με χαμηλή φυσική δεξιότητα ή οι οποίοι δεν μπορούν να πληκτρολογήσουν, μπορούν να χρησιμοποιήσουν την ομιλία για να εισάγουν εντολές στον υπολογιστή.
- Χρήστες με εξασθένιση όρασης μπορούν να ακούν τον υπολογιστή καθώς μιλάει.
- Όταν γραφικές ή/και μηχανικές διεπαφές δεν αποτελούν επιλογή, οι χρήστες μπορούν να πραγματοποιήσουν συναλλαγές λέγοντας τι θέλουν να κάνουν και παρέχοντας τις κατάλληλες πληροφορίες.
- Αν ένα άτομο με κάποιου είδους ανικανότητα, μπορεί να μιλήσει και να ακούσει, αυτό το άτομο μπορεί να χρησιμοποιήσει έναν υπολογιστή.

Παράκαμψη των περιορισμών των μικρών πληκτρολογίων και οθονών

- Αν και οι συσκευές γίνονται μικρότερες σε μέγεθος δε συμβαίνει το ίδιο με τα δάχτυλά μας. Τα κουμπιά στο τηλεφωνικό καντράν συχνά είναι τόσο μικρά που οι άνθρωποι με χοντρά δάχτυλα πατούν με το ένα δάχτυλο δύο ή και περισσότερα κουμπιά.
- Οι μικρές οθόνες σε μερικά κινητά τηλέφωνα δεν φαίνονται καλά, ειδικά υπό δυνατό φωτισμό.
- Με την ομιλία οι χρήστες του τηλεφώνου μπορούν να παρακάμψουν το καντράν (εκτός ίσως από την περίπτωση που βρίσκονται σε ένα θορυβώδες περιβάλλον και θέλουν να καταχωρήσουν προσωπικά δεδομένα).
- Με την ομιλία και την ακοή οι χρήστες του τηλεφώνου μπορούν να παρακάμψουν τη μικρή οθόνη.

Χρήση συσκευής χωρίς πληκτρολόγιο (1 από 2)

- Πολλές συσκευές δεν έχουν πληκτρολόγιο ή άλλο πίνακα ελέγχου. Για παράδειγμα οι ηλεκτρικές κουζίνες, τα ψυγεία και τα κλιματιστικά δεν έχουν πληκτρολόγιο. Οι συσκευές αυτές διαθέτουν ένα μικρό πίνακα ελέγχου με μερικά κουμπιά και μια τηλεφωνική πλάκα.
- Ο φυσικός έλεγχος είναι απαραίτητος για ενέργειες όπως: η έναρξη και ο τερματισμός λειτουργίας μιας συσκευής, για τη ρύθμιση της ταχύτητας ή της θερμοκρασίας.
- Χωρίς την ομιλία ο χρήστης δεν μπορεί να διατυπώσει πολύπλοκες οδηγίες όπως «ανέβασε τη θερμοκρασία του φούρνου στους 350 C για 30' λεπτά, μετά χαμήλωσέ την στους 240 C για 15' λεπτά και μετά άσε το φούρνο ανοικτό σε χαμηλή θερμοκρασία».
- Χωρίς ομιλία η συσκευή δεν μπορεί να κάνει ερωτήσεις όπως «πότε ακριβώς πρέπει να ανάψει ο φούρνος το Σάββατο το πρωί;»

Χρήση συσκευής χωρίς πληκτρολόγιο (2 από 2)

- Οποιοσδήποτε σύνθετος διάλογος με κάποια οικιακή συσκευή απαιτεί χρήση ομιλίας
- Μικρές συσκευές με πληκτρολόγιο, όπως τα τηλέφωνα και τα PDA (Personal Digital Assistants) είναι χρήσιμες για την εισαγωγή αριθμητικών στοιχείων, αλλά δύσχρηστες και αργές για την εισαγωγή κειμένου.
- Ακόμη και τα PDA με πληκτρολόγιο τύπου QWERTY δεν είναι εύκολα.
- Οι χρήστες κρατούν τη συσκευή με το ένα χέρι και πληκτρολογούν με το δείκτη του άλλου χεριού.
- Η πληκτρολόγηση με τυφλό σύστημα είναι αδύνατη.
- Τα πλήκτρα είναι πολύ μικρά σε μέγεθος και κάποιοι χρήστες, λόγω του ότι έχουν χονδρά δάχτυλα, μπορούν να τα πατήσουν μόνο με στυλό.
- Ομιλία μπορεί επίσης να χρησιμοποιηθεί στα τηλέφωνα με κυκλικό καντράν, τα οποία δεν έχουν πληκτρολόγιο.

Την ώρα που τηλεφωνούν χρησιμοποιούν τα χέρια και τα μάτια (1 από 2)

- Οι ικανότητες ομιλίας και ακοής είναι ιδιαίτερα χρήσιμες όταν τα μάτια και τα χέρια του χρήστη είναι απασχολημένα.
- Οι οδηγοί δεν πρέπει να απομακρύνουν το βλέμμα τους από το δρόμο και τα χέρια τους από το τιμόνι. Αν πρέπει να χρησιμοποιήσουν κατά την οδήγηση Η/Υ, η διεπαφή θα πρέπει να βασίζεται μόνο σε ομιλία.
- Οι χειριστές μηχανών θα πρέπει επίσης να χρησιμοποιούν ομιλία για να επικοινωνούν με τον υπολογιστή, επειδή όταν εργάζονται με τη μηχανή χρειάζονται τα χέρια τους για να χειρίζονται τον πίνακα ελέγχου και τα μάτια τους επικεντρωμένα στις εργασίες που επιτελεί το μηχάνημα.

Την ώρα που τηλεφωνούν χρησιμοποιούν τα χέρια και τα μάτια (2 από 2)

- Οι μητέρες και ο καθένας που κρατάει ένα παιδί στην αγκαλιά του θα εξυπηρετούνταν ιδιαίτερα να συνομιλεί με την ιστοσελίδα ενός γιατρού ή μια ιστοσελίδα ιατρικής φροντίδας.
- Δεδομένου ότι κάποιος μπορεί να ακούει και να μιλάει σε άλλους ενώ δουλεύει, τότε μπορεί επίσης να ακούει και να μιλάει ταυτόχρονα στον Η/Υ.

Οποιαδήποτε ώρα της ημέρας

- Πολλές τηλεφωνικές γραμμές βοήθειας και ρεσεψιονίστ είναι διαθέσιμοι μόνο τις ώρες εργασίες.
- Μέσω των υπολογιστών θα ήταν δυνατή η αυτοματοποίηση υπηρεσιών όπως η παραλαβή κλήσεων, η παροχή πληροφοριών και η απάντηση ερωτήσεων.
- Όσοι τηλεφωνούν θα μπορούν να χρησιμοποιούν σε 24ωρή βάση, 7 ημέρες την εβδομάδα τέτοιου είδους αυτοματοποιημένες υπηρεσίες, μιλώντας και ακούγοντας έναν απομακρυσμένο υπολογιστή.
- Κάποιος που διαθέτει τις ικανότητες της ομιλίας και της ακοής μπορεί να αλληλεπιδράσει με ένα Η/Υ οποιαδήποτε ώρα της ημέρας και της νύχτας.

Με άμεση σύνδεση χωρίς αναμονή

- Όσοι τηλεφωνούν εκνευρίζονται όταν ακούν τη φράση «το τηλεφώνημά σας είναι πολύ σημαντικό για εμάς», επειδή αυτό σημαίνει ότι θα πρέπει να περιμένουν.
- Η φράση «Ευχαριστούμε που περιμένετε. Όλες μας οι γραμμές είναι κατειλημμένες» σημαίνει πρόσθετη αναμονή.
- Όταν όμως χρησιμοποιείται η ομιλία κατά την αλληλεπίδραση με μια εφαρμογή, δεν υπάρχει αναμονή. Ο υπολογιστής απαντά αμέσως. (Βέβαια κάποιες πύλες μεταξύ τηλεφωνικών γραμμών και υπολογιστών, καθώς και μεταξύ κάποιων υπολογιστών μεταξύ τους είναι δυνατόν να υπερφορτωθούν και σε αυτή την περίπτωση να υπάρξουν καθυστερήσεις. Αυτό όμως συμβαίνει λιγότερο συχνά από την περίπτωση της αναμονής για συνομιλία με φυσικό πρόσωπο).
- Επειδή οι περισσότεροι πελάτες μπορούν να εξυπηρετηθούν από την εφαρμογή, οι ανθρώπινοι χειριστές ελευθερώνονται και μπορούν να εξυπηρετούν τα πιο δύσκολα αιτήματα.

Χρήση γλωσσών που δεν υποστηρίζονται από το πληκτρολόγιο

- Οι ασιατικές γλώσσες δεν υποστηρίζονται από το παραδοσιακό πληκτρολόγιο τύπου QWERTY.
- Προκειμένου να αναγκάζονται οι χρήστες ασιατικών γλωσσών να μεταφράζουν τις λέξεις και τις φράσεις τους σε φωνητικούς ήχους και έπειτα να τις πληκτρολογούν, πατώντας τα αντίστοιχα πλήκτρα στο QWERTY, η καλύτερη λύση είναι να μιλούν και να ακούν.
- Η αναγνώριση ομιλίας και γραφής είναι το κλειδί με το οποίο οι ομιλούντες ασιατικές γλώσσες θα αποκτήσουν πλήρη χρήση του Η/Υ τους.
- Ένα άτομο που μιλά και καταλαβαίνει μια ασιατική γλώσσα μπορεί να αλληλεπιδράσει με έναν υπολογιστή που «μιλά» αυτή τη γλώσσα.

Δυνατότητα έκφρασης συναισθημάτων

- Σε μια προσπάθεια εμπλουτισμού ενός γραπτού κειμένου με περισσότερο συναίσθημα, οι αποστολείς μηνυμάτων συχνά σχηματίζουν φατσούλες με τα σύμβολα του πληκτρολογίου.
- Για παράδειγμα τη φατσούλα :) για να εκφράσουν χαρά ή για να τονίσουν ένα αστείο, τη φατσούλα : (για να εκφράσουν λύπη.
- Στην ομιλία τέτοιου είδους συναισθήματα εκφράζονται με αλλαγή της προσωδίας (π.χ. η αλλαγή στον τόνο της φωνής, στην ταχύτητα και στην ένταση της φωνής).

Πρόσβαση σε χρονικά ευαίσθητα δεδομένα

- Κάποιες φορές αυτοί που πραγματοποιούν ένα τηλεφώνημα χρειάζονται πρόσβαση σε χρονικά ευαίσθητα δεδομένα διαθέσιμα και ανανεωμένα όταν χρειάζονται.
- Για παράδειγμα όταν ξαφνικά συμβεί κάτι με μια εταιρία, στην οποία αυτός που καλεί έχει μετοχές ή υπάρχει είδηση για την καθυστέρηση μιας πτήσης.
- Οι προφορικές ανακοινώσεις αυτού του είδους μπορούν να αντικαταστήσουν το σύστημα των γραπτών καταχωρήσεων.
- Μέσα από το κινητό τηλέφωνο μπορεί κανείς να έχει πρόσβαση σε κάθε είδους ευαίσθητα χρονικά δεδομένα

Πρόσβαση σε χωρικά ευαίσθητα δεδομένα (1 από 2)

- Τα κινητά τηλέφωνα, οι υπολογιστές παλάμης και τα laptop δίνουν τη δυνατότητα στον χρήστη όπου κι αν βρίσκεται να έχει πρόσβαση σε ευαίσθητα δεδομένα τα οποία χρειάζεται, όπως όταν κάποιος επιστρέφει σπίτι του και θέλει να μάθει την κατάσταση της κυκλοφοριακής κίνησης ή το που βρίσκεται το κοντινότερο ξενοδοχείο ή νοσοκομείο.
- Αυτή η δυνατότητα μπορεί να οδηγήσει στη μείωση των τηλεφωνικών κλήσεων για τέτοια θέματα.

Πρόσβαση σε χωρικά ευαίσθητα δεδομένα (2 από 2)

- Οι πωλητές, οι μεταφορείς και γενικά οι υπάλληλοι που βρίσκονται συνεχώς εν κινήσει μπορούν να έχουν πρόσβαση σε δεδομένα της εταιρίας, σχετικά με στοιχεία επικοινωνίας, καταχωρήσεις παραγγελιών, κατάσταση παραγγελιών, διαθεσιμότητα προϊόντος καταλόγου, πελατολόγιο και κάθε είδους πληροφορία που αφορά στην εταιρία. Έτσι οι υπάλληλοι δεν θα χρειάζεται πλέον να μαντεύουν τη διεύθυνση ενός πελάτη –σε άλλη περίπτωση θα έπρεπε να την αναζητήσουν στον Η/Υ της εταιρίας.
- Με τις φορητές αυτές συσκευές ο χρήστης έχει πρόσβαση σε χωρικά ευαίσθητα δεδομένα.

Πρόσβαση σε δημόσιες και προσωπικές πληροφορίες

- Το κινητό τηλέφωνο δίνει στους χρήστες τη δυνατότητα πρόσβασης σε πλήθος δημόσιων πληροφοριών, π.χ. οδηγίες κατεύθυνσης, πληροφορίες για την κυκλοφοριακή κίνηση, τον καιρό, πληροφορίες τηλεφωνικού καταλόγου, τραπεζικών λογαριασμών, μετοχών, ειδήσεις, διασκέδαση, αποτελέσματα αθλητικών αγώνων, ωροσκόπιο και άλλες δραστηριότητες.
- Οι χρήστες εξυπηρετούνται εντοπίζοντας τις πληροφορίες που χρειάζονται ή παραγγέλλοντας διάφορα προϊόντα και υπηρεσίες, π.χ. ο χρήστης μπορεί να πει στο σύστημα «στείλε λουλούδια στη γυναίκα μου».
- Το κινητό τηλέφωνο δίνει επίσης στους καταναλωτές τη δυνατότητα να έχουν πρόσβαση σε πληθώρα προσωπικών στοιχείων, όπως e-mail, υπενθυμίσεις επαγγελματικών ραντεβού και ιατρικές πληροφορίες.

Έλεγχος υπολογιστικών διαδικασιών και δραστηριοτήτων (1 από 2)

- Οι χρήστες τηλεφώνου μπορούν από απόσταση να ελέγχουν την ασφάλεια του σπιτιού τους και τις οικιακές τους συσκευές.
- Μπορούν να εκτελέσουν συναλλαγές, όπως αγορές, πληρωμές και μεταφορές.
- Εάν ένας χρήστης μπορεί να μιλάει και να ακούει, τότε μπορεί να ελέγχει από απόσταση διάφορες λειτουργίες που είναι συνδεδεμένες με υπολογιστή.

Έλεγχος υπολογιστικών διαδικασιών και δραστηριοτήτων (2 από 2)

- Υπάρχουν περίπου 1.3 δισεκατομμύρια τηλέφωνα στον κόσμο και μόνο 250 εκατομμύρια Η/Υ με πρόσβαση στο διαδίκτυο
- Πολύ περισσότεροι χρήστες μπορούν να έχουν πρόσβαση στο διαδίκτυο μέσα από το τηλέφωνο τους παρά από τον Η/Υ
- Η περίπτωση της χρήσης ομιλίας είναι ισχυρή. Υπάρχουν όμως και μειονεκτήματα.
- Η ομιλία έχει επίσης δυσκολίες

Μειονεκτήματα του να μιλάς και να ακούς έναν Η/Υ

Οι χρήστες τηλεφώνου δεν μπορούν

- να ακούν γραφικά
- να επεμβαίνουν με κατάδειξη
- να ακούν ενεργά για μεγάλα διαστήματα
- να υπερβούν τους περιορισμούς της ανθρώπινης μνήμης
- να ακουστούν σε ένα χώρο με θόρυβο
- να ακούσουν σε ένα χώρο με θόρυβο
- να θέλουν να τους ακούν τρίτοι ενώ μιλούν

Να ακούν γραφικά

- Κάποιες εφαρμογές απαιτούν γραφικά ή άλλα οπτικά στοιχεία ώστε να παρουσιαστούν σε αυτόν που καλεί.
- Αυτό είναι δύσκολο να επιτευχθεί μέσω ενός κινητού τηλεφώνου ή ενός τηλεφώνου που δεν έχει οθόνη.
- Χάρτες, πίνακες πληροφοριών, και κατάλογοι ηλεκτρονικού εμπορίου με εικονογραφήσεις και φωτογραφίες δεν μπορούν να μεταφερθούν μέσω φωνητικών –μόνο– διεπαφών.

Να επέμβουν με κατάδειξη (1 από 2)

- Για να αλλάξει πληροφορίες σε ένα λογιστικό φύλλο, ένας χρήστης (π.χ. του Excel) πρέπει να τοποθετήσει τον κέρσορα σε ένα κελί και να πληκτρολογήσει τη νέα τιμή.
- Παρόλο που είναι σίγουρα δυνατό να εισάγει την νέα τιμή εκφωνώντας την, το να αναγνωρίσει το επιθυμητό κελί είναι δύσκολο και μάλλον αλλόκοτο να πραγματοποιηθεί χρησιμοποιώντας ομιλία - ο χρήστης πρέπει να γνωρίζει και να εκφωνεί τις συντεταγμένες του κελιού. Αυτό είναι σχεδόν αδύνατο χωρίς μια οπτική διεπαφή χρήσης.
- Για να επέμβει σε ένα έγγραφο κειμένου, ένας χρήστης πρέπει να μετακινήσει τον κέρσορα στη λέξη ή την πρόταση που θέλει να τροποποιήσει και στη συνέχεια να εκφωνήσει τη μετατροπή. Παρόλο που η εκφώνηση της μετατροπής είναι εύκολη, ο χρήστης μπορεί να χρειαστεί να πει αρκετές λέξεις προκειμένου να τοποθετήσει τον κέρσορα στη θέση της λέξης που θέλει να τροποποιήσει.

Να επέμβουν με κατάδειξη (2 από 2)

- Μια θεμελιώδης διάσταση κάθε προγράμματος σχεδίασης και εικονογράφησης είναι η τοποθέτηση του κέρσορα σωστά προτού τροποποιηθεί το επιλεγμένο αντικείμενο. Είναι ξεκάθαρο ότι κάτι τέτοιο δεν είναι δυνατό χωρίς μια οπτική γραφική διεπαφή.
- Οποιαδήποτε εφαρμογή απαιτεί από τον χρήστη να τοποθετεί συχνά τον κέρσορα δεν αποτελεί ιδανική υποψηφιότητα για μετατροπή χρήση μόνο μέσω διεπαφής ομιλίας.

Να ακούν προσεκτικά για μεγάλο χρονικό διάστημα

- Η «πλοήγηση» όπως τελείται σήμερα στα PCs απαιτεί την επιλογή ενός «συνδέσμου» και την ενεργοποίηση του μέσω των πλήκτρων του ποντικιού.
- Το λεκτικό ισοδύναμο της επιλογής του συνδέσμου μπορεί να είναι δύσκολο μέσω μιας διεπαφής ομιλίας.
- Μια δυνατότητα που έχει ο χρήστης είναι να ακούσει κάποιο εκφωνούμενο κείμενο, να εντοπίσει που περιέχεται κάποιος σύνδεσμος στο κείμενο και στη συνέχεια να εκφωνήσει κάποια εντολή όπως «μετάβαση» ή «πήγαινε εκεί».
- Ο χρήστης πρέπει να ακούσει προσεκτικά το πρόγραμμα φωνητικής φυλλομέτρησης (voice browser), να αναγνωρίσει τότε εκφωνείται κάποιος σύνδεσμος και να καθορίσει αν θα πρέπει να μεταβεί στην τοποθεσία του συνδέσμου – όλα αυτά σε πολύ μικρό χρονικό διάστημα. Οι περισσότεροι χρήστες βρίσκουν κουραστική μια τέτοια διαδικασία.

Να υπερβούν τους περιορισμούς της ανθρώπινης μνήμης (1 από 2)

- Η ομιλία είναι κάτι το φευγαλέο. Γίνεται αντιληπτή κατά την εκφώνηση, αλλά στη συνέχεια χάνεται πέραν από κάποιες έννοιες που παραμένουν στην βραχύχρονη μνήμη.
- Για να το δοκιμάσετε αυτό μιλήστε σε κάποιο φίλο για μερικά λεπτά και στη συνέχεια ζητήστε του να επαναλάβει τις τελευταίες πέντε προτάσεις που είπατε.
- Ελάχιστοι άνθρωποι μπορούν να αποστηθίσουν τις ακριβείς λέξεις που είπατε, και πολύ λίγοι να αναφέρουν το υποκείμενο και το ρήμα κάθε πρότασης.
- Ευτυχώς, οι περισσότεροι άνθρωποι μπορούν, με προσπάθεια, να επαναλάβουν τη γενική ιδέα αυτών που ειπώθηκαν.

Να υπερβούν τους περιορισμούς της ανθρώπινης μνήμης (2 από 2)

- Οι οθόνες, από την άλλη μεριά, δρουν ως δευτερεύουσα μνήμη. Αυτό επιτρέπει στο χρήστη να σαρώσει την οθόνη ψάχνοντας για την επιθυμητή πληροφορία ή σύνδεσμο.
- Για να ισορροπήσουν την έλλειψη βοηθήματος δευτερεύουσας μνήμης – όπως ένα οπτικό μενού ή μια φόρμα με πεδία – οι διάλογοι ομιλίας πρέπει να κατασκευάζονται προσεκτικά ώστε να επιτρέπουν στον χρήστη να διαλέγει την επιθυμητή επιλογή καθώς αυτή εκφωνείται, αντί να χρειάζεται να επαναλάβει μεγάλες λίστες με επιλογές.

Να ακουστούν σε θορυβώδες περιβάλλον

- Πολλοί άνθρωποι, ιδιαίτερα αυτοί που είναι μεγαλύτεροι, έχουν δυσκολία να ακούσουν κάτι όταν υπάρχει θόρυβος στο βάθος.
- Τα θορυβώδη περιβάλλοντα είναι ιδιαίτερα προβληματικά για αναγνώριση ομιλίας.
- Οι σχεδιαστές διαλόγων πρέπει να σχεδιάσουν προσεκτικά το διάλογο ώστε να αντιπαρέλθουν σε κακής ποιότητας αναγνώριση ομιλίας (λόγω του θορύβου).
- Μπορούν να σχεδιάσουν κάποιες επιλογές ώστε να διαφοροποιούνται όταν εκφωνούνται. Π.χ., αντί να ζητείται από το χρήστη να πει "Austin" ή "Boston," να ζητείται να πει "Austin, Texas" ή "Boston, Massachusetts.»

Να ακούσουν σε θορυβώδες περιβάλλον

- Κάποιοι πιστεύουν ότι το να μιλάς σε έναν υπολογιστή θα προσθέσει στην υπάρχουσα ηχορύπανση του χώρου εργασίας.
- Η βελτίωση του σχεδιασμού ενός γραφείου μπορεί να ελαχιστοποιήσει την ηχορύπανση – εγκαθιστώντας ηχοαπορροφητικά υλικά στους τοίχους, τα πατώματα, τις πόρτες, και άλλα φράγματα ήχου και αντικαθιστώντας τα διαχωριστικά γραφείων με κανονικούς τοίχους.
- Ενθαρρύνετε τους ανθρώπους να μιλούν χαμηλόφωνα χρησιμοποιώντας μικρόφωνα μικρής απόστασης αντί για μικρόφωνα γραφείου.
- Σαν έσχατο μέτρο, προμηθεύστε τους εργαζόμενους με εσωτερικά ακουστικά αυτιού.

Να τους ακούν τρίτοι ενώ μιλούν

- Αυτοί που καλούν δεν θέλουν άλλους να τους ακούν να κάνουν λάθη ενώ ομιλούν στους υπολογιστές τους.
- Ακόμα περισσότερο, οι χρήστες μπορεί να μη θέλουν να ξέρουν οι άλλοι τι κάνουν όταν μιλούν στους υπολογιστές τους.
- Ζητήματα μυστικότητας μπορούν να ελαχιστοποιηθούν παρέχοντας μια διεπαφή πληκτρολογίου για χρήση σε περιπτώσεις που απαιτούν μυστικότητα, όπως η περίπτωση εισαγωγής κωδικών ή άλλων ιδιωτικών πληροφοριών.

Περιορισμοί των διεπαφών ομιλίας

Οι δυσκολίες για την ανάπτυξη εφαρμογών ομιλίας περιλαμβάνουν τις ακόλουθες:

- οι τεχνολογίες ομιλίας δεν είναι τέλειες
- οι χρήστες έχουν εσφαλμένες προσδοκίες
- η ομιλία είναι ένα παροδικό μέσο
- η ομιλία αποτελεί μια μετατόπιση πλαισίου χρήσης Η/Υ (paradigm shift)

Οι τεχνολογίες ομιλίας δεν είναι τέλειες (1 από 2)

- Παρά την τεράστια προσπάθεια από ερευνητές και μηχανικούς, οι τεχνολογίες ομιλίας ενίοτε αποτυγχάνουν.
- Από καιρό εις καιρό, οι αλγόριθμοι αναγνώρισης ομιλίας εμφανίζουν λάθη.
- Οι αλγόριθμοι σύνθεσης ομιλίας παρουσιάζουν λάθη στην προφορά.
- Υπάρχουν πολλές τεχνικές για να ελαχιστοποιούνται αυτά τα λάθη, παρόλα αυτά τα λάθη συμβαίνουν.
- Όταν οι άνθρωποι μιλούν ο ένας στον άλλο, συμβαίνουν τα ίδια λάθη – ένα άτομο προφέρει λανθασμένα μια λέξη ή ένα άτομο δεν καταλαβαίνει μια λέξη.

Οι τεχνολογίες ομιλίας δεν είναι τέλειες (2 από 2)

- Δεν πείθεστε; Ακούστε άλλους να συνομιλούν και μετρήστε τις φορές που θα ακούσετε «Ορίστε;», «Τι;» «Το ξαναλές;». Πάραυτα, η συζήτηση συνεχίζεται. Πολλές φορές συνεχίζεται και χωρίς διακοπές.
- Πολλοί άνθρωποι μπορεί να μην έχουν συνείδηση τέτοιων ειδών λαθών και των ενεργειών που πρέπει να πράξουν προκειμένου να επανορθώσουν.
- Οι άνθρωποι που αναπτύσσουν διεπαφές ομιλίας σπαταλούν ένα μεγάλο μέρος του χρόνου και της δουλειάς τους ρυθμίζοντας με ακρίβεια τους αλγόριθμους αναγνώρισης και σύνθεσης ομιλίας ώστε να δουλεύουν καλύτερα. Επίσης σχεδιάζουν εφαρμογές και διάλογους ώστε να αντισταθμίσουν τις περιστασιακές αποτυχίες των αλγορίθμων αναγνώρισης και σύνθεσης ομιλίας.

Οι χρήστες έχουν ψευδείς προσδοκίες (1 από 3)

- Ο υπολογιστής HAL στην ταινία «Οδύσσεια του Διαστήματος: 2001», το ρομπότ C3PO στις 3 ταινίες «Οι Πόλεμοι των Άστρων», και ο υπολογιστής του ασρόπλοιου Enterprise στην τηλεοπτική σειρά «Star Trek», διαθέτουν και επιδεικνύουν διεπαφές ομιλίας που συμπεριλαμβάνουν εκλεπτυσμένη κατανόηση φυσικής γλώσσας, η οποία δεν είναι τεχνολογικά δυνατή σήμερα.
- Πολλοί χρήστες περιμένουν μία διεπαφή ομιλίας να υποστηρίξει εκλεπτυσμένη κατανόηση όσων λένε και ζητούνται.
- Πολλοί άνθρωποι, όταν αντιλαμβάνονται ότι οι μηχανικοί προσπαθούν να κάνουν τους υπολογιστές να κατανοούν τον προφορικό λόγο, σκέφτονται ότι οι υπολογιστές είναι τόσο περίπλοκοι που εύχονται να μπορούσαν απλά να πουν στον υπολογιστή όσα χρειάζονταν και ο υπολογιστής να τα έπραττε. Αυτό πιθανό δεν είναι εφικτό στο κοντινό μέλλον.

Οι χρήστες έχουν ψευδείς προσδοκίες (2 από 3)

- Παρόλο που ένας υπολογιστής είναι ικανός να μεταφράσει εκφωνήσεις σε κείμενο, αυτό δεν υπονοεί ότι μπορεί να κατανοήσει το κείμενο και να πράξει ανάλογα.
- Υπάρχει μια πληθώρα προβλημάτων κατανόησης φυσικής γλώσσας που πρέπει να λυθούν προτού να είναι εφικτό κάτι τέτοιο. Αυτά τα προβλήματα περιλαμβάνουν την άρση της αμφισημίας από λέξεις με πολλαπλά νοήματα, την αναγνώριση των ουσιαστικών στα οποία αναφέρονται οι αντωνυμίες, επαναφορά παρεξηγημένων νοημάτων, και επεξεργασία φωνητικών διορθώσεων –του τύπου «Θέλω μία ροζ, όχι, λάθος, θέλω μία κόκκινη μπλούζα». Η επίλυση αυτών των προβλημάτων απαιτεί γνώση του πραγματικού κόσμου, ή αλλιώς, αυτό που ονομάζεται «κοινή λογική».

Οι χρήστες έχουν ψευδείς προσδοκίες (3 από 3)

- Για να μεταβιβάσουμε κατάλληλες προσδοκίες στους χρήστες, μέσω τεκμηρίωσης και εκπαίδευσης πρέπει να εξηγήσουμε προσεκτικά ποιες δυνατότητες έχει ένας χρήστης μιας εφαρμογής ομιλίας.
- Επίσης οι χρήστες πρέπει να κατανοήσουν, ότι πρέπει να συμβάλλουν στην διαδικασία, απαντώντας προσεκτικά στις ερωτήσεις του υπολογιστή έτσι ώστε να πετύχουν το επιθυμητό αποτέλεσμα.
- Θα υπάρξουν φορές κατά τις οποίες ο υπολογιστής δεν θα καταλάβει τον χρήστη και ο χρήστης θα πρέπει να είναι πρόθυμος να ξεπεραστεί αυτή η δυσκολία.

Η ομιλία είναι ένα παροδικό μέσο

- Σε αντίθεση με τις οπτικές διεπαφές, η φωνή είναι εφήμερη. Τη μια στιγμή είναι εδώ και την άλλη χάνεται.
- Σε αντίθεση με τις οπτικές διεπαφές, οι οποίες επεκτείνουν την ανθρώπινη μνήμη παρουσιάζοντας μια οθόνη γεμάτη με πληροφορίες τις οποίες ο χρήστης μπορεί να αναθεωρήσει, να επιλέξει, και να διαχειριστεί με άνεση, οι φωνητικές διεπαφές είναι φευγαλέες.
- Οι χρήστες πρέπει να ακούσουν, να αναγνωρίσουν και να αντιδράσουν σε φωνητικές διεπαφές με ταχύτητα.

Speech is a paradigm shift

- Οι εφαρμογές ομιλίας διαφέρουν από τις παραδοσιακές διαδικτυακές εφαρμογές.
- Οι χρήστες διαβαίνουν τον παγκόσμιο οπτικό ιστό, ενώ συνδιαλέγονται με τον παγκόσμιο φωνητικό ιστό.
- Οι δύο διεπαφές χρήστη, οι γραφικές (GUI) και οι φωνητικές (VUI), είναι πολύ διαφορετικές.
- Οι οδηγίες για την σχεδίασή τους είναι διαφορετικές, καθώς και ο τρόπος που χρησιμοποιούνται.
- Η μετατόπιση από την HTML στην VoiceXML είναι τόσο σημαντική όσο και η μετάβαση από γλώσσες προγραμματισμού όπως η C και η Java στην HTML.

Πολυτροπικότητα (1 από 2)

- Τα πολυτροπικά συστήματα επεξεργάζονται δύο ή και συνδυασμό περισσότερων εισόδων από το χρήστη:
 - ομιλία, γραφή, αφή, βλέμμα, κίνηση κεφαλιού και σώματος.
- Ενσωματώνουν μία ή περισσότερες τεχνολογίες αναγνώρισης, π.χ.: “Βάλε αυτό εκεί”
- Έχουν τη δυνατότητα να λειτουργήσουν πιο «στιβαρά».
- Η έρευνα επικεντρώνεται κυρίως στους συνδυασμούς:
 - ομιλία και είσοδος μέσω γραφής
 - ομιλία και αναγνώριση κίνηση χειριών

Πολυτροπικότητα (2 από 2)

- Ένα πολυτροπικό σύστημα δύναται να υποστηρίξει αμοιβαία αποσαφήνιση, π.χ.:
 - Ο χρήστης λέει «χαντάκια» και ζωγραφίζει μερικά γραφικά σύμβολα.
 - Αλλά το σύστημα αναγνώρισης ομιλίας επικυρώνει το «χαντάκι» ως την καλύτερη εικασία.
 - Το σύστημα αναγνώρισης γραφής επικυρώνει την σωστή ερμηνεία του πληθυντικού.
 - Το σύστημα ελέγχει δευτερεύουσες επιλογές στη λίστα με τις καλύτερες δυνατές επιλογές.
 - Το σύστημα βρίσκει τη λέξη «χαντάκια» και άρει την πιθανή αμφισημία.
- Η σημασιολογική πληροφορία από κάθε τρόπο εισόδου μπορεί να αποσαφηνίσει μερικώς τις άλλες εισόδους.

Άσκηση 2.1 (1 από 3)

Επισκεφτείτε δύο φωνητικές πύλες της επιλογής σας (παραδείγματα δίδονται στην ιστοσελίδα του μαθήματος eclass.di.uoa.gr στη θέση ΣΥΝΔΕΣΜΟΙ). Καλέστε τον τηλεφωνικό αριθμό και συνομιλήστε με τις δύο πύλες. Γράψτε μια περίληψη απαντώντας στις ακόλουθες ερωτήσεις:

1. Ποιο είναι το όνομα και το τηλέφωνο των 2 φωνητικών πυλών;
2. Ποιο είναι το κοινό στο οποίο απευθύνεται η κάθε μίας;
3. Ποια είναι τα προϊόντα, οι πληροφορίες και οι υπηρεσίες που παρέχουν;
4. Ποιες εφαρμογές ή υπηρεσίες είναι χρονικά ευαίσθητες; (π.χ., «λήγουν» σε διάστημα μερικών λεπτών, ωρών, ημερών, εβδομάδων ή μηνών;)
5. Ποιες εφαρμογές εξαρτώνται από την τοποθεσία; (π.χ., είναι χρήσιμες μόνο στο σπίτι, στο γραφείο, σε ένα αυτοκίνητο εν κινήσει, ή σε μια πόλη σε άλλη χώρα;)

Άσκηση 2.1 (2 από 3)

6. Κάντε έναν συγκριτικό πίνακα για τις δύο πύλες που να περιλαμβάνει τουλάχιστον τα εξής χαρακτηριστικά:
- Αν πρόκειται για εφαρμογές δημόσιες ή ιδιωτικές,
 - αν μπορείτε να θυμηθείτε τις οδηγίες της εφαρμογής και τις εντολές των μενού ενώ χρησιμοποιείται την εφαρμογή,
 - αν πρόκειται για συναλλαγο-στραφείς εφαρμογές ή ελεγκο-στραφείς,
 - αν ζητούν από τον χρήστη να εισαγάγουν μια προσωπική πληροφορία;
 - αν πρέπει ο χρήστης να απαγγείλει την πληροφορία ή μπορεί να την εισαγάγει χρησιμοποιώντας την αφή μέσω του πληκτρολογίου;
 - τις προσωπικές ρυθμίσεις που μπορείτε να κάνετε με τη φωνή σε κάθε πύλη

Άσκηση 2.1 (3 από 3)

- ονομάστε τις τρεις κορυφαίες υπηρεσίες που παρέχονται από την κάθε μία
- περιγράψτε συνοπτικά πώς οι επισκέπτες πλοηγούν μεταξύ των υπηρεσιών στις συγκεκριμένες πύλες
- ποιο επιχειρησιακό πρότυπο χρησιμοποιεί κάθε πύλη; (π.χ., πώς οι προμηθευτές της πύλης πληρώνονται;)
- βαθμολογείστε σε μια κλίμακα από το 1 έως 10 (με 10 να είναι το καλύτερο αποτέλεσμα) όσον αφορά το πόσο αποδοτικό και εύκολο είναι κάποιος να χρησιμοποιήσει την κάθε μία πύλη. Εξηγήστε την εκτίμησή σας.

Άσκηση 2.2

Υπηρεσία τηλεανάγνωσης CallingWeb www.CallingWeb.org

Το CallingWeb αποτελεί μία δράση του **Συλλόγου Τεχνολογικής Ανάπτυξης Τυφλών** <http://www.stat.net.gr> και λειτουργεί με τις καινοτόμες τεχνολογίες και την υποστήριξη της e-rhetor <http://www.e-rhetor.gr/>

α) Περιγράψτε τις λειτουργικές δυνατότητες της υπηρεσίας συγκριτικά με τον παραδοσιακό τρόπο πρόσβασης σε ιστοσελίδες

β) Αξιολογήστε τη χρηστικότητα και την αποτελεσματικότητα της υπηρεσίας

Τέλος Ενότητας

Αρχιτεκτονική Συστημάτων Φωνητικών Διεπαφών
Χρήστη

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «**Εκπαίδευση και Δια Βίου Μάθηση**» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών 2015, Γεώργιος Κουρουπέτρογλου 2015. Γεώργιος Κουρουπέτρογλου.
«Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής. Αρχιτεκτονική Συστημάτων Φωνητικών Διεπαφών Χρήστη». Έκδοση: 1.0. Αθήνα 2015.
Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/DI37/>.

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων (1 από 2)

Οι φωτογραφίες που περιέχονται στην παρουσίαση αποτελούν πνευματική ιδιοκτησία τρίτων. Απαγορεύεται η αναπαραγωγή, αναδημοσίευση και διάθεσή τους στο κοινό με οποιονδήποτε τρόπο χωρίς τη λήψη άδειας από τους δικαιούχους. Στην παρουσίαση περιέχεται περιεχόμενο από τις ακόλουθες πηγές:

- KEIO University, SFC Global Campus, www.sfc.keio.ac.jp
- Encyclopedia Britannica Inc., English Phonetics, 1994, www.britannica.com
- Intel Drill – Information Flow, How to Read Lips, Nov 29th, 2011 by Markers.

Σημείωμα Χρήσης Έργων Τρίτων (2 από 2)

Η δομή και οργάνωση της παρουσίασης, καθώς και το υπόλοιπο περιεχόμενο, αποτελούν πνευματική ιδιοκτησία του συγγραφέα και του Πανεπιστημίου Αθηνών και διατίθενται με άδεια Creative Commons Αναφορά Μη Εμπορική Χρήση Παρόμοια Διανομή Έκδοση 4.0 ή μεταγενέστερη.