Επιμορφωτικό Σεμινάριο
Ξενοδοχείο Caravel, 10-11 Μαΐου 2003

Η Ποιότητα της Συνθετικής Ομιλίας στην Ακουστική Αναπαράσταση της Πληροφορίας

Γεράσιμος Ξύδας
BSc, MSc Πληροφορικής, υποψ. διδάκτωρ

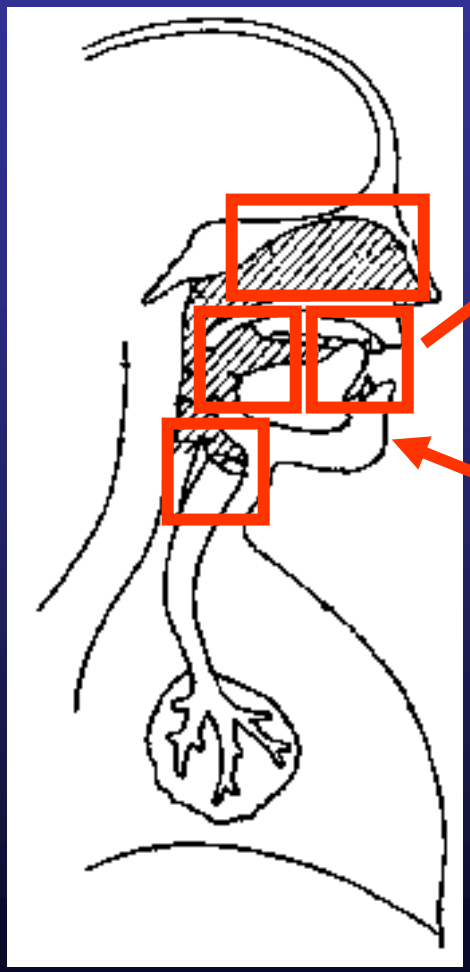
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών,
Τμήμα Πληροφορικής και Τηλεπικοινωνιών,

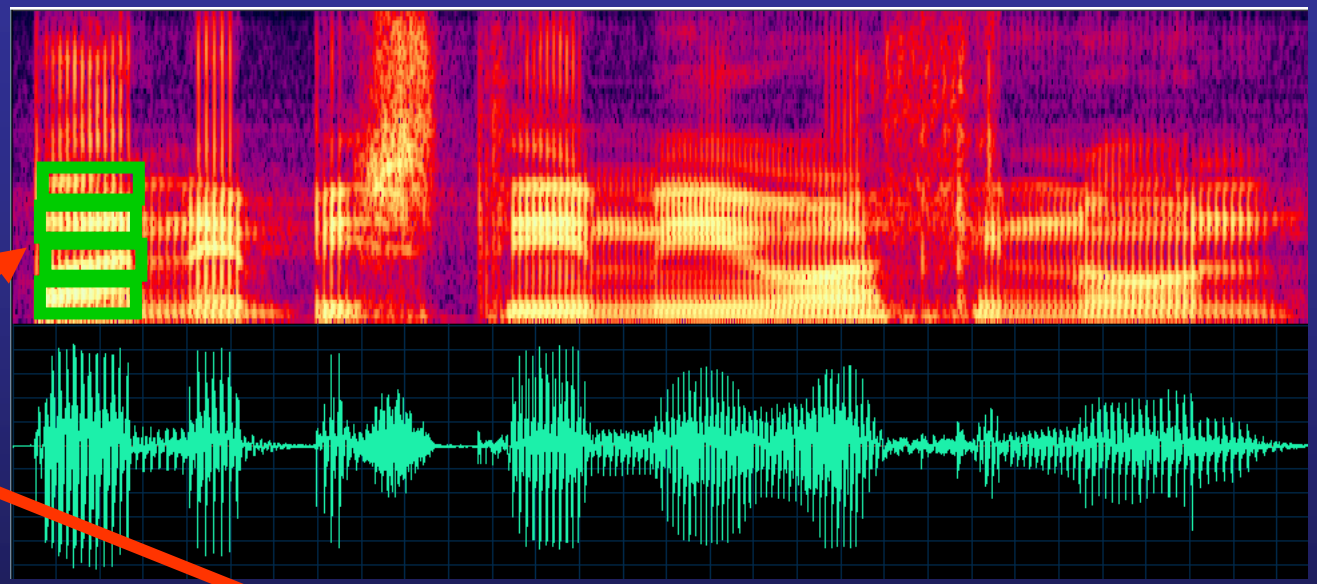
gxydas@di.uoa.gr

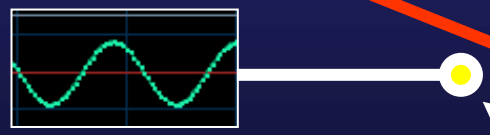
Δομή

- Μηχανισμός παραγωγής ομιλίας
- Τι είναι η συνθετική ομιλία;
- Μετατροπή κειμένου σε ομιλία - ΔΗΜΟΣΘέΝΗΣ
- Προσεγγίσεις στην ακουστική αναπαράσταση εγγράφων

Μηχανισμός Παραγωγής Ομιλίας

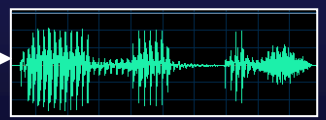


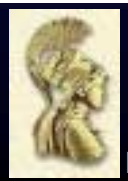




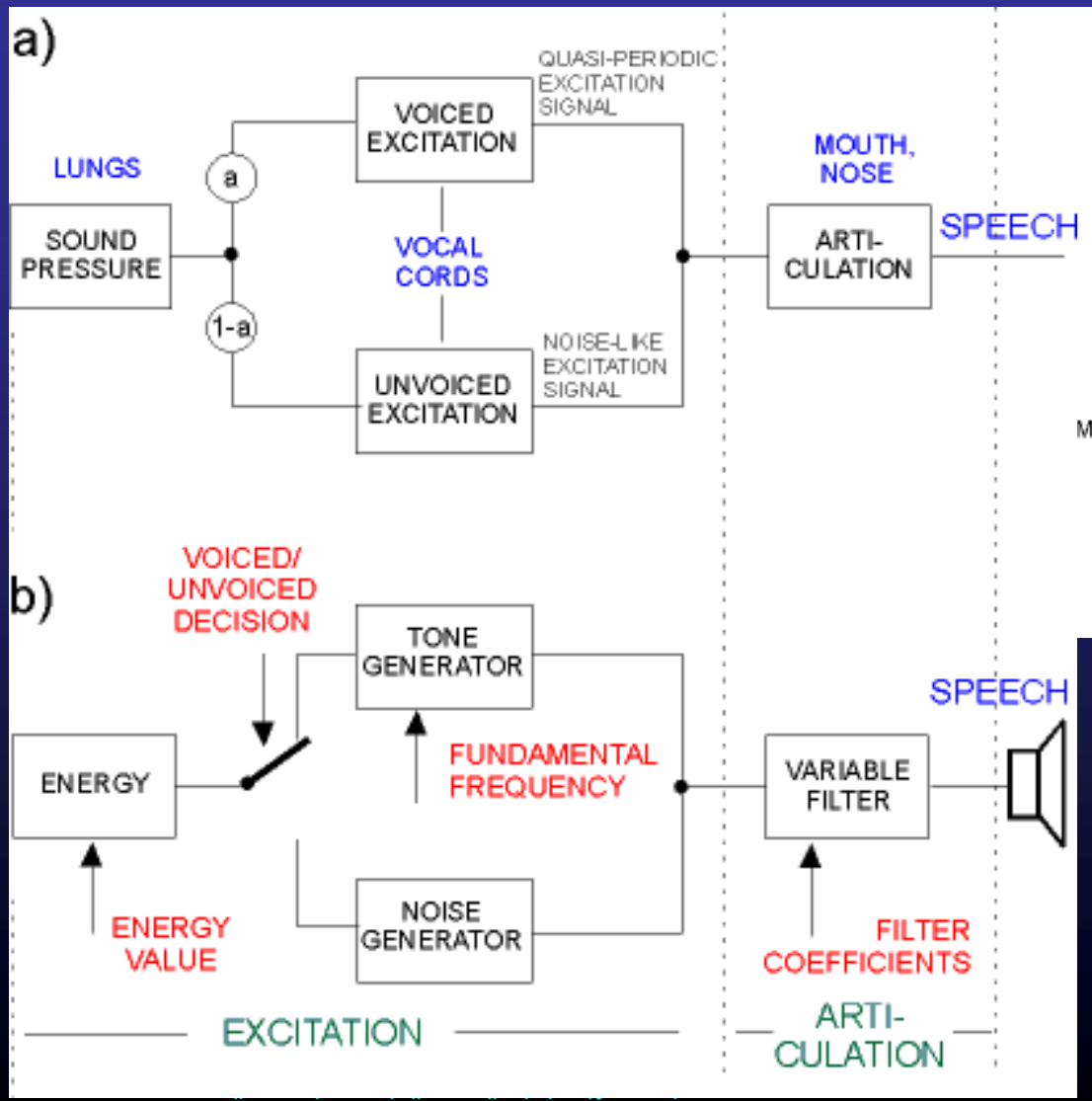
πηγές

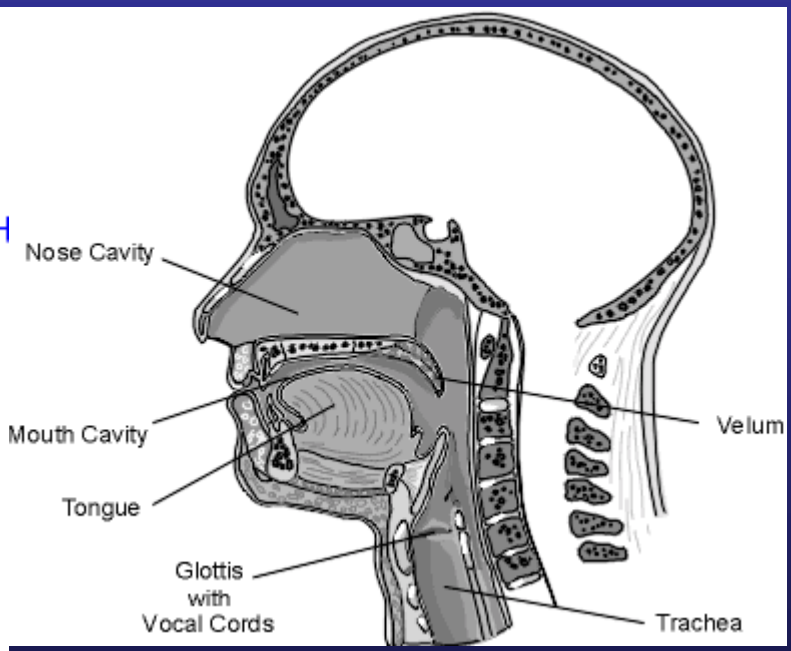
ηχεία





Μοντέλο Πηγής-Φίλτρου





Τι είναι η «συνθετική ομιλία»

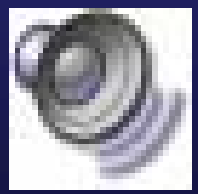
- Ομιλία που παράγεται με αυτόματες διαδικασίες από έναν ηλεκτρονικό υπολογιστή και προσομοιάζει την συμπεριφορά της ανθρώπινης ομιλίας.
- Σαν τεχνολογία υφίσταται πειραματικά από τα τέλη ’50 και εμπορικά από το ’70.
- Διεπαφή με τον χρήστη τελευταίας γενιάς.
- Από hardware (κουτάκια, χαμηλή ρομποτική ποιότητα) σε software (ευέλικτα, υψηλή ποιότητα)

Ποιότητα συνθετικής ομιλίας

- **Καταληπτότητα**: ήταν κατανοητό το περιεχόμενο της ομιλίας;
- **Φυσικότητα**: πόσο κοντά στην φυσική χροιά ήταν η ομιλία;
- Επιπλέον, μεταδόθηκε σωστά η **πληροφορία**;

Συνθέτης Φωνοσυντονισμών Ομιλίας

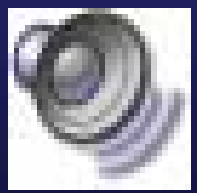
- Ελέγχεται από έναν **πίνακα 40 παραμέτρων** που ανανεώνουν την συμπεριφορά **πηγών** και **φίλτρων** κάθε 5 msec.
- ***«Είμαι ο πρώτος συνθέτης ομιλίας του Πανεπιστημίου Αθηνών»*** (1998)



  - **Μέτρια καταληπτότητα**
  - **Χαμηλή φυσικότητα**

Βελτιώνοντας την ποιότητα

- Οι παράμετροι της ομιλίας εμπεριέχονται σε μικρά ηχογραφημένα τμήματα (φωνήματα, δίφωνα, λέξεις, φράσεις).
- ***«Είμαι η πρώτη φωνητική βάση διφώνων του Πανεπιστημίου Αθηνών»*** (2001)



  - **Υψηλή καταληπτότητα**
  - **Μέτρια φυσικότητα**

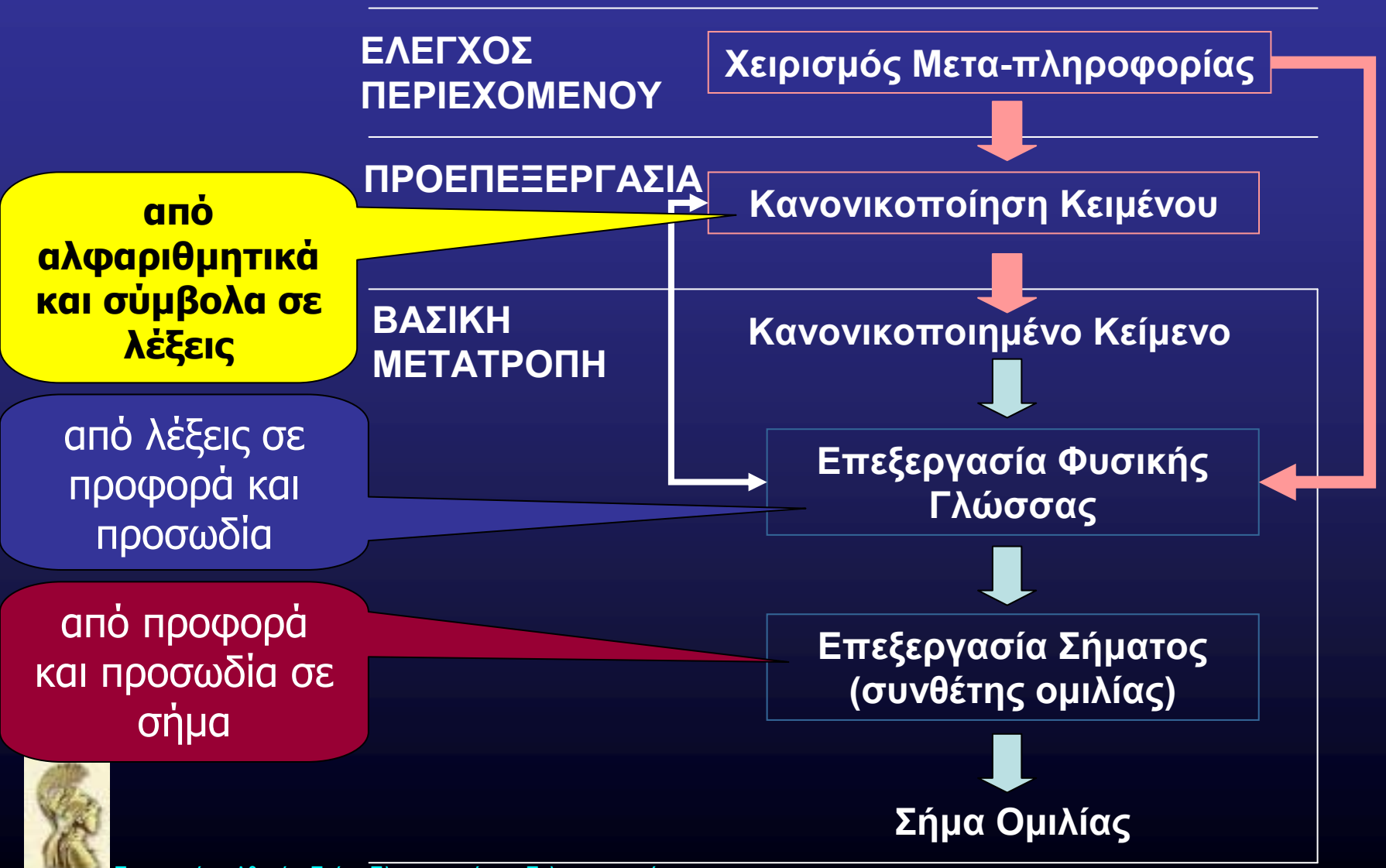
Πληροφορία στην ομιλία

- *«Ο Νίκος ήρθε από τη Νάξο με πλοίο»*
- *«Ο **Νίκος** ήρθε από τη Νάξο με πλοίο»*
- *«Ο Νίκος ήρθε από τη **Νάξο** με πλοίο»*
- *«Ο Νίκος ήρθε από τη Νάξο με **πλοίο**»*

Από τον συνθέτη ομιλίας στο σύστημα μετατροπής κειμένου σε ομιλία (ΜΚσΟ)

- Ο συνθέτης ομιλίας αποτελεί πλέον το τελικό στάδιο σε μία πληθώρα διαδικασιών.
- Η είσοδος σε ένα σύστημα ΜΚσΟ είναι κείμενο και όχι παράμετροι:
  - Ηλεκτρονικές εφημερίδες, e-mail, επεξεργαστής κειμένου, ηλεκτρονικά βιβλία, σελίδες διαδικτύου,...
- Το ΜΚσΟ ανάλογα με το κείμενο, συνθέτει την κατάλληλη εκφορά λόγου (προφορά & προσωδία)

Μετατροπή κειμένου σε ομιλία



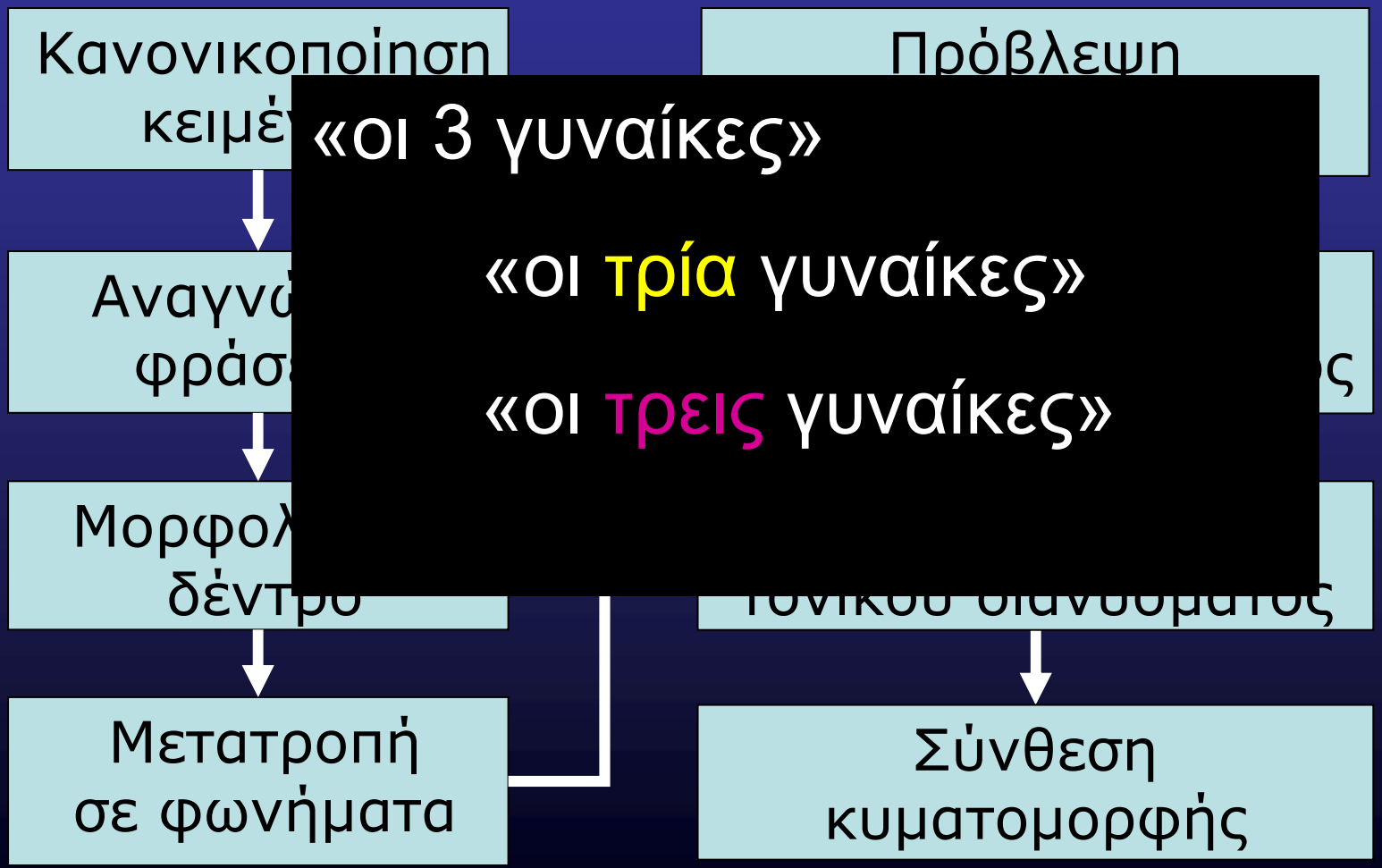
ΔΗΜΟΣΘέΝΗΣ (1/5)

- Σύστημα μετατροπής κειμένου σε ομιλία.
- Εξ'ολοκλήρου ανάπτυξη από την Ομάδα Φωνής του ΕΚΠΑ.
- Υποστηρίζει την ελληνική γλώσσα και την αγγλική.

ΔΗΜΟΣΘέΝΗΣ (2/5)

- **http://demosthenes.di.uoa.gr**
- Διατίθενται ελεύθερα (για μη εμπορικούς σκοπούς): κάποιες εκδόσεις του ΔΗΜΟΣΘέΝΗ, φωνητικοί & γλωσσολογικοί πόροι για τα Ελληνικά.
- Λίστα συζητήσεως ηλεκτρονικού ταχυδρομείου

ΔΗΜΟΣΘΈΝΗΣ (3/5)



ΔΗΜΟΣΘΈΝΗΣ (4/5)

- ***«Η θέλησή του Δημοσθένη ήταν τόσο μεγάλη, ώστε, όπως μας αναφέρει ο Πλούταρχος, έβαζε στο στόμα του μικρά χαλίκια την ώρα που απήγγειλε λόγους, προκειμένου να βελτιώσει την άρθρωσή του. Οι προσπάθειες του αυτές, απέδωσαν καρπούς και εξελίχθηκε σε σπουδαίο ρήτορα και πολιτικό.»***

Από εμπειρικά μοντέλα...  (2002)

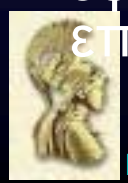
σε εκπαιδευόμενα...  (2003)

ΔΗΜΟΣΘΈΝΗΣ (5/5)

- Βαθμωτό σύστημα, ευέλικτο σε μετατροπές, τόσο για πειράματα όσο και για εφαρμογές.
  - Server mode (200*realtime)
  - Συνεργάζεται με τρίτες εφαρμογές (MS-OFFICE, Internet Explorer,…)
  - Συμβατό με το πρωτόκολλο Microsoft SAPI 4 & 5 (σύνδεση με screen-readers, talking agents, συστήματα IVR, ...)

Εφαρμογές

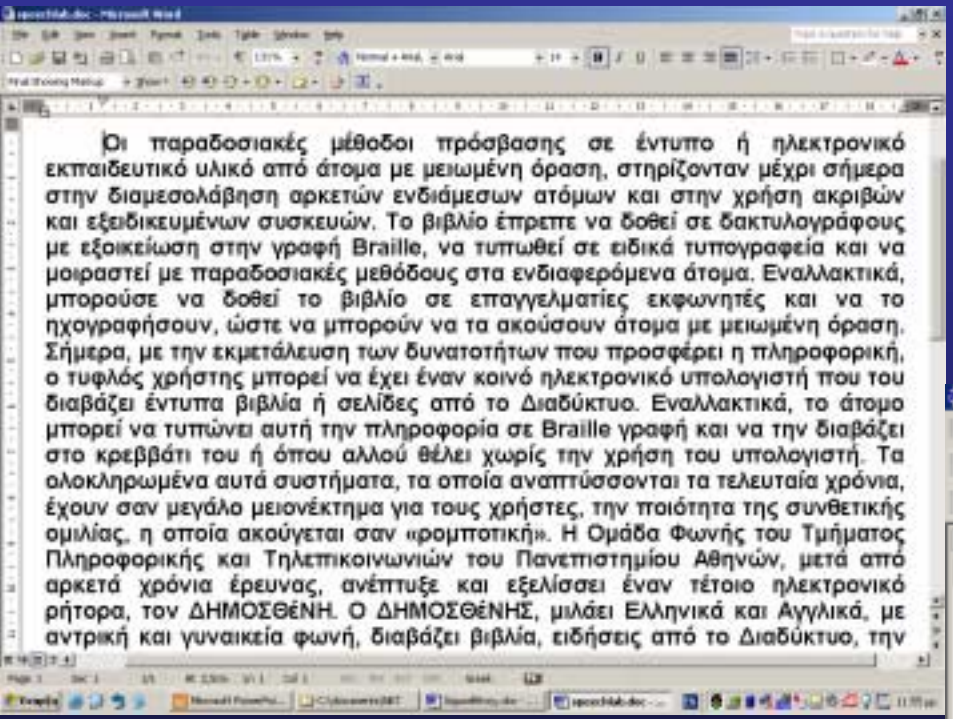
- **Πολυμέσα**: παιχνίδια, εκπαιδευτικό λογισμικό, λεξικά, εγκυκλοπαίδειες, ηλεκτρονικά βιβλία, εκμάθηση ξένων γλωσσών.
- **Τηλεπικοινωνίες**: αλληλεπιδραστικά συστήματα απόκρισης με φωνή - IVR, κέντρα εξυπηρέτησης κλήσεων – Call Centers, υπηρεσίες καταλόγων, ανάγνωση e-mail μέσω τηλεφώνου, μετατροπή SMS σε ομιλία, φωνητικές πύλες (voice portals)
- **Βιομηχανία**: συστήματα διάγνωσης βλαβών, συστήματα ρομποτικής, παρακολούθηση παραγωγής
- **Δημόσιοι χώροι**: περιηγήσεις μουσείων-αρχαιολογικών χώρων-εκθέσεων, ηλεκτρονικά κιόσκια ενημέρωσης για αεροδρόμια, εμπορικά κέντρα
- **Άτομα με ειδικές ανάγκες**: αναγνώστες οθόνες (screen readers) για άτομα με τύφλωση ή χαμηλή όραση, τεχνητά στόματα για περιπτώσεις αλαλίας/ δυσαρθρίας, συστήματα ανάγνωσης εφημερίδων βιβλίων για ηλικιωμένους, βοηθήματα διαπροσωπικής επικοινωνίας



Ποιότητα Μετατροπής Κειμένου σε Ομιλία

- Η ποιότητα της εξαρτάται κυρίως από το μέγεθος του θεματικού πεδίου το οποίο προσομοιάζεται.
  - 141 (ώρα), αποτελέσματα ΟΠΑΠ, δρομολόγια, υπηρεσίες καταλόγου, εκφώνηση δελτίου καιρού, ...
  - Πολιτικές ειδήσεις, αθλητικές ειδήσεις, πολιτιστικά, νομικά κείμενα, τεχνικά κείμενα, ...
  - Λογοτεχνικά κείμενα, ...

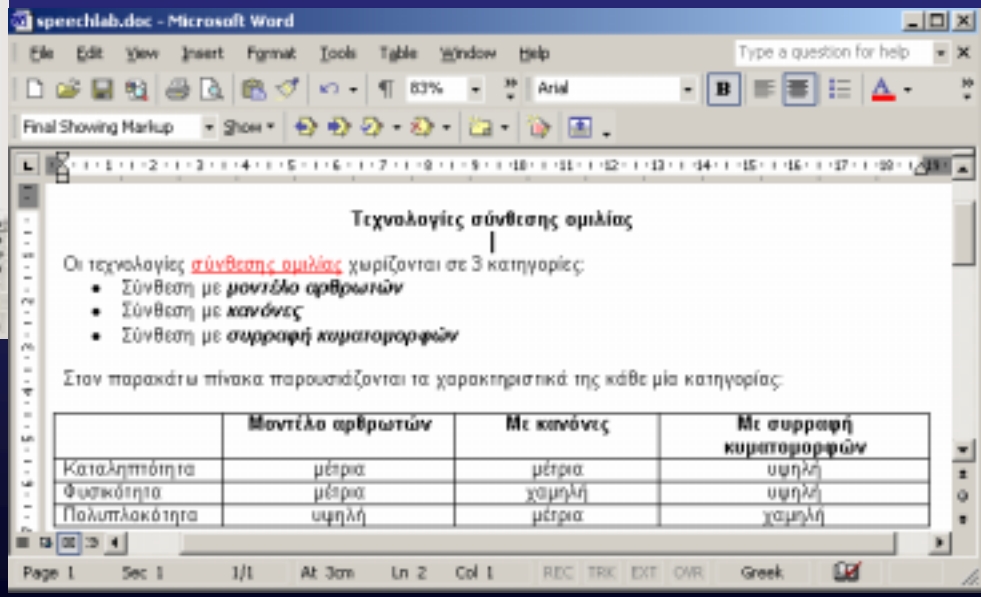
Ηλεκτρονικά Έγγραφα



Οι παραδοσιακές μέθοδοι πρόσβασης σε έντυπο ή ηλεκτρονικό εκπαιδευτικό υλικό από άτομα με μειωμένη όραση, στηρίζονταν μέχρι σήμερα στην διαμεσολάβηση αρκετών ενδιάμεσων ατόμων και στην χρήση ακριβών και εξειδικευμένων συσκευών. Το βιβλίο έπρεπε να δοθεί σε δακτυλογράφους με εξοικείωση στην γραφή Braille, να τυπωθεί σε ειδικά τυπογραφεία και να μοιραστεί με παραδοσιακές μεθόδους στα ενδιαφερόμενα άτομα. Εναλλακτικά, μπορούσε να δοθεί το βιβλίο σε επαγγελματίες εκφωνητές και να το ηχογραφήσουν, ώστε να μπορούν να τα ακούσουν άτομα με μειωμένη όραση. Σήμερα, με την εκμετάλευση των δυνατοτήτων που προσφέρει η πληροφορική, ο τυφλός χρήστης μπορεί να έχει έναν κοινό ηλεκτρονικό υπολογιστή που του διαβάζει έντυπα βιβλία ή σελίδες από το Διαδύκτυο. Εναλλακτικά, το άτομο μπορεί να τυπώνει αυτή την πληροφορία σε Braille γραφή και να την διαβάζει στο κρεββάτι του ή όπου αλλού θέλει χωρίς την χρήση του υπολογιστή. Τα ολοκληρωμένα αυτά συστήματα, τα οποία αναπτύσσονται τα τελευταία χρόνια, έχουν σαν μεγάλο μειονέκτημα για τους χρήστες, την ποιότητα της συνθετικής ομιλίας, η οποία ακούγεται σαν «ρομποτική». Η Ομάδα Φωνής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Αθηνών, μετά από αρκετά χρόνια έρευνας, ανέπτυξε και εξελίσσει έναν τέτοιο ηλεκτρονικό ρήτορα, τον ΔΗΜΟΣΘΕΝΗ. Ο ΔΗΜΟΣΘΕΝΗΣ, μιλάει Ελληνικά και Αγγλικά, με αντρική και γυναικεία φωνή, διαβάζει βιβλία, ειδήσεις από το Διαδύκτυο, την

Απλό κείμενο

Εμπλουτισμένο κείμενο



Τεχνολογίες σύνθεσης ομιλίας

Οι τεχνολογίες σύνθεσης ομιλίας χωρίζονται σε 3 κατηγορίες:

- Σύνθεση με ***μοντέλο αρθρωτών***
- Σύνθεση με ***κανόνες***
- Σύνθεση με ***συρραφή κυματομορφών***

Στον παρακάτω πίνακα παρουσιάζονται τα χαρακτηριστικά της κάθε μία κατηγορίας:

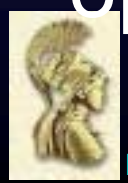
	Μοντέλο αρθρωτών	**Με κανόνες**	**Με συρραφή κυματομορφών**
Καταληπτότητα	μέτρια	μέτρια	υψηλή
Φυσικότητα	μέτρια	χαμηλή	υψηλή
Πολυπλοκότητα	υψηλή	μέτρια	χαμηλή

Κατηγορίες μετα-πληροφορίας

- **Οδηγίες οπτικοποίησης**: **έντονα γράμματα**, *πλάγια*, πίνακες, bullets, μέγεθος γραμμάτων κλπ (π.χ. HTML, MS-Word)
- **Οδηγίες δομής**: επικεφαλίδα, τίτλος, παράγραφος, record content (π.χ. HTML, XML, SQL), σύνθετα συνολα (π.χ. μαθηματικές πράξεις – MathML)
- **Γλωσσολογική πληροφορία**: γραμματική, σύνταξη, μορφολογία, ρητορικές σχέσεις (π.χ. SOLE-ML, plain κείμενο)
- **Δομή κειμένου**: παρενθέσεις, σημάδια, κλπ
- **Φωνητικές Οδηγίες**: <prosody>, <emp>, <rate>, <pitch> κλπ (π.χ. SABLE, VoiceXML, SSML, ACSS)

Ποιά είναι η ακουστική αξία της μετα-πληροφορίας;

- Ανάλογα με τον τύπο του εγγράφου αλλάζει και ο ρόλος της μετα-πληροφορίας.
  - Σε κείμενα με οπτική διαμόρφωση βοηθάει στον εντοπισμό εστιακών (focus) σημείων.
  - Σε έγγραφα με ιεραρχική δομή βοηθάει στην απόδοση του σωστού νοήματος προς τον ακροατή.
- Η μετα-πληροφορία μπορεί να αποδοθεί με έλεγχο προσωδιακών χαρακτηριστικών, αλλαγή ομιλητή, παρεμβολή άλλων ήχων κλπ.



Πανεπιστημιο Αθηνων
Τμημα Πληροφορικης

ΔΗΜΟΣΘΕΝΗΣ
Συνθέτης Ομιλίας
Πληροφορίες
Χαρακτηριστικά
Το σύστημα
Δείγματα
Downloads
Δημοσιεύσεις
Ο ρήτορας
Επικοινωνία

Γεράσιμος Ξύδας
gxydas@di.uoa.gr

Τελευταία ενημέρωση:
Μάρτιος 2003

DEMOSTHeNES
Speech Composer

Συνθέτης Ομιλίας ΔΗΜΟΣΘΕΝΗΣ

έκδοση 2

Ο συνθέτης ομιλίας ΔΗΜΟΣΘΕΝΗΣ είναι ένα πολυγλωσσικό (multilingual και polyglot) σύστημα λογισμικού που μετατρέπει οποιοδήποτε κείμενο σε ομιλία και που υποστηρίζει πλήρως την ελληνική γλώσσα. Ο ΔΗΜΟΣΘΕΝΗΣ στοχεύει στην παραγωγή καταληπτής ανθρωπομορφικής συνθετικής ομιλίας από ένα ευρύ φάσμα ηλεκτρονικών κειμένων. Η καινοτόμος, ανοικτή και αρθρωτή αρχιτεκτονική του προσφέρει μεγάλες δυνατότητες προσαρμογής και επέκτασης.

Το σύστημα ΔΗΜΟΣΘΕΝΗΣ έχει υιοθετήσει με επιτυχία μερικές από τις κορυφαίες φωνητικές τεχνολογίες, ενώ παράλληλα εισάγει και καινούργιες μεθοδολογίες. Οι γεννήτριες προφοράς και προσωδίας που διαθέτει παράγουν ομιλία αρκετά κοντά στη φυσική. Ο ΔΗΜΟΣΘΕΝΗΣ διαχειρίζεται τα κείμενα με ποικίλους τρόπους, αναγνωρίζοντας και αναλύοντας διάφορες μορφές τους, όπως ακρώνυμα, ημερομηνίες, αριθμητικά, κ.α.

Ο ΔΗΜΟΣΘΕΝΗΣ είναι ιδανικός για συστήματα πολύμεσων (ομιλούσες εγκυκλοπαίδειες, παρουσιάσεις κ.ά.), εφαρμογές τεχνολογιών φωνής (π.χ. τηλεφωνικές υπηρεσίες, υπηρεσίες καταλόγου), βοηθήματα για άτομα με ειδικές ανάγκες κ.α. Επιπλέον, μπορεί να ενσωματωθεί ή να συνδεθεί με άλλες εφαρμογές παρέχοντάς τους έξοδο σε μορφή ομιλίας. Η προτοποριακή του σχεδίαση είναι πολύ αποδοτική (περίπου **200** φορές realtime στην έκδοση 2), με αποτέλεσμα να προσφέρει πολλά κανάλια σε εξυπηρετητές (servers). Επιπλέον, η υποστήριξη πρωτοκόλλων όπως το MS-SAPI επιτρέπουν την εύκολη διασύνδεσή του με πληθώρα εφαρμογών.

World Wide Web Consortium (W3C)

- Οι συστάσεις της W3C (ACSS, VoiceXML) αποτελούν οδηγό για κάποιον που θέλει να συμπεριλάβει συγκεκριμένα φωνητικά χαρακτηριστικά στα έγγραφά του.
- Όμως:
  - η πλειονότητα των κειμένων είναι και εξακολουθεί να γράφεται χωρίς φωνητική αντίληψη (εξοικίωση συγγραφέων με οπτικά εφφέ).
  - Οι συστάσεις αφορούν το Web.
  - Δεν χειρίζονται φωνητικά τη μετα-πληροφορία με την ιεραρχική δομή με την οποία έχει συντεθεί.

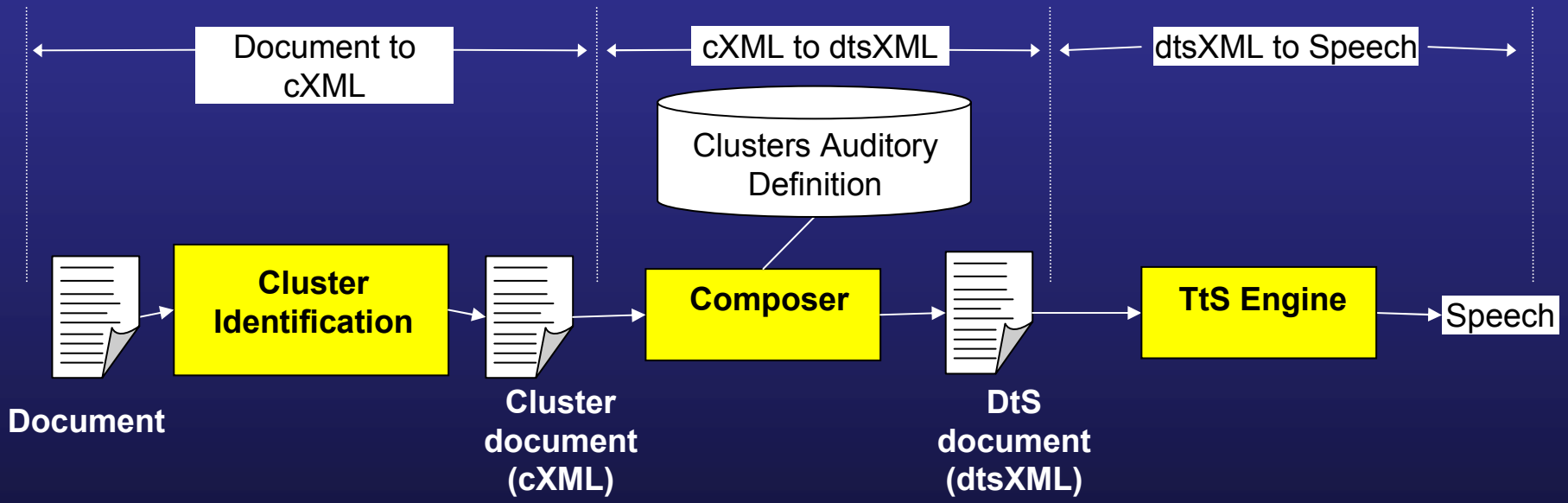
Σκοπός

- Προκειμένου η μετα-πληροφορία να έχει διακριτή ακουστική αναπαράσταση, αναθέτουμε ελεγχόμενα προσωδιακά χαρακτηριστικά σε αυτή.
- Έγγραφα χωρίς φωνητική αντίληψη αποκτούν ένα φωνητικό τρόπο αναπαράστασης.
- Ενισχύεται η «μνήμη» του ακροατή κατά την ακρόαση ενός εγγράφου (στην οπτική μορφή ο αναγνώστης μπορεί γρήγορα να μεταφερθεί μπρος-πίσω. Στην ακουστική...;).

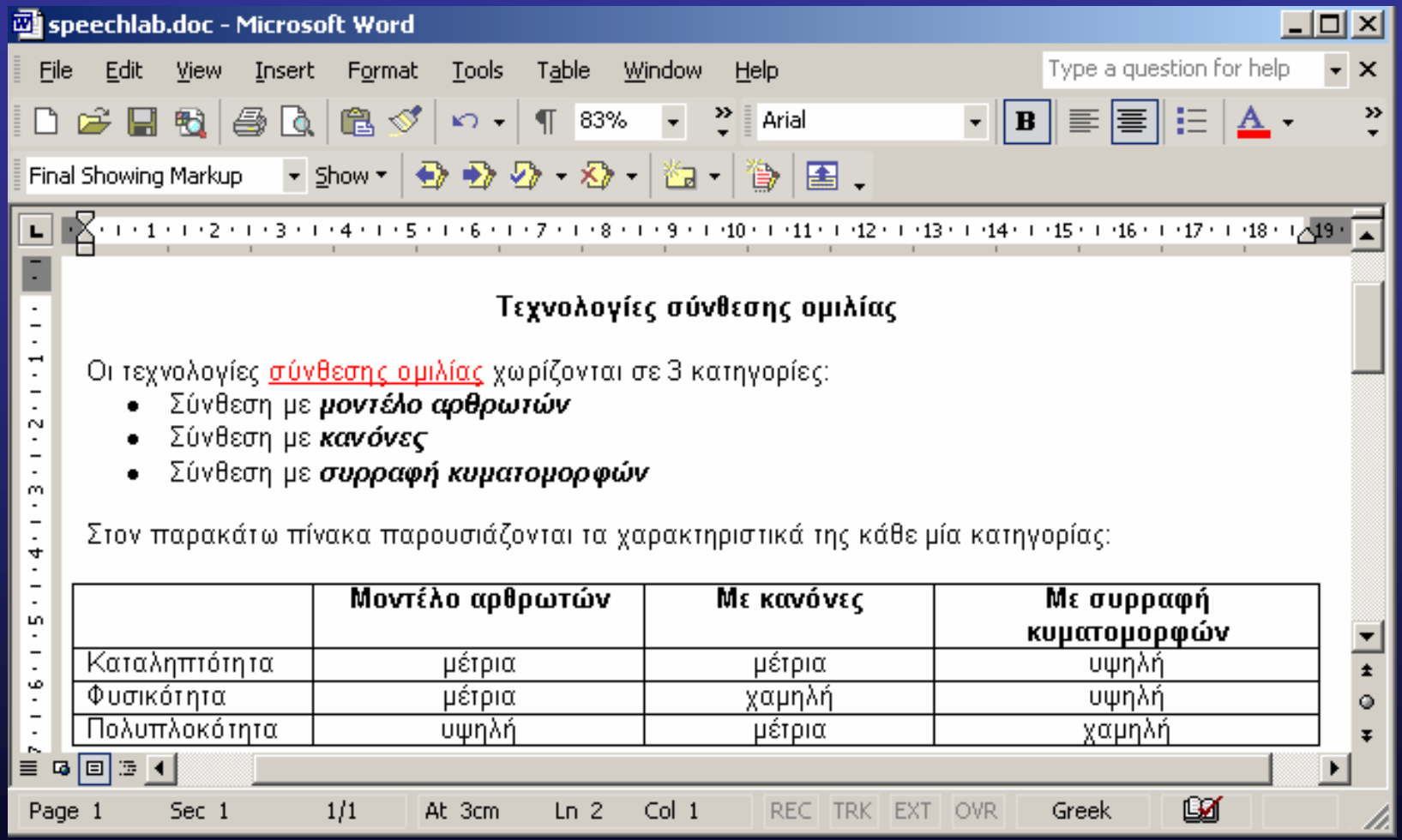
Document-to-Speech (1/2)

- Υπο-σύστημα ενός συστήματος μετατροπής κειμένου σε ομιλία που:
  - Αναγνωρίζει την μετα-πληροφορία
  - Επιτρέπει την συγγραφή σεναρίων ακουστικής συμπεριφοράς ανάλογα με την μετα-πληροφορία
- Επαυξημένη σε ποιότητα και πληροφορία ακουστική αναπαράσταση εγγράφων.
- Πεδία εφαρμογής: τηλεφωνική πρόσβαση σε ηλεκτρονικές βάσεις ή στο Web, πιστότερη απεικόνιση εγγράφων σε άτομα με προβλήματα όρασης, συστήματα IVR, υπηρεσίες καταλόγου, e-mail μέσω τηλεφώνου, ...

Document-to-Speech (2/2)

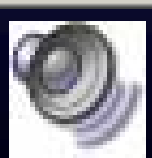


Παράδειγμα

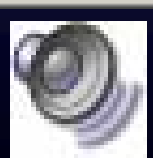


	Μοντέλο αρθρωτών	Με κανόνες	Με συρραφή κυματομορφών
Καταληπτότητα	μέτρια	μέτρια	υψηλή
Φυσικότητα	μέτρια	χαμηλή	υψηλή
Πολυπλοκότητα	υψηλή	μέτρια	χαμηλή

Text-to-Speech

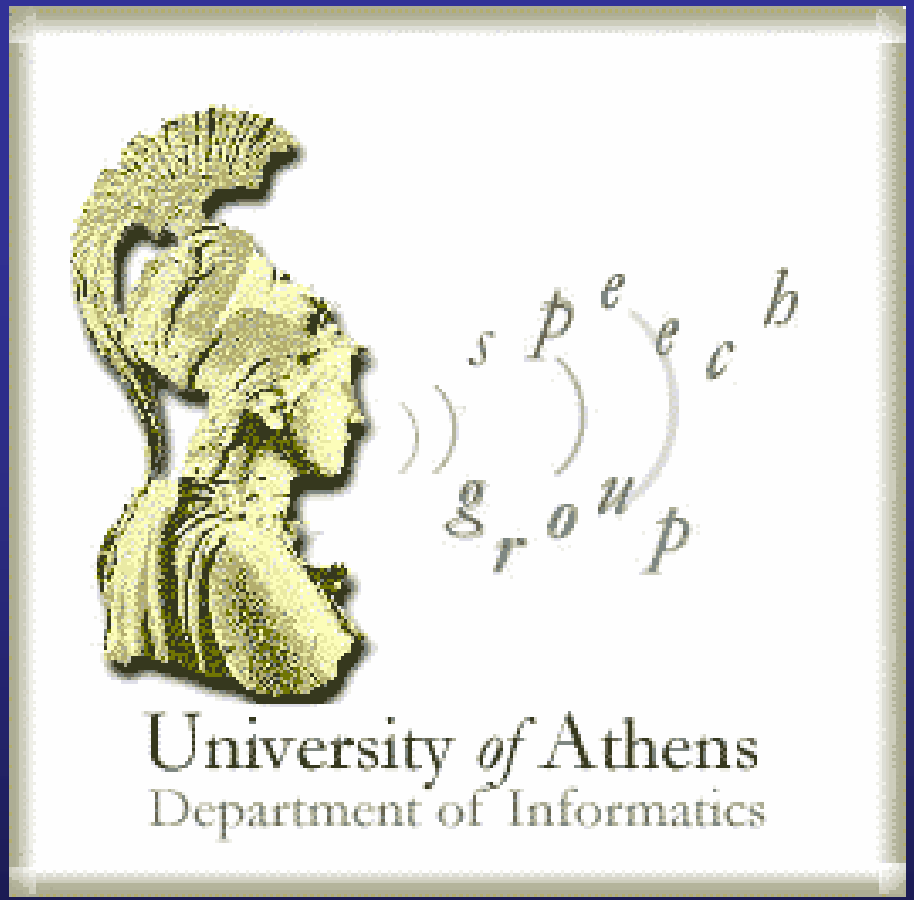


Doc-to-Speech



Συμπεράσματα

- Ποιότητα συνθετικής ομιλίας: Καταληπτότητα, φυσικότητα και μετάδοση πληροφορίας. Εξαρτάται από το μέγεθος του θεματικού πεδίου.
- ΔΗΜΟΣΘέΝΗΣ: Μία νέα αρχιτεκτονική συστήματος μετατροπής κειμένου σε ομιλία που μπορεί να φιλοξενήσει τις παραδοσιακές τεχνικές σύνθεσης καθώς και καινοτόμες.
- Μία προσέγγιση για D-t-S: Δομημένα Έγγραφα ή Έγγραφα με οπτική πληροφορία μπορούν να αναπαρασταθούν ακουστικά με μεγαλύτερη ακρίβεια.



http://www.di.uoa.gr/speech
http://www.di.uoa.gr/~gxydas