Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής

Ενότητα 4: **Αναγνώριση Ομιλίας και Κατανόηση Γλώσσας**

Γεώργιος Κουρουπέτρογλου
koupe@di.uoa.gr

Περιεχόμενα ενότητας

Μέθοδοι αναγνώρισης ομιλίας και κατανόησης γλώσσας

Συστατικά ενός Συστήματος Προφορικού Διαλόγου – Είσοδος και Έξοδος Ομιλίας

- **αναγνώριση ομιλίας**
- **κατανόηση γλώσσας**
- παραγωγή γλώσσας
- έξοδος ομιλίας
- βαθμίδα διαχείρισης διαλόγου

Επισκόπηση ενός Συστήματος Προφορικού Διαλόγου (1 από 4)

Παράδειγμα διαλόγου:

1. Σύστημα: Καλώς ήρθατε στην Υπηρεσία Πληροφοριών. Πού θα θέλατε να ταξιδέψετε? *(χαιρετίζει τον χρήστη και τον προτρέπει για κάποιες πληροφορίες)*
2. Επισκέπτης: Θα ήθελα να πετάξω για Λονδίνο την Παρασκευή φτάνοντας γύρω στις 9 το πρωί.
3. Σύστημα: Υπάρχει μια πτήση που αναχωρεί στις 7.45 π.μ. και φτάνει στις 8.50 π.μ.

Επισκόπηση ενός Συστήματος Προφορικού Διαλόγου (2 από 4)

Για να επεξεργαστεί την εκφώνηση 2, το σύστημα πρέπει να εκτελέσει τις ακόλουθες διεργασίες:

1. Να αναγνωρίσει τις λέξεις που εκφωνεί ο επισκέπτης (**Αναγνώριση Ομιλίας**).
2. Να αποδώσει ένα νόημα σε αυτές τις λέξεις (**Κατανόηση Γλώσσας**).
3. Να προσδιορίσει πως η εκφώνηση ταιριάζει στον διάλογο και να αποφασίσει τι να κάνει στη συνέχεια (**Διαχείριση Διαλόγου**) π.χ. : να διασαφηνίσει την εκφώνηση αν είναι ασαφής, να την επιβεβαιώσει (**θεμελίωση**), να ζητήσει περισσότερες πληροφορίες, ή να συμβουλευτεί μια πηγή πληροφοριών.
4. Να ανακτήσει όσες πτήσεις ταιριάζουν τις απαιτήσεις του χρήστη (**Εξωτερική Επικοινωνία**).
5. Να επιλέξει τις λέξεις και φράσεις που θα χρησιμοποιηθούν στην απάντηση (**Παραγωγή Γλώσσας**).
6. Να εκφωνήσει την απάντηση (**Μετατροπή κειμένου σε συνθετική ομιλία**).

Επισκόπηση ενός Συστήματος Προφορικού Διαλόγου (3 από 4)

- Σε αυτό το παράδειγμα κάθε διεργασία ολοκληρώθηκε ομαλά.
- Παρόλα αυτά, θα μπορούσαν να υπάρξουν προβλήματα:
  - Το σύστημα αναγνώρισης ομιλίας μπορεί να αποτύχει να αναγνωρίσει τις λέξεις του επισκέπτη ορθά
  - Το συστατικό κατανόησης γλώσσας:
    - Μπορεί να αποδώσει λανθασμένο νόημα, ή
    - Να είναι αδύνατο να επιλέξει το σωστό νόημα σε περιπτώσεις αμφισημίας.

Επισκόπηση ενός Συστήματος Προφορικού Διαλόγου (4 από 4)

- Διαχείριση διαλόγου:
  - που μπορεί να αφορά τη λήψη αποφάσεων σχετικά με το αν το σύστημα είναι σε θέση να ανακτήσει πληροφορίες από την εξωτερική πηγή πληροφοριών, ή
- Αν κάποια από τα αντικείμενα από τα δεδομένα εισόδου του επισκέπτη χρειάζονται διευκρίνιση.
- Το συστατικό παραγωγής της απάντησης πρέπει να σχηματοποιήσει τις εκφωνήσεις του συστήματος κατά τέτοιο τρόπο ώστε οι πληροφορίες να μην παρουσιαστούν καθαρά και χωρίς αμφισημίες,
- Το συστατικό μετατροπής κειμένου σε συνθετική ομιλία πρέπει να προφέρει τις λέξεις κατά τέτοιο τρόπο ώστε το μήνυμα να μην είναι κατανοητό.

Βασική Αρχιτεκτονική ενός Συστήματος Προφορικού Διαλόγου

σειριακή ροή ελέγχου

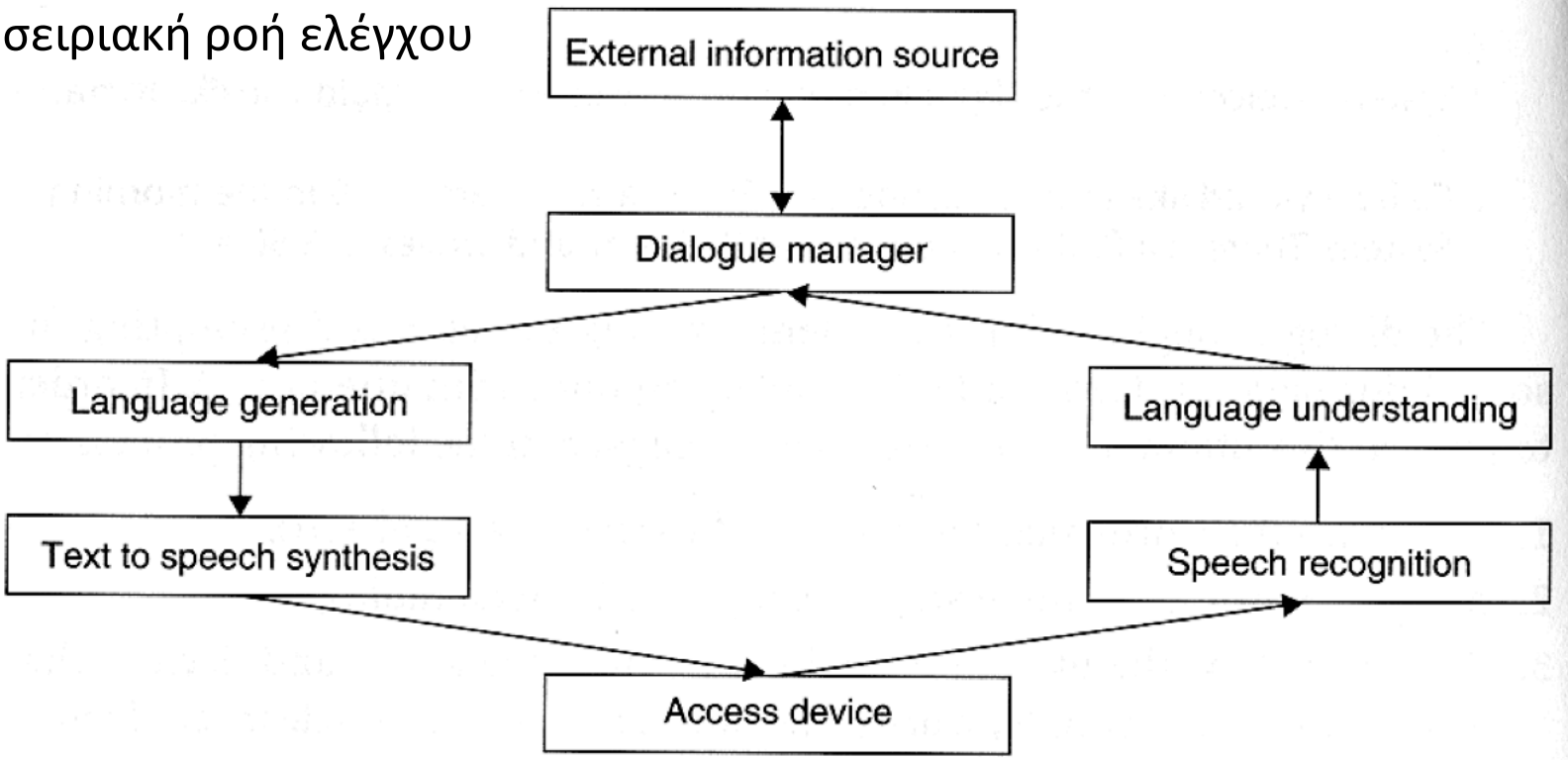
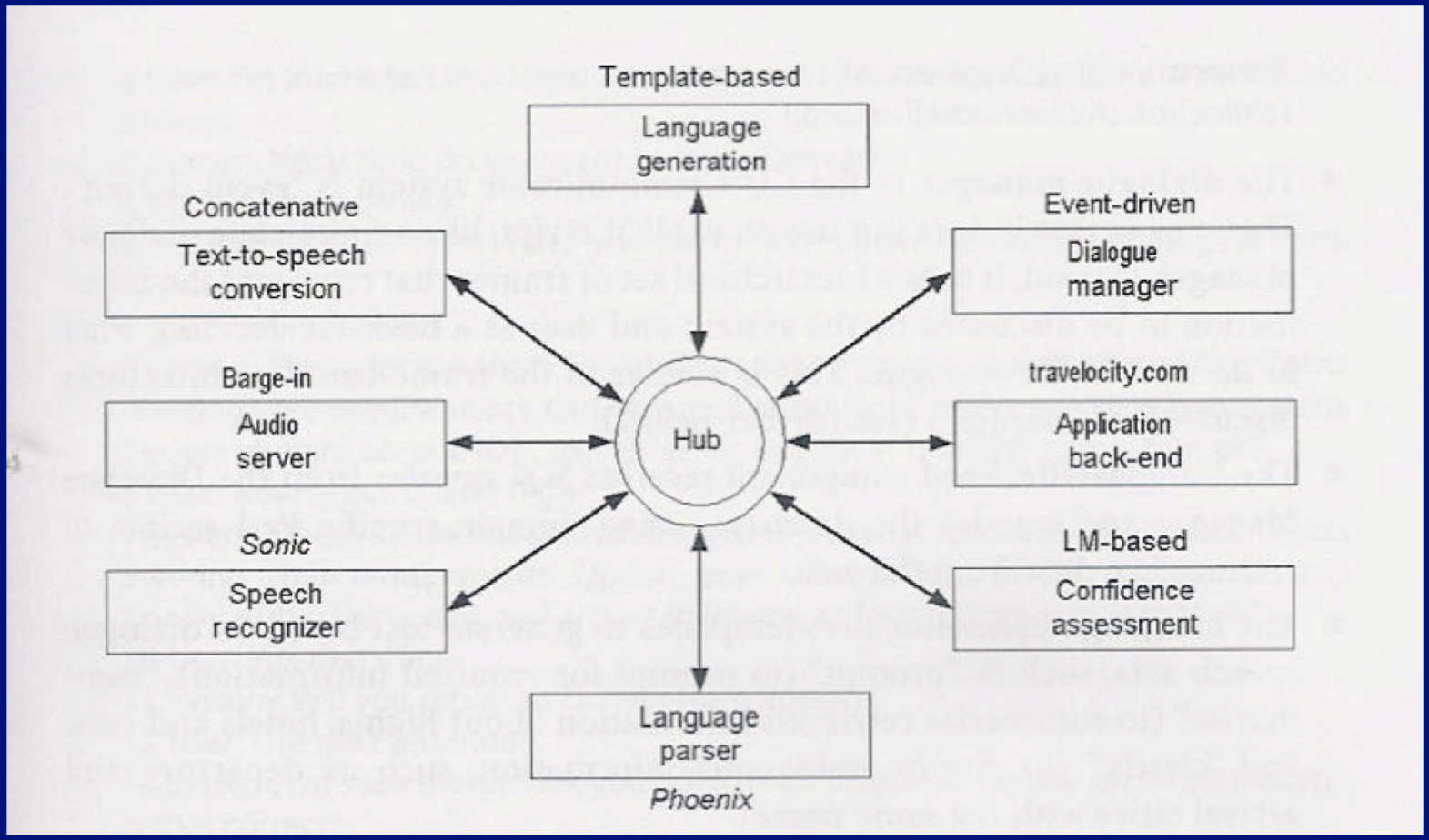


Figure 4.1. Architecture of a spoken dialogue system.

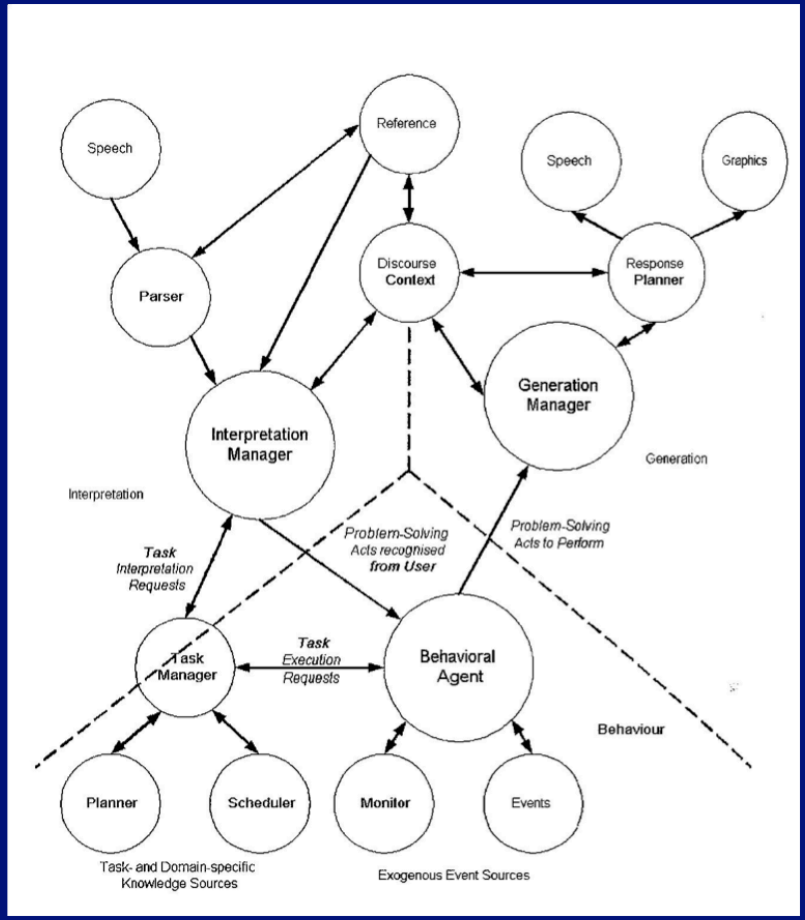
Αρχιτεκτονικές με μη-σειριακή ροή ελέγχου

- Σε προηγμένα ερευνητικά συστήματα
- Η βαθμίδα διαχείρισης διαλόγου επιτελεί μια κεντρική λειτουργία ελέγχου και αλληλεπιδρά άμεσα με τα άλλα συστατικά. (με χρήση πολλαπλών πηγών γνώσης σε διαφορετικά στάδια) π.χ. :
  - αν υπάρχει πρόβλημα στο στάδιο αναγνώρισης ομιλίας: η βαθμίδα διαχείρισης διαλόγου πιθανόν να δύναται να αποδώσει κάποια συναφή γνώση που δεν θα ήταν αλλιώς διαθέσιμη στο συστατικό αναγνώρισης ομιλίας.

Αρχιτεκτονική Communicator Πανεπιστημίου Colorado



Αρχιτεκτονική TRIPS Πανεπιστημίου Rochester



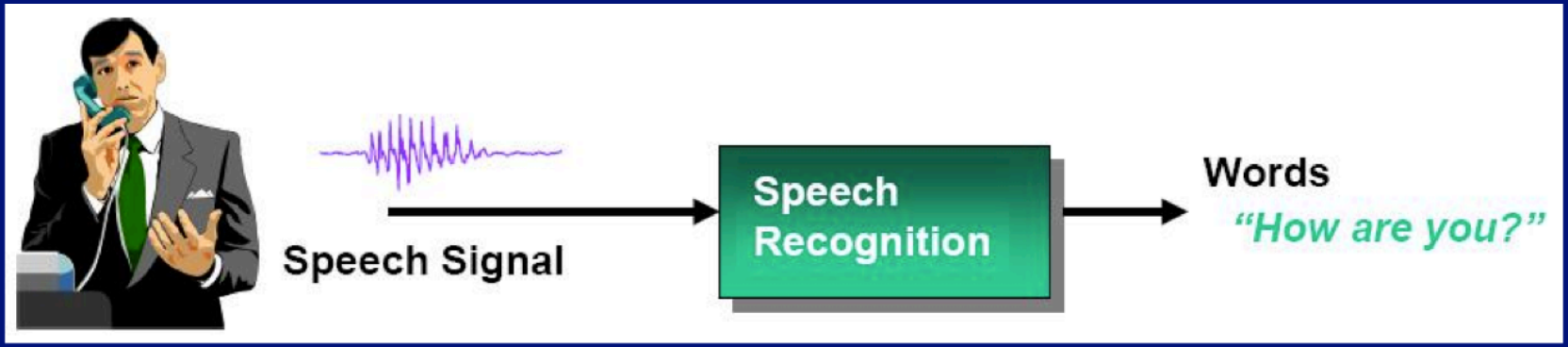
Αναγνώριση Ομιλίας (1 από 2)

- Κύριο έργο:
  - να «συλλάβει» την είσοδο από το χρήστη
  - να τη μετατρέψει σε ακολουθία λέξεων.
- Αξιοσημείωτη πρόοδος τις τελευταίες δεκαετίες, που οφείλεται σε:
  - βελτιωμένους αλγορίθμους
  - προόδους στην τεχνολογία των υπολογιστών
- *Οι λεπτομέρειες των μεθοδολογιών αναγνώρισης ομιλίας είναι πέραν του σκοπού αυτού του μαθήματος*

Αναγνώριση Ομιλίας (2 από 2)

- Οι μηχανικοί ανάπτυξης πρέπει να έχουν κάποια κατανόηση της τεχνολογίας αναγνώρισης ομιλίας ώστε να εκτιμούν:
  - γιατί η αναγνώριση ομιλίας είναι δύσκολη,
  - γιατί τα λάθη αναγνώρισης είναι αναπόφευκτα και
  - πώς να σχεδιάζουν συστήματα με τέτοιο τρόπο ώστε να ελαχιστοποιούν την εμφάνιση λαθών και να ασχολούνται επιτυχώς με όσα προκύπτουν.

Τι είναι αναγνώριση ομιλίας; Τι δεν είναι; (1 από 2)



Τι είναι αναγνώριση ομιλίας; Τι δεν είναι; (2 από 2)

Η αναγνώριση ομιλίας **δεν** καθορίζει:

- Ποιός είναι ο ομιλητής (αναγνώριση ομιλητή)
- Την έξοδο σε ομιλία (σύνθεση ομιλίας)
- Τι σημαίνουν οι λέξεις (κατανόηση ομιλίας)

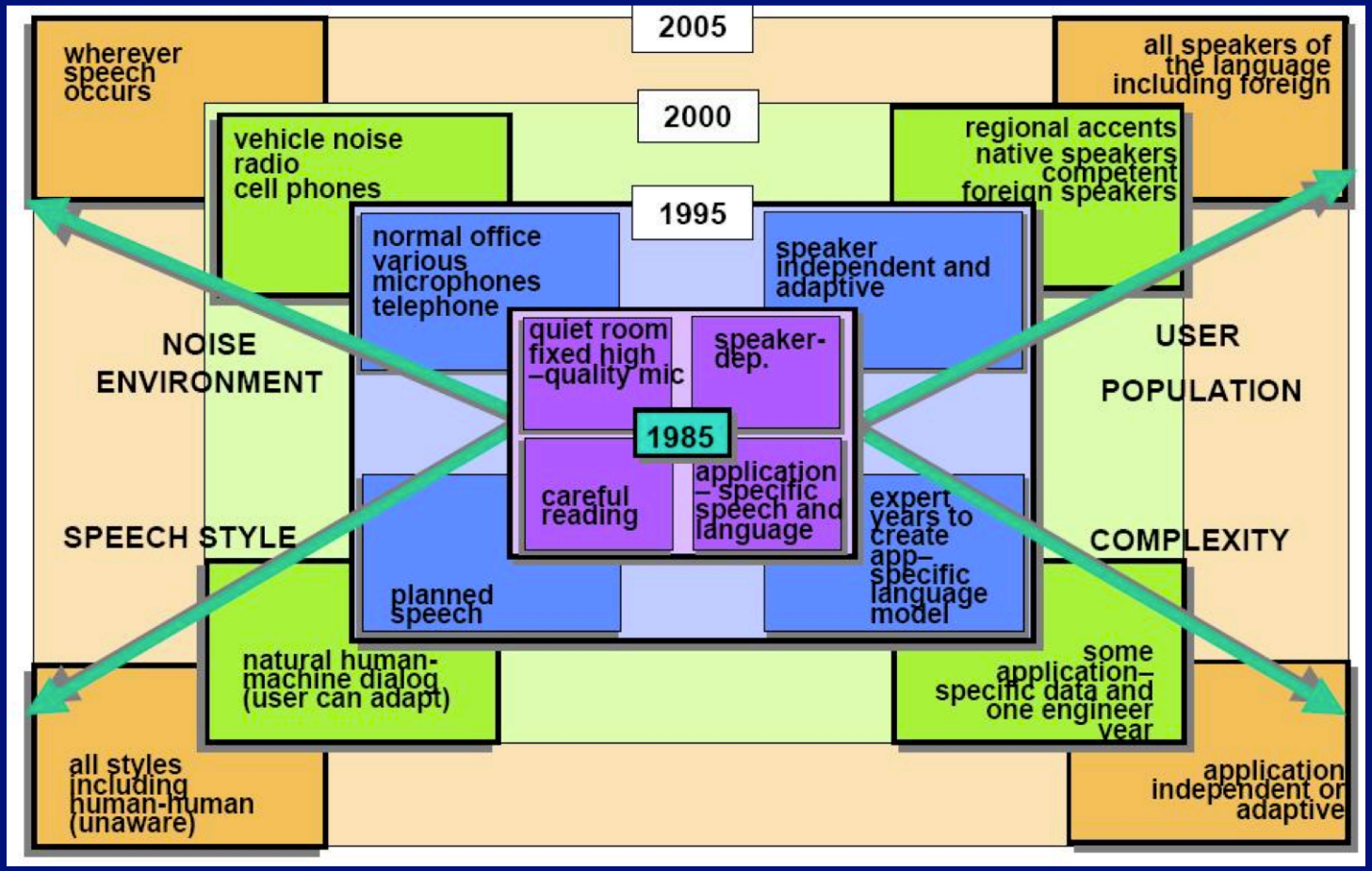
Γιατί είναι δύσκολη η αναγνώριση ομιλίας;

- Κύριο πρόβλημα: δεν μπορεί να εγγυηθεί μια σωστή απόδοση της εισόδου σε σύγκριση με είσοδο από το πληκτρολόγιο ή το ποντίκι.
- Μπορεί να περιγραφεί ως:
  - θορυβώδες κανάλι επικοινωνίας: η εκφώνηση που δίνει ο χρήστης μπορεί να είναι «φθαρμένη» καθώς περνάει από ένα θορυβώδες κανάλι.
  - αναγνώριση: αποκωδικοποίηση της θορυβώδους εκφώνησης μαντεύοντας ποια ήταν η αρχική της κατάσταση.
- Με χρήση προηγμένων στατιστικών μεθόδων:
  - αναθεωρεί όλες τις λέξεις στο λεξιλόγιο, και
  - επιστρέφει την πιο πιθανή λέξη (ή μια ταξινομημένη λίστα λέξεων) με βάση την ακουστική είσοδο.

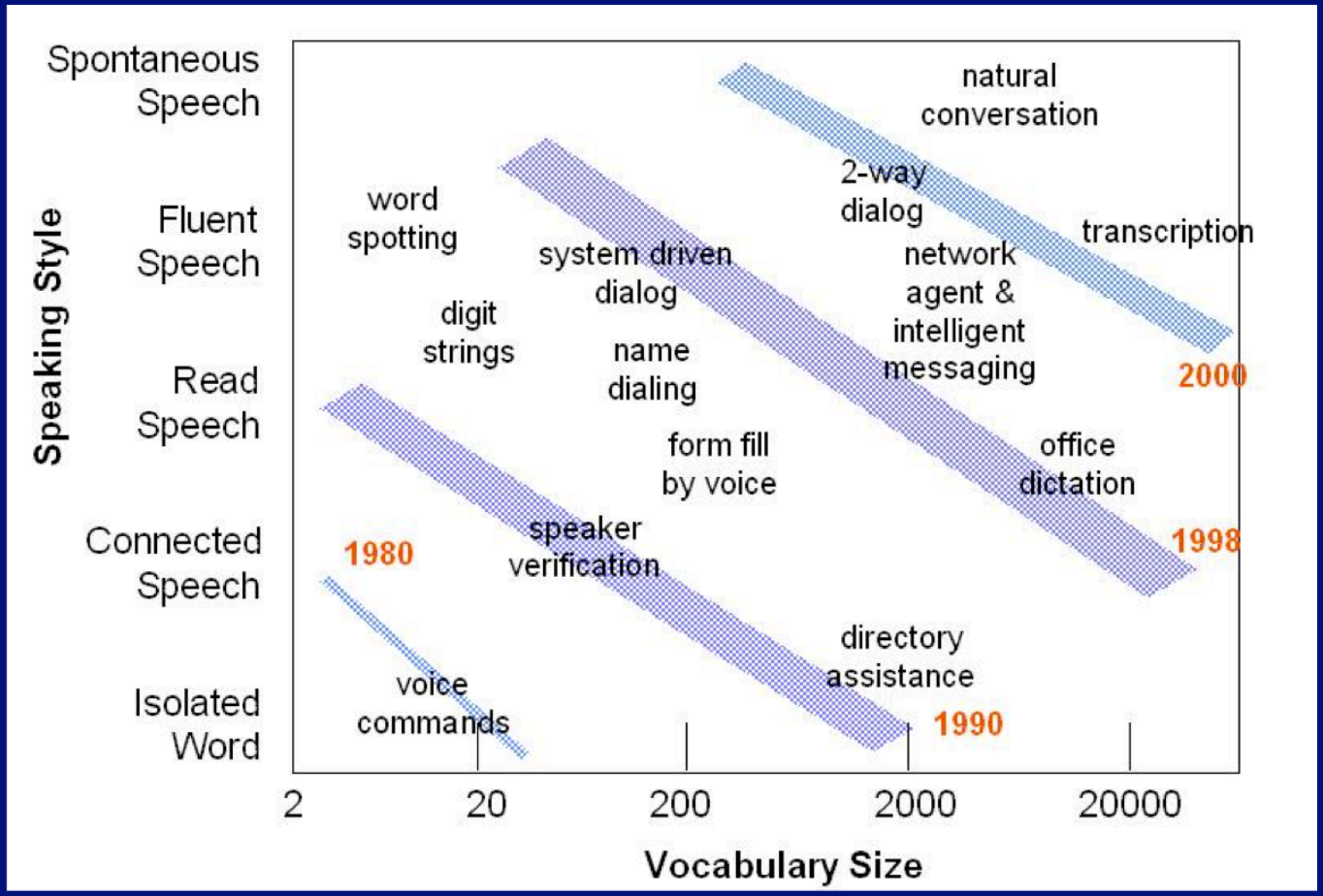
Κύριοι παράγοντες που επιδρούν στην απόδοση ενός συστήματος αναγνώρισης ομιλίας

- Στύλ / είδος ομιλίας
- Πλήθος χρηστών / ομιλητών
- Μέγεθος λεξιλογίου – πολυπλοκότητα εφαρμογής
- Περιβαλλοντικός θόρυβος

Πίνακας 4.1: Παράγοντες που επιδρούν στην απόδοση (χρονική εξέλιξη state of the art)



Εξελίξεις στην τεχνολογία αναγνώρισης ομιλίας



Λόγοι για τους οποίους η διαδικασία αναγνώρισης ομιλίας είναι δύσκολη

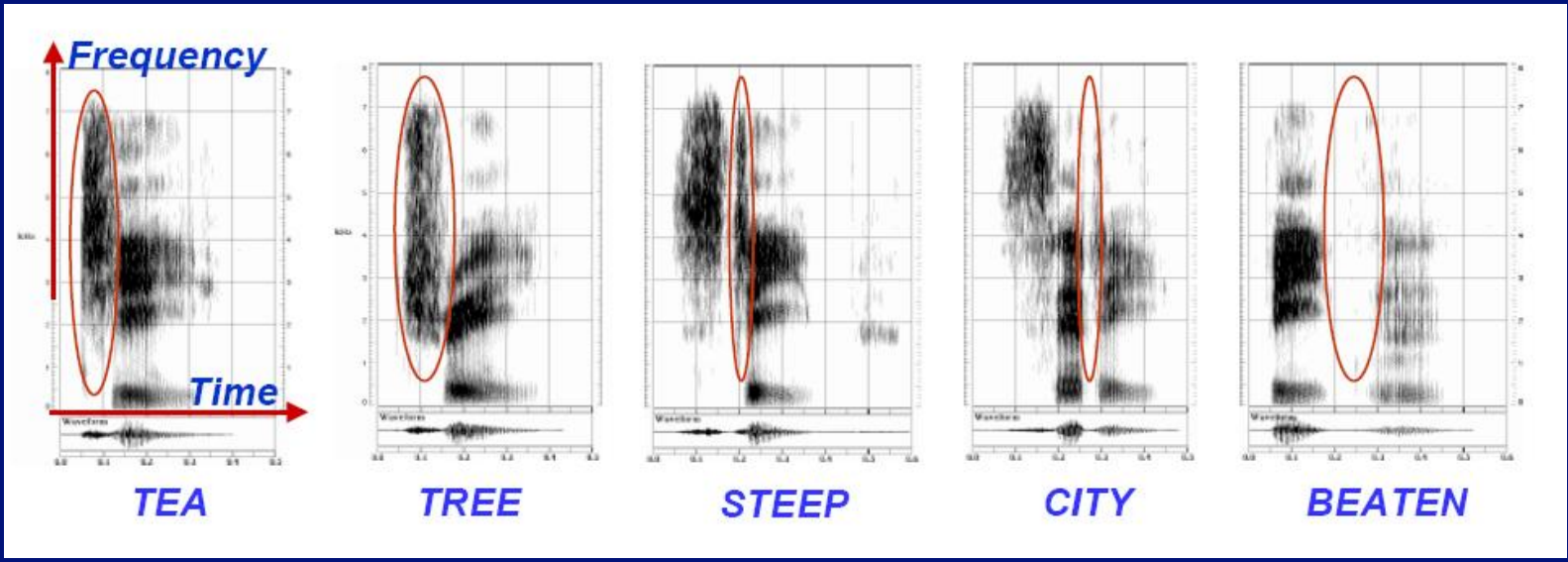
- Ενώ υπάρχει ένα πεπερασμένο σύνολο φωνημάτων που αναπαριστά τους ήχους μιας γλώσσας,
- Δεν υπάρχει μια 1-1 σχέση μεταξύ των φωνημάτων και των ακουστικών προτύπων που προκύπτουν στην ομιλία.

π.χ.: το φώνημα /t/ θα έχει διαφορετική ακουστική αναπαράσταση ανάλογα με τη θέση του σε μια λέξη:

- σε αρχική θέση (όπως στο «τύπος"),
- μεταξύ ενός /s/ και ενός φωνήεντος (όπως στο «στέμμα"),
- μέσα σε μια ομάδα συμφώνων (όπως στο «στρέφω"),
- σε τελική θέση (όπως στο «κατ’ ").

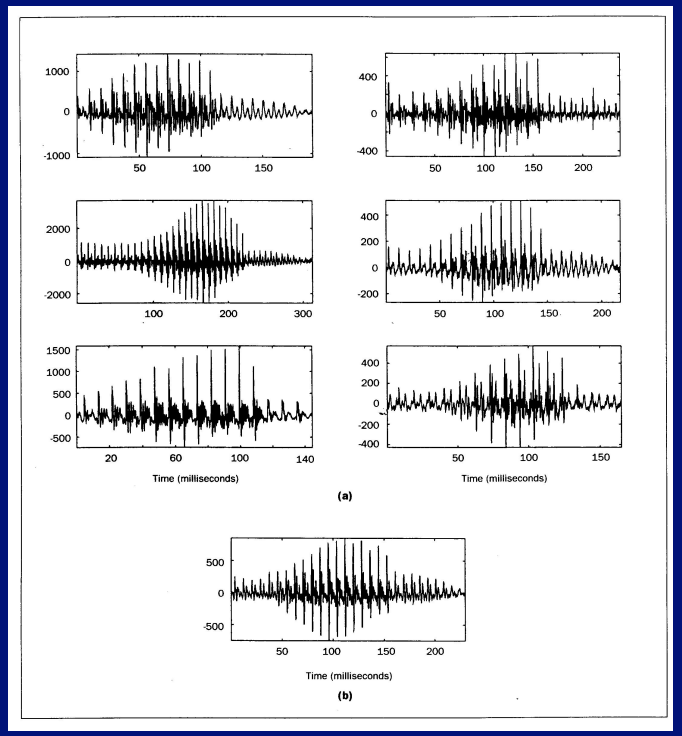
Παράδειγμα φωνολογικής μεταβλητότητας

- Η ακουστική πραγμάτωση ενός φωνήματος εξαρτάται πάρα πολύ από το φωνητικό του περιβάλλον. π.χ. /t/



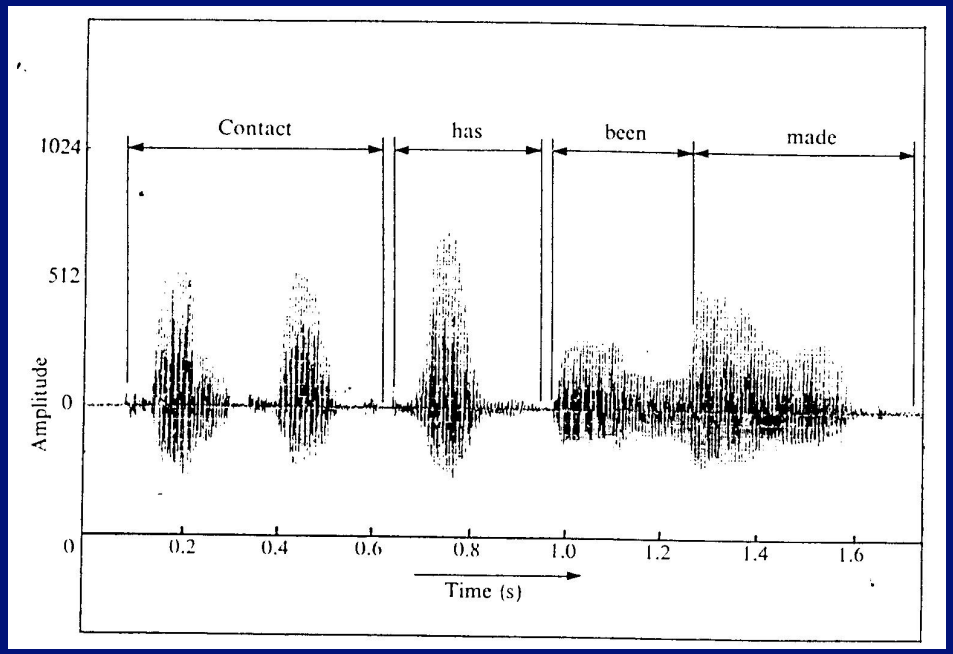
Εκφωνήσεις της λέξης “one”

(α) Έξι διαφορετικά σήματα που παριστάνουν εκφωνήσεις της λέξης “one”
(β) Τυπικό σήμα εκφώνησης της λέξης “one”
Ένα σύστημα αναγνώρισης ομιλίας θα πρέπει να μάθε να αποδίδει τις παραλλαγές (α) στο τυπικό σήμα (β).

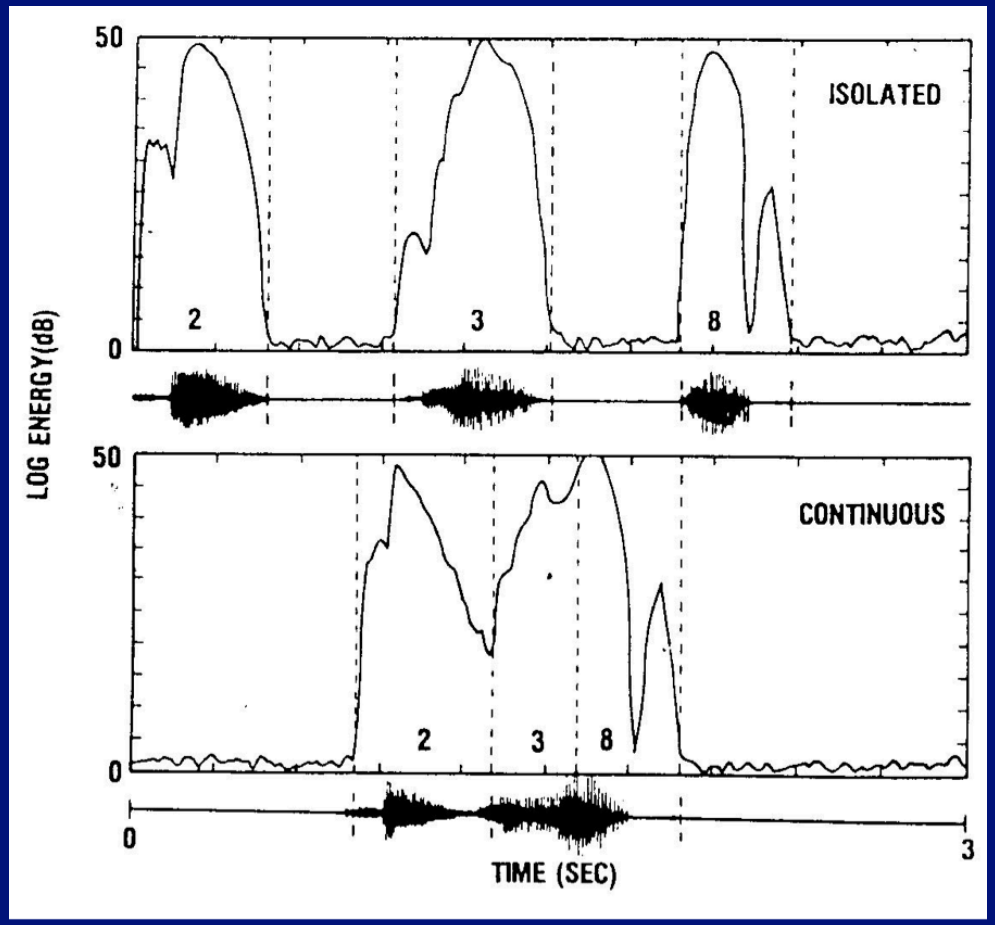


Παράδειγμα του προβλήματος διαχωρισμού των λέξεων σε συνεχή ομιλία

Η λέξη contact εμφανίζεται σαν δύο σχεδόν χωριστά σήματα ενώ οι λέξεις been made αποτελούν συνεχές σήμα.



Παράδειγμα εκφώνησης των λέξεων two three eight μεμονωμένα και συνδεδεμένα ή συνεχόμενα



Γλωσσολογικοί παράγοντες που επηρεάζουν την προφορά των λέξεων (1 από 3)

- Συνάρθρωση:
  - η προφορά των φωνημάτων επηρεάζεται από το περιβάλλον τους
  - επηρεάζει ήχους στα όρια λέξεων.

    π.χ.: το "the" μπορεί να προφερθεί ως "ni" όπως στην φράση "in the event»

    αφομοίωση του /th/ στο προηγούμενο ρινικό φώνημα /n/ και

    αφομοίωση του /uh/ φωνήεντος στο επακόλουθο φωνήεν στη λέξη "event»
- Σε μερικές περιπτώσεις ένα φώνημα μπορεί να παραλειφθεί εντελώς:

  Π.χ.: "good boy" όταν προφερθεί γρήγορα, το /d/ μπορεί να παραλειφθεί εντελώς "gu boy".

Γλωσσολογικοί παράγοντες που επηρεάζουν την προφορά των λέξεων (2 από 3)

- Η συγκεκριμένη προφορά μιας λέξης μπορεί να διαφέρει σημαντικά εξαρτώμενη από το αν η λέξη είναι τονισμένη ή όχι σε μια εκφώνηση

  π.χ.: το "the" μπορεί να προφερθεί "thee" όπως στην φράση "in the event" αν ο εκφωνητής προφέρει πολύ αργά και ηθελημένα, αλλά είναι πιο πιθανό να προφερθεί "ni" σε πιο φυσιολογική ομιλία.

Γλωσσολογικοί παράγοντες που επηρεάζουν την προφορά των λέξεων (3 από 3)

Οι πιο πολλές από τις διεργασίες:

- Είναι προβλέψιμες βάσει του γλωσσολογικού περιβάλλοντος, μπορούν να μοντελοποιηθούν χρησιμοποιώντας τρίφωνα (ή φωνήματα εντός πλαισίου), στα οποία κάθε φώνημα αναπαρίσταται με όρους του προηγούμενου και του επακόλουθου φωνητικού περιβάλλοντος.

  π.χ: το /t/ στη λέξη "strong" αναπαρίσταται ως s/t/r.

- Κάποιες από αυτές τις διαδικασίες είναι πιθανολογικές: μοντελοποιούνται με βάση τη στατιστική που προέρχεται από την ανάλυση μεγάλων σωμάτων δεδομένων ομιλίας

Άλλοι τύποι μεταβλητότητας στην αναγνώριση ομιλίας (1 από 3)

Μεταβλητότητα μεταξύ ομιλητών:

- Υπάρχουν διαφορές μεταξύ των ομιλητών στον τρόπο που μιλούν και προφέρουν τις λέξεις.
- Κάποιες από αυτές τις διαφορές οφείλονται σε φυσικούς παράγοντες, όπως είναι το σχήμα της φωνητικής οδού.
- Άλλες διαφορές οφείλονται σε παράγοντες όπως η ηλικία, το φύλο και η τοπική καταγωγή (προφορά).

Άλλοι τύποι μεταβλητότητας στην αναγνώριση ομιλίας (2 από 3)

Μεταβλητότητα στον ίδιο ομιλητή:

- Οι ίδιες λέξεις εκφωνημένες σε διαφορετικές περιστάσεις από τον ίδιο ομιλητή, μπορεί να έχουν διαφορετικές ακουστικές ιδιότητες.
- Παράγοντες όπως η κούραση, φραγμένοι αεραγωγοί λόγω κρυώματος, και διαφοροποιήσεις στη διάθεση επιδρούν στον τρόπο που προφέρονται οι λέξεις.
- Ακόμα και όταν ένα άτομο κάνει μια συνειδητή προσπάθεια να προφέρει μια λέξη με τον ίδιο ακριβώς τρόπο, μπορεί να υπάρχουν αισθητές διαφορές στο ακουστικό σήμα που θα μπορούσαν να προκαλέσουν σφάλμα στην αναγνώριση ομιλίας.

Άλλοι τύποι μεταβλητότητας στην αναγνώριση ομιλίας (3 από 3)

Μεταβλητότητα καναλιού

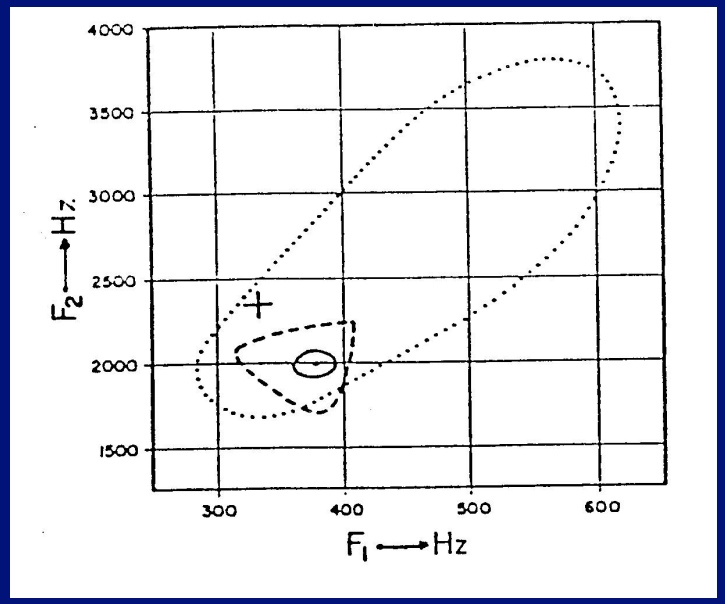
- Διαφορές στα κανάλια μετάδοσης - μικρόφωνο ή τηλέφωνο (σταθερό ή κινητό) - εγείρουν διαφορές στο ακουστικό σήμα.

Θόρυβος υπόβαθρου

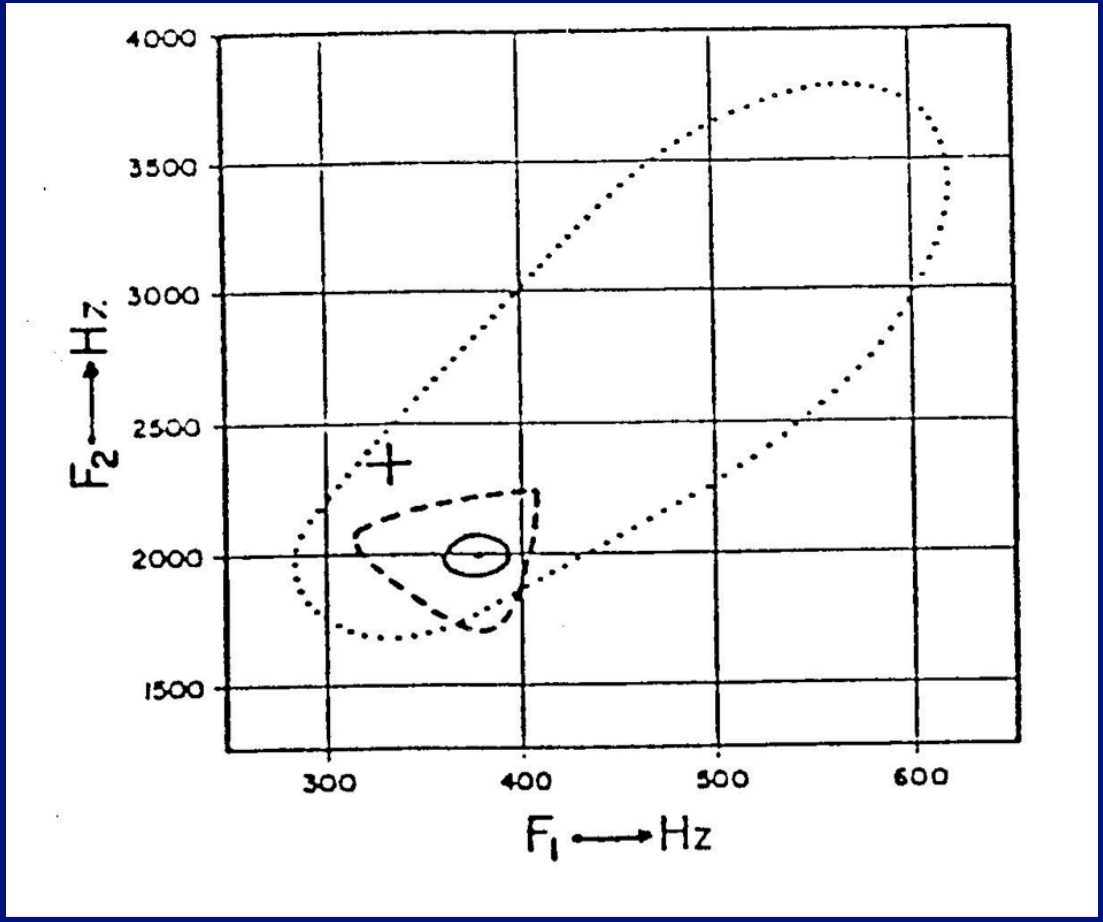
- Το ακουστικό σήμα μπορεί να αλλοιωθεί λόγω επίδρασης θορύβου από το περιβάλλον,
  - είτε αυτός είναι συνεχής, όπως στην περίπτωση του βουητού ενός ανεμιστήρα υπολογιστή,
  - είτε παροδικός, όπως στην περίπτωση ενός φταρνίσματος ή μιας πόρτας που κλείνει με δύναμη.

Μεταβλητότητα μεταξύ ομιλητών και στον ίδιο ομιλητή για διαφορετικά φωνητικά περιβάλλοντα (1 από 2)

Μέρος του φωνητικού χώρου F1, F2
που δείχνει την κατανομή του φωνήεντος
/I/ σε περιβάλλον συφώνω [h-d], όπως
μετρήθηκε από 76 ομιλητές
(μεγάλη εστιγμένη γραμμή).Η κατανομή
του ίδιου φωνήεντος στο ίδιο φωνητικό
περιβάλλον όπως εκφωνήθηκε από τον
ίδιο ομιλητη (μικρή συνεχής γραμμή).
Η κατανομή του ίδιου φωνήεντος όπως
Εκφωνήθηκε από τον ίδιο ομιλητή σε 576
Διαφορετικά φωνητικά περιβάλλοντα συμφώνων (μεσαία γραμμή -----).



Μεταβλητότητα μεταξύ ομιλητών και στον ίδιο ομιλητή για διαφορετικά φωνητικά περιβάλλοντα (2 από 2)



Αναγνώριση Ομιλίας: που είμαστε σήμερα;

- Είναι δυνατή η αναγνώριση ομιλίας υψηλής ποιότητας, ανεξάρτητη του ομιλητή:
  - Μεγάλου λεξιλογίου (για συνεργαζόμενους ομιλητές σε μη εχθρικά περιβάλλοντα)
  - Μέσου λεξιλογίου (για αυθόρμητη ομιλία μέσω τηλεφώνου)
- Υπάρχουν διαθέσιμα εμπορικά προϊόντα:

  π.χ. IBM, Scansoft (Dragon, L&H, Philips, Nuance) για εφαρμογές υπαγόρευσης και για διαλογικές εφαρμογές
- Στην περίπτωση καλοστημένων εφαρμογών, η τεχνολογία είναι σε θέση να βοηθήσει στην περαίωση πραγματικών εργασιών π.χ. Τραπεζικών συναλλαγών

Βασικές μετρικές αξιολόγησης συστημάτων αναγνώρισης ομιλίας

- Word Error Rate (WER)
- Graph Error Rate (GER)
- Phoneme Error Rate (PER)
- Sentence Error Rate (SER)

Word Error Rate

- The most widely used measurement in speech recognition research.
- Evaluates the output of a speech recognition system on a word –by–word basis
- The words in the output are aligned against a given reference transcription of the spoken utterance. With this alignment each word in the output is categorized into four classes: correct, substitution, insertion and deletion (σωστό, υποκατάσταση, εισαγωγή, απαλοιφή).
- The word error rate is than computed as:

$$\mathrm{WER} = \frac{\#del + \#ins + \#sub}{\#spoken\ words}$$

Παράδειγμα υπολογισμού WER (I:insertion, S=Substitution, D=Deletion)

REF:	i	***	**	UM	the	PHONE	IS	i	LEFT	THE	portable
HYP:	i	GOT	IT	TO	the	*****	FULLEST	i	LOVE	TO	portable
Eval:		I	I	S		D	S		S	S	

REF:	****	PHONE	UPSTAIRS	last	night	so	the	battery	ran	out
HYP:	FORM	OF	STORES	last	night	so	the	battery	ran	out
Eval:	I	S	S							

This utterance has 6 substitutions, 3 insertions, and 1 deletion:

$$\text{Word Error Rate} = 100\frac{6+3+1}{18} = 56\%$$

Graph Error Rate

- Most speech recognition systems do not only produce a single transcription of a spoken utterance but a network of possible transcriptions with different probabilities at a given word position. These so called "word graphs" are used for post−processing the recognition output, e.g. rescoring with a refined language model.
- To compute the graph error rate, the transcription with the least error rate of all possible transcriptions in the word graph is used for scoring. This transcription may not be the most likely one (which would be produced if only a single transcription is wanted) because of sub optimal acoustic and/or language models. So the graph error rate is always equal or less than the word error rate.

Phoneme Error Rate

- For this measure not only an orthographic transcription of the spoken words is needed but also a phonetic transcription.
- This measure is the most detailed of the measures described here.
- In comparison to the word error rate the phoneme error rate puts more emphasis on the acoustic confusability of the words in the vocabulary of the speech recognizer.
- For scoring, the recognition output and reference transcriptions are reduced to the phonemes. Again, the reference transcription and the recognizer output are aligned, but on a phoneme−by−phoneme basis.
- Afterwards, the PER is computed analogous to the WER:

$$\mathrm{PER} = \frac{\#del + \#ins + \#sub}{\#spoken\ phonems}$$

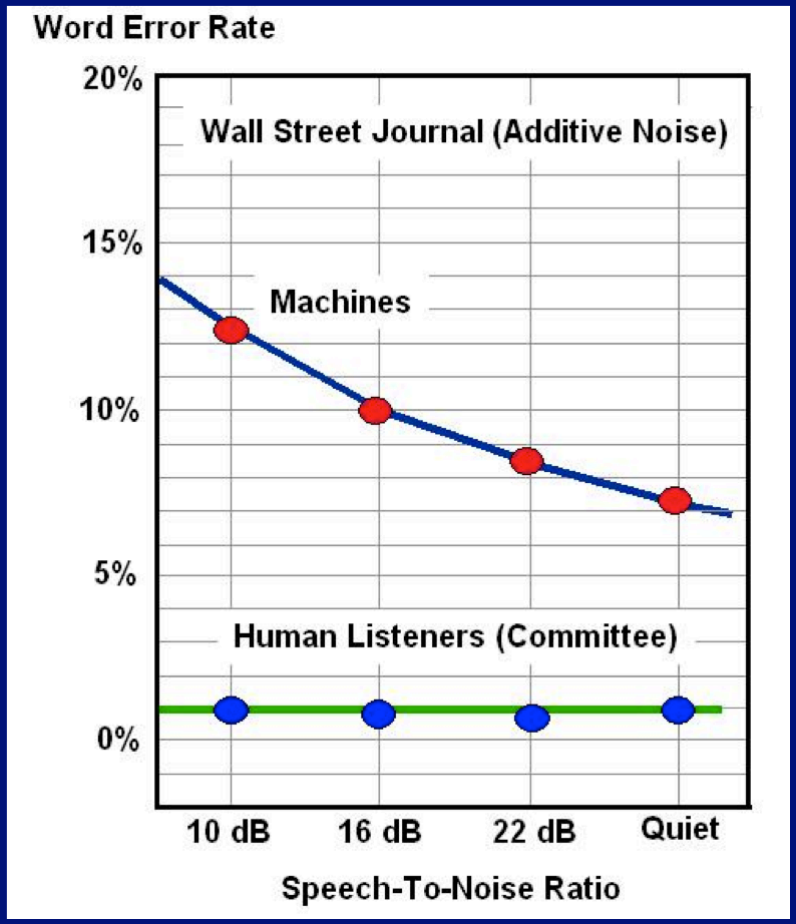
Sentence Error Rate (1 από 3)

- A sentence is considered to have an error if one or more word errors occur in this sentence:

$$\mathrm{SER} = \frac{\#\,erroneous\ sentences}{\#\,spoken\ sentences}$$

- This measure is sub−optimal (and less importance), as it does not take into account the length of a sentence.

Sentence Error Rate (2 από 3)



Sentence Error Rate (3 από 3)

- Human performance exceeds machine performance by a factor ranging from 4x to 10x depending on the task.
- On some tasks, such as credit card number recognition, machine performance exceeds humans due to human memory retrieval capacity.
- The nature of the noise is as important as the SNR (e.g., cellular phones).
- A primary failure mode for humans is inattention.
- A second major failure mode is the lack of familiarity with the domain (i.e., business terms and corporation names).

Βασικές κλάσεις μεθόδων Αυτόματης Αναγνώρισης Ομιλίας (ΑΑΟ) (1 από 5)

α) Ταίριασμα προτύπων ιχνών (template matching) ή δομικές μέθοδοι ή ντετερμινιστικές μέθοδοι.

β) Στοχαστικές: Κρυφά Μοντέλα Markov HMM (Hidden Markov Models) και Τεχνητά Νευρωνικά Δίκτυα – ANN (Artificial Neural Networks).

γ) Βασισμένες σε γνώση (φωνητική και γλωσσολογική) (knowledge based)

Οι α) και β) ονομάζονται και **Μέθοδοι Βασισμένες σε Δεδομένα** (data-based approaches): το σήμα της ομιλίας μοντελοποιείται με αλγορίθμους που μπορούν να εξάγουν γνώση αυτόματα από τα δεδομένα.

Βασικές κλάσεις μεθόδων Αυτόματης Αναγνώρισης Ομιλίας (ΑΑΟ) (2 από 5)

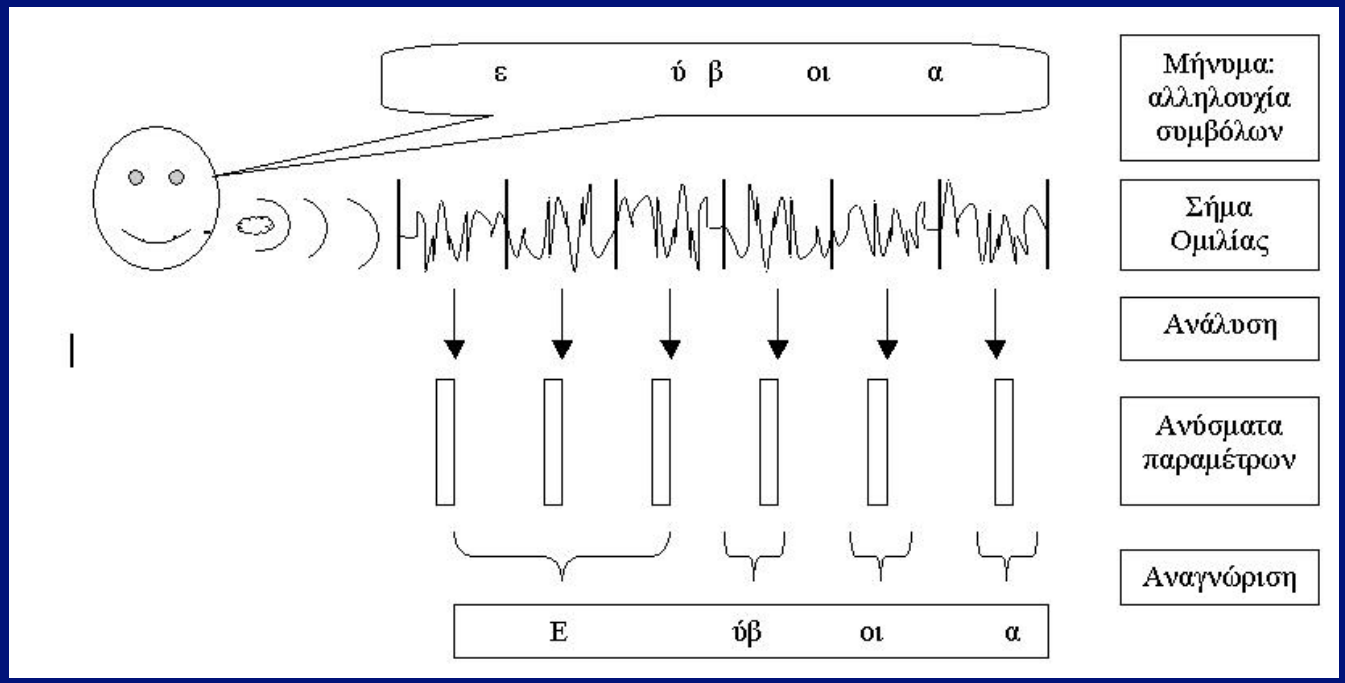
Στη γ) μέθοδο σκοπός είναι να εκφραστεί η γνώση που έχει ο άνθρωπος για την ομιλία με ένα σύνολο αποκλειστικών κανόνων, όπως: ακουστικοί-φωνητικοί κανόνες, κανόνες που περιγράφουν τις λέξεις ενός λεξικού, κανόνες που περιγράφουν τη σύνταξη της γλώσσας, και ούτω καθεξής. Οι ευριστικοί (εμπειρικοί) κανόνες συλλέγονται από ανθρώπους εμπειρογνώμονες.

Υβριδικές μέθοδοι: Κάθε συνδυασμός δύο ή περισσοτέρων από τις βασικές μεθόδους

Βασικές κλάσεις μεθόδων Αυτόματης Αναγνώρισης Ομιλίας (ΑΑΟ) (3 από 5)

Κωδικοποίηση/ αποκωδικοποίηση μηνύματος ομιλίας

Η αλληλουχία συμβόλων μπορεί να περιλαμβάνει: φωνήματα ή λέξεις ή μέρη λέξεων όπως (διφωνήματα, τριφωνήματα, συλλαβές, κλπ)



Βασικές κλάσεις μεθόδων Αυτόματης Αναγνώρισης Ομιλίας (ΑΑΟ) (4 από 5)

- Γενικά όλες οι μέθοδοι ΑΑΟ υποθέτουν ότι το σήμα ομιλίας είναι μια πραγματοποίηση της κωδικοποίησης κάποιου μηνύματος σε μια αλληλουχία ενός ή περισσοτέρων συμβόλων.
- Το πρόβλημα της ΑΑΟ είναι να αναγνωρίζει την αλληλουχία των συμβόλων από μια δεδομένη εκφώνηση.
- Στην αρχή το συνεχές σήμα της ομιλίας μετατρέπεται σε μια αλληλουχία διακριτών και σε ίσες (χρονικές) αποστάσεις ανυσμάτων παραμέτρων.
- Η αλληλουχία των ανυσμάτων παραμέτρων θεωρείται ότι αποτελεί μια ακριβή αναπαράσταση του σήματος ομιλίας στη βάση ότι για τη χρονική διάρκεια που αναπαριστά ένα άνυσμα (που ονομάζεται πλαίσιο και τυπικά είναι 10-30 msec) το σήμα της ομιλίας μπορεί να θεωρηθεί στατικό.

Βασικές κλάσεις μεθόδων Αυτόματης Αναγνώρισης Ομιλίας (ΑΑΟ) (5 από 5)

Τυπικές παράμετροι που συνήθως χρησιμοποιούνται στην ΑΑΟ είναι:

- οι συντελεστές λειασμένου φάσματος,
- οι συντελεστές γραμμικής πρόβλεψης
- άλλες αναπαραστάσεις που προκύπτουν από αυτούς τους δύο, όπως: συντελεστές cepstrum, formants, έξοδοι τράπεζας φίλτρων, καθώς και όσες αναφέρθηκαν στο κεφάλαιο της ανάλυσης ομιλίας.

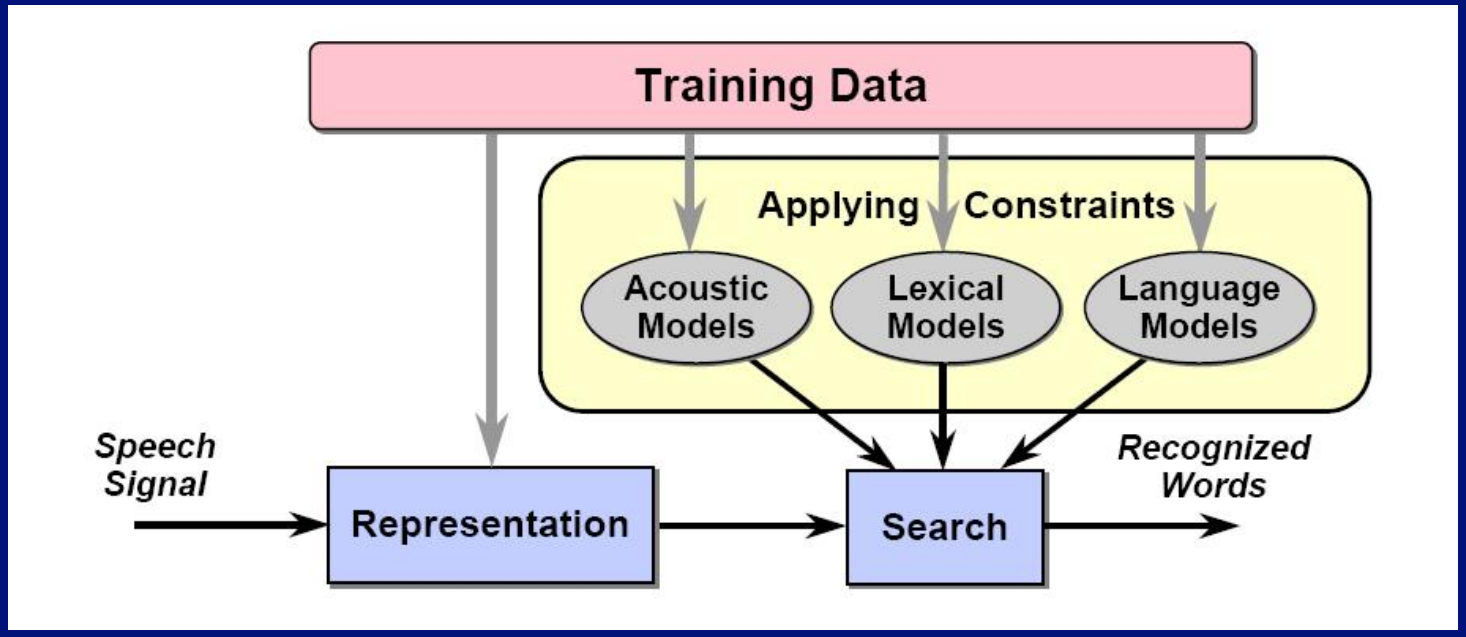
Το άνυσμα παραμέτρων είναι Ν διαστάσεων, όπου Ν α αριθμός των παραμέτρων που χρησιμοποιείται κάθε φορά.

Το μέρος αυτό κάθε συστήματος ΑΑΟ ονομάζεται πολλές φορές **Μετωπικός Επεξεργαστής** (front-end processor) και η αντίστοιχη διαδικασία **μετωπική επεξεργασία** (front-end processing).

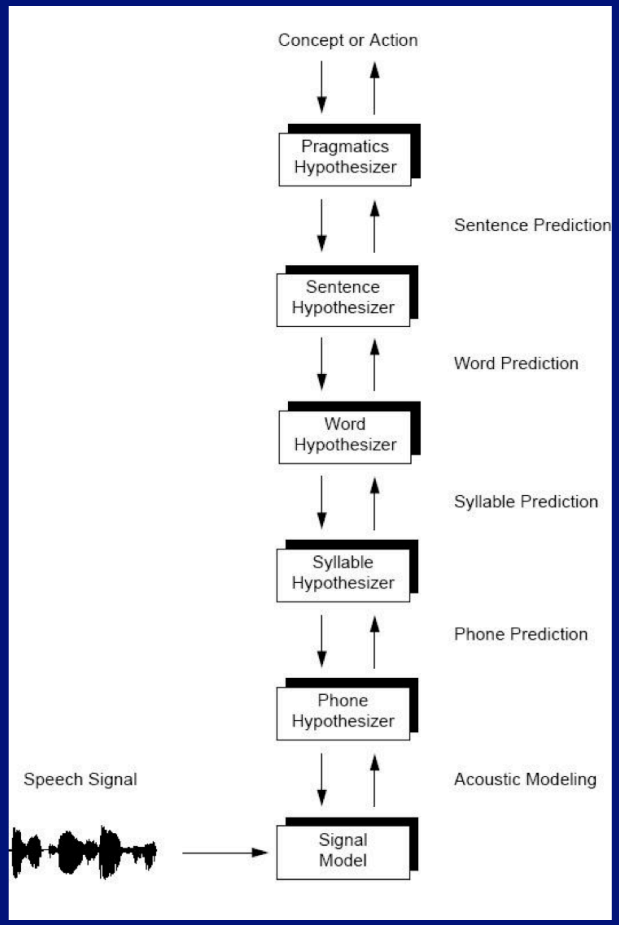
Major Components in a Speech Recognition System

Speech recognition is the problem of deciding on:

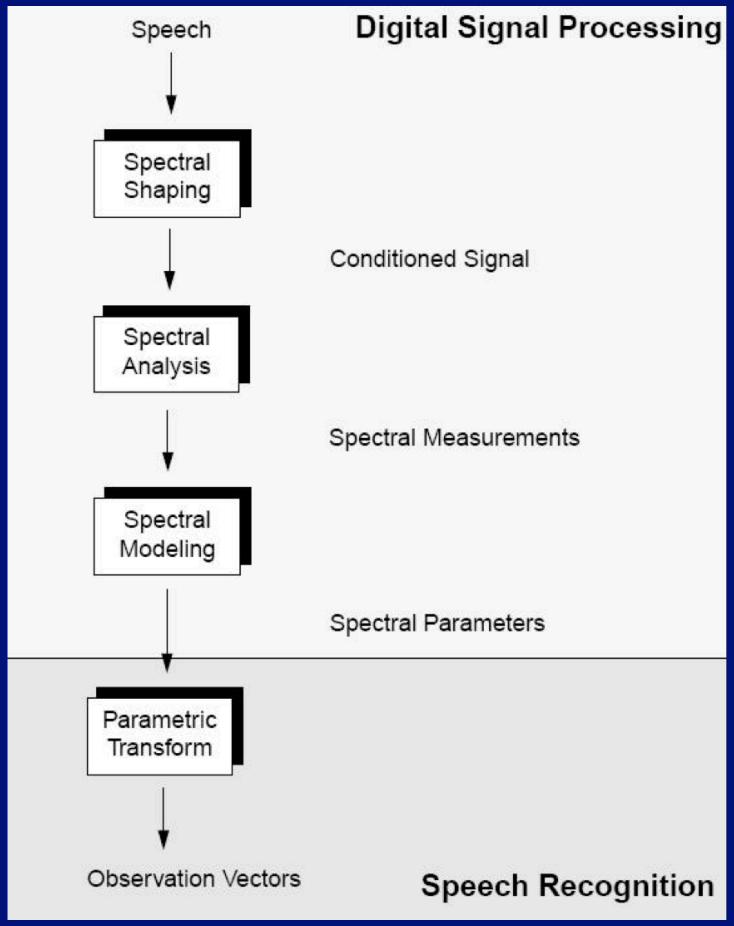
- How to r**epresent** the signal
- How to **model** the constraints
- How to **search** for the most optimal answer



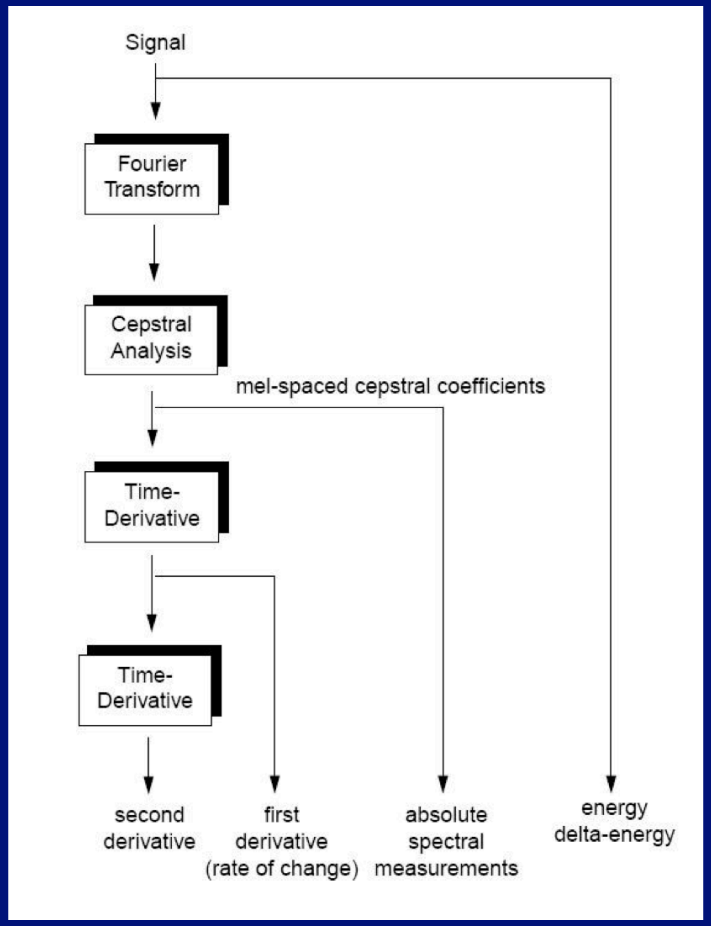
Γενικό Διάγραμμα αναγνώρισης και κατανόησης ομιλίας



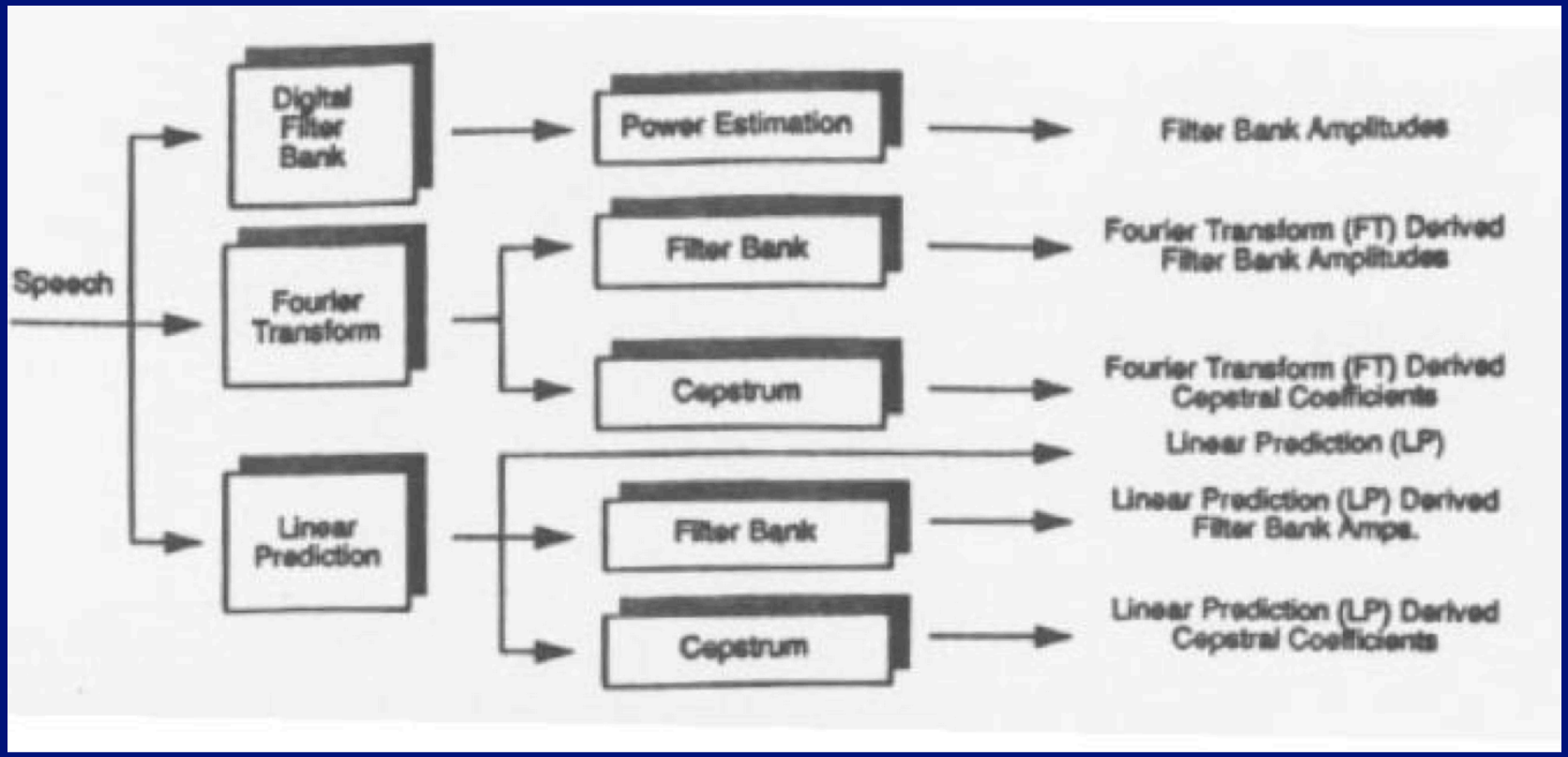
Γενικό Μετωπικό Σύστημα αναγνώρισης ομιλίας



Μετωπικό Σύστημα αναγνώρισης ομιλίας



Μέθοδοι υπολογισμού παραμέτρων ομιλίας



Αναγνώριση Φωνημάτων και Λέξεων

- Λαμβάνει ένα σύνολο χαρακτηριστικών που εξήχθησαν από το ακουστικό σήμα και:
  - Τα ταξινομεί ως φωνήματα
  - Συνδυάζει ακολουθίες φωνημάτων σε λέξεις.

Αυτή η διαδικασία περιλαμβάνει δύο μοντέλα:

- Ένα **ακουστικό μοντέλο** που δείχνει:
  - πως κάθε λέξη αποτελείται από μια ακολουθία φωνημάτων, και
  - πως κάθε φώνημα σχετίζεται με τις τιμές των χαρακτηριστικών που εξήχθησαν από το ακουστικό σήμα.
- Ένα **γλωσσικό μοντέλο** που προσδιορίζει επιτρεπτές ακολουθίες λέξεων.

Το ακουστικό μοντέλο

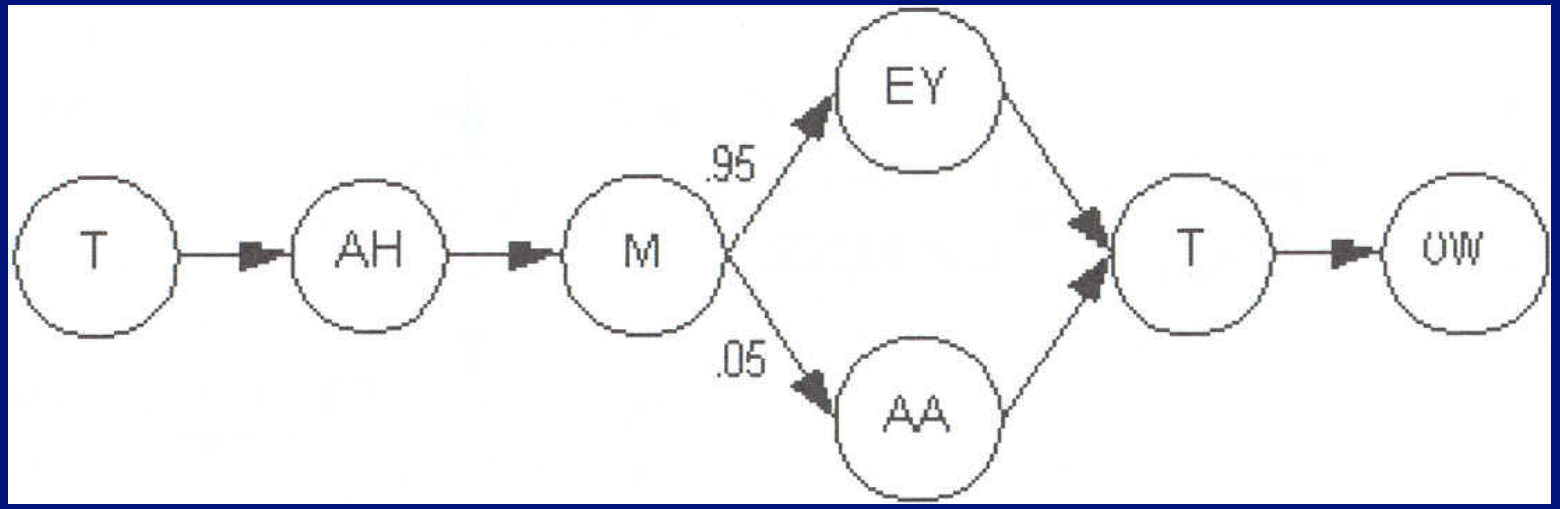
- Το ακουστικό μοντέλο «συλλαμβάνει» τη μεταβλητότητα στην προφορά χρησιμοποιώντας πιθανότητες.
- Ένα μοντέλο λέξεων αποτελείται από τα φωνήματα που αποτελούν τη λέξη.
- Παρόλα αυτά, μια λέξη μπορεί να προφερθεί με ποικίλους τρόπους.

π.χ., το Λεξικό Προφοράς του πανεπιστήμιου Carnegie Mellon (CMU) (*διαθέσιμο στην διεύθυνση: http://www.speech.cs.cmu.edu/cgi-bin/cmudict*) παράγει τις ακόλουθες δύο προφορές για τη λέξη "tomato":

1. T AH M EY T OW (*American English*)
2. T AH M AA T OW (*British English*)

Ακουστικό μοντέλο της λέξης «tomato»

- Στην πραγματικότητα το μοντέλο είναι περισσότερο πολύπλοκο:
  - Πρώτο φωνήεν: AH ή OW
  - Τελευταίο σύμφωνο: “t” ή “d”



Κρυφά Μοντέλα Markov (HMM) για Ακουστική Μοντελοποίηση (1 από 4)

1. Αποτελείται από έναν αριθμό **καταστάσεων** που αναπαριστούν την χρονική πρόοδο μιας λέξης, από την αρχική κατάσταση, και μέσω όλων των φωνημάτων, στην τελική κατάσταση.

Καθώς η διάρκεια κάθε φωνήματος μπορεί να ποικίλει λόγω διαφορών στο ρυθμό εκφώνησης, τα τόξα έχουν μεταβάσεις επαναφοράς(που συμβολίζονται με all, α22 , ...) που επιτρέπουν στο μοντέλο να παραμείνει στην ίδια κατάσταση ώστε να εκφράσει μια πιο αργή μορφή εκφώνησης.

Με αυτό τον τρόπο ένα HMM αντιμετωπίζει τη χρονική μεταβλητότητα σε ένα σήμα ομιλίας.

Κρυφά Μοντέλα Markov (HMM) για Ακουστική Μοντελοποίηση (2 από 4)

2. Κάθε κατάσταση έχει μια **κατανομή από πιθανές εξόδους** (συμβολίζονται με b1(o1), b1(o2), b2(o3) ....).

Σε κάθε κατάσταση το σύστημα ταιριάζει ένα μέρος της εισόδου (ένα πλαίσιο διάρκειας 10ms, που συμβολίζεται με o1, o2, o3.....), με όλες τις πιθανές εξόδους, κάθε μια από τις οποίες έχει διαφορετική πιθανότητα.

Η έξοδος που ταιριάζει (καλύτερα) επιστρέφεται μαζί με την πιθανότητά της.

Με αυτό τον τρόπο αντιμετωπίζεται η ακουστική μεταβλητότητα του σήματος ομιλίας.

Κρυφά Μοντέλα Markov (HMM) για Ακουστική Μοντελοποίηση (3 από 4)

3. Κάθε έξοδος μπορεί να βρίσκεται σε περισσότερες από μια καταστάσεις.

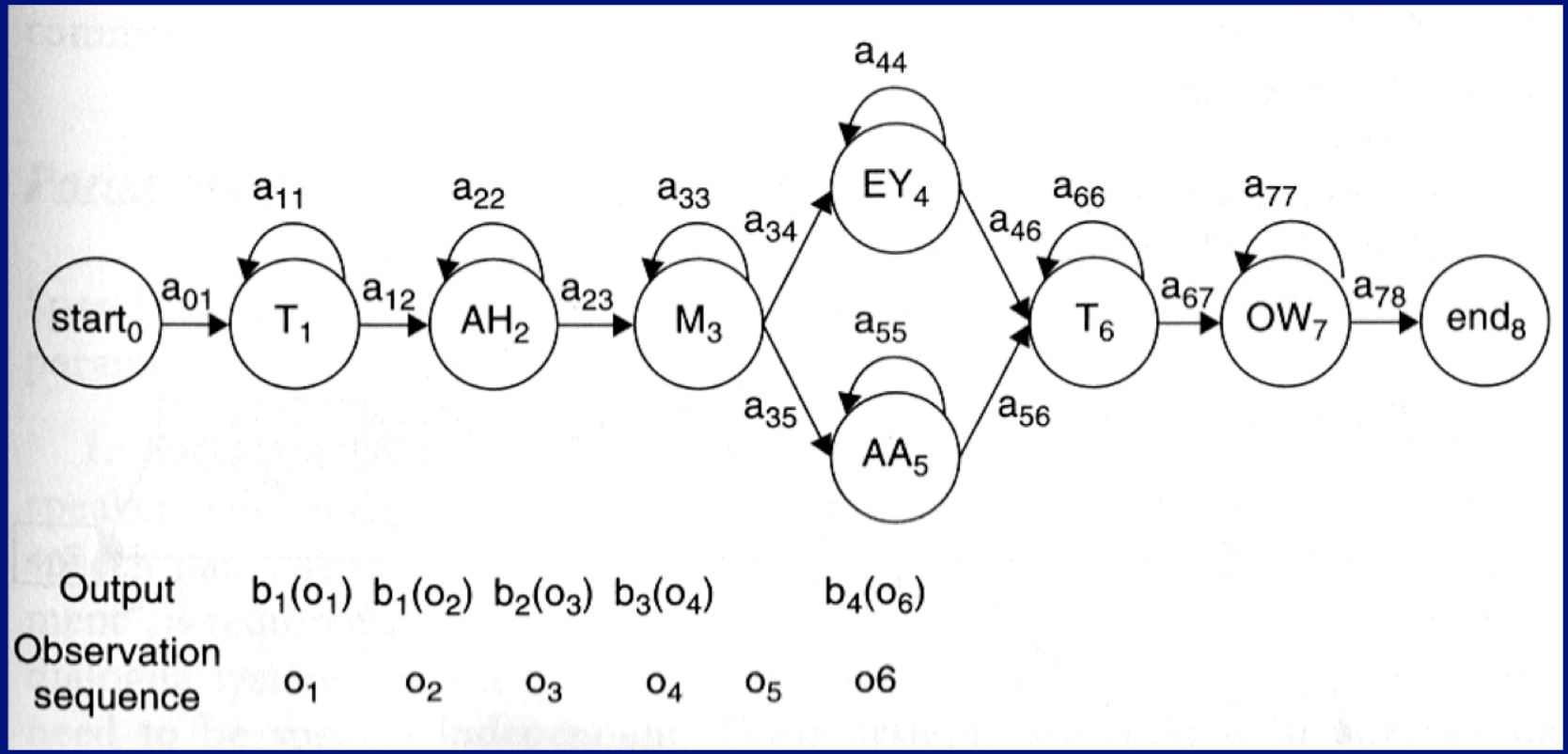
Είναι πιθανό να δει κανείς ότι το μοντέλο παράγει ένα συγκεκριμένο σύμβολο σαν έξοδο αλλά δεν είναι δυνατό να γνωρίζουμε από ποια κατάσταση προήρθε αυτό το σύμβολο.

- Το ταίριασμα προτύπων με χρήση HMM περιλαμβάνει υπολογισμό της πιθανότητας μιας ακολουθίας καταστάσεων.

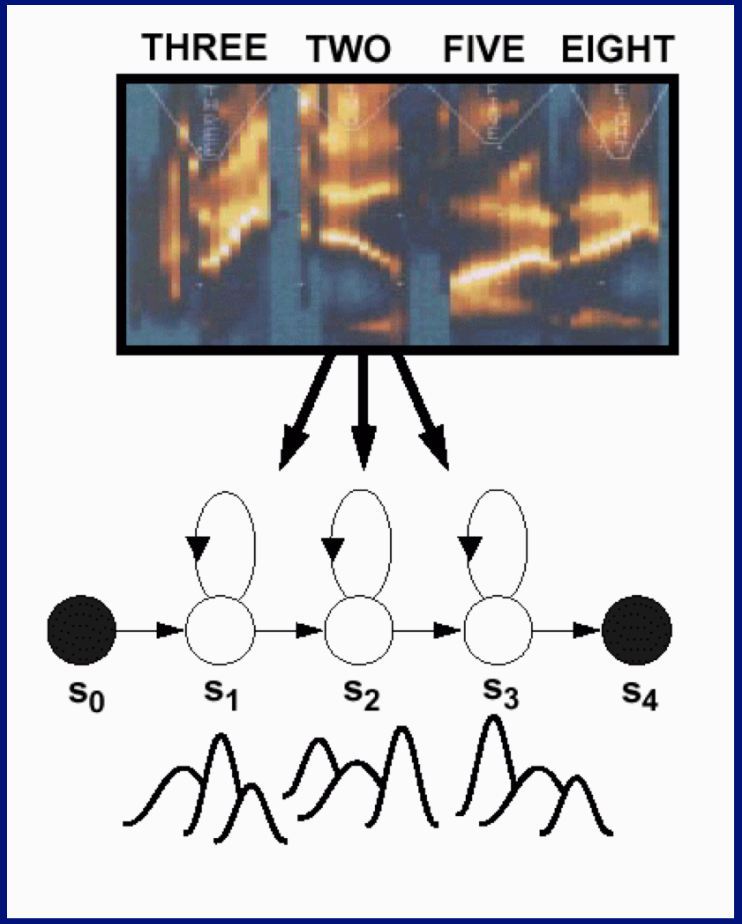
Κρυφά Μοντέλα Markov (HMM) για Ακουστική Μοντελοποίηση (4 από 4)

- Καθώς ο αριθμός των πιθανών ακολουθιών καταστάσεων, για μια συγκεκριμένη ακολουθία πλαισίων που αντιπροσωπεύουν το σήμα ακουστικής εισόδου, είναι πολύ μεγάλος, μια διαδικασία που ονομάζεται δυναμικός προγραμματισμός χρησιμοποιείται για να εκτελεστεί βέλτιστα ο υπολογισμός.
- πιο συχνά χρησιμοποιούμενοι αλγόριθμοι: Viterbi και A* αλγόριθμοι

Κρυφό Μοντέλο Markov της λέξης "tomato» (1 από 2)



Κρυφό Μοντέλο Markov της λέξης "tomato" (2 από 2)



Το γλωσσικό μοντέλο

- Περιέχει γνώση σχετικά με:
  - επιτρεπτές ακολουθίες λέξεων, και
  - το ποιες λέξεις είναι πιο πιθανές σε συγκεκριμένες ακολουθίες.

Υπάρχουν δύο τύποι γλωσσικών μοντέλων που χρησιμοποιούνται συνήθως:

1. Μια γραμματική (ή δίκτυο πεπερασμένων καταστάσεων):

Όλες οι επιτρεπτές ακολουθίες λέξεων στην εφαρμογή προσδιορίζονται.

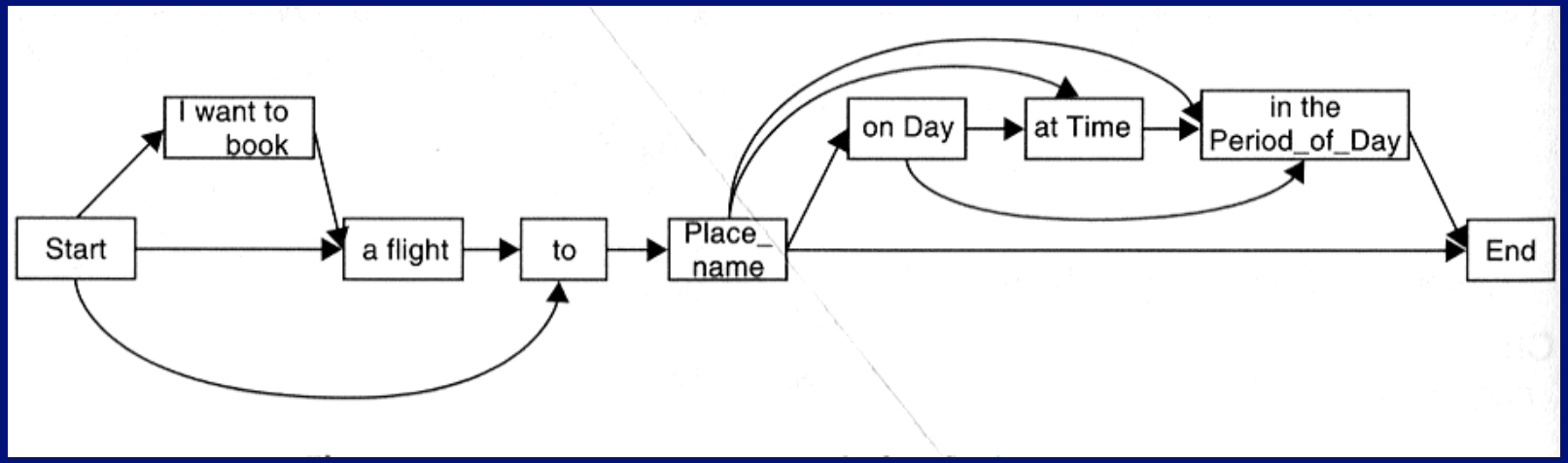
2. Ένα N-gram γλωσσικό μοντέλο:

Παρέχει στατιστικές πληροφορίες για ακολουθίες λέξεων.

(Παρατήρηση: η λέξη «γραμματική» προκαλεί σύγχυση, καθώς χρησιμοποιείται επίσης για αναφορά στην ανάλυση προτάσεων με όρους συντακτικών κατηγοριών, όπως ουσιαστικά, ρήματα και προθέσεις)

Απλό δίκτυο γραμματικής (1 από 4)

Ένα απλό δίκτυο γραμματικής που θα μπορούσε να χρησιμοποιηθεί για την γραμματική ανάλυση της απάντησης ενός χρήστη σε ένα αυτόματο σύστημα αεροπορικών κρατήσεων.



Απλό δίκτυο γραμματικής (2 από 4)

Αυτή η γραμματική καλύπτει διάφορους τρόπους με τους οποίους θα μπορούσε να απαντήσει ο χρήστης, όπως:

1. Θέλω να κλείσω μια πτήση για Λονδίνο την Παρασκευή στις 11 το πρωί.
2. Μια πτήση για Λονδίνο;
3. Για Λονδίνο το πρωί.
4. Για Λονδίνο στις 11 το πρωί.

Απλό δίκτυο γραμματικής (3 από 4)

Δεδομένου ότι είναι πιθανό να υπάρχει πεπερασμένο πλήθος τρόπων με τους οποίους μπορεί ένας χρήστης να προσδιορίσει απαιτήσεις πτήσεων, θα μπορούσε να κατασκευαστεί μια γραμματική ώστε να καλύψει αυτές τις πιθανότητες.

π.χ., μπορεί να αναγνωρισθεί ένας αριθμός ομαλά επαναλαμβανόμενων φράσεων, όπως:

για + Όνομα_Προορισμού

στις + Ημερομηνία

το + Περίοδος_της_Ημέρας (π.χ. μεσημέρι, απόγευμα κ.α.)

στις | περίπου τις | πριν τις | μετά τις + Ώρα

Απλό δίκτυο γραμματικής (4 από 4)

Άλλες φράσεις μπορούν να ταξινομηθούν ως προαιρετικές, καθώς δεν συμβάλλουν στο συνολικό νόημα:

- (Θέλω μια πτήση)
- (Θέλω να κλείσω μια πτήση)

Τέτοιες προαιρετικές εκφράσεις αναγνωρίζονται στην είσοδο αλλά δεν χρησιμοποιούνται καθώς δεν είναι χρήσιμες για την επακόλουθη επεξεργασία.

N-γράμματα (N-grams)

- Χρησιμοποιούνται σε μεγάλα συστήματα λεξιλογίων στα οποία δεν είναι δυνατό να προσδιοριστούν όλες οι επιτρεπτές προτάσεις και οι συνδυασμοί λέξεων εκ των προτέρων.
- Παρέχουν στατιστικές πληροφορίες πάνω σε ακολουθίες λέξεων, δείχνοντας πόσο πιθανή είναι μια λέξη σε ένα συγκεκριμένο πλαίσιο.

π.χ. η ακουστική πρόταση "ni" θα μπορούσε να αντιπροσωπεύει διάφορες λέξεις:

"need", "knee", "neat" και "new"

- στηριζόμενοι μόνο σε ακουστικές μετρήσεις, η πιο πιθανή λέξη θα ήταν η "knee".
- αν η προηγούμενη λέξη ληφθεί υπ’ όψιν και είναι η λέξη "I", τότε πιο πιθανή λέξη γίνεται η λέξη "need".

Παράμετροι στην Αναγνώριση Ομιλίας (1 από 5)

1. Αναγνώριση:

- εξαρτώμενη από τον χρήστη
  - πρέπει να εκπαιδευτεί στα πρότυπα ομιλίας ενός συγκεκριμένου χρήστη
  - για σύγχρονα συστήματα υπαγόρευσης με μεγάλα λεξιλόγια

- ανεξάρτητη από το χρήστη
  - για Συστήματα Προφορικού Διαλόγου,
  - εκπαιδεύονται εκ των προτέρων σε δείγματα συγκεντρωμένα από ποικίλους χρήστες των οποίων τα πρότυπα ομιλίας πρέπει να είναι αντιπροσωπευτικά των δυνητικών χρηστών του συστήματος.
  - είναι πιο επιρρεπής σε λάθη, ειδικά όταν εμπλέκονται χρήστες με μη-τοπική ή, κατά οποιονδήποτε τρόπο, ασυνήθιστη προφορά.

Παράμετροι στην Αναγνώριση Ομιλίας (2 από 5)

2. Μέγεθος Λεξιλογίου

- ποικίλει ανάλογα με την εφαρμογή και το ιδιαίτερο σχεδιασμό του συστήματος διαλόγου.
- ένας προσεκτικά ελεγχόμενος διάλογος μπορεί να «δεσμεύσει» το χρήστη σε ένα λεξιλόγιο περιορισμένο σε μερικές λέξεις που εκφράζουν τις επιλογές που είναι διαθέσιμες μέσα στο σύστημα, ενώ σε ένα πιο εύκαμπτο σύστημα το λεξιλόγιο μπορεί να περιλαμβάνει αρκετές χιλιάδες λέξεις.
- Συστήματα υπαγόρευσης έχουν πολύ μεγάλα λεξιλόγια, άνω των 200,000 λέξεων και μπορεί επίσης να περιλαμβάνουν επιπλέον λεξιλόγια για ειδικευμένα πεδία, όπως νομική ή ιατρική ορολογία.

Παράμετροι στην Αναγνώριση Ομιλίας (3 από 5)

3. Τρόπος ομιλίας

- Απομονωμένες ή διακριτές λέξεις
  - Μέχρι πρόσφατα, πολλά συστήματα ομιλίας απαιτούσαν από τον χρήστη να σταματάει στιγμιαία ανάμεσα σε λέξεις.
- συνεχής λόγος
  - ο χρήστης μπορεί να μιλάει φυσικά.
  - Η βασική δυσκολία που αντιμετωπίζουν συνεχή συστήματα είναι πώς να κατακερματίσουν την ομιλία εντοπίζοντας, ορθά, τα όρια των λέξεων π.χ. : It's hard to recognise speech. / It's hard to wreck a nice beach.

Παράμετροι στην Αναγνώριση Ομιλίας (4 από 5)

- Κάποια συστήματα επιτρέπουν συνδεδεμένες φράσεις που έχουν καθοριστεί σε μορφή γραμματικής που επιτρέπει μικρές παραλλαγές μέσα στις φράσεων. Η αναγνώριση ενωμένων φράσεων χρησιμοποιείται σε συστήματα εντολής-και-ελέγχου,

π.χ., «Άνοιξε το Χ", όπου Χ μπορεί να είναι μία οποιαδήποτε εφαρμογή από ένα πλήθος εφαρμογών σε έναν υπολογιστή.

Παράμετροι στην Αναγνώριση Ομιλίας (5 από 5)

4. Στιλ ομιλίας.

- Κάποια συστήματα αναγνώρισης ομιλίας είναι εκπαιδευμένα σε ομιλία που αναγιγνώσκεται (από ένα κείμενο – Read Speech).
  - συνήθως είναι πιο ομαλή και με λιγότερα λάθη
- Άλλα εκπαιδεύονται σε αυθόρμητο λόγο.
  - συνήθως λιγότερο εύγλωττα, χαρακτηρίζονται από δισταγμούς, λάθος εκκινήσεις, και ποικίλα εξω-γλωσσικά φαινόμενα, όπως π.χ. βήχας.
  - πιο κατάλληλα για ένα σύστημα προφορικού διαλόγου

Θέματα για τους Μηχανικούς Ανάπτυξης

- Γενικά δεν είναι απαραίτητο για τους μηχανικούς ανάπτυξης συστημάτων προφορικού διαλόγου να κατασκευάσουν ένα σύστημα αναγνώρισης ομιλίας από την αρχή
- Οι μηχανές αναγνώρισης ομιλίας γενικά παρέχονται ως μέρη των περιβαλλόντων ανάπτυξης προφορικών διαλόγων.
- Αυτές οι μηχανές συνήθως περιλαμβάνουν ακουστικά μοντέλα για λεξικά των πιο συχνά εμφανιζόμενων λέξεων σε μία γλώσσα.
- Σε περιπτώσεις που μια τέτοια μηχανή δεν παρέχεται, είναι συχνά εφικτό να αποκτηθεί μια κατάλληλη μηχανή για την ζητούμενη γλώσσα από κάποιον άλλο προμηθευτή.

Τα κύρια καθήκοντα ενός μηχανικού ανάπτυξης είναι:

- να ρυθμίσει λεπτομερώς και να επεκτείνει τα ακουστικά μοντέλα
- να αναπτύξει ή να επεκτείνει τις γραμματικές που αποτελούν τα γλωσσικά μοντέλα.

Ρυθμίζοντας την αναγνώριση ομιλίας (1 από 2)

- Οι μηχανές αναγνώρισης ομιλίας συνήθως παρέχουν υπηρεσίες για την λεπτομερή ρύθμισή τους.
- Η λεπτομερής ρύθμιση περιλαμβάνει την πρόσθεση διαφορετικών προφορών στα ακουστικά μοντέλα.

π.χ.: αν το ακουστικό μοντέλο για την "tomato" βασιζότανε μόνο στην προφορά "tom-eh-to", και προβλεπόταν ότι κάποιοι χρήστες μπορεί να πρόφεραν τη λέξη σαν "tom-ah-to", τότε αυτή η εναλλακτική προφορά θα έπρεπε να προστεθεί.

Ρυθμίζοντας την αναγνώριση ομιλίας (2 από 2)

Η λεπτομερής ρύθμιση είναι συνήθως μια επαναληπτική διαδικασία:

- κατασκευάζεται ένα αρχικό σύστημα
- συλλέγονται δεδομένα από ένα σύνολο χρηστών
- τα δεδομένα αναλύονται
- γίνονται αλλαγές στο ακουστικό μοντέλο

Συχνά αυτή η διαδικασία είναι αυτοματοποιημένη, αλλά σε μερικές περιπτώσεις οι αλλαγές γίνονται με το χέρι.

Επεκτείνοντας το Ακουστικό Μοντέλο

- Περιλαμβάνει την προσθήκη νέων λέξεων στο λεξιλόγιο.
- Γενικά, τα λεξιλόγια μιας καλής μηχανής αναγνώρισης ομιλίας θα περιλαμβάνουν τις πιο συχνά χρησιμοποιούμενες λέξεις στη γλώσσα.
- Μπορεί να χρειαστεί να προστεθούν νέες λέξεις, όπως τεχνικοί όροι που δεν περιλαμβάνονται στο γενικό λεξιλόγιο και, ειδικότερα, ονόματα τόπων και ανθρώπων.

Δημιουργώντας Γραμματικές Αναγνώρισης Ομιλίας

- Τα περισσότερα εργαλεία για προφορικούς διαλόγους απαιτούν από τους μηχανικούς ανάπτυξης να δημιουργούν γραμματικές αναγνώρισης ομιλίας που προσδιορίζουν την επιτρεπτή είσοδο για κάθε σημείο μέσα στο διάλογο.
- Χρησιμοποιείται ένα πλήθος διαφορετικών σημειογραφιών.

Οι γραμματικές αναγνώρισης ομιλίας και κάποιες από τις σημειογραφίες που χρησιμοποιούνται πιο συχνά θα συζητηθούν περαιτέρω τις επόμενες εβδομάδες.

Σχεδιασμός Συστημάτων Ομιλίας

Οι μηχανικοί ανάπτυξης μπορούν να λάβουν μέτρα ώστε να αποτρέψουν ή να ελαχιστοποιήσουν λάθη αναγνώρισης ομιλίας μέσω προσεκτικού σχεδιασμού με τις ακόλουθες δύο μεθόδους:

- *Σχεδιασμός Βασισμένος στις Φωνητικές Ιδιότητες των Λέξεων*
- *Χρήση της Τεχνικής «απρόσκλητης διακοπής»*

Σχεδιασμός Βασισμένος στις Φωνητικές Ιδιότητες των Λέξεων (1 από 2)

Κάποιοι ήχοι προκαλούν προβλήματα σε ένα σύστημα αναγνώρισης ομιλίας. Μια συνετή στρατηγική θα ήταν να σχεδιαστούν οι προτροπές του συστήματος με τέτοιο τρόπο ώστε τέτοιες λέξεις να μην προκύπτουν στην είσοδο από το χρήστη.

π.χ. Τυρβώδη σύμφωνα (όπως τα /h/, /s/ and /f/) και άηχα σύμφωνα (όπως τα /p/ or /t/) θεωρούνται «θορυβώδη» και πιο δύσκολο να αναγνωριστούν με ακρίβεια.

Έτσι μια λέξη όπως η "help" θα ήταν προβληματική επειδή το αρχικό /h/ μπορεί να χανόταν και το τελικό /p/ να αναγνωριστεί λανθασμένα.

Δυστυχώς, η λέξη "help" είναι πιθανό να είναι χρήσιμη είσοδος σε πολλά συστήματα, με αποτέλεσμα ώστε η σκοπιμότητα να υπερισχύει των συλλογισμών περί των φωνητικών ιδιοτήτων της λέξης.

Σχεδιασμός Βασισμένος στις Φωνητικές Ιδιότητες των Λέξεων (2 από 2)

Παραδείγματα:

- η λέξη "cancel", στην οποία ο ήχος του /s/ στην τελική συλλαβή μπορεί να χαθεί όπως και ολόκληρη η συλλαβή μιας και δεν είναι τονισμένη και είναι χαμηλής ενέργειας.
- λέξεις όπως η λέξη "six" είναι προβληματικές, καθώς, πέραν από την ύπαρξη του φωνήματος /s/, το φωνήεν έχει μικρή διάρκεια.
- "Repeat" είναι μια ακόμα συχνά χρησιμοποιούμενη λέξη που έχει μια αδύναμη αρχική συλλαβή. Είναι πιθανή η σύγχυση με λέξεις που μοιράζονται κάποια από τα φωνήεντα και έχουν επίσης μη-τονισμένες αρχικές συλλαβές, π.χ. η λέξη "delete".

Γενικότερα, η προτεινόμενη στρατηγική είναι να ζητείται είσοδος αποτελούμενη από αρκετές συλλαβές, όπως π.χ. φράσεις, γιατί είναι πιθανό να αποτελούνται από πλουσιότερα φωνητικά χαρακτηριστικά.

Χρησιμοποιώντας την Τεχνική της «απρόσκλητης διακοπής» - (barge-in) (1 από 3)

- Ο χρήστης μπορεί να διακόψει την προτροπή του συστήματος
- Είναι μια χρήσιμη λειτουργία, ειδικά για έμπειρους χρήστες που δεν χρειάζεται να ακούν γνωστές προτροπές και που μπορούν να διακόπτουν την προτροπή με τις απαντήσεις τους και να ολοκληρώνουν το διάλογο γρηγορότερα.

Χρησιμοποιώντας την Τεχνική της «απρόσκλητης διακοπής» - (barge-in) (2 από 3)

- Θέματα που πρέπει να μελετηθούν σχολαστικά στο στάδιο της σχεδίασης:

1. Η «απρόσκλητη διακοπή» περιλαμβάνει:

α) Την εξάλειψη της ηχούς (ένα υπολογιστικά επίπονο έργο)

β) Η μηχανή Αυτόματης Αναγνώρισης Ομιλίας να είναι ενεργή για τη μεγαλύτερη διάρκεια της κλήσης

2. Λανθασμένη αποδοχή, όπου ήχοι που δεν αποτελούν ομιλία ή ομιλία που δεν απευθύνεται στο σύστημα (π.χ., ομιλία που απευθύνεται σε κάποιον τρίτο) αντιμετωπίζεται ως σκόπιμη είσοδος, και μπορεί να είναι αιτία για να διακοπεί η προτροπή του συστήματος και ο χρήστης να μείνει σαστισμένος.

Χρησιμοποιώντας την Τεχνική της «απρόσκλητης διακοπής» - (barge-in) (3 από 3)

3. Στο σημείο που ο χρήστης παρεμβαίνει, αν το σύστημα δεν διακόπτει την προτροπή αμέσως, ο χρήστης μπορεί να μιλήσει δυνατότερα ώστε να ακουστεί πιο δυνατά από την προτροπή (φαινόμενο "Lombard"). Η παραγόμενη ενισχυμένη ομιλία μπορεί να είναι πιο δύσκολη ως προς την επεξεργασία καθώς μπορεί να μην ταιριάζει με τα εκπαιδευμένα μοντέλα για ομιλία που εκφωνείται με φυσιολογική ένταση ήχου.
4. Στην περίπτωση επικάλυψης μεταξύ χρήστη και συστήματος ο χρήστης μπορεί να επαναλάβει μέρος κάποιας εκφώνησης που αντιλήφθηκε ότι επικαλύφθηκε (φαινόμενο «τραυλίσματος"), καθώς η παραγόμενη συμβολοσειρά που δίδεται στο συστατικό αναγνώρισης μπορεί να περιέχει δισταγμούς και λανθασμένες εκκινήσεις, με άλλα λόγια, μια συμβολοσειρά που πιθανόν δεν θα συμμορφώνεται με τη γραμματική που προσδιορίζει την είσοδο.

Διαφορετικοί τύποι «απρόσκλητης διακοπής» (1 από 2)

1. «Απρόσκλητη διακοπή» με ανίχνευση ενέργειας.

Η προτροπή του συστήματος διακόπτεται μόλις εντοπιστεί οποιοσδήποτε ήχος.

- Πλεονεκτήματα:
  - Περιορίζει την παραμόρφωση που προκαλείται από την «απρόσκλητη διακοπή» των πρώτων συλλαβών της ομιλίας του χρήστη.
  - Τα φαινόμενα «Lombard» και «τραυλίσματος» ελαχιστοποιούνται.
- Μειονέκτημα:
  - Αύξηση στην λανθασμένη αποδοχή ήχων από το περιθώριο που δεν προορίζονται για έγκυρη είσοδο.

Διαφορετικοί τύποι «απρόσκλητης διακοπής» (2 από 2)

2. «Απρόσκλητη διακοπή» σε σίγουρη αναγνώριση λέξεων.

Η είσοδος ήχου διακόπτεται μόνο αφού το σύστημα καθορίσει ότι ο χρήστης εκφώνησε μια πλήρη και έγκυρη λέξη ή φράση, καθορισμένη από μια ενεργή γραμματική αναγνώρισης.

- **Πλεονέκτημα:** ελαχιστοποιεί την λανθασμένη αποδοχή που οφείλεται σε μη-σκόπιμες διακοπές ή θόρυβο από το περιθώριο.
- **Μειονέκτημα:** αυξάνει τις λανθασμένες απορρίψεις λόγω αύξησης της εμφάνισης των φαινομένων του «Lombard» και του «τραυλίσματος».

Κατανόηση Γλώσσας (1 από 2)

- Ρόλος: να αναλύσει την έξοδο του συστατικού αναγνώρισης ομιλίας και να αποδώσει αναπαράσταση νοήματος που μπορεί να χρησιμοποιηθεί από τη βαθμίδα διαχείρισης διαλόγου.
- Παραδοσιακά, περιλαμβάνει δύο διαδικασίες:
  - **συντακτική ανάλυση**, να προσδιορίσει τη συστατική δομή της αναγνωρισμένης συμβολοσειράς (δηλαδή, πως οι λέξεις ομαδοποιούνται μεταξύ τους),
  - **σημασιολογική ανάλυση**, να προσδιορίσει το νόημα των συστατικών.

Κατανόηση Γλώσσας (2 από 2)

- Σε πολλά σύγχρονα συστήματα προφορικού διαλόγου το νόημα των εκφωνήσεων παράγεται άμεσα από την αναγνωρισμένη συμβολοσειρά χρησιμοποιώντας μια «σημασιολογική γραμματική».
- Η κατανόηση γλώσσας στα συστήματα προφορικού διαλόγου είναι προβληματική για δύο λόγους:
  - Λόγω της αμφισημίας στην φυσική γλώσσα.
  - Λόγω κακοσχηματισμένης εισόδου.

Αμφισημία στη φυσική γλώσσα

- **Λεκτική αμφισημία**

Μια λέξη μπορεί να ανήκει σε περισσότερες από μια κατηγορίες μερών του λόγου, π.χ, η λέξη "book" μπορεί να είναι ουσιαστικό ή ρήμα. Συνήθως μπορεί να διαλευκανθεί μέσω του πλαισίου των άλλων λέξεων της πρότασης, όπως στην πρόταση "book a flight to London".

- **Αμφισημία νοήματος**

Μια λέξη μπορεί να έχει διαφορετικά νοήματα, π.χ., η λέξη "bank" μπορεί να είναι ένα οικονομικό ίδρυμα ή η όχθη του ποταμού.

- **Δομική Αμφισημία**

Η σχέση μεταξύ των φράσεων της πρότασης είναι αμφίσημη, π.χ., «μια πτήση για το Λονδίνο φτάνει στις 9". Με βάσει καθαρά συντακτική ανάλυση, υπάρχουν δύο πιθανές αναγνώσεις, μία στην οποία η πτήση φτάνει στις 9, και μια άλλη στην οποία το Λονδίνο φτάνει στις 9.

Παραγωγή Γλώσσας (1 από 3)

- Να κατασκευαστεί ένα μήνυμα σε φυσική γλώσσα που να μεταφέρει όσες από τις αιτηθείσες πληροφορίες ανακτήθηκαν από την εξωτερική πηγή.
- Η ανακτηθείσα πληροφορία μπορεί να πάρει ποικίλες μορφές: πίνακες αριθμητικών δεδομένων, εγγραφές βάσεων δεδομένων, ακολουθίες από οδηγίες για την εκτέλεση/ολοκλήρωση κάποιου έργου.

Παραγωγή Γλώσσας (2 από 3)

Οι απλούστερες μέθοδοι:

- **«Κονσερβοποιημένο» κείμενο**

Ένα πεδίο βάσης δεδομένων που θα περιείχε περιγραφικές (σε μορφή κειμένου) για αντικείμενα, θα ήταν επαρκές για παροχή απαντήσεων σε βασικές επερωτήσεις.

Άκαμπτα ή χωρίς ευελιξία.

- **Γέμισμα προτύπων αναφοράς** (templates)
  - Μεγαλύτερος βαθμός ευελιξίας σε περιπτώσεις που ένα μήνυμα μπορεί να παραχθεί πολλές φορές με μικρές παραλλαγές.
  - Τα περισσότερα σύγχρονα Συστήματα Προφορικού Διαλόγου που αφορούν ανάκτηση πληροφοριών, χρησιμοποιούν την τεχνική του γεμίσματος προτύπων αναφοράς για να παράγουν το μήνυμα που θα ακούσει ο χρήστης.

Παραγωγή Γλώσσας (3 από 3)

- Προηγμένη μέθοδος:

Η σχεδίαση ξεκινά με έναν επικοινωνιακό στόχο και καταλήγει με ένα γλωσσικό μήνυμα.

Μπορεί να χωριστεί σε τρία στάδια:

- Σχεδίαση εγγράφων.
- Μικροσχεδίαση.
- Υλοποίηση Επιφάνειας.

Σχεδίαση (πλάνα) εγγράφων (1 από 2)

Προσδιορίζει:

- Τι είδους πληροφορίες θα έπρεπε να συμπεριληφθεί στο μήνυμα (επιλογή περιεχομένου)
- Πώς πρέπει να δομηθεί το μήνυμα (δομή λόγου).
- Δεν είναι κατάλληλες όλες οι πληροφορίες που ανακτήθηκαν για να εκφωνηθούν στο χρήστη.
- Οι πληροφορίες που πρέπει να μεταφερθούν μπορεί να διαφέρουν ανάλογα με τους χρήστες: Ένα μοντέλο χρηστών βοηθάει το σύστημα να παράγει έξοδο που είναι η κατάλληλη για έναν συγκεκριμένο χρήστη.
- Η δομή του λόγου του μηνύματος είναι σημαντική για να υποβοηθήσει την κατανόηση από τον χρήστη, ειδικά σε ένα μήνυμα που περιλαμβάνει πολλές προτάσεις

Σχεδίαση (πλάνα) εγγράφων (2 από 2)

Καλοσχηματισμένα κείμενα μπορούν να μοντελοποιηθούν χρησιμοποιώντας:

- **Σχήματα:** Ένα σχήμα θέτει τα βασικά συστατικά ενός κειμένου, χρησιμοποιώντας στοιχεία όπως «ταυτοποίηση», «αναλογία», «σύγκριση» και «συγκεκριμένο παράδειγμα», που ανακλούν την σειριακή οργάνωση του κειμένου.
- **Θεωρία Ρητορικής Δομής:** περιγράφει τις σχέσεις μεταξύ στοιχείων του κειμένου. Το κεντρικό στοιχείο ενός κειμένου («πυρήνας») μπορεί να σχετίζεται με ένα πιο περιφερειακό στοιχείο («δορυφόρο») μέσω ρητορικών σχέσεων όπως η «επεξεργασία» και η «αντίθεση».
  - Η επεξεργασία παρέχει επιπλέον πληροφορίες σχετικά με το περιεχόμενο του πυρήνα.
  - Η αντίθεση παρουσιάζει αντικείμενα που είναι παρόμοια σε κάποια σημεία αλλά διαφέρουν σε άλλα.

Μικροσχεδίαση (1 από 3)

Δέχεται την έξοδο των πλάνων συνομιλίας από τον σχεδιαστή εγγράφου και προετοιμάζει την είσοδο που θα σταλεί στον υλοποιητή επιφάνειας.

- Υπάρχουν τρεις κύριες εργασίες στη μικροσχεδίαση:

1. **Αναφορικές Εκφράσεις**

Καθορισμός του τρόπου αναφοράς σε μια οντότητα σε ένα κείμενο μέσα σε ένα συγκεκριμένο πλαίσιο. π.χ., αν κάποια οντότητα που έχει ήδη αναφερθεί θα αναφερθεί ξανά, θα πρέπει να είναι δυνατή η αναφορά σε αυτή την οντότητα χρήσει μιας αντωνυμίας

Μικροσχεδίαση (2 από 3)

2. Συσσώρευση

Το θέμα του συνδυασμού περιεχομένου σε τμήματα με νόημα, π.χ.,με χρήση συζεύξεων ή ελλείψεων,

- The flight departs at 9. It arrives at 10. (Χωρίς συσσώρευση.)
- The flight departs at 9 and it arrives at 10. (Συσσώρευση με σύζευξη.)
- The flight departs at 9 and [ ] arrives at 10. (Συσσώρευση με σύζευξη και έλλειψη.)

Μικροσχεδίαση (3 από 3)

3. Λεκτική επιλογή

Επιλογή των κατάλληλων λέξεων ώστε να εκφράσουν το περιεχόμενο. Στα πιο απλά συστήματα ένα απλό λεκτικό αντικείμενο σχετίζεται κάθε οντότητα στην βάση δεδομένων. Παρόλα αυτά, η χρήση διαφορετικών εκφράσεων παρέχει περισσότερη ποικιλία στο κείμενο όπως στο παράδειγμα:

- The first flight departs at 9. The second flight departs at 10. The third flight departs at 11. (Χωρίς παραλλαγή.)
- The first flight departs at 9. The departure times of the next flights are 10 and 11.

Υλοποίηση Επιφάνειας (1 από 2)

Η διαδικασία της μετατροπής της εξόδου προδιαγραφών κειμένου από τον μικροσχεδιαστή σε γλωσσικό κείμενο.

Περιλαμβάνει δύο δραστηριότητες:

1. **Δομική Υλοποίηση**

Περιλαμβάνει τη χρήση επισημειώσεων ώστε να μεταφερθεί η δομή του εγγράφου.

Η γλώσσα XML επικρατεί ως πρότυπο για επισημείωση εγγράφων.

Υλοποίηση Επιφάνειας (2 από 2)

2. Γλωσσική Υλοποίηση

Περιλαμβάνει: την επιλογή των λέξεων με τις συντακτικές τους δομές που εκφράζουν το επιθυμητό νόημα.

Επιπλέον περιλαμβάνει: την εισαγωγή λειτουργικών λέξεων, την επιλογή της σωστής κλίσης των λέξεων περιεχομένου, την τοποθέτηση σε σωστή σειρά των λέξεων μέσα στην πρόταση, και την εφαρμογή κανόνων ορθογραφίας.

Χρησιμοποιείται μια γραμματική που παρέχει ένα σύνολο επιλογών για γλωσσική υλοποίηση, π.χ., μεταξύ ενεργητικών και παθητικών προτάσεων, όπως:

- Bad weather has delayed the flight.
- The flight has been delayed by bad weather.

Άσκηση 4.1

Α) Με βάση το δείγμα διαλόγου που γράψατε στην άσκηση 2.1, προσδιορίστε τους ρόλους των ακόλουθων συστατικών και εκτιμήστε τι είδους τεχνολογίες χρησιμοποιούνται:

1. Αναγνώριση ομιλίας.
2. Επεξεργασία γλώσσας.
3. Παραγωγή γλώσσας.
4. Σύνθεση ομιλίας από κείμενο.
5. Εξωτερική επικοινωνία (π.χ., με μια βάση δεδομένων).

Άσκηση 4.2

Β) Κατατάξτε το σύστημα που χρησιμοποιήσατε στην άσκηση 2.1 ως προς τους άξονες (με βάση και τα στοιχεία του Πίνακα 4.1):

- Στύλ / είδος ομιλίας
- Πλήθος χρηστών / ομιλητών
- Μέγεθος λεξιλογίου – πολυπλοκότητα εφαρμογής
- Περιβαλλοντικός θόρυβος

Απαιτείται νέα δοκιμασία με το σύστημα

Τέλος Ενότητας

Αναγνώριση Ομιλίας και Κατανόηση Γλώσσας

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

Σημείωμα Αδειοδότησης

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:

- “Potential Synergies Between Speech Recognition and Proteomics”, Joseph Picone PhD, Life Sciences and Biotechnology Institute, Department of Electrical
- Lib4U, Automatic Speech Recognition, Spring 2003, globlib4u.wordpress.com

Σημείωμα Χρήσης Έργων Τρίτων (2 από 2)