



Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Επεξεργασία Ομιλίας και Φυσικής Γλώσσας

Ενότητα 7: Αναγνώριση ομιλίας

Γεώργιος Κουρουπέτρογλου

koupe@di.uoa.gr



Αυτόματη Αναγνώριση Ομιλίας (ΑΑΟ) (1/2)

Κύρια Πλεονεκτήματα

- Η είσοδος πληροφορίας με ομιλία σε μία μηχανή είναι εύκολο να εκτελεσθεί διότι δεν απαιτεί εξειδικευμένη ικανότητα όπως θα απαιτούσε η δακτυλογράφηση ή οι λειτουργίες πληκτρολόγησης.
- Η εισαγωγή πληροφοριών με ομιλία είναι 3 με 4 φορές πιο γρήγορη απ' ό,τι η δακτυλογράφηση και 8 με 10 φορές πιο γρήγορη απ' ό,τι η συνήθης γραφή.
- Οι πληροφορίες μπορούν να εισαχθούν με ομιλία ακόμη και όταν ο χρήστης μετακινείται ή εκτελεί άλλες δραστηριότητες, που αφορούν τα χέρια, τα πόδια, τα μάτια ή τα αυτιά.
- Αφού κατά τη ΑΑΟ μπορεί να χρησιμοποιηθεί ένα μικρόφωνο ή ένα τηλέφωνο ως τερματικό εισόδου, η εισαγωγή πληροφοριών είναι οικονομική και εφικτή κα από απόσταση με τη χρήση των υπάρχοντων τηλεφωνικών δικτύων.

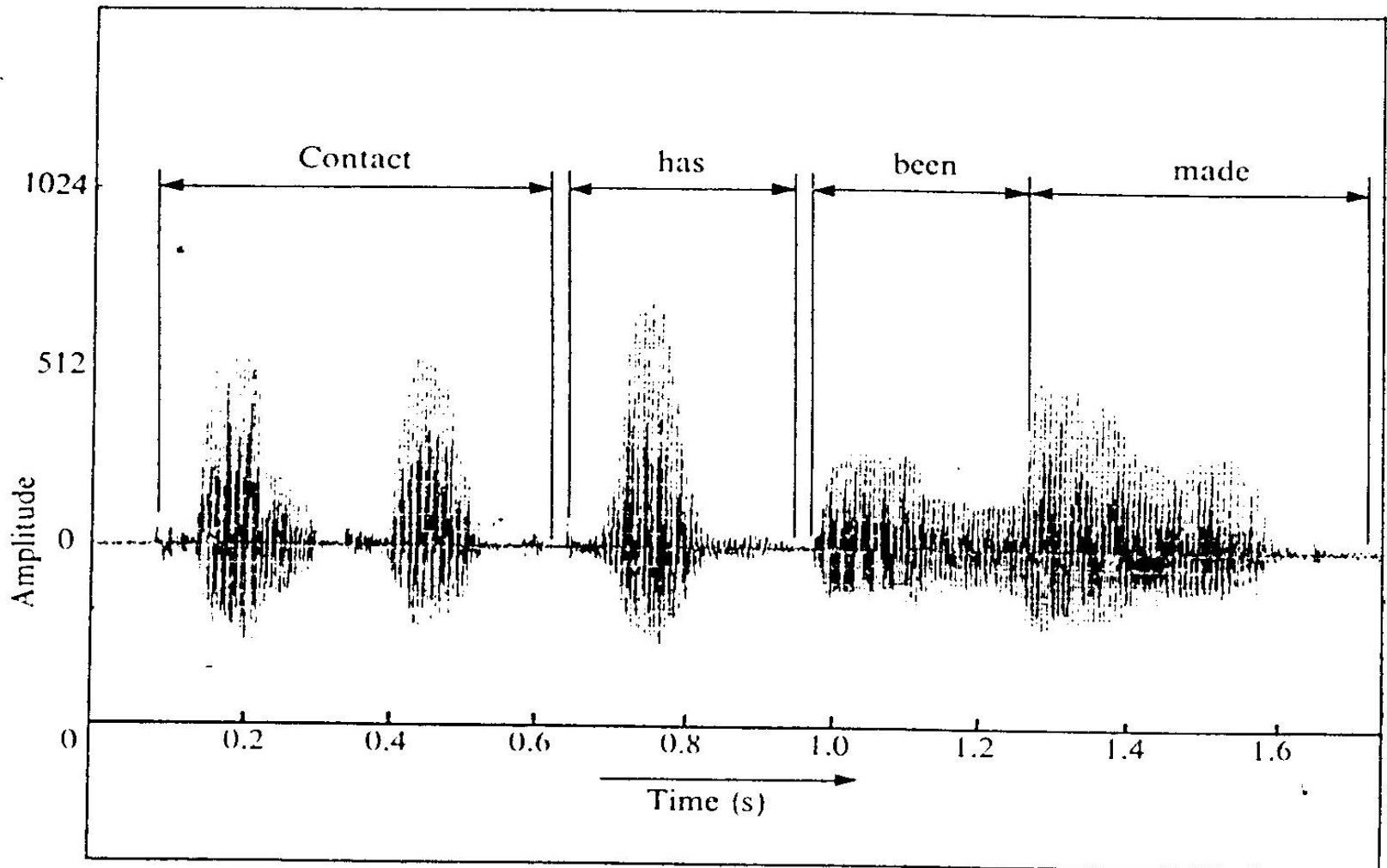
Αυτόματη Αναγνώριση Ομιλίας (ΑΑΟ) (2/2)

Κύρια Μειονεκτήματα

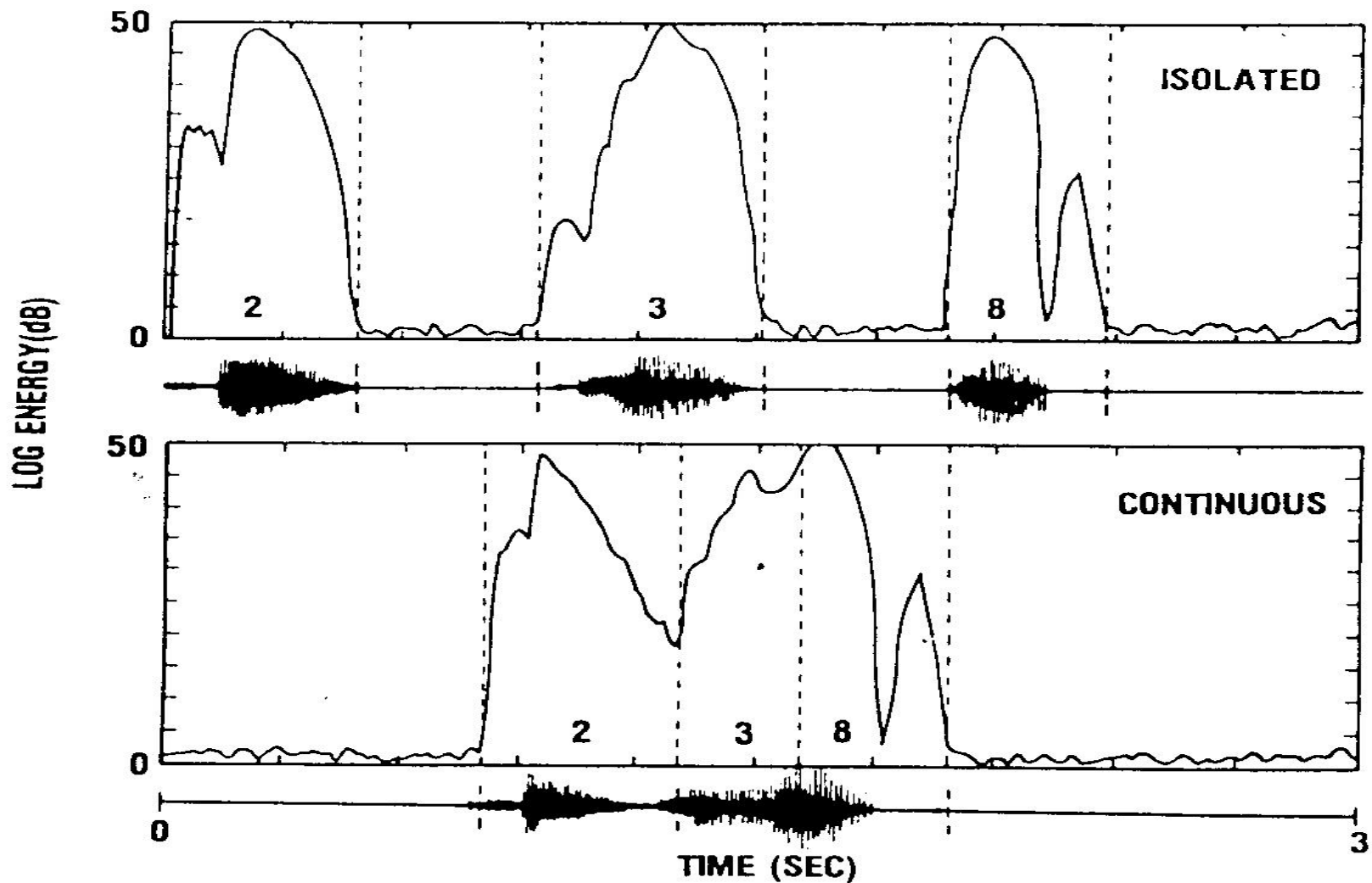
- Η είσοδος ή η συνομιλία δεν είναι εκτυπωμένη.
- Απαιτούνται μέθοδοι αντιμετώπισης του θορύβου όταν η ΑΑΟ χρησιμοποιείται σε θορυβώδες περιβάλλον.

Δυσκολίες κατά την Αυτόματη αναγνώριση Ομιλίας

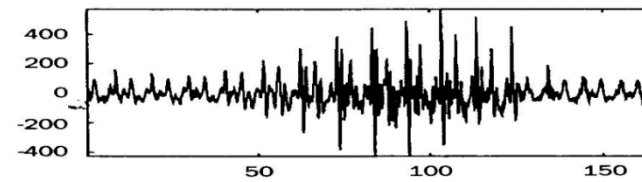
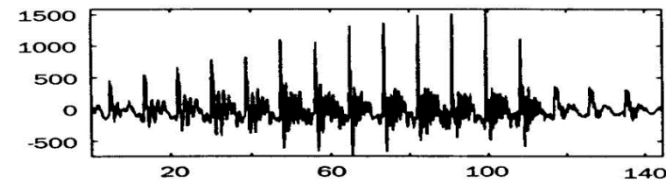
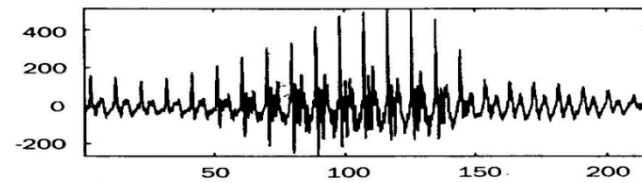
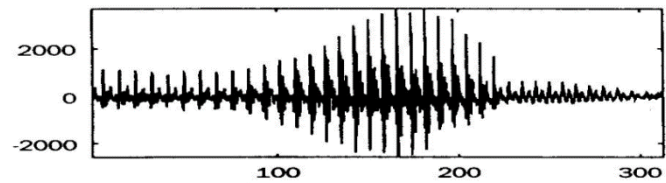
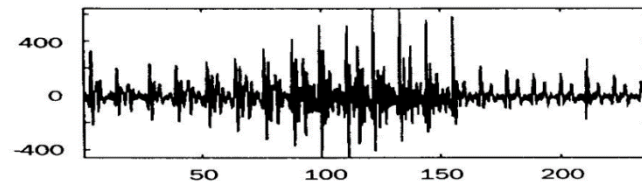
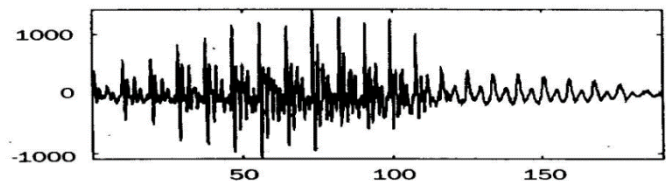
- 1) Προβλήματα συνάρθρωσης και μείωσης / έκθλιψης (αποκοπής συλλαβών).
- 2) Δυσκολίες στην κατάτμηση.
- 3) Ατομικότητα και άλλα προβλήματα διαφοροποίησης.
- 4) Ανεπαρκής γλωσσολογική γνώση.
- 5) Θόρυβος περιβάλλοντος.



Πρόβλημα Κατάτμησης



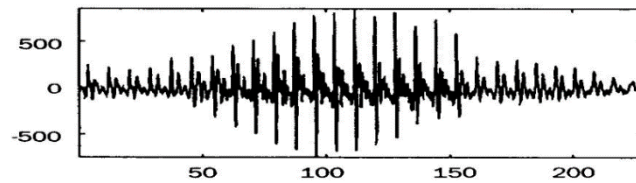
Παράδειγμα εκφώνησης των λέξεων two three eight
μεμονωμένα και συνδεδεμένα ή συνεχόμενα.



Time (milliseconds)

Time (milliseconds)

(a)



Time (milliseconds)

(b)

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

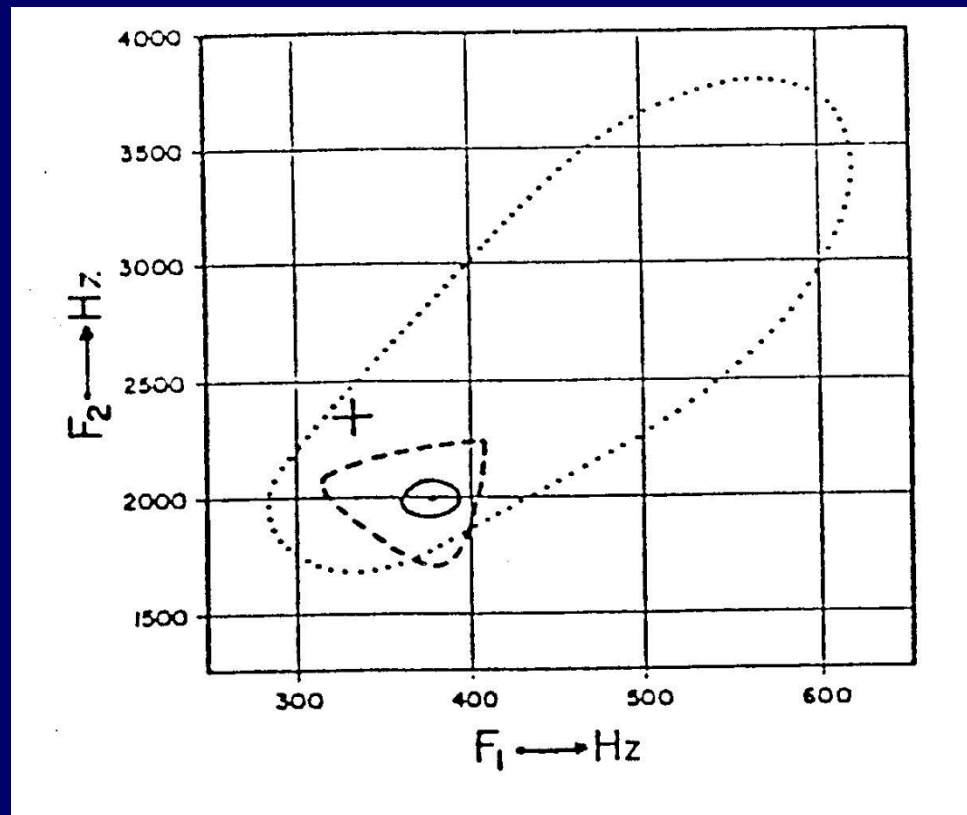


Κατηγοριοποίηση Αυτόματης Αναγνώρισης Ομιλίας (1/2)

Αντικείμενο	Εξαρτώμενη από ομιλητή	Ανεξάρτητη από ομιλητή
<i>Αναγνώριση μεμονωμένων λέξεων</i>		
<i>Αναγνώριση συνεχόμενης ομιλίας: Αναγνώριση συνδεόμενων λέξεων</i>		
<i>Αναγνώριση ομιλίας σε μορφή συζήτησης (κατανόηση ομιλίας)</i>		

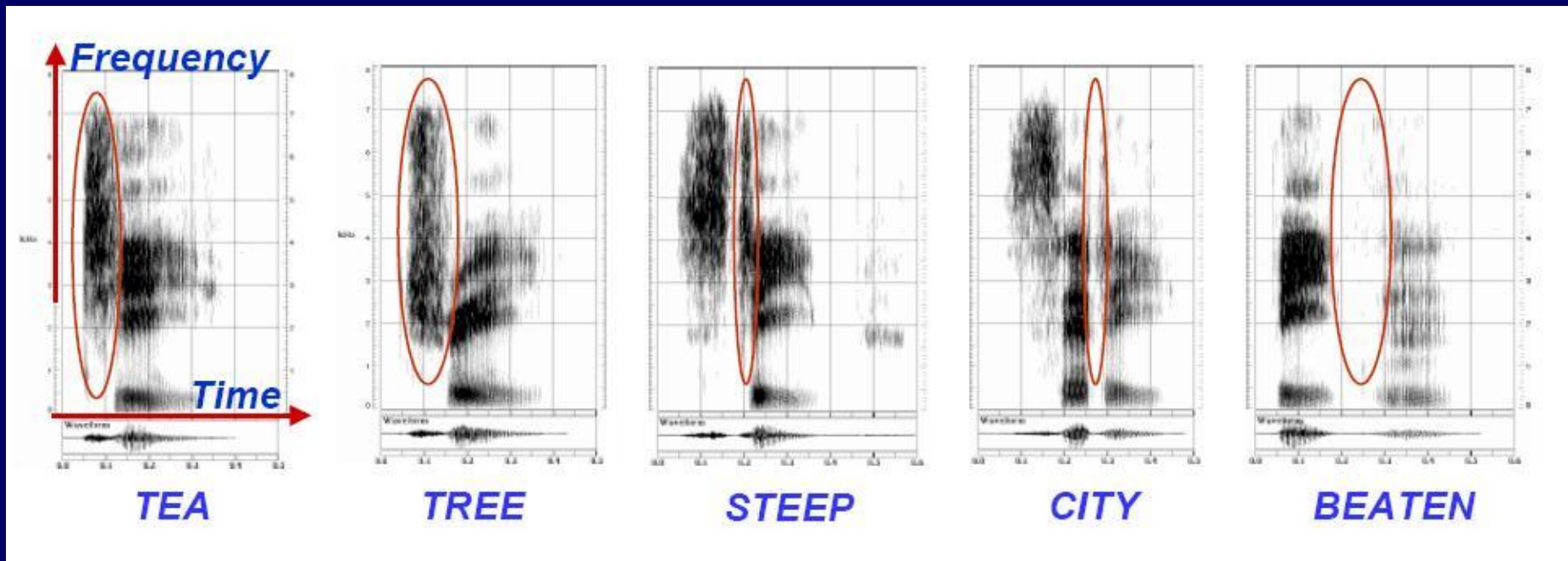
Κατηγοριοποίηση Αυτόματης Αναγνώρισης Ομιλίας (2/2)

Μέρος του φωνητικού χώρου F_1 , F_2 που δείχνει την κατανομή του φωνήεντος /i/ σε περιβάλλον συμφώνων [h-d], όπως μετρήθηκε από 76 ομιλητές (μεγάλη εστιγμένη γραμμή). Η κατανομή του ίδιου φωνήεντος στο ίδιο φωνητικό περιβάλλον όπως εκφωνήθηκε από τον ίδιο ομιλητή (μικρή συνεχής γραμμή). Η κατανομή του ίδιου φωνήεντος όπως εκφωνήθηκε από τον ίδιο ομιλητή σε 576 διαφορετικά φωνητικά περιβάλλοντα συμφώνων (μεσαία γραμμή).



Παράδειγμα φωνολογικής μεταβλητότητας

Η ακουστική πραγμάτωση ενός φωνήματος εξαρτάται πάρα πολύ από το φωνητικό του περιβάλλον.



Βασικές Κλάσεις Μεθόδων ΑΑΟ (1/4)

- A. Ταίριασμα προτύπων ιχνών (template matching) ή δομικές μέθοδοι ή **ντετερμινιστικές μέθοδοι**.
- B. **Στοχαστικές**: Κρυφά Μοντέλα Markov HMM (Hidden Markov Models) και Τεχνητά Νευρωνικά Δίκτυα – ANN (Artificial Neural Networks).
- C. **Μέθοδοι βασισμένες σε γνώση** (φωνητική και γλωσσολογική) (knowledge based).

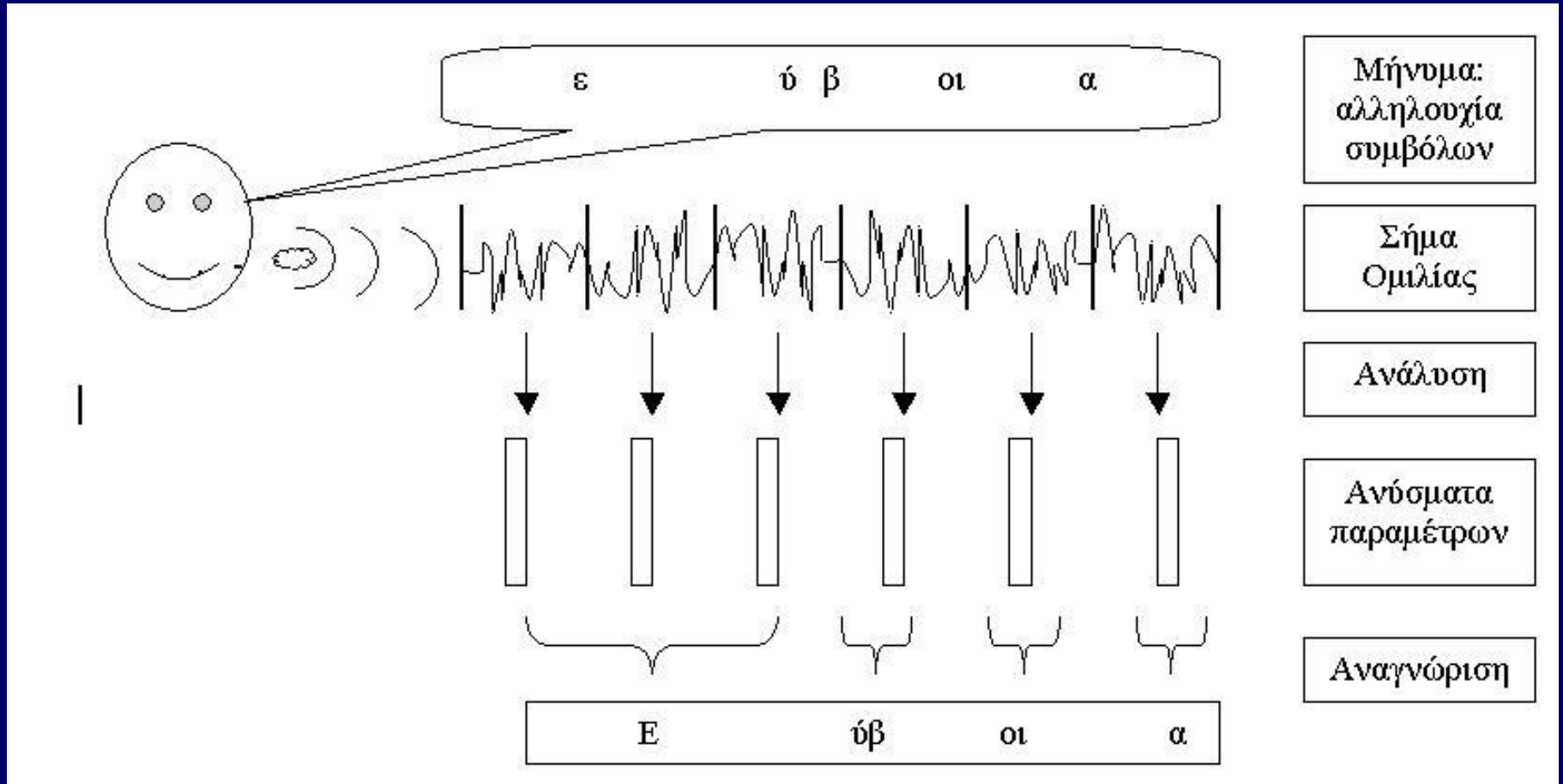
Οι Α. και Β. ονομάζονται και **Μέθοδοι βασισμένες σε Δεδομένα (data based approaches)**: το σήμα της ομιλίας μοντελοποιείται με αλγορίθμους που μπορούν να εξάγουν γνώση αυτόματα από τα δεδομένα.

Στη C. 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 περισσότερων από τις βασικές μεθόδους.



Βασικές Κλάσεις Μεθόδων ΑΑΟ (2/4)



Κωδικοποίηση / Αποκωδικοποίηση Μηνύματος Ομιλίας.

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

Βασικές Κλάσεις Μεθόδων ΑΑΟ (3/4)

- ❖ Γενικά όλες οι Μέθοδοι ΑΑΟ υποθέτουν ότι το σήμα ομιλίας είναι μία πραγματοποίηση της κωδικοποίησης κάποιου μηνύματος σε μία αλληλουχία ενός ή περισσότερων συμβόλων.
- ❖ Το πρόβλημα της ΑΑΟ είναι να αναγνωρίζει την αλληλουχία των συμβόλων από μία δεδομένη εκφώνηση.
- ❖ Στην αρχή το συνεχές σήμα της ομιλίας μετατρέπεται σε μία αλληλουχία διακριτών και σε ίσες (χρονικές) αποστάσεις ανυσμάτων παραμέτρων.
- ❖ Η αλληλουχία των ανυσμάτων παραμέτρων θεωρείται ότι αποτελεί μία 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.

Τυπικές παράμετροι που συνήθως χρησιμοποιούνται στην ΑΑΟ είναι:

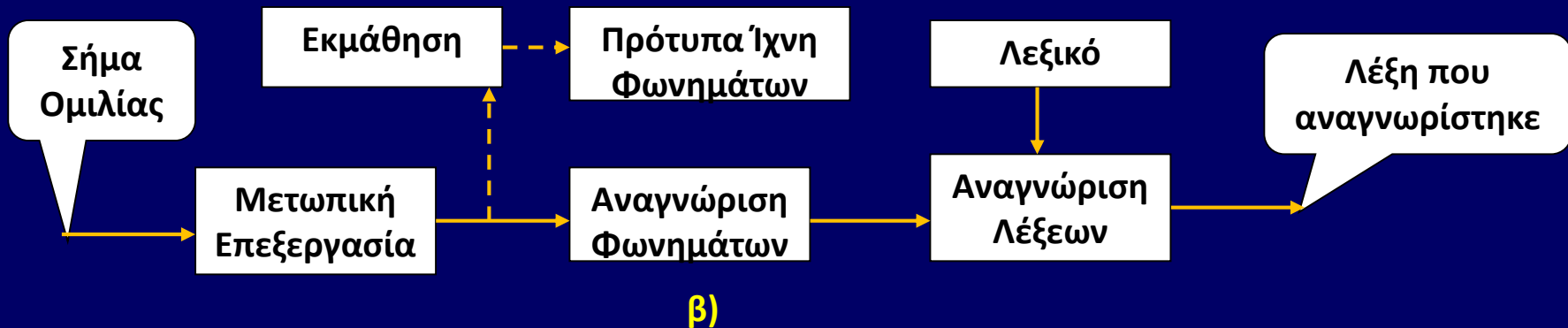
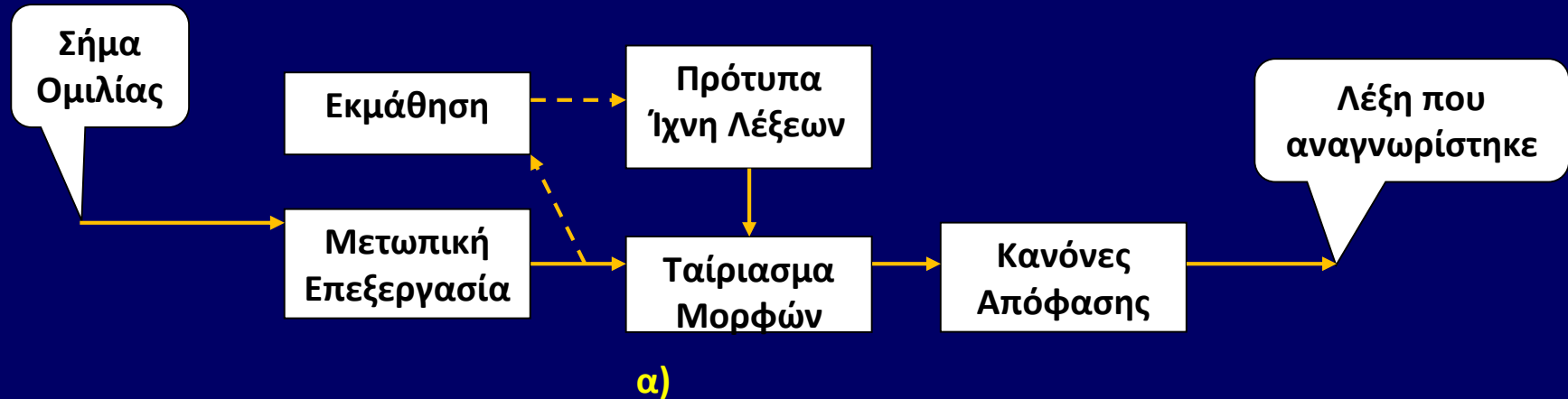
- ❖ Οι συντελεστές λειασμένου φάσματος.
- ❖ Οι συντελεστές γραμμικής πρόβλεψης.
- ❖ Άλλες αναπαραστάσεις που προκύπτουν από αυτούς τους δύο, όπως: συντελεστές cepstrum, formants, έξοδοι τράπεζας φίλτρων, καθώς και όσες αναφέρθηκαν στο κεφάλαιο της ανάλυσης ομιλίας.

Βασικές Κλάσεις Μεθόδων ΑΑΟ (4/4)

Το άνυσμα παραμέτρων είναι N διαστάσεων, όπου N ο 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).

Ταίριασμα Προτύπων Ακουστικών Ιχνών για ΑΑΟ (1/2)



Δομή συστήματος ΑΑΟ με ταίριασμα προτύπων ακουστικών ιχνών:

- αναγνώριση βασισμένη σε λέξεις.
- αναγνώριση βασισμένη σε φωνήματα.

Ταίριασμα Προτύπων Ακουστικών Ιχνών για ΑΑΟ (2/2)

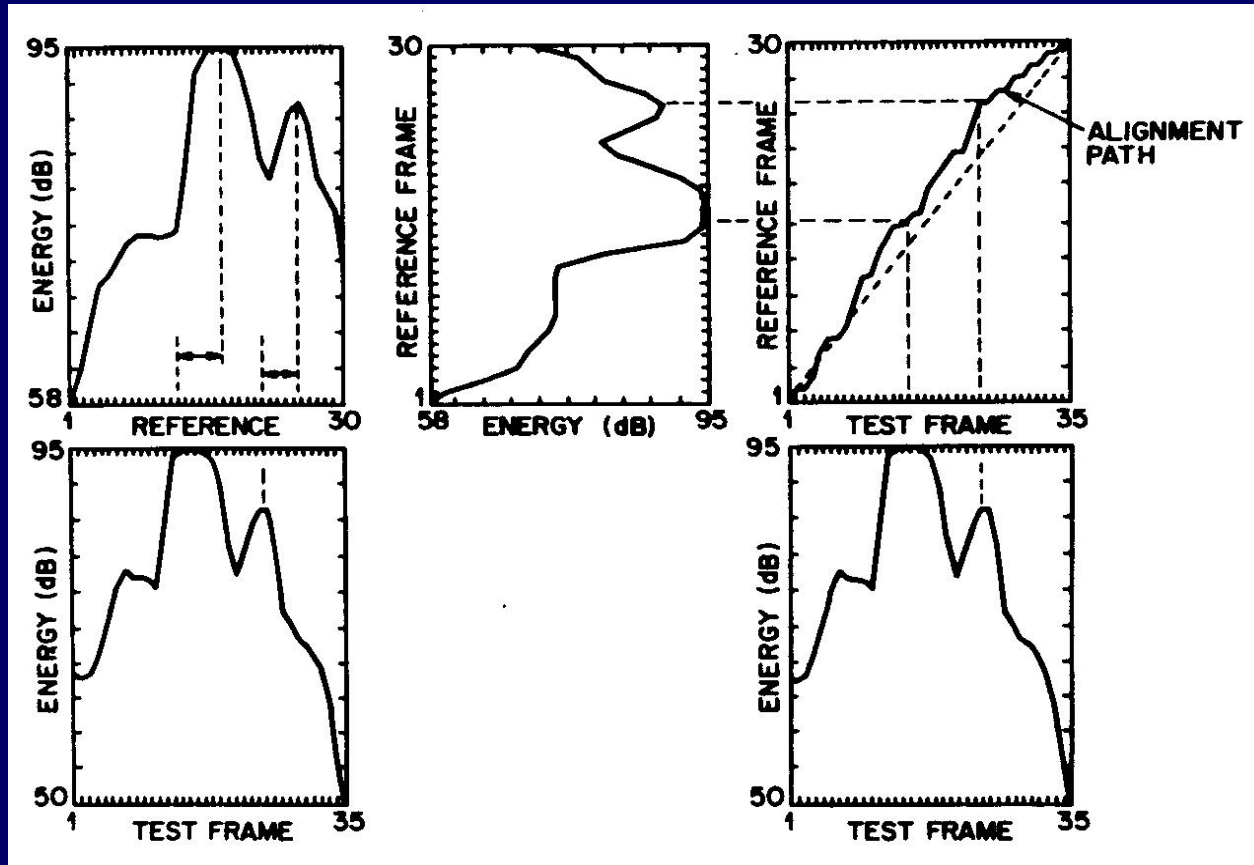
Στα συστήματα ΑΑΟ που εξαρτώνται από τον ομιλητή, τα ίχνη, που αποτελούν πρότυπα ανύσματα παραμέτρων που ισαπέχουν χρονικά και καθένα τους αντιστοιχεί σε ένα πλαίσιο ανάλυσης, λαμβάνονται χωριστά για κάθε νέο ομιλητή με τη διαδικασία της εκμάθησης. Στα συστήματα που είναι ανεξάρτητα του ομιλητή τα ίχνη αποτελούν τη μέση τιμή ενός αντιπροσωπευτικού δείγματος ομιλητών ή/ και λαμβάνονται πολλαπλά διαφορετικά ίχνη για κάθε λέξη ή φώνημα. Υπάρχουν και συστήματα προσαρμόσιμα στον ομιλητή τα οποία προσαρμόζουν τα ίχνη κάθε φορά που ένα καινούριο πρόσωπο χρησιμοποιεί το σύστημα με επιτυχία.

Αντί για φωνήματα στην πράξη λαμβάνονται πολλές φορές μέρη λέξεων όπως διφωνήματα, τριφωνήματα, συλλαβές, κλπ.

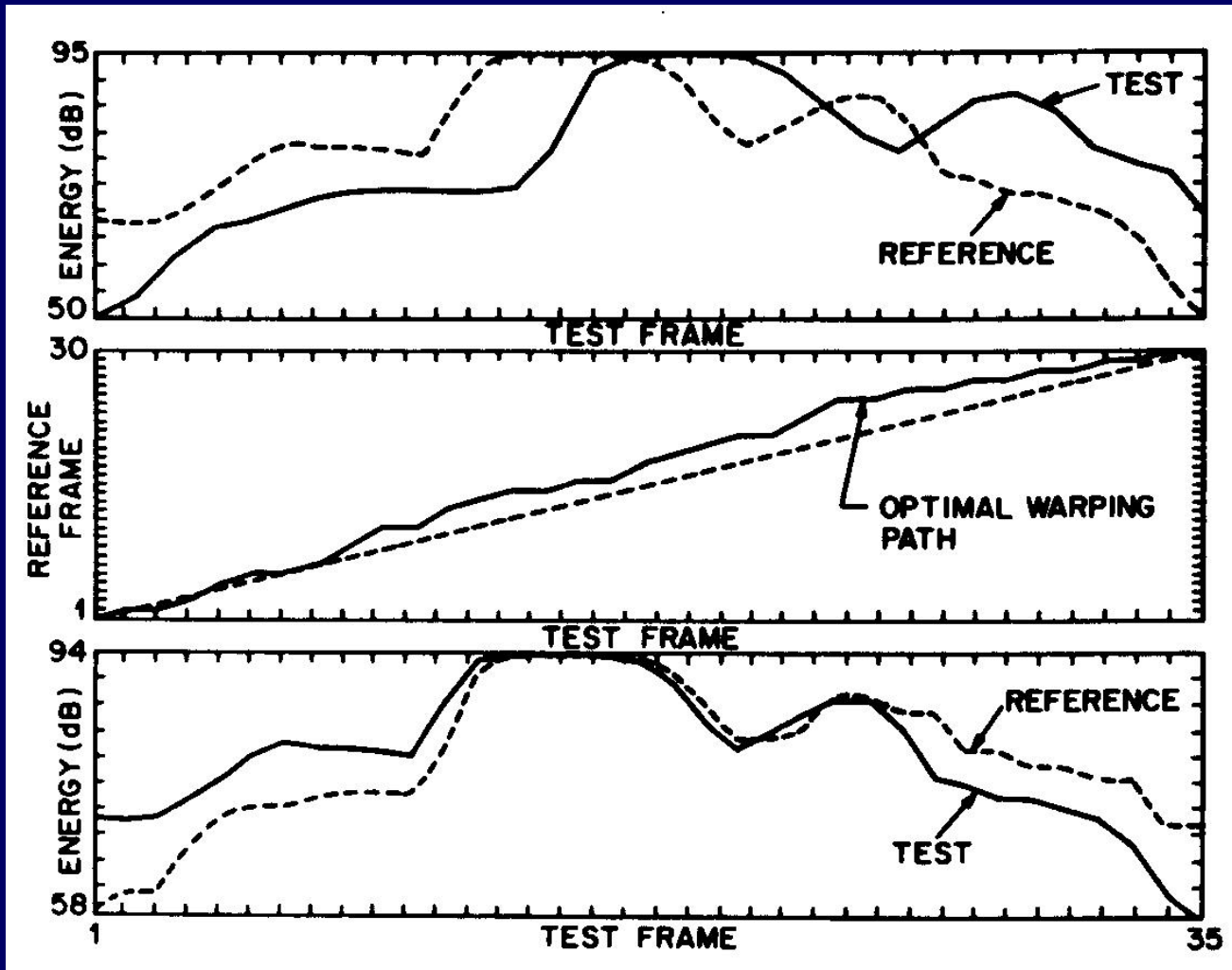
Ερωτήματα – προβλήματα:

- 1) Χρονική σύμπτωση ιχνών και αγνώστων εκφωνήσεων.
- 2) Μέθοδος σύγκρισης συνόλου διανυσμάτων ή μέτρησης απόστασης διανυσμάτων παραμέτρων.
- 3) Κανόνες απόφασης.

Δυναμική Χρονική Στρέβλωση – Dynamic Time Warping (DTW) (1/3)



Παράδειγμα που δείχνει την αναγκαιότητα για μη γραμμική χρονική στρέβλωση.



Παράδειγμα που δείχνει το αποτέλεσμα της δυναμικής χρονικής στρέβλωσης.

Δυναμική Χρονική Στρέβλωση – Dynamic Time Warping (DTW) (2/3)

Εστω δύο χρονικές ακολουθίες διανυσμάτων τα οποία θα πρέπει να συγκριθούν:

$$A = a_1, a_2, \dots, a_I \text{ και } B = b_1, b_2, \dots, b_J$$

Η συνάρτηση χρονικής στρέβλωσης μπορεί να αναπαρασταθεί από μία ακολουθία σημείων πλέγματος στο επίπεδο $C = (I, j)$:

$$F = c_1, c_2, \dots, c_k, \dots, c_K \quad c_k = (i_k, \dots, j_k)$$

Όταν η φασματική απόσταση ανάμεσα σε δύο κύρια διαστήματα a_I και b_i

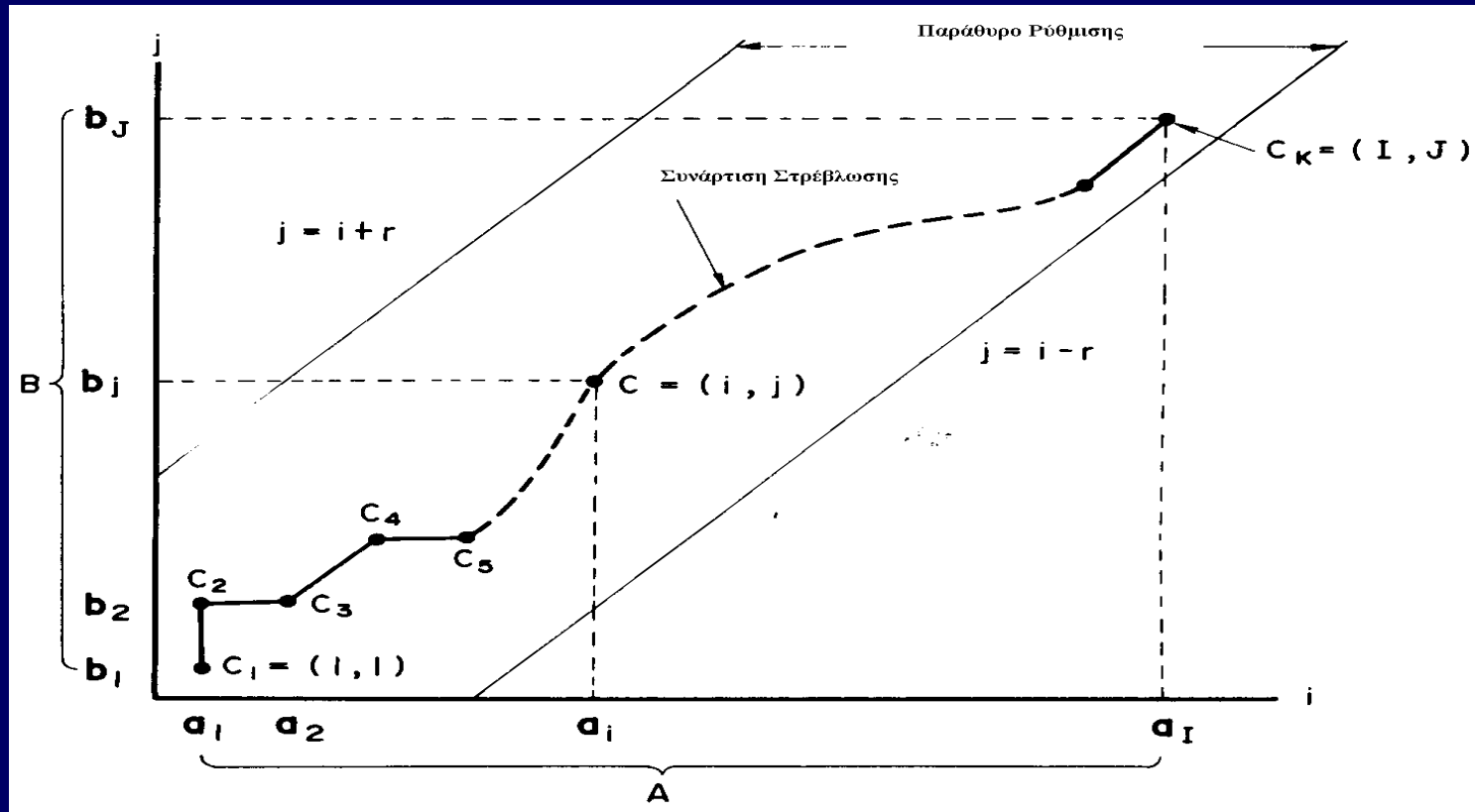
αναπαρίσταται με: $d(c) = d(i, j)$ το άθροισμα των αποστάσεων από την αρχή ως το τέλος της ακολουθίας κατά μήκος του F μπορεί να αναπαρασταθεί από την εξίσωση:

$$D(F) = \sum_{k=1}^K d(c_k) w_k / \sum_{k=1}^K w_k$$

Όπου w_k είναι μία θετική συνάρτηση βάρους η οποία σχετίζεται με το F .

Δυναμική Χρονική Στρέβλωση – Dynamic Time Warping (DTW) (3/3)

Όσο μικρότερη είναι η τιμή του $D(F)$ τόσο καλύτερη είναι η ταύτιση ανάμεσα στο A και στο B.



DTW ανάμεσα σε δύο χρονικές ακολουθίες A και B.

Δυναμική Χρονική Στρέβλωση

Ελαχιστοποίηση της εξίσωσης: $D(F) = \sum_{k=1}^K d(c_k)w_k / \sum_{k=1}^K w_k$
κάτω από τις ακόλουθες συνθήκες:

- 1) Συνθήκη μονοτονίας και συνέχειας: $0 \leq i_k - i_{k-1} \leq 1, 0 \leq j_k - j_{k-1} \leq 1$
- 2) Συνθήκη ορίων: $i_1 = j_1 = 1, i_k = I, j_k = J$
- 3) Συνθήκη παραθύρου ρύθμισης: $|i_k - j_k| \leq r, r = \text{σταθερά}$

Ορίζουμε το w_k έτσι ώστε ο παρονομαστής της εξίσωσης να γίνει μία ανεξάρτητη του F σταθερά. Για παράδειγμα αν:

$$w_k = (i_k - i_{k-1}) + (j_k - j_{k-1}) (i_0 = j_0 = 0) \text{ τότε } \sum_{k=1}^K w_k = I + J$$

$$\text{Άρα } D(F) = \sum_{k=1}^K d(c_k)w_k / I + J$$

Από τη στιγμή που η συνάρτηση η οποία ελαχιστοποιείται γίνεται προσθετική, η ελαχιστοποίηση μπορεί να λυθεί αποτελεσματικά χωρίς εξαντλητική εξέταση όλων των πιθανοτήτων για F .

Κανόνες Απόφασης ή Ταξινόμηση Μορφών

Αναγνώριση θεωρείται η επιλογή της λέξης που ταιριάζει περισσότερο με το πρότυπο ίχνος.

Συνήθως εφαρμόζεται ο κανόνας του πλησιέστερου γείτονα:

Έστω ότι υπάρχουν αποθηκευμένα N ίχνη R_I . Υπολογίζονται όλες οι αποστάσεις D_I μεταξύ της άγνωστης εισόδου και κάθε ίχνους. Επιλέγεται το ως άριστο R_I^* εκείνο για το οποίο ισχύει:

$$i^* = \arg \min |D_i|$$

Δηλαδή επιλέγεται ως λέξη ή φώνημα που αναγνωρίστηκε το ίχνος R_I^* με τη μικρότερη μέση απόσταση από το άγνωστο ίχνος.



Μετρήσεις Αποστάσεων Ανυσμάτων

Παραμέτρων στην ΑΑΟ (1/5)

Είναι επιθυμητό η μέτρηση απόστασης $d(x, y)$ ανάμεσα σε δύο διανύσματα x και y να ικανοποιεί τις ακόλουθες εξισώσεις για αποδοτική χρήση στην αναγνώριση ομιλίας:

- a) Συμμετρία: $d(x, y) = d(y, x)$
- b) Θετικός ορισμός: $d(x, y) > 0, x \neq y$
 $d(x, y) = 0, x = y$

Η απλή ευκλείδεια απόσταση χρησιμοποιείται σε πολλές περιπτώσεις μέτρησης της $d(x, y)$.

Αποστάσεις με βάση μη παραμετρική φασματική ανάλυση

1) Μέθοδος τράπεζας ζωνοπερατών φίλτρων

Οι διαφορές της λογαριθμικής εξόδου για κάθε ζωνοπερατό ισαπέχον φίλτρο:

- Αθροίζονται για όλα τα φίλτρα ή
- Αθροίζονται τα τετράγωνα των τιμών τους



Μετρήσεις Αποστάσεων Ανυσμάτων Παραμέτρων στην ΑΑΟ (2/5)

Αποστάσεις με βάση μη παραμετρική φασματική ανάλυση

2) Μέθοδος FFT

Οι φασματικές τιμές λαμβάνονται σε ίσα διαστήματα σε ένα γραμμικό άξονα συχνότητας αλλά συνήθως επαναδειγματοληπτούνται σε ίσα διαστήματα:

- Σε λογαριθμική κλίμακα συχνότητας
- Στην κλίμακα Bark
- Στην κλίμακα Mel

Η κλίμακα Bark αντιστοιχεί στην κλίμακα της βασικής μεμβράνης στο περιφερειακό ακουστικό σύστημα. Ορίζεται ως:

$$B = 13 \arctan(0,76f) + 3,5 \arctan(f/7,5)^2$$

(B, f αναπαριστούν την κλίμακα Bark και τη συχνότητα σε KHz)

Η κλίμακα Mel αντιστοιχεί στην ακουστική αίσθηση του ύψους τόνου.

Η σχέση ανάμεσα στη συχνότητα f (KHz) και **Mel** είναι:

$$Mel = 1000 \log_2(1 + f)$$

Οι κλίμακες Bark και Mel σχεδόν συμπίπτουν με τη λογαριθμική κλίμακα συχνότητας στο διάστημα συχνότητας που υπερβαίνει το 1 KHz.



Μετρήσεις Αποστάσεων Ανυσμάτων Παραμέτρων στην ΑΑΟ (3/5)

Αποστάσεις με βάση μη παραμετρική φασματική ανάλυση

3) Μέθοδος cepstrum

- Η ευκλείδεια απόσταση ανάμεσα σε διαστήματα που αποτελούνται από συντελεστές cepstrum χαμηλότερης τάξης αντιστοιχεί στην απόσταση ανάμεσα σε εξομαλυμένα λογαριθμικά φάσματα.
- Συντελεστές cepstrum σε Mel οι οποίοι μετατρέπονται από το λογαριθμικό φάσμα το οποίο επαναδειγματοληπτείται σε συχνότητα κλίμακας Mel.

4) Μέθοδος συνάρτησης αυτοσυσχέτισης

Η απόσταση ανάμεσα στα διανύσματα της συνάρτησης αυτοσυσχέτισης πολλαπλασιαζόμενη από το παράθυρο καθυστέρησης αντιστοιχεί στην απόσταση ανάμεσα σε εξομαλυμένα φάσματα.



Μετρήσεις Αποστάσεων Ανυσμάτων Παραμέτρων στην ΑΑΟ (4/5)

Αποστάσεις με βάση παραμετρική ανάλυση LPC

Αν $f(\lambda)$ και $g(\lambda)$ αναπαριστούν τις φασματικές περιβάλλουσες με βάση το μοντέλο LPC για ένα πρότυπο ίχνος αναφοράς και την ομιλία εισόδου αντίστοιχα, τότε:

$$f(\lambda) = u^{(f)} R^{(f)} / 2\pi \sum_{k=-p}^{k=p} A_k^{(f)} e^{jk\lambda} \text{ και } g(\lambda) = u^{(g)} R^{(g)} / 2\pi \sum_{k=-p}^{k=p} A_k^{(g)} e^{jk\lambda}$$

Όπου u είναι η ενέργεια, r_i οι συντελεστές αυτοσυσχέτισης, a_i οι συντελεστές πρόβλεψης, A_i είναι η παράμετρος μέγιστης πιθανότητας, R είναι το κανονικοποιημένο υπόλοιπο και c_n είναι οι συντελεστές cepstrum.

Για τον καθορισμό της απόστασης ανάμεσα στο $f(\lambda)$ και $g(\lambda)$ έχουν προταθεί διάφορες αποστάσεις εκ των οποίων οι κυριότερες δύο είναι:

1) Φασματική απόσταση μέγιστης πιθανότητας (Itakura - Saito)

$$\begin{aligned} E &= 1/2\pi \int_{-\pi}^{\pi} (\log f(\lambda)/g(\lambda) + g(\lambda)/f(\lambda) - 1) d\lambda = \\ &= \log u^{(f)} R^{(f)} / u^{(g)} R^{(g)} + u^{(g)} / u^{(f)} R^{(f)} \sum A_k^{(f)} r_k^{(g)} - 1 \end{aligned}$$

Όταν το $d\lambda = \log f(\lambda) - \log g(\lambda)$ είναι μικρό, η απόσταση E διαφέρει ελάχιστα από την τετραγωνική λογαριθμική φασματική απόσταση.



Μετρήσεις Αποστάσεων Ανυσμάτων Παραμέτρων στην ΑΑΟ (5/5)

Αποστάσεις με βάση παραμετρική ανάλυση LPC

2) Απόσταση LPC Cepstrum

$$L^2 = \sum_{N=-\infty}^{\infty} \left(c_n^{(f)} - c_n^{(g)} \right)^2$$
$$= \left(\log u^{(f)} R^{(f)} / u^{(g)} R^{(g)} \right)^2 + 2 \sum_{N=1}^{\infty} \left(c_n^{(f)} - c_n^{(g)} \right)^2 = 1/2\pi \int_{-\pi}^{\pi} \{d(\lambda)\}^2 d\lambda$$

Η απόσταση αυτή είναι χρήσιμη μέτρηση απόστασης για τρεις κυρίως λόγους:

1. Μπορεί εύκολα να υπολογισθεί από τους συντελεστές γραμμικής πρόβλεψης.
2. Αντιστοιχεί άμεσα στη λογαριθμική απόσταση ανάμεσα στις φασματικές περιβάλλουσες LPC.
3. Ικανοποιεί τις απαιτήσεις για συμμετρία και θετικά ορισμένο χαρακτηριστικό.

Ανυσματική Κβάντιση (VQ)

- Το σήμα της ομιλίας αναλύεται σε μία αλληλουχία χρονικών πλαισίων.
- Κάθε πλαίσιο αναπαριστάται από ένα σύνολο k αριθμητικών τιμών ή ένα άνυσμα με k διαστάσεις σε ένα χώρο k διαστάσεων.
- Βαθμωτή κβάντιση: κάθε παράμετρος στο άνυσμα κβαντίζεται χωριστά σε ένα ορισμένο αριθμό bits.

Μειονέκτημα: υποθέτει ότι οι τιμές έχουν μία ομοιόμορφη κατανομή στον ανυσματικό χώρο, πράγμα που δεν συμβαίνει.

- Ανυσματική κβάντιση: ο χώρος χωρίζεται σε N μη ομοιόμορφες κλειστές περιοχές και κάθε περιοχή αναπαριστάται με ένα άνυσμα που αντιστοιχεί στο κέντρο της περιοχής.
- Η συλλογή των N κεντροειδών ανυσμάτων ονομάζεται κωδικοβιβλίο (codebook).
- Σε κάθε ένα μέλος V του κωδικοβιβλίου δίδεται μία ετικέτα ή μία διεύθυνση A_q η οποία αντιστοιχεί στο πλησιέστερο κέντρο, μέσω σύγκρισης με όλα τα κεντροειδή V_i :

$$A_q = \text{Argmin}[d(V_i - \bar{V}_i)]$$

(όπου A_q δηλώνει την τιμή του δείκτη i για την οποία η έκφραση μέσα στην αγκύλη είναι ελάχιστη)



Παράδειγμα Μηχανής Αναγνώρισης Φωνηέντων (1/2)

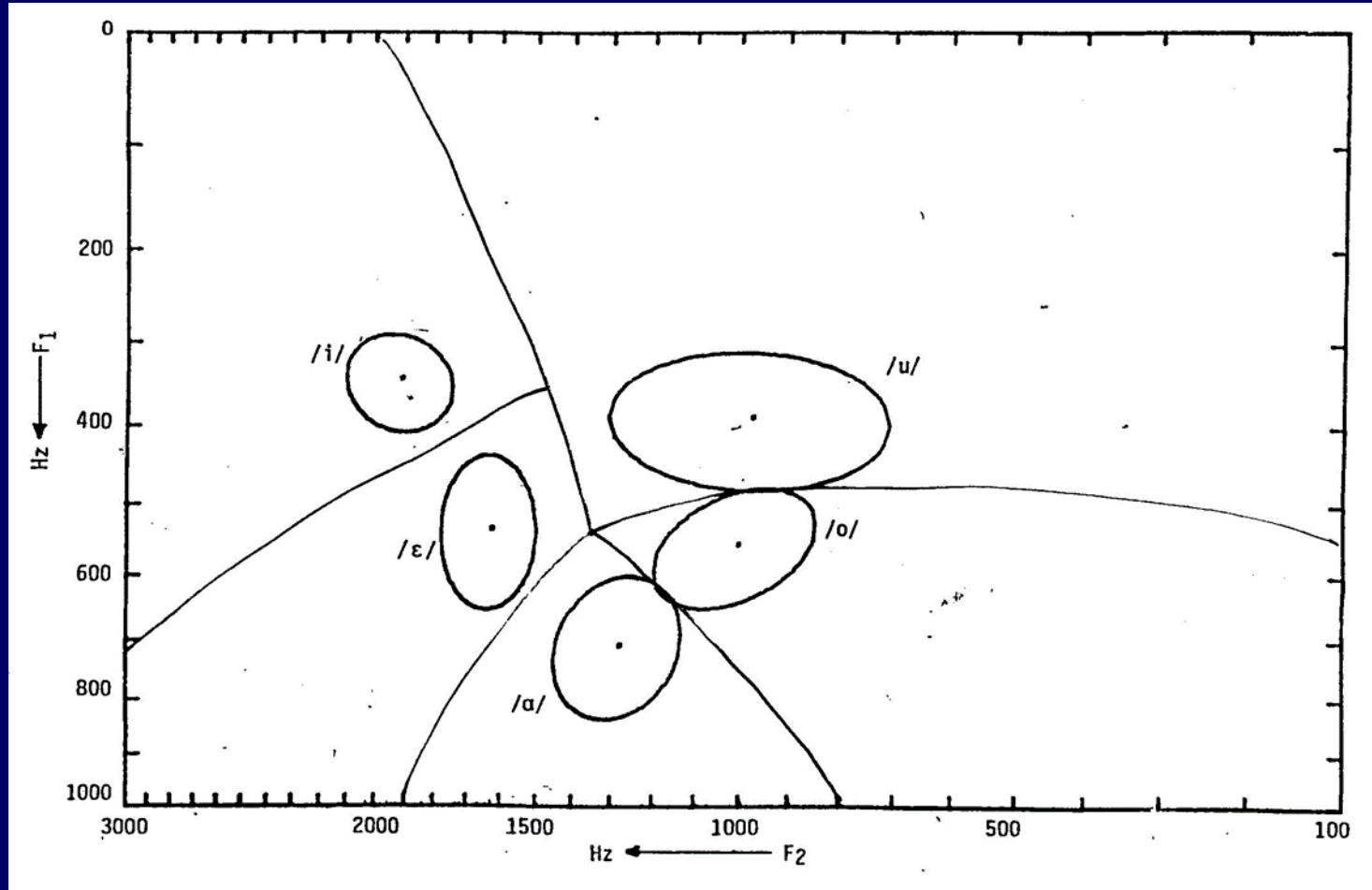
ΕΚΜΑΘΗΣΗ

- Φωνήεντα: [α], [ε], [ι], [ο], [υ]
- Επιλογή Λεξιλογίου: ΣΦΣ (Σ=Σύμφωνο, Φ=Φωνήεν)
- Επιλογή ομιλητών
- Εκφώνηση
- Ψηφιοποίηση
- Επιλογή παραμέτρων

ΑΝΑΓΝΩΡΙΣΗ

- Ομιλητής
- Εκφώνηση άγνωστου φωνήεντος
- Ψηφιοποίηση
- Εξαγωγή παραμέτρων
- Σύγκριση με δεδομένα εκμάθησης
- Απόφαση

Παράδειγμα Μηχανής Αναγνώρισης Φωνηέντων (2/2)



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:

1. 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.
2. Ένα γλωσσικό μοντέλο που προσδιορίζει επιτρεπτές ακολουθίες λέξεων.

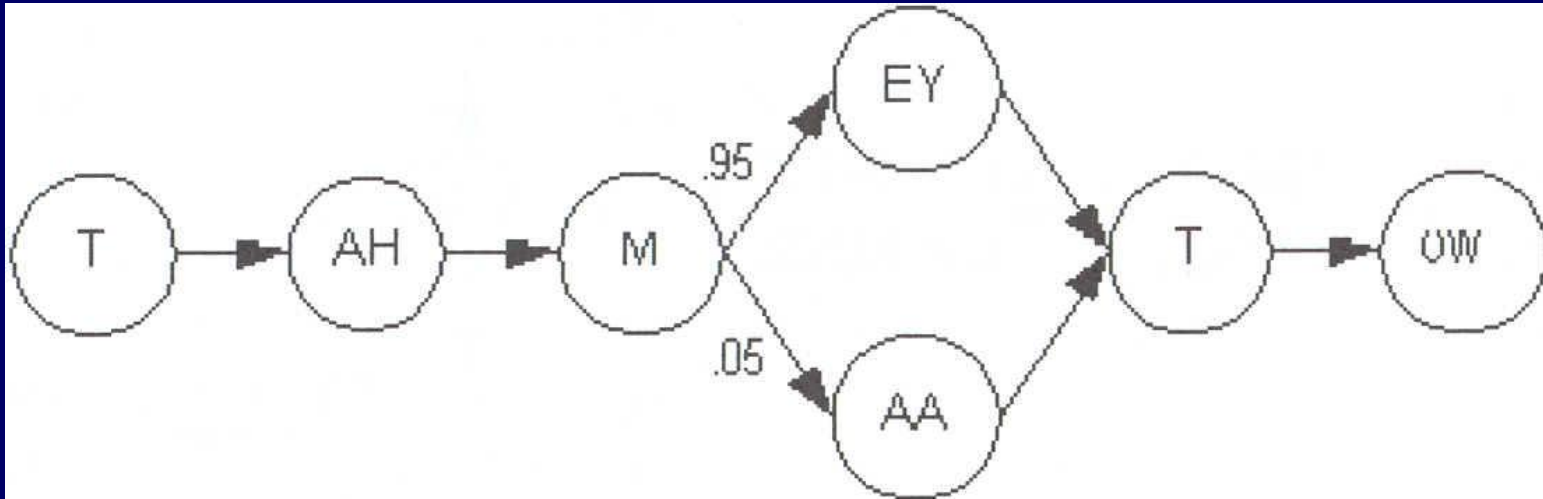
Το ακουστικό μοντέλο

- 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 πανεπιστημίου 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 (1/2)

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 a_{11}, a_{22}, \dots) 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.

2. κάθε κατάσταση έχει μια κατανομή από πιθανές εξόδους (συμβολίζονται με $b_1(o_1), b_1(o_2), b_2(o_3), \dots$).

σε κάθε κατάσταση το σύστημα ταιριάζει ένα μέρος της εισόδου (ένα πλαίσιο διάρκειας $10ms$, που συμβολίζεται με o_1, o_2, o_3, \dots), 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.



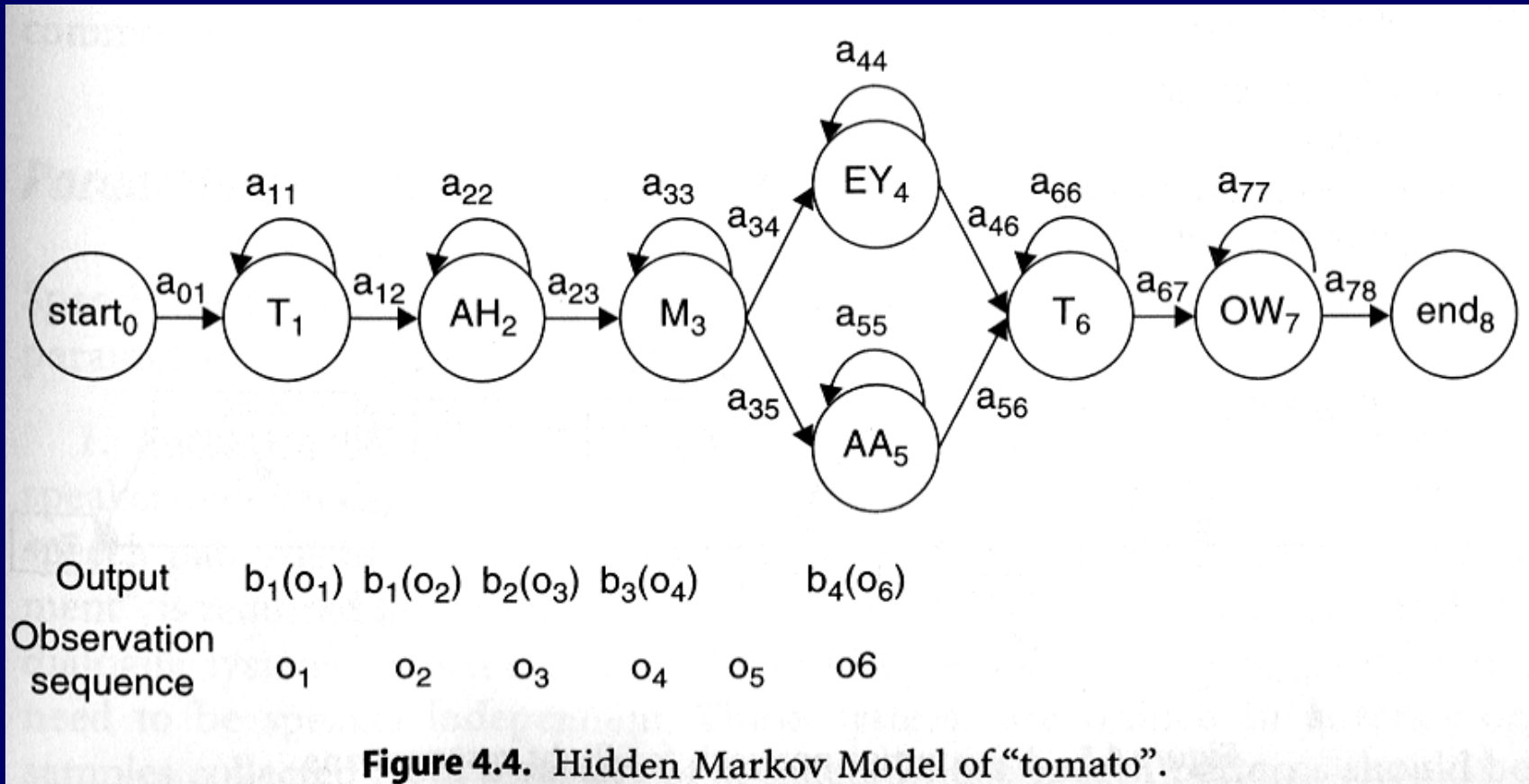
Κρυφά Μοντέλα Markov (HMM) για Ακουστική Μοντελοποίηση (2/2)

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

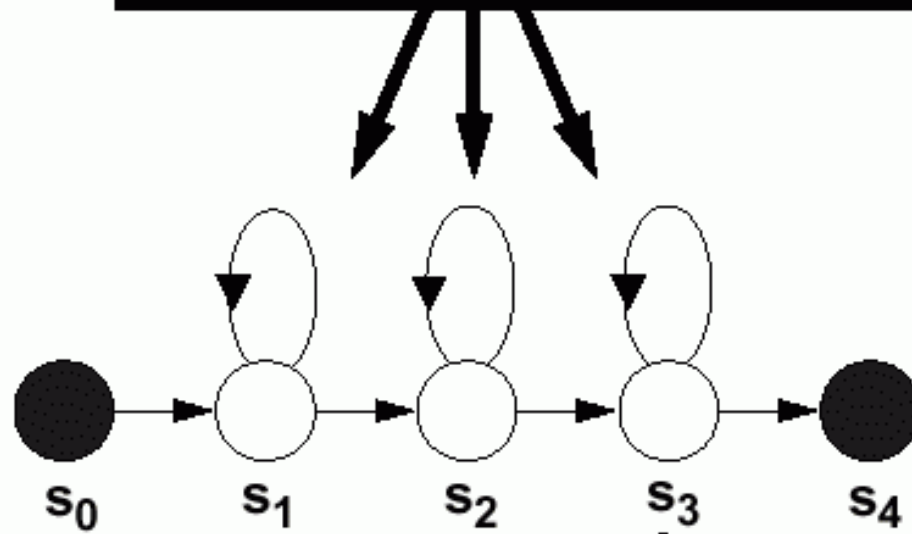
Είναι πιθανό να δει κανείς ότι το μοντέλο παράγει ένα συγκεκριμένο σύμβολο σαν έξοδο αλλά δεν είναι δυνατό να γνωρίζουμε από ποια κατάσταση προήρθε αυτό το σύμβολο.

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



THREE TWO FIVE EIGHT



Το γλωσσικό μοντέλο

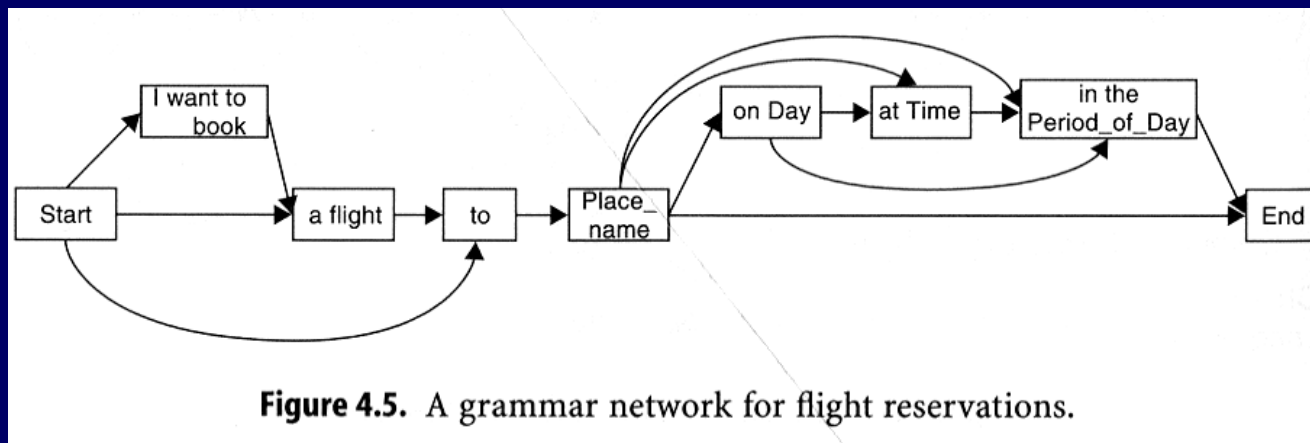
- περιέχει γνώση σχετικά με:
 - επιτρεπτές ακολουθίες λέξεων, και
 - το ποιες λέξεις είναι πιο πιθανές σε συγκεκριμένες ακολουθίες.

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).

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



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.

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., ένας αριθμός ομαλά επαναλαμβανόμενων φράσεων μπορούν να αναγνωρισθούν, όπως:

- για + Όνομα_Προορισμού
- στις + Ημερομηνία
- το + Περίοδος_της_Ημέρας (π.χ. μεσημέρι, απόγευμα κ.α.)
- στις | περίπου τις | πριν τις | μετά τις + Ώρα

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".

HMM – Πρόβλημα 1

Ποια είναι η πιθανότητα για ένα συγκεκριμένο μοντέλο HMM να παράγει μία αλληλουχία παρατηρήσεων (σαν την AAB);

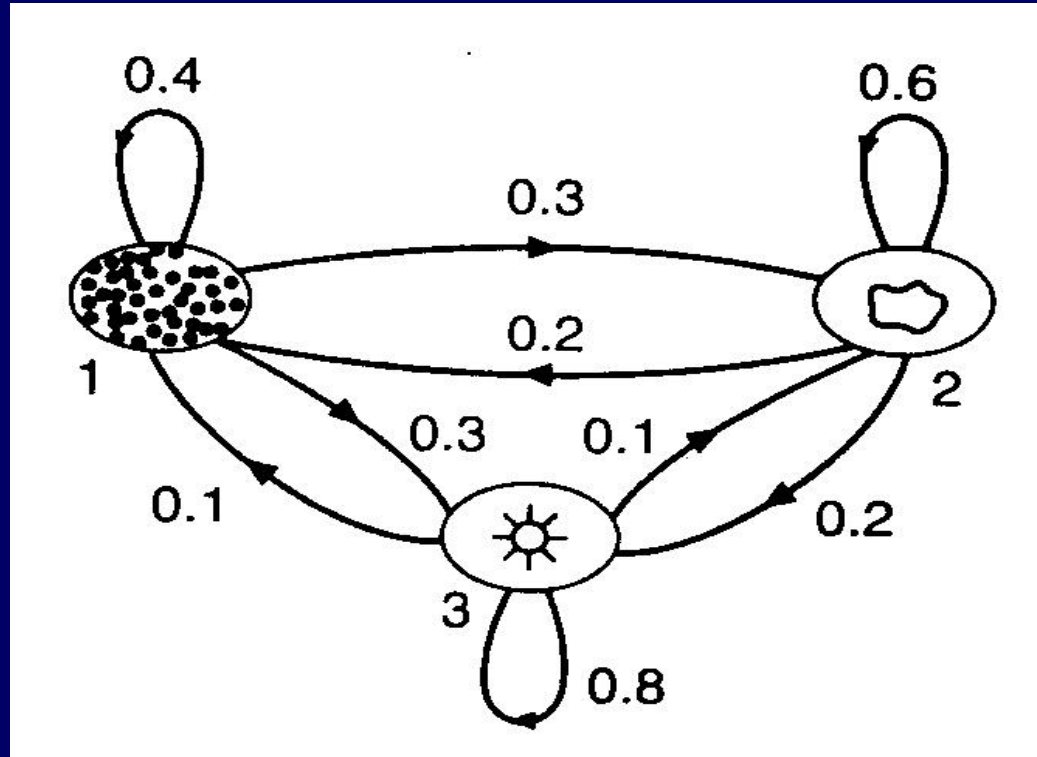
Εφαρμογή του προβλήματος σε αναγνώριση ομιλίας

- Θεωρούμε αναγνώριση μεμονωμένων λέξεων.
- Για κάθε λέξη του λεξιλογίου σχεδιάζουμε ένα χωριστό HMM.
- Ένα σήμα ομιλίας αναπαριστάται (ύστερα από ανυσματική κβάντιση) σαν μία χρονική αλληλουχία φασματικών διανυσμάτων τα οποία κωδικοποιούνται σε ένα κωδικοβιβλίο.
- Μπορούμε να εφαρμόσουμε τον παραπάνω αλγόριθμο σε εκφωνήσεις γνωστών λέξεων για τη διαδικασία εκμάθησης του συστήματος.
- Η αναγνώριση άγνωστης εκφώνησης γίνεται με το να υπολογίσουμε την πιθανότητα που γεννά κάθε HMM για την άγνωστη εκφώνηση και επιλογή του μοντέλου που δίνει το μεγαλύτερο σκορ.



Εισαγωγή στα HMM (1/2)

$$A = \{a_{ij}\} = \begin{pmatrix} 0.4 & 0.2 & 0.1 \\ 0.3 & 0.6 & 0.1 \\ 0.3 & 0.2 & 0.8 \end{pmatrix}$$



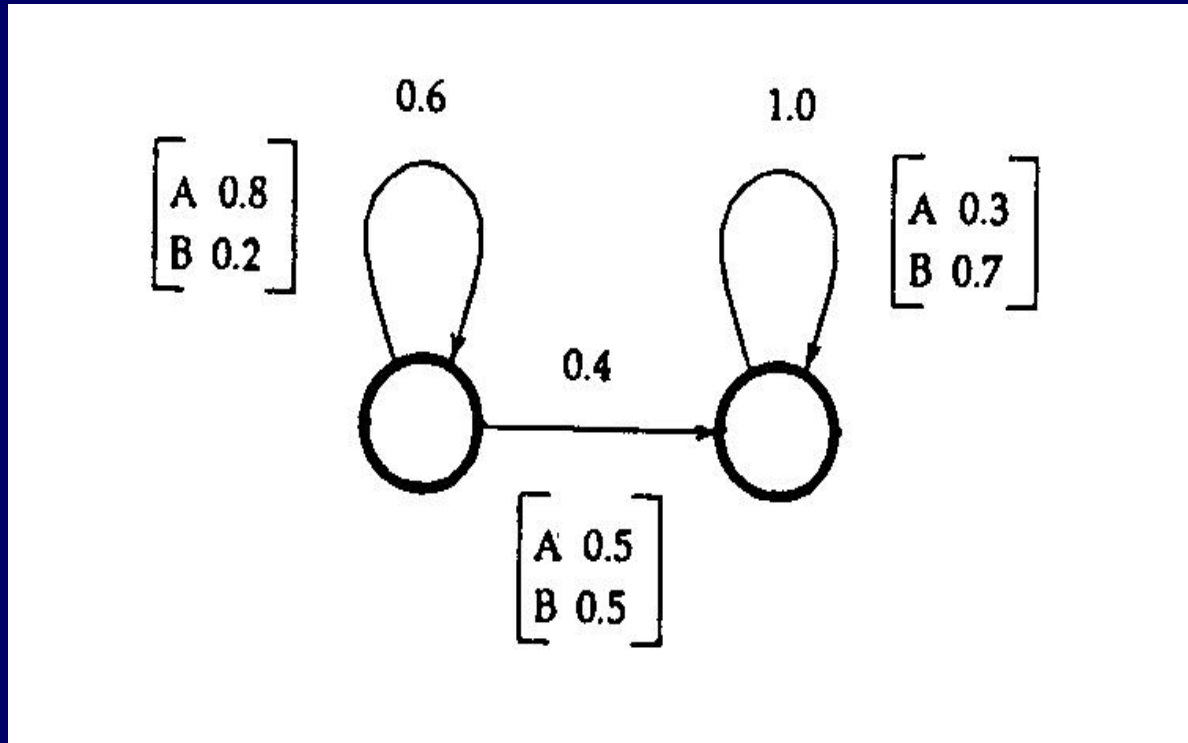
Απλό Μοντέλο Markovon πρώτης τάξης

Ο καιρός μια μέρα t χαρακτηρίζεται από μία από τις τρεις καταστάσεις και από τον πίνακα μετάπτωσης καταστάσεων A που δημιουργήθηκε παρατηρώντας τον καιρό κάθε μέρα μια συγκεκριμένη ώρα.

Ερωτήσεις για το παράδειγμα MM καιρού

- 1) Ποια είναι η πιθανότητα σύμφωνα με το μοντέλο, ο καιρός τις επόμενες 8 μέρες να είναι: ήλιος – ήλιος – ήλιος – βροχή – βροχή – ήλιος – σύννεφα – ήλιος
απάντηση: $1,53 \cdot 10^{-4}$
- 2) Θεωρούμε ότι το σύστημα ευρίσκεται σε κάποια δεδομένη κατάσταση. Ποιος είναι ο αναμενόμενος αριθμός ημερών να παραμείνει στην ίδια κατάσταση;
απάντηση: ήλιος \rightarrow 5 ημέρες
σύννεφα \rightarrow 2,5 ημέρες
βροχή \rightarrow 1,67 ημέρες

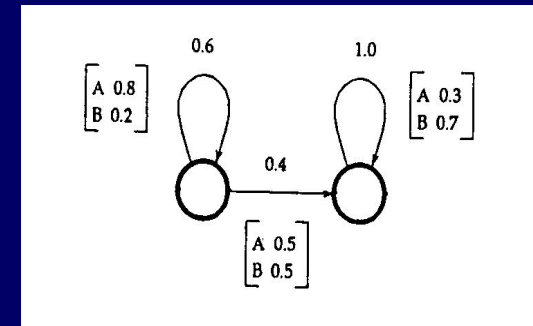
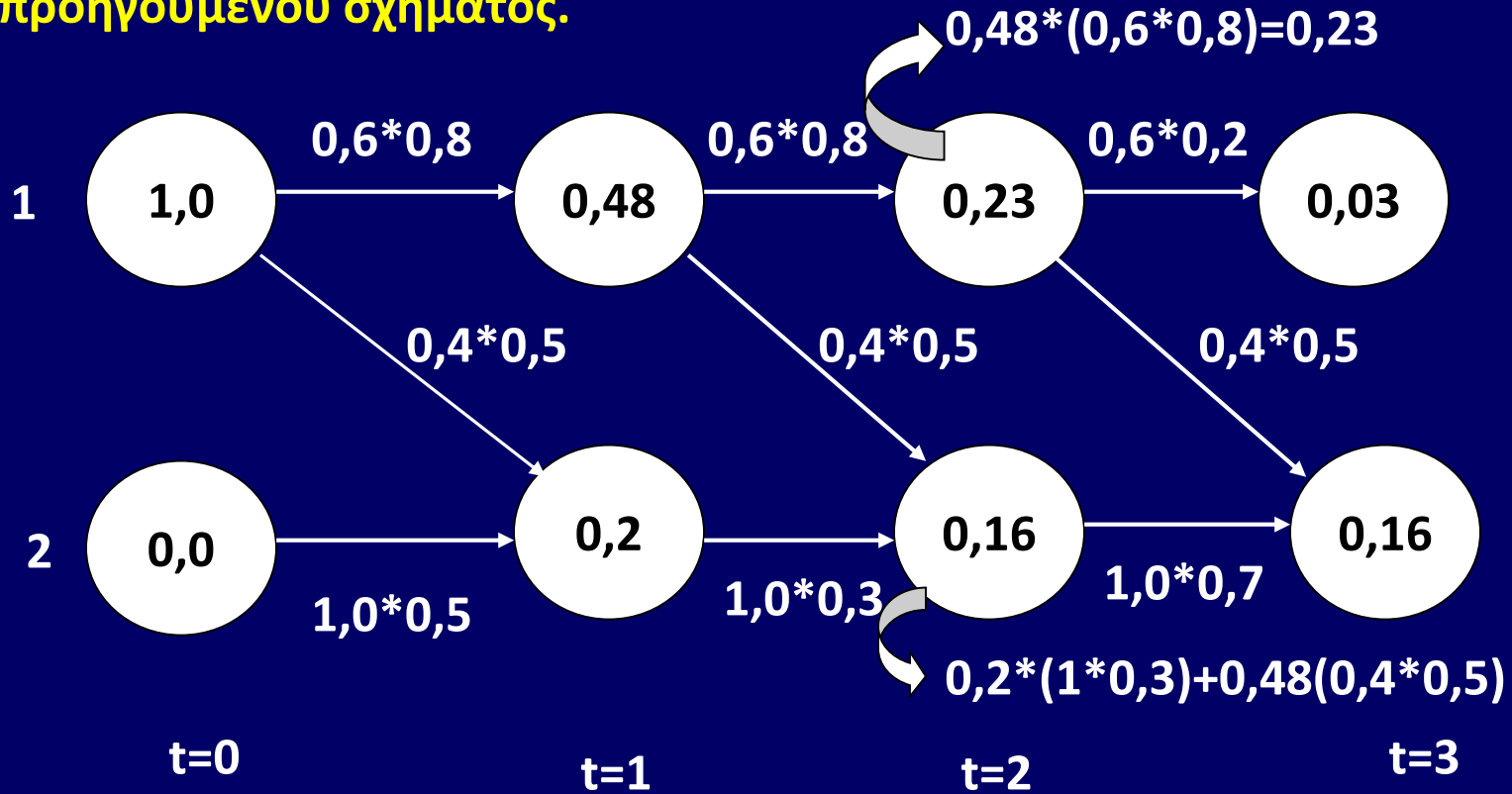
Εισαγωγή στα HMM (2/2)



Απλό HMM. Υπάρχουν δύο πιθανές παρατηρήσεις (A & B) σε κάθε μετάπτωση, επομένως τώρα οι παρατηρήσεις συνδέονται με τις μεταπτώσεις και όχι με τις καταστάσεις.

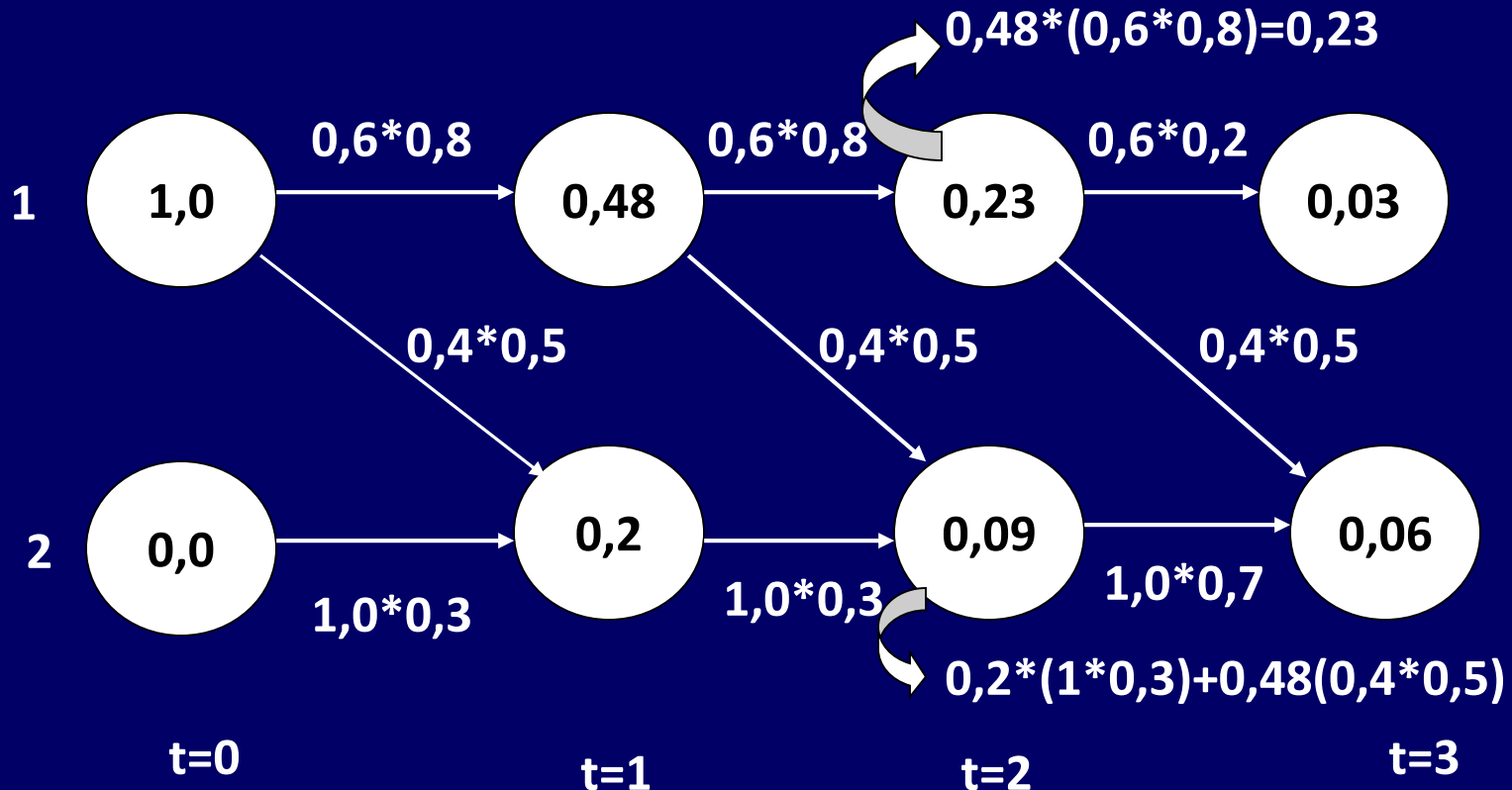
Πρόβλημα 1

Παράδειγμα Υπολογισμού της πιθανότητα εμφάνισης της αλληλουχίας παρατηρήσεων AAB δεδομένου του HMM δύο καταστάσεων του προηγούμενου σχήματος.



Πρόβλημα 2

Αλγόριθμος Viterbi για τον υπολογισμό της καλύτερης αλληλουχίας καταστάσεων που αντιστοιχεί σε μία αλληλουχία παρατηρήσεων.



ΑΑΒ: αναγνώριση, Αποτέλεσμα: 1-1-2-2



HMM - Πρόβλημα 2

Καθορισμός της καλύτερης αλληλουχίας καταστάσεων

Εφαρμογή του προβλήματος σε αναγνώριση ομιλίας συνδεδεμένων λέξεων.

Υποθέτουμε ότι έχουμε στη διάθεσή μας τα HMM για τις λέξεις του λεξιλογίου, αλλά τώρα μπορούμε να κατασκευάσουμε προτάσεις με συνδυασμό τους.

Το να βρούμε τον πιο πιθανό (κρυφό) δρόμο αλληλουχιών καταστάσεων μέσα σε ένα σύνθετο μοντέλο μας επιτρέπει να καθορίσουμε την πιο πιθανή αλληλουχία λέξεων που αντιστοιχεί στην αλληλουχία παρατηρήσεων.

Ο αλγόριθμος Viterbi είναι παρόμοιος με τον πρόσω. Η μόνη διαφορά είναι ότι αντί να αθροίζω τις πιθανότητες μεταπτώσεων για όλες τις καταστάσεις του προηγούμενου χρονικά βήματος, κρατώ μόνο την πιο πιθανή μετάπτωση.

Για το παράδειγμα είναι: 1 – 1 – 2 – 2.

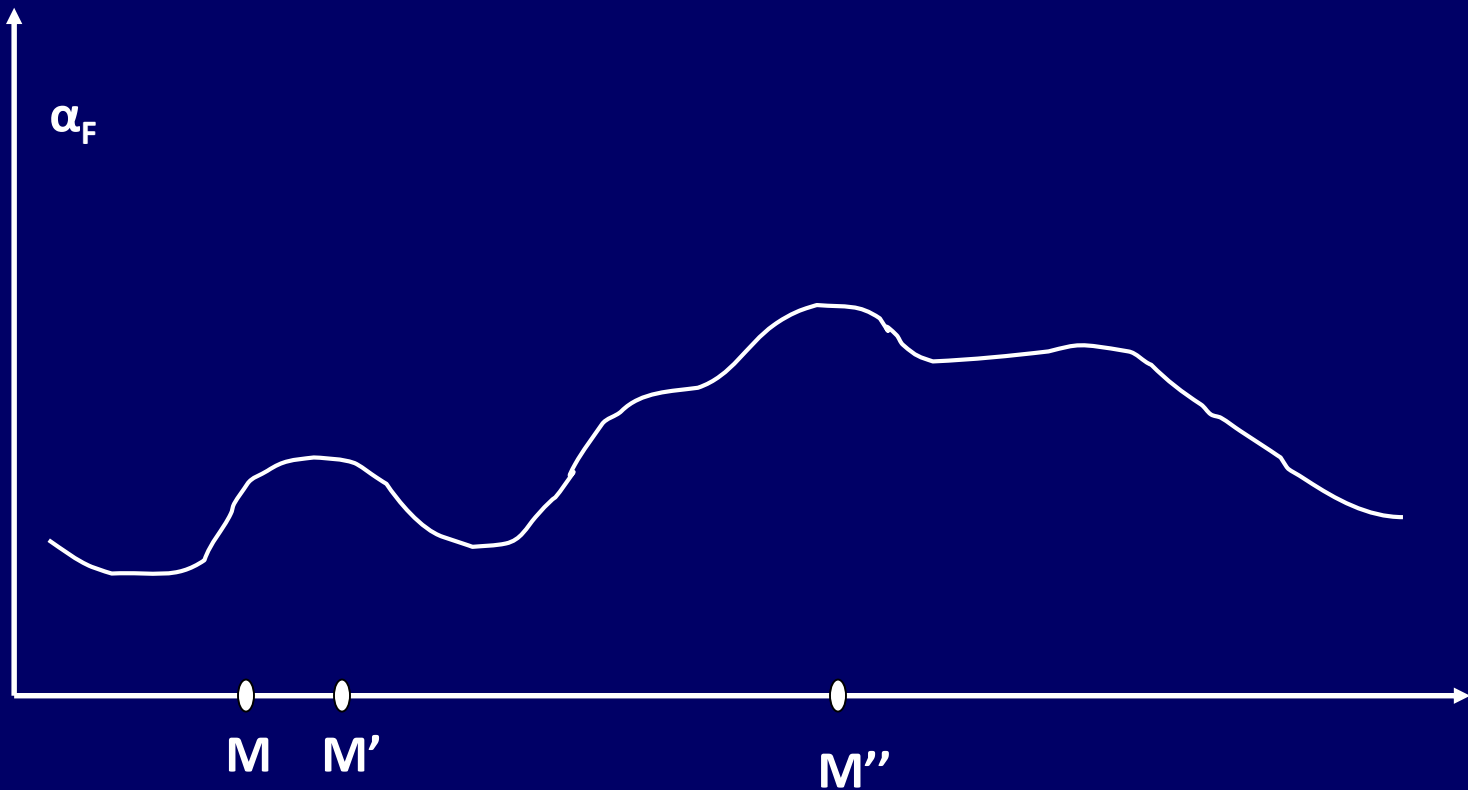
Σημείωση: Για μέρος της ακολουθίας παρατηρήσεων, η καλύτερη αλληλουχία καταστάσεων μπορεί να μην είναι η ίδια από εκείνη του συνόλου.

Παράδειγμα: για την AA η καλύτερη αλληλουχία καταστάσεων είναι η 1-1-1.

Πρόβλημα 3

Ο χώρος των παραμέτρων κατά τη διαδικασία εκμάθησης σε HMM.

Διαδικασία επανεκτίμησης Baum – Welch.



HMM - Πρόβλημα 3

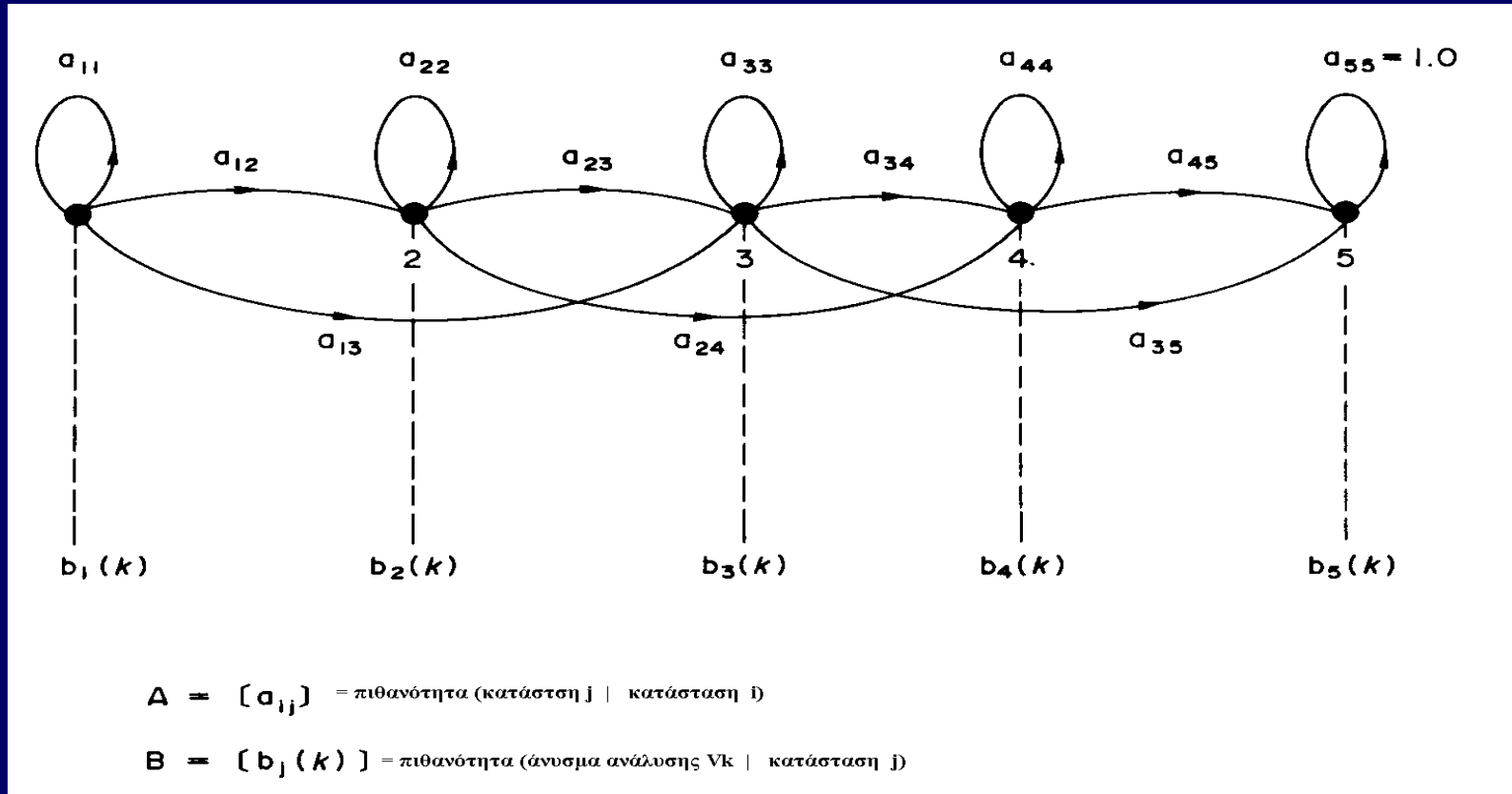
Πώς να ρυθμιστούν οι παράμετροι του μοντέλου για μεγιστοποίηση της πιθανότητας των αλληλουχιών παρατηρήσεων εκμάθησης, δεδομένης μιας δομής του μοντέλου

- Το πιο δύσκολο πρόβλημα από τα τρία
- Δεν υπάρχει γνωστός αναλυτικός τρόπος για τη λύση του προβλήματος
- Μπορούμε όμως να επιλέξουμε παραμέτρους κατά τέτοιο τρόπο έτσι ώστε η πιθανότητα αυτή να είναι τοπικά μέγιστη.
- Η επιλογή μπορεί να γίνει με τη μέθοδο επανεκτίμησης Baum – Welch.
- Οι παλιές παράμετροι αντικαθιστώνται από νέες, οι οποίες υπολογίζονται με βάση τις προηγούμενες καθώς και από παρατηρήσεις εκμάθησης.

Δηλαδή: υποθέτουμε ότι το M παριστά τις παρούσες τιμές παραμέτρων, οι νέες τιμές M κάνουν τις παρατηρήσεις πιο πιθανές και οδηγούν σε μία μεγαλύτερη πρόσω πιθανότητα α_F .



Παράδειγμα HMM 5 καταστάσεων για αναγνώριση λέξεων (1/3)



Τη χρονική στιγμή t (που αντιστοιχεί σε συγκεκριμένο πλαίσιο) το μοντέλο παραμένει στην κατάσταση ή μετακινείται σε μία νέα και παράγει ένα συγκεκριμένο ίχνος (ή μία παρατήρηση) ομιλίας O με πιθανότητα b_{ij} . Η διαδικασία αυτή επαναλαμβάνεται έως ότου παραχθούν όλη η αλληλουχία των παρατηρήσεων.

Παράδειγμα HMM 5 καταστάσεων για αναγνώριση λέξεων (2/3)

$O = \{O_1, O_2, \dots, O_r\}$ = ακολουθία παρατήρησης (εκφώνηση εισόδου).

T = μήκος (διάρκεια) της ακολουθίας παρατήρησης.

$Q = \{q_1, q_2, \dots, q_N\}$ = (κρυφές) καταστάσεις στο μοντέλο.

N = αριθμός καταστάσεων.

$V = \{v_1, v_2, \dots, v_m\}$ = διακριτό σύνολο πιθανών παρατηρήσεων κωδικοβιβλίου συμβόλων.

M = αριθμός συμβόλων παρατήρησης (μέγεθος κωδικοβιβλίου).

$A = \{a_{ij}\}, a_{ij} = Prob(q_j \text{ σε } t + 1 | q_i \text{ σε } t)$ = κατανομή πιθανοτήτων μετάπτωσης από την κατάσταση i τη χρονική στιγμή t στην κατάσταση j τη χρονική στιγμή $t+1$.

$B = \{b_j(k)\}, c = Prob(v_k \text{ σε } t | q_j \text{ σε } t)$ = κατανομή πιθανοτήτων παρατήρησης των συμβόλων στην κατάσταση j .

$\pi = \{\pi_i\}, \pi_i = Prob(q_j \text{ σε } t = 1)$ = κατανομή της αρχικής κατάστασης

Παράδειγμα HMM 5 καταστάσεων για αναγνώριση λέξεων (3/3)

Μια ακολουθία παρατήρησης O προκύπτει ως εξής:

Βήμα 1: θέστε $t = 1$

Βήμα 2: Επιλέξτε μία αρχική κατάσταση I σύμφωνα με την κατανομή της αρχικής κατάστασης π .

Βήμα 3: Επιλέξτε O_t σύμφωνα με το $b_j(k)$.

Βήμα 4: Επιλέξτε j σύμφωνα με $\{a_{ij}\}$ ($j = 1, 2, \dots, N$).

Βήμα 5: θέστε $t \leftarrow t + 1$. Επέστρεψε στο βήμα 3 εφόσον $t < T$, διαφορετικά τερμάτισε τη διαδικασία.

Ο προσδιορισμός ενός HMM αφορά την επιλογή του αριθμού των καταστάσεων N , καθώς και του αριθμού των διακριτών συμβόλων M , και τον προσδιορισμό των τριών πυκνοτήτων πιθανότητας A , B και π .

Τα A , B και $\pi = \pi_i$ αναπαριστούν χρονικές και φασματικές διαφοροποιήσεις σε κάθε λέξη. Υπολογίζονται με τη χρήση των δεδομένων εκμάθησης.

Τα Τρία Προβλήματα ΗΜΜ

Πρόβλημα 1: Πρόβλημα αξιολόγησης

Δεδομένης της ακολουθίας παρατήρησης $O = \{O_1, O_2, \dots, O_T\}$ και του μοντέλου $\lambda = (A, B, \pi)$, πως μπορεί να υπολογιστεί η πιθανότητα $Prob(O|\lambda)$ της ακολουθίας παρατήρησης;

Χρησιμοποιείται για τη βαθμολόγηση κάθε μοντέλου λέξεων με βάση μία δεδομένη δοκιμαστική ακολουθία παρατήρησης για την αναγνώριση μίας άγνωστης λέξης.

Πρόβλημα 2: Πρόβλημα αποκάλυψης της ακολουθίας κρυφών καταστάσεων

Δεδομένης της ακολουθίας παρατήρησης $O = \{O_1, O_2, \dots, O_T\}$ πως μπορεί να επιλεγεί μία ακολουθία καταστάσεων $I = \{i_1, i_2, \dots, i_T\}_I$ η οποία να είναι, κατά τρόπο, βέλτιστη;

Χρησιμοποιείται για την κατανόηση του φυσικού νοήματος των καταστάσεων του μοντέλου.

Πρόβλημα 3: Πρόβλημα εκμάθησης

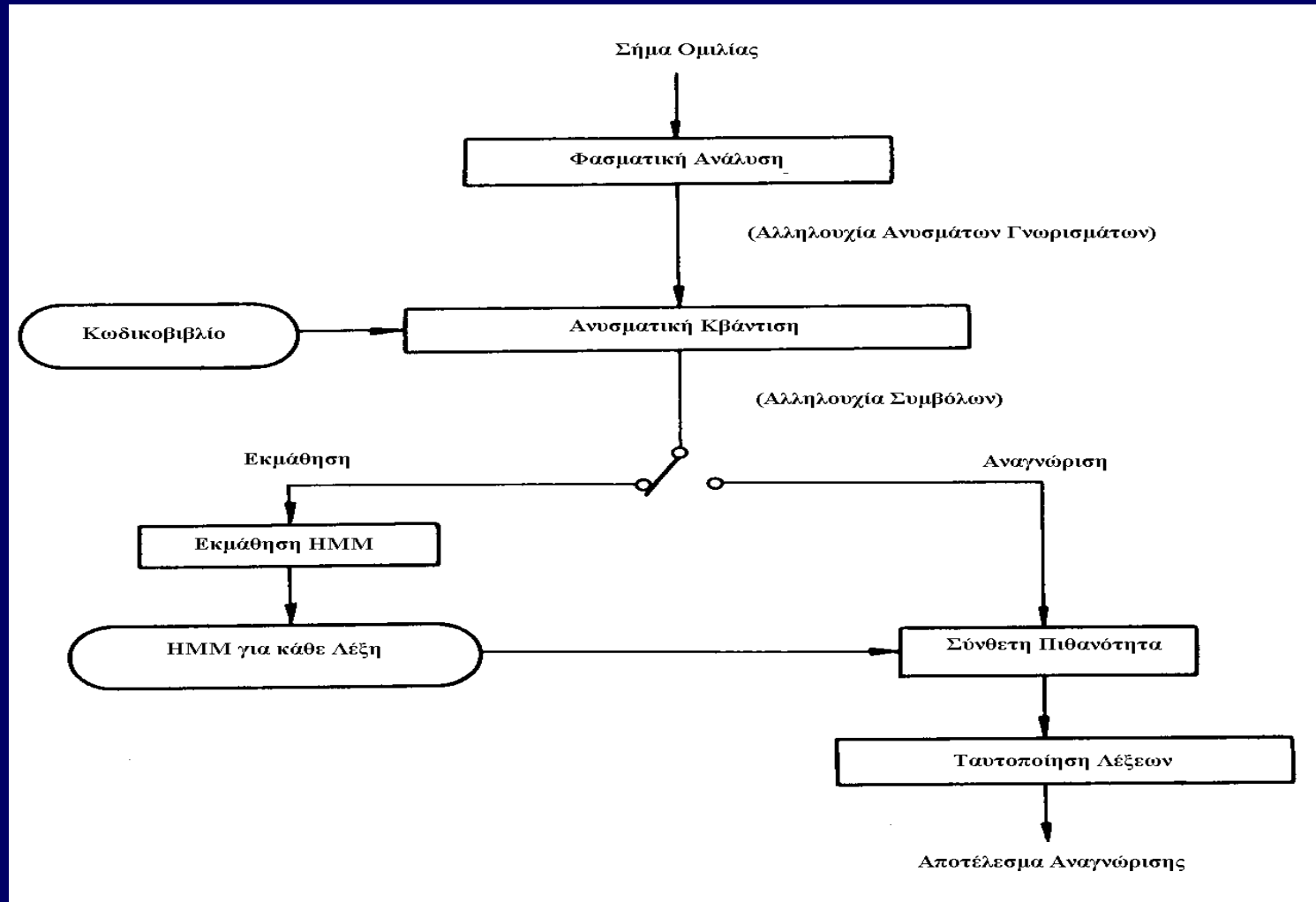
Πως μπορούν οι παράμετροι $\lambda = (A, B, \pi)$ του μοντέλου να ρυθμιστούν έτσι ώστε να μεγιστοποιούν την $Prob(O|\lambda)$;

Εφαρμόζεται για τη βέλτιστη εξαγωγή παραμέτρων του μοντέλου για κάθε μοντέλο λέξης χρησιμοποιώντας εκφωνήσεις εκμάθησης.



Στοχαστικές Μέθοδοι ΑΑΟ (1/5)

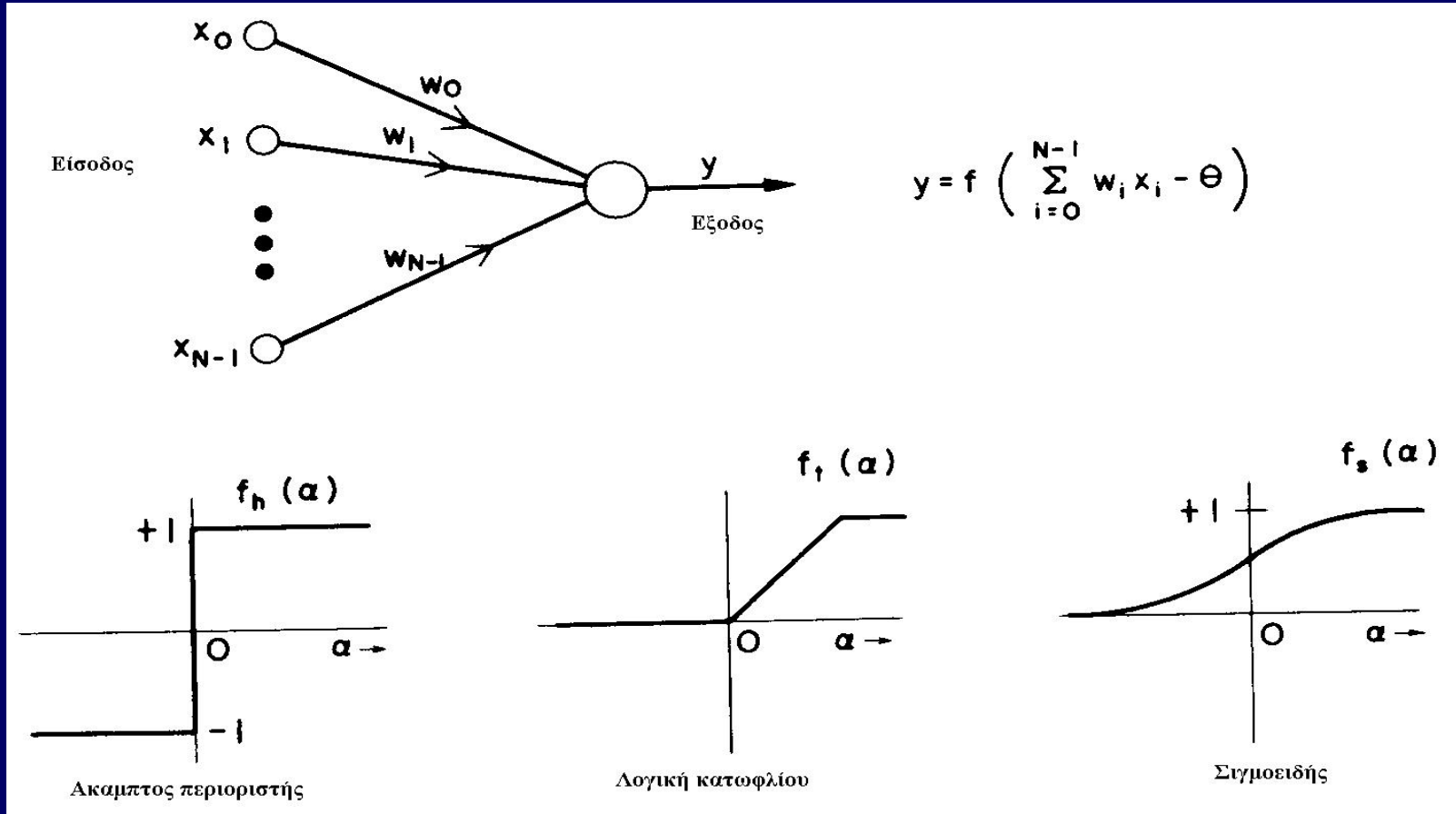
ΑΑΟ με Κρυφά Μοντέλα Markov (HMM)



Βασική Δομή ενός αναγνωριστή λέξεων με βάση το μοντέλο HMM.

Στοχαστικές Μέθοδοι ΑΑΟ (2/5)

Τεχνητά Νευρωνικά Δίκτυα



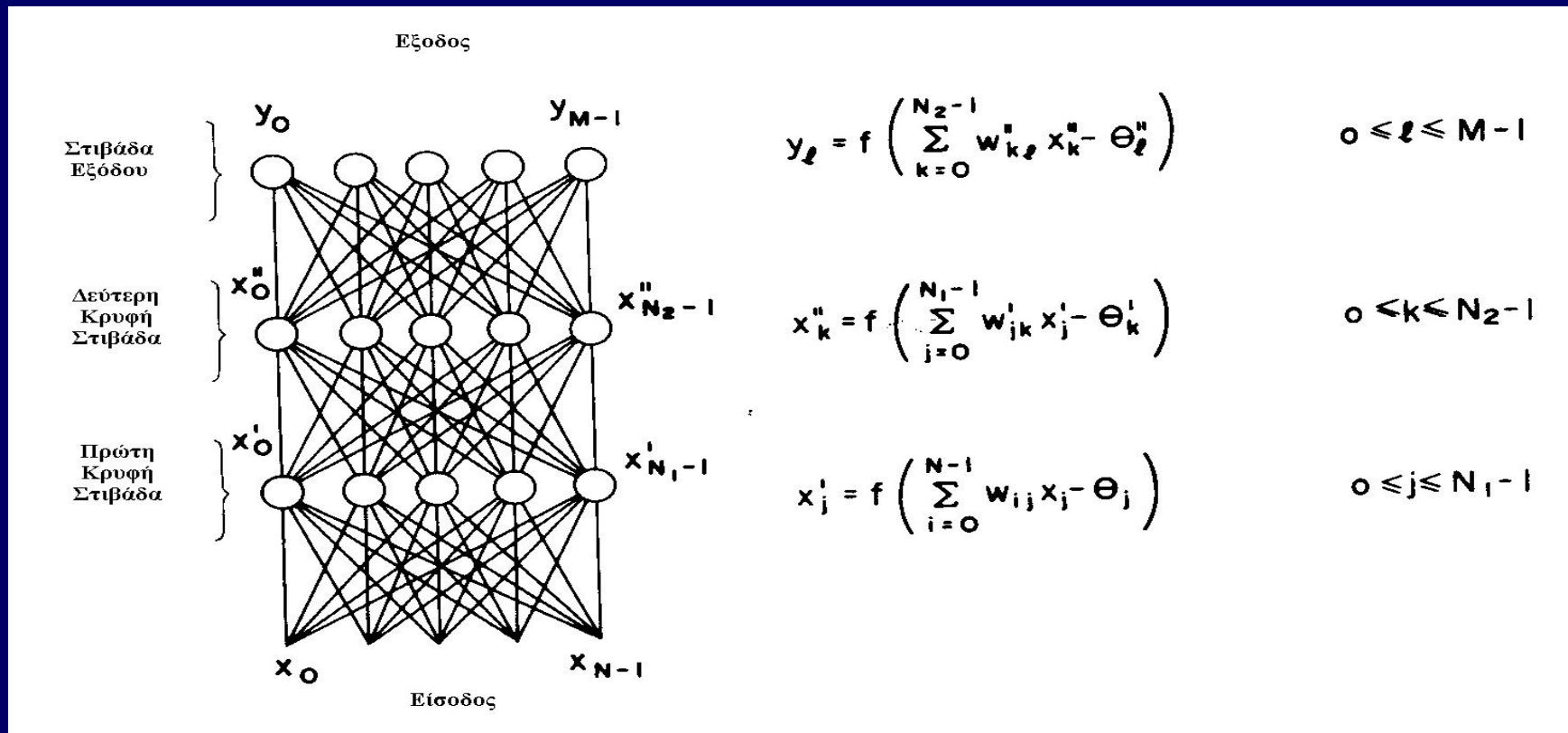
Στοιχείο νευρώνα και τρεις αντιπροσωπευτικές μη γραμμικότητες

Στοχαστικές Μέθοδοι ΑΑΟ (3/5)

Τα μοντέλα των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) δομούνται από πολλούς απλούς μη γραμμικούς υπολογιστικούς κόμβους (νευρώνες) οι οποίοι λειτουργούν παράλληλα και τίθενται σε μορφές οι οποίες εξομοιάζουν τα βιολογικά νευρωνικά δίκτυα. Ο κόμβος αθροίζει τις N εισόδους εφαρμόζοντας στην κάθε μία ένα βάρος και περνά το αποτέλεσμα μέσω μίας μη γραμμικότητας όπως φαίνεται στο προηγούμενο σχήμα. Ο κόμβος χαρακτηρίζεται από ένα εσωτερικό κατώφλι ή αρχικό σημείο μετάπτωσης θ και από το είδος της μη γραμμικότητας (μη γραμμικός μετασχηματισμός).


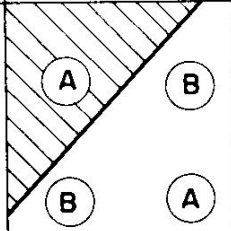
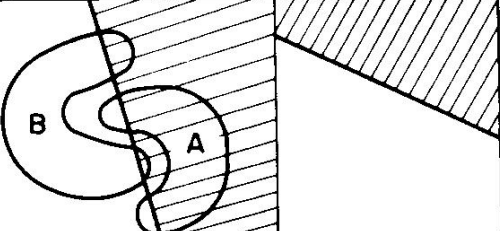
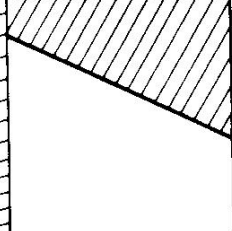
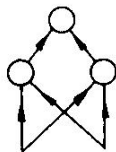
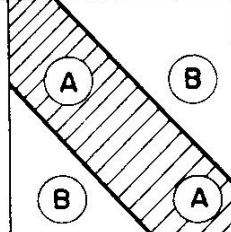
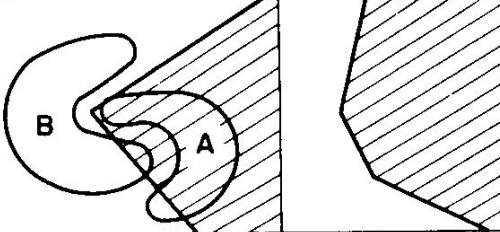
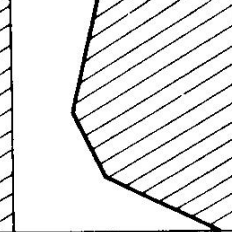
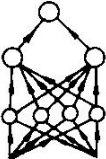
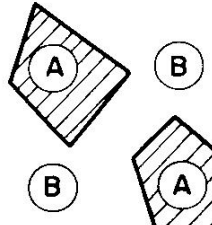
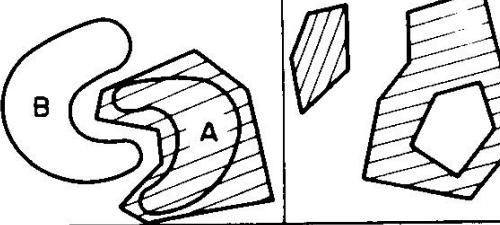
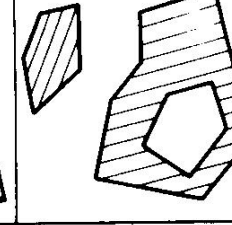
Στοχαστικές Μέθοδοι ΑΑΟ (4/5)

Ανάμεσα στα διάφορα είδη ΤΝΔ, τα perceptrons πολλαπλών στοιβάδων / στρωμάτων έχουν αποδειχθεί ικανά στην αντιμετώπιση πολλών προβλημάτων. Τα perceptrons πολλαπλών στρωμάτων είναι δίκτυα εμπρόσθιας ανάδρασης με ένα ή περισσότερα στρώματα κόμβων ανάμεσα στους κόμβους εισόδου και εξόδου.



Ένα perceptron τριών στρωμάτων με N εισόδους τιμών, M εξόδους και δύο στοιβάδες κρυφών μονάδων.

Στοχαστικές Μέθοδοι ΑΑΟ (5/5)

Δομή	Τύποι Περιοχών Απόφασης	Πρόβλημα XOR	Κλάσεις με περιοχές περιπλεγμένων βρόχων	Ποιά γενικά σχήματα περιοχών
<p>Μιάς Στιβάδας</p> 	<p>Ημιεπίπεδα περικλειόμενα από υπερεπίπεδα</p>			
<p>Δύο Στιβάδων</p> 	<p>Κυρτές ανοικτές ή κλειστές περιοχές</p>			
<p>Τριών Στιβάδων</p> 	<p>Αυθαιρέτες περιοχές (η πολυπλοκότητα περιορίζεται από τον αριθμό των κόμβων)</p>			

Είδη περιοχών απόφασης οι οποίες μπορούν να μετασχηματιστούν με perceptrons ενός ή πολλαπλών στρώματων με ένα και δύο στρώματα κρυφών μονάδων και δύο εισόδους. Η σκίαση υποδεικνύει τις περιοχές απόφασης για την τάξη A. Οι ομαλές κλειστές ισουψείς οριοθετούν τις κατανομές των εισόδων για τις τάξεις A και B. Οι κόμβοι σε όλα τα δίκτυα χρησιμοποιούν μη γραμμικότητες άκαμπτου περιοριστή.

Αναγνώριση – Κατανόηση Ομιλίας

Υψηλότερα επίπεδα αναγνώρισης (γλωσσολογική πληροφορία)

Επίπεδα:

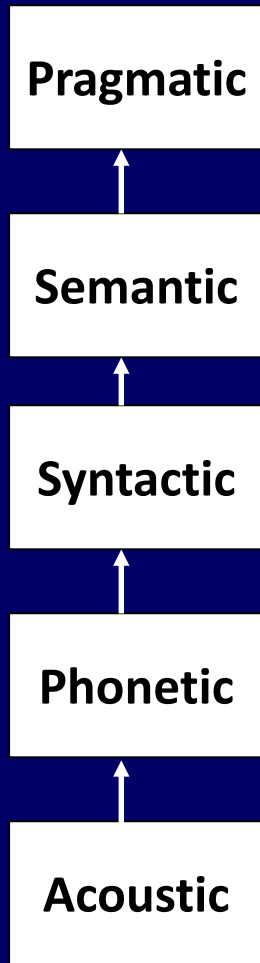
- Λεκτικό (ήχοι)
- Γραμματικό (λέξεις)
- Συντακτικό (φράσεις)

Συστήματα κατανόησης ομιλίας

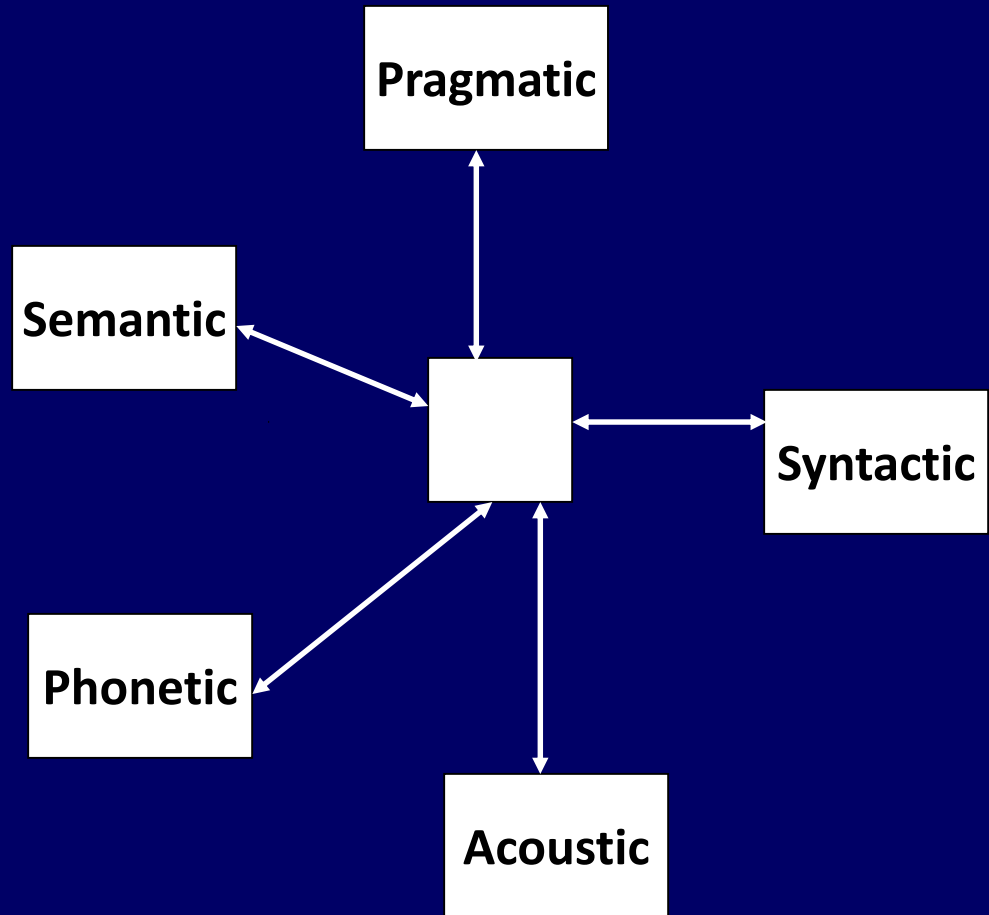
Επιπλέον πηγές γνώσης:

- Φωνολογικοί φορείς (phonology) → τρόπος προφοράς – διαφορές από μεμονωμένες λέξεις.
- Προσοδικοί φορείς (prosodics) → σημεία στίξης.
- Σημαντικοί φορείς (semantics) → σχέση μεταξύ λέξεων με την εργασία που θα εκτελεστεί.
- Πραγματικοί φορείς (pragmatics) → υπόδειγμα διαλόγου με τη μηχανή σε συγκεκριμένο περιβάλλον.

Αρχιτεκτονικές για Κατανόηση Ομιλίας (1/2)

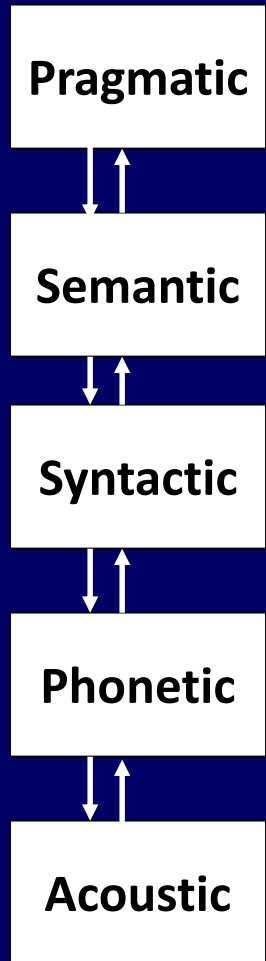


Hierarchical (a)

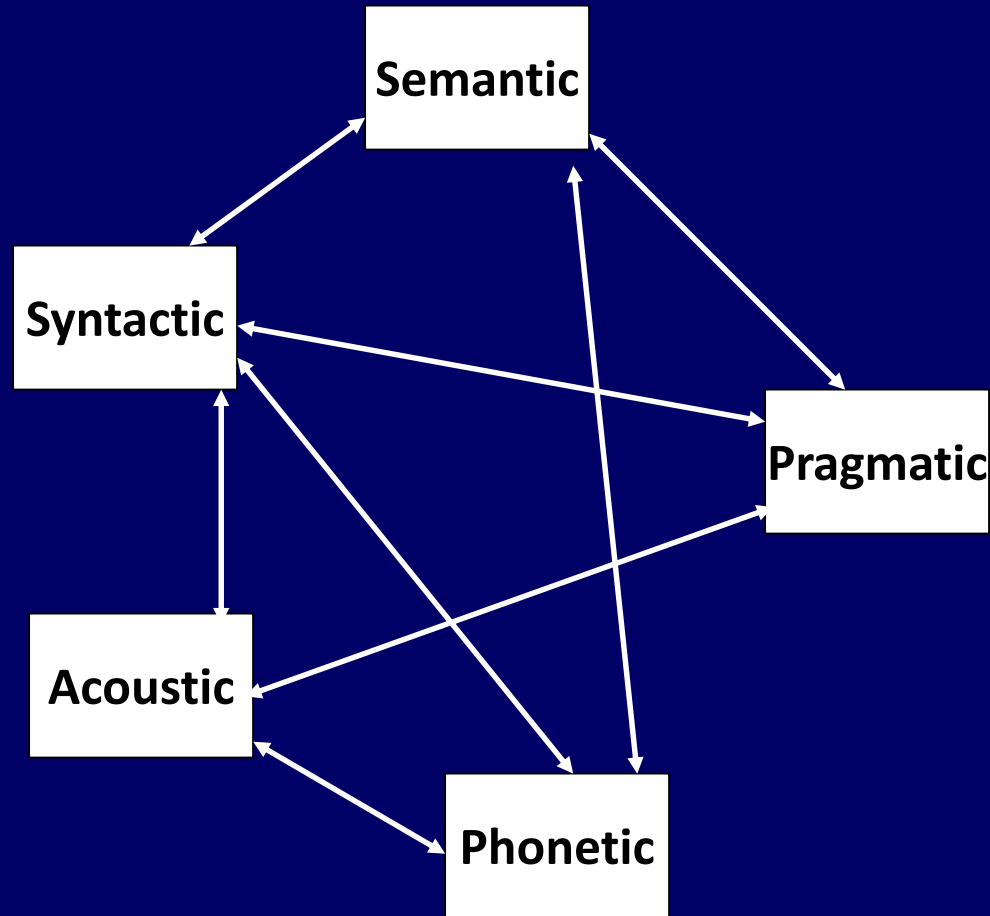


Blackboard (b)

Αρχιτεκτονικές για Κατανόηση Ομιλίας (2/2)

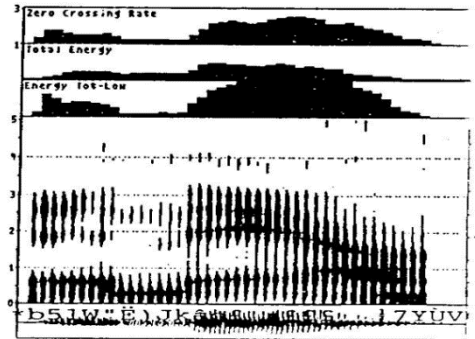
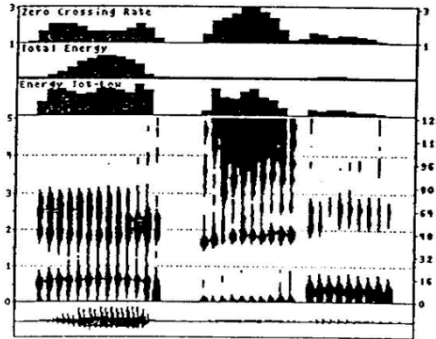
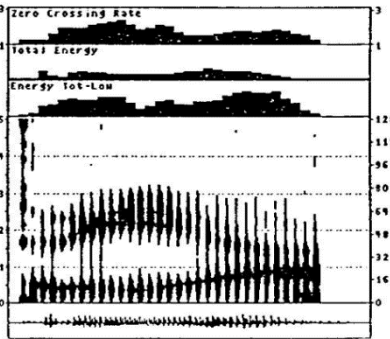
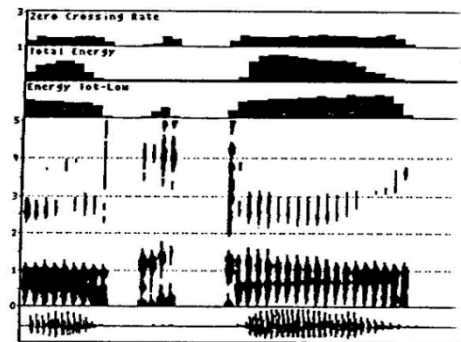
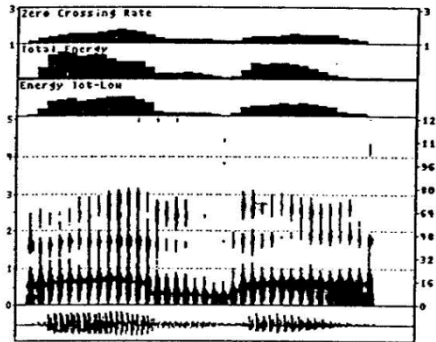
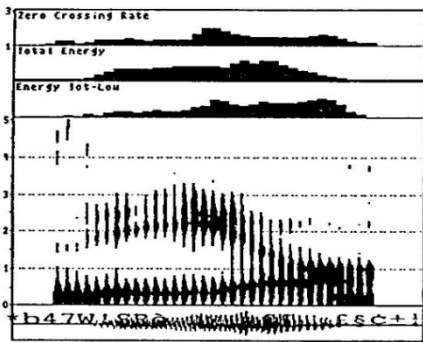
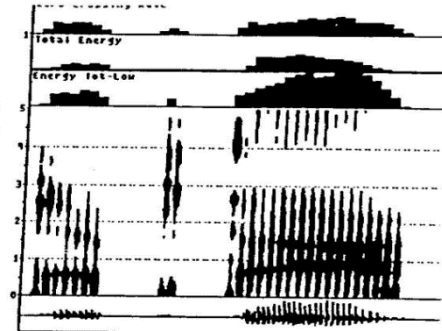
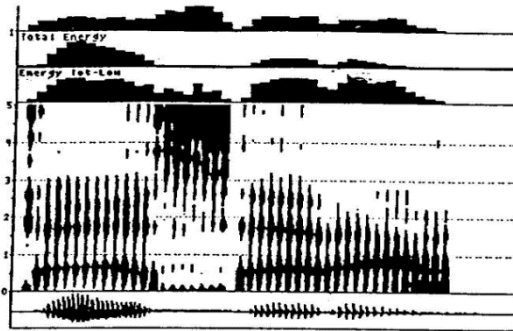
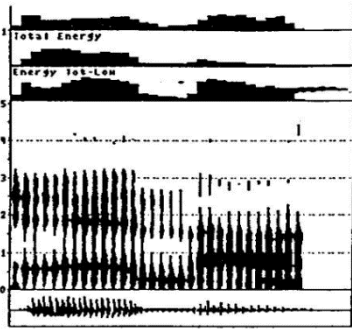


Generative (c)



Heterarchical (d)





AAO βασισμένη σε ακουστική γνώση (knowledge based)



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, Γεώργιος Κουρουπέτρογλου 2015. «Επεξεργασία ομιλίας και φυσικής γλώσσας. Αναγνώριση ομιλίας.». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/DI36/>.

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

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