



Εθνικόν και Καποδιστριακόν  
Πανεπιστήμιον Αθηνών

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

# Επεξεργασία Ομιλίας και Φυσικής Γλώσσας

Ενότητα 2: Βασικά χαρακτηριστικά της ομιλίας

Γεώργιος Κουρουπέτρογλου

[koupe@di.uoa.gr](mailto:koupe@di.uoa.gr)



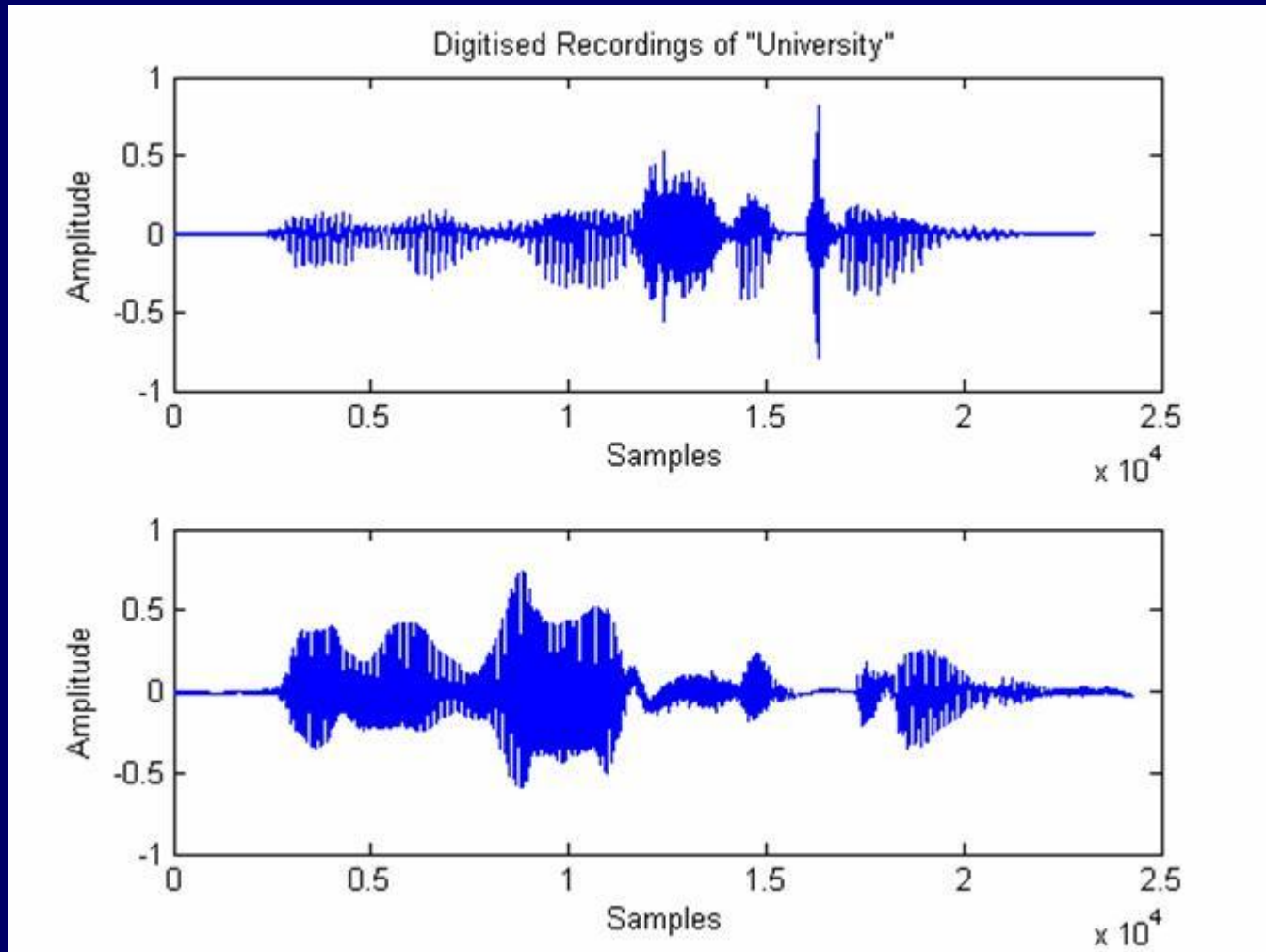
# Βασικά Χαρακτηριστικά της Ομιλίας (1/3)

## ΓΛΩΣΣΟΛΟΓΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ

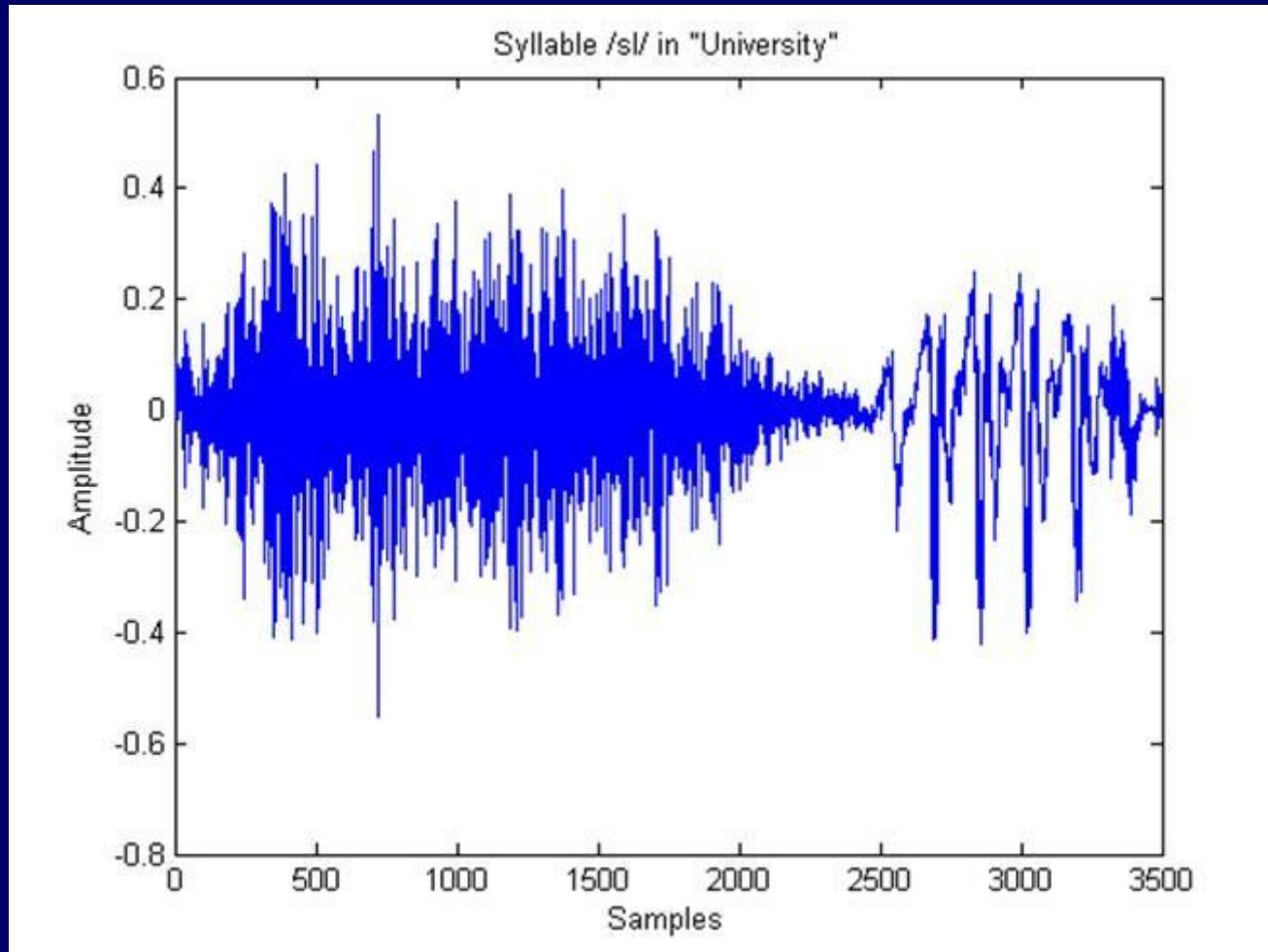
- ✓ κύμα (σήμα) ομιλίας: μεταφέρει διάφορα είδη πληροφοριών: κυριότερη η γλωσσολογική.
- ✓ πρόταση: λέξεις, λέξη: συλλαβές, συλλαβή: φωνήματα (φωνήεντα ή σύμφωνα).
- ✓ αριθμός φωνημάτων ποικίλει από γλώσσα σε γλώσσα (σπάνια >50).
- ✓ Αγγλική: 12 φωνήεντα και 24 σύμφωνα, Ιαπωνική: 5 φωνήεντα και 20 σύμφωνα.
- ✓ αριθμός συλλαβών: ένα μέρος όλων των πιθανών συνδυασμών.
- ✓ ο αριθμός των λέξεων σε κάθε γλώσσα είναι πολύ μεγάλος (<από όλους τους δυνατούς συνδυασμούς συλλαβών ή φωνημάτων).



# Σήμα της λέξης university από άνδρα και γυναίκα ομιλητή

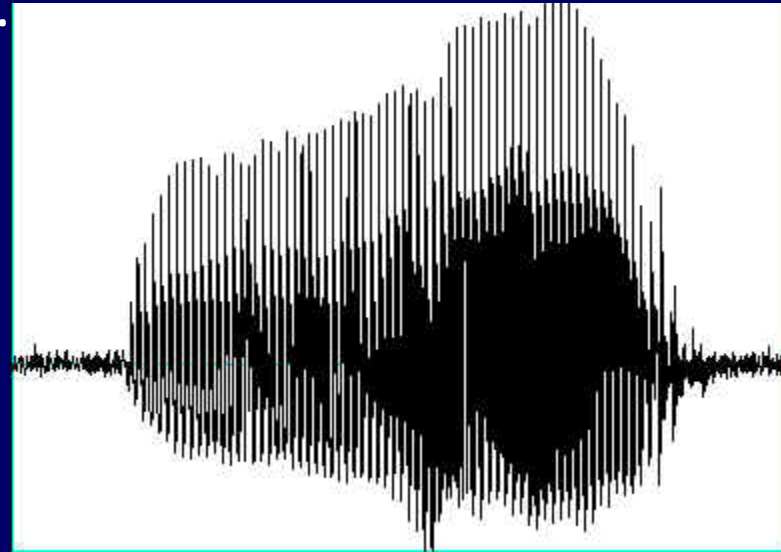


# Σήμα της συλλαβής /si/ στην εκφώνηση: “university”



# Πληροφορίες που μεταφέρει το σήμα της ομιλίας

- Γλωσσολογική πληροφορία: το νόημα.
- Γλώσσα (διάλεκτος) που μιλιέται.
- Ομιλία παιδιού, άνδρα, γυναίκας, ηλικιωμένου.
- Ποιος μιλά: ταυτότητα ομιλητή.
- Συναισθηματική κατάσταση ομιλητή.
- Αν είναι μητρική η γλώσσα του ομιλητή.
- ....



# Βασικά Χαρακτηριστικά της Ομιλίας (2/3)

## ΓΛΩΣΣΟΛΟΓΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ

- ✓ συνεχώς προστίθενται καινούριες λέξεις.
- ✓ αριθμός λέξεων που χρησιμοποιούνται συχνά: 2000 – 3000.
- ✓ αριθμός λέξεων που χρησιμοποιεί μέσος άνθρωπος: 5000 – 10000.
- ✓ ο τονισμός των συλλαβών και των λέξεων μέσα στην πρόταση παίζει ουσιαστικό ρόλο:
  - ✓ στο να υποδείξουμε τη θέση σημαντικών λέξεων,
  - ✓ στο να κάνουμε ερωτηματικές προτάσεις και
  - ✓ στο να μεταφέρουμε το συναίσθημα του ομιλητή.

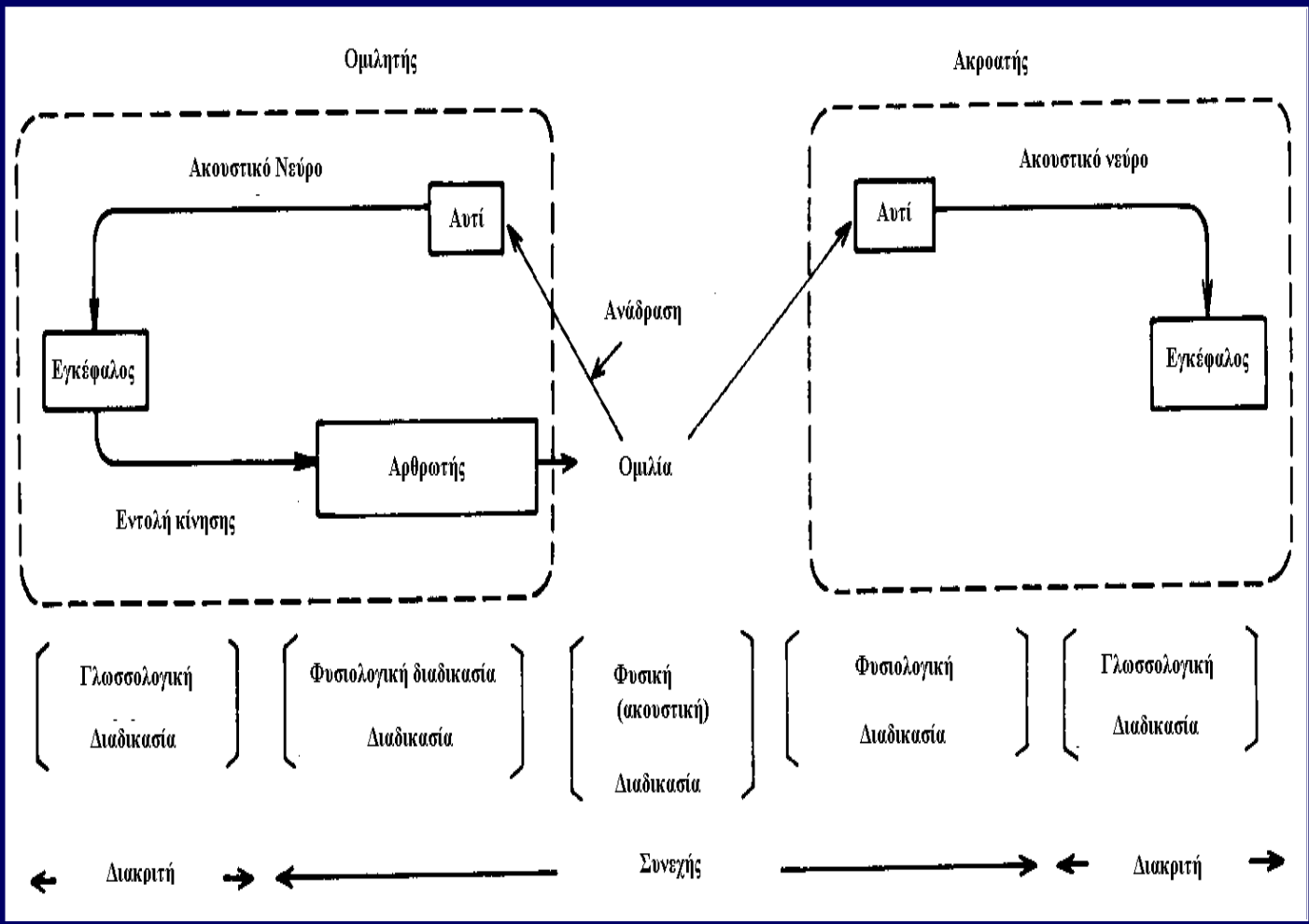


# Βασικά Χαρακτηριστικά της Ομιλίας (3/3)

## ΓΛΩΣΣΟΛΟΓΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ

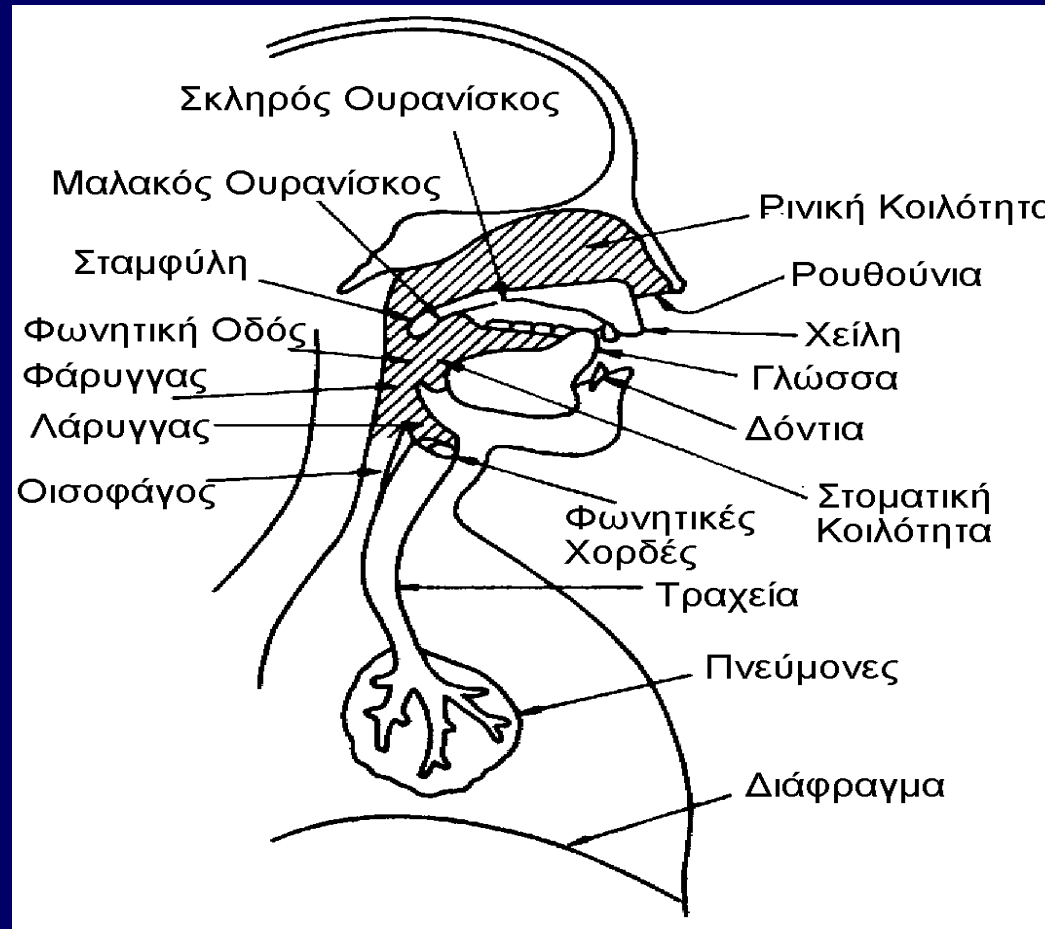
- ✓ Ομιλία: σκοπός μετάδοση νοημάτων, λέξεις ή προτάσεις δεν ενδιαφέρουν από μόνες τους.
- ✓ Νοητική δραστηριότητα και εκμάθηση γλώσσας αναπτύσσονται παράλληλα.
- ✓ Στην πραγματικότητα οι άνθρωποι σχηματίζουν νοήματα σε λέξεις και προτάσεις.
- ✓ Δύο ομιλητές διαφορετικών γλωσσών κωδικοποιούν το ίδιο νόημα με διαφορετικές λέξεις.
- ✓ Δύο ομιλητές της ίδιας γλώσσας ίσως αποδίδουν διαφορετικό νόημα στην ίδια λέξη.





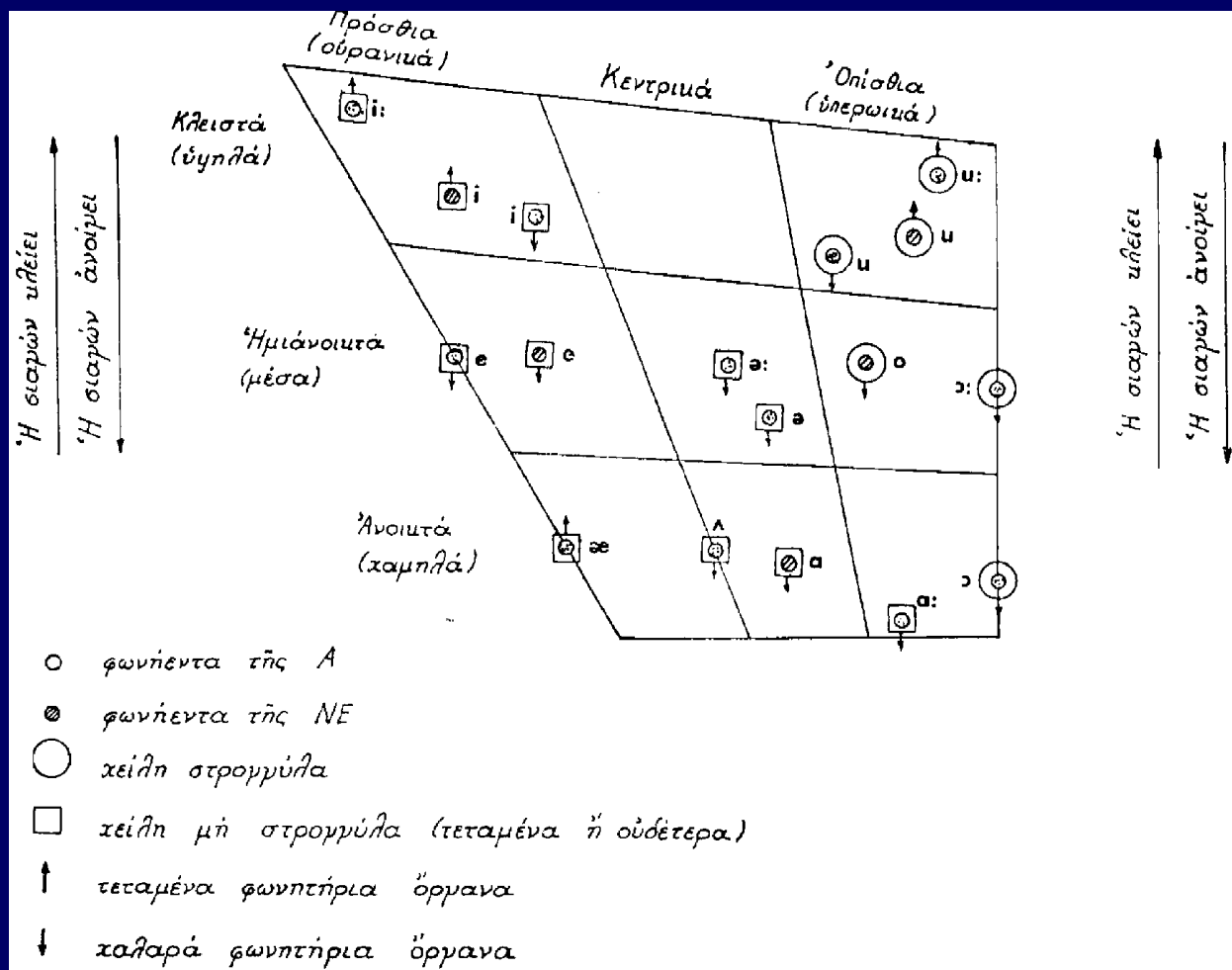


# Μηχανισμός Παραγωγής Ομιλίας

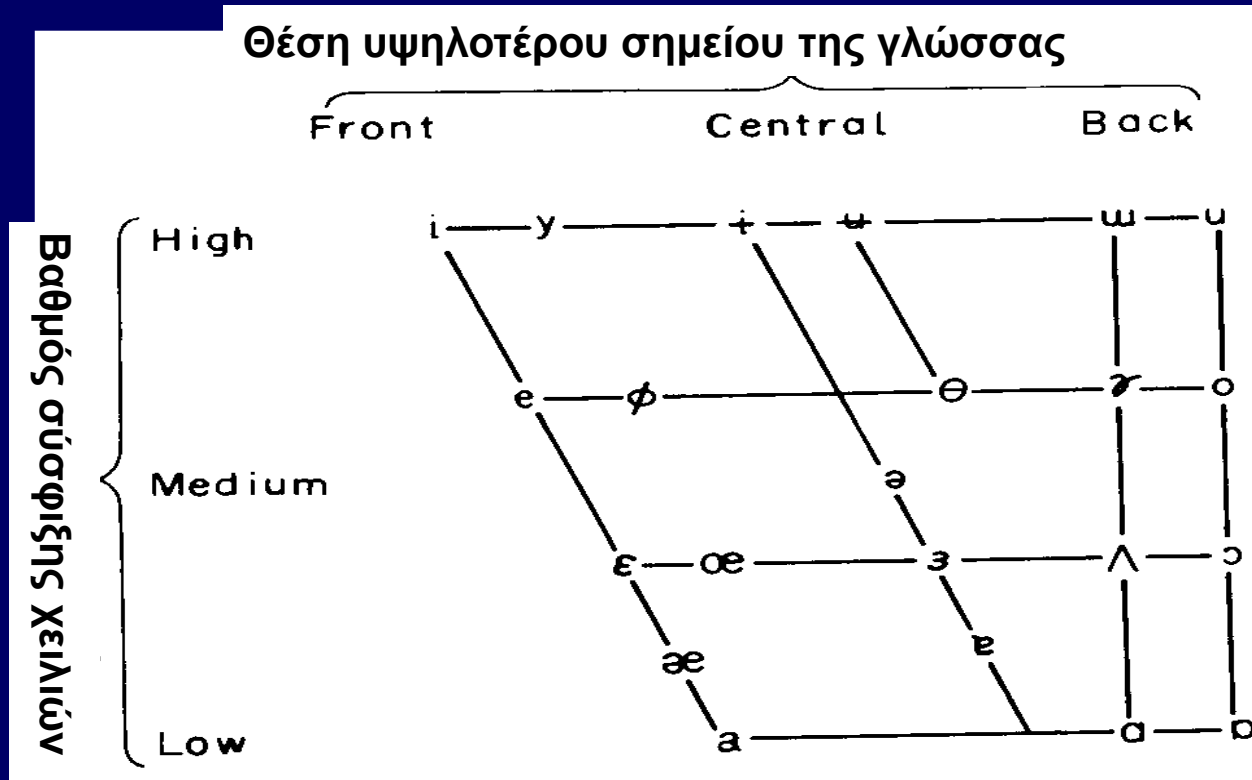


Φωνητικά όργανα

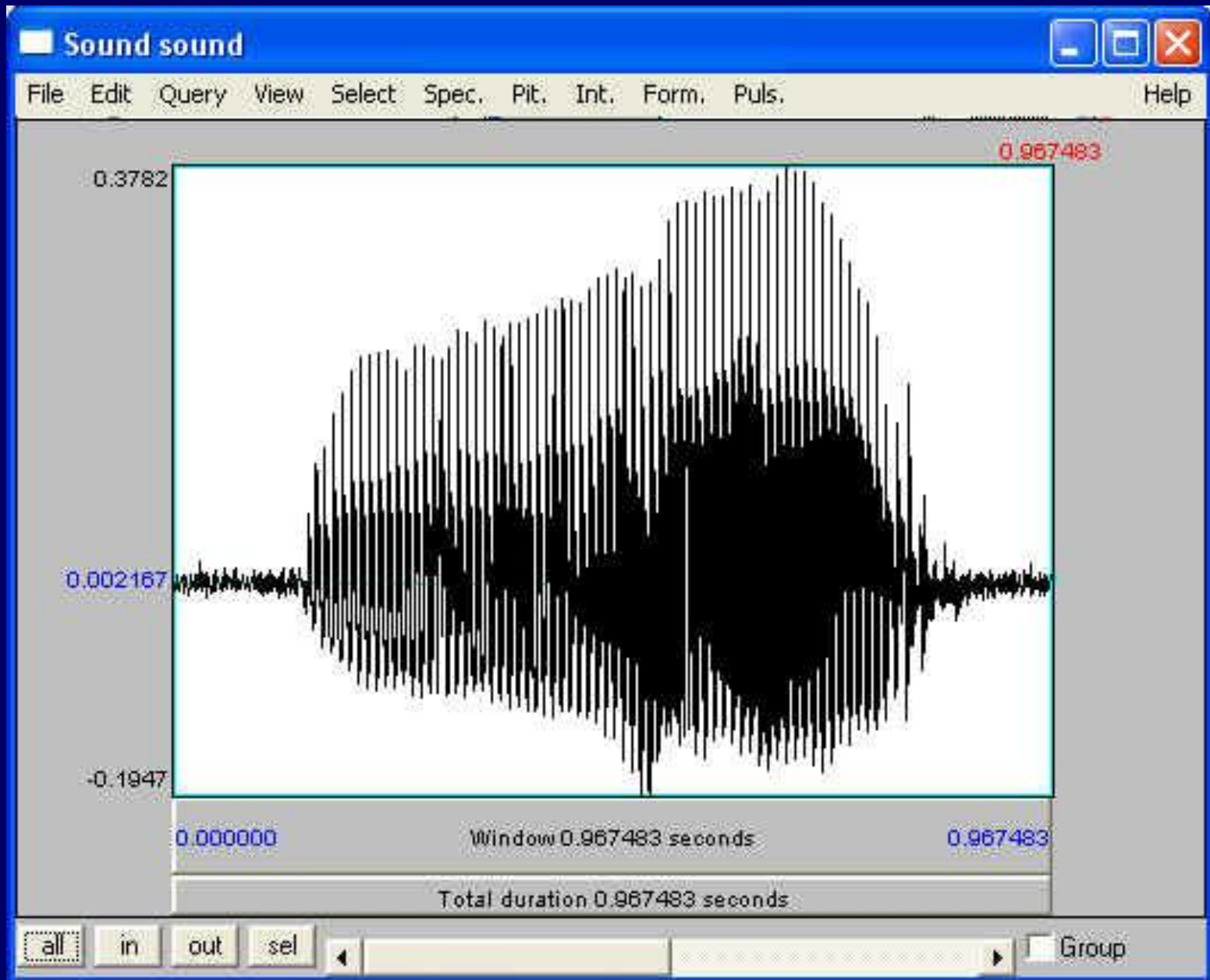
# Κατάταξη των φωνηέντων της Νέας Ελληνικής (ΝΕ) και της Αγγλικής (Α) από την κατά προσέγγιση αναπαράσταση των φωνητικών οργάνων.

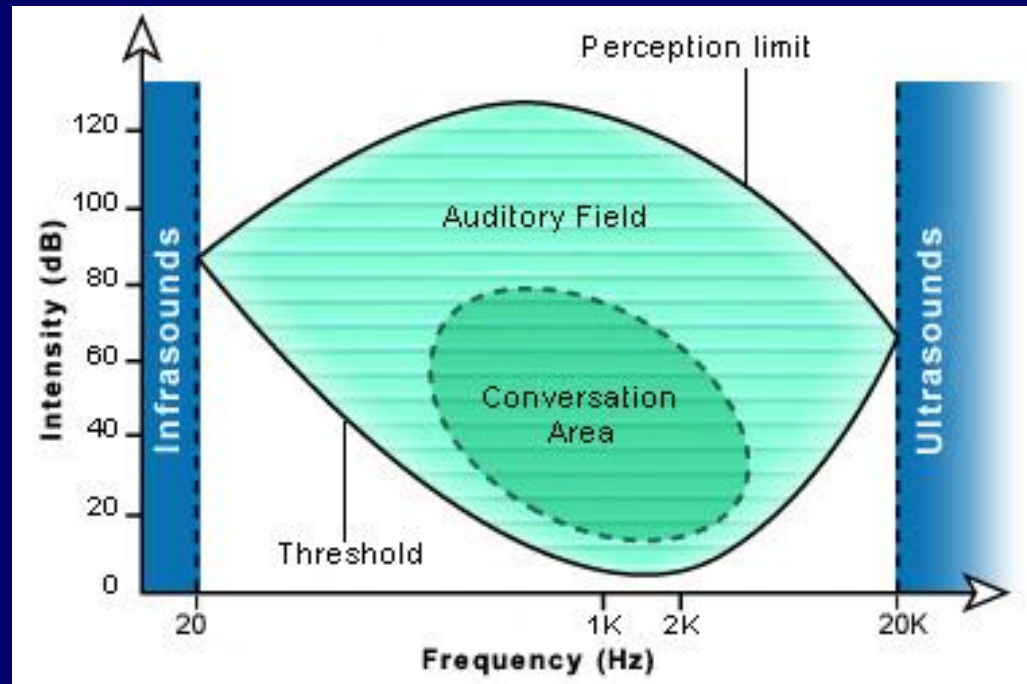


# Κατάταξη των φωνηέντων της Αγγλικής (Α) από την κατά προσέγγιση αναπαράσταση των φωνητικών οργάνων.



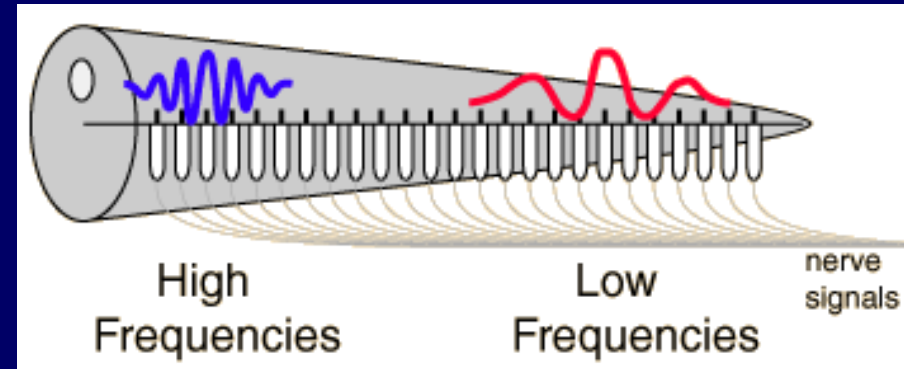
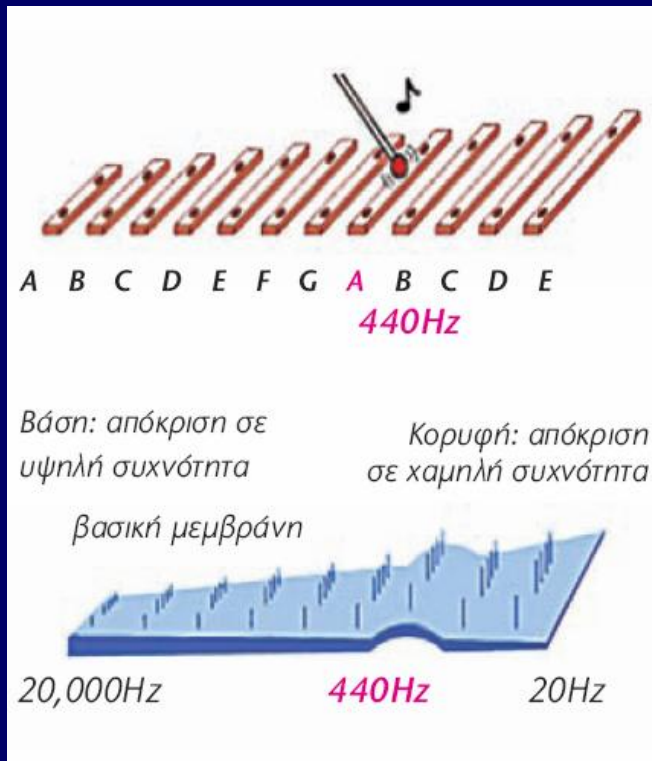
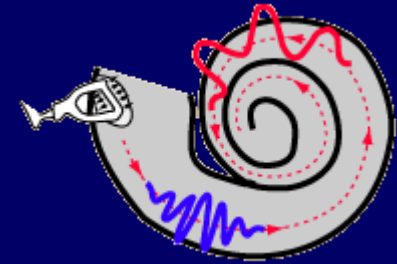
# Χρονικά μεταβαλλόμενο σήμα ομιλίας



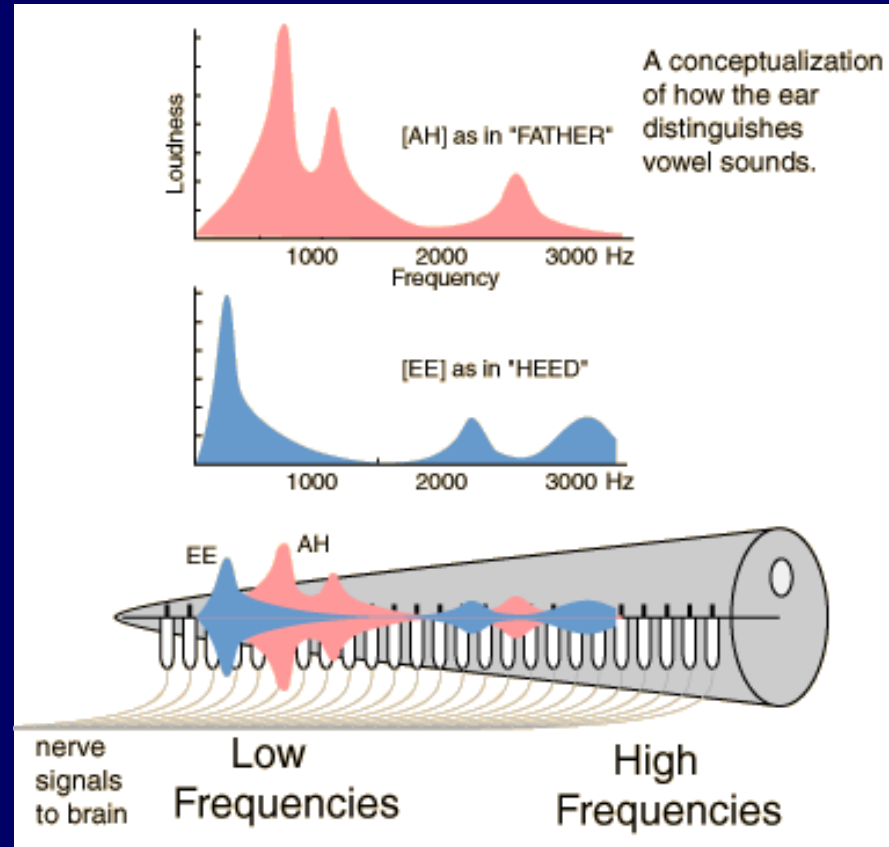
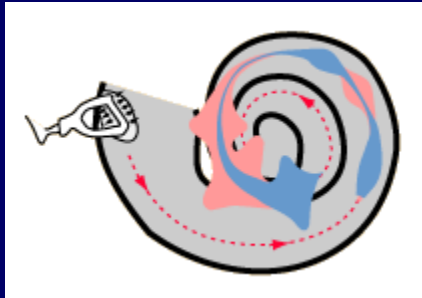


# Αντίληψη ήχων στο αυτί – The Place Theory

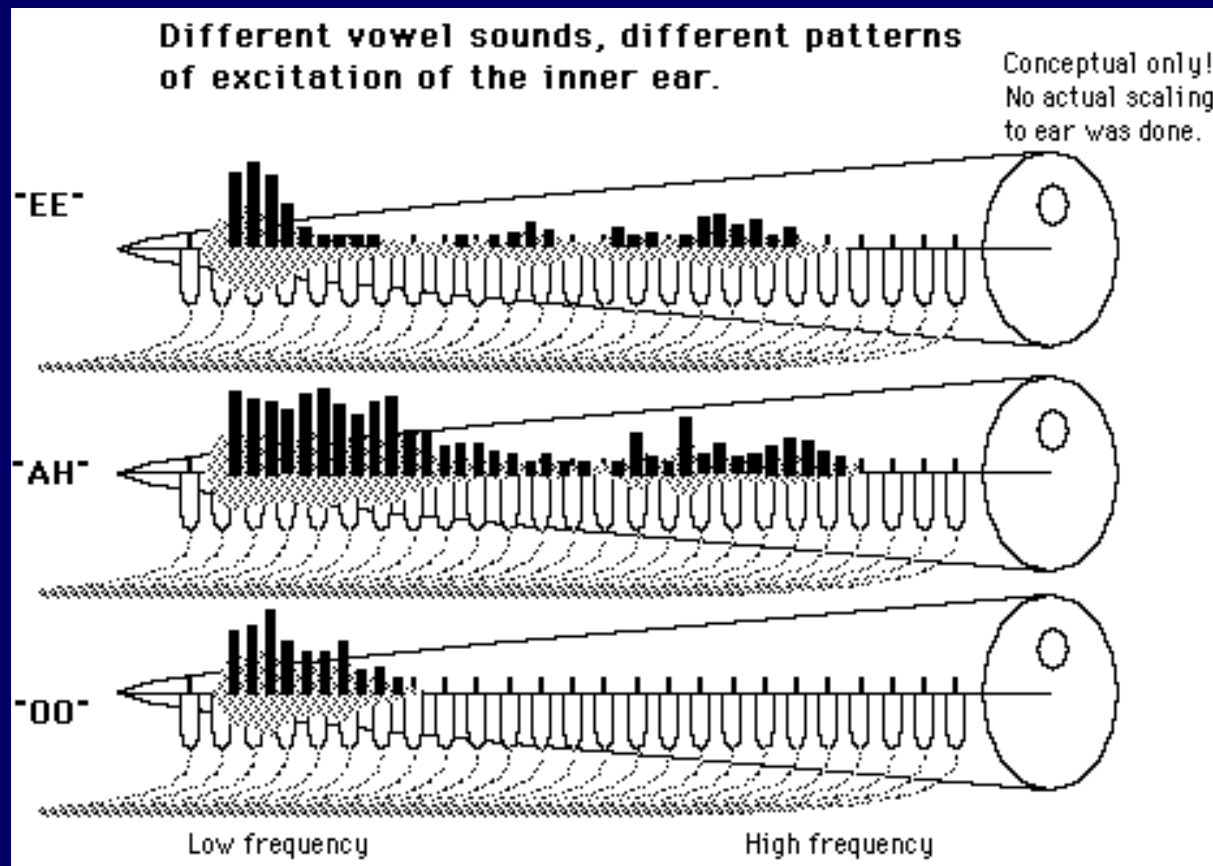
- Τα ηχητικά κύματα διαφορετικών συχνοτήτων διεγείρουν τους υποδοχείς σε διαφορετικές θέσεις στη βασική μεμβράνη. Ο εγκέφαλος αναγνωρίζει τη συχνότητα ενός ήχου από την διάκριση της θέσης των τριχωτών κυττάρων (στο όργανο του Corti) που αποστέλλουν το αντίστοιχο νευρωνικό σήμα.



# Πως το αυτί διακρίνει τους ήχους των φωνηέντων- Place theory



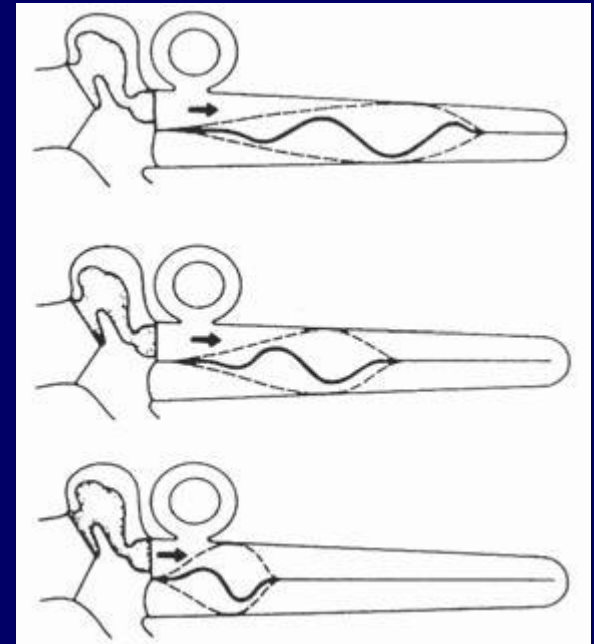
# Το αυτί σαν φασματικός αναλυτής- Place theory





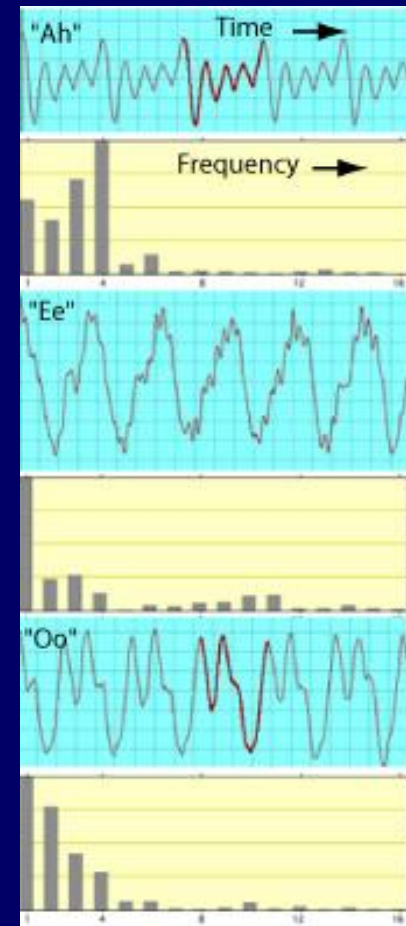
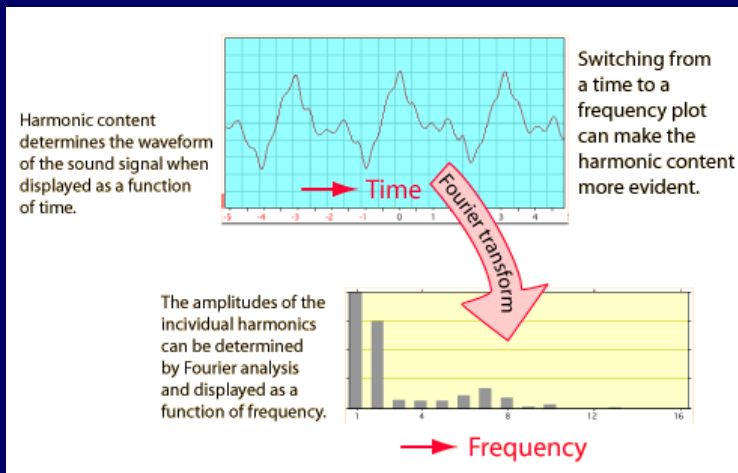
# Αντίληψη ήχων στο αυτί – The Frequency theory

Τα ηχητικά κύματα διαφορετικών συχνοτήτων αναγκάζουν τη βασική μεμβράνη να πάλλεται με διαφορετικούς ρυθμούς και έτσι να παράγει νευρωνικές διεγέρσεις να αποστέλλονται με διαφορετικούς ρυθμούς. Η συχνότητα καθορίζεται από το πόσο γρήγορά μετακινούνται τα νευρωνικά σήματα προς τον εγκέφαλο.

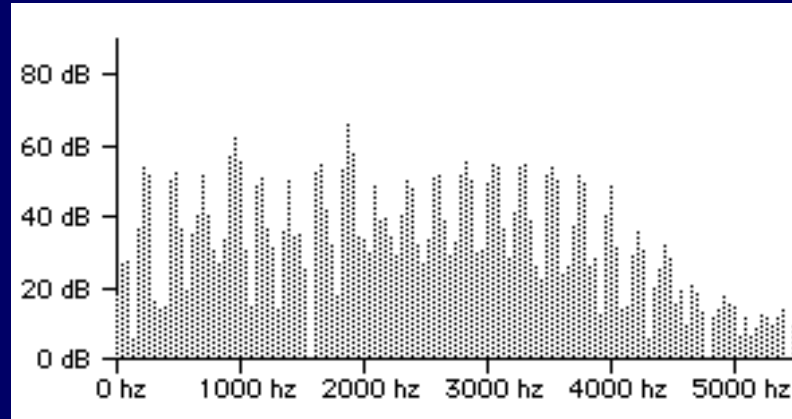


Α. Ψηφίδα «Η κλασική και η σύγχρονη αντίληψη για τη λειτουργία του κοχλίου», Ωτορινολαρυγγολογία, σελ. 42-49, 2006

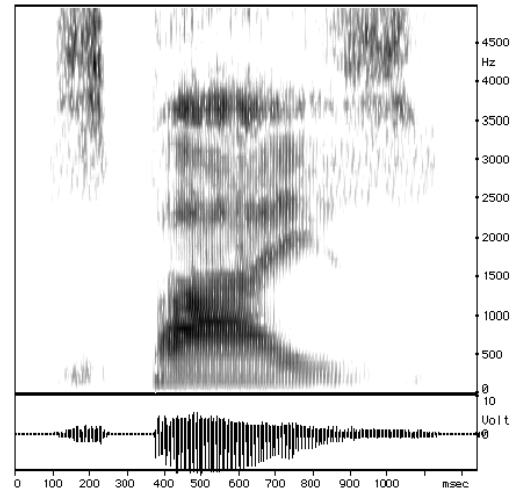
# Σήματα φωνηέντων και αντίστοιχα φάσματα



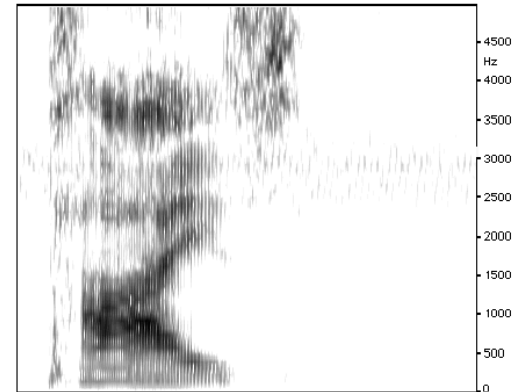
# Φάσμα σήματος ομιλίας (μικρής χρονικής διάρκειας)



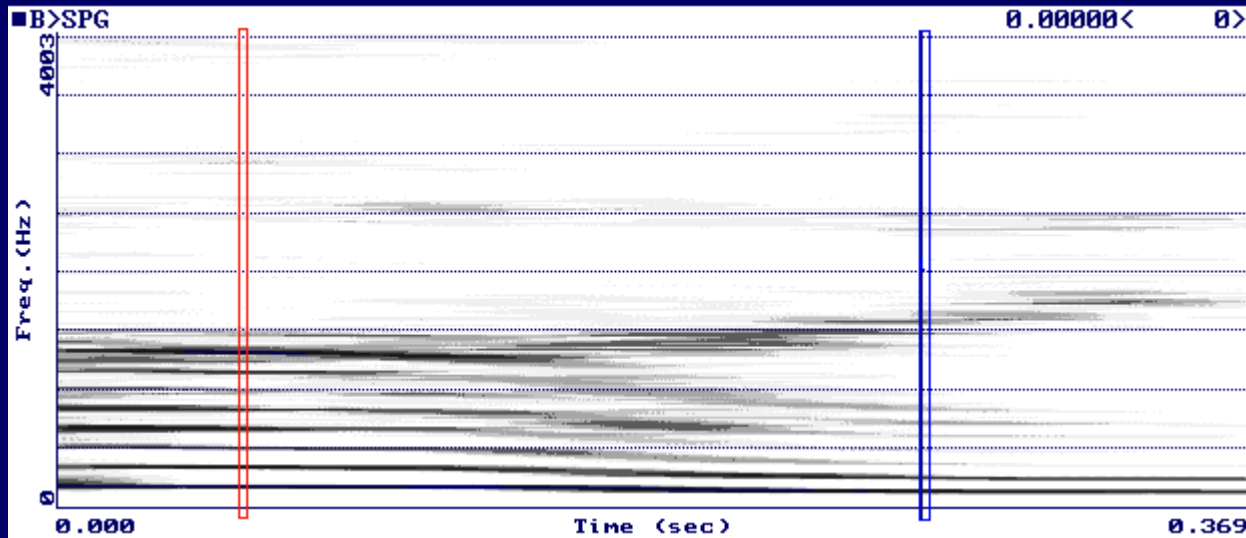
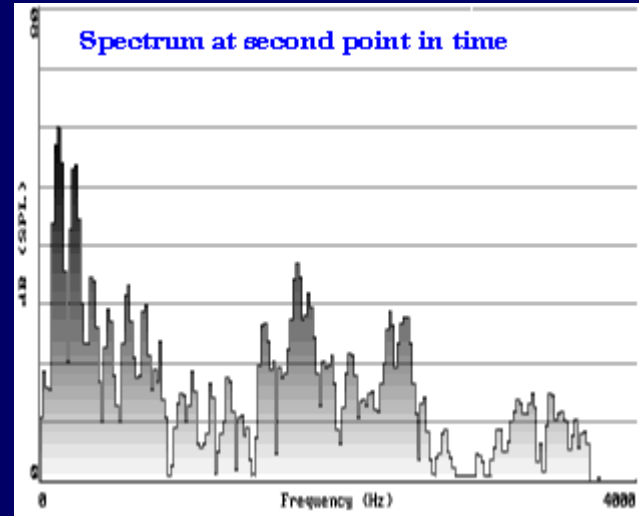
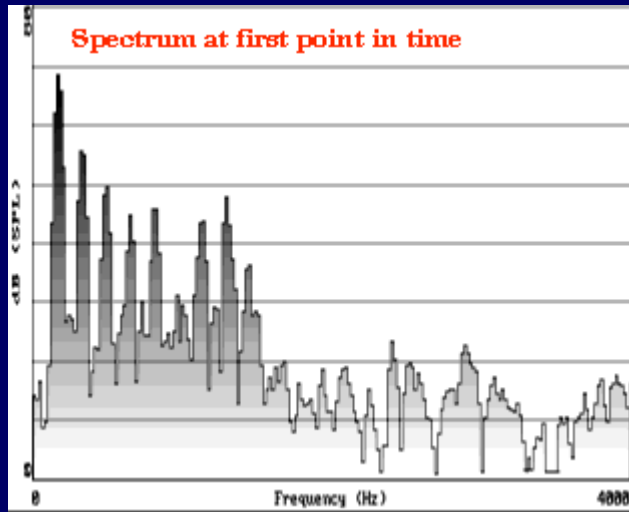
# Φασματογράφημα σήματος ομιλίας



[ s p a j z ]  
'spies'



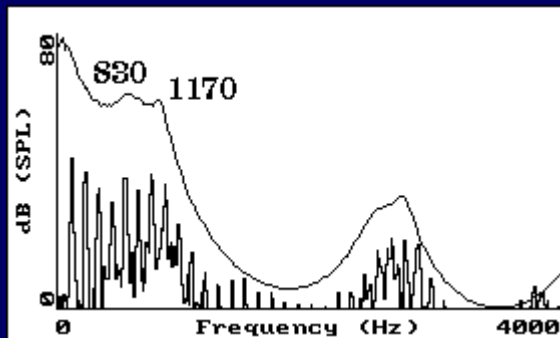
[ p<sup>h</sup> a j z ]  
'pies'



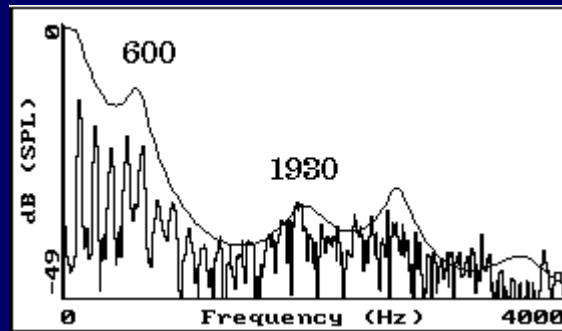
# Φωνοσυντονισμοί (Formants):

(χρονικά μεταβλλόμενα) ενεργειακά μέγιστα στο φάσμα ομιλίας που καθορίζουν την ποιότητα των ήχων (ιδιαίτερα των φωνηέντων) και τα οποία οφείλονται σε ακουστικούς συντονισμούς στη φωνητική οδό του σήματος της πηγής (π.χ. φωνητικών χορδών για τους ηχηρούς ήχους)

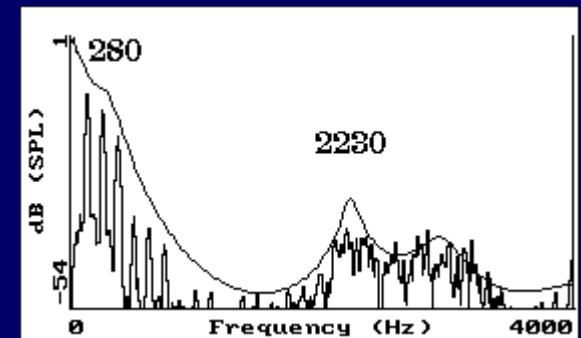
α



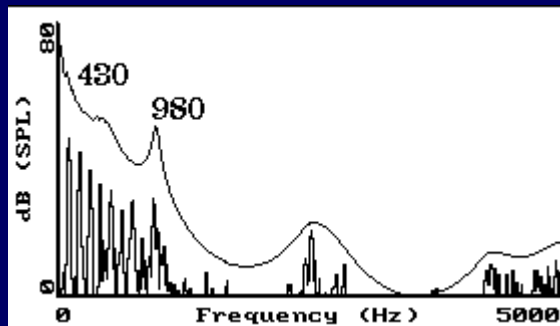
ε



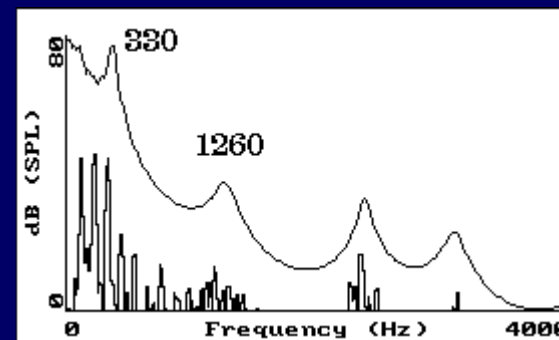
ι



ο

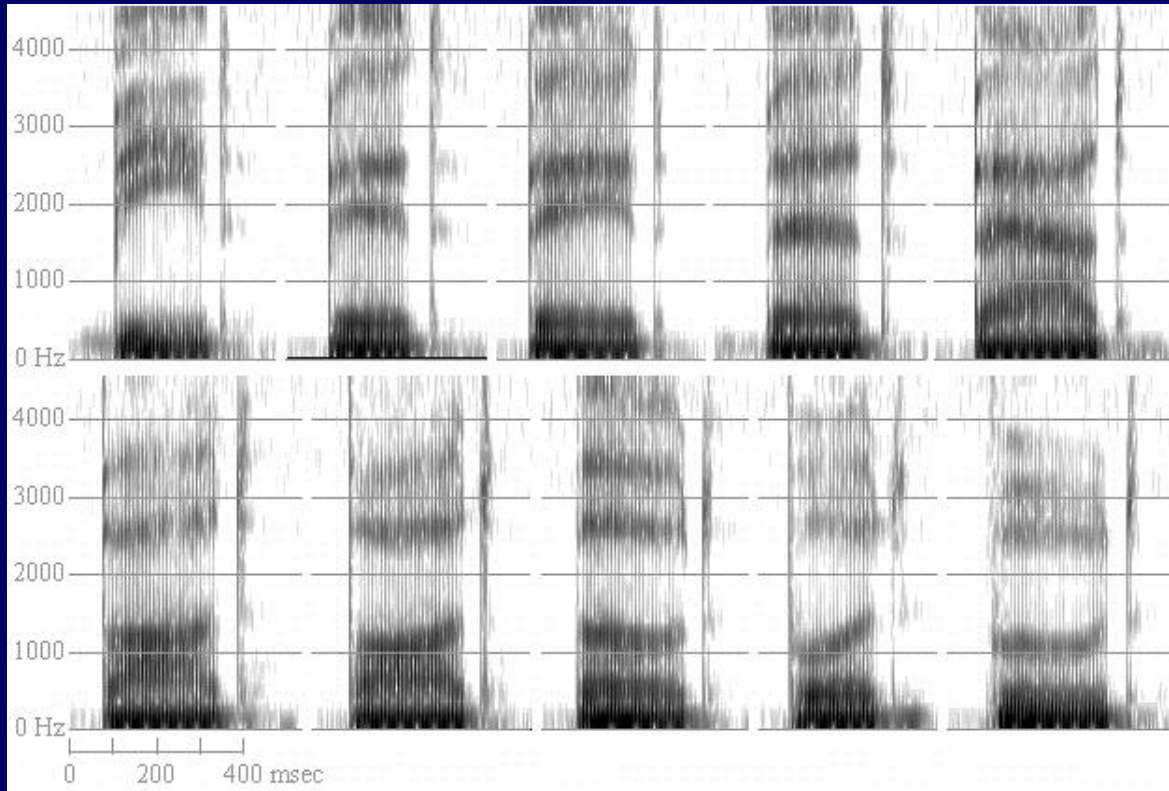


υ



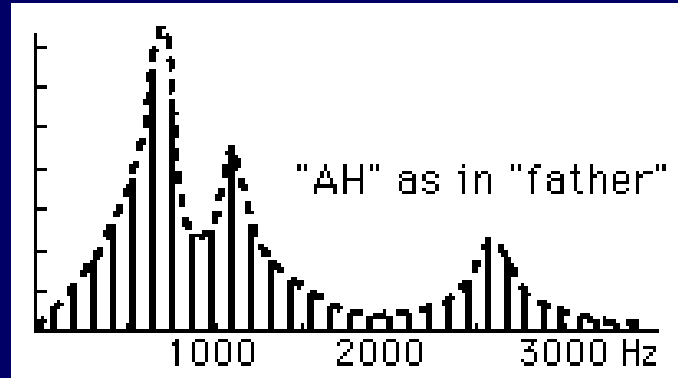
# Φασματογραφήματα των φωνηέντων της Αμερικάνικης Αγγλικής όταν εκφωνούνται σε φωνητικό περιβάλλον /b\_\_d/ .

"bead" "bid" "bade" "bed" "bad"



"bod" "bawd" "bode" "buhd" "booed"

# Φωνοσυντονισμοί (Formants)



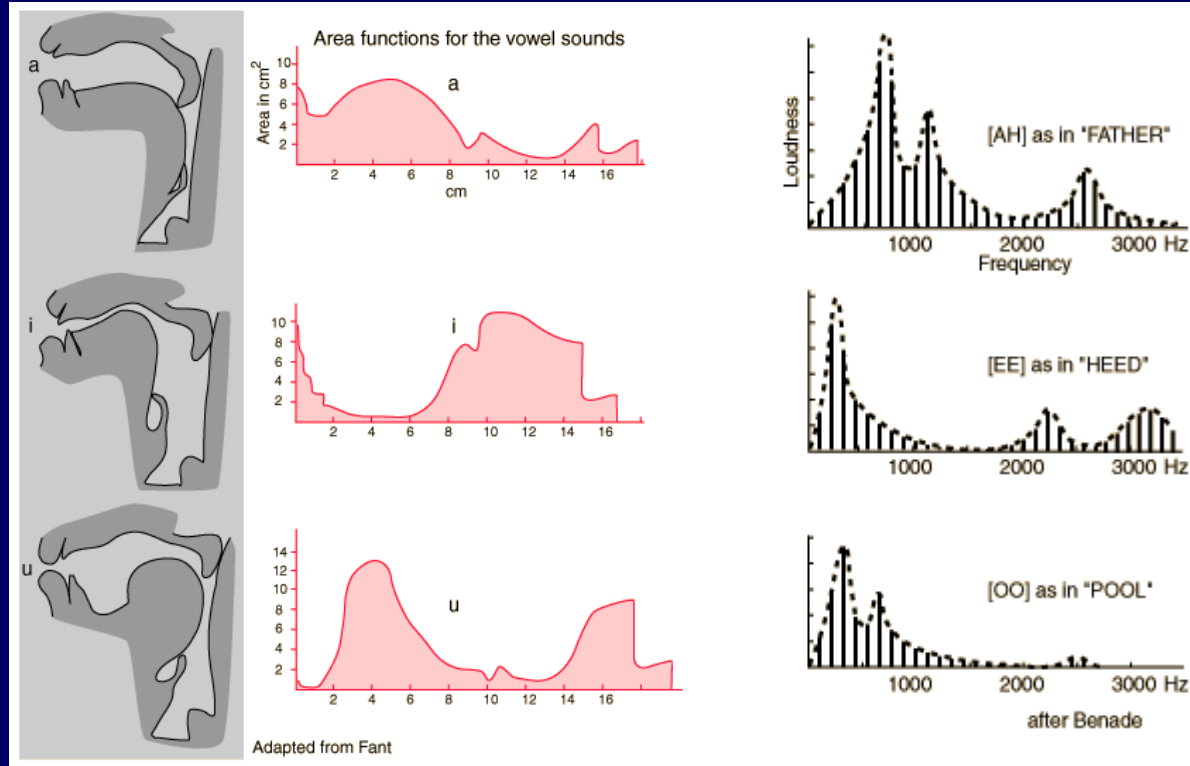
Φωνοσυντονισμοί: Μέγιστα στο φάσμα της ομιλίας τα οποία οφείλονται σε συντονισμούς του σήματος που παράγει το σήμα ομιλίας (σήμα πηγής) οι οποίοι συμβαίνουν στη φωνητική οδό.

Οι φωνοσυντονισμοί τείνουν να μένουν οι ίδιοι (δηλαδή σταθεροί) όταν η συχνότητα της πηγής αλλάζει.

- 1<sup>ο</sup> formant 150-850 Hz
- 2<sup>ο</sup> formant 500-2500 Hz
- 3<sup>ο</sup> formant 1500-3500 Hz
- 4<sup>ο</sup> formant 2500-4800 Hz

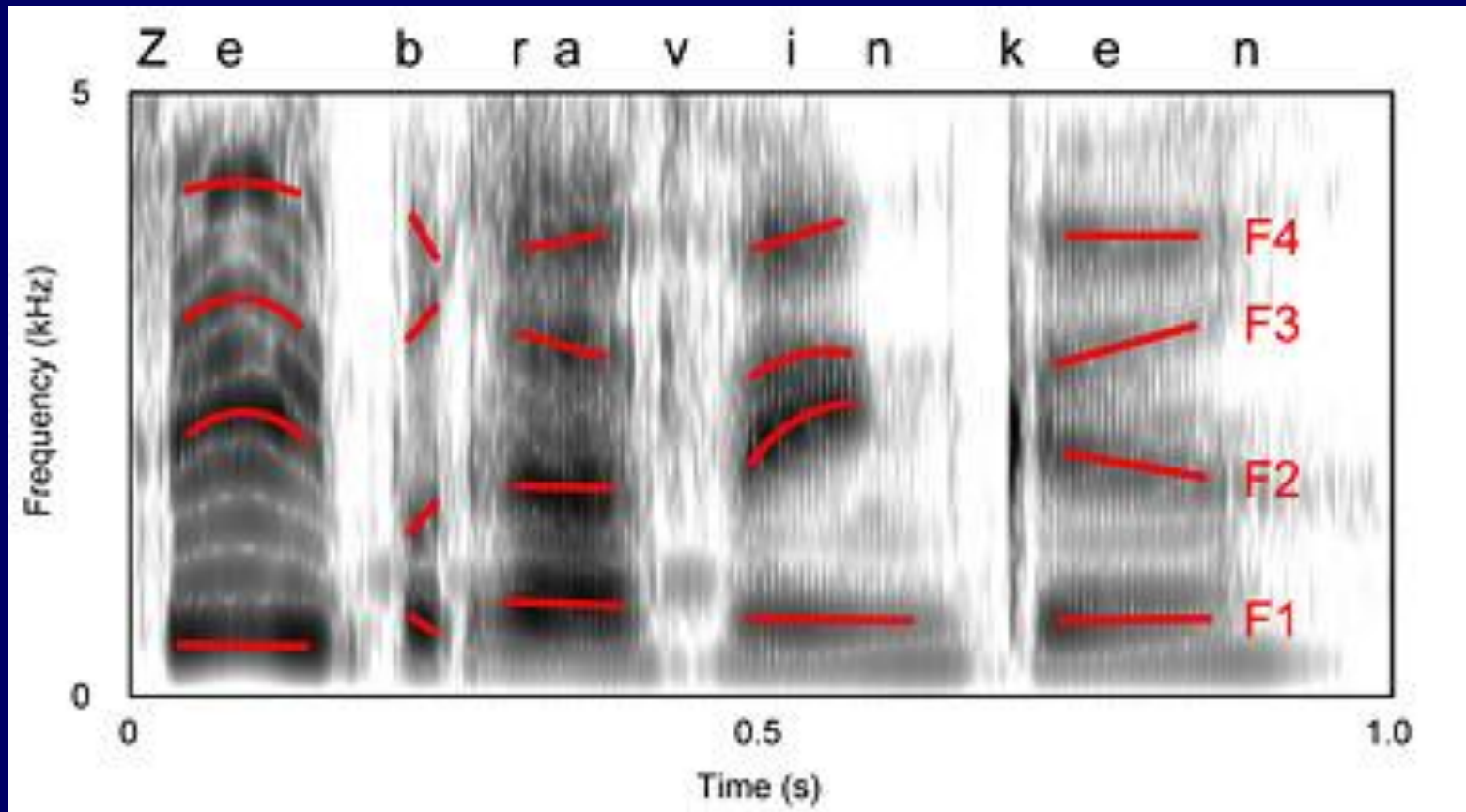


# Σχέσεις Φωνοσυντονισμών

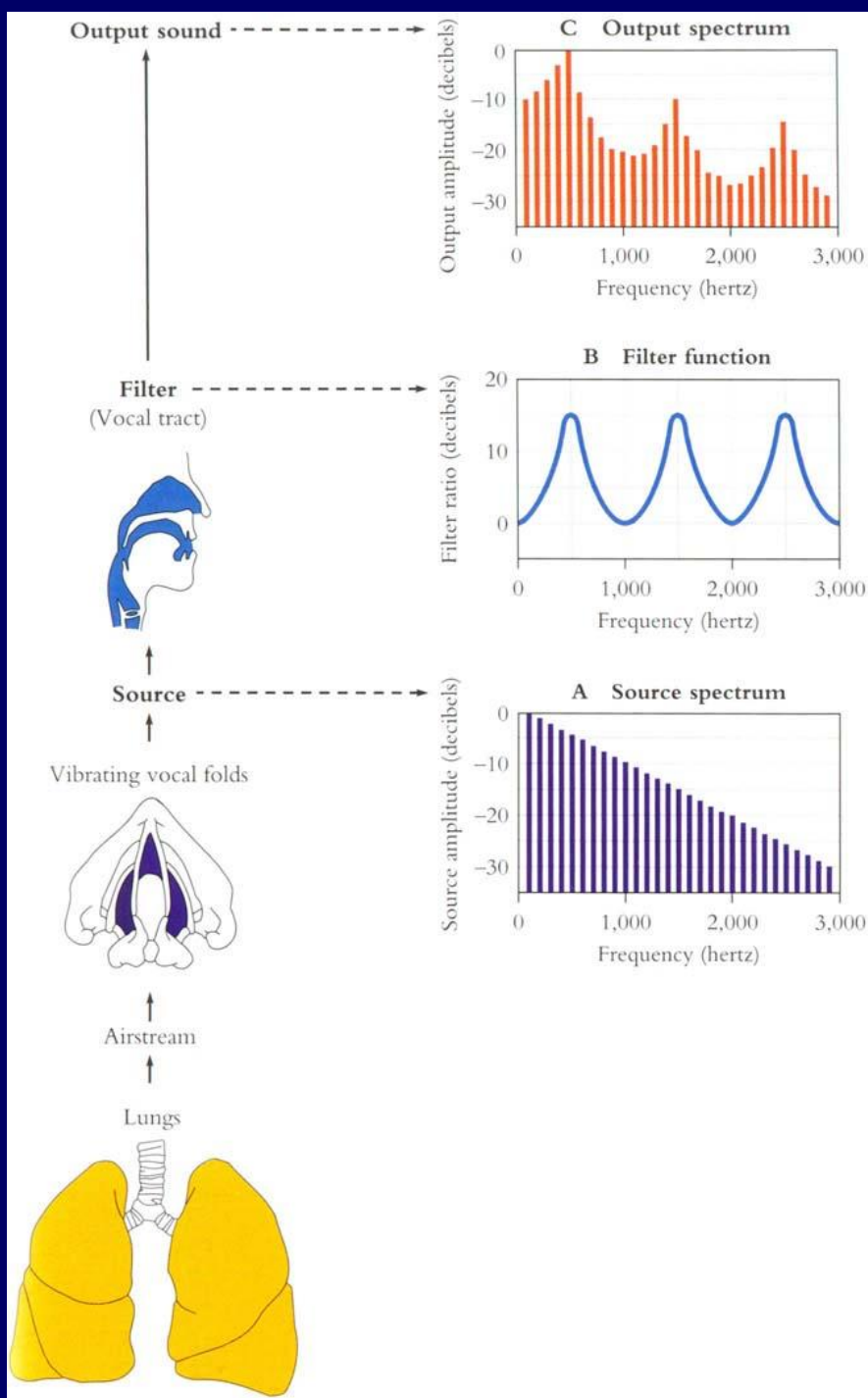


- Θέσεις αρθρωτών κατά την παραγωγή 3 φωνηέντων.
- Αντίστοιχες συναρτήσεις επιφάνειας φωνητικής οδού.
- Αντίστοιχοι φωνοσυντονισμοί στο φάσμα ομιλίας.

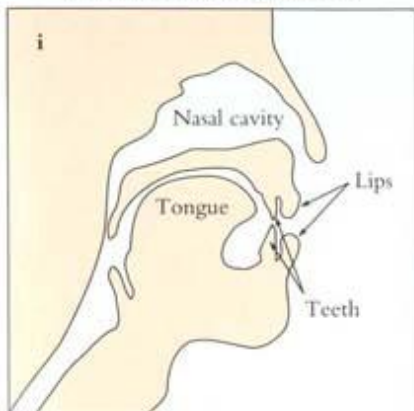
# Τροχιές των formants στο φασματογράφημα



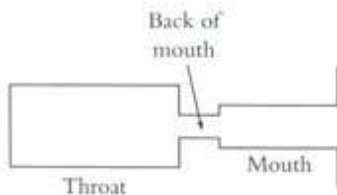
Φασματογράφημα γυναικείας ομιλίας που λέει 'zebra finch'.



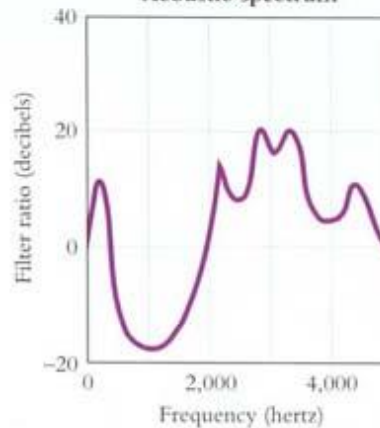
Cross section of vocal tract



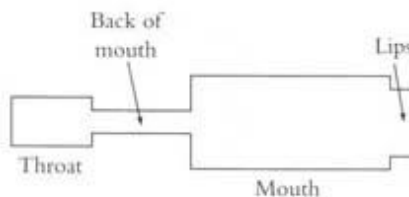
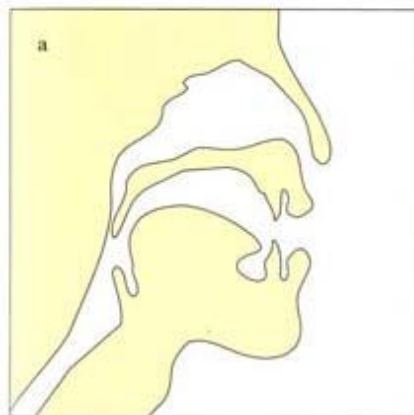
Model of vocal tract



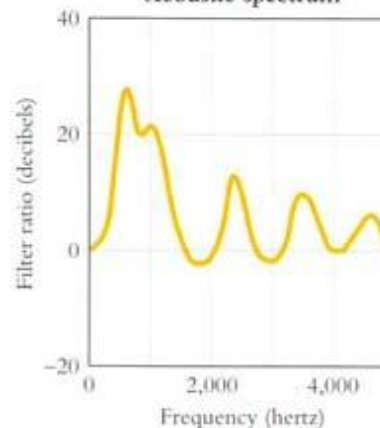
Acoustic spectrum



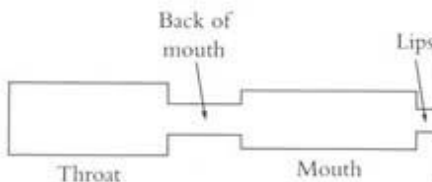
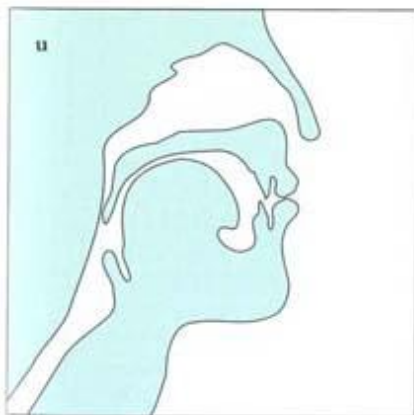
a



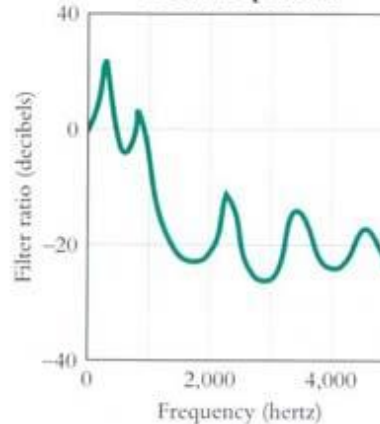
Acoustic spectrum

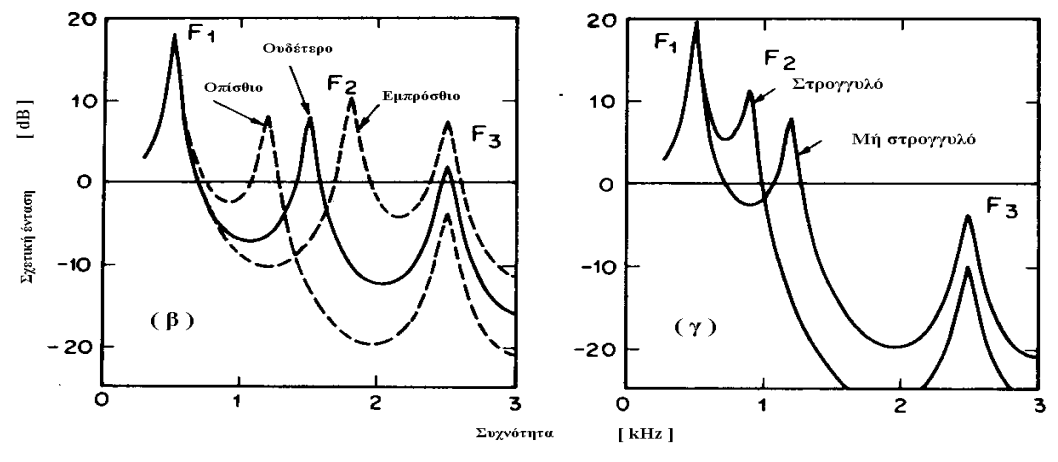
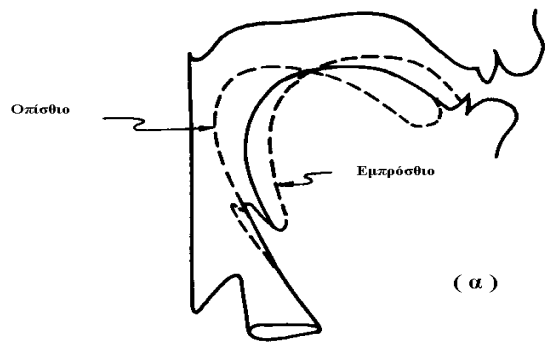


u



Acoustic spectrum





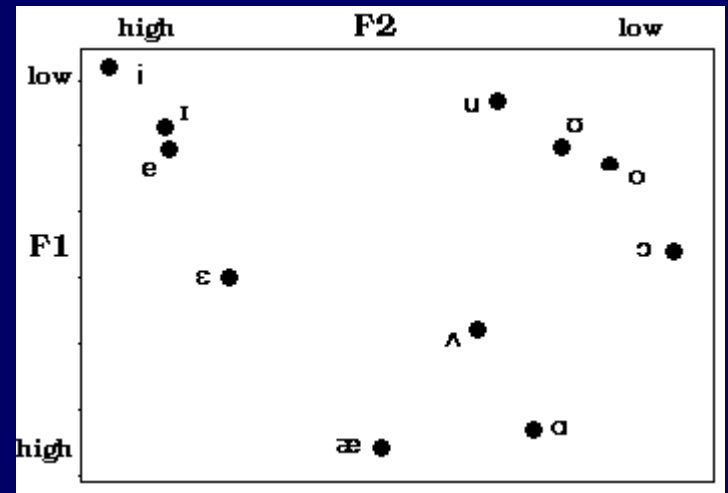
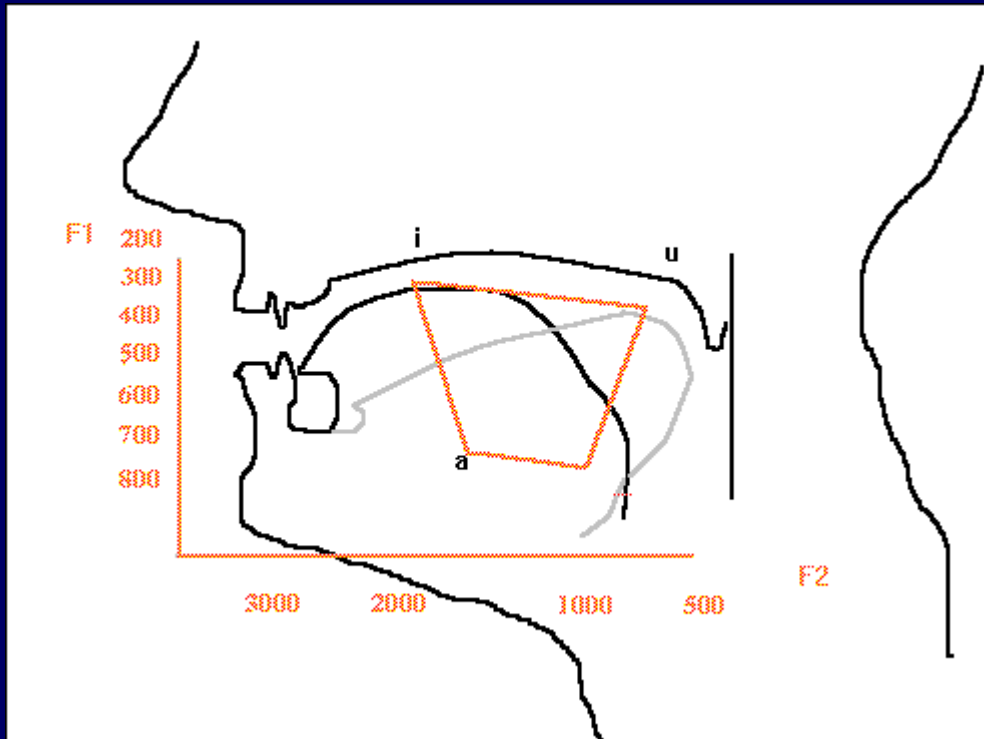
Παραδείγματα της σχέσης ανάμεσα στις μορφές της φωνητικής οδού και τις φασματικές περιβάλλουσες φωνηέντων:

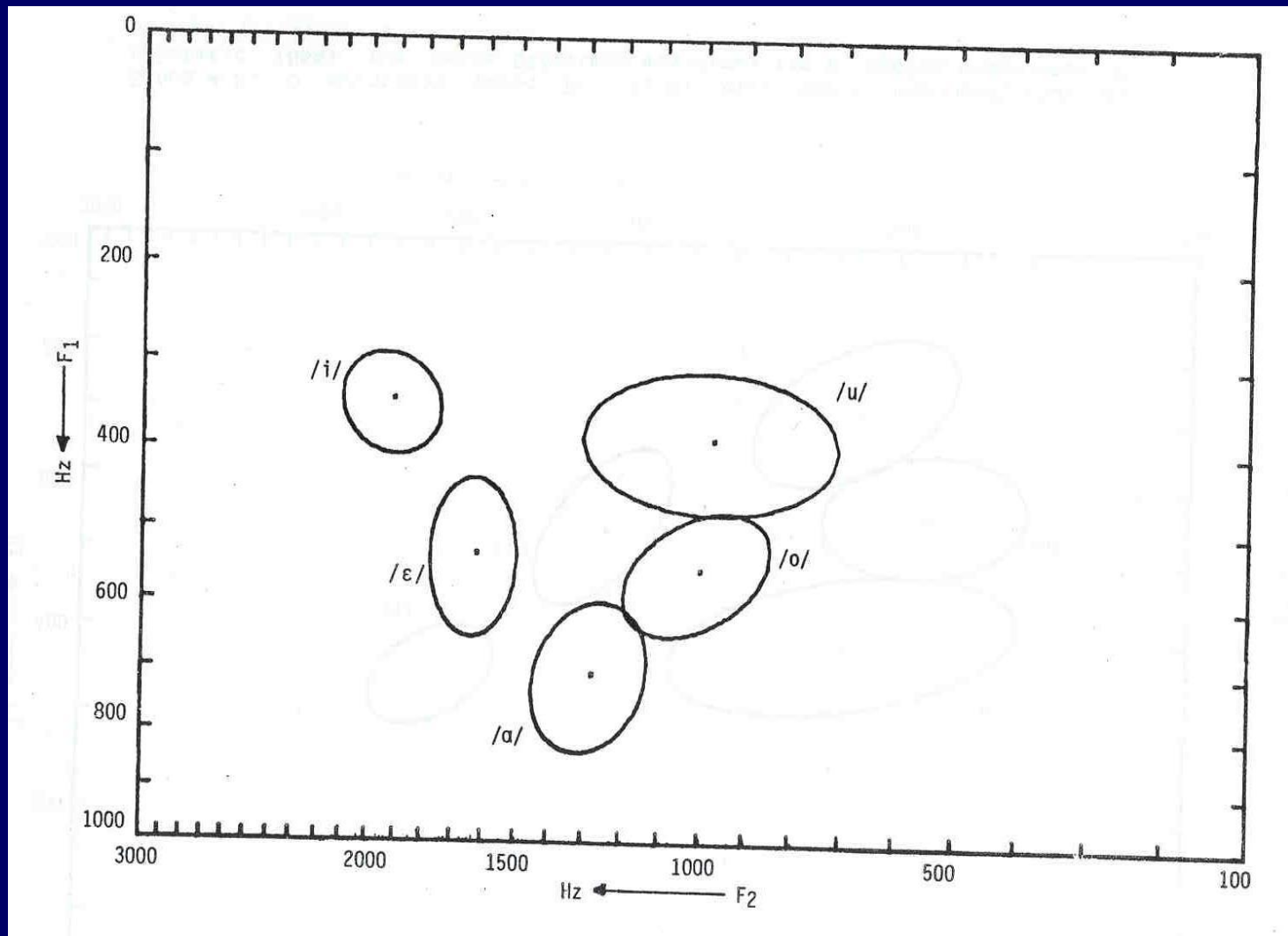
(α) σχηματική αναπαράσταση της τοξοειδούς διατομής της φωνητικής οδού για ένα ουδέτερο φωνήεν (συνεχόμενη καμπύλη και για οπίσθιες και εμπρόσθιες θέσεις του σώματος της γλώσσας)

(β) ιδανικές φασματικές περιβάλλουσες οι οποίες αντιστοιχούν στις τρεις διαμορφώσεις της γλώσσας στο (α),

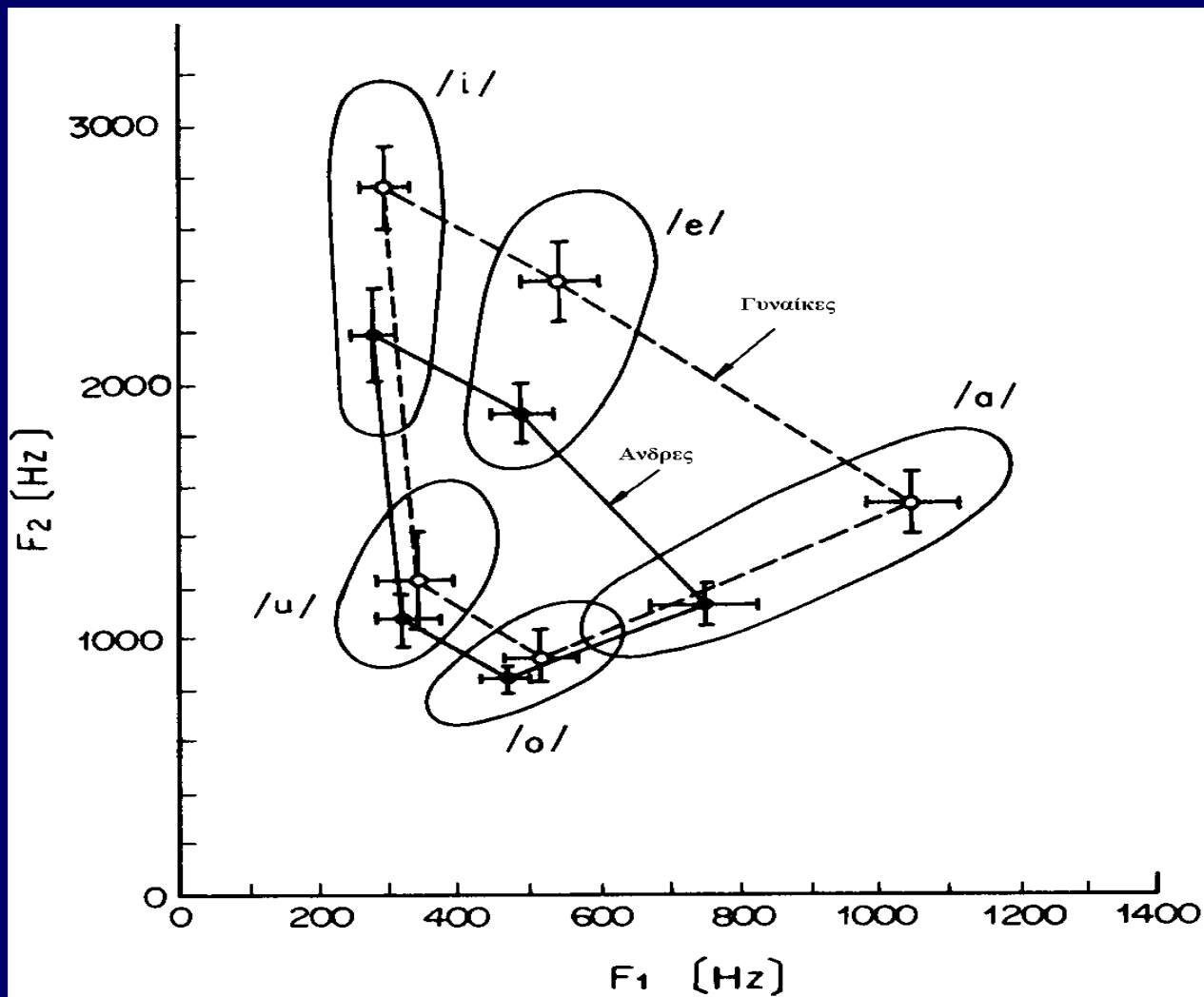
(γ) μέση επίδραση του στρογγυλέματος των χειλιών για ένα οπίσθιο φωνήεν.





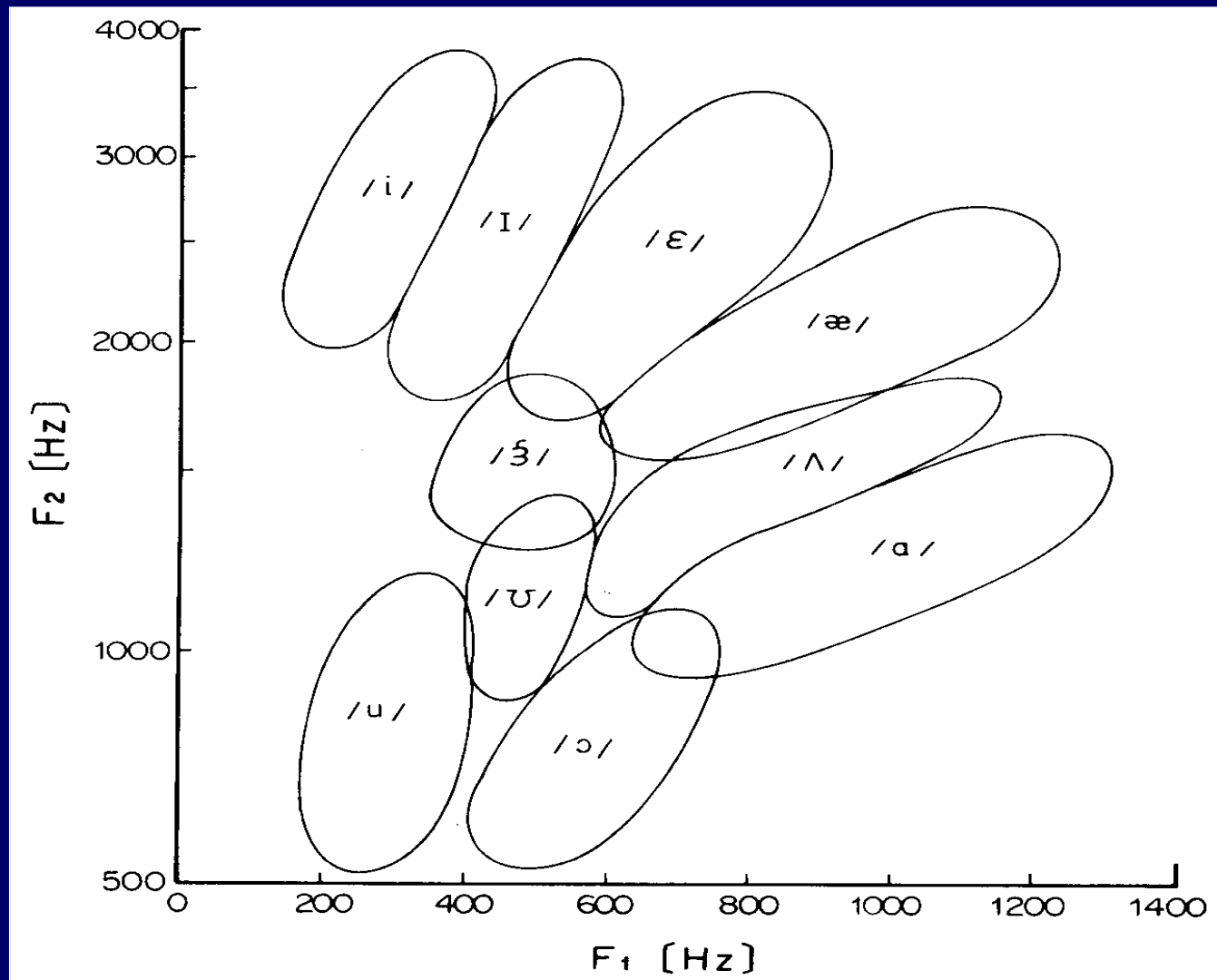


**Διάγραμμα διασποράς των συχνοτήτων των δύο πρώτων formants για τα πέντε φωνήεντα της Ελληνικής όπως εκφωνήθηκαν από πέντε άνδρες ομιλητές σε διάφορα φωνητικά περιβάλλοντα.**

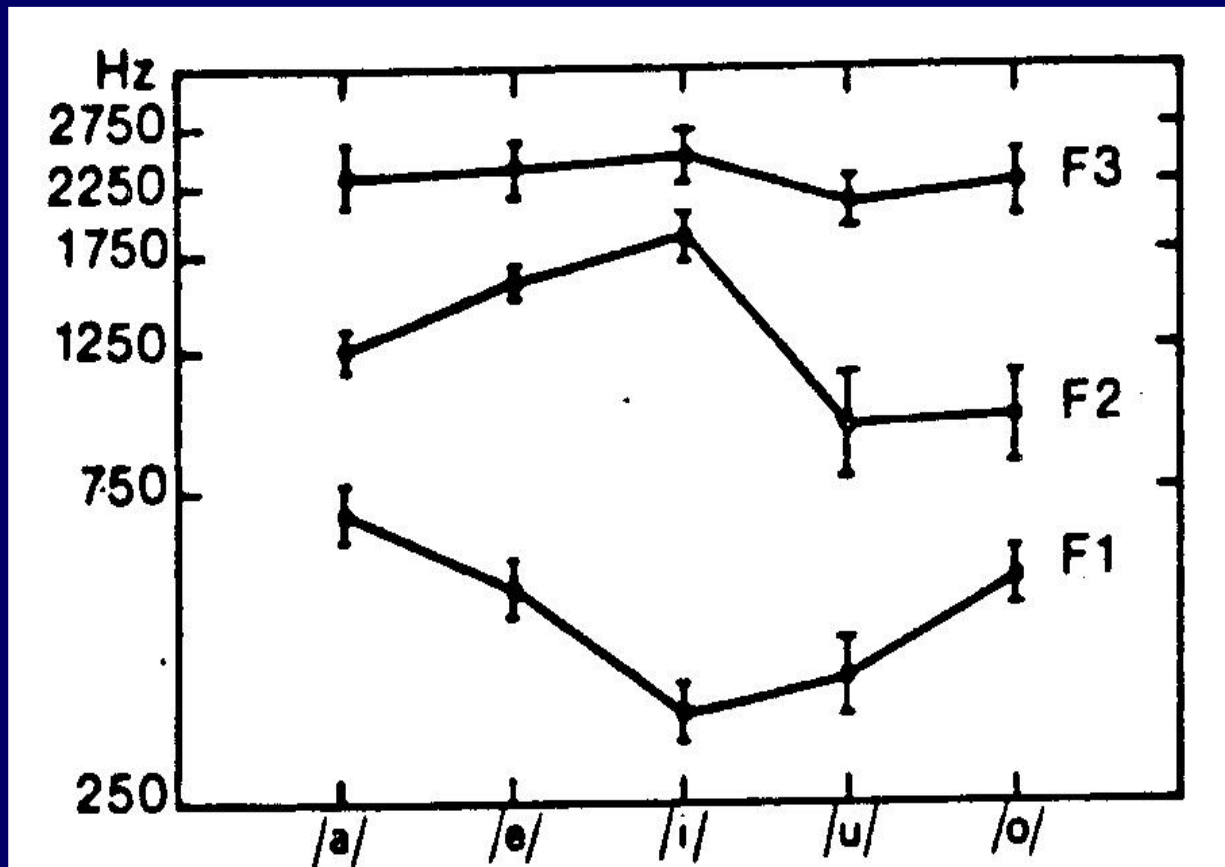


**Διάγραμμα διασποράς των συχνοτήτων formant για πέντε ιαπωνικά φωνήεντα τα οποία εκφωνούνται από 60 ομιλητές (30 άνδρες και 30 γυναίκες) στο επίπεδο F1-F2.**

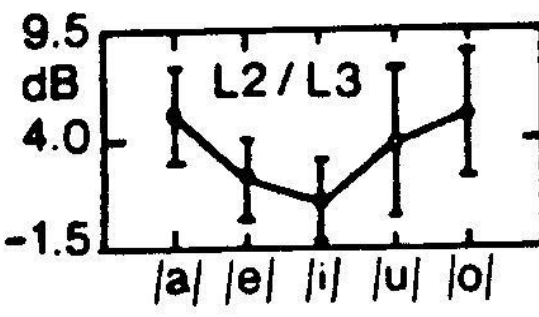
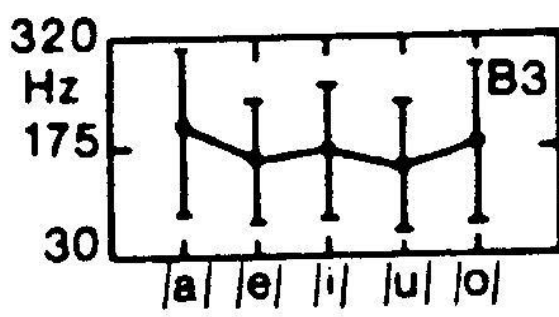
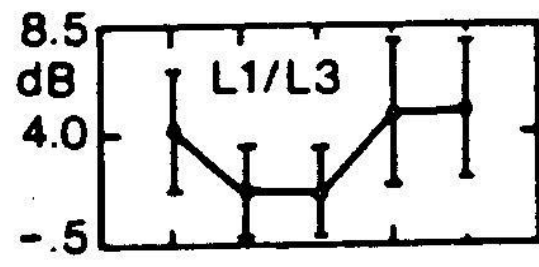
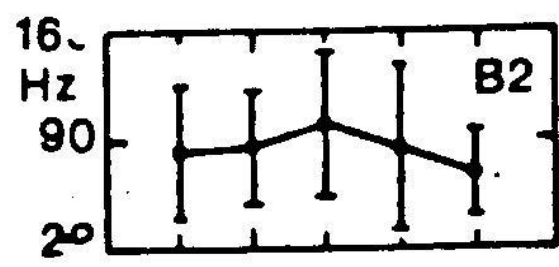
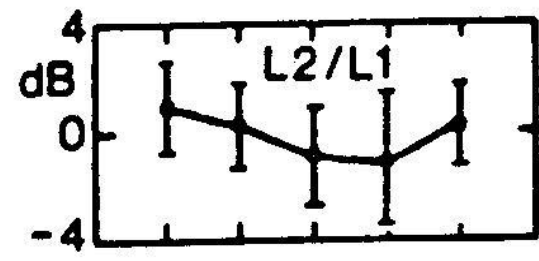
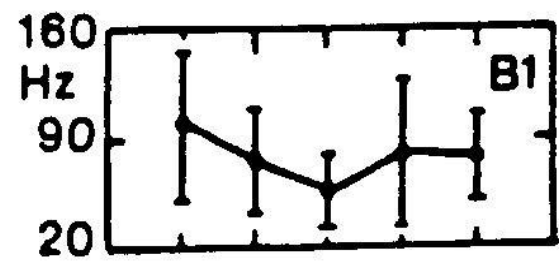




**Διάγραμμα διασποράς των συχνοτήτων formant για 10 αγγλικά φωνήεντα τα οποία εκφωνούνται από 76 ομιλητές (33 άνδρες, 28 γυναίκες και 15 παιδιά) στο επίπεδο F1-F2.**



Τιμές συχνοτήτων των formants των Ελληνικών φωνηέντων.



Τιμές εύρους ζώνης (B) και σχετικού πλάτους (L) για τα φωνήεντα της Νέας Ελληνικής όπως μετρήθηκαν από πέντε άνδρες ομιλητές σε διάφορα φωνητικά περιβάλλοντα.

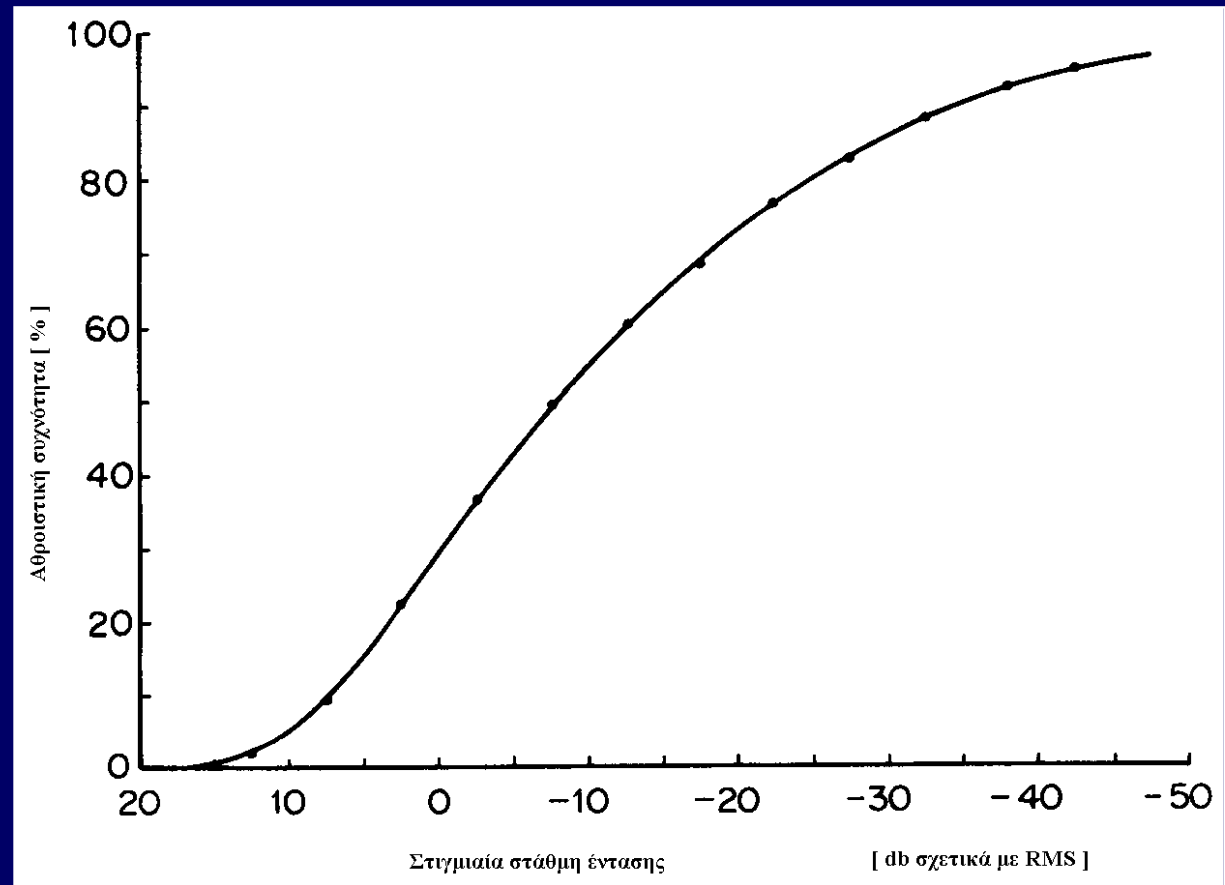


# Στατιστικά Χαρακτηριστικά Στατιστικής της Ομιλίας

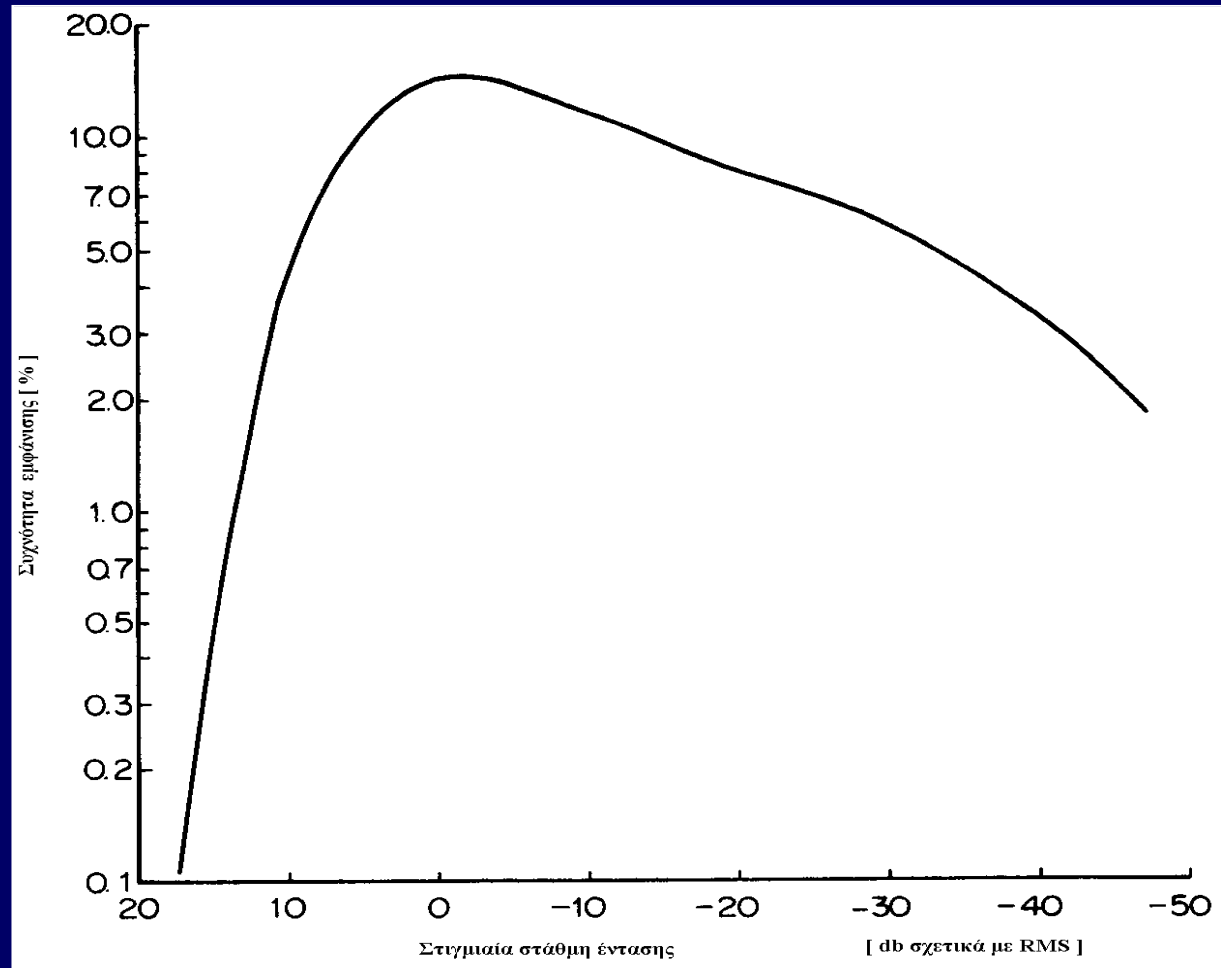
Κατανομή του πλάτους κύματος  $p(x) = 1/2^{1/2} \cdot \sigma \cdot e^{(-2^{1/2}x/\sigma)}$

$\sigma$ : η ενεργός τιμή ( $\sigma^2$  αντιστοιχεί στη μέση ενέργεια)

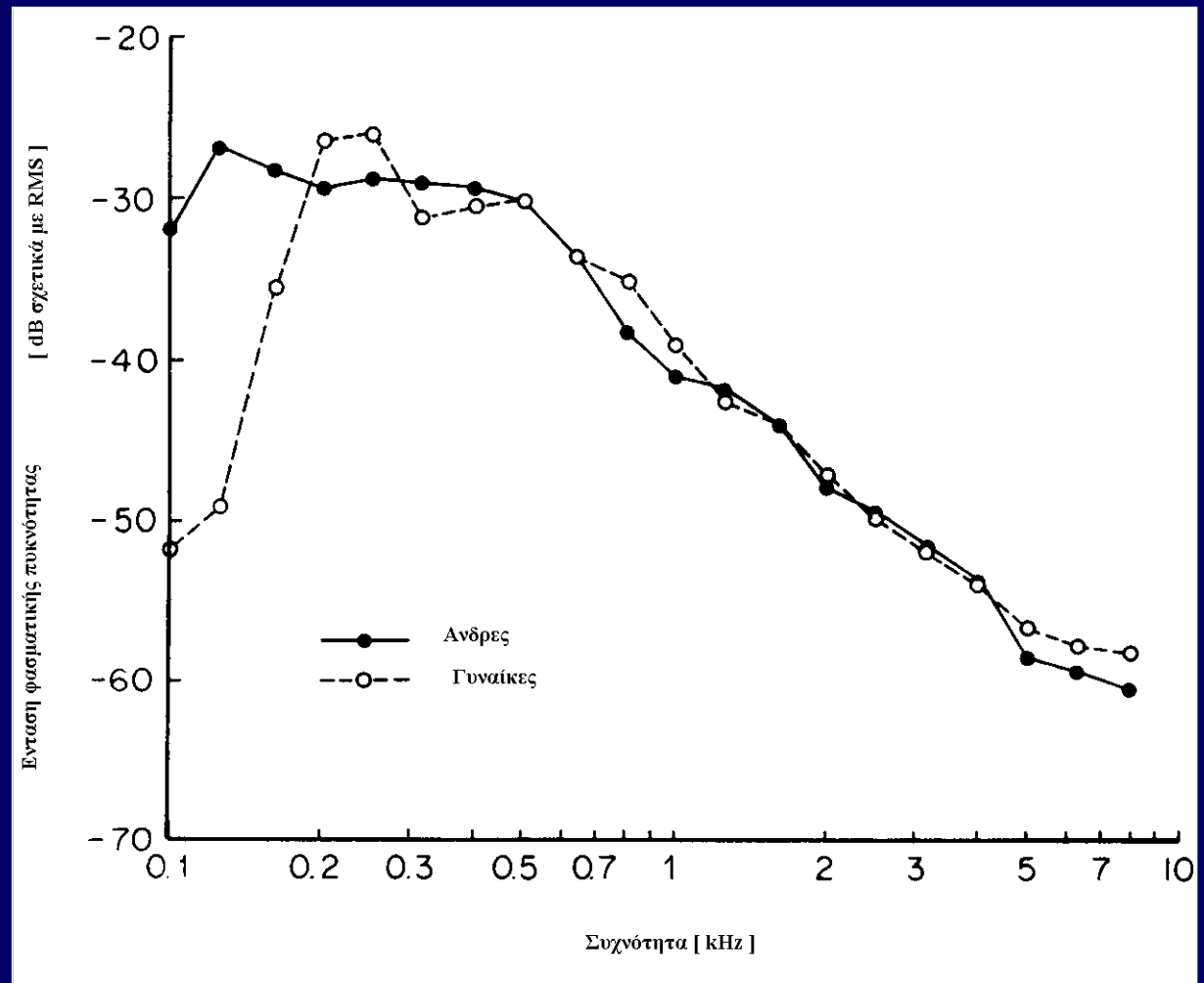
Συγκεντρωτική  
κατανομή της  
στάθμης του πλάτους  
κύματος ομιλίας η  
οποία υπολογίσθηκε  
για εκφωνήσεις 80  
ομιλητών που  
μίλησαν περίπου 37  
λεπτά.



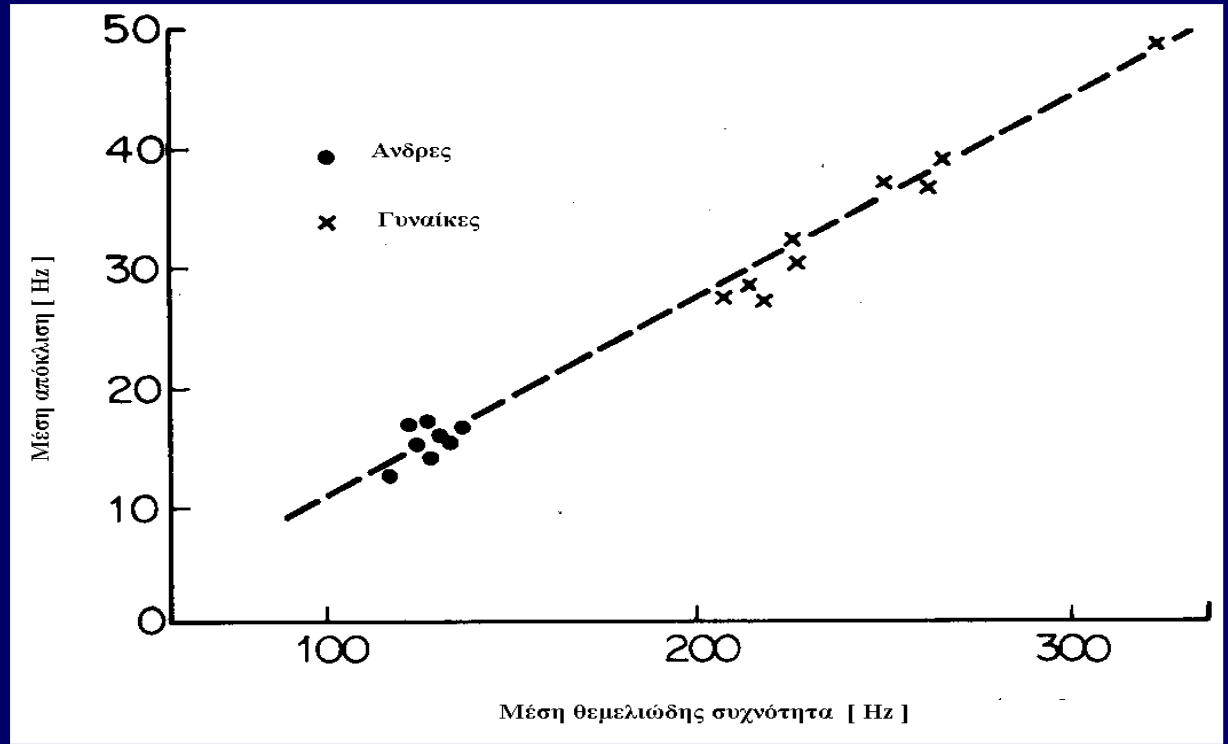
**Συνάρτηση της  
κατανομής της  
πυκνότητας πλάτους  
κύματος η οποία  
παράγεται από το  
Σχήμα της  
προηγούμενης  
διαφάνειας.**



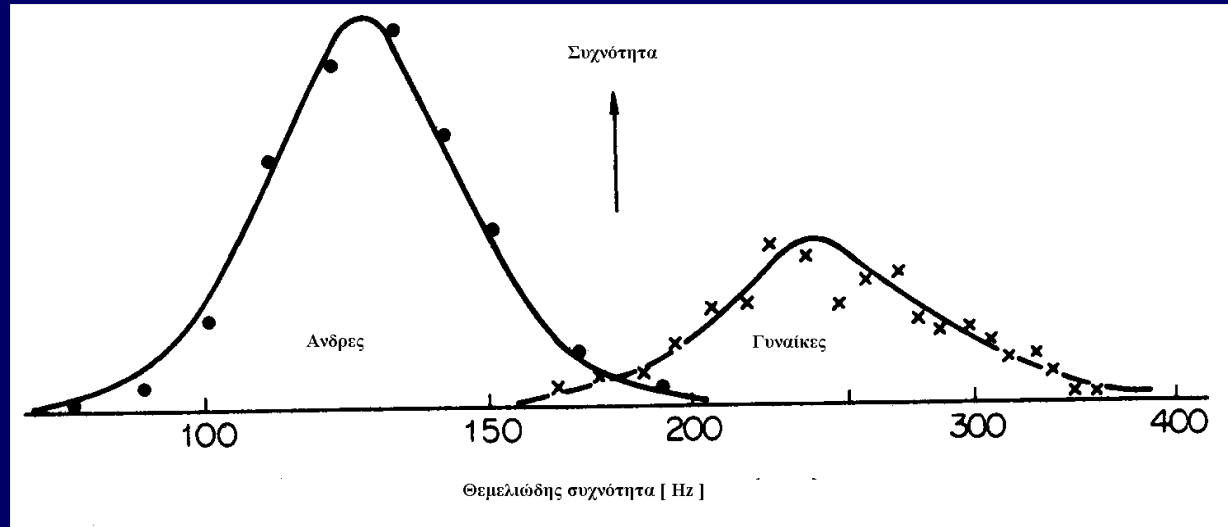
**Μέσο φάσμα  
μεγάλης  
διάρκειας ομιλίας  
υπολογιζόμενο  
για εκφωνήσεις 80  
ομιλητών.**



Μέσες τιμές του ρυθμού μεταβολής  $\Delta f/\Delta t$  κατά τη διάρκεια διαλεκτικής ομιλίας για διάφορους ομιλητές.



Κατανομή της μέσης θεμελιώδους συχνότητας στους ομιλητές.



# Βασικές Ιδιότητες Παραγωγής Ομιλίας

## A. Μη Στατικότητα Μεγάλης Διάρκειας $\geq 0.5$ sec

- ✓ Μη στατικότητα πλάτους  $> 30$  dB
- ✓ Περιοχές: Ηχηρές, Άφωνες, Ησυχίας
- ✓ Φωνητική Δομή και Δομή Γλώσσας

## B. Στατικότητα Μικρής Διάρκειας $\sim 20 - 40$ msec

- ✓ Δομή formant
- ✓ Δομή Τόνου (Ηχηρά)
- ✓ Δομή Θορύβου (Άφωνα)



# 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. Βασικά χαρακτηριστικά της 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. "