

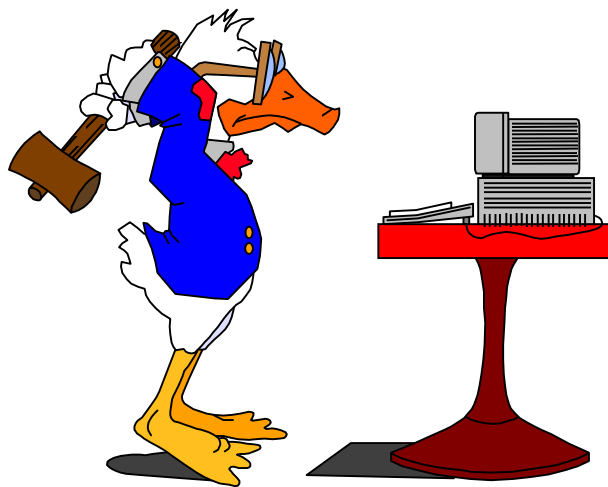
**Μάθημα :** « Ρομποτική – Μαθήματα επεξεργασίας ομιλίας »

**Project :** « Αρθρωτικά μοντέλα στη σύνθεση ομιλίας »

Υπό την επίβλεψη του κ. Κουρουπέτρογλου

Μανδράκης Χρήστος , Α.Μ. 96514 (Η/Α)

Αθήνα 1999



1.1.1.1.1.1.1.1

1.1.1.1.1.1.1.2 ΚΕΦΑΛΑΙΟ 10

## 2 ΑΡΘΡΩΤΙΚΑ ΜΟΝΤΕΛΑ ΣΤΗ ΣΥΝΘΕΣΗ ΟΜΙΛΙΑΣ

3

4

5 *Bernard Gabioud*

*Laboratoire d'analyse informatique de la parole (LAIP)*

*Universite de Lauseanne, CH-1015 Lausanne, Switzerland*

### Περίληψη

Το κεφάλαιο αυτό αφορά στα μοντέλα άρθρωσης που χρησιμοποιούνται στη σύνθεση της ομιλίας. Μετά το πρώτο μέρος όπου αναλύεται η ανάγκη για μοντέλα άρθρωσης, περιγράφεται η αλυσίδα μοντέλων που απαιτούνται για την παρουσίαση της αρθρωτικής σύνθεσης. Στο επόμενο μέρος δίνονται παραδείγματα διαφο-ρετικών ειδών αρθρωτικών μοντέλων που έχουν αναπτυχθεί μέχρι τώρα. Μετά περιγράφεται χονδρικά το μοντέλο της φωνητικής οδού του S. Maeda: πώς δηλαδή το σχήμα μίας φωνητικής οδού μπορεί πραγματικά να περιγραφεί με έναν μικρό αριθμό παραμέτρων. Τέλος, αναφέρονται οι περικοπές της παρούσης στατιστικής προσέγγισης, ενώ προτείνονται οι κατευθύνσεις για μελλοντικές εξελίξεις.

Ο πρωταρχικός στόχος ενός μοντέλου αρθρωτικής σύνθεσης είναι να αναπαράγει το φωνητικό σήμα αποδίδοντας τους μηχανισμούς του φυσικού τρόπου παραγωγής του. Η ομιλία παράγεται μέσω της φωνητικής οδού και είναι το αποτέλεσμα της ταλάντωσης των φωνητικών χορδών, το σήμα των οποίων διαμορφώνεται από την κίνηση της γλώσσας, των χειλιών και άλλων αρθρωτών. Επειδή η μοντελοποίηση του συνόλου αυτής της διαδικασίας αποτελεί μία περίπλοκη εργασία, μιά και συμμετέχουν πολλά κόκκαλα, μύες, μαλακά και σκληρά τοιχώματα και κοιλότητες, είναι απαραίτητη η εφαρμογή απλοποιήσεων. Έχουν προταθεί πολλά διαφορετικά μοντέλα, τα οποία βασίζονται γενικά σε παρόμοια, αλλά και πολλές φορές σε εντελώς διαφορετικά συμπεράσματα. Αυτό το κεφάλαιο ασχολείται με την στατική προσέγγιση των αρθρωτικών μοντέλων.

Σήμερα, η αρθρωτική σύνθεση πρέπει να θεωρείται περισσότερο σαν ένα βασικό εργαλείο έρευνας για την κατανόηση της παραγωγής του λόγου, παρά ως μία μέθοδος παραγωγής εμπορικών εφαρμογών σύνθεσης ομιλίας. Με λίγα λόγια, η αρθρωτική σύνθεση με το να μας δίνει τη δυνατότητα να κατανοήσουμε τους μηχανισμούς παραγωγής της ομιλίας, βοηθά στη μακροπρόθεσμη δυνατότητα επίλυσης των προβλημάτων που επηρεάζουν τις τρέχουσες προσεγγίσεις της σύνθεσης της ομιλίας.

6

### 7 Η ανάγκη για αρθρωτικά μοντέλα

Θα μπορούσε να διαφωνήσει κανείς με την άποψη ότι τα αρθρωτικά μοντέλα δεν είναι το απαραίτητο βήμα για τη σύνθεση ομιλίας. Εφόσον οι στόχοι της παραγωγής ομιλίας είναι κυρίως ακουστικοί (βλέπε Stevens, 1989), κάποιος μπορεί να συμπεράνει ότι η παρεμβολή των αρθρωτών είναι μία άχρηστη επιπλοκή. Η σωστή σύνθεση επιτυγχάνεται είτε με καταταμημένη συρραφή ή με την εφαρμογή κανόνων φωνολογίας σε ακουστικές προσεγγίσεις .

Παρόλα αυτά και οι δύο προσεγγίσεις θεωρούνται περιορισμένες. Οι τεχνικές συρραφής βασίζονται σε λεξικά τα οποία είναι εξαρτώμενα από τον ομιλητή και από το ρυθμό ομιλίας, διότι η συνάρθρωση διαφέρει ανάλογα με το ρυθμό ομιλίας. Σήμερα, δεν είναι γνωστό πώς θα μεταφραστούν τα δείγματα από ένα ρυθμό ομιλίας και από έναν ομιλητή σε μία άλλη συναφή φωνή με σωστή συνάρθρωση. Μπορεί να υποστηρίξει κανείς ότι θα ήταν πολύ καλύτερα να δουλέψει στο επίπεδο των συνεπαγομένων αιτιών, παρά στο επίπεδο των συνεπειών της άρθρωσης.

Τα ακουστικά σε φωνητικά μοντέλα που βασίζονται σε κανόνες είναι περιορισμένα με την έννοια ότι πάρα πολλοί κανόνες χρησιμοποιούνται για να παράγουν τα σωστά αποτελέσματα. Ακόμα και αν αυτοί οι κανόνες είναι σωστά τεκμηριωμένοι φωνολογικά, η φωνητική τους απόδοση είναι πολύ εμπειρική. Αυτό υποδεικνύει επίσης την ανάγκη για μία σωστή θεωρία άρθρωσης για τη σωστή θεμελίωση των φωνητικών φαινομένων.

Προσεγγίσεις που βασίζονται μονάχα στην περιγραφή της φωνητικής οδού σαν ένα σύνολο από  $n$  φωνητικούς σωλήνες διαφορετικής διατομής δε φαίνεται να αποδίδουν ένα σωστό και επαρκές μοντέλο παραγωγής ομιλίας. Εάν ο αριθμός των  $n$  φωνητικών σωλήνων διαφορετικής διατομής είναι μικρός, (Fant 1960), το μοντέλο είναι ανθρωπομορφικά εξωπραγματικό. Εάν ο αριθμός των  $n$  φωνητικών σωλήνων διαφορετικής διατομής είναι επαρκής, ο συνδιασμός των παραμέτρων δε δίνει ικανοποιητικό αποτέλεσμα. Με την έννοια ότι επιτρέπει την ανάλυση και την προσομοίωση των ακουστικά σχετικών φαινομένων, είναι περισσότερο συνετό να θεωρήσει κανείς την προσέγγιση των  $n$  φωνητικών σωλήνων διαφορετικής διατομής σαν το χρήσιμο εργαλείο για την μετάβαση από τη γεωμετρία της φωνητικής οδού στο ακουστικό επίπεδο, παρά να θεωρηθεί σαν ένα αληθινό μοντέλο παραγωγής.

Οι συνθέτες ομιλίας που χρησιμοποιούν τους φωνητικούς συντονισμούς (formants) σαν είσοδο (πχ. Klatt, 1980) δεν έχουν αυτόματες δεσμεύσεις για να συνδέσουν τις διάφορες τιμές των formants μεταξύ τους. Κάθε συγκεκριμένο σχήμα της φωνητικής οδού επιβάλλει μία σύνθετη σχέση μεταξύ όλων των formants, των συχνοτήτων τους και των πλατών τους. Αυτή η εξαρτώμενη από τον ομιλητή σχέση είναι απαραίτητο να ληφθεί υπόψη για σύνθεση υψηλής ποιότητας. Αυτή η σχέση είναι δεδομένη σε ένα αρθρωτικό μοντέλο, ενώ η αναμενόμενη ποιότητα εξαρτάται από την συνολική ποιότητα του μοντέλου.

Ουσιαστικά, τα αρθρωτικά μοντέλα ολοκληρώνονται σφαιρικά και με διάφορους βαθμούς, μορφολογικούς και αρθρωτικούς περιορισμούς. Τέτοια μοντέλα προσφέρουν επίσης τη δυνατότητα εναλλαγής του σχήματος της φωνητικής οδού ενεργοποιώντας

αντίστοιχα τις παραμέτρους ελέγχου αυτών των μοντέλων. Τότε γίνεται δυνατή η σύνδεση αυτών των γεωμετρικών εναλλαγών και των σχετικών αλλαγών στα formants σε πραγματική συμπεριφορά στο σύστημα παραγωγής ομιλίας.

### **7.1 Η αλυσίδα των μοντέλων**

Εξ' ορισμού τα αρθρωτικά μοντέλα προσπαθούν να αναπαράγουν με τον καλύτερο δυνατό τρόπο το σήμα της ομιλίας, τείνουν να προσομοιάσουν τους φυσικούς μηχανισμούς που δίνουν το σήμα αυτό. Αυτού του είδους τα μοντέλα είναι μια απλοποιημένη παρουσίαση του τι πραγματικά συμβαίνει μεταξύ της ταλάντωσης της γλωττίδος και της ακτινοβολίας του ήχου από τα χείλη.

Το πρόβλημα μπορεί να χωριστεί σε διαφορετικά στάδια. Παρόλο που τα στάδια αυτά δεν είναι εντελώς ανεξάρτητα μεταξύ τους, αυτός ο διαχωρισμός κάνει το πρόβλημα πιο εύκολο.

#### ***7.1.1.1.1 Αναγνώριση των Μοντέλων***

#### ***7.1.1.1.2 Επιδίωξη***

Μοντέλο Πηγής	7.1.1.1.2.1 Ταλάντωση φωνητικών χορδών
Τοξοειδής συνάρτηση	Δισδιάστατο σχήμα φωνητικής οδού
Συνάρτηση επιφανείας	Τριών διαστάσεων φωνητική οδός
Μοντέλο χειλιών	Σχήμα χειλιών
Συνάρτηση μεταφοράς	Φάσμα συχνοτήτων
Μοντέλο τριβής	Παραγωγή τυρβώδους θορύβου

**Σχήμα 1.** Η αλυσίδα αρθρωτικής σύνθεσης, η οποία δεν συμπεριλαμβάνει την ίδια τη σύνθεση.

Πρώτα από όλα θα πρέπει να παρουσιαστεί η πρωταρχική πηγή της ανθρώπινης φωνής, δηλαδή η ταλάντωση των φωνητικών χορδών. Πολλές μελέτες έχουν αφιερωθεί σε αυτό τον τύπο προσέγγισης της παραγωγής ομιλίας (π.χ. βλέπε Ishizaka και Flanagan, 1972; Fabt και Liljencrants, 1985; Fant, 1986; Titze, 1989). Υπάρχουν επίσης επιπρόσθετες πηγές θορύβου και εξάρσεων όπου η σύσφιξη είναι περιορισμένη, όπως στα εξακολουθητικά και κλειστά σύμφωνα (βλέπε Klatt, 1980; Stevens και Bickley, 1986; Badin, 1989).

Κατά δεύτερο λόγο, χρειάζεται ένα μοντέλο για το σχήμα της φωνητικής οδού. Αυτό το στάδιο χωρίζεται σε δύο στάδια: Πρώτα, κατασκευάζεται ένα δισδιάστατο μοντέλο του τοξοειδούς τμήματος της φωνητικής οδού (πχ. βασισμένο σε παρατηρήσεις ακτινογραφιών, βλέπε αναφορές στο επόμενο τμήμα). Έπειτα υπολογίζεται μία συνάρτηση επιφανείας, η οποία προσθέτει την τρίτη διάσταση μετατρέποντας το τμήμα του πλάτους σε τμήμα επιφάνειας (βλέπε Chiba και Kajiyama, 1941; Heinz και Stevens, 1964; Perrier κα, 1992; Beautemps κα, 1993).

Τελικά, για να ολοκληρωθεί το μηχανικό μέρος της διαδικασίας κατασκευής του μοντέλου, είναι απαραίτητο ένα μοντέλο χειλιών, ιδιαίτερα εάν η σύνθεση σκοπεύει να συμπεριλάβει την οπτική σύνθεση της κίνησης των χειλιών, η οποία συμβάλλει σημαντικά στην αναγνώριση της ανθρώπινης ομιλίας σε περιβάλλον θορύβου (βλέπε Lindblom και Sundberg, 1971; Arby και Boe, 1986; Guiard-Marigny, 1992).

Με όλα τα αθρωτικά κομμάτια στη θέση τους είμαστε έτοιμοι για τη διαδικασία κατασκευής ακουστικού μοντέλου, το οποίο πάλι συμπεριλαμβάνει έναν αριθμό σταδίων: παραγωγή ηχητικού σήματος στη γλωττίδα, μετάδοση κυμάτων κατά μήκος της φωνητικής οδού, παραγωγή χαρακτηριστικών αντήχησης (resonances) και formants, υπολογισμό των απωλειών στα μαλακά τοιχώματα, διάμετρος της ρινικής οδού, και τέλος ακτινοβολία ήχου από τα χείλη (βλέπε Fant 1960; Badin και Fant 1984; Κεφάλαιο 6).

Το παρόν κεφάλαιο επικεντρώνεται μόνο σε ένα κεντρικό κομμάτι αυτής της αλυσίδας: το πλευρικό μοντέλο της φωνητικής οδού.

## **7.2 Σύγκλιση ιδεών**

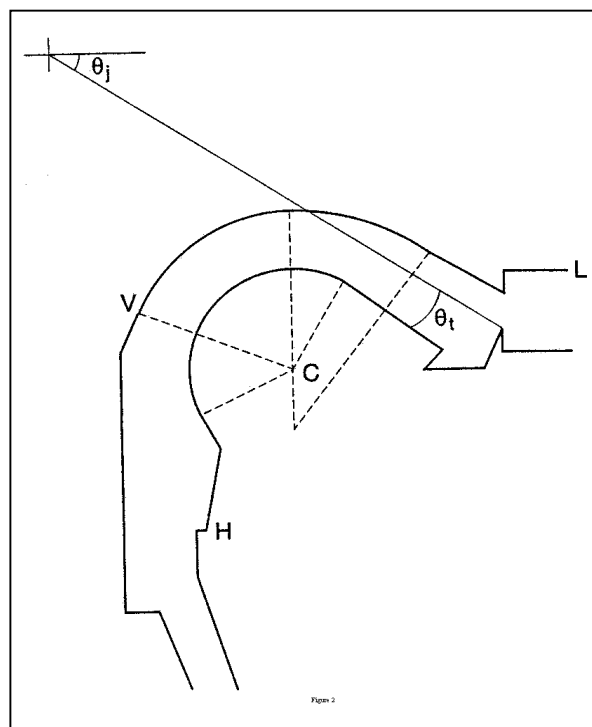
Επειδή έχει προταθεί ένας μεγάλος αριθμός μοντέλων, είναι καθαρά πέρα από το σκοπό της παρούσης εργασίας να παρουσιάσει μία ανάλυση όλων όσων έχουν επιτευχθεί σε αυτό το πεδίο. Είναι απαραίτητο να περιοριστούμε σε μερικά παραδείγματα έτσι ώστε να δείξουμε τις τρεις βασικές κατευθύνσεις τις οποίες θα μπορούσαμε να ονομάσουμε *γεωμετρικά μοντέλα*, *φυσιολογικά μοντέλα* και *στατιστικά μοντέλα*. Επίσης, σε αυτό το ιστορικό κομμάτι θα πρέπει να δώσουμε έμφαση στην προέλευση των ιδεών, τις οποίες ακολούθησαν τα σύγχρονα αρθρωτικά μοντέλα.

### **7.2.1 Γεωμετρικά Μοντέλα**

Μία σημαντική αρχική μελέτη είναι αυτή του Öhman (1966, 1967) που βασίζεται σε κινηματοραδιογραφικά φιλμ στα οποία η κίνηση της γλώσσας περιγράφεται από τρία διαφορετικά μηχανικά συστήματα που επικοινωνούν με έναν κορυφαίο αρθρωτή (apical articulator), έναν νωτιαίο αρθρωτή (dorsal articulator) και έναν αρθρωτή του σώματος της γλώσσας (tongue body articulator). Τα φωνήεντα είναι αποτέλεσμα ενός γραμμικού συνδυασμού των ακραίων φωνητικών θέσεων των /i/,/a/ και /u/. Παρόλα αυτά το σχήμα της γλώσσας παρουσιάζεται σαν κύκλος με σταθερή ακτίνα και οι αρθρωτικές κινήσεις δεν είναι τόσο πραγματικές που να αναπαράγουν τη ραχιαία κίνηση της γλώσσας.

Στην έρευνα των Coker και Fujimura (1966) πέντε παράμετροι καθορίζουν τη διαμόρφωση της φωνητικής οδού. Έτσι το σχήμα της γλώσσας είναι ένας κύκλος με σταθερή ακτίνα, αλλά μία πρόσθετη παράμετρος επιτρέπει καλύτερη προσομοίωση του εμπρόσθιου τμήματος της γλώσσας, με το να αποκλίνει από τον κύκλο. Η κίνηση των χειλιών περιγράφεται από **μία παράμετρο ανοίγματος** και από **μία παράμετρο προώθησης**. **Οι τιμές αυτού του μικρού αριθμού παραμέτρων χρησιμοποιούνται από ένα πρόγραμμα ηλεκτρονικού υπολογιστή για να υπολογιστούν οι συναρτήσεις επιφανείας και οι συχνότητες των formants.**

**Σχήμα 2.** Μία απλουστευμένη κάτοψη του μοντέλου του Mermelstein. Η συνεχής γραμμή δείχνει ένα τοξοειδές τμήμα, το οποίο αποτελείται από 3 τόξα και 18 γραμμικά τμήματα. Οι διακεκομμένες γραμμές δείχνουν τη μέγιστη και την ελάχιστη ακτίνα των τόξων. Ο ουρανίσκος και το εξωτερικό μέρος του φαρυγγικού τοιχώματος είναι σταθερά, ενώ τα άλλα μέρη είναι κινητά. Τα διάφορα γράμματα αναφέρονται στα τόξα και τα σημεία που χρησιμοποιούνται ως παράμετροι εντολών.



Το 1973, ο Mermelstein πρότεινε έναν αρθρωτή, ένα συνθέτη ομιλίας βασισμένο σε κανόνες. Το σχήμα της πλευρικής αναπαράστασης της φωνητικής οδού, αναπαρίσταται από τρία τόξα και από 18 γραμμικά τμήματα. Για τα φωνήεντα ορίζεται ένα σύνολο αρθρωτικών παραμέτρων, οι τιμές των οποίων καθορίζονται από ακτινογραφίες, και συμπληρώνονται από ένα σύνολο ακουστικών παραμέτρων από δεδομένα φυσικής ομιλίας. Τα σύμφωνα εισάγονται σαν μεταβατικά στάδια μεταξύ φωνηέντων που ικανοποιούν συγκεκριμένες συσφίξεις, και έτσι συναρθρώνονται με τα φωνήεντα. Παρόλο που ενώ επιτρέπει μία πιο λεπτομερή περιγραφή του σχήματος της φωνητικής οδού από αυτής του μοντέλου των Coker και Fujimura, αυτό το μοντέλο παρουσιάζει δυσκολίες στο να προσδιορίσει μία κατανοητή κίνηση επειδή χρειάζεται μεγάλος αριθμός παραμέτρων.

### 7.2.1.1 Φυσιολογικά Μοντέλα

Η τάση των φυσιολογικών μοντέλων έχει εκφραστεί από τον Perkell (1969,1974) ο οποίος εκπροσωπεί τη γλώσσα με «16 μυικά σημεία στο υπόβαθρο της μάζας της, τα οποία συνδέονται μεταξύ τους αλλά και με το κυρίως σώμα με 38 ενεργά και 47 παθητικά σημεία παραγωγής». Το πλεονέκτημα των φυσιολογικών μοντέλων είναι ότι επιτυγχάνουν μία σύνδεση μεταξύ της δράσης συγκεκριμένων μυών και της άρθρωσης της γλώσσας.

Στην ίδια τάση των μυικών μοντέλων, κάποιος πρέπει επίσης να αναφερθεί και στην έρευνα των Hiki και Oizumi (1974) και του Kiritani κα. (1975), παρόλο που αυτά τα μοντέλα είναι πολύ περίπλοκα για να εφαρμοστούν στις περισσότερες μορφές της σύγχρονης σύνθεσης ομιλίας.

### 7.2.1.2 Στατιστικά Μοντέλα

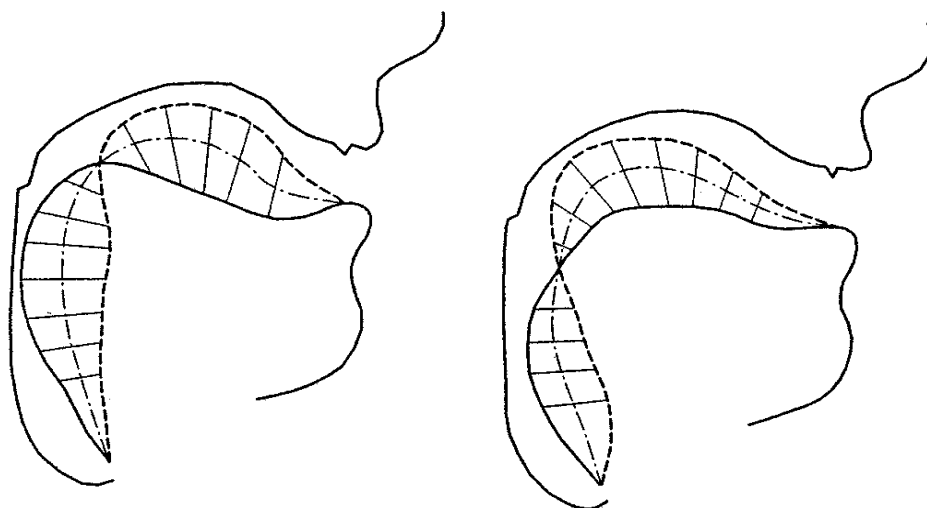
Για να ληφθεί υπόψη η σημασία της περιοχής της σύσφιξης, οι Lindblom και Sundberg (1971) χρησιμοποίησαν την ιδέα της μοντελοποίησης του σχήματος και των θέσεων του σώματος της γλώσσας με παράγοντες που επεξηγούν τις παρεκλίσεις από ένα ουδέτερο βασικό σχηματισμό. Ο παράγοντας C ο οποίος διαφέρει από 0 έως 1, επηρεάζει το σχήμα της γλώσσας και το βαθμό της σύσφιξης. Ο παράγοντας D καθορίζει την μπρος-πίσω θέση της γλώσσας: Το φωνήεν /i/ ανταποκρίνεται στο C = 1, D = -1, το φωνήεν /u/ ανταποκρίνεται στο C = 1, D = 0, το φωνήεν /a/ ανταποκρίνεται στο C = 1, D = +1. Όλα τα υπόλοιπα σχήματα λαμβάνονται με παρεμβολή. Μετά τον υπολογισμό των συχνοτήτων των formants χρησιμοποιώντας το ηλεκτρικό αναλογικό κύκλωμα του Fant (1960), οι Lindblom και Sundberg συνέδεσαν τις ευαισθησίες των formants στις παραμέτρους του μοντέλου.

Αυτού του είδους το μοντέλο περιλαμβάνει τη βασική ιδέα ότι πρέπει να καθοριστούν οι παράμετροι, οι οποίοι ερμηνεύουν την παρέκκλιση από τον ουδέτερο βασικό σχηματισμό. Έκτοτε αυτή η ιδέα έχει υιοθετηθεί από πολλούς άλλους ερευνητές μεταξύ των οποίων και οι Kiritani και Sekimoto (1977), Shirai και Honda (1977), Harshman κα. (1977), και Maeda (1979,1988).

Το 1977, οι Harshman, Kadefoged, Goldstein και Rice πρότειναν μία ανάλυση παραγόντων δέκα αγγλικών φωνηέντων που τα πρόφεραν πέντε ομιλητές. Όλες οι κινήσεις της γλώσσας περιγράφονται από δύο παραμέτρους: από το εμπρόσθιο ανασηκωμένο μέρος το οποίο είναι υπεύθυνο για την προώθηση της βάσεως της γλώσσας και της ανύψωσης του εμπρόσθιου μέρους της γλώσσας και το πίσω ανασηκωμένο μέρος το οποίο ακολουθεί την ανύψωση της γλώσσας προς τη σταμφύλη. Οι ανατομικές διαφορές ανάμεσα στους ομιλητές είναι συνδεδεμένες με διαφορετικές διακυμάνσεις αυτών των παραμέτρων. Στην έρευνά τους το 1978 (Ladefoged κα.,1978) οι ίδιοι συγγραφείς ανακατασκευάζουν το πλευρικό σχήμα της φωνητικής οδού από τα τρία πρώτα formants και γι' αυτό προσθέτουν την παράμετρο του ανοίγματος των χειλιών στο μοντέλο.

Η βασική αδυναμία αυτού του μοντέλου εντοπίζεται στη μη ακριβή αναπαράσταση των κινήσεων του άκρου της γλώσσας, πράγμα που είναι απαραίτητο για τα σύμφωνα. Η θετική όψη του μοντέλου είναι ότι αποδεικνύει πόσο επιτυχημένη μπορεί να είναι η τεχνική της ανάλυσης παραμέτρων στο να παράγει ένα μοντέλο με μικρό αριθμό παραμέτρων και με σωστή προσαρμογή δεδομένων.

Παρόλο που τα φυσιολογικά μοντέλα έχουν το πλεονέκτημα του να προτείνουν μία λεπτομερή περιγραφή της μηχανικής δομής των αρθρωτών και των αλληλεπιδράσεών τους, οι παράμετροι ελέγχου παραμένουν περίπλοκες. Τα αμιγώς γεωμετρικά μοντέλα είναι σχετικά απλά, αλλά οι παράμετροι εισόδου είναι δυστυχώς δύσκολο να ερμηνευτούν με φυσικούς όρους. Τα στατιστικά μοντέλα έχουν ένα καθαρό πλεονέκτημα, εφόσον βασίζονται σε αρκετά δεδομένα έτσι ώστε να εκπροσωπήσουν ικανοποιητικά όλα τα φωνήεντα και τα σύμφωνα μίας γλώσσας και να έχουν ένα μικρό αριθμό από παραμέτρους εντολών που συνδέονται απόλυτα με όλες τις σημαντικές κινήσεις των αρθρωτών.



**Σχήμα 3.** Η επίδραση των δυο παραγόντων στο σχήμα της γλώσσας όπως προβλέφθηκε από το μοντέλο των Harshman, Ladefoged, Goldstein και Rice. Η συνεχής γραμμή αντιπροσωπεύει τη μικρότερη τιμή του παράγοντα, η διακεκομμένη γραμμή τη μεγαλύτερη τιμή και η διακεκομμένη με τελείες γραμμή την ουδέτερη θέση της γλώσσας.

Τέτοια μοντέλα έχουν παρουσιαστεί από τον S. Maeda το 1979 και ξανά το 1988 με διαφορετική επιλογή δεδομένων. Το ακόλουθο κείμενο είναι αφιερωμένο στο πιο πρόσφατο μοντέλο του Maeda.

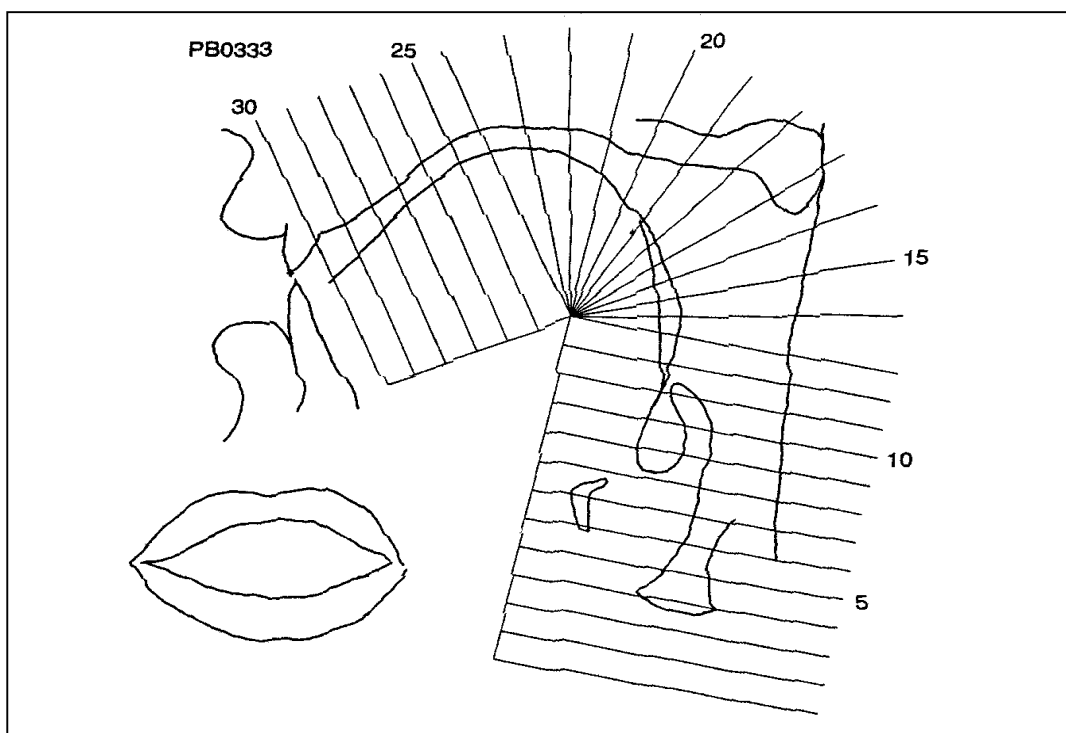
### Το Στατιστικό Μοντέλο του Maeda



Το γραμμικό μοντέλο του Maeda (1988) βασίζεται στην ανάλυση δεδομένων κινηματογραφημάτων (cineradiographic data) από το Ινστιτούτο Φωνολογίας του Στρασβούργου από τον Bothorel κ.α., που εκδόθηκε το 1986. Χρησιμοποιήθηκε μία συλλογή ηχητικών δεδομένων διάρκειας 10,4 δευτερολέπτων πραγματικής ομιλίας αποτελούμενη από δέκα προτάσεις, εκφωνούμενες από γυναίκα, Γαλλίδα ομιλήτρια. Με ταχύτητα 50 πλαισίων ανά δευτερόλεπτο, αποδόθηκαν 519 πλαίσια συνολικά.

Χρησιμοποιώντας ημισφαιρικό σύστημα συντεταγμένων, το πλευρικό προφίλ μετρήθηκε κατά μήκος 30 γραμμικών πλεγμάτων (κυψελών) πάνω στις ψηφιοποιημένες και ιχνογραφημένες ακτινογραφίες (Σχήμα 4). Αντίθετα με τον τύπο πλέγματος που χρησιμοποιήθηκε παλαιότερα, το πλέγμα του 1988 έχει ένα ευθυγραμμισμένο τμήμα για καλύτερη περιγραφή του μπροστινού μέρους της γλώσσας.

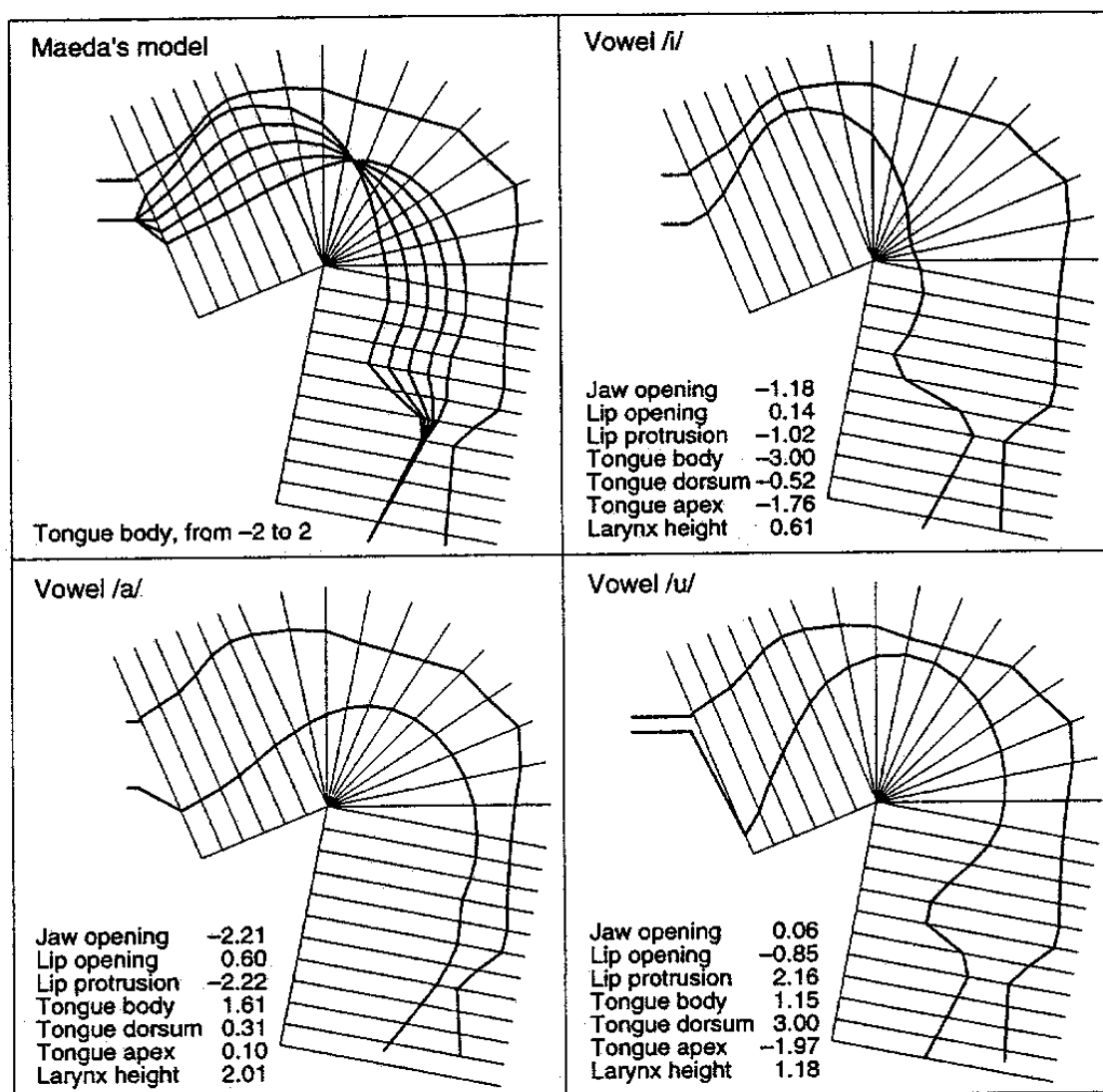
Εδώ θα περιγράψουμε σύντομα πώς έχει δημιουργηθεί αυτό το γραμμικό μοντέλο της γλώσσας. Βασικά ακολουθεί το μοντέλο που προτάθηκε από τους Lindblom και Sundberg το 1971. Πρώτα από όλα η μέση τιμή των δεδομένων είναι καθορισμένη και μόνο οι αποκλίσεις από αυτό το ουδέτερο σχήμα μοντελοποιούνται. Το άθροισμα των τετραγώνων των κανονικοποιημένων αποκλίσεων ονομάζεται ολική απόκλιση των δεδομένων (*the total variance of data*).



**7.2.1.2.1.1.1 Σχήμα 4.** Ένα χειρόγραφο σχέδιο της φωνητικής οδού, ραδιογραφική διατομή με το ημιπολικό πλέγμα συντεταγμένων του Maeda. Κάτω αριστερά η μπροστινή όψη των χειλιών. Το φωνήεν είναι το /i/ ακριβώς πριν την ουρανική φραγή του φωνήματος /k/ στην πρόταση στα γαλλικά «Donne un petit coup»

Κατόπιν, όπως το κάτω χείλος και η γλώσσα είναι προσκολλημένες στο σαγόι, είναι τεκμηριωμένο ότι πρέπει να συνυπολογίζεται η κίνηση του σαγονιού πριν να ληφθούν υπόψη οι άλλες κινήσεις. Επιπλέον, από τα δεδομένα αφαιρείται πρώτα η επίδραση από το άνοιγμα της σιαγόνος. Σε αυτή τη συλλογή ηχητικών σημάτων αυτή η παράμετρος εξηγεί μόνο το 15% της διακύμανσης, γιατί η επιλεγμένη ομιλήτρια δεν κινεί πολύ το σαγόι της κατά τη διάρκεια της ομιλίας της.

Κατόπιν όλων αυτών, ερμηνεύεται η βασική ανάλυση του συστήματος των δεδομένων της γλώσσας, που οδηγεί σε τρεις σημαντικούς παράγοντες. Ο πιο σημαντικός παράγοντας ακολουθεί την μπρος-πίσω θέση του σώματος της γλώσσας, και εξηγεί το 43% της διακύμανσης (βλέπε σχ. 5). Ο δεύτερος παράγοντας μεταφράζεται ως το ύψος του ραχιαίου μέρους της γλώσσας και εξηγεί το 23% της διακύμανσης. Ο τρίτος παράγοντας συνδέεται με το ύψος της κορυφής της γλώσσας και εξηγεί το 7% της διακύμανσης. Έτσι το μοντέλο αυτό ερμηνεύει συνολικά το 88% της διακύμανσης των δεδομένων της γλώσσας. Το πιο σημαντικό είναι ότι οι παράγοντες που παίρνει ο Maeda έχουν μία εξήγηση, η οποία αντιστοιχεί σε αυτά που περιμένουμε από άλλες μελέτες για τις κινήσεις της γλώσσας. (Ohman, 1966; Coker και Fujimura, 1966).



**Σχήμα 5.** Σχεδιαγράμματα τεσσάρων φωνητικών οδών κατά το μοντέλο του Maeda. Η επάνω αριστερή φωτογραφία δείχνει την επίδραση της παραμέτρου του σώματος της γλώσσας που ποικίλει από -2 έως +2. Τα άλλα δείχνουν παραδείγματα των φωνηέντων /i/, /a/ και /u/. Το φωνήεν /i/ έχει το σώμα της γλώσσας μπροστά και χαμηλή κορυφή. Το φωνήεν /a/ έχει το σώμα της γλώσσας όπισθεν και το σαγόρι ανοιχτό. Το φωνήεν /u/ έχει έντονη προεξοχή και το ραχιαίο της γλώσσας στο ψηλότερο σημείο του.

Τα δεδομένα των χειλιών, συμπεριλαμβανομένου του πλάτους, μπορούν να περιγραφούν επαρκώς από δύο παραμέτρους: το άνοιγμα και την προώθηση των χειλιών. Μία παράμετρος ακολουθεί το μήκος του λάρυγγα. Η κίνηση της σταμφύλης δεν περιλαμβάνεται, καθώς επίσης και η ρινική οδός.

Με το μοντέλο του Maeda κάθε σχήμα της φωνητικής οδού μπορεί να καθοριστεί από τις τιμές των επτά κανονικοποιημένων παραμέτρων, οι οποίες συνήθως κυμαίνονται μεταξύ  $-3$  και  $+3$  (βλέπε σχ. 5)

Στην εργασία του, του 1990 ο Maeda χρησιμοποιεί δεδομένα από δύο ομιλητές και παρατηρεί τη χρονική διακύμανση των παραμέτρων του μοντέλου του, που ερμηνεύονται ως οι βασικοί αρθρωτές ομιλίας. Το πιο σημαντικό αποτέλεσμα είναι η απόδειξη της αρθρωτικής αντιστάθμισης. Διαφορετική διαμόρφωση των αρθρωτών χρησιμοποιείται για να παραχθεί ο ίδιος ήχος. Η προσαρμογή της θέσης του ραχιαίου τμήματος της γλώσσας αντισταθμίζει το διαφορετικό άνοιγμα των σιαγόνων. Στην περίπτωση των στρογγυλών φωνέντων τα διαφορετικά ανοίγματα των σιαγόνων αντισταθμίζονται από μία κατάλληλη προσαρμογή των ανοιγμάτων των χειλιών. Αυτό δείχνει καθαρά ότι οι στόχοι της παραγωγής της ομιλίας είναι ακουστικοί και όχι αρθρωτικοί, οι αρθρωτές προσαρμόζονται με τις συσφίξεις τους για να παράγουν τον επιθυμητό ήχο.

Αν και τα αποτελέσματα προκύπτουν από την ανάλυση της φωνητικής οδού μίας γυναίκας ομιλήτριας, το μοντέλο του Maeda αποδίδει το ίδιο καλά την παραγωγή των γαλλικών φωνέντων όπως αυτά προφέρονται από έναν άνδρα ομιλητή. Εφαρμόζοντας απλά μία ολική αύξηση των παραμέτρων της φωνητικής οδού (εκτός από το μήκος του σκληρού τμήματος του ουρανίσκου), Boe κα. (1992) παράγεται ικανοποιητική ανδρική φωνή, με τα formants όλων των προτευόντων Γαλλικών φωνέντων στη σωστή περιοχή. Ουσιαστικά στο σχήμα 5 χρησιμοποιείται η ανδρική έκδοση αυτού του μοντέλου.

Το αρθρωτικό μοντέλο του Maeda συμπεριλαμβάνει επιτυχώς τα κύρια χαρακτηριστικά της φωνητικής οδού ενός πραγματικού ανθρώπου, που οδηγεί στη σύνθεση μίας παρόμοιας φωνής. Όλα τα φωνήεντα που παράγονται από αυτό το μοντέλο ακούγονται σα να είχαν προφερθεί από την ίδια φωνή.

### **7.2.2 Μελλοντικές Τάσεις**

Πρώτα από όλα θα ήταν πολύ ενδιαφέρον να γενικεύσουμε αυτό το μοντέλο τού ενός ομιλητή σε άλλα άτομα, με το να προσθέσουμε και άλλες παραμέτρους επιτρέποντας αλλαγές στο μήκος και στο πλάτος διαφόρων τμημάτων της φωνητικής οδού.

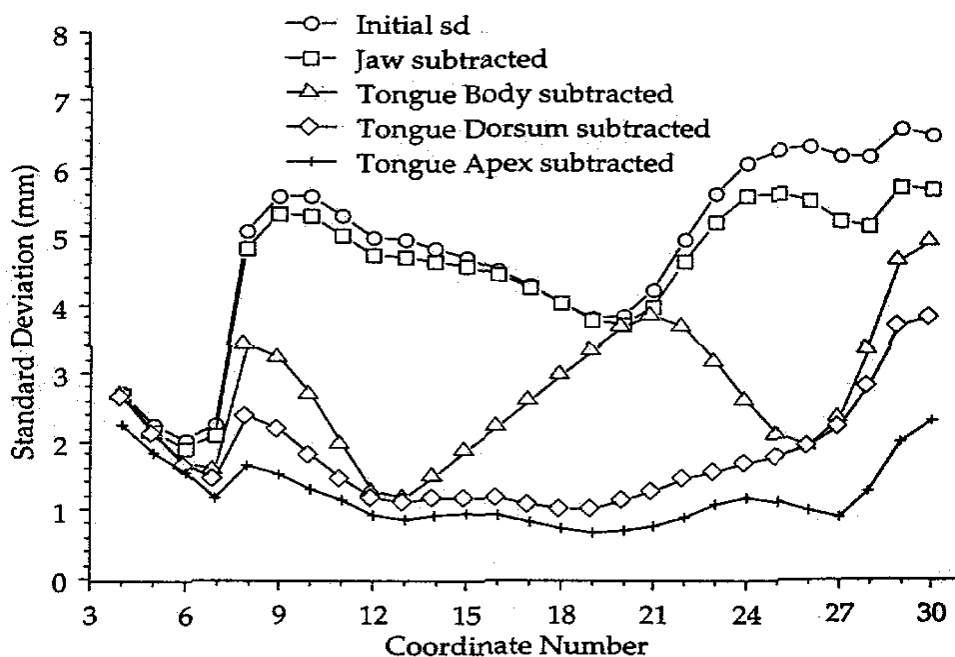
Μία ακόμη σημαντική εργασία θα ήταν εάν χρησιμοποιούνταν το μοντέλο του Maeda σ' ένα συνθέτη ομιλίας. Παρόλα αυτά στις προσπάθειές μας να επανασυνθέσουμε τις αρχικές προτάσεις που χρησιμοποιήθηκαν στην ανάλυση του Maeda, αντιμετωπίσαμε πολλά δισεπίλυτα προβλήματα, μερικά από τα οποία δεν είναι τόσο δύσκολο να επιλυθούν, ενώ άλλα είναι πολύ πιο δύσκολα. Σα συμπλήρωμα στο μοντέλο πλευρικής λειτουργίας του Maeda χρησιμοποιήθηκε η συνάρτηση επιφανείας του Perrier κα. (1992) και η ακουστική απόκριση της φωνητικής οδού υπολογίζεται με τη χρήση του μοντέλου των Badin και Fant (1984).

### 7.2.2.1

#### 7.2.2.2 Το άκρο της γλώσσας δεν έχει συμπεριληφθεί από τις στατιστικές

Η ανάλυση βασικών τμημάτων είναι επαναληπτική. Βρίσκει τον παράγοντα (βλέπε το γραμμικό συνδυασμό μεταβλητών) ο οποίος εξηγεί (μειώνει) το μεγαλύτερο μέρος της διακύμανσης και τότε ανακαλύπτει έναν άλλο παράγοντα ο οποίος ξανά «μειώνει» το μεγαλύτερο μέρος της υπόλοιπης διακύμανσης κ.ο.κ. Συνήθως μετά από έναν μικρό αριθμό επαναλήψεων το μεγαλύτερο μέρος των δεδομένων της διακύμανσης εξηγείται και τότε μπορεί να σταματήσει η διαδικασία. Αυτή η μέθοδος έχει το πλεονέκτημα να εξηγεί το μεγαλύτερο μέρος της διακύμανσης με το μικρότερο αριθμό παραμέτρων.

Για να χρησιμοποιηθούν σε ένα αρθρωτικό μοντέλο οι παράμετροι χρειάζεται να έχουν καλή φυσική σημασία, έτσι ώστε κάποιος να μπορεί να προβλέψει μία συγκεκριμένη δυναμική συμπεριφορά τους. Μετά την αφαίρεση της επίπτωσης της κίνησης του σαγονιού, ο Maeda δεν είχε πρόβλημα με αυτή την απαίτηση και έτσι κράτησε τις τρεις πρώτες βασικές συνιστώσες ως τις παραμέτρους του σχήματος της γλώσσας.



**Σχήμα 6.** Η σταθερή απόκλιση στο μοντέλο γλώσσας του Maeda σε συνάρτηση με τις τιμές των συντεταγμένων (χαμηλές τιμές συντεταγμένων: φάρυγγας, μεγάλες τιμές συντεταγμένων: άκρο της γλώσσας). Η πάνω γραμμή δείχνει την αρχική σταθερή απόκλιση. Οι άλλες γραμμές δείχνουν τη συνολική επίδραση των παραμέτρων που λαμβάνονται υπόψη. Η διαφορά μεταξύ δύο γραμμών δείχνει την επίδραση, αν συμπεριληφθεί μία νέα παράμετρος. Η κάτω γραμμή αποτελεί την τελική σταθερή απόκλιση που απομένει για το μοντέλο του Maeda.

Δυστυχώς η ακρίβεια της θέσης του μπροστινού άκρου της γλώσσας που καταγράφεται από αυτό το μοντέλο δεν είναι επαρκής. Παρόλο που επεξηγείται πολύ καλά η διακύμανση των δεδομένων στη μέση της γλώσσας, η λύση των τριών παραγόντων δε συμπεριλαμβάνει καθόλου τα δύο άκρα της γλώσσας. Τα υπόλοιπα αποκλίσεων δεν είναι ομοιόμορφα κατανομημένα κατά μήκος της φωνητικής οδού. Είναι ιδιαίτερα μικρά στη μέση και συμμετρικά μεγάλα στα δύο άκρα της γλώσσας. Παρόλο που μία ανακρίβεια στην περιοχή του φάρυγγα δεν έχει μεγάλη ακουστική επίπτωση, μία ανακρίβεια στην μπροστινή κοιλότητα επηρεάζει σημαντικά την ποιότητα των εμπρόσθιων φωνηέντων.

Η κάτω γραμμή του σχεδιαγράμματος 6 δείχνει τα RMS υπόλοιπα ως συνάρτηση του αριθμού της συντεταγμένης. Η ανάλυση στο άκρο της γλώσσας (συντεταγμένη 30) είναι χειρότερη κατά έναν παράγοντα 3 από ό,τι στο μέσο της γλώσσας (συντεταγμένη 19). Μία ανάλυση 2mm σημαίνει ουσιαστικά ότι το 95% των στοιχείων που προκύπτουν είναι μεταξύ -4mm και +4mm.

Ο λόγος για αυτή τη συμπεριφορά είναι ότι οι συντεταγμένες έχουν ένα δυνατό τοπικό συσχετισμό. Ο πίνακας 7 δείχνει μέρος του συσχετισμού, ο οποίος αναφέρεται στο μεσαίο τμήμα της γλώσσας. Στατιστικά η θεωρία υποθέτει στατιστική ανεξαρτησία: οι παρατηρήσεις θεωρούνται ότι είναι μεταξύ τους ανεξάρτητες. Από την ανάλυση αυτών των δεδομένων βλέπουμε ότι ο ισχυρός συσχετισμός μεταξύ γειτονικών σημείων εισάγει μεγάλους περιορισμούς, και έτσι έχουμε μία καλύτερη εφαρμογή στο μέσον της γλώσσας παρά στα δύο άκρα της, όπου οι περιορισμοί είναι ασθενέστεροι. Συνεπώς μία επιπλέον κύρια παράμετρος θα χρειαζόνταν για τα δύο άκρα της γλώσσας.

	d10	d12	d14	d16	d18	d20	d22	d24	D26
d10	1.000								
d12	0.924	1,000							
d14	0,766	0,936	1,000						
d16	0,600	0,820	0,960	1,000					
d18	0,346	0,604	0,809	0,929	1,000				
d20	-	0,069	0,331	0,540	0,791	1,000			
	0,170								
d22	-	-0,607	-0,398	-	0,123	0,668	1,00		
	0,714			0,186			0		
d24	-	-0,845	-0,735	-	-	0,246	0,867	1,000	
	0,833			0,586	0,333				
d26	-	-0.847	-0.826	-	-	-0.039	0.667	0.929	1.000
	0.740			0.741	0.561				

**Σχήμα 7.** Ο συσχετισμός γινομένου-ροπής (product-moment) του Pearson αποτελούμενος από (μέρος του) συντεταγμένες της γλώσσας παραλείποντας κάθε δεύτερη γραμμή. Εκτός από έναν πολύ ισχυρό τοπικό συσχετισμό των δεδομένων μπορούμε να παρατηρήσουμε (γύρω στο d12-d24) τον πολύ ισχυρό αντισυσχετισμό (anticorrelation) που απορρέει από τη φυσική μπρος-πίσω κίνηση του σώματος της γλώσσας.

Όπως ο ίδιος ο Maeda υποστηρίζει, δε χρειάζεται κάποιος να χρησιμοποιεί κύριες παραμέτρους ως παράγοντες. Τυχαίοι παράμετροι, οι οποίοι έχουν μία καθαρά φυσική σημασία μπορούν να χρησιμοποιηθούν αντ'αυτών. Αυτό έκανε ο Maeda στο μοντέλο του το 1979 για όλους τους παράγοντες της γλώσσας, και στο μοντέλο του 1988 για τον παράγοντα του ανοίγματος της σιαγόνας. Τέτοιοι παράγοντες προκύπτουν από απευθείας μετρήσεις παρά από στατιστικές αναλύσεις. Το μειονέκτημα τέτοιων αυθαίρετων παραγόντων είναι ότι συνήθως δεν εξηγούν τόσο μεγάλο μέρος της διακύμανσης όσο οι κύριοι παράμετροι. Αλλά αυτό δε μας απασχολεί εδώ κατά κύριο λόγο.

Ακολουθώντας την αιτιολόγηση του 1979 του Maeda είναι πιθανό να δημιουργηθεί ένα μοντέλο γλώσσας με τρεις τυχαίους παράγοντες, οι οποίοι έχουν την ίδια φυσική απόδοση όπως οι κύριες παράμετροι, αλλά οι οποίοι επιπλέον ελαχιστοποιούν τα σφάλματα υπολοίπων στην περιοχή όπου χρειαζόμαστε την καλύτερη εφαρμογή.

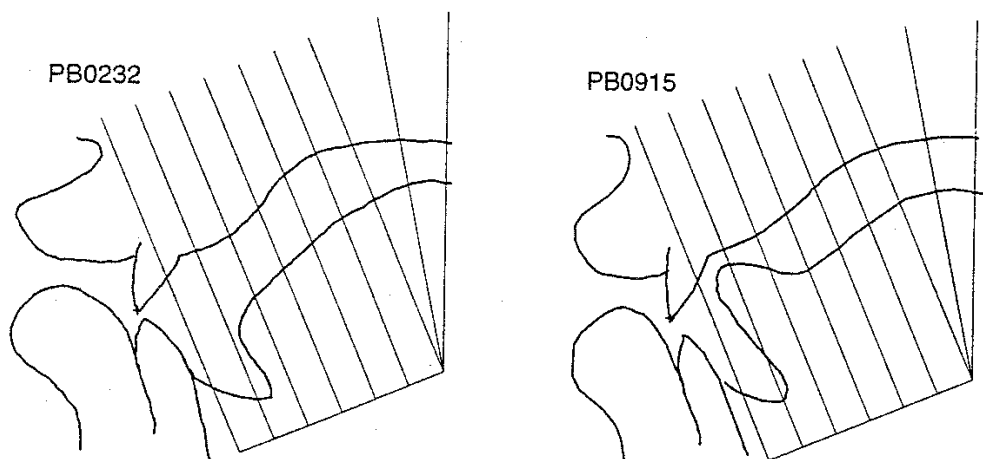
Εδώ αναφέρουμε μία μέθοδο. Μετά από την αφαίρεση της επίδρασης της σιαγόνας η εμπρός-πίσω κίνηση του κυρίως σώματος της γλώσσας μπορεί να μετρηθεί από τη διαφορά των συντεταγμένων 11 και 25 (αυτό είναι καλύτερο από το να χρησιμοποιηθούν μετρήσεις μόνο στο πίσω μέρος). Τότε, αφού αφαιρεθούν οι επιδράσεις της παραμέτρου από τα δεδομένα, η ανύψωση του σώματος της γλώσσας μπορεί να καθοριστεί με τη συντεταγμένη 18. Τελικά, το ύψος του άκρου της γλώσσας μπορεί να καθοριστεί με τη συντεταγμένη 29. Επιπλέον, για να μειωθεί η επίπτωση από την ανακρίβεια των μετρήσεων, μπορεί κανείς να χρησιμοποιήσει το μέσο όρο πολλών γειτονικών συντεταγμένων από το να χρησιμοποιήσει μόνο μία.

Μία ακόμα καλύτερη λύση είναι η ακόλουθη. Μετά την αφαίρεση της επίδρασης της σιαγόνας, για τη βασική ανάλυση των μερών της γλώσσας χρησιμοποιούνται οι συντεταγμένες 9 έως 27 (εκτός από το άκρο της γλώσσας) καταλήγοντας σ' έναν παράγοντα σώματος της γλώσσας και σ' έναν παράγοντα του ραχιαίου της γλώσσας. Μετά από την αφαίρεση της συνδυασμένης επίδρασης αυτών των δύο παραγόντων μπορεί να υπολογιστεί από τις συντεταγμένες 27 έως 30 μία αξιόπιστη παράμετρος για το άκρο της γλώσσας.

Χρησιμοποιώντας μία από τις δύο αυτές λύσεις περιορίζουμε την παράμετρο του άκρου της γλώσσας στο 1mm. Όταν οι αρχικές προτάσεις που χρησιμοποιούνται για την ανάλυση έχουν επανασυντεθεί, η διαφορά είναι αναμφιβόλως αντιληπτή και εντοπίζεται κυρίως στα μπροστινά φωνήεντα όπως τα /i/ και /y/ .

Θα πρέπει να αναφερθεί ότι για τον αποκλειστικό σκοπό της απευθείας παραγωγής σήματος ομιλίας, το μέγεθος της προόδου που παρουσιάζεται σε αυτό το τμήμα δεν είναι ιδιαίτερα σημαντικό σε σχέση με την ανάλυση του Maeda. Εφόσον αυτή η ανακρίβεια εφαρμογής μπορεί να ρυθμιστεί με μία μικρή διαφορά της τιμής των παραμέτρων, οι ίδιες φωνητικές οδοί μπορούν να δημιουργηθούν και από τα δύο μοντέλα. Παρόλα αυτά το νέο μοντέλο μπορεί να δώσει καλύτερη ερμηνεία στις παραμέτρους: αντί για γραμμικούς συνδυασμούς όλων των μετρήσεων, οι νέοι

παράγοντες βασίζονται στις πιο σχετικές μετρήσεις των αρθρωτών. Με αυτό τον τρόπο, η κίνηση της γλώσσας μπορεί να περιγραφεί πολύ καθαρά σαν μία κίνηση της σιαγόνας και μιάς μπρός-πίσω κίνησης του σώματος της γλώσσας σε σχέση με τη σιαγόνα συν μιας ανύψωσης του πίσω μέρους της γλώσσας σε σχέση με το σώμα της γλώσσας συν μιας τελικής ρύθμισης του υψηλότερου σημείου. Αυτό είναι πολύ σημαντικό για τη δυναμική της κίνησης της γλώσσας.



**Σχήμα 8.** Δύο παραδείγματα που δείχνουν το άκρο της γλώσσας σε τέτοια θέση που η κατάλληλη επεξεργασία του είναι αδύνατη. Η αριστερή φωτογραφία είναι ένα /u/ και η δεξιά φωτογραφία είναι ακριβώς μετά το φωνητικό /ʌ/. Βλέπε επίσης σχήμα 4 για σύγκριση με την κανονική (και πιο συνήθη) περίπτωση.

### Το σχήμα του άκρου της γλώσσας

Ακόμα και με όλες τις προαναφερόμενες βελτιώσεις δεν έχουμε ακόμα κατορθώσει ένα πραγματικά, ικανοποιητικό πλευρικό μοντέλο: το άκρο της γλώσσας χρειάζεται περισσότερη έρευνα. Η προηγούμενη ανάλυση υποθέτει ότι το περίγραμμα της γλώσσας τέμνει όλες τις γραμμές των πλεγμάτων. Οι δύο εικόνες που ακολουθούν δείχνουν ότι αυτό δεν είναι πάντοτε σωστό.

Εφόσον υπάρχει, το σχήμα της κοιλότητας μπροστά από το άκρο της γλώσσας μπορεί χονδρικά να υπολογιστεί λανθασμένα και να χρησιμοποιηθεί έτσι στην ανάλυση. Για να εξαληφθεί αυτό θα πρέπει κανείς μάλλον να σταματήσει τη χρήση των πλεγμάτων τουλάχιστον 1cm πριν από τα δόντια, και να χρησιμοποιήσει στην ανάλυση την πραγματική θέση του άκρου της γλώσσας από τη θέση της στους άξονες x και y. Αυτό δεν μπορεί να πραγματοποιηθεί με τα δεδομένα του Στρασβούργου, επειδή το άκρο της γλώσσας δεν έχει καθοριστεί αρκετά καλά από τις ακτινογραφίες. Θα πρέπει να σημειώσουμε ότι ο Lasefoged το 1979 στον κατάλόγό του με τις 17 απαραίτητες και ικανές παραμέτρους περιλαμβάνει το μέγεθος της προώθησης του άκρου της γλώσσας, τις προεξοχές της γλώσσας, τη στενότητα της γλώσσας και την κοιλότητα της γλώσσας.



Δεκαπέντε χρόνια μετά, επιβεβαιώνεται για άλλη μια φορά ότι όλες οι έρευνες πρέπει να γίνουν σ' αυτή την κατεύθυνση.

### 7.2.2.3 Ο έλεγχος της τριβής

Ένα ακόμη ερώτημα που αφορά στα εξακολουθητικά σύμφωνα είναι: πόσο ακριβές είναι το μοντέλο για την πρόβλεψη μικρών περιοχών σύσφιξης; Η απάντηση είναι: σίγουρα όχι ιδιαίτερα ακριβής. (Beautemps, 1993). Σε αυτό συμβάλλει η ενυπάρχουσα ανακρίβεια της μεθόδου ανάλυσης και πρόβλεψης.

Θα πρέπει να θυμόμαστε ότι είναι δύσκολο ακόμη και για μας τους ίδιους. Η παραγωγή των σωστών τυρβώδων ή εξακολουθητικών είναι η τελευταία αρθρωτική ικανότητα που αποκτούν τα παιδιά και μερικές φορές ακόμη και μερικοί ενήλικες δεν μπορούν να τη χρησιμοποιήσουν πολύ επιδέξια. Επίσης, το σύστημα ελέγχου της κίνησης δεν είναι μόνο ένα απευθείας σύστημα παραγωγής. Χρησιμοποιεί ένα συνδυασμό διαφορετικών τεχνικών για να φτάσει στο στόχο, όπως αυτόν της αντίληψης των συνεπειών της κίνησης σε συνδυασμό με τις προκαθορισμένες κινήσεις. Περαιτέρω χρησιμοποιούνται διάφορα τεχνάσματα όπως το γύρισμα και των δύο πλευρών της γλώσσας προς τον ουρανίσκο, έτσι ώστε να ασκηθεί ο σωστός μηχανικός έλεγχος σ' ένα φατνιακό εξακολουθητικό φώνημα.

### 7.2.2.4 Συμπεράσματα

7.2.2.4.1 Η στατιστική μέθοδος εμπεριέχει τη δύναμη και την απλότητα στο αρθρωτικό μοντέλο: με τη δημιουργία ενός μόνιμου μεγάλου πίνακα συντελεστών, που χρησιμοποιεί πραγματικά δεδομένα ένας μικρός αριθμός παραμέτρων επιτρέπει μιά επαρκή περιγραφή όλων των σχημάτων που μπορεί να πάρει η φωνητική οδός.

#### 7.2.2.4.2

7.2.2.4.3 Το αρθρωτικό μοντέλο που παρουσιάζεται εδώ δίνει ένα πολύ καλό αποτέλεσμα στην παραγωγή των γαλλικών φωνηέντων, σε σχέση με τις ιδανικότερες συναρτήσεις επιφανείας, τις ακουστικές συναρτήσεις μεταφοράς και τη σύνθεση των formants. Ειδικότερα αξιοσημείωτη είναι η πρόβλεψη των συχνοτήτων των formants.

Η εργασία επανασύνθεσης έδειξε ότι τα σταθερά πλέγματα είναι πολύ περιορισμένα σαν τρόπος μέτρησης από κινηματοραδιογραφικά δεδομένα. Επίσης, η χρήση βασικών χαρακτηριστικών από τις συντεταγμένες σταθερών πλεγμάτων αποτυγχάνουν στο να δώσουν μία ικανοποιητική ανάλυση των περιοχών της φωνητικής οδού, όπου αυτό απαιτείται.

Περισσότερη έρευνα θα πρέπει να γίνει για την παραγωγή των συμφώνων. Επίσης, εάν η επέκταση σε άλλους ομιλητές και σε άλλες γλώσσες αποτελεί μελλοντικό αντικειμενικό στόχο, ένα εξελιγμένο αρθρωτικό μοντέλο είναι αναμφισβήτητα το σωστό εργαλείο.

Οι επιστήμονες ονειρεύονται μία απλούστερη εξίσωση για να εξηγήσουν όλες τις παρατηρήσεις τους. Ίσως υπάρχει κάποια για να ανακαλυφθεί το βασικό αξίωμα της φυσικής του σύμπαντος. Όσον αφορά στο λόγο, θα πρέπει να δεχθούμε το γεγονός ότι αυτή είναι μία από τις πολύπλοκότερες ανθρώπινες ενέργειες. Ένα απλό, συμμετρικό, όμορφο μοντέλο απλά δεν είναι αρκετό.

## ΒΙΒΛΙΟΦΡΑΦΙΑ

- Abry,C.,& Boë, L.-J.(1986) “Laws” for Lips. *Speech Communication*, 5,97-104.
- Badin, P.,&Fant,G.(1984).Notes on vocal tract computation. *STQL-QPSR*,2-3,53-108.
- Badin, P.(1989).Acoustics of voiceless fricatives: Production theory and data. *Speech Transmission Lab. QPSR*, 3,33-55.
- Beautemps, D.,Badin,P.,& Laboissiere,R.(1993). Recovery of vocal-tract midsagittal and area functions from speech signal for vowels and fricative consonants. In *Eurospeech '93 Proceedings, 3<sup>rd</sup> European Conference on Speech Communication and Technology*, 1, 279-302.Berlin.
- Boë, L.-J., Perrier, P., & Bailly, G. (1992). The geometric vocal tract variables controlled for vowel production: Proposals for constraining acoustic-to-articulatory inversion. *Journal of Phonetics*, 20, 27-38.
- Bothorel, A., Simon, P., Wioland, F., & Zerling, J.-P. (1986). Cinéradiographie des voyelles et des consonnes du français. Recueil de documents synchronisés pour quatre sujets: vues latérales du conduit vocal, vues frontales de l'orifice labial, données acoustiques. Strasbourg, France: Institut de Phonetique.
- Chiba,T.,& Kajiyama,M.(1941).The vowel, its nature and structure. Tokyo: Tokyo-Kaseidan Pub. Co.
- Coker,C.,& Fujimura,O.(1966) Model for the specification of the vocal tract area function. *Journal of the Acoustical Society of America*,40, 1271.
- Fant,G.(1960). *Acoustic theory of speech production*. The Hague: Mouton.
- Fant,G.,& Liljencrants,J.(1985). A four parameter model of glottal flow. *Speech Transmission Lab. QPSR* 2, 18-24.
- Fant,G.(1986). Glottal flow: Models and interaction. *Journal of Phonetics*, 14, 393-399.
- Harshman,R., Ladefoged, P.,& Goldstein, L.(1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62,693-707.

- Heinz, J.M., & Stevens, K.N. (1964). On the derivation of the area functions and acoustic spectra from cineradiographic films of speech. *Journal of the Acoustical Society of America*, 36, 1037-1038.
- Hiki, S., & Oizumi, J. (1974). Speech synthesis by rule from neurophysiological parameters. In *Speech Communication Seminar*. Stockholm, Vol. 1-2, pp. 219-225.
- Ishizaka, K., & Flanagan, J.L. (1972). Synthesis of voiced sounds from a two mass model of the vocal chords, *Bell Systems Tech. Journal*, 51, 1233-1268.
- Kiritani, S., Itoh, S.K., & Fujimura, O. (1975). Tongue-pellet tracking by a computer controlled X-ray microbeam system. *Journal of the Acoustical Society of America*, 57, 1516-1520.
- Kiritani, S., & Sekimoto, S. (1977). Parameter description of the tongue movements in vowel production. In *Modèles articulatoire et Phonétique*. GALF – Actes du Symposium de Grenoble, pp. 202-222.
- Klatt, D. H. (1980). Software for a cascade/parallel formant synthesiser. *Journal of the Acoustical Society of America*, 67, 971-995.
- Ladefoged, P. (1979). Articulatory parameters. *9<sup>th</sup> International Congress of Phonetic Sciences, Copenhagen*, U.C.L.A. WPP 45, 25-52.
- Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America*, 64, 1027-1035.
- Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America*, 50, 1166-1179.
- Maeda, S. (1979). An articulatory model based on statistical analysis. *Journal of the Acoustical Society of America*, 65, S1, S22 (A).
- Maeda, S. (1988). Improved articulatory model. *Journal of the Acoustical Society of America*, 84, Sup. 1, S146.
- Maeda, S. (1989). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W.J. Hardcastle & A. Marchal (Eds.), *Speech Production and Modelling* (pp. 131-149). Kluwer: Academic Publishers.
- Mermelstein, P. (1973). Articulatory model for the study of speech production. *Journal of the Acoustical Society of America*, 53, 1073-1082.
- Öhman, S.E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 41, 310-320.
- Öhman, S.E. (1967). Numerical model for coarticulation. *Journal of the Acoustical Society of America*, 36, 1038.
- Perkell, J.S. (1969). Physiology of speech production: Results and implications of a quantitative cineradiographic study. *Res. Monogr.* 53. Cambridge, MA: M.I.T. Press.
- Perkell, J.S. (1974). *A physiologically oriented model of tongue activity in speech production*. Ph.D. Thesis. M.I.T. Cambridge, MA: M.I.T. Press.
- Perrier, P., Boé, L.-J., & Sock, R. (1992). Vocal tract area functions estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modelling the transition with two sets of coefficients. *Journal of Speech and Hearing Research*, 35, 53-67.
- Shirai, K., & Honda, M. (1977). Estimation of articulatory motion. In Masayuki, Sawashima & Cooper (Eds.), *Dynamic aspects of speech production* (pp. 279-302). Tokyo: University of Tokyo Press.

- Stevens, K.N.(1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.
- Stevens, K.N., & Bickley, C.A. (1986). Effect of vocal tract constriction on the glottal source: Experimental and modelling studies. *Journal of Phonetics*, 14, 373-382.
- Titze, I.R. (1989). A four parameter model of the glottis and vocal fold contact area. *Speech Communication*, 8, 191-201.

anticorrelation	αντισυσχετισμός
area function	συνάρτηση επιφανείας
cineradiographic data	κινηματοραδιογραφικά δεδομένα
correlation	συσχέτιση
fricatives	εξακολουθητικά, τυρβώδη φωνήματα
friction model	μοντέλο τριβής
jaw opening	άνοιγμα σιαγόνας
larynx	λάρυγγας
lip opening	άνοιγμα χειλιών
lip protrusion	προώθηση χειλιών
residuals	υπόλοιπα
sagittal function	τοξοειδής συνάρτηση
source model	μοντέλο πηγής
tongue apex	κορυφή της γλώσσας
tongue body	σώμα της γλώσσας, κορμός της γλώσσας
tongue dorsum	ραχιαίο της γλώσσας

tongue model	μοντέλο γλώσσας
tongue tip	άκρο της γλώσσας
vocal tract	φωνητική οδός

1.1.1.1.1.1.1.1.1 ΑΡΘΡΩΤΙΚΑ ΜΟΝΤΕΛΑ

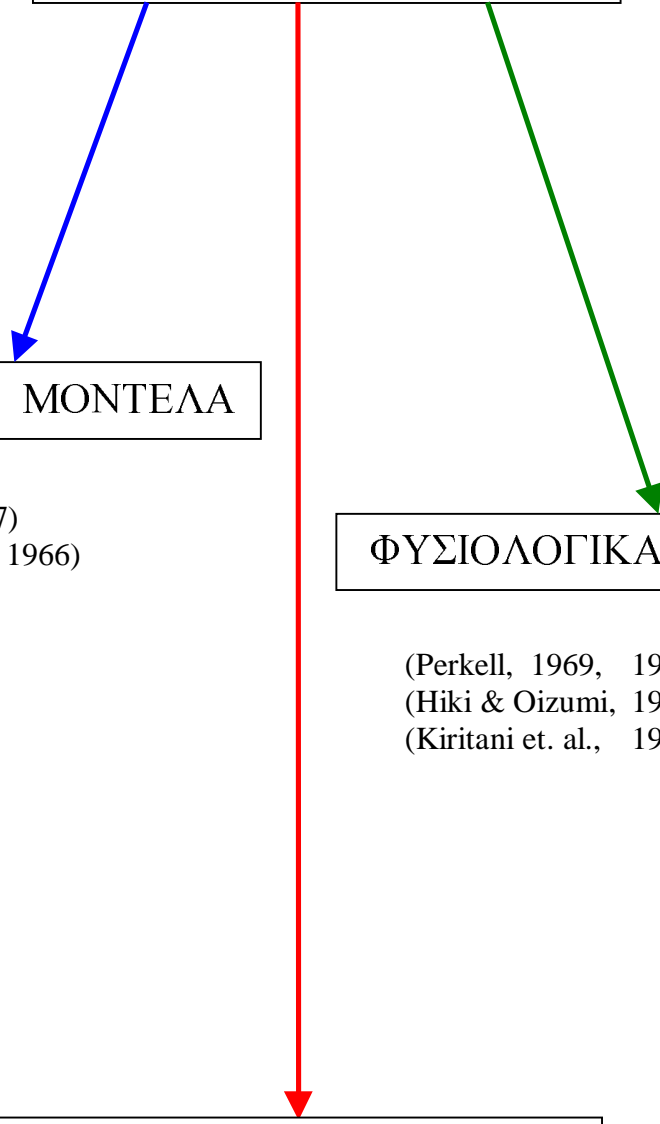
ΓΕΩΜΕΤΡΙΚΑ ΜΟΝΤΕΛΑ

(Öhman, 1966, 1967)  
(Coker & Fujimura, 1966)

ΦΥΣΙΟΛΟΓΙΚΑ ΜΟΝΤΕΛΑ

(Perkell, 1969, 1974)  
(Hiki & Oizumi, 1974)  
(Kiritani et. al., 1975)

ΣΤΑΤΙΣΤΙΚΑ ΜΟΝΤΕΛΑ



(Harshman, Kadefors,  
Goldstein & Rice, 1977)  
(Maeda, 1979, 1988)