



*Μεταπτυχιακό στους
Αλγόριθμους Υψηλής Απόδοσης*

Ρομποτική

*Επικοινωνία ανθρώπου μηχανής
με ομιλία*

Διδάσκων:

Γεώργιος Θ. Κουρουπέτρογλου

**προσομοίωση γυναικείας φωνής στον
συνθέτη ομιλίας
του Τμήματος Πληροφορικής
του Πανεπιστημίου Αθηνών**

Φοιτητές

**Δημήτριος Προβατάς Γιαννακού Σοφία
Α.Μ.: M207 Α.Μ.: M175**

Φεβρουάριος 1999

Εισαγωγή.

Είναι γενικά παραδεκτό πως βασική προϋπόθεση για την κοινωνική συμβίωση οποιασδήποτε ανθρώπινης ομάδας είναι η δυνατότητα επικοινωνίας ανάμεσα στα μέλη της. Στις ανεπτυγμένες κοινωνίες η ουσιαστικότερη επαφή των μελών της πραγματοποιείται άμεσα με τον προφορικό λόγο και έμμεσα με τον γραπτό.

Ο προφορικός λόγος που φορέας του είναι η φωνή, προσφέρει τη δυνατότητα άμεσης επικοινωνίας. Ακούγοντας ένα λόγο μπορούμε να εξάγουμε περισσότερες πληροφορίες όσον αφορά την ταυτότητα του ομιλητή. Είμαστε σε θέση να προσδιορίσουμε τον ομιλητή από την φωνή και την προφορά του, να αναγνωρίσουμε την τοπική του διάλεκτο, τη διάθεσή του και αρκετά άλλα χαρακτηριστικά.

Γενικά, μπορούμε να προσδιορίσουμε το φύλο του ομιλητή, από την φωνή του και την προφορά του. Για παράδειγμα οι γυναίκες γενικώς μιλούν με υψηλότερο pitch από ότι οι άνδρες. Οι διαφορές που αφορούν το ύψος του pitch σχετίζονται με διαφορές μεταξύ των φύλων ως προς την ανατομία και τη φυσιολογία του φωνητικού συστήματος.

Ωστόσο, εκτός από το ύψος του pitch δεν γνωρίζουμε πολλά γύρω από τις φωνητικές διαφορές γυναικών και ανδρών. Η αιτία για το ότι μερικοί άνθρωποι μιλούν πιο γρήγορα ή άλλοι πιο μελωδικά ή δίνουν μεγαλύτερη έμφαση στο λόγο τους από ότι κάποιοι άλλοι, φαίνεται να καθορίζεται από περιβαλλοντολογικούς παράγοντες, παρά από βιολογικούς. Σε πολλές περιπτώσεις αυτό οφείλεται στο ότι προσαρμόζουν τον κοινωνικό τους ρόλο όσον αφορά την ένδυση, τον τρόπο δράσης και ομιλίας τους. Είναι γενικά παραδεκτό πως οι άνδρες και οι γυναίκες παίζουν εν γένει διαφορετικούς ρόλους στην κοινωνία. Τέτοιες επεκτάσεις ή νόρμες επηρεάζουν την παραγωγή ομιλίας και την συμπεριφορά του ατόμου κατά την ομιλία.

Στην παρούσα μελέτη θα αναφερθούμε σε αυτές τις διαφορές ομιλίας ανάμεσα στους άνδρες και τις γυναίκες όσον αφορά τις τιμές των formants και της θεμελιώδους συχνότητας F0.

Φασματικές κορυφές.

Η φυσικότητα ή η ποιότητα ενός φωνήεντος καθορίζεται από το σχήμα της φωνητικής οδού, που αποτελεί συνδυασμό θέσης και σχήματος της γλώσσας και των χειλιών, καθώς και της θέσης του ουρανίσκου. Κάθε φωνήεν είναι δυνατόν να περιγραφεί ως συνδυασμός αρκετών αντηχήσεων ή επιθυμητών συχνοτήτων. Μετρώντας αυτές τις συχνότητες που αναφέρονται ως formants, μπορούμε να περιγράψουμε τον τρόπο με τον οποίο ηχούν τα φωνήεντα. Ο πιο συνηθισμένος τρόπος μελέτης των φωνηέντων είναι να μετρήσουμε τα δύο ή τρία χαμηλότερα formants.

Τα formants δίνουν μία ισχυρή ένδειξη του φυσικού χώρου στο στόμα. Το χαμηλότερο formant δίνει μία ένδειξη του πόσο ψηλά στο στόμα βρίσκεται η γλώσσα κατά τη διάρκεια παραγωγής του φωνήεντος, ενώ το δεύτερο formant δίνει μία ένδειξη για το πόσο εμπρός-πίσω βρίσκεται η γλώσσα.

Στην ομιλία οι αντηχήσεις της φωνητικής οδού ή των formants παράγουν πληροφορίες για τα φωνήματα και για την προσωπικότητα του ομιλητή.

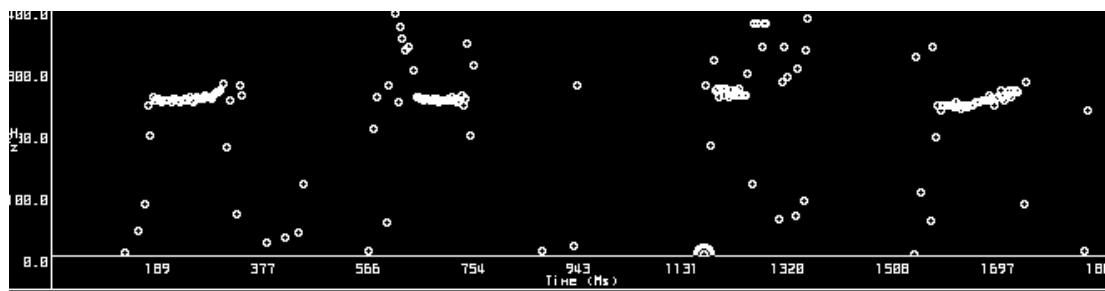
Οι αντιθέσεις των φωνηέντων σηματοδοτούνται πρωταρχικά από τις διαφορές των formants. Οι κορυφές των συχνοτήτων των τριών χαμηλότερων formants και οι μεταβολές τους ανά περίοδο, καθορίζουν την ταυτότητα του φωνήματος. Η παρουσίαση των formants μας δίνει μία συμπαγή περιγραφή του φασματογραφήματος ομιλίας. Δοθέντος ενός αρχικού συνόλου υποθέσεων για τη γλωττιδική πηγή για να καθορίσουμε τα εύρη ζώνης των formants, η φασματική περιβάλλουσα μπορεί να περιγραφεί από τις συχνότητες των φασματικών κορυφών (Fant 1960). Μία αλλαγή στη συχνότητα ενός formant οδηγεί σε αλλαγές του φασματογραφήματος.

Οι ακροατές αρχικά δίνουν μεγαλύτερη προσοχή στις φασματικές κορυφές για να διακρίνουν την ποιότητα διαφορετικών φωνηέντων. Ένας λόγος για τον οποίο οι φασματικές κορυφές είναι σημαντικές στην ομιλία είναι ότι μπορούμε να μελετήσουμε ή να αποφανθούμε σχετικά με τον θόρυβο που ενδεχομένως βρίσκεται σε αυτές.

Μία πρώτη επισήμανση στην ομιλία ανδρών-γυναικών.

- *Η θεμελιώδης συχνότητα.*

Η ουσιώδης διαφορά στην ομιλία ανδρών και γυναικών έγκειται στη θεμελιώδη συχνότητα. Από ακουστικές μελέτες που έχουν γίνει αναφέρεται ότι το pitch είναι ο σπουδαιότερος παράγων για να διακρίνουμε τον λόγο μεταξύ ανδρών και γυναικών. Έχει επίσης αναφερθεί πως η μελωδικότητα που παρουσιάζει κύρια η γυναικεία φωνή, συνεπάγεται μεγάλες εναλλαγές στη θεμελιώδη συχνότητα. Στο παρακάτω σχήμα είναι δυνατόν να παρατηρηθεί αυτή η διακύμανση.



Ακουστικές αναλύσεις που έγιναν πάνω σε δείγματα ανδρικών και γυναικείων ομάδων, επισημαίνουν μία διαφορά στη θεμελιώδη συχνότητα ± 120 Hz, έναντι ± 200 Hz. Στην ανάλυση αυτή δεν βρέθηκαν σημαντικές διαφορές στη θεμελιώδη συχνότητα μεταξύ ατόμων με διαφορετικό μορφωτικό επίπεδο ή διαφορετικό επάγγελμα.

Βέβαια η θεμελιώδης συχνότητα δεν είναι ικανή από μόνη της να μας δώσει πληροφορίες για το φύλο του ομιλητή. Είναι λοιπόν σαφές ότι δεν αρκεί να αυξήσουμε τη θεμελιώδη συχνότητα για να προσομοιώσουμε τη γυναικεία φωνή.

Αρχικά, κοιτάζοντας την πηγή διέγερσης υποθέτουμε ότι μία πηγή που προσομοιώνει πιο στενά τον γλωττιδικό παλμό θα συνεισέφερε στην προσομοίωση γυναικείας φωνής, αλλά δεν θεωρείται και αυτή από μόνη της σημαντική διαφορά.

Ειδικότερα, στην ανθρώπινη ομιλία υπάρχουν σημαντικές μεταβολές στις τιμές των amplitudes και formants 4,5,6, το οποίο παίζει σημαντικό ρόλο στο να καθορίσουμε τα χαρακτηριστικά των διαφορετικών ανθρωπίνων φωνών, συμπεριλαμβανομένης και της γυναικείας φωνής.

- *Ta Formants.*

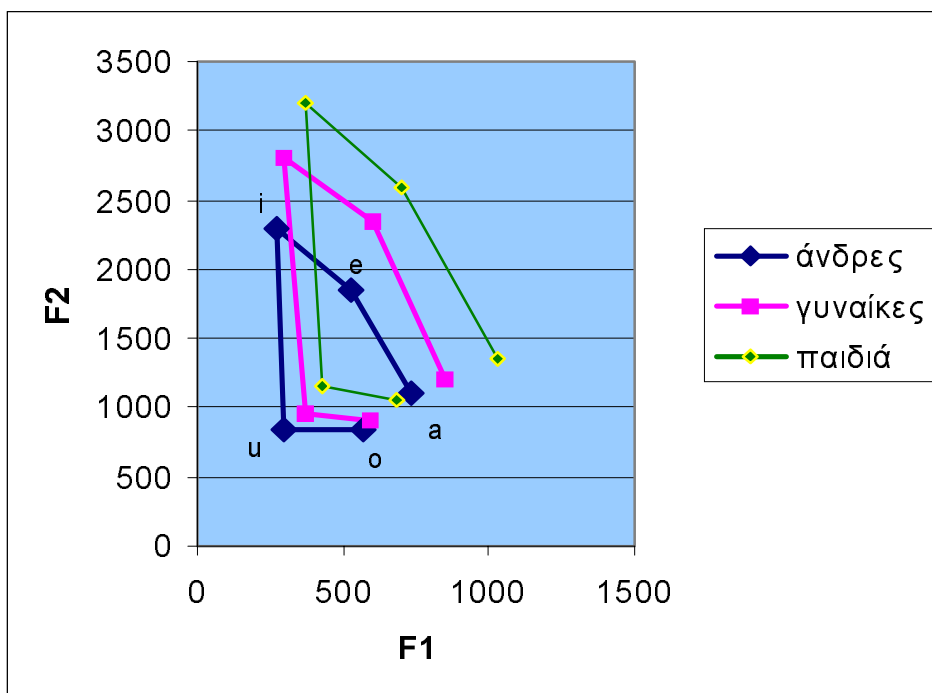
Ως γνωστόν, η μέση διάσταση της γυναικείας φωνητικής οδού είναι μικρότερη από αυτή του ανδρός. Αυτό όχι μόνο αυξάνει τις συχνότητες των formants στα φωνήεντα και στους ήχους που προσομοιάζουν με φωνήεντα, αλλά και αλλάζει το φασματογράφημα των περιοδικών ήχων που παράγονται στην στοματική κοιλότητα. Σαν γενικό συμπέρασμα, μπορεί να θεωρηθεί ότι για να υλοποιήσουμε τη γυναικεία φωνή μπορούμε να συμπεριλάβουμε στο σχεδιασμό ενός συστήματος συνθέτη ομιλίας μία δεύτερη περιοδική πηγή διέγερσης.

Γενικά και στους άνδρες αλλά και στις γυναίκες, τα φωνήεντα παρουσιάζονται αμετάβλητα κατά τη διάρκεια της εκφώνησης και ορίζονται από ένα στατικό σχήμα της φωνητικής οδού και μπορούν έτσι να παρασταθούν με ένα σημείο στο επίπεδο F1-F2. Η παραδοχή βέβαια αυτή που κάναμε, ότι δηλαδή τα φωνήεντα παρουσιάζονται αμετάβλητα κατά τη διάρκεια εκφώνησής τους, είναι σχετική. Αλλά σε αυτό θα αναφερθούμε αργότερα.

Στον κάτωθι πίνακα παρατηρούμε τις F1 F2 και F3 συχνότητες για φωνήεντα που παρήχθησαν από ένα δείγμα 76 ομιλητών, συμπεριλαμβανομένων ανδρών, γυναικών και παιδιών.

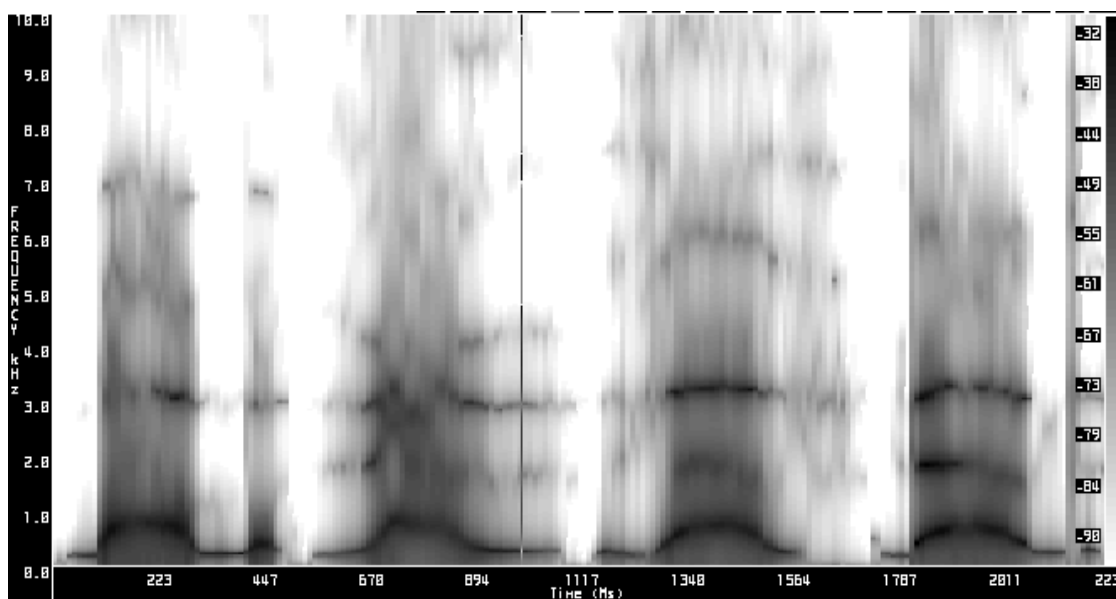
Φώνημα	άνδρες			γυναίκες			Παιδιά		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
a	730	1100	2450	850	1200	2800	1030	1350	3200
e	530	1850	2500	600	2350	3000	700	2600	3550
i	270	2300	3000	300	2800	3300	370	3200	3700
u	300	850	2250	370	950	2650	430	1150	3250
o	570	850	2400	590	900	2700	680	1050	3200

Τα αποτελέσματα τα παρατηρούμε στο F1-F2 επίπεδο όπου από αυτό μπορούμε να διαπιστώσουμε πως οι τιμές για τις γυναίκες είναι σαφώς υψηλότερες από ότι στους άνδρες, αλλά και αυτές των παιδιών υψηλότερες από ότι των γυναικών. Αυτό οφείλεται όπως προείπαμε στο μικρότερο σχήμα της φωνητικής οδού. Όσο μικρότερη είναι, τόσο και υψηλότερες συχνότητες παρουσιάζονται.



Ένα άλλο φαινόμενο, βέβαια, που παρουσιάζεται και στους άνδρες αλλά και στις γυναίκες είναι το φαινόμενο του target undershoot. Στο φαινόμενο αυτό παρουσιάζονται για το ίδιο φώνημα διαφορετικές τιμές στο F1-F2 επίπεδο, όταν αυτό το φώνημα εκφωνηθεί μόνο του ή σε CVC συλλαβή, αλλά και σε διαφορετικές συλλαβές που περιέχουν το ίδιο φώνημα.

Παρατηρούμε στο κάτωθι σχήμα το φώνημα «α» όταν εκφωνείται από ελληνίδα γυναίκα ομιλήτρια σε CVC μορφή, όπως «μπαμπ», «γαγγ», «θαθ» και «νταντ».



- *Πώς μετράμε τα formants;*

Για την εξαγωγή των formants θεωρήσαμε πως αυτά παραμένουν σταθερά κατά τη διάρκεια της εκφώνησης του φωνήματος. Ένα όμως φωνήεν είναι σπάνια στατικό και αμετάβλητο. Η γλώσσα αλλά και τα χείλη κινούνται κατά την παραγωγή ενός φωνήεντος παράγοντας έναν ήχο δυναμικής ποιότητας.

Σε αυτό το σημείο μία κρίσιμη ερώτηση που προκύπτει κατά τη μελέτη των φωνηέντων είναι το πρόβλημα: «πού μέσα στο χρόνο μετράμε τα formants του φωνήεντος; την αρχή; τη μέση; ή το τέλος;».

Ένας αριθμός από μελέτες έδειξε πως η ακουστική φύση των φωνηέντων και ο τρόπος με τον οποίο οι ακροατές τα αντιλαμβάνονται μπορεί να περιγραφεί προσπαθώντας να χαρακτηρίσουμε αυτόν τον δυναμισμό **μετρώντας τα formants σε περισσότερες από μία χρονικές στιγμές του φωνήεντος.**

Σε μελέτη που διεξήγαγαν οι Marie Huffman και Elyse Tamberino του Τμήματος Γλωσσολογίας του Πανεπιστημίου της Νέας Υόρκης, διαπίστωσαν ότι πολλοί αγγλόφωνοι Αμερικάνοι πρόφεραν 5 διακριτά μπροστινά φωνήεντα σε φωνητικά περιβάλλοντα όπως 'beet', 'bit', 'bait', 'bet', και 'bat', ενώ ομιλητές του Long Island και σε άλλες περιοχές παράγουν ένα επιπλέον φωνήεν με τη θέση της σιαγόνας και της γλώσσας υψηλότερα όταν προφέρουν το φωνήεν στο φωνητικό περιβάλλον 'bat'.

Έτσι ενώ πολλοί αγγλόφωνοι ομιλητές παρουσίασαν την ίδια ποιότητα φωνήεντος κατά την εκφώνηση του τριφώνου 'bat' και 'bad', πολλοί Άγγλοι ομιλητές παρουσίασαν διαφορετική ακουστική ποιότητα σε αυτές τις λέξεις επειδή τα φωνήεντα εκφωνούνται στο τρίφωνο 'bad' υψηλότερα. Εάν οι ομιλητές του Long Island έχουν ένα μεγαλύτερο αριθμό φωνηέντων είναι πραγματικά δύσκολο να εντοπίσουμε τις ακουστικές διαφορές.

Τότε πως οι ομιλητές του Long Island διατηρούν την ευδιακριτότητα των φωνηέντων τους; Ποια είναι τα ακουστικά χαρακτηριστικά των φωνηέντων ούτως ώστε οι ακροατές να μπορούν να τα ξεχωρίζουν; Σε αυτή την περίπτωση υπέθεσαν μια δυναμική ποιότητα των φωνηέντων τα οποία παράγονται στο ίδιο τμήμα της στοματικής κοιλότητας, δηλαδή μέσω αλλαγής των formants επιτυγχάνουμε το ευδιάκριτο άκουσμα τους. Κατέγραψαν έτσι ομιλητές από το Long Island που μιλούσαν μόνο μία γλώσσα και που εκφωνούσαν λέξεις που περιείχαν τους ίδιους ήχους συμφώνων αλλά διαφορετικά φωνήεντα.

Προσδιόρισαν και μέτρησαν τα δύο πρώτα formants σε αυτές τις λέξεις. Η χρονική στιγμή μέτρησης για κάθε φωνήεν ήταν αντίστοιχα στο 20%, στο 50% και στο 80% της διάρκειας του φωνήματος. **Υπολόγισαν πως οι τιμές των formants των φωνηέντων αλλάζουν στις λέξεις κατά διάφορες χρονικές στιγμές μέτρησης.**

Το γενικό συμπέρασμα από αυτές τις μετρήσεις ήταν ότι λαμβάνοντας υπόψιν τις τιμές των formants που υπολογίστηκαν στο 50% της διάρκειας του φωνήεντος, τα φωνήεντα δεν παρουσίαζαν καμμία διαφορά μεταξύ τους. Ωστόσο εάν λάβουμε υπόψιν τις τιμές των formants στο 20% σημείο και στο 80% , τότε μπορούμε σαφώς να καθορίσουμε την λέξη που ακούστηκε.

Στη μελέτη αυτή των Marie Huffman και Elyse Tamberino αναφέρεται πως τα φωνήεντα στη γυναικεία φωνή σαφώς γίνονται πιο ευκρινή όταν οι τιμές των formants υπολογιστούν από 11 τέτοια σημεία.

Πάντως οι Fry, Abramson, Eimas και Liberman υποστήριξαν ότι όταν χρησιμοποιούμε τα formants για τη σύνθεση των φωνηέντων, τα αποτελέσματα είναι σε γενικές γραμμές ικανοποιητικά. Επιπλέον οι Rakerd και Verbrugge ισχυρίζονται πως υπάρχει συσχέτιση μεταξύ των τιμών των formants και τη θέση της γλώσσας.

Συγκεκριμένα, το ύψος της γλώσσας καθορίζει την τιμή του F1, ενώ η πίσω-εμπρός θέση της καθορίζει την τιμή του F2. Γενικά, τα χαμηλά φωνήεντα παρουσιάζουν υψηλό F1, ενώ τα υψηλά φωνήεντα χαμηλό F1. Τα πίσω φωνήεντα έχουν ένα χαμηλό F2 και μικρή διαφορά F2-F1, ενώ τα μπροστινά φωνήεντα έχουν υψηλότερο F2 και μεγάλη διαφορά F2-F1.

- *Τα εύρη ζώνης των formants.*

Τα εύρη ζώνης των formants αυξάνουν ανάλογα με τον αριθμό του formant. Πειραματικές μελέτες έχουν δείξει ότι μεταβάλλοντας το εύρος ζώνης των formants, υπεισέρχεται μικρή μεταβολή στην αντίληψη ενός φωνήεντος. Αυτό ισχύει διότι η ανθρώπινη ακοή δεν είναι ευαίσθητη σε τέτοιες μεταβολές. Αλλά και όταν ακόμη η ελάττωση του εύρους ζώνης γίνει αισθητή, ο ακροατής μπορεί ακόμα και τότε να προσδιορίσει το φωνήεν το οποίο ακούει. Η μόνη διαφορά έγκειται στο γεγονός πως ασυνήθιστα στενό εύρος ζώνης συνεπάγεται τεχνητό άκουσμα του φωνήεντος

Από την άλλη αυξάνοντας υπερβολικά τα εύρη ζώνης των formants, ελαχιστοποιείται η ευδιακριτότητα των φωνηέντων, λόγω των επικαλύψεων των formant.

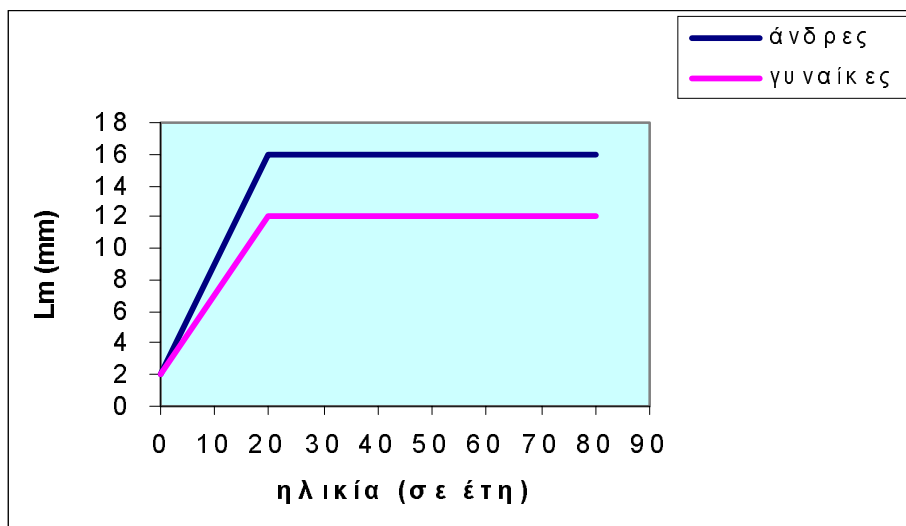
Οι Dunn, White και Fant, υποστήριξαν πως τα δύο πρώτα formants έχουν εύρη ζώνης της τάξης των 30-100 c/s, ενώ τα F3 και F4 παρουσιάζουν εύρη ζώνης της τάξεως των 40-200 c/s.

Ανέφεραν έτσι εύρη ζώνης μικρότερα από αυτά στα οποία κατέληξε ο Bogert. Πάντως, πρόσφατες μελέτες των House και Stevens έδειξαν εύρη εύρος ζώνης που έρχονται πιο κοντά στα αποτελέσματα του Van de Berg.

Κύριες διαφορές στην ομιλία ανδρών-γυναίκων.

Τα ακουστικά χαρακτηριστικά των φωνηέντων διαφέρουν σε άνδρες, γυναίκες και παιδιά και είναι φανερό ότι εξαρτώνται από την ηλικία και το φύλο του ομιλητή. Όπως προαναφέραμε η φωνή των γυναικών παρουσιάζει υψηλότερο pitch κατά μία οκτάβα έναντι των ανδρών ή περίπου 1.7 φορές υψηλότερο των ανδρών.

Αυτή η διαφορά στη θεμελιώδη συχνότητα σχετίζεται με το μέγεθος των φωνητικών χορδών (Titze 1989). Το κάτωθι σχήμα δείχνει το μέγεθος της γλωττίδας ανάλογα με την ηλικία και το φύλο.



Ο Titze βασισμένος στο μέγεθος της μεμβράνης των φωνητικών χορδών L_m , προκειμένου να υπολογίσει τις διαφορές στη μέση θεμελιώδη συχνότητα, τη μέση εκροή του αέρα και την αεροδυναμική ισχύ, χρησιμοποιεί έναν συντελεστή κλίμακος, ο οποίος υπολογίζεται ότι είναι περίπου 1.6. Επιπλέον πρέπει να προστεθεί και ένας συντελεστής κλίμακας της τάξης του 1.2, ο οποίος βασίζεται στο πλάτος και είναι απαραίτητος για να υπολογίσει τις διαφορές ισχύος μεταξύ των ανδρικών και γυναικείων φωνών.

Όπως αναφέραμε και στην εισαγωγή, οι διαφορές μεταξύ γυναικών και ανδρών βασίζονται και σε άλλους παράγοντες. Κάποιοι από αυτούς όσο αναφορά τη φωνή των γυναικών, μπορεί να θεωρηθεί ότι είναι

- ψιθυριστή
- αδύναμη
- γλωττιδική διαρροή (που σημαίνει ότι διαφεύγει αέρας από τη γλωττίδα ακόμα και κατά τη διάρκεια της κλειστής φάσης).
- περισσότεροι συμμετρικοί φωνητικοί παλμοί
- μικρότερης διάρκειας παλμοί
- υψηλότερη θεμελιώδη συχνότητα
- διαφορετικό εύρος θεμελιώδους συχνότητας
- περισσότερος θόρυβος στις περιοχές μεταξύ των formants
- υψηλότερες συχνότητες των formants
- μεγαλύτερα εύρη ζώνης των formants
- διαφορετική σύζευξη ή αλληλεπίδραση μεταξύ υπογλωττιδικών και υπεργλωττιδικών κοιλοτήτων
- μεγαλύτερη αλληλεπίδραση μεταξύ πηγής και φίλτρου

Τα χαρακτηριστικά αυτά δεν είναι ανεξάρτητα μεταξύ τους, αλλά συσχετισμένα. Έτσι εάν προσπαθήσουμε να παράγουμε τη γυναικεία φωνή από την ανδρική αυξάνοντας απλά και μόνο τις τιμές της θεμελιώδους και των υπόλοιπων συχνοτήτων, δεν θα έχουμε την αναμενόμενη επιτυχία και η φωνή δεν θα ηχεί γυναικεία.

Εξάλλου, η υψηλότερη θεμελιώδης συχνότητα που παρουσιάζουν οι γυναίκες παρουσιάζει δυσκολίες στην ακουστική ανάλυση. Όσο αυξάνεται η θεμελιώδης συχνότητα, τόσο αυξάνονται τα αντίστοιχα διαστήματα μεταξύ των αρμονικών στο φασματογράφημα λαρυγγικής πηγής. Σε μερικά διαστήματα αρμονικών δύσκολα διακρίνουμε τη θέση των formants στο φασματογράφημα. Το πρόβλημα ουσιαστικά είναι πρόβλημα δειγματοληψίας, καθότι πλατειά διεσπαρμένες αρμονικές δεν δίνουν την κατάλληλη πληροφορία για την φασματική περιβάλλουσα.

Τα πρώτα φασματογραφήματα περιοριζόντουσαν στην ανάλυση ομιλίας γυναικών με υψηλό pitch λόγω του ότι ήταν εφοδιασμένα με ένα σταθερό φίλτρο 300Hz για ανάλυση ευρείας ζώνης. Αυτό το φίλτρο ήταν αποτελεσματικό για τις περισσότερες ανδρικές φωνές επειδή περιέκλειε τουλάχιστον δύο αρμονικές και επομένως ανέλυε περισσότερο formants παρά αρμονικές. Βέβαια για πολλές γυναικείες φωνές το εύρος του φίλτρου αντιστοιχούσε σε διάστημα μιάς αρμονικής. Σαν αποτέλεσμα τα φασματογραφήματα έδειχναν μία αλληλεπίδραση αρμονικής και formant.

Αυτό καθιστούσε δύσκολο την αναγνώριση μίας ζώνης στο εάν πρόκειται για αρμονική ή για formant. Μία λύση ήταν να αυξηθεί το εύρος ζώνης ανάλυσης. **Σαν γενικός κανόνας το εύρος ζώνης του φίλτρου ανάλυσης θα έπρεπε να είναι δύο ή τρεις φορές μεγαλύτερο από τη θεμελιώδη συχνότητα του ομιλητή.** Για παράδειγμα το εύρος ζώνης για μία γυναίκα που έχει θεμελιώδη συχνότητα 300Hz, θα έπρεπε να είναι 600 Hz.

Υπάρχουν άνω όρια στο μέγεθος του φίλτρου που ανατρέπουν το σκοπό της ακουστικής ανάλυσης. Για παράδειγμα ένα φίλτρο πλάτους 1000Hz θα περιέβαλλε όχι μόνο αρμονικές αλλά και κοντινά μεταξύ τους formants. Μία προσέγγιση για να αναλύσουμε τον γυναικείο και παιδικό λόγο με σταθερού εύρους φίλτρα ανάλυσης είναι η μαγνητική αναπαραγωγή του σήματος ομιλίας με μικρότερη ταχύτητα από την ταχύτητα που χρησιμοποιήθηκε για την εγγραφή. Όταν το πιο αργό σήμα εμφανιζόταν στο φασματογράφημα, το αποτέλεσμα ήταν να αλλάζει το εύρος του φίλτρου αναλογικά με την ταχύτητα εγγραφής-αναπαραγωγής.

Ένα άλλο άρθρο αναφέρει πως η τιμή της συχνότητας της γυναίκας πρέπει να είναι 20% υψηλότερη από αυτή των ανδρών. Είναι βέβαια αυτός ένας καθαρά εμπειρικός κανόνας ο οποίος πρέπει να λαμβάνεται υπόψιν για φωνήματα που παρουσιάζουν κυρίως υψηλές συχνότητες.

Για παράδειγμα ένα εύρος συχνότητας 8kHz το οποίο θεωρείται ικανοποιητικό για την ανάλυση της τυρβώδους ενέργειας για τους άνδρες, ίσως να μην είναι επαρκές για τις γυναίκες.

Ένας αριθμός μελετών οδήγησε στο συμπέρασμα πως οι γυναικείες φωνές διαφέρουν από τις ανδρικές ως προς άλλες διαστάσεις παρά ως προς τη θεμελιώδη συχνότητα. Αυτές οι διαστάσεις είναι σχετικές με τη βέλτιστη ανάλυση της γυναικείας ομιλίας. Ένα χαρακτηριστικό γνώρισμα της γυναικείας φωνής είναι ότι είναι πιο ψιθυριστές από ότι των ανδρών.

Για παράδειγμα οι Henton και Bladon, ισχυρίστηκαν ότι για τους Βρετανούς ομιλητές το πλάτος της πρώτης αρμονικής σε σχέση με το πλάτος της δεύτερης αρμονικής ήταν περίπου 6dB ισχυρότερο στις γυναίκες από ότι στους άνδρες. Οι Klatt και Klatt ανέφεραν μία ανάλογη διαφορά ανάμεσα στους αγγλόφωνους αμερικανούς άνδρες και γυναίκες, αλλά παρατήρησαν ότι υπάρχει μία σημαντική διαφορά. Από στροβοσκοπικές μελέτες του λάρυγγα που διεξήγαγαν οι Bless, Biever και Shaikh κατέληξαν στο συμπέρασμα πως οι γυναίκες που έχουν μπροστινή γλωττιδική σχισμή κατά τη διάρκεια της κλειστής περιόδου του κύκλου είναι 4πλάσιες των ανδρών.

Οι Klatt και Klatt ανακάλυψαν στις γυναικείες φωνές μία τάση για μεγαλύτερη διέγερση στο F3 formant εξαιτίας δασώδους θορύβου (κατά την εκπνοή, εφόσον η γυναικεία φωνή είναι πιο ψιθυριστή). Συμπέραναν επίσης, ότι το γλωττιδικό άνοιγμα των ψιθυριστών φωνών προκαλεί μία αύξηση του εύρους του πρώτου formant, που μερικές φορές εξαφανίζει τελείως τη φασματική κορυφή του F1.

- *Χαρακτηριστικά ομιλίας των παιδιών.*

Σε αντιστοιχία με τις διαφορές που παρουσιάζει η γυναικεία με την ανδρική φωνή, είναι δυνατόν να γίνουν και κάποιες επισημάνσεις σχετικά με τα χαρακτηριστικά ομιλίας των παιδιών. Επειδή τα παιδιά έχουν μικρότερου μήκους φωνητική οδό και κοντύτερες φωνητικές χορδές, παρουσιάζουν υψηλότερη τη θεμελιώδη και τις υπόλοιπες συχνότητες. Εκείνο που επιπλέον ισχύει είναι ότι τα παιδιά έχουν τα δικά τους χαρακτηριστικά ομιλίας. Η θεμελιώδης συχνότητα επί παραδείγματι στα παιδιά μειώνεται κατά μία οκτάβα. Σύμφωνα με μελέτες τα νήπια έχουν τις υψηλότερες τιμές στη θεμελιώδη και τις λοιπές συχνότητες, δεδομένου ότι έχουν τη μικρότερη φωνητική οδό και τις κοντύτερες φωνητικές χορδές.

Ακουστική έρευνα που έγινε το 1982 από τους Kent και Murray, έδειξε ότι για τα νήπια οι τιμές των formants για ένα ημικεντρικό φωνήεν είναι :

- θεμελιώδης συχνότητα = 400 Hz.
- F1 = 1000 Hz
- F2 = 3000 Hz
- F3 = 5000 Hz

Η θεμελιώδης περίοδος στα 400 Hz δεν πρέπει βέβαια να ληφθεί αυστηρά υπόψιν μιάς και τα νήπια παρουσιάζουν μεγάλες διακυμάνσεις στη θεμελιώδη συχνότητα. Οι μικρότερες τιμές προσεγγίζουν τη θεμελιώδη συχνότητα των ανδρών ενώ οι μέγιστες επεκτείνονται στα 1000Hz ή και περισσότερο. Τα νήπια επίσης μπορούν να παράγουν πολλούς φωνητικούς τύπους και συνεπώς παρουσιάζουν μεγάλες διαφορές στα φωνητικά χαρακτηριστικά τους.

Οι Bauer και Kent ανέφεραν ότι το φάσμα ενέργειας για τα τυρβώδη στα νήπια φτάνει τα 8kHz που είναι το άνω όριο συχνότητας ενός συμβατικού φασματογραφήματος. Καθώς τα παιδιά μεγαλώνουν οι φωνητικές χορδές τους μακραίνουν, με αποτέλεσμα οι συχνότητες των formants να ελαττώνονται. Αυτό συμβαίνει επειδή μεταβάλλονται τα χαρακτηριστικά του προσώπου τους με γρηγορότερο ρυθμό από ότι σε μεγάλη ηλικία.

Η ηλικία στην οποία τα νήπια αρχίζουν να παράγουν λέξεις ποικίλει μεταξύ των παιδιών, αλλά η κανονικοποίηση φράσεων εμφανίζεται μεταξύ 6 και 15 μηνών. Σε σχέση με τους ενήλικες τα παιδιά έχουν αργότερους ρυθμούς ομιλίας και μεγαλύτερη μεταβλητότητα όταν προφέρουν κάτι. Καθώς αποκτούν επιδεξιότητα η εκτέλεση γίνεται γρηγορότερη και κατανοητή.

- **Ομιλητές άνδρες και γυναίκες διαφόρων γλωσσών.**

Τα φωνήεντα τα οποία εκφωνούνται από διαφορετικούς ομιλητές έχουν διαφορετικά formants. Εάν κάποιος θέλει να συγκρίνει την ποιότητα των φωνηέντων που εκφωνούνται από φυσικούς ομιλητές διαφόρων γλωσσών μετρώντας τις συχνότητες των formants των φωνηέντων θα προκύψουν οι ακόλουθες ερωτήσεις:

- Ποιές ακουστικές μετρήσεις αντανakλούν γνήσιες φωνητικές διαφορές μεταξύ των δύο γλωσσών;
- Ποιές απόψεις αντανakλούν σημασιολογικές διαφορές οι οποίες μπορούν να ελλατωθούν;
- Και εφόσον οι σημασιολογικές διαφορές ελλατωθούν ποιές είναι θεωρητικά οι διαφορές που παραμένουν;

Η αδυναμία να απαντήσουμε στις δύο πρώτες ερωτήσεις καθιστά αδύνατον να συγκρίνουμε τα φωνητικά χαρακτηριστικά των φωνηέντων μεταξύ δύο ομιλητών διαφορετικών γλωσσών.

Η απάντηση αυτών των ερωτήσεων απαιτεί να διαγνώσουμε τις πηγές διαφοροποίησης του ομιλητή. Η διαφοροποίηση του ομιλητή έγκειται κατά πρώτον σε γλωσσολογικούς παράγοντες όπως είναι η διάλεκτος και οι κοινωνικές διαφορές και κατά δεύτερον σε μη γλωσσολογικούς παράγοντες όπως είναι η φυσική ανατομία, η ηλικία και συναισθηματική κατάσταση του ομιλητή.

Οι επιδράσεις των μη γλωσσολογικών χαρακτηριστικών στην ομιλία μπορούν να εξαλειφθούν με κατάλληλες μετατροπές.

Σκοπός της εξάλειψης των μη γλωσσολογικών παραγόντων από την ομιλία είναι ο ακουστικός προσδιορισμός ενός φωνήεντος σε μία δεδομένη γλώσσα. Η διαδικασία αυτή ονομάζεται κανονικοποίηση (normalization).

Οι προτάσεις για κανονικοποίηση της ομιλίας είναι ακουστικού και αρθρωτού τύπου. Οι δύο μέθοδοι επικεντρώνονται στην μετατροπή των δεδομένων χρησιμοποιώντας λογαριθμική κλίμακα του εύρους ζώνης (κλίμακα Bark, Koenig). Μέσω κατάλληλων μετατροπών που επιτυγχάνονται με την κανονικοποίηση είναι δυνατόν να κατορθώσουμε να έχουμε τις γενικευμένες τιμές των formants, έχοντας εξαλείψει τις παραπάνω διαφορές.

Επιδιώκοντας μια γλωσσολογική σύγκριση των φωνηέντων θα αναφέρουμε τρεις μη γλωσσολογικές διαφορές που παίζουν ρόλο σε αυτή την σύγκριση, όπως είναι η θεμελιώδης συχνότητα, το μήκος της φωνητικής οδού και η αναλογία 'μπρος κοιλότητα' προς 'πίσω κοιλότητα'.

Κατά πρώτον, η αναλογία δόνησης της φωνητικής οδού (ακουστικά F0 εύρος) είναι αντιστρόφως ανάλογη με την μάζα και το μέγεθος της φωνητικής οδού και ανάλογη με το τέντωμα των χορδών.

Ο Negus αναφέρει ότι το μήκος της φωνητικής χορδής κυμαίνεται κατά μέσον όρο από 12.5 μέχρι 17 mm στις ενήλικες γυναίκες και από 17 μέχρι 23 mm στους άνδρες.

Ο μέσος όρος παλμικής δόνησης των ανδρικών χορδών είναι περίπου 125 Hz ενώ των γυναικείων κυμαίνεται περίπου στα 200 Hz.

Επίσης, η θεμελιώδης συχνότητα μεταβάλλεται καθώς μεταβάλλονται αφενός το τέντωμα των λαρυγγικών μυών του ομιλητή και αφετέρου ως κάποιο βσθμό η υπογλωτιδική πίεση.

Ο Boothroyd (1986) παρατήρησε ότι η θεμελιώδης συχνότητα ποικίλει μεταξύ

- 70 και 200 Hz για τους άνδρες,
- 140-400 Hz για τις γυναίκες και
- 180-500 Hz για τα παιδιά,

τιμές που συμφωνούν σε γενικές γραμμές με όσα προείπαμε παραπάνω για τους άνδρες, τις γυναίκες και τα παιδιά.

Επιπλέον, όπως έχουμε αναφέρει, **οι συχνότητες των formants σχετίζονται με το μήκος της φωνητικής οδού του ομιλητή το οποίο ποικίλλει ανάλογα με την ηλικία και το φύλο.**

Το συνολικό μήκος της φωνητικής οδού μπορεί να υπολογισθεί από μετρήσεις των formants.

Χρησιμοποιώντας την εξίσωση του Fant

$$L = \frac{5 \cdot C}{4 \cdot F_3} \quad (1)$$

όπου $C=34,000\text{cm/sec}$, (δηλαδή την ταχύτητα ήχου του αέρα),

οι αναλογίες της φωνητικής οδού γυναίκας προς άνδρα σε τρεις Ευρωπαϊκές γλώσσες βρέθηκαν να είναι 0,89 για Σουηδούς, 0.89 για Δανούς και 0.86 για Άγγλους.

Αυτές οι μετρήσεις έδειξαν ότι η γυναικεία φωνητική οδός είναι 11-14% κοντύτερη από αυτήν του ανδρός.

Οι Nordstrom και Lindlom (1975) πρότειναν μια μέθοδο για την κανονικοποίηση του φύλου. Η μέθοδός τους βασιζόταν στον υπολογισμό του συνολικού μήκους της φωνητικής οδού του ατόμου από τον μέσο όρο του F3 στα φωνήεντα, με την προϋπόθεση το F1 να είναι μεγαλύτερο των 600 Hz.

Τότε ο λόγος k που δίνει την αναλογία της μέσης ανδρικής φωνητικής οδού (L_m) προς την μέση γυναικεία φωνητική οδό (L_f) καθορίζεται από:

$$k = \frac{F3_{m.av}}{F3_{f.av}} \quad (2)$$

όπου $F3_{m.av}$, $F3_{f.av}$ δείχνουν το μέσο όρο του τρίτου formant στους άνδρες και στις γυναίκες αντίστοιχα.

Η συχνότητα του v -οστού γυναικείου formant δίνεται από τον τύπο :

$$F_{nf.sc} = k \cdot F_{nf} \quad (3)$$

Τελικά η μέθοδος των Nordstrom και Lindblom καθίσταται μία από τις απλούστερες διαδικασίες που παρέχουν ικανοποιητικό αποτέλεσμα κανονικοποίησης.

Ένας τρίτος μη γλωσσολογικός παράγοντας που διαφοροποιεί την ομιλία είναι η αναλογία μήκους φάρυγγα προς στοματική κοιλότητα. Παρατηρήθηκε ότι το μήκος της στοματικής κοιλότητας ενός οκτάχρονου κοριτσιού ήταν 30% κοντύτερο από αυτό ενός άνδρα ενήλικα, ενώ το μήκος του φάρυγγα του κοριτσιού ήταν 56% κοντύτερο απ αυτό του άνδρα.

Το μήκος του φάρυγγα και της στοματικής κοιλότητας μπορεί να υπολογισθεί από τις συχνότητες του φωνήεντος [i]. Σε ένα απλοποιημένο μοντέλο για το [i] το F2 εξαρτάται από την πίσω κοιλότητα ή τον φάρυγγα και το F3 από την μπροστινή ή από την στοματική κοιλότητα.

Το μήκος της πίσω κοιλότητας (LB) και της μπροστινής (LF) δίνεται προσεγγιστικά ως εξής

$$LB = \frac{C}{2 \cdot F2} \quad , C = \text{ταχύτητα ήχου} \quad (4)$$

$$LF = \frac{C}{2 \cdot F3} \quad , C = \text{ταχύτητα ήχου} \quad (5)$$

Για τους Σουηδούς ομιλητές, σύμφωνα με τους παραπάνω τύπους, ο Fant, ανέφερε, ότι οι γυναικείες πρόσθιες και πίσω κοιλότητες ήταν αντίστοιχα 1.2cm και 1.25cm κοντύτερες των ανδρικών .

Από την μελέτη του Byunggon Yang που αφορά την εύρεση των formants σε σχέση με τη θεμελιώδη συχνότητα, μεταξύ ανδρών-γυναικών και αγγλόφωνων αμερικανών και κορεατών, που έγινε σε 40 άτομα (10 άνδρες, 10 γυναίκες για κάθε γλώσσα), παρατηρούμε τα αποτελέσματα που δίνονται στους παρακάτω πίνακες:

- αγγλόφωνοι αμερικάνοι

φώνημα	Άνδρες				Γυναίκες			
	F0	F1	F2	F3	F0	F1	F2	F3
a	125	638	1051	2318	205	782	1287	2563
e	128	469	2082	2636	209	521	2536	2991
i	136	286	2317	3033	221	390	2826	3416
o	129	498	1127	2375	207	528	1206	2824
u	135	333	1393	2282	228	417	1511	2796

- Κορεάτες

φώνημα	Άνδρες				Γυναίκες			
	F0	F1	F2	F3	F0	F1	F2	F3
a	162	738	1372	2573	264	986	1794	2957
e	167	490	1968	2644	263	650	2377	3068
i	172	341	2219	3047	271	344	2814	3471
o	170	453	945	2674	269	499	1029	3068
u	174	369	981	2565	278	422	1021	3024

Η ποσοστιαία διαφορά στη θεμελιώδη συχνότητα, καθώς και στις άλλες συχνότητες, μεταξύ ανδρών και γυναικών που υπολογίζεται χρησιμοποιώντας τον τύπο:

$$\text{Diff\%} = \frac{F_{nf} - F_{nm}}{F_{nm}} \times 100$$

βρέθηκε ότι είναι:

- για τους άγγλους ομιλητές άνδρες και γυναίκες και την F0 63%
- για τους Κορεάτες ομιλητές άνδρες και γυναίκες και την F0 59%

- για τους άγγλους ομιλητές η ποσοστιαία διαφορά μεταξύ ανδρών-γυναικών
 - F1 = 18%
 - F2 = 16%
 - F3 = 15%

- για τους Κορεάτες ομιλητές η ποσοστιαία διαφορά μεταξύ ανδρών-γυναικών
 - F1 = 18%
 - F2 = 20%
 - F3 = 17%

Εάν χρησιμοποιήσουμε τη διαδικασία κανονικοποίησης που προτείνεται από τον Fant, τότε μειώνοντας επί παραδείγματι την τιμή του F2 κατά 16% για τις γυναίκες αμερικανίδες, τότε η τιμή του F2 θα συμπέσει με αυτή του ανδρός, ενώ το F1 θα παρουσιάσει μία απόκλιση 2% και το F3 μία απόκλιση 1%.

Ο Byunggon Yang εφαρμόζοντας στα παραπάνω δεδομένα τους τρεις μη γλωσσολογικούς παράγοντες, διεπίστωσε ότι η μέση θεμελιώδη συχνότητα διαμορφώνεται ως εξής:

- Για τους αγγλόφωνους Αμερικανούς

- άνδρες 130 Hz
- γυναίκες 212 Hz

- Για τους Κορεάτες ομιλητές

- άνδρες 169 Hz
- γυναίκες 269 Hz

Για ανοιχτά φωνήεντα με υψηλό F1 η θεμελιώδης συχνότητα ήταν χαμηλότερη.

Ο Miller υποστήριξε ότι η θεμελιώδης συχνότητα μπορεί να αποτελέσει σημαντικό παράγοντα για να επιτύχουμε την κανονικοποίηση και το ευδιάκριτο των φωνηέντων. Στατιστικά, τα δεδομένα έδειξαν ότι υπάρχει μία ισχυρή αρνητική συσχέτιση μεταξύ της F0 και της F1 στους άνδρες.

(το $r = -0.84$ για τους Αμερικάνους και $r = -0.92$ για τους Κορεάτες).

Η συσχέτιση αυτή για τις γυναίκες είναι ασθενέστερη ($r = 0.79$ για τις Αμερικάνες και $r = -0.73$ για τις Κορεάτισες). Έτσι το F0 ίσως μπορεί να χρησιμοποιηθεί για τον υπολογισμό του F1 με πιθανότητα ορθότητας της τάξης του 70-84%. Αλλά η συσχέτιση αυτή είναι δεν είναι ικανοποιητική για να βρούμε τα F2 ή F3, σε σχέση με το F0.

Ο δεύτερος παράγοντας είναι το μήκος της φωνητικής οδού. Χρησιμοποιώντας την εξίσωση (1) το συνολικό μήκος της φωνητικής οδού υπολογίστηκε από το F3 του φωνήεντος στη λέξη hudd για τους Αμερικανούς και από το F3 του φωνήεντος Ω για τους Κορεάτες. Το μέσο μήκος της φωνητικής οδού ήταν 17.1 cm για τους Αμερικάνους άνδρες και 14.7 cm για τις γυναίκες, ενώ για τους Κορεάτες ήταν 15.9 cm και 14.2 cm για άνδρες και γυναίκες αντίστοιχα.

Έτσι τα groups με τα διαφορετικά γλωσσικά χαρακτηριστικά διέφεραν κατά 1 cm οι άνδρες και λιγότερο από 1 cm οι γυναίκες. Ωστόσο 1 cm διαφορά στη φωνητική οδό, όπως προκύπτει από την εξίσωση (1) οδηγεί σε 200 Hz διαφορά στη συχνότητα F3. Για παράδειγμα η F3 είναι 3036 Hz για μια φωνητική οδό 14cm αν την συγκρίνουμε με 2833 Hz για μια φωνητική οδό μήκους 15 cm.

Συνεπώς το ότι συμπεριλαμβάνουμε στη σύγκριση των δύο ομάδων την διαφορά μεγέθους της φωνητικής οδού είναι απόλυτο βάσιμο.

Η τρίτος παράγοντας έγκειται στη διαφορά μεγέθους της στοματικής και φαρυγγικής κοιλότητας. Τα μεγέθη των όπισθεν και πρόσθιων κοιλοτήτων υπολογίσθηκαν για το «ι» για Άγγλους και Κορεάτες ομιλητές χρησιμοποιώντας τις εξισώσεις (4) και (5).

Το LF ήταν 7.3 cm και 7.7 cm για αμερικάνους και Κορεάτες άνδρες αντίστοιχα και 6.0 cm και 6.0cm για τις γυναίκες αντίστοιχα. Το μέσο LB ήταν 5.0 cm και 4.9 cm για Αμερικάνες και Κορεάτισες αντίστοιχα και 5.6 cm για αμερικάνους και Κορεάτες άνδρες.

Άνδρες και γυναίκες ομιλητές έχουν την όπισθεν κοιλότητα μακρύτερη από την πρόσθια, αλλά η διαφορά δεν είναι μεγάλη για τις γυναίκες. Παρατηρείται ότι η μέση διαφορά μεγέθους της πρόσθιας κοιλότητας είναι μικρή και στα δύο φύλα, σε αντίθεση με την όπισθεν κοιλότητα, όπου η διαφορά μεγέθους της είναι διπλάσια από τη διαφορά μεγέθους της πρόσθιας κοιλότητας.

Επίσης, η μέση διαφορά για το φωνήεν {i} το οποίο εκφωνήθηκε από τους Κορεάτες ομιλητές είναι της τάξης το 1% για το formant F1, ενώ για το F2 είναι περίπου 27%.

Ένα άλλο σημείο που παρατηρούμε είναι ότι το συνολικό μέγεθος των LB και LF είναι ελάχιστα μεγαλύτερο για τους Κορεάτες άνδρες σε σύγκριση με τους Αμερικάνους άνδρες, ενώ ισχύει το αντίθετο για τα τα groups των γυναικών. Από αυτό συμπεραίνουμε ότι η φωνητική οδός για τους Αμερικανούς άνδρες είναι μεγαλύτερη από αυτή των Κορεατών.

Παρατηρώντας τις μεγάλες διαφορές στη φωνητική οδό διαφόρων ομάδων ατόμων, υιοθετήθηκε το μοντέλο των Nordstrom και Lindblom.

Τα αγγλικά φωνήεντα για τους άνδρες απετέλεσαν την κύρια πηγή αναφοράς για την κανονικοποίηση. Κατ' αρχήν τα δεδομένα για τις γυναίκες με αγγλική και κορεατική καταγωγή δεν ήταν κανονικοποιημένα σε σχέση με αυτά των ανδρών.

Κατά δεύτερον τα δεδομένα που αφορούσαν άνδρες και γυναίκες κορεατικής καταγωγής δεν ήταν κανονικοποιημένα σε σχέση με αυτά των αμερικανών.

Το F3 για ανοιχτά φωνήεντα με F1 μεγαλύτερο των 600 Hz, για τους αγγλόφωνους Αμερικάνους (F3mAE) και το ισοδύναμο για τα γυναικεία δεδομένα (F3fAE) υπολογίζονται ως εξής

$$F3_{mAE} = (2318+2527+2497+2548) / 4 = 2472.5$$

$$F3_{fAE} = (2968 + 2901 + 2895 + 2563 + 2928 + 2877) / 6 = 2855.3$$

Τότε ο συντελεστής κλίμακος k καθορίζεται ως εξής

$$k = \frac{f3_{mAE}}{f3_{fAE}} = \frac{2472.5}{2855.3} = 0.86593$$

Ο συντελεστής κλίμακος εφαρμόζεται ανομοιόμορφα στα AE (American English) γυναικεία δεδομένα. Όπως δηλώνουν οι συντελεστές κλίμακος, αναμένεται να έχουμε μία μείωση στις τιμές των γυναικείων φωνηέντων. Ένας συντελεστής κλίμακος για τα κορεάτικα, ανδρικά και γυναικεία δεδομένα υπολογίστηκε ότι είναι της τάξης του 0.8696.

Πως όμως η uniform scaling method μειώνει τα μη γλωσσολογικά χαρακτηριστικά;

Μετρήσεις στα αυθεντικά δεδομένα έδειξαν και στις δύο γλώσσες μια τάση των γυναικείων formants να είναι υψηλότερα.

Παρατηρώντας ότι οι συντελεστές κλίμακος ήσαν 0.86593 για Άγγλους και 0.8696 για Κορεάτες υπολογίζουμε ότι οι γυναικείες φωνητικοί οδοί ήσαν περίπου 14% κοντύτεροι των ανδρικών.

Βρέθηκε ότι μερικά γυναικεία φωνήεντα πέφτουν αναφορικά με τα ανδρικά. Η μέση διαφορά συχνότητας μεταξύ αγγλόφωνων αμερικανών ανδρών και γυναικών είναι:

- * 26 Hz όσο αναφορά το F1 , με τυπική απόκλιση 15 Hz ,
- * 64 Hz όσο αναφορά το F2 , με τυπική απόκλιση 35 Hz και
- * 61 Hz για το F3, με τυπική απόκλιση 39 Hz.

Εξάλλου η μέση διαφορά μεταξύ ανδρών και γυναικών κορεατικής καταγωγής κυμαινόταν:

- 39Hz για το F1 με τυπική απόκλιση 39 Hz ,
- 118Hz για το F2 με τυπική απόκλιση 77Hz,
- 62 Hz για το F3 με τυπική απόκλιση 49Hz.

Για τα κορεάτικα μπροστινά φωνήεντα τα F2 και F3 παρουσιάζουν συστηματική μεταβολή. Όταν το F2 αυξάνεται, τότε η τιμή του F3 μειώνεται.

Τελικά, παρατηρούμε πως η απόσταση μεταξύ ανδρών και γυναικών είναι σχεδόν ίδια στις διαστάσεις F1/F2 και F1/F3.

Οι μεγαλύτερες αποστάσεις στην διάσταση F1/F2 βρέθηκαν στο “u” και στο “a”, ενώ στην F1/F3 διάσταση η μικρότερη απόσταση παρατηρήθηκε για το φωνήεν «Ω».

Μη γραμμικοί συσχετισμοί μεταξύ F0 και των χαμηλότερων formants.

Ως γνωστόν οι φωνητικές χορδές παράγουν ήχο που μετατρέπεται μέσω των δονήσεων της φωνητικής οδού. Έτσι για κάθε φώνημα το ηχητικό κύμα που παράγεται απεικονίζει τα formants του φωνήματος.

Πειραματικά, δεν βρέθηκε καμμία συσχέτιση ανάμεσα στο φώνημα και στο formant. Σύμφωνα με στατιστική αναφορά των Peterson και Barney παρατηρούνται δύο είδη μεταβολών στα formants:

- 1) οι συχνότητες των formants ποικίλλουν ανάλογα με τα χαρακτηριστικά του ομιλητή
- 2) για φωνήεντα τα οποία δείχνουν μεγάλες φασματολογικές διαφορές, οι συχνότητες τους εξαρτώνται από την συνάρθρωση.

Επομένως προκειμένου να μελετηθεί η σχέση ανάμεσα στο φώνημα και τα φυσικά χαρακτηριστικά του σήματος πρέπει να ληφθούν υπόψιν δύο διαφορετικές θεωρητικές προσεγγίσεις.

Η πρώτη προσέγγιση αφορά το πρόβλημα κανονικοποίησης των formants για ένα φωνήεν, ενώ η δεύτερη προσέγγιση αφορά την δυναμική περιγραφή των ήχων.

Καμμία από τις δύο προσεγγίσεις δεν δίνει απάντησεις για το εύρος των φασματολογικών χαρακτηριστικών των φωνηέντων.

Μεθοδολογικά προβλήματα υπολογισμού των formants

Υποστηρίζεται ότι είναι δύσκολο να καθορίσουμε τις συχνότητες των formants όταν η θεμελιώδης συχνότητα κυμαίνεται στο μισό της συχνότητας του πρώτου formant.

Ωστόσο σε στατιστική αναφορά για 15 Σουηδικά φωνήεντα η συχνότητα του πρώτου formant για γυναίκες διαπιστώθηκε ότι είναι < 450 Hz .

Σύμφωνα με αυτό ο υπολογισμός της συχνότητας των formants καθίσταται προβληματικός όταν η θεμελιώδης συχνότητα είναι μεγαλύτερη των 225 Hz. Επιπλέον, όταν η F0 είναι μεγαλύτερη των 300 Hz δεν μπορεί να δοθεί κάποιος κανόνας βάσει του οποίου εκχωρούνται οι τιμές στα formants, αν και μπορεί να αποδειχθεί ότι τα φωνήεντα ακούγονται καθαρά όταν η θεμελιώδης συχνότητα κυμαίνεται μέχρι και 700 Hz.

Μία άλλη μελέτη για τους Γερμανούς άνδρες, γυναίκες και παιδιά ομιλητές, που διεξήχθη από τους Andreas Klinkert και Dieter Maurer, δείχνει τα τη σχέση των formants με τη θεμελιώδη συχνότητα. Στη μελέτη αυτή έγιναν 18674 εγγραφές από 35 άνδρες, 44 γυναίκες και 20 παιδιά, που εκφώνησαν CVC συλλαβές που περιείχαν τα φωνήεντα i, y, e, o, και u. Η θεμελιώδης συχνότητα κατά τη μελέτη αυτή κυμάνθηκε από 70Hz μέχρι 870Hz και με τρεις διαφορετικές εντάσεις. Για την ψηφιοποίηση του ήχου χρησιμοποιήθηκε η συχνότητα δειγματοληψίας 22050 Hz και 16bits ανάλυση έντασης.

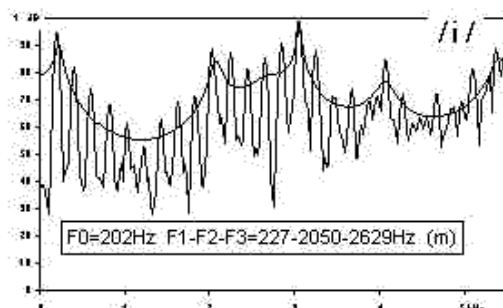
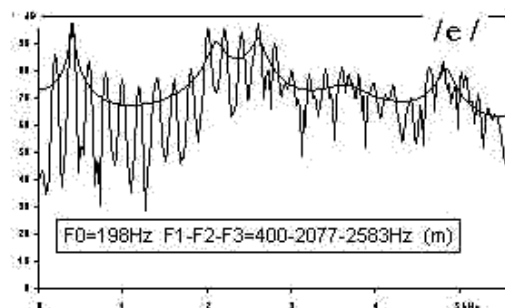
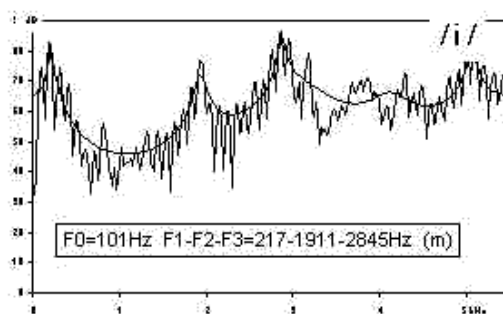
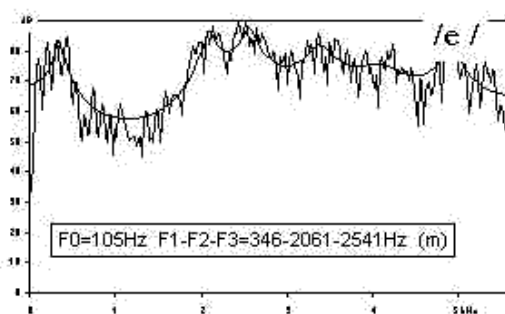
Τα αποτελέσματα της μελέτης αυτής, έδειξαν τη μετατόπιση των formants F1 και F2 < 1.5 kHz, σε συνδυασμό με την αύξηση της F0.

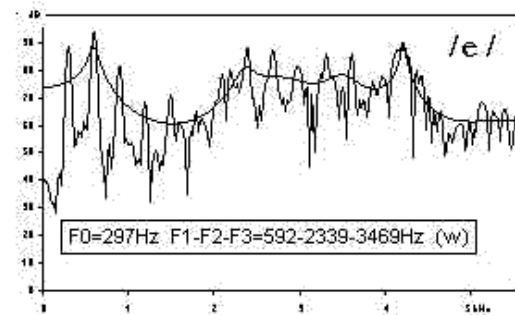
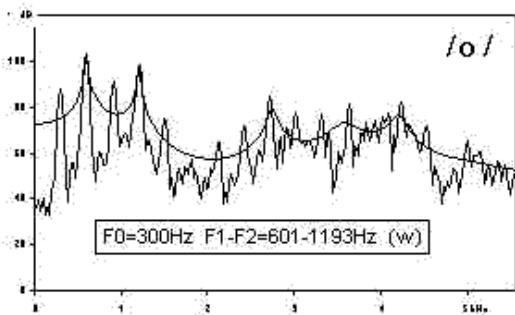
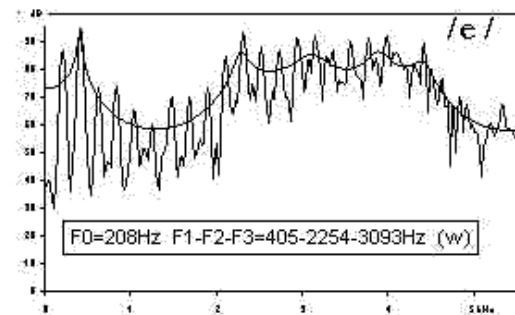
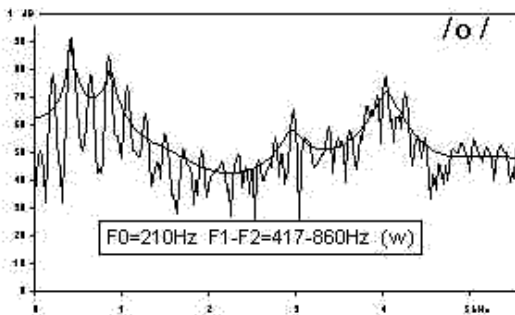
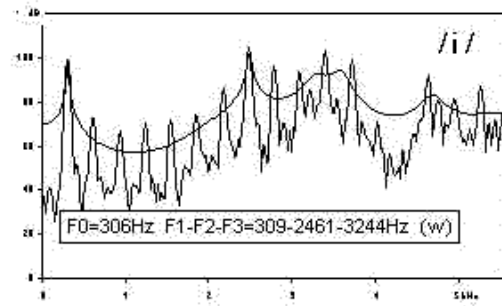
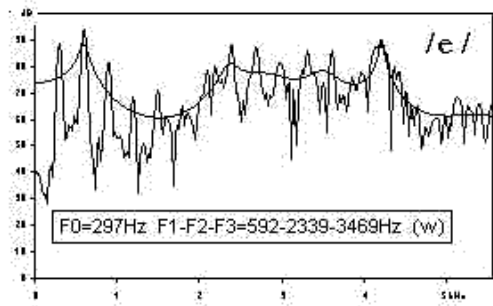
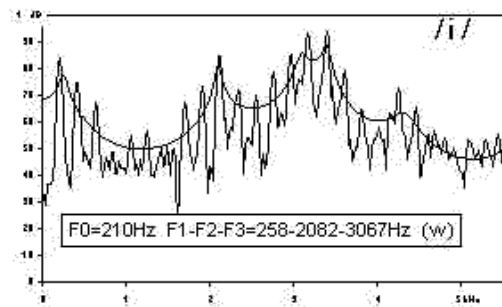
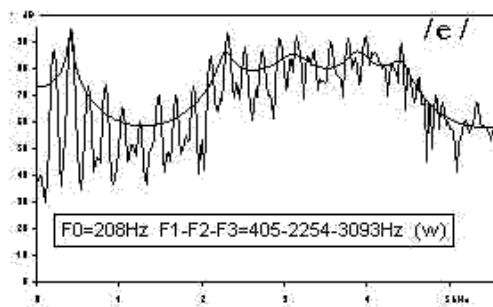
Ωστόσο, το αριθμητικό μέγεθος τέτοιων μετατοπίσεων δεν υπόκειται αναγκαστικά σε ένα τέτοιο περιορισμό. Από μελέτες έχει αποδειχθεί ότι καθώς αυξάνεται η F0 έχουμε μεγαλύτερες μετατοπίσεις συχνοτήτων. Ειδικότερα μόνο τα formants $< 1.5 - 2.0$ kHz συσχετίζονται με την F0 ενώ η μετατόπιση της F1 πολλές φορές υπερβαίνει την μετατόπιση της F0. Επίσης μεγάλες μεταβολές στην F2 δημιουργούν πρόβλημα στον υπολογισμό των formants.

Αποδείχθηκε πως ο συσχετισμός μεταξύ της F0 και των χαμηλότερων formants είναι μη γραμμικός και μη μονοτονικός.

- Οι μετατοπίσεις ήταν μη-γραμμικές. Για παράδειγμα δεν ήταν αξιοσημείωτες για $F0 < 175\text{Hz}$, αλλά κάποιες φορές ήταν ιδιαίτερα εμφανείς για $F0 > 250\text{ Hz}$, πράγμα που μπορεί να εξηγήσει γιατί σε προηγούμενες μελέτες που έγιναν για χαμηλότερη F0 δεν γινόταν αντιληπτή η σχέση θεμελιώδους συχνότητας και formants.
- Οι μετατοπίσεις δεν ήταν μονοτονικές. Για παράδειγμα για διαφορετικά φωνήεντα βρέθηκαν διαφορετικές μετατοπίσεις.

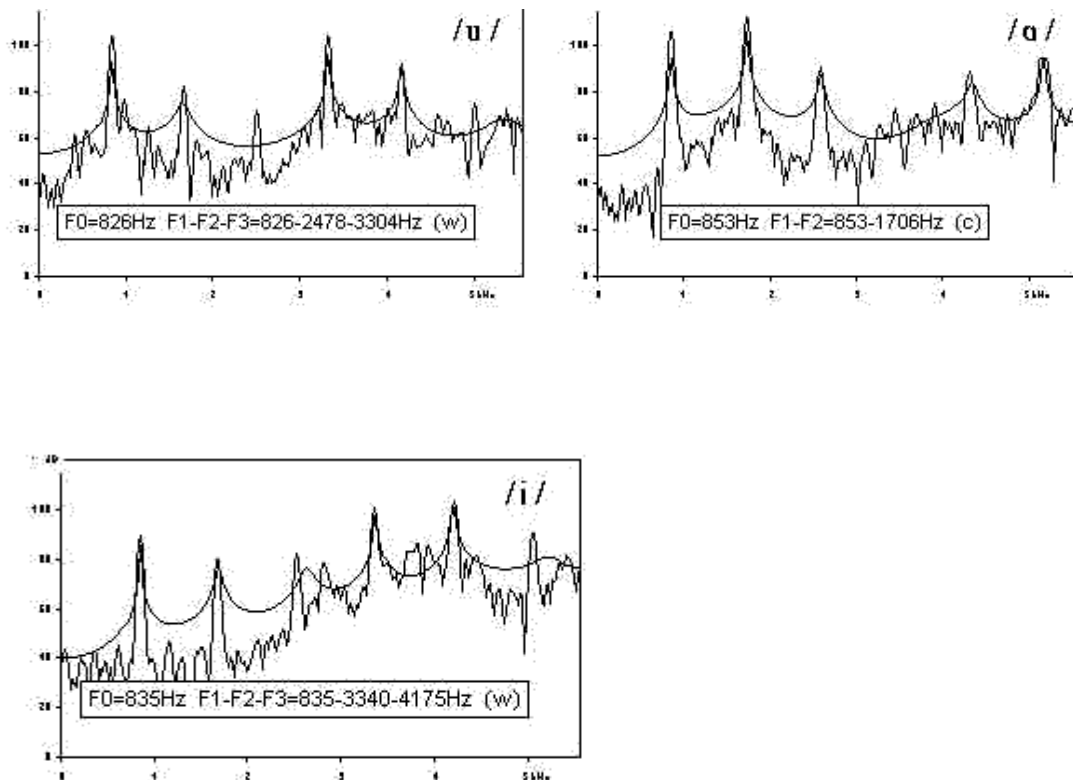
Στα παρακάτω σχήματα παρατηρούμε τις μετατοπίσεις των formants F1 και F2 για διαφορετικές θεμελιώδεις συχνότητες των 100Hz και 200Hz για τους άνδρες και 200Hz και 300Hz για τις γυναίκες.





Όταν η θεμελιώδης συχνότητα F0 ήταν μικρότερη των 200Hz, τότε η μετατόπιση των χαμηλών συχνοτήτων των formants ήταν μικρή, ενώ για F0 μεγαλύτερη των 200Hz, η μετατόπιση ήταν σαφώς μεγαλύτερη για τα φωνήεντα «ο» και «ε», από ότι για το «ι».

Βέβαια η αύξηση της F0 δεν μπορεί να υπερβεί κάποιο άνω φράγμα. Για παράδειγμα, η συγκεκριμένη μελέτη στην οποία αναφερόμαστε έδειξε πως για αύξηση της F0 στα επίπεδα των 728Hz, η εκφώνηση έγινε κατανοητή από το 80% των ακροατών, ενώ για θεμελιώδη συχνότητα της τάξης των 843Hz, έγινε κατανοητή από το 100% των ακροατών μόνο όμως για τα φωνήεντα «υ» και «ί».



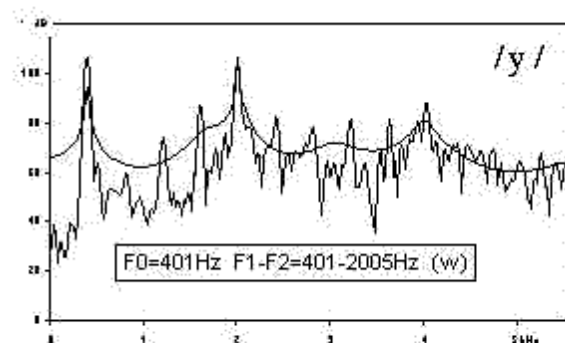
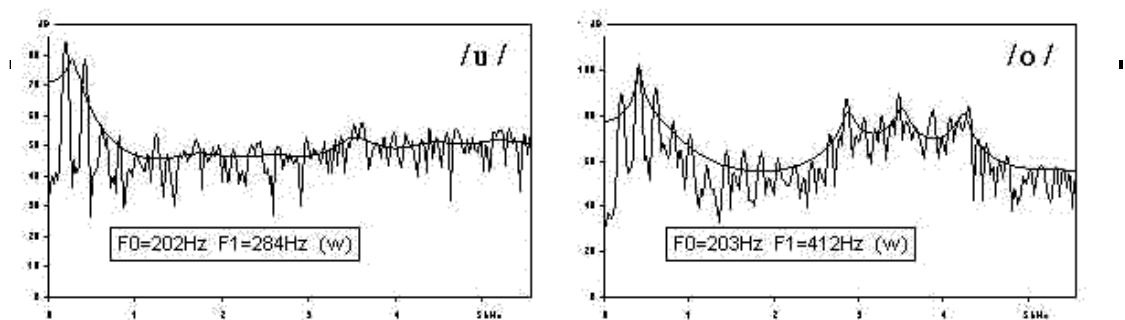
Επιπλέον, βρέθηκε ότι το άνω όριο της F0 για τους ομιλητές ήταν μεγαλύτερο από αυτό των τραγουδιστών.

Είναι ευρέως γνωστό, ότι αυξάνοντας την F0, ο υπολογισμός των συχνοτήτων των formants καθίσταται προβληματικός. Για $F0 > 300$ Hz δεν μπορούμε να εκχωρήσουμε συχνότητες στα formants, παρ' όλο που τα φωνήεντα προσδιορίζονται καθαρά.

Υποστηρίζεται ότι ένα όπισθεν φωνήεν μπορεί να χαρακτηριστεί ικανοποιητικά από ένα formant, όπως επίσης ένα πρόσθιο φωνήεν μπορεί να προσδιοριστεί ικανοποιητικά από τις τιμές δύο formants.

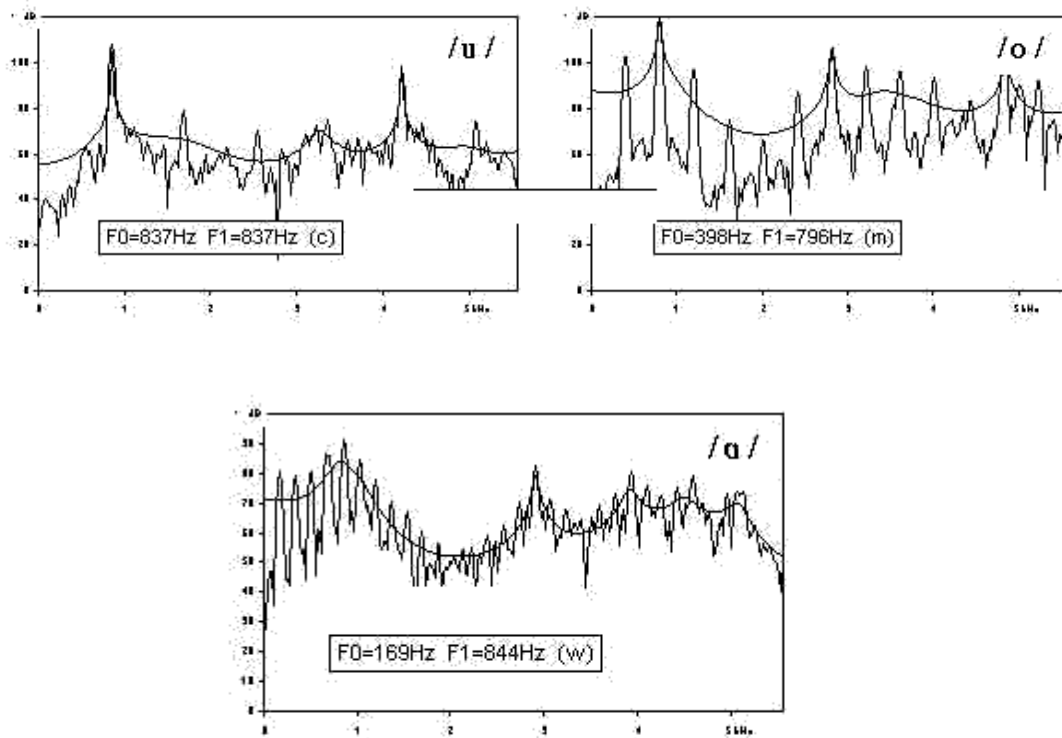
Σε εργαστήρια ομιλίας όμως παρατηρείται συχνά ότι χάνονται formants. Η απώλεια των formants ερμηνεύεται σαν συγχώνευση των formants. Ερευνώντας τα φωνήεντα στη γερμανική γλώσσα βρίσκουμε παραδείγματα φωνηέντων με ένα ή δύο το πολύ formants και άρα δεν μπορούμε να έχουμε εμπειριστατωμένη άποψη για την φασματική ενσωμάτωση.

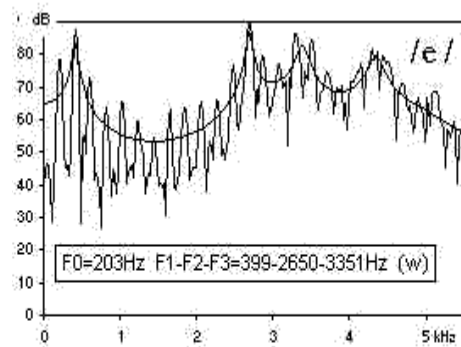
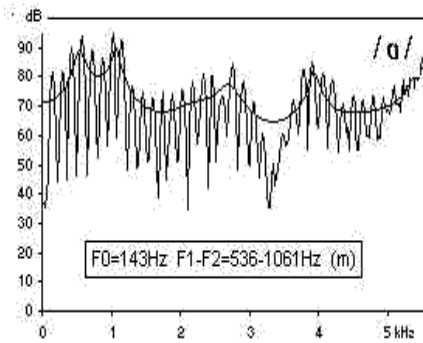
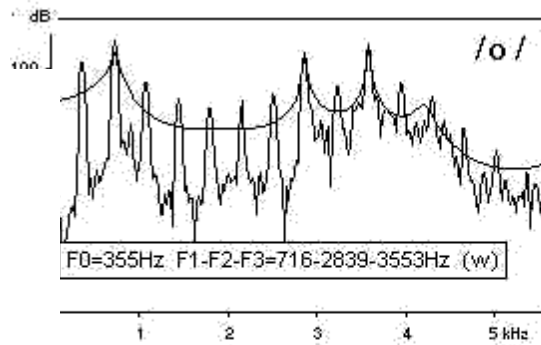
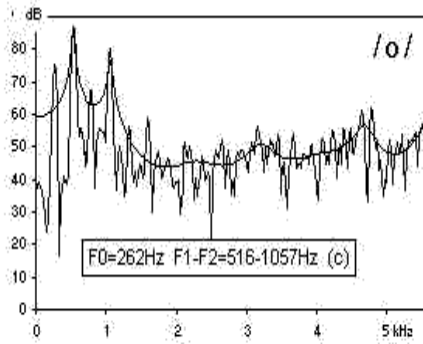
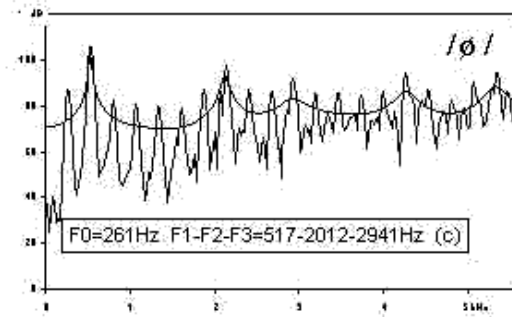
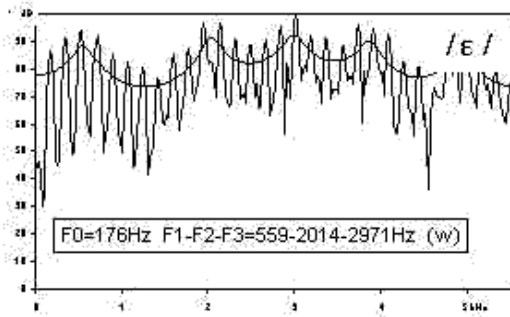
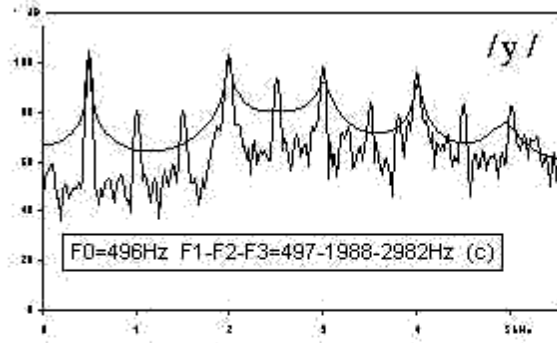
Συνάγεται λοιπόν ότι ο αριθμός των formants που προσδιορίζουν ένα φωνήεν δεν είναι σταθερός



Μπορούν όμως δύο γειτονικά ή μη φωνήεντα να έχουν τις ίδιες τιμές για τα F1, F2 και F3; Τέτοιες ομοιότητες βρέθηκαν σύμφωνα με έρευνα και για πρόσθια και για όπισθεν φωνήεντα και αποδείχθηκε ότι μια τέτοια ασαφопоίηση αποδεικνύει ότι δεν υπάρχει μια ακριβής ή προσεγγιστική σχέση μεταξύ φωνηέντων και formants. Τότε όμως δεν ισχύει η συνάρθρωση.

Επειδή η F0 και τα χαμηλότερα formants συσχετίζονται, επειδή οι συχνότητες των υψηλότερων formants μεταβάλλονται για ένα φωνήεν και επειδή οι εντάσεις των formant επηρεάζουν το φωνήεν που ακούμε, τότε όμοιες συχνότητες για τα formants δεν μας δείχνουν ξεκάθαρα τις διαφορές μεταξύ των φωνηέντων. Για παράδειγμα, για τα πίσω φωνήεντα ο προσδιορισμός τους επιτυγχάνεται από formants με συχνότητες μικρότερες των 2kHz, ενώ τα εμπρόσθια φωνήεντα προσδιορίζονται από τις συχνότητες των τριών χαμηλότερων formants.





Τα formants στους άνδρες-γυναίκες τραγουδιστές.

Στον παρακάτω πίνακα μπορούμε να δούμε τις τιμές των δύο πρώτων formants που έδωσε η μελέτη για τους tenor, alto, bass, countertenor και soprano

Tenor	F1	F2
a	650	1080
e	400	1700
i	290	1870
o	400	800
u	350	600

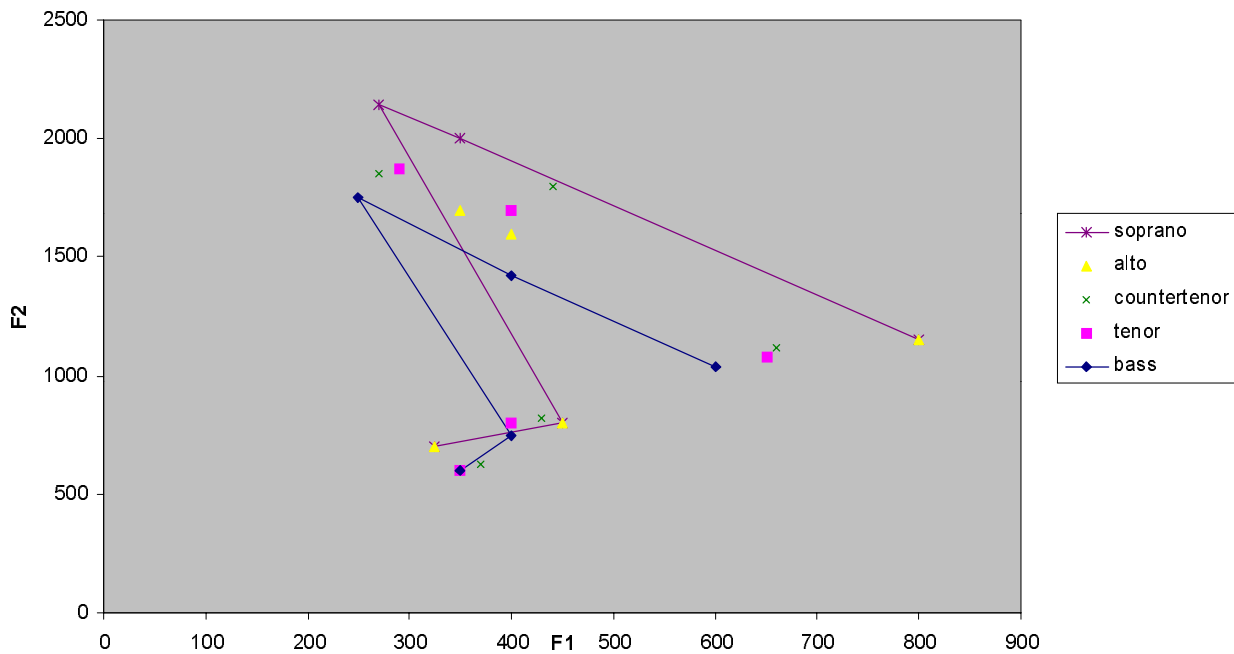
Alto	F1	F2
a	800	1150
e	400	1600
i	350	1700
o	450	800
u	325	700

Bass	F1	F2
a	600	1040
e	400	1420
i	250	1750
o	400	750
u	350	600

Countertenor	F1	F2
a	660	1120
e	440	1800
i	270	1850
o	430	820
u	370	630

Soprano	F1	F2
a	800	1150
e	350	2000
i	270	2140
o	450	800
u	325	700

Μπορούμε έτσι να διαπιστώσουμε και στους τραγουδιστές τις διαφορές στα formants μεταξύ ανδρών και γυναικών. Συγκεκριμένα, στο σχήμα της επόμενης σελίδας, χρησιμοποιώντας τις τιμές των F1-F2, μπορούμε να δούμε τις μετατοπίσεις αυτές. Δώσαμε ιδιαίτερη έμφαση στις soprano και στους bass, που θεωρούνται οι ακραίες περιπτώσεις και ενώσαμε τις τιμές τους με ευθύγραμμα τμήματα για να γίνουν πιο ευκρινείς οι μετατοπίσεις αυτές.



Αλλά και σύμφωνα με τη μελέτη του Sundburg, τα αποτελέσματα της οποίας βλέπουμε στον κάτωθι πίνακα, παρατηρούμε πως η συχνότητα του πρώτου formant στις soprano, είναι υψηλότερη από την αντίστοιχη των ανδρών. Ο ίδιος ισχυρίστηκε ότι η θεμελιώδης συχνότητα στις γυναίκες soprano είναι υψηλότερη από την τιμή του πρώτου formant, με συνέπεια να υποβαθμίζεται η ποιότητα της φωνής και να μην είναι ευδιάκριτη. Ωστόσο, ο ίδιος παρατήρησε πως οι γυναίκες soprano είχαν μία τεχνική για να αντιμετωπίσουν το πρόβλημα. Διευρύνοντας το άνοιγμα της σιαγόνας όταν η θεμελιώδης συχνότητα αυξάνεται, τότε αυξάνεται αντίστοιχα και η συχνότητα του πρώτου formant, καθώς επίσης παρατηρείται και μετατόπιση και των υπολοίπων formant.

	άνδρες			soprano		
φώνημα	F1	F2	F3	F1	F2	F3
a	609	1000	2450	650	1100	2860
e	400	1700	2300	500	1750	2450
o	325	700	2550	400	840	2800

Συμπεράσματα – Τελικές τιμές στην παρούσα εργασία

Στόχος της παρούσας εργασίας ήταν να μεταβληθούν οι τιμές του συνθέτη ομιλίας του Πανεπιστημίου Αθηνών του Τμήματος Πληροφορικής του Τομέα Επικοινωνιών και Επεξεργασίας Σήματος, έτσι ώστε να προσομοιωθεί η ομιλία του με γυναικεία φωνή. Αυτό θα μπορούσε να γίνει εφικτό μεταβάλλοντας τις μέσες τιμές αναφοράς των παραμέτρων του Klatt. Ωστόσο, μεταβάλλοντας τις παραμέτρους Klatt, το αποτέλεσμα δεν μπορεί να είναι απόλυτα ικανοποιητικό, μιάς και η προσομοίωση γυναικείας ομιλίας, απαιτεί και ένα πλήθος άλλων παραμέτρων όπως προαναφέραμε ή και ενδεχομένως να απαιτούσε ένα άλλο μοντέλο. Πάντως, η προσπάθεια επικεντρώθηκε στην όσο το δυνατόν εύστοχη μεταβολή των τριών πρώτων formants και των αντίστοιχων εύρων ζώνης τους, καθώς επίσης και στην κατάλληλη τιμή που έπρεπε να θέσουμε στη θεμελιώδη συχνότητα. Παράγοντες, βέβαια, που δεν είναι σαφώς αρκετοί για να παραχθεί το επιθυμητό αποτέλεσμα.

Βασιζόμενοι στις μέσες τιμές των formants των φωνηέντων για τις γυναίκες, - όπου και χρησιμοποιήσαμε- , που υπολογίστηκαν από τον Ιωάννη Κονταξάκη στην φασματογραφική μελέτη των φωνηέντων a,e,i,o,u, της ελληνικής γλώσσας και στις υπάρχουσες τιμές Klatt του συνθέτη ομιλίας, χρησιμοποιήσαμε την κάτωθι αναλογία, για να υπολογίσουμε τις τιμές των formants για τα τονούμενα φωνήεντα της ελληνίδας ομιλήτριας:

$$\varepsilon.γ.τ. = \frac{\xi.α.τ. \bullet \varepsilon.γ.α.}{\xi.α.α.},$$

όπου	ε.γ.τ. :	ελληνίδας γυναίκας τονούμενο φωνήεν
	ε.γ.α. :	ελληνίδας γυναίκας άτονο φωνήεν
	ξ.α.τ. :	τονούμενο φωνήεν συνθέτη Klatt
	ξ.α.α. :	άτονο φωνήεν συνθέτη Klatt

Για τα εύρη ζώνης B1, B2, B3 χρησιμοποίησαμε τον τύπο

$$B_n = 50 + 0.05 F_n ,$$

όπου μας δίνει το B_n δηλαδή το n-οστό εύρος ζώνης σε σχέση με το F_n την τιμή δηλαδή του n-οστού formant.

οι τιμές του συνθέτη Klatt για τα ανδρικά φωνήεντα δίνονται στον παρακάτω πίνακα:

φώνημα	F1	F2	F3	B1	B2	B3
a	715	1290	2397	102	83	203
e	544	1645	2474	76	87	159
i	350	1941	2584	56	101	172
o	565	900	2392	79	68	178
u	400	989	2584	82	85	145
a'	765	1340	2397	102	83	203
e'	594	1695	2474	76	87	159
i'	400	1991	2584	56	101	172
o'	615	1068	2392	79	68	178
U	450	1039	2584	82	85	145

οι τιμές για τα γυναικεία φωνήεντα που υπολόγισε ο Ιωάννης Κονταξάκης είναι:

φώνημα	F1	F2	F3	B1	B2	B3
a	857	1648	2678	71	101	183
e	665	2089	2811	84	114	161
i	525	2351	2942	61	161	144
o	639	1152	2783	140	139	141
u	500	1007	2645	45	87	144

Έτσι με τα στοιχεία αυτά και την παραπάνω μεθοδολογία βρήκαμε πως τα τονούμενα φωνήεντα της ελληνίδας ομιλήτριας θα είναι:

	F1	F2	F3	B1	B2	B3
a'	916.93	1711.9	2678	95.85	135.59	183.9
e'	726.12	2152.5	2811	86.31	157.62	190.55
i'	600	2411.6	2942	80	170.58	197.1
o'	695.55	1367	2783	84.78	118.35	189.15
U	562.5	1057.9	2645	78.13	102.9	182.25

Επιπλέον, όσον αφορά τα σύμφωνα, βασισμένοι στις τιμές του Klatt, για να πετύχουμε τη γυναίκα αυξήσαμε

- το F1 κατά 18%,
- το F2 κατά 20% και
- το F3 κατά 17%.

Για το εύρος ζώνης χρησιμοποιήσαμε την ίδια τεχνική. Στις περιπτώσεις που βρήκαμε μικρότερα εύρη ζώνης για κάποια formant, αφήσαμε αυτά του συνθέτη Klatt.

Έτσι, ενώ ο συνθέτης Klatt είχε για τα σύμφωνα τις παρακάτω τιμές:

φώνημα	F1	F2	F3	B1	B2	B3
v	220	1100	2080	180	90	120
g	260	2070	3020	40	250	500
d	270	1290	2540	500	500	40
z	240	1390	2530	70	60	180
th	320	1290	2540	200	90	200
k	300	1990	2850	300	250	300
l	310	1050	2880	40	40	40
m	480	1270	2130	40	200	200
n	480	1340	2470	40	300	300
p	400	1300	2350	300	150	280
r	310	1060	1380	70	100	120
s	320	1390	2530	200	80	200
τ	400	1600	2600	300	120	250
f	340	1100	2080	200	120	150
x	350	1941	2584	356	101	172
gk	300	1990	2850	300	250	300
b	400	1300	2350	300	150	280
c	240	1390	2530	70	60	180
j	240	1390	2530	70	60	180

χρησιμοποιώντας την τεχνική που περιγράψαμε καταλήξαμε πως οι τιμές για τα σύμφωνα στην ελληνίδα ομιλήτρια θα πρέπει να είναι:

φώνημα	F1	F2	F3	B1	B2	B3
v	259.6	1320	2433.6	180	116	171.68
g	306.8	2484	3533.4	65.34	250	500
d	318.6	1548	2971.8	500	500	198.59
z	283.2	1668	2960.1	70	133.4	198.005
th	377.6	1548	2971.8	200	127.4	198.59
k	354	2388	3334.5	300	250	300
l	365.8	1260	3369.6	68.29	113	218.48
m	566.4	1524	2492.1	78.32	200	200
n	566.4	1608	2889.9	78.32	300	300
p	472	1560	2749.5	300	150	280
r	365.8	1272	1614.6	68.29	113.6	130.73
s	377.6	1668	2960.1	200	133.4	198.005
t	472	1920	3042	300	146	250
f	401.2	1320	2433.6	200	120	171.68
x	413	2329.2	3023.28	356	166.46	201.164
gk	354	2388	3334.5	300	250	300
b	472	1560	2749.5	300	150	280
c	283.2	1668	2960.1	70	133.4	198.005
j	283.2	1668	2960.1	70	133.4	198.005

Για τις σταθερές τιμές χρησιμοποιήσαμε:

θεμελιώδης συχνότητα $F_0=230\text{Hz}$

συχνότητα 5^{ου} formant $F_5=3750\text{Hz}$

συχνότητα 6^{ου} formant $F_6=4900\text{Hz}$

εύρος ζώνης 4^{ου} formant $B_4=250\text{Hz}$

εύρος ζώνης 5^{ου} formant $B_5=200\text{Hz}$

εύρος ζώνης 6^{ου} formant $B_6=1000\text{Hz}$

αριθμός σειριακών formants $NFC=5$

συχνότητα έρρινου πόλου $FNP=200\text{Hz}$

εύρος ζώνης έρρινου μηδενικού $BNZ=200\text{Hz}$

εύρος ζώνης έρρινου πόλου $BNP=150\text{Hz}$

συχνότητα 1^{ου} γλωττιδικού ηχείου $FGP=600\text{Hz}$

εύρος ζώνης 1^{ου} γλωττιδικού ηχείου $BGP=500\text{Hz}$

συχνότητα γλωττιδικού μηδενικού $FGZ=1500\text{Hz}$

εύρος ζώνης γλωττιδικού μηδενικού $BGZ=6000\text{Hz}$

εύρος ζώνης 2^{ου} γλωττιδικού ηχείου $BGS=100\text{Hz}$

ρυθμός δειγματοληψίας $SR=17500$

αριθμός δειγμάτων ανά πλαίσιο $NWS=25$

ρυθμιστικό συνολικής απολαβής $G_0=35\text{dB}$

Στον κάτωθι πίνακα παρουσιάζονται συνολικά οι τιμές των formants F1, F2, F3 και τα αντίστοιχα εύρη ζώνης B1, B2, B3 που χρησιμοποιήθηκαν και για φωνήεντα και για σύμφωνα.

φώνημα	F1	F2	F3	B1	B2	B3
v	259.6	1320	2433.6	180	116	171.68
g	306.8	2484	3533.4	65.34	250	500
d	318.6	1548	2971.8	500	500	198.59
z	283.2	1668	2960.1	70	133.4	198.005
th	377.6	1548	2971.8	200	127.4	198.59
k	354	2388	3334.5	300	250	300
l	365.8	1260	3369.6	68.29	113	218.48
m	566.4	1524	2492.1	78.32	200	200
n	566.4	1608	2889.9	78.32	300	300
p	472	1560	2749.5	300	150	280
r	365.8	1272	1614.6	68.29	113.6	130.73
s	377.6	1668	2960.1	200	133.4	198.005
t	472	1920	3042	300	146	250
f	401.2	1320	2433.6	200	120	171.68
x	413	2329.2	3023.28	356	166.46	201.164
gk	354	2388	3334.5	300	250	300
b	472	1560	2749.5	300	150	280
c	283.2	1668	2960.1	70	133.4	198.005
j	283.2	1668	2960.1	70	133.4	198.005
a	857	1648	2678	71	101	183
e	665	2089	2811	84	114	161
i	525	2351	2942	61	161	144
o	639	1152	2783	140	139	141
u	500	1007	2645	45	87	144
a'	916.93	1711.9	2678	95.85	135.59	183.9
e'	726.12	2152.5	2811	86.31	157.62	190.55
i'	600	2411.6	2942	80	170.58	197.1
o'	695.55	1367	2783	84.78	118.35	189.15
U	562.5	1057.9	2645	78.13	102.9	182.25

Ποιοτική μέτρηση

Έγινε το εξής πείραμα:

Σε 10 ακροατές 5 άνδρες και 5 γυναίκες ηλικίας 29-40 ετών, βάλαμε να ακούσουν 10 προτάσεις που συνετέθησαν με γυναικεία φωνή (αυτή τουλάχιστον που προσπαθήσαμε να προσομοιάσουμε) και να μας γράψουν αναλυτικά τη φράση που άκουσαν. Τις φράσεις αυτές τις άκουσαν μία και μόνη φορά.

Τα αποτελέσματα που είχαμε φαίνονται στον κάτωθι πίνακα:

Φράση/πρόσωπο	α1	α2	α3	α4	α5	γ1	γ2	γ3	γ4	γ5	
1	v	v		v	v	v	v	v	v	v	90%
2	v	v	v	v	v	v	v	v	v	v	100%
3	v	v		v	v	v	v	v		v	80%
4	v		v	v	v	v	v	v	v	v	90%
5	v	v	v	v	v	v	v	v	v	v	100%
6	v	v	v	v	v	v	v	v	v	v	100%
7	v	v	v	v	v	v	v	v	v	v	100%
8	v	v	v	v	v	v	v	v	v	v	100%
9	v	v	v	v	v	v	v	v	v	v	100%
10	v	v	v	v	v	v	v	v	v	v	100%
	100%	90%	80%	100%	100%	100%	100%	100%	90%	100%	

όπου α1,α2,α3,α4,α5 είναι οι άνδρες και
γ1,γ2,γ3,γ4,γ5 είναι οι γυναίκες

με συνολική επιτυχία κατανόησης 96%

Στη συνέχεια άκουσαν 10 προτάσεις εκ των οποίων 5 ήταν με τη γυναικεία φωνή και 5 με την ανδρική με τυχαία όμως σειρά.

Οι ακροατές ερωτήθηκαν εάν ο ομιλητής (=συνθέτης) που άκουσαν ήταν άνδρας ή γυναίκα.

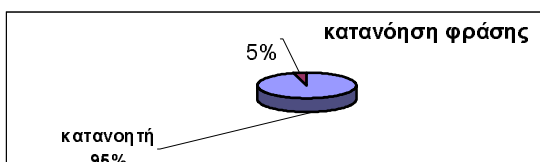
Η επιτυχία στην διάκριση του φύλου ήταν 100% εάν και κάποιοι από τους ακροατές ήταν επιφυλακτικοί στον εάν πρόκειται για γυναικεία ή παιδική φωνή, πράγμα που σήμαινε ελάττωση της θεμελιώδους συχνότητας.

Ποιοτική ανάλυση του συνθέτη ομιλίας του Τμήματος Πληροφορικής του Πανεπιστημίου Αθηνών με προσέγγιση γυναικείας φωνής.

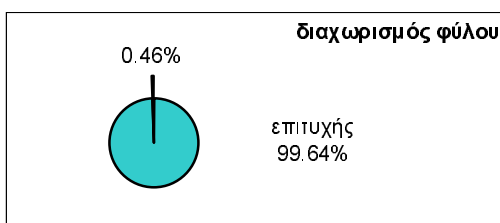
Εικοσιδύο συμμετοχτές μας, ηλικίας 23 έως 45 ετών άκουσαν από τον συνθέτη ομιλίας του Τμήματος Πληροφορικής του Πανεπιστημίου Αθηνών με γυναικεία φωνή 10 φράσεις και τις κατέγραψαν. Η κατανόηση του συνθέτη ανέρχεται στο 95.5%.

Στη συνέχεια, άκουσαν 10 φράσεις του συνθέτη ομιλίας, εκ των οποίων οι 5 με ανδρική και οι υπόλοιπες 5 με γυναικεία φωνή, με τυχαία σειρά. Ζητήθηκε να καταγράψουν εάν κάθε φράση που ακουγόταν προερχόταν από γυναικεία ή ανδρική φωνή. Το ποσοστό ευστοχίας του διαχωρισμού του φύλου ήταν 99.54%.

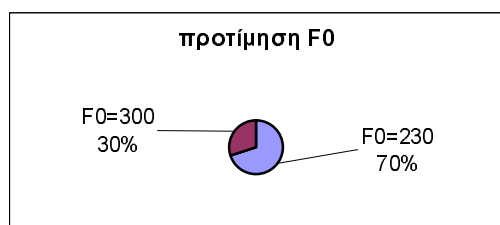
Τέλος, ακούστηκαν δύο εκδοχές της γυναικείας φωνής οι οποίες στο μόνο που διέφεραν ήταν στην Θεμελιώδη συχνότητα (230Hz και 300Hz) και ζητήθηκε εάν τους ικανοποιεί η γυναικεία φωνή ή όχι και εάν η απάντησή τους είναι καταφατική σε ποιά από τις δύο περιπτώσεις. Το 72% έμεινε ικανοποιημένο από την ποιότητα της γυναικείας φωνής και από αυτούς το 70% προτίμησε τη γυναικεία φωνή με $F_0=230\text{Hz}$.



κατανόηση φράσης	μη κατανόηση
95,50%	4,50%



διαχωρισμός φύλου	
επιτυχής	ανεπιτυχής
99,54%	0,46%



προτίμηση F0	
$F_0= 230\text{Hz}$	$F_0=300\text{Hz}$
70%	30%

Βιβλιογραφία-Σημειώσεις-Μελέτες.

- Μαθήματα Επεξεργασίας Ομιλίας

Γεώργιος Θ. Κουρουπέτρογλου
Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών,
Τμήμα Πληροφορικής.

- Φασματογραφική Μελέτη χαρακτηριστικών παραμέτρων των φωνημάτων της ελληνικής γλώσσας

Πτυχιακή εργασία Ιωάννη Κοντζάκη
(Επιβλέπων Καθηγητής Γεώργιος Θ. Κουρουπέτρογλου)

- Ανάπτυξη Συστήματος μετατροπής κειμένου σε ομιλία βάσει κανόνων Εφαρμογή στην ελληνική γλώσσα

Πτυχιακή εργασία Γεράσιμου Ξυδά
(Επιβλέπων Καθηγητής Γεώργιος Θ. Κουρουπέτρογλου)

- Acoustic theory of speech production
G.Fant Mouton 1970

- The acoustic analysis of speech
Ray D. Kent & Charles Read

- Acoustics of American English Speech
A dynamic approach
Joseph P. Olive & Alice Greenwood & John Coleman

- Compute Music
C. Dodge & T. Jezse
1985/Schirmer Books

Διευθύνσεις στο Διαδίκτυο

- Vocal Analysis and synthesis

<http://debussy.music.ubc.ca/~courses/Notes/VocalSynthesis.html>

- Voice of computer

http://isw3.aist-nara.ac.jp/IS/Shikano-ia...voice_of_computer/e-voice_of_computer

- Bell Labs Text-to-Speech Synthesis

<http://www.bell-labs.com/project/tts/voices.html>

- Text-toAudiovisual-speech
- Face model

<http://ophale.icp.grenet.fr/2.1.html>

- Festival: pre-synthesized demos

<http://www.cstr.ed.ac.uk/projects/festival/demos.html>

- Model of speech

<http://svr-www.eng.cam.ac.uk/~ajr/SA95>

- Swedish vowels

<http://www.ling.su.se/STAFF/hartmut/svok.html>

- Voices – singing voices – formants

<http://www.cs.ust.hk/faculty/layers/subpage/342notes/soprano2.html>

- Acoustic Phonetics (Dieter Maurer)

<http://www.neurol.unizh.ch/psychologie/associates/maurer/formants>

- Manipulations in speaker age and sex

<http://www.ling.su.se/staff/hartmut/>

- The perception of speech under adverse acoustic conditions

<http://www.cs.indiana.edu/hyplan/pkienze/greenburg/Chapter5.html>

- Long Island Vowels

<http://www.acoustics.org/134th/huffman.htm>

- Fundamental of experimental phonetics

<http://sp.cis.iwate-u.ac.jp/sp/jp/phonetics>

- The role F0 in vowel perception

<http://www.ling.su.se/staff/hartmut/i.htm>

- Table of formant vowel sounds in Chant

<http://www.computermusic.ch/files/misc/Table-of-formant-values>

- Low level synthesis interface

<http://www.asel.udel.edu/speech/tutorials/synthesis/Klatt.html>

- A comparative study of American English and Korean vowels produced by male and female speakers

<http://hyomin.donggeui.ac.kr/~bgyang/96jp.htm>

Δημοσιεύσεις

- The role of formant synchrony in the coherence of vowels

Peter C. Gordon
Erika Manning
Dept. of Psych., Univ. of North Carolina

- Effects of speaking rate and stress on vowel durations and formant structures

Janet W. Stack
Dept. of Psychol., Univ. of South Florida

- Variability of vowel formant frequencies and speaker normalization: the cases of Mandarin and Taiwanese

Kuo-You Huang
Psych. Dept. Natl. Chung-Cheng Univ. Taiwan

Chin-Hsing Tseng
Natl. Kaohsiung Normal Univ. Taiwan

- vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques

W. Tecumseh Fitch

- A comparison between the acoustics of vowel and consonant reduction

R.JJ.H. van Son and Louis C.W. Pols
Institute of Phonetic Sciences
University of Amsterdam

- Is babbling language-specific? A listening test using vocalizations produced by Swedish and American 12-18 month olds

Karen Williams and Francisco Lacerda
Department of Linguistics, Stockholm Univ.

- Speaker Database Test and Fundamental Frequency in Speech

Tuija Niemi-Laitinen and Poivikki Eskelinen
Department of Phonetics Finland

- Acoustic Phonetics

Dieter Maurer
Neurology, University Hospital Zurich

- The perception of speech under adverse acoustic conditions

Peter Assmann and Quentin Summerfield

- Long Island Vowels

Marie K. Huffman
Elyse Tamberino
Department of Linguistics
State University of New York

- A comparative study of American English and Korean vowels produced by male and female speakers

Byunggon Yang
Department of English, College of Humanities, Dongeui University

