



**ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**  
Μεταπτυχιακό Πρόγραμμα Σπουδών

**Δημητρίου Οικονόμου (Α.Μ. 198)**  
**Ιωάννη Πρίγγουρη (Α.Μ. 206)**

**Speech Viewer vs Snack: Δύο πακέτα για την  
επεξεργασία και ανάλυση ομιλίας**

Εργασία στο μάθημα: Επικοινωνία με Ομιλία  
Διδάσκων: Γεώργιος Κουρουπέτρογλου

Αθήνα 1999

## ΠΕΡΙΕΧΟΜΕΝΑ

<b><u>ΜΕΡΟΣ ΠΡΩΤΟ:SpeechViewer</u></b>	3
Εισαγωγή	3
Περίληψη	4
Κεφάλαιο 1: Ξεκινώντας τον Speech Viewer	5
1.1)Menus	5
1.2)Toolbar	5
Κεφάλαιο 2 : Φόρτωμα ενός σήματος	7
2.1)Groups και Windows	8
2.2)Μαρκάρισμα	9
Κεφάλαιο 3 : Επεξεργασία Κυματομορφής	10
3.1)Options	10
3.2)Wave Cut	11
3.3)Play File	12
3.4)Play Window	12
3.5)Play Highlighted	12
Κεφάλαιο 4 : Επεξεργασία Φασματογραφήματος	13
4.1)Φασματογράφημα	13
4.2)Παράμετροι ενός φασματογραφήματος	14
Κεφάλαιο 5 : Επισημείωση(Labelling)	17
Κεφάλαιο 6 : Παραδείγματα-Εφαρμογές	19
6.1)Παραδείγματα	19
6.1.1)Ηχογράφηση κυματομορφής	19
6.1.2)Φασματογράφημα	19
6.1.3)Επισημείωση	21
6.2)Εφαρμογές	21
6.2.1)Μεταβολή του σήματος με τον χρόνο	21
6.2.2)Μεταβολή του σήματος με τον τόνο του ομιλητή	22
6.1.3)Εκφώνηση μιας λέξης από διαφορετικούς ομιλητές	24
<b><u>ΜΕΡΟΣ ΔΕΥΤΕΡΟ:Snack</u></b>	27
Εισαγωγή	27
Περίληψη	28
Κεφάλαιο 1: Ξεκινώντας με τον Snack	29
1.1)Παράθυρα	29
1.2)Μαρκάρισμα	30
Κεφάλαιο 2 : Επεξεργασία κυματομορφής	31
2.1)Toolbar	32
2.2)Menus	32
2.2.1)File	32
2.2.2)Edit	32
2.2.3)Audio	33
2.2.4)Tools	33
2.2.5)Options	33
2.2.6)Window	34
Κεφάλαιο 3 : Επισημείωση(Labelling)	35
Κεφάλαιο 4 : Φασματογράφημα και Φάσμα	37
4.1)Φασματογράφημα	37
4.1.1)Εμφάνιση φασματογραφήματος	37
4.1.2)Δημιουργία Φασματογραφήματος	38
4.2)Φάσμα(spectrum)	39
4.2.1)Παράμετροι δημιουργίας φάσματος	40
Κεφάλαιο 5 : Παραδείγματα-Εφαρμογές	41

5.1) Παραδείγματα	41
5.1.1) Ηχογράφηση κωματομορφής	41
5.1.2) Φασματογράφημα	41
5.1.3) Επισημείωση	43
5.1.4) Spectrum	43
5.2) Εφαρμογές	44
5.2.1) Λεπτή δομή του φάσματος 2 φωνημάτων	44
<b><u>ΜΕΡΟΣ ΤΡΙΤΟ: SpeechViewer vs Snack</u></b>	46
Βιβλιογραφία-πηγές	49
Πίνακας αγγλικών όρων	50

## **ΜΕΡΟΣ ΠΡΩΤΟ : SnackViewer**

### **ΕΙΣΑΓΩΓΗ**

Το CSLU speech Toolkit είναι ένα ολοκληρωμένο πακέτο για την επεξεργασία ομιλίας το οποίο αναπτύχθηκε από το Κεντρο Κατανόησης Ομιλούμενης Γλώσσας (Center of Spoken Language Understanding) του Πανεπιστημίου του Oregon. Για τον σκοπό αυτό περιέχει μια σειρά από εργαλεία για την αναγνώριση, ανάλυση και σύνθεση ομιλίας. Ο speechViewer είναι ένα από τα πολλά αυτά εργαλεία και ασχολείται με την επεξεργασία και ανάλυση της ομιλίας. Στη παρούσα εργασία θα γίνει μια αναλυτική παρουσίαση του SpeechViewer καθώς και των δυνατοτήτων που προσφέρει στον χρήστη του για την επεξεργασία ενός σήματος ομιλίας.

Εδώ θα ήταν, επίσης, χρήσιμο να αναφέρουμε τις απαιτήσεις που έχει το πρόγραμμα για να λειτουργήσει ικανοποιητικά. Αυτές είναι:

Απαιτήσεις συστήματος:

Intel x86 200MHz Pentium class 200MHz επεξεργαστής ή ανώτερος  
32MB ή παραπάνω μνήμη (64MB συστήνεται)  
100MB χώρος στον σκληρό δίσκο  
Soundcard (SoundBlaster συμβατή)

Προσοχή οι παραπάνω απαιτήσεις προτείνονται για την ικανοποιητική λειτουργία ολόκληρου του πακέτου. Κάποια από τα επιμέρους εργαλεία πιθανόν να έχουν χαμηλότερες απαιτήσεις.

Τέλος, θα πρέπει να αναφέρουμε, εδώ, ότι ο SpeechViewer δεν είναι ένα αυτόνομο πρόγραμμα που μπορεί να εκτελεστεί μόνο του, αλλά προκειται για μια επέκταση γραμμένη σε Tcl/Tk. Για όσους δεν γνωρίζουν πρόκειται για έναν μεταφραστή που μπορεί να μεταφράζει ένα text αρχείο γραμμένο ειδικά για αυτόν σε εκτελέσιμο κώδικα. Έτσι κάθε φορά που θέλεις να καλέσεις μια εφαρμογή γραμμένη στη συγκεκριμένη γλώσσα πρέπει πρώτα να καλέσεις τον μεταφραστή (wish80.exe) και εν συνεχεία μέσα από αυτόν την εφαρμογή.

## ΠΕΡΙΛΗΨΗ

Ο Speech Viewer είναι ένα πρόγραμμα για την εμφάνιση ,αναπαραγωγή ,εγγραφή,κοιμμό και επικόλληση ηχητικών κυματομορφών.Μπορεί να χρησιμοποιηθεί επίσης για την κατασκευή του φασματογραφήματος της ηχητικής κυματομορφής , σύνταξη και εμφάνιση ετικετών που αντιστοιχούν σε αυτή και τέλος για την εμφάνιση συνοδευτικών αρχείων αυτής.

Από το κυρίως μενού,μπορούμε να ανοίξουμε μια υπάρχουσα κυματομορφή ή μπορούμε να ηχογραφήσουμε μια καινούργια.Αυτή η κυματομορφή εμφανίζεται μαζί με ένα παράθυρο απεικόνισης (που περιέχει όλη την κυματομορφή) και με μια μπάρα κύλισης.Κάτω από το παράθυρο απεικόνισης εμφανίζεται ένα τμήμα της κυματομορφής αυτό που είναι μαρκαρισμένο με πράσινο χρώμα στο παράθυρο αντιστοίχισης.Μπορούμε να προβάλλουμε και περισσότερες από μια κυματομορφές συνχρόνως με το πρόγραμμα αυτό.Το Speech View παρέχει δυνατότητες για την επεξεργασία της κυματομορφής (αφαίρεση κάποιου τμήματος του ηχητικού σήματος,επικόλληση μιας κυματομορφής σε κάποιο τμήμα μιας άλλης κτλ).

Μια από τις κύριες εργασίες που μπορεί να επιτελέσει το συγκεκριμένο πρόγραμμα είναι η εύρεση του φασματογραφήματος της κυματομορφής και η προβολή του κάτω ακριβώς από αυτή.Η απεικόνιση του φασματογραφήματος μπορεί να γίνει με δύο διαφορετικούς τρόπους.Ο χρήστης μπορεί να μεταβάλλει κάποιες παραμέτρους για την κατασκευή του φασματογραφήματος (όπως μήκος παραθύρου-window size-) καθώς και μερικές άλλες για την επιλογή των χρωμάτων και των αποχρωσεών τους που χρησιμοποιούνται στο φασματογράφημα.

Εκτός από αυτά που αναφέρθηκαν έως τώρα , με τον Speech Viewer μπορούμε να επισυνάψουμε ετικέτες (labels) σε ένα ηχητικό σήμα.Για την εμφάνιση των ετικετών υπάρχει κάτω από κάθε κυματομορφή ένας ειδικός χώρος.Με τον όρο ετικέτα εννοούμε ένα σύνολο φωνημάτων που αντιστοιχίζονται σε ένα τμήμα της κυματομορφής με σκοπό να μας δείχνουν αυτό που θα ακούσουμε στο συγκεκριμένο κομμάτι του σήματος.

Στα κεφάλαια που ακολουθούν ο αναγνώστης μπορεί να βρει έναν λεπτομερή εγχειρίδιο χρήσης του προγράμματος καθώς και αναλυτική περιγραφή των όρων του.

## 1) ΞΕΚΙΝΩΝΤΑΣ ΤΟΝ SPEECH VIEWER

Ο Speechviewer σου παρέχει τη δυνατότητα είτε να φορτώσεις ένα αρχείο ήχου που είναι ήδη αποθηκευμένο στο δίσκο είτε να ηχογραφήσεις επιτόπου τη φωνή σου μέσω ενός μικροφώνου, εφόσον ο υπολογιστής σου διαθέτει. Τα αρχεία που μπορείς να φορτώσεις πρέπει να είναι τυποποιημένα σύμφωνα με το πρότυπο wave.



Σχήμα 1.1

Ξεκινώντας το πρόγραμμα εμφανίζεται ένα παράθυρο που περιλαμβάνει 3 δυνατά menus επιλογών καθώς και μια Toolbar με 5 εικονίδια όπως ακριβώς φαίνεται στο παραπάνω σχήμα (σχήμα 1.1).

### 1.1) Menus

#### 1) File

Περιλαμβάνει τις στοιχειώδεις λειτουργίες του προγράμματος. Οι επιπλέον υποεπιλογές που έχει είναι οι ακόλουθες 3:

**Open Wave File** για το φόρτωμα ενός ήδη αποθηκευμένου στο δίσκο αρχείου.

**Record Wave** για την ηχογράφιση της φωνής σου με ένα μικρόφωνο.

**Exit** για έξοδο από το πρόγραμμα.

#### 2) Options

Διάφοροι παράμετροι που αφορούν τον Speech Viewer. Διεξοδικότερα περιλαμβάνει τα ακόλουθα:

**Recording Sample Rating** που σου επιτρέπει να ορίσεις τον επιθυμητό ρυθμό δειγματοληψίας σε samples/sec. Η παράμετρος αυτή επηρεάζει τη ποιότητα του ηχογραφημένου σήματος καθώς μεγάλοι ρυθμοί συνεπάγονται μεγαλύτερη πιστότητα σήματος.

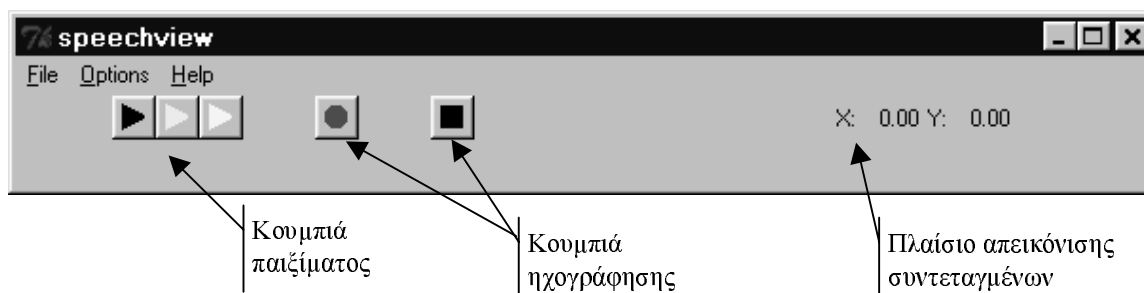
**Labeling Scroll Options**

**Save Current Parameters As Defaults** με την οποία σώζεις τις επιλογές που έχεις κάνει στα a) και b) ώστε κάθε φορά που ανοίγεις τον Speech Viewer αυτός να προεπιλέγει τις παραμέτρους που έχεις ορίσει.

#### 3) Help

Βοήθεια πάνω στη χρήση του Speech Viewer

### 1.2) Toolbar



Σχήμα 1.2

Περιλαμβάνει 2 κατηγορίες κουμπιών και ένα πλαίσιο απεικόνισης συντεταγμένων.

### **Κουμπιά παιξίματος**

Είναι τρία στον αριθμό, όπως φαίνεται και στο σχήμα 1.2 και χρησιμεύουν για να ακούσεις το σήμα που επεξεργάζεσαι. Συνοπτικά η λειτουργία του καθενός είναι η ακόλουθη:

- **Μαύρο** : Παίξει ολόκληρη την υπό επεξεργασία επιλεγμένη κυματομορφή.
- **Πράσινο** : Παίξει το κομμάτι εκείνο της επιλεγμένης κυματομορφής που είναι μαρκαρισμένο με πράσινο χρώμα.
- **Κίτρινο** : Παίξει το κομμάτι εκείνο της επιλεγμένης κυματομορφής που είναι μαρκαρισμένο με κίτρινο χρώμα.

Το πώς ακριβώς γίνεται το μαρκάρισμα των κυματομορφών θα αναλυθεί διεξοδικότερα στο επόμενο κεφάλαιο (§2.2).

### **Κουμπιά ηχογράφησης**

Είναι δύο, όπως φαίνεται και στο σχήμα 1.2:

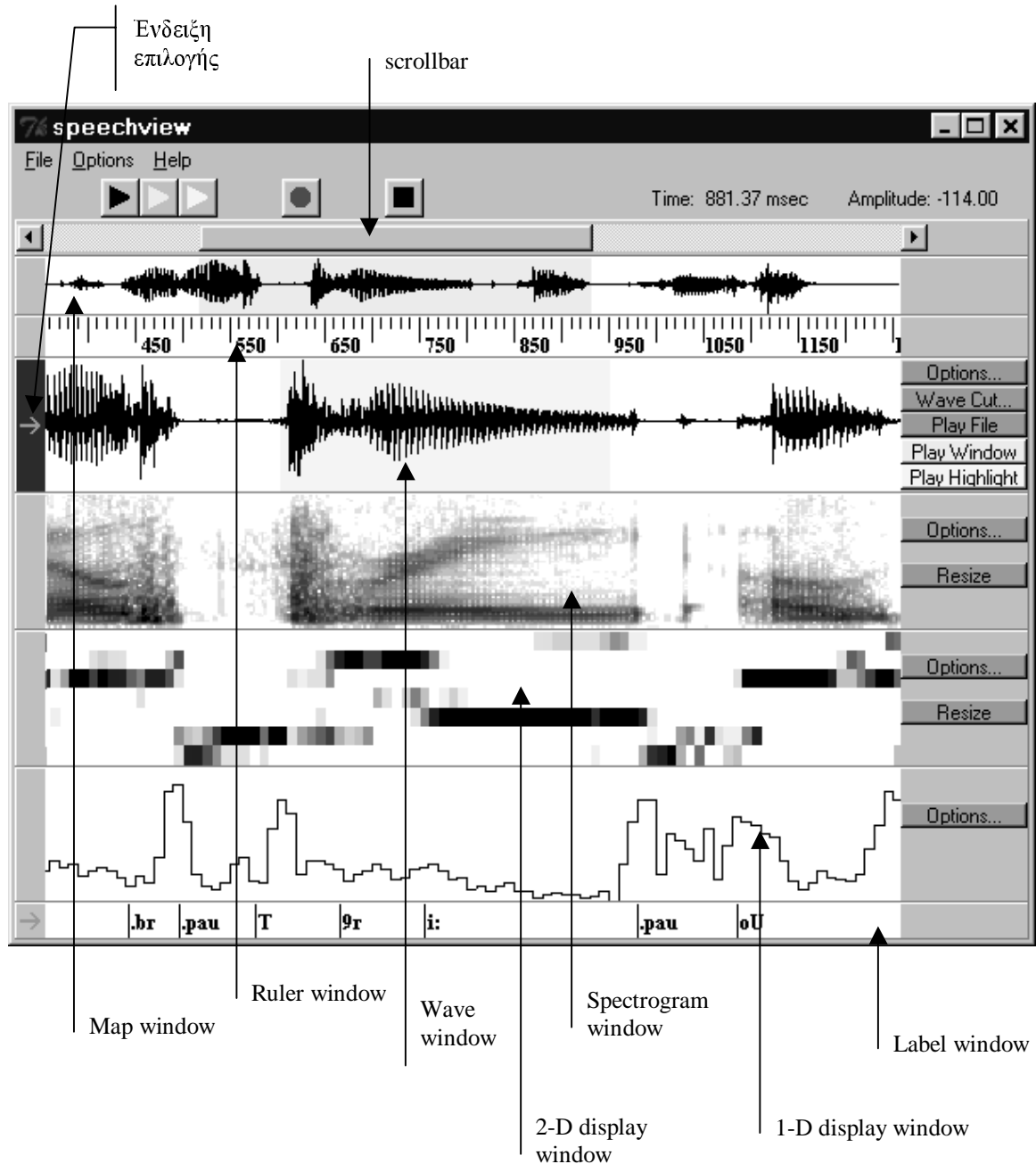
- **Κόκκινο** : Το πάτημά του σηματοδοτεί την έναρξη μιας ηχογράφησης. Κατά τη διάρκεια της ηχογράφησης το κουμπί αναβοσβήνει ενώ στα δεξιά φαίνεται και η χρονική διάρκεια της ηχογράφησης.
- **Μαύρο** : Το πάτημά του διακόπτει την τρέχουσα ηχογράφηση.

### **Πλαίσιο απεικόνισης συντεταγμένων**

Απεικονίζει τις συντεταγμένες ενός παραθύρου απεικόνισης(βλέπε κεφάλαιο 2).Η μορφή που έχει στο σχήμα 1.2 είναι η default μορφή που έχει όταν πρωτοανοίγουμε τον Speech Viewer.

## 2) ΦΟΡΤΩΜΑ ΕΝΟΣ ΣΗΜΑΤΟΣ

Το φόρτωμα ενός σήματος μπορεί να γίνει είτε από κάποιο αρχείο που είναι ήδη αποθηκευμένο στο δίσκο είτε με ηχογράφηση από ένα μικρόφωνο. Και στις 2 περιπτώσεις το αποτέλεσμα είναι η δημιουργία 3 παραθύρων (windows) τα οποία αποτελούν μια ομάδα (group). Παρακάτω θα δούμε τι είναι το καθένα από αυτά τα παράθυρα διεξοδικότερα.



Σχήμα 2.1



## 2.1) Groups και Windows

### Group(ομάδα)

Πρόκειται για ένα σύνολο από παράθυρα που είναι συνδεδεμένα με την ίδια κυματομορφή. Ένα group μπορεί να περιέχει 7 ειδών παράθυρα και μια μπάρα κύλισης (scrollbar). Πρακτικά κάθε νέα κυματομορφή που φορτώνουμε ή ηχογραφήσουμε ανήκει σε ξεχωριστό group.

#### παρατήρηση

Στη περίπτωση που επεξεργαζόμαστε πολλές κυματομορφές ταυτόχρονα με τον Speech Viewer το ποια χειριζόμαστε κάθε φορά καθορίζεται από την **ένδειξη επιλογής** όπως φαίνεται και στο ΣΧΗΜΑ 2.1 Η κυματομορφή, δηλαδή, που έχει την ένδειξη αυτή κοκκινισμένη αποτελεί την κυματομορφή που βρίσκεται υπό επεξεργασία. Επιλογή μπορεί να γίνει μόνο στα παράθυρα που φέρουν τον βελάκι επιλογής.

### Windows(παράθυρα)

- 1) **Map window (παράθυρο απεικόνισης)** : Περιέχει ολόκληρη την κυματομορφή του σήματος που πρόκειται να επεξεργαστούμε. Με dragging του αριστερού κουμπιού του mouse μπορούμε να μαρκάρουμε ένα κομμάτι του σήματος με πράσινο χρώμα και το οποίο κομμάτι απεικονίζεται μεγεθυμένο στο wave window (βλ. παρακάτω). Εφόσον έχει μαρκαριστεί κάποιο κομμάτι πατώντας το αριστερό κουμπί του mouse οπουδήποτε μέσα στο παράθυρο ο Speech Viewer θα παίξει αυτό το κομμάτι. Κάθε group μπορεί να έχει ένα μόνο τέτοιο παράθυρο.
- 2) **Ruler window (παράθυρο χρονικού άξονα)** : Ο χρονικός άξονας που αντιστοιχεί στο wave window (βλ. αμέσως μετά). Κάθε group περιέχει ένα μόνο τέτοιο παράθυρο.
- 3) **Wave window (παράθυρο επεξεργασίας)** : Το παράθυρο αυτό απεικονίζει το κομμάτι εκείνο της κυματομορφής που είναι μαρκαρισμένο με πράσινο χρώμα στο Map window. Μέσω της scrollbar μπορούμε να μετακινηθούμε σε οποιαδήποτε περιοχή της κυματομορφής αλλά το παράθυρο θα απεικονίζει πάντοτε ένα κομμάτι χρονικής διάρκειας ίσης με το πράσινα μαρκαρισμένο τμήμα. Το παράθυρο αυτό είναι το βασικό παράθυρο μέσω του οποίου γίνεται όλη η επεξεργασία του σήματος που έχουμε φορτωμένο. Η επεξεργασία αυτή πραγματοποιείται μέσα από τα menus που υπάρχουν στο δεξιό άκρο του παραθύρου και τα οποία αναλύονται σε επόμενη παράγραφο. Όπως και στο map window έτσι και εδώ μπορούμε να μαρκάρουμε κομμάτι του σήματος με αριστερό dragging του mouse. Το μαρκαρισμένο τμήμα, στη περίπτωση αυτή έχει κίτρινο χρώμα και με ένα αριστερό click του mouse οπουδήποτε μέσα στο παράθυρο ο Speech Viewer θα το παίξει. Με το παράθυρο αυτό και τις λειτουργίες του θα ασχοληθούμε αναλυτικά παρακάτω (κεφάλαιο 3). Υπάρχει ένα μόνο wave window σε κάθε group.
- 4) **Spectrogram window (παράθυρο φασματογραφήματος)** : Το παράθυρο αυτό περιέχει το φασματογράφημα του σήματος που επεξεργαζόμαστε. Στο παράθυρο παρουσιάζεται κάθε φορά ένα κομμάτι, μόνο, του συνολικού φασματογραφήματος και συγκεκριμένα αυτό που αντιστοιχεί στο σήμα που απεικονίζεται στο wave window. Και αυτό το παράθυρο θα μας απασχολήσει περισσότερο παρακάτω (κεφάλαιο 3), ενώ εδώ πρέπει να τονίσουμε ότι ένα group μπορεί να περιλαμβάνει περισσότερα από ένα τέτοια παράθυρα.
- 5) **Label window (παράθυρο ετικέτας)** : Το παράθυρο αυτό μας επιτρέπει να βάλουμε ετικέτες, υπό τη μορφή κειμένου, σε συγκεκριμένα σημεία της κυματομορφής μας. Ένα group δύναται να περιέχει πολλά label windows. Το ποιο label window χειριζόμαστε κάθε φορά καθορίζεται από μια ένδειξη επιλογής αντίστοιχη με αυτήν που προαναφέραμε. Και αυτό το παράθυρο θα μας απασχολήσει αναλυτικότερα παρακάτω (κεφάλαιο 5).
- 6) **Two-dimensional display window (παράθυρο διδιάστατης απεικόνισης).**
- 7) **One-dimensional display window (παράθυρο μονοδιάστατης απεικόνισης)** : Δείχνει την μεταβολή των LPC παραμέτρων με το χρόνο.

## 2.2) Μαρκάρισμα

Αν και πιθανότατα θα έχει γίνει ήδη αντιληπτό σημειώνουμε ότι υπάρχουν δύο ειδών μαρκάρια. Και τα δύο μαρκάρια πραγματοποιούνται με αριστερό dragging του mouse πάνω στο αντίστοιχο παράθυρο:

- a) **πράσινο μαρκάρισμα** στο map window
- b) **κίτρινο μαρκάρισμα** στο wave window.

Στο σχήμα 2.1 φαίνονται και οι δύο παραπάνω μορφές μαρκάριατος.

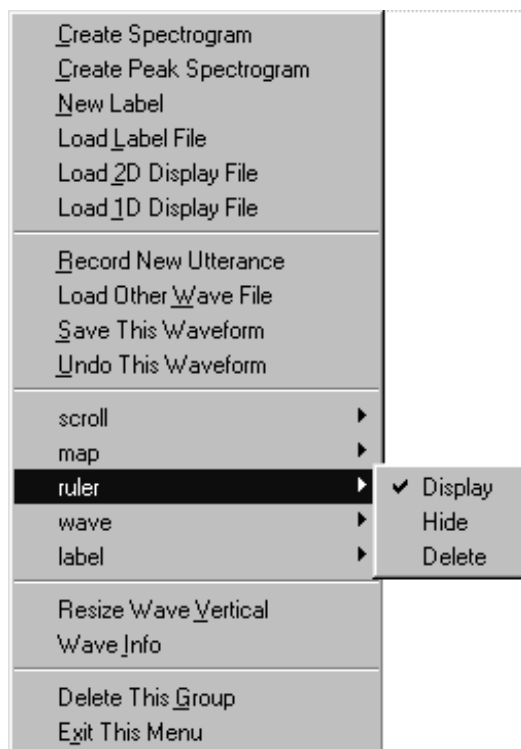
### 3) ΕΠΕΞΕΡΓΑΣΙΑ ΚΥΜΑΤΟΜΟΡΦΗΣ

Στο πρώτο μέρος αυτού του κεφαλαίου θα ασχοληθούμε με την λειτουργία που επιτελούν τα κουμπιά του SpeechViewer που βρίσκονται δεξιά από το παράθυρο της κυματομορφής. Τα κουμπιά αυτά φέρουν τις ακόλουθες επιγραφές:

- 1)Options
- 2)Wave Cut
- 3)Play
- 4)Play Window
- 5)play Highlighted Region

#### 3.1) Options

Επιλέγοντας αυτό το κουμπί εμφανίζεται το μενού του σχήματος 3.1



Σχήμα 3.1

#### Περιγραφή των επιλογών:

**Create Spectrogram:** Με την επιλογή αυτή ο SpeechViewer υπολογίζει και απεικονίζει το φασματογράφημα της κυματομορφής. Η απεικόνιση γίνεται σε ένα spectrogram window το οποίο ανοίγει ειδικά γι' αυτόν το σκοπό κάτω από όλα τα αλλά ήδη ανοιγμένα παράθυρα.

**Create Peak Spectrogram:** Υπολογίζει το φασματογράφημα κορυφής και το προβάλλει σε ένα spectrogram window.

**New Label:** Η αναλυτική περιγραφή αυτής της επιλογής γίνεται στο κεφάλαιο 5.

**Load Label File:** Η επιλογή αυτή ανοίγει ένα παράθυρο διαλόγου για το φόρτωμα ενός ήδη αποθηκευμένου αρχείου ετικετών (label file).

**Load 2D Display File:** Ανοίγει ένα παράθυρο όπουσε προτρέπει να φορτώσεις ένα ήδη αποθηκευμένο 2D αρχείο.Μετά το επιλογή του αρχείου αυτό εμφανίζεται στην οθόνη σε ένα νέο παράθυρο (two dimensional display window).

**Load 1D Display File:** Ανοίγει ένα παράθυρο όπουσε προτρέπει να φορτώσεις ένα ήδη αποθηκευμένο 2D αρχείο..Μετά την επιλογή του αρχείου , αυτό εμφανίζεται στην οθόνη σε ένα νέο παράθυρο (one dimensional display window).

**Record New Utterance:**σηματοδοτεί την έναρξη μιας νέας ηχογράφησης στο τρέχον παράθυρο.

**Load Other Wave File:**Ανοίγει ένα παράθυρο διαλόγου για το φόρτωμα μιας νέας κυματομορφής.

**Undo this Waveform :** Η επιλογή αυτή εμφανίζεται μόνο εάν έχουμε τροποποιήσει τη τρέχουσα κυματομορφή (βλ. §3.2) και σου δίνει τη δυνατότητα να ακυρώσεις την τελευταία αλλαγή που έκανες.

**Save this Waveform:**Αποθηκεύει την τρέχουσα κυματομορφή.

**Scroll:** Με επιλογή αυτή μπορούμε να εμφανίσουμε και να κρύψουμε το Scroll Window.Οι τέσσερις ακόλουθες επιλογές (Map - Ruler - Wave – Spectrogram ) εμφανίζουν και κρύβουν τα τέσσερα υπόλοιπα παράθυρα.

**Resize Wave Vertical:**Η επιλογή αυτή τροποποιεί το μέγεθος του παραθύρου που απεικονίζεται το ηχητικό σήμα.Το ελάχιστο μέγεθος του παραθύρου είναι 50 pixel και το μέγιστο 300 pixel.

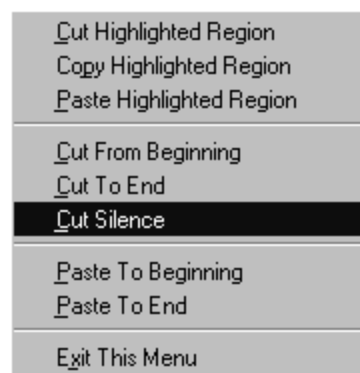
**Wave Info:**Η επιλογή αυτή παρέχει πληροφορίες για την κυματομορφή όπως μέγιστη τιμή ,ελάχιστη τιμή , μέση τιμή , συχνότητα δειγματοληψίας και αριθμό δειγμάτων.

**Delete This Group:**Διαγράφει το αντίστοιχο Group.

**Exit This Menu:**Κλείνει το συγκεκριμένο μενού.

### 3.2) Wave Cut

Πατώντας αυτό το κουμπί εμφανίζεται το μενού του σχήματος 3.2



Σχήμα 3.2

#### Περιγραφή των επιλογών:

**Cut Highlighted Region:**Με την επιλογή αυτή ο χρήστης μπορεί να αφαιρέσει το τμήμα της κυματομορφής που έχει μαρκάρει και συνχρόνως να το τοποθετήσει στον Clipboard.

**Copy highlighted Region:**Αντιγράφει το μαρκαρισμένο τμήμα της κυματομορφής στον Clipboard.

**Paste Highlighted Region:**Τοποθετεί το τμήμα της κυματομορφής που είναι αποθηκευμένο στον Clipboard στο σημείο της κυματομορφής που είναι μαρκαρισμένο με κίτρινο χρώμα.

**Cut From Beginning:**Αφαιρεί το τμήμα της κυματομορφής από την αρχή του σήματος ως και την μαρκαρισμένη περιοχή.

**Cut to End:** Αφαιρεί το τμήμα της κυματομορφής από την μαρκαρισμένη περιοχή ως το τέλος του σήματος.

**Cut Silence:** Αφαιρεί το τμήμα της κυματομορφής από την αρχή του σήματος ως το πρώτο φώνημα.

**Paste To Beginning:**Επισυνάπτει το περιεχόμενο του Clipboard στην άρχη του σήματος.

**Paste To End:** Επισυνάπτει το περιεχόμενο του Clipboard στο τέλος του σήματος.

**Exit This Menu:**Κλείνει το συγκεκριμένο μενού.

### 3.3) Play File

Το κουμπί αυτό αναπαράγει το σήμα του αρχείου.

### 3.4) Play Window

Το κουμπί αυτό αναπαράγει το σήμα που είναι μαρκαρισμένο με πράσινο χρώμα στο παράθυρο αντιστοίχισης (Map Window).

### 3.5) Play Highlighted

Το κουμπί αυτό αναπαράγει το σήμα που είναι μαρκαρισμένο με κίτρινο χρώμα στο παράθυρο αντιστοίχισης (Wave Window).

## 4) ΕΠΕΞΕΡΓΑΣΙΑ ΦΑΣΜΑΤΟΓΡΑΦΗΜΑΤΟΣ

Στο προηγούμενο κεφάλαιο είδαμε πώς, ακριβώς μπορούμε να φτιάξουμε το φασματογράφημα ενός ηχητικού σήματος το οποίο και απεικονίζεται στο spectrogram window. Εδώ θα προχωρήσουμε σε μια πιο διεξοδική ανάλυση των επιλογών που αφορούν το συγκεκριμένο παράθυρο ενώ θα δώσουμε και μια περιγραφή του τι αντιπροσωπεύει ένα φασματογράφημα.

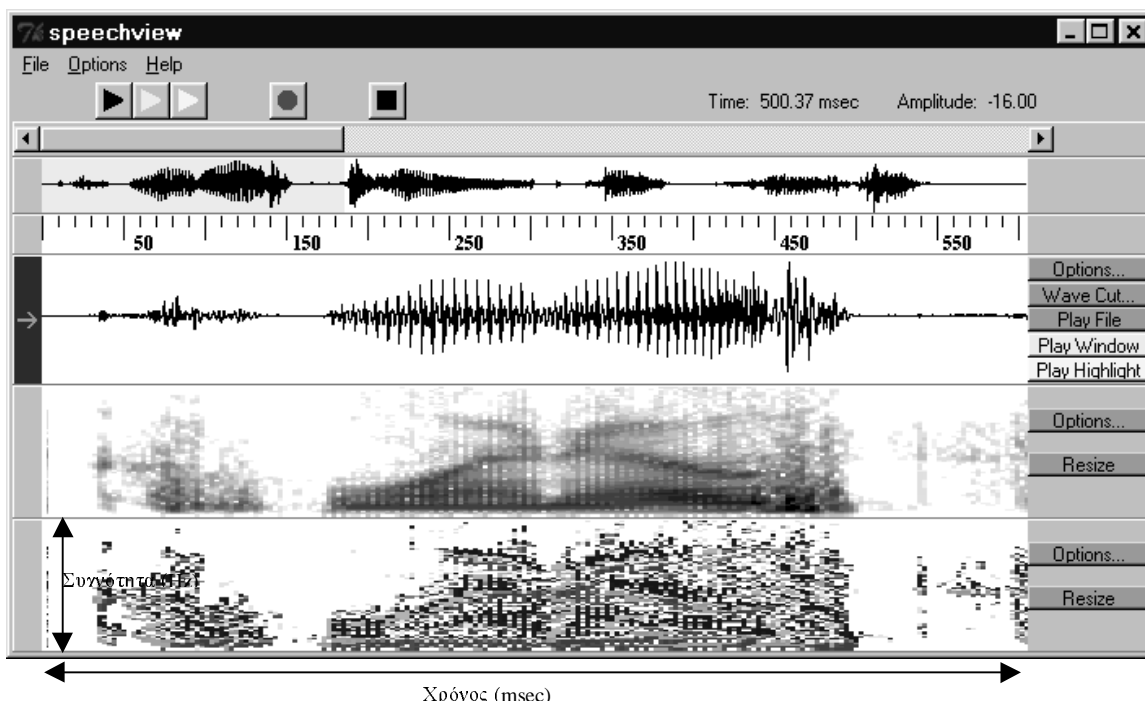
### 4.1) Φασματογράφημα

Ένα φασματογράφημα είναι, ουσιαστικά, μια τρισδιάστατη γραφική παράσταση απεικονισμένη σε δύο διαστάσεις. Όπως φαίνεται και στο σχήμα 4.1 ο οριζόντιος άξονας είναι ο άξονας του χρόνου ενώ ο κάθετος άξονας περιέχει τις συχνότητες. Η τιμή της φασματικής ισχύος του σήματος για συγκεκριμένη τιμή του χρόνου και της συχνότητας αποτελεί την τρίτη διάσταση η οποία απεικονίζεται ως μια χρωματική απόχρωση. Σκούρα απόχρωση σημαίνει υψηλή τιμή φασματικής ισχύος ενώ ανοιχτότερες αποχρώσεις ισοδυναμούν με χαμηλότερες τιμές ισχύος. Στην πράξη αυτό που απεικονίζει η χρωματική ένταση δεν είναι η ισχύς του φάσματος αυτή καθαυτή αλλά ο λογάριθμος της ισχύος.

Παρατηρείστε, επίσης, ότι όταν μετακινούμε τον mouse cursor πάνω από ένα spectrogram window αυτός μετατρέπεται σε ένα σταυρόνυμα. Αυτό γίνεται επειδή, όπως ήδη αναφέραμε, το φασματογράφημα έχει δύο ενεργούς άξονες. Παρατηρείστε, ακόμα, ότι στο πλαίσιο απεικόνισης συντεταγμένων το X έχει πάρει την τιμή **Time** (χρόνος) και το Y την τιμή **Frequency** (συχνότητα) ενώ έχει εμφανιστεί και μια τρίτη μεταβλητή που ονομάζεται **Value**, και η οποία αντιστοιχεί στη λογαριθμημένη φασματική ισχύ του σήματος. Μετακινώντας, τώρα, το σταυρό σε κάποιο σημείο του φασματογραφήματος κάθε μεταβλητή παίρνει την τιμή της αντίστοιχης παραμέτρου στο σημείο αυτό.

Ο Speech Viewer υποστηρίζει δύο ειδών φασματογραφήματα:

- typical spectrogram (τυπικό φασματογράφημα)** : το κλασσικό φασματογράφημα που περιγράφηκε παραπάνω. Η φασματική ισχύς απεικονίζεται ως μια απόχρωση του γκρι.
- Peak spectrogram (φασματογράφημα κορυφής)** : ένα φασματογράφημα όπου τονίζονται χρωματικά οι αλλαγές στις τιμές του φασματικής ισχύος. Συγκεκριμένα οι τιμές της φασματικής ισχύος κβαντίζονται χονδρικά (σε ακέραιες τιμές dB) και οι περιοχές όπου η ισχύς είναι στο ίδιο επίπεδο κβαντισμού απεικονίζεται με το ίδιο χρώμα.

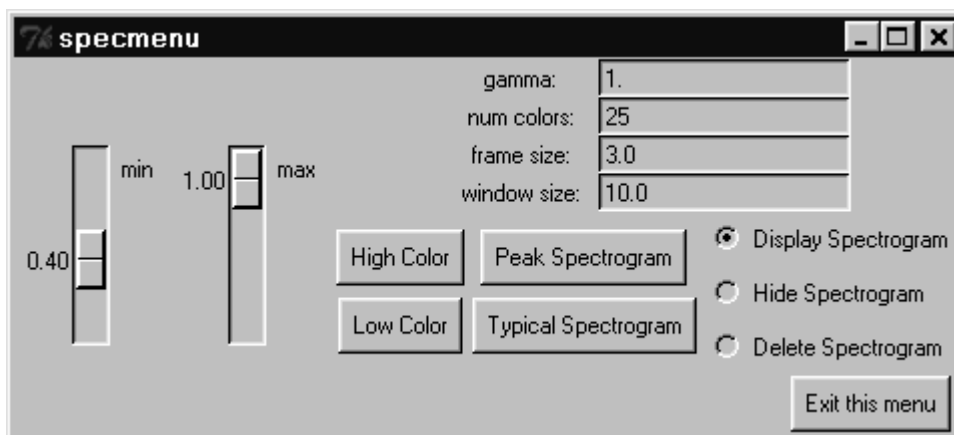


Σχήμα 4.1

## 4.2) Παράμετροι ενός φασματογραφήματος

Όπως φαίνεται και από το σχήμα 4.1 στα δεξιά ενός spectrogram window υπάρχουν 2 μπάρες επιλογής:

- a) **options** : ανοίγει ένα νέο παράθυρο (βλ. σχήμα 4.2) όπου καλείσαι να ορίσεις τις παραμέτρους εμφάνισης και δημιουργίας του φασματογραφήματος. Ακολουθεί αναλυτική περιγραφή του παρακάτω.



Σχήμα 4.2

Οι δύο κάθετες μπάρες ολίσθησης στα αριστερά που φέρουν τον τίτλο **min** και **max** αντίστοιχα σχετίζονται με την οπτική αναπαράσταση του φασματογραφήματος. Πιο συγκεκριμένα έχουμε:

**min** : ορίζει την ελάχιστη τιμή της μεταβλητής Value που θα απεικονιστεί στο spectrogram window. Με τον όρο απεικονιστεί εννοούμε ότι θα αντιστοιχηθεί σε κάποια απόχρωση διαφορετική από τη λευκή. Παραδειγματοσχάρι η τιμή 0,40 που φαίνεται παραπάνω υποδηλώνει ότι οποιαδήποτε τιμή του φασματογραφήματος μικρότερη από το 40% της μέγιστης τιμής της φασματικής ισχύος (σε dB) που εμφανίζεται σε όλο το φασματογράφημα θα απεικονιστεί με άσπρο χρώμα.

**max** : ορίζει τη μέγιστη τιμή της μεταβλητής Value που θα απεικονιστεί στο spectrogram window. Κι εδώ με τον όρο απεικονιστεί εννοούμε αντιστοίχιση σε μια απόχρωση διαφορετική από τη μαύρη, ενώ η τιμή που ορίζουμε, όπως και προηγουμένως, είναι σε ποσοστό, σε σχέση με την μέγιστη τιμή της φασματικής ισχύος, που εμφανίζεται στο φασματογράφημα. Έτσι η τιμή 1.00 που φαίνεται στο παραπάνω σχήμα σημαίνει ότι στη μαύρη απόχρωση θα απεικονιστεί μία μόνο τιμή: η μέγιστη τιμή φασματικής ισχύος.

Και οι δύο παραπάνω επιλογές έχουν ως στόχο να βοηθήσουν τον χρήστη του προγράμματος να διακρίνει καλύτερα τις περιοχές υψηλής και χαμηλής φασματικής ισχύος.

Δίπλα, ακριβώς, από τις μπάρες αυτές υπάρχουν δυο κουμπιά που φέρουν τον τίτλο high color και low color αντίστοιχα. Η λειτουργία τους αναλύεται παρακάτω:

**High color** : ανοίγει ένα νέο παράθυρο όπου σε καλεί να ορίσεις την χρώμα και την απόχρωση ειδικότερα στο οποίο θα απεικονίζονται οι υψηλές τιμές της μεταβλητής Value του φασματογραφήματος(βλ. σχήμα 4.3).

**Low color** : ανοίγει ένα νέο παράθυρο όπου σε καλεί να ορίσεις την χρώμα και την απόχρωση ειδικότερα στην οποία θα απεικονίζονται οι χαμηλές τιμές της μεταβλητής Value του φασματογραφήματος(βλ. σχήμα 4.3).

Ορίζοντας κάποιο διαφορετικό high color ή low color μπορούμε να ξεφύγουμε από τη κλασική ασπρόμαυρη αναπαράσταση και να χρησιμοποιήσουμε για την απεικόνιση του φασματογραφήματος διαφορετικές αποχρώσεις από αυτές του γκρι. Τέλος οι επιλογές **typical spectrogram** και **peak spectrogram** σου επιτρέπουν να εναλλάξεις οποιαδήποτε από τις δύο μορφές φασματογραφήματος στο ίδιο παράθυρο.

#### Παρατήρηση

Όλες οι παραπάνω επιλογές αφορούν την εμφάνιση του τυπικού φασματογραφήματος ενώ δεν επηρεάζουν καθόλου την αντίστοιχη του φασματογραφήματος κορυφής.

Τέλος υπάρχουν κάποιες παράμετροι που έχουν σχέση με τη δημιουργία του φασματογραφήματος και τις οποίες ο χρήστης καλείται να ορίσει μόνος του. Κατά τη δημιουργία ενός φασματογραφήματος οι παράμετροι αυτοί παίρνουν κάποιες default τιμές οι οποίες φαίνονται στο σχήμα 4.2. Οι παράμετροι αυτοί είναι:

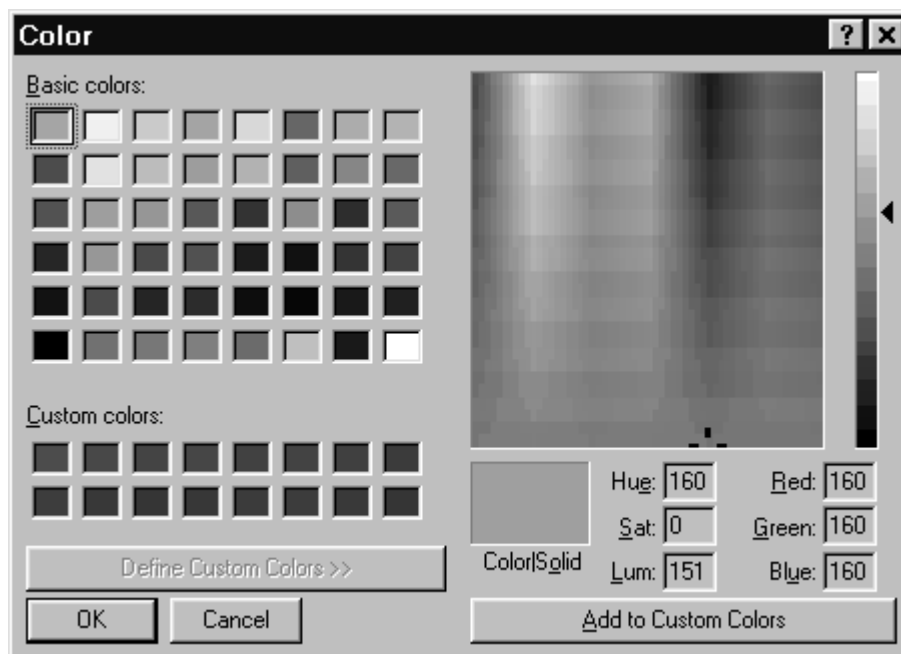
**gamma** : αυξάνει (μικρή τιμή) ή μειώνει (μεγάλη τιμή) το contrast σε ένα τυπικό φασματογράφημα.

**num colors** : ορίζει το πλήθος των αποχρώσεων σε ένα τυπικό φασματογράφημα.

**frame size** : ορίζει το μήκος του πλαισίου (σε msec) κατά το οποίο θα λαμβάνονται δείγματα της κυματομορφής για τη δημιουργία του φασματογραφήματος.

**window size** : ορίζει το μήκος του χρονικού παραθύρου (σε msec) που χρησιμοποιείται κατά τη δημιουργία ενός φασματογραφήματος. Η τιμή του πρέπει να είναι υποχρεωτικά μεγαλύτερη από αυτήν του frame, ώστε να υπάρχει κάποια επικάλυψη, ή το λιγότερο ίση, ώστε να μην έχουμε απώλειες δειγμάτων. Έτσι για παράδειγμα οι τιμές 3 και 10 αντίστοιχα που φαίνονται στο σχήμα 4.2 δηλώνουν ότι κάθε 3 msec θα παίρνουμε ένα κομμάτι του σήματος μήκους 10 msec για την παραγωγή του φασματογραφήματος.

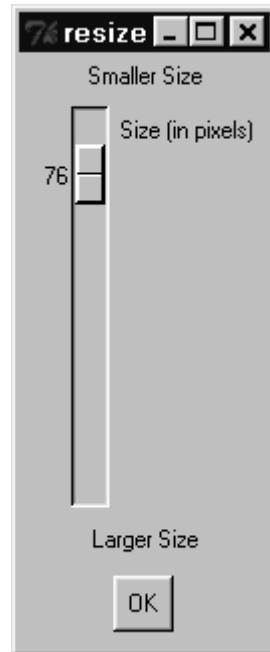
Οι δύο τελευταίες επηρεάζουν την ποιοτική εμφάνιση τόσο του τυπικού φασματογραφήματος όσο και του φασματογραφήματος κορυφής.



Σχήμα 4.3



- b) **resize** : ανοίγει ένα παράθυρο όπως αυτό του διπλανού σχήματος στο οποίο καλείσαι να ορίσεις το ύψος του spectrogram window σε pixels. Η default τιμή της παραμέτρου αυτής είναι 150 pixels, ενώ η ελάχιστη δυνατή είναι 50 pixels. Η αλλαγή του ύψους ισοδυναμεί στην ουσία με ένα scaling στον άξονα των συχνοτήτων.



Σχήμα 4.4

## 5) ΕΠΙΣΗΜΕΙΩΣΗ (LABELLING)

Με τον όρο ετικέτα εννοούμε ένα σύνολο με ένα ή περισσότερα φωνήματα τα οποία αντιστοιχούν σε ένα τμήμα της κυματομορφής και δείχνουν τι ήχο παριστάνει το σήμα. Οι ετικέτες ενός σήματος εμφανίζονται στο παράθυρο της ετικέτας. Με το φόρτωμα ενός αρχείου τύπου wav το παράθυρο της ετικέτας δεν εμφανίζεται στην οθόνη. Σε ένα σήμα είναι δυνατόν να βάλουμε περισσότερες από μια ετικέτες χωρίζοντας το παράθυρο σε αντίστοιχους τομείς. Είναι δυνατόν να τοποθετηθούν και περισσότερα από ένα παράθυρα με ετικέτες σε κάθε σήμα. Οι ετικέτες είναι δυνατόν να αποθηκευτούν σε αρχεία και να φορτωθούν ξανά από αυτά. Τα παραπάνω γίνονται ως εξής:

### 1) Δημιουργία ενός παράθυρου ετικετών (label window):

Για να δημιουργηθεί ένα τέτοιο παράθυρο κάνουμε 'κλικ' στο κουμπί **Options** (που βρίσκεται δεξιά από το wave window) και στην συνέχεια επιλέγουμε **New Label**.

### 2) Χωρισμός ενός παραθύρου σε τομείς:

Αφού δημιουργήσουμε ένα παράθυρο ετικετών κάνουμε 'κλικ' στην ένδειξη επιλογής και το παράθυρο αλλάζει χρώμα (από άσπρο σε γκρι). Όταν έχουμε δημιουργήσει το παράθυρο τοποθέτησης ετικετών τότε μπορούμε να το χωρίσουμε σε τομείς πατώντας το πλήκτρο Enter. Από τους τομείς αυτούς ένας έχει γκρι χρώμα και ονομάζεται ενεργός και οι υπόλοιποι έχουν άσπρο. Κάθε φορά που το πατάμε δημιουργείται και ένας τομέας μήκους 100 msec και γίνεται αυτόματα ενεργός. Κάνοντας 'κλικ' σε έναν τομέα με το αριστερό κουμπί του ποντικιού γίνεται αυτόματα ενεργός. Πατώντας το πλήκτρο Insert μπορούμε να διαιρέσουμε έναν τομέα σε δυο τμήματα. Τέλος με το πλήκτρο Delete μπορούμε να διαγράψουμε ένα τομέα. Για να αλλάξουμε το χρονικό διάστημα που αντιστοιχεί ένας τομέας πηγαίνουμε πάνω από την κάθετη γραμμή που χωρίζει δυο τομείς το ποντίκι και ο κέρσορας μετατρέπεται αυτόματα σε δύο βέλη. Μετακινώντας αριστερά ή δεξιά τα βέλη μεταβάλλουμε το μέγεθος ενός τομέα.

### 3) Ενεργοποίηση ενός τομέα:

Η ενεργοποίηση ενός τομέα μπορεί να γίνει με τους ακόλουθους τρόπους:

- A) πατώντας το 'Enter' γίνεται αυτόματα ενεργός ο επόμενος τομέας (αν δεν υπάρχει δημιουργείται).
- B) πατώντας το 'Tab' γίνεται αυτόματα ενεργός ο επόμενος τομέας (αν δεν υπάρχει δημιουργείται).
- Γ) Κάνοντας 'κλικ' με το αριστερό κουμπί του ποντικιού σε έναν οποιοδήποτε τομέα.

### 4) Τοποθέτηση κειμένου σε έναν τομέα:

Εάν ένας τομέας είναι ενεργός, ό,τι πληκτρολογήσουμε στο πληκτρολόγιο εμφανίζεται αυτόματα σε αυτόν. Μπορούμε να διαγράψουμε το περιεχόμενο ενός τομέα εάν είναι ενεργός με το 'backspace'. Ένας τομέας δεν είναι απαραίτητο να έχει κείμενο.

### 5) Αποθήκευση μιας ετικέτας:

Κάνοντας 'κλικ' το δεξιό κουμπί του mouse σε μια ετικέτα εμφανίζεται ένα μενού επιλογών στο οποίο υπάρχει και η αποθήκευση. Η αποθήκευση ενός τέτοιου παραθύρου γίνεται σε ένα αρχείο με κατάληξη phn που έχει την ακόλουθη δομή:

```

Millisecondsperframe 1.0      (Επικεφαλίδα αρχείου)
END OF HEADER
0 20 .pau                      (Κύριο τμήμα αρχείου)
20 45 no

```

Στην επικεφαλίδα του αρχείου αποθηκεύεται η τιμή του frame. Το κύριο μέρος του αρχείου είναι χωρισμένο σε τόσα τμήματα όσα και οι τομείς του παραθύρου. Κάθε τμήμα του κυρίου μέρους έχει δύο αριθμητικές τιμές που αντιστοιχούν στην χρονική στιγμή έναρξης του τομέα και το τέλος του. Περιέχει επίσης το κείμενο που περιέχει ο τομέας.

### 6) Φόρτωμα ενός label file:

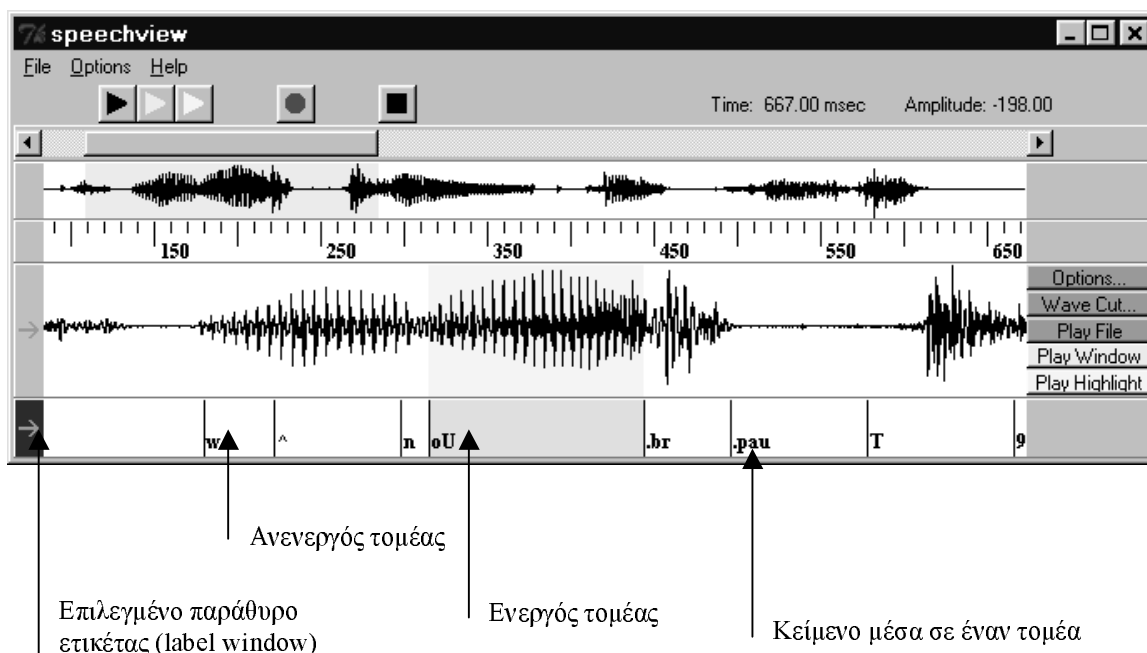
Μπορούμε να φορτώσουμε ένα τέτοιο αρχείο με δύο τρόπους. Ο πρώτος τρόπος είναι να πάμε στο κουμπί **Options** και να χρησιμοποιήσουμε την επιλογή **Load label File**. Ο δεύτερος τρόπος είναι να κάνουμε 'κλικ' στο δεξί κουμπί του πάνω σε μια υπάρχουσα ετικέτα και επιλέξουμε το **Load label File**.

### 7) Αναπαραγωγή σήματος με την βοήθεια των ετικετών:

Από την στιγμή που ενός τομέας γίνει ενεργός τότε μπορούμε να ακούσουμε την κυματομορφή που αντιστοιχεί σε αυτόν κάνοντας 'κλικ' με το αριστερό κουμπί του ποντικιού πάνω του.

### ΠΡΟΣΟΧΗ!

Θα πρέπει να επισημάνουμε το γεγονός ότι ο Speech Viewer διαθέτει και άλλους τρόπους να εκτελέσει κάποιες από τις εργασίες που αναφέρθηκαν έως τώρα και αφορούν τις ετικέτες. Για παράδειγμα αναφέρουμε ότι με τα πλήκτρα 'Ctrl - d' διαγράφεται ο ενεργός τομέας. Για ένα πλήρη κατάλογο των επιπλέον αυτών επιλογών παραπέμπουμε τον χρήστη στον κατάλογο 'Label Bindings'. Αυτός εμφανίζεται αν από την επιλογή 'Help' του κυρίως μενού του προγράμματος επιλέξουμε την υποεπιλογή 'Bindings'. Θα πρέπει να διευκρινίσουμε ότι ο κατάλογος αυτός δεν περιέχει επιπλέον δυνατότητες αλλά περιέχει επιπλέον τρόπους για την πραγματοποίηση όλων όσων αναφέρθηκαν έως τώρα.



Σχήμα 5.1

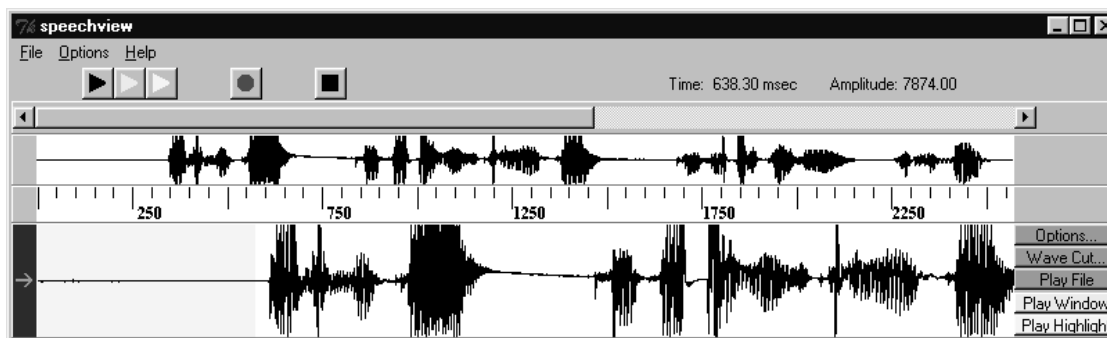
## 6) ΠΑΡΑΔΕΙΓΜΑΤΑ – ΕΦΑΡΜΟΓΕΣ

Στην παράγραφο αυτή θα παρουσιαστούν παραδείγματα εφαρμογής του προγράμματος και θα γίνει επεξεργασία ηχητικών κυματομορφών με βάση τα όσα αναφέρθηκαν στα προηγούμενα κεφάλαια του παρόντος εγχειριδίου. Στην παράγραφο 6.1 θα παρουσιαστούν τα παραδείγματα και στην παράγραφο 6.2 θα παρουσιαστούν οι εφαρμογές.

### 6.1 ΠΑΡΑΔΕΙΓΜΑΤΑ

#### 6.1.1) Ηχογράφηση κυματομορφής

Αρχικά ηχογραφήθηκε η φράση “Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών” (με τον τρόπο που περιγράφηκε στο κεφάλαιο 1) και με το τέλος της ηχογράφησης η κυματομορφή εμφανίζεται στο παράθυρο επεξεργασίας όπως φαίνεται στο σχήμα 6.1. Ο ρυθμός δειγματοληψίας του σήματος επιλέχτηκε στα 16000 bits/sec. Θα πρέπει να παρατηρήσουμε ότι στο συγκεκριμένο σήμα είναι ξεκάθαρο από που αρχίζει και που τελειώνει η κάθε λέξη (εκτός ίσως από την λέξη “και” που αρχίζει την χρονική στιγμή 800msec και τελειώνει την χρονική στιγμή 850 msec. Εκτός όμως από τις λέξεις είναι ξεκάθαρη η έναρξη και το τελείωμα και των περισσότερων συλλαβών της φράσης. Αυτό οφείλεται στο τρόπο ομιλίας του εκφωνητή.



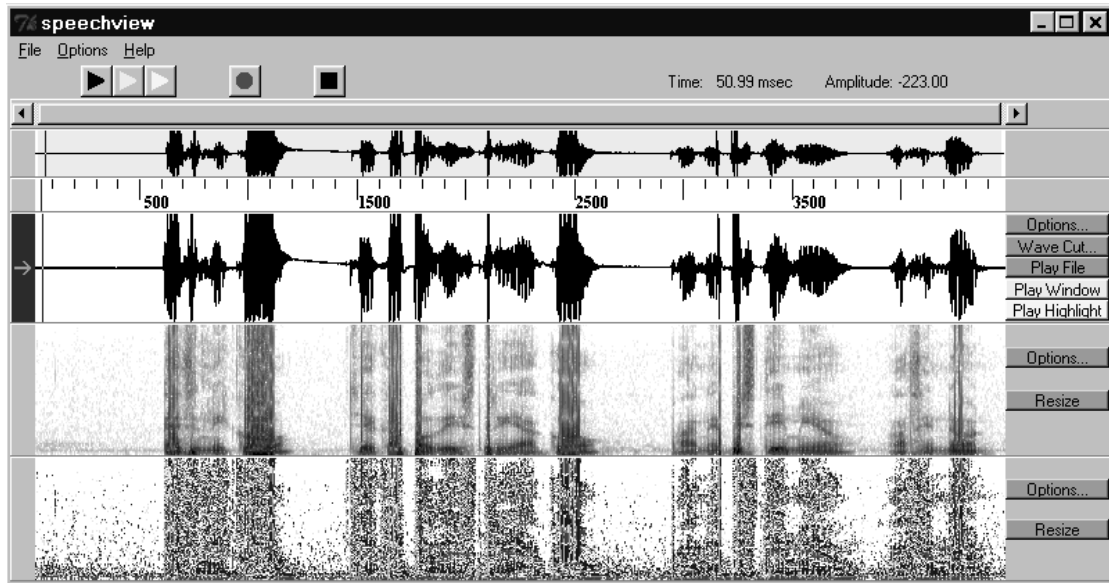
Σχήμα 6.1

#### 6.1.2) Φασματογράφημα

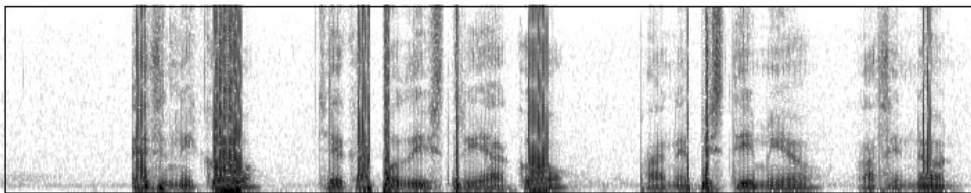
Θα κάνουμε αρχικά το φασματογράφημα του σήματος για διάφορες τιμές των παραμέτρων του και θα συγκρίνουμε τα αποτελέσματα. Στο σχήμα 6.2 έχει γίνει το τυπικό φασματογράφημα και το φασματογράφημα κορυφής του σήματος για τιμές των παραμέτρων Frame=3 και Window=10msec (οι τιμές αυτές είναι οι “default” τιμές που χρησιμοποιεί το πρόγραμμα). Θα πρέπει να επισημάνουμε ότι:

Στο τυπικό φασματογράφημα με μαύρο χρώμα απεικονίζονται οι υψηλές τιμές ενώ με άσπρο οι χαμηλές τιμές. Επίσης τιμές που είναι μικρότερες από το 40% τις μέγιστης τιμής του φάσματος απεικονίζονται με λευκό χρώμα (αυτό συμβαίνει διότι οι τιμές των παραμέτρων min και max είναι 0.40 και 1.0 αντίστοιχα). Στο φασματογράφημα κορυφής οι υψηλές συχνότητες απεικονίζονται με μπλέ χρώμα ενώ οι χαμηλές με άσπρο.

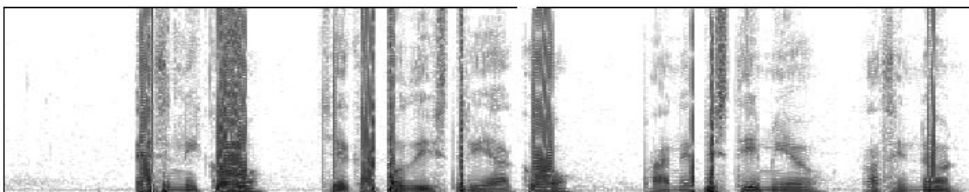
Στο σχήμα 6.3 παρουσιάζονται τρία τυπικά φασματογραφήματα για διαφορετικές τιμές των παραμέτρων Frame και Window. Στο πρώτο φασματογράφημα έχουμε επιλέξει Window=10.0msec και Frame=3.0msec, στο δεύτερο Window = 30.0 και Frame=10.msec και τέλος στο τρίτο οι τιμές των Window είναι 100.0msec και Frame είναι 50.0msec. Από τα τρία αυτά σχήματα συμπεραίνουμε ότι μια αύξηση της τιμής των δύο παραμέτρων οδηγεί σε χαμηλότερης ποιότητας φασματογραφήματα με την έννοια ότι χάνεται η λεπτομέρεια των σχημάτων 6.3(α) και 6.3(β).



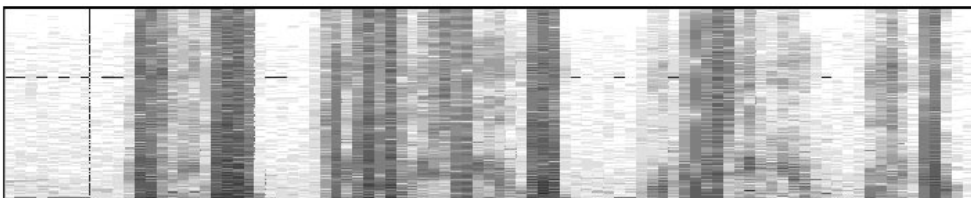
Σχήμα 6.2



Σχήμα 6.3(α): Window = 10.0msec Frame = 3.0msec



Σχήμα 6.3(β): Window = 20.0msec Frame = 10.0msec

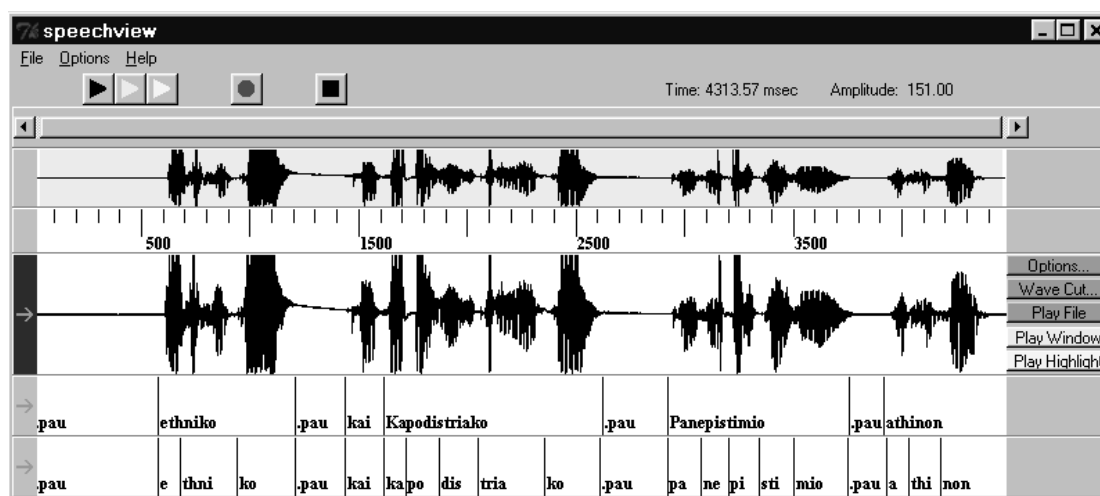


Σχήμα 6.3(γ): Window = 100.0msec Frame = 50.0msec

Σχήμα 6.3

### 6.1.3) Επισημείωση

Στο σημείο αυτό θα επισημειώσουμε το σήμα μας με ετικέτες. Θα φτιάξουμε δύο διαφορετικές επισημειώσεις η πρώτη με βάση τις λέξεις του σήματος και η δεύτερη με βάση τις συλλαβές που στην συγκεκριμένη κυματομορφή είναι όλες σχεδόν εμφανείς που αρχίζουν και που τελειώνουν. Στο σχήμα 6.4 παρουσιάζεται το σήμα επισημειωμένο ακολουθώντας την διαδικασία που περιγράφηκε σε προηγούμενο κεφάλαιο. Η πρώτη ετικέτα δείχνει τις λέξεις που αποτελείται το σήμα μας και η δεύτερη δείχνει τις συλλαβές του. Με “.pau” σημειώνονται οι παύσεις μεταξύ των λέξεων. Σημειώνεται ότι το πρόγραμμα δεν υποστηρίζει την ελληνική γλώσσα για την επισημείωση και για αυτό σύμφωνα όπως το “δ” αποδίδεται με το “d” πράγμα που προφανώς δεν είναι σωστό.

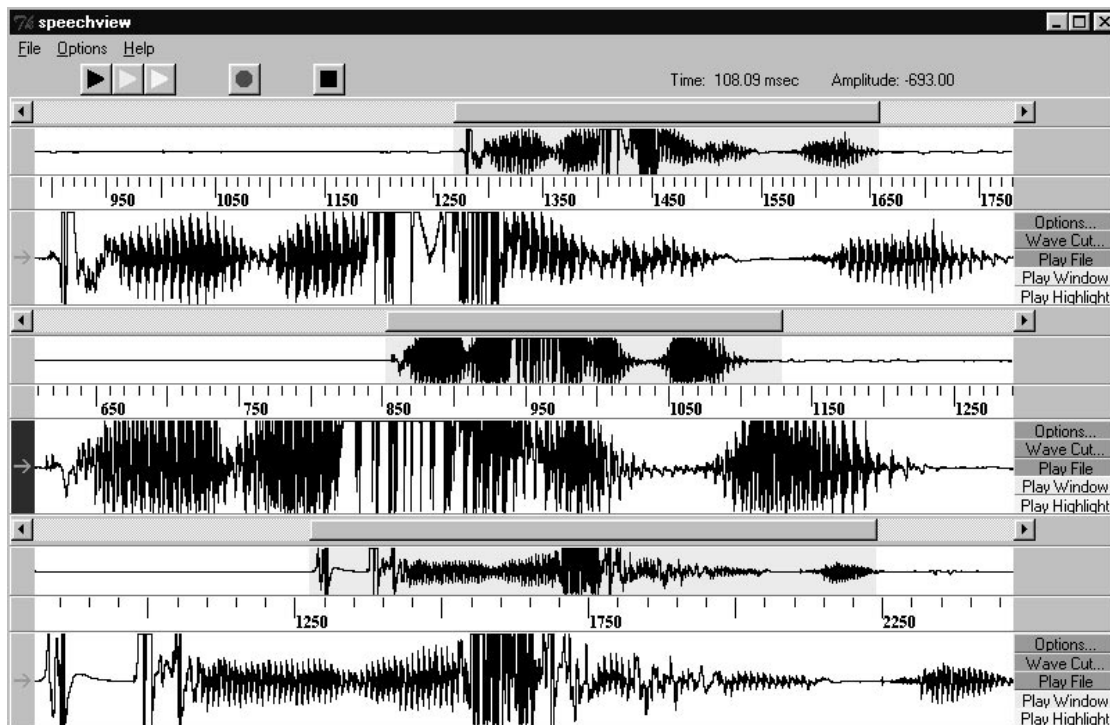


Σχήμα 6.4

## 6.2) ΕΦΑΡΜΟΓΕΣ

### 6.2.1) Μεταβολή του σήματος με τον χρόνο

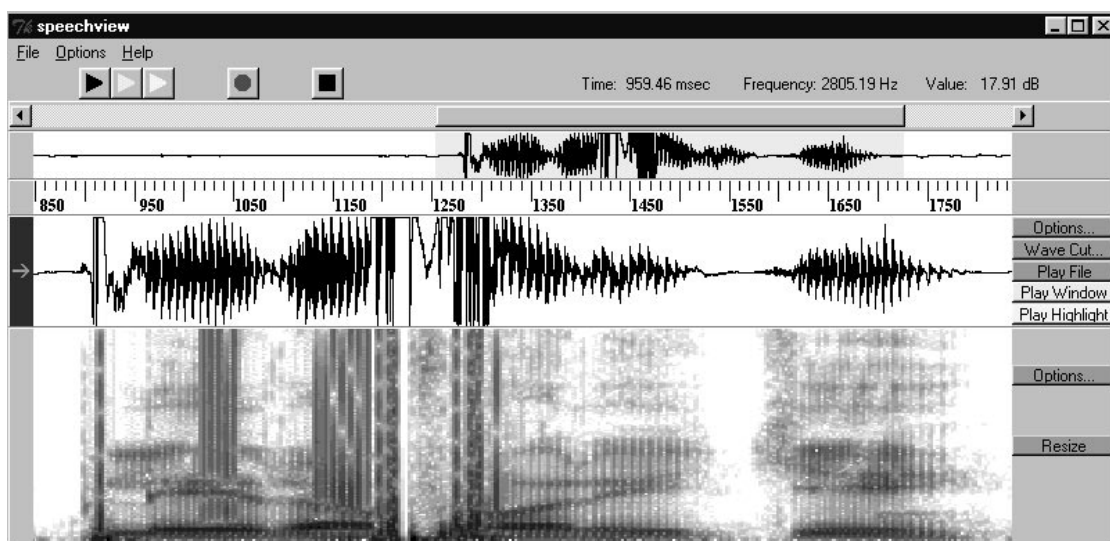
Με το παράδειγμα αυτό θέλουμε να δείξουμε ότι η χρονική διάρκεια του σήματος ομιλίας του ίδιο εκφωνητή μεταβάλλεται με τον χρόνο. Με τον Speech Viewer ηχογραφήθηκαν τρία διαφορετικά ηχητικά σήματα που ο ίδιος ομιλητής εκφωνεί την λέξη «ΠΛΗΡΟΦΟΡΙΚΗ». Τα σήματα αυτά παρουσιάζονται στο σχήμα 6.5. Ο ομιλητής την στο πρώτο σχήμα εκφωνεί την λέξη φυσιολογικά την στο δεύτερο σχήμα εκφωνεί την λέξη γρήγορα και στο τρίτο αργά. Η χρονική διάρκεια του σήματος με φυσιολογική εκφώνηση είναι 400 msec με γρήγορη 250 msec ενώ με αργή είναι 970 msec.



Σχήμα 6.5

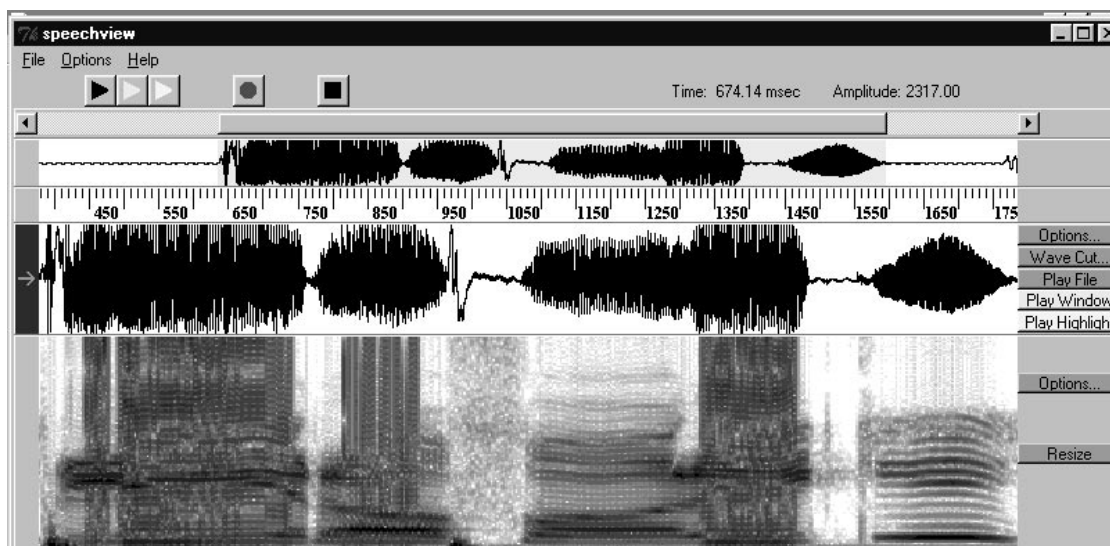
### 6.2.2) Μεταβολή του σήματος με τον τόνο του ομιλητή:

Θα παρουσιάσουμε τον τρόπο που μεταβάλλεται το σήμα με τον τόνο του ομιλητή δηλαδή με το αν κάνει την φωνή του ‘‘ψιλή’’ ή ‘‘χοντρή’’. Ηχογραφήθηκαν τρεις διαφορετικές εκφωνήσεις της λέξης «ΠΛΗΡΟΦΟΡΙΚΗΣ» από τον ίδιο εκφωνητή και παρουσιάζονται στο σχήμα 6.6 μαζί με το φάσμα τους. Το βασικό συμπέρασμα αυτής της συγκρίσης είναι η διατήρηση των φασματικών χαρακτηριστικών. Μπορεί να υπάρχει κάποια μεταβολή στις τιμές του φάσματος (σε κάποιες συχνότητες) αλλά δεν υπάρχει αξιόλογη αλλοίωση. Στο σχήμα 6.6 (α) είναι η κανονική εκφώνηση της λέξεως.



Σχήμα 6.6 (α)

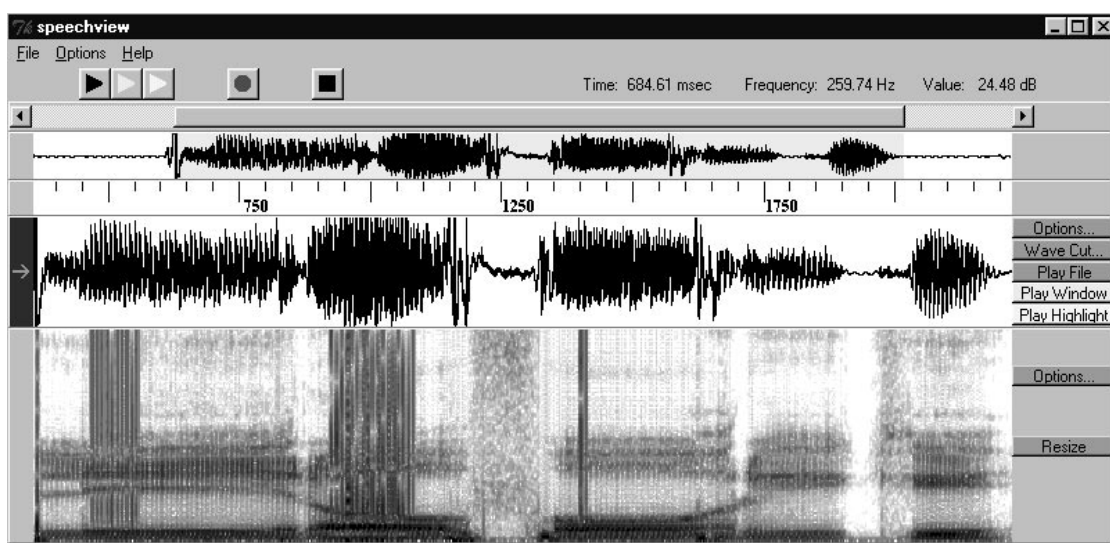
Στο σχήμα 6.6 (β) παρουσιάζεται το σήμα στο οποίο ο εκφωνητής έχει κάνει την φωνή του ψιλή. Στο φάσμα αυτού του σήματος παρατηρούμε αύξηση στην τιμή του φάσματος στις υψηλές συχνότητες ορισμένων τμημάτων του.



Σχήμα 6.6 (β)

Στο σχήμα 6.6 (γ) παρουσιάζεται το σήμα στο οποίο ο εκφωνητής έχει κάνει την φωνή του χοντρή. Στο φάσμα αυτού του σήματος παρατηρούμε αύξηση στην τιμή του φάσματος στις χαμηλές συχνότητες σε σχέση με το σχήμα 6.6 (α).





Σχήμα 6.6 (γ)

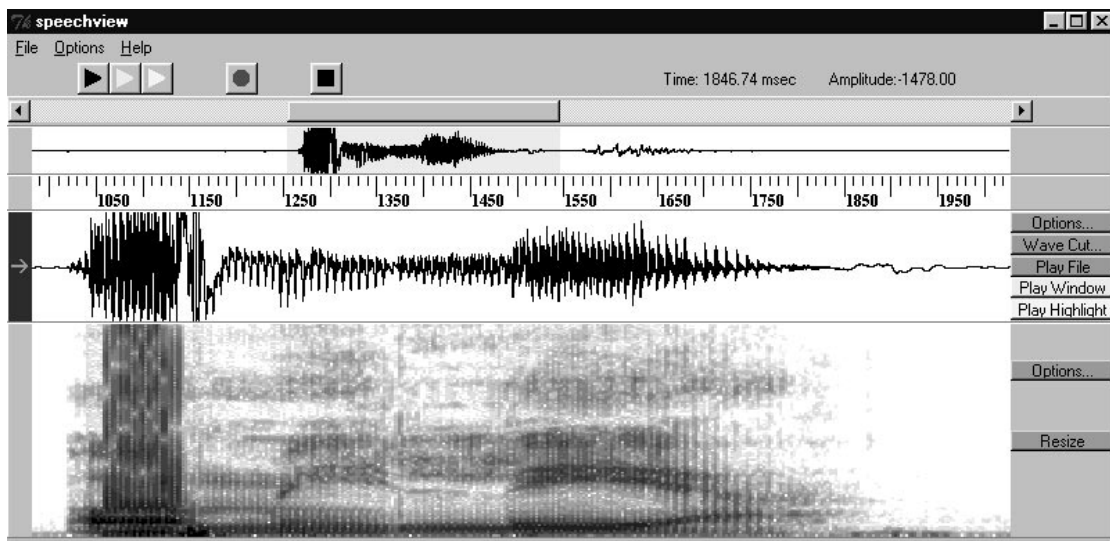
### 6.2.3) Εκφώνηση της ίδιας λέξης από διαφορετικούς ομιλητές

Στο πείραμα αυτό τρεις διαφορετικοί άνδρες ομιλητές εκφώνησαν την λέξη «ΟΜΙΛΙΑ». Οι γραφικές παραστάσεις και τα φάσματα των σημάτων που ηχογραφήθηκαν παρουσιάζονται στα σχήματα 6.7 (α), (β) και (γ). Από αυτά συμπεραίνουμε τα ακόλουθα:

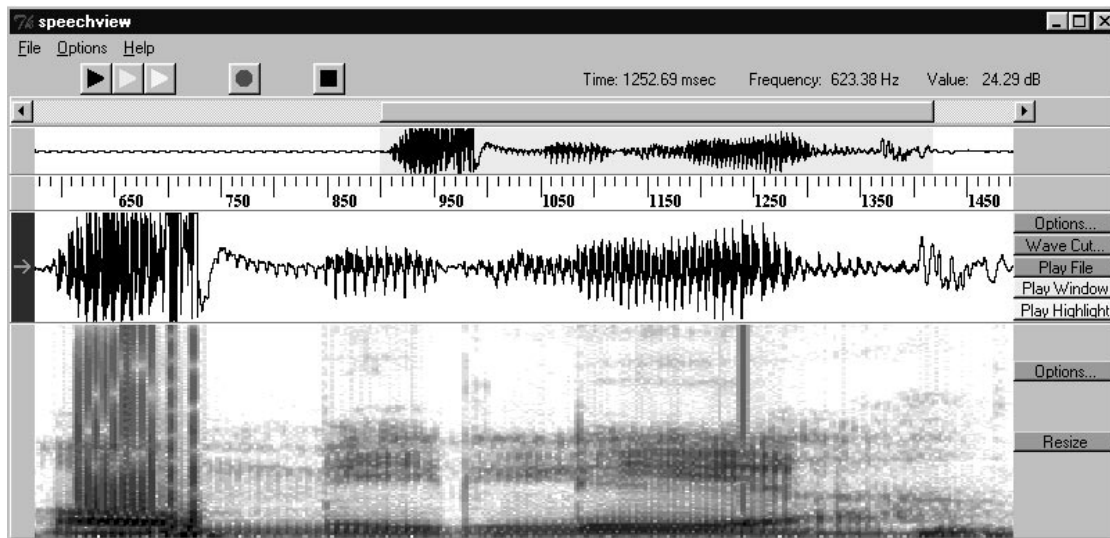
1) Η χρονική διάρκεια των σημάτων δεν είναι σταθερή (780 msec, 650 msec και 570 msec). Επίσης διαφέρει και ο χρόνος εκφώνησης του κάθε φωνήματος.

2) Η ένταση του σήματος ομιλίας διαφέρει από άνθρωπο σε άνθρωπο.

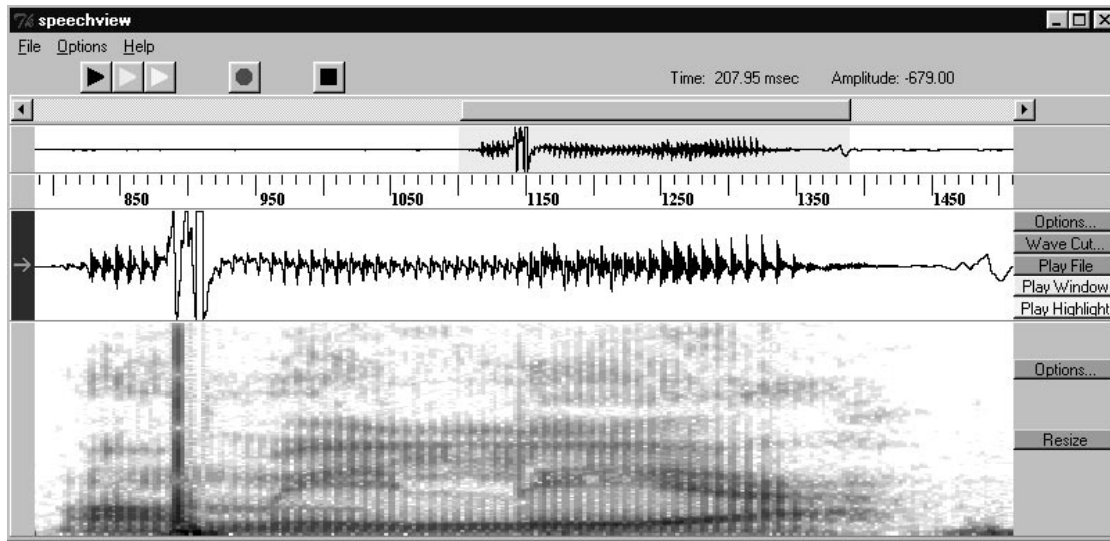
3) Τα φασματικά χαρακτηριστικά των τριών σημάτων είναι τα ίδια.



Σχήμα 6.7 (α)



Σχήμα 6.7 (β)



Σχήμα 6.7 (r)

## **ΜΕΡΟΣ ΔΕΥΤΕΡΟ : Snack**

### **ΕΙΣΑΓΩΓΗ**

Ο Snack είναι ένα πακέτο για την επεξεργασία ομιλίας, το οποίο αναπτύχθηκε από το τμήμα λόγου μουσικής και ακουστικής του Πανεπιστημίου της Στοκχόλμης. Ο Snack ασχολείται με την επεξεργασία και ανάλυση της ομιλίας. Στη παρούσα εργασία θα γίνει μια αναλυτική παρουσίαση του Snack καθώς και των δυνατοτήτων που προσφέρει στον χρήστη του για την επεξεργασία ενός σήματος ομιλίας.

Εδώ θα ήταν, επίσης, χρήσιμο να αναφέρουμε τις απαιτήσεις που έχει το πρόγραμμα για να λειτουργήσει ικανοποιητικά. Αυτές είναι:

Απαιτήσεις συστήματος:

Intel x86 200MHz Pentium class 200MHz επεξεργαστής ή ανώτερος  
32MB ή παραπάνω μνήμη (64MB συστήνεται)  
100MB χώρος στον σκληρό δίσκο  
Soundcard (SoundBlaster συμβατή)

Προσοχή οι παραπάνω απαιτήσεις προτείνονται για την ικανοποιητική λειτουργία του πακέτου. Αυτό δε σημαίνει ότι το πακέτο δε θα λειτουργεί σε χαμηλότερης ισχύος μηχανήματα. Απλά, πιθανόν κάποιες από τις δυνατότητες του να μη λειτουργούν σε τόσο ικανοποιητικό βαθμό.

Τέλος, θα πρέπει να αναφέρουμε, εδώ, ότι ο Snack δεν είναι ένα αυτόνομο πρόγραμμα που μπορεί να εκτελεστεί μόνο του, αλλά προκειται για μια επέκταση γραμμένη σε Tcl/Tk. Για όσους δεν γνωρίζουν πρόκειται για έναν μεταφραστή που μπορεί να μεταφράζει ένα text αρχείο γραμμένο ειδικά για αυτόν σε εκτελέσιμο κώδικα. Έτσι κάθε φορά που θέλεις να καλέσεις μια εφαρμογή γραμμένη στη συγκεκριμένη γλώσσα πρέπει πρώτα να καλέσεις τον μεταφραστή (wish80.exe) και εν συνεχεία μέσα από αυτόν την εφαρμογή.

## **ΠΕΡΙΛΗΨΗ**

Ο Snack είναι ένα πρόγραμμα για την εμφάνιση, αναπαραγωγή, εγγραφή, κοψιμό και επικόλληση ηχητικών κυματομορφών. Μπορεί να χρησιμοποιηθεί επίσης για την κατασκευή του φασματογραφήματος της ηχητικής κυματομορφής, σύνταξη και εμφάνιση ετικετών που αντιστοιχούν σε αυτή και τέλος για την κατασκευή του spectrum του σήματος.

Από την toolbar, μπορούμε να ανοίξουμε μια υπάρχουσα κυματομορφή ή μπορούμε να ηχογραφήσουμε μια καινούργια. Αυτή η κυματομορφή εμφανίζεται κάτω από την toolbar μαζί με μια μπάρα κύλισης. Μπορούμε να προβάλλουμε και περισσότερες από μια κυματομορφές συγχρόνως με το πρόγραμμα αυτό. Ο Snack παρέχει δυνατότητες για την επεξεργασία της κυματομορφής (αφαίρεση κάποιου τμήματος του ηχητικού σήματος, επικόλληση μιας κυματομορφής σε κάποιο τμήμα μιας άλλης κτλ).

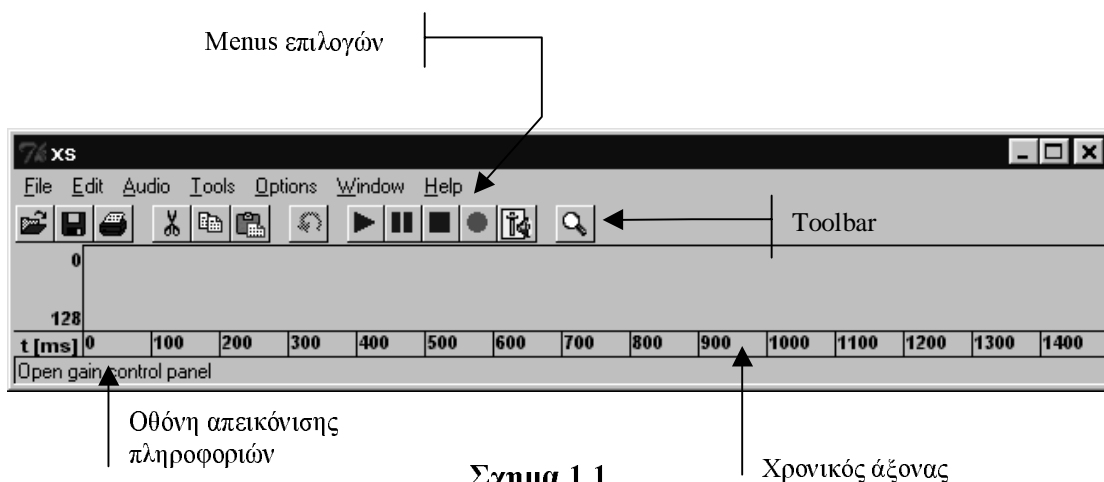
Μια από τις κύριες εργασίες που μπορεί να επιτελέσει το συγκεκριμένο πρόγραμμα είναι η εύρεση του φασματογραφήματος της κυματομορφής και η προβολή του κάτω ακριβώς από αυτή. Το φασματογράφημα που υποστηρίζει το Snack είναι το τυπικό φασματογράφημα δηλ χρησιμοποιεί την κλίμακα του γκρί για την απεικόνιση των μεταβολών των τιμών της φασματικής ισχύος στις διάφορες συχνότητες. Ο χρήστης μπορεί να μεταβάλλει ένα μεγάλο αριθμό παραμέτρων για την κατασκευή του φασματογραφήματος. Εκτός όμως από το φασματογράφημα ο Snack έχει την δυνατότητα κατασκευής και του φάσματος (spectrum) του σήματος.

Τέλος, με τον Snack μπορούμε να επισυνάψουμε ετικέτες (labels) σε ένα ηχητικό σήμα. Για την εμφάνιση των ετικετών υπάρχει κάτω από κάθε κυματομορφή ένας ειδικός χώρος. Με τον όρο ετικέτα εννοούμε ένα σύνολο φωνημάτων που αντιστοιχίζονται σε ένα τμήμα της κυματομορφής με σκοπό να μας δείχνουν αυτό που θα ακούσουμε στο συγκεκριμένο κομμάτι του σήματος.

Στα κεφάλαια που ακολουθούν ο αναγνώστης μπορεί να βρει έναν λεπτομερή εγχειρίδιο χρήσης του προγράμματος.

## 1) ΞΕΚΙΝΩΝΤΑΣ ΜΕ ΤΟΝ SNACK

Ο snack σου παρέχει τη δυνατότητα είτε να φορτώσεις ένα ήδη αποθηκευμένο αρχείο ήχου από τον δίσκο είτε να ηχογραφήσεις την δικιά σου ομιλία μέσω ενός μικροφώνου εάν ο υπολογιστής διαθέτει κάτι τέτοιο. Τα υποστηριζόμενα πρότυπα (formats) ήχου είναι αρκετά με πιο γνωστά τα .wav και .au. Το σύνολο των υποστηριζόμενων προτύπων μπορεί να το δει κανείς από τις επιλογές κατά το φόρτωμα ενός αρχείου (βλέπε κεφάλαιο 2).



Σχήμα 1.1

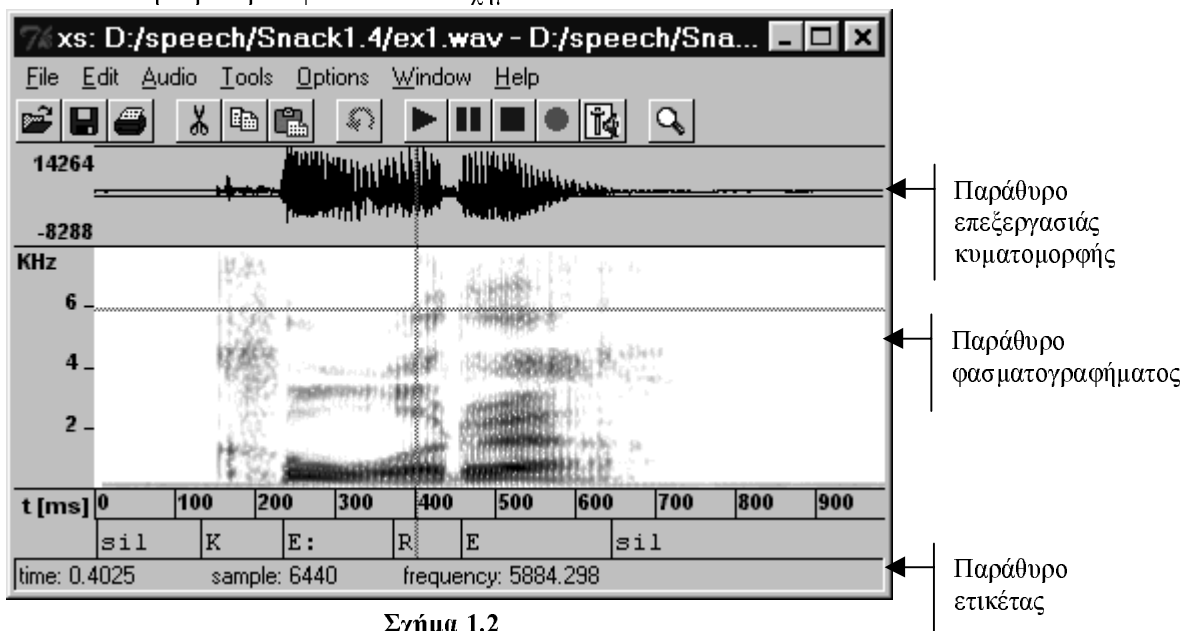
Ο snack φορτώνει από την επιλογή xs.tcl και με το που ξεκινάει καλεί το χρήστη να φορτώσει μια αποθηκευμένη κυματομορφή. Στη περίπτωση που ο χρήστης δεν επιθυμεί κάτι τέτοιο επιλέγει cancel οπότε ο snack ξεκινάει χωρίς κάποιο σήμα να απεικονίζεται. Το κυρίως παράθυρο του snack πριν το φόρτωμα κάποιου σήματος φαίνεται στο σχήμα 1.1.

### 1.1) παράθυρα

Όπως φαίνεται και στο σχήμα 1.1 ο snack διαθέτει ένα πλήθος από menus επιλογών καθώς και μια Toolbar. Επίσης διαθέτει έναν χρονικό άξονα και μια οθόνη απεικόνισης πληροφοριών. Επίσης ο snack διαθέτει μια σειρά από παράθυρα για την επεξεργασία της κυματομορφής. Τα παράθυρα αυτά είναι 3 ειδών:

- 1) παράθυρο επεξεργασίας κυματομορφής (βλ. κεφάλαιο 2)
- 2) παράθυρο φασματογραφήματος (βλ. κεφάλαιο 4)
- 3) παράθυρο ετικέτας (βλ. κεφάλαιο 3)

Τα 3 αυτά είδη παραθύρων φαίνονται στο σχήμα 1.2.



Σχήμα 1.2

Το παράθυρο επεξεργασίας της κυματομορφής είναι αυτό στο οποίο απεικονίζεται το σήμα μας στο χρονικό πεδίο και αυτό στο οποίο γίνεται όλη η επεξεργασία του. Είναι δε το βασικό παράθυρο απεικόνισης το οποίο ανοίγει όταν φορτώνουμε ένα σήμα. Το πώς γίνεται η επεξεργασία της ομιλίας αναλύεται στο επόμενο κεφάλαιο.

Το παράθυρο φασματογραφήματος απεικονίζει το φασματογράφημα της κυματομορφής που απεικονίζεται στο βασικό παράθυρο ενώ, τέλος, το παράθυρο ετικέτας είναι αυτό στο οποίο εμφανίζεται η επισημείωση που έχουμε κάνει στην κυματομορφή. Όπως φαίνεται και στο σχήμα 1.2 ο mouse cursor έχει τη μορφή ενός σταυρονήματος. Αυτή η μορφή του cursor δεν είναι η default μορφή αλλά μπορεί να εμφανιστεί από αντίστοιχη επιλογή του menu options(βλ. §2.2.5). Μετακινώντας τώρα το σταυρόνημα πάνω από ένα παράθυρο στην οθόνη απεικόνισης πληροφοριών εμφανίζονται οι τιμές του χρόνου(time), της συχνότητας(frequency) και του αριθμού του δείγματος(sample) στο αντίστοιχο σημείο.

## 1.2) Μαρκάρισμα

Το μαρκάρισμα της κυματομορφής μπορεί να γίνει σε οποιαδήποτε από τα παράθυρα τύπου 1 και 2 ταυτόχρονα. Δύο μύαρες κατακόρυφες γραμμές σημειώνουν την έναρξη και το τέλος ενός μαρκαρισμένου τμήματος. Το μαρκάρισμα πραγματοποιείται με τη γνωστή μέθοδο του dragging. Δηλαδή πατάμε το αριστερό κουμπί του mouse στο σημείο της κυματομορφής που επιθυμούμε να μαρκάρουμε κι εν συνεχεία το κρατάμε πατημένο μετακινώντας το μέχρι το σημείο που θέλουμε να επιλέξουμε. Παρατηρούμε ότι τώρα υπάρχουν πάνω στο παράθυρο οι 2 κατακόρυφες ευθείες που προαναφέραμε. Για την καλύτερη οπτικοποίηση του μαρκαρισμένου τμήματος υπάρχει η επιλογή χρωματισμού αυτού με κατάλληλη επιλογή από το menu options(βλ. §2.2.5).

Άλλος τρόπος μαρκαρίσματος είναι από το ειδικό menu που εμφανίζεται όταν κάνουμε δεξιό 'κλικ' σε κάποιο παράθυρο(σχήμα 1.3) και πιο συγκεκριμένα από τις επιλογές Mark start και Mark end.



**Σχήμα 1.3**

## 2) ΕΠΕΞΕΡΓΑΣΙΑ ΚΥΜΑΤΟΜΟΡΦΗΣ

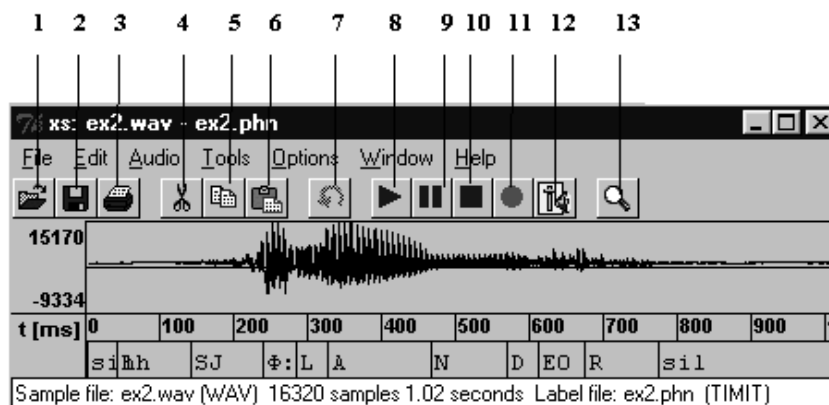
Στο κεφάλαιο αυτό θα παρουσιάσουμε θέματα όπως ηχογράφιση ενός σήματος, αποθήκευση του, αναπαραγωγή του κ.λ.π. Οι λειτουργίες αυτές γίνονται από την toolbar του προγράμματος. Πριν μιλήσουμε για αυτήν θα αναφέρουμε το σταυρόνημα (Crooss-Hair). Για να εμφανιστεί το σταυρόνημα ο χρήστης θα πρέπει να πάει στην επιλογή Options του menu του προγράμματος και να επιλέξει Cross Hair. Στο σημείο που δείχνει το ποντίκι εμφανίζεται ένας κόκκινος σταυρός. Αν ο χρήστης πάει στην επιλογή Options και επιλέξει Fill Between Marks τότε το τμήμα της κυματομορφής που υποδεικνύεται με το σταυρόνημα μαρκάρεται αυτόματα και αποκτά κίτρινο χρώμα.

### 2.1) Toolbar:

Η toolbar του Snack φαίνεται στο σχήμα 2.1. Η λειτουργία κάθε κουμπιού της είναι:

- (1) Ανοίγει ένα πλαίσιο διαλόγου για το φόρτωμα ενός σήματος.
- (2) Αποθηκεύει την κυματομορφή.
- (3) Εκτυπώνει την κυματομορφή.
- (4) Κόβει το μαρκαρισμένο τμήμα της κυματομορφής.
- (5) Αντιγράφει το μαρκαρισμένο τμήμα της κυματομορφής στον Clipboard.
- (6) Επισυνάπτει το τμήμα της κυματομορφής που είναι στον clipboard στην θέση που δείχνει το σταυρόνημα.
- (7) Αναίρει (undo) κάποια προηγούμενη επιλογή.
- (8) Αναπλάθει την κυματομορφή από το μαρκαρισμένο σημείο και κάτω.
- (9) Παγώνει την αναπαραγωγή της κυματομορφής.
- (10) Σταματάει την αναπαραγωγή της κυματομορφής.
- (11) Ηχογραφεί μια καινούργια κυματομορφή.
- (12) Μεταβάλλει την ένταση της φωνής.
- (13) Με το κουμπί αυτό ο χρήστης μπορεί να κάνει μεγένθυση (zoom) της κυματομορφής. Όταν ο χρήστης πατήσει αυτό το κουμπί θα του εμφανιστεί ένα καινούργιο παράθυρο στο οποίο θα απεικονίζεται η μεγένθυση του μαρκαρισμένου τμήματος του σήματος. Στο παράθυρο αυτό ο χρήστης μπορεί να κάνει μεγαλύτερη ή μικρότερη μεγένθυση στον κατακόρυφο άξονα από τις αντίστοιχες επιλογές (Y-zoom in και Y-zoom-out).





Σχήμα 2.1

## 2.2) Menus

Οι περισσότερες από τις λειτουργίες της toolbar μπορούν να γίνουν και με τις τρεις πρώτες επιλογές (File – Edit – Audio) του menu του προγράμματος.

### 2.2.1) File

Κάνοντας 'κλικ' στην επιλογή **File** εμφανίζεται ένα υπομενού με τις εξής επιλογές

- 1) **Open** Ανοίγει ένα πλαίσιο διαλόγου για το φόρτωμα ενός σήματος.
- 2) **Get Url** Δίνει την διεύθυνση στο διαδίκτυο όπου ο χρήστης μπορεί να πάρει περισσότερες πληροφορίες για το πρόγραμμα
- 3) **Save** Σώζει την τρέχουσα κυματομορφή στο τρέχων αρχείο.
- 4) **Save as** Σώζει την τρέχουσα κυματομορφή σε ένα νέο αρχείο το οποίο καλείται να ορίσει ο χρήστης
- 5) **Save Mark** Σώζει το μαρκαρισμένο τμήμα του σήματος
- 6) **Close** Κλείνει την τρέχουσα κυματομορφή
- 7) **Print** Τυπώνει την κυματομορφή
- 8) **Info** Δίνει πληροφορίες για την τρέχουσα κυματομορφή.
- 9) **Exit** Τερματίζει το πρόγραμμα

### 2.2.2) Edit

Με την επιλογή αυτή ο χρήστης μπορεί να κάνει τα ακόλουθα

- 1) **Undo** Αναίρει κάποια προηγούμενη επιλογή.
- 2) **Cut** Κόβει το μαρκαρισμένο τμήμα της κυματομορφής.
- 3) **Copy** Αντιγράφει το μαρκαρισμένο τμήμα της κυματομορφής στον Clipboard.
- 4) **Paste** Επισυνάπτει το τμήμα της κυματομορφής που είναι στον clipboard στην θέση που δείχνει το σταυρόνημα.
- 5) **Crop** Κόβει ολοκληρη την κυματομορφή εκτός από το μαρκαρισμένο τμήμα.
- 6) **Mark all** Μαρκάρει όλη την κυματομορφή

### 2.2.3) Audio

Με την επιλογή **Audio** ο χρήστης μπορεί να ρυθμίσει παραμέτρους που αφορούν την ηχογράφιση και αναπαραγωγή της ομιλίας. Πιο συγκεκριμένα υπάρχουν οι ακόλουθες επιλογές:

**1) Play Range** Αναπαράγει (κοινώς “παίζει”) το μαρκαρισμένο τμήμα της κυματομορφής

**2) Play All** Αναπαράγει ολόκληρη την κυματομορφή

**3) Stop Play** Σταματάει την αναπαραγωγή της κυματομορφής

**4) Gain Control** Μεταβάλλει την ένταση της φωνής

**5) Play Settings** Με την επιλογή αυτή ο χρήστης μπορεί να μεταβάλλει τις παραμέτρους αναπαραγωγής του σήματος. Οι παράμετροι αυτοί είναι :

α) συχνότητα αναπαραγωγής

β) κωδικοποίηση (format) των δειγμάτων. Τα υποστηριζόμενα formats είναι: Lin16(16-μπιτο) , Lin8(8-μπιτο) , muLaw και Alaw.

γ) εναλλαγή από στερεοφωνική σε μονοφωνική αναπαραγωγή και αντίστροφα.

δ) swap bytes

Είναι προφανές ότι η τροποποίηση αυτή των παραμέτρων μπορεί να οδηγήσει σε αλλοίωση της ακουστικότητας του σήματος. Για παράδειγμα υποθέτουμε μια κυματομορφή η οποία είναι αποθηκευμένη σε ένα αρχείο με συχνότητα δειγματοληψίας 16000 bits/sec . Τροποποιούμε την συχνότητα αναπαραγωγής στα 8000 bits/sec. Είναι προφανές ότι κάτι τέτοιο θα οδηγήσει σε αλλοίωση της ομιλίας αφού είναι σαν να «παίζουμε» τα δείγματα με πιο χαμηλό ρυθμό.

**6) Conversions** Η επιλογή αυτή δίνει τη δυνατότητα στον χρήστη να τροποποιήσει τις παραμέτρους α,β,γ που αναφέρθηκαν και στο 5 με τη διαφορά ότι τώρα ο Snack θα αναλάβει να κάνει όλες τις αναγκαίες προσαρμογές ώστε η αναπαραγωγή του σήματος να παραμείνει αναλόγως. Με τις επιλογές αυτές μπορούμε να προσαρμόσουμε ένα αρχείο στο επιθυμητό format (συχνότητα, bits/sample κοκ) και εν συνεχεία να το σώσουμε. Για παράδειγμα αναφέρουμε ο χρήστης μπορεί να κατασκευάσει ένα αρχείο με συχνότητα δειγματοληψίας 16000 bits/sec από ένα αρχείο με συχνότητα 8000 bits/sec.

### 2.2.4) Tools

Κάνοντας ‘κλικ’ στην επιλογή Tools εμφανίζεται ένα υπομενού με τις ακόλουθες επιλογές:

**1) Plot data**

**2) Beep Generator** Η επιλογή αυτή αντικαθιστά το μαρκαρισμένο τμήμα του σήματος με το σήμα μιας γεννήτριας παλμών τις παραμέτρους τις οποίας καλείται να ορίσει διαλογικά ο χρήστης.

#### Παρατήρηση

Οι επιλογές που εμφανίζονται στο συγκεκριμένο menu αποτελούν plug-ins και ο χρήστης μπορεί να φτιάξει δικά του και να τα καλεί κάθε φορά. Η κλήση τους γίνεται από συγκεκριμένη επιλογή του menu settings.

### 2.2.5) Options

Κάνοντας ‘κλικ’ στην επιλογή options εμφανίζεται ένα υπομενού με τις ακόλουθες επιλογές:

**1) Settings** Η επιλογή αυτή τροποποιεί ορισμένες παραμέτρους για τον τρόπο που απεικονίζεται το σήμα, το φασματογράφημα και το φάσμα .

**2) Plug – in** Κάνοντας ‘κλικ’ σε αυτή την επιλογή εμφανίζεται ένα παράθυρο διαλόγου με το οποίο ο χρήστης μπορεί να προσθέσει και να αφαιρέσει τα plug-ins που θα εμφανιστούν στην επιλογή Tools που περιγράφηκε παραπάνω.

**3) Label File Format - Label Alignment** Οι επιλογές αυτές θα παρουσιάσουν σε επόμενο κεφάλαιο με τίτλο ετικέτες.

**4) Record Button** Ενεργοποιεί και απενεργοποιεί το κουμπί της ηχογράφησης(βλ. 2.1)

**5) Show Spectrogram** Εμφανίζει το φασματογράφημα του σήματος.

- 6)Auto Load** Φορτώνει αυτόματα το label file που αντιστοιχεί σε ένα σήμα.
- 7)Cross Hairs** Εμφανίζει το σταυρόνημα
- 8)Fill Between Marks** Αν η επιλογή αυτή είναι ενεργή η περιοχή του σήματος που υποδεικνύεται με το σταυρόνημα μαρκάρεται αυτόματα.
- 9)Set Default Options** Επαναφέρει τις επιλογές (4)-(8) στις default τιμές τους. Προεπιλεγμένη μόνο η (4).
- 10)Save Options** Σώζει τις επιλογές που έκανε ο χρήστης στα (4) έως (8) ώστε κάθε φορά που ο Snack επανεκκινείται να είναι προεπιλεγμένες. Έτσι για παράδειγμα εάν ο χρήστης έχει μαρκάρει την επιλογή Show Spectrogram και κάνει 'κλικ' στο Save Options τότε με κάθε κυματομορφή που θα φορτώνει το πρόγραμμα θα φτιάχνει αυτόματα και το φασματογράφημα της.

## 2.2.6) Window

Κάνοντας 'κλικ' στην επιλογή αυτή εμφανίζεται το ακόλουθο υπομενού :

- 1)New window** Η επιλογή αυτή ανοίγει ένα καινούργιο παράθυρο για την απεικόνιση μιας νέας κυματομορφής.
- 2)Refresh** Η επιλογή αυτή ανανεώνει το υπάρχον παράθυρο.
- 3)Waveform Zoom** Μεγενθώνει την κυματομορφή.
- 4)Spectrum Section** Η επιλογή αυτή εμφανίζει το φάσμα της ομιλίας και περιγράφεται σε επόμενο κεφάλαιο (κεφάλαιο 4)

### 3)ΕΠΙΣΗΜΕΙΩΣΗ (Labeling)

Με τον όρο ετικέτα εννοούμε ένα σύνολο με ένα ή περισσότερα φωνήματα τα οποία αντιστοιχούν σε ένα τμήμα της κυματομορφής και δείχνουν τι ήχο παριστάνει το σήμα.Οι ετικέτες ενός σήματος εμφανίζονται στο παράθυρο της ετικέτας.Με το φόρτωμα ενός αρχείου το παράθυρο της ετικέτας εμφανίζεται στην οθόνη μονο εαν επιλογή Auto Load του Option είναι μαρκαρισμένη.Σε ένα σήμα είναι δυνατόν να βάλουμε περισσότερες από μια ετικέτες χωρίζοντας το παράθυρο σε αντίστοιχους τομείς.Δεν είναι δυνατόν να τοποθετηθούν περισσότερα από ένα παράθυρα με ετικέτες σε κάθε σήμα.Οι ετικέτες είναι δυνατόν να αποθηκευτούν σε αρχεία και να φορτωθούν ξανά από αυτά.Τα παραπάνω γίνονται ως εξής

#### 1)Δημιουργία ενός παράθυρου ετικετών

Για να δημιουργηθεί ένα τέτοιο παράθυρο κάνουμε ‘κλικ’ με το δεξιό κουμπί του ποντικιού στο χώρο μεταξύ της μπάρας κύλισης και του παραθύρου εμφάνισης πληροφοριών.Θα εμφανιστεί τότε ένα menu με τις ακόλουθες επιλογές απο τις οποίες επιλέγουμε Insert Label.Το παράθυρο ετικετών θα εμφανιστεί στην οθόνη.



Σχήμα 3.1

Εκτός όμως από την διαδικασία που περιγράψαμε υπάρχει και ένας δεύτερος τρόπος.Μπορούμε να δημιουργήσουμε ένα παράθυρο ετικετών με την χρήση ενός επεξεργαστή κειμένου (όπως οι Notepad και Wordpad).Με τον επεξεργαστή κειμένου φτιάχνουμε ένα αρχείο με την ακόλουθη δομή:

```
0      1000  .pau
1000  2000  a
2500  2850  b
```

Οι αριθμοί που εμφανίζονται σε κάθε γραμμή αντιστοιχούν στην αρχή και τέλος της ετικέτας και αντιστοιχούν όχι σε msec αλλά σε αριθμό δείγματος, ενώ μετά από αυτούς γράφουμε την ετικέτα.Στην συνέχεια αποθηκεύουμε το αρχείο με την κατάληξη .rhn στον ίδιο υποκατάλογο που είναι αποθηκεύμενο και το αρχείο.Τέλος μαρκάρουμε την επιλογή Auto Load στο Options του menu.Με το φόρτωμα κάθε αρχείου η ετικέτα θα φορτώνεται αυτόματα.

#### 2)Χωρισμός ενός παραθύρου σε τομείς

Όταν έχουμε δημιουργήσει το παράθυρο τοποθέτησης ετικετών τότε μπορούμε να το χωρίσουμε σε τομείς κάνοντας ‘κλικ’ σε αυτό και στην συνέχεια από το menu επιλέγουμε την ένδειξη Insert Label. Ο καινούργιος τομέας που δημιουργείται έχει μέσα του το γράμμα ‘χ’.Μπορούμε να διαγράψουμε ένα τομέα κάνοντας ‘κλικ’ στο δεξί κουμπί του ποντικιού και επιλέγοντας την ένδειξη Delete Label.Για να αλλάξουμε το χρονικό διάστημα που αντιστοιχεί ένας τομέας πηγαίνουμε πάνω από την κάθετη γραμμή που χωρίζει δύο τομείς το ποντίκι και αυτόματα εμφανίζεται μια μικρή κόκκινη γραμμή.Μεταφέροντας την αριστερά ή δεξιά μεταβάλλουμε το διάστημα που αντιστοιχεί ο τομέας.

### 3) Τοποθέτηση κειμένου σε έναν τομέα

Για να τοποθετήσουμε κείμενο σε έναν τομέα θα πρέπει να κάνουμε 'κλικ' σε αυτόν με το αριστερό κουμπί του ποντικιού. Τότε εμφανίζεται ο κέρσορας και ότι πληκτρολογήσουμε αυτόματα τοποθετείται στον τομέα.

### 4) Αποθήκευση μιας ετικέτας:

Η αποθήκευση γίνεται αυτόματα με την επιλογή Save File. Η δομή του αρχείου που αποθηκεύεται η ετικέτα περιγράφει παραπάνω.

### 5) Φόρτωμα ενός label file:

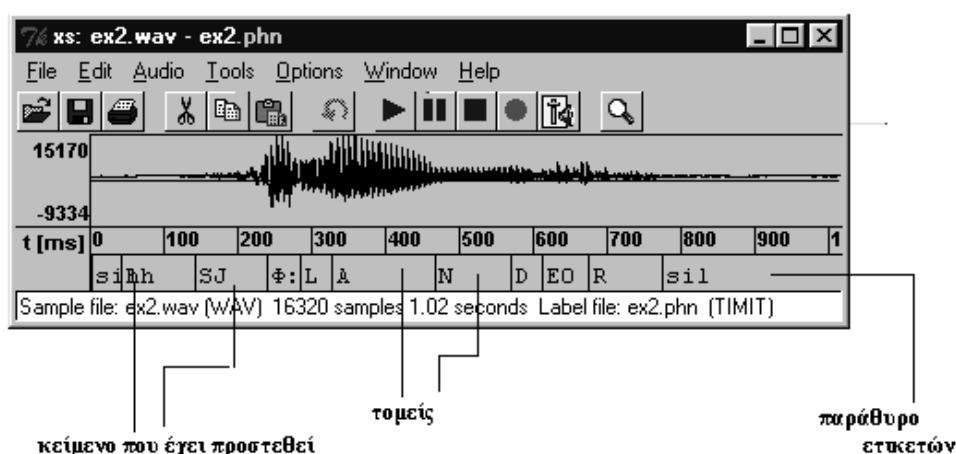
Μπορούμε να φορτώσουμε ένα τέτοιο αρχείο μαζί με το φόρτωμα του αρχείου εάν στην επιλογή Options έχουμε μαρκάρει το Load File ή να το φορτώσουμε από την επιλογή load file του menu File.

### 6) Αναπαραγωγή σήματος με την βοήθεια των ετικετών:

Από την στιγμή που ένας τομέας γίνει ενεργός τότε μπορούμε να ακούσουμε την κυματομορφή που αντιστοιχεί σε αυτόν Κάνοντας 'κλικ' με το δεξιά κουμπί του ποντικιού πάνω στο παράθυρο ετικετών εμφανίζεται ένα menu με την ένδειξη Play Label. Κάνοντας 'κλικ' σε αυτή το πρόγραμμα αναπαράγει το σήμα που βρίσκεται ανάμεσα στα όρια της ετικέτας.

### 7) Ευθυγράμμιση ετικετών (Label Alignment):

Το Snack παρέχει την δυνατότητα ευθυγράμμισης των ετικετών. Με τον όρο αυτό εννοούμε τον τρόπο τοποθέτησης της ετικέτας ανάμεσα στα όρια της. Οι δυνατότητες που παρέχονται είναι τρεις: αριστερή ευθυγράμμιση (το κείμενο των ετοκετών τοποθετείται δίπλα ακριβώς στο αριστερό άκρο της ετικέτας), ευθυγράμμιση στο κέντρο (κείμενο στο κέντρο της ετικέτας) και δεξιά ευθυγράμμιση (κείμενο στο δεξιά άκρο της ετικέτας). Η ευθυγράμμιση μπορεί να γίνει είτε από την επιλογή Label Alignment στο Options είτε από το menu που εμφανίζεται κάνοντας 'κλικ' με το δεξιά κουμπί στο παράθυρο ετικετών.

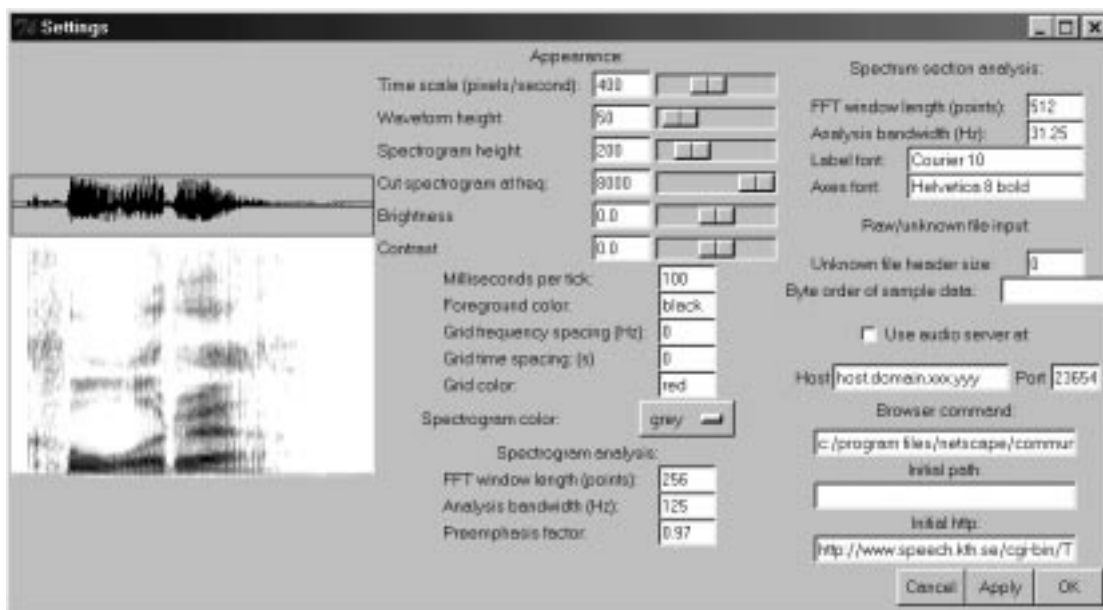


Σχήμα 3.2

## 4)ΦΑΣΜΑΤΟΓΡΑΦΗΜΑ ΚΑΙ ΦΑΣΜΑ

### 4.1)Φασματογράφημα

Το παράθυρο του φασματογραφήματος εμφανίζεται όπως είδαμε και παραπάνω από αντίστοιχη επιλογή του menu options(πιο συγκεκριμένα από την επιλογή show spectrogram). Το φασματογράφημα που υλοποιεί ο snack είναι ένα τυπικό φασματογράφημα όπου οι χαμηλές τιμές απεικονίζονται σε άσπρο χρώμα ενώ οι υψηλές σε κάποια απόχρωση του γκρι. Οι παράμετροι δημιουργίας και εμφάνισης του φασματογραφήματος ρυθμίζονται από την επιλογή **settings** του menu options.Η συγκεκριμένη επιλογή ανοίγει το παράθυρο του σχήματος 4.1.



Σχήμα 4.1

Από τις επιλογές του παραθύρου αυτού μπορεί ο χρήστης να ρυθμίσει:

- 1)την εμφάνιση του φασματογραφήματος
- 2)τη δημιουργία του φασματογραφήματος

#### 4.1.1)Εμφάνιση φασματογραφήματος

Ρυθμίζεται από μια σειρά παραμέτρων που βρίσκονται κάτω από τον γενικότερο τίτλο **Appearance**(βλ. σχήμα 4.1).Οι παράμετροι αυτοί δεν χρήζουν ιδιαίτερης ανάλυσης.Απλά θα αναφέρουμε ότι μπορούμε να τροποποιήσουμε παραμέτρους όπως : τη χρονική προσαρμογή του σήματος(time scale) το ύψος του φασματογραφήματος, αλλά και της κυματομορφής (waveform-spectrogram height σε pixels), τη συχνότητα μέχρι την οποία θα απεικονιστεί το φασματογράφημα (Cut Spectrogram at Freq), τη φωτεινότητα(brightness) και την αντίθεση(contrast).

Οι 5 παράμετροι που ακολουθούν δεν επηρεάζουν την εμφάνιση του φασματογραφήματος αυτού καθαυτού αλλά την αντίστοιχη των χρονικού άξονα και του άξονα συχνότητων ενώ παράλληλα παρέχουν τη δυνατότητα προσθήκης πλέγματος πάνω στο φασματογράφημα στην επιθυμητή μορφή.Πιο διεξοδικά αναφέρουμε:

**1)milliseconds per tick** Ρυθμίζει την επισημείωση του χρονικού άξονα.Π.χ. μια τιμή 200 σε αυτήν τη παράμετρο υποδεικνύει ότι ο άξονας του χρόνου θα εμφανιστεί σε κλίμακα των 200 msec.

**2)foreground color** Ασήμαντη παράμετρος που αλλάζει απλά τόν χρωματισμό των αξόνων και της κυματομορφής.

**3)Grid frequency Spacing** Η παράμετρος αυτή ορίζει την κλίμακα του πλέγματος στον άξονα των συχνότητων.Έτσι για παράδειγμα δίνοντας στην παράμετρο αυτή τη τιμή 1000, στο παράθυρο του

φασματογραφήματος θα εμφανιστούν ευθείες κάθε 1KHz(1000 Hz).Στη περίπτωση που έχει τη τιμή 0 δεν υπάρχει πλέγμα.

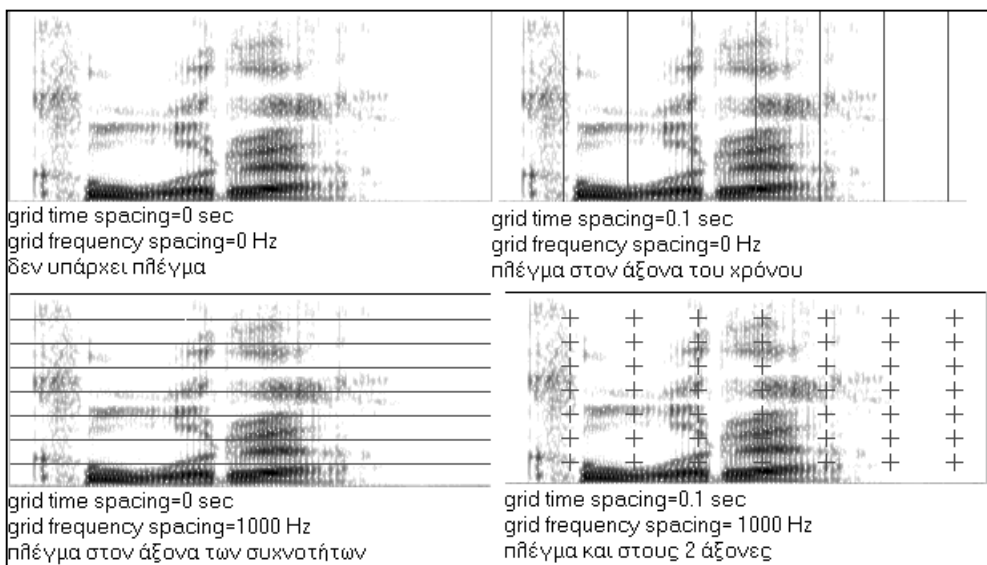
**4)Grid time Spacing** Όπως και η προηγούμενη έτσι και αυτή ορίζει την κλίμακα του πλέγματος στον άξονα του χρόνου.Έτσι μια τιμή 0.2 της παραμέτρου χωρίζει το φασματογράφημα σε λωρίδες των 200 msec(0.2 sec).Τιμή 0 δηλώνει, όπως και πριν απουσία πλέγματος.

**5)grid color** Αλλά αλλάζει το χρώμα εμφάνισης του πλέγματος

**6)spectrogram color** Η παράμετρος αυτή σου δίνει τη δυνατότητα 3 διαφορετικών μορφών απεικόνισης του φασματογραφήματος.Την κλασική της γκρι κλίμακας συν 2 έγχρωμες.

### Παρατήρηση

Στη περίπτωση που οριστεί κλίμακα πλέγματος και στους 2 άξονες η εμφάνιση του γίνεται με σταυρούς στα σημεία τομής μονάχα(βλ. σχήμα 4.2)



Σχήμα 4.2

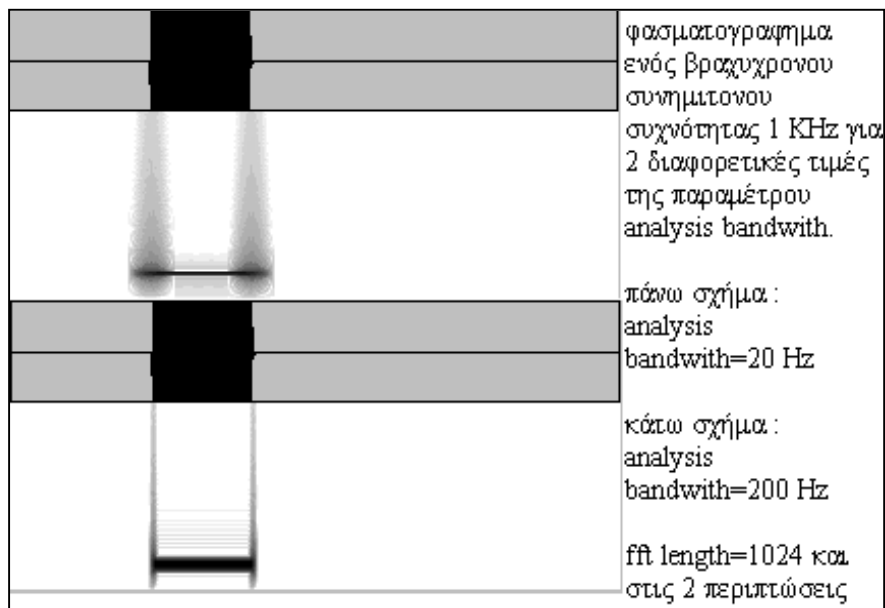
#### 4.1.2)Δημιουργία Φασματογραφήματος

Ρυθμίζεται από τις παραμέτρους που φέρουν τον τίτλο **Spectrogram analysis**(βλ. σχήμα 4.1). Οι παράμετροι που μπορούν να τροποποιηθούν και πώς επηρεάζουν το φασματογράφημα αναφέρονται ακολούθως:

**1)FFT window length** Ορίζει το πλήθος των σημείων του FFT για κάθε παράθυρο του σήματος(window).Υψηλή τιμή στην παράμετρο αυτή συνεπάγεται καλύτερο ποιοτικά φασματογράφημα.Οι τιμές που μπορεί να πάρει η παράμετρος είναι δυνάμεις του 2(από 8 έως 4096).

**2)Analysis bandwidth** Η παράμετρος αυτή επηρεάζει την ανάλυση στον άξονα των συχνοτήτων. Προφανώς όσο πιο μικρή είναι η τιμή αυτής της παραμέτρου τόσο καλύτερη είναι η ανάλυση στις συχνότητες.Όμως μικρή τιμή συνεπάγεται άπλωμα του φασματογραφήματος στο χρονικό πεδίο.Ακριβώς αντίθετα αποτελέσματα έχουμε στην περίπτωση που δώσουμε στη παράμετρο μια αρκετά μεγάλη τιμή.Τότε υπάρχει καλύτερη αντιστοιχία στον χρονικό άξονα αλλά εμφανίζεται ένα άπλωμα στο πεδίο των συχνοτήτων. (βλ. σχήμα 4.3).

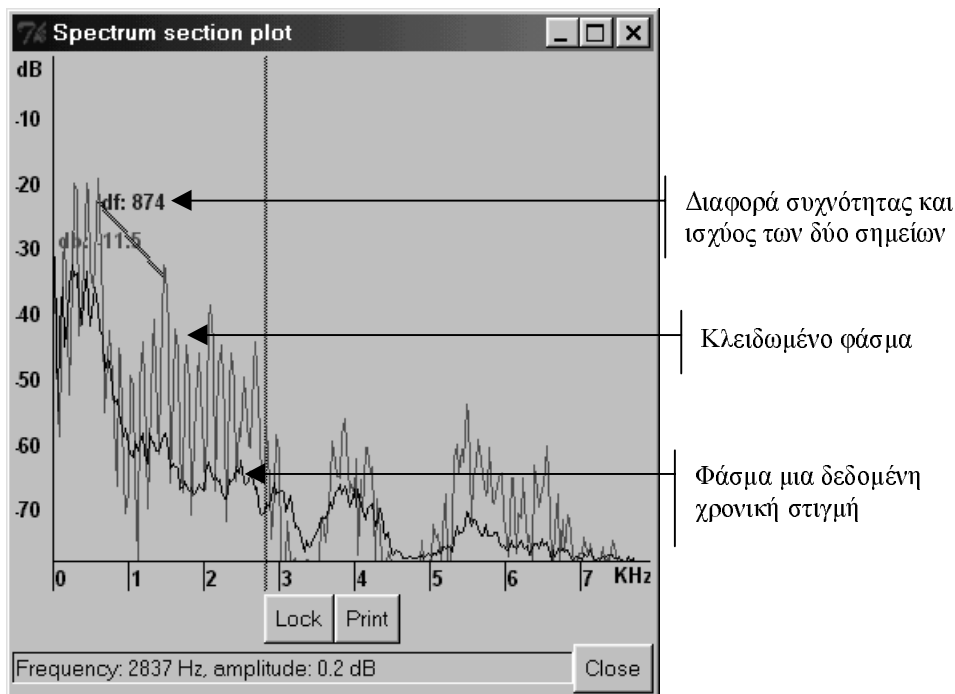
**3)preemphasis factor** Οι τιμές του φασματογραφήματος πολλαπλασιάζονται με αυτήν την τιμή πριν απεικονιστούν.Στην ουσία χρησιμεύει για ανάδειξη της διαφοράς ισχύος ανάμεσα σε διαφορετικές περιοχές του φασματογραφήματος.



Σχήμα 4.3

#### 4.2) Φάσμα(spectrum)

Το φάσμα εμφανίζεται σε ένα ειδικό παράθυρο που ανοίγει από την επιλογή spectrum section του menu **window**. Το παράθυρο αυτό έχει την μορφή του σχήματος 4.4.



Σχήμα 4.4

Στο παράθυρο του φάσματος απεικονίζεται το φάσμα του σήματος σε μια συγκεκριμένη χρονική στιγμή, δηλαδή το πώς κατανέμεται η φασματική ισχύς στις διάφορες συχνότητες μια δεδομένη χρονική



στιγμή. Μετακινώντας τον mouse cursor πάνω στο φάσμα μπορούμε στο κάτω μέρος του παραθύρου να δούμε πληροφορίες για το σημείο στο οποίο βρίσκεται αυτός κάθε φορά. Το φάσμα που απεικονίζεται αντιστοιχεί στην χρονική στιγμή της μύτης γραμμής που έχουμε σημειώσει πάνω στο φασματογράφημα. Είναι δηλαδή σαν πέρνουμε μια τομή του φασματογραφήματος κατά μήκος της ευθείας και τη κοιτάμε από τον άξονα των συχνοτήτων. Στη περίπτωση που έχουμε μαρκάρει κάποιο τμήμα, οπότε υπάρχουν 2 μαύρες γραμμές, το παράθυρο του φάσματος υπολογίζει έναν μέσο όρο για το μαρκαρισμένου τμήμα και το απεικονίζει (αυτό καθίσταται σαφές από την ομαλοποίηση των κορυφών σε αυτήν τη περίπτωση).

Πατώντας το δεξιό κουμπί του mouse σε κάποιο σημείο του φάσματος και με την γνωστή, τώρα, τακτική του dragging (αναφέρθηκε στο κεφάλαιο 1) εμφανίζεται μια γραμμή που ενώνει την αρχικό με το τελικό σημείο εμφανίζοντας επιπλέον πληροφορίες για την διαφορά συχνοτήτων (με μπλέ χρώμα) και ισχύος (κόκκινο χρώμα) μεταξύ τους (βλ. σχήμα 4.4).

Τέλος στο κάτω μέρος του παραθύρου, εμφανίζονται 2 επιλογές:

**1) Lock** Η επιλογή αυτή κλειδώνει το φάσμα που απεικονίζεται τη δεδομένη χρονική στιγμή που θα πατηθεί το κουμπί. Το κλειδωμένο φάσμα θα εμφανίζεται τώρα πια με κόκκινο χρώμα μόνιμα (ή μέχρι να ξαναπατηθεί το Lock) ενώ ο χρήστης μπορεί, πλέον, να επιλέξει το φάσμα σε κάποια άλλη χρονική στιγμή για να το συγκρίνει με το κλειδωμένο. Αυτό φαίνεται παραστατικά και στο σχήμα 4.4.

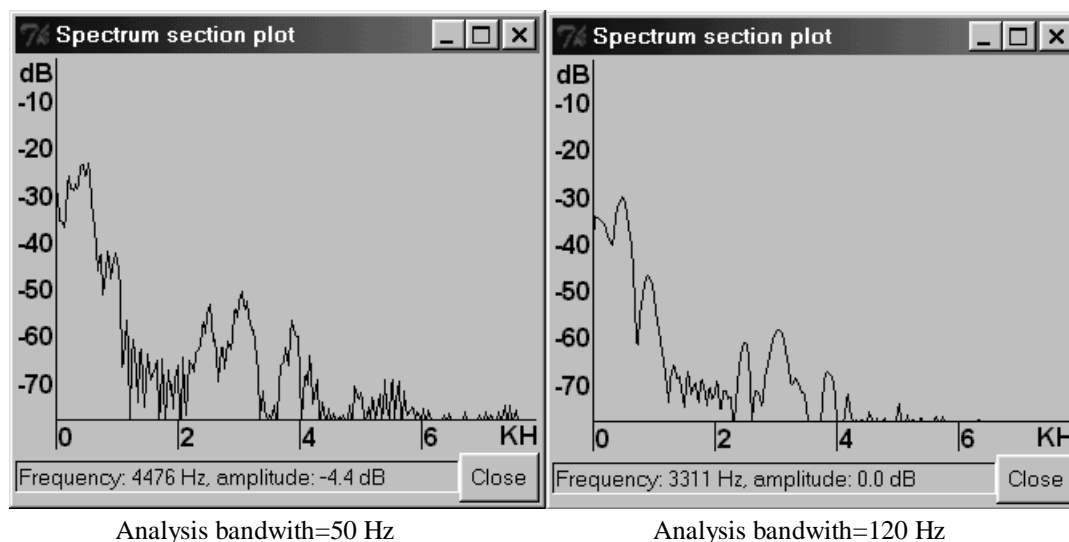
**2) Print** Εκτυπώνει ότι εμφανίζεται στο παράθυρο.

#### 4.2.1) Παράμετροι δημιουργίας φάσματος

Οι παράμετροι δημιουργίας του φάσματος ρυθμίζονται όπως και οι αντίστοιχες του φασματογραφήματος (βλ. σχήμα 4.1) και φέρουν τον τίτλο **spectrum section analysis**.

**1) FFT window length** Ορίζει το μήκος του FFT παραθύρου (πλήθος σημείων). Όσο μεγαλύτερη τιμή έχει τόσο πιο πιστή αναπαράσταση του φάσματος έχουμε. Οι επιτρεπόμενες τιμές του είναι δυνάμεις του 2 (από 8 έως 4096).

**2) Analysis bandwidth** Επειρεάζει την ανάλυση στον άξονα των συχνοτήτων. Μικρές τιμές της παραμέτρου κάνουν το φάσμα πιο ακριβές αλλά με μεγάλη συχνότητα μεταβολών. Αντίθετα μικρές τιμές ομαλοποιούν το φάσμα αλλά μειώνουν την ακριβειά του (βλ. σχήμα 4.5).



Σχήμα 4.5

#### Παρατήρηση

Η μεγαλύτερη χρησιμότητα του φάσματος είναι αναμφισβήτητη η δυνατότητα που δίνει στον χρήστη να προσδιορίσει το φάσμα ενός φωνήματος για έναν συγκεκριμένο ομιλητή.

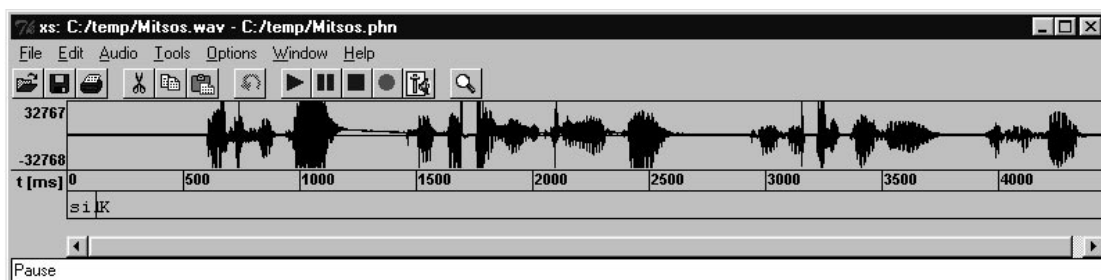
## 5) ΠΑΡΑΔΕΙΓΜΑΤΑ – ΕΦΑΡΜΟΓΕΣ

Στην παράγραφο αυτή θα παρουσιαστούν παραδείγματα εφαρμογής του προγράμματος και θα γίνει επεξεργασία ηχητικών κυματομορφών με βάση τα όσα αναφέρθηκαν στα προηγούμενα κεφάλαια του παρόντος εγχειριδίου. Στην παράγραφο 5.1 θα παρουσιαστούν τα παραδείγματα και στην παράγραφο 5.2 θα παρουσιαστούν οι εφαρμογές.

### 5.1) ΠΑΡΑΔΕΙΓΜΑΤΑ

#### 5.1.1) ΗΧΟΓΡΑΦΗΣΗ ΚΥΜΑΤΟΜΟΡΦΗΣ

Αρχικά ηχογραφήθηκε η φράση “Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών” (με τον τρόπο που περιγράφηκε στο κεφάλαιο 1) και με το τέλος της ηχογράφησης η κυματομορφή εμφανίζεται στο παράθυρο επεξεργασίας όπως φαίνεται στο σχήμα 5.1. Ο ρυθμός δειγματοληψίας του σήματος επιλέχτηκε στα 16000 bits/sec. Θα πρέπει να παρατηρήσουμε ότι στο συγκεκριμένο σήμα είναι ξεκάθαρο από που αρχίζει και που τελειώνει η κάθε λέξη (εκτός ίσως από την λέξη “και” που αρχίζει την χρονική στιγμή 1460msec και τελειώνει την χρονική στιγμή 1620 msec. Εκτός όμως από τις λέξεις είναι ξεκάθαρη η έναρξη και το τέλος και των περισσότερων συλλαβών της φράσης. Αυτό οφείλεται στο τρόπο ομιλίας του εκφωνητή. Ακόμα θα πρέπει να επισημάνουμε και το γεγονός ότι το σήμα ομιλίας παραμένει σταθερό για χρονικά διαστήματα της τάξης των 10.0msec.

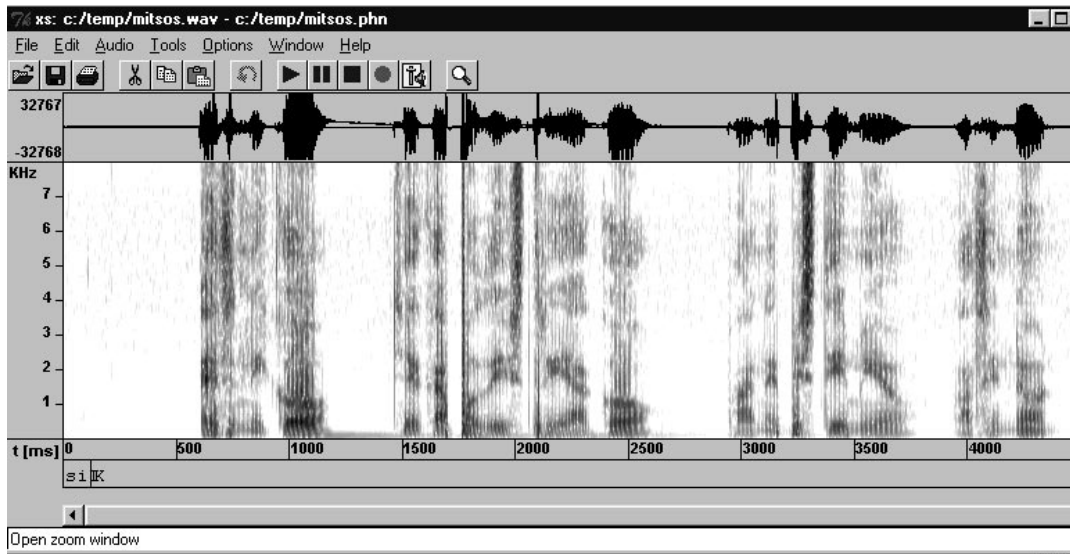


Σχήμα 5.1

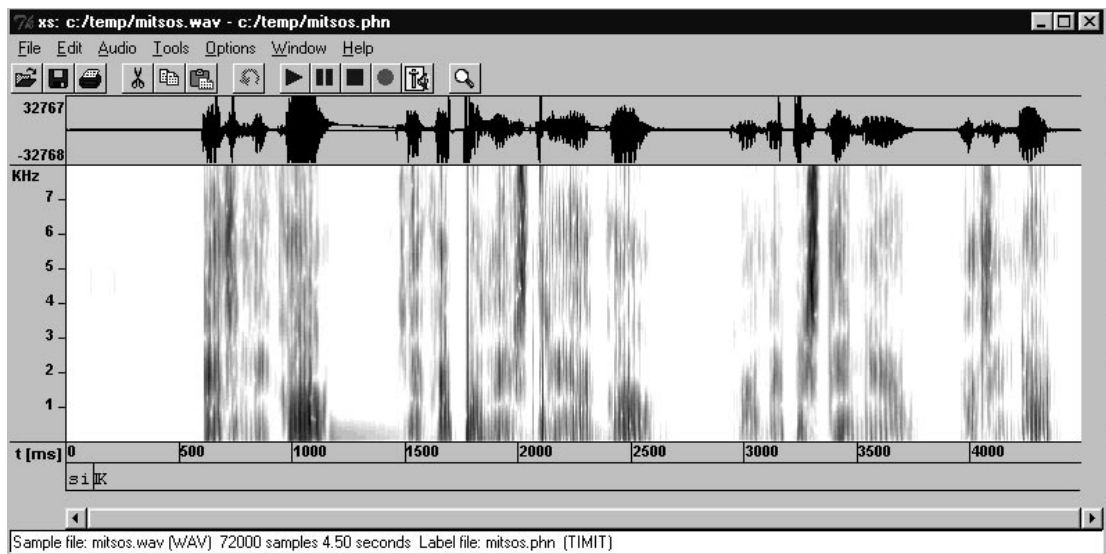
#### 5.1.2) ΦΑΣΜΑΤΟΓΡΑΦΗΜΑ

Θα κάνουμε αρχικά το φασματογράφημα του σήματος για διάφορες τιμές των παραμέτρων του και θα συγκρίνουμε τα αποτελέσματα. Στο σχήμα 5.2 έχει γίνει το τυπικό φασματογράφημα για τιμές των παραμέτρων FFT Window=256, preemphasis factor 0.97 και bandwidth 125 Hz (οι τιμές αυτές είναι οι “default” τιμές που χρησιμοποιεί το πρόγραμμα). Θα πρέπει να επισημάνουμε ότι στο τυπικό φασματογράφημα με μαύρο χρώμα απεικονίζονται οι υψηλές τιμές ενώ με άσπρο οι χαμηλές τιμές.

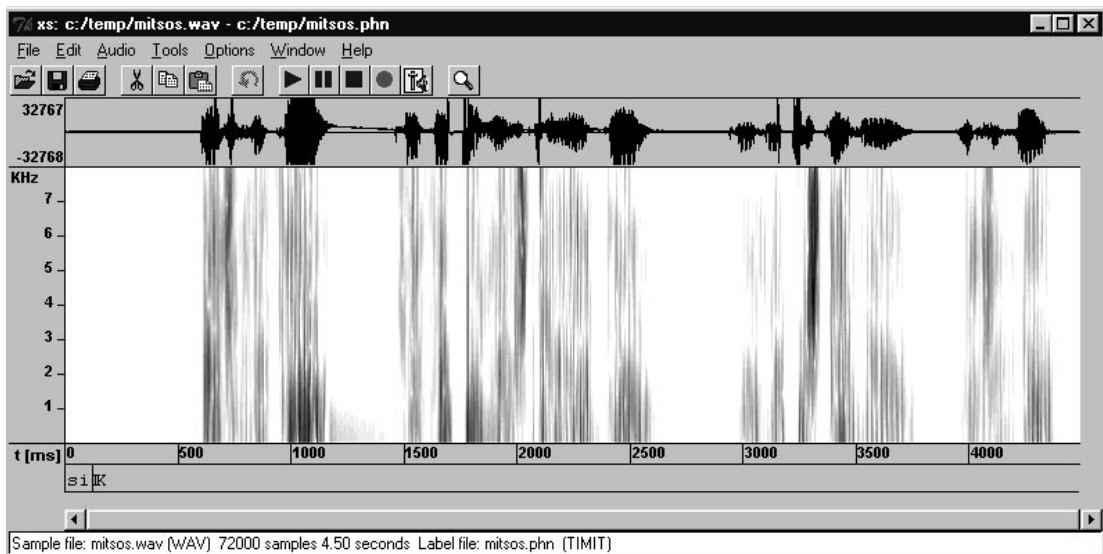
Στο σχήμα 5.3 παρουσιάζονται δύο τυπικά φασματογραφήματα για διαφορετικές τιμές των παραμέτρων Bandwidth και FFT Window τα οποία θα συγκριθούν με αυτό το σχήματος 5.2. Στο πρώτο φασματογράφημα (5.2) έχουμε επιλέξει Window=256 και Bandwidth=125 Hz, στο δεύτερο (5.3(α)) Window = 512 και bandwidth=500 Hz και τέλος στο τρίτο (5.3 (β)) οι τιμές των Window είναι 1024 και Frame είναι 1000 Hz. Από τα τρία αυτά σχήματα συμπεραίνουμε ότι μια αύξηση της τιμής της παραμέτρου Bandwidth οδηγεί σε χαμηλότερης ποιότητας φασματογραφήματα με την έννοια ότι χάνεται η λεπτομέρεια του σχήματος 5.2 στον άξονα των συχνοτήτων.



Σχήμα 5.2



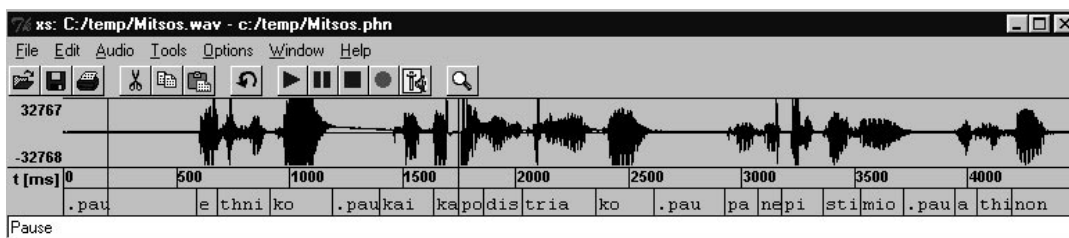
Σχήμα 5.3 (α)



Σχήμα 5.3 (β)

### 5.1.3) ΕΠΙΣΗΜΕΙΩΣΗ

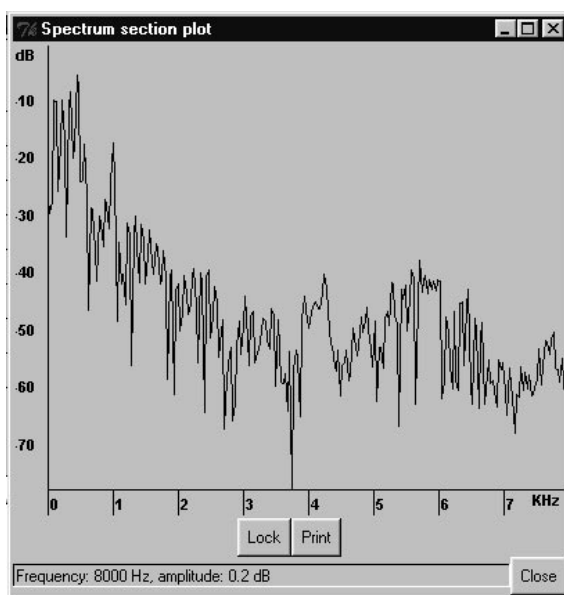
Στο σημείο αυτό θα επισημειώσουμε το σήμα μας με ετικέτες. Θα φτιάξουμε την επισημείωση με βάση τις συλλαβές που στην συγκεκριμένη κυματομορφή είναι όλες σχεδόν εμφανείς που αρχίζουν και που τελειώνουν. Στο σχήμα 5.4 παρουσιάζεται το σήμα επισημειωμένο ακολουθώντας την διαδικασία που περιγράφηκε σε προηγούμενο κεφάλαιο. Με “.pau” σημειώνονται οι παύσεις μεταξύ των λέξεων. Σημειώνεται ότι το πρόγραμμα δεν υποστηρίζει την ελληνική γλώσσα για την επισημείωση και για αυτό σύμφωνα όπως το “δ” αποδίδεται με το “d” πράγμα που προφανώς δεν είναι σωστό.



Σχήμα 5.4

### 5.1.4) SPECTRUM

Μαρκάροντας την χρονική στιγμή 1000 msec ο Snack μας δίνει το Spectrum του σχήματος 5.5. Υπάρχουν αρκετές παραμετροί που μπορούμε να μεταβάλλουμε για την γραφική παράσταση του Spectrum. Από αυτές η παράμετρος Bandwidth αν αυξηθεί θα επιφέρει υποβάθμιση της γραφικής του παράστασης.

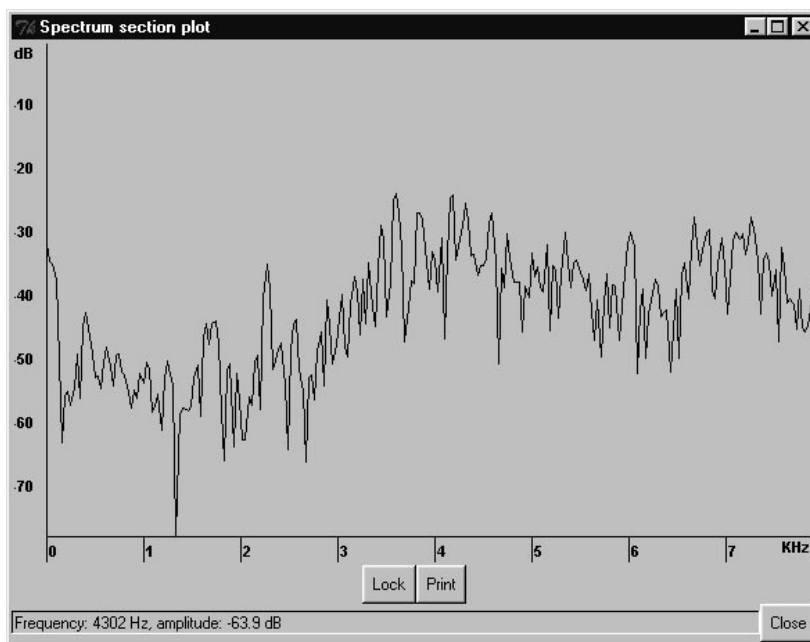


Σχήμα 5.5

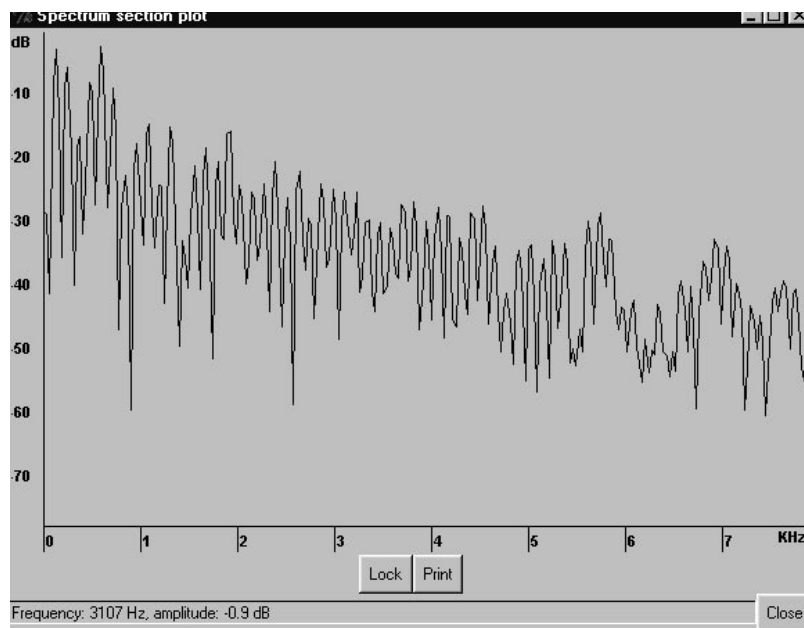
## 5.2) ΕΦΑΡΜΟΓΕΣ

### 5.2.1) ΛΕΠΤΗ ΔΟΜΗ ΤΟΥ ΦΑΣΜΑΤΟΣ ΔΥΟ ΦΩΝΗΜΑΤΩΝ

Με την βοήθεια του Snack θα δείξουμε το φασματογράφημα δύο φωνημάτων της ελληνικής γλώσσας. Αρχικά ηχογραφήσαμε τα δύο φωνήματα με τις διαδικασίες που περιγράφηκαν σε προηγούμενο κεφάλαιο. Τα φωνήματα είναι το /α/ και το /s/ και εκφωνήθηκαν από τον ίδιο ομιλητή. Στην κατασκευάσαμε το SPECTRUM των δύο αυτών φωνημάτων στο σχήμα 5.6 (α) είναι για το /s/ και στο σχήμα 5.6 (β) είναι για το /α/. Και στα δύο φωνήματα διακρίνεται με ευκολία η λεπτή δομή τους. Αντίθετα η φασματική περιβάλουσα δεν είναι εύκολο να προσδιοριστεί.



Σχήμα 5.6 (α)



Σχήμα 5.6 (β)

## ΜΕΡΟΣ ΤΡΙΤΟ : SpeechViewer vs Snack

Στην εργασία αυτή θα συγκρίνουμε δύο πακέτα για επεξεργασία ομιλίας. Πιο συγκεκριμένα θα συγκρίνουμε το πρόγραμμα **Speech Viewer** του CSLU (Center for Spoken Language Understanding) του πανεπιστημίου Oregon με τον **Snack** του πανεπιστημίου της Στοκχόλμης. Η σύγκριση θα γίνει σε δύο διαφορετικά επίπεδα. Στο πρώτο επίπεδο θα γίνει παρουσίαση και σύγκριση των γενικών δυνατοτήτων των προγραμμάτων και στο δεύτερο θα γίνει σύγκριση των παρόμοιων λειτουργιών τους. Τέλος τα συμπεράσματα θα παρουσιαστούν και σε ένα συνοπτικό πίνακα.

Βασικές λειτουργίες των προγραμμάτων – Σύγκριση σε πρώτο επίπεδο

Οι βασικές δυνατότητες του Speech Viewer είναι

- A) Εγγραφή, αναπαραγωγή και αποθήκευση μιας ηχητικής κυματομορφής.
- B) Γραφική παραστασή κυματομορφής.
- Γ) Επεξεργασία κυματομορφής
- Δ) Κατασκευή του φασματογραφήματος της κυματομορφής.
- Ε) Επισημείωση κυματομορφής.
- ΣΤ) Εύρεση της μεταβολής των συντελεστών της μεθόδου Linear Predicting Coding (LPC).

Οι βασικές δυνατότητες του Snack είναι

- A) Εγγραφή, αναπαραγωγή και αποθήκευση μιας ηχητικής κυματομορφής.
- B) Γραφική παράσταση κυματομορφής.
- Γ) Επεξεργασία κυματομορφής
- Δ) Κατασκευή του φασματογραφήματος της κυματομορφής.
- Ε) Επισημείωση κυματομορφής.
- ΣΤ) Εμφάνιση του φάσματος (spectrum) σε κάποια χρονική στιγμή.

Είναι προφανές από τα παραπάνω ότι οι βασικές λειτουργίες των δύο προγραμμάτων είναι παρόμοιες. Τα προγράμματα είναι σχεδιασμένα να επιτελούν τις ίδιες σχεδόν εργασίες, αλλά καθένα έχει και κάποιες επιπλέον δυνατότητες που δεν έχει το άλλο.

Αναλυτική παρουσίαση των λειτουργιών – Σύγκριση σε δεύτερο επίπεδο

A) Εγγραφή, αναπαραγωγή και αποθήκευση μιας ηχητικής κυματομορφής.

Με τα δύο προγράμματα μπορούμε να ηχογραφήσουμε ένα ηχητικό σήμα το οποίο στην συνέχεια μπορούμε να το παραστήσουμε γραφικά, να το αποθηκεύσουμε σε ένα αρχείο και το φορτώσουμε από εκεί στο μέλλον. Τα δύο προγράμματα για τις λειτουργίες αυτές έχουν μία toolbar και κατάλληλες επιλογές των menu. Ο Snack όμως υποστηρίζει αποθήκευση πολλών διαφορετικών format ενώ ο SpeechViewer υποστηρίζει το πρότυπο wav μόνο. Πέρα από αυτό δεν μπορούμε να πούμε ότι κάποιο από τα δύο υπερτερεί του άλλου.

B) Γραφική παράσταση κυματομορφής.

Τα δύο προγράμματα διαθέτουν επιλογές για το φόρτωμα μιας κυματομορφής από ένα αρχείο. Συγχρόνως με το φόρτωμα τα δύο προγράμματα κάνουν την γραφική παράσταση του σήματος σε ένα παράθυρο απεικόνισης. Οι γραφικές παραστάσεις αναπαρίσταν ένα σήμα με την ίδια ποιότητα. Μαζί με την γραφική παράσταση του σήματος τα προγράμματα εμφανίζουν και ένα παράθυρο αντιστοίχισης χρονικών τιμών ώστε ο χρήστης να βρίσκει σε ποια χρονική στιγμή αντιστοιχεί ένα κομμάτι του σήματος ή πότε αρχίζει και τελειώνει μια λέξη στο σήμα. Ο Speech Viewer εκτός από το παράθυρο αντιστοίχισης δημιουργεί και ένα δεύτερο παράθυρο στο οποίο μπορεί να εμφανιστεί ένα τμήμα του σήματος (δηλ με το παράθυρο αυτό κάνουμε zoom στον οριζόντιο άξονα). Επίσης στις επιλογές του προγράμματος είναι και μια για την μεταβολή της κλίμακας του κατακόρυφου άξονα (zoom στον κατακόρυφο άξονα). Το πακέτο του

πανεπιστημίου της Στοκχολμης μπορεί να κάνει μεταβολή της κλίμακας στους δύο άξονες και αυτό, σε διαφορετικό παράθυρο ,όμως, το οποίο δεν εμφανίζεται αυτόματα .

#### Γ)Επεξεργασία κυματομορφής

Ο Snack και ο Speech Viewer έχουν την δυνατότητα να επεξεργαστούν ένα ηχητικό σήμα.Με τον όρο επεξεργασία της κυματομορφής εννοούμε ότι με τα δύο προγράμματα μπορούμε να κόψουμε ένα τμήμα του σήματος , να το αντιγράψουμε και τέλος να το επισυνάψουμε σε ένα άλλο σημείο της ίδιας κυματομορφής ή ακόμα και διαφορετικής.Για τον λόγο αυτό τα προγράμματα έχουν στα menu τους την επιλογή EDIT με τις ίδιες υποεπιλογές.Στον τομέα αυτό δεν μπορούμε να θεωρήσουμε ότι κάποιο πρόγραμμα υπερτερεί του άλλου. Ένα ελαφρύ προβάδισμα αποκτά ο Snack εξαιτίας της δυνατότητας που παρέχει για προσαρμογή της συχνότητας δειγματοληψίας και της κωδικοποίησης των δειγμάτων ενός ήδη ηχογραφημένου σήματος από ειδικό υπομενού.

#### Δ)Κατασκευή του φασματογραφήματος της κυματομορφής.

Στην κατασκευή του φασματογραφήματος ο Snack είναι καλύτερος από τον Speech Viewer .Τα δύο πακέτα κάνουν την επίπεδη απεικόνιση του φασματογραφήματος (δηλ ο χρόνος είναι στον οριζόντιο άξονα , οι συχνότητες στον κατακόρυφο άξονα και οι τιμές του φάσματος στις διάφορες συχνότητες παρίστανονται με διαφορετικές χρωματικές αποχρώσεις).Ο Speech Viewer δίνει την δυνατότητα οι διαφορετικές τιμές των συχνοτήτων να παρασταθούν με διαφορετικούς τόνους του γκρι και με την χρωματική παλέτα ενώ ο Snack μόνο με διαφορετικούς τόνους του γκρι.Ο χρήστης μπορεί να μεταβάλει ορισμένες παραμέτρους για την κατασκευή του φασματογραφήματος και στα δύο προγράμματα.Ο Snack έχει σαν παράμετρος το μήκος του παραθύρου , το παράγοντα προέμφασης (ένας αριθμός που πολλαπλασιάζεται με την τιμή του σήματος πριν την εύρεση του φάσματος) και το Analysis Bandwith (ένας αριθμός που επηρεάζει την παρουσίαση της λεπτομέρειας στον άξονα των συχνοτήτων).Ο Speech Viewer έχει σαν παραμετρούς το μήκος παραθύρου και το μήκος πλαισίου (Frame Size). Το φάσμα του σήματος βρίσκεται και στα δύο προγράμματα με τον FFT (Fast Fourier Transformation).Το σημείο που κάνει τον Snack να ξεχωρίζει είναι η εύρεση σε πραγματικό χρόνο του φασματογραφήματος αλλά και η κατασκευή του SPECTRUM από αυτόν.Μαρκάροντας μία χρονική στιγμή του φασματογραφήματος με τον Snack και ζητώντας του να δημιουργήσει φάσμα(spectrum) του σήματος , παίρνουμε μια γραφική παράσταση που στον οριζόντιο άξονα βρίσκονται οι συχνότητες και στον κατακόρυφο το πλάτος του φάσματος (σε λογαριθμική κλίμακα).Έτσι μπορούμε να έχουμε την ακριβή τιμή του φάσματος σε κάποια συχνότητα αλλά και τον τρόπο που μεταβάλλονται αυτές.Αντίθετα με το κανονικό φασματογράφημα δεν μπορούμε να έχουμε την λεπτομέρεια του spectrum όσο και αν αυξήσουμε την κλίμακα στον κατακόρυφο άξονα.Επίσης με το spectrum μπορούμε να πάρουμε το φάσμα διαφόρων φωνημάτων με τόσο καλή λεπτομέρεια ώστε να φαίνεται η λεπτή δομή τους .

#### Ε)Επισημείωση κυματομορφής.

Τα δύο προγράμματα διαθέτουν την λειτουργία της επισημείωση.Με τον όρο επισημείωση καλούμε την αντιστοίχιση λέξεων , χαρακτήρων ή φωνημάτων σε τμήματα του σήματος ώστε ο χρήστης να γνωρίζει τι έχει εκφωνηθεί.Ο Speech Viewer κάνει την επισημείωση με μεγάλη ευκολία και λειτουργικότητα.Αρχικά ο χρήστης δημιουργεί ένα παράθυρο (ή και περισσότερα εάν επιθυμεί) για την τοποθέτηση των ετικετών.Με το πληκτρολόγιο ή το ποντίκι μπορεί να το χωρίσει σε όσους τομείς θέλει και να μεταβάλει τα όρια τους.Αντίθετα ο Snack παρουσιάζει κάποια προβλήματα στον τομέα αυτό.Το παράθυρο των ετικετών δημιουργείται δύσκολα γιατί ο χρήστης πρέπει να κάνει 'κλικ' με το δεξί κουμπί του ποντικιού σε ένα πολύ μικρό χώρο μεταξύ της κυματομορφής και ενός παραθύρου εμφάνισης πληροφοριών.Βέβαια υπάρχει και δεύτερος τρόπος δημιουργίας του συγκεκριμένου παραθύρου με την χρήση ενός επεξεργαστή κειμένου (όπως το Wordpad) και αυτός όμως είναι πιο χρονοβόρος και απαιτεί καλή γνώση του τρόπου αποθήκευσης των ετικετών σε αντίστοιχα αρχεία. Ο χωρισμός του παραθύρου σε τομείς γίνεται μόνο με την βοήθεια του ποντικιού.Η λειτουργικότητα του Speech Viewer του δίνει ένα σημαντικό πλεονέκτημα στον τομέα αυτό.



Συνοψίζοντας τα δύο προγράμματα εκτελούν παρόμοιες λειτουργίες. Η εγγραφή αναπαραγωγή και αποθήκευση, γραφική παράσταση κυματομορφής και η επεξεργασία κυματομορφής γίνονται και από τα δύο προγράμματα εξίσου καλά. Για επεξεργασία ενός ηχητικού σήματος με το φασματογράφημα θα πρέπει να χρησιμοποιήθει ο Snack ενώ για επισημείωση και εύρεση της μεταβολής των συντελεστών της μεθόδου LPC ο Speech Viewer. Τα αποτελέσματα της παραπάνω σύγκρισης παρουσιάζονται στον ακόλουθο πίνακα.

<div style="text-align: right;">Πακέτο</div> <div style="text-align: left;"><u>Λειτουργία</u></div>	<u>SpeechViewer</u>	Snack
Εγγραφή-αναπαραγωγή-αποθήκευση	Υποστήριξη κυρίως του wav format	Υποστήριξη πολλών διαφορετικών formats ήχου
Γραφική παράσταση σήματος	Τα δύο πακέτα έχουν πάνω-κάτω τις ίδιες δυνατότητες	
Επεξεργασία Σήματος	Παρέχει όλες τις βασικές λειτουργίες για την επεξεργασία του σήματος (cut, paste κ.ο.κ.)	Παρέχει την πρόσθετη δυνατότητα της αλλαγής της συχνότητας και της κωδικοποίησης των δειγμάτων ακόμα και μετά την ηχογράφηση
Ανάλυση ομιλίας	1)Τυπικό φασματογράφημα 2)Φασματογράφημα κορυφής 3)LPC	1)Τυπικό φασματογράφημα 2)Φασμα (Spectrum)
Επισημείωση	Τα 2 πακέτα παρουσιάζουν τις ίδιες δυνατότητες πάνω-κάτω. Παρολαυτά ο Speech Viewer είναι πιο εύχρηστος	

## **Βιβλιογραφία-Πηγές**

CSLU homepage : <http://CSLU.cse.ogi.edu/>

Snack homepage : <http://www.speech.kth.se/snack/>

## ΠΙΝΑΚΑΣ ΑΓΓΛΙΚΩΝ ΟΡΩΝ

Default	Αρχικά προεπιλεγμένη
Frequency	Συχνότητα
Frame	Πλαίσιο
Group	Ομάδα
Label	Ετικέτα
Labelling	Επισημείωση
Lock	κλειδώνω
Menu	μενού
Pixel	Στοιχείο απεικόνισης
Plug in	πρόσθετο
Sample	Δείγμα
Sample rate	Ρυθμός δειγματοληψίας
Scrollbar	Μπάρα κύλισης
Text	κείμενο(π.χ text file: αρχείο κειμένου)
Time	Χρόνος
Toolbar	Μπάρα εργαλείων
Window	Παράθυρο
1D display	μονοδιάστατης απεικόνισης
2D display	διδιάστατης απεικόνισης
Label	ετικέτας
Map	απεικόνισης(αντιστοίχισης)
Ruler	χρονικού άξονα
Spectrogram	φασματογραφήματος
Peak	κορυφής
Typical	τυπικού(κλασσικού)
Wave	επεξεργασίας
Spectrum	Φάσμα