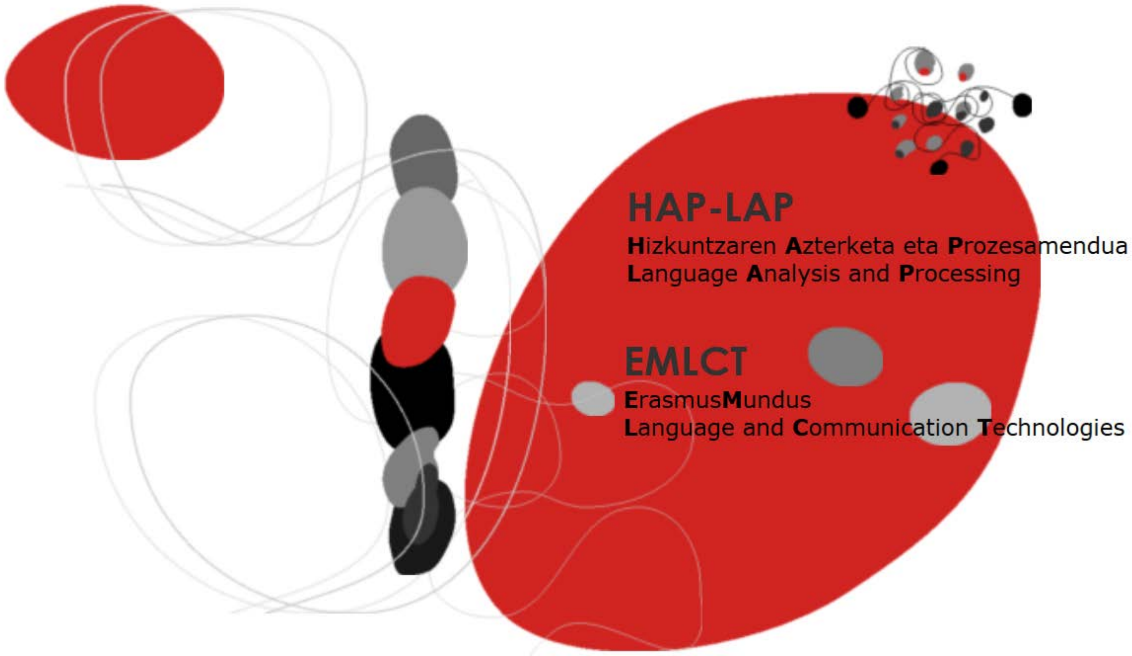
# *Introduction to audio managing software*




**Inmaculada Hernáez** inma.hernaez@ehu.es

**Eva Navas** eva.navas@ehu.es

**Ibon Saratxaga** ibon.saratxaga@ehu.es

# Outline

# 1 Software for the analysis of signals

When managing speech and audio signals we need the help of software tools in order to record, analyze or edit the signals. There are of course many different possibilities, and some of them are described in this document.

When working with speech signals it is also of interest the use of a tool to help with the transcription task. This task consists on assigning labels to different intervals in the signal. For example, we may want to assign different labels to the time intervals corresponding to two different phonemes. Some of these programs are also useful to manage different levels of transcriptions (for example orthographic transcription and phonetic transcription).

If we want to record a signal using the PC audio card we must pay attention to use the correct input connector. Section 6 in this document gives some indications to correctly perform a recording. It is also important to set the recording levels such as not to produce saturation. The recording device level together with the signal amplitude level must be carefully set so that no saturation is produced while a high input signal level is kept (a level around -3dB is recommended).

# 2 Audacity

Visualizing audio signals can be done with the audio editor *Audacity*. This software is available at:

http://audacity.sourceforge.net/download/

The user manual can be found at http://manual.audacityteam.org/o/.

*Audacity* is a software tool to record and edit audio signals in an easy way. It is designed for several platforms (Windows, Linux, IOS), free and open source. Apart from recording and playing, it can import and export audio signals from multiple formats (WAV, AIFF, MP3 and many more).

With Audacity you can:

- Record from a microphone, input line or other audio sources.
- Create multi-track recordings by copying over existing tracks.
- Record up to 16 channels (requires multichannel recording hardware).

You can also create audio files by cutting, copying, pasting and deleting. Unlimited levels of undoing and redoing can be used to return to a previous state. It also allows the edition of big files, editing and mixing an unlimited number of tracks and the modification of individual samples using the drawing tool.

To record an audio signal, the classic recording icons are used. Configuration of the recording parameters is performed from the *Edit->Preference* menu, selecting the appropriate values at

Devices and Quality, as shown in Figure 1. The recording input level is shown at the upper line of icons, as shown in Figure 2 and care must be taken that the level is kept under the saturation level (in green).

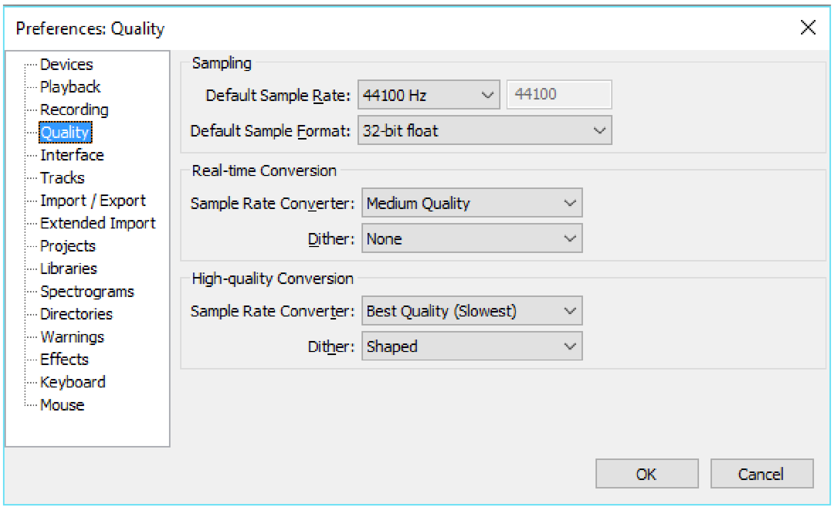



Figure 1: Configuring sample rate in Audacity

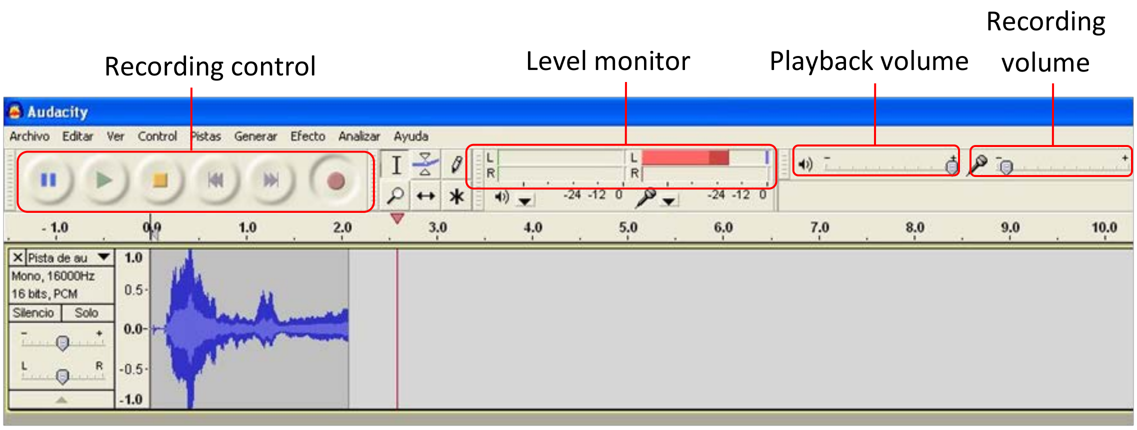



Figure 2: Recording window in Audacity

Record a mono speech signal at 44,1 kHz and using 16 bits per sample and save the file using the option *export* from the *File* menu (export as WAV) . Use a long sentence with multiple different sounds. If you cannot record a signal, you can use the "*Long speech signal"* provided in eGela.

The default view in Audacity is the signal over time. We can see the spectrogram (a time-frequency representation of the signal that we will study in topic 2 Basic concepts about signals & systems) by selecting the option *Spectrogram* from the available views at the lateral dropdown menu as shown in Figure 3. The parameters to calculate the spectrogram are controlled from the *Edit->Preferences* menu, under the section *Spectrograms*.

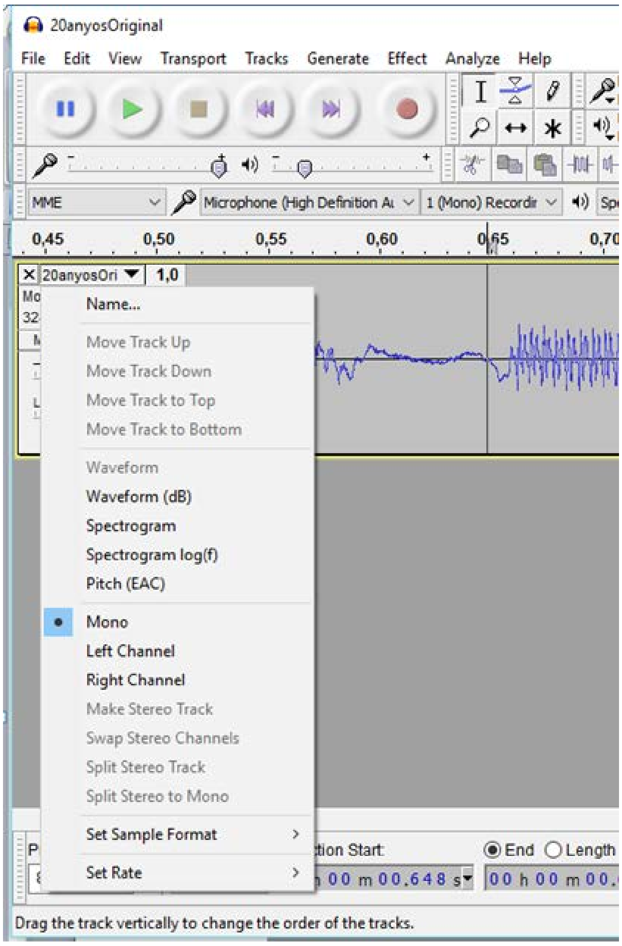
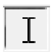



*Figure 3: View selection window in Audacity*

With Audacity we can also edit the audio file: Select one of the words around the center of the recorded sentence with the selection tool (with this icon: ). To delete the word, use *Edit->Cut* or type Ctrl+X, just as you would do in a text editor. Listen to the sentence and verify that the selected word has been eliminated. If the performed deletion is perceived, undo the changes using Ctrl+Z and try to carefully select the word limits, such that the cut is not perceived. Likewise, you can insert audio segments (from the same file or from a different one) by selecting the interval and typing Ctrl+C, inserting the cursor at the desired insertion point and then typing Ctrl+V. Try to copy the last word of your sentence at the beginning of the sentence, trying to avoid undesired noises.

*Audacity* also allows applying different effects to the audio signal. For example, we can amplify the signal, eliminate noise, change the pitch, speed and tempo and many others. To apply the amplification operation to the signal (obtaining kx(t) as explained in theory), select the first word of your sentence and amplify it by 6 dB using *Effect->Amplify*. Check the change in volume in this part of the sentence.

Recover the original signal and change the speed with the corresponding effect. This corresponds to the time scaling operation seen in theory (x(at)). Increase the speed by 30% and listen to the result. What is the effect of reducing speed? Do you perceive any additional difference?

To modify the duration without altering the frequency content *Audacity* provides the *Tempo modification* effect (*Cambiar ritmo* in the Spanish version of the menu). Recover the original signal and reduce the tempo by 20%. Listen to the result and check the new duration of the signal. Is there any other perceived difference besides the duration expansion? Try now reducing the tempo by 80%. Has the signal been distorted? Which is the limit value for the reduction percentage in order to avoid distortion?

Record a speech signal containing a phonetic palindrome (like "say yes", "easy" or "funny enough") and save it using the option *Export Audio* in the *File* menu. If you cannot record the signal, use the signal *"Phonetic palindrome"* from eGela. Apply the effect *Reverse* (*Revertir* in Spanish) to the palindrome. This is equivalent to the time inversion operation seen in theory (x(-t)). Listen to the resulting signal and compare it with the original one (remember that you can undo all your changes with Ctrl+Z and redo them with Ctrl+Y). Can you understand the message after applying the reverse effect?

Change the pitch of the first half of the signal: Click on *Effect->Change Pitch* and choose either a Frequency *Percent Change* (positive or negative) or move the sliding bar at the lower part of the window (see Figure 4). You can also directly enter the desired frequency value. If the recorded signal corresponds to a female choose a percent change between -25 and -15. If it is a male voice make a positive change. Listen to the results. Test with different frequency values, using higher and lower percent changes. Remember that you can undo all your changes using Ctrl+Z. Which are the final frequency values (minimum and maximum) that you can apply without perceivable distortion? How much higher or lower is this final value than the original initial frequency?

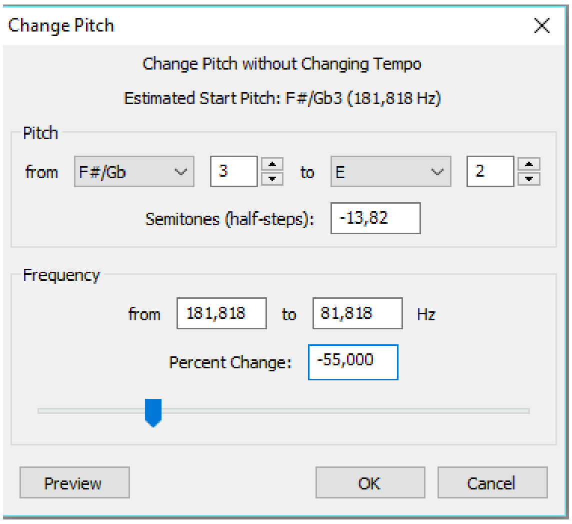



*Figure 4: Selection of the values to apply a pitch change in Audacity*

Another interesting effect is *Noise Removal*. Open the file 'aeiou_uxoa.wav' which contains five sustained vowels pronounced by a female speaker. Listen to the signal and observe the considerable level of noise. In order to clean the signal, use the Noise Reduction effect and apply the following instructions: get a Noise Profile by first selecting a noisy interval on the signal and then clicking on the Get Noise Profile button (Step 1); then apply a 12 dB noise reduction to the signal by first selecting the interval to clean and then selecting again the Noise Removal effect to apply the effect (Step 2). Experiment with different segments of the signal to obtain better results. It is also possible to observe the noise reduction effect over the signal spectrogram.

# 3 Wavesurfer

*Wavesurfer* is an open source software tool for the visualization and manipulation of audio signals. Its use is very popular for speech analysis and transcription. The program is available from:

https://sourceforge.net/projects/wavesurfer/

With this program a new speech file can be recorded using the recording control buttons shown in Figure 5. The recording parameters are configured through the *File->Preferences* menu, at the tab *Sound I/O* shown in Figure 6.

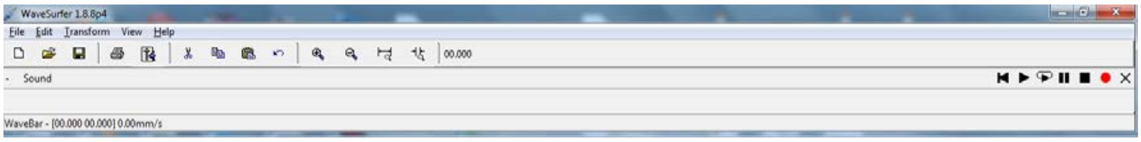



*Figure 5: Opening window in Wavesurfer*

Using *Wavesurfer,* record a speech signal with similar characteristics as the one recorded with *Audacity*. By right-clicking on the main window, select the *Apply Configuration* option and you will be offered a list of possible configurations. Select *Waveform* configuration.

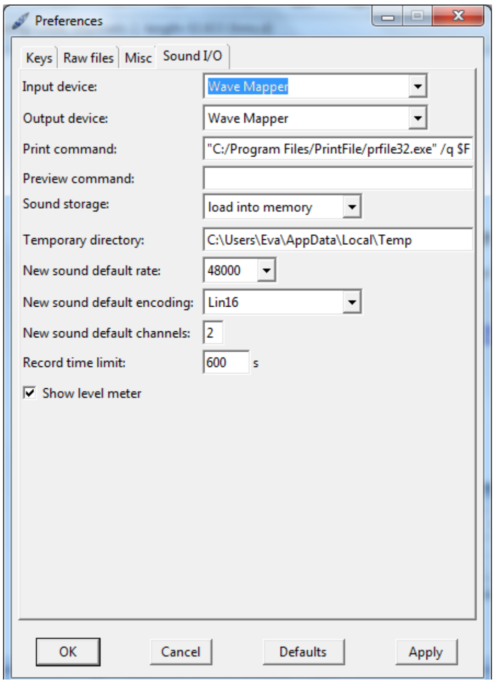


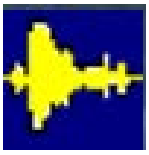

*Figure 6: Configuration of the sampling frequency in Wavesurfer*

Wavesurfer is able to apply some effects to the sound signal. Obtain the signal kx(t) by using the option *Amplify* in the *Transform* menu. You can select the amplification in the linear scale by fixing a percentage value or in the logarithmic scale by choosing a dB value. Attenuate the first word of the signal to get half the power of the original one in this segment. Check the decrease in volume in the first part of the signal. You can undo the changes done to the signal clicking the icon .

Open the phonetic palindrome signal recorded with Audacity and build the signal x(-t) using the *Reverse* option of the *Transform* menu.

Wavesurfer also allows labeling the signal at different levels: phoneme, syllable, word, etc. Add a new panel for the transcription and create the word level transcription of the recorded sentence. To do that, put the cursor at the selected label insertion point on the transcription panel and write the corresponding text. Use the label 'sil' for the intervals of silence. You can save the transcription in a file by right-clicking over the transcription panel and selecting the *Save Transcription As* option. Using a text editor open the file just created and analyze its format. Which units are used to measure the word limits?

## 4 Speech Analyzer



*Speech Analyzer* is yet another program to record and analyze the speech signal. It is freely distributed and can be downloaded and installed from here:

http://www.sil.org/computing/sa/sa_download.htm

The recording is started by clicking on the icon. Clicking on the Settings button you can select the sampling frequency, number of bits per sample and the convenience of using or not a filter to eliminate the frequencies under 70 Hz (We will study filters in topic 2 Basic concepts about signals & systems). As shown in Figure 7, there is a recording level meter that reaches the red zone the moment the signal is saturated.

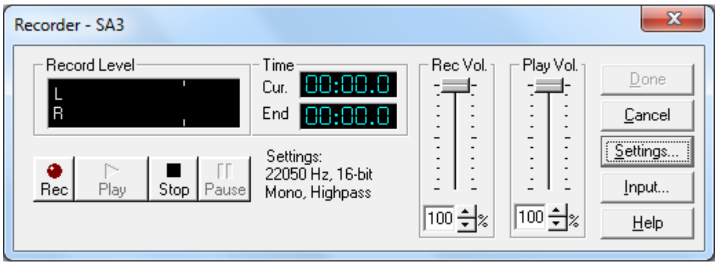



*Figure 7: Recording interface in* Speech Analyzer

Record a speech signal in mono 44,1 kHz and 16 bits per sample with Speech Analyzer and pay attention to unselect the high pass filter. Save the file. At the lateral panel we can choose the configuration of the graphics to be shown. Choose *Waveform* to watch the speech signal in time.

*Speech Analyzer* allows for analyzing the speech signal and visualizing different speech related parameters like pitch, energy, formants that we will study in topic 3 *"Speech signal: representations"*.

It is also possible to label and segment the signal on the bar named *Phonetic* that appears over the waveform representation. After selecting an interval with the left and right cursors, click on the icon to introduce the phoneme label, changing the symbol with the corresponding phone label. The phonetic level should be the first level to be labeled, because the higher levels (such as orthographic or word levels) will be aligned with it. In this exercise, we will use the phonetic level to label the signal at word label. To do this, put the left cursor (green) at the beginning of the first word and the right cursor (red) at the end of the same word. When labeling it is very useful to see also the spectrogram to ease the labeling task. Listen to the selected interval (click on the icon). When the word limits are well positioned add the corresponding label. To introduce the next label, the program will position the new beginning of the word at the end of the last labeled word. The borders can be moved later by pressing the icon. Repeat the process until all the words in the signal are labeled.

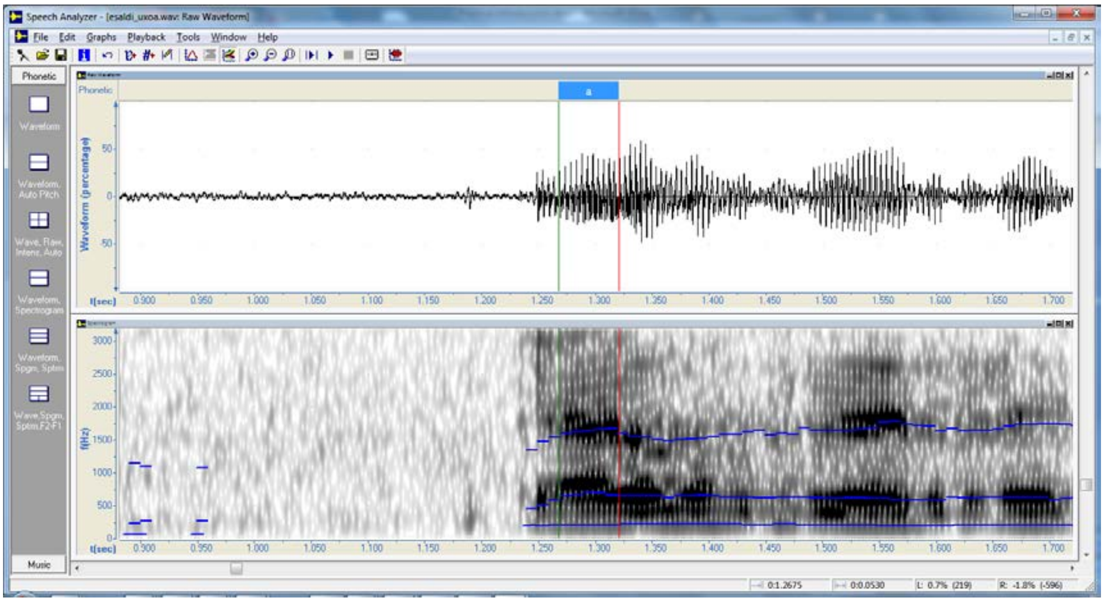



*Figure 8: Phonetic labeling in Speech Analyzer*

Save the produced word label segmentation using *File->Export->SFM Time Table (.SFT)* selecting the region corresponding to the entire file. Open the created file with a text editor and observe the format used. Which units are used to save the time labels?

# 5 Praat

*Praat* is a program for the phonetic analysis of the speech signal that can be obtained from the page: http://www.fon.hum.uva.nl/praat/ and whose user manual can be found here:

http://www.fon.hum.uva.nl/praat/manual/Intro.html

When opening *Praat* two windows appear:

- *Praat Objects*, in which the recorded (or file opened) signals are listed, together with other objects created by the program. It also contains the options for the analysis.
- *Praat Picture* where the graphs corresponding to the different objects from the object window are shown and can be saved in different image formats.

The range of values for the sampling frequency offered by *Praat* is wider than in *Speech Analyzer*. At the *Praat Objects* window select the option *Record mono Sound…* o *Record stereo Sound…* as desired from the menu *New*. Figure 9 shows the recording interface, where the saturation levels are shown in red color.


*Figure 9: Audio recording window in Praat*

Record a new speech signal at 44,1kHz and name it by changing the literal untitled from the field *Name*. Press *Save to list & Close* to go back to the object list. A new object of type Sound with the assigned name appears in the list of objects. The spectrogram and the waveform can be seen by pressing *View & Edit* button, as shown in Figure 10.

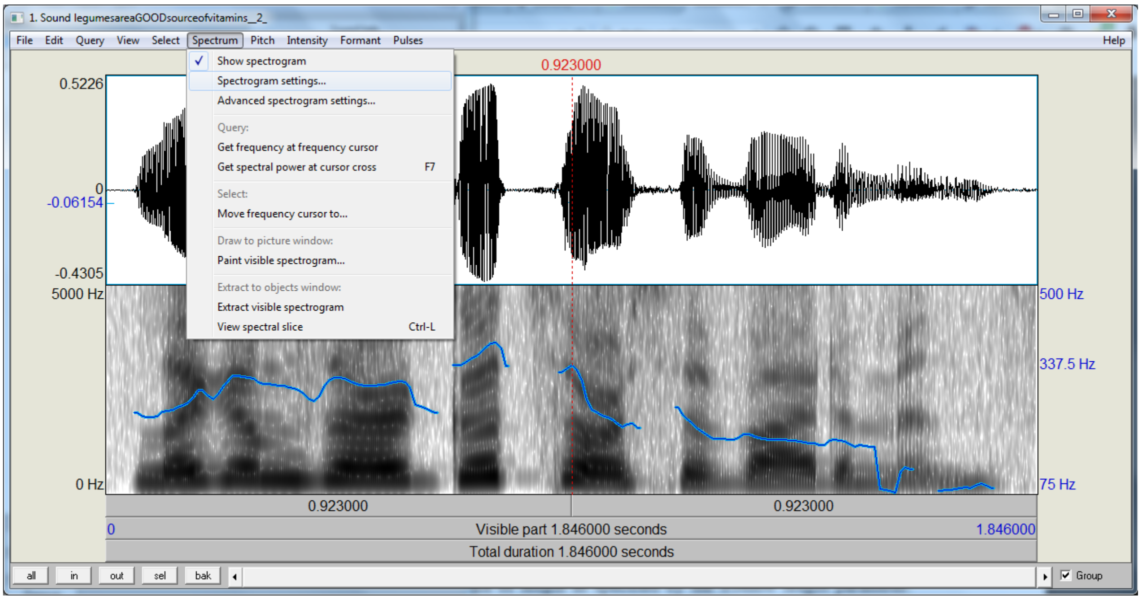



*Figure 10: Signals display window in Praat*

There are many different processing possibilities in *Praat*, like time reversing the signal (building x(-t)). Open the phonetic palindrome by using the option *Read from file…* in the *Open* menu or by pressing Crtl O. You can listen to your signal with the *Play* button. Select the object corresponding to the palindrome in the object list and click *Modify->Reverse*. Listen to the time reversed signal. There is no possibility to undo operations in Praat, so to recover the original palindrome signal apply time reversing to it again. It is also possible to amplify the signal (obtain kx(t)) with the option *Multiply* in the *Modify* menu and to scale the time axis (obtain x(at)) with the option *Modify->Modify Times->Scale times by*. Produce a faster version of your signal using a factor of 0.5. Check that the duration of the signal is now half the previous duration. Is there any other difference between the original signal and the time compressed one?

*Praat* is also used to label audio signals. Select the signal recorded with *Praat* and press the button *Annotate* and *To TextGrid* to label the signal at word level. In the pop-up window you must write the name of the level or levels (*Tiers*) of labeling into the field *Tier names* (words in this case). If at any labeling level the labels are points in time (instead of time intervals), this must also be included in the field *Which of these are point tiers?*. In our case, as words have non-zero duration we will use time-interval labels, so that field must be left blank. Select the corresponding *Sound* object together with *TextGrid* (using Ctrl) and then press *View & Edit* to open the labeling window. A yellow area where the word limits and labels must be added appears at the bottom of the bottom of the graphic, as shown in Figure 11. Display the signal and the spectrogram and click over the signal at the beginning of the first word of your sentence. The first time bar will appear divided from that point. By clicking on it we can listen to the signal interval up to or from that point, in order to reassure that the word boundary has been correctly set. When the point is the desired one, we can set the boundary with the Enter key, and go forward to set next boundary. Once the interval corresponding to a word has been fixed, we can select it in a way that it is shown in yellow, the word label can be directly written. The boundaries can be moved by clicking on them and dragging them to the new position.

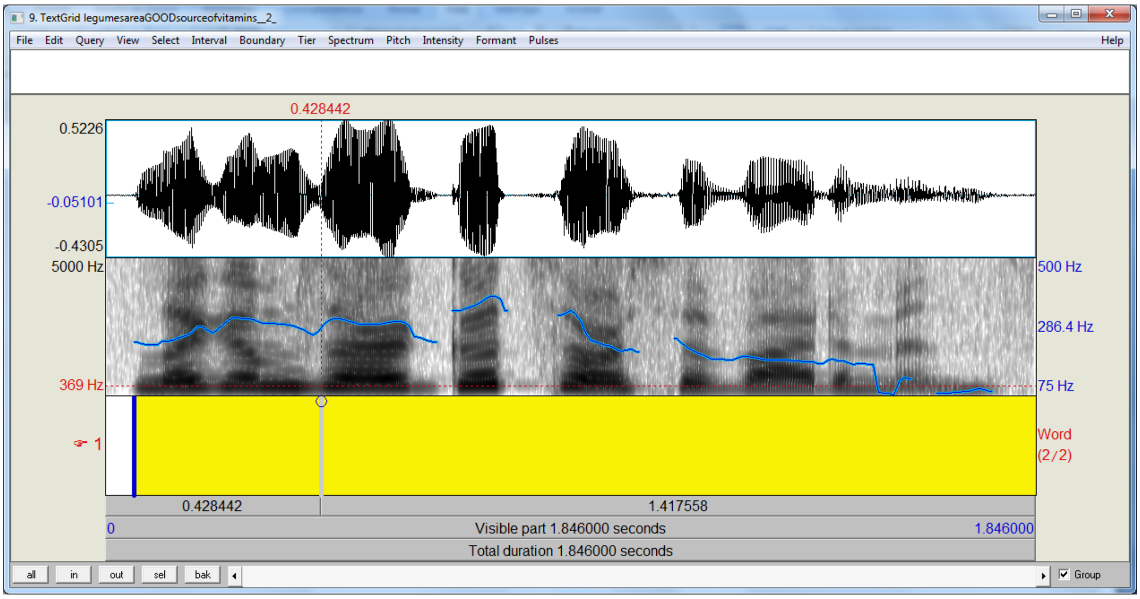



*Figure 11: Segmentation and labeling window in Praat*

*Praat* allows us for extracting the segments corresponding to each label by selecting the object *Sound* and the corresponding *TextGrid*, pressing *Extract* and selecting if we want to divide the signal into the labeled intervals for all or only some of the labels (those fulfilling certain conditions). Extract the signals that correspond to each word labeled in your sentence. What happens with the unlabeled intervals of silence?.

The created labels can be saved in two different ways:

- From the label display window (window *TextGrid*) with *File -> Save TextGrid as text file*
- From the object window by selecting the *TextGrid* object and choosing *Save -> Save as text file*

The created labels files can be read and added to the list of objects with *Open -> Read from file*.

Save your word labels and open them with a text editor. Observe the file format and find the fields corresponding to the time points and labels. Which units are used for the time points in this case?

# 6 Annex: Selection of the audio input and setting the recording volume in Windows.

The recording of the signals can be done through the *line input* or the *microphone input* of the sound card. Therefore, it is necessary to select the corresponding device as input recording device. This selection must be done from the Windows *Volume Control*, right clicking in the speaker icon and selecting Recording devices, as shown in Figure 12.

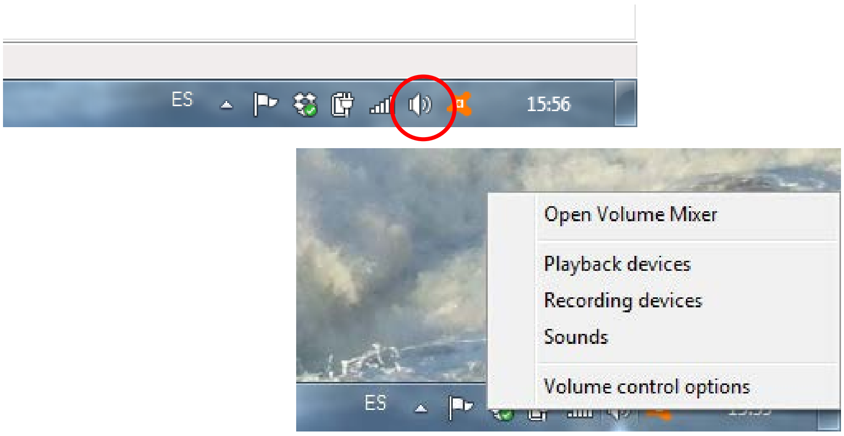



*Figure 12: Opening the Control of the Recording devices*

Once the Control of the Recording devices window is open, we must select the suitable device in the Recording tab and click *Properties* to open the window shown in Figure 13. In the *Levels* tab you can adjust the volume of the recording in order to avoid signal saturation.

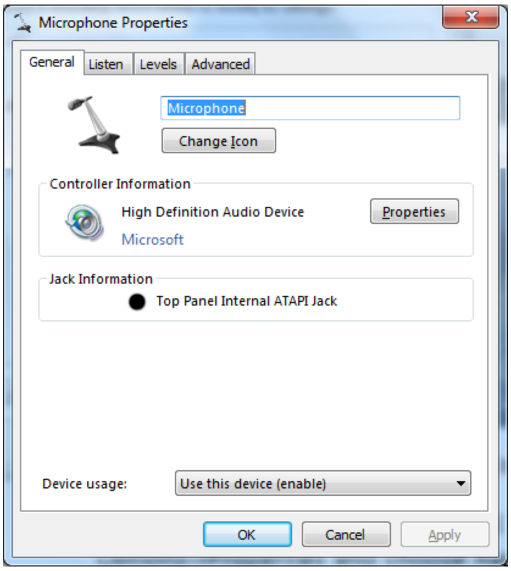



*Figure 13: Selection of the recording devices*