"Gheorghe Dima" Academy of Music, Cluj-Napoca

Music informatics

Processing of digital sounds

Course handbook

Lect. univ. dr. Cristian Mihăescu

Cluj-Napoca, 2010

I. General notions

Sounds can be stored in two **formats**:

- analogical format (used in the past): tapes (cassette players and tape recorders), vinyl records, etc;

- digital format (used currently): data stored on computers, DAT (*digital audio tape*), CDs or DVDs.

The sounds stored in analogical format can be converted into the digital format through the process named **digitalization**, using the computer's sound board and some dedicated software such as *Audacity,* recommended due to its high quality and the fact that it is open source.

The analogical source is connected through an audio cable to the sound board through the *Line-In* input.

Digitalization

Analog sound – Line-In – soundboard – Audacity – digital sound

The digitalization process will be described in detail in a separate chapter.

The formats in which the sounds can be stored in a computer are, in general:

- the MIDI format (such as a score);

- the WAVE format (digital audio).

The MIDI format consists in score files (with the extension MID, or extensions specific to the software that generates them). These files store only information about the

score files (pitch and duration of sounds, nuances, tempo, etc.), without elements of timbre, which confers them a high flexibility. Thus, the MIDI files will have the following features:

- the small size of the files – an advantage for portability and storage;

- the possibility of editing elements characteristic of a score (format used in composition);

- does not include the instruments' timbre or other audio events. It uses the timbres of the sound board or of the external modules;

- the sound is somewhat artificial, as it depends on the quality of the timbres used (which is also the big disadvantage of the MIDI format);

- compatibility with most synthesizers, composition software and music score editing software.

The audio digital format consists of files that most often have the *wav* extension, which contain music recordings. These files have the following features:

- they contain the real timbres and all the audio events present during recording;

- large file sizes[1];

- the tune (or other elements of the score) cannot be edited;

- one can perform operations specific to sound engineering, such as quality processing (elimination of background noise and impurities, resampling), mixing, etc.
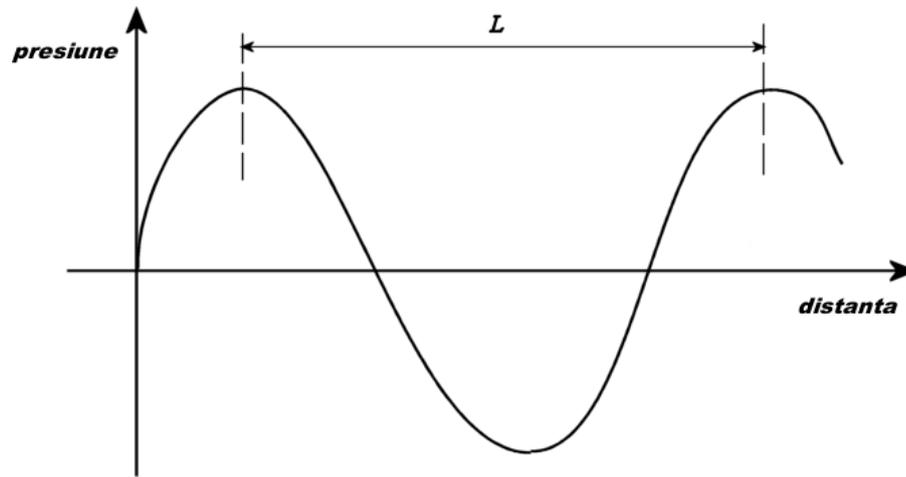
The features of the digital sound are:

- frequency;

- resolution;

- number of channels.

1. The frequency of the digital sound

Sound is a variation that can be detected by the human ear. The image below depicts the wavelength of sound:

---

[1] These sizes can be reduced by compressing them into MP3 formats

Pressure   **Distance**

The frequency of the digital sound is defined as the number of pressure variations within a unit of time and is measured in Hertz (Hz).

The frequencies are classified in three categories: low, middle, and high frequencies.

The human ear generally perceives the sounds ranging from 20 Hz to 20 kHz. All the frequencies outside this interval are not perceivable to the human ear.

As examples, the high frequency level of the tape recorder does not exceed, in general, 22 kHz, the cassette player reaches 20 kHz only on metal music tapes, while a phone conversation takes place at low frequency thresholds (ca. 10 KHz), which leads to a very low audio quality.

The frequency standard for audio CDs is 44,1 kHz.

The quality and quantity (the occupied space) of a recording grow directly proportionally to the frequencies' interval.

2. The resolution of the digital sound

As a consequence of the digitalization process, or after a recording in digital format, we can notice that the sound is not recorded continuously, so that a digital recording contains a number of samples, which succeed one another so fast that the ear perceives a continuous sound (a process resembling the running of a movie at 30 frames per second, which the naked eye perceives as a continuous movement).

The following image depicts a very short fragment (under a millisecond) from a two-channel (stereo) recording, in order to observe the samples.
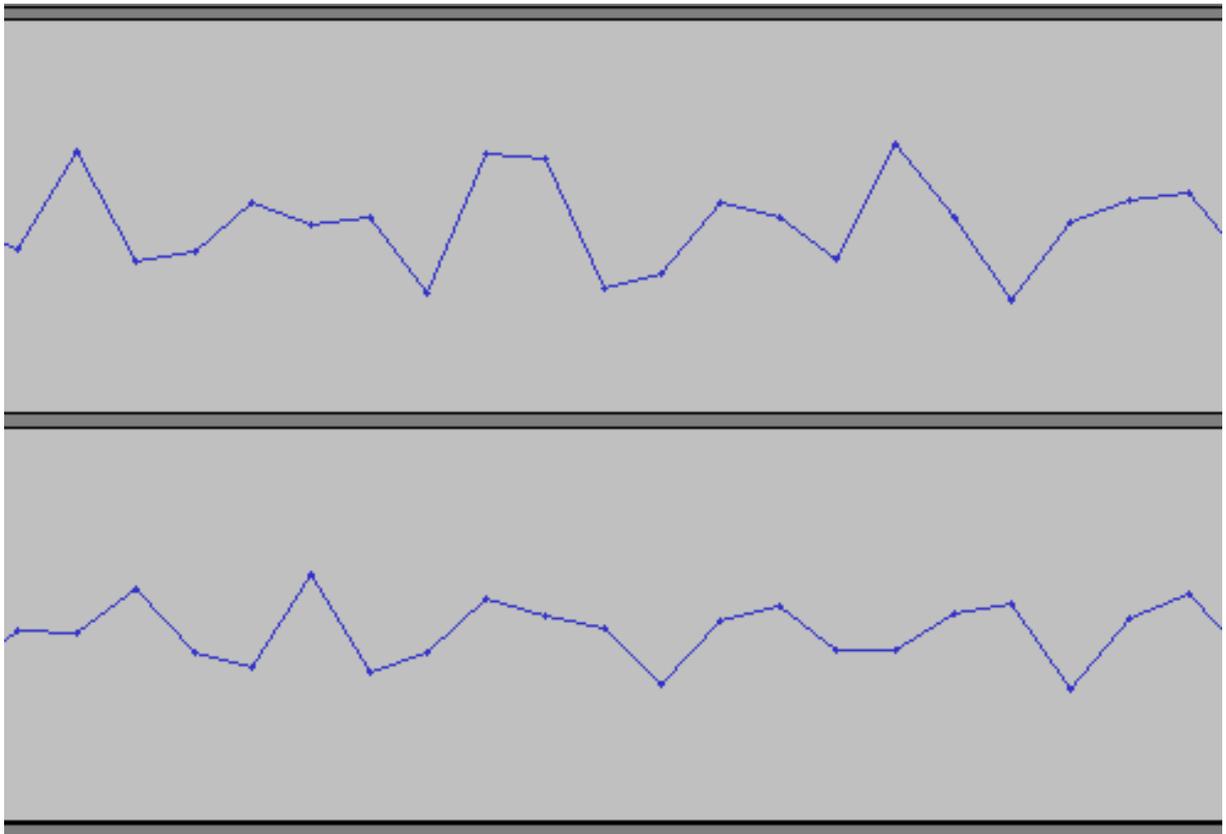
The resolution of the digital sound represents the number of samples comprised in a unit of time. Usually, the recordings are done at resolutions of 8, 16 or 32 bits. A 16 bits recording contains twice as much information as the 8 bits one (a double number of samples), having as a result a higher quality, but on the other hand it takes up more space.

The image below depicts the same musical fragment recorded at two different resolutions: 8 bits and 16 bits.
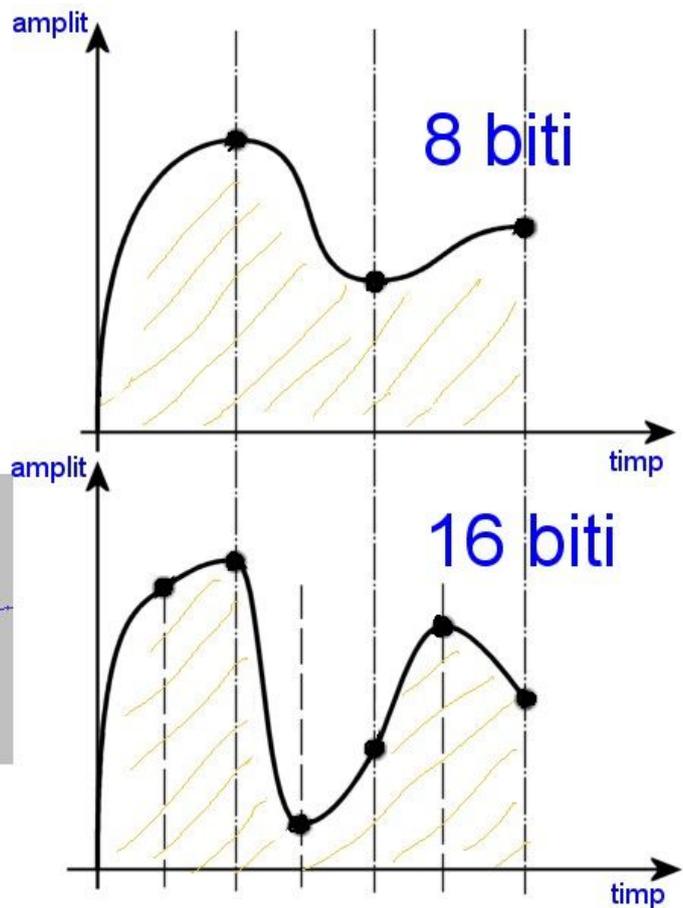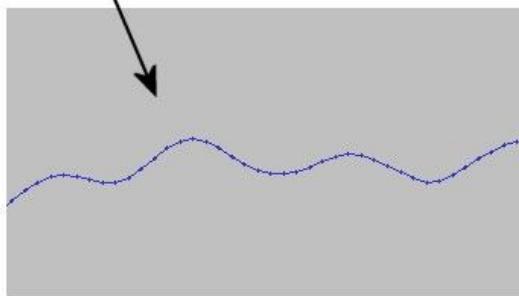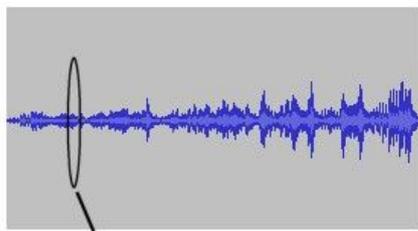
THE RESOLUTION OF THE DIGITAL SOUND

We can notice that the supplementary samples of the second chart (the 16 bit recording) offer a higher level of detail, leading to an increased quality of the recording. It becomes obvious, thus, that a 16 bit recording will much better reflect the reality of sounds than an 8 bits one.

The resolution of the digital sound grows in direct proportion to the quality and quantity (the occupied space) of the audio material.

Thus, the following parameters were set as standard for audio CDs: the maximal frequency threshold of 44 kHz, a resolution of 16 bits and two (stereo) channels, all of these ensuring a quantity that can play 74 minutes (650 Mb) or 80 minutes (700 Mb) of good quality music. At higher values of these parameters, the human ear no longer perceives differences and the occupied space grows, reducing the length of the recording time. At low values of these parameters, the human ear perceives the decreasing quality.

3. The number of channels

There are three main ways to classify sound according to the number of channels:

- mono (1 channel);

- stereo (2 channels);

- surround (5 channels).

A stereo recording will take up a double space as compared to a mono recording. Moreover, the stereo recording bestows a certain spatiality on the sound, while the panning one can direct on a left-right axis the sources of emission (instruments, voices, etc.). The result is an impression that they are placed in space (correlated with subtle changes of the sound level in order to simulate closeness and distance – an axis that lacks from the setting due to the limited number of only two channels).

Panning is a technique that confers, on each particular instrument, a different sound level on each channel. For instance, an instrument will be perceived as being in the centre if it has 50% level on the left channel and 50% on the right channel (for a stereo recording). The flutes are positioned on a Philharmonic's stage to the left of the listeners (and of the stage), so that the panning of the flutes has to be performed with a high level on the left channel (ex. 80%) and a lower level on the right channel.

In the case of moving sound sources (in films) the panning must be dynamic, by increasing/decreasing the sound level in time.

For the surround recordings we need to add the close-distance axis simulated by means of the back-front speakers. For example, the flutes, being placed farther away from the audience than the double basses, will have a high sound level on the front channels (speakers) and less on the back ones.

The image displays the position of the instruments in space (for simplification we chose only the two instrument groups) in order to calculate the panning.

FLAUTI

CONTRABASI

POZITIONAREA IN SPATIU A INTRUMENTELOR
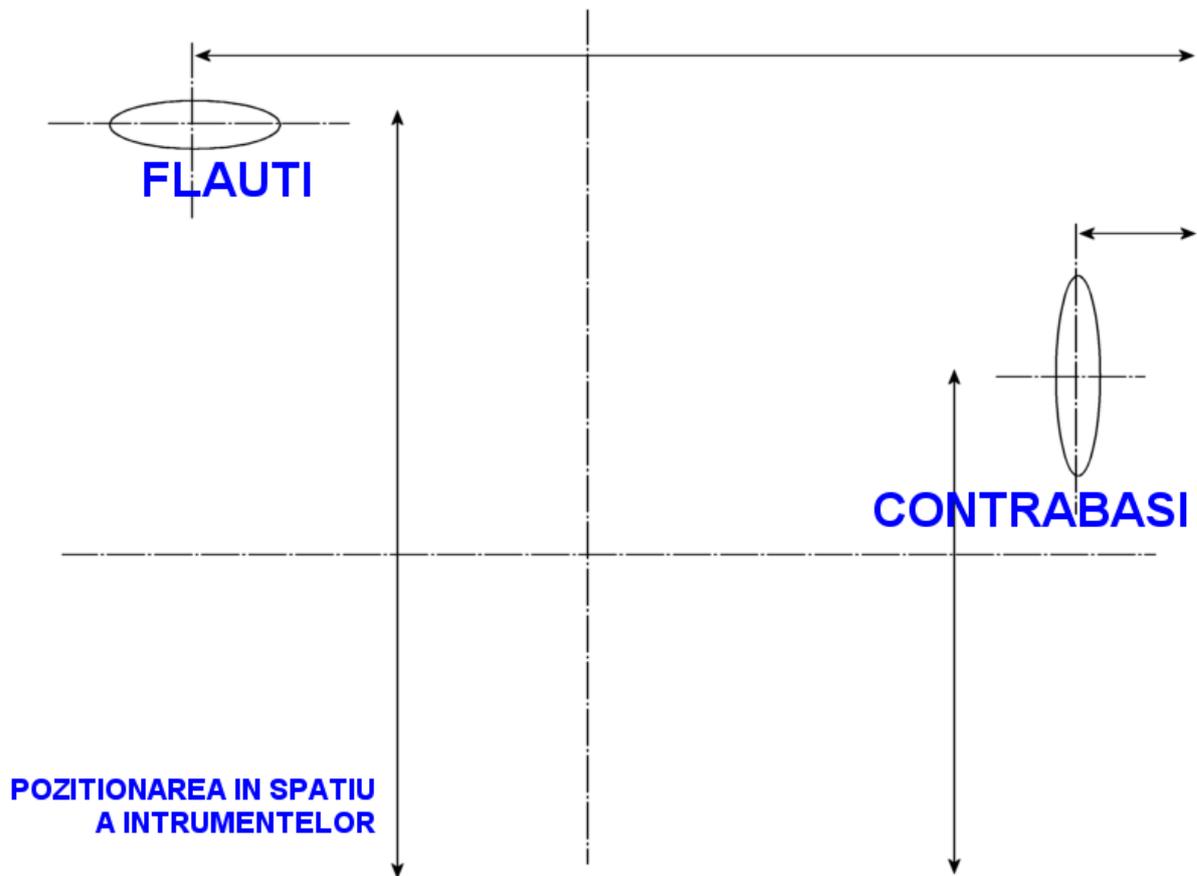
THE INSTRUMENTS' POSITION IN SPACE

In order to accomplish the panning and many other post-processing studio techniques that are necessary in order to edit a CD, the recording of a symphonic concert must be done with a large number of microphones, one for each instrumental or vocal group and mikes that will capture the general sound. Each of these mikes will represent a different (mono or stereo) track (channel), which will be processed, panned and, eventually, mixed.

Regarding the process of recording, we can have the following three **types of recording**:

- low signal to noise ratio recording;
- overdrive recording;
- correct recording.
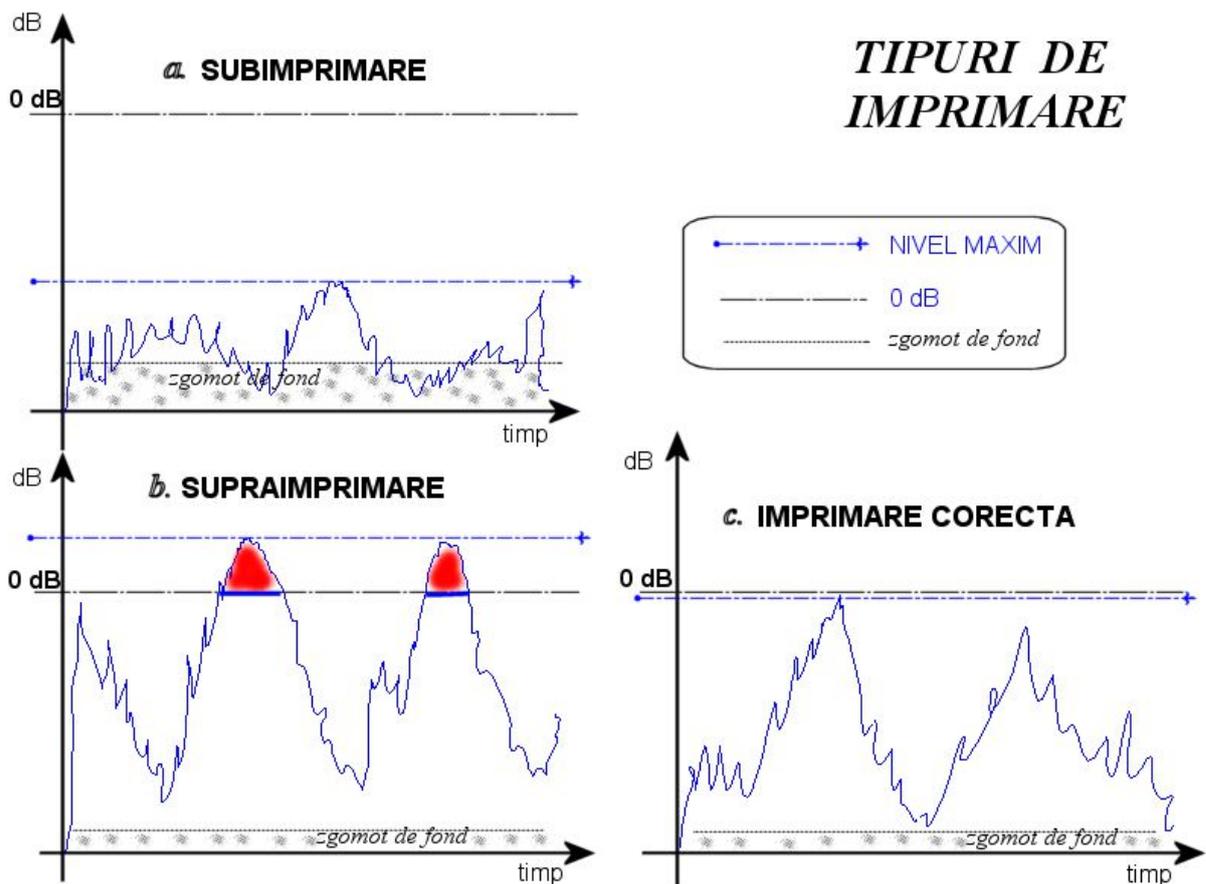
The **Low signal to noise ratio recording** is characterized by a weak signal, where the maximal amplitude is much below the ideal -3 db threshold. It can be reached only if the *Line-In* potentiometer of the sound board mixer was set at a too low level during the recording process. The great disadvantage of this type of recording is the very high level of background noise, the percentage of which will be too high in relation to the amplitude of the sound source.

In the case of **overdrive recording** the amplitude level sometimes exceeds the accepted 0 db threshold, so that the respective fragments are transformed into distortions and perceived by the human ear as unpleasant. The distortion area is actually an area of lost (or ruined) information. This type of recording occurs when the level of the *Line-In* potentiometer has been erroneously set too high.

A **correct recording** must get as close as possible, without exceeding it, to the ideal -3 db threshold. This will provide a high-quality sound level, a good high amplitude ratio and a very low background noise.

When doing a high-quality recording from an external source (radio, cassette player, tape recorder, turntable, mike, etc.) we must set, to start with, the correct level of the sound board potentiometer that controls the level of the recording. We must test the moment with the highest sound level, which must get as close as possible to the ideal threshold. This potentiometer can be *Line-In* for analogical sources, *Mic* for the mike, *Auxiliary* for devices inside the computer (ex. *tuner TV*), etc.

The maximal level can be followed visually on the *Monitoring panel*, described below (see *Audacity*), on the two red bars assigned to recording.

Most types of digital sound procession software are capable of calculating and rendering (some of them in real time) **spectrographic analyses of the sound signals.** The spectrographic analysis is very useful in the timbral analysis of the instruments and in the vocal analysis.

In the image below you can see the timbral spectrogram of a violin group. Even if the same sound is played by all the violins, we can notice a much richer timbrality that that of a single violin, due to the different construction of the (individual) violins involved.
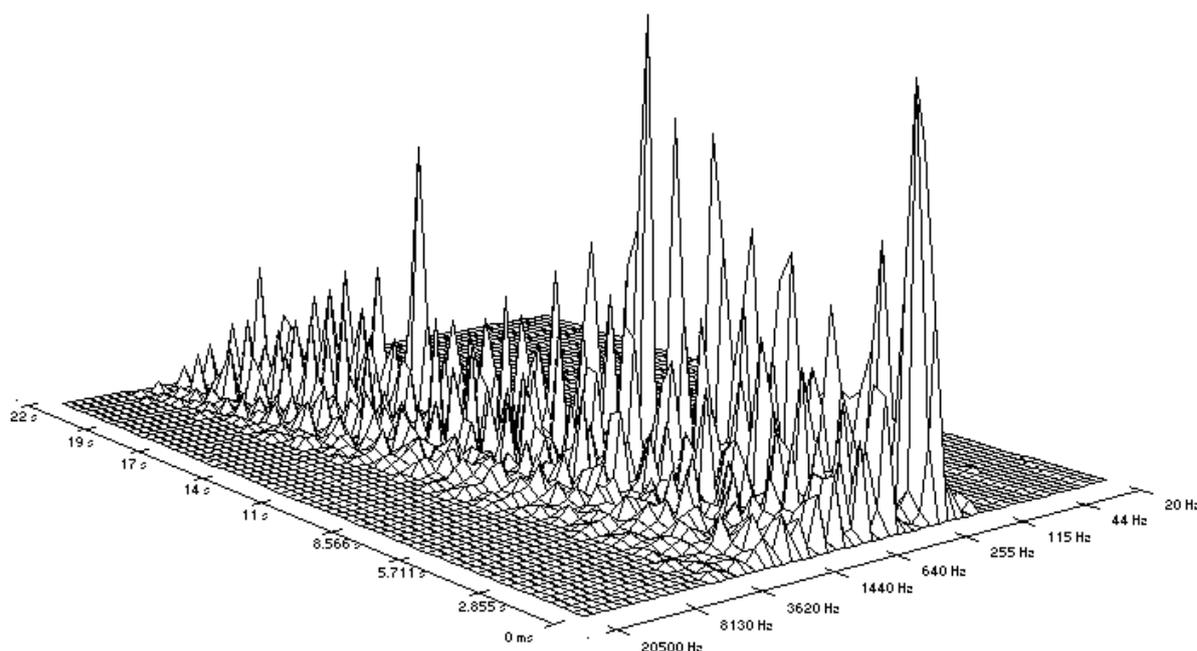
In order to analyse an instrument's quality, we can also use the spectrographic analysis. The following image displays the analysis of the same sound on two different violins (using the same type of strings).
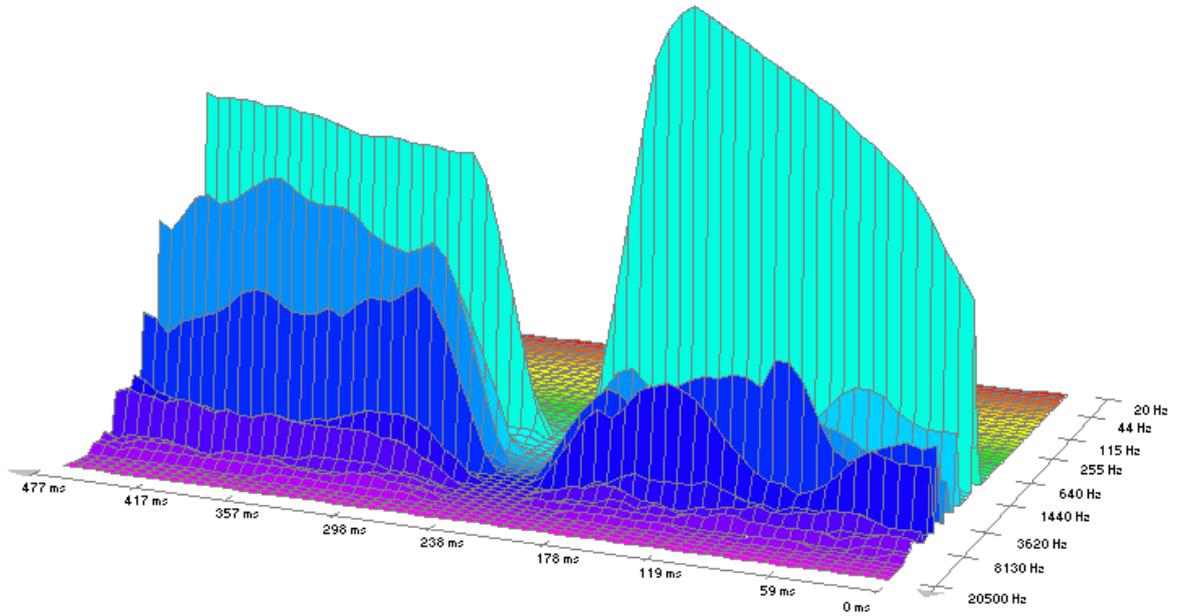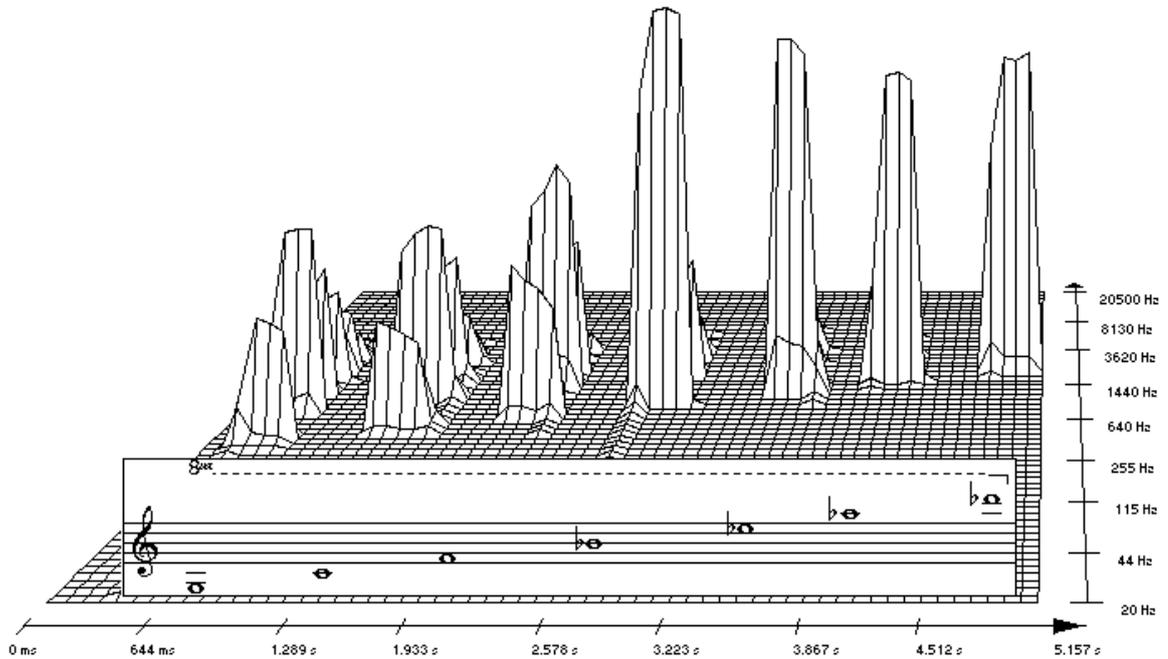
The image below displays the spectrogram of the same sound played on four different strings of the same violin.
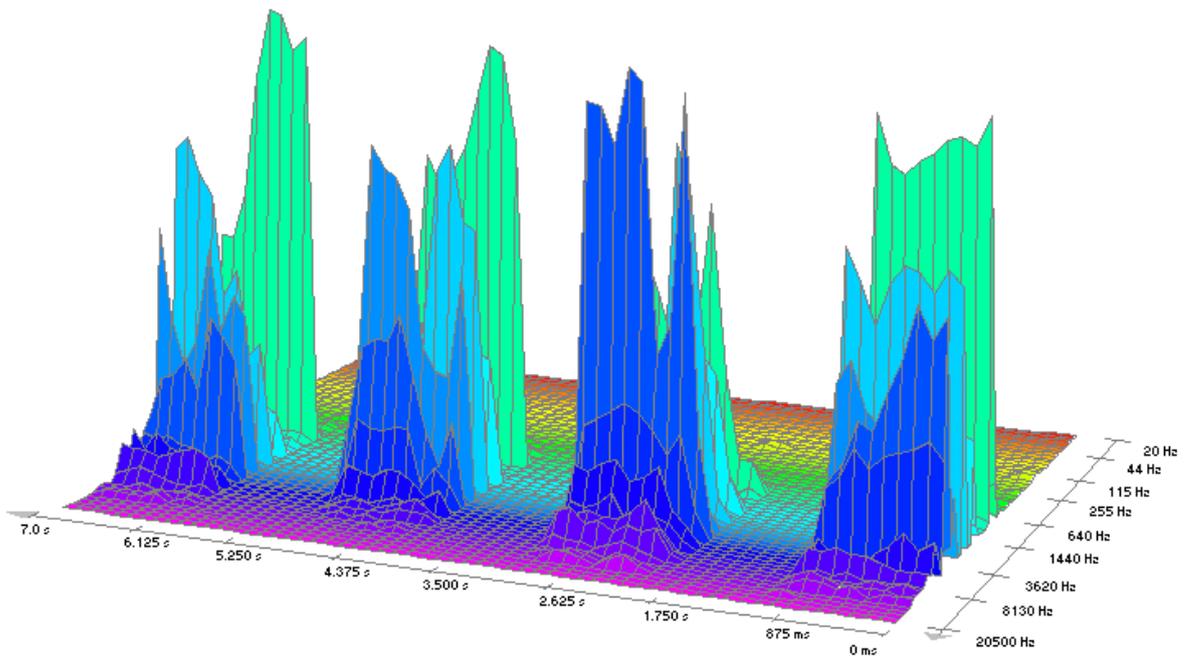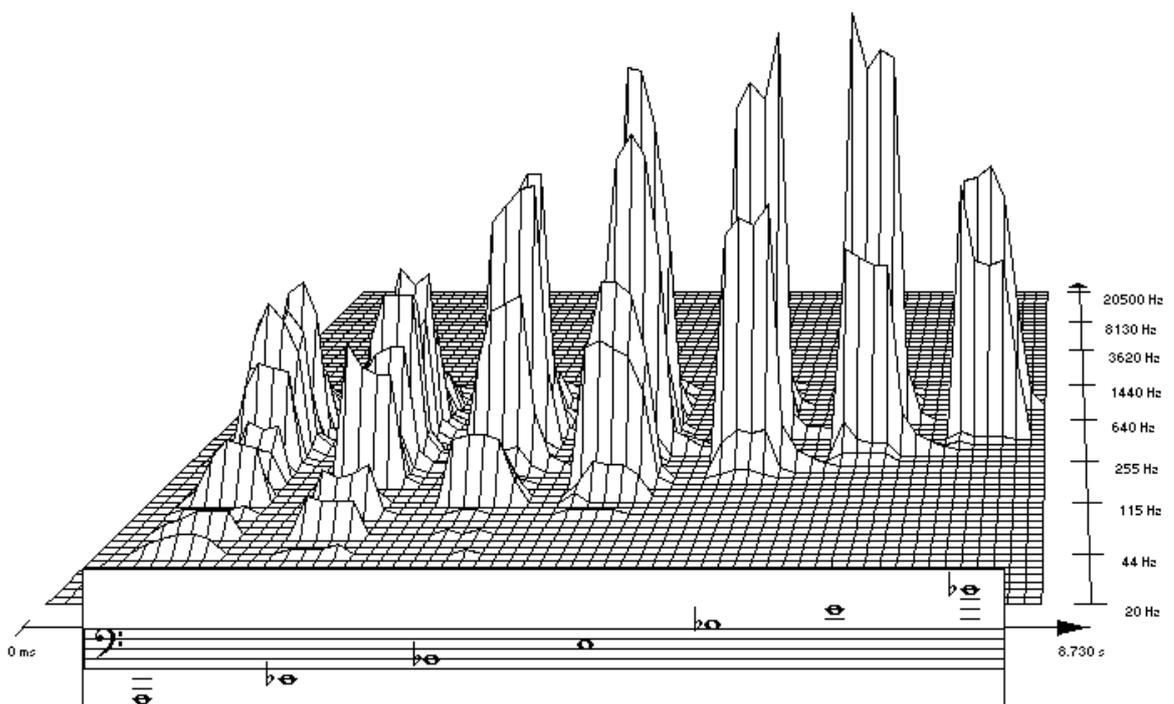

The spectrographic analyses can also be used in order to create virtual instruments. As a first step, we must record all the sounds of a real instrument. They are analysed, establishing the various sound registers of the real instrument.

The image below displays the flute's spectrogram along its entire range. We can notice the existence of two distinct registers of the flute, according to the richness of the harmonics and the amplitudes. We can also analyse the flute's timbrality.


As a comparison of timbral analysis, the image below shows the spectrogram of the tuba along the entire ambitus of the instrument. The tuba's timbral richness in the low register is easily discernible (even in the other registers as compared to the flute).

After performing the spectrographic analyses and deciding how the assign the real instruments according to their registers, we can go on to a draft of the virtual instruments. We must establish the number of timbral samples to be allotted to the virtual instrument. We cannot use one timbral sample for each semitone – to the detriment of a high-quality timbre – since the space allotted to the instrument would be too large. We only use a few representative samples which will be expanded over a wider sound interval, taking care no to lose the instrument's timbral identity. In the areas where the register changes we shall use several samples, which will enable the smoothest possible transition form one instrument to another.

The image below depicts the pattern of the register division of a virtual flute – according to the spectral analysis of the real flute from an image we provided above. We used six samples (s1 - s6) in order to cover the two registers (R1, R2), of which four will cover the buffer area "z" that marks the passage from one register to another.
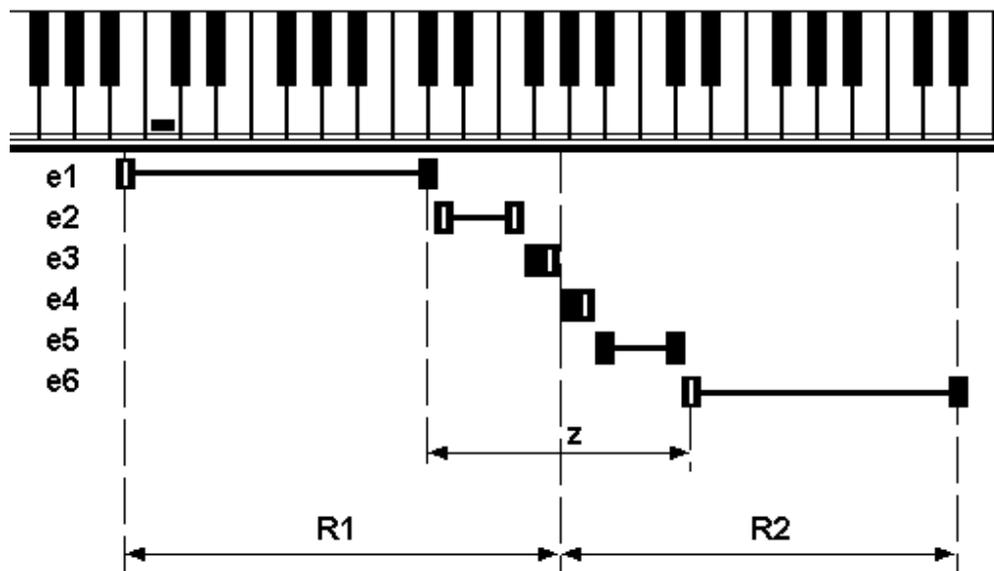
In order to perform the digital processing of these captured samples, we shall use a type of digital sound processing software such as *Audacity*.

In the end, the virtual instrument will be created with the help of some dedicated software, such as the *freeware* called *Vienna* in the image below. This application enables samples translations, frequency tuning, volume adjustment, as well as various specific effects, such as *Attack*, *Delay*, *Hold*, *Sustain*, etc.

Clarinet Bas
Alto Sax
Tenor Sax
Baritone Sax
Oboe
English Horn
Bassoon
Clarinet
Bb ff Clar G#2
Bb ff Clar G#2
Bb ff Clar D3
Bb ff Clar F3
Bb ff Clar B3
BB ff Clar D4
Bb ff Clar F4
Bb ff Clar B4
Bb ff Clar Gb6
Piccolo
Flute

Pitch	Value	Unit
Coarse Tune	0	semit...
Fine Tune	-28	cents
Scale Tune	100	cents

Effects	Value	Unit
Filter Q	0	dB
Filter Cutoff	999	Hz
Reverb	7	%
Chorus	0	%
Pan	0	%

Volume Envelope	Value	Unit
Delay	0.001	sec
Attack	0.001	sec
Hold	0.001	sec
Decay	2.718	sec
Sustain	97.7	dB
Release	0.36	sec
Initial Attenuation	10.5	dB
Keynum To Hold	1	X
Keynum To Decay	1	X

Modulation Envelope	Value	Unit
Delay	0.001	sec
Attack	0.001	sec
Hold	0.001	sec
Decay	3.279	sec
Sustain	78.8	%
Release	1.056	sec
To Pitch	0	cents
To Filter Cutoff	2587	cents
Keynum To Hold	1	X
Keynum To Decay	1	X

Modulation LFO	Value	Unit
Delay	0.537	sec
Frequency	5.379	Hz
To Pitch	0	cents
To Filter Cutoff	0	cents
To Volume	0	dB

Vibrato LFO	Value	Unit
Delay	0.001	sec
Frequency	5.379	Hz
To Pitch	0	cents

Ready

38 75 86

14

II. Audacity

Audacity[2] is currently one of the *open source - freeware* software types most widely used for digital sound processing (audio editing). It can be downloaded from: *http://audacity.sourceforge.net/download/*
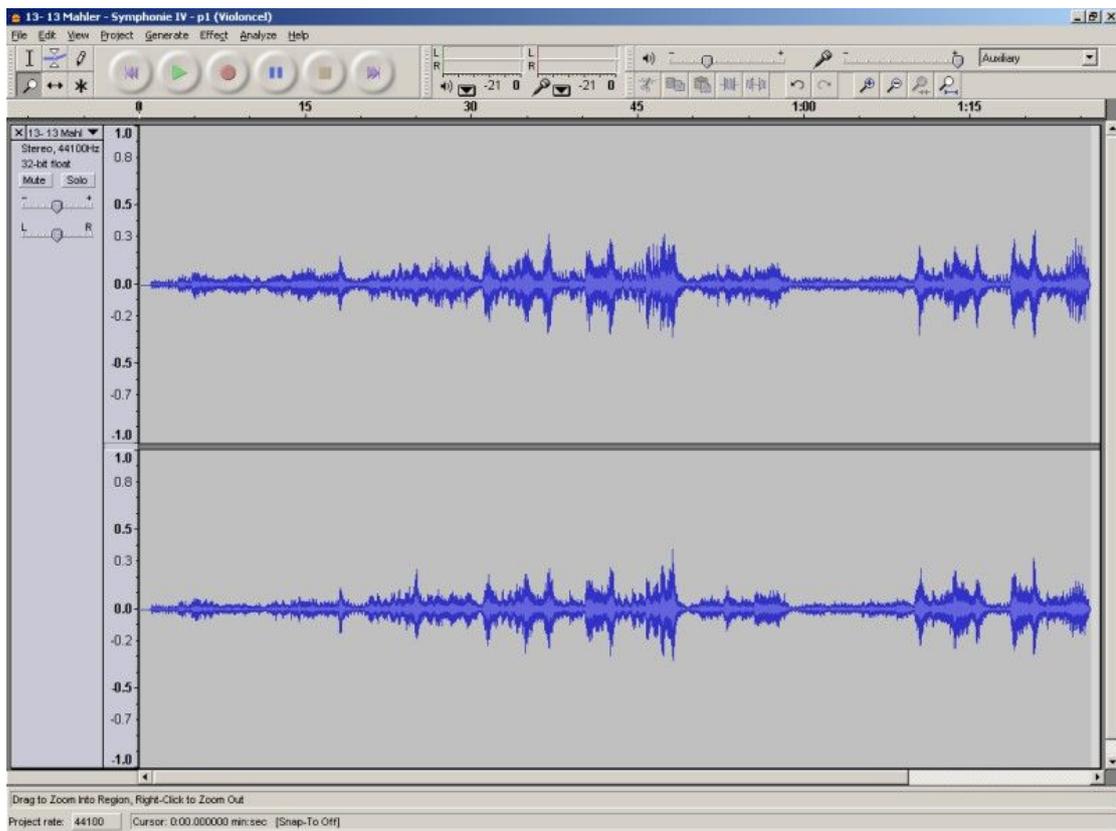
Some of the software's **characteristics** (abilities) are, according to Wikipedia[3]:

- it records and renders sounds;

- it imports and exports WAV, AIFF, MP3 files (the free LAME format), OGG Vorbis, FLAC (compression format without loss of quality) and others (WMA, AC3, etc.);

- editing with the help of the operations *Cut*, *Copy* or *Paste* with unlimited *Undo*;

- mixing on several channels (*multitrack*);

- it is compatible with a large number of effects and plug-ins;

---

[2] http://audacity.sourceforge.net/
[3] http://en.wikipedia.org/wiki/Audacity

- editing amplitude curves (*Envelope tool*);

- elimination of background noise;

- spectral audio analysis;

- multi-channel support with a sampling rate of up to 96 kHz and a 32 bits resolution;

- you can modify tempo without altering pitch (very useful, for instance, in video synchronization);

- pitch can be modified without changing the playback speed;

- the software contains most of the functions specific to any professional multi-track software;

- it uses a multitude of destructive and non-destructive effects;

- converting the analogue sound (cassettes, records, etc.) into digital format through diverse automatic procedures;

- it is compatible with the main computer systems: Windows, Linux or Mac OS X;

- *Audacity* can be used for post-processing as well (for various types of audio files, as, for instance, podcasts). The usual operations are: normalization, *trimming* or *fade-in*, *fade-out*.

As far as limitations (disadvantages) are concerned, we can mention the lack of controls for dynamic equalization, real-time effects and limitations in processing MIDI files.

The main **panels** are:

1. The control panel
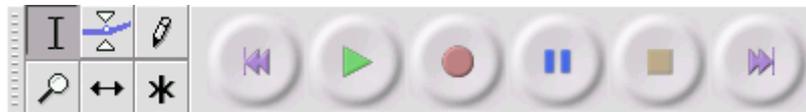
It is used for:

- selecting an audio fragment (*Selection tool*);

- changing volume in time (*Envelope tool*);

- modification (by zoom-in) of samples' amplitude (*samples*)(*Draw tool*);

- zoom-in and zoom-out (*Zoom tool*);

- shifting (moving) of audio fragments (tracks) in time (*Timeshift tool*);

- all of the above can be grouped in a multi-functional tool (*Multi tool*).
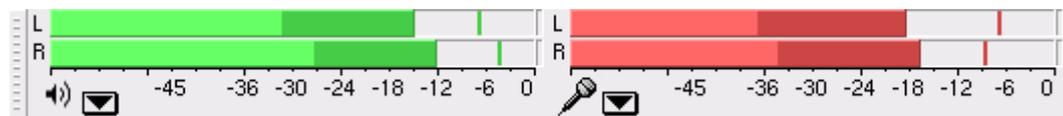
2. The mixing panel

It contains three volume controls, the input volume used for recording (*Record level meter*) and selection of the input source (microphone, line-in, audio CD, etc.).

3. The editing panel

It is used for simple editing-visualizing actions.

4. The monitoring panel

The panel is used to monitor the audio level while recording (red colour) and playing (green colour). The scale beneath the control bars represents the interval of peaks admissible for a recording without distortions (between minus-infinity and 0 db). In case it exceeds the 0 db threshold (correlated with the appearance of the distortions) the overdrive signals light up.

Normally, only the green monitoring bars are active. In order to activate, during pre-listening, the red bars used to monitor the recording, press *Monitor Input* (see image below).
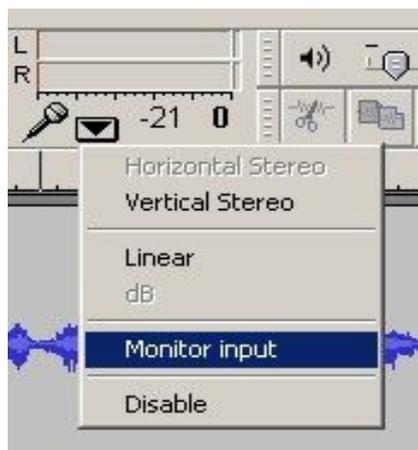
We can thus control the optimal audio level of the future recording (from the Line-In or Microphone potentiometer) in the pre-listening stage, so that no other distortions will emerge.

5. Audio track panel

It contains the audio recording, the volume and panning controls as well as a pop-up menu.
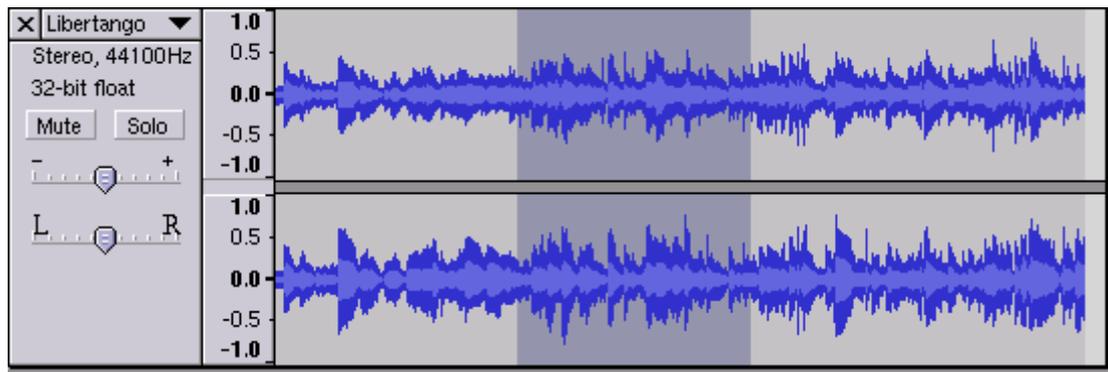
The pop-up menu (see the image below) enables us to choose the order (vertically) and to mix tracks, change resolution, visualise the spectrogram, etc.
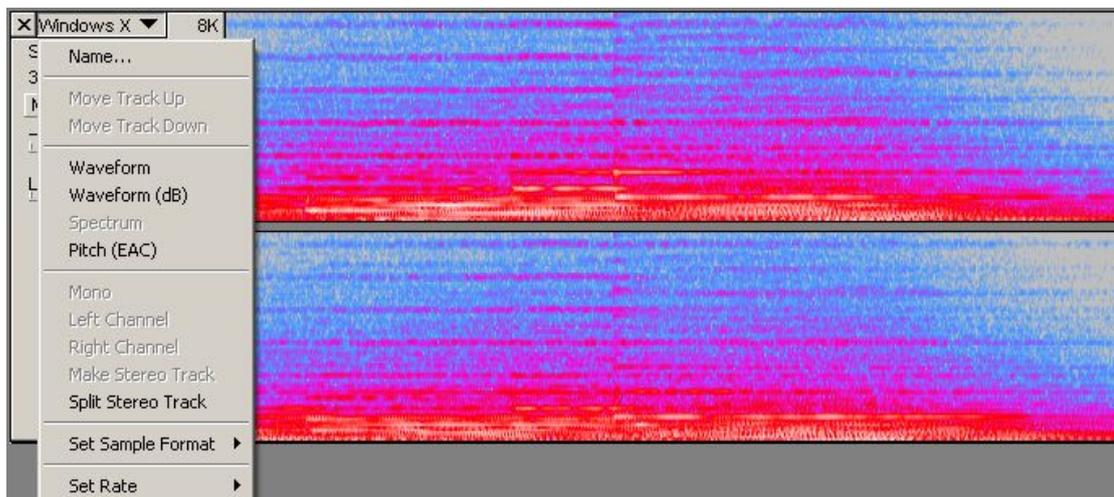
The image above shows the spectrogram of a recording.

6. Other available panels are the MIDI panel (being presently developed in the current software version) or the labelling panel in the image below.

Menu-specific operations

The **File menu** enables imports and exports from/into the best known formats (mentioned above), as well as specific saving and printing operations, etc.

We would like to mention a few commands specific to the **Edit menu**:

- Trim: keep only the selected fragment;
- Silence: replace selected fragment with silence;
- Split: move the selected region into another track (see the image below):
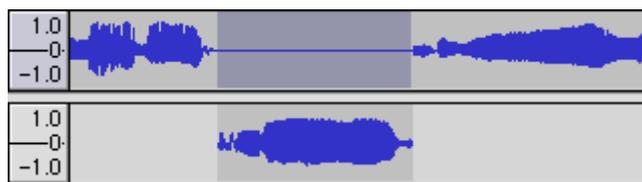
- Duplicate: copy the selected region into another track (see the image below):
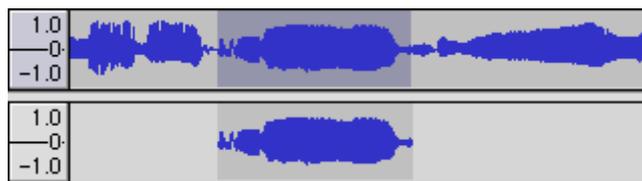
The menu also contains commands for the precise positioning of the cursor (in order to facilitate editing), various selection methods and commands specific to editing software such as *Find Zero Crossings*, etc.

The **View** menu comprises specific visualisation commands (*zoom*, *fit to*..., etc.) and detach commands used to position the detachable panels (described above) on the screen.

The **Project** menu includes the main import commands (audio, RAW, MIDI), operations of editing and arranging tracks.

The **Generate** menu enables the insertion of a silence fragment, the generation and insertion of a tone (the simple wave type), or the insertion of noise.

The **Analyze** menu is used to analyse the audio material (the energy on any frequency threshold). We can notice the quantity of high/middle/low frequencies.

The Effects menu

The effects can be classified into two categories: destructive and non-destructive effects. All are available only after the selection of a fragment or of the entire audio material.

The main effects available in Audacity are:

1. Amplify
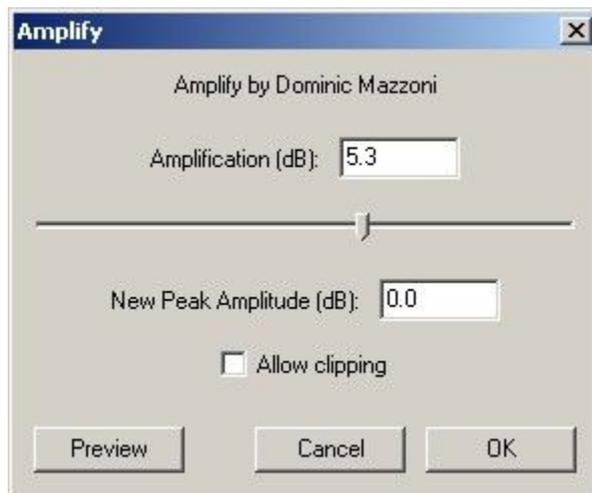
The command Amplify modifies the volume of the selected fragment (in decibels). In general it is used in order to amplify a low signal recording - preferably - up to the standard (distortion) limit of -3 db.

If we tick the option "Allow clipping", the amplifying will be done without taking into consideration the appearance of possible distortions in case we pass the 0 db limit.

2. Bass Boost
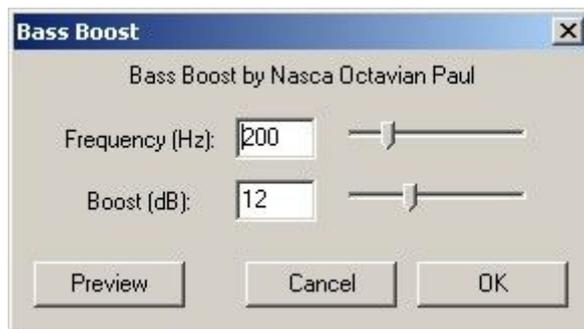
It amplifies the low frequencies of the selected fragment. It is used to correct the low bass level in recordings or on certain sequences.
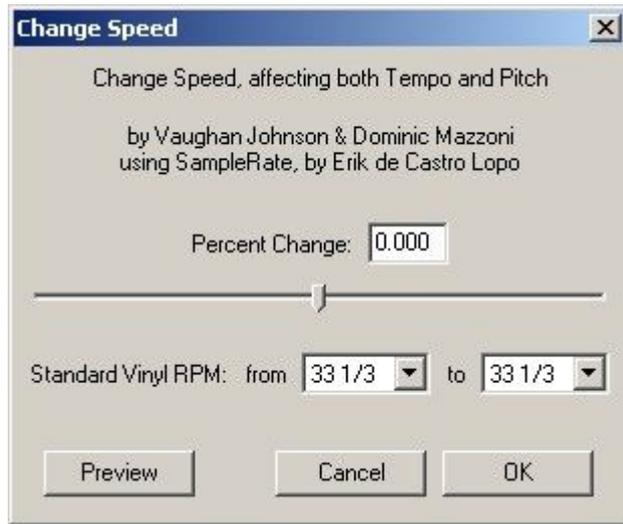
3. Change Pitch
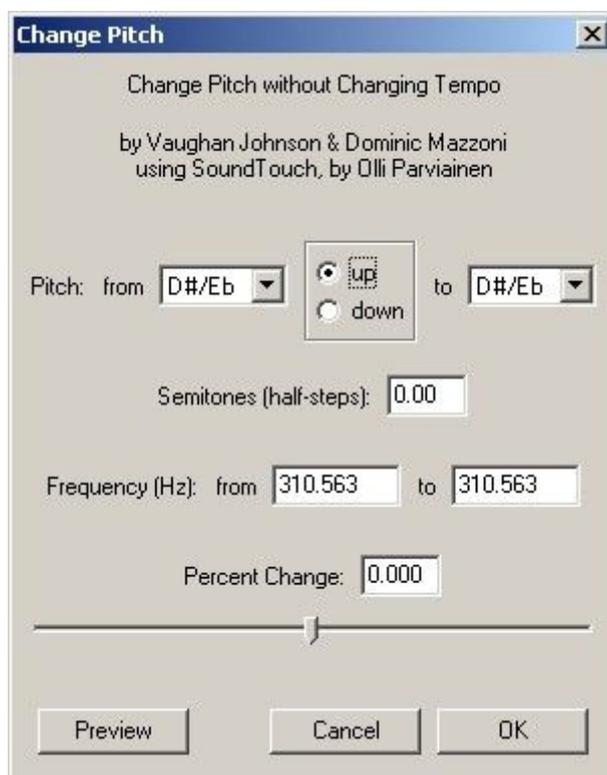
It changes the pitch (frequency) of the selected material without changing the tempo. It is used to transpose voice or an instrument in the desired interval. This transposition can be done in four ways:

- indicating a musical note;
- choosing the interval relying on semitones;
- choosing a certain frequency;
- changing the percentage.

4. Change Speed

It changes the playback speed (the tempo) together with the pitch (frequency) of the sounds. We use it when we want to change the duration of the recording. The disadvantage of this method consists in the alteration of the frequencies.

5. Change Tempo
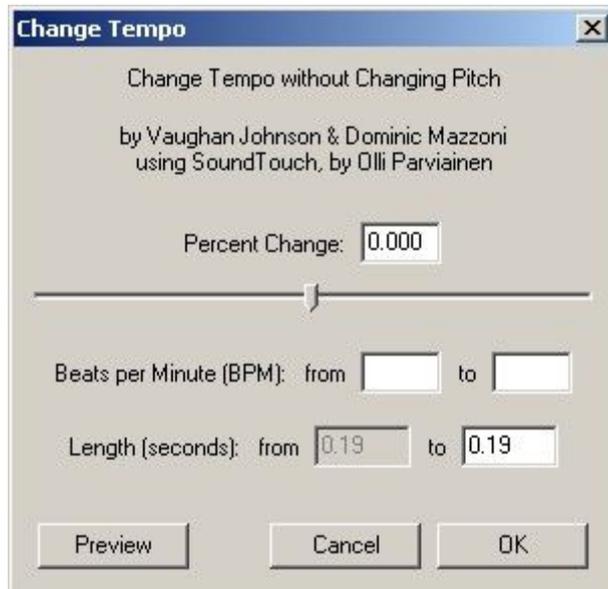
It changes the playback speed (the tempo) of the recording, without changing the pitch (frequency) of the sounds. We use it when we want to change the duration of the recording. This method is preferable in comparison to the previous one.

6. Click and Pop Removal
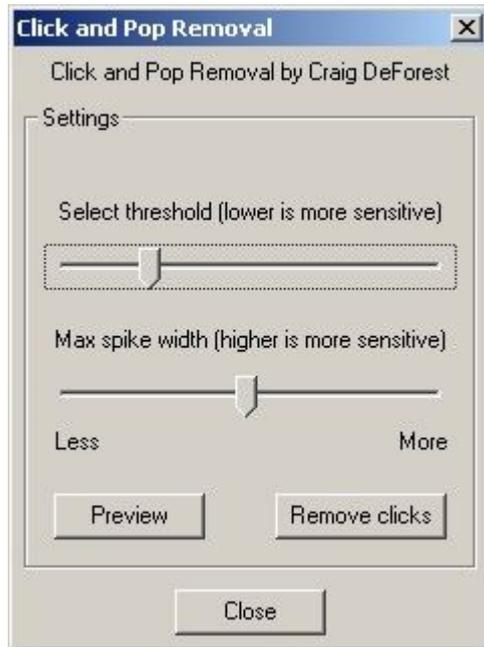
A plug-in included in Audacity which automatically eliminates pops or other unpleasant sound events occurring in this type of recordings. For instance, these pops can emerge after the digitalization of an audio material from an old vinyl record. They can also be eliminated manually, by using the zoom-in command for the desired fragment, selecting the event and removing it. When using this plug-in it is recommendable to work in several stages, using a low sensitivity, as an increase in sensitivity can alter the quality of the audio material.

7. (Dynamic Range) Compressor
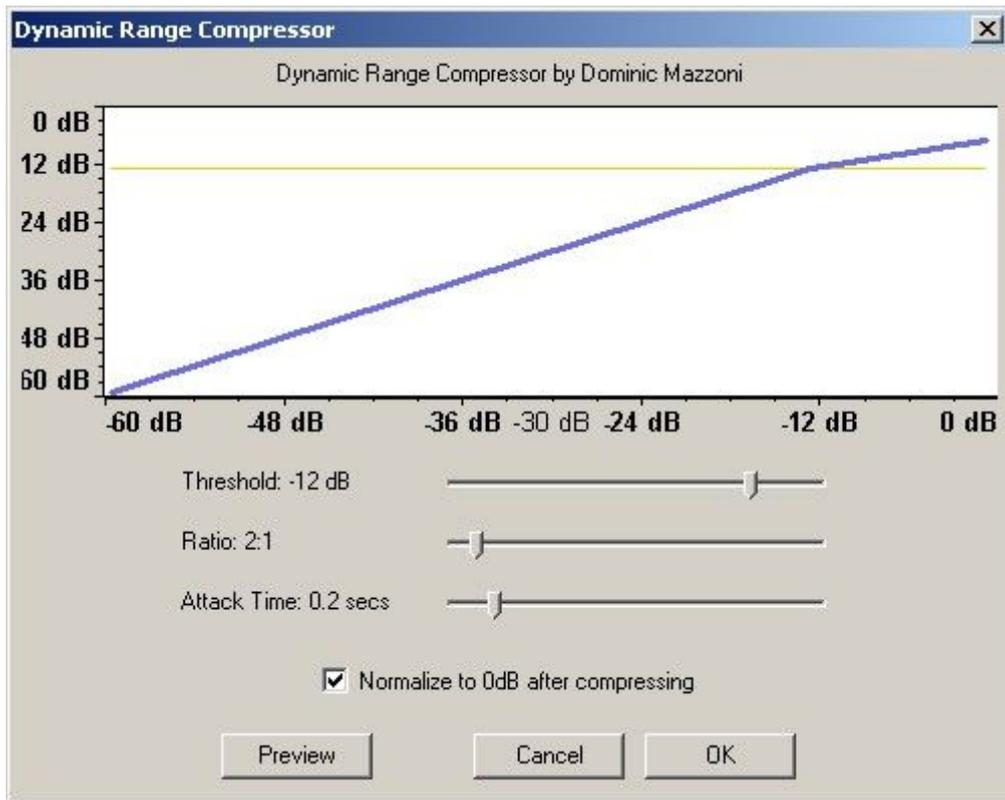
It is a dynamic range compression effect in which the fragments with higher amplitude will be reduced, while the ones with low amplitude remain the same, without getting diminished. It helps to decrease the difference between the sound levels of the material.
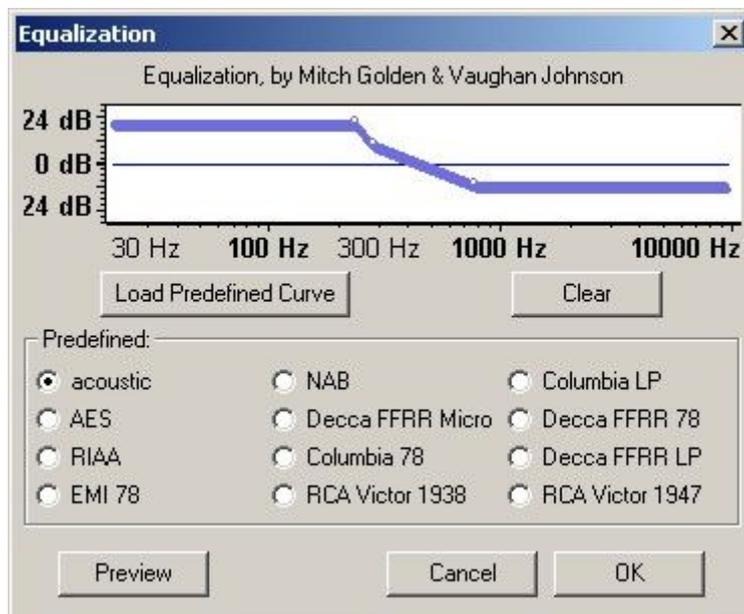
8. Echo
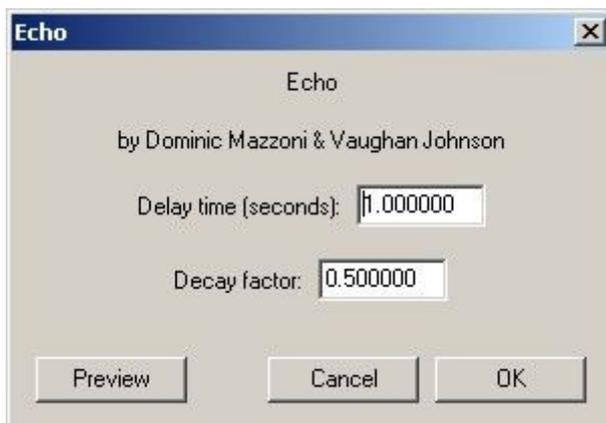
This effect repeats the selected fragment at the desired intervals with an amplitude reduction factor, having as a result a series of echoes. Due to the fact that the effect does not

change the length of the selection it is recommendable to add a fragment of silence (using the Generate menu) at the end, before activating the command.

9. Equalization

The command is used to mix the amplitude of frequencies, amplifying or reducing the audio level on various frequency thresholds. With the help of the mouse we can trace the curve we want or we can choose a predefined settings pattern.
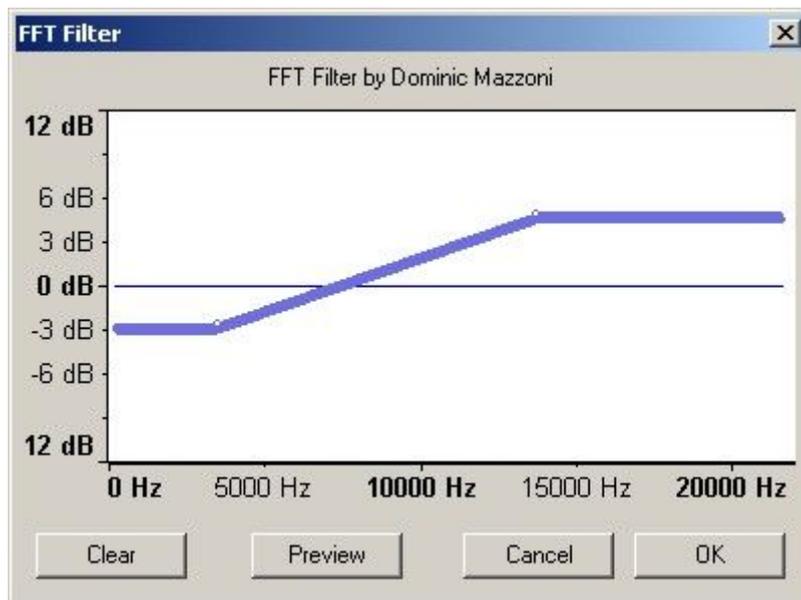
10. Fade In

The command is usually applied at the beginning of a song and provides a gradual (linear) fading of the sound level, from absolute zero to the reference value.

11. Fade Out

The command is usually applied at the end of a tune, as it provides a linear fading of the sound level until the value of absolute zero.
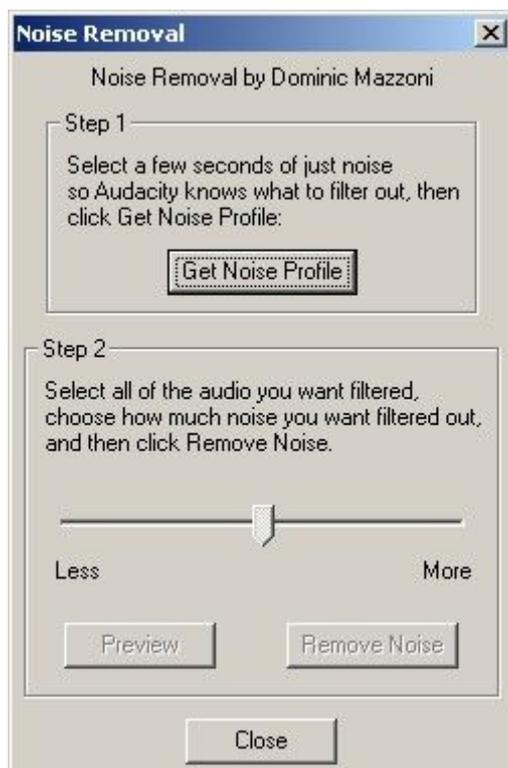
12. FFT Filter

It is used, just like Equalization, to amplify or reduce the level of frequencies.

13. Invert

It inverts the sound wave of the selected fragment on the vertical line.

14. Noise Removal

It is an effect that removes background noise and requires the following operations:

- select a "silence" fragment which contains only background noise (such fragments usually appear at the pause between two tracks, at their beginning or end);

- click *Get Noise Profile* to capture the "print" specific to background noise;

- select the entire audio material you want;

- adjust the potentiometer to remove more or less noise (by clicking *Preview* you can pre-visualise the audio result). Setting the potentiometer too high may have an unwanted effect of altering the audio material, eliminating other frequencies as well which do not belong to the background noise. That is why we recommend a moderate setting and, if necessary, repeating the process several times;

- click *Remove Noise* to start the final process.

15. Normalize
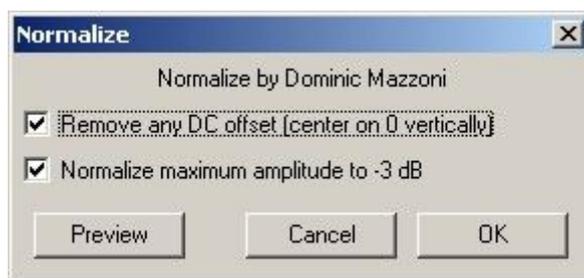
The command enables the adjustment (amplification) of the sound material's amplitude until the ideal value is reached in which the maximal level should be as close as possible to -3 db, without exceeding it. It is used to amplify low signal recordings (without any danger of overdrive recording with distortions). It is the operation that precedes the audio mixing process.
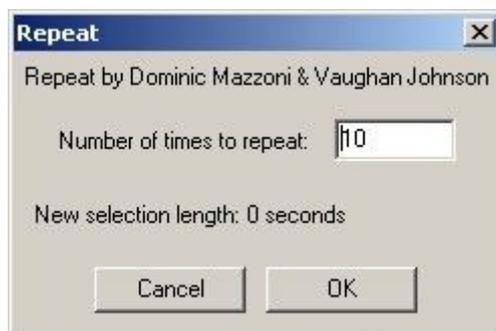
16. Nyquist Prompt

This is an interface for advanced users, especially for programmers, and adds new functions to Audacity.

17. Phaser

It is used in order to change the timbres of instruments (or of the human voice) in time, by using oscillators that change frequencies, and results in a change of the timbre colour in time.
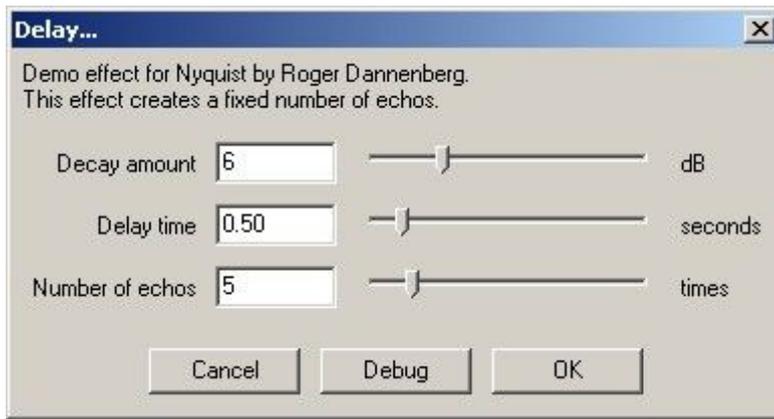
18. Repeat

It repeats the selected fragment several times. It is used in music (composition) to create repetitions to a preset tempo.
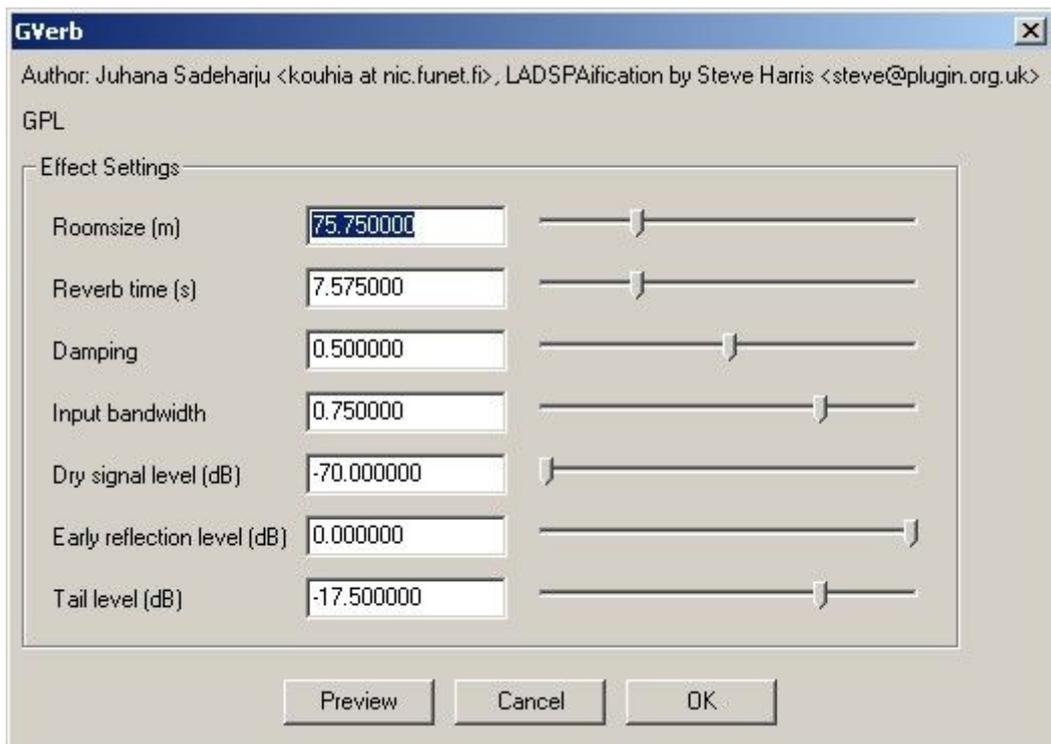
19. Reverse

It reverses the sound fragment in time, so that the end will become the beginning and vice versa.

20. Delay

It is used in order to improve the timbre of the instruments or of the human voice, by creating small echoes which overlap, enriching the timbres. *Delay* and *Chorus* are indispensable effects in any synthesizer's arsenal.

21. GVerb

It is a *Reverb* effect. It is useful for improving the timbre of an instrument (or of the human voice) and for producing a "hall effect", which creates the impression of reverberations produced in concert halls.

22. Hard Limiter
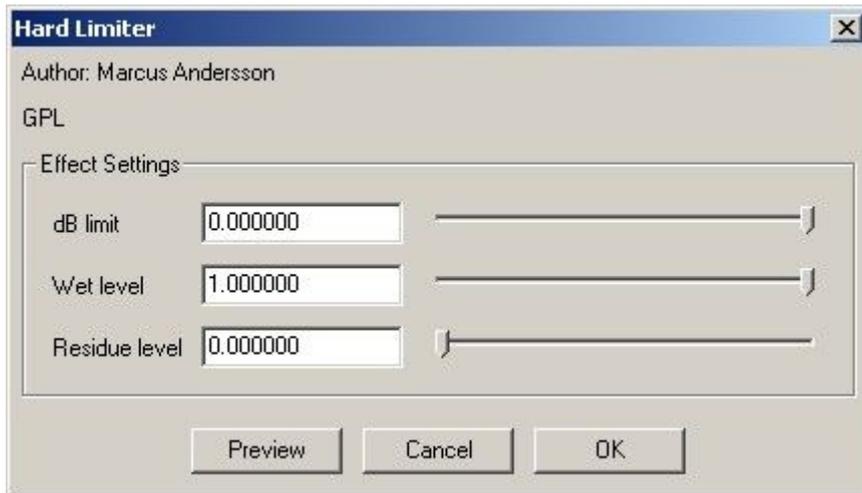
It resembles the effect *Dynamic Range Compressor*.

23. High Pass Filter

It cuts off (removes) all the frequencies under a chosen threshold. It is used to listen only to the high frequencies of an audio material or in order to remove low frequencies.

24. Low Pass Filter

It cuts off (removes) all the frequencies over a chosen threshold. It is used to listen only to the low frequencies of an audio material or in order to remove high frequencies.

25. Tremolo
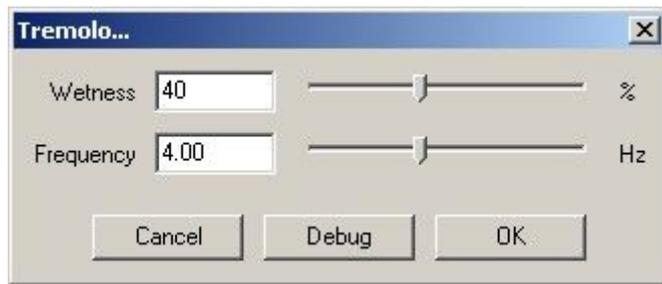
The Tremolo effect is used in music in order to alter the instruments' timbre in time (especially for electric guitars), or the human voice's timbre.

For **additional information**, very useful documentary sources can be found at:

- *http://audacity.sourceforge.net/manual-1.2/* - instruction manual;

- *http://wiki.audacityteam.org/wiki/Category:Tutorial* - tutorials;

- *http://audacity.sourceforge.net/help/faq* - FAQ;

- *http://wiki.audacityteam.org/wiki/Category:Tips* – useful examples.