

## SOUND EDITING ON THE SONOGRAM

Niels Bogaards

Analysis-Synthesis Team, IRCAM, Paris, France  
niels.bogaards@ircam.fr

### ABSTRACT

Sound editing applications commonly use a waveform display to graphically represent a sound signal. This representation not only conveys little information relevant to sound design, it also severely limits the ways in which the user can interact with the sound. By placing the sonogram at the center, AudioSculpt provides a more intuitive and insightful visualization, while at the same time allowing new ways of interacting with the sound's content, such as copy/paste of parts of the spectrum.

### 1. SOUND VISUALIZATION

To represent a sound file graphically, the sound must be visualized. The most common way is to plot the sound's samples on a time-amplitude plane: the ubiquitous waveform. Waveform display is logical from a technological point of view; direct and unambiguous link between sampled data and the graphical representation. While the waveform can convey an impression of loudness on global scale and hint at periodicity when zoomed in, it becomes increasingly difficult to interpret when the signal is more complex, and for typical real world signals provides hardly any useful information.

Much more insightful is the spectral representation of the sonogram. To obtain a sonogram, a Short Time Fourier Transform (STFT) is made, and its results plotted on a time-frequency plane, typically with a grayscale or color ramp function to visualize the energy intensity in a spectral bin. As longtime the calculation of sonograms was considered too costly, they are often an afterthought, and only provided to produce 'pretty pictures'. Nowadays, the processing speed of modern computers is such that sonograms can be calculated many times faster than real-time, and their use as sound visualization should be reconsidered.

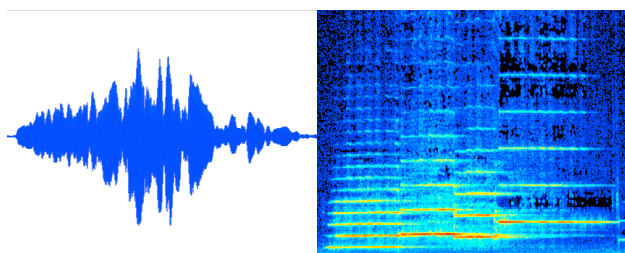


Figure 1. A 4 second flute phrase, in waveform and in a sonogram representation

To facilitate the study of musical signals, AudioSculpt's sonogram can also be displayed with a logarithmical frequency scale. Among the advantages of a logarithmical scale are the fact that equal musical intervals will have a regular visual distance and the relatively larger part of the image that is allocated to the perceptually more relevant registers.

### 1.1. Alternative sonogram representations

While the STFT is by far the most widely used method for the calculation of sonograms, other spectral visualisations exist, which highlight specific features of the sound. In AudioSculpt, alternative sonograms can be made using the LPC, Discrete Cepstrum or True Envelope spectral envelope estimators, or the Reassigned FFT method [1,2].

### 2. INTERACTION WITH THE SONOGRAM

While the use of sonograms is relatively common in scientific analysis of sound, for instance in speech research, few applications take advantage of the powerful editing capabilities a spectral representation can offer. In AudioSculpt, interaction takes place on the sonogram, where one can listen to selected zones and directly modify the spectral content.

### 2.1. Selection in the time-frequency plane

Selection of time-frequency zones on the sonogram is reminiscent of the 2-dimensional selections used in graphics applications. Typical are the rectangular and polygonal or lasso selection tools that can be used to draw an outline around the region that is to be selected.

More advanced is the Magic Wand tool, popularized by Adobe's Photoshop program: a click on a pixel with this tool selects all pixels with a similar color, as defined by a threshold of variance (tolerance). The Magic Wand allows for quick selection of a relevant cluster of pixels, and is very useful for selection on the sonogram as well. AudioSculpt's Magic Wand tool selects all contiguous bins on the sonogram within a range in dB around the clicked bin. This makes it easy to for instance select a frequency-modulated partial, or a cloud of noise.

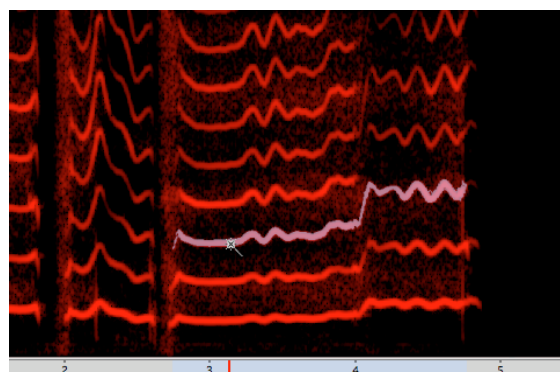


Figure 2. Selection of a single partial using the Magic Wand tool and a tolerance setting of 32 dB.

## 2.2. Sound Design on the Sonogram

An additional possibility of the selection of parts of a sound (*snippets*) on the sonogram is that it allows for the cut/copy of the selected regions and paste in a different location in time and frequency. This way, for instance a partial can be selected and moved to a different frequency, or shifted in time, or the call of a bird removed from a recording.

When moving snippets of sound around on in the linear time-frequency plane, it is important to realize that shifting in frequency can have dramatic effects on the timbre, as the relations between partials will change. Therefore, AudioSculpt has two ways of treating a copied snippet that's moved in frequency: frequency shift or transposition. Where the former moves the copied sound in a linear way, applying the same shift in hertz to all bins, the latter transposes by a factor in cents, relative to an anchor point.

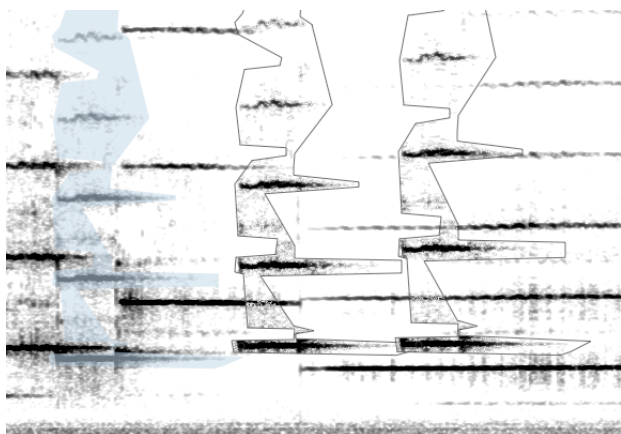


Figure 2. Copy and paste of a single note; the original (left) is pasted using frequency shift (middle) and transposition (right)

## 3. CONCLUSION

AudioSculpt's use of the sonogram as main sound representation provides a useful interface to both sound analysis and sound editing. The ability to copy/paste and drag and drop part of sound defined on the time-frequency plane opens up an innovative way of sound design.

AudioSculpt is part of IRCAM's Forum Studio package. The current version is a Universal Binary application, and works on MacOSX 10.4 and higher. [3]

## 4. REFERENCES

- [1] Roebel, A. and X. Rodet, "Efficient Spectral Envelope Estimation and its application to pitch shifting and envelope preservation," *DAFx*, Madrid, Spain, 2005.
- [2] Bogaards, N. and A. Roebel, "An interface for analysis-driven sound processing," *AES 119th Convention*. New York, USA, 2005
- [3] <http://forumnet.ircam.fr>