

Sound Analysis and Processing with AudioSculpt 2

Niels Bogaards, Axel Röbel and Xavier Rodet

Analysis-Synthesis Team, IRCAM
{bogaards, roebel, rodet}@ircam.fr

Abstract

AudioSculpt is an application for the musical analysis and processing of sound files. The program allows very detailed study of a sound's spectrum, waveform, fundamental frequency and partial contents. Multiple algorithms provide automatic segmentation of sounds. All analyses can be edited, stored and used to guide processing within the application, such as spectral filtering, time-stretching and noise removal, or serve as input for compositional environments. The current version is a complete revision, introducing many new features, a real-time mode, new analysis, processing and segmentation algorithms plus a significantly enhanced user interface. The introduction of SDIF as data format for analysis data and the ability to use long and multichannel sound files, take the application to a new level of usability.

1 Introduction

From its first conception in 1993, AudioSculpt's goal has been to provide musicians with a tool to analyze and process sound files in the frequency domain. Earlier research at the IRCAM had already suggested working in the frequency domain to be a musically intuitive method (Eckel, 1992), and AudioSculpt continues to elaborate and refine the ways in which to 'sculpt' a sound by directly editing its spectral contents. The current version brings the application to a new level of usability, with a drastically enhanced user interface, real-time processing and new algorithms for sound segmentation and transformation.

At the heart of the program lies a very flexible sonogram representation of the signal's spectrum. The sonogram view features sophisticated zooming and contrast adjustments controls, which provide a means for detailed and accurate signal inspection. Furthermore, the parameters for the STFT analysis that produces the sonogram can be freely modified.

Once the sonogram has been obtained, filters or treatments can be drawn directly onto it. AudioSculpt features a unique class of spectral gain filters, called surface filters, which allow the amplification or attenuation of an arbitrarily shaped time-frequency region in the sound.

Other treatments that can be applied to a section or the whole of the sound include bandpass/reject, transposition, timestretch, noise removal, spectral freeze, clipping, breakpoint and formant filtering.

AudioSculpt currently contains three automatic segmentation methods that produce series of markers. These markers, visible on the waveform as well as on the sonogram and the sequencer, can be edited by hand and used in AudioSculpt itself, for instance to align treatments to a certain time region of the sound, or exported to other applications, such as the OpenMusic composition environment (Agon, Stroppa and Assayag, 2000) or Max/MSP (Wright, et al., 1999).

For both analysis and synthesis/processing, AudioSculpt uses the SuperVP kernel. SuperVP is the IRCAM tool for time/frequency signal analysis and processing. It has been under continuous development since 1989 (Depalle and Poirrot, 1991). SuperVP incorporates algorithms for spectral envelope estimation (LPC and discrete cepstrum), transient detection, cross synthesis, mixing, time/frequency domain filtering, noise removal, phase vocoder based algorithms for time variant time stretching including transient preservation, time variant transposition and many others.

2 Application Design

Oriented towards musicians, AudioSculpt strives to provide an ergonomical and friendly interface, while preserving the ability to focus on minute details of the sound and the parameter settings for the often complex analysis and processing algorithms. With the transition to Mac OSX, existing interaction elements have been reevaluated and adapted to take advantage of new interfacing paradigms. To keep the focus on the creative process at all times, temporary files and smart presets are used as much as possible, minimizing the interruption of the workflow by dialogs.

While AudioSculpt is a Macintosh-only application, the processing kernels it uses are developed as cross-platform command line-based tools. With command line functionality readily available on OSX, the same kernel can be used for work within AudioSculpt as for command line use from the Macintosh's Terminal application. This separation between processing kernel and user interface application results in an efficient development cycle, where algorithms are designed and tested on UNIX and Linux workstations, using tools like Matlab and Xspect (Rodet, François and Levy, 1996), and new versions of the kernel can be directly and transparently used by AudioSculpt.

2.1 Overview of the interface

AudioSculpt's document window, as shown in fig. 1, consists of various views on the sound, plus a multitrack sequencer which holds graphical objects representing the treatments created to process the sound with. From top to bottom:

- Overview: shows a zoom invariant overview of the sum of all channels
- Waveform: fully zoomable view of the sound channel's waveform(s)
- Sonogram: a view of the STFT of the sound signal using log amplitude representation in dB. Contrast and color thresholds can be dynamically adjusted with sliders
- Spectrum: A log amplitude spectrum in dB of the frame currently played or selected with the diapason tool (see paragraph 2.2). As here frequency alignment is required instead of time alignment, the panel is placed next to the sonogram
- Sequencer: a multi-track view where treatments are organized

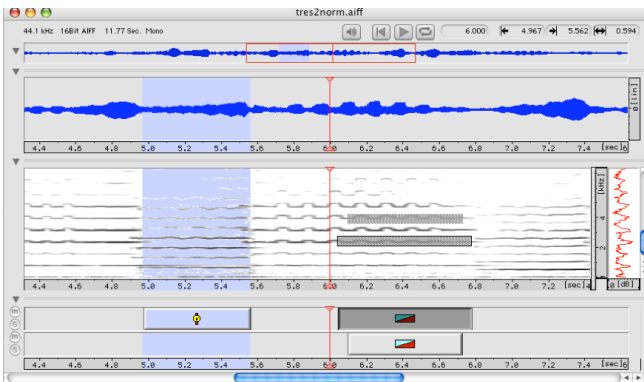


Fig 1. The document window

All treatments and markers can be edited in the Inspector window, as well as in their respective editors. The Inspector makes it particularly easy to copy and paste exact values to and from parameter values.

2.2 Diapason

The diapason is a special tool, designed to interact directly with the sound's spectrum. It can serve either to auditorily inspect a sound, or to compare two spectral frames.

Pointed at a certain time-frequency spot in the sonogram, the diapason will play the sinusoid associated with the pointed frequency, with an amplitude corresponding to that frequency's presence in the sound at the chosen time. Moving the diapason in a horizontal way a single frequency band's time envelope can be followed, moving vertically sweeps through the spectral contents at a single time frame.

When the diapason is pointed, the spectral contents for the selected time frame are displayed in the single-frame spectrum view. By moving the playback cursor to a second reference frame the two spectra can be compared.

2.3 SDIF

As AudioSculpt is meant to analyze and synthesize individual sounds, rather than produce entire compositions, communication and exchange with other applications is very important. Therefore, SDIF (Schwarz and Wright, 2000) was chosen as format for all analysis data read and produced by AudioSculpt2 and SuperVP, and also as the storage format for AudioSculpt's tracks and treatments. The SDIF format yields significant advantages over other existing formats. SDIF is extensible, which means new values, types or text information can be added without affecting the portability of the file and its contents. Moreover, SDIF is a binary format, so it is efficient, compact and precise, even for large data sets, like STFT signal representation.

3 Analysis

Sounds can be analyzed in various ways using AudioSculpt. While not aspiring to be a scientific analysis tool, the program permits very accurate and detailed inspection of a sound and its properties.

All analyses can be exported and imported, either individually or combined using the SDIF format.

3.1 Sonogram

Most work starts out with creating a sonogram to get an overview of the sound's spectral content. The sonogram is based on the STFT representation of the sound. As there is no one setting that gives the ideal analysis for all sounds and all uses, AudioSculpt allows detailed manipulation of the STFT's parameters, such as the size and shape of the analysis window, the size of the step with which the STFT slides over the signal and the resolution of the FFT. The resulting sonogram permits the visual inspection of the Fourier representation, and the selection of settings that provide an optimal match between the sound's characteristics and the desired analysis or transformation. To facilitate this process, presets are available with typical settings, and the user can store custom presets.

3.2 Fundamental Frequency Estimation

Fundamental Frequency or F0 analysis estimates the fundamental frequency of sounds supposing a harmonic spectrum. The result is a breakpoint function of frequency over time, which can be laid over the sonogram, as in Fig. 2, to be inspected and edited. This fundamental frequency can serve as a guide for subsequent treatments, or be exported to other applications, for instance to serve as compositional material.

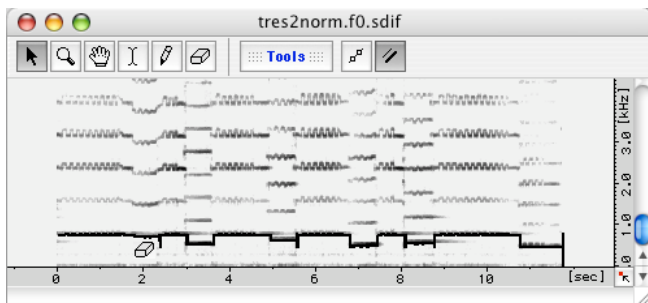


Fig.2 Editing the fundamental frequency

3.3 Partial Trajectory Analysis

In Partial Trajectory Analysis multiple breakpoint functions are created for sinusoidal components in the sound. Like with the fundamental frequency estimation, the resulting trajectories can be overlaid on the sonogram, edited and exported. For this analysis, instead of SuperVP, the Pm2 Kernel is used, which implements a standard additive analysis that can either work in harmonic or inharmonic mode.

3.4 Segmentation and Markers

There are three automatic segmentation methods available in AudioSculpt, one based on the transient detection algorithm that is also used in the time stretch's transient preservation mode, and two based on the difference in spectral flow between FFT frames. An energy variation threshold can be interactively adjusted to filter less significant transients. The transitions produce markers in AudioSculpt (see Fig.3), which can be edited and used in various ways. In a purely analytical use, the markers can be adjusted, edited and consequently exported for use in other applications. In AudioSculpt itself, the markers serve as alignment boundaries for treatments, for instance to have a filter work just up until a certain transition, or to start a spectral freeze just after a transient. Using markers in this way provides a much more significant grid than a one solely based on tempo or amplitude/zero-crossing methods.

4 Transformations

AudioSculpt features a wealth of treatments that can be applied to transform a sound, all of which operate in the frequency domain.

4.1 Types of treatments

Filters. Several types of filters are provided to change the gain of certain frequency ranges in the signal, their difference laying in the way the user creates them.

- the *surface filter* is a rectangular or polygon (not necessarily convex) region, drawn on the sonogram. The resulting region is assigned a value of amplification or attenuation in decibel. In Fig.5 polygon filters are used to attenuate unwanted frequencies from a sound.

- the *pencil filter* allows drawing arbitrary shapes onto the sonogram, where the pencil's tip can be assigned a width in seconds and a height in Hz. As with the surface filter, a gain factor in decibel determines the effect of the filter.
- the *band filter* can function as a brickwall bandpass or bandreject, with multiple bands and time-varying band edges.
- the *breakpoint filter* applies a static frequency/amplification breakpoint function, comparable in effect to a graphic equalizer.
- the *formant filter*, creates a series of second order resonant band filters at a harmonic frequency spacing.

Transposition. AudioSculpt features constant and dynamic transposition, with or without time-correction. For the dynamic version, a breakpoint function determines the transposition factor over time.

Timestretch. Time stretching in AudioSculpt uses the phase vocoder algorithm including a recent extension for transient preservation (Roebel 2003). Time stretching with transient preservation achieves surprisingly high quality of the transformed signals as long as the attack transients are not too dense in time to be individually detected (see Fig.4).

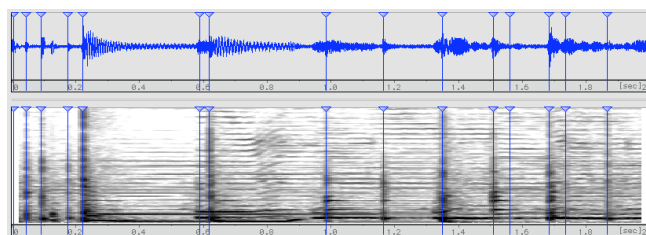


Fig 3. Markers pinpoint detected transients

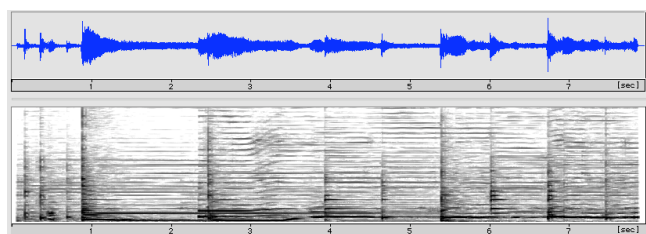


Fig 4. The same sound, stretched 4 times, using transient preservation

Freeze. Spectral Freezing is a technique, comparable to time stretching, where a single frame of the spectrum is 'frozen' for an amount of time.

Clipping. The clipping filter performs a non-linear dynamic stretch, whereby frequency amplitudes under a threshold are set to zero, the ones above an upper limit set to maximum and the range between the two thresholds rescaled to fit the dynamic range.

4.2 Applying treatments: the sequencer

Treatments can be applied to a time interval in the sound, or to the entire sound. All treatments create a

graphical object in both the sonogram view as well as an object in the treatment sequencer. Before processing the objects can be moved, modified, copied, pasted or replicated in time or frequency.

The concept of the sequencer is useful, as it provides a means to focus on certain treatments and try them out, while other treatments are muted, which means that these are currently not applied. The consequent final combined processing will thus be at maximum fidelity.

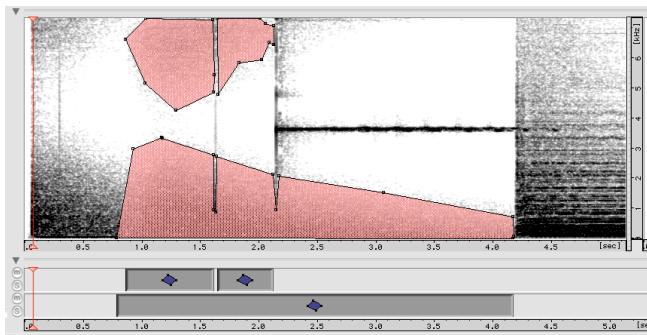


Fig 5. Surface filters to attenuate unwanted frequencies

5 Recent developments

Both the kernels and the AudioSculpt application are under constant development, and features and algorithms are continually added and refined. Below is an overview over some of the recent changes.

5.1 Multichannel sounds

A recent addition to AudioSculpt is its ability to analyze and process multichannel sounds. Among the possibilities to be explored are detailed comparison of the spectral content of spatialized sounds, use of transient markers on different channels, masking and unmasking.

5.2 Real-time processing

An exciting new feature, which has become available with the continuing increase in processing speed of personal computers, is the ability to listen to the processed result in real-time before creating a new soundfile. This makes it significantly faster and easier to experiment with settings for treatments, combinations of different treatments and compare their effects using the sequencer's track mute and solo capabilities.

5.3 Unlimited soundfile length

Previous versions of AudioSculpt suffered from a limitation in length of the sounds that could be used. This limitation no longer exists, which facilitates for instance the use of the noise removal algorithm on a one-hour recording, or the subtle timestretch or pitch transposition of a song.

5.4 Audio quality

Where the Macintosh's SoundManager limited samplersates to 48 kHz, the new CoreAudio framework provided the possibility for high definition audio, and smooth integration of third party hardware devices for sound playback.

6 Conclusion

Over the years, AudioSculpt has proven to be an artistically relevant concept, serving as a musician's interface for the IRCAM's research into the analysis and synthesis of sound.

Current developments, like the advent of OSX for high quality sound and graphics rendering, as well as seamless integration with UNIX command line tools and the processing power available on new consumer level computers, inspired a version of AudioSculpt that invites more than ever the creative experimentation with sounds and their spectra. Meanwhile, new features continue to be added, and work is taking place on for instance the resynthesis of sound from imported sonograms and images, gradient slope filters, frequency shifting based on fundamental frequency analysis and a harmonic pencil tool to suppress or emphasize partials.

AudioSculpt is available for members of IRCAM's Forum (<http://forumnet.ircam.fr>), as part of the Analysis-Synthesis Tools.

References

- Eckel, G. (1992). "Manipulation of Sound Signals Based on Graphical Representation", In *Proceedings of the 1992 International Workshop on Models and Representations of Musical Signals*, Capri, September 1992.
- Agon C., M. Stroppa M. and G. Assayag (2000), "High Level Musical Control of Sound Synthesis in OpenMusic", In *Proceedings of the ICMC*, pp. 332 - 335.
- Wright, M, S. Khoury, R. Wang, D. Zicarelli, R. Dudas (1999), "Supporting the Sound Description Interchange Format in the Max/MSP Environment" In *Proceedings of the International Computer Music Conference*, pp. 182-185
- Depalle, P. and G. Poirrot (1991): "SVP: A modular system for analysis, processing and synthesis of sound signals" In *Proceedings of the International Computer Music Conference*
- Schwarz, D., and M. Wright (2000), "Extensions and Applications of the SDIF Sound Description Interchange Format". In *Proceedings of the International Computer Music Conference*.
- Rodet, X., D. François and G. Levy (1996), "Xspect: a New Motif Signal Visualisation, Analysis and Editing Program." In *Proceedings of the International Computer Music Conference*,
- Roebel, A. (2003). "Transient detection and preservation in the phase vocoder." In *Proceedings of the ICMC*, pp. 247-250.