Audio Engineering Society

Convention Paper

Presented at the 119th Convention
2005 October 7–10 New York, New York USA

# An interface for analysis-driven sound processing

Niels Bogaards and Axel Röbel

IRCAM Analysis-Synthesis Team, Paris, 75004, France
bogaards@ircam.fr,
roebel@ircam.fr

## ABSTRACT

AudioSculpt is an application for the musical analysis and processing of sound files. The program unites a very detailed inspection of sound, both visually and auditorily, with high quality analysis-driven effects, such as time-stretch, transposition and spectral filtering. Multiple algorithms provide automatic segmentation to guide the placement of sound treatments and steer processing parameters.
By designing transformations directly on the sonogram, very precise spectral modifications can be made, allowing both intuitive sound design as well as sound restoration and source separation.

## 1. INTRODUCTION

To provide a convenient and high quality analysis and processing tool to composers and musicians as well as researchers, the IRCAM started development of AudioSculpt in 1995.

The central notion of the program is that viewing, editing and manipulating a sound in the frequency domain yields significant advantages over the more common time domain processing and waveform display. Interacting with a sonogram and using the phase vocoder as processing engine is both musically intuitive and scientifically insightful.

Over the years the application has seen steady development in its user interface as well as in the processing algorithms. AudioSculpt's modular setup, with separate processing kernels has meant that many of the algorithms and programs developed at the IRCAM's various research departments can be seamlessly integrated, with data being exchanged in the standardized Sound Description Interchange Format (SDIF)[1].

The combination of analysis and processing into an integrated package provides a powerful way to inspect and treat sounds, as specific time and frequency ranges can be targeted, allowing detailed and high quality processing.

To a large extent sound analysis and sound processing are complementary, where analysis serves to guide various forms of processing, and processing is the ultimate form of validating the analytic results. In AudioSculpt there is also a close technological link between the analysis and processing stages; almost all transformations make use of analysis phases, which are also available separately, and the possibility to view and evaluate these analysis steps facilitates the selection of optimal processing parameters.

## 2. THE FREQUENCY DOMAIN

Both AudioSculpt's analysis and processing take place in the frequency domain, as opposed to the more common time-domain. To be able to work in the frequency domain, the sound needs to be transformed to obtain its spectrum. When all treatments are processed the spectrum is converted back to the time-domain to produce the resulting sound.

Various methods to obtain spectral information exists, usually relying on the discrete Fourier transform, or its computationally more efficient variant the Fast Fourier Transform (FFT). The biggest problem when using the Fourier transform is that it is meant only for stationary and infinite signals. Since musical signals per definition never follow these two constraints, a modified version of the general technique is used, the Short-Time Fourier Transform (STFT)[2].

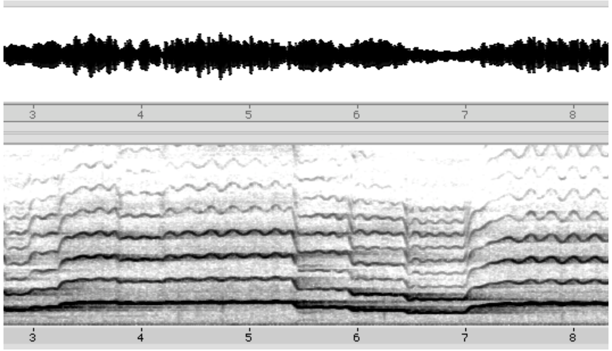


Fig 1. The waveform and sonogram in AudioSculpt

If no effects are applied, the conversion to and from the time-frequency domain can in theory be lossless, provided the resynthesis from frequency to time-domain performs is the exact inverse of the analysis stage. However, as soon as we start modifying the sound in the spectral domain, artifacts are likely to be introduced. The main causes for these artifacts lie in the fact that the STFT models the sound with a finite number of sinuoids that are assumed to be stationary within the windowed segment. For rapidly changing sounds a high time resolution and therefore small windowsize is required. However, a small windowsize has a lower frequency resolution, limiting the accuracy in frequency that the transformation can use. A similar trade-off is made in the choice of the windowing function or window type: there is no single solution that produces the best results on all kinds of signals [3].

The continuing increase in the processing power of personal computers has made it more feasible to use the short-time Fourier transform (STFT) on medium to long soundfiles.

## 3. SPECTRAL ANALYSIS

AudioSculpt features both 'general' spectral analysis methods, and ones specific for musical analysis. Various types of sonogram analysis provide insight into the sound's spectral evolution over time, providing information that is often more relevant than the sound's waveform. Fundamental Frequency estimation, Partial Tracking and automatic segmentation all rely on an interpretation of the sound's spectral components and are used to show specific musically relevant aspects of the spectrum.

### 3.1. Sonogram

A sonogram is a plot of the spectral evolution over time. The sonogram plays a central role in AudioSculpt as point of reference for further spectral analyses or the placement of treatments and transformations. The most common sonogram is the one based on the STFT, which shows the amplitude part of the sound's Fourier transform in equally spaced frequency bins (see Fig.1).

Whereas the theoretical Fourier transform is infinitely accurate in frequency, it holds no time information (as opposed to the time-domain signal which has optimal time resolution but no frequency information). To represent a frequency response that varies over time, the time-windowed STFT presents a useful compromise. However, it is important to be aware of the various trade-offs (due to windowsize, FFT size, window shape etc.) involved in order to correctly interpret an FFT sonogram [4].

#### 3.1.1. Spectral envelopes

LPC, Discrete Cepstrum and True Envelope analysis all generate sonograms that represent a spectral envelope, rather than a discretely sampled spectrum in bins, as the STFT does. These envelopes display a smoother spectral shape, making it easier to distinguish certain qualities in the sound, such as formants (Fig. 2) [5,6].

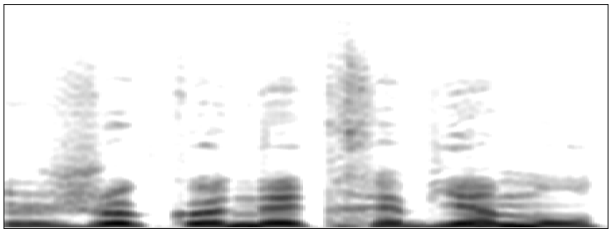


Fig 2. True Envelope spectrum of a speech signal

### 3.1.2. Reassigned Spectrum

The reassigned spectrum is a recent technique relying on phase information to improve the time-frequency location of the represented energy. By interpreting the phases of successive bins, the dominant frequencies can be deduced in finer detail, which for certain sounds yields a more accurate representation.

## 3.2. Fundamental Frequency

Fundamental Frequency or F0 analysis estimates the fundamental frequency of sounds, supposing a harmonic spectrum. The resulting breakpoint function can serve as a guide for subsequent treatments, as a basis for partial trajectory analysis, or be exported to other applications, for instance to serve as compositional material. The fundamental frequency is plotted onto the sonogram, and can be edited. Furthermore it is possible to analyze different sections of the sound with different parameters, according to the nature of the sound.

## 3.3. Partials

Partials are the individual frequencies present in a sound. For harmonic sounds, partials are integer multiples of the fundamental frequency and are also called overtones. In inharmonic sounds, such as gongs, there is a non-integer relationship between the constituent frequencies.

### 3.3.1. Partial Tracking

In Partial Trajectory Analysis breakpoint functions are created for individual sinusoidal components in the sound. As with the fundamental frequency estimation, the resulting trajectories can be overlaid on the sonogram, edited and exported. Partial Tracking, which implements a standard additive analysis, can either work in harmonic or inharmonic mode [7]. The harmonic partial tracking mode uses a previously analyzed fundamental frequency function as guidance for harmonic partials. The inharmonic mode finds peaks in spectral frames and forms partials from subsequent peaks, using a sophisticated connection algorithm. This algorithm can also find partials for polyphonic and inharmonic sounds. Both analysis methods yield very precise results, and can be exported and used in other programs using the SDIF format, to serve for instance as compositional material. Alternatively, the partials can be resynthesized to create a new sound containing just the selected partials.

### 3.3.2. Chord Sequence Analysis

Chord Sequence Analysis is special kind of Partial Tracking, where within temporal bounds (zones delimited by markers) stable frequencies are found (Fig. 3). Using one of the automatic segmentation methods or by demarcating zones by hand musically relevant events can be isolated and their spectral content analyzed.

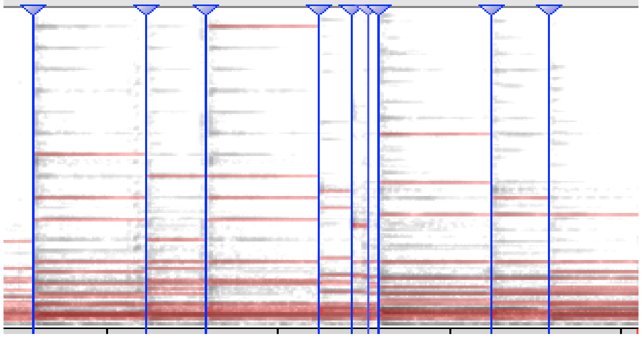


Fig. 3 Chord Sequence partials between markers

## 3.4. Segmentation

Segmentation serves to define temporal zones in the sound. In AudioSculpt, time markers can be placed by hand, or by three automatic segmentation methods; 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 energy between adjacent FFT frames An energy variation threshold can be interactively adjusted to filter less significant transients.

The generated markers can be edited and used in various ways. In a purely analytical setting, the markers can be adjusted, edited and consequently exported for use in other applications. Within AudioSculpt, the markers can 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. All marker types can also serve as zone delimiters for the Chord Sequence and the Noise Removal algorithms.

### 3.5. Tools

AudioSculpt's sonogram is meant to be the sound's main representation, whether it be for the inspection of a sound's spectral content, measurement of specific features or the precise placement, setting or alignment of transformations. To this end, a very flexible and powerful zoom system is in place, allowing the user to focus on minute details or overall trends, and interactive contrast controls facilitate the interpretation of the sound's spectral representation.

The specialized diapason (Fig. 4) and harmonics tools allow interactive and exact measurement and comparison of frequency, as well as the ability to listen to separate bins and partials [7]. A recent addition is the scrub mode, which performs a real-time resynthesis of the instantaneous spectrum, making it possible to listen to single stationary time-windows in the sound, or search for subtle spectral changes by moving through the file at a very slow speed. By modifying the transformation parameters, such as windowsize and FFT size one can listen to the differences in temporal and frequency resolution.

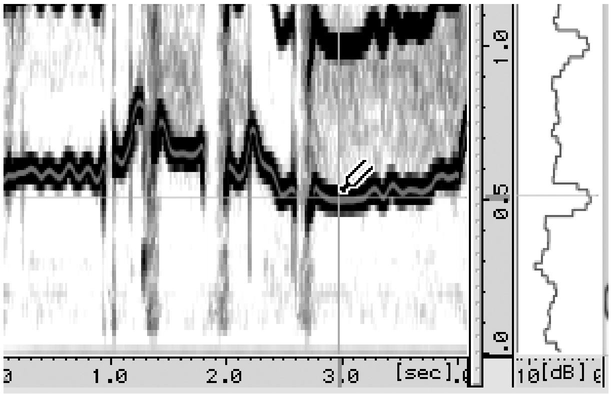


Fig. 4 Inspection of a frequency with the diapason tool

## 4. PROCESSING AND TRANSFORMATIONS

Most transformations available in AudioSculpt are based on phase vocoder techniques. This means that in the effects delicate and musically relevant algorithms can be applied, for example envelope preservation and time correction when doing a transposition, transient preservation when time-stretching and spectral subtraction for noise reduction [8]. Furthermore, detailed and very accurate filters operating on single frequency bins can be used in sound restoration or to subtly change the spectral balance of a sound.

Since all the advanced processing options rely on analyses that are also available separately in AudioSculpt, a visual inspection can help to find optimal settings to be used in the processing. For instance, the markers produced by the Transient Detection segmentation algorithm correspond to the transients that will be preserved in dilating treatments, such as time-stretch and time-corrected transposition. Likewise, the sonograms produced by LPC or True Envelope analysis method show the envelope that can be preserved when doing a transposition.

A detailed study of the visual representation of the resulting sound can also help to identify which artifacts where introduced in the processing, as a result of the choice of windowsize, FFT size and type of window, and to iteratively find the settings that best match the sound's characteristics.

AudioSculpt contains both 'classic' phase vocoder-based effects, like dynamic time-stretching, transposition and band filters, and more exotic treatments, such as spectral freeze, clipping and the pencil filter, with which arbitrarily shaped filters can be designed, that change over time, for instance to follow a partial frequency.

A novel effect, based on the transient detection algorithm as well, is Transient Remix, in which the balance between transients and stationary parts of the sound can be readjusted.

Perhaps the most extreme form of using spectral analysis to control processing parameters is in Cross Synthesis, where the spectral envelope of one sound is used to filter a second, or two spectra are cross-multiplied.

## 4.1. Filters

Filtering in the frequency domain, using the phase vocoder, has significantly different properties than more traditional time-domain filters. As time-domain filters can be compared to the parametric equalizer as found on a mixing desk, where a specific frequency is boosted at the expense of others, and therefore always the response of the entire frequency range is affected, frequency-domain filters are more like large graphic equalizers with a vast number of parallel band filters that do not interfere with neighboring frequencies.

Typical processing settings in AudioSculpt use a windowsize of over a thousand points, which one could compare to a graphic equalizer with several hundred bands (the main difference being that equalizer bands are usually at logarithmic intervals and the FFT uses a linear scale). This high resolution means that very precise filters are possible, acting on individual frequency bands. In AudioSculpt various types of filters can be drawn or overlaid directly on the sonogram, providing a convenient workspace for interaction with the sound's spectrum.

### 4.1.1. Pencil and Surface filters

The Pencil and Surface filters are closest to the direct manipulation of the sound's time-frequency representation. In a manner analogous to drawing programs, filters are drawn onto the sonogram and can be moved, scaled, copied, etc. By selectively filtering out specific partials, sources can be separated or timbres subtly changed (see Fig. 5).

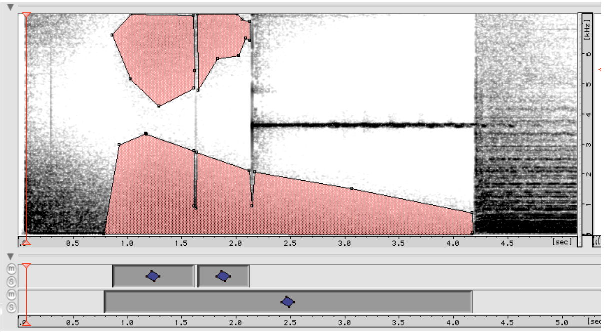


Fig. 5 Surface filters to separate two sources

### 4.1.2. Band Filter

The band filter applies multiband dynamic brickwall filters that either pass or reject.

### 4.1.3. Image Filter

AudioSculpt's generalized frequency domain approach means that in fact any table of data could be used as a time-frequency filter. The image filters imports a standard image file (jpeg, png, tiff etc.) and overlays it on the sonogram. Thus, for instance sketched partitions can be imported and applied as filter to a source sound.

### 4.1.4. Breakpoint Filter

Close in concept to the graphic equalizer, the breakpoint filter is applies a drawn frequency response to a section of the sound.

## 4.2. Time stretch and Transposition

Two 'classic' phase vocoder effects are time stretching and transposition, whose quality is generally much higher than their time-domain counterparts. In AudioSculpt, the traditional phase vocoder algorithm is significantly enhanced, yielding more musical results and a wider range of practical applicability. Breakpoint functions can control the time-stretch and transposition factors over time.

### 4.2.1. Time stretch

A common issue with time stretching is the handling of transients, short noisy sections of a sound that occur typically at the start of a note. Where stretching the sound's stationary part often yields convincing results, a stretched transient is immediately perceived as unnatural, limiting the usability of the classic algorithm. AudioSculpt's transition detection and preservation methods successfully address this problem by leaving detected transients untouched while stretching the non-transient components (see Fig. 6 and 7). Thus even very large stretch factors (> 10 times) or sounds with many transients, such as percussion, produce natural sounding results [9].

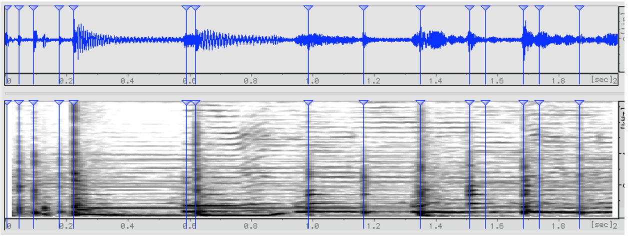


Fig 6. Markers pinpoint detected transients

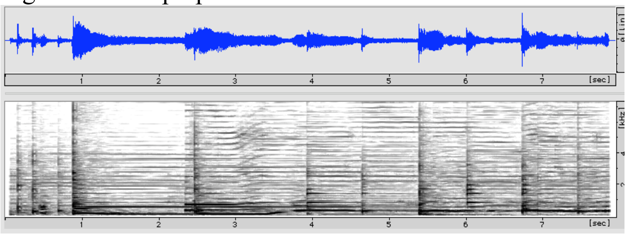


Fig 7. The same sound, stretched 4 times, using transient preservation

### 4.2.2. Transposition

With transposition, another problem usually limits the perceived sound quality. Natural sounds are often ‘shaped’ by specific formant frequencies, peaks in the amplitude response due to the nature of the resonating body. The formants contribute significantly to the sound’s timbre, and their frequencies should ideally not be shifted when a transposition is applied. For example, when a guitar note’s pitch is to be transposed, the transposition should not shift the amplitude peaks that are due to the guitar’s resonant body, but only the frequencies that are input into the body, namely those produced by the vibrating string.

To preserve the original timbre, or spectral envelope, a spectral analysis is made using the LPC, Discrete Cepstrum or True Envelope methods. Before the transposition is applied, this overall spectral shape is subtracted form the sound, and afterwards the envelope is reapplied, so that the formant frequencies were not affected.

In practice, the envelope preservation methods found in AudioSculpt can produce very satisfactory results, especially noticeable in sounds where formants play a large role, such as the voice.

## 4.3. Spectral Clipping

The processing counterpart of the sonogram’s contrast sliders, the clipping filter passes only frequencies whose amplitude is within a specified dynamic range, cutting all frequencies below it and leveling the ones above it. The Renormalize function then re-expands the passed amplitudes to take advantage of the full dynamic range.

## 4.4. Freeze

The Freeze algorithm sustains a selection for a given time. The selected region’s spectrum is repeated or ‘frozen' for the duration of the effect, which can produce abstract semi-stationary sounds or subtly prolong a section of a sound.

## 4.5. Reverse and Repeat

A big advantage of working in the frequency domain is that all phases can be corrected in the resynthesis. In the Reverse/Repeat transformations this is used to generate smooth transitions when collating various time slices of the sound. Where a time-domain copy-paste or reverse of a section of the sound is likely to generate clicks or discontinuities, AudioSculpt’s transitions are smooth and musical.

# 5. PROCESSING

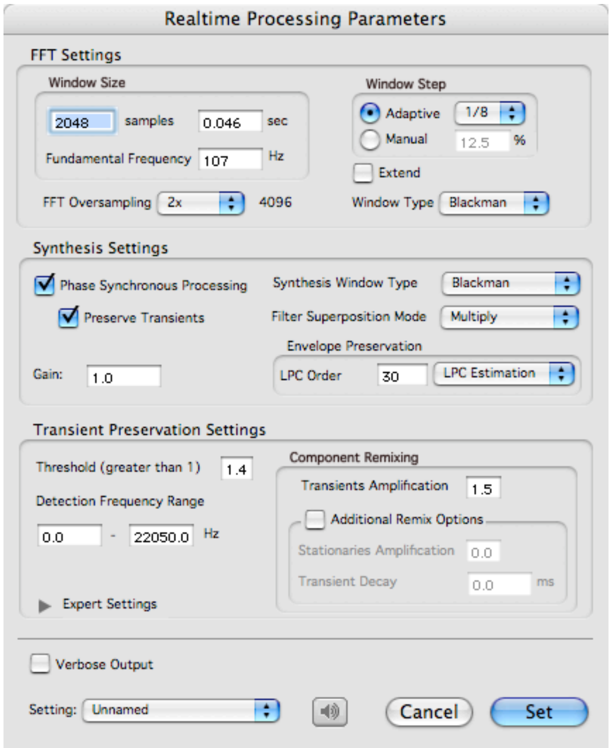
AudioSculpt’s analysis and processing stages share many concepts, as most analysis methods also play some role in the processing phase. For the phase vocoder, which underlies most processing, the concepts of window size, overlap and FFT size are important, as well as settings for transformation specific features, such as transient preservation or envelope preservation.

## 5.1. Treatments and the sequencer

Most transformations or effects in AudioSculpt take the form of Treatments, movable objects that are placed on tracks of a sequencer, similar to those found in multitrack recording or MIDI sequencers. This approach not only makes it feasible to manage a large number of filters or effects, but more importantly ensures the highest possible sound quality in complex transformations involving many treatments. The sequencer concept is useful, as it provides a means to focus on certain treatments and try them out, while other effects are ‘muted’, which means these are currently not applied. The consequent final combined processing will thus be at maximum fidelity.

#### 5.1.1. Realtime Processing

With the continuing increase in processing speed of personal computers, realtime audio processing has become ever more feasible. Currently, the conversion to and from the spectral domain using the STFT can on most computers easily be done in realtime. Although it remains possible to choke processing when applying many costly algorithms at once, most treatments in AudioSculpt can be carried out in realtime, which significantly speeds up the trial-and-error process that is often involved when finding ideal settings for the transformations (Fig. 8). Even though AudioSculpt's realtime mode is meant for pre-flight experimentation rather than as an interactive performance tool, no compromises are made to the sound quality.



Realtime Processing Parameters

FFT Settings

Window Size

2048 samples 0.046 sec

Fundamental Frequency 107 Hz

FFT Oversampling 2x 4096

Window Step

● Adaptive 1/8

○ Manual 12.5 %

□ Extend

Window Type Blackman

Synthesis Settings

☑ Phase Synchronous Processing

☑ Preserve Transients

Gain: 1.0

Synthesis Window Type Blackman

Filter Superposition Mode Multiply

Envelope Preservation

LPC Order 30 LPC Estimation

Transient Preservation Settings

Threshold (greater than 1) 1.4

Detection Frequency Range

0.0 - 22050.0 Hz

▶ Expert Settings

Component Remixing

Transients Amplification 1.5

□ Additional Remix Options

Stationaries Amplification 0.0

Transient Decay 0.0 ms

□ Verbose Output

Setting: Unnamed

Fig. 8. The Realtime Processing Parameters dialog

### 5.2. Sound Restoration

A specific use of AudioSculpt is in the field of sound restoration. The very precise filters and flexible Fourier transform settings make it possible to modify or remove selections in the time-frequency plane with a minimum of unwanted side effects. The scrub mode can help the identification of problematic areas and the interactive tuning of compensating filters.

#### 5.2.1. Noise Reduction

Besides the use of filters, another sound restoration tool is the Noise Reduction function, where marked zones are defined as noise keys, and their spectrum then subtracted from the sound. Optionally, sinusoidal noise (like hum) can also be removed.

### 5.3. Partial Synthesis

The partials found using either Partial Tracking or Chord Sequence analysis can be resynthesized to create a new sound that only contains the selected partial frequencies. Also available in real-time, this method can be used to isolate harmonic content.

### 5.4. Cross Synthesis

Perhaps the most extreme form of using spectral analysis to control effects parameters is in Cross Synthesis, where the spectral envelope of one sound is used to filter a second, or two spectra are cross-multiplied. In the first case, called Source Filter Synthesis, the spectral envelope, as calculated using LPC, Discrete Cepstrum or True Envelope is applied to a second sound, effectively filtering the second sound by the first. As it is also possible to use the inversion of the envelope, Source Filter Synthesis can also be used to remove the spectral envelope from a sound itself, in order to neutralize its timbre [10].

In Generalized Cross Synthesis both soundfiles are analyzed using the STFT, and their spectra combined according to specified factors for the amplitudes and phases. In this way a new sound can be created using one file's phases and the other sound's amplitudes.

## 6. KERNELS

For the actual analysis and processing of sound data, AudioSculpt uses different processing kernels. These kernels are developed at the IRCAM as cross-platform command line-based tools, often on Linux platforms. 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 by researchers on UNIX and Linux work-stations, using tools like Matlab and Xspect, and new versions of the kernel can be directly and transparently used by AudioSculpt [11].

Currently, most analysis and processing is handled by the SuperVP kernel, an enhanced version of the phase vocoder, that's been under continual development since 1989. For partial analysis and synthesis the Pm2 kernel implements an additive model [12, 10].

As the kernels are in fact commandline tools, AudioSculpt features console windows in which the commandlines sent to the kernels are printed. It is possible to modify and then execute these commandlines within AudioSculpt, or from a shell such as OSX's Terminal.app.

Analysis and sound files generated with AudioSculpt contain a string with the exact command-line used to create them, so that the complex and subtle settings remain available for later reference.

## 7. SDIF

The large number of different analysis methods present in AudioSculpt and other programs developed at research institutes like the IRCAM prompted the need for a flexible, extensible file format to describe information extracted from sounds. The Sound Description Interchange Format (SDIF) has proven to be an excellent way to exchange analysis data between AudioSculpt, processing kernels like SuperVP and Pm2, composition software like OpenMusic and Max and purely scientific tools such as Matlab [13]. Currently, all analysis data made with AudioSculpt is stored using the SDIF file format.

As SDIF is a binary format, it is precise and efficient for large datasets such as FFT analyses of long sounds. The extensibility facilitates the addition of new fields to an existing data type, without compromising its compatibility.

An Open Source project, programs can add SDIF support at a low cost, and then take advantage of AudioSculpt's extensive viewing and editing capabilities [14].

## 8. FUTURE WORK

Research at the IRCAM is ongoing, and constant feedback from composers and musicians means that new features continue to be added while existing ones are being enhanced.

Current areas of development include the research into various forms of annotation, the detection of tempo and vibrato and a system to define time trajectories that allow sophisticated collage techniques.

## 9. 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. The continuing development of the user interface, the analysis and processing algorithms and the modernization efforts to take advantage of emerging possibilities have lead to an application that is both mature and stable as well as fresh and innovative.

Recent 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, have significantly increased the program's practical usability.

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

## 10. TECHNICAL SPECIFICATIONS

- Apple Macintosh computer running OSX 10.3 or higher
- Full CoreAudio compatibility
- AIFF, WAV and SDII soundfiles, with a samplerate up to 192 kHz, multichannel, in integer and floating points samples up to 32 bits
- AltiVec optimized

## 11. REFERENCES

[1] Schwarz, D., and M. Wright, "Extensions and Applications of the SDIF Sound Description Interchange Format", in Proceedings of the International Computer Music Conference, 2000.

[2] Arfib, D., F. Keiler and U. Zölzer, "Time-frequency Processing", in DAFX - Digital Audio Effects, J.Wiley & Sons, 2002

[3] Rabiner, L. and B. Gold, "Theory and Application of Digital Signal Processing", pp. 88-105, Prentice Hall, 1975.

[4] N. Bogaards, "Analysis-assisted Sound Processing with AudioSculpt", in Proceedings of the 8th International Conference on Digital Audio Effects (DAFx'05), Madrid, 2005

[5] Cappé, O., J. Laroche and E. Moulines, "Regularized estimation of cepstrum envelope from discrete frequency points", in Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 1995

[6] Röbel, A. and X. Rodet, "Spectral envelope estimation using the true envelope estimator and its application to signal transposition", submitted for publication to DAFX2005.

[7] Serra, X., and J. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition" in Computer Music Journal, vol. 14, no. 4, pp. 12-24, 1990.

[8] Bogaards, N., A. Röbel and X. Rodet, "Sound Analysis and Processing with AudioSculpt 2" in Proceedings of the International Computer Music Conference, Miami, 2004.

[9] Röbel, A.: "A new approach to transient processing in the phase vocoder", in Proceedings of the 6th International Conference on Digital Audio Effects (DAFx'03), pp.344-349, London, 2003.

[10] Serra, M-H, "Introducing the Phase Vocoder", in Musical Signal Processing, Swets & Zeitlinger, Lisse, 1997

[11] Rodet, X., D. François and G. Levy, " Xspect: a New Motif Signal Visualisation, Analysis and Editing Program", in Proceedings of the International Computer Music Conference, 1996

[12] Depalle, P. and G. Poirrot, "SVP: A modular system for analysis, processing and synthesis of sound signals", in Proceedings of the International Computer Music Conference, 1991

[13] 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

[14] SDIF website, http://www.ircam.fr/sdif