ANALYSIS-ASSISTED SOUND PROCESSING WITH AUDIOSCULPT

Niels Bogaards

▶ To cite this version:

Niels Bogaards. ANALYSIS-ASSISTED SOUND PROCESSING WITH AUDIOSCULPT. 8th International Conference on Digital Audio Effects (DAFX-05), Sep 2005, Madrid, Spain. pp.269-272. hal-01161331

HAL Id: hal-01161331
https://hal.science/hal-01161331

Submitted on 8 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANALYSIS-ASSISTED SOUND PROCESSING WITH AUDIOSCULPT

Niels Bogaards

Analysis-Synthesis Team, IRCAM, Paris, France

niels.bogaards@ircam.fr

ABSTRACT

Digital audio effects using phase vocoder techniques are currently in widespread use. However, their interfaces often hide vital parameters from the user. This fact, and the generally limited ways in which sound designing and compositional tools can represent sounds and their spectral content, complicates the effective use of the full potential of modern effect processing algorithms.
This article talks about ways in which to use analysis to obtain better processing results with phase vocoder based effects, and how these techniques are implemented in IRCAM's AudioSculpt application. Also discussed are the advantages of using the SDIF format to exchange analysis data between various software components, which facilitates the integration of new analysis and processing algorithms.

1. INTRODUCTION

Digital effects and signal processing are omnipresent today, and the steady increase in available processing power has meant that ever higher quality and computationally intensive algorithms can be used, including those that operate directly in the spectral domain as obtained by the Short-Time Fourier Transform (STFT)[1]. While modern algorithms, such as effects based on the phase vocoder, can produce phenomenal results, treating these effects as black boxes, with just two or three controls, as seen in many plugins today, hampers the full exploitation of current signal processing techniques.
For many reasons, the careful spectral analysis of a sound can benefit the quality of subsequent processing or treatment. Modern phase vocoder based digital audio effects often depend on complex parameters, that have a large impact on the resulting sound quality. Moreover, for effects that are applied in the spectral domain, appropriate settings need to be selected for the transform [2]. These parameters may be perceived as musically unintuitive and difficult to understand. As a result, many applications and plugins choose to hide vital settings from the user, or try to automatically adapt their values. These compromises may introduce undesired artifacts, resulting in suboptimal sound quality. By allowing the user to first analyze the sound and based on the results define the processing settings, a more predictable and higher quality output can be achieved.
In AudioSculpt, analysis and processing go hand in hand. Under development since 1993, AudioSculpt contains a wealth of analysis and processing methods, most of which are depending on the STFT and its processing counterpart, the phase vocoder [3]
Various analysis methods, such as the detection and demarcation of musical events, like note-onsets, the analysis of spectral content and harmonicity, the detection of spectral changes and transients provide distinctive descriptions of the sound's content. This information can be used to place and align treatments and adapt parameters to the content found.
AudioSculpt provides a powerful toolset to analyze sounds in a detailed way, both visually through zoomable sonograms and partial representations, and auditorily, with the playback of single analysis frames and frequencies and a real-time interactive time-stretch. Likewise, the visual and auditory evaluation of settings to be used in the digital analysis and processing stages, such as windowsize, FFT size and windowstep, allows the selection of optimal representations in the spectral domain for the particular sound.

2. SOUND ANALYSIS IN AUDIOSCULPT

AudioSculpt features many analysis methods to extract information from sound, and ways to interactively inspect the results in a visual or auditory way, as well as modify them. The types of information that can be extracted from the sound include spectral composition, transient detection, masked and perceived frequency, fundamental frequency, harmonic and inharmonic partials.
These (possibly edited) analyses can also serve as input for new analysis, such as is the case with the Chord Sequence analysis, which takes a series of time markers as delimiters for subsequent partial analysis, or the fundamental frequency analysis, which serves as a guide for Harmonic Partial Tracking analysis [4,5].
Spectral analyses are displayed on the versatile sonogram, where they can be inspected and edited. Reanalysis of parts of the sound with different settings, or correction by hand helps to obtain accurate results, fine-tuned for the particular task at hand.

2.1. Spectral Analysis

Central to most analysis and processing done with AudioSculpt is the sonogram representation. Using various analysis methods, like STFT, Linear Prediction Coding (LPC), Discrete Cepstrum, True Envelope or Reassigned FFT, a sonogram gives a visual representation of the sound's spectral content over time, which often serves as a point of orientation for subsequent analysis and processing [6,7].
For a meaningful evaluation of sounds, it is important that the sonogram's display is flexible and interactive. To this end AudioSculpt features a very powerful zoom, which works independently in time and frequency, as well as sliders to change the sonogram's dynamic range in real-time. An adjacent instantaneous spectrum view can be used for the inspection of

single analysis frames, or the comparison of spectra at two discrete times.
Fundamental Frequency or F0 analysis estimates the fundamental frequency of sounds supposing a harmonic spectrum. This fundamental frequency can serve as a guide for subsequent treatments, as a basis for harmonic 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.

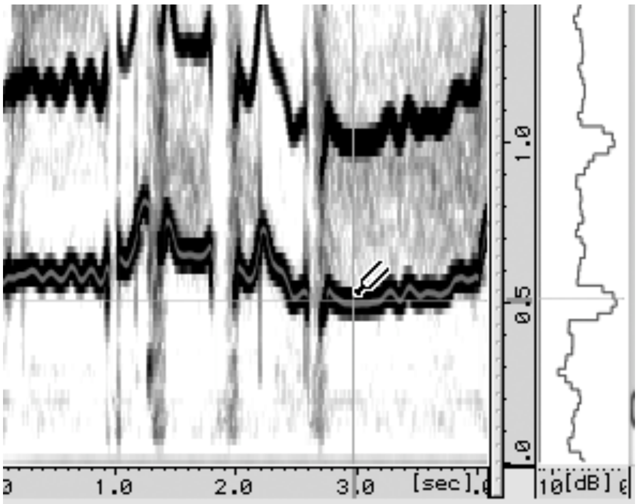

Fig 1. Sonogram with overlaid F0 analysis and the diapason tool

AudioSculpt features multiple methods for the estimation of partial or harmonic content of a sound. Using an additive analysis model partial trajectories can be found, which can also serve as control input to an additive synthesizer [5].

Other algorithms available to evaluate the spectral content are the Masking Effects, Peak and Formant analysis. Masking Effects uses the psycho-acoustical algorithm developed by Terhardt to estimate which spectral peaks are masked by other frequencies, and which pitch is perceived [8].
Formant and Peak analysis search for peaks in the spectral envelope, as obtained by LPC or Discrete Cepstrum analysis.

2.2. Segmentation

Segmentation serves to delimit 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 flow between FFT frames [6]. Detected events can be filtered according to their significance using interactive sliders, and hand-editing allows for precise fine-tuning to obtain a desired result.

2.3. Analysis Tools

The special *diapason* and *harmonics* tools allow the interactive and exact measurement and comparison of frequency, as well as the ability to listen to separate bins and partials [4]. A new *scrub* mode 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. Modifying the transformation parameters, such as windowsize and FFT size provides insight into the significance of temporal and frequency resolution in an auditory way.

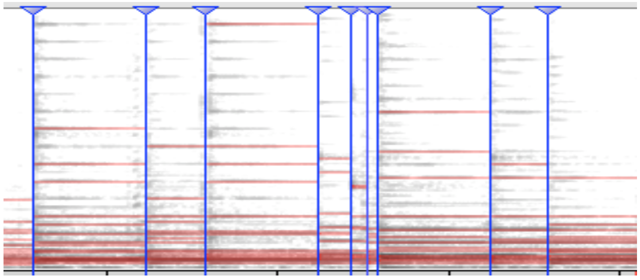
Fig. 2. A Chord Seq analysis with transient detection markers

3. HIGH QUALITY PROCESSING

Sound processing algorithms can largely be divided into two categories; those that operate directly on the sampled values in the time-domain, and those that are applied in the time-frequency or spectral domain. A significant advantage of the application and design of effects in the spectral domain is that spectral representations are much closer to the perceived content of a sound then their time-domain counterparts.
To be able to work in the spectral domain, the sound needs to be transformed using techniques such as the STFT. After the transformation, the effects are applied, and the sound is reconverted to the time-domain. [1]
If no effects are applied, the conversion to and from the time-frequency domain can in theory be transparent, provided the resynthesis from frequency to time-domain performs is the exact inverse of the analysis stage. However, as soon as the sound is modified in the spectral domain, the transformation will always introduce artifacts. The main causes for these artifacts lie in the fact that the STFT works upon windowed segments of the sound, therefore a trade-off is always made: a larger windowsize permits a higher frequency resolution, but has a larger 'spill' in time, and therefore less time accuracy. Conversely, a small windowsize will correctly preserve the timing of spectral events, but limits the resolution in frequency that the effect 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 [9].
Because of these inherent and unavoidable artifacts, it is of great importance to choose the optimal windowsize according to the sound's content and the desired result (see Fig. 3).

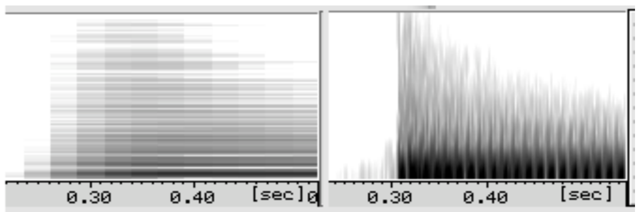

Fig.3. The attack part of a guitar tone, analyzed with a windowsize of 10000 samples (left) and 100 samples (right)

3.1. Sound transformations using AudioSculpt

All 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 spectral envelope preservation and time correction when doing a transposition, transient preservation when time-stretching and spectral subtraction for noise reduction [10]. Furthermore, detailed and very accurate filters can work on single frequency bins, which can be used for instance in sound restoration or subtle changes in the spectral balance of a sound.
Since all the advanced processing options rely on analyses also available separately in AudioSculpt, a visual analysis 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 sonogram produced by LPC or the True Envelope analysis method shows the envelope that can be preserved when doing a transposition, or the filter response for use in cross-synthesis [6].
A detailed visual representation can also help to identify which artifacts where introduced in the processing, due to the choice of windowsize, FFT size and type of window, and to iteratively find the settings that best match the sound's characteristics.

3.2. Effects available in AudioSculpt

AudioSculpt contains both 'classic' phase vocoder-based effects, such as dynamic time-stretching, transposition, band filters and cross-synthesis, as well as 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 also on the transient detection algorithm, is Transient Remix, in which the balance between transients and stationary parts of the sound can be readjusted.

3.3. Processing Settings

When applying an effect to a sound, one needs to decide on when to apply it, how to set the effect's parameters, and possibly how to control these parameters over time. While these questions seem trivial and are often answered by a trial-and-error process, the use and careful inspection of various analyses of the sound can help to apply the effect in a more optimal way, thus yielding superior quality results. For instance, applying a filter right after a sound's transient phase may produce a more natural sound than just processing the whole sound with the filter.
Sound is typically not spectrally stationary over time, therefore one often wants a treatment's parameters to change over time as well. Designing a filter by drawing directly onto the sonogram assures that the filter will act only on the desired frequencies, introducing as few artifacts as possible. Similarly, a curve drawn onto the sonogram could steer subtle pitch changes.

3.4. Obtaining high quality effects

Since AudioSculpt strives to be an application for use by musicians and composers, sound quality is of extreme importance. Besides supporting a wide range of sample formats and frequencies, up to 32-bit floating point and a samplerate of 192 kHz, a system has been devised to limit the number of processing passes needed, even for complex treatments, that involve many filters, time dilations or frequency transpositions. Treatments are grouped onto tracks, which as in a sequencer can be muted and soloed [4]. Therefore, it is possible to listen to the effects separately or grouped, either using the real-time processing mode or by generating a file, and apply and combine all of them together to create the final result, thus limiting the total amount of transformations, and thus the artifacts introduced by the phase vocoder.

3.5. Sound restoration

A specific use of AudioSculpt is in the field of sound restoration.
By drawing filters directly onto the sonogram, it is possible to exactly eliminate or accentuate certain frequency bands, for instance to remove an unwanted instrument or noise from a soundfile. The pencil and surface tools can be used for the design of very detailed filters, with fine control over the attenuation factor and bandwidth.

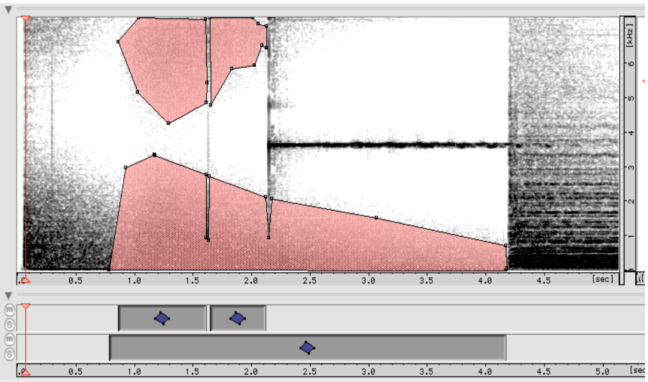

Fig 4. Surface filters to isolate a single sound

A recent addition is the Noise Removal module, which allows the definition of noisy zones in a sound, which can then function as keys for spectral subtraction. This way both sinusoidal noise, such as hum, and a noise spectrum can be removed.

4. AUDIOSCULPT FEATURES

The design and implementation of AudioSculpt has been going on for over 10 years, maturing slowly but steadily. The continuing input from composers, researchers and musicians, as well as the improved capabilities and speed of affordable computers has lead to a flexible and useable program. [3,4]

4.1. Parameter control

AudioSculpt is designed to facilitate the musical use of sophisticated analysis and processing algorithms without compromising on flexibility, adjustability and transparency. Rather than hiding behind default values and leaving it up to the users to discover the limits of these settings, all parameters are modifiable and users are expected to modify them in order to best match their current needs. To achieve this, parameters can be stored in presets, are passed between different analysis and synthesis modules and ranges are automatically adapted to the sound.

4.2. Kernels

For the actual analysis and processing of sound data, AudioSculpt uses external 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 Mac 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 Linux workstations, using tools like Matlab and Xspect [11], and new versions of the kernel can be directly and transparently used by AudioSculpt.
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 the Pm2 kernel implements an additive model [5].
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.

4.3. 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, signal-processing kernels like SuperVP and Pm2, composition software like OpenMusic and Max and purely scientific tools such as Matlab. 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 [13,14].

5. FUTURE WORK

Future work on AudioSculpt and its underlying kernels will include the selection of time-frequency regions on the sonogram, using such tools as a Magic Wand (a tool famous from Adobe's Photoshop application that allows the selection of pixels of similar value by a mouse-click), and the ability to copy, paste and displace these zones.
To profit even more from the various available analyses, a new class of filters will be able to automatically 'follow' the fundamental frequency, selected partials or harmonics.
At the same time, the increasing computational power of (multi-processor) personal computers will permit a more advanced use of real-time processing, such as the interactive manipulation of filters and effect parameters.
Furthermore, a MIDI input and output will facilitate the auditory evaluation of various analysis, as well as improved interaction and integration with other applications.

6. CONCLUSIONS

For phase vocoder based effects processing, the combination of analysis and sound processing in an iterative process allows for the selection of optimal effect and STFT parameters. By making the various analyses visible and verifiable, the results of the effect processing become more predictable and much easier to fine-tune. The use of SDIF as a standard to exchange sound analysis data has made it possible to conveniently integrate a large number of analysis methods into AudioSculpt, making it possible to visualize sound in many different ways.
AudioSculpt is available for members of IRCAM's Forum (http://forumnet.ircam.fr), as part of the Analysis-Synthesis Tools.

7. REFERENCES

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

[2] Todoroff, T., "Control of Digital Audio Effects", in *DAFX - Digital Audio Effects*, J.Wiley & Sons, 2002.

[3] Eckel, G., "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.

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

[5] 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.

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

[7] 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

[8] Terhardt, E., "Pitch perception and frequency analysis", in *Proc. 6th FASE Symposium*, Sopron, Budapest, 1986

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

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

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

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

[14] Agon C., M. Stroppa and G. Assayag, “High Level Musical Control of Sound Synthesis in OpenMusic", in *Proceedings of the ICMC*, pp. 332 - 335.