INTRODUCING ASANNOTATION: A TOOL FOR SOUND ANALYSIS AND ANNOTATION

Niels Bogaards *Chunghsin Yeh* *Juan José Burred*
Analysis-Synthesis Team, IRCAM, Paris, France
{niels.bogaards, chunghsin.yeh, juan.jose.burred}@ircam.fr

ABSTRACT

The ASAnnotation application provides many sophisticated features for the analysis and annotation of sound files. Its sound descriptions can be used in a research as well as in a composition environment. This paper describes several methods and tools available in ASAnnotation to display, produce, align and verify annotations, as well as how the SDIF format is used to produce reliable and portable sound descriptions.

1. INTRODUCTION

ASAnnotation is based on AudioSculpt [2], an application for the analysis and transformation of sound files. While ASAnnotation does not provide any sound transformation, it does provide full control over analysis parameters, like AudioSculpt, allowing high quality analysis of sound features for musicians as well as for researchers. The built-in sound analysis is mainly done by IRCAM's SuperVP and Pm2 kernels [2], but ASAnnotation can also be used to display, edit and export sound description data coming from environments such as Matlab using SDIF [12].

The past couple of years have seen an increased interest in sound and music annotation, and the introduction of several specialized tools. Popular applications originating from speech research are WaveSurfer [13, 8] and Praat[1]. In the open source domain both Audacity[2] and Sonic Visualiser [3] have sound analysis and annotation capabilities and the CLAM environment [1] allows the rapid development of customized annotation programs. A more musicological approach is taken in the Acousmographe [6].

While each of these tools has its particular strengths, ASAnnotation offers some unique features that set it apart, such as the systematic use of SDIF, the high degree of control over analysis parameters, the interactive auditory tools and the ergonomics of its user interface. It should also be mentioned that ASAnnotation is the only one of the aforementioned annotation tools that, in our opinion, performs well on Mac OSX.

The reasons for analyzing and annotating sound files are manifold, ranging from the study of low-level features such as partial frequencies to high-level descriptions of song structure. In a research context, annotation is often the only means to define a ground truth, against which algorithms are verified [14], whereas annotation is also used in tasks like score following [4]. A combination of automatically generated annotation and the ability to edit and correct manually is currently the most efficient means to achieve high quality annotation [14, 5]. This article presents the analysis and annotation capabilities of ASAnnotation, and discusses how the application can be integrated in various workflows.

2. INTERACTIVE SOUND ANALYSIS AND ANNOTATION

ASAnnotation can be used for the annotation of a wide range of sound and music features. Built-in extractors estimate descriptors for low-level features of sound, such as fundamental frequency and partial trajectories, voiced-unvoiced separation and transient positions. Higher level annotation is provided in the ability to overlay notes and text from Standard MIDI Files on a sonogram and place markers for temporal events and regions.

Various specialized tools allow interactive inspection of parts of the sound or its representations, which facilitates the task of enhancing a sound's transcription and the verification of the analysis data.

2.1. Sonogram

Central to ASAnnotation is a very flexible sonogram display, which in many cases is much more intuitive and explanatory representation of a sound than the ubiquitous waveform.

What sets ASAnnotations's sonogram apart from those found in most other applications is the very precise control of parameters like window size, window step, window function and FFT size as well zooming in both time and frequency, which makes it possible to focus on specific features of the sound, such as the exact frequency of low notes or the spectral evolution of transients. The flexible switching between linear, logarithmic and mel frequency scales allows the selection of the most suitable spectral representation for a specific sound and context.

The sonogram's graphic rendering can be interactively controlled with sliders specifying upper and lower thresholds for the various color mappings. In this way, it is for instance possible to exactly determine the point at which a partial's intensity surpasses the noise floor.

Besides the common STFT (Short Time Fourier Transform) , several other spectral analyses can be used to generate sonograms, such as LPC (Linear Predictive Coding) and Discrete Cepstrum analysis, Reassigned Spectrum and the recently developed True Envelope method for spectral envelope estimation [11].

[1] http://www.praat.org/

[2] http://audacity.sourceforge.net/

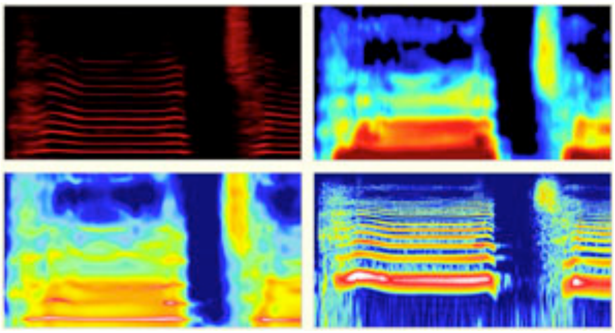

Figure 1. Different sonogram representations for the same sound, clockwise: STFT, True Envelope, logarithmical STFT and LPC

2.2. Partial Tracking

A common sound analysis method is to model the sound as a sum of sinusoids and a residual. The Pm2 kernel used by ASAnnotation can perform partial tracking both harmonically (using a previously calculated fundamental frequency track as guidance) or inharmonically, using a bias corrected estimation technique [10]. In combination with regions defined by markers, Pm2's Chord Sequence analysis produces "averaged partials" with a demodulated frequency.

2.3. Visual And Auditory Tools

Intuitive and useful as a sonogram and various analyses may be, it can be extremely insightful to listen to details of a sound in order to perceptually appreciate spectral changes, specific pitches etc. Complementing the sonogram's intuitive representation of a sound's content, ASAnnotation provides a variety of tools for detailed visual or auditory interaction with the sonogram.

ASAnnotation can also play back arbitrarily shaped time-frequency selections of the sonogram. Selected by hand or using the amplitude- threshold-driven Magic Wand tool, these **Time-Frequency zones** allow detailed inspection of the evolution of frequency ranges or for instance the level of noise.

The **Diapason** tool shows an instantaneous spectrum graph at a specific time, and allows accurate measurement at a time-frequency point. Clicking the mouse will synthesize a sine tone with the amplitude and frequency of the selected bin. The diapason can also be used to compare two spectra, facilitating for instance the measurement of a vibrato's width.

To evaluate the harmonic relations between frequencies on the sonogram, the **Harmonic Tool** displays horizontal reference lines on the sonogram. The index in the harmonic series of the frequency under the mouse can be changed to look for harmonic relations both below and above the targeted frequency. In the time domain, the Harmonic Tool indicates the periods corresponding to the harmonic intervals.

Dragging the **Scrub** tool over either the sonogram or the waveform plays back individual resynthesized spectral frames, so one can move through the sound file at arbitrary speed and direction, without altering the pitch. This feature is especially useful to find, by ear, the exact point in time at which certain perceptual features of a sound become apparent.

Another auditory tool is the **Partial Synthesis**, which can synthesize in real time one or more selected partials, as detected by a previous partial tracking analysis.

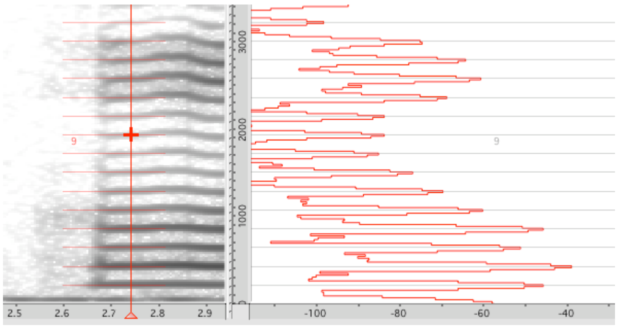

Figure 2. The Harmonic Tool can help find harmonic relations between partials, on the sonogram and on the instantaneous spectrum view.

2.4. Note Annotation in ASAnnotation

Added specifically for annotation purposes is the possibility to overlay notes and text from Standard MIDI Files on the sonogram, edit them and export the result to MIDI or SDIF files. This allows very precise alignment of MIDI notes to a sound file, for instance to prepare them to serve as reference for score following systems, or to align transcription files to their sound complements [7]. Other uses are the verification of the results of feature extraction algorithms [14] or the annotation of speech, lyrics or musicological analysis.

The annotated notes can be overlaid sonically as well, where each note is transformed into a filter with a bandwidth of 100 cents, passing or rejecting exactly the region of the sonogram that the note covers.

A specific issue with the alignment of real-world sound files and their MIDI equivalents is that real recordings might not be tuned to a 440 Hz middle-A. To accommodate this, the MIDI notes' tuning frequency in ASAnnotation can be “detuned” to match the recorded pitches.

2.5. Grid

Sonograms based on STFT analysis have a fixed time-frequency resolution, relative to the window step and the size of the FFT. In a musical context, it might be convenient to have alternative references, such as the frequencies of an equally-tempered scale or time indications relative to a given tempo. ASAnnotation contains a flexible and detailed grid system to cater to these needs. The time grid can be adjusted in seconds, beats per minute (BPM) and samples, with an offset to finely align the grid to musical events. For the frequency grid, there is a choice between linearly and harmonically spaced gridlines, a piano-roll-style MIDI grid and a rendition of musical staves. In the latter two cases the grid can be "detuned" to fit a specific sound file. The ability to space gridlines at for instance 50 cents intervals facilitates the analysis of quartertone music. Alignment of annotation objects is facilitated by "magnetic" snap modes, which adjust a note's start, end, length or frequency to the gridlines.

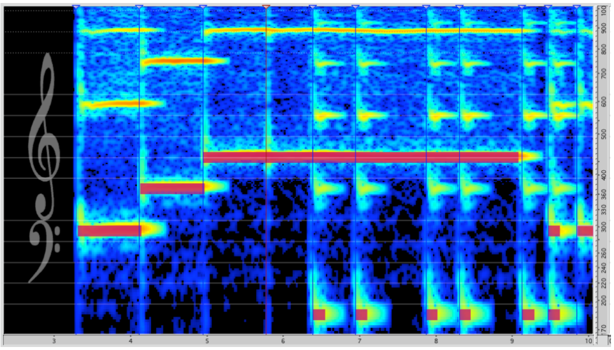

Figure 3. Markers, notes and staff grid on a sonogram

2.6. Markers

To mark events in time, such as note starts, transient positions, or musical sections, markers can be used. In ASAnnotation markers can be placed automatically using SuperVP's transient detection algorithm [9], posed by hand or imported from other applications using either SDIF or the .lab format used by WaveSurfer [13]. Specialized marker types facilitate the distinction between markers for transients, phonemes, chords, etc.

A useful feature is the "threshold slider", which can interactively adjust the number of markers based on a threshold value criterion. As with the grid, annotations can be made to snap magnetically to markers, which is especially useful for the alignment of MIDI notes to detected transients.

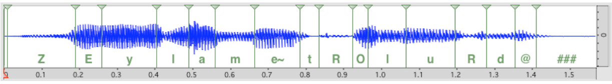

Figure 4. Phoneme markers imported from a .lab file

3. APPLICATION DESIGN

As ASAnnotation is a specialized version of the AudioSculpt application, the design of both programs resembles closely. It is useful to note that all annotation features described for ASAnnotation are also present in the current version of AudioSculpt.

3.1. Kernels

Besides sonogram analysis, a number of other analysis methods are available, such as fundamental frequency estimation, Voiced/Unvoiced analysis and Partial Tracking. For its signal processing and analysis, ASAnnotation uses kernels developed by IRCAM's Analysis-Synthesis Team; the SuperVP kernel for phase vocoder based processing and the Pm2 kernel for additive modeling of partials.

3.2. SDIF

Communication between ASAnnotation and the kernels is done using the Sound Description Interchange Format (SDIF)[12]. As opposed to XML, another popular format for conveying content analysis data, SDIF is a binary format, making it very efficient for large data sets, such as FFT frames, partial trajectories and short-time low-level descriptors. It is also very accurate as it can use up to 64-bits floating-point precision.

SDIF is an open source standard and libraries are available for C, C++, Java, Matlab and scripting languages via SWIG, making it easy to add support to existing and new programs[1]. Data contained in an SDIF file is organized as a sequence of time frames containing one or more data matrices, each one identified by a signature that denotes the specific audio feature being stored. The size of the matrices and the meaning of its elements are defined by an SDIF type declaration associated with that particular feature.

The SDIF libraries and interfaces contain a pre-defined set of types for several commonly used features such as fundamental frequency and partial trajectories. However, the standard provides with a mechanism to create user-defined types, or to extend existing ones, and thus it can be adapted to a large set of applications. The declaration of user-defined types is included as a special frame to each file, so that no additional information is needed for an application to handle them.

ASAnnotation can display and edit a large number of SDIF standard and extended types, providing a convenient environment for the evaluation of annotations and analysis results. Based on the above-mentioned type extension capabilities, a large set of new SDIF types have been proposed to accommodate several low- and mid-level temporal, spectral and perceptual features, some of them widely used in the fields of Audio Content Analysis and Music Information Retrieval. Examples include, among many others, loudness envelopes, spectral shape descriptors such as centroid and flatness, or harmonicity measures. Furthermore, on a longer time scale, other types are defined to store mid- and long-term descriptions derived from the short-term features, such as statistical moments or feature modulations. In this context, superimposing the temporal trajectory of such features on the signal or sonogram by loading such descriptor files into ASAnnotation can further help and extend annotation in the content analysis, information retrieval or sound and music perception domains.

3.3. Workflows

ASAnnotation can be used in a wide range of environments, from compositional and music oriented tasks involving Max/MSP or OpenMusic to pure science using Matlab.

[1] http://sdif.sourceforge.net/

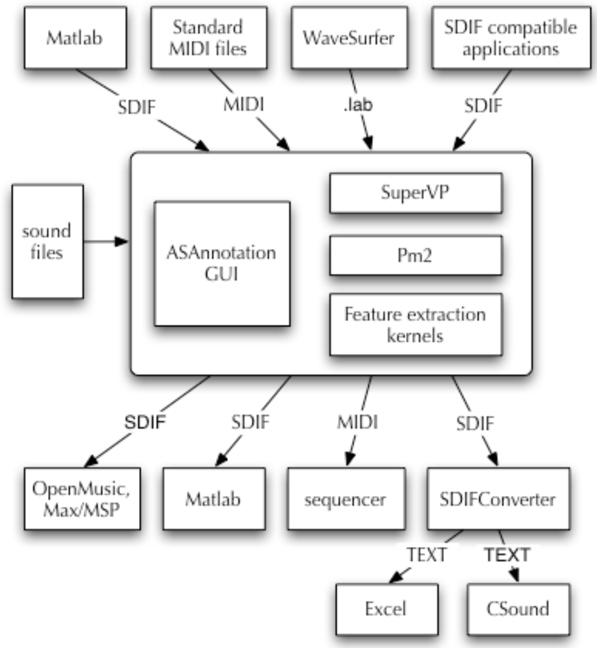

Figure 5. Various ways in which ASAnnotation can be integrated into musical and scientific workflows

4. AVAILABILITY

ASAnnotation is available for free. It can be downloaded from http://www.ircam.fr/anasyn/ASAnnotation/ For more information about IRCAM software and AudioSculpt, visit: http://forumnet.ircam.fr .

Technical Specifications: Mac OSX 10.4 or later, AIFF, WAVE or SdII sound files, up to 192 kHz, 32-bit, mono, stereo and multichannel.

5. CONCLUSION

As a recent spin-off of the AudioSculpt application, ASAnnotation is targeted primarily to researchers in the domain of music analysis and annotation and composers wanting to explore spectral characteristics of sounds. The combination of ease of use and precise control can greatly speed up annotation and analysis tasks, while at the same time providing a high level of accuracy. A key aspect in ASAnnotation's flexibility is the use of the SDIF standard, which allows researchers to generate data with tools such as Matlab and inspect and edit this data in an environment that is optimized for the study of sound and music.

6. REFERENCES

[1] Amatriain, X., Massaguer, J., García, D. and Mosquera, I. “The CLAM Annotator: A cross-platform audio descriptors editing tool”, *Proc. of the 6th International Conference on Music Information Retrieval*, London, UK, 2005.

[2] Bogaards, N and A. Roebel, "An interface for analysis-driven sound processing", presented at *119th AES Convention*, New York, USA, 2005.

[3] Cannam, C. et al., "The Sonic Visualiser: A visualisation Platform for Semantic Descriptors from Musical Signals", *Proceedings of 7th International Conference on Music Information Retrieval*, Victoria, Canada, 2006

[4] Cont A., D. Schwarz and N. Schnell, "Evaluation of Real-time Audio-to-Score Alignment", *Proceedings of 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007

[5] Fabiani, M. and A. Friberg, "Expressive Modifications of Musical Recordings: Preliminary Results", *Proceedings of the International Computer Music Conference*, Copenhagen, Denmark, 2007.

[6] Geslin Y. and A. Levebre, "Sound and musical representation: the Acousmographe software", *Proceedings of the International Computer Music Conference*, Miami, USA, 2004.

[7] Goto, M., “AIST Annotation for the RWCMusic Database”, *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Canada, 2006.

[8] Herrera, P. et al., " MUCOSA: a Music Content Semantic Annotator ", *Proceedings of the 6th International Conference on Music Information Retrieval*, London, UK, 2005.

[9] Röbel A., "Onset Detection in Polyphonic Signals by means of Transient Peak Classification|", *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Canada, 2006.

[10] Röbel, A., "Frequency-Slope Estimation and Its Application to Parameter Estimation for Non-Stationary Sinusoids", *Computer Music Journal*, Vol. 32, No. 2, pp.68-79, 2008

[11] Roebel, A. and X. Rodet, "Efficient Spectral Envelope Estimation and its application to pitch shifting and envelope preservation", *Proceedings of the 8th Int. Conference on Digital Audio Effects*, Madrid, Spain, 2005.

[12] Schwarz, D. and M. Wright, "Extensions and Applications of the SDIF Sound Description Interchange Format", *Proc. of Int. Computer Music Conference*, Berlin, Germany, 2000.

[13] Sjölande, K. and J. Beskow, "Wavesurfer - An open source speech tool", *Proceedings of the International Conference on Spoken Language Processing*, Bejing, China, 2000.

[14] Yeh, C., N. Bogaards and A. Roebel, "Synthesized Polyphonic Music Database with Verifiable Ground Truth for Multiple F0 Estimation", *Proceedings of 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007.