Sonification Platform for Interaction with Real-Time Particle Collision Data from the ATLAS Detector

Juliana Cherston
Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge, MA 02139, USA
cherston@media.mit.edu

Steven Goldfarb
Department of Physics
University of Michigan
Ann Arbor, MI 49201, USA
steven.goldfarb@cern.ch

Ewan Hill
Dept. of Physics and Astronomy
University of Victoria
Victoria, BC, Canada
TRIUMF
Vancouver, BC, Canada
e4hill@uvic.ca

Joseph A. Paradiso
Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge, MA 02139, USA
joep@media.mit.edu

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892295

Abstract
This paper presents a platform that enables composers to generate unique auditory representations of real-time particle collision data from the ATLAS experiment at CERN. An associated web page then enables the public to listen to real-time experimental data through the aesthetic lens of selected artists. The current tool is built in collaboration with the ATLAS Outreach team and is designed to increase public engagement in high energy physics by exposing the data through a novel interaction mode. More broadly, it is part of a larger vision to better harness audio as a medium to interact with big data from ever more prevalent real-time sensors.

Author Keywords
auditory display; sonification; real-time sensor data; physics

ACM Classification Keywords
H.5.5 [Information interfaces and presentation (e.g.,HCI)]: Sound and Music Computing

Project Overview
Non-speech audio is heralded by researchers as a promising tool for communicating complex data to the listener, and the field of sonification has recently emerged in order to more rigorously study auditory display of data in both academic and artistic settings [11]. The prevalence of com-

ATLAS Experiment

Overview The LHC collides particles at extremely high energies. The ATLAS experiment studies the particles created in these collisions to investigate some of the deepest questions of nature: What is Dark Matter? Are there extra dimensions? What is the origin of mass?

Detector Layers The ATLAS detector consists of an inner detector surrounded by electromagnetic (EM) and hadron calorimeters and a muon spectrometer. The inner detector tracks the trajectories of charged particles. The calorimeters measure energy produced from electromagnetic and strong interactions. The muon spectrometer is used to reconstruct the trajectories of muons, the only charged particles that escape the detector [8].

Triggering System ATLAS collects collision data at a nominal rate of 40 MHz [8]. A triggering system selects a subset of collisions to save to disk at a rate of $\sim$ kHz.

mon tools like stethoscopes and Geiger counters as well as technical advances in real-time audio processing software suggest that there are emerging opportunities to better harness audio for the communication of information. In parallel, real-time data streaming from increasingly prevalent sensors can provide a new medium for artistic expression that is yet to be fully explored [13].

The presented platform provides a novel method for interacting with real-time experimental data from the ATLAS Experiment. Our current research goal is to use this platform to explore how creative musical pieces driven by a rich, real-time data stream from the ATLAS Experiment can educate and inspire artists and the general public. The ATLAS detector is one of two general-purpose detectors built along the Large Hadron Collider (LHC) at CERN. A tiny subset of particle collision event data is routed through a sonification engine designed to map incoming data to sound properties in real-time. Users can then listen to real-time particle collision audio streams through the aesthetic lens of their preferred composer, an experience that is much like listening to different music channels on a live radio station.

This project consists of two parts: the composition engine and the website featuring real-time audio streams. The composition engine is built using Python and the Open Sound Control (OSC) networking protocol [17]. Composers can either use the project's graphical user interface (GUI) to control the data-to-audio mapping or can choose to design their own compatible audio synthesizers.

While composers have thus far focused predominantly on the audio aesthetics, they also have a chance to learn about the physics in the ATLAS data by creatively using some of the underlying patterns. As an example, they may discover that the number of charged particles decreases with increasing momentum, or that there are clusterings of

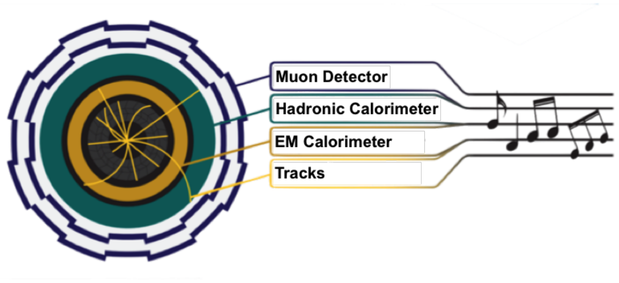

Figure 1: Sketch of the translation of data from different detector layers to audio

energy deposits in individual collision events. These patterns, if highlighted effectively by the composer in the audio, can also teach listeners about the physics of the collisions.

An associated web page currently streams three selected composition mappings. During ATLAS data taking, the recorded data control the audio produced in these streams according to rules set by the composers.

System Architecture

Event Stream

This project has been granted direct access to a special event stream designated for outreach purposes. During ATLAS data taking, this tiny subset of XML-formatted particle collision data ($\sim$1 event / 25 seconds) is routed to the sonification engine, at which point certain physics parameters from some sub-detectors are sonified. The real-time data access is restricted to the audio streams supported by the project's website.

The event stream contains low-level event data (such as calorimeter energy deposits) and only partially reconstructed data (such as tracks) rather than full ATLAS reconstructions

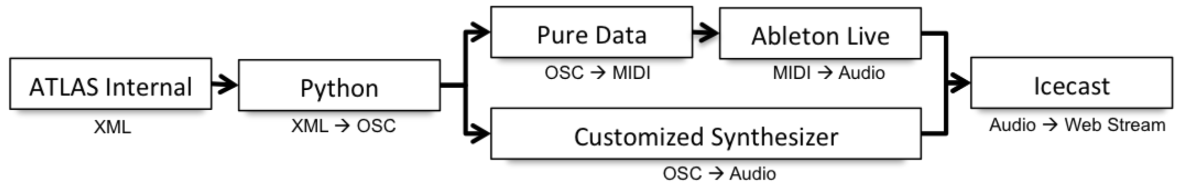

Figure 2: Data flow through platform. Software and data formats are defined for each stage. Composers can choose whether to produce sonifications using a Pure Data (PD) [2] GUI or whether to read OSC messages into custom-built synthesizers

Physics Data Streamed as OSC Messages

Liquid Argon EM Calorimeter Positions and magnitudes of energy deposits

Hadronic Endcap Calorimeter Positions and magnitudes of energy deposits

Particle Tracks Directions of track trajectory, and track momentum

Resistive Plate Chamber [part of the muon spectrometer] Positions of detector hits

Event Data Missing transverse momentum (from particles, e.g. neutrinos, that pass through ATLAS undetected), sum of all particle track energies, and sum of all particle track momenta

of particles (such as electrons). A summary of the detector layers is available in the “ATLAS Experiment” sidebar. This lower level event stream was selected for use because:

- this is the only live event stream that the experiment has approved for automatic displaying to the public,
- the partially reconstructed data files are in XML, a format that can be parsed more quickly than other data formats used by the experiment, and
- full particle reconstructions are likely to introduce larger delays, whereas this platform is meant to operate in close to real-time.

The process of checking for new events is performed every few seconds, and if one is found, the new XML data file is transferred to a dedicated machine responsible for performing the data-to-audio conversion and added to an event queue. This process can be parallelised to produce multiple different compositions using the same data. The overall data flow is shown in Figure 2.

Data-to-Audio Conversion
A set of Python scripts extract a preset list of physics parameters from the incoming data, which are described in the “Physics Data Streamed as OSC Messages” sidebar and depicted in Figure 1.

The scripts also apply some physics-based selections to extract a smaller set of interesting data from within each physics parameter, as well a set of music-based selections to add musical structure to the data. The composer can control multiple parameters that influence these selection algorithms, including the duration of audio per event and the detector geometry with respect to which physics parameters are streamed, among others. The composer can also choose to apply a “beat structure” to the sonification, in which case the data are geometrically binned and streamed at fixed time intervals, mimicking musical beat structures in traditional audio.

The resulting parameters are streamed as a time series of Open Sound Control messages, a common protocol used in audio synthesis. Our multithreaded implementation of OSC allows for simultaneous streaming of all supported physics parameters within each event. The data-to-audio mapping is described further in the section “Additional Detail on Mapping Paradigms Used”.

If composers wish to develop custom audio synthesizers using the OSC message stream, they can do so by reading in the OSC messages using signal processing software like Pure Data (PD) [2] and Max MSP [3]. Alternatively, we have

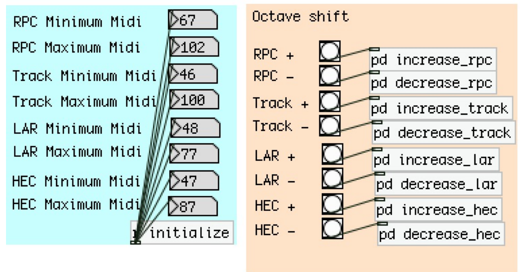

Figure 3: Extract from the Pure Data graphical interface used by certain composers as well as in a real-time DJ'ing performance to control midi ranges of each OSC stream and shift octaves in real-time. Other interface controls include setting musical scale, toggling OSC streams on and off, setting tempo, defining calorimeter beat structures, and setting velocities and durations

developed a graphical interface in PD that maps incoming OSC messages to MIDI notes and gives users control over the tempo, MIDI ranges, and rhythms, among other mapping properties (see Figure 3). In this case, commercially available audio processors like Ableton Live [1] can be used to synthesize sound. The GUI has been built with particular use cases in mind including real-time DJ'ing to data.

Audio Stream

The resulting audio is routed in real-time to an Icecast streaming server using Jack [4] and BUTT [14], at which point it is embedded into a web page using HTML5 audio.

Previous Work

There are several other projects that have explored offline sonification of high-energy physics detector data [15, 6, 9]. For example, in one project, ionisation traces of charged particles from the ALICE experiment at CERN were sonified using a parameter mapping technique whereby pitch, time, amplitude, and timbre of the spatialised audio were varied based on detector data [16]. This project used pre-recorded data for sonification.

There have also been projects exploring sonification of real-time physics data. For example, the cosmic piano project sonifies incoming cosmic radiation detected using a set of Geiger counters [7].

A small number of sonification platforms have also been built with broader applications in mind. For example, the Tidmarsh sonification platform provides composers with access to real-time environmental data from a network of sensors deployed in a former cranberry bog [10], and the SonART platform is built for sound-based data exploration, targeting the exploration of graphical data with sound [18].

The platform described in this paper is the first known effort to broadcast real-time data from a high energy physics experiment to a sonification platform. The web portion of the project, in which the public can listen to real-time audio streams generated from physics data, provides an entirely novel method for engaging the public with live experimental data and is depicted in Figure 4.

Additional Detail on Mapping Paradigms Used

Timing information is very important to the experiment for event reconstruction in part because it is used to extract precise coordinates by drift technology detectors and to veto spurious data. However, timing information has already been used to partially reconstruct the data and so there is little time information left in the data files that this project uses as input. Therefore, more creative or approximate mappings of this data to time must be adopted.

The composers are given the option to stream physics information in time with respect to different detector geometries.

> **Summary of Events Held**
>
> **Composition Workshop** July, 2015. Conducted for roughly 20 composers at the International Conference on Auditory Display (ICAD) in order to test usability of the composition engine in its early stages.
>
> **Performance at Montreux Jazz Festival** July, 2015. A pianist improvised alongside audio generated from ATLAS detector data, and the PD GUI was used to "DJ" to the physics data by adjusting the parameter mapping, tempo, and subset of physics data streamed.
>
> **Website Alpha Launch** November, 2015. Access granted to roughly 50 users. Three real-time audio streams were featured, powered by protons collided by the LHC. Feedback was collected and incorporated in preparation for the website's 2016 public launch.

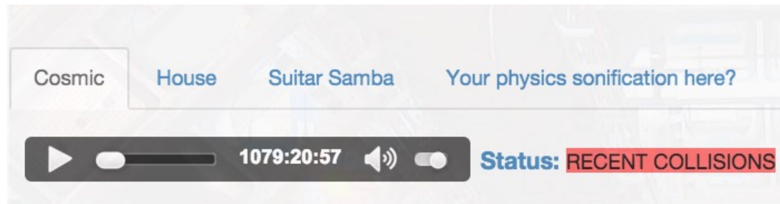

Figure 4: Snippet of website where real-time audio streams are featured. The status indicator changes based on whether real-time data are currently available. The website also includes a set of real-time plots and collision event displays (not pictured) approximately corresponding to data from the same collision event driving the live audio

One available setting approximates the propagation of particles through the detector by sequentially transferring the different detector layer information in time (inner detector, followed by calorimeters, followed by muon spectrometer). Another method supported by the platform involves a linear scanning of the detector as a function of the polar angle assuming a cylindrical coordinate system centred on the middle of the detector.

In all timing mapping approaches used by this platform, event durations are stretched because a collision event itself only lasts a very short period of time: the time for the particles to finish propagating through the detector is only a few tens of nanoseconds and the detector signals themselves can last a few nanoseconds to microseconds.

A set of data cuts are applied based on the expected frequency of the data in the collision event and based on whether the composer has enabled a beat structure. In all cases, energy and momentum cuts are applied such that only the highest energy hits are preserved. In cases where the composer has chosen to impose a beat structure, further adjustments are made: for calorimeter energy deposit data, for which there is a large amount of geometrically dense data, only a single energy deposit nearest to each bin boundary is selected. For sparser data (e.g. track and RPC hits), each data point that passes the initial energy and momentum cuts is associated to the nearest bin.

Evaluation

Evaluation of Public Events Held

Observations from events were used to evaluate the tool's potential to engage audiences. Specifically, we held a workshop at ICAD 2015 for feedback on the composition platform, a performance to experiment with the tool as an artistic medium, and a website alpha launch for feedback from listeners (See "Summary of Events Held" sidebar).

The workshop provided a useful perspective on the composition experience. As detailed in table 1, participants had a wide range of prior composition experience and tool comfort. Verbal feedback indicated great interest in the premise of the tool (at least one attendee of the workshop shared their work on social media) but the installation process was lengthy, especially for a workshop of short duration. Prior experience working in PD predisposed workshop participants to develop custom synthesizers, which shortens the installation time but greatly increases the composition de-

Operating System Used

	Mac	PC	Ubuntu
Users	6	4	1

Music Composition Experience

	A bit	Some	A lot
Users	4	3	4

Average Tool Comfort (Out of 5)

Tool	Average
Ableton Live	1.82
Pure Data	3.36
Python	2.45
Max MSP	3.18

Table 1: Self-reported statistics from participants of a pre-workshop survey (n=11). Tool comfort was reported on an integer scale of 1 to 5, with 5 indicating highest level of comfort

Stage	Delay
Collision→Data File	~minutes
Data Transfer	~seconds
Queuing	~10's of seconds
Sonification	~seconds
Stream to Web	~10's of seconds

Table 2: Order of magnitude of delays introduced at each stage of data transfer from time of collision to time of audio streaming

velopment time. We have since opted to work over longer time periods with individual composers, and we are working on shortening the installation process.

The Montreux Jazz Festival performance drew media attention and interest by students of sonification [12]. The website alpha launch drew in some positive feedback and useful constructive criticism on the website functionality.

Evaluation of time-delays in the "real-time" audio streams
A key feature of the platform is the live nature of the data, and design decisions were made to optimise for real-time behaviour. It has previously been recognised that the real-time nature of ATLAS data visualizations increases public engagement and we hypothesize that the same will be true for real-time audio. Table 2 is a summary of remaining time-delays. The first delay is introduced between the time of particle collision and the time that the corresponding data file is made accessible. Once the data file is available, a small delay is introduced for copying the file to the appropriate machines. Next, the data file is queued for conversion to audio. Currently, the queue has a maximum length of two files, meaning that the worst case queuing time is equal to 2x the event audio duration. Next the input file is read, the data extracted and filtered, and the audio generated. This file access and sonification process is relatively quick. A final audio buffering delay is introduced by the Icecast streaming server as well as browser caching behaviours.

Based on the data in table 2, the sum of all delays can be treated in the following ways:

1. Time of collision → time of web-stream: minutes
2. Time files are accessible → time of web-stream: tens of seconds
3. Time files are accessible → audio generated: seconds

The total delay (1) is a bit slow, but this is largely due to the necessary time to perform partial physics reconstructions, and generating and uploading the resulting data. These tasks are performed by ATLAS and are beyond our control. The audio buffering delay (2) of the order of 10's of seconds is common for existing Internet radio stations [5]. Finally, a few-second processing delay (3) is negligible for the current application since new data files are only made available to us roughly every 25 seconds. The queuing delay is easily minimised by tuning the event audio duration to closely match the time between events entering the queue.

Conclusion and Future Work

We have streamed live ATLAS collision data through a sonification engine that composers can use to generate aesthetically diverse auditory representations of the physics. We have also alpha-launched a corresponding website where the generated audio is streamed in real-time.

The public launch of the web component of the project will be available at quantizer.media.mit.edu in 2016 when ATLAS data taking restarts, after which we anticipate broadening the reach of this tool to composers.

Acknowledgements

We would like to thank the ATLAS experiment for letting us use the data and for its continued support. Thank you to CERN for the invitation to perform at the Montreux Jazz Festival, and to Al Blatter for his performance collaboration. We also wish to acknowledge Media Lab composers Evan Lynch and Akito van Troyer for developing creative custom synthesizers. Thank you to Domenico Vicinanza for providing design feedback in the project's early stages as well as helping to present the project at different venues, including organising the ICAD workshop. Finally, thank you to Felix Socher for sharing his ATLAS data parsing code.

References

[1] 2014. Ableton Live. (2014). Retrieved January 11th, 2016 from http://ableton.com/.

[2] 2015. Pure Data. (2015). Retrieved January 11th, 2016 from http://puredata.info.

[3] 2016. Cycling '74 Max MSP. (2016). Retrieved January 11th, 2016 from https://cycling74.com/products/max/.

[4] 2016. Jack audio connection kit. (2016). Retrieved January 11th, 2016 from http://www.jackaudio.org/.

[5] 2016. The Zone. (2016). Retrieved January 11th, 2016 from http://cjzn.streamon.fm/.

[6] Lily Asquith. 2011. LHCSound: Sonification of the ATLAS Detector Data. (2011). Retrieved January 7, 2016 from http://lhcsound.wordpress.com.

[7] Panos Charitos. 2013. A Cosmic Piano for Alice. (2013).

[8] The ATLAS Collaboration. 2008. The ATLAS Experiment at the CERN Large Hadron Collider. *J. Instrum.* 3 (2008), S08003. 437 p. https://cds.cern.ch/record/1129811 Also published by CERN Geneva in 2010.

[9] Antonella Del Rosso. 2012. Higgs at 3.5 Seconds into the Melody. (2012). http://cds.cern.ch/journal/CERNBulletin/2012/28/News%20Articles/1460881.

[10] Responsive Environments Group. 2015. Tidmarsh Living Observatory. (2015). http://tidmarsh.media.mit.edu.

[11] Thomas Hermann, Andy Hunt, and John G Neuhoff. *The Sonification Handbook*. Berlin,GE.

[12] Matilda Heron. 2015. Physics and music collide at the Montreux Jazz Festival. (2015). Retrieved January 7, 2016 from http://home.cern/about/updates/2015/07/physics-and-music-collide-montreux-jazz-festival.

[13] Gregory Kramer, Bruce Walker, Terri Bonebright, Perry Cook, John H Flowers, Nadine Miner, and John Neuhoff. 2010. Sonification Report: Status of the Field and Research Agenda. (2010).

[14] Daniel Nothen. 2014. butt - broadcast using this tool. (2014). Retrieved January 11th, 2016 from http://butt.sourceforge.net/.

[15] Katharina Vogt and Robert Höldrich. 2010. A Metaphoric Sonification Method-Towards the Acoustic Standard Model of Particle Physics. In *in Proc. of the International Conference on Auditory Display*.

[16] Katharina Vogt, Robert Holdrich, David Pirro, Martin Rumori, Stefan Rossegger, Werner Riegler, and Matevz Tadel. 2010. A Sonic Time Projection Chamber: Sonified Particle Detection at CERN. (2010).

[17] Matthew Wright. 2005. Open Sound Control: an enabling technology for musical networking. *Organised Sound* 10, 03 (2005), 193–200.

[18] Woon Seung Yeo, Jonathan Berger, and Zune Lee. 2004. SonART: A framework for data sonification, visualization and networked multimedia applications. In *Proceedings of the 2004 International Computer Music Conference*. 180–184.