Reviews: Digital and Multimedia Scholarship — Sonic Visualiser

Marc R. Thompson
Finnish Centre of Excellence in Music, Mind, Body and Brain
Senior Researcher, Musicology (University of Jyväskylä)
Docent of Interdisciplinary Music Research (Arts University Helsinki; Sibelius Academy)
marc(dot)thompson(at)jyu(dot)fi

Published in: Journal of the American Musicological Society (2021) 74 (3): 701–714.
https://doi.org/10.1525/jams.2021.74.3.701

Sonic Visualiser: Visualisation, Analysis, and Annotation of Music Audio Recordings. *Centre for Digital Music at Queen Mary (University of London) and the AHRC Researcher Centre for the History and Analysis of Recorded Music (CHARM)*. URL: https://www.sonicvisualiser.org/

What If Music Could Be Seen?

Music recordings have become increasingly important tools for music anal- ysis and many forms of musicological research.1[1] Compared to music analysis using notation, a recording allows the researcher to investigate music from the performer’s perspective. Much can be learned by exploring the expres- sive choices a musician makes to elevate the written score. Additionally, re- cordings are advantageous when analyzing musical styles that might not lend themselves comfortably to standard notational systems (such as some forms of electronic, pop/rock, or folk music).

The analysis of music via recording need not be a highly sophisticated process. Insightful investigation can be carried out using a pencil, paper, and stopwatch to track temporal characteristics such as tempo changes and loudness dynamics. However, the process is greatly enriched by using software that displays audio waveform and spectrum visualizations. Waveforms allow one to observe temporal characteristics at a glance, while spectrum displays the music’s frequency components. Significantly, an analysis of music based on waveform and spectrum enables one to investigate music’s acoustical properties outside of a theoretical context, potentially making it possible to analyze a wider variety of musical forms and styles using the same analytical framework and tools.

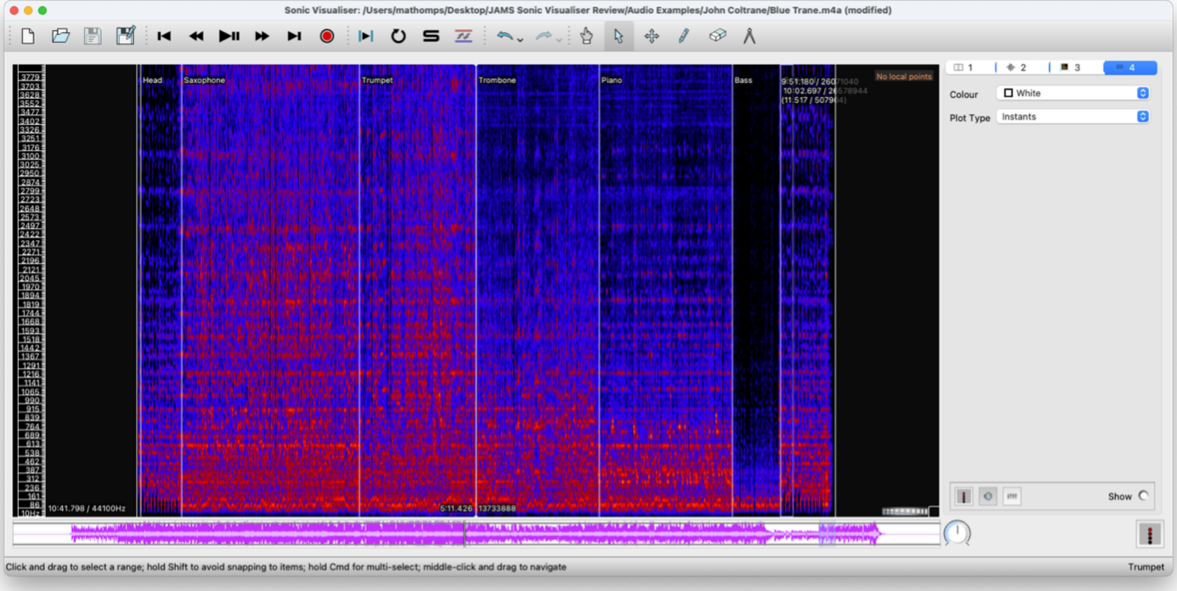
For over a decade, Sonic Visualiser (SV) has provided users with the abil- ity to truly see the music. Designed for musicologists, archivists, and signal- processing researchers alike, SV is a graphical user interface for computational music analyses that would otherwise require experience in digital signal proc- essing and music information research. Upon installing the software, users can load audio files, or make their own recordings, and quickly begin exploring the sonic characteristics and features through an array of visualization and an- notation tools.

SV’s main modes of viewing sound are the waveform and the spectro- gram. Figure 1 shows a spectrogram of the classic John Coltrane recording of “Blue Train.”[2] In the image, annotations have been made to indicate the start of each instrument’s solo. With time displayed on the horizontal axis and frequency on the vertical, the image is read as a heat map whereby the presence of frequencies is represented by color variation and intensity. This variation enables rapid identification of the various sections of the recording. For example, one can easily see the stark contrast in colors between the saxophone solo (displaying a wide range of middle frequencies) and bass solo (narrow range of low frequencies).

[1] See Daniel Leech-Wilkinson, *The Changing Sound of Music: Approaches to Studying Re-corded Musical Performances*, CHARM, 2009, https://www.charm.rhul.ac.uk/studies/chapters/intro.html .

[2] John Coltrane, *Blue Train*, Blue Note BLP 1577, 1958, compact disc.

SV was developed in the mid-2000s by researchers at the Centre for Digital Music at Queen Mary (University of London) in collaboration with the AHRC Researcher Centre for the History and Analysis of Recorded Music (CHARM). A conference proceedings article, which the developers suggest researchers use to reference the software in publications, has been cited hundreds of times and indicates the extent to which SV has been used in music-related research.[3] SV is free to download across main platforms (Windows, Mac, and Ubuntu), and is distributed through the GNU open public license. Compared with other freely available analysis tools, SV is one of the few that is a graphical user interface (GUI). Others are distributed as toolboxes, which require scripting knowledge in MATLAB or Python to be used properly.



Sonic Visualiser: /Users/mathomps/Desktop/JAMS Sonic Visualiser Review/Audio Examples/John Coltrane/Blue Trane.m4a (modified)

Colour: White

Plot Type: Instants

Click and drag to select a range; hold Shift to avoid snapping to items; hold Cmd for multi-select; middle-click and drag to navigate

Figure 1 Screenshot of Sonic Visualiser spectrogram of "Blue Train" by John Coltrane.

So, what is SV exactly? It is many things, but perhaps most of all, it is not a DAW (digital audio workstation), at least not in the traditional sense. Recording is possible, although this is not SV's primary purpose. In fact, once you record something it is impossible to edit the audio. This is a fact that often disappoints and confuses new users. What then is it for? many new users ask. SV is a tool for exploring sound, and for extracting and analyzing music's multidimensional features. The SV website provides excellent back- ground information, videos, tutorials, and links to past and current projects. The material is extensive, especially if one begins to explore the resources listed on the documentation page, which leads to the CHARM website.[4] The tutorial "A Musicologist's Guide to Sonic Visualiser" is particularly instructive for new users,[5] as is Leech-Wilkinson's free online book The Changing Sound

[3] Chris Cannam, Christian Landone, and Mark Sandler, "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files," in *Proceedings of the 18th ACM International Conference on Multimedia, Florence, Italy, October 25–29, 2010*, 1467–68, https://doi.org/10.1145/1873951.1874248 .

[4] https://www.sonicvisualiser.org/documentation.html

[5] Nicholas Cook and Daniel Leech-Wilkinson, "A Musicologist's Guide to Sonic Visualiser," 2009, https://charm.rhul.ac.uk/analysing/p9_1.html

of Music.[6] The unique extended documentation gives the SV user much more than an overview of features—it provides an entire the- oretical framework from which to draw research questions and methods.

SV was designed as a community project. Third parties can contribute by developing so-called Vamp plug-ins, which are software add-ons incorporated into SV. The Vamp download page lists about thirty plug-ins with features such as automatic melody extraction, beat finding, chord analysis, and com- putation of common audio descriptors such as spectral flux (measure of how quickly the frequency spectrum changes) and spectral centroid (a characteristic of sound spectrum analogous to tone brightness). The plug-ins greatly en- hance SV's appeal, although many users will find the native features adequate for most tasks. Interestingly, the plug-ins page has not been updated for some time, which might indicate that development has slowed in recent years. SV itself continues to be updated regularly, albeit with incremental improve- ments. At the time of writing, Sonic Visualiser v4.3, released on January 18, 2021, is the most recent version.

This review begins with a survey of SV's main features, followed by demonstrations of possible applications for musicological research and teaching. The last section discusses how SV stacks up against other computational music analysis tools.

SV's Main Features

SV's interface centers on the concept of panes and layers. A pane is a section of the screen that displays graphical information such as the audio waveform or its spectrogram. For new users with experience of regular DAWs the pane might seem like an audio track. But a pane is different from a track in that multiple panes can show information related to the same audio file. When an audio file is loaded, a Waveform Pane is created displaying the audio file in waveform representation. From the Pane menu, the user can create a new pane to display a spectrogram or spectrum. As mentioned above, the spec- trogram shows the audio file in its entirety with frequency on the vertical axis and time on the horizontal. The spectrum is a frequency analysis for the audio file's current point in time. When playback is activated, the spectrum changes to match the changes occurring in the audio file.

The layer allows the user to conceptually stack pieces of information onto a single pane. There are numerous layers to choose from for annotating the panes, such as time instants, spectrograms, text, boxes, and images. There is also the Note Layer, which enables the user to add pitched sounds onto the spectrogram. The notes can be added to match the location of frequencies on the vertical axis. This is very useful for analyzing recordings in which the singer is slightly off key, or for annotating sound recordings with nondistinct pitches, like those of songbirds.

Each layer type has its set of display properties, which can be altered in the panel that shows up on the right side of the window when a new layer is cre- ated. The user navigates the layers by clicking on labeled tabs at the top of the Display Properties box. The editable properties include color schemes for all types of layers, and for spectrogram layers users can change window and overlap sizes. The spectrogram is calculated by segmenting the audio file into windows and performing a series of fast Fourier transforms to obtain the frequency components present in the audio stream, and the overlap refers to the number of overlapped samples between windows. These properties might be confusing for nontechnical users but are adequately explained in the documentation. In any case, a deep understanding of fast Fourier trans- form is not required for the use of SV. As far as the nontechnical user is concerned, these parameters function as image modifiers, and allow the user to edit the spectrogram's sharpness or smoothness.

Keyboard shortcuts are intuitive and easy to learn. New users are advised to memorize them as they enable quick navigation throughout the interface, which requires quite a bit of switching between modes (e.g., scrolling, selection, editing). Zooming in and out of the image can be done with scrolling wheels or via mouse/trackpad gestures.

[6] Danel Leech-Wilkinson, *Changing Sound of Music*, https://www.charm.rhul.ac.uk/studies%20/chapters/intro.html

Lastly, a very simple yet valuable feature is the variable playback speed, which by default will not alter the sound's pitch. This alone would make SV a useful tool for someone simply needing to make transcriptions of recorded music or the spoken word.

On a personal note, I have used SV for about a decade and have found it to be a valuable resource for small and large projects alike. Some personal pet peeves nonetheless persist. For one, SV can become confusing to use when multiple audio files are loaded into a single session. In a regular DAW, when the user imports audio to a new track, there is an option to name the new track. Strangely, SV lacks this kind of option. Although the software automatically assigns a new color for each new waveform loaded, this simply does not replace naming the waveform. This became evident when I once had sixteen performances of the same piece open in a single session. Through experience, I have realized that a work-around is to add a Text Layer upon loading new audio and annotating the name of the file onto the audio display.

Another issue pertains to navigating between the layers. It can be difficult to keep track of which layer is active. Each layer is identified by number (signi- fying the order in which it was created) and icon (signifying the type of layer) on the right side of the screen. A layer is activated when one clicks on its icon, but it takes some time to remember what each icon refers to. So the user is left having to click on a few of the icons before accessing the desired layer. Inter- estingly, the default numbers given to layers can be changed to a name of the user's choice (not the case for waveforms). Finally, labeling time instants can be confusing for new users. Most of the edits for timing instants are found in the Edit menu. But it would surely be intuitive to have access to these proper- ties (such as the ability to rename time instants) through the Time Instant's Property Box on the right side of the main window.

Minor gripes aside, SV remains a great tool. Ease of use is something that comes with practice and new users should not be discouraged. The next sec- tion examines how SV has been used in music-related research.

SV for Research

The research field most often affiliated with SV is Music Information Re- search/Retrieval (MIR). This wide-ranging field focuses on the computational analysis of music and sound, including the design of algorithms for extracting meaningful features from audio signals, and then indexing those features for search and retrieval schemes.[7] Such features can include perceivable and non- perceivable dimensions of music. An example of a perceivable dimension would be observing the time onsets of notes to determine how a performer executes a tempo rubato section in Chopin's Nocturne op. 9, no. 2. This can be done using a stopwatch. However, if one wished to analyze how the choice of tem- po varies across hundreds of recordings of the piece, one could program a computer to construct tempo curves for each recording, and then calculate descriptive statistics (e.g., mean, standard deviation) to form an idea of the most likely approach to the piece. Examples of nonperceivable features might be those characteristics that describe music's frequency spectrum. One spectral feature, for example, is centroid, which indicates the frequency at which a spectrum's energy is centered. Unlike tempo, spectral features require computational methods in order to be calculated. However, many of these features do have perceptual analogues—centroid is commonly associated with a sound's brightness. MIR is interested in cataloging musical features on a grand scale and building recommendation systems. Programs like Apple Music and Spotify make use of MIR to recommend music to listeners on the basis of their previ- ous listening patterns, and popular apps like Shazam use MIR to identify songs playing close to the user's phone.

[7] For an excellent overview, see Markus Schedl, Emilia Gómez, and Julián Urbano, "Music Information Retrieval: Recent Developments and Applications," *Foundations and Trends in Information Retrieval* 8, nos. 2–3 (September 12, 2014): 127–261, https://doi.org/10.1561/1500000042

As mentioned above, SV has been cited in hundreds of journal articles, conference proceedings, and books since 2010. I have picked out a number of publications from the literature to see how SV has been used, and notice that many of the studies merely mention it in passing. This is often the case in articles introducing analytical toolboxes.[8] When articles feature SV as part of the reported research, it is mostly used for event onset detection at the data preprocessing phase.[9] Onset detection is carried out either manually by the researchers and/or assistants, or automatically using various Vamp plug- ins that can detect onsets and/or beats.

In some cases, SV was part of the experimental design, and human anno- tators were asked to segment musical excerpts to establish a ground truth that could be incorporated into an analysis.[10] SV is well suited to this kind of data collection because it is relatively easy to learn to use for manual event detection and annotation, such that training the participants to complete the task should not take too much time. Onset detection can be achieved by adding time points in the Edit mode or by playing the track and tapping in time points on the computer keyboard as the audio plays. The spectrogram visualization is ideal for checking the accuracy of one's tapping as the onsets of notes are clearly visible, especially if the recording is of a single instrument. Altering the audio playback speed also helps the user ensure that the onset appears at the correct spot.[11]

In my own research, SV has been employed in music performance studies where musicians performed the same piece multiple times using different performance intentions (deadpan, normal, exaggerated).[12] In these studies, SV was used to annotate note onsets in each performance. The onsets were exported from SV into MATLAB, where they were used as timing models to temporally align the recording's corresponding motion capture data.

Recently, however, my primary use of SV has been as a pedagogical tool for teaching, particularly in undergraduate Musicology courses. The next section provides a brief overview of how SV can be used in teaching situations.

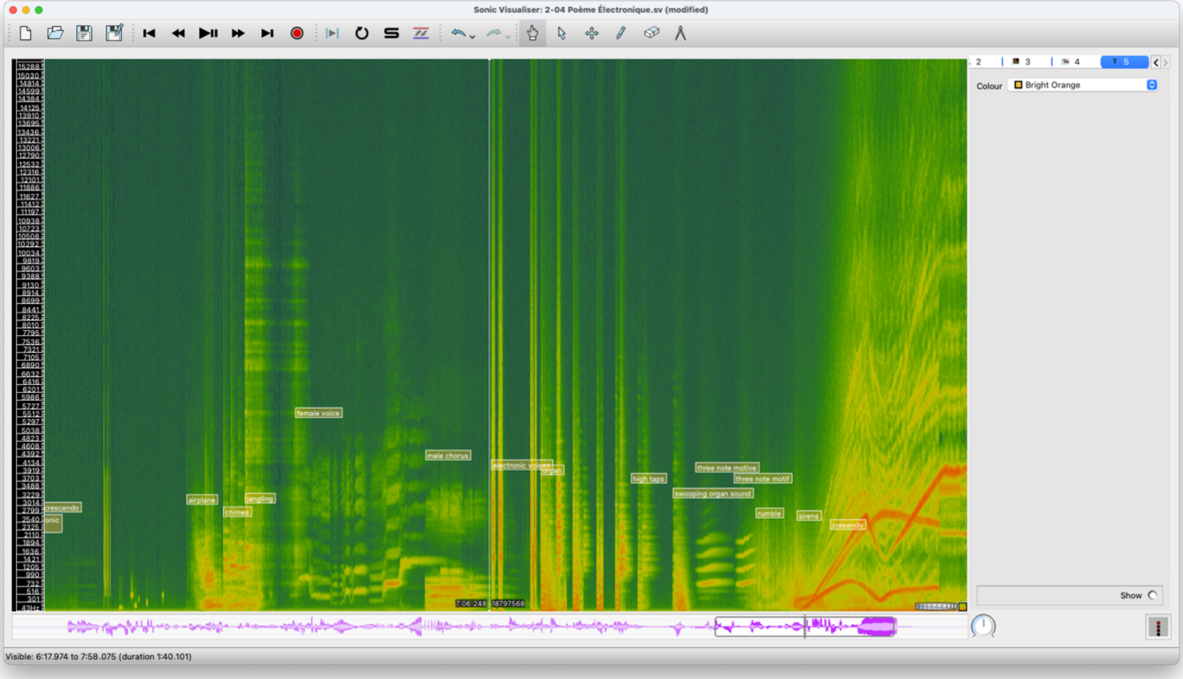
[8] For example, Brian McFee, Colin Raffel, Dawen Liang, Daniel P. W. Ellis, Matt McVi- car, Eric Battenberg, and Oriol Nieto, "Librosa: Audio and Music Signal Analysis in Python," in *Proceedings of the 14th Python in Science Conference*, 2015, 18–24, https://doi.org/10.25080/Majora-7b98e3ed-003 .

[9] For example, Dorottya Fabian, "Analyzing Difference in Recordings of Bach's Violin Solos with a Lead from Gilles Deleuze," *Music Theory Online* 23, no. 4 (2017), https://www .mtosmt.org/ojs/index.php/mto/article/view/132; Meghan Goodchild, Jonathan Wild, and Stephen McAdams, "Exploring Emotional Responses to Orchestral Gestures," *Musicae scientiae* 23, no. 1 (2019): 25–49, https://doi.org/10.1177/1029864917704033 .

[10] For example, Martin Hartmann, Olivier Lartillot, and Petri Toiviainen, "Musical Feature and Novelty Curve Characterizations as Predictors of Segmentation Accuracy," in *Proceedings of the 14th Sound and Music Computing Conference*, 2017, ed. T. Lokki, J. Pätynen, and V. Välimäki, 365–72, https://jyx.jyu.fi/bitstream/handle/123456789/54912/hartmannlar tillottoiviainenmusicalfeature.pdf?sequence=1&isAllowed=y; Thassilo Gadermaier and Gerhard Widmer, "A Study of Annotation and Alignment Accuracy for Performance Comparison in Complex Orchestral Music," in Proceedings of the 20th International Society for Music Informa- tion Retrieval Conference, Delft, The Netherlands, November 4–8, 2019, 769–75, http://arxiv.org/abs/1910.07394

[11] SeeMichaelRector,"HistoricalTrendsinExpressiveTimingStrategies:Chopin'sEtude, Op. 25 No. 1," *Empirical Musicology Review* 15, nos. 3–4 (2020): 176–201, https://doi.org/10.18061/emr.v15i3-4.7338

[12] Marc R. Thompson and Geoff Luck, "Exploring Relationships between Pianists' Body Movements, Their Expressive Intentions, and Structural Elements of the Music," *Musicae scientiae* 16, no. 1 (2012): 19–40; Jonna K. Vuoskoski, Marc R. Thompson, Eric F. Clarke, and Charles Spence, "Crossmodal Interactions in the Perception of Expressivity in Musical Perfor- mance," *Attention, Perception, and Psychophysics* 76, no. 2 (February 2014): 591–604.



Sonic Visualiser: 2-04 Poème Électronique.sv (modified)

Colour: Bright Orange

Visible: 6:17.974 to 7:58.075 (duration 1:40.101)

Figure 2 Screenshot of Sonic Visualiser spectrogram of Poème électronique by Edgard Varèse.

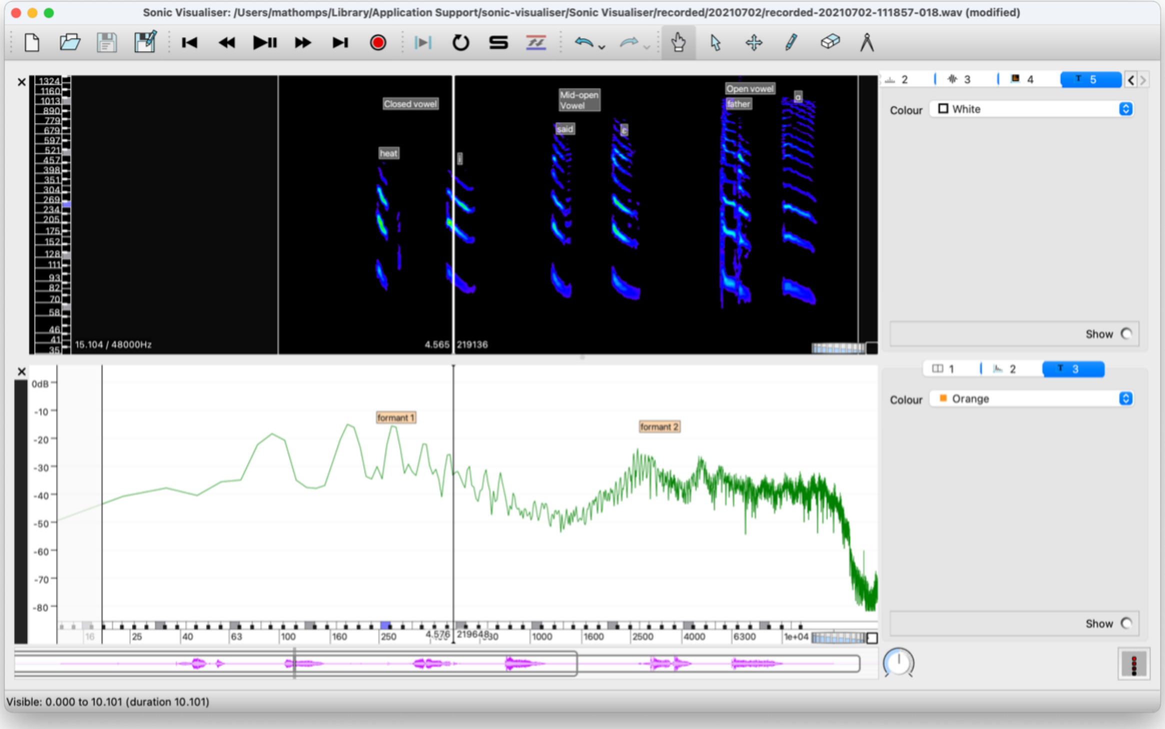
a phenomenological analysis by Lawrence Ferrara.[13] The spectrogram is striking, as it displays the variety in dynamics and timbres within the piece. Natural sounds of human voices and instruments clearly show which harmonics are present in the signal, while percussive noisy sounds contain non- harmonic partials that span the full range of frequencies. The end of the piece shows a synthetic crescendo that not only rises in loudness but sweeps across the frequency spectrum.

A first exercise for students to complete in class introduces the concept of vowel formants and lets them practice recording their own voices into SV. Figure 3 demonstrates what an outcome of the exercise might look like. The students are given a list of words containing either closed, middle-opened, or opened vowel sounds (e.g., "heat," "said," and "father," respectively). The first step is to record the words, followed by each word's main vowel sound, into SV. It is at this point that students realize that although audio can be recorded into SV, the recorded audio cannot be edited. This gives them ample opportunity to practice recording audio, as it will take several takes to get the words right. Once the words are recorded, students can see the audio display of their recording. To better visualize the recording, they create a new layer called a Melodic Range Spectrogram. The visualization settings for this layer allow students to change the spectrogram's colors and sharpen the image by fiddling with the window, hop, and bin parameters on the right side of the screen. As the image sharpens, students become aware that the recorded words contain different frequency components because the vowels within the words have unique formants. Students then add a sec- ond pane to display the spectrum of their recording. The Spectrum Pane dis- plays how the recording's frequency components change as the recording plays. Finally, students practice annotation by adding a Text Layer, which al- lows them to label different portions of their recording. Once the recordings are made, students can sit with a partner and compare results, to see if the words' formants appear at similar frequencies. For instance, voices with low- er fundamental frequencies uttering the vowel [i], as in "heat," will see a for- mant crescent at around 240

[13] Lawrence Ferrara, "Phenomenology as a Tool for Musical Analysis," *Musical Quarterly* 70, no. 3 (Summer 1984): 355–73.

Hz (formant 1) and 2400 Hz (formant 2). The exercise can then move on to a group discussion about singers' formants.

A more advanced exercise combines learning to use SV tools with con- ducting a small music performance analysis. Students are given recordings of four renditions of a short piece (see figure 4). The piece was composed as part of an emotional validation study in which participants rated musical excerpts for self-perceived happiness, sadness, threat, and peacefulness.[14]


Sonic Visualiser: /Users/mathomps/Library/Application Support/sonic-visualiser/Sonic Visualiser/recorded/20210702/recorded-20210702-111857-018.wav (modified)

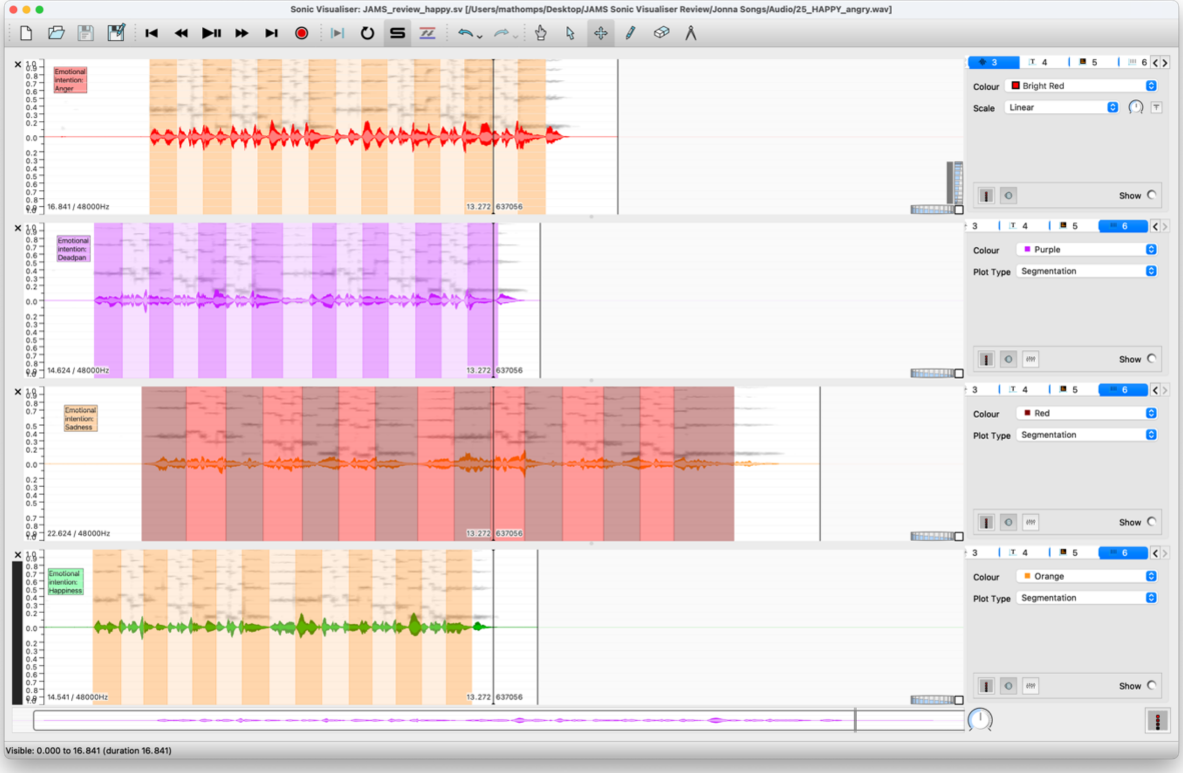
Colour □ White

Colour ■ Orange

Visible: 0.000 to 10.101 (duration 10.101)

Figure 3 Screenshot of Sonic Visualiser vowel formant class exercise.

[14] Sandrine Vieillard, Isabelle Peretz, Nathalie Gosselin, Stéphanie Khalfa, Lise Gagnon, and Bernard Bouchard, "Happy, Sad, Scary and Peaceful Musical Excerpts for Research on Emotions," *Cognition and Emotion* 22, no. 4 (2008): 720–52.



Sonic Visualiser: JAMS_review_happy.sv [/Users/mathomps/Desktop/JAMS Sonic Visualiser Review/Jonna Songs/Audio/25_HAPPY_angry.wav]

Colour: Bright Red
Scale: Linear

Colour: Purple
Plot Type: Segmentation

Colour: Red
Plot Type: Segmentation

Colour: Orange
Plot Type: Segmentation

Visible: 0.000 to 16.841 (duration 16.841)

Emotional intention: Anger
16.841 / 48000Hz
13.272 637056
Emotional intention: Deadpan
14.624 / 48000Hz
Emotional intention: Sadness
22.624 / 48000Hz
Emotional intention: Happiness
14.541 / 48000Hz
3
4
5
6
Show

Figure 4 Screenshot of Sonic Visualiser performance analysis class exercise.

The piece used in the class exercise had been the most highly rated for per- ceived happiness. The four renditions given to students were recordings from a project on which I have collaborated, in which a violinist performed the "happiness" piece employing different emotional intentions: angry, sad, happy, and deadpan. The students are not informed that the jaunty melody (in 2/4, with dotted rhythmic figures and in a major key) has been validated as a happy piece. The students must first annotate the note onsets of each performance, and then investigate the spectrograms. They answer questions about the ways in which the performances differ in terms of dynamics, timing, and timbre. They are then asked to rate each performance for happiness, sadness, and anger. The exercise emulates a music perception study that investigates how musicians' expressive intentions impact listeners' reactions to music. In giving this assignment several times, I have observed that the composition's construction acts as a main influence on perception. For all the renditions, ratings for perceived happiness are generally higher than ratings for perceived sadness or anger. The students can then discuss their results in groups.

Future Directions of Musicology and Music Analysis

The above teaching examples demonstrate how SV can be used in a class- room situation where the goal is to learn about musicological research, as op- posed to learning about MIR. For this purpose, SV is an excellent solution, despite its age. For conducting MIR research, better, more powerful tools ex- ist. One well-known tool developed by colleagues within my research group is the MIRToolbox, which recently celebrated its first ten years of development.[15] Like SV, MIRToolbox can detect pitch information, perform onset detection, and conduct dynamics

[15] Olivier Lartillot and Petri Toiviainen, "A Matlab Toolbox for Musical Feature Extrac- tion from Audio," in *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07),* Bordeaux, France, September 10–15, 2007, 237–44, https://dafx.labri.fr/main/papers/p237.pdf

and timbre analysis. The difference is that MIRToolbox works in MATLAB and is intended to analyze large data sets at once, whereas SV generally looks at one audio file at a time. MIRToolbox requires the knowledge of scripting in order to be used effectively, which is not difficult to learn but has a steeper learning curve than starting on SV. It should be noted that there are Python packages that allow programmers to use SV for large data sets as well, but to my knowledge these have not had the same impact as MIRToolbox.

SV's advantage over a tool like MIRToolbox has been its focus on serving musicologists, MIRToolbox being more useful for researchers interested in topics like music cognition. Future projects might combine these interests, however. MIRToolbox's main author is working on a new project called MIRAGE, which is a comprehensive AI-based system for music analysis.[16] In June 2021, a first symposium dedicated to MIRAGE took place, at which its impressive abilities to conduct real-time analysis of musical features and produce musical visualizations specific to the music were demonstrated. The visualizations are intuitive and are aimed at nontechnical researchers.

Conclusion

To sum up, although SV was developed over a decade ago and has gone through only minor updates, it continues to be a useful tool for music research and teaching. Its greatest advantages are that (i) it is free to download and use and (ii) it has a relatively smooth learning curve for people with non- technical backgrounds. For research, its greatest asset has been the ease with which one can view spectrograms and annotate time instants. This feature has been used for tasks in music perception studies (for example, participants annotate performances into meaningful segments), or as a data preprocess- ing step for extracting event onsets for statistical analyses. In teaching situations, SV allows students to learn psychoacoustics, and acts as an entry point for learning basic concepts of MIR (such as spectral feature extraction). The conceptual learning that SV facilitates is a stepping-stone for more technical courses focused on MIR.

Lastly, SV is a means to consider music analysis frameworks that go beyond score-based music. It is particularly appropriate for analyzing music as a living, embodied phenomenon. Suddenly, variations between performances can be identified, as one can access the expressive parameters of timing, dynamics, articulation, and timbre.[17] To put it plainly, SV is powerful because it allows one to see music (pun intended) as a truly multidimensional phenomenon.

[16] Olivier Lartillot, "MIRAGE—A Comprehensive AI-Based System for Advanced Music Analysis," RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, 2021, https://www.uio.no/ritmo/english/projects/mirage/index.html

[17] See John Rink, Neta Spiro, and Nicholas Gold, "Motive, Gesture and the Analysis of Performance," in *New Perspectives on Music and Gesture*, ed. Anthony Gritten and Elaine King (Milton Park: Routledge, 2016), 293–318.