L. Fyfe, D. Bedoya and E. Chew, "Annotation and Analysis of Recorded Piano Performances on the Web"
J. Audio Eng. Soc., vol. 70, no. 11, pp. 962–978, (2022 November).
DOI: https://doi.org/10.17743/jaes.2022.0057

Annotation and Analysis of Recorded Piano Performances on the Web

LAWRENCE FYFE,[1] DANIEL BEDOYA,[1] AND ELAINE CHEW[1, 2]
(lawrence.fyfe@ircam.fr) (daniel.bedoya@ircam.fr) (elaine.chew@ircam.fr)

[1]STMS Laboratoire (UMR9912) – CNRS, IRCAM, Sorbonne Université, Ministère de la Culture, Paris 75004, France
[2]Department of Engineering, King's College London, London WC2R 2LS, United Kingdom

Advancing knowledge and understanding about performed music is hampered by a lack of annotation data for music expressivity. To enable large-scale collection of annotations and explorations of performed music, the authors have created a workflow that is enabled by CosmoNote, a Web-based citizen science tool for annotating musical structures created by the performer and experienced by the listener during expressive piano performances. To enable annotation tasks with CosmoNote, annotators can listen to the recorded performances and view synchronized music visualization layers including the audio waveform, recorded notes, extracted audio features such as loudness and tempo, and score features such as harmonic tension. Annotators have the ability to zoom into specific parts of a performance and see visuals and listen to the audio from just that part. The annotation of performed musical structures is done by using boundaries of varying strengths, regions, comments, and note groups. By analyzing the annotations collected with CosmoNote, performance decisions will be able to be modeled and analyzed in order to aid in the understanding of expressive choices in musical performances and discover the vocabulary of performed musical structures.

0 INTRODUCTION

In the course of performing notated compositions, performers add their own expressive manipulations that may not be scripted in the score and, if transcribed back into music notation, could be shown to be far from the written score [1]. These expressive manipulations, including variations in timing, loudness, articulation, and timbre [2], convey groupings and prominence of notes, forms of performed structures, to listeners [3]. These structures may be conceived first in the mind of the performer, developed on the fly as they make sense of the music, or made while performing the piece of music.

In any of these cases, the structures are then eventually transmitted to the minds of listeners. Although these structures may be perceptible by listeners, consciously or unconsciously, they may, however, be difficult to discern with automated analysis. For example, whether an accented note marks the beginning or the end of a note grouping may be ambiguous with automated analysis but not to a human listener. In the absence of reliable automated analysis, citizen science was used to study performed music structures because they are created in recorded music performances and in the minds of listeners.

The Computational Shaping and Modeling of Musical Structures (COSMOS) project [4] was created to study such performed musical structures in performances of classical piano music. The video (in French), "Le piano virtuose" [5], explains this research in general terms. To enable this research, a workflow that allowed citizen scientists to annotate those perceived structures and a software tool that enabled that workflow were needed. The workflow needed to start with presenting the recorded piano performances. Citizen scientists would then listen to the recorded performances and see the notes played and see expressive features extracted from the recorded audio. A variety of annotation types would be provided including the marking of boundaries, regions, comments, and groups of notes. Finally, annotations collected from citizen scientists, ranging from musical novices to professional musicians, would then form the basis for studying how performance shapes or re-shapes perceived musical structures.

The research question that is addressed in this paper is: how can a workflow for presenting recorded piano performances, creating annotations of those performances, and then analyzing those annotations be created? In answering this question, the authors developed CosmoNote [6], a Web-based citizen science tool for visualizing and annotating

expressive piano performances, to embody the workflow that was created. A previous CosmoNote paper [7] discussed the basic design of the software. This paper goes beyond a description of the software to describe the CosmoNote workflow in detail while including more detailed descriptions of the software functionality and new features and significant changes since the publication of the previous paper.

The next section (SEC. 1) describes related work and approaches to music annotation. The next four sections are organized according to the CosmoNote workflow. First, performance data are obtained for inclusion in CosmoNote (SEC. 2). Second, the performance data are presented to annotators as both audio and visuals (SEC. 3). Third, based on the performance data, annotators create their annotations (SEC. 4). Fourth, the annotations are collected and analyzed as shown in two pilot studies (SEC. 5). To conclude, the contributions in CosmoNote, the results of the two pilot studies, and some future work are summarized (SEC. 6).

1 RELATED WORK

In this section, for clarity, the related work is provided as included in the authors' earlier CosmoNote paper [7] but with some additional references. As in the earlier paper, rather than providing a comprehensive list of all the audio annotation projects, only the music annotation tools most relevant to the authors' own work will be described. Annotation applications tend to involve either human or automated annotations, and sometimes a combination of both. Because citizen science is a core part of the workflow, the authors are only looking at human or combination human-computer applications, and, because CosmoNote was envisioned as a Web-based project from the beginning, the projects reviewed are divided into non–Web-based projects and, most relevant to the CosmoNote workflow, Web-based projects.

1.1 Non–Web-Based

Tzanetakis and Cook [8] wanted to determine whether a computer-assisted human temporal segmentation annotation task benefited from automated segmentation. As part of a pilot user study, they presented users with an annotation application, based on their MARSYAS software framework [9], and asked them to mark temporal boundaries based on what they called sound "texture" or changes in instrument or speaker, etc. Similarly, Amatriain et al. developed, using their CLAM audio framework [10], the CLAM Annotator [11], a combination human-computer annotation application with which users could edit descriptors that were created automatically.

Notess and Swan [12] developed Timeliner, a human-based annotation application, to allow users of a digital music library to create annotations for audio files in the library. Annotations included marking time regions and specific time points of interest. Text labels could be created for each of these annotations and the playback of the audio files could be tied to the annotations.

Herrera et al. [13] developed the Music Content Semantic Annotator (MUCOSA) project to enable a variety of annotation workflows. MUCOSA was built on top of WaveSurfer [14], a speech annotation tool that allowed for plugins. Annotations like structure markings were shown as squares in a panel vertically stacked below separate spectrogram and waveform visualizations. Li et al. [15] wanted to establish a set of ground truth data for the segmentation of songs, i.e., by annotating regions for chorus and non-chorus parts of the songs. To do that, they built an annotation system on top of the Audacity audio editor [16] by adding separate tracks below Audacity's normal waveform visualization track. The annotation tracks contained region markers and labels for regions, and, like the MUCOSA annotation system, the waveform visualization and the annotations were stacked vertically.

Cannam et al. [17] developed Sonic Visualiser to be "the first program you reach for when want to study a musical recording rather than simply listen to it." In addition to featuring different visualization layers like waveforms, spectrograms, and notes, Sonic Visualiser allowed annotations to be placed directly over the visualization layers, saving screen space. Of all of the non-Web tools that were examined, Sonic Visualiser had the set of features that matched most closely with the overall workflow.

1.2 Web-Based

The projects described in the previous section had interesting features but were not Web tools, making them difficult to use in the citizen science–based workflow. So, Web-based annotation tools and the capabilities they provided for music annotation were specifically looked at.

Cartwright et al. [18], as part of their CrowdCurio project, wanted users to annotate soundscapes of varying complexity based on waveform visualizations, spectrograms, or no visualization at all. To do that, they created Audio-annotator [19], built on top of the WaveSurfer.js [20] waveform visualization library. Audio-annotator allowed users to create annotations by selecting a sound region. Annotations could be edited, and users could listen to the sounds from the selected regions of their annotations. To allow users to annotate radio recordings with the goal of detecting music in radio broadcasts, Melendez-Catalan et al. [21] created BAT [22], another WaveSurfer.js-based Web annotation tool. Annotators were asked to distinguish between music and speech in the recordings by selecting time regions and then identifying them as music or speech.

Wang et al. [23] used the CAQE toolkit [24] to create a Web interface for crowd-sourcing a music segmentation task in which they asked annotators on Amazon's Mechanical Turk to listen for changes between one part of a song and another and to mark a boundary there. Annotators would listen to 20-s clips from the songs and mark boundaries by adjusting a slider that could be positioned anywhere in the time frame of the clip provided. Other than the slider and an audio progress bar, both of which only appear after the first listen, there is no other visual indicator because the authors wanted annotators to focus on what they heard in the clip.

1.3 Relevance

Although each of the projects described above offered some useful features, they did not offer enough features to be relevant for the workflow. In particular, all of the Web-based projects displayed their sound or music selections as waveforms and/or spectrograms, failing to include the note-based information that is crucial for a more fine-grained analysis of performed structures. The other projects were also limited in their annotation types with all of the projects offering region selections and only some offering boundaries and/or comments. For this workflow, all of these annotation options, including boundaries, regions, comments, and, along with the note visuals, the ability to select groups of notes, were needed. The need for a more-customized Web tool for annotations with a particular workflow led to the development of the authors' own citizen science annotation tool, CosmoNote.

2 OBTAINING PERFORMANCE DATA

The CosmoNote workflow starts with obtaining performance data for the citizen scientists to annotate. Piano performance data are obtained in the form of both audio and MIDI recordings. These recordings are obtained via two paths: from existing recordings of audio and MIDI or from recordings made by the CosmoNote team. The recordings made by the CosmoNote team have the advantage of having synchronized audio and MIDI data, something that is not necessarily true for other recordings.

2.1 Recording the Performances

For recordings made by the CosmoNote team, the authors use a reproducing piano, the Bösendorfer Enspire [25], that is capable of recording performances as MIDI data, producing high-quality acoustic sounds, and enabling fine control of musical expression. During the recording process, the MIDI data are streamed from the piano to the computer simultaneously with the audio, recorded via microphones, ensuring proper synchronization. Fig. 1 shows, via CosmoNote, a performance of Christian Petzold's Minuet in G minor (originally attributed to J.S. Bach because of its inclusion in Bach's Little Notebook for Anna Magdalena Bach). The performance, which was by co-author Elaine Chew and recorded on the Bösendorfer piano, will be, unless otherwise noted, used throughout the remainder of the article to illustrate the different aspects of CosmoNote.

2.2 Preparing the Audio, Note, and Pedal Data

Audio files, whether recorded by the CosmoNote team or obtained from another source, start as uncompressed WAV or AIFF. To make the audio faster to download, the audio files were initially compressed using both FLAC [26] and OPUS [27]. FLAC files proved to be quite large for fast downloading, and OPUS was not completely supported by all of the browsers tested. In the end, the authors decided to use MP3 for compressed audio because it provided a good trade-off between fast downloading and audio quality and it was supported by all of the browsers tested.

Note data in CosmoNote come from recorded MIDI files, either recorded by the CosmoNote team with the recording piano or from existing recordings. In MIDI, individual notes are split into pairs of note-on and note-off events. To visualize the notes in CosmoNote, each note needs to be a single event with a start and end time. To that end, a custom Python script that uses the Mido MIDI library [28] is used to convert the MIDI data to single-event note data.

Pedal data from the sustain, soft, and sostenuto pedals are also taken from MIDI files using the same script. The pedal data from MIDI are a series of control change events rather than pairs on on-off events as with the note events. For recordings from CosmoNote's recording piano, the extent of pedal depression is recorded with an 8-bit resolution of 0–127. For certain pre-existing MIDI files (not recorded with CosmoNote's recording piano), the control change messages for the pedals are only on-off events because the extent of pedal depression was not recorded.

2.3 Computing Feature Data

In CosmoNote, feature data are defined as the data computed from audio files, MIDI files, or the score for a given performance. Loudness data (in sones) are computed from the audio file, using a custom Python script ported from the MATLAB MA Toolbox [29], as a global representation of the perceived intensity of the notes. It corresponds roughly to velocity data shown for the notes, although the perceived loudness is influenced by the number of notes played, their pitches, and how quickly the key is depressed for individual notes as with velocity. Loudness data can be used to, for example, locate a group of notes highlighted by the performer to make a melody more salient than the contextual background or a note more prominent than its neighbors.

Tempo is computed using timestamps of the onset of each beat—in which the beats are located by alignment to a MusicXML score using an alignment tool from Nakamura et al. [30]—throughout a performance and is computed as the inverse of the time between beats. The tempo can also be computed from manual beat annotations using the same approach. As an example of using tempo data, the parts in which the curve shows a steep descent followed by an ascent could help in the identification of phrase boundaries [31] or tipping points [32] (devices for heightening suspense in expressive performance).

Harmonic tension is computed from the score using the XmlTensionVisualiser [33] tool and has three dimensions [34]: cloud diameter representing dissonance, cloud momentum representing the rate of chord changes, and tensile strain representing the distance from the global key. Harmonic tension, for example, cloud momentum, could be used by musically trained annotators to visualize large movements in tonality in which non-diatonic chords are used.

2.4 Storing the Data

Once all of the performance data are collected, it is loaded into a CouchDB [35] database server using custom Python scripts. CouchDB was chosen because it works well

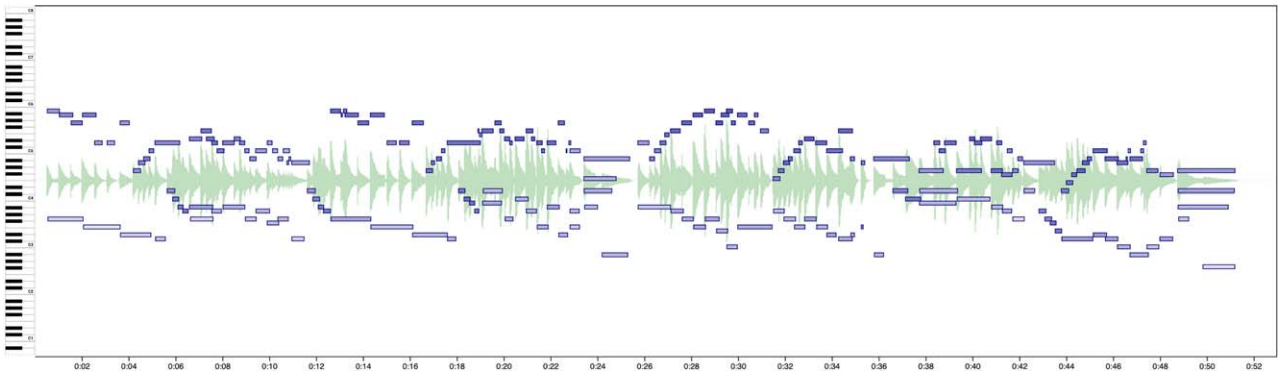

Fig. 1. CosmoNote presenting a performance by co-author Elaine Chew of the Minuet in G minor from the Little Notebook for Anna Magdalena Bach as recorded on a Bösendorfer Enspire reproducing piano.

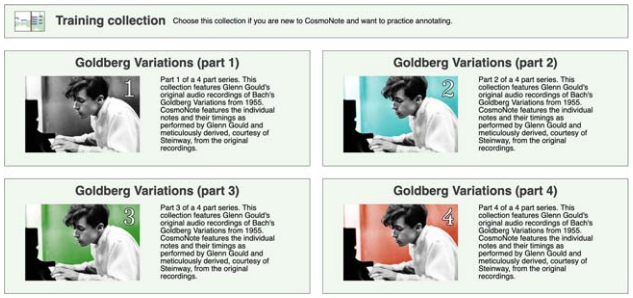

Fig. 2. The CosmoNote collections page allows annotators to select a collection to annotate including an always-available training collection for new annotators.

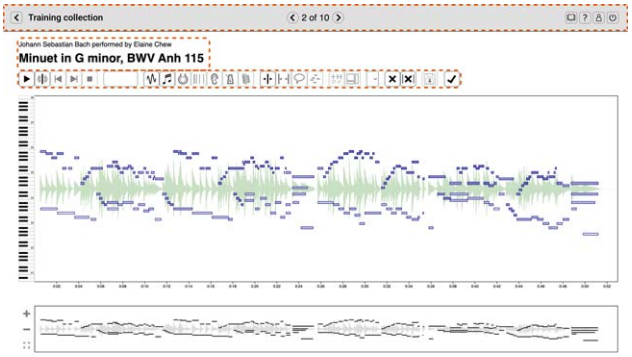

Fig. 3. A complete view of CosmoNote's interface with three elements highlighted by dashed rectangles: 1) the navigation bar (top), 2) information about the performance (middle), and 3) controls (bottom).

with Web applications by supporting HTTP for transactions (with no database drivers needed) and by using JSON [36], a data format widely supported by a variety of tools and languages, for data storage.

3 PRESENTING THE PERFORMANCES

Next in the CosmoNote workflow, after the CosmoNote performance data are collected and stored, the data are presented to annotators as both audio and visuals via a client-side Web application. CosmoNote audio uses the Web Audio API [37] and CosmoNote visuals use D3 [38], a highly-customizable, Scalable Vector Graphics–based [39] visualization and interaction library. CosmoNote clients get the performance data (as JSON) from the CouchDB server using PouchDB [40].

Performances are presented to annotators either as whole pieces of music or fragments of pieces. In either case, all of the performances are grouped into collections, whereby collections can be based around a performer, composer, or some other theme. For example, the first sets of collections released in CosmoNote, starting in December 2021, were built around the 1955 recordings of Bach's Goldberg Variations by Glenn Gould and a collection of simple practice pieces for training new annotators. Fig. 2 shows the collection selection page featuring four collections of performances by Glenn Gould along with the training collection that always appears at the top of the page. The performance of the Minuet in G minor by co-author Elaine Chew, used as an example throughout this article, is included in the training collection.

Annotators select a collection before beginning to annotate its performances. Once they have selected a collection, annotators can access the performances in that collection with each recorded performance having its own page. A gray bar at the top of the page shows the name of the current collection and the current performance number (out of the total) along with forward and backward buttons for navigating the performances in the collection. Performances in a collection have a set order, but there is an option to present the performances in a given collection to each annotator in a randomized order with the order being stored per annotator for later analysis.

Below the collection navigation bar is information about the performance, usually title, performer, and composer. The display of performance information (or lack of display) is customizable, depending on the task. The controls, mostly buttons, are also highly customizable depending on the task. For example, there are buttons to turn various visuals like the waveform or notes on and off. These buttons can be removed by the researchers if, for a given task, annotators always see those visuals. Fig. 3 shows an example

(a) The CosmoNote audio controls *before* playback has started.

(b) The CosmoNote audio controls *after* playback has started.

Fig. 4. A close-up view of the CosmoNote audio controls before (a) and after (b) playback has started.

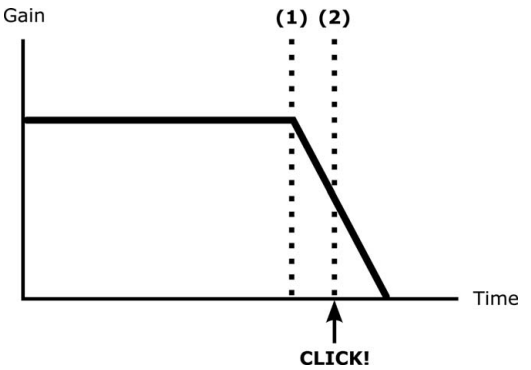

Fig. 5. A click happens when (1) the `stop()` function is called on an `AudioBufferSourceNode` and then (2) the same reference is pointed to a new `AudioBufferSourceNode` as required to restart playback.

performance with the navigation bar, performance information, and controls highlighted. The remainder of the page layout for each recorded performance is taken up by the main visualization pane with a smaller zoom pane beneath.

CosmoNote presents the recorded performances to annotators via both audio and visual data, as described in the following subsections. In presenting the performances, CosmoNote has the option to combine these presentations into one of three forms: 1) with audio only, 2) with visuals only, or 3) with both audio and visuals, depending on the nature of the annotation task. Which of these three forms is presented to annotators is recorded per annotator for later analysis.

3.1 Listening to the Performances

When an annotator wants to listen to the audio for a performance, they can start with the controls shown in Fig. 4(a) and click play at any time. The audio playback can then be paused or stopped, again, at any time. The pause button replaces the play button, whereas playback is ongoing as shown in Fig. 4(b), reverting to the play button when paused. The stop button is only active during playback, and stopping resets the playback position to the beginning of the piece. The three buttons in between the play and stop buttons are for boundary annotations and are described in SEC. 4.1.1.

During playback, annotators can jump to any time point in the audio by clicking on that time position anywhere in the main visualization (vertical position makes no difference). The current audio playback will stop and then immediately start from the new time position. Furthermore, when playback is paused, an annotator can click on any time point, and then, when restarted, playback will begin from that time point. CosmoNote has a zoom feature (see Fig. 8) that works for both visuals and audio. When a piece is zoomed in to a particular time range and an annotator hits the play button, only the corresponding time range in the audio will be played.

To show playback progress and pause time points, when playback begins, a green vertical line appears over the visuals as a play head. To keep the position of the play head synchronized with the audio playback, the play head is animated using `requestAnimationFrame()` [41], which enables animations to run at the frame rate currently used by the browser. On each animation frame request, a check is made on the amount of time that has passed in the audio file by subtracting the start time of the file from the current time (both obtained via the `AudioContext` object). The time passed is then converted into a position within the main note visualization. Using this technique, based on an idea by Wilson [42], the play head is always in sync with the audio playback, and the time increments for the movement of the play head are small enough to ensure smooth animation even for smaller audio files.

For audio file playback, CosmoNote uses the `AudioBufferSourceNode` interface. One quirk of this interface is that playback on a particular file can only be started once [43]. This is not a problem if the audio file is played through and only once. However, it is sometimes necessary to stop playback on the file and restart it again. For example, CosmoNote allows an annotator to click on the timeline during playback, causing the audio to stop and restart at that point in time. This feature requires that the `stop()` function be called on an `AudioBufferSourceNode`, which, in turn, requires that a reference be kept to that `AudioBufferSourceNode`. However, because of the quirk in that interface, to start the audio file playback again, a new `AudioBufferSourceNode` needs to be created. When the existing source node reference (required for stopping) is pointed to a new `AudioBufferSourceNode`, a click would normally be heard because the previous `AudioBufferSourceNode` is now out of scope and cannot continue to play during a gain ramp down created specifically to avoid clicks. This problem is shown graphically in Fig. 5.

To solve this clicking problem, CosmoNote uses an array of `AudioBufferSourceNode` objects to separate a currently playing `AudioBufferSourceNode` object that is being stopped from a new `AudioBufferSourceNode` object that is being started. This allows for a currently playing `AudioBufferSourceNode` that is ramping down after a stop to be kept in scope until the ramping down is completed. A new `AudioBufferSourceNode` can be assigned, started, and ramped up before the ramp down of

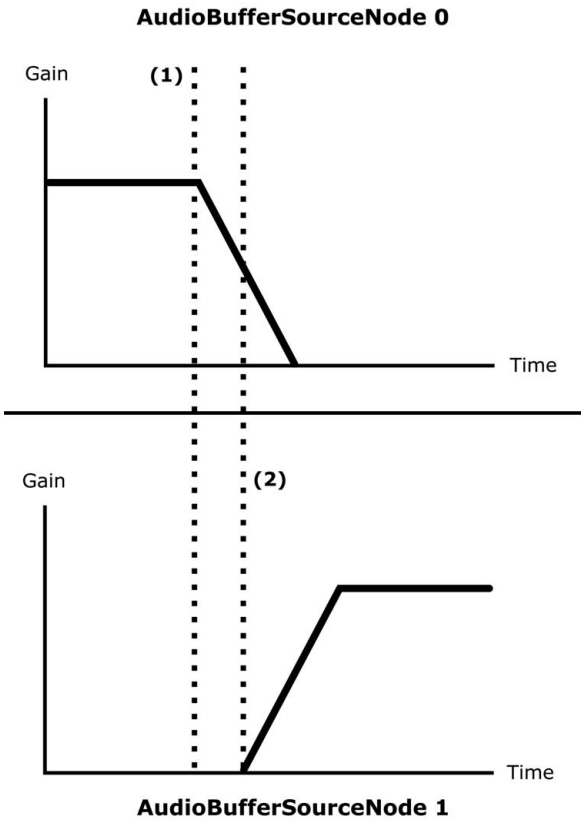

Fig. 6. An array holds two AudioBufferSourceNode objects and then **(1)** the `stop()` function is called on `AudioBufferSourceNode` 0 and **(2)** a new `AudioBufferSourceNode`, 1, is created and then started. This avoids any clicks by creating a new `AudioBufferSourceNode` instead of reassigning an existing reference to one.

the stopping `AudioBufferSourceNode` is complete. The stopped `AudioBufferSourceNode` can then go out of scope while the new `AudioBufferSourceNode` ramps up. The latest addition to the `AudioBufferSourceNode` array is always the one that is stopped, immediately followed by the addition of a newly created `AudioBufferSourceNode` that is started. Whenever playback needs to be stopped then started again, the process is repeated. This solution is shown graphically in Fig. 6.

3.2 Visualizing the Performances

CosmoNote shows musical data in its main visual pane via waveforms, note and pedal data, and curves based on extracted feature data including loudness, tempo, and harmonic tension (all described in SEC. 2). Each type of visual data is a distinct layer in the overall visualization. With all of the data layers displayed at the same time, the visualization can become dense and difficult to read, so transparency is used to varying degrees in all of the layers. To further mitigate the problem of too much visual data, each information layer can be turned on and off individually. This also allows for a focus on particular data types as shown for different layers in Fig. 7. For all of the layers, the design of the visuals emphasizes general trends in each kind of data because annotators are tasked with annotating expressive structures in the performance rather than finding exact values for any given data type.

Beneath the main visual pane is a zoom pane in which annotators can select a time range that zooms the main visualization panel to the selected time range, allowing for a more detailed look at the notes and other data layers for that part of the performance. When zooming, all of the visible data layers in the main pane are zoomed. Zooming into specified time ranges not only zooms the visuals but the audio as well, allowing annotators to listen to the corresponding time range in the audio for a given performance. Fig. 8 shows a time range selection (the gray area) in the zoom pane while showing the corresponding notes in that time range in the main visualization panel. The zoom pane shows only the waveform and notes because it is meant to provide a basic context map of the full performance while the main visualization pane is zoomed.

3.2.1 Audio Data

The lowest layer for CosmoNote visuals is a basic waveform taken from the audio file data from a given recorded performance. To allow for efficient drawing of the waveform, especially with the zoom function, the amount of audio data is reduced down to match the pixel width of the main and zoom panes. When zoomed, the waveform behaves as expected, showing just the waveform from the selected time range. The waveform layer is shown in Fig. 7(a).

3.2.2 Note and Pedal Data

The note visual layer includes note data derived from MIDI recordings with the note value on the vertical axis (from lowest to highest note on the piano) and the length of each note in seconds on the horizontal (time) axis. Because the actual values of the notes played are not needed for the annotation task, the MIDI note value is not shown. However, because all of the performance data in CosmoNote come from piano recordings, a piano graphic is shown along the left vertical axis that allows annotators to loosely infer the values of the notes. The MIDI velocity, associated with approximate loudness, of each note is depicted via the note's transparency with more opaque notes being louder and more transparent notes being quieter. Note data are shown in Fig. 7(b).

Sustain, soft, and sostenuto pedal data are taken from the MIDI recordings and shown as area curves. The displacement of the each pedal is shown on the vertical axis with the distance from the top of the graph representing the pedal displacement distance from the rest position. That is, the closer the line is to the horizontal (time) axis, the more the pedal is being depressed. With this orientation, it makes sense to show the data as an area graphic in which the area shows both that the pedal is being used and how much it is depressed at a glance. Fig. 7(c) shows sustain pedal data as the teal-colored area curve.

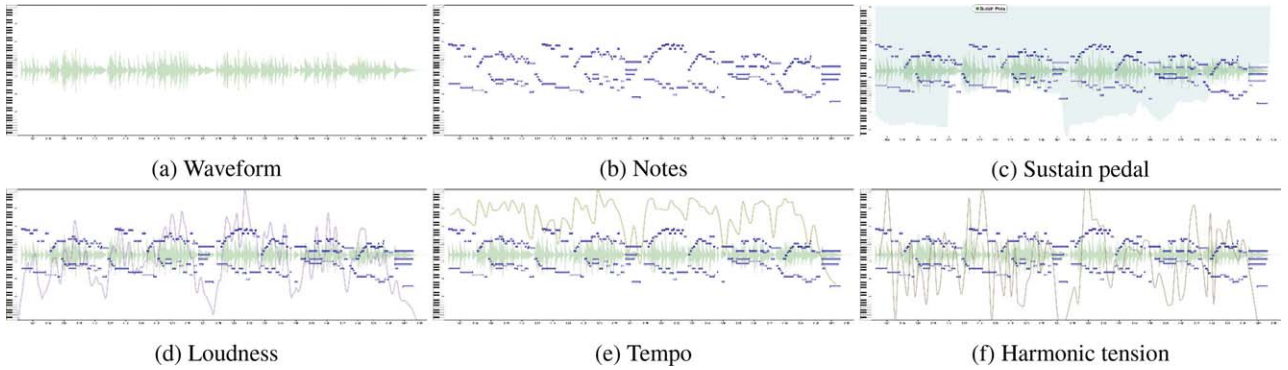

Fig. 7. CosmoNote data visualization layers: (a) waveform, (b) notes, (c) sustain pedal, (d) loudness (in sones), (e) tempo (in beats per minute), and (f) harmonic tension (the rate of chord changes). The first two layers, waveform and notes, are shown as backdrop for the rest and are maintained throughout the other examples.

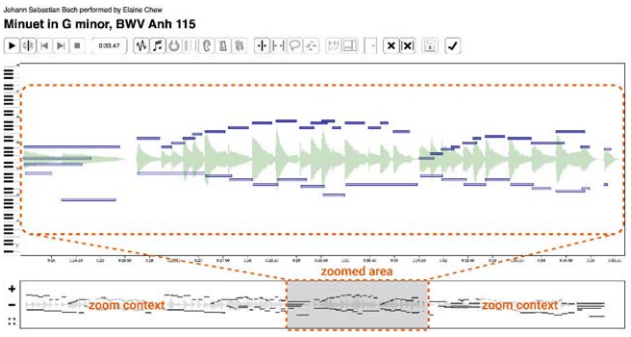

Fig. 8. CosmoNote zoomed to show a short time range. The smaller visualization pane below shows the context and zoomed notes (inside the gray square). The controls on the left, from top to bottom, increase the zoom, decrease it, and reset it.

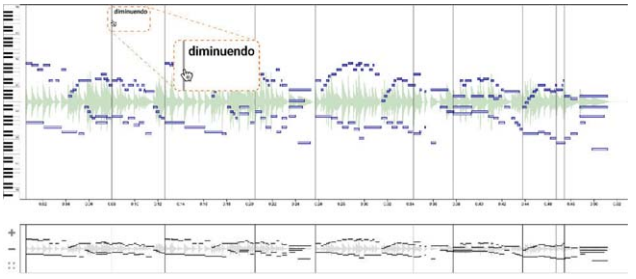

Fig. 9. A set of instants, shown as gray vertical lines, for a performance. Labels are displayed when the mouse cursor hovers over the instants as with the "diminuendo" label shown here.

3.2.3 Feature Data

Feature data, described in SEC. 2.3, is shown as curves layered over the waveform and note layers. Fig. 7 shows the various information layers with the note layer as backdrop. The loudness curve is the purple curve in Fig. 7(d). Tempo (in beats per minute) is given as the green curve in Fig. 7(e). The tension curve is the brown curve in Fig. 7(f). More specifically, Fig. 7(f) shows cloud momentum (changing dissonance) though all three dimensions of tension can be shown with each dimension being assigned its own color.

3.2.4 Instants Data

Instants are a type of annotation for pre-marking areas of interest, like score structures or other landmarks, for annotators. Because instants are a part of the recorded performance data, in contrast to the annotations created by annotators themselves (which are described next in SEC. 4), they cannot be edited or deleted. They mark specific times in a piece visually with a vertical line and a text label that appears when annotators mouse over the line. Fig. 9 shows a series of instants (showing only one label) for a performance of the Minuet in G minor by co-author Elaine Chew. In this case, the instants describe dynamic and tempo markings from the score labeled, in order, mezzo forte, diminuendo, forte, diminuendo, piano, mezzo forte, piano, crescendo, forte, and poco ritardando.

The lines representing instants are shown in both the main visual pane and in the zoom pane below it (as seen in the zoom pane at the bottom of Fig. 9). In the zoom pane, the instants delineate a series of clickable boxes that represent time ranges. By double-clicking in one of the boxes in the zoom pane, the main visual pane will zoom exactly to that time range, giving annotators a convenient way to zoom into the boxes delineated by instants.

3.2.5 Supplementary Data

In CosmoNote, supplementary data are defined as data that are not directly computed from a recording of a performance or even the score but may still be related to the performance. Essentially any time-based data can be included as supplementary data, although those data are most useful if synchronized with the audio or note data. Supplementary data are presented as curves with customizable appearances and any number of curves can be added. For example, CosmoNote can incorporate related data such as physiological data from performers or listeners. Fig. 10 shows two electrocardiographic signals recorded during a piano performance with the red signal being the player and blue signal the listener.

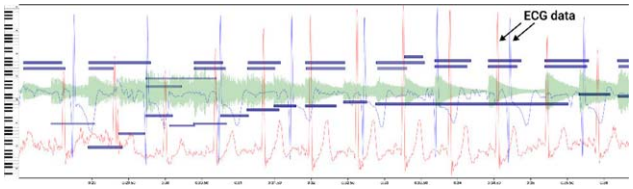

Fig. 10. Supplementary data in the form of two electrocardiographic (ECG) signals recorded synchronously with the performance, by co-author Elaine Chew, of Prokofiev's "Juliet As A Young Girl" from *Romeo And Juliet*, Op. 75.

4 ANNOTATING THE PERFORMANCES

Once annotators have listened to and studied the visuals of the performance data, the next step in the workflow, at the heart of CosmoNote, is the annotation of the perceived structures in the performances.

4.1 Annotation Types

CosmoNote features four types of annotations: boundaries, regions, comments, and note groups. To create one of the annotation types, annotators select the corresponding button on the toolbar above the main visualization pane (shown in Fig. 3) to set the annotation type mode. Once an annotation type mode is selected, annotators can place as many of that type of annotation as desired. For each type of annotation, annotators can create a label with custom text. Labels are created for annotations by accessing an inspector pane (see Fig. 11 for example labels for regions) and typing in the desired text for each annotation. For boundaries, regions, or comments, annotators can, when that annotation type mode is selected, hover the mouse over the annotation to see the label. Fig. 12 shows examples of each the four annotation types.

4.1.1 Boundaries

Boundaries represent time points that separate the performed music into segments of coherent chunks of music, e.g., a complete musical idea or a musical thought. Boundaries communicated through performance not only separate a larger piece of music into smaller, meaningful units but also help listeners make sense of the music. Annotators can place any number of boundaries for a given recorded performance, and they can be placed (or removed) at any time. Once placed, boundaries can be moved in time by clicking on the lines and dragging them or, when selected, by clicking and dragging the arrows that appear at the top of the boundary [as shown in Fig. 12(b)]. Boundary labels are created or edited via the annotation inspector like the region inspector shown in Fig. 11. The red vertical lines in Fig. 12(b) are examples of boundaries.

Annotators can place boundaries of four different levels that represent segmentation of musical ideas at different time scales with the exact definition of each level, depending on the specific annotation task. The boundary levels are indicated visually via the thickness of the boundary and by transparency with opacity increasing as the level increases as shown in Fig. 12(b). During audio playback, boundaries can be placed at the current location of the play head by selecting one of the numbers 1–4 from the keyboard, with the numbers corresponding to the four levels.

When an annotator listens to a performance after boundaries have been placed, they can hear (if enabled) a small woodblock sound effect when the play head reaches a boundary. The gain of the sound effect is louder for each of the four boundary levels with level one being fairly quiet and level four being the loudest. When the audio is playing, an annotator can use the skip forward or backward buttons to skip the playback to the next boundary or back the previous boundary, allowing annotators to hear the results of their boundary placement (see Fig. 4).

4.1.2 Regions

Regions delineate entire sections or areas of interest in a performance, and, although they perform a function similar to boundaries, they instead encompass all of the notes between two boundaries rather than simply denoting the boundaries themselves. They can be used, for example, to mark transitions or lead up to a tipping point. Any number of regions can be placed, and they can be moved, resized, or deleted after placement. Regions can overlap each other, enabling the annotation of, for example, overlapping phrases in which one phrase begins before the other ends. Because regions use transparency, when regions do overlap, the overlapping area will be darker, showing the overlap clearly at a glance. Regions are shown as the semi-transparent red squares in Fig. 12(c) and region labels are shown in Fig. 11.

4.1.3 Comments

Comments enable annotators to mark elements of interest and write custom text about them. Comments, though similar to the labels associated with the other annotation types, provide a means to point at something of interest that is not captured by the other types. Comments are presented as dotted lines to distinguish them from boundaries as shown in Fig. 12(d). The text can be edited from the annotation inspector, in comment mode, similar to what is shown in Fig. 11.

4.1.4 Groups

Groups provide a way for annotators to select and highlight a group of notes of interest. Any number of notes can be selected to create a group and any number of groups can be created. To create a group, annotators create a selection square with the mouse that approximately encompasses the notes that they want to select for their group. Once a group selection is created, annotators can add or remove notes individually from the group. This is especially useful if the selection square, because of its shape, did not exactly create the desired group. As with the other annotations, annotators can create and edit labels for groups by enabling the inspector when in group mode. Fig. 12(e) shows two note groups that are highlighting the upper and lower melodies played by the right hand of the performer.

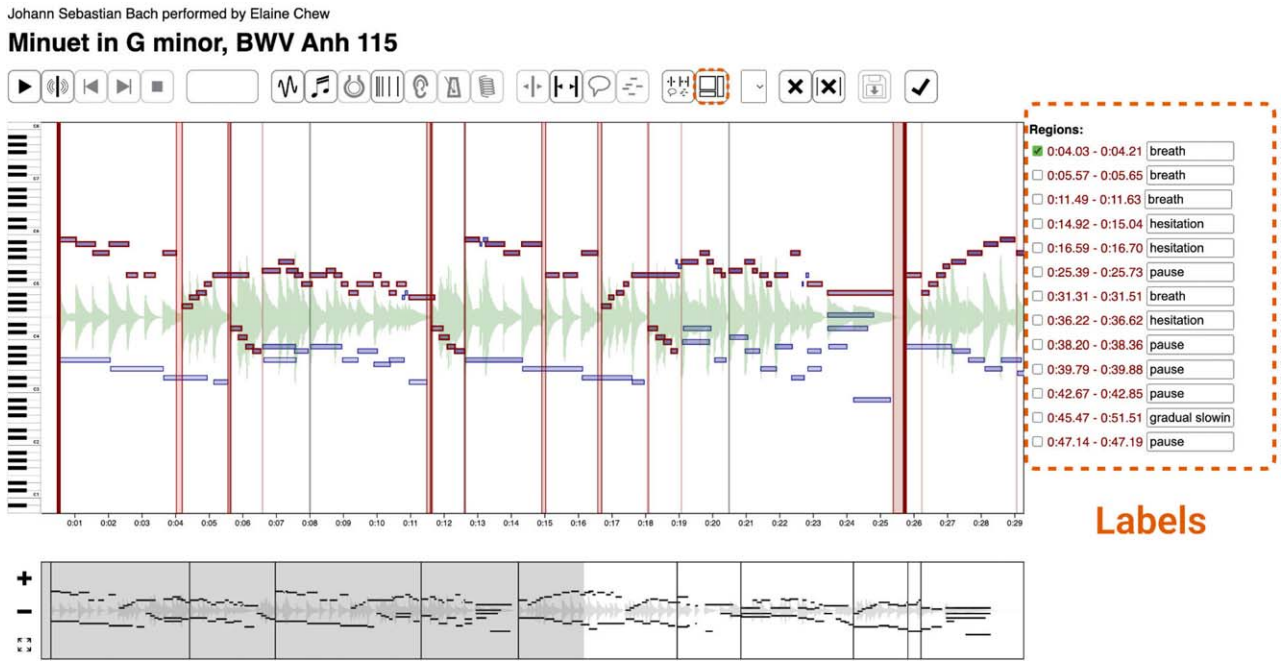

Fig. 11. Various annotations with the annotation inspector showing labels for regions. All annotation labels appear to the right of the main pane when the inspector is activated. The annotation's starting time (and ending time for regions) is shown to the left of each label.

4.2 Annotation Tasks

Once annotators have understood the different annotation types, they can move on to the annotation task itself. Annotators are provided with instructions for specific annotation tasks, accessible via a link at the top of the page, which can change depending on the nature of the collection or study. Although the task instructions are variable, they are essential for helping annotators focus on the significant elements of any given task. The difficulty, though, is to balance the task instructions such that they are neither too specific nor too general. To evaluate the task instructions, annotators were presented with instructions in two pilot studies, described below.

4.2.1 Pilot Study 1 Task

For the first pilot study task, eight music and audio researchers were presented a series of excerpts from Chopin's Ballade No. 2 as performed by co-author Elaine Chew. The ballade was split into eight musically coherent excerpts and presented to annotators in a shuffled order. For the annotation task, participants were asked to place boundaries as communicated in the performed music and indicate the strength of each boundary (levels 1–4) with the levels defined as

1. Motives: smallest indivisible successions of notes that may be delineated by accents;
2. Sub-Phrases: parts of a phrase, e.g., antecedent or consequent phrases;
3. Phrases: complete self-contained musical statements; and
4. Sections: major structural units comprising a complete musical idea.

To further help annotators with the task, the following suggestion about boundaries was included with the instructions:

> Performers may mark boundaries using pauses, stress, or contrast. For example, accents could mark the beginnings of groups of notes, pauses can separate musical ideas, phrases may be expressed by increasing then decreasing tempo and/or loudness, a change of timbre and loudness may mark the beginning of a new section.

The interface presented annotators with visual layers from note, pedal, loudness, tempo, and harmonic tension data. Participants were allowed to toggle any of these visualizations on and off at any time but were instructed to "let their ear be their main guide."

4.2.2 Pilot Study 2 Task

For the second pilot study task, seven music and audio researchers were presented with a single short music excerpt, Variation XXXII in Beethoven's "32 Variations in C minor, WoO 80." In contrast to the freeform nature of the task from the first pilot study, for this task, the annotators were asked, in the task instructions, to annotate *segmentation* and *prominence* as paraphrased (for stylistic consistency) below. Both definitions are followed by lists of examples, which were also provided in the task instructions.

Segmentation is the process of dividing something, in this case music, into meaningful units.

- Boundaries: time points that separate a music stream into segments representing meaningful chunks of music, e.g., a musical idea or thought. Boundaries not only separate

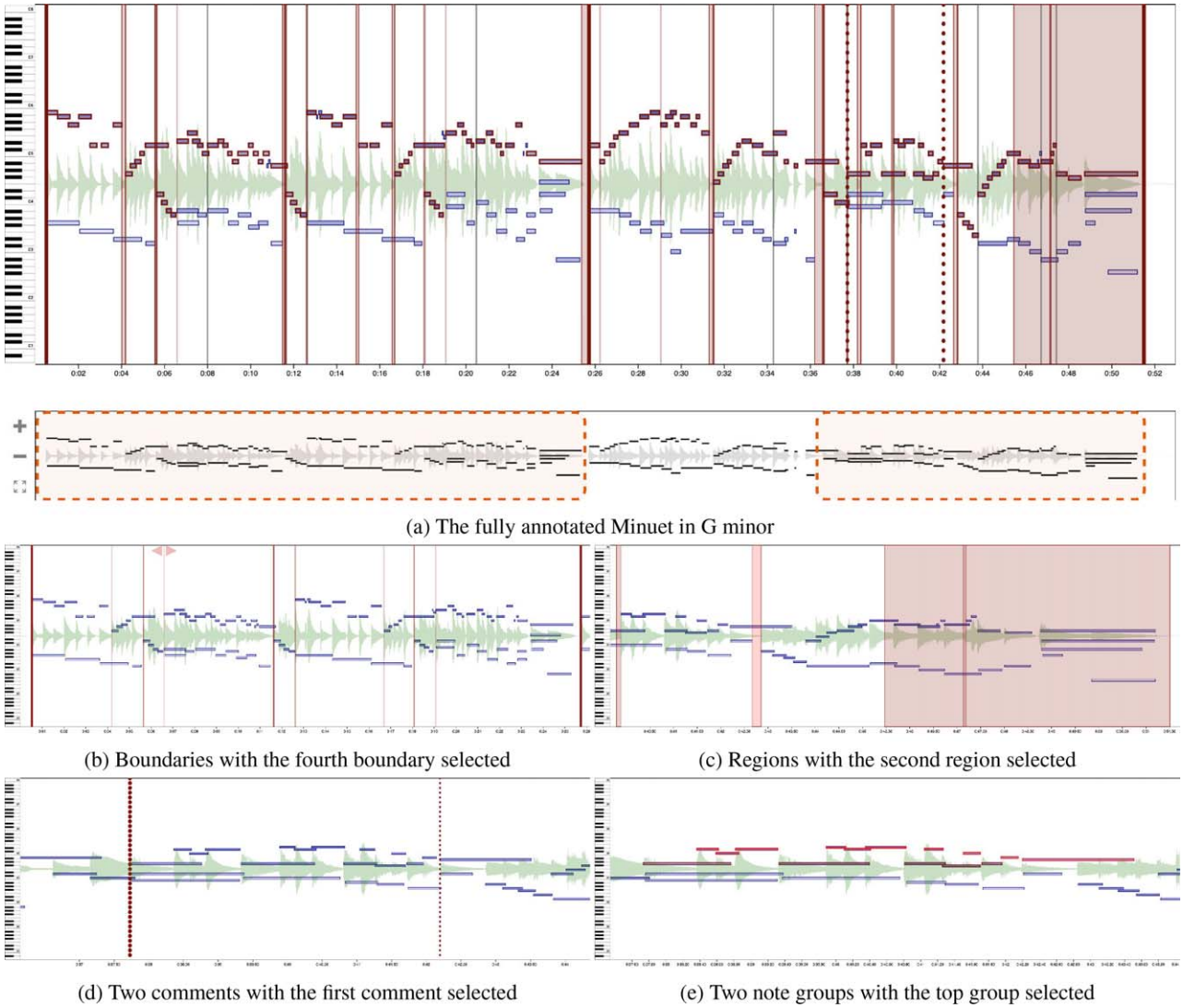

Fig. 12. The four types of annotations with (a) showing a complete piece with every annotation type. The two dashed areas in the zoom pane are the selected time ranges for boundaries and regions shown in (b) and (c), respectively. Comments (d) and groups (e) are further zoomed to better show their placement.

a larger piece of music into smaller, coherent units; they also help listeners make sense of the music. There is a boundary annotation type in CosmoNote, with four levels of boundaries, defined from 1 (weakest) to 4 (strongest).

- Transitions: musical passages that set up a change coming in the music or blur changes in the music by moving slowly through them, linking musical ideas. Transitions may be annotated using regions.
- Pauses: segments of time that add space between two adjacent structures. Executed by the performer by lingering on notes or using silence. Pauses may be annotated using regions.

Prominence characterizes an emphasis drawn toward a certain part of a whole in the music.

- Stress: an emphasis of a particular element to make it more prominent than those around it. Stress may be indicated by a combination of performer actions like an increase in sound intensity, duration, or change in timbre. Stress may be marked using boundaries or note groups.
- Melodic salience: a special case of prominence dedicated to a sequence of notes that may be recognized by an increase in loudness and duration or a variation in the timbre of the melody notes. Melodic salience may be marked using note groups.
- Tipping points: moments when musical time is suspended/stretched to a point beyond which a return to the pulse is inevitable. Tipping points may be marked using regions, boundaries, or note groups.

5 ANALYZING THE ANNOTATIONS

The last step in the CosmoNote workflow is the collection and analysis of the annotation data. The data are exported from the CosmoNote database with a custom Python script that extracts all of annotation data as JSON. Data for each annotation contain a pseudonymous annotator iden-

tifier, the collection, the performance, the annotation type (boundary, region, comment, and group), a time (two times for regions), the boundary strength (for boundaries only), a label (if set), a creation timestamp, and an updated timestamp (if subsequently edited after creation). Additionally, note groups also have the MIDI note number and the start and end times for every note in the group.

To evaluate how citizen scientists use CosmoNote for annotations tasks, two pilot studies were conducted as described in SEC. 4.2. For the first pilot study, annotators were asked to focus strictly on boundaries. For the second, annotators were asked to use all four of the annotation types: boundaries, regions, comments, and note groups. In both pilot studies, the annotations of participants were compared with those of the performer.

5.1 Pilot Study 1 Results

For the first pilot study, whose task is described in SEC. 4.2.1, the results were compared to annotations made by the performer that served as a baseline for comparison. First, the number of boundaries of each level per excerpt was compared. Fig. 13 shows a distribution of the number of boundaries (by level) participants placed for each excerpt compared to the performer's annotations of the same excerpts. The *x* axis shows the count (the number of boundaries placed), and the *y* axis shows results for each of the eight excerpts. Individual box plots represent the boundary distribution of the eight listeners, and the red dot represents the performer's distribution. The panels contain the different distributions of the boundary levels from 1 to 4.

The results indicate a convergence for the higher boundary levels, 3 and 4; these levels were, overall, less used by both the listeners and the performer. Level 2 boundaries were also broadly comparable between them. In contrast, level 1 presented the highest variability in the number of boundaries and the lowest correspondence between how the performer and listeners placed boundaries on each excerpt. In general, the performer placed fewer boundaries than the average listener for levels 1, 3, and 4.

The location of boundaries was analyzed for every level over time. An aggregated representation of the placement of boundaries, by level, for all listeners is compared to that of the performer in Fig. 14. The boundary profile curves by level were obtained using kernel density estimation [44, 45] to indicate the concentration of boundaries marked by many listeners around a given time point. Boundaries of level 4 occur more often close to score markings, which are often large-scale sectional boundaries, as is also the case with some level 3 boundaries. Density profiles for levels 2 and 1, which may correspond to finer subdivisions into phrases or subphrases, are more spread over time and are distinct from score markings. Overall, all listeners marked boundaries close to the performer's annotations, although they sometimes used different boundary levels to demarcate the segmentation. For example, many listeners' level 1 markings can be seen after the score section marked *agitato*, whereas the performer used boundary level 2 annotations on that passage. Thus, although listeners concur on the existence of a segmentation boundary, the precise boundary level varied.

5.2 Pilot Study 2 Results

For the second pilot study (task described in SEC. 4.2.2) listeners' annotations were also compared to those of the performer with, for this study, comparisons of boundaries (by level), regions, and groups for the same piece.

To analyze the boundary annotation results, the same technique as demonstrated in Fig. 14 was applied. Fig. 15 shows the distribution of boundaries for the second study. Results for this piece were similar to those described in SEC. 5.1 with more boundaries of lower levels (1 and 2) than those of higher levels (3 and 4) for both the listeners and performer. However, although an overall agreement in the location of these boundaries is observed, agreement is not observed for their strength level. For example, almost all of the performers markings are represented by listeners' annotations yet the performer did not place any boundary level 4 annotations, whereas listeners did. For the performer, this excerpt was one of 32 variations (plus the preceding tema for a total of 33) and, as such, did not require any level 4 boundary whereas the listener likely viewed the excerpt as a standalone piece, hence the level 4 boundaries.

The region annotations results are shown in Fig. 16. The results showed that listeners marked regions according to quite different conceptualizations of their meaning. For example, listener 5 placed regions throughout the whole piece, essentially dividing it into chunks, whereas all other listeners (including the performer) used regions to mark only specific time selections of interest. Region starts and endings, which sometimes correspond to dynamic markings, were shared between most listeners and the performer.

The group annotations are shown in Fig. 17. Listeners were not required to use note group annotations, and only five out of seven listeners did so for this task. As shown in Fig. 17, all of the listeners marked the series of ascending notes of the left hand melody in the first 20 s of the piece, whereas the performer only marked the last note. This and other notes the performer marked tended to be notes made prominent for structural significance, like an outline of the listeners' combined note groups. Common groups were marked by listeners 3 and 6 at around 70–80 s and again around 90–110 s. These corresponded with a few notes marked by the performer. The performer's note groups that coincided with listeners' note groups are outlined in red in Fig. 17. The number of groups created (differentiated by the markers in Fig. 17) is also of interest. Although listener 4 marked 12 different groups, listener 3 and the performer created eight, and listener 6 created five, and listeners 2 and 7 only created one group, with corresponding notes. This difference in group numbers and the discrepancy between listeners and the performer are larger than for the other annotation types, which may be an indication of how listeners are understanding the concept of note groups and using them in their annotations.

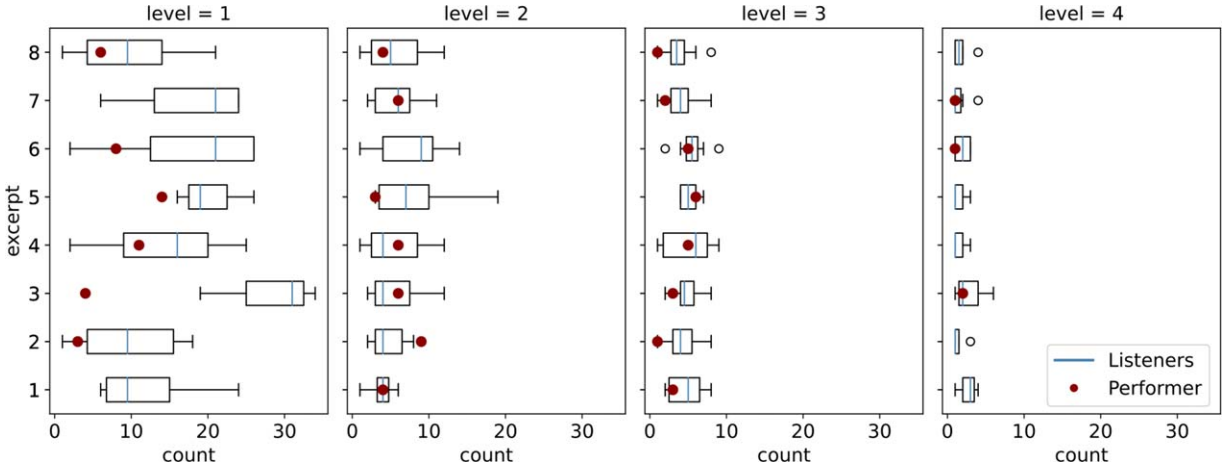

Fig. 13. Pilot study results showing the number of boundaries of each level placed, per excerpt, by participants compared with the boundaries placed by the performer.

6 CONCLUSION

This paper described CosmoNote, a Web-based citizen science annotation tool, and the workflow that it enables. In the course of developing and testing CosmoNote, the following novel features were created.

- Display of discrete note information for the performance visualization that is synchronized with the audio playback;
- Visual display of note velocity via transparency corresponding to loudness;
- Ability to start audio playback from any arbitrary time point with a simple click;

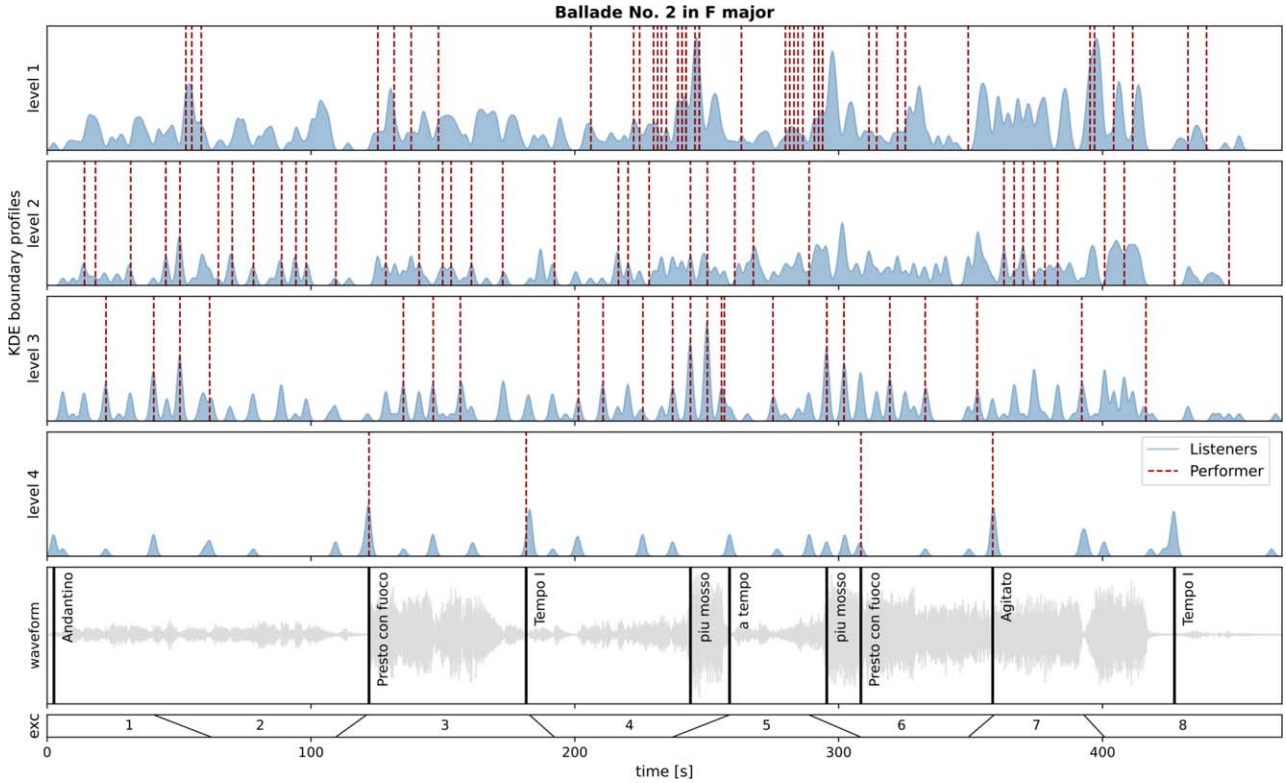

Fig. 14. Boundary annotation placement profiles comparing the listeners (blue density curves over time) to the performer (red vertical dashed lines). Annotations are split into four levels; boundaries were aggregated per excerpt. The bottom panel shows how Chopin's Ballade No. 2 was split into eight musically coherent excerpts; diagonal lines mark overlapping zones between excerpts, where the widest possible range of each trapezoid represents the time range of the excerpt. Score dynamics and tempo markings are overlayed on the waveform.

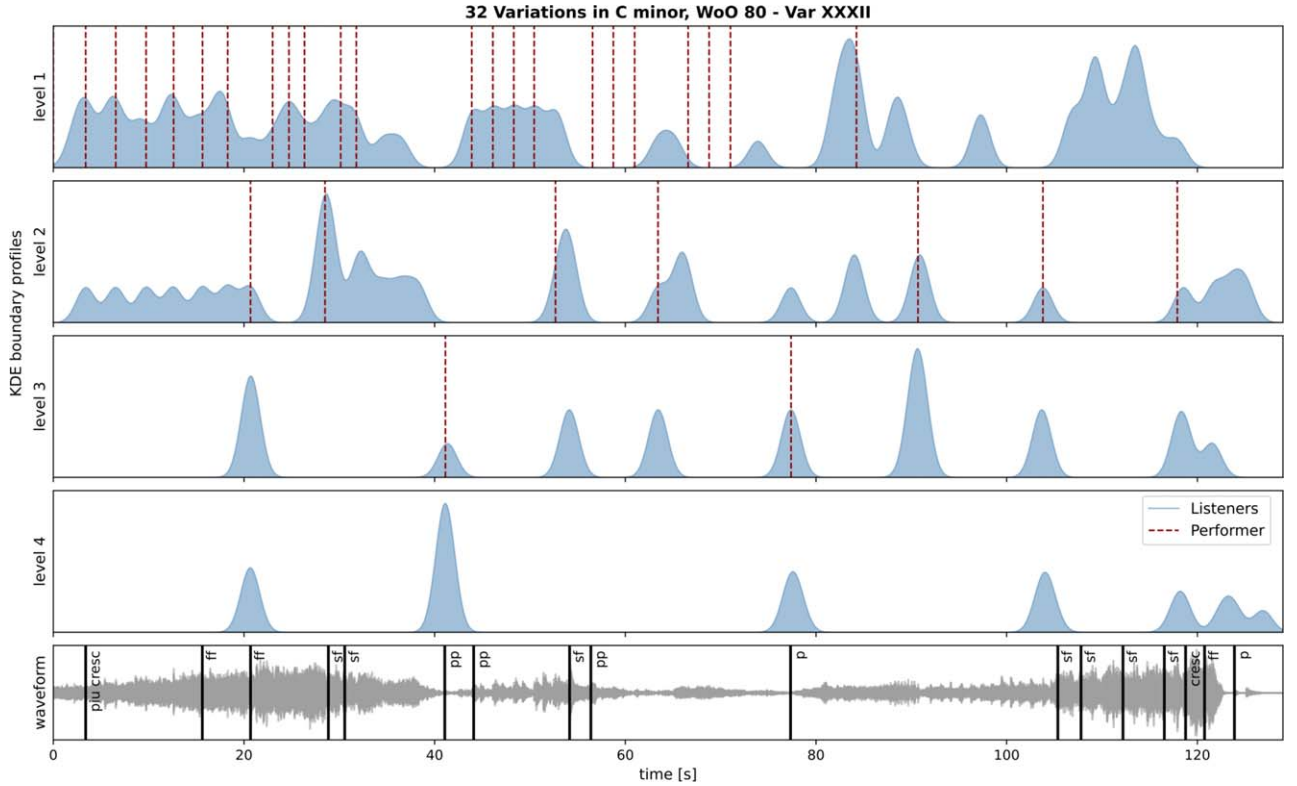

Fig. 15. Boundary annotation placement profiles comparing the listeners (blue density curves over time) to the performer (red vertical dashed lines). Annotations are split into 4 levels. The bottom panel shows Beethoven's Variation XXXII waveform with score dynamic markings overlayed.

- Ability to zoom into a set of notes and to see and hear just the audio for those notes;
- Option to view information layers like piano pedals, loudness, tempo, and harmonic tension;
- Ability to zoom all data visuals including the waveform, notes, pedal data, loudness, tempo, and harmonic tension;
- Ability to easily zoom into sections of performances delineated by instants;
- Ability to flexibly incorporate other time series data as additional data visualization layers for analysis;
- Ability to create multiple types of annotations including boundaries, regions, groups, and comments; and
- Ability to skip the audio playback from boundary to boundary.

CosmoNote is available to the public and features four collections of performances of Bach's Goldberg Variations

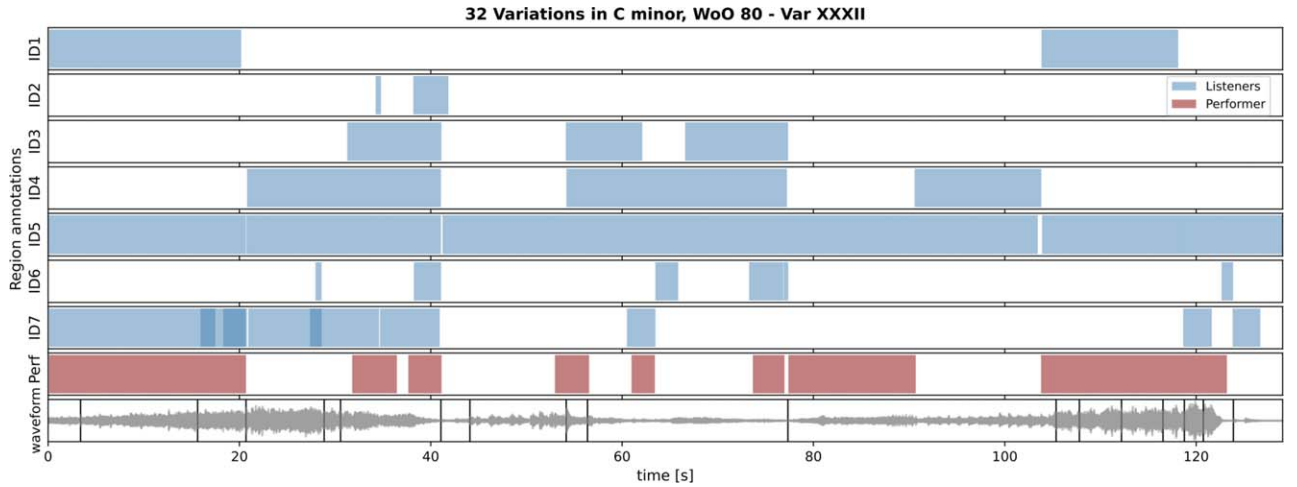

Fig. 16. Region annotations comparing the listeners (blue patches over time) to the performer (red patches over time). Each row shows data for one listener ID. The bottom panel shows Beethoven's Variation XXXII waveform with score dynamic markings overlayed.

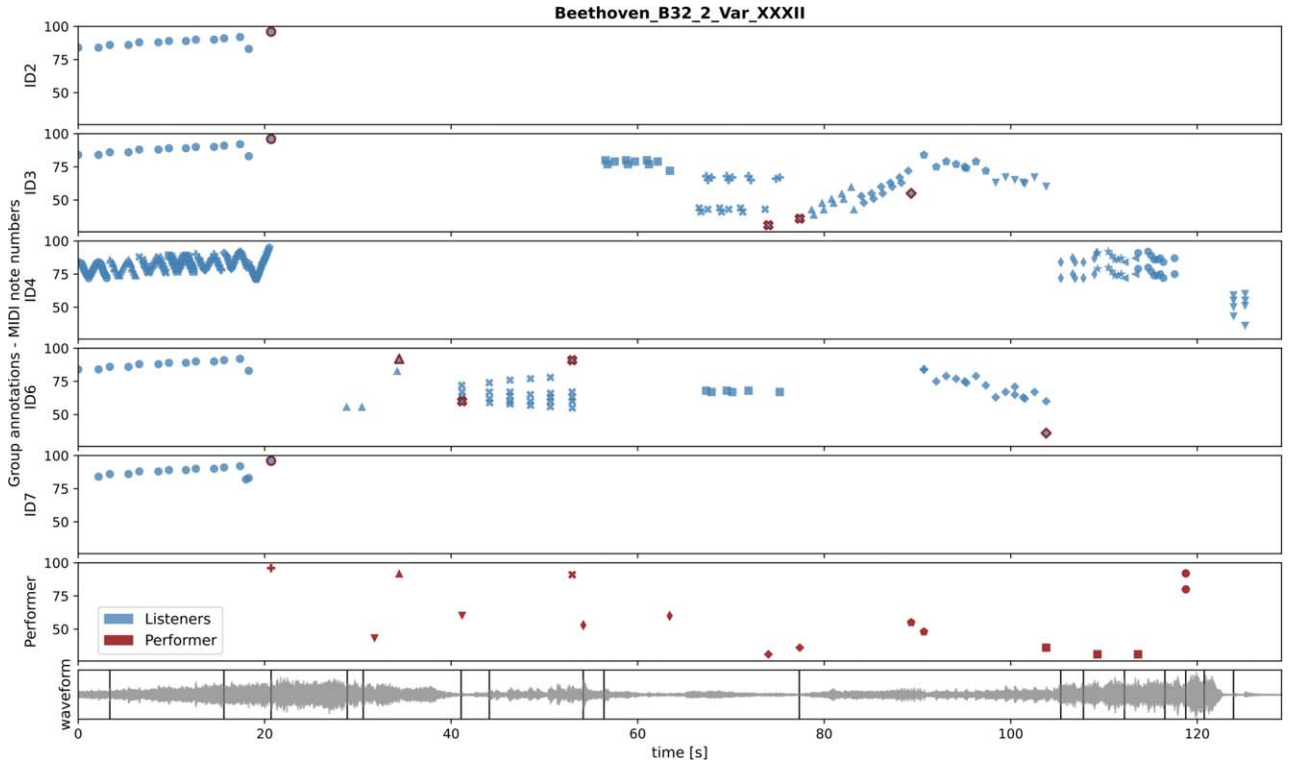

Fig. 17. Group annotations comparing the listeners (blue markers) to the performer (red markers). Each row shows data for one listener ID; with different markers per group. The bottom panel shows Beethoven's Variation XXXII waveform with score dynamics markings overlayed.

by Glenn Gould. Following this initial public release, it will be updated periodically with other thematic campaigns featuring different performance collections and study the efficacy of this method of collecting annotations on musical expression.

In addition to a public release, two pilot studies have been conducted to evaluate the use of CosmoNote for annotating performances.

1. In the first pilot study, the authors compared annotations of excerpts from a Chopin performance from study participants with those of the performer. Aggregating data from all participants allowed for smoothing out annotation behaviors of specific participants to concentrate on the global tendencies. Annotations from individual boundary levels provided consistent information showing listeners and the performer focusing on different time scales of the performance. The highest levels coincided more with score markings and matched more closely between participants and performer, whereas performance subtleties were most evident at the lowest boundary levels, and it was found that the most significant divergence between participants and performer were at these lowest levels, especially level 1. The placement of level 1 boundaries could be caused by a mismatch in listener's perception of subtle segmentation cues. This first pilot study provided feedback to help with understanding how annotators used boundaries in CosmoNote and has given clues about how to approach the annotation task instructions for future annotators, especially for marking the more subtle aspects of performances.
2. The second pilot study also compared listener's annotations to those of the performer. In this case however, listeners were free to use all of CosmoNote's annotation types and they were tasked with marking segmentation and prominence on a shorter (2 min) piece. Boundary and region annotations were consistent with those of the performer on most occasions (with varying strength levels for boundaries and longer/shorter selections for regions). On the contrary, group annotations tended to be clustered and were most similar among participants but were different from those of the performer, whose individual markings were more sparse. Although these results conveyed information about the structures listeners and performers were perceiving, more importantly for this analysis, they provided clues to help fine-tune the annotation task instructions.

For further studies, annotation data will be collected from citizen scientists, and, using the collected data, particularly the annotation times, relationships will be established among the aggregated annotation times and high-level prosodic features, low-level musical features, and acoustic

properties. To do this, innovative techniques like change point analysis and multidimensional techniques such as cluster analysis or multiple regressions will be used.

Ultimately, the citizen science annotation data will be used to develop a vocabulary of expressive musical gestures to establish a common standard for transcribing expressive elements in performed music and facilitate the sharing of annotated databases for research and technological development so as to advance understanding of decision making in expressive musical performances.

7 ACKNOWLEDGMENT

This result is part of the project COSMOS that has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant number 788960).

8 REFERENCES

[1] E. Chew, "Notating Disfluencies and Temporal Deviations in Music and Arrhythmia," *Music Sci.*, vol. 1, no. 3, pp. 1–22 (2018 Sep.). https://doi.org/10.1177/2059204318795159.

[2] C. Palmer and S. Hutchins, "What is Musical Prosody?" *Psychol. Learn. Motiv.*, vol. 46, pp. 245–278 (2006 May). https://doi.org/10.1016/S0079-7421(06)46007-2.

[3] E. Chew, "From Sound to Structure: Synchronizing Prosodic and Structural Information to Reveal the Thinking Behind Performance Decisions," in Cristine MacKie (Ed.), *New Thoughts on Piano Performance: Research at the Interface Between Science and the Art of Piano Performance*, pp. 143–144 (London International Piano Symposium, London, UK, 2016).

[4] "COSMOS," https://cosmos.cnrs.fr/ (Accessed Oct. 19, 2022).

[5] "Le Piano Virtuose," https://www.youtube.com/watch?v=yXkwusNyte4 (Accessed Oct. 19, 2022).

[6] "CosmoNote," https://cosmonote.ircam.fr/ (Accessed Oct. 19, 2022).

[7] L. Fyfe, D. Bedoya, C. Guichaoua, and E. Chew, "CosmoNote: A Web-Based Citizen Science Tool for Annotating Music Performances," in *Proceedings of the International Web Audio Conference* (Barcelona, Spain) (2021 Jul.).

[8] G. Tzanetakis and P. R. Cook, "Experiments in Computer-Assisted Annotation of Audio," in *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 111–115 (Atlanta, GA) (2000 Apr.).

[9] G. Tzanetakis and F. Cook, "A Framework for Audio Analysis Based on Classification and Temporal Segmentation," in *Proceedings of the 25th EUROMICRO Conference. Informatics: Theory and Practice for the New Millennium*, pp. 61–67 (Maspalomas, Spain) (1999 Aug.). https://doi.org/10.1109/EURMIC.1999.794763.

[10] X. Amatriain, P. Arumí, and M. Ramírez, "CLAM, yet Another Library for Audio and Music Processing?" in *Proceedings of the ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications*, pp. 46–47 (Seattle, WA) (2002 Nov.). https://doi.org/10.1145/985072.985097.

[11] X. Amatriain, J. Massaguer, D. Garcia, and I. Mosquera, "The CLAM Annotator: A Cross-Platform Audio Descriptors Editing Tool," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 426–429 (London, UK) (2005 Sep.).

[12] M. Notess and M. B. Swan, "Timeliner: Building a Learning Tool Into a Digital Music Library," in *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp. 603–609 (Lugano, Switzerland) (2004 Jun.).

[13] P. Herrera, Ò. Celma, J. Massaguer, et al., "MUCOSA: A Music Content Semantic Annotator," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 77–83 (London, UK) (2005 Sep.).

[14] K. Sjölander and J. Beskow, "Wavesurfer-An Open Source Speech Tool," in *Proceedings of the 6th International Conference on Spoken Language Processing*, vol. 4, 464–467 (Beijing, China) (2000 Oct.).

[15] B. Li, J. A. Burgoyne, and I. Fujinaga, "Extending Audacity for Audio Annotation," in *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, pp. 379–380 (Victoria, Canada) (2006 Oct.).

[16] "Audacity," https://www.audacityteam.org/ (Accessed Oct. 19, 2022).

[17] C. Cannam, C. Landone, and M. Sandler, "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files," in *Proceedings of the ACM International Conference on Multimedia*, pp. 1467–1468 (Firenze, Italy) (2010 Oct.). https://doi.org/10.1145/1873951.1874248.

[18] M. Cartwright, A. Seals, J. Salamon, et al., "Seeing Sound: Investigating the Effects of Visualizations and Complexity on Crowdsourced Audio Annotations," in *Proceedings of the ACM Conference on Human-Computer Interaction*, vol. 1, paper 29 (Denver, CO) (2017 Nov.). https://doi.org/10.1145/3134664.

[19] "Audio-Annotator," https://github.com/CrowdCurio/audio-annotator (Accessed Oct. 19, 2022).

[20] "Wavesurfer.js," https://wavesurfer-js.org/ (Accessed Oct. 19, 2022).

[21] B. Meléndez-Catalán, E. Molina, and E. Gómez, "BAT: An Open-Source, Web-Based Audio Events Annotation Tool," in *Proceedings of the International Web Audio Conference* (London, UK) (2017 Aug.).

[22] "BAT - BMAT Annotation Tool," https://github.com/BlaiMelendezCatalan/BAT (Accessed Oct. 19, 2022).

[23] C. Wang, G. J. Mysore, and S. Dubnov, "Re-Visiting the Music Segmentation Problem With Crowdsourcing," in *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, pp. 738–744 (Suzhou, China) (2017 Oct.).

[24] M. Cartwright, B. Pardo, G. J. Mysore, and M. Hoffman, "Fast and Easy Crowdsourced Perceptual Audio Evaluation," in *Proceedings of the IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 619–623 (Brighton, UK) (2016 May). https://doi.org/10.1109/ICASSP.2016.7471749.

[25] "Bösendorfer Disklavier Edition," https://www.boesendorfer.com/en/pianos/disklavier-edition (Accessed Oct. 19, 2022).

[26] "FLAC," https://xiph.org/flac/ (Accessed Oct. 19, 2022).

[27] J.-M. Valin, G. Maxwell, T. B. Terriberry, and K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec," presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), paper 8942. https://doi.org/10.48550/arXiv.1602.04845.

[28] "Mido - MIDI Objects for Python," https://mido.readthedocs.io/en/latest/index.html (Accessed Oct. 19, 2022).

[29] E. Pampalk, "A Matlab Toolbox to Compute Music Similarity From Audio," in *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, paper 180 (Barcelona, Spain) (2004 Oct.).

[30] E. Nakamura, K. Yoshii, and H. Katayose, "Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment," in *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, pp. 347–353 (Suzhou, China) (2017 Oct.).

[31] D. Stowell and E. Chew, "Maximum a Posteriori Estimation of Piecewise Arcs in Tempo Time-Series," in M. Aramaki, M. Barthet, R. Kronland-Martinet, and S. Ystad (Eds.), *From Sounds to Music and Emotions*, Lecture Notes in Computer Science, vol. 7900, pp. 387–399 (Springer, Berlin, Germany, 2013).

[32] E. Chew, "Playing With the Edge: Tipping Points and the Role of Tonality," *Music Percept.*, vol. 33, no. 3, pp. 344–366 (2016 Feb.). https://doi.org/10.1525/mp.2016.33.3.344.

[33] "XmlTensionVisualiser," https://dorienherremans.com/tension (Accessed Oct. 19, 2022).

[34] D. Herremans and E. Chew, "Tension Ribbons: Quantifying and Visualising Tonal Tension," in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR)*, pp. 8–18 (Cambridge, UK) (2016 May).

[35] "CouchDB," https://couchdb.apache.org/ (Accessed Oct. 19, 2022).

[36] T. Bray, "The JavaScript Object Notation (JSON) Data Interchange Format," RFC 8259 (2017 Dec.). https://datatracker.ietf.org/doc/html/rfc8259 (Accessed Oct. 19, 2022).

[37] "Web Audio API," https://developer.mozilla.org/en-US/docs/Web/API/Web_Audio_API (Accessed Oct. 19, 2022).

[38] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 Data-Driven Documents," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2301–2309 (2011 Dec.). https://doi.org/10.1109/TVCG.2011.185.

[39] "W3C SVG Working Group," https://www.w3.org/Graphics/SVG/ (Accessed Oct. 19, 2022).

[40] "PouchDB," https://pouchdb.com/ (Accessed Oct. 19, 2022).

[41] "Window.requestAnimationFrame()," https://developer.mozilla.org/en-US/docs/Web/API/window/requestAnimationFrame (Accessed Oct. 19, 2022).

[42] C. Wilson, "A Tale of Two Clocks," https://www.html5rocks.com/en/tutorials/audio/scheduling/ (2013 Jan.).

[43] "AudioBufferSourceNode," https://developer.mozilla.org/en-US/docs/Web/API/AudioBufferSourceNode.

[44] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Routledge, New York, NY, 1998). https://doi.org/10.1201/9781315140919 (Accessed Oct. 19, 2022).

[45] M. A. Hartmann, *Modelling and Prediction of Perceptual Segmentation*, Ph.D. thesis, University of Jyväskylä, Jyväskylä, Finland (2017 Jan.).

THE AUTHORS


Lawrence Fyfe


Daniel Bedoya


Elaine Chew

Lawrence Fyfe is a research engineer who creates Web-based visualization software and database infrastructure. Lawrence received his Ph.D. in Computational Media Design from the University of Calgary and a Master's degree in Music, Science and Technology from the Centre for Computer Research in Music and Acoustics (CCRMA) at Stanford University. Before joining the COSMOS project, he worked on a binaural telepresence system for the Digiscope project at the Institut National de Recherche en Informatique et en Automatique (INRIA). The Digiscope project connected various visualization labs around Paris via telepresence (audio and video conferencing) to facilitate collaboration.

•

Daniel Bedoya is a Ph.D. student who designs citizen science experiments to help understand musical structures created in performance and analyzes the perception of musical structures in performed music and physiological responses to these performed structures. He has an undergraduate degree in Sound Engineering [Universidad de Las Américas (UDLA) Quito, Ecuador] and a Master's degree in Computer Science, Acoustics and Signal Processing Applied to Music [Acoustique, Traitement du Signal, Informatique, Appliqués à la Musique (ATIAM); Institute for Research and Coordination in Acoustics/Music (IRCAM); Sorbonne Université]. Previously, he was a research assistant with Jean-Julien Aucouturier in the Perception and Sound Design (PDS) Team at IRCAM working on the relationship between music and emotions in the European Research Council (ERC) project CREAM and explored the influence of smiled speech in dyadic interactions in the REFLETS project.

•

Elaine Chew is the principal investigator of the European Research Council (ERC) Advanced Grant project COSMOS. Her research centers on the mathematical and computational modeling of musical structures, with a present focus on structures as they are communicated in performance and in electrocardiogram traces of cardiac arrhythmias. As a pianist, she has collaborated with composers to create and premiere new works, and she frequently designs and performs in concerts that present visualizations and compositions created by her research team. She is a past recipient of Presidential Early Career Award for Scientists and Engineers (PECASE)/CAREER awards and fellowships at the Radcliffe Institute for Advanced Studies at Harvard. Her research has been supported by the ERC, Engineering and Physical Sciences Research Council (EPSRC), Arts and Humanities Research Council (AHRC), and National Science Foundation (NSF) and featured on BBC World Service/Radio 3, *Smithsonian*, *Philadelphia Inquirer*, WIRED Blog, *MIT Technology Review*, and *LA Phil: Inside the Music*.