Quasar Spectroscopy Sound: Analyzing Intergalactic and Circumgalactic Media via Data Sonification

BRIAN HANSEN, JOSEPH N. BURCHETT, AND ANGUS G. FORBES
(brmhanse@ucsc.edu) (burchett@ucolick.org) (angus@ucsc.edu)

University of California, Santa Cruz, USA

In this paper, we present sonification approaches to support research in astrophysics, using sound to enhance the exploration of the intergalactic medium and the circumgalactic medium. Astrophysicists often analyze matter in these media using a technique called absorption line spectroscopy. Our sonification approaches convey key spectral features identified via this technique, including the presence and width of spectral absorption lines within a region of the Universe, the relationship of a particular redshift location with respect to the absorption peak of a spectral absorption line, and the density of gas at various regions of the Universe. In addition, we introduce *Quasar Spectroscopy Sound*, a novel software tool that enables researchers to perform these sonification techniques on cosmological datasets, potentially accelerating the discovery and classification of matter in the intergalactic medium and circumgalactic medium.

0 INTRODUCTION

Astronomical observations peering beyond the Milky Way have revealed billions of other galaxies in addition to our own. These galaxies are primarily detected by the starlight they emit. However, the matter detectable via emission from stars represents a small fraction of the matter that composes the Universe, even when we neglect the contribution of dark matter, which dominates the matter budget [1]. Indeed, the majority of "normal" matter (consisting of protons, neutrons, and electrons) resides *outside* the luminous starlit regions of galaxies, within the circumgalactic medium (CGM) and intergalactic medium (IGM) [2, 3, 4, 5].

This material is comprised of ionized gas that is so diffuse that it is undetectable through the light it emits. Instead, this tenuous gas must be detected via absorption line spectroscopy [6, 7]. This technique involves observing a bright emitting source in the distant universe, such as a quasar, and obtaining a spectrum wherein the object's light is dispersed into its constituent wavelengths (see Fig. 1). Foreground complexes of gas leave a signature on the spectrum wherein light from the background source is removed at specific wavelengths corresponding to the atomic energy transitions of chemical species (atoms and ions) found in the gas "cloud". Because these transition wavelengths are set by well-understood atomic physics, species such as neutral hydrogen have characteristic spectral "fingerprints" that enable the chemical composition of intergalactic and circumgalactic gas to be identified.

Astrophysicists study the IGM and CGM for a myriad of reasons. The majority of non-dark matter is purported to reside in the IGM and CGM, and there is an ongoing initiative to obtain a census of this ionized gas to validate predictions from the prevailing cosmological theories [8, 9, 10]. Furthermore, understanding the chemical composition of this gas can constrain models of how elements such as carbon, oxygen, magnesium, and iron (those with atomic number greater than helium) propagate throughout the Universe [11, 12]. Lastly, IGM and CGM gas play critical roles in the evolution of galaxies. Gas flows from the IGM and CGM are galaxies' "lifeline", as they cannot form new generations of stars without them [13, 14]. Conversely, the explosive deaths of stars expel matter and energy from galaxies into the CGM and IGM, modulating the mass, temperature, and density of gas within galaxies [15, 16]. Any self-consistent theory of how galaxies form and evolve must reproduce conditions in these media.

From the observational perspective, the task at hand centers on first identifying and then analyzing the absorption lines in spectra. Identifying IGM and CGM spectral lines essentially involves two key properties: the chemical species and the velocity at which the detected gas complex is moving. The former is generally determined by the aforementioned spectral energy transitions set by atomic physics. As the gas is predominately ionized, one typically

encounters species such as O VI, C IV, or Mg II. In this nomenclature, an ion is denoted by the atomic symbol of the element then a numeral indicating the ionization state. For example, O I represents neutral oxygen (equal numbers of electrons and protons), O II is singly ionized oxygen (one electron stripped away), and O VI denotes an oxygen nucleus with five electrons missing.

The second key property of an absorption line is the velocity of the gas it traces. The velocity is reflected in analogy to the Doppler effect, where sound waves emitted from a source in relative motion to an observer appear to decrease or increase in frequency. In spectroscopy, lines from an absorbing or emitting source with relative motion to the observer will appear *redshifted* or *blueshifted* and occur at longer or shorter wavelengths, respectively, than the characteristic wavelengths of their atomic transitions. For this reason, we will refer to these characteristic wavelengths as *rest-frame wavelengths*.

Quasar spectral analysis is exceedingly challenging in terms of size and complexity. Currently, there exists an ever-growing repository of over 600 quasar spectra obtained via observation from the Hubble Space Telescope. Each spectral dataset is saturated with features that contain information on the entirety of physical space that lies between us (planet Earth) and distant quasars observed in the universe. However, the richness of the data also makes it difficult to identify and extract specific features within the spectrum (see Fig. 1). As the data are analyzed for a myriad of purposes in the astrophysics community, there is an ongoing effort to develop and utilize novel approaches for analysis and discovery.

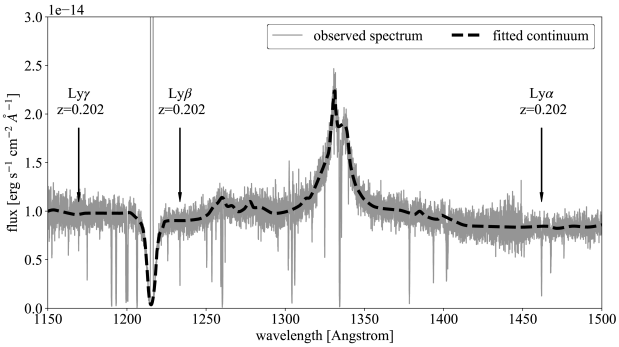

Fig. 1. Segment of the TON580 quasar spectrum, the fiducial quasar used in developing our sonification analysis platform. The observed spectrum, obtained with the Hubble Space Telescope, is plotted in gray, while the fitted continuum is overplotted in red. The continuum is a model of the intrinsic shape of the quasar emission, from which absorption signatures appear as strong spikes protruding downward. The key absorption lines we sonify for neutral hydrogen, Lymanα, Lymanβ, and Lymanγ, are marked at a redshift of 0.202. For reference, these same spectral transitions are also featured in Figs. 2 and 3. A primary task for the spectroscopist is to identify absorption systems such as the one labelled among the multitude of lines arising from gas complexes all along the line of the sight to the quasar. As in this example, multiple spectral features in an absorption system are related via known relationships between their locations in wavelength space (horizontal axis). We leverage sonification to superpose related absorption features (expressed visually in Fig. 2) and to reveal the presence of and characterize absorption systems, a traditionally challenging and time consuming task.

In this paper, we present sonification approaches for the analysis of quasar spectroscopy, emphasizing how sonification can augment and accelerate the study of IGM and CGM. Specifically, our contribution introduces four interrelated techniques to: *a)* identify the presence of spectral absorption lines, *b)* characterize the width of spectral absorption lines within a region of the Universe, *c)* compare the distance of a particular redshift location with respect to the absorption peak of a specified spectral absorption line, and *d)* highlight the density of gas at various regions of the Universe. Further, we present an interactive software analysis tool, *Quasar Spectroscopy Sound* (QSS), that encapsulates these sonification techniques, enabling researchers to effectively investigate quasar spectrum datasets.

1 BACKGROUND AND RELATED WORK

Sonification has made numerous appearances in the realms of both astronomy and spectroscopy. In both fields, it has been shown to provide a complementary perspective to existing analysis techniques and to enhance discovery.

1.1 Astronomy Sonification

In the field of astronomy, sonification has demonstrated the ability to lead to new discoveries, enhance visual representations, and broaden the accessibility of data analytics. For example, as the Voyager 2 space probe was traveling across the rings of Saturn there was a problem with the space craft. The problem could not be discerned with visual controls due to the large amount of noise present in the signal. However, when realized aurally through a digital synthesizer, a salient sonic pattern emerged described as a "machine gun" sound. This led to the discovery that the problem was caused by collisions with electromagnetically charged micrometeoroids [17].

In 2009, Greg Laughlin, an astronomer at the University of California, Santa Cruz developed a sonification tool called *Systemic* that helps researchers to detect extra-solar planets from data acquired by the Kepler Telescope. The tool uses a range of data analysis techniques in tandem with sonification, where potential planetary orbits can be identified from the harmonics in the sonified data [18].

In 2012, Alexander et al. demonstrated success audifying data from the Solar Wind Ion Composition Spectrometer. While listening to the raw solar data, they detected an underlying "hum" at a frequency of 137.5 Hz. Upon closer inspection, it was discovered that the sound corresponded to solar rotations and had implications for the carbon ion composition of solar winds. The harmonics in the tone present indicated periodic changes in temperature and hence solar wind type, allowing for the study of the temporal evolution of the Sun's wind source regions [19].

The blind astronomer Wanda Diaz Merced has performed considerable research in data sonification related to astrophysics data [20]. Diaz Merced sonified data using the tool *xSonify* [21], first created at the NASA Goddard

Space Flight Centre and later converted to a user-centered prototype at the University of Glasgow in Scotland. In her 2013 PhD thesis, she summarizes the organization of focus groups to test her sonifications with the tool, finding that it gives scientists a better understanding of the data [22].

1.2 Spectroscopy Sonification

Sonification has also exhibited many unique representations in the field of spectroscopy, ranging from pure audification to musical composition. Newbold et al. [23] and Morowitz [24] have sonified the structure of molecules through the audification of Nuclear Magnetic Resonance (NMR) spectra. An NMR spectrum plots the spectral intensity of a chemical shift, and peaks present in the spectrum are converted directly into the frequencies of sinusoids in an audio signal. Each chemical has a unique spectrum and, thus, a unique sound when audified. This procedure has shown to be advantageous as it allows for fast and efficient detection of chemicals present in large amounts of data.

Similar to the above approach, Terasawa et al. [25] present a sonification of ECoG Seizure Data where the data are parameter mapped to overtones in a sound spectrum. The authors label this approach as "gestalt formation" operating as a means of applying semantics to sound. In the approach, the overall timbre of a sonority is a representation or display of features, where the spectral characteristics of the sound signify the characteristics of the source. In their implementation, fifty-six channels of ECoG data were monitored. To sonify the data, a fundamental frequency of 180 Hz was selected, and sixteen harmonics of sinusoids were realized. Each harmonic was then amplitude modulated by each channel of ECoG data.

Pietrucha [26] takes a musically based parameter mapping approach in his sonification of spectra. In this approach, plots of absorption vs wavelength are presented, where spectral peaks in terms of absorption strength are mapped to frequency. The wavelength at which the peak is present is represented by a chord, where the harmonic quality of the chord indicates its location along the wavelength axis. The domain of wavelengths is divided into bands, where wavelengths residing in the shortest band are represented by a tonic chord, and subsequent bands utilize other diatonic chords, such as the subdominant, dominant, or submediant. The result is highly musical, with a melodic succession representing peak intensities that is accompanied by a chord progression representing wavelength location.

The *Atom Tone* project explores aesthetic possibilities for spectral sonification through music composition. Sonic material is generated through synthesis and modulation. Tones are generated via the audification of atomic emission spectra similar to the NMR approach above. The generated tones are then modified via effects processing techniques such as waveshaping, frequency shifting, or filtering, where the modification parameters are based on data taken from the periodic table of elements. The modulation is conceptually more open-ended, with the goal being for the musician to find the most aesthetically pleasing result and to have many options to explore [27].

Martin et al. [28] devised an approach to enable the visually impaired to examine audio spectrograms. With this sonification, a user is instructed to first select a frequency band to be monitored, and then set a threshold level in decibels. If a spectral peak in the selected band exceeds the threshold, a short beep will sound, where the frequency of the beep corresponds to the decibel level above the threshold.

2 QUASAR SPECTROSCOPY SONIFICATION

Here, we present a sonification approach for the analysis of quasar spectroscopy, with particular application to studying the IGM and CGM. Despite the rich history of sonifcation in the field of astronomy, there is currently no tool or sonifcation approach specifically tailored to the analysis of quasar sightline spectroscopy. Our design follows best practices in sonification related to the effective parameter mapping of spectral features to the audible domain [29]. Specifically, we identify and sonify parameters in quasar spectrum data sets to make it easier to identify the presence, magnitude, symmetry, and alignment of spectral absorption lines, and to highlight their characteristic strengths. As we describe below, this information enhances the astrophysicist's ability to find meaningful patterns within quasar sightline spectroscopy data.

2.1 Data Configuration

Quasar spectrum data sets are available from the Hubble Spectroscopic Legacy Archive (HSLA) [30], which serves as a database of science-ready spectra obtained with the spectrographs on board the Hubble Space Telescope. Each spectrum includes the intrinsic signature of the quasar itself, which 1) adds a "bumpy" underlying shape to regions local to the features of interest for analysis and 2) reflects the intrinsic brightness of the quasar, which for our purposes serves only as a background "light bulb". These bumpy features need to be removed in order to reveal the net absorption or emission signatures in the spectrum and putting those signatures on a common normalized scale so that the flux values at different positions in the spectrum may be compared to one another. Thus each spectrum, once downloaded from the HSLA, is preprocessed using the *linetools* software package [31]. We fit a continuum over the full wavelength range of the spectrum, then normalize the data by dividing by the continuum, and extract the normalized spectrum as an ASCII table with the parameters of wavelength, flux, and error.

Our main goal in sonifying quasar spectra is to detect gas in the IGM and CGM and localize it in velocity. Our primary observational technique for studying the IGM/CGM is via absorption lines in spectra to background quasars, where we search for absorption signatures of spectral lines that correspond to particular ions we seek. Each quasar spectrum probes the IGM/CGM gas along the line of sight, where the observables for each spectrum include a flux and

uncertainty value for each wavelength recorded at the detector. Before sonification, the data are reduced, all exposures are coadded, and we normalize the flux by dividing by an approximation to the intrinsic shape of the quasar itself (the continuum). The normalized flux will take on values ranging from 0 to 1, where smaller flux values indicate stronger absorption signatures. The location of each absorption feature is determined by the velocity of the gas. This velocity is dominated[1] by the expanding Universe, which shifts spectral lines to longer (redder) wavelengths, and we generally quantify this velocity first by a *redshift z*:

$$\lambda_{obs} = \lambda_{rest}(1 + z) \quad (1)$$

where λ_{obs} is the wavelength at which a feature is observed and λ_{rest} is the rest frame wavelength of the spectral transition.

The ions available for selection are derived from the atomic data set compiled by Verner et al. [34]. This data set contains ion names and their spectral transition properties of rest-frame wavelength, oscillator strength, and a dimensionless absorption strength parameter,[2] which we use for data filtration. As our aim is to focus on absorption systems most commonly encountered outside of our galaxy, we filter out ions and spectral transitions that are least likely to be found (informed by past experience analyzing similar data through traditional techniques) in the IGM/CGM gas we are studying. We omit ions that are considered to be fine-structure and excited-state, as these materials are mainly observed residing in our galaxy. In addition, we filter on the properties of absorption strength and rest-frame wavelength, only importing spectral transitions with an absorption strength parameter greater than 9.75 and with wavelengths between 900 and 1800 Angstroms. These values are commensurate with the sensitivity and spectral coverage of the *Hubble Space Telescope* instrument that collected our initial dataset used in this application; i.e., spectral lines meeting these criteria are potentially detectable given the signal-to-noise ratio and wavelength coverage of our sample spectra. Finally, we aim to analyze only three spectral transitions for a given element. If a particular element has more than three spectral transitions that meet our initial criteria, we only preserve the spectral transitions with the three highest absorption probabilities.

2.2 Presence of Absorption

The primary feature we aim to sonify is the presence of spectral absorption lines. Certain ions present within a gas cloud may have multiple signature transitions (λ_{rest}) for which our data have spectral coverage; each spectrum covers a particular range of (λ_{obs}). To corroborate the identification of a given ion, we attempt to scan for up to three lines from the ion. Certain ions may have more than three lines covered in our spectra, and in this case, we scan for the three lines with the highest absorption strength (determined by atomic physics). Through sonifying absorption presence, our aim is to conduct a first exploratory pass with the goal of finding a potential ion candidate, which can then be verified using additional sonification techniques, described in Sections 2.3-2.5 below.

Our approach is inspired by Terasawa et al. [25], in that we define a harmonic construct above a fundamental frequency for our parameter mapping and rely on the notion of "gestalt formation" as a means to identify features in the data. In our case, the harmonic construct we define is a root position major triad with a selected fundamental frequency of 220 Hz. The triad is in closed voicing with chord tone intervals of the third and fifth in just intonation having frequencies of 275 Hz and 330 Hz respectively. The voicing of the triad is closed to mitigate the perception of tonal fusion [35], as we aim to make each tone within the chord more clearly distinguishable. As our triad sounds as a continuous drone, we selected these frequency because they are low enough to mitigate irritation and fatigue for prolonged listening. In addition, they are high enough to be minimally impacted from low frequency roll-off of an average laptop speaker.

Our approach focuses our search to one ion at a time. We select a particular ion of interest and map each of its spectral lines to one of the tones in the triad, which we refer to as the *presence frequency*. Since we are only concerned with the three most prominent spectral lines of an ion (determined by absorption strength), the spectral line with the highest absorption strength is assigned to the root of the triad, the second highest strength is mapped to the interval a perfect fifth above the root, and the third highest strength line is mapped to the major third of the triad.

The triad is a useful sonic construct as it not only allows for a 1-to-1 mapping between pitches and the spectral transitions we seek, but also, as supported by Huron [36] and Parncutt [37], has a limited pitch density (being 3 tones) to allow for the perception of each individual tone. Additionally, each tone in the triad is assigned a unique timbre. As noted by Huron [36] and demonstrated by Wessel [38], assigning unique timbres to tones enables them to be more easily perceived and distinguished, which is especially important to facilitate listening for the presence of multiple tones simultaneously. The expectation is that the first spectral transition will have the strongest absorption signature and thus yield the strongest amplitude. The second transition will have a weaker signature and the third transition the weakest. Thus, we *increase* spectral richness to each tone as the expected amplitude *decreases* so that a weaker tone can be better perceived among its neighbors. Specifically, the root is designated as a sine tone, the fifth is a square wave, and the third is a sawtooth wave [39].

Next, the amplitude for each tone is determined by the amount of absorption present in a corresponding line. At a

[1] There are effectively three components to the velocity of the gas: 1) that imparted by the expansion of the Universe, wherein the velocity increases with the distance of the object from Earth; 2) motions of gas clouds due to, e.g., gravitational attraction to nearby galaxies; and 3) random thermal motions of particles within a gas cloud [32, 33].

[2] Verner et al. [34] define this parameter P combining the intrinsic probability of the particular atomic transition and the corresponding ion's abundance.

given redshift, we map the flux at that pixel to amplitude as follows:

$$A = 1 \quad F_{rs} \tag{2}$$

where F_{rs} is the flux obtained for a spectral line at a given pixel, denoted by the redshift (see Fig. 2). Thus, as the absorption increases, the chord tone for the associated line becomes more pronounced.

We also consider the perception of loudness as it is mapped to each tone of the triad. First, for each absorption line, we convert the amplitudes derived from flux to sound pressure level based on an assumed reference amplitude of 10 ^{5} [39]. Next, we determine the corresponding loudness levels in phons, using our selected root frequency of 220 Hz as a base. Then, we use equal loudness contours [40] to adjust the sound pressure levels for perceived loudness of the chord tone frequencies of 275 and 330 Hz. The resulting sound pressure levels are then re-converted to amplitude values. This process allows the audible perception of each chord tone to more accurately reflect the measured proportional differences in flux among spectral transitions.

As a user scans through the spectrum in redshift space searching for the presence of a particular ion, each chord tone varies in intensity. For ions with three spectral lines covered by the data, the goal is for the user to find a location (in redshift) where they hear the full presence of the major triad. That is, hearing the simultaneous presence of every chord tone indicates to the user that they have identified the location of the ion they seek.

It is important for the user not only to listen for the presence of the absorption lines but also to listen to the balance of amplitude among the tones present. The spectral transitions exhibit a unique proportion of absorption among the lines (once again set by atomic physics), and the overall mix of the triadic tones reflect this relationship. The mix of amplitude and timbral mapping in a properly aligned spectrum results in a uniquely perceived gestalt formation. When scanning through other localities of the spectrum, absorption patterns that do not form proper alignments yield a distinctly different sonority due to the different mix of amplitude and timbral emphasis of the triadic tones. This enables the listener to quickly identify the target gestalt signature, increasing the efficiency of ion detection.

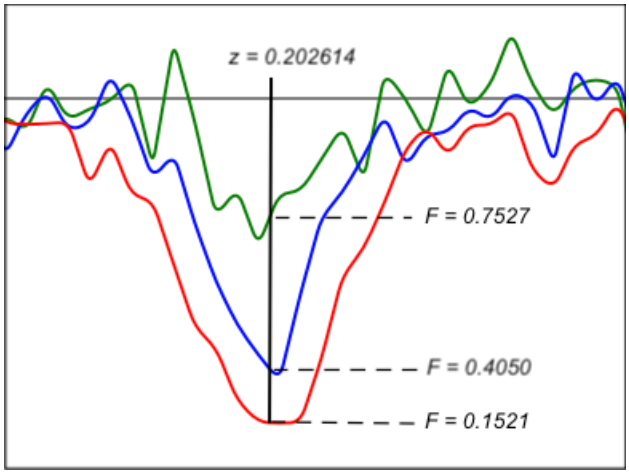

Fig. 2. Here we see the absorption feature alignment for neutral hydrogen, where the lignment of absorption lines (shifted to their rest frames and superimposed) for H I is at a redshift z = 0.202614. The flux value for the first spectral line (red) is 0.1521 (amp = 0.8479), the second line (blue) is 0.4050 (amp = 0.595), and the third (green) is 0.7527 (amp = 0.2473). Note: realized amplitude values for the root, 5th, and 3rd of the triad adjusted for equal loudness perception are 0.8479, 0.4794, and 0.2073 respectively.

2.3 Absorption Linewidth

Ancillary to absorption presence, the width of each absorption line is also sonified, reflecting the overall amount of absorption present and asymmetry in the absorption profile. We determine the absorption width as the upper and lower bound wavelengths that correspond to half the flux deviation from 1.0 observed at a given redshift (see Fig. 3).

For this sonification, a dual tone, bi-directional glissando is constructed that represents the overall width and skew of the absorption signature. As opposed to the continuous drone of the absorption presence, this sonic construct has an ADSR envelope (up to 1 second total duration) and dynamic frequency movement. These features yield a unique pitch motion and onset synchrony [36], producing a distinct temporal coherence [41] that results in a gestalt formation that is easily distinguishable for listeners.

First, we determine the bounding tones of the glissando. For each line, the spectral width upper and lower bounds are mapped to frequencies that are above and below the presence frequency. The glissando tones simultaneously emerge from this base and then sweep the frequency space until arriving at their respective boundaries. The width of the absorption, as represented by the glissandi, is characterized by duration and distance. A broader absorption line will have glissandi that cover a wider sonic frequency range and that is longer in duration. Moreover, each glissando differs in duration. This is an indication of asymmetry (or skew) in the absorption width. For example, if one glissando is longer in duration traversing the higher fre-

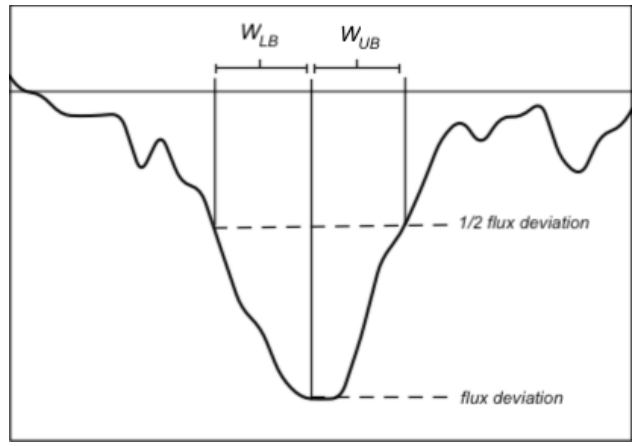

Fig. 3. Here we see the spectral absorption width for an absorption line with width at 1/2 the flux deviation from 1 for a given redshift. The total width is divided into upper and lower sections centered about the current red shift.

quency than the lower, then the upper bound of the absorption is proportionally further from the line center than the lower bound. The skewness of the absorption is an important feature to perceive, as it may indicate that our scan is not centered at the absorption peak, or it may indicate contamination from absorption of a different ion at a different redshift.

2.4 Centroiding

Upon detecting the presence of a particular spectral transition, we can perform centroid sonification. This allows us to hear how well aligned the currently scanned redshift location is with respect to the absorption peak. To accomplish this, for each line we introduce a centroid tone and compare it to the presence tone. As with the spectral linewidth, we compute a neighborhood about the current redshift and then determine upper and lower bound wavelengths at half of the flux deviation at our redshift location. With the absorption width calculated, we then compute the ratio between the widths above and below the current redshift. The centroid tone is then introduced, whose frequency is the presence frequency multiplied by the calculated ratio:

$$f_{centroid} = f_p \frac{W_{UB}}{W_{LB}} \tag{3}$$

where f_p is the presence frequency for the spectral line, W_{UB} is the upper bound absorption width, and W_{LB} is the lower bound absorption width.

If our redshift is centralized about the current absorption area, the ratio between upper and lower bound widths will equal 1, causing the centroid frequency and presence frequency to sound in unison. If not, the added tone will generate a dissonance that is above or below our tone of reference. If the centroid tone sounds higher in pitch, then we are positioned above the absorption center and can adjust downwards to find the line centroid. If the centroid tone sounds lower, then we can re-position upwards to centroid. In addition, our framework allows us to scan redshift space at various velocity intervals, the coarsest being 25 km/sec and the finest being 0.1 km/sec (recall that the redshift is merely an expression of velocity due to the Universe's expansion). The added tone heard specifically produces the perception of sensory dissonance (dubbed "tonal consonance" by Plomp and Levelt [42]), which is characterized by beating or roughness between the two tones. When adjusting the velocity of the reference tone, a change the quality of the dissonance between the tones can be heard. As the user moves away from the centroid, the beating increases in frequency, causing an increasingly rougher sonority. The user can finely adjust the velocity position of our reference tone, listening for the beating to slow until the roughness diminishes and ultimately transforms into a unison.

2.5 Apparent Column Density

As stated above, the various absorption lines associated with a particular ion will have differing characteristic strengths. Although absorption line strength is mapped to amplitude in the presence sonification (described above in Section 2.2), the precise differences among lines can be difficult to perceive for comparison purposes when represented via amplitude. Thus, we devise a sonification that allows a user to evaluate and compare specifically the absorption line strengths for a given ion, and as a result validate potential ions discovered via the presence sonification.

We can account for the differing intrinsic absorption strengths using the apparent optical depth at each pixel, from which we then calculate the Apparent Column Density (ACD) [43]. The ACD describes the density of gas intercepted along the line of sight in particular region of space. Comparing the ACD in each pixel for multiple spectral lines can be used to verify that our detected absorption profiles match for the ions we seek.

When observing a given spectral transition at a particular redshift, ACD relates its oscillator strength, restframe wavelength, and absorption such that when this relationship is compared among all spectral transitions for a given ion, they should be equal. We calculate the ACD for each absorption line as follows:

$$ACD \propto \frac{1}{O_s \lambda_{rest}} ln(\frac{1}{F}) \tag{4}$$

where, for a given ion, O_S is the oscillator strength, λ_{rest} is the rest-frame wavelength, and F is the absorbed flux, in terms of flux at a given redshift. If the calculated ACD for all spectral lines covered for a given ion are equal, then this is strong validation we have accurately identified an ion.

To sonify the ACD, we first compare the percentage change in ADC between successive absorption lines as follows:

$$\Delta ACD_n = \frac{ACD_n}{ACD_{n\ 1}} \quad 1 \tag{5}$$

where ACD_n is the apparent column density for spectral line n. Then, the ACD frequency for a given absorption line is calculated as:

$$f_{ACD} = f_p \quad 2^{\Delta ACD \frac{1}{6}} \tag{6}$$

where f_p is the presence frequency for a given spectral line, and ΔACD is the ACD percentage change for the line.

The user hears the ACD frequency in comparison to the presence frequency assigned to the spectral transition. If the ACD for a particular transition aligns with the others, then the ACD frequency will sound in unison with the base. If it is not, the presence of sensory dissonance [42] will be heard. Thus overall, when listening to the presence of all spectral lines at a given redshift, if they all align in presence as well as in absorption (proportionally), a fully in-tune triad will sound.

3 QSS SOFTWARE

Quasar Spectroscopy Sound (QSS) is an interactive software tool for detecting and characterizing gaseous ions in the intergalactic and circumgalactic medium. Development of the software resulted from a six-month long iterative process informed by expert users with clearly defined analysis tasks. In its current state, QSS is intended for astrophysicists investigating absorption line spectroscopy. How-

ever, we expect that the soni cation approach enabled by QSS will be useful to researchers in other domains that investigate spectroscopy data. The software utilizes the JUCE C++ application framework for the development of the user interface and for audio processing. The soni cation audio results from an ensemble of direct digital signal processing techniques. The tool was designed to facilitate ef cient interactive analysis, and as such it has a compact, intuitive interface consisting of a single window with only essential controls (see Fig. 4). The interface consists of three main components—Data Con guration, Audio Controls, and Data Navigation—which are clearly marked to highlight functionality.

The QSS software facilitates a work ow for ion discovery. After importing a dataset and selecting an ion for analysis, the user should rst enable the "Presence" sonifcation for each spectral transition. With this enabled, the user then uses the navigation tools to scan the redshift location of the spectral data until a candidate for the ion signature is detected. Once a candidate is detected, the user may then utilize the remaining sonifcation tools to further evaluate the absorption signature and to con rm the candidacy of the ion location. This is accomplished by enabling the "ACD" sonifcation, which veri es that absorption lines detected are in the expected relation for the ion. In addition, enabling the "Centroid" soni cation enables a user to hone in on the location of the ion absorption signature. Finally, triggering the "Width" glissandi is used to convey a sense of the overall skew of the absorption signature.

Fig. 4. The QSS software user interface with sections including Data Con guration (top), Audio Controls (upper middle), Data Navigation (bottom middle), and spectral absorption plot (bottom).

3.1 Data Con guration

The Data Con guration component allows a user to import a desired data set and select an ion for spectral analysis. Loaded data sets obtained from the Hubble Spectroscopic Legacy Archive (HSLA) must be formatted to the speci cations as detailed in Section 3.1 above. Upon importing the data set the user may select an individual ion for analysis, where QSS performs the ion ltering criteria also de ned in Section 3.1. Once the data is correctly imported, boxes for soni cation controls appear below, displaying the name and rest frame rate associated with each ion.

3.2 Audio Controls

The Audio Controls component includes standard controls to enable playback and control the overall volume level of the audio. To the right of these controls are boxes that allow the user to control audio features of the spectral soni cation. The rst of the controls is a "Presence" toggle button for listening to the absorption presence. Enabling this allows the users to hear the absorption strength, perceived in terms of amplitude, of the spectral line present at a selected redshift. Next, there is a "Centroid" button that, when enabled, allows the user to hear the positioning of the selected redshift with respect to the spectral width as described in Section 2.4. Third, an "ACD" toggle button enables the soni cation for apparent column density, allowing the user to hear how a given spectral transition is aligned with respect to others of the same ion. Finally, there is a "Width" trigger button. Clicking on this button triggers a glissando that represents the absorption linewidth for given neighborhood in redshift.

3.3 Data Navigation

The Data Navigation component consists of user interface widgets that allow the user to scan the spectral data and a visual spectral plot. First and foremost, at the top of the section, the user sets a redshift anchor by either entering a value in the number box or adjusting the horizontal slider that encompasses the redshift range of the ion being analyzed. Scrubbing the data via redshift alone can be problematic as even small increments in redshift account for large changes in velocity, which can cause the user to skip over sections in the data too quickly and thus miss many absorption features. Thus, in addition to the slider, we have included a velocity jog wheel that allows the user to scrub the data at ner level of detail. With this tool, the user can explore the absorption in the neighborhood of a local spectral region at varying intervals ranging from velocity increments of 0.1 to 25 kilometers per second. The jog wheel is essential for exploring the centroid, width, and apparent column density soni cations.

Directly underneath the navigation widgets is a visual plot of the spectral data. The plot appearing in the window shows the spectral neighborhood about the currently selected red shift. The absorption lines for (up to) all three spectral transitions are superimposed, allowing the user to more easily see areas of absorption alignment. This visual-

ization appeals to more traditional spectral analysis techniques and enhances data navigation while serving as a complement to the soni cation.

3.4 Development Process

The development of our software and its corresponding analysis methodology proceeded via an iterative interdisciplinary design process. Our team includes experts in computer music, audio engineering, data science, and astrophysics. Our astrophysics collaborator selected the dual spectrum (TON580) employed in this study from the HSLA based on this spectrum's high signal-to-noise ratio and wide spectral coverage. TON580 also has been analyzed previously through classical (primarily visual) techniques.

After the initial sound design and version of the QSS software, our astrophysics collaborator immediately identi ed two key advantages over commonly employed methods. First, users can scan through redshift space to identify absorption presence much more quickly via soni cation, than they can via more traditional visual analysis, which can be time consuming and can lead to visual fatigue. Second, users can quickly identify even weak spectral features which "pop out" from among the noise using our approach, indicating that soni cation is more effective in identifying these features than is visual inspection of the spectrum. During this initial exploration, the absorption system at $z = 0.202$ (featured in Figs. 1–4) could be identi ed due to its aural prominence, and we then expanded the functionality to incorporate additional analysis tasks beyond merely identifying an absorption system. The second stage of development involved more detailed analyses of absorption systems, such as apparent column density pro les, centroiding, and line widths. Finding appropriate soni cation techniques for these additional spectroscopic features involved several design iterations before nding and honing effective sonic constructs, such as, for example, the glissando mapping to line width.

3.5 User Feedback

In addition to the ongoing feedback from our astrophysics collaborator during the development process, we sought feedback from other domain experts regarding the QSS software tool, including those without previous musical training. For example, one astrophysicist— a Professor of Astronomy and Astrophysics who regularly works with spectroscopy data— immediately reported an enhanced perception of absorption features and, after only a few minutes of training, could also successfully perceive meaningful characteristics of the centroid, ACD, and width modalities. They especially appreciated the "clever" glissando representation for linewidths, noting that while it can be dif cult to perceive subtle asymmetries in line pro les when examining them visually, it was easy to identify the asymmetry using our glissando approach. This has important implications, as the cutting-edge limit of resolution for the most sensitive spectrograph on the Hubble Space Telescope remains below that capable of resolving bulk motions of gas complexes in galaxies. Overall, they were positive about the soni cation approach, nding QSS to be "an excellent addition" to existing visual analysis tools. They also came up with additional analysis tasks for our soni cation tool, such as redshift validation, and indicated that they believed our tool could be useful for pedagogical purposes, topics we plan to explore in the future.

4 CONCLUSION

Initial analyses using quasar spectroscopy soni cation with the QSS software have shown potential to be a valuable resource for astrophysicists conducting research into intergalactic and circumgalactic media. Currently, the prevalent tools available for analysis reside in the visual realm, and the introduction of a sonic perspective not only enhances visually-oriented tools, but also can provide unique advantages as a standalone analysis methodology. For example, it became immediately clear early in the development process that QSS can enable more rapid discovery and identi cation of IGM/CGM system candidates than visually scanning through spectra. In addition, early testing indicates that hearing the presence of absorption features can improve sensitivity to weaker lines, as it can be challenging to visually identify these weak lines among the "noise" in the spectra. These advantages increase the discovery space in re-analyzing already existing archival data, and anticipate the next generation of spectroscopic surveys on the horizon.

Although our soni cation approaches and software have already undergone substantial development based on expert feedback, we continue to seek additional feedback from a broader user community made up of astrophysicists with varying specializations and at different career stages to further enhance the methodology and its applications. In addition, we plan to build upon the framework developed here to extend spectral soni cation approaches to other areas of scienti c analysis. For example, we plan to connect the soni ed CGM absorption properties to other properties of the galaxies themselves, such as star formation rate and mass. The QSS software is available via our open source project repository located at https://github.com/CreativeCodingLab/QuasarSonify .

5 REFERENCES

[1] M. Fukugita, P. J. E. Peebles, "The Cosmic Energy Inventory," The Astrophysical Journal, vol. 616, no. 2, pp. 643–668 (2004), https://doi.org/10.1086/425155 .

[2] J. Tumlinson, M. S. Peeples, J. K. Werk, "The Circumgalactic Medium," Annual Review of Astronomy and Astrophysics, vol. 55, pp. 389–432 (2017), https://doi.org/10.1146/annurev-astro-091916-055240 .

[3] M. McQuinn, "The Evolution of the Intergalactic Medium," Annual Review of Astronomy and Astrophysics,

vol. 54, pp. 313–362 (2016), https://doi.org/10.1146/annurev-astro-082214-122355 .

[4] J. N. Burchett, D. Abramov, J. Otto, C. Artanegara, J. X. Prochaska, A. G. Forbes, "IGM-Vis: Analyzing Intergalactic and Circumgalactic Medium Absorption Using Quasar Sightlines in a Cosmic Web Context," Computer Graphics Forum, vol. 38, no. 3, pp. 491–504 (2019), https://doi.org/10.1111/cgf.13705 .

[5] J. N. Burchett, O. Elek, N. Tejos, J. X. Prochaska, T. M. Tripp, R. Bordoloi, A. G. Forbes, "Revealing the Dark Threads of the Cosmic Web," The Astrophysical Journal Letters, vol. 891, no. 2, p. L35 (2020), https://doi.org/10.3847/2041-8213/ab700c .

[6] L. Spitzer, F. R. Zabriskie, "Interstellar research with a spectroscopic satellite," Publications of the Astronomical Society of the Paci c, vol. 71, no. 422, pp. 412–420 (1959), https://doi.org/10.1086/127416 .

[7] N. H. Dieter, W. M. Goss, "Recent work on the interstellar medium," Reviews of Modern Physics, vol. 38, no. 2, p. 256 (1966), https://doi.org/10.1103/RevModPhys.38.256 .

[8] R. Cen, J. Miralda-Escudé, J. P. Ostriker, M. Rauch, "Gravitational collapse of small-scale structure as the origin of the Lyman-alpha forest," The Astrophysical Journal Letters, vol. 437, pp. L9–L12 (1994), https://doi.org/10.1086/187670 .

[9] R. Davé, R. Cen, J. P. Ostriker, G. L. Bryan, L. Hernquist, N. Katz, D. H. Weinberg, M. L. Norman, B. O'Shea, "Baryons in the Warm-Hot Intergalactic Medium," The Astrophysical Journal, vol. 552, pp. 473–483 (2001), https://doi.org/10.1086/320548 .

[10] J. Tumlinson, C. Thom, J. K. Werk, J. X. Prochaska, T. M. Tripp, N. Katz, R. Davé, B. D. Oppenheimer, J. D. Meiring, A. B. Ford, J. M. O'Meara, M. S. Peeples, K. R. Sembach, D. H. Weinberg, "The COS-Halos Survey: Rationale, Design, and a Census of Circumgalactic Neutral Hydrogen," The Astrophysical Journal, vol. 777, no. 59 (2013), https://doi.org/10.1088/0004-637X/777/1/59 .

[11] L. L. Cowie, A. Songaila, T.-S. Kim, E. M. Hu, "The metallicity and internal structure of the Lyman-alpha forest clouds," Astronomical Journal, vol. 109, pp. 1522–1530 (1995), https://doi.org/10.1086/117381 .

[12] J. N. Burchett, T. M. Tripp, J. X. Prochaska, J. K. Werk, J. Tumlinson, J. M. O'Meara, R. Bordoloi, N. Katz, C. N. A. Willmer, "A Deep Search For Faint Galaxies Associated With Very Low-redshift C IV Absorbers: II. Program Design, Absorption-line Measurements, and Absorber Statistics," The Astrophysical Journal, vol. 815, no. 2 (2015), https://doi.org/10.1088/0004-637X/815/2/91 .

[13] R. B. Larson, "Infall of Matter in Galaxies," Nature, vol. 236, pp. 21–23 (1972), https://doi.org/10.1038/236021a0 .

[14] D. Kereš, N. Katz, D. H. Weinberg, R. Davé, "How do galaxies get their gas?" Monthly Notices of the Royal Astronomical Society, vol. 363, pp. 2–28 (2005), https://doi.org/10.1111/j.1365-2966.2005.09451.x .

[15] S. Veilleux, G. Cecil, J. Bland-Hawthorn, "Galactic Winds," Annual Review of Astronomy and Astrophysics, vol. 43, pp. 769–826 (2005), https://doi.org/10.1146/annurev.astro.43.072103.150610 .

[16] D. Fielding, E. Quataert, M. McCourt, T. A. Thompson, "The impact of star formation feedback on the circumgalactic medium," Monthly Notices of the Royal Astronomical Society, vol. 466, pp. 3810–3826 (2017), https://doi.org/10.1146/annurev-astro-091916-055240 .

[17] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner, J. Neuhoff, "Soni cation report: Status of the eld and research agenda," (2010), https://digitalcommons.unl.edu/psychfacpub/444 .

[18] G. Laughlin, "Systemic: Characterizing Planets," (2012), available online: http://oklo.org .

[19] R. L. Alexander, J. A. Gilbert, E. Landi, M. Simoni, T. H. Zurbuchen, D. A. Roberts, "Audi cation as a diagnostic tool for exploratory heliospheric data analysis," presented at the International Community on Auditory Display (ICAD) (2011), http://hdl.handle.net/1853/51574 .

[20] W. L. Diaz Merced, R. M. Candey, N. Brickhouse, M. Schneps, J. C. Mannone, S. Brewster, K. Kolenberg, "Soni cation of astronomical data," Proceedings of the International Astronomical Union, vol. 7, no. S285, pp. 133–136 (2011), https://doi.org/10.1017/S1743921312000440 .

[21] R. M. Candey, A. M. Schertenleib, W. L. Diaz Merced, "xSonify Soni cation Tool for Space Physics," presented at the International Conference on Auditory Display (ICAD) (2006), http://hdl.handle.net/1853/50697 .

[22] W. L. Diaz Merced, Sound for the exploration of space physics data, Ph.D. thesis, University of Glasgow (2013), http://theses.gla.ac.uk/id/eprint/5804 .

[23] J. W. Newbold, A. Hunt, J. Brereton, "Chemical spectral analysis through soni cation," presented at the International Community on Auditory Display (ICAD) (2015), http://hdl.handle.net/1853/54197 .

[24] F. Morawitz, "Molecular soni cation of nuclear magnetic resonance data as a novel tool for sound creation," presented at the International Computer Music Conference (ICMC), pp. 6–11 (2016), http://hdl.handle.net/2027/spo.bbp2372.2016.002 .

[25] H. Terasawa, J. Parvizi, C. Chafe, "Sonifying ECOG seizure data with overtone mapping: A strategy for creating auditory gestalt from correlated multichannel data," presented at the International Community on Auditory Display (ICAD) (2012), http://hdl.handle.net/1853/44445 .

[26] M. Pietrucha, Soni cation of Spectroscopy Data, Master's thesis, Worcester Polytechnic Institute (2019), https://digitalcommons.wpi.edu/etd-theses/1277/ .

[27] J. Suchánek, "ATOM TONE v2.0—Software for sonification of atomic data for purpose of electroacoustic music," presented at the *International Symposium on Sound* (2018).

[28] F. Martin, O. Metatla, N. Bryan-Kinns, T. Stockman, "Accessible Spectrum Analyser," presented at the *International Community on Auditory Display (ICAD)* (2016), http://hdl.handle.net/1853/56586.

[29] T. Hermann, A. Hunt, J. G. Neuhoff, *The Sonification Handbook* (Logos Verlag Berlin) (2011), http://sonification.de/handbook.

[30] M. Peeples, J. Tumlinson, A. Fox, A. Aloisi, S. Fleming, R. Jedrzejewski, C. Oliveira, T. Ayres, C. Danforth, B. Keeney, *et al.*, "The Hubble Spectroscopic Legacy Archive," *Instrument Science Report COS*, vol. 4 (2017), https://www.stsci.edu/hst/instrumentation/cos.

[31] J. X. Prochaska, et al., "linetools: Third Minor Release," (2017), https://doi.org/10.5281/zenodo.168270.

[32] E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," *Proceedings of the National Academy of Sciences*, vol. 15, no. 3, pp. 168–173 (1929), https://doi.org/10.1073/pnas.15.3.168.

[33] N. A. Bahcall, "Hubble's Law and the expanding universe," *Proceedings of the National Academy of Sciences*, vol. 112, no. 11, pp. 3173–3175 (2015), https://doi.org/10.1073/pnas.1424299112.

[34] D. Verner, G. J. Ferland, K. Korista, D. Yakovlev, "Atomic data for astrophysics. II. New analytic fits for photoionization cross sections of atoms and ions," *arXiv preprint astro-ph/9601009* (1996), https://doi.org/10.1086/177435.

[35] L. A. DeWitt, R. G. Crowder, "Tonal fusion of consonant musical intervals: The oomph in Stumpf," *Perception & Psychophysics*, vol. 41, no. 1, pp. 73–84 (1987), https://doi.org/10.3758/BF03208216.

[36] D. Huron, "A derivation of the rules of voice-leading from perceptual principles," *The Journal of the Acoustical Society of America*, vol. 93, no. 4, pp. 2362–2362 (1993), https://doi.org/10.1525/mp.2001.19.1.1.

[37] R. Parncutt, "Pitch properties of chords of octave-spaced tones," *Contemporary Music Review*, vol. 9, no. 1-2, pp. 35–50 (1993), https://doi.org/10.1080/07494469300640331.

[38] D. L. Wessel, "Timbre space as a musical control structure," *Computer Music Journal*, vol. 3, no. 2, pp. 45–52 (1979), https://doi.org/10.2307/3680283.

[39] M. Puckette, *The Theory and Technique of Electronic Music* (World Scientific Publishing Company) (2007), http://msp.ucsd.edu/techniques.htm.

[40] Y. Suzuki, H. Takeshima, "Equal-loudness-level contours for pure tones," *The Journal of the Acoustical Society of America*, vol. 116, pp. 918–33 (2004), https://doi.org/10.1121/1.1763601.

[41] L. P. A. S. van Noorden, "Temporal coherence and the perception of temporal position in tone sequences," *IPO Annual Progress Report*, vol. 10, pp. 4–18 (1975), https://pure.tue.nl/ws/files/3389175/152538.pdf.

[42] R. Plomp, W. J. M. Levelt, "Tonal consonance and critical bandwidth," *The Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560 (1965), https://doi.org/10.1121/1.1909741.

[43] B. D. Savage, K. R. Sembach, "The analysis of apparent optical depth profiles for interstellar absorption lines," *The Astrophysical Journal*, vol. 379, pp. 245–259 (1991), https://doi.org/10.1086/170498.