SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

WINTER 2025/2026

Sonification of Exoplanet Transit Light Curves from the Kepler Mission

By Emily Liao

AUTHOR BIOGRAPHY

Emily Liao is a high school student based in Toronto, Canada with academic and creative interests in both music and astronomy. Her interdisciplinary work focuses on the sonification of astrophysical data, particularly exoplanet light curves, with the goal of enhancing public engagement and accessibility in science communication. She is especially interested in how auditory methods can convey complex scientific phenomena without compromising data integrity. She will graduate in 2027 and plans to pursue both physics and music in her further studies.

ABSTRACT

There is a plethora of publicly accessible data on the light curves of exoplanet transits that are more often represented using visual graphs. This project aims to make the data more accessible by sonifying the light curve data, presenting it through audio means and picking up patterns not visible through data visualization. Data was taken from the Kepler mission from the NASA Exoplanet Archive, and Python libraries such as lightkurve and MIDIutil were used to sonify the data. The flux was mapped to frequency, amplitude, note duration, and musical pitches. In each of the sonifications, the transits were clearly audible. This project could aid in accessibility of scientific data for the visually impaired, education, fusing of music and science, and pattern detection.

Keywords: *Exoplanet, Data sonification, Kepler mission, Transit, Light curves, Auditory display, Astronomy, Music*

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

WINTER 2025/2026

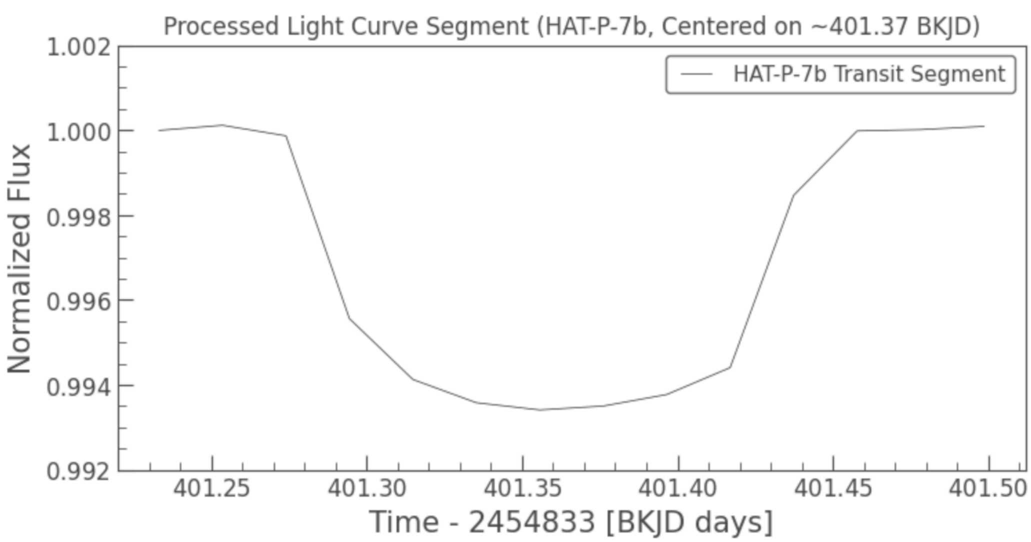
INTRODUCTION

The question of whether other planets existed beyond the Solar System has been on the minds of astronomers and philosophers for millennia. However, it is not until the last 30 years that concrete evidence for the existence of exoplanets has been found. Research on the discovery of exoplanets has expanded rapidly in the last two decades, transforming it into one of the most prominent fields in astronomical research. Among the various methods used to detect exoplanets, the transit method is one of the most widely used. This paper focuses on a new method of interpreting transit data: light curve sonification, which has benefits in education, accessibility, outreach, and pattern detection (Maskey and Maroune, 2023).

Transit Method for Exoplanet Detection

The transit method observes the change in brightness of a star or multiple stars in a side-on planetary system caused by one or more planets passing in front of the star (Seager 2010). This method is biased towards the detection of short-period planets, as they are easier to detect due to their more frequent transits and larger dips in brightness. The data points mapping flux of the star and time on a graph is called a light curve. Through such data, one is able to process the change in brightness and calculate the period of the planet's orbit and its size.

Figure 1.



Processed light curve segment displaying a single transit of HAT-P-7 b. See section 2.3 for information on the generation of this visualization.

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

WINTER 2025/2026

Data Sonification

Despite its widespread use and the immense impact the visualization of data has had on the field of data analysis, it does have its shortcomings - such as subtle anomalies not standing out and more difficulty in detecting periodicity - that hinder the processing of data, which is where data sonification comes in. Data sonification refers to the process of transforming data points into audio input, using parameters such as frequency, amplitude, or timbre (Bornmann, 2024). Datasets with numerous subtle variations can be difficult for the human eye to detect, whereas translating the same data into sound allows the human ear—highly sensitive to minor changes—to perceive these variations more effectively. In addition, the sonification of data allows visually impaired people to process data. Lastly, sonification of data holds the potential to spark interest in students and the public for the topic.

Research has already been done in this area by Matt Russo and SYSTEM Sounds, who created a sonification and a visualization of how human knowledge of exoplanets has progressed, in order to commemorate the discovery of the 5000th exoplanet by NASA. Furthermore, SYSTEM Sounds has also been sonifying images from the Hubble Space Telescope, matching pitch and volume to the brightness in areas of the image. Despite the many steps that have already been taken in such research, research was found to be lacking in the sonification of exoplanet light curves. Only two conference papers, *Sonigrapher. Sonified Light Curve Synthesizer*, and *Planethesizer: Approaching Exoplanet Sonification* by A. G. Riber, as well as one peer-reviewed journal article *Evaluating the Efficacy of Sonification for Signal Detection in Univariate, Evenly Sampled Light Curves Using Astronify* by J. T. Brown, C. M. Harrison, A. Zanella, and J. Trayford, exist for research in this area, and are focused on outreach and education.

Light Curves Used

This project sonified the light curves of three star systems: HAT-P-7, Kepler-10, and KOI-94. HAT-P-7 is a star with a planet HAT-P-7 b, which is larger than Jupiter in size and orbits 0.038 au away from its star. The Kepler-10 binary system contains three exoplanets all orbiting Kepler-10 A, with Kepler-10 b being approximately 1.5 times the radius of Earth and orbiting at 0.017 au. KOI-94 is a binary system with four planets orbiting KOI-94 a. KOI-94 b is located 0.051 au from its star and has a radius of 1.71 that of Earth. KOI-94 c is located 0.10 au from its star and has a radius of 4.3 times that of Earth.

This paper aims to design and produce a method of data sonification for light curves of exoplanets, matching the flux of the star to musical frequency. This study proposes a reproducible pipeline that transforms exoplanet light curve data into sonified audio representations which accurately represent data points. This approach ensures the preservation of artistic expression through data while not sacrificing scientific integrity.

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

WINTER 2025/2026

METHODS

Data

Data for this project was collected from the NASA Exoplanet Archive, specifically from the Kepler mission. All data used were obtained between 2009-2013, with various quarters taken. Selection of PDCSAP data was based on several factors, including the indication of clear transits and low-levels of noise. This introduced a bias towards star systems with only one planet, due to overlapping transits causing individual transits to be less distinguishable.

The goal of data processing in this project was to prepare the light curve data for sonification by cleaning, normalizing, and flattening the flux values. Data was processed using the Python lightkurve library due its compatibility with Kepler data, as well as its built-in tools for processing light curves.

Amplitude and Frequency Mapping

The objective of the sonification was to convert data points into audio representations. This study experimented with mapping higher fluxes to higher frequencies, as well as mapping the lower fluxes to higher frequencies. It was found that the latter method places more emphasis on the exoplanet itself, making it better at communicating how the planet transits across the star. Aside from frequency, flux was also mapped to amplitude in two ways. Higher flux was mapped to higher amplitude, as well as lower flux to higher amplitude. It was found that mapping lower flux to higher amplitude once again emphasizes the planet better for the same reason. In addition, the ear perceives the higher frequencies used here as louder than lower ones, which makes it more sensible to pair higher frequencies with higher amplitude, as the opposite pairing would make the effect less noticeable. Lastly, data was mapped to musical pitch directly by creating and exporting MIDI files. This allows timbre to indicate the transit, with different timbres being used to represent the base flux and the transit dip.

Audio Generation

This study also experimented with note lengths, and found that a note duration of around 0.33 allowed for the clearest perception of sound. The transit can also be indicated through varying note durations. The MIDI files can be played in a digital audio workstation (DAW) or notation software, allowing it to be easily used for music projects. Audio was mapped to two or more light curves to be played sequentially using a variety of the above methods. Comparisons were made between two different transits of the same planet and two different planets.

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

WINTER 2025/2026

First, the procedure remove_nans() was applied, which cleans missing data by removing unusable points. Then, normalize() is applied, which scales the flux values so that the median values are set to 1.0, mapping the changes in flux to the star's normal brightness. Next, flatten(window_length=x) is applied to remove stellar variability, ensuring changes in flux come from the transit itself. Lastly, remove_outliers(sigma=x) is applied in order to remove outliers. In order to pick a single instance of a transit, the BKJD time values taken from the NASA Exoplanet Archive and the period of the exoplanet were used to calculate a date within the observed quarter for which a transit occurs. See Figure 1 for a visualization of a light curve processed using this method.

For the sonification section of the process, the sample rate for the audio output was set to be 44100 samples per second, and specified a note duration. Using the numpy library, data points were mapped from the processed data to their corresponding frequencies, at first for the minimum flux to minimum frequency, and for maximum flux to maximum frequency, with minimum frequency and maximum frequency being predefined parameters. The same process was repeated for an inverse mapping, with minimum flux being mapped to maximum frequency and vice versa. Then, the amplitude was mapped using numpy, with the minimum flux to minimum amplitude, and for maximum flux to maximum amplitude, again with minimum amplitude and maximum amplitude being predefined parameters. For the inverse mapping, the process is the same. These parameters, amplitude and frequency, are used to generate a sine wave for each note, and then concatenated into a single file. Finally, the Python Sounddevice library is used to play the file.

For the generation of the MIDI file, the Python MIDIutil library was used. First, set a maximum and minimum MIDI note and amplitude, using Numpy to map them to flux. Then, each note is added to the MIDI object. Timbre is indicated by taking the number of the instrument in General MIDI, and changing timbre requires the creation of a new channel. To differentiate between the base flux of the star and the transit dip using timbre, a threshold was set, and the timbre changed once the flux dipped below the threshold. To indicate the transit using note duration, note duration was mapped to flux, making higher fluxes correspond to longer note durations using the same process.

RESULTS

Mapping

This research aimed to produce sonifications of light curves of exoplanet transits. Sine wave audio that mapped the points to raw frequency values was generated, without any regard for note pitches. This produced a series of stationary tones. In addition, MIDI that mapped data to note pitches rather than frequency was also generated. This also enabled us to vary the timbre, which can change depending on the flux, and note duration,

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

WINTER 2025/2026

with longer notes in higher fluxes overlapping and causing high flux to sound more resonant. Lastly, the results included comparative sonifications, which compared two different transits by playing their sonifications sequentially.

Audio Results

Direct frequency mapping for HAT-P-7 yielded audio which featured a sharp drop in frequency and then a rise back to the original frequency, indicating a transit. Inverse mapping for Kepler-10 yielded audio with a sharp rise in frequency and then a drop, though fluctuations between increasing and decreasing frequency indicated higher levels of noise. Direct mapping was also applied to KOI-94, yielding similar results as HAT-P-7, with longer note durations making the data points more defined. MIDI sonification produced similar results with regards to pitch. HAT-P-7 featured a direct frequency mapping with a change in timbre between violins and cello, starting with violin and switching to cello once the flux dropped below a certain threshold. Kepler-10 featured an inverse frequency mapping with a change in timbre between bassoon and flute, with the flute switching to bassoon past the threshold. KOI-94 featured a direct frequency mapping with variable note durations and a bowed strings timbre. Comparative sonification yielded an audio file playing a series of five transits of HAT-P-7 b across quarters 4-6. The transits stayed mostly uniform in that period, with a smooth descent in volume followed by a rise for all transits. In addition, the transits of KOI-94 b and KOI-94 c were compared, with those of KOI-94 c being in a higher register and having shorter note durations. The result of this was a smooth U-shaped dip for KOI-94 b and a more sporadic series of pitches for KOI-94 c. A transit of HAT-P-7 b was compared with a transit of Kepler-10 b using a similar method, except without the adjusting of register. This resulted in a clear transit for HAT-P-7 b, and a barely perceptible lowering of frequency for Kepler-10 b, indicating that the transit dip for Kepler-10 b is much smaller than HAT-P-7 b.

The audio files, along with the code documentation can be found here: https://github.com/emilyliao15/lightcurvesonification.

Assessment of Effectiveness

Mapping points to frequency values may be objectively more accurate and authentically perceived than mapping points to note values, as note values are just certain predefined frequencies and do not increase at a consistent rate. However, when perceiving sounds, many humans are accustomed to hearing each note as a "step", which may help in assisting understanding of the data. Note values may require rounding of values, as there are less values available to plot, meaning tiny changes may not be shown, whereas raw frequency allows for more precision. In addition, mapping to notes has the potential to introduce bias in perception, because those who are not familiar with western notation and pitch may perceive the data differently.

It is important to note that some DAWs or notation software (notably, Garageband and Musescore) ignore the program change function, which changes the timbre. This is due to MIDI instrumentation not translating directly to the instruments and sounds available in those programs, which means that the instruments have to be manually selected within the program. Garageband ignores the procedure altogether, while Musescore separates the audio into specified channels and may sometimes recognize the instrument specified.

DISCUSSION

Challenges and Limitations

In the process of sonifying these light curves, there were several challenges in both data processing and sonification. Sometimes the data contained a large amount of noise or too few points during the transit period, which made it difficult to hear the transit when sonified. In addition, removing outliers occasionally caused the data to have even fewer data points. Another challenge was that timing variations in multi-planet systems made transits more difficult to locate. Lastly, as mentioned above, many DAWs and notation software programs did not recognize the MIDI program change, making it difficult to change the timbre directly from MIDIutil. This method of sonification also faces limitations. One limitation is light curves which do not contain a clear transit may not be mapped in a way such that is clear to the listener what they are listening to. This means that this method of sonification can only be reliably applied to certain light curves.

Impact and Evaluation

One of the most meaningful aspects of this project is its relevance to accessibility. The sonification of light curves allows visually impaired individuals to engage with astronomical data, and hear the changes in brightness rather than see it. This could open up pathways for such individuals interested in astronomy but face a barrier due to the lack of accessible data.

Another meaningful aspect of this project is its educational potential. Sonifying data gives students and prospective astronomers a fresh perspective on exoplanet light curves, enhancing educational experiences. Through sonification, educators can present data in a multisensory manner, potentially allowing students to grasp the concepts quicker.

Overall, the results procured can be seen as successful, as they clearly convey their respective transits through sound. Out of the three light curves, the sonifications from HAT-P-7 were the clearest and most indicative of a transit (for example, this one). The light curves from Kepler-10 tended to contain more noise and anomalies (example). Many insights were gained from this project. Certainly, the most insight was gained from the comparative sonifications. The continuous sonifications of the light curves of HAT-P-7 showed us that transits stayed consistent throughout, but with tiny variations in baseline flux (especially in the third transit) that would not have been visible through solely visualization. In addition, the transit from KOI-94 c was a lot less smooth than KOI-94 b, which indicates the presence of anomalies or interference from other factors. Furthermore, one can get an idea of how much the transit dip of Kepler-10 b is shallower than HAT-P-7 b, when the transit of the former would have been barely visible on a graph. Lastly, all sonifications give listeners a stronger intuition of the transit and its shape.

Future Improvements

This method can be improved in the future in several ways. One such way is to experiment with continuous sonifications of consecutive transits, allowing listeners to detect patterns and anomalies which may reveal characteristics of the planet or the system. In addition, multi-planet systems can be sonified in such a way so that two or more overlapping transits can be played simultaneously and be differentiated by parameters such as timbre. Lastly, the method can be tested through obtaining feedback from the target audience, which include visually impaired people who are interested in astronomy, and students with an interest in astronomy. Their feedback could be invaluable for improving the method and its function.

CONCLUSION

This project explored methods of sonifying exoplanet light curves as a way to transform astronomical data into sound. The project demonstrates strong potential both in fields of accessibility and education, as well as in its use in astronomical data processing. Sonification could provide equal access to data for the visually impaired, providing individuals with the tools they need to engage in the field of astronomy and the study of exoplanets. In addition, the project allows students to be presented with a multisensory way of experiencing data, aiding their understanding of exoplanet transits. Moreover, it could allow scientists to detect patterns in the data, potentially revealing patterns in the data that were missed in graphs. Lastly, data sonification is a unique intersection point between music and science, allowing creative opportunities to arise, demonstrating that science can be heard and felt and promoting interdisciplinary thinking. In conclusion, this project can help open pathways not only for visually impaired individuals interested in astronomy, but also students, scientists, and artists alike.

REFERENCES

Bornmann, L. (2024, September). *The sound of science: Data sonification has emerged as possible alternative to data visualization*. EMBO reports. https://pmc.ncbi.nlm.nih.gov/articles/PMC11387736/

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

WINTER 2025/2026

Maskey , M., & Maroune, A. (2023, December 7). *From data to melody: Data sonification and its role in Open science*. NASA. https://www.earthdata.nasa.gov/news/blog/from-data-melody-data-sonification-its-role-open-science

Riber, A. G. (2018). p. 219-226. In *The 24th International Conference on Auditory Display (ICAD 2018)*. Retrieved May 23, 2025, from https://www.researchgate.net/profile/Adrian_Garcia_Riber/publication/329333261_Planethesizer_Approaching_Exoplanet_Sonification/links/601b116c92851c4ed5490474/Planethesizer-Approaching-Exoplanet-Sonification.pdf.

Riber, A. G. (2019). p. 62-66. In *The 25th International Conference on Auditory Display (ICAD 2019)*. Retrieved May 23, 2025, from https://www.researchgate.net/profile/Adrian_Garcia_Riber/publication/334856051_Sonifigrapher_Sonified_Light_Curve_Synthesizer/links/601b128492851c4ed549048e/Sonifigrapher-Sonified-Light-Curve-Synthesizer.pdf.

Russo, M., & Santaguida , A. (2022). (rep.). *5000 EXOPLANETS: LISTEN TO THE SOUNDS OF DISCOVERY* (pp. 64–68). Atlanta, Georgia: Georgia Institute of Technology.

Seager, S. (2010). *Exoplanets*. University of Arizona Press.

Tucker Brown, J., Harrison, C. M., Zanella, A., & Trayford, J. (2022). Evaluating the efficacy of sonification for signal detection in univariate, evenly sampled light curves using astronify. *Monthly Notices of the Royal Astronomical Society*, *516*(4), 5674–5683. https://academic.oup.com/mnras/article/516/4/5674/6698731