



Using GPT-3 to Achieve Semantically Relevant Data Sonification for an Art Installation

Rodolfo Ocampo^{1(✉)}, Josh Andres², Adrian Schmidt², Caroline Pegram³, Justin Shave³, Charlton Hill³, Brendan Wright^{1,3}, and Oliver Bown¹

¹ The University of New South Wales, Kensington, Australia
r.ocampo.blanco@unsw.edu.au

² The Australian National University, Canberra, Australia

³ Uncanny Valley, Canberra, Australia

Abstract. Large Language Models such as GPT-3 exhibit generative language capabilities with multiple potential applications in creative practice. In this paper, we present a method for data sonification that employs the GPT-3 model to create semantically relevant mappings between artificial intelligence-generated natural language descriptions of data, and human-generated descriptions of sounds. We implemented this method in a public art installation to generate a soundscape based on data from different systems. While common sonification approaches rely on arbitrary mappings between data values and sonic values, our approach explores the use of language models to achieve a mapping not via values but via meaning. We find our approach is a useful tool for musicification practice and demonstrates a new application of generative language models in creative new media arts practice. We show how different prompts influence data to sound mappings, and highlight that matching the embeddings of texts of different lengths produces undesired behavior.

Keywords: data sonification · machine learning · generative music · generative AI · large language models · word embeddings

1 Introduction

Data sonification transforms data into sound to convey information sonically for data exploration, entertainment, or artistic expression [17] [10] [6].

The authors of this paper were commissioned with a data sonification art installation to produce a generative soundscape for the foyer of the building of The School of Cybernetics at The Australian National University. The brief requested that the sonified data represented the complex systems that the building is using a Cybernetic lens, where humans, technology and the environment adaptively interact [1, 7].

Sonification approaches often involve a numeric mapping of data values to sonic values [6]. This mapping determines the extent to which the sonic properties convey data properties and the extent to which the sonified data results

in musical outputs acceptable for the purpose of the data sonification project. Usually, the mappings involve either continuous or discreet symbolic mappings. In a continuous approach, numerical data is mapped to sonic qualities such as sound frequency, attack, release or effects like delay and reverb. In a symbolic approach, categorical data or discreetly segmented data is mapped to symbolic music representations such as specific notes, MIDI values or instrument types. Sonification approaches involving more than one data source commonly employ continuous and symbolic approaches. A common challenge of these approaches is the potential arbitrariness or lack of meaning of such value mappings [6].

To address the installation’s brief, we experimented with the creative potential of generative language models to enable a mapping not via value mapping but via semantic matching between data interpretations and sound descriptions. In other words: mapping data to sound via language.

In a preliminary examination, we found that GPT-3 could generate natural language interpretations of alphanumeric data and then match these to human-labeled sounds using word embeddings. Sounds were labeled according to what they evoked in the listener, thus our approach’s objective is to match machine-generated subjective interpretations of data with human-generated subjective interpretations of sounds.

[10] proposed a three-level distinction for sonification approaches that is useful to situate our approach. The first level is generative, where the contents of the data are not relevant: it is merely used to achieve compositions that would be impossible or difficult manually. The second level, allusive sonification, aims to create a bridge between complex information with a more accessible artistic domain that appeals to sensory and emotional experiences. Lastly, curatorial sonification is mostly a scientific tool that aims to accurately convey salient features of the data through sound in a meaningful way.

Our approach can be placed within the Allusive level: our intention is to transform complex data into a sensory and emotional experience through the use of a semantic matching that captures expressed emotional information.

For this project, the authors, University of New South Wales academics, partnered with Uncanny Valley, music technology company and built upon their existing generative music engine, MEMU (<https://memu.live/>), which has been deployed previously in different contexts.

By developing a novel creative use of one of the most prominent machine learning models to date, GPT-3, we provide what we believe is a useful new tool for the data sonification and creative new media practice communities. We outline the challenges faced and the strategies used to address them. Lastly, we discuss limitations and qualitative and quantitative learnings that can be useful for future work.

2 Related Work

The sonification of data, particularly environmental and atmospheric, has fascinated researchers for decades to extend beyond text, tables, and visualisations on how humans can engage with data. For example, [15] explored an artistic approach to newly collected data on the earth’s Ice Core, Seismic and Solar Wind

Data to map data attributes to musical elements for a one-hour pre-recorded performance. This work is an example of data sonification in a non-audio generative context because it uses static multimodal time series data where the musical elements are controlled, tweaked and finessed to achieve second by second a desired musical expression. Interestingly, in this creative decision-making process to select data attributes to musical elements, researchers can emphasise temporal data aspects for listeners that are often non-appreciated using numbers [4] [5] [14]. [18] developed a sonification approach rooted in Cybernetics by integrating urban open data streams in live musical performances, where live data was matched to audio effects and sonic parameters affecting music improvised by a human.

Recent data sonification explorations benefit from computational architectures that use real-time data streams and generative machine learning models to create audio experiences. From using atmospheric environmental data with generative Japanese music melodies data sets to query varying data periods [8] to experimental approaches combining multiple generative models, such as large language text models, to query creative music composition decisions for composer-to-machine dialogue opportunities [9]. Mubert is a recently developed text-to-music system that takes a user prompt and matches it to labeled pieces of sounds composed by humans via latent embeddings in order to build a finished track. Our approach to text-to-music matching is highly similar in the sense that sounds are composed by humans, and combined to create music tracks, and that matchings between text and sound are achieved via prompt-label word embedding similarity.

Our work contributes to the recent efforts in exploring the creative potential of Transformer and large language models, and provides a new tool to the creative new media and data sonification communities.

Next, we describe the presented architecture and report on the audio generative learnings.

3 Implementation

3.1 Conceptual Development

To create a continuous data sonification soundscape of the building as a complex cybernetic system, we drew on the concept of Pace Layers by Stewart Brand [2], which describes complex systems as composed of six layers with distinct paces of change: nature, culture, infrastructure, governance, commerce and fashion/media. We carefully selected a data stream to each layer as follows:

- **Nature layer:** Global CO2 concentration in the atmosphere (ppm) per month for the last 36 months, updated monthly.
- **Culture layer:** The most read articles of the day from Wikipedia, updated daily.
- **Infrastructure layer:** Latest indicator values from the Infrastructure topic from the WorldBank open data API, for Australia, updated daily.

- **Governance layer:** Latest indicator values from the Public Sector topic from the WorldBank open data API, for Australia, updated daily.
- **Commerce layer:** Latest indicator values from the Economy and Growth topic from the WorldBank open data API, for Australia, updated daily.
- **Fashion/media layer:** The latest tweet from the ANU School of Cybernetics Twitter handle, updated daily.

Each data source is mapped to a sonic layer in the soundscape that reflects a similar rate of change. For example, atmospheric sounds are the slowest changing sounds, thus they are mapped to Nature. On the other hand, accents are the fastest-changing, thus they are mapped to fashion. Therefore, we chose the mapping so that the soundscape, comprised of stacked layers of differently paced sounds, reflects the stacking of the pace layers diagram.

- Fashion layer: Accents
- Commerce layer: Bass
- Infrastructure layer: Timekeepers
- Governance layer: Pads and chords
- Culture layer: Scales
- Nature layer: Atmospheric

3.2 Human Labeling of Sounds by Audio Elicitation

We invited the building occupants, composed of academics, students, professional staff, and also friends, family, and the installation developers, to label a library of 1050 sounds created by Uncanny Valley, totalling approximately 80 labellers. Labellers were asked to enter a label description of multiple words based on what the sound elicited from them via a custom-built interface showing an audio player and a text field. The sound labelling interface randomly loaded sounds to each labeller as a strategy to ensure all sounds were labelled. From the 1050 sounds, 219 sounds received more than one label. The labels provided were succinct, averaging 2.9 words per label. Examples of labels include:

- A blue light traveling through a vacuum chamber, serious
- The dopplered siren fades to silence as it merges with the horizon.
- bladerunner electric android dreams

3.3 Generating Natural Language Interpretations of Data with GPT-3

The next step of our pipeline involves retrieving each data source at regular interval periods and then passing the data to GPT-3 as part of a prompt to generate natural language interpretations of the data.

A Python script retrieves each data source via web APIs. The data is converted into a string to be appended to the GPT-3 prompt, which has the following structure.

Prompt = data explanation + string representation of data + request.

Our intention was to use GPT-3 to produce an interpretation of the data that contained subjective descriptors and emotional language, since this is the type of language contained in the sound labels we aimed to match to. Thus one of our approaches involved requesting the generation of poems from the data. We tried different prompting approaches and the effects of each are shown in the results section.

An example prompt and GPT-3 response are provided below:

This data contains the last 36 months of CO2 measurements in the atmosphere. The first column is the year, the second the month and the third is the CO2 level in parts per million.

Data:

```
[[ '2019', '11', '410.48'],
 [ '2019', '12', '411.98'],
 [ '2020', '1', '413.61'],
 [...],
 [ '2022', '8', '417.19'],
 [ '2022', '9', '415.95'],
 [ '2022', '10', '415.78']]
```

Write a poem about the data:

GPT-3 response:

```
The CO2 in the atmosphere
Is rising day by day
The levels are getting higher
And higher all the time
[...]
```

To obtain the GPT-3 generations, we used the OpenAI GPT-3 Python API [12] with the following settings:

- **Engine:** text-davinci-002
- **Temperature:** 0.9
- **Max Tokens:** 256
- **Top P:** 1
- **Frequency penalty:** 0
- **Presence penalty:** 0

3.4 Matching GPT-3's Data Interpretations to Sound Labels for Soundscape Generation

For the soundscape generation, we extended an existing music engine developed by Uncanny Valley called MEMU. This engine contains a music library of multiple thousands of sounds, and it is able to combine them in musically coherent

ways to produce nearly infinite musical variations. This engine works by selecting and stacking different types of sounds (pads, atmospheric). The selected sounds in the engine are constrained by a globally defined chroma value as well as valence, and then sounds matching these constraints are selected at random. Our work consisted on modifying the engine so that instead of selecting sounds at random, sounds are selected based on the incoming data and the associated sound labels.

At regular intervals, the engine queries the generated GPT-3 text for each data source, including each embedding representation, which is a single 1024-valued vector for each text. The engine then matches these embeddings to pre-computed sound label embeddings via cosine similarity, a common measure used to calculate semantic similarity between word embeddings [19].

The word embeddings are generated using the “text-similarity-ada-001” model provided through the OpenAI embeddings Python API [13], a Transformer word embedding model derived from GPT [11].

The sounds that each data source can match to are constrained by three factors: 1) the layer-to-sound type mapping described in Sect. 3.1., 2) the global musical key, determined by the engine at random and 3) the musical mode, determined by the local weather outlook obtained from a weather API, according to the following mapping.

- Sunny outlook: Ionian mode
- Partly cloudy outlook: Dorian mode
- Cloudy or rainy outlook: Aeolian mode

4 Results

In order to evaluate our approach to semantic data musification, we are interested in understanding the following:

- **Interpretation quality:** how well does the GPT-3 generated interpretation reflect characteristics from the data
- **Matching quality:** can a clear semantic connection be drawn between the GPT-3 generated natural language interpretations of the data and the matched sound label?
- **Breadth:** how many different sounds are selected?

4.1 Experiment 1: Preliminary Testing of Matching GPT-3 Data Interpretations to Sound Labels

As described above, the prompt we passed to GPT-3 has the following structure:

Prompt = explanation + string version of data + generation request.

An example prompt with these three components, for the Culture layer data would be as follows.

"These data contains the titles for the most read wikipedia articles today.

Data:

Aaron Carter
Crown Jewel (2022)
"ICC Mens T20 World Cup"
Leslie Carter Cleopatra
"Weird Al" Yankovic
Elon Musk
Guy Fawkes Night
Nick Carter (singer)
Sally McNeil

Interpret these data."

We evaluated the effect of different prompts by varying the request component. We wrote six different requests:

- Request 1: Describe the data in only three words
- Request 2: Describe how these data make you feel in only three words
- Request 3: Describe in three words how these data make you feel and why
- Request 4: Write in one sentence how these data make you feel and why:
- Request 5: Write a poem about the data
- Request 6: Interpret the data

To test them, we isolated one data source: the most-read Wikipedia articles of the day corresponding to the Culture layer. We generated a hundred variants for each request, one for each of the last hundred days of data, thus obtaining six hundred generations and corresponding matchings.

We qualitatively evaluated the interpretation quality by observing if the generated text reflects the characteristics of the data. We evaluated the the matching quality qualitatively by examining if a clear semantic connection could be drawn by the GPT-3 generated text and the matched label. Lastly, we evaluated the sound selection breadth by calculating how many unique labels were selected for each request and obtaining the sound selection distribution of each selected sound.

A representative example of text generation and corresponding matching is provided in Table 1.

For each request, we calculate the number of unique sound labels matched, shown below:

- Request 1 (Describe the data in only three words): 21
- Request 2 (Describe how these data make you feel in only three words): 19
- Request 3 (Describe in three words how these data make you feel and why): 11
- Request 4 (Write in one sentence how these data make you feel and why): 10

Table 1. Generated text by GPT-3 when asked to generate poems from the data and resulting sound label matchings

Request	Data	GPT-3 generated request	Matched Sound Label
Describe the data in only three words:	["Manti Te'o", 'Omegle', 'Casemiro' [...]]	Titles of Wikipedia articles.	exotic and interesting
Describe how these data make you feel in only three words:	["Manti Te'o", 'Omegle', 'Casemiro' [...]]	Curious, intrigued, satisfied.	cold, curious, Interested
Describe in three words how these data make you feel and why:	["Manti Te'o", 'Omegle', 'Casemiro' [...]]	Curious, because these are all topics that I am not familiar with. Intrigued, because I am curious to learn more about them. And entertained, because it is always interesting to learn new things.	Interesting, curious, Spatial, night, Open, airy, blue, whimsical, otherworldly, calling me
Write in one sentence how these data make you feel and why:	["Manti Te'o", 'Omegle', 'Casemiro' [...]]	The data make me feel curious about why Manti Te'o, Omegle, and Casemiro are the most read Wikipedia articles today.	eerie, suspenseful, sexy, inquisitive, reflective
Write a poem about the data:	["Manti Te'o", 'Omegle', 'Casemiro' [...]]	These are the titles for the most read wikipedia articles today. Manti Te'o is a football player who was caught in a hoax. Omegle is a website where you can chat with strangers. Casemiro is a Brazilian footballer. Sanna Marin is the Prime Minister of Finland. [...]	death and dying, death and dying, death and dying, fear death and dying, fear death and dying, fear death and dying, fear death and dying
Interpret the data:	["Manti Te'o", 'Omegle', 'Casemiro' [...]]	The data shows the titles of the most read Wikipedia articles as of today. The top article is about Manti Te'o, followed by Omegle, Casemiro, Sanna Marin, and Andrew Tate.	death and dying, death and dying, death and dying, fear death and dying, fear death and dying, fear death and dying, fear death and dying, fear death and dying

- Request 5 (Write a poem about the data): 6
- Request 6 (Interpret the data): 6

We observed that requests 5 and 6 were producing text that was being matched to only six unique sounds, out of the 100 generations, thus displaying a degree of sound selection breadth that was undesirable given the installation’s objectives. In fact, most times, these were being matched to a single label “death and dying [...]”, as shown in Fig. 1 and Table 1. A very similar distribution was observed for request 6.

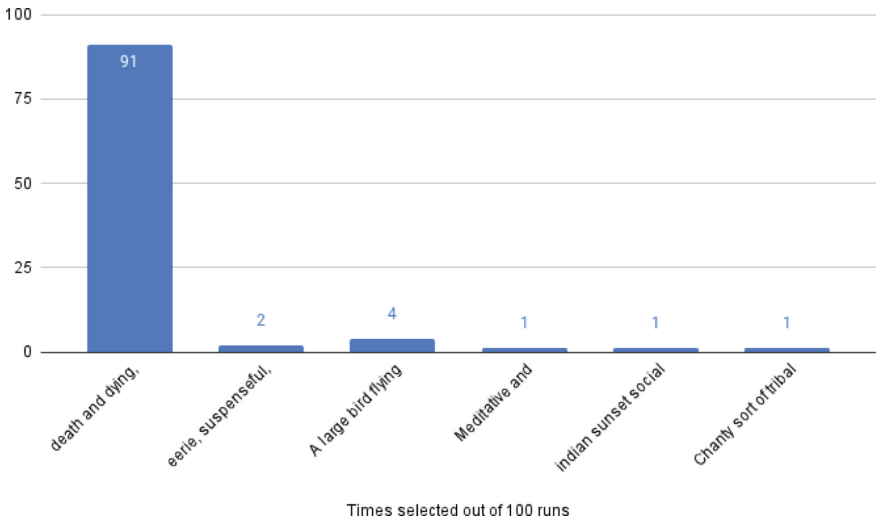


Fig. 1. Sound labels selected from Request 5 (“Write a poem about the data”).

While some of the generated text had references to death, not all of them did but they were still being matched to the same label, and it was not clear why this was the case.

Meanwhile, requests 1 and 2 were producing text that matched to a larger number of unique labels, which also had clearer semantic relationships.

Matching Text of Different Lengths Produced Undesired Behavior.

Upon closer examination, we found that requests 5 and 6 were producing the longest text out of all requests. The “death and dying [...]” label was the longest label in the library. While most labels in the library were around three words long, the death and dying label is 29 words long. Thus, we hypothesised that length was becoming a salient feature, perhaps due to word embedding averaging effects. While short GPT-3 generated text was producing semantically relevant matchings to sound labels, longer text seemed to only be match to longer text.

We sought to test this further by testing 6 hand-crafted test texts and evaluate what they matched to. We tested four short test texts and one long test text, corresponding to the first three paragraphs of the song "Stayin' Alive". The matching results are shown in Table 2.

Table 2. Testing sound label matchings using human written test inputs

Test input text	Matched sound label
A happy day	happy and warm
Sad and gloomy	depressing
Exciting and adventurous	exotic and interesting
Ethereal and spacey	exotic and ethereal
Well, you can tell by the way I use my walk I'm a woman's man [...] You're stayin' alive, stayin' alive Feel the city breakin' and everybody shakin' And we're stayin' alive, stayin' alive Ah, ha, ha, ha, stayin' alive, stayin' alive [...]	death and dying,death and dying,death and dying,fear death and dying,fear death and dying,fear death and dying,fear death and dying,fear death and dying

The results above provide supporting evidence that short text was producing clear semantic matchings, while long text was just matching to the longest label, even if the meaning was radically different. Thus using the text-similarity-ada-001 produced embeddings to match text of largely different lengths produces undesired behavior.

Repetitive Generation with Low Data Interpretation Quality. While the shorter text produced by Requests 1 and 2 produced clearer semantic matchings, when examining the generated text from these requests for interpretation quality, we found they did not reveal properties of the data, and instead were fixating around a small number of three-word descriptions. A subset of 10 generations for Request 2 showing this behavior is provided below.

1. Curious, intrigued, motivated.
2. Informative, curious, excited.
3. Curious, Engaged, Intrigued.
4. Curious, Engaged, Intrigued.
5. hopeful, inspired, curious.
6. Curious, interested, engaged.
7. Curious, intrigued, excited.

8. Intrigued, curious, interested.
9. Curious, intrigued, excited.
10. Happy, curious, and excited.

This repetitive generation constrains the variety of matched labels, as shown in Fig. 2, where we plot the distribution of each of the selected labels for the 100 runs using Request 2 (“Describe how these data make you feel in only three words”). Out of the 19 unique sounds matched, 64 were matched to just one label: “question everything”. A similar behaviour was observed for Request 1 (“Describe the data in only three words”), where out of the 21 unique labels selected for the 100 trials, 35 times it was matched to just one label: “cold, curious interested”.

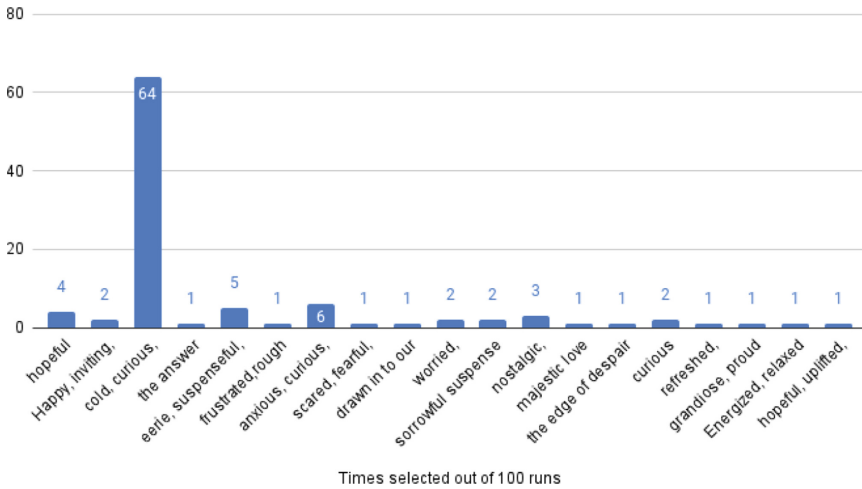


Fig. 2. Sound labels selected from Request 2 (“Describe how these data make you feel in only three words”).

We hypothesised this was the result of asking the model to describe how the data made it “feel”. We experimented with producing three-word descriptions that leverage the observed improvements of short text on matching quality and sound selection breadth while generating text that more closely captures characteristics of the data (better data interpretation quality). We tested this in experiment 2.

4.2 Experiment 2: Extracting Emotional Qualities of the Data for Better Matching

In the previous experiment, we found that short interpretations provide more varied and clearer sound label matchings than longer generations. However, these

short texts mostly fixate around a few descriptors that don't reflect the data accurately. In order to address this, instead of how the data makes the model "feel", we experimented with requesting interpretations from GPT-3 that extract emotional qualities of the data. We trialled the following three phrasings:

- Request 1: Write three words that capture the emotional quality of the data
- Request 2: Write three words that summarise these titles
- Request 3: Write three words that summarise the vibe of these titles

Again, we generated 100 prompts for each request, one for each of the last 100 d of data for the Wikipedia data. We found these requests yield short data interpretations that reflect the data more closely, while still providing highly relevant semantic matchings, as shown in the examples in Table 3. Moreover, these generations more closely reflect qualities in the data.

We found this prompting approach yields text that is more varied than the approach in experiment 1, which fixated around curiosity related descriptors, A few generations are provided below.

- Readership, Popularity, Notability
- Sorrowful, triumphant, celebratory.
- Famous, Joyful, Controversial
- Gory, Terrifying, Shocking

We also we found that this approach yields a higher breadth in sound selection compared to the prompting approach in Experiment 1. Below, we list the unique sounds selected for each request, after generating 100 times for each.

- Write three words that capture the emotional quality of the data: 43
- Write three words that summarise these titles: 43
- Write three words that summarise the vibe of these titles: 45

These requests produce GPT-3 data interpretations that yield more than double the amount of unique sound matchings than the approach used in Experiment 1. Moreover, they provide flatter sound label selection distributions that are less dominated by particular labels, as shown in Fig. 3.

This prompting approach proved to be the best approach in terms of our musification objectives: matching quality, interpretation quality and sound selection breadth, thus it became our adopted approach in the installation.

In the supplementary materials, this approach is applied to all data sources.

Lastly, in the following link: <https://github.com/rodolfoocampo/EvoMusArt2023-SemanticSonification>, we provide a repository data for the experiments, example code, a link to a live demo in the browser of the application and a link to an interface to play the selected sounds alongside with labels and matched GPT-3 generated text so that the reader gets a sense of the data-to-sound mappings. In the supplementary materials, we provide a video of the working installation.

Table 3. Comparison of generated text for different requests to GPT-3 and resulting sound label matchings

Request	Data	GPT-3 response	Matched Label
Write three words that capture the emotional quality of the data:	['Aaron Carter', 'Nick Carter (singer)', 'Leslie Carter' [...]]	Readership, Popularity, Notability	attention grabbing empowering
Write three words that capture the emotional quality of the data:	['Manti Te'o', 'Omegle', 'Casemiro' [...]]	humorous, lighthearted, fun	fun playful
Write three words that summarise these titles:	['Aaron Carter', 'Nick Carter (singer)', 'Leslie Carter' [...]]	Musicians, death, crime.	broken jazz band
Write three words that summarise these titles:	['Manti Te'o', 'Omegle', 'Casemiro' [...]]	1. People 2. Culture 3. Technology	confused people
Write three words that summarise the vibe of these titles:	['Aaron Carter', 'Nick Carter (singer)', 'Leslie Carter' [...]]	celebrities, murder, internet	extremely disturbing psycho
Write three words that summarise the vibe of these titles:	['Manti Te'o', 'Omegle', 'Casemiro' [...]]	humorous, chatty, lighthearted	fun playful

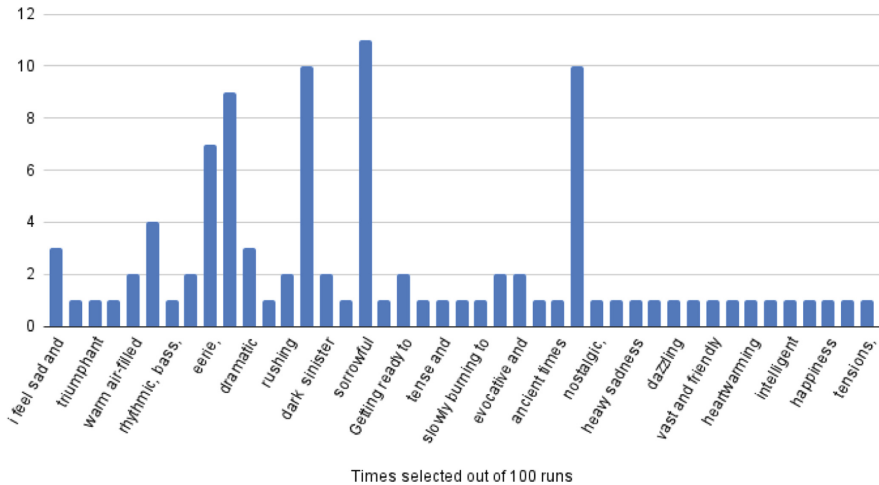


Fig. 3. Sound labels selected from Request 1b (“Write three words that capture the emotional quality of the data”).

5 Discussion

Our sonification approach did not focus on precisely conveying properties of the data (curatorial scientific sonification [10]). Instead, our approach focused on allusive sonification [10] with the intention of transforming complex data into

a more accessible artistic representation that appeals to emotional responses by matching data and sound through language interpretations.

For this, we set three objectives: interpretation quality, matching quality, and breadth of sound selection. We found that interpretation quality is highly influenced by prompting approaches as well as the model’s own capabilities. While GPT-3 sometimes can produce factually accurate interpretations of data, such as recognising when values have gone up or down, or are positive or negative, this was not always the case, particularly when recognising more complex numerical relationships. We believe this is due to the well-documented weakness in GPT-3’s numerical reasoning capabilities [3]. Musification or text-to-sound mappings that require higher objective descriptiveness might benefit from chain-of-thought few short prompting or fine-tuning. In our case, the intention was to produce short interpretations of data using subjective and emotional language that could be semantically mapped to sounds described in similar terms. We found that prompting approaches that require the model to extract the emotional qualities of the data best served this purpose. Moreover, the shortness of the generated text was important since we found that matching text of different lengths via word embedding cosine similarity produced undesired behavior, such that text appeared to be matched on the basis of length and not meaning.

We found that breadth of sound selection was influenced by the prompting approach. Initially, we wanted to create text that reflected a first person subjective evaluation of the data, so we asked the model to describe how the data made it “feel”. However, this produced repetitive responses that limited the breadth of sound selection. As mentioned above, this was addressed by requesting the model to instead extract emotional qualities of the data, which yielded responses that more closely reflected the data and its variability. From a sonification perspective, breadth of sound selection was important for the soundscape to be interesting. However, maximum breadth was not the objective, but a degree that reflected the data’s variability.

This paper does not involve any evaluation of the musicality of the MEMU engine. MEMU’s sounds were created by professional musicians, and uses generative techniques to mix and arrange music. Our approach simply selects elements that MEMU will play, and could be easily adapted to work with other generative music engines.

Lastly, we believe this approach is relevant in the context of multimodal generative AI, which maps different modes of data, like text, image and music via latent embeddings. These approaches usually rely on datasets labelled in natural language [16]. While training text-to-image multimodal models benefit from image datasets with highly descriptive captions, music is often described in more subjective, abstract and emotional terms. A future direction for our work could be to use our labeled library to train a generative text-to-music model. This would require expanding the dataset and more standardised labelling.

6 Conclusion

We presented a novel machine learning pipeline built to create a continuous data sonification soundscape using large language models. Our work shows that large language models enable data sonification by semantic mapping between data interpretations and sound descriptions. We demonstrated via two experiments the potentials and limitations of the presented pipeline to support creative endeavours in data sonification practices.

Acknowledgements. This research was made possible by a commission from the School of Cybernetics at the Australian National University for music studio Uncanny Valley (UV). The development of the novel concept for a semantically relevant sonification using Large Language Models is an original contribution from Rodolfo Ocampo, who also led the technical development of the system, in collaboration with members of the UV team. The artwork uses UV's MEMU generative music system, developed by Justin Shave and Brendan Wright. The design and development of the visual user interface were led by Adrian Schmidt and Josh Andres. Oliver Bown and Rodolfo Ocampo's research is supported by an Australian Research Council Discovery Project (DP200101059).

References

1. Andres, J.: Adaptive human bodies & adaptive built environments for enriching futures. In: *Frontiers in Computer Science, Special Issue Inbodied Interaction* (2022)
2. Brand, S.: Pace layering: how complex systems learn and keep learning. *J. Design Sci.* (Jan 2018)
3. Floridi, L., Chiriatti, M.: GPT-3: Its nature, scope, limits, and consequences. *Minds Mach.* **30**(4), 681–694 (2020)
4. Flowers, J.H., Whitwer, L.E., Grafel, D.C., Kotan, C.A.: Sonification of daily weather records: issues of perception, attention and memory in design choices. *Faculty Publications, Department of Psychology*, p. 432 (2001)
5. Hermann, T., Drees, J.M., Ritter, H.: Broadcasting auditory weather reports-a pilot project (2003)
6. Hermann, T., Hunt, A., Neuhoﬀ, J.G.: *The sonification handbook*. Logos Verlag Berlin (2011)
7. Andres, J., et al.: Cybernetic lenses for designing and living in a complex world. In: *Extended Abstracts of the 2022 OzCHI Conference on Human Factors in Computing Systems (OZCHI EA 2022)*. Association for Computing Machinery, New York, NY, USA. (2022)
8. Kalonaris, S.: Tokyo Kion-On: query-Based generative sonification of atmospheric data (Aug 2022)
9. Krol, S.J., Llano, M.T., McCormack, J.: Towards the generation of musical explanations with GPT-3. In: *Artificial Intelligence in Music, Sound, Art and Design: 11th International Conference, EvoMUSART 2022, Held as Part of EvoStar 2022, Madrid, Spain, April 20–22, 2022, Proceedings*, pp. 131–147. Springer-Verlag, Berlin, Heidelberg (Apr 2022)
10. Mardakheh, M.K., Wilson, S.: A strata-based approach to discussing artistic data sonification. *Leonardo* **55**(5), 516–520 (2022)

11. Neelakantan, A., et al.: Text and code embeddings by contrastive Pre-Training (Jan 2022)
12. OpenAI: OpenAI API. <https://beta.openai.com/docs/introduction> (2022). Accessed 17 Nov 2022
13. OpenAI: OpenAI API. <https://beta.openai.com/docs/guides/embeddings> (2022). Accessed 21 Aug 2022
14. Polli, A.: Atmospherics/weather works: a multi-channel storm sonification project (2004)
15. Quinn, M.: Research set to music: the climate symphony and other sonifications of ice core, radar, DNA, seismic and solar wind data (2001)
16. Ramesh, A., et al.: Zero-shot text-to-image generation (Feb 2021)
17. Rocchesso, D., et al.: Sonic interaction design: sound, information and experience. In: CHI'08 Extended Abstracts on Human Factors in Computing Systems, pp. 3969–3972 (2008)
18. Roddy, S.: Signal to noise loops: a cybernetic approach to musical performance with smart city data and generative music techniques. Leonardo, pp. 525–532 (2022)
19. Singhal, A.: Modern information retrieval: a brief overview. <http://160592857366.free.fr/joe/ebooks/ShareData/Modern%20Information%20Retrieval%20-%20A%20Brief%20Overview.pdf> (2001). Accessed 17 Nov 2022