



OPEN ACCESS

EDITED BY

Tae Yeon Kim,
Khalifa University, United Arab Emirates

REVIEWED BY

Zihao Ding,
KLA, United States
Shaoping Xiao,
The University of Iowa, United States

*CORRESPONDENCE

O. Barrera,
✉ olga.barrera@eng.ox.ac.uk

SPECIALTY SECTION

This article was submitted to
Computational Materials Science, a
section of the journal *Frontiers in
Materials*

RECEIVED 08 November 2022

ACCEPTED 20 March 2023

PUBLISHED 09 May 2023

CITATION

Waghorne J, Howard C, Hu H, Pang J,
Peveler WJ, Harris L and Barrera O
(2023), The applicability of
transperceptual and deep learning
approaches to the study and mimicry of
complex cartilaginous tissues.
Front. Mater. 10:1092647.
doi: 10.3389/fmats.2023.1092647

COPYRIGHT

© 2023 Waghorne, Howard, Hu, Pang,
Peveler, Harris and Barrera. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

The applicability of transperceptual and deep learning approaches to the study and mimicry of complex cartilaginous tissues

J. Waghorne¹, C. Howard², H. Hu³, J. Pang³, W. J. Peveler⁴,
L. Harris² and O. Barrera^{1,5*}

¹School of Engineering, Computing and Mathematics, Oxford Brookes University, Oxford, United Kingdom, ²School of Culture and Creative Arts, University of Glasgow, Glasgow, United Kingdom, ³Department of Computer Science, University of Luxembourg, Luxembourg, United Kingdom, ⁴School of Chemistry, University of Glasgow, Glasgow, United Kingdom, ⁵Department of Engineering Science, University of Oxford, Oxford, United Kingdom

Introduction: Complex soft tissues, such as knee meniscus, play a crucial role in mobility and joint health but are incredibly difficult to repair and replace when damaged. This difficulty is due to the highly hierarchical and porous nature of the tissues, which, in turn, leads to their unique mechanical properties that provide joint stability, load redistribution, and friction reduction. To design tissue substitutes, the internal architecture of the native tissue needs to be understood and replicated.

Methods: We explore a combined audiovisual approach, a so-called transperceptual approach, to generate artificial architectures mimicking the native architectures. The proposed methodology uses both traditional imagery and sound generated from each image to rapidly compare and contrast the porosity and pore size within the samples. We have trained and tested a generative adversarial network (GAN) on 2D image stacks of a knee meniscus. To understand how the resolution of the set of training images impacts the similarity of the artificial dataset to the original, we have trained the GAN with two datasets. The first consists of 478 pairs of audio and image files for which the images were downsampled to 64 × 64 pixels. The second dataset contains 7,640 pairs of audio and image files for which the full resolution of 256 × 256 pixels is retained, but each image is divided into 16 square sections to maintain the limit of 64 × 64 pixels required by the GAN.

Results: We reconstructed the 2D stacks of artificially generated datasets into 3D objects and ran image analysis algorithms to characterize the architectural parameters statistically (pore size, tortuosity, and pore connectivity). Comparison with the original dataset showed that the artificially generated dataset based on the downsampled images performs best in terms of parameter matching, achieving between 4% and 8% of the mean of grayscale values of the pixels, mean porosity, and pore size of the native dataset.

Discussion: Our audiovisual approach has the potential to be extended to larger datasets to explore how similarities and differences can be audibly recognized across multiple samples.

KEYWORDS

hierarchical porous media, audiovisual, sonification, machine learning, meniscus, pore network extraction

1 Introduction

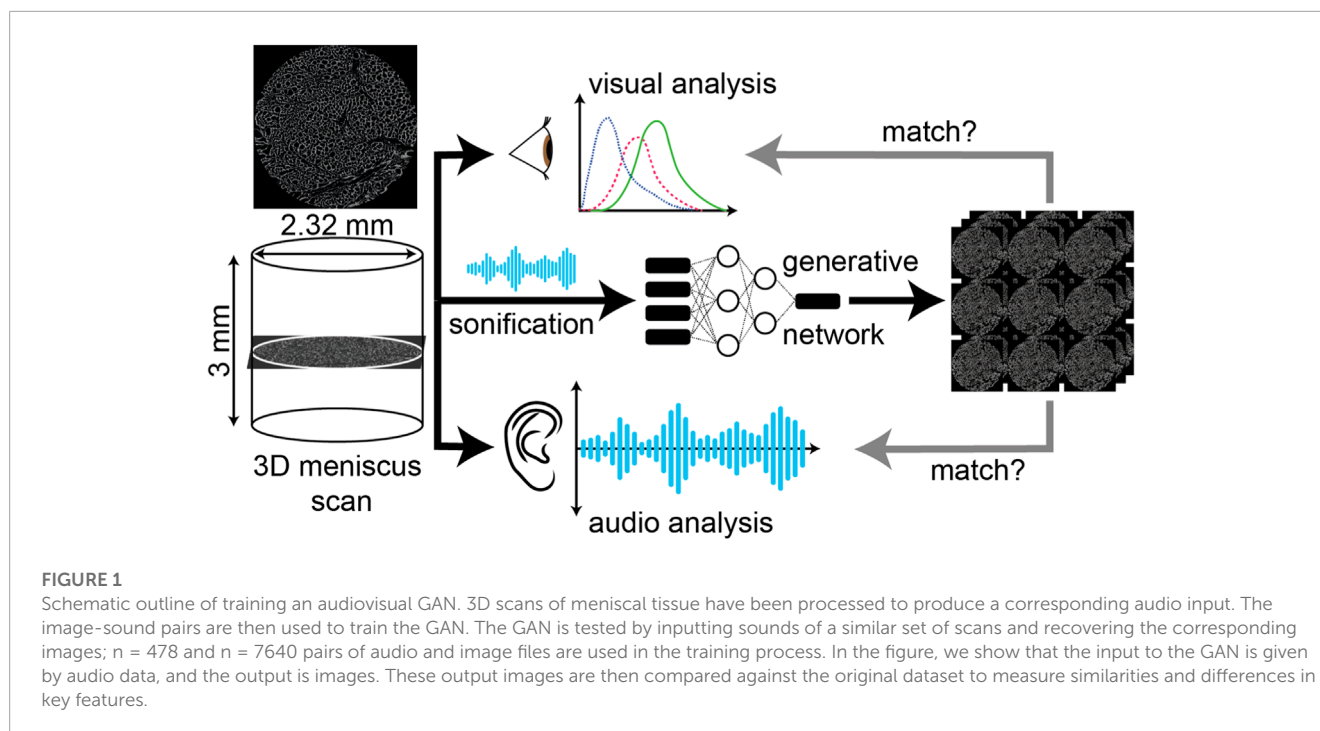
Natural materials, such as soft tissues, present spatially inhomogeneous architectures often characterized by a hierarchical distribution of pores (Vetri et al., 2019; Bonomo et al., 2020; Maritz et al., 2020; Agustoni et al., 2021; Maritz et al., 2021; Elmukashfi et al., 2022). Due to the shape of pores observed at different scales and the resulting intricate network of pore connectivity, the characterization of architectural parameters — porosity, pore size, pore connectivity, and tortuosity — is a non-trivial task and is the current object of an extensive research effort (Cooper et al., 2014; An et al., 2016; Rabbani et al., 2021). It has been observed that the knee meniscus is one such example of a soft tissue with remarkable properties in terms of its ability to transfer a load from the upper to the lower part of the body. The structure is similar to a sandwich structure, with a stiff outside layer and a softer internal layer that can accommodate deformation and dissipate energy. This tissue can be seen as an effective damping system designed and optimized by nature (Waghorne et al., 2023a). The secret of the internal layer is the arrangement of the collagen in a non-uniform network of channels oriented in a preferential direction guiding the fluid flow paths (Vetri et al., 2019; Waghorne et al., 2023a; Waghorne et al., 2023b). Understanding how mechanical parameters such as permeability are linked to the morphology of these channels and their interconnection/tortuosity is essential to the design of suitable biomimetic systems that can be adopted for the replacement or repair of meniscal tissue. To date, there are no such suitable artificial material systems.

Designing artificial structures that can resemble both the internal architecture and the material properties of such natural objects is currently an area of ongoing study (Shah and Sitti, 2004; Libonati et al., 2021; Zolotovskiy et al., 2021), but there are a number of difficulties. *In vivo* 3D visualization and quantification of biological microstructural parameters are difficult to obtain as the tissue often changes during observation, and image segmentation techniques—needed to locate image features and boundaries (lines, curves, etc.)—carry errors. The process requires high pixel-level accuracy and intensive manual intervention. A single image can take up to 30 min or more to process. With the development of image acquisition techniques such as microcomputed tomography (microCT) scans with lab-scale or synchrotron x-ray sources and multi-photon microscopy, it is possible to reach an extremely high resolution (up to 300 nm/voxel). However, this results in large datasets over 0.5TB, which are difficult to store and analyze. Analyzing these large datasets and quantifying distributions of architectural and material parameters is computationally expensive. Recent advances from our group on image analysis methods (Waghorne et al., 2023a; Waghorne et al., 2023b) allow the statistical

quantification of architectural features – porosity, pore diameter, pore connectivity, and tortuosity – of natural and highly porous materials. However, the aim of this work is to lay the foundations for an innovative approach to revealing the content of this data whilst finding suitable lower-order representations of it. The idea is to achieve this by moving through sensorial spaces, particularly sound, and exploring how the relationships between those spaces can enhance our perception of the data. In particular, our aim is to develop a transperceptual method to detect architectural features that might be missed by the analysis of segmented images (Figure 1).

Sonification is the presentation of data via the medium of sound, in contrast to the typical visual display of data (visualization). Sonified data utilize aural processing to try and rapidly understand large datasets that may be difficult or time-consuming to visualize and view on-screen (Rimland et al., 2013) and is an appealing approach to enable data analysis by those who struggle to visually process data (Ballora, 2014; Beans, 2017). Sonification has previously been applied to data from fields as diverse as medicine (Ballora et al., 2000; Ahmad et al., 2010), astronomy (Zanella et al., 2022), and chemistry [Mahjour et al., 2023], with many different approaches being taken to map sound to data and produce an audio output. Challenges in the sonification field include identifying the best mappings of sound onto a dataset to render subtle changes and differences audible and building the tools to render numerical or visual data into sound quickly and efficiently. The key advancement in the method developed here is the adoption of a transperceptual approach (Harris, 2021) to data exploration in which sound and image are used in combination and reciprocally as opposed to in isolation, which can be typical of standalone sonification projects. Research points to the value of sound and image in enhancing one another, and the aim of a transperceptual approach is to capture that added value (Chion, 2019) and harness it for the purposes of data exploration.

To unify our audio and visual data, we have also explored the use of audio-to-image and image-to-audio generative adversarial networks (GANs). GANs have achieved excellent performance in many domains, such as image synthesis (Goodfellow et al., 2014; Karras et al., 2020), image-to-image translation (Isola et al., 2017; Zhu et al., 2017), and audio-to-image generation (Duarte et al., 2019; Wan et al., 2019). In this work, we utilize audio-to-image GANs to learn the relationship between audio and image data. We train the sound domain with the audio files of each image. More specifically, given a piece of audio, we utilize an audio-to-image GAN to generate the corresponding image. From the perspective of human perception, this method provides us with a way of analyzing audio data from visual data. Here, the potential reciprocity between image analysis, sonification, and subsequent training of the GAN



as part of a transperceptual approach offers the potential to expand the method to the point where only the GAN is needed to reverse engineer the initial image analysis. Although at an early stage, the results thus far are promising. Here, we start with samples of knee meniscus, providing an overview of the imaging data (microCT scans), and explore image analysis techniques for the quantitative analysis of the architectural properties. We then apply a sonification process to convert the data (image) into various sounds. These images and sounds were then used to train and test a GAN. Our results show that the 3D structure of complex tissues can be understood in the sound domain, and we can successfully generate an artificial dataset of soft tissues with biomimetic architectures with our GAN.

2 Materials and methods

2.1 Image acquisition, processing, and statistical analysis

A sample of dimension $9\text{ mm} \times 10\text{ mm} \times 12\text{ mm}$ (Figure 2B) was extracted from the centre part of the medial meniscus (Figure 2A). Micro-computed tomography (μCT) analyses were carried out with an image resolution of $6.25\mu\text{m}/\text{px}$. Four different volumes of interest (VOIs) were extracted from different regions of the scanned sample (Figure 2B) (Waghorne et al., 2023a). The scanning dataset and procedure are the objects of another study (Waghorne et al., 2023a). We have analyzed in detail here VOI4: a cylinder with a diameter of 2.32 mm and height of 3 mm; in total, this dataset contains 478 images. It was necessary to crop the circular cross section of all images to rectangular due to the background pixels present on the scan that would impair the sonification. The dataset was cropped to an image size of 256×256 pixels, as shown in Figure 2C. From here, two separate preprocessing techniques

were employed, with each method used for a different experiment. In the first instance, the images were compressed to fit within a 64×64 frame using the “Resize” function from the Pillow library (Clark, 2015) within Python; this was due to a limitation of the GAN that will be discussed later. A workflow for this downsampled data can be seen in Figure 6, along with an example of its generated counterpart. As the original images have only been downsampled for this process, the number of images input remains at 478. For the second experiment, rather than downsampling the images to compress them to the required size, the whole images are instead segmented into 64×64 squares. This method increases the data input to 7468 images and avoids the loss of any information due to downsampling; however, this comes at the expense of no longer interpreting each sample holistically. As the samples were deconstructed in preprocessing, the corresponding generated images needed to be reconstructed in post-processing. A workflow for this full-resolution dataset can be seen in Figure 9.

2.2 3-Dimensional analysis

To quantify the characteristic features of the microstructures, a MATLAB code was written for the image processing of the 3D dataset and performing statistical analysis of the architecture (Waghorne et al., 2023b). The main steps of the software are as follows:

- The images are initially preprocessed by binarisation using Ostu’s method, followed by a “majority” transform to remove insignificant features that can interfere with the segmentation (Rabbani et al., 2014).
- A distance transform is performed on the binary 3D image along with median filtering to improve the quality of segmentation. Watershed segmentation is then applied to

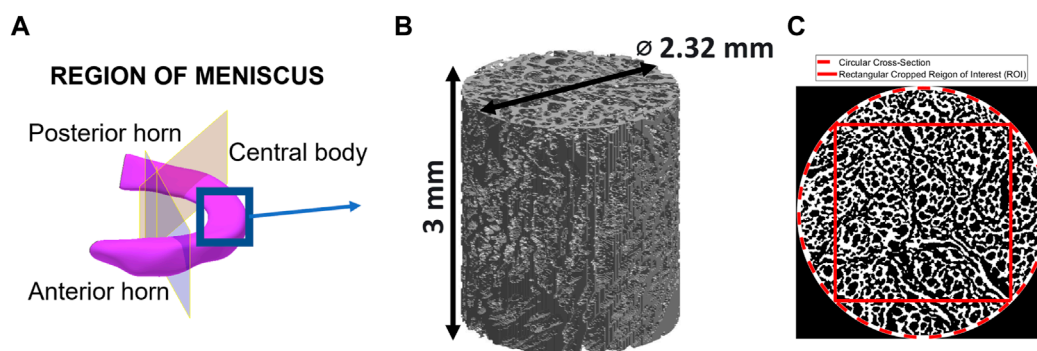


FIGURE 2

(A) Schematic representation of lateral meniscus. (B) The cylindrical dataset extracted, $2.32\text{mm} \times 3\text{mm}$, is denoted here as VOI(4). (C) An example of an image from the 2D stack—in total, 478 images—are reconstructed in 3D.

this distance map, leading to the labelling and quantification of pores (Rabbani et al., 2014). This approach is seen to consistently yield results similar to more classical methods at a fraction of the computation time (Baychev et al., 2019).

- Morphological characteristics can be acquired from the segmented pore space, most importantly, pore size. The size is described by creating an “equivalent sphere” of equal volume to the pore. The diameter of the created sphere is used to describe its size.
- The connectivity of the pore network can be identified using a methodology taken from (Rabbani et al., 2016). This is accomplished by dilation of the 3D binary image, in combination with the labelled pore matrix attained by segmentation, to interpret the connection of pores.
- Lastly, the pore network can be modelled as a graph, with the centroid of pores acting as nodes and edges representing the distance between connected pores. With inspiration taken from Sobieski (2016), regularly spaced start and finish points are assigned to the graph. The shortest path between all possible start and finish points is then calculated, providing a distribution of geometric tortuosities (Ghanbarian et al., 2013).

2.3 Sonification

The sonification involved the creation of a 255-voice synthesizer mapped to the luminance (brightness) of the pixels in the images. The process involved sorting each image in turn to assess the luminance of each pixel, then using the combined number of pixels at each luminance level to control the amplitude of the corresponding voice of the 255-voice synthesizer. The amplitude was scaled for each synthesizer voice to avoid overloading the audio buffer. However, this scaling was fixed for each voice as opposed to dynamic, with each voice being scaled in relation to the others. This model could run in real time and did not require modification of the images but did involve significant amounts of phase interference due to the proximity of the frequencies in the synthesizer. This can be seen in the audio data files attached as supplementary data. The model, at this stage, was unable to retain spatial information regarding the location of areas of luminance in

the image. This information was included in an earlier approach to the sonification process that required excessive downsampling of the images to a smaller size ($16 \times 16\text{px}$). It retained spatial information by conducting per-pixel, pitch-based sonification up to a maximum of 1024 voices. While the retention of spatial information would have been beneficial for the GAN model to learn, having images at such a low resolution prevented the extraction of any meaningful sample characteristics. While the conversion from image to the audio domain used in this study can be viewed as a limitation rather than a benefit, it is important to bear in mind that the objective of this study is to explore the potential that the audio data have to offer while laying the foundations of a method for others to do the same. As such, the sonification presented in this study is simplistic, making it accessible for others to imitate and build upon. Future work should aim toward developing a sonification method that can incorporate spatial information without a detrimental loss of resolution and the inclusion of a dynamic scale to represent the preponderance of luminance values.

2.4 Audio-image GANs

The audio domain presents a novel way to represent and transform data. Through the addition or reduction of noise, filtering, or direct editing of the spectral display, to name a few examples, there are numerous ways that an audio dataset can be altered, either very subtly or more dramatically. These altered datasets are computationally very cheap to produce and can quickly create a huge range of artificial samples to insert into a GAN model once trained. In this study, we utilize the audio-image generative adversarial network (GAN) proposed by Duarte et al. (2019) to model the relationship between audio and image data. Figure 3 shows the training and inference phase for the audio-image GAN. The training phase mainly consists of three neural network modules: an audio encoder, an image generator, and a discriminator. The audio encoder takes a piece of audio as an input and generates corresponding audio features. The audio features are fed into the image generator, and a fabricated image can be generated. The discriminator is responsible for differentiating between the fabricated image from the generator and the real image from the training set. After finishing the training

process, only the audio encoder and the image generator are utilized in the inference phase. More specifically, given any piece of audio, the audio encoder transforms it using the audio features. Then, a generated image can be obtained through the image generator. The initial dataset contained 478 pairs of image and audio data. We randomly chose 400 pairs of data as training data, and the remaining pairs were used for testing. For the larger segmented image dataset, we randomly chose 7000 pairs as training data and 648 pairs for testing. We have based our approach on a similar implementation of work by Duarte et al. (2019) for the audio-image GAN. Specifically, we set the number of channels as one because images are grayscale in our work. The number of epochs (learning iterations) is set as 2,000 for the GAN model trained on downsampled images and 1000 for the model trained on full-resolution images. The learning rates of the generator and the discriminator are fixed as 0.00001 and 0.00004. In Figure 4, we show that the MSE reduces more rapidly in the case of the full-resolution dataset—Figure 4B—than in the case of the downsampled dataset shown in Figure 4A. Therefore, the learning iterations (epoch) for the full-resolution image training set were set to 1,000.

3 Results and discussion

The sonification process outputs (audio files), extracted using the methods described in Section 2.3 from both these experiments, along with their corresponding images, are provided in the Supplementary Material section. The audio-image GAN resulted in the images shown in Figure 6C and Figure 9D. We use mean squared error (MSE) to measure the difference between original and generated images to evaluate the quality of reconstructed images. The design choice to use MSE was made as this is a foundation study of this novel methodology; it was imperative to use standard and simple metrics that will be easy for others to replicate. Future developments of this method could use more advanced metrics, such as structural similarity index measure (SSIM), which has been shown to be more robust at differentiating between similar images, rather than using luminance values like MSE. This method could be further improved by incorporating the statistical parameters of the image calculated beforehand using image analysis, either through the sonification process or as an input to the GAN directly. However, this study set out what relationship could be obtained from the images alone, without the need for excessive preprocessing or analysis that introduces its own error. Given original images $X_{ori} = \{x_{ori}^1, x_{ori}^2, \dots, x_{ori}^n\}$, the corresponding reconstructed images $X_{res} = \{x_{res}^1, x_{res}^2, \dots, x_{res}^n\}$, the MSE is computed by

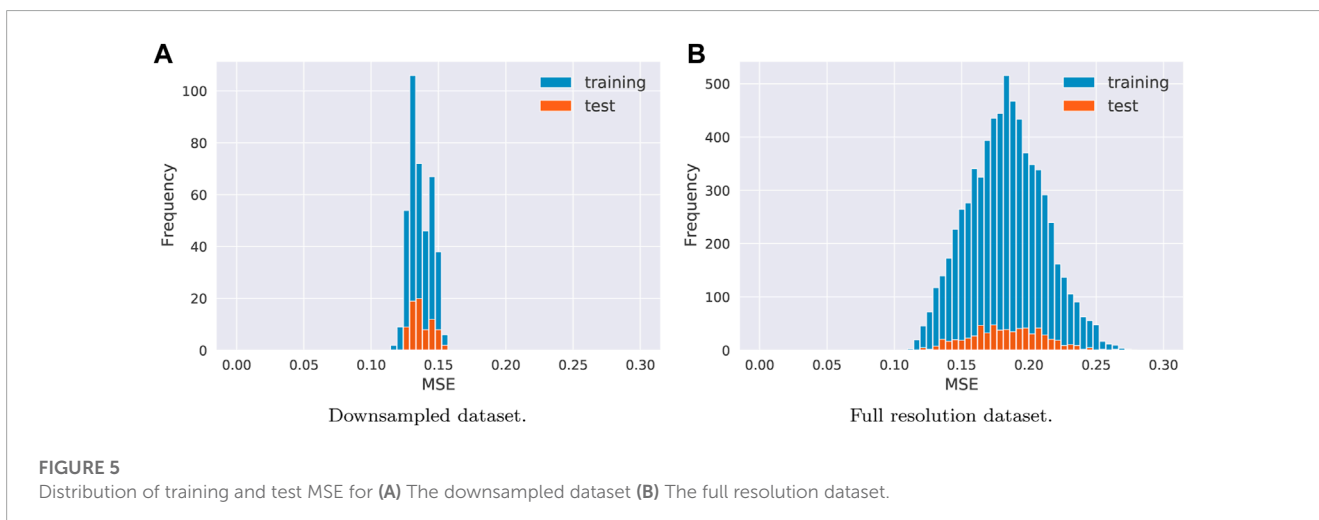
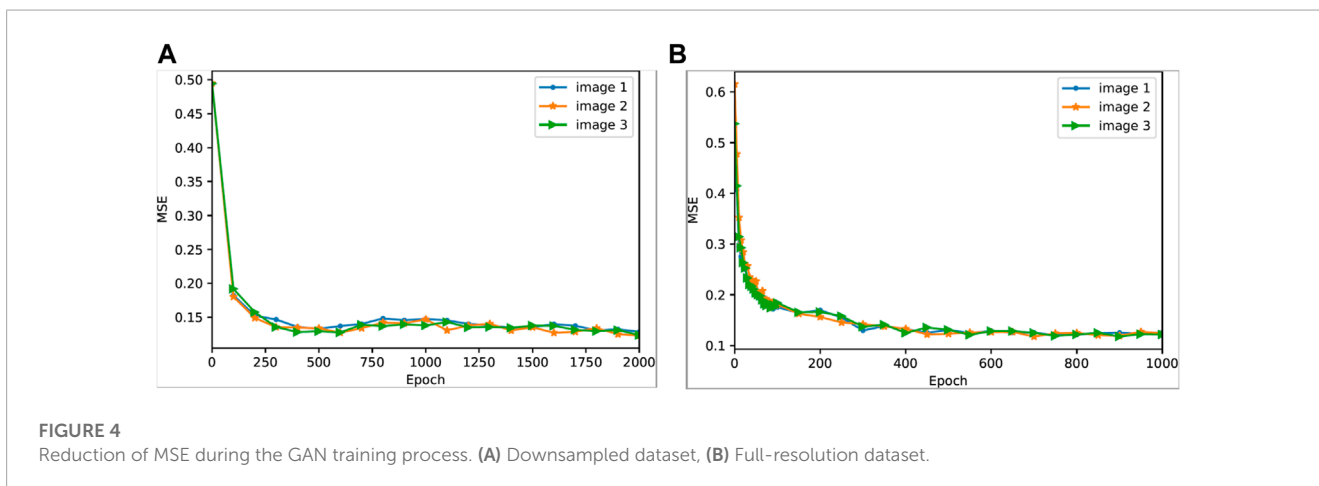
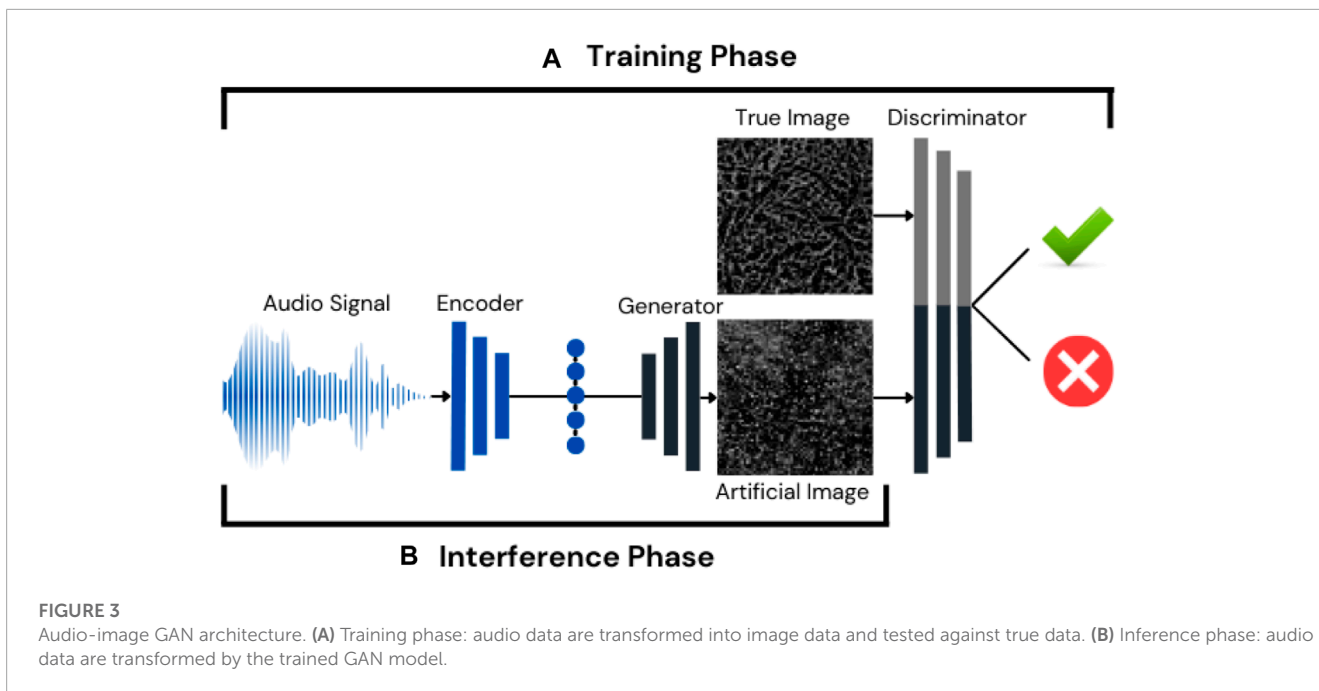
$$MSE = \frac{1}{n} \sum_{i=1}^n (x_{res}^i - x_{ori}^i)^2. \quad (1)$$

Figure 5 shows the frequency of MSE on training and test data. We can observe that the MSE distribution of training and test data is similar. Specifically, the mean MSE of training data is 0.1365, while the test data have a mean MSE of 0.1372 for the downsampled dataset. The full-resolution dataset has a mean MSE of training data of 0.1831 and a mean MSE of test data of 0.1829. Figure 6 shows some examples of original images and reconstructed images. Overall, given audio samples as input, the GAN can roughly predict the trends of the original images.

Downsampled dataset: 478 pairs of images and audio

As discussed in Section 2.1, a first attempt at producing mimetic native tissues was met by means of downsampling the original images to fit within a 64×64 frame, as seen in Figure 6. Before analyzing the results produced by the GAN, we present the GAN's learning process. Before any information has been passed through the GAN, the image produced appears as solid grey (Figure 7A). At the start of the learning process, the distribution of luminance values seems to be closer to the original image. The pixels are, however, seemingly randomly distributed across the image frame, as seen in Figure 7B. By the end of the process, the GAN appears to have learnt some of the key characteristics of the images, most noticeably clustering. The final images in the training stage (Figures 7C, D) appear to have more condensed regions of darker pixels. This is the GAN's method of emulating pores, not because there has been any spatial information provided by the audio but because having clusters of dark regions increases the chance of producing a lower MSE by matching with pores.

These datasets have been analyzed first in the 2D version, highlighting the discrepancy in parameters between the corresponding original and generated image pairs. The parameters evaluated are first related to the information provided to the GAN via the audio; in this instance, quantities of pixel luminance are shown in Figures 8A, B. Figure 8A is created by simply calculating the mean average of all the luminance values within an image. It is clear that the generated images stably follow the same trend as the native dataset. On average, there is an 8% error between the images. It is worth noting that, for legibility, a moving average window of 25 has been applied to Figures 8A, C, D; this is also true for Figure 11. Figure 8B shows the distribution of the average luminance values and demonstrates good agreement between them, in both range and values, with the exception of 0 luminance. The model appears to have difficulty recognising black pixels. This is likely to be due to the relative scaling of amplitudes during the sonification process. There is a large discrepancy between the number of black pixels and all other grayscale pixels. This difficulty interpreting black pixels also contributes to the offset in average image luminance seen in Figure 8A. The second set of results (Figures 8C, D) required some post-processing before it could be analyzed. The filtering and segmentation methodology laid out for 2D images in Rabbani et al. (2014) was followed to provide a quantified 2D pore space. It was found that a median filter produced the best results (Chion, 2019). The porosity seen in Figure 8C also presented very good agreeability, on average only showing a 4% error between the values. Interestingly, it is seen that the GAN tends to create images that overpredict porosity, despite having higher average luminance values. This is likely due to the binarisation method used. While the similar luminance distributions produce a similar average threshold value, as seen in Figure 8B, the threshold of the generated images is slightly higher, which will lead to more frequent black pixels and higher porosity. This also explains why the porosity tends to be better in regions where the mean luminance value is closer to the threshold. Lastly, we present the average pore size in Figure 8C. While this demonstrates an average 8% error in pore diameters, this parameter requires spatial information to match successfully. The resemblance in trends and low error values we see has been entirely learned by the GAN network by trying to minimise the MSE. The results show an overestimation, likely because having a



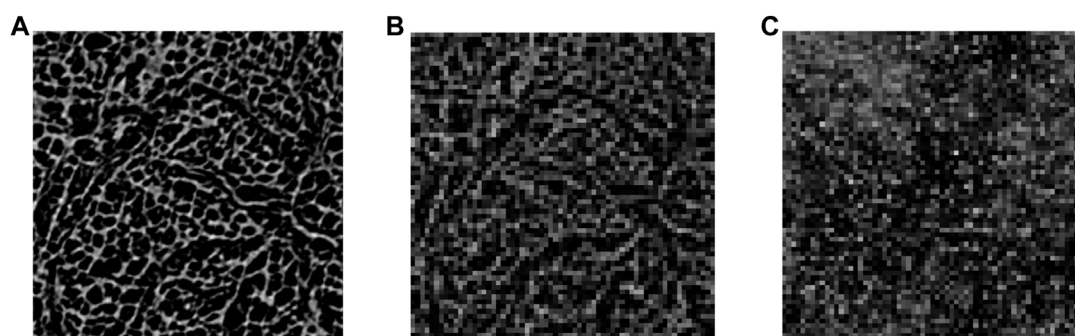


FIGURE 6

Workflow of downsampling analysis: (A) Original dataset (256×256), (B) Downsampled image (64×64), and (C) GAN-generated image (64×64).

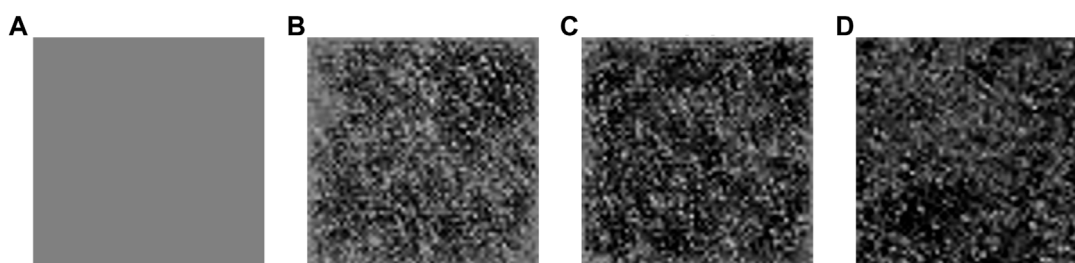


FIGURE 7

Training progression of GAN with downsampled images. (A) Learning iteration = 0, (B) Learning iteration = 50, (C) Learning iteration = 100, and (D) Learning iteration = 2000.

slightly larger cluster of dark pixels allows for the crossover of more than one pore. The large clusters of dark pixels would likely decrease the MSE as there are more black pixels present in the original data.

Full-resolution dataset: 7648 pairs of images and audio

The second experiment revolves around maintaining as much of the native morphology as possible by segmenting the initial full-resolution images (256×256) instead of downsampling them. This experiment comes with the additional step of reconstruction for analysis, as seen in [Figures 9C, D](#). The additional step consists of segmenting the image into 16 squares of size 64×64 , assigning unique names for each segmented image based on the location in the image stack, and maintaining this through all processing stages. Barring the reconstruction, all other aspects of the analysis remained the same as the downsampled dataset. The first step, again, was to interpret the GAN's learning process to ascertain an understanding of what additional features have been detected compared to the downsampled dataset. The learning process begins in a relatively similar fashion to that of the downsampled experiment, with [Figures 10A, B](#) demonstrating the formation of clusters; however, toward the end of the training, the images seem to take on a different form. [Figures 10C, D](#) present blurring within some sections of the generated images and a cross-hatch pattern in others. One of the possible reasons is that, due to retaining the original resolution of the image, there are more frequent black pixels (luminance = 0) and more intensely bright pixels. The GAN's way of minimising the MSE is to not include clusters or distinct boundaries but instead to have smoother areas with less intense luminance values. This avoids

incurring a high MSE if pixels were to fall within the wrong side of the pore boundary.

The first two results of [Figure 11](#) support this theory, as [Figure 11B](#) shows a clear shift of the GAN's luminance distribution to the left, with hardly any pixels with luminance > 75 . This accommodates the now-huge discrepancy of zero luminance values to all others in the original data. This shift is reflected in [Figure 11A](#). With the GAN consistently producing pixels skewed to be darker, the average luminance value drops considerably, even lower than the natural morphology, albeit with its spike in 0 pixels. This difference means that the error in mean luminance value across the sample in this experiment increases to almost $4\times$ that of the downsampled instance. The porosity in this experiment is not only considerably different in values, but, as [Figure 11C](#) shows, the GAN is now underestimating the porosity rather than overestimating. Again, this is due to the binarisation process as the distributions of the luminance values are very dissimilar. The original data have clearly defined background and foreground pixels with a much more contrasted range, while the generated data are not so dichotomic. This causes the threshold for the generated images to be much lower than that of the original. This discrepancy in threshold, along with the original data's heavy right skew, explains the large discrepancy in porosity. The pore size comparison does not perform well in this experiment; the GAN does not emphasise clustering. Therefore, the pore size is considerably smaller than the native morphology, on average being 30% lower.

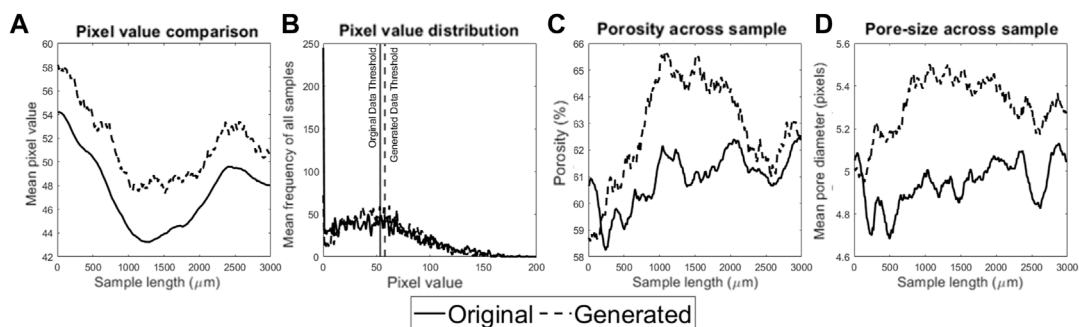


FIGURE 8 2D Analysis of the downsampled dataset: (A) Mean luminance across the sample, (B) Luminance distribution, (C) Porosity across the sample, and (D) Pore size across the sample.

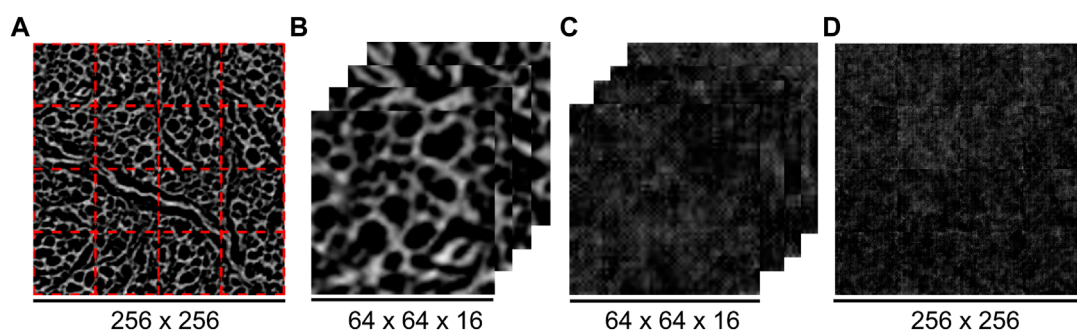


FIGURE 9 Workflow of full-resolution analysis: (A) Original image showing segmentation lines (red), (B) Deconstructed original image, (C) Deconstructed generated image, and (D) Reconstructed generated image.

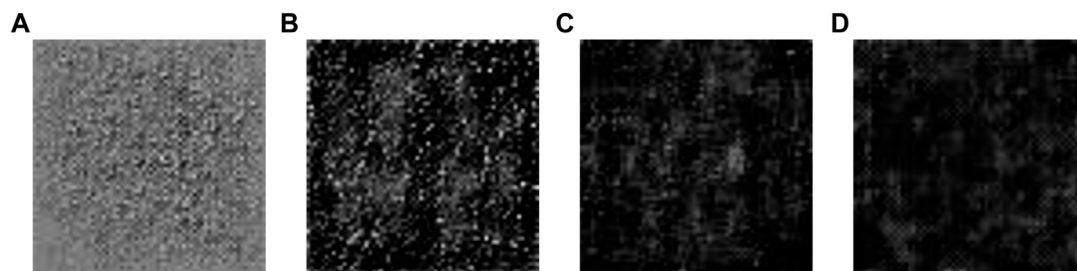


FIGURE 10 Training progression of GAN for the full-resolution dataset. (A) Learning iteration = 5, (B) Learning iteration = 100, (C) Learning iteration = 750, and (D) Learning iteration = 1,000.

The results of the 3D analysis describe the effect of how the GAN model learning process can affect the generation of images. It is first important to note that these graphs are all normalised on the Y-axis. This is because we are interpreting images of varying sizes, which will have a large effect on the frequency of occurrence. In other words, the results have been normalised for comparison purposes. In [Figure 12A](#), we can see the variation in distributions of the pore size. The largest and widest distribution belongs to the original data, with the spatial relation of these pores larger in 3D space. This is followed closely by the downsampled dataset, which is only 1/4 the size of the original. If the resolution of the image were to be scaled up, it would

likely have clusters considerably larger than the natural morphology. Lastly, as very little clustering occurs in the full-resolution dataset, further intensified by a lack of realization between planes, this results in a very thin pore size distribution. In [Figure 12B](#), the distribution of pore connectivity appears relatively similar among the original, downsampled, and full-scale images. There is a relationship between the pore size and connectivity due to the available surface area. This is likely the reason for the full-resolution dataset presenting the thinnest distribution. The downsampled and original connectivities are relatively close, with the downsampled seemingly having a wider spread; this is perhaps due to the lack of conventional boundaries

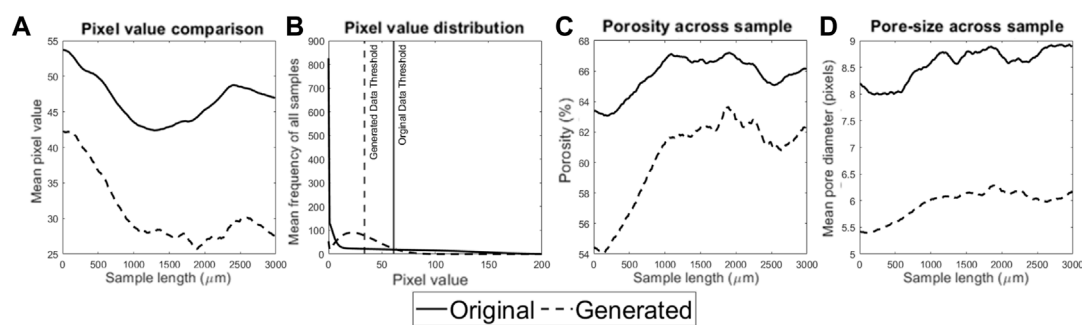


FIGURE 11

2D analysis of the full-resolution dataset: (A) Mean luminance across the sample, (B) Luminance distribution, (C) Porosity across the sample, and (D) Pore size across the sample.

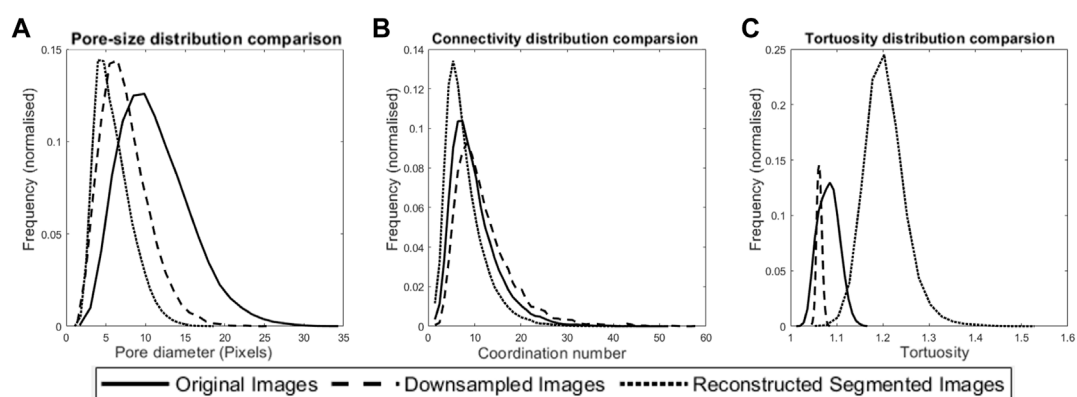


FIGURE 12

3D analysis of all datasets: (A) 3D pore size distribution, (B) Pore connectivity distribution, and (C) Tortuosity distribution.

between pores that typical pore structures see in the original data. The tortuosity distributions in [Figure 12C](#) clearly demonstrate where the analysis differs the most. The downsampled dataset is likely highly limited by the reduced size; there are far fewer start and finish points, and the channel is restricted to a very small area, leading to smaller tortuosities. The full-resolution dataset, however, presents smaller pores with fewer connections; therefore, it leads to a more uniform tortuosity and to the wider distribution in [Figure 12C](#).

4 Conclusion and future perspectives

From the results presented, we have demonstrated that it is possible for a GAN to produce an artificial image dataset mimicking a parameter described to it through audio information. Both this and other characteristics can be taught during to the GAN through the learning process. While the audio information and results presented thus far may seem primitive, the work lays the foundation for increased complexity. With audio information that can describe more features, results will improve, as will the potential for the GAN to uncover a hidden characteristic not clearly visible with visual perception alone. These kinds of developments could be used

in a wide range of applications, such as assisting with diagnoses, exploring new materials, and generating artificial structures, to name a few.

This investigation has also highlighted some potential weaknesses in the methodology. The interaction between the audio information and GAN is complicated and difficult to evaluate directly. A standardized methodology for evaluating weighting factors of the GAN's neural network would be highly beneficial for improving results and uncovering relationships faster. To further this, careful thought and consideration must go into selecting what data would be most suitable for use. Two different sets of data were used in these experiments: one containing downsampled images, and the other is full-resolution data with a unit of magnitude greater volume of data, with the latter set performing considerably worse. This is a paradigm example of how situations that may typically look good in theory do not behave so in practicality. However, if these precautions are taken, then this methodology has the potential to be very powerful.

The method proposed in this paper has the potential to be adopted for analyzing and predicting a wide variety of biomimetic architectures. The original dataset analyzed consisted of a stack of 478 circular 2D images (371×371 pixels) that have been reconstructed as a 3D object whose architectural properties have

been statistically characterized, sonified, and then reconstructed. 2D and 3D analyses of the reconstructed data show that the audiovisual GAN performed better in generating artificial datasets when trained on the downsampled images. The mean of the luminance values, porosity, and pore size are within 4%–8% of the original dataset.

The proposed method has the following limitations:

- The current GAN architecture can only deal with image sizes up to 64×64 pixels, while the sonification process can deal with any size image.
- The sonification method currently can underestimate the sound of black pixels (pore space). The downsampled dataset contains fewer black pixels than the full-resolution images. Therefore, despite the improved resolution quality of the full-resolution set to train the GAN, the actual sonification methodology is more suitable for the downsampled dataset.

Our future development will focus on how to use the GAN to automatically generate audio datasets from the sample scans and investigate how audio data might be perceived differently for different samples by humans and computers.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

References

- Agustoni, G., Maritz, J., Kennedy, J., Bonomo, F. P., Bordas, S., and Barrera, O. (2021). High resolution micro-computed tomography reveals a network of collagen channels in the body region of the knee meniscus. *Ann. Biomed. Eng.* 49 (9), 2273–2281. doi:10.1007/s10439-021-02763-6
- Ahmad, A., Adie, S. G., Wang, M., and Boppart, S. A. (2010). Sonification of optical coherence tomography data and images. *Opt. Express* 18 (10), 9934–9944. doi:10.1364/OE.18.009934
- An, S., Yao, J., Yang, Y., Zhang, L., Zhao, J., and Gao, Y. (2016). Influence of pore structure parameters on flow characteristics based on a digital rock and the pore network model. *J. Nat. Gas Sci. Eng.* 31, 156–163. doi:10.1016/j.jngse.2016.03.009
- Ballora, M., Pennycook, B., Ivanov, P., Goldberger, A., and Glass, L. (2000). Detection of obstructive sleep apnea through auditory display of heart rate variability. *Comput. Cardiol.* Vol. 27, 739–740. doi:10.1109/CIC.2000.898630
- Ballora, M. (2014). Sonification, science and popular music: In search of the 'wow'. *Organised Sound*. 19 (1), 30–40. doi:10.1017/S1355771813000381
- Baychev, T. G., Jivkov, A. P., Rabbani, A., Raeini, A. Q., Xiong, Q., Lowe, T., et al. (2019). Reliability of algorithms interpreting topological and geometric properties of porous media for pore network modelling. *Transp. Porous Media* 128, 271–301. doi:10.1007/s11242-019-01244-8
- Beans, C. (2017). Musicians join scientists to explore data through sound. *Proc. Natl. Acad. Sci.* 114 (18), 4563–4565. arXiv: doi:10.1073/pnas.1705325114
- Bonomo, F. P., Gregory, J. J., and Barrera, O. (2020). A procedure for slicing and characterizing soft heterogeneous and irregular-shaped tissue. *Mater. Today Proc.* 33, 2020–2026. doi:10.1016/j.matpr.2020.07.624
- Chion, M. (2019). "Audio-vision: Sound on screen," in *Audio-Vision: Sound on screen* (Columbia University Press).
- Clark, A. (2015). Pillow (pil fork) documentation. Available at: <https://pillow.readthedocs.io/en/stable/Refstyle.html>
- Cooper, S., Eastwood, D., Gelb, J., Damblanc, G., Brett, D., Bradley, R., et al. (2014). Image based modelling of microstructural heterogeneity in lifepo4 electrodes for li-ion batteries. *J. Power Sources* 247, 1033–1039. doi:10.1016/j.jpowsour.2013.04.156
- Duarte, A. C., Roldan, F., Tubau, M., Escur, J., Pascual, S., Salvador, A., et al. (2019). "Wav2pix: Speech-conditioned face generation using generative adversarial networks," in *IEEE international conference on acoustics, speech and signal processing (ICASSP)* (IEEE), 8633–8637. doi:10.1109/ICASSP.2019.8682970
- Elmukashfi, E., Marchiori, G., Berni, M., Cassiolas, G., Lopomo, N. F., Rappel, H., et al. (2022). Chapter five – model selection and sensitivity analysis in the biomechanics of soft tissues: A case study on the human knee meniscus. *Adv. Appl. Mech.* 55, 425–511. doi:10.1016/bs.aams.2022.05.001
- Ghanbarian, B., Hunt, A. G., Ewing, R. P., and Sahimi, M. (2013). Tortuosity in porous media: A critical review. *Soil Sci. Soc. Am. J.* 77, 1461–1477. doi:10.2136/sssaj2012.0435
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Proceedings of annual conference on neural information processing systems (NeurIPS)* (Curran Associates, Inc.), 2672–2680.
- Harris, L. (2021). *Composing Audiovisually: Perspectives on audiovisual practices and relationships*. CRC Press.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *Proceedings of IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (IEEE), 1125–1134.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). "Analyzing and improving the image quality of stylegan," in *Proceedings of IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (IEEE), 8110–8119.
- Libonati, F., Graziosi, S., Ballo, F., Mognato, M., and Sala, G. (2021). *3d-printed architected materials inspired by cubic bravais lattices*. ACS Biomaterials Science and Engineering.
- Mahjour, B., Bench, J., Zhang, R., Frazier, J., and Cernak, T. (2023). Molecular sonification for molecule to music information transfer. *Digital Discov.* doi:10.1039/d3dd00008g
- Maritz, J., Agustoni, G., Dragnevski, K., Bordas, S., and Barrera, O. (2021). The functionally grading elastic and viscoelastic properties of the body region of the knee meniscus. *Ann. Biomed. Eng.* 49 (9), 2421–2429. doi:10.1007/s10439-021-02792-1

Author contributions

JW material characterisation, writing, CH and LH audio-visuals files production, JP and HH GAN training and testing, WP conceptualisation and editing, OB data acquisition, characterization, conceptualization and funding acquisition.

Acknowledgments

The authors are grateful for the financial support provided by the SFC and FNR-funded European Crucible Prize.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Maritz, J., Murphy, F., Dragnevski, K., and Barrera, O. (2020). Development and optimisation of micromechanical testing techniques to study the properties of meniscal tissue. *Mater. Today Proc.* 33, 1954–1958. doi:10.1016/j.matpr.2020.05.807
- Rabbani, A., Ayatollahi, S., Kharrat, R., and Dashti, N. (2016). Estimation of 3-d pore network coordination number of rocks from watershed segmentation of a single 2-d image. *Adv. Water Resour.* 94, 264–277. doi:10.1016/j.advwatres.2016.05.020
- Rabbani, A., Fernando, A. M., Shams, R., Singh, A., Mostaghimi, P., and Babaei, M. (2021). Review of data science trends and issues in porous media research with a focus on image-based techniques. *Water Resour. Res.* 57 (10), e2020WR029472. doi:10.1029/2020wr029472
- Rabbani, A., Jamshidi, S., and Salehi, S. (2014). An automated simple algorithm for realistic pore network extraction from micro-tomography images. *J. Petroleum Sci. Eng.* 123, 164–171. doi:10.1016/j.petrol.2014.08.020
- Rimland, J., Ballora, M., and Shumaker, W. (2013). “Beyond visualization of big data: A multi-stage data exploration approach using visualization, sonification, and storification,” in *Next-generation analyst*. Editors B. D. Broome, D. L. Hall, and J. Llinas (SPIE: International Society for Optics and Photonics), 8758, 87580K. doi:10.1117/12.2016019
- Shah, G. J., and Sitti, M. (2004). “Modeling and design of biomimetic adhesives inspired by gecko foot-hairs,” in *2004 IEEE international conference on robotics and biomimetics* (IEEE), 873–878.
- Sobieski, W. (2016). The use of path tracking method for determining the tortuosity field in a porous bed. *Granul. Matter* 18 (8), 72. doi:10.1007/s10035-016-0668-3
- Vetri, V., Dragnevski, K., Tkaczyk, M., Zingales, M., Marchiori, G., Lopomo, N., et al. (2019). Advanced microscopy analysis of the micro-nanoscale architecture of human menisci. *Sci. Rep.* 9, 18732. doi:10.1038/s41598-019-55243-2
- Waghorne, J., Bonomo, F., Bell, D., and Barrera, O. (2023a). *On the characteristics of natural hydraulic dampers: An image based approach to study the fluid flow behavior inside the human meniscal tissue*. To be submitted to PNAS.
- Waghorne, J., Scheper, T. O., and Barrera, O. (2023b). *Statistical quantification of morphological and topological characteristics in heterogeneous, higher-porosity media*. To be submitted to computer methods in applied mechanics and engineering.
- Wan, C.-H., Chuang, S.-P., and Lee, H.-Y. (2019). “Towards audio to scene image synthesis using generative adversarial network,” in *IEEE international conference on acoustics, speech and signal processing (ICASSP)* (IEEE), 496–500.
- Zanella, A., Harrison, C. M., Lenzi, S., Cooke, J., Damsma, P., and Fleming, S. W. (2022). Sonification and sound design for astronomy research, education and public engagement. *Nat. Astron.* 6, 1241–1248. doi:10.1038/s41550-022-01721-z
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (IEEE), 2223–2232.
- Zolotovskiy, K., Varshney, S., Reichert, S., Arndt, E. M., Dao, M., Boyce, M. C., et al. (2021). Fish-inspired flexible protective material systems with anisotropic bending stiffness. *Commun. Mater.* 2 (1), 35–10. doi:10.1038/s43246-021-00140-3