



Article

SonoNERFs: Neural Radiance Fields Applied to Biological Echolocation Systems Allow 3D Scene Reconstruction through Perceptual Prediction

Wouter Jansen ^{1,2,*}  and Jan Steckel ^{1,2} 

¹ Cosys-Lab, Faculty of Applied Engineering, University of Antwerp, 2020 Antwerpen, Belgium; jan.steckel@uantwerpen.be

² Flanders Make Strategic Research Centre, 3920 Lommel, Belgium

* Correspondence: wouter.jansen@uantwerpen.be

Abstract: In this paper, we introduce SonoNERFs, a novel approach that adapts Neural Radiance Fields (NeRFs) to model and understand the echolocation process in bats, focusing on the challenges posed by acoustic data interpretation without phase information. Leveraging insights from the field of optical NeRFs, our model, termed SonoNERF, represents the acoustic environment through Neural Reflectivity Fields. This model allows us to reconstruct three-dimensional scenes from echolocation data, obtained by simulating how bats perceive their surroundings through sound. By integrating concepts from biological echolocation and modern computational models, we demonstrate the SonoNERF's ability to predict echo spectrograms for unseen echolocation poses and effectively reconstruct a mesh-based and energy-based representation of complex scenes. Our work bridges a gap in understanding biological echolocation and proposes a methodological framework that provides a first-order model of how scene understanding might arise in echolocating animals. We demonstrate the efficacy of the SonoNERF model on three scenes of increasing complexity, including some biologically relevant prey–predator interactions.

Keywords: Neural Radiance Fields; Neural Reflectance Fields; echolocating bats; ultrasound sensing; sonar; biosonar



Citation: Jansen, W.; Steckel, J. SonoNERFs: Neural Radiance Fields Applied to Biological Echolocation Systems Allow 3D Scene Reconstruction through Perceptual Prediction. *Biomimetics* **2024**, *9*, 321. <https://doi.org/10.3390/biomimetics9060321>

Academic Editor: Andrew Adamatzky

Received: 20 April 2024

Revised: 24 May 2024

Accepted: 25 May 2024

Published: 28 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Echolocating bats exhibit a strong ability to localize and discriminate prey objects using sound as the primary sensing modality. A subset of bats called gleaning bats are especially adept at finding and identifying prey in dense clutter [1–6]. One of the main theories that explains this behavior is that these animals make clever use of physical interactions of their echolocation signals and the clutter in which their prey is perched upon [7–11]. Another class of bats called nectar-feeding bats locates the flowers from which they nourish themselves using a special kind of leaf that is co-evolved by the pitcher plants that bear these flowers [12–14]. Similar traits of co-evolution have also been observed in other plant–bat relationships, such as bat-pollinated cacti [15]. In these plant–bat interactions, physical interactions between the emitted sound signals and the objects of interest give rise to emergent, stable spectral cues, which the bat can utilize to localize these leaves [14]. These insights have led to robotic applications, building a specific set of sonar reflectors that can be localized reliably by manufactured sonar sensors [16,17].

Although there is a growing body of evidence of the prevalence of these robust emergent cues, which bats utilize to solve their prey localization and identification task, the hypothesis of more in-depth scene understanding and modeling in bats still lingers. Indeed, when visiting bat research conferences, talks about 3D scene models in bats can often be overheard, giving rise to intense discussions about the sense or nonsense of such models existing. This is understandable, as humans like to reason in high-level 3D models

of the environment, as this representation is natural to us. However, it is essential to understand that there are significant differences between the sensory data originating from a complex scene when sensed using optical sensors (with wavelengths in the range of hundreds of nanometers) or with acoustic sensors (using wavelengths in the order of several millimeters [18]).

This prevalence of the concept of an ‘acoustic image’ or ‘acoustic 3D model’ is not unsurprising, given the vast amount of literature by the pioneering researchers of bat echolocation [19–23]. Furthermore, some researchers have proposed systems that perform tomographic reconstruction of complex scenes using echolocation-like signals and use these generated images to explain certain phenomena observed in bat–prey interactions [24–26]. Many of these previous works consider the problem of image formation based on the reception time-domain reception signal, which represents the acoustic wave field impinging on the external ears of the bats. However, an important note here is that phase information is lost as these pressure waves pass through the inner ear structures of the bat [27–29]. Indeed, as the bat’s cochlea can be modeled as a set of band pass filters, followed by an envelope detection step, it is apparent that the phase information is effectively lost from the reflected signals. Therefore, the assumption can be made that the inputs into the bat’s auditory system can be adequately modeled by the magnitude of the short-time Fourier transform of the impinging sound pressure waves [30], ignoring the logarithmic spacing of the frequency components in the bat’s auditory system.

In this paper, we aim to build upon this early research and try to lay the foundations for a model for effective 3D scene reconstruction in bats using the required phase-less information. To do this, we let ourselves be inspired by the seminal work achieved in novel view synthesis using deep learning networks. In particular, Neural Radiance Fields (NeRFs), which were first introduced in the seminal paper by Mildenhall et al. at the ECCV conference in 2020 [31]. In that paper, a novel approach for view synthesis is proposed based on building a differentiable visual rendering pipeline, which queries a radiance field represented by a multilayer perceptron (MLP) neural network. The MLP is trained for a specific scene based on images taken from multiple viewpoints, using the differentiable rendering pipeline (DRPL). Novel viewpoints can be generated using the DRPL and the learned radiance function. The original NeRF paper has inspired many researchers and sparked a whole new research field into improving upon the method proposed in the original paper [32–34]. Furthermore, the usage of NeRFs has been expanded to other application domains such as magnetic resonance imaging [35,36], ultrasonic medical imaging [37,38], multimodal acoustic/visual scene representation [39–41], and acoustic room impulse response prediction [42], with many more other examples to be found.

Based on the success of the underlying concept of NeRFs, more specifically, the differentiable rendering equation combined with a learnable radiance function, we try to explain 3D scene representation by echolocating bats using a NeRF-inspired model. Our model is called a SonoNERF, in which Sono represents ‘sound’ or ‘sonar’, and NERF stands for Neural Reflectivity Field instead of Radiance field, as reflectivity is more appropriate in the context of acoustic echolocation problems. In the remainder of this paper, we will introduce the underlying model of the SonoNERFs and explain how the differentiable rendering pipeline is tailored to the problem of echolocation in bats. Then, we will illustrate the model’s performance on various scenes, and discuss the performance of the model (Figure 1). Finally, we will draw some conclusions and highlight the shortcomings of our model at this point.

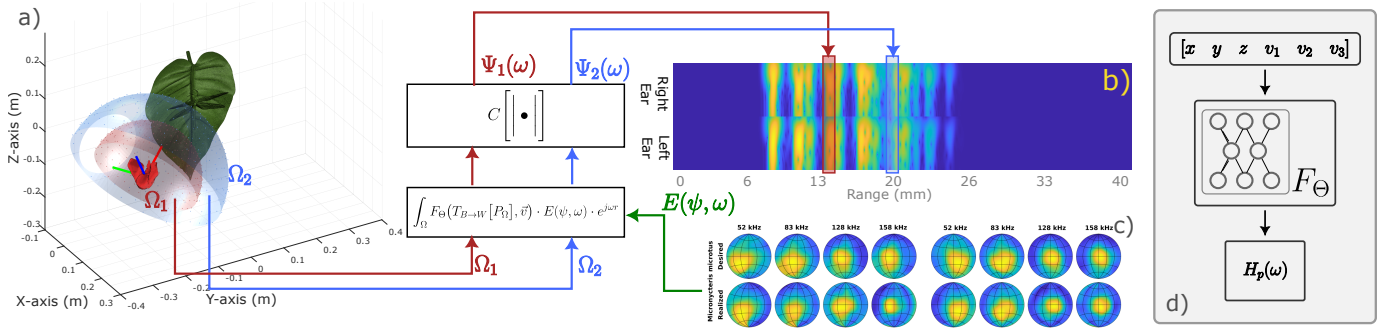


Figure 1. Illustration of the processing flow of the SonoNERF model. Panel (a) depicts the bat positioned at a single pose, showcasing two hemispheres Ω used for reflectivity function querying. The reflectivity function is sampled at 600 uniformly distributed points over Ω , which are then used to generate the received spectrum through the rendering equation. This spectrum undergoes nonlinear compression and is placed in the corresponding range slice within the predicted spectrogram panel (b). Panel (c) displays the Echolocation-Related Transfer Function (ERTF) for a *Micronycteris microtus* bat call, calculated following the methodology outlined in [43]. Panel (d) offers a schematic overview of the Neural Reflectivity field F_{Θ} , responsible for predicting the reflected spectrum $H_p(\omega)$ based on input position and incidence vector, all represented in world coordinates. The symbols Ω_1 and Ω_2 are two hemispheres over which the acoustic rendering equation is calculated, corresponding to two separate range bins in the spectrogram.

2. Echo Formation in Echolocating Bats

In this section, we will briefly explain the echo formation process in bat echolocation, as this is a requirement to understand the reasoning behind the operating principle of our proposed SonoNERFs. Without loss of generality, we assume that bats emit a broadband signal $s_e(t)$, which typically is some multi-harmonic chirp [44]. We are aware of the existence of so-called constant-frequency bats. Still, these typically do not perform the gleaning behavior that underlies the SonoNERF principle, so we assume a broadband call. This call is filtered by the external facial features of the bat by a direction-dependent transfer function $h_e(\psi, t)$, in which ψ is a direction vector in 3D, typically represented by the azimuth angle θ and elevation angle φ . The filtered emitted signal is then reflected by the environment, which we assume to be a collection of point-like reflectors (which, in the Huygens approximation of acoustics, is an acceptable assumption [45]). Each point-reflector filters the impinging signal with a filter $h_p(\eta, t)$, characterized by the impinging angle η (to allow non-isotropic reflection functions to exist in this model). Upon reflection and subsequent reception, the then filtered and reflected signals are filtered by the outer ears of the bats, with a filter $h_L(\psi, t)$ for the left ear and a filter $h_R(\psi, t)$ for the right ear. All these filtered paths are delayed by the round trip range (i.e., from the emitter to the point reflector and from the point reflector to the respective ear).

$$h_{LT}(p_i, t) = h_L(\psi_i, t) * h_{p_i}(\eta_i, t) * h_e(\psi_i, t) * \delta(t - \Delta t_i) \quad (1)$$

$$h_{RT}(p_i, t) = h_R(\psi_i, t) * h_{p_i}(\eta_i, t) * h_e(\psi_i, t) * \delta(t - \Delta t_i) \quad (2)$$

In this equation, the total filters for a specific point p_i for the left ear is $h_{i,L}(p_i, t)$ and for the right ear is $h_{i,R}(p_i, t)$, and consist of the convolution of a delayed dirac function ($\delta(t - \Delta t_i)$), the emission filter $h_e(\psi_i, t)$, the point reflector function $h_{p_i}(\eta_i, t)$, and the receiver Head Related Transfer Function $h_L(\psi_i, t)$. The received signal for the left and the right ear is the linear sum over all N point-like reflectors, as follows:

$$s_L(t) = \sum_{i=1}^N h_{LT}(p_i, t) * s_e(t) \quad (3)$$

$$s_R(t) = \sum_{i=1}^N h_{RT}(p_i, t) * s_e(t) \quad (4)$$

This time-domain representation can be transformed into the Fourier domain, in which the convolutions become multiplications, which facilitates discovering the underlying structure of the physical echo formation process. For this, we first transform the total filters, as follows:

$$H_{LT}(p_i, \omega) = H_L(\psi_i, \omega) \cdot H_{p_i}(\eta_i, \omega) \cdot H_e(\psi_i, \omega) \cdot e^{jkr_i} \quad (5)$$

$$H_{RT}(p_i, \omega) = H_R(\psi_i, \omega) \cdot H_{p_i}(\eta_i, \omega) \cdot H_e(\psi_i, \omega) \cdot e^{jkr_i} \quad (6)$$

We then plug these into the equations for the total signals in the left and right ear, as follows:

$$S_L(\omega) = \sum_{i=1}^N H_{LT}(p_i, \omega) \cdot S_e(\omega) \quad (7)$$

$$S_R(\omega) = \sum_{i=1}^N H_{RT}(p_i, \omega) \cdot S_e(\omega) \quad (8)$$

Often, it makes sense to combine the HRTFs of the ears (i.e., $H_L(\psi_i, \omega)$) with the transfer functions of the emitter filter ($H_e(\psi_i, \omega)$) into an object called the ERTF (echolocation-related transfer function, called $E(\psi_i, \omega)$), as follows:

$$E_L(\psi_i, \omega) = H_e(\psi_i, \omega) \cdot H_L(\psi_i, \omega) \quad (9)$$

$$E_R(\psi_i, \omega) = H_e(\psi_i, \omega) \cdot H_R(\psi_i, \omega) \quad (10)$$

Which in turn reduces the equation for the total signal filter for point p_i to the following:

$$H_{LT}(p_i, \omega) = E_L(\psi_i, \omega) \cdot H_{p_i}(\eta_i, \omega) \cdot e^{jkr_i} \quad (11)$$

$$H_{RT}(p_i, \omega) = E_R(\psi_i, \omega) \cdot H_{p_i}(\eta_i, \omega) \cdot e^{jkr_i} \quad (12)$$

This subtle difference between the HRTF and the ERTF becomes important later in this paper when we describe the differentiable rendering pipeline we implemented for our SonoNERF model. The received signals $s_L(t)$ and $s_R(t)$ still contain the phases of the signals. However, as we have argued before, phase information is not considered readily available to bats due to the processing that happens in the bat's cochlea. Therefore, we approximate the effects of the cochlea by taking the magnitude of the short-term Fourier transform of these received signals, as follows:

$$\Psi_L(t, \omega) = \left| \mathcal{F}_{\text{STFT}}(s_L(t)) \right| \quad (13)$$

$$\Psi_R(t, \omega) = \left| \mathcal{F}_{\text{STFT}}(s_R(t)) \right| \quad (14)$$

in which $\mathcal{F}_{\text{STFT}}$ represents the short-term Fourier transform using adequate windowing and overlap values. Subsequently, we concatenate the left and right short-term spectrogram magnitudes into a binaural spectrogram magnitude $\Psi_B(t, \omega)$, as follows:

$$\Psi_B(t, \omega) = \begin{bmatrix} \Psi_L(t, \omega) \\ \Psi_R(t, \omega) \end{bmatrix} \quad (15)$$

This binaural spectrogram magnitude is then dechirped [46] to remove the time-frequency dependence of the call, which is equivalent to a semi-coherent matched filter, as follows (semi-coherent because phase information is not used) [47]:

$$\Psi(t, \omega) = C \left[\Psi_B(t - \Delta t_\omega, \omega) \right] \quad (16)$$

In which the delays Δt_ω are calculated based on the time-frequency distribution of the emitted signal $s_e(t)$ [46,48]. With this, we have arrived at the input data for our SonoNERF: the binaural dechirped magnitude of the short-term Fourier transform $\Psi_D(t, \omega)$. A nonlinear compression function C is applied to map the high dynamic ranges natural to

echolocation signals adequately. In the SonoNERF model, a logarithmic compression with linear rescaling was used as a compression function. The matrix $\Psi(t, \omega)$, together with the pose information of the sensor in the scene, will be the input data for the computations happening inside the SonoNERF model.

3. SonoNERFs

3.1. Neural Acoustic Rendering

The NeRF model proposed in [33] solves the task of novel view synthesis. In this task, the model receives a set of observation images I_t and a set of corresponding sensor poses ζ_t consisting of the Cartesian coordinates X , Y , and Z , and the three Euler angles α , β , and γ . The challenge is synthesizing novel views I_u (or ‘observations’) for previously unseen poses ζ_u . Similarly, the SonoNERF model tries to predict new and unseen dechirped binaural spectrograms $\Psi_u(t, \omega)$ for previously unseen poses ζ_u based on a training ensemble $E_t = [\zeta_t \ \Psi_t]$. This prediction step in NeRFs is solved by implementing a differentiable rendering pipeline that uses an underlying radiance field F_Θ to represent the scene. In many practical applications, this field is represented by a deep multilayer perceptron neural network. Similarly, we will develop a reflectance field F_Θ for our SonoNERF model, which will, in cooperation with the SonoNERF-DRPL, allow the prediction of novel observations Ψ_u .

In the previous section, we laid the foundation of the principles governing signal formation in echolocating bats, which we will now adapt to build the SonoNERF-DRPL. The observation Ψ_u is represented by a 2D matrix of size $[2N_f \times N_t]$, where N_f is the number of frequency components taken from the STFT (times two because of the binaural concatenation), and N_t is the number of time slices in the STFT, which is the result of the choice of window length and overlap in the STFT computation. In order to calculate the spectrum ψ_ω at range r , we propose the following differentiable rendering equation:

$$\psi(\omega, r) = C \left[\left| \int_{\Omega} F_{\Theta}(T_{B,W}[P_{\Omega}], \vec{v}) \cdot E(\psi, \omega) \cdot e^{\frac{j\omega r}{v_s}} \right| \right] \quad (17)$$

This rendering equation calculates the received spectrum ψ_ω using an integration of hemispheres Ω_i . Figure 1, panel a shows two of these hemispheres for a certain pose of the bat/sensor. Inside of the integral, the term $E(\psi, \omega)$ denotes the ERTF of the bat, $e^{\frac{j\omega r}{v_s}}$ is the Fourier transform of the delay function with r_Ω , the range for the current hemisphere that is being integrated, and v_s is the speed of sound in air. The term $F_{\Theta}(T_{B,W}[P_{\Omega}], \vec{v})$ is the neural reflectance field. This function takes as input the position of the voxel for which the reflectivity is to be calculated (P_{Ω} , converted to world coordinates by transform $T_{B \rightarrow W}$), and a direction vector \vec{v} , which indicates the direction from which the voxel is being sonified (which allows non-isotropic reflection surfaces to be modeled with the SonoNERF, similar to the concept of a BRDF in optical rendering [49]). The function $C()$ is a nonlinear compression function (as explained in Equation (16)).

The overall processing flow of this equation can be seen in Figure 1, panel a shows the bat in a single pose, with two hemispheres Ω . The reflectivity function is queried over a discretized version of this hemisphere (typically, 600 points, distributed uniformly over Ω). The spectrum is calculated through the rendering equation, passed through the nonlinear compression function, and placed at the corresponding range slice in the predicted spectrogram (panel b). Panel c shows the ERTF for a bat called *Micronycteris microtus* [50], calculated using the method described in [43]. Figure 1, panel d shows a schematic representation of the Neural Reflectivity field F_{Θ} , which predicts the reflected spectrum $H_p(\omega)$ from the input position and incidence vector, all represented in world coordinates.

3.2. Training of a SonoNERF

In the previous section, we developed the SonoNERF model, in which a rendering equation was developed to calculate the binaural spectrum received from a scene for a specific range, expressed as $\psi(\omega, r)$. In this rendering equation, a neural reflectivity field represents the scene, called F_{Θ} , parameterized by a parameter set Θ . In practice,

this reflectivity function is implemented using a multi-layer perceptron (MLP), and the parameters Θ are the set of weights and biases for this MLP. In this section, we will detail how this MLP is trained.

Figure 2 shows the overall training process. Data, in the form of a binaural and dechirped spectrogram, are recorded from a scene from different poses, as shown in panels a, c, e, and g. The corresponding spectrograms are shown in panels b, d, and f. For a certain pose of the bat and a certain range slice in the spectrogram, we can discretize a hemisphere in front of the bat. This is shown by the blue and red hemispheres (panels a, c, e), which are linked to the corresponding range slices in the spectrogram (panels b, d, f). For each of these hemispheres, we calculate the coordinates of the points on that hemisphere in the world coordinate system, which are then used to query the reflectivity function. This reflectivity function yields a reflectivity value for each of these 3D points and incidence angles, which then is passed through the rendering Equation (17) to yield a spectrum $\psi(\omega)$ for a range slice r . This predicted spectrum yielded by the rendering function is then compared to the measured spectrum in the spectrogram, from which a loss is calculated (depicted by \mathcal{L} in Figure 2). As both the rendering equation and the reflectivity function are fully differentiable, back-propagation can be used to calculate a gradient between the loss \mathcal{L} and the parameters Θ of the reflectivity function, which allows stochastic gradient descent with back-propagation to be used to optimize these parameters Θ . This approach of training our reflectance function is entirely analogous to the training process of the radiance function described in the original implementation of NeRFs in [31], with the main difference being the type of rendering equation that is used. The versatility of this approach shows the brilliance of the original idea of learnable radiance functions by the authors of [31]. The following section will discuss details on the implementation of the reflectivity function and the training setup.

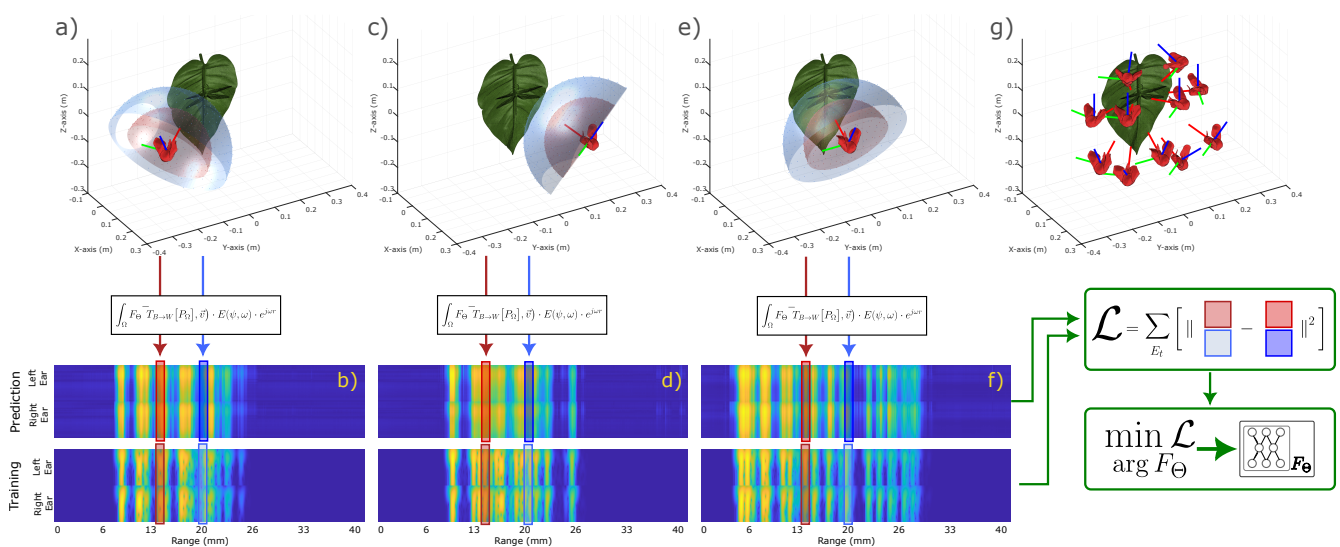


Figure 2. Overview of the SonoNERF Training Process. Data, comprising binaural and dechirped spectrograms, are recorded from various poses within a scene, as depicted in panels (a,c,e,g). The corresponding spectrograms are displayed in panels (b,d,f). Each pose of the bat and range slice in the spectrogram corresponds to a discretized hemisphere in front of the bat, depicted by the blue and red hemispheres in panels (a,c,e), linked to their respective range slices in the spectrogram. Coordinates of points on these hemispheres in the world coordinate system are calculated and used to query the reflectivity function, generating reflectivity values for each 3D point and incidence angle. These values are then processed through rendering Equation (17) to produce a spectrum $\psi(\omega)$ for a given range slice r . The predicted spectrum is compared to the measured spectrum in the spectrogram, yielding a loss \mathcal{L} . This loss \mathcal{L} is minimized using stochastic gradient descent by adapting the learnable parameters in function F_{Θ} . The two hemispheres in blue and red indicate two range slices of the spectrograms for which the acoustic rendering equation is calculated.

4. Experimental Validation

In this section, we will detail the implementation of the SonoNERF setup we developed and show some experimental validation. At this point, we rely purely on simulation to validate the SonoNERF model, as simulation allows us to iterate over ensonification setups rapidly and gives us access to reliable ground truth data. We use SonoTraceLab, a raytracing simulator for simulating acoustic echolocation scenes [51]. The simulator has been extensively validated with real-world experiments, yields reliable data of complex scenes, and allows accurate simulation of biologically relevant acoustic cues. In the remainder of this section, we will provide details on the effective implementation of our SonoNERF model and show its efficacy in multiple experimental scenes. We acknowledge that providing experiments using real-world measured data would be beneficial in illustrating the performance of the SonoNERF approach. However, the authors also believe that frequent scientific communication of relevant bodies of work is crucial to advancing science. Implementing real-world experiments for SonoNERFs is a non-trivial task we will address in future work.

4.1. Sononerf Model Implementation

Inside the acoustic rendering Equation (17), the neural network F_{Θ} represents the neural reflectance field, which encodes the scene being modeled. Figure 3 shows the architecture of this neural network. It consists of a directed acyclic graph and has six input variables: the X, Y, and Z position of the voxel under investigation and the directional vector from where the voxel is being ensonified. The output of the network has 94 values, representing a complex reflectivity function discretized into 47 frequency bins. The first 47 elements of the output of F represent the real components of the reflectivity function, and the last 47 elements represent the imaginary components of the reflectivity function. We lift the six input variables inside the network through a Fourier Embedding layer into a higher-dimensional space ($6 \times 30 = 180$ variables), based on the reasoning provided in [52]. Indeed, embedding low-frequency inputs into a high-frequency representation allows neural networks to learn high-frequency representations better. It should be noted that this step is equivalent to the approach taken in the original NeRF implementation [33].

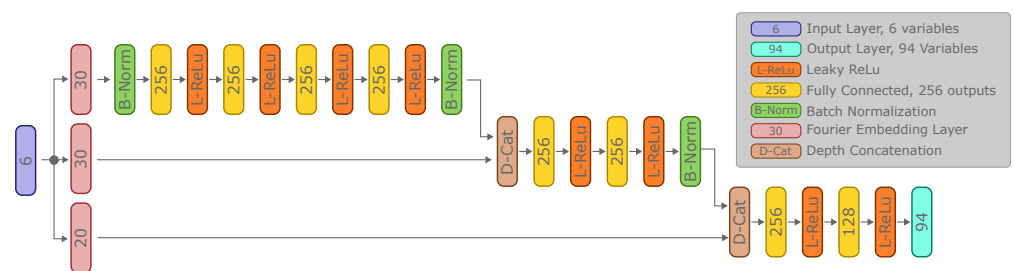


Figure 3. Overview of the SonoNERF reflectivity function F_{Θ} , which encodes a scene’s acoustic properties within a neural reflectance field framework. The network takes six input variables: the X, Y, and Z coordinates of a voxel, along with a directional vector indicating the angle of ensonification. These inputs undergo a Fourier Embedding process, expanding them into a higher-dimensional space (180 variables) to capture high-frequency details more effectively. The network architecture features multiple fully connected layers with Leaky ReLU activation functions and incorporates skip connections for enhanced gradient flow during training. The network output consists of 94 values, which describe a complex reflectivity function across 47 frequency bins, with the first 47 values representing the real components and the last 47 the imaginary components of the reflectivity function.

After the Fourier embedding step, the network consists of several fully connected layers, combined with Leaky ReLU activation functions as non-linearity [53]. Skip connections are added to the network to allow for more efficient gradient flow, allowing faster

learning during training. Skip connections are concatenated through depth concatenation, increasing the size of the inputs of the subsequent layer following the depth concatenation.

We implemented the SonoNERF model, including the rendering equation and the neural reflectance function, in Matlab using object-oriented programming methods and used the Deep Learning toolbox [54] to perform automatic differentiation on the complete rendering equation. This allows rapid iteration over multiple versions of the rendering equation without manually calculating the gradients of the rendering equation. We optimized the performance of the rendering equations by using the GPUArray objects in Matlab, which allow rapid evaluation of matrix and vector operations on the GPU of the system. This allows data to stay in GPU memory during the evaluation of the rendering equation, which yields a significant speed-up compared to a naive CPU implementation (in our tests, speedups of around $100\times$ were achieved).

Training of the SonoNERF models was performed using Adam [55], with a batch size of 512, a learning rate of 0.01, and a learning rate drop factor of 0.97. We trained the networks for 100 epochs using a single NVidia RTX4090 (NVIDIA Corporate, Santa Clara, CA, USA) which took around 7 h to complete. The network described in Figure 3 has around 500,000 learnable parameters, which consists of the weights and biases of the fully connected layers in the reflectivity function.

4.2. From Spectrograms to 3D Scene Description

The SonoNERF model is a method to predict the magnitude of spectrograms that would be observed for previously unseen poses using a model that is trained using measured magnitude spectrograms observed from a set of training poses. Once trained, the SonoNERF model can predict a new spectrogram for an arbitrary pose and nothing more. Indeed, from the point of view of the training algorithm, the concept of a 3D scene description is not made explicit during the training process. However, it is possible to query the trained reflectance function to obtain a 3D voxel description of the environment [33]. To achieve this reconstruction of the 3D scene geometry, we discretize the volume of interest in a cubic voxel with a side of 0.5 mm and use 100 ensonification directions distributed uniformly over the sphere for each voxel. We then query the trained reflectance F_{Θ} for each of these voxel and ensonification directions. The result is a matrix R with the following dimensions:

$$[R(x, y, z, i)] = [n_x \times n_y \times n_z \times n_{\bar{v}}] \quad (18)$$

with n_x , n_y , and n_z , the number of voxels in the X, Y, and Z dimensions, respectively, and $n_{\bar{v}}$ being the number of sampled ensonification directions (100 in our examples). Next, we integrate this reflectivity matrix over the following direction dimension:

$$R_s(x, y, z) = \sum_{i=1}^{n_{\bar{v}}} R(x, y, z, i) \quad (19)$$

Next, using a fixed threshold, we calculate the isosurface over this function $R_s(x, y, z)$. This yields a surface V_r , which can then be visualized. It should be noted that various other visualization methods could be used, such as maximum intensity projection [56] or a plethora of other techniques [57]. However, for the scope of this paper, we opted for applying the isosurface method.

4.3. Typical Data Setup

To train a SonoNERF, we generated 400 ensonifications distributed uniformly over a sphere centered at the target with an appropriate radius depending on the scene size (ranging from 0.3 m to 0.6 m). Each of the ensonifications was simulated using our SonoTraceLab simulator, and the resulting spectrograms were calculated. Each spectrogram consists of around 500 time (i.e., range) samples. So, the total input for training the reflectivity function is around 200,000 spectrum measurements. These data were fully used for training the SonoNERF, as is standard practice in NERF rendering. Then, for evaluation, we

sampled the reflectivity function using 100 points on a sphere as explained in the previous section. In what follows, we will illustrate the performance of SonoNERF on four scenes. During development, several other simpler scenes (consisting of one sphere) were used, as the simulation of these scenes can be performed more rapidly.

4.4. SonoNERF Trained on a Simple Scene

In order to validate the proposed SonoNERF model, we experimented using a simple scene. The model used in this experiment is shown in Figure 4, panel a. This panel shows three spheres with a diameter of 3 cm, arranged in an L-shaped configuration. The figure also shows three poses (which were not part of the training set), ensonifying the scene. We used 100 observations spaced uniformly around the scene and calculated the received spectrograms using our SonoTraceLab simulator [51]. This simulator uses a ray acoustics-based approach and is ideally suited to simulate complex echolocation scenes, and has been validated using real-world experiments. The simulated spectrograms can be seen in the top row of panels b, d, and f. One can see various direct and multi-path reflections occurring later in time, which is expected based on the scene geometry. SonoTraceLab is capable of simulating complex multi-path reflections occurring in challenging scenes. After training, the SonoNERF model is capable of reconstructing the spectrograms for previously unseen poses, shown in panels b, d, and f, bottom row. Both the location as well as spectral content of these reconstructed spectrograms correspond to the ground truth that was simulated. Panel c shows a magnified view of the scene, and panel e shows the isosurface V_r , reconstructed from the queried reflectance function F_{Θ} accumulated into R . Panel g shows a maximum intensity projection of the same reflectivity matrix R . The main shape and location of the individual spheres are reconstructed, whereas the size of the spheres appears to be inflated. This inflation most likely happens due to the absence of phase information in the input data, which diminishes range resolution because only semi-coherent matched filters can be used [30].

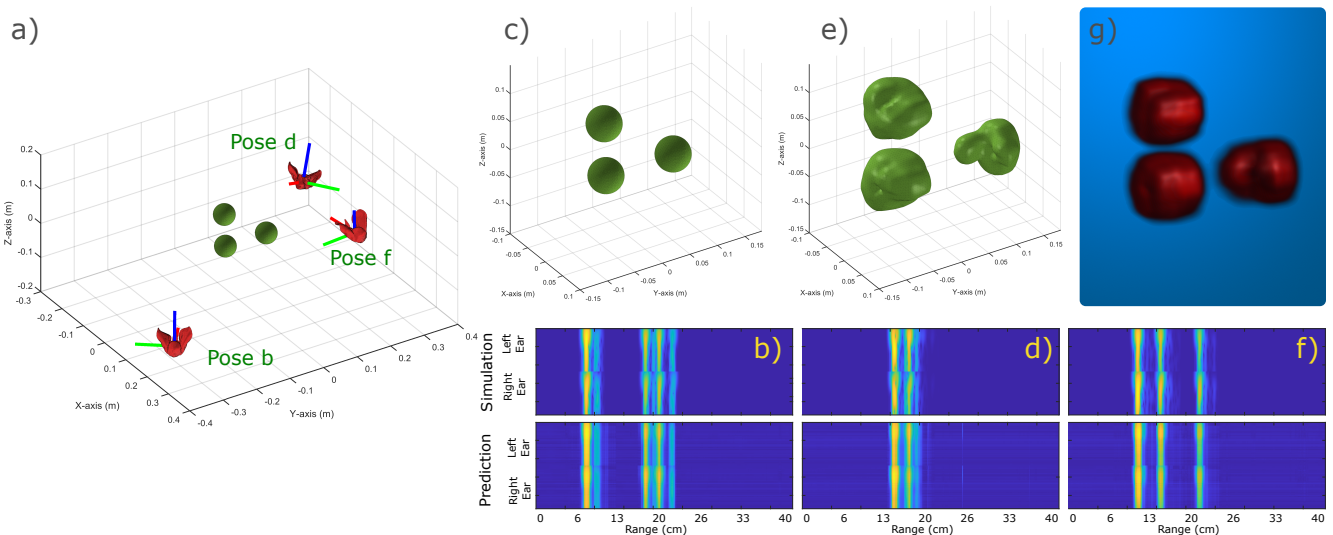


Figure 4. Overview of the result of a trained SonoNERF model on a simple scene. Panel (a) shows the scene, which consists of three spheres in an L-shaped configuration. Three poses (b,d,f) are shown from which the scene is ensonified. The corresponding received spectrograms are shown in panels (b,d,f), top panels (called ‘simulation’). We trained the described SonoNERF model using 100 observations. The resulting predicted spectrograms for poses b, d, and f (not part of the training set) are shown in the top row of panels (b,d,f). Panel (c) shows a more magnified scene plot, and panel (e) shows the reconstructed isosurface V_r . Panel (g) shows a maximum intensity projection of the same reconstruction. It should be noted that the chosen poses b, d, and f are arbitrarily chosen, and are not specifically cherry-picked.

4.5. SonoNERF Trained on a Complex Scene

After our initial validation of the SonoNERF model, we constructed a more complicated scene. This scene can be seen in Figure 5. This scene consists of 19 spheres arranged to form the letters 'UA', corresponding to the name of our institute, University of Antwerp. Similar to the previous subsection, we show the simulated and reconstructed spectrograms. It becomes clear that the time separation between the individual reflections in the spectrograms is less clear, causing the echoes to overlap more. This is also reflected in the reconstructed isosurface, shown in panel e. The individual spheres are no longer separated but form a continuous surface. However, despite these shortcomings, the letters UA are recognizable, with the important features of the letters being conserved (such as the gap between the vertical parts of the U or the hole inside of the A).

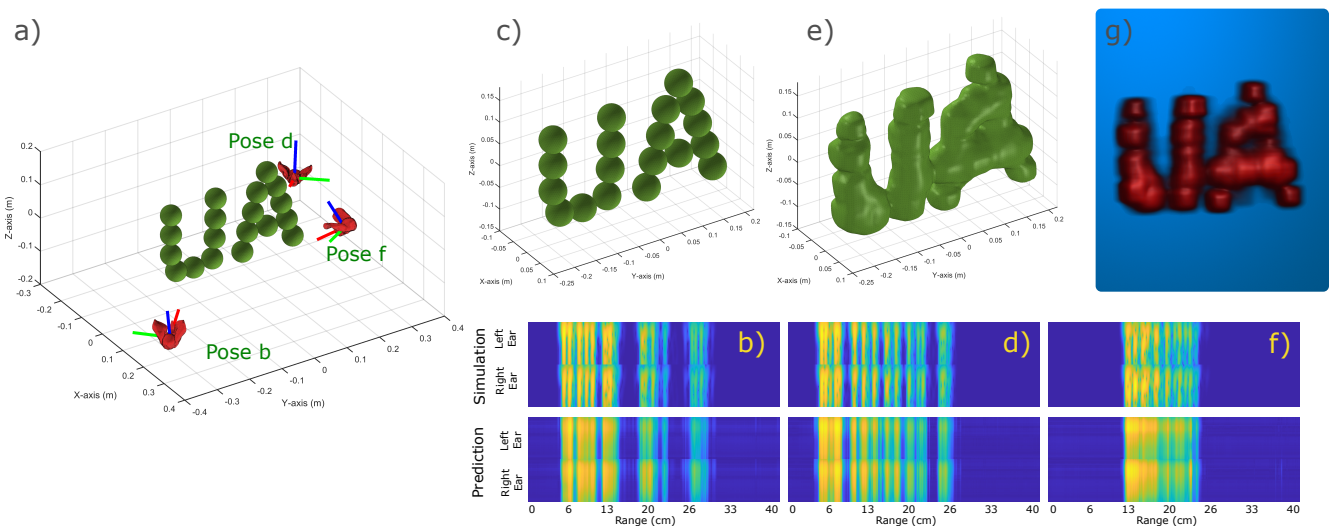


Figure 5. Overview of the result of a trained SonoNERF model on a more complex scene. Panel (a) shows the scene, which consists of 19 spheres arranged to form the letters 'UA', for the University of Antwerp. Three poses (b,d,f) are shown from which the scene is ensounded. The corresponding received spectrograms are shown in panels (b,d,f), top panels (called 'simulation'). We trained the described SonoNERF model using 100 observations. The resulting predicted spectrograms for the poses b, d, and f (not part of the training set) are shown in the top row of panels (b,d,f). Panel (c) shows a more magnified scene plot, and panel (e) shows the reconstructed isosurface V_r . Panel (g) shows a maximum intensity projection of the same reconstruction. It should be noted that the chosen poses b, d, and f are arbitrarily chosen, and are not specifically cherry-picked.

4.6. SonoNERF Trained on a Biologically Relevant Scene

As a final demonstration of the capabilities of our SonoNERF model, we constructed a scene consisting of a leaf on which we perched a dragonfly, which is a scenario that is performed by hunting *Micronycteris microtus* bats in the neotropics [7]. We performed the same simulation and SonoNERF training as in the two previous scenarios, and show the results in Figure 6. The SonoNERF is capable of reconstructing the spectrogram representations (panels b, d, f), as well as the overall leaf geometry (panels e and g). To further investigate this scenario, we modeled the leaf with and without a dragonfly present on the leaf surface. The result of this reconstruction can be seen in Figure 7. Panel a shows the leaf without a dragonfly, and panel c shows the leaf with a dragonfly. Panels b and d show the isosurface reconstruction of these two scenarios. Here, a clear bulge can be seen in the reconstruction of the leaf with a dragonfly (panel d), which is absent in the reconstruction of the leaf without a dragonfly (panel b). Panel e shows the maximum intensity projection of energy matrix R_s (Equation (19)) into the direction of the camera viewpoint, which has been normalized to the maximum of both R_s for the two reconstructions. It becomes apparent that the overall reflection strength of the leaf with a dragonfly is much larger compared to

the leaf without a dragonfly. The top-view of this representation (panel f) further details this. Finally, panel g shows the difference in energy between the reflected energy matrix of the leaf with and without a dragonfly. Here, a strong energy peak can be seen in the middle of the leaf surface. To further illustrate this energy difference, we plotted the energy difference using maximum intensity projection on top of the STL model of the leaf with a dragonfly. This can be seen in Figure 8, panel c. In this overlay, a strong peak can be observed around the location of the dragonfly on the leaf.

These results illustrate a potential biological implication of our SonoNERF model. As shown, the SonoNERF reconstruction allows for the reconstruction of biologically relevant information from the scene: it allows the bat to localize the dragonfly on the leaf, and to distinguish between a filled and an empty leaf, and whereas the model proposed in [7] for leaf occupancy state discrimination has not lost its validity, the model therein proposed does not explain how the bat might be able to localize the prey item on the leaf, and whereas we make no claims on which model is implemented in the bat's brain, we do think that with our proposed SonoNERF model, we have shown that prey localization against background clutter is enabled by such a data-aggregation task. More specifically, through learning a prediction task of novel spectrograms from unseen poses, the surface reconstruction, and therefore the prey localization, emerges as a byproduct of the computational graph that is used to solve the prediction task.

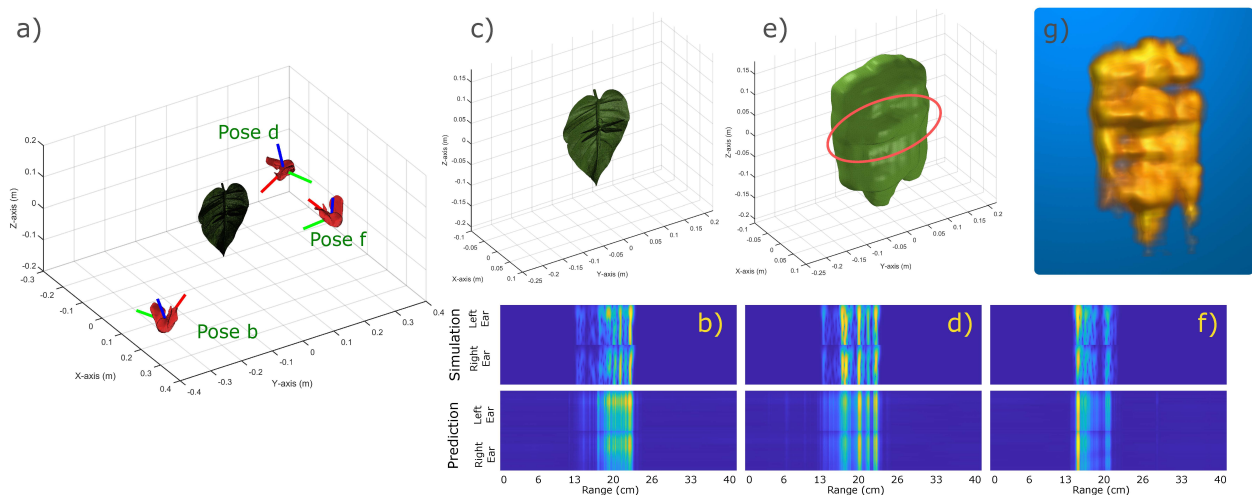


Figure 6. Overview of the result of a trained SonoNERF model on a biologically relevant scene of an insect perched on a leaf. Panel (a) shows the scene, a leaf with a dragonfly, similar to the approaches of *Micronycteris microtus* like in [7]. Three poses (b,d,f) are shown from which the scene is ensounded. The corresponding received spectrograms are shown in panels (b,d,f), top panels (called 'simulation'). We trained the described SonoNERF model using 100 observations. The resulting predicted spectrograms for the poses b, d, and f (not part of the training set) are shown in the top row of panels (b,d,f). Panel (c) shows a more magnified scene plot, and panel (e) shows the reconstructed isosurface V_r . Panel (g) shows a maximum intensity projection of the same reconstruction. Panel (e) shows a thickening in the volume mesh on the location of the dragonfly, hinting for detailed scene information to be present in the reconstructed mesh. It should be noted that the chosen poses b, d and f are arbitrarily chosen, and are not specifically cherry-picked.

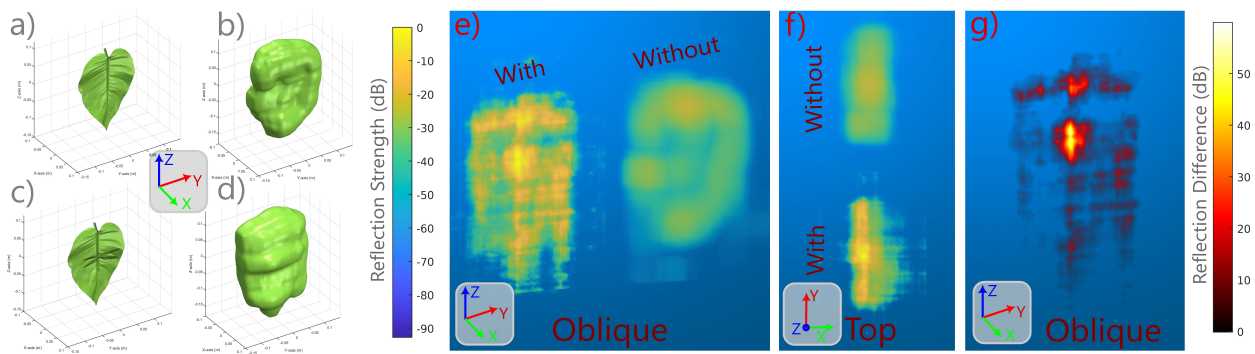


Figure 7. Overview of the SonoNERF volume reconstructions for a leaf without an insect (panels (a,b)) and for a leaf with an insect (panels (c,d)). A significant bulge in the mesh surface can be observed on the location of the dragonfly in panel (d). Panel (e) shows the reflectivity function through maximum intensity projection (MIP) into the camera, normalized to the strongest reflection across the two instances. Panel (f) shows the top view of the same MIP visualization. Finally, panel (g) shows the difference in reflectivity function for a leaf with and without a dragonfly.

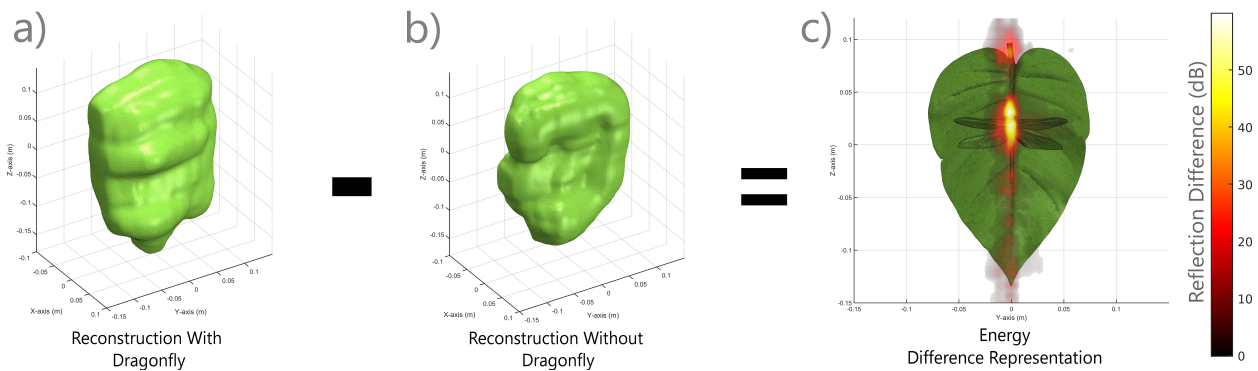


Figure 8. Detail of the difference in reflectivity function between a leaf with a dragonfly and without a dragonfly. The largest differences can be observed around the location of the dragonfly where the difference function peaks strongly. This could be used as a cue for prey localization on the leaf.

5. Data and Code Availability

We provide the full source code for our SonoNERF approach, which can be used to perform a full simulation, training, and reconstruction. The data and source code can be found on our public Github page: <https://github.com/Cosys-Lab/SonoNERF> (accessed on 24 May 2024).

6. Discussion and Conclusions

In this paper, we proposed SonoNERFs, a novel model for 3D scene reconstruction in a biologically plausible manner. We use the concept of Neural Radiance Fields to solve the problem of predicting echo spectrograms that would be obtained from a scene when ensonified from previously unseen poses, without access to the phase information of the received echoes. As explained, phase coherence and phase reception are unlikely to exist in echolocating bats because of how vocalization and reception (most notably the cochlea) behave in real animals. After we provided a solution for the spectrogram prediction problem, we detailed how the learned reflectivity model can be used to perform 3D scene reconstruction of complex shapes, which we demonstrated using three scenes of varying complexity.

One could argue that the fact that 3D scene reconstruction becomes possible when combining measurements from different poses is not that surprising. Indeed, computed tomography techniques exist and are already applied to echolocation scenarios. However, these systems utilize phase-coherent measurements and need the phase information of

the echoes to work well. What, in the opinion of the authors, is not so trivial is the fact that 3D scene reconstruction emerges when solving a completely different task, namely predicting sensor data for a novel scene, and whereas it has been observed that bats can predict aspects of the scene by accumulating sensor data [58], to the best of our knowledge, no concrete model on how this prediction might operate has been proposed in previous works. Other technical systems have been proposed to produce 3D scene reconstruction and semantic interpretation [59,60], but these proposed techniques utilize a teaching modality like LIDARs or cameras to perform a form of modality translation. Our SonoNERF model relies solely on acoustic data without the need for an additional supervision modality. Furthermore, reference [60] does not use an acoustic sensing modality, causing the title to be somewhat misleading. Our approach follows the approach called ‘self-supervised learning’, which has received much attention in recent years [61,62]. Through self-supervision, computational graphs can learn how to predict their own inputs, and, based on the structure of the implemented computational graph, can lead to emergent insights into the underlying sensor data (such as demonstrated in our SonoNERF model).

In addition to the capability of SonoNERF to predict spectrograms and to perform surface reconstruction, we also showed how the SonoNERF model could be used to solve a relevant task for a biological echolocator, namely, finding motionless prey items against strong clutter backgrounds, and whereas we postulated a model in [7] on how the bat *Micronycteris microtus* might be able to distinguish between an empty or a full leaf, no concrete postulate was made on how localization might take place. In the previous section, we demonstrated how a SonoNERF model might be used to perform exactly this task, through the examination of the reconstructed reflectivity function. It should be noted that whereas our proposed SonoNERF model might be one solution to how scene reconstruction takes place, we nowhere claim that this is the exact model that is implemented in the brain of a bat. More specifically, what we do claim is that the SonoNERF model is a first-order hypothesis to how bats might be able to solve this problem, similarly to how we proposed our BatSLAM algorithm as a first-order model on how bats might perform large-scale localization [63]. In-depth biological behavior experiments need to be performed using testable hypotheses generated through our SonoNERF model to gain insight into the true underlying behavioral mechanisms.

We acknowledge the lack of real-world measurements in this paper. However, as we argued before, using our validated SonoTraceLab simulator is a strong alternative approach to producing experimental data, as this simulator has been demonstrated to be able to produce biologically relevant acoustic data from complex scenes without the arising complexities of performing real-world data. We also acknowledge that real-world experiments are relevant, which is why these experiments are part of our short-term future work. Using SonoTraceLab to provide us with data allows us to rapidly iterate over experiments and allows complex visualizations such as the one in Figure 8, where we overlay the reconstructed reflectivity function over the base model, which then allows the discovery of important features such as the difference in a leaf with and without an insect.

Next to performing real-world experimentation, several improvements can be proposed to the proposed model. For example, at this point, our SonoNERF model is trained from a randomly initialized reflectivity model each time. One could argue that, over time, bats learn priors on how scenes might be encoded, which could then be reflected into prior models or transfer-learned models later. Furthermore, the acoustic rendering Equation (17) has no concept of multi-path signal propagation. One natural extension would be to expand this rendering equation to encompass multi-path reflections. However, this augmented rendering equation should still be fully differentiable; it is currently unknown to the authors whether this will be the case. In addition to the proposed improvements, we will perform more quantitative experiments on the accuracy of the reconstructed 3D scenes. From the experiments shown in this paper, it is already indicative that the size of the reconstruction is larger than the original scene (i.e., some kind of ‘thickening’ effect). This can be explained by the fact that phase information is lost, causing a delay in the peak of the reflections in the

spectrograms. Possible solutions here can be developed to reduce these delays, which then in turn would reduce the thickening effect. However, the core of the paper is introducing the SonoNERF concept, relating that to bat echolocation, which is why we did not perform extensive quantitative evaluations at this stage. We acknowledge that we simulated scenes (albeit complex scenes) in empty space, i.e., no surrounding clutter was introduced, which of course is not realistic in real-world echolocation scenes. In future work, we will assess the performance of SonoNERF in more complex scenes with background clutter, as this is perfectly simulatable by our SonoTraceLab simulator.

Author Contributions: Conceptualization, J.S.; methodology, J.S.; software, W.J., J.S.; validation, J.S., investigation, J.S.; resources, W.J. and J.S.; writing—original draft preparation, J.S.; writing—review and editing, J.S. and W.J.; visualization, J.S.; supervision, J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: We provide the full source code for our SonoNERF approach, which can be used to perform a full simulation, training, and reconstruction. The data and source code can be found on our public Github page: <https://github.com/Cosys-Lab/SonoNERF> (accessed on 24 May 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of this study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Bell, G.P. Behavioral and Ecological Aspects of Gleaning by a Desert Insectivorous Bat *Antrozous Pallidus* (Chiroptera: Vespertilionidae). *Behav. Ecol. Sociobiol.* **1982**, *10*, 217–223. [[CrossRef](#)]
- Entwistle, A.C.; Racey, P.A.; Speakman, J.R. Habitat Exploitation by a Gleaning Bat, *Plecotus Auritus*. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **1996**, *351*, 921–931.
- Geipel, I.; Jung, K.; Kalko, E.K. Perception of Silent and Motionless Prey on Vegetation by Echolocation in the Gleaning Bat *Micronycteris Microtis*. *Proc. R. Soc. B Biol. Sci.* **2013**, *280*, 20122830.
- Razak, K.A. Adaptations for Substrate Gleaning in Bats: The Pallid Bat as a Case Study. *Brain Behav. Evol.* **2018**, *91*, 97–108. [[CrossRef](#)] [[PubMed](#)]
- Stoffberg, S.; Jacobs, D.S. The Influence of Wing Morphology and Echolocation on the Gleaning Ability of the Insectivorous Bat *Myotis Tricolor*. *Can. J. Zool.* **2004**, *82*, 1854–1863. [[CrossRef](#)]
- Swift, S.; Racey, P. Gleaning as a Foraging Strategy in Natterer's Bat *Myotis Nattereri*. *Behav. Ecol. Sociobiol.* **2002**, *52*, 408–416.
- Geipel, I.; Steckel, J.; Tschapka, M.; Vanderelst, D.; Schnitzler, H.U.; Kalko, E.K.; Peremans, H.; Simon, R. Bats Actively Use Leaves as Specular Reflectors to Detect Acoustically Camouflaged Prey. *Curr. Biol.* **2019**, *29*, 2731–2736. [[CrossRef](#)] [[PubMed](#)]
- Verreycken, E.; Simon, R.; Quirk-Royal, B.; Daems, W.; Barber, J.; Steckel, J. Bio-Acoustic Tracking and Localization Using Heterogeneous, Scalable Microphone Arrays. *Commun. Biol.* **2021**, *4*, 11. [[CrossRef](#)] [[PubMed](#)]
- Arlettaz, R.; Jones, G.; Racey, P.A. Effect of Acoustic Clutter on Prey Detection by Bats. *Nature* **2001**, *414*, 742–745. [[CrossRef](#)]
- Siemers, B.M.; Baur, E.; Schnitzler, H.U. Acoustic Mirror Effect Increases Prey Detection Distance in Trawling Bats. *Naturwissenschaften* **2005**, *92*, 272–276. [[CrossRef](#)]
- Zsebok, S.; Kroll, F.; Heinrich, M.; Genzel, D.; Siemers, B.M.; Wiegrebe, L. Trawling Bats Exploit an Echo-Acoustic Ground Effect. *Front. Physiol.* **2013**, *4*, 65. [[CrossRef](#)] [[PubMed](#)]
- Grafe, T.U.; Schöner, C.R.; Kerth, G.; Junaidi, A.; Schöner, M.G. A Novel Resource–Service Mutualism between Bats and Pitcher Plants. *Biol. Lett.* **2011**, *7*, 436–439. [[CrossRef](#)] [[PubMed](#)]
- Schöner, M.G.; Schöner, C.R.; Simon, R.; Grafe, T.U.; Puechmaille, S.J.; Ji, L.L.; Kerth, G. Bats Are Acoustically Attracted to Mutualistic Carnivorous Plants. *Curr. Biol.* **2015**, *25*, 1911–1916. [[CrossRef](#)] [[PubMed](#)]
- Simon, R.; Bakunowski, K.; Reyes-Vasques, A.E.; Tschapka, M.; Knoernschild, M.; Steckel, J.; Stowell, D. Acoustic Traits of Bat-Pollinated Flowers Compared to Flowers of Other Pollination Syndromes and Their Echo-Based Classification Using Convolutional Neural Networks. *PLoS Comput. Biol.* **2021**, *17*, 20. [[CrossRef](#)] [[PubMed](#)]
- Simon, R.; Matt, F.; Santillan, V.; Tschapka, M.; Tuttle, M.; Halfwerk, W. An Ultrasound-Absorbing Inflorescence Zone Enhances Echo-Acoustic Contrast of Bat-Pollinated Cactus Flowers. *J. Exp. Biol.* **2023**, *226*, jeb245263. [[CrossRef](#)] [[PubMed](#)]
- Simon, R.; Rupitsch, S.; Baumann, M.; Wu, H.; Peremans, H.; Steckel, J. Bioinspired Sonar Reflectors as Guiding Beacons for Autonomous Navigation. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1367–1374. [[CrossRef](#)] [[PubMed](#)]

17. de Backer, M.; Jansen, W.; Laurijssen, D.; Simon, R.; Daems, W.; Steckel, J. Detecting and Classifying Bio-Inspired Artificial Landmarks Using in-Air 3D Sonar. In Proceedings of the 2023 IEEE SENSORS, Vienna, Austria, 29 October–1 November 2023; pp. 1–4. [[CrossRef](#)]
18. Denny, M. The Physics of Bat Echolocation: Signal Processing Techniques. *Am. J. Phys.* **2004**, *72*, 1465–1477. [[CrossRef](#)]
19. Altes, R.A. Sonar for Generalized Target Description and Its Similarity to Animal Echolocation Systems. *J. Acoust. Soc. Am.* **1976**, *59*, 97–105. [[CrossRef](#)] [[PubMed](#)]
20. Saillant, P.A.; Simmons, J.A.; Dear, S.P.; McMullen, T.A. A Computational Model of Echo Processing and Acoustic Imaging in Frequency-modulated Echolocating Bats: The Spectrogram Correlation and Transformation Receiver. *J. Acoust. Soc. Am.* **1993**, *94*, 2691–2712. [[CrossRef](#)]
21. Simmons, J.A.; Stein, R.A. Acoustic Imaging in Bat Sonar: Echolocation Signals and the Evolution of Echolocation. *J. Comp. Physiol.* **1980**, *135*, 61–84. [[CrossRef](#)]
22. Simmons, J.A.; Moss, C.F.; Ferragamo, M. Convergence of Temporal and Spectral Information into Acoustic Images of Complex Sonar Targets Perceived by the Echolocating Bat, *Eptesicus Fuscus*. *J. Comp. Physiol. A* **1990**, *166*, 449–470. [[CrossRef](#)]
23. Simmons, J.A. A View of the World through the Bat’s Ear: The Formation of Acoustic Images in Echolocation. *Cognition* **1989**, *33*, 155–199. [[CrossRef](#)] [[PubMed](#)]
24. Balleri, A.; Griffiths, H.D.; Woodbridge, K.; Baker, C.J.; Holderied, M.W. Bat-Inspired Ultrasound Tomography in Air. In Proceedings of the 2010 IEEE Radar Conference, Arlington, VA, USA, 10–14 May 2010; pp. 44–47.
25. Clare, E.L.; Holderied, M.W. Acoustic Shadows Help Gleaning Bats Find Prey, but May Be Defeated by Prey Acoustic Camouflage on Rough Surfaces. *Elife* **2015**, *4*, e07404. [[CrossRef](#)]
26. Neil, T.R.; Shen, Z.; Robert, D.; Drinkwater, B.W.; Holderied, M.W. Moth Wings Are Acoustic Metamaterials. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 31134–31141. [[CrossRef](#)] [[PubMed](#)]
27. Chitradurga Achutha, A.; Peremans, H.; Firzlaff, U.; Vanderelst, D. Efficient Encoding of Spectrotemporal Information for Bat Echolocation. *PLoS Comput. Biol.* **2021**, *17*, e1009052. [[CrossRef](#)]
28. Kim, S.Y.; Allen, R.; Rowan, D. The Simulation of Bat-Oriented Auditory Processing Using the Experimental Data of Echolocating Signals. *J. Acoust. Soc. Am.* **2008**, *123*, 3621. [[CrossRef](#)]
29. Kössl, M.; Vater, M. The Cochlear Frequency Map of the Mustache Bat, *Pteronotus Parnellii*. *J. Comp. Physiol. A* **1985**, *157*, 687–697. [[CrossRef](#)]
30. Peremans, H.; Hallam, J. The Spectrogram Correlation and Transformation Receiver, Revisited. *J. Acoust. Soc. Am.* **1998**, *104*, 1101–1110. [[CrossRef](#)]
31. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing Scenes as Neural Radiance Fields for View Synthesis. *Commun. ACM* **2021**, *65*, 99–106. [[CrossRef](#)]
32. Zhang, K.; Riegler, G.; Snively, N.; Koltun, V. Nerf++: Analyzing and Improving Neural Radiance Fields. *arXiv* **2020**, arXiv:2010.07492.
33. Gao, K.; Gao, Y.; He, H.; Lu, D.; Xu, L.; Li, J. Nerf: Neural Radiance Field in 3d Vision, a Comprehensive Review. *arXiv* **2022**, arXiv:2210.00379.
34. Zhu, F.; Guo, S.; Song, L.; Xu, K.; Hu, J. Deep Review and Analysis of Recent Nerfs. *APSIPA Trans. Signal Inf. Process.* **2023**, *12*, e6. [[CrossRef](#)]
35. Iddrisu, K.; Malec, S.; Crimi, A. 3D Reconstructions of Brain from MRI Scans Using Neural Radiance Fields. In Proceedings of the International Conference on Artificial Intelligence and Soft Computing, Zakopane, Poland, 18–22 June 2023; Springer: Cham, Switzerland, 2023; pp. 207–218.
36. Jang, T.J.; Hyun, C.M. NeRF Solves Undersampled MRI Reconstruction. *arXiv* **2024**, arXiv:2402.13226.
37. Wysocki, M.; Azampour, M.F.; Eilers, C.; Busam, B.; Salehi, M.; Navab, N. Ultra-Nerf: Neural Radiance Fields for Ultrasound Imaging. In Proceedings of the Medical Imaging with Deep Learning, PMLR, Nashville, TN, USA, 25 January 2023; pp. 382–401.
38. Zou, Y.; Lin, Y.; Zhu, Q. PA-NeRF, a Neural Radiance Field Model for 3D Photoacoustic Tomography Reconstruction from Limited Bscan Data. *Biomed. Opt. Express* **2024**, *15*, 1651–1667. [[CrossRef](#)]
39. Chen, C.; Richard, A.; Shapovalov, R.; Ithapu, V.K.; Neverova, N.; Grauman, K.; Vedaldi, A. Novel-View Acoustic Synthesis. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 6409–6419.
40. Chen, Z.; Gebru, I.D.; Richardt, C.; Kumar, A.; Laney, W.; Owens, A.; Richard, A. Real Acoustic Fields: An Audio-Visual Room Acoustics Dataset and Benchmark. *arXiv* **2024**, arXiv:2403.18821.
41. Guo, Y.; Chen, K.; Liang, S.; Liu, Y.J.; Bao, H.; Zhang, J. Ad-Nerf: Audio Driven Neural Radiance Fields for Talking Head Synthesis. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 5784–5794.
42. Luo, A.; Du, Y.; Tarr, M.; Tenenbaum, J.; Torralba, A.; Gan, C. Learning Neural Acoustic Fields. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 3165–3177.
43. De Mey, F.; Reijniers, J.; Peremans, H.; Otani, M.; Firzlaff, U. Simulated Head Related Transfer Function of the Phyllostomid Bat *Phyllostomus Discolor*. *J. Acoust. Soc. Am.* **2008**, *124*, 2123–2132. [[CrossRef](#)] [[PubMed](#)]
44. Jones, G.; Holderied, M.W. Bat Echolocation Calls: Adaptation and Convergent Evolution. *Proc. R. Soc. B Biol. Sci.* **2007**, *274*, 905–912. [[CrossRef](#)] [[PubMed](#)]

45. Pierce, A.D. *Acoustics: An Introduction to Its Physical Principles and Applications*; Springer: Berlin/Heidelberg, Germany, 2019.
46. Wang, J.; Cai, D.; Wen, Y. Comparison of Matched Filter and Dechirp Processing Used in Linear Frequency Modulation. In Proceedings of the 2011 IEEE 2nd International Conference on Computing, Control and Industrial Engineering, Wuhan, China, 20–21 August 2011; Volume 2, pp. 70–73.
47. Wiegrebe, L. An Autocorrelation Model of Bat Sonar. *Biol. Cybern.* **2008**, *98*, 587–595. [[CrossRef](#)]
48. Steckel, J.; Vanderelst, D.; Peremans, H. BatSLAM: Combining Biomimetic Sonar with a Hippocampal Model. In Proceedings of the Robotica Conference, Guimaraes, Portugal, 11 April 2012.
49. Matusik, W.; Pfister, H.; Brand, M.; McMillan, L. Efficient Isotropic BRDF Measurement. In Proceedings of the 14th Eurographics Workshop on Rendering, Leuven, Belgium, 25–27 June 2003; ACM International Conference Proceeding Series Volume 44.
50. Vanderelst, D.; De Mey, F.; Peremans, H.; Geipel, I.; Kalko, E.; Firzlauff, U. What Noseleaves Do for FM Bats Depends on Their Degree of Sensorial Specialization. *PLoS ONE* **2010**, *5*, e11893. [[CrossRef](#)]
51. Jansen, W.; Steckel, J. SonoTraceLab-A Raytracing-Based Acoustic Modelling System for Simulating Echolocation Behavior of Bats. *arXiv* **2024**, arXiv:2403.06847.
52. Tancik, M.; Srinivasan, P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.; Ng, R. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 7537–7547.
53. Banerjee, C.; Mukherjee, T.; Pasiliao, E., Jr. An Empirical Study on Generalizations of the ReLU Activation Function. In Proceedings of the Proceedings of the 2019 ACM Southeast Conference, Kennesaw, GA, USA, 18–20 April 2019; pp. 164–167.
54. Deep Learning Toolbox. Available online: <https://nl.mathworks.com/products/deep-learning.html> (accessed on 24 May 2024).
55. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
56. Napel, S.; Marks, M.P.; Rubin, G.D.; Dake, M.D.; McDonnell, C.H.; Song, S.M.; Enzmann, D.R.; Jeffrey, R.B., Jr. CT Angiography with Spiral CT and Maximum Intensity Projection. *Radiology* **1992**, *185*, 607–610. [[CrossRef](#)]
57. Kaufman, A.E.; Mueller, K. Overview of Volume Rendering. *Vis. Handb.* **2005**, *7*, 127–174.
58. Salles, A.; Diebold, C.A.; Moss, C.F. Echolocating Bats Accumulate Information from Acoustic Snapshots to Predict Auditory Object Motion. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 29229–29238. [[CrossRef](#)]
59. Christensen, J.H.; Hornauer, S.; Stella, X.Y. Batvision: Learning to See 3d Spatial Layout with Two Ears. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 1581–1587.
60. Zhang, C.; Yang, Z.; Xue, B.; Zhuo, H.; Liao, L.; Yang, X.; Zhu, Z. Perceiving like a Bat: Hierarchical 3D Geometric–Semantic Scene Understanding Inspired by a Biomimetic Mechanism. *Biomimetics* **2023**, *8*, 436. [[CrossRef](#)]
61. Hendrycks, D.; Mazeika, M.; Kadavath, S.; Song, D. Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
62. Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; Tang, J. Self-Supervised Learning: Generative or Contrastive. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 857–876. [[CrossRef](#)]
63. Steckel, J.; Peremans, H. BatSLAM: Simultaneous Localization and Mapping Using Biomimetic Sonar. *PLoS ONE* **2013**, *8*, e54076. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.