JASA ARTICLE

Wave equation-informed generative networks for the design of speech recognition acoustic materials

Huilan Wu,[1] Yijun Liu,[1,a)] and Han Zhang[2,b)]
[1]*Department of Mechanics and Aerospace Engineering, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China*
[2]*Key Laboratory of Noise and Vibration, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China*

ABSTRACT:
The inverse problem of designing materials to achieve desired acoustic functionality while strictly adhering to acoustic principles remains an unresolved scientific challenge. This work introduces a generative adversarial network based on the wave equation (Wave-GAN), in which the governing equation is directly embedded into the training process to iteratively optimize the material density distribution. The physics-guided framework enables the model to learn acoustic patterns directly from sound and generate material distributions with specified functionalities. As a verification example, a speaker recognition task was conducted. The results demonstrate that Wave-GAN produces physically consistent materials, achieving a recognition accuracy of 95.6%. This opens a promising direction for fully physics-driven material design at the interface of acoustics and machine learning.
 https://doi.org/10.1121/10.0043580
(Received 10 October 2025; revised 19 March 2026; accepted 25 March 2026; published online 15 May 2026)
[Editor: Yun Jing] Pages: 4223–4237

I. INTRODUCTION

Traditional functional material design mainly relies on expensive, time-consuming, and labor-intensive trial-and-error methods. This "Edison-style" experimental process not only heavily depends on researchers' intuition and domain knowledge but also involves a considerable degree of chance.[1–3] Therefore, rational design of functional materials based on physical principles is the ultimate goal of modern engineering applications.[4–7]

In recent years, with the rapid development of artificial intelligence and computing power, machine learning has made significant progress in the field of material design.[8–11] By learning the relationship between material structure and performance from experimental or simulation data, data-driven methods can accelerate the discovery of new materials.[12–14] However, although pure deep learning models (e.g., generative adversarial networks) have strong generative capabilities, their "black-box" nature means the generated results lack physical guarantees, potentially producing structures that cannot be practically fabricated or have unstable acoustic performance, such as deviations in target frequency responses, sound absorption coefficients, or bandgap characteristics.[15]

Some studies have attempted to combine machine learning with acoustic simulations, such as using wave equation solvers or physics-informed neural networks (PINNs) for sound field prediction[16] and reconstruction.[17] Optimization methods and machine learning surrogate models have also been used for the inverse design of acoustic metamaterials.[18,19] In these applications, the primary focus has predominantly been on selecting appropriate machine learning algorithms or material descriptors to map the input space to the output space. Once trained, the machine learning model can serve as a screening tool to perform an enumerative search for promising candidate materials from a predefined design space based on specific criteria. Although predicting the properties of new materials using machine learning models is far more efficient than laboratory synthesis and measurement, exhaustive enumeration of all possible combinations within the vast design space remains computationally intractable, even with high-performance computing. A more fundamental bottleneck lies in the current lack of a robust method for designing materials with specific acoustic functionalities. This method would need to operate without relying on large-scale data while strictly adhering to acoustic principles. Consequently, effectively addressing this inverse design problem remains a significant challenge.[20–23]

In this work, we propose a wave equation-based generative adversarial network (Wave-GAN), aimed at establishing an inverse material design framework that strictly adheres to acoustic wave principles. Wave-GAN directly embeds the wave equation into the generator during the training process, ensuring that the generated materials not only conform to acoustic principles but also accurately achieve the target performance of candidate designs. As a practical demonstration, two acoustic recognition material design problems are selected as case studies to illustrate the proposed inverse design method. We compare Wave-GAN with other common neural network methods and demonstrate its advantages over alternative optimization strategies. Beyond merely functioning as a black-box neural network, the model explicitly aims to align the resonant frequencies of the material with the dominant spectral

a)Electronic mail: liuyj3@sustech.edu.cn
b)Email: zhanghan@mail.ioa.ac.cn

components of the input speech signal. This study establishes a new paradigm for generative material design based on wave physics, providing a novel pathway for physics-driven material discovery and the design of intelligent acoustic systems. This work addresses the inverse problem of designing materials that strictly comply with acoustic principles to achieve desired acoustic effects, rather than merely predicting the outcomes of predefined materials.

II. DESIGN METHOD OF WAVE-GAN GENERATIVE NETWORK

A. Generative model based on wave equations

The wave equation is the foundation of sound wave propagation and the basis for all sound field analysis and calculation. Consider a vector wave field distribution, where the dynamic behavior of sound waves can be described as follows:

$$\frac{\partial^2 \boldsymbol{u}}{\partial t^2} - c(\rho, \boldsymbol{r})^2 \nabla^2 \boldsymbol{u} = f, \tag{1}$$

where $\boldsymbol{u}$ denotes the acoustic pressure field, $c(\rho, \boldsymbol{r})$ is the speed of sound determined by the density of the medium, $\boldsymbol{r} = (x, y)$ represents the spatial position vector, and $\boldsymbol{f}$ is the external acoustic excitation.

When a speech signal $s(t)$ acts as an external excitation to the system, the sound field distribution at time $t+1$ depends on the previous t steps of the sound field state and the excitation. As shown in Fig. 1, the Wave-GAN processes the acoustic signal split into time slices as excitation at each time step. The sound field states from the previous t steps are stored in the Wave-GAN's memory and updated at each step. This allows the Wave-GAN to retain past information and learn temporal patterns and long-term dependencies in the data. The wave equation is discretized with time step Δt as follows:

$$\frac{\boldsymbol{u}_{t+1} - 2\boldsymbol{u}_t + \boldsymbol{u}_{t-1}}{\Delta t^2} - c(\rho, \boldsymbol{r})^2 \cdot \nabla^2 \boldsymbol{u}_t = \boldsymbol{f}_t, \tag{2}$$

where the subscripts denote vectors at a given time step, we express the wave Eq. (3) in matrix form as follows:

$$\begin{bmatrix} \boldsymbol{u}_{t+1} \\ \boldsymbol{u}_t \end{bmatrix} = \begin{bmatrix} 2 + \Delta t^2 \cdot c(\rho, \boldsymbol{r})^2 \cdot \nabla^2 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \Delta t^2 \cdot \begin{bmatrix} \boldsymbol{f}_t \\ 0 \end{bmatrix}. \tag{3}$$

The hidden state of the wave system is defined as the concatenation of the field distributions at the current and immediately previous time step, $\boldsymbol{h}_t \equiv [\boldsymbol{u}_t, \boldsymbol{u}_{t-1}]^T$, where $\boldsymbol{u}_t$ and $\boldsymbol{u}_{t-1}$ are the acoustic field vectors, represented on a discrete grid in the spatial domain. The update of the wave equation can then be written as follows:

$$\begin{aligned} &\boldsymbol{h}_{t+1} = \boldsymbol{w} \cdot \boldsymbol{h}_t + \sigma \cdot \boldsymbol{x}_t, \\ &\boldsymbol{h}_{t+1} \equiv [\boldsymbol{u}_{t+1}, \boldsymbol{u}_t]^T, \\ &\boldsymbol{w} = \begin{bmatrix} 2 + \Delta t^2 \cdot c(\rho, \boldsymbol{r})^2 \cdot \nabla^2 & -1 \\ 1 & 0 \end{bmatrix}, \\ &\boldsymbol{x}_t = \boldsymbol{f}_t, \quad \text{and} \quad \sigma = \Delta t^2 \end{aligned} \tag{4}$$

In this generator, the wave-equation-driven network takes the input speech signal $s(t)$ as the excitation source $\boldsymbol{f}_t$, iteratively learns the density distribution $c(\rho, \boldsymbol{r})$, and subsequently generates the acoustic field $\boldsymbol{u}_{t+1}$ corresponding to different excitation disturbances. This part corresponds to the generator component in Fig. 1. The generated acoustic field $\boldsymbol{u}_{t+1}$ is subsequently evaluated by the discriminator and

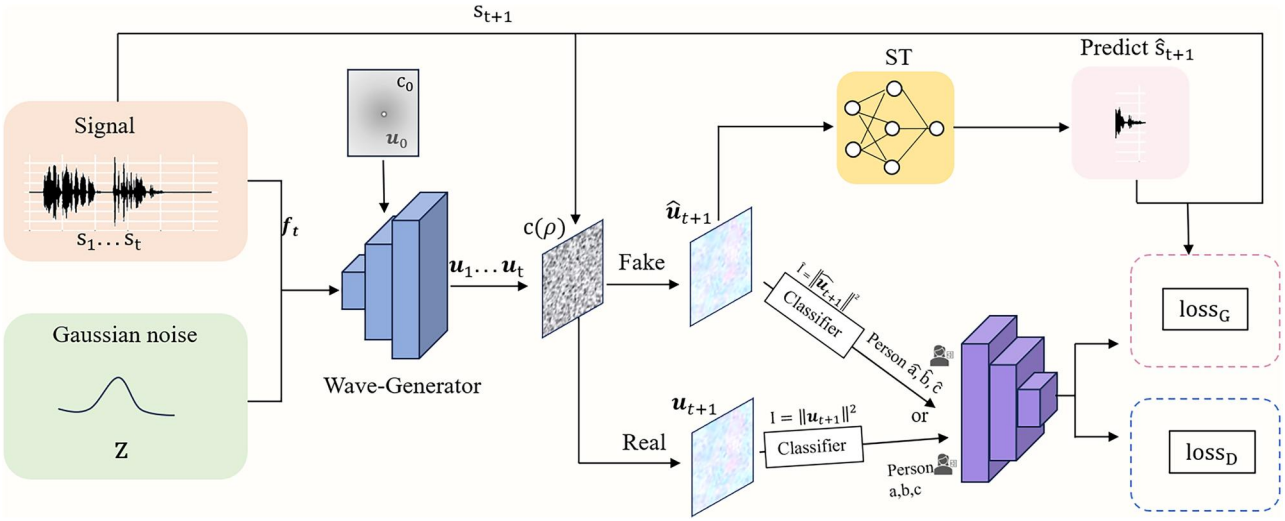

FIG. 1. Schematic illustration of the proposed Wave-GAN framework. The generator takes discretized speech signals (s_1, …, s_t) and Gaussian noise z as inputs and outputs of the predicted acoustic field $\hat{\mathbf{u}}_{t+1}$ through a wave-equation-based update. The learnable material density distribution $\rho(\mathbf{r})$ determines the local sound speed field $c(\rho, \mathbf{r})$. The discriminator distinguishes generated (Fake) and ground-truth (Real) acoustic fields $\mathbf{u}_{t+1}$ based on the physics residual. The ST, which denotes the signal transformation module used to obtain the predicted signal $\hat{s}_{t+1}$, is incorporated to convert the generated acoustic field back into a temporal signal. A classifier predicts speaker identities from the acoustic response. The generator and discriminator are optimized using loss_G and loss_D, respectively.

https://doi.org/10.1121/10.0043580

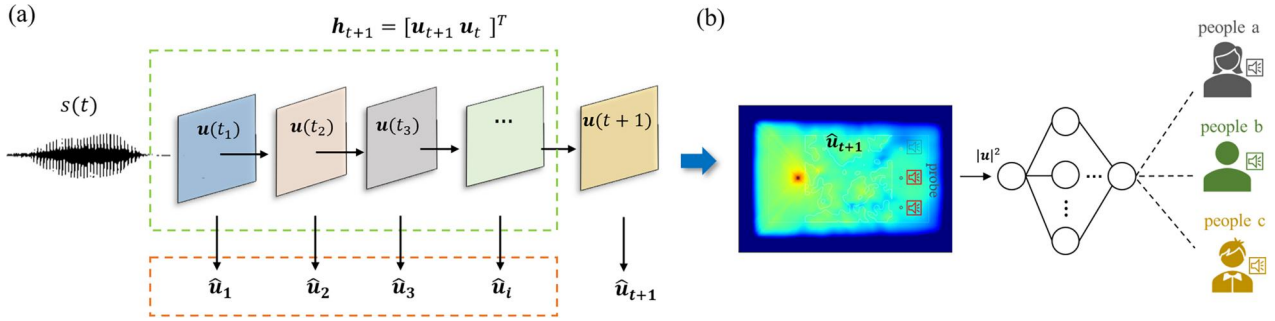

FIG. 2. The internal structures of the generator and the classifier presented in Fig. 1. (a) Wave-Generator. (b) Classifier.

classifier to inversely optimize the wave speed distribution $c(\rho, \boldsymbol{r})$, as detailed in the following section.

Figure 1 is the schematic illustration of the proposed Wave-GAN framework. Speech signals from different speakers are discretized in the time domain and injected as input to the generator. The generator embeds the discrete wave equation into its recurrence update. Given the input speech, it iteratively optimizes the material density distribution $\rho(\boldsymbol{r})$, which determines the local sound speed field $c(\rho, \boldsymbol{r})$. This density field acts as the "design variable" to be learned. The generated material distribution is used to simulate acoustic wave propagation via the time-discrete wave equation, producing the predicted acoustic field $\boldsymbol{u}_t$. Physical consistency is enforced through a physics-based loss [Eq. (6)], distinguishing between physically valid and invalid solutions. In parallel, a classifier probes the acoustic response at selected positions and predicts the speaker identity. The recognition accuracy serves as a task-specific objective. Gradients from the discriminator (physics residual), the classifier (recognition accuracy), and the regularization term (density smoothness) are backpropagated to update the density distribution through the Adam optimizer. Figure 2 shows the internal structures of the generator; and the classifier, discriminators, and optimizers will be introduced in detail in Secs. II B and II C [Fig. 2(a) corresponds to the Wave-Generator section on the left side of Fig. 1]. The architectures and hyperparameters of the related networks are provided in Table III. All simulations are conducted in a two-dimensional (2D) acoustic domain, where the material properties and wave fields are defined on a planar spatial grid. The dominant learnable parameters in the generator correspond to the spatial distribution of the material-dependent wave speed $c(\rho, \boldsymbol{r})$, discretized on a 150×100 grid, while the classifier and discriminator are implemented as lightweight multilayer perceptrons (MLPs).

B. Discriminators and classifiers

In the generator, after producing acoustic fields corresponding to different speech inputs, we employ both a discriminator and a classifier to evaluate the accuracy of the results. The classifier performs speaker classification based on the energy distribution ($\hat{I} = \|\hat{\boldsymbol{u}}_{t+1}\|^2$) of the generated acoustic field. The classification results are then passed to the discriminator D, which evaluates the classification accuracy by calculating a residual between the predicted and expected outcomes. Simultaneously, a state transformation network (ST) is incorporated to convert the generated acoustic field back into a temporal signal, assessing whether the generated field is physically plausible. This process is illustrated in Fig. 1 (right) and Fig. 2(b).

The residual from the generator enforces that the generated acoustic field must obey the wave equation as follows:

$$\mathcal{L}_{physics} = \frac{1}{T}\sum_{t=1}^{T}\left\| \frac{\boldsymbol{u}_{t+1} - 2\boldsymbol{u}_t + \boldsymbol{u}_{t-1}}{\Delta t^2} - c(\rho,\boldsymbol{r})^2 \cdot \nabla^2 \boldsymbol{u}_t - \boldsymbol{f}_t \right\|^2. \quad (5)$$

After generating $\boldsymbol{u}_{t+1}$, the ST model converts the generated acoustic field into a time-domain signal and assists in evaluating the plausibility and authenticity of the generated field at the signal level. The signal loss (accuracy between the generated sound and the real sound) is defined as $\mathcal{L}_s = 1/N \sum (\hat{s}_{t+1} - s_{t+1})^2$. The ST model is implemented as a simple MLP with three hidden layers, each containing 128 units.

Then, the classifier identifies which specific individual or category the sound originates from. The discriminator distinguishes between the "classification results from the real acoustic field" and those from the "generated acoustic field." Its loss function (sound recognition accuracy) is defined as follows:

$$\mathcal{L}_D = -\left[\log\left(D(a_{real})\right) + \log(1 - D(\hat{a}_{fake}))\right], \quad (6)$$

where $D(a_{real})$ denotes the probability that the discriminator judges the classification result of the real acoustic field as "accurate." $D(\hat{a}_{fake})$ represents the probability that it judges the classification result of the generated acoustic field as "accurate." The discriminator attempts to maximize this loss $\mathcal{L}_D$, driving $D(a_{real})$ toward 1 and $D(\hat{a}_{fake})$ toward 0. When this loss reaches a Nash equilibrium, it indicates that the generated material can accurately discriminate between speakers. The discriminator is also implemented using a standard MLP with three hidden layers, each containing 128 units, with ReLU activation functions applied after each hidden layer. No batch normalization is employed in this network.

Therefore, the overall optimization goal is to minimize the total loss function $\mathcal{L}_{total}$. Since all training outcomes depend on the acoustic fields corresponding to different sounds, and the field distribution is directly related to the material density, the total loss becomes a function of density. Thus, the backward optimization process involves

gradient-based optimization of the loss with respect to density. Accordingly, the following optimization objective is formulated as follows:

$$\min_{\rho} \mathcal{L}_{total} = \min_{\rho} \left(\mathcal{L}_{physics} + \mu\mathcal{L}_s + \beta\mathcal{L}_D\right), \tag{7}$$

where β and μ are trade-off coefficient. The specific implementation of this optimization process relies on the adjoint method. This is an efficient technique for computing gradients, particularly suited for solving optimization problems with physical field constraints, as it avoids the direct computation of the intractable partial derivatives $\partial \boldsymbol{u}/\partial\rho$.

After aggregating the current total loss $\mathcal{L}_{total}$, a virtual adjoint field $\lambda(\boldsymbol{r}, t)$ is introduced.[24] Mathematically, this field acts as a language multiplier to embed the physical constraints. The adjoint field satisfies an equation that is the time-reversed form of the original wave equation, with its source term being the gradient of the loss function with respect to the acoustic field $\boldsymbol{u}$, i.e., $\partial\mathcal{L}_{total}/\partial\boldsymbol{u}$:

$$\nabla^2\lambda - \frac{1}{c^2}\frac{\partial^2\lambda}{\partial t^2} = -\frac{\partial\mathcal{L}_{total}}{\partial\boldsymbol{u}}. \tag{8}$$

Using the forward acoustic field $\boldsymbol{u}$ and the solved adjoint field λ, the gradient of the total loss $\mathcal{L}_{total}$ with respect to the density ρ can be efficiently computed as follows:

$$\frac{\partial\mathcal{L}_{total}}{\partial\rho} = \int_0^T \left\langle \lambda(\boldsymbol{r}, t), \frac{\partial\boldsymbol{F}}{\partial\rho}u(\boldsymbol{r}, t) \right\rangle dt. \tag{9}$$

Here, $\boldsymbol{F}$ denotes the wave equation operator, i.e., $\boldsymbol{F}(\boldsymbol{u}) = \partial^2\boldsymbol{u}/\partial t^2 - c(\rho)^2 \cdot \nabla^2\boldsymbol{u} - \boldsymbol{f}$. This integral quantifies the influence of infinitesimal changes in the density ρ at any point in space on the total loss $\mathcal{L}_{total}$. After obtaining the gradient $\partial\mathcal{L}_{total}/\partial\rho$, gradient descent algorithms (e.g., Adam) are used to update the material density distribution as follows:

$$\rho_{t+1} = \rho_t - \eta\frac{\partial\mathcal{L}_{total}}{\partial\rho_t}, \tag{10}$$

where η is the learning rate, controlling the step size of the update. Iterative training continues until the total loss $\mathcal{L}_{total}$ converges or a preset maximum number of iterations is reached. The final optimized material distribution design is denoted as ρ.

C. Binarization strategy for fabrication

To obtain physically realistic and manufacturable acoustic designs composed of only two distinct materials, we adopt a density-based material representation combined with filtering and projection schemes. Rather than directly optimizing the spatial distribution of the wave speed, which would result in a non-differentiable optimization problem, a continuous design density field $\rho(\boldsymbol{r}) \in [0, 1]$ is introduced as the optimization variable.

To suppress numerical artifacts and enforce a minimum feature size, the raw density field is first processed using a spatial low-pass filter, yielding a smoothed density field $\tilde{\rho}(\boldsymbol{r})$. This filtering operation improves numerical stability and ensures manufacturable feature sizes. To drive the design toward a two-material configuration, a smoothed Heaviside projection is applied to the filtered density field. The projected density $\overline{\rho}$ is defined as follows:[7,24]

$$\overline{\rho}_i = \frac{\tanh(\beta\eta) + \tanh(\beta[\tilde{\rho}_i - \eta])}{\tanh(\beta\eta) + \tanh(\beta[1 - \eta])}, \tag{11}$$

where $\eta = 0.5$ specifies the projection threshold and β controls the sharpness of the projection. The parameter β is gradually increased during optimization to a final value of approximately 100, resulting in a numerically near-binary density distribution. The spatially varying acoustic wave speed is then obtained from the projected density via a linear interpolation as follows:

$$c(\boldsymbol{r}) = (c_2(\boldsymbol{r}) - c_1(\boldsymbol{r}))\overline{\rho} + c_1(\boldsymbol{r}), \tag{12}$$

where c_1 and c_2 denote the acoustic wave speeds of the two constituent materials. Specifically, $c_1 = 314\,\mathrm{m/s}$ corresponds to air, and $c_2 = 150\,\mathrm{m/s}$ corresponds to porous silicone rubber. Although this mapping is formally continuous, the strong projection ensures that the resulting wave speed field is numerically concentrated at the two target values, effectively yielding a binarized material distribution suitable for practical implementation. All forward simulations and gradient computations in the Wave-GAN optimization framework are performed using this filtered and projected wave speed field, with gradients propagated through both the filtering and projection operations via the adjoint method.

D. Resonance matching principle

While the proposed Wave-GAN framework provides an end-to-end optimization pipeline, its efficacy is grounded in a fundamental physical principle of resonance matching.[25] The core physical objective is to design a material density distribution ρ such that its intrinsic resonant frequencies ω_0 satisfy the characteristic equation as follows:[26]

$$det[K(\rho) - \omega_0^2 M(\rho)] = 0, \tag{13}$$

where $K(\rho)$ and $M(\rho)$ are the stiffness and mass matrices of the system, respectively, both dependent on the density distribution. Here, "matching" refers to the precise coordination between the wave propagation dynamics, governed by $c(\rho)$, and the frequency response of the material, with the spectral features of the speech excitation.

When the incident speech frequency ω approaches a resonant frequency ω_r of the material, the system enters a state of resonance. Mathematically, this corresponds to the condition where the determinant of the system matrix approaches zero. Under this condition, standing waves form efficiently

https://doi.org/10.1121/10.0043580

within the structure, concentrating acoustic energy and significantly enhancing the specific frequency components pertinent to the input speech. This selective amplification creates a unique and discriminative acoustic energy distribution at the probe points for different speakers, thereby enabling highly accurate classification.

Therefore, the optimization process is not a blind trial-and-error search. Instead, it is a purposeful endeavor to inversely design a material distribution ρ whose resonant response is tailored to the spectral content of the target speech signals. This physics-driven approach provides a clear and interpretable rationale for the model's operation, elevating it from a purely data-driven model to a physics-informed intelligent system.

III. RESULTS AND DISCUSSION

A. Data process

The dataset used in this study is sourced from https://openslr.org/38/, specifically the ST-CMDS-20170001_1: Free ST Chinese Mandarin Corpus. It contains a total of 930 speech recordings from 10 speakers (5 male and 5 female). To demonstrate the fundamental design capability of the proposed Wave-GAN framework, we adopted a repeated randomized sampling strategy. Specifically, for each independent training run, a subset of three speakers was randomly selected from a pool of ten speakers, and their corresponding recordings were used to form the dataset. The data were randomly split into 80% for training and 20% for testing, where the test set consists of previously unseen recordings. This repeated random selection across multiple training runs ensures that the model is exposed to diverse acoustic realizations and speaker combinations, thereby reducing potential bias associated with any fixed speaker subset. This simulation design allows the model to assess speaker identity based on learned acoustic patterns alone, without introducing the additional complexity of a larger speaker set. Such a controlled yet acoustically diverse setting is well suited for proof-of-concept validation, while keeping the computational cost tractable during the physics-embedded optimization stage.

The input to the material system is provided as a raw time-domain speech waveform, which is directly used as the excitation signal in the wave-equation-based training. No explicit preprocessing is applied, including amplitude normalization, spectral feature extraction, filtering, or other signal conditioning operations. This choice reflects the physics-based nature of the proposed framework, in which the excitation signal represents a physical source term in the governing wave equation, and numerical stability is ensured through the physics-constrained loss formulation rather than signal normalization.

Each recording contains a unique sentence per speaker, meaning that the model cannot rely on fixed textual or semantic features for training. Instead, it must learn to identify speakers based on their distinct acoustic energy distribution patterns. The model input is a randomly selected 1000-sample segment (0.1 s at 10 000 Hz) taken from the stable middle section of each original recording. These recordings are 2–3 s long on average, ensuring sufficient signal for analysis. The model was trained for a total of 500 epochs. Detailed statistics are provided in Table I.

TABLE I. Summary of the dataset and experimental setup used for the three-speaker classification task.

Item	Description
Dataset source	ST-CMDS-20170001 1 (Free ST Chinese Mandarin Corpus)
Original number of speakers	10 (5 males + 5 females)
Total utterances	930
Selected speakers	3 (denoted as person a, b, c)
Utterance content	Each sentence is unique; semantic recognition is not applicable
Sampling rate	10 000 Hz
Input feature	1000 sampling points per utterance (0.1 s window) (s_1, …, s_t)
Label type	One-hot encoding (3-class classification)
Train/test split	80% training, 20% testing
Total training epochs	500

B. Training the Wave-GAN to classify speech from different speakers

After constructing the modeling and optimization framework for trainable material distributions, we further evaluate the system's practical performance in a speech recognition task. Specifically, we train the wave-based system to classify speech signals from three different speakers, thereby validating the effectiveness of the proposed generative design framework in a classification scenario. To analyze the acoustic response of the material, a representative sample from each speech category is selected for training, and the probability distribution of energy at the measurement probe is computed. The results are shown in Figs. 3 and 4.

Figure 3 illustrates the performance of the wave-based system trained to recognize speech from three different speakers, including confusion matrices, loss curves, and accuracy trends for both training and testing phases. Figures 3(a) and 3(d) show the confusion matrices of the training and test sets at the initial stage of training (epoch = 0), respectively. In each confusion matrix, the horizontal axis represents the input (true speaker label), and the vertical axis represents the prediction (predicted speaker label). The numbers in each cell denote the classification percentage (%), with darker shades of blue indicating higher accuracy values and the exact numerical values are provided within the cells. It is evident that, before training, the system's ability to correctly recognize speakers is essentially random, with little to no ability to distinguish between them.

Figures 3(b) and 3(e) present the confusion matrices obtained after 500 training epochs. On the training set, the system achieves nearly perfect classification accuracy, with

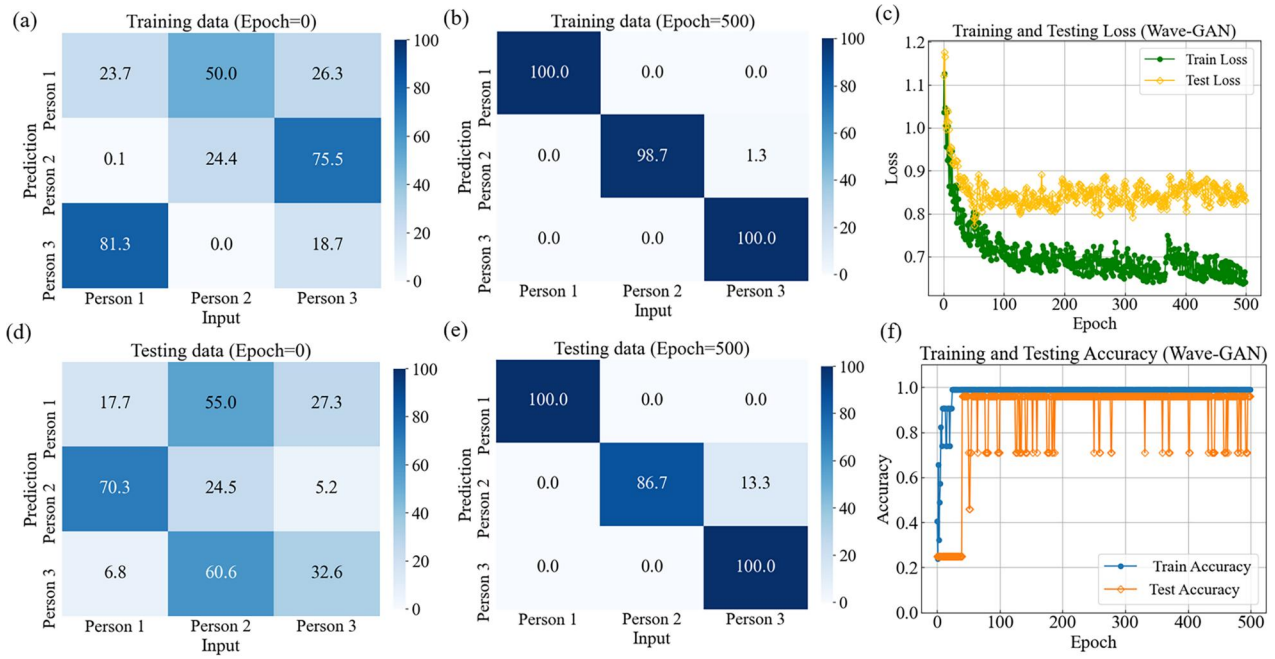

FIG. 3. Training results of the Wave-GNN-based three-person speech recognition system. (a) The initial confusion matrix at the beginning of training (epoch = 0). (b) The confusion matrix after training is completed (epoch = 500). (c) The evolution of loss curves for both training and testing phases. (d) The initial confusion matrix during testing. (e) The confusion matrix after training. (f) The classification accuracy curves for both training and testing.

99.6% of the predictions lying on the diagonal, indicating that the model has successfully learned the distinguishing acoustic features of the three speakers. The prediction accuracy on the test set is slightly lower, reaching 95.6%. Occasional misclassifications suggest that the generalization ability for certain categories can still be improved; nevertheless, as an initial attempt at developing a material-based speech recognizer, this level of performance is sufficient to demonstrate the feasibility of the proposed approach. In addition, the macro-averaged F1-score computed from the final test confusion matrix [Fig. 3(e)] reaches approximately 0.96, indicating balanced precision and recall across all speaker classes and confirming that the high classification accuracy is not dominated by any single class.

Figure 3(c) shows the average cross-entropy loss over five independent training runs, with a different set of three speakers randomly selected for each run. The training loss decreases steadily from 1.13 to approximately 0.64 after 500 epochs, while the testing loss follows a similar trend. The corresponding evolution of training and testing accuracy is shown in Fig. 3(f). The training accuracy rapidly increases and approaches unity, whereas the testing accuracy gradually stabilizes around 95% after several epochs, consistent with the results observed in the test confusion matrix. Together, these results demonstrate that the proposed Wave-GAN framework can effectively learn optimized material distributions that enable reliable discrimination between different speakers, thereby validating the feasibility of wave-

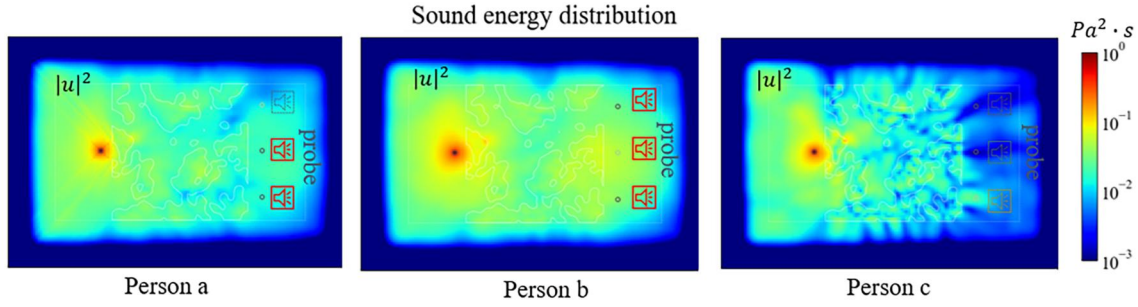

FIG. 4. Color maps show the distribution of acoustic energy within the material region in response to different speech signal perturbations (corresponding to the $\hat{\boldsymbol{u}}_{t+1}$ to classifier section in Figs. 1 and 2). The small speaker icons on the right indicate the positions of acoustic probes, where darker colors correspond to higher acoustic energy (stronger pressure fluctuations), and lighter colors indicate lower energy levels. After training, the sound velocity in the material region has been optimized to produce distinct acoustic field responses for different speakers. The classifier extract features from the average energy measured at the probe locations to determine the identity of the speaker.

https://doi.org/10.1121/10.0043580

equation-informed material optimization for speech recognition tasks.

Figure 4 presents color maps of the acoustic energy distribution within the material domain in response to perturbed speech signals. The optimized acoustic velocity distribution leads to distinct field responses for different speakers, resulting in measurable variations at the probe locations. This demonstrates that the proposed framework does not rely on textual content but instead learns to reshape the acoustic propagation environment such that speaker-specific energy signatures emerge naturally. By extracting average probe responses as discriminative features, the classifier achieves reliable speaker identification. This demonstrates that our Wave-GAN generative framework can design materials according to requirements. The designed materials can amplify subtle acoustic differences for robust recognition. To clearly show the material distribution generated after training, we present it separately in Fig. 5. We used non-dimensionalized spatial coordinates, normalized to the range [0, 1]. This highlights the distribution pattern itself and makes the results more general and scalable.

As shown in the Fig. 5, the initial density distribution at the beginning of training (epoch = 0) is uniform. A common approach to finding the global minimum of the loss function with respect to the material density is to use gradient descent, often implemented with the Adam optimizer to compute gradients. However, it is well known that gradient descent can be limited by the presence of local minima, making the optimization outcome sensitive to the choice of the initial starting point. In our case, the material starts with a uniform density distribution, which helps avoid this issue. After training, the Wave-GAN uses backpropagation to compute the density gradients and search for the global minimum. Therefore, compared with other numerical techniques, such as genetic algorithms, or particle swarm optimization, Wave-GAN can perform more optimization steps using the same computational resources. This computational advantage in gradient evaluation will become increasingly important for highly complex design problems, where the cost of computing numerical gradients is particularly high.

Figure 5 shows the training results of the material density distribution. As training progresses (e.g., Epoch 2 and Epoch 10), the density distribution gradually changes and begins to exhibit structural features, indicating that the gradient information derived from the wave equation has started to influence the evolution of the material density.

Between Epoch 50 and Epoch 150, the density distribution progressively converges toward a more stable pattern. Distinct high- and low-density regions begin to emerge locally within the material, reflecting the adaptation process between the speech signal and the material's acoustic response. As training continues further (Epoch 200, Epoch 300, and Epoch 400), the overall morphology of the density distribution stabilizes, showing little variation. This indicates that the system has reached an optimized state under physical constraints, and the density configuration of the material has become well adapted to the propagation characteristics of the speech signals.

The continuous density field shown in Fig. 5 is the direct output of the Wave-GAN optimization process. To obtain a manufacturable binary structure, we followed the standard post-processing procedure in topology optimization,[24] applying a thresholding operation to the continuous

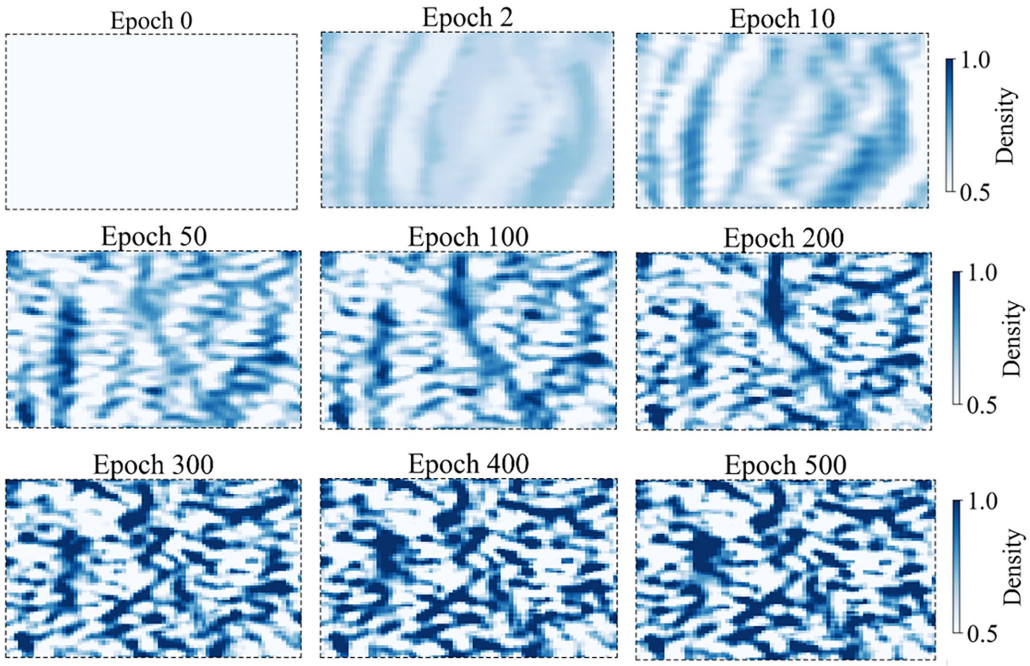

FIG. 5. Wave-GAN-generated material density distribution. The color represents the value of the density in the training region, with darker colors indicating higher density values.

field. The final binary design (Ref. 24, their Fig. 11 in Appendix 3) was obtained by setting the threshold at $(c(\rho) = 0.75)$, which corresponds to the intermediate value between the two material properties (c_1) and (c_2). This choice is a common practice in topology optimization to achieve a clear interface between phases.[24]

The stable density field in Fig. 5 is not a black-box result. It serves as physical evidence of the optimal resonance between sound waves and the material density gradient. Through gradient descent, the system automatically maps the information of the input speech into the material distribution. The final structure (see Fig. 11) thus becomes a task-adaptive acoustic recognizer. The significance of this approach is that it shifts feature learning from parameter optimization in the digital domain to material distribution optimization in the acoustic domain, laying the foundation for highly interpretable physical intelligent devices.

C. Noise robustness evaluation

To evaluate the robustness of the proposed Wave-GAN framework under realistic recording conditions, we conducted additional experiments by introducing additive noise into the input speech signals. This evaluation aims to assess the stability of the learned physics-constrained model when the excitation signals are corrupted by measurement noise, which commonly occurs in practical acoustic acquisition scenarios. Specifically, zero-mean additive white Gaussian noise was added to the time-domain speech signals used as excitation inputs. The noisy signal $s_n(t)$ is defined as follows:

$$s_n(t) = s(t) + \epsilon(t), \tag{14}$$

where $s(t)$ denotes the original clean speech signal and $\epsilon(t) \sim \mathcal{N}(0, \sigma^2)$ represents Gaussian noise. The noise level is controlled by the noise ratio, defined as the relative amplitude of the added noise with respect to the clean signal. Noise ratios of 0%, 1%, 5%, and 10% were considered, covering moderate to relatively severe noise conditions.

TABLE II. Recognition accuracy of the proposed Wave-GAN framework under different noise levels.

Noise ratio	0%	1%	5%	10%
Accuracy	95.8%	92.2%	88.7%	82.2%

The trained Wave-GAN model, without any retraining or noise-specific fine-tuning, was directly evaluated on these noisy test signals. Importantly, the test set consists of previously unseen speech recordings from known speakers, ensuring that the evaluation reflects generalization to novel excitation patterns rather than memorization of training samples.

The recognition accuracy under different noise levels is summarized as follows: 95.8% for clean signals (0% noise), 92.2% at 1% noise, 88.7% at 5% noise, and 82.2% at 10% noise. As expected, the recognition performance degrades gradually as the noise level increases. Nevertheless, the proposed framework maintains relatively high accuracy under moderate noise conditions, indicating that the incorporation of wave-equation-based physical constraints helps stabilize the learned acoustic field representations against perturbations in the input signal. The recognition accuracy under different noise levels is summarized in Table II.

D. Comparison of Wave-GAN, standard PINN, and conventional recurrent neural network (RNN)

To evaluate the modeling capability and generalization performance of Wave-GAN in the inverse design of wave fields, we conducted comparative experiments against PINNs[27] and a conventional RNN model. The network diagram is shown in Figs. 6 and 7. The training process and test results are shown in Fig. 8.

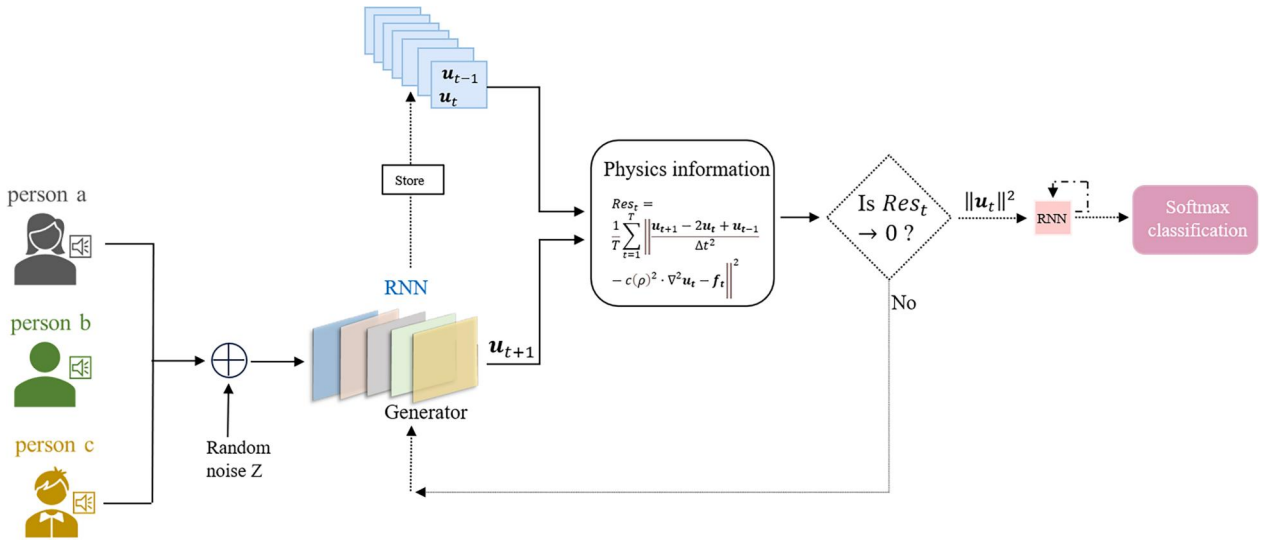

FIG. 6. Diagram of PINN architectures.

https://doi.org/10.1121/10.0043580

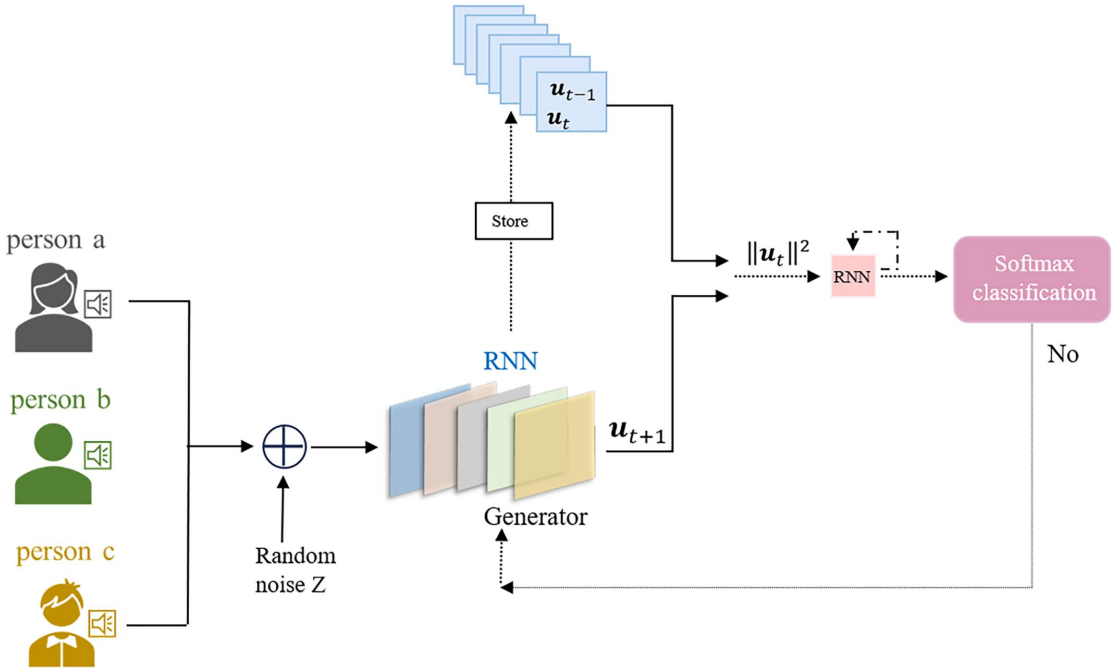

FIG. 7. Diagram of RNN architectures.

Figures 8(a) and 8(b) present the training and testing loss and accuracy curves of the PINN model. It can be observed that the model converges rapidly in the early stages, with training loss gradually decreasing and stabilizing. The test loss exhibits slight fluctuations but remains generally stable. Meanwhile, the training accuracy quickly reaches 1.0, and the test accuracy stabilizes around 0.75, indicating that PINN effectively captures the relationship between structural geometry and acoustic response and demonstrates a certain level of generalization. However, its accuracy is not sufficient for precise prediction.

Similarly, as shown in Figs. 8(c) and 8(d), the performance of the traditional RNN model reveals clear limitations. Although both training and testing losses decrease with the number of training epochs, they start from relatively large values (initially exceeding 8) and remain high even after training, with final losses still above 1. Meanwhile, the test accuracy starts at approximately 50% but fails to improve significantly during training, remaining in the low range of 0.2 to 0.3. These results indicate that the RNN model is unable to establish an effective mapping between structural parameters and energy response in the inverse design task, and it suffers from severe overfitting.

By comparison, Wave-GAN [Figs. 8(c) and 8(d)] incorporates physical consistency constraints through the wave equation during training, enabling it to explicitly encode local energy propagation pathways. This allows for more precise modeling of the temporal recursion and spatial distribution of internal structural energy. Additionally, Wave-GAN introduces a discriminator-based feasibility supervision module, which acts as a discriminative constraint to evaluate whether the generated structure satisfies the desired acoustic functionality.

Overall, Wave-GAN jointly optimizes two objectives: physical consistency and classification discriminability. This dual-objective approach alleviates the reliance of traditional supervised models on globally accurate acoustic field ground truth and improves adaptability in practical, function-driven acoustic material design. Experimental results show that, compared to traditional RNNs that only perform temporal modeling, Wave-GAN significantly outperforms in terms of accuracy, stability, and generalization capability, demonstrating stronger engineering applicability and greater potential for broader deployment.

E. Perceptual characteristics of speech

Different speaker traits are not perceived independently but are highly interdependent. Given the strong correlation among speaker characteristics, we performed principal component analysis on the speaker traits extracted from the average scores of each speech recording. Figure 9 shows the average energy spectra of three speaker categories after downsampling to 10 kHz. We observed that most of the energy for all speakers lies below 1 kHz, and there is substantial overlap in the average peak energy across different utterances. Furthermore, the female speaker's average peak energy is significantly lower than that of the other two speakers. This phenomenon suggests that, although different speakers may exhibit varying pronunciation habits or speech characteristics, their frequency and energy distributions are generally similar. This also indicates that it is difficult for a speech recognition system to rely on a universal model that captures these frequency and energy features effectively.

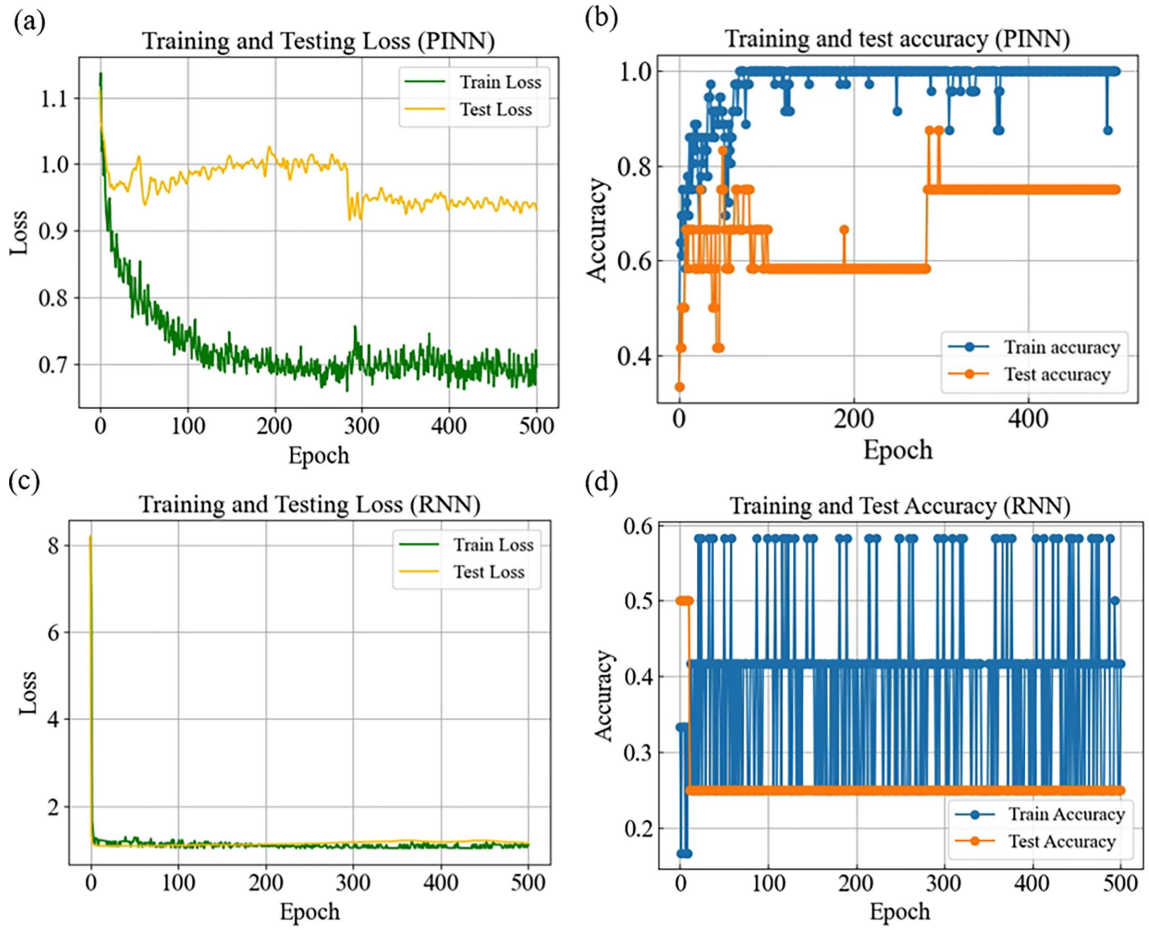

FIG. 8. Training results of PINN and traditional RNN for speech recognition material design. (a) Prediction accuracy of speech recognition using PINN. (b) Loss function during PINN training. (c) Prediction accuracy of speech recognition using RNN. (d) Loss function during RNN training.

Therefore, the role of speech recognition materials becomes particularly important. By employing suitable acoustic materials, we can ensure that the system performs robustly under various speech conditions and can handle subtle differences between speakers. Enhancing the robustness and accuracy of such systems through material design is an inevitable trend for future research. No amplitude normalization or energy scaling is applied to the input speech signals. Variations in signal power across speakers are intentionally preserved, as they constitute physically meaningful excitation differences that are naturally handled by the wave-equation-based propagation and material interaction.

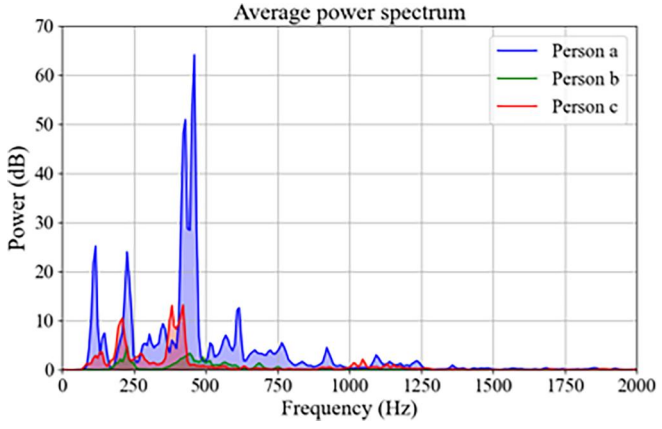

FIG. 9. Frequency content of different speakers. The plotted quantity represents the average energy spectrum of the recordings from person a, person b, and person c.

Figure 10 shows the Short-time Fourier transform (STFT) of vowel waveforms, used to analyze the time-frequency characteristics of speech signals. In the STFT matrix, rows represent speech recordings from different speakers, while columns correspond to different frequency components. By applying STFT to the speech signals, we can observe the spectral distribution over time, revealing the speech characteristics of different speakers and the dynamic changes in frequency components.

The recognition of speech signals relies not only on signal processing techniques (e.g., FFT and convolutional transforms), but more fundamentally on the physical essence of a deeper matching process. Traditional speech recognition focuses on matching between speech signals themselves, but in this study, we explore the matching relationship between speech signals and materials. From a broader perspective, whether in material design, wave

https://doi.org/10.1121/10.0043580

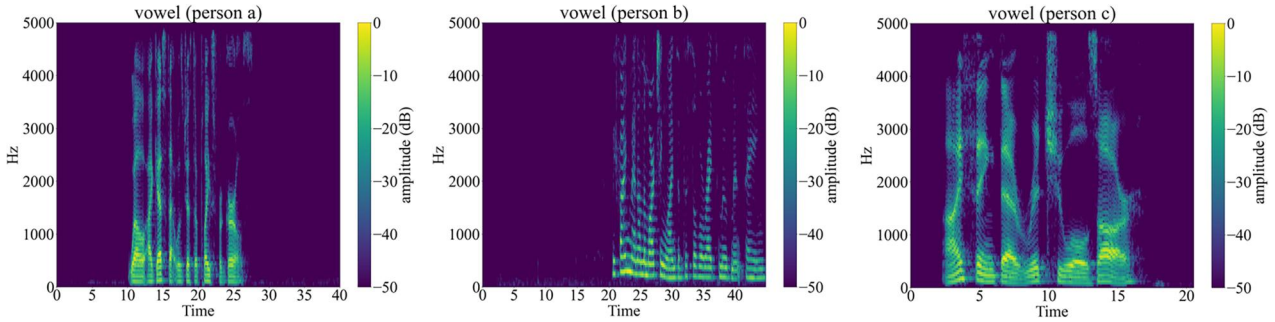

FIG. 10. The STFT of vowel waveforms. The time-domain signals (waveforms) are transformed into time-frequency representations, displaying the frequency components within different time segments (short-time windows). The formants (resonant frequencies) of the vowels appear as energy-concentrated bands in the spectrum.

propagation, or speech signal processing, the optimization of complex systems follows the same physical principle — the system tends toward an optimal matching state under multivariable optimization. This finding not only deepens our understanding of the speech recognition process but also offers a new research direction for physics-driven material optimization design.

IV. CONCLUSION

This study proposed a generative material design framework for speech recognition based on the wave equation, termed Wave-GAN. By deeply integrating wave physics with generative network techniques, the framework leveraged backpropagation to compute gradients of the objective function with respect to design variables, enabling the generation of nonlinear material density distributions tailored for speech recognition. The research established a connection between material design and speech recognition by constructing a wave-equation-governed neural network architecture. The wave equation was discretized into a trainable generative model, and efficient optimization of the material density distribution was achieved through adjoint-field-based gradient techniques. In a three-speaker classification task, numerical experiments demonstrated that the optimized nonlinear materials achieved a speaker recognition accuracy of 95.6%. This approach enabled the extraction of physically meaningful features at the microscale by regulating material density gradients through a wave-equation-informed generative neural network, thereby producing materials capable of accurately classifying different speakers. Overall, the proposed method provided a novel and interpretable paradigm for the design of intelligent acoustic materials.

Future work will further explore the potential of the proposed framework in other acoustic tasks. While the present study focuses on a controlled three-class speaker identification problem as a proof of concept, aimed at demonstrating the fundamental feasibility of embedding wave-equation constraints into a generative learning framework for the inverse design of acoustic materials, extending the proposed approach to more complex speech-related tasks, such as large-vocabulary speech recognition or keyword-based recognition, remains an important and challenging research direction. Achieving this goal will require systematic extensions of the current framework, including but not limited to modifications of the material representation, redesign of the signal-to-wavefield mapping strategy, and the ability to handle higher-dimensional classification spaces and larger datasets under physics-based constraints. These challenges define a clear path for future research. Our long-term objective is to develop the proposed framework

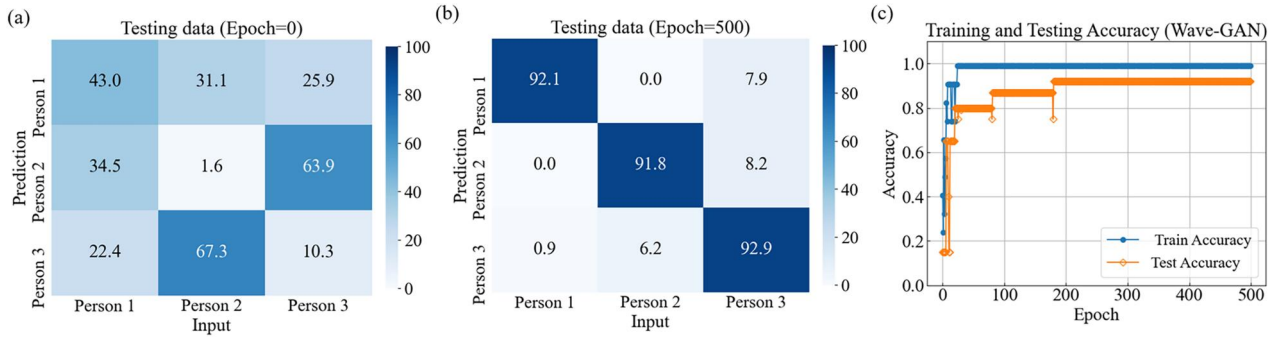

FIG. 11. Training results of the Wave-GNN-based three-person speech recognition system using the projected binary material representation. (a), (b) The initial confusion matrix at the beginning of training (epoch = 0), the confusion matrix after training is completed (epoch = 500), and the evolution of loss curves for both training and testing phases. (c) The classification accuracy curves for both training and testing, respectively.

into a physics-informed generative design tool with a certain degree of generality, applicable to a broad range of wave manipulation tasks.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China Grant No. 12372198.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The training and validation datasets used in this study are sourced from the OpenSLR speech recognition corpus, which has been appropriately cited in the manuscript. Some of the computational results presented in this paper are available upon request by contacting the corresponding author.

APPENDIX

1. Grid resolution sensitivity analysis

To assess the numerical reliability of the proposed Wave-GAN framework with respect to spatial discretization, we conducted a grid resolution sensitivity analysis by varying the resolution of the material design grid in the forward wave simulation and optimization process.

The acoustic domain was discretized using uniform Cartesian grids with different resolutions in the x and y directions, resulting in varying numbers of spatial design variables. All other model settings, loss formulations, and training procedures were kept identical across experiments to isolate the effect of grid resolution.

Table III summarizes the recognition accuracy and computational cost associated with different grid resolutions. The training time is reported per epoch and measured under the same hardware configuration.

As shown in Table III, increasing the grid resolution leads to improved recognition accuracy, particularly when moving from coarse to moderate resolutions. This improvement can be attributed to the increased representational capacity of the material layout and reduced numerical dispersion in the wave simulation. However, beyond the resolution used in the main text (150×100), the performance gains become marginal relative to the rapidly increasing computational cost.

TABLE III. Grid resolution sensitivity analysis.

Grid resolution ($N_x \times N_y$)	No. of design variables	Recognition accuracy (%)	Training time per epoch (min)
75×50	3750	80.1	0.12
100×75	7500	88.3	0.26
150×100 (main)	15 000	95.6	0.34
175×125	21 875	95.4	0.57
200×150	30 000	96.2	0.77

Specifically, while the finest grid (200×150) achieves the highest accuracy, it incurs more than twice the per-epoch training time compared to the baseline resolution. In contrast, the chosen resolution of 150×100 provides a favorable balance between numerical accuracy, structural expressiveness, and computational efficiency, making it well suited for physics-embedded optimization and repeated training runs. Therefore, considering the trade-off between recognition performance and computational cost, the grid resolution of 150×100 is retained as the primary design setting throughout the main text.

2. Model-architecture table

Table IV indicates the model architecture and neural network hyperparameters of Wave-GAN.

3. Binary manufacturability verification

In the main text, the Wave-GAN framework is formulated using a density-based material representation combined with filtering and smooth Heaviside projection. This strategy enables gradient-based optimization while driving the design toward a near-binary material distribution. To further address manufacturability concerns and verify that the optimized designs do not rely on intermediate material properties, we conduct an additional post-processing evaluation using a more strictly binarized material representation.

Importantly, this binarization is applied only at the testing and visualization stage and does not affect the training, optimization, or gradient computation processes described in the main paper. In the Wave-GAN framework, material distributions are represented using a density-based formulation combined with spatial filtering and a smooth Heaviside projection, as described in the main text. To further verify manufacturability, we apply the same projection operator used during training, but with a significantly increased projection parameter, to obtain a more strictly binarized material distribution.

Specifically, let $\tilde{\rho}_i$ denote the filtered density field obtained after spatial smoothing. The projected density is defined as follows:

$$\overline{\rho}_i = \frac{\tanh(\beta\eta) + \tanh(\beta(\tilde{\rho}_i - \eta))}{\tanh(\beta\eta) + \tanh(\beta(1-\eta))}, \tag{A1}$$

where $\eta = 0.5$ is the projection threshold and β controls the sharpness of the projection. During training, β is gradually increased to promote convergence toward a near-binary material distribution while maintaining differentiability. In this supplementary evaluation, applying the same projection formula with a larger value $\beta_{\text{bin}} = 300$, yielding a numerically near-binary density field. The corresponding wave

https://doi.org/10.1121/10.0043580

TABLE IV. The model architecture and neural network hyperparameters of Wave-GAN.

Module	—	—	Remark
Generator (G)	Input	Speech signal at the pre-t moment	$(s_1, ...s_t)$
	Control equation	$\frac{\partial^2 \boldsymbol{u}}{\partial t^2} - c(\rho)^2 \cdot \nabla^2 \boldsymbol{u} = \boldsymbol{f}$	
	Grid size (Δx, Δy)	1.5 mm	adjusted according to the actual situation
	Time step (Δt)	1×10^{-5}	
	Sound speed	$\bar{\rho}_i = \frac{\tanh(\beta\eta) + \tanh(\beta[\tilde{\rho}_i - \eta])}{\tanh(\beta\eta) + \tanh(\beta[1-\eta])}$ $c(r) = (c_2(r) - c_1(r))\bar{\rho} + c_1(r)$	
	Output	u_{t+1}	
	Trainable parameters	$c(\rho, r)$	Spatial distribution, dimension: (150, 100)
Classifier (C)	Input	Probe point acoustic energy	—
	Structure	MLP	Linear(128) → ReLU →Linear(128)…SoftMax
	Output	Speaker recognition rate	—
Discriminator (D)	Input	Classification results	Person a or b or c
	Structure	MLP	Linear (128) → Leaky ReLU → …
	Output	True and false probability	—
Hyperparameter	Optimizer	Adam	—
	Learning rate	1×10^{-4}	—
	Batch size	32	—
	Epochs	500	
	Loss weight (β, μ)	(0.5, 0.5)	—

speed distribution is then obtained via the same linear interpolation as in the main text:

$$c_i = (c_2 - c_1)\overline{\rho}_i + c_1, \tag{A2}$$

where c_1 and c_2 denote the acoustic wave speeds of the two constituent materials. This post-processing step does not affect the training or optimization procedure and is applied only for testing and visualization.

Figures 11 and 12 illustrates representative optimized geometries after applying the sharpened projection described above. Compared with the projected designs shown in the main text, the binarized geometries exhibit a clear two-phase structure with sharp material interfaces and negligible intermediate regions. Although the underlying wave speed field used for simulation is continuous during training, the post-processed geometries demonstrate that the learned designs are inherently compatible with a binary material realization.

To assess whether strict binarization degrades the functional performance, we additionally evaluate the trained

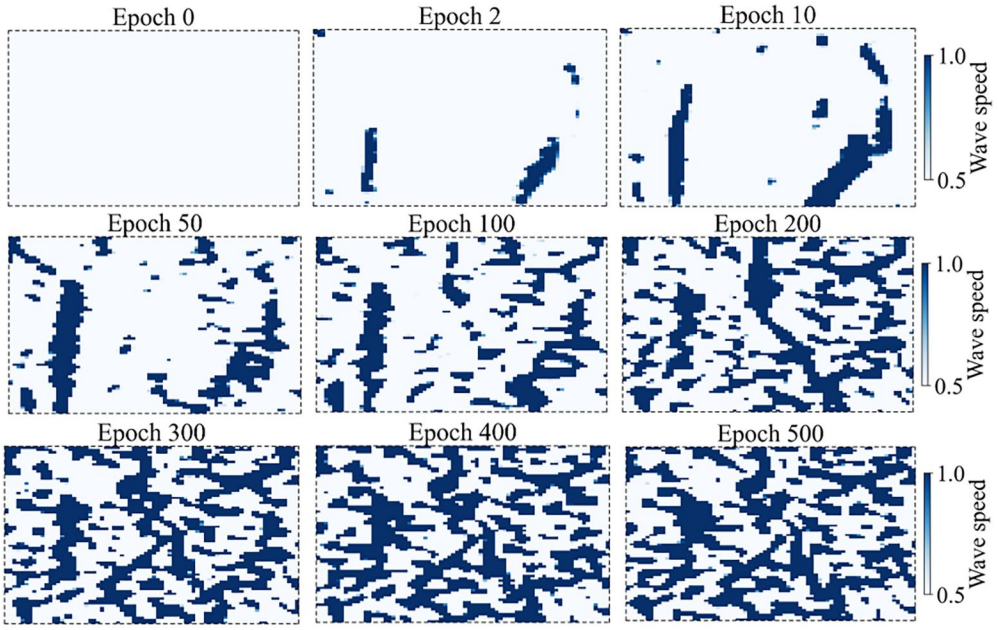

FIG. 12. The optimized geometries after applying the same Heaviside projection operator used during training, but with a larger projection parameter $\beta_{\text{bin}} = 300$. The resulting designs exhibit a clear two-phase structure with sharp material interfaces.

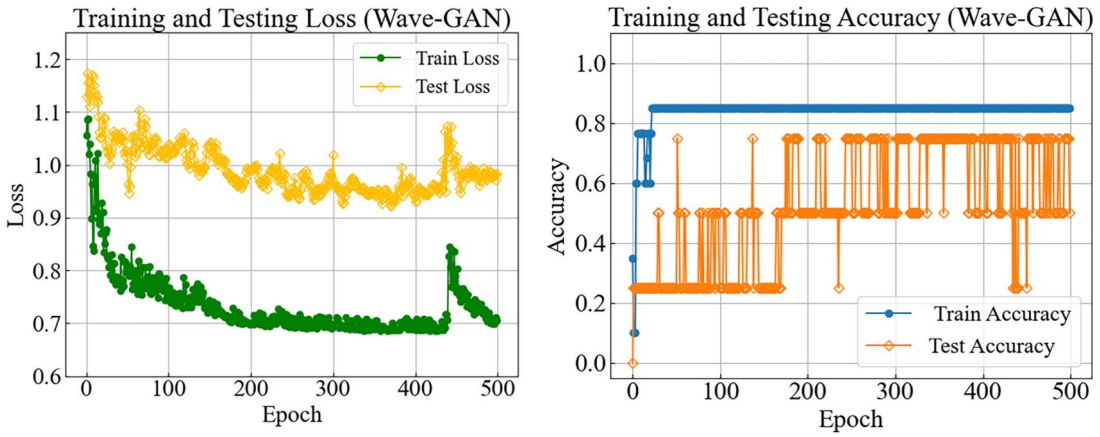

FIG. 13. Training and testing performance of the Wave-GAN framework on the Speech Commands version 0.01 dataset. Left panel: The evolution of training and testing loss over epochs. Right panel: The corresponding classification accuracy.

Wave-GAN using the binarized wave speed field $c_{\text{bin}}(\mathbf{r})$ during inference. The resulting speech-recognition accuracy is 92.3% remains comparable to that obtained using the smoothly projected field reported in the main text 95.6%, with only a marginal difference observed. This confirms that the optimized designs do not critically depend on intermediate material values and that the near-binary solutions produced by the training process are robust to strict binarization.

This supplementary study demonstrates that

1. the density-based optimization with smooth projection naturally produces designs that are numerically concentrated near the two target material properties;
2. a more strictly binarized representation, obtained via post-processing, preserves both the geometric characteristics and functional performance of the optimized designs;
3. the Wave-GAN framework therefore yields acoustic structures that are compatible with practical two-material fabrication constraints.

Together with the results in the main paper, these findings support the claim that the proposed approach enables the inverse design of manufacturable, binary acoustic materials for speech-recognition applications.

4. Preliminary evaluation on a larger speech corpus

To assess the task transferability of the proposed Wave-GAN framework beyond speaker identification, we conducted additional exploratory experiments using a larger and more diverse public speech dataset, Speech Commands version 0.01(Google Speech Commands), which contains short utterances corresponding to multiple spoken keywords from a broad range of speakers.

In these tests, the core structure of the proposed framework was kept unchanged, including the wave-equation-based generator, the material parameterization, and the optimization strategy. The speech signals from the Speech Commands dataset were directly used as excitation sources in the wave equation, following the same signal-to-wave-field mapping procedure described in the main text. For each training run, three keywords were randomly selected for recognition. No task-specific architectural modifications, large-scale hyperparameter tuning, or redesign of the material representation were introduced.

As shown in Fig. 13, the results indicate that the framework is still able to capture coarse discriminative acoustic patterns when exposed to more diverse speech inputs, suggesting that the underlying physics-informed optimization mechanism remains functional beyond the original dataset. However, the recognition performance in this setting is notably lower and less stable than that achieved in the controlled three-class experiments presented in the main text.

This behavior is expected and highlights several current limitations of the framework. In particular, reliable and accurate recognition for larger vocabularies would require substantial extensions, including but not limited to the following: (1) redesigning the material layout to accommodate a significantly larger number of output classes; (2) modifying the signal-to-field transformation strategy to better preserve fine-grained spectral–temporal features; and (3) performing extensive hyperparameter tuning and dataset-specific optimization.

We therefore emphasize that the present study does not claim immediate applicability to large-vocabulary speech recognition. Instead, these preliminary tests serve to illustrate the potential extensibility of the framework and to clarify the challenges associated with scaling physics-informed generative material design to more complex speech tasks. Developing a truly general-purpose, wave-based acoustic recognition material remains an important direction for future work.

[1]C. T. Chen and G. X. Gu, "Generative deep neural networks for inverse materials design using backpropagation and active learning," Adv. Sci. **7**(5), 1902607 (2020).
[2]N. Gao, J. Wu, K. Lu, and H. Zhong, "Hybrid composite meta-porous structure for improving and broadening sound absorption," Mech. Syst. Signal Process. **154**, 107504 (2021).
[3]X. Pan, X. Fang, X. Yin, Y. Li, Y. Pan, and Y. Jin, "Gradient index metamaterials for broadband underwater sound absorption," APL Mater. **12**(3), 031111 (2024).

[4]R. Dong, D. Mao, X. Wang, and Y. Li, "Ultrabroadband acoustic ventilation barriers via hybrid-functional metasurfaces," Phys. Rev. Appl. **15**(2), 024044 (2021).
[5]S. Huang, X. Fang, X. Wang, B. Assouar, Q. Cheng, and Y. Li, "Acoustic perfect absorbers via Helmholtz resonators with embedded apertures," J. Acoust. Soc. Am. **145**(1), 254–262 (2019).
[6]T. Srivastava, R. M. Winters, T. Gable, Y. T. Wang, T. LaScala, and I. J. Tashev, "Whispering wearables: Multimodal approach to silent speech recognition with head-worn devices," in *Proceedings of the 26th International Conference on Multimodal Interaction* (2024).
[7]T. W. Hughes, I. A. D. Williamson, M. Minkov, and S. Fan, "Wave physics as an analog recurrent neural network," Sci. Adv. **5**(12), eaay6946 (2019).
[8]R. Vasudevan, G. Pilania, and P. V. Balachandran, "Machine learning for materials design and discovery," J. Appl. Phys. **129**(7), 070401 (2021).
[9]K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. L. Billinge, E. Holm, S. P. Ong, and C. Wolverton, "Recent advances and applications of deep learning methods in materials science," npj Comput. Mater. **8**(1), 59 (2022).
[10]H. Zhang, Y. Wang, H. Zhao, K. Lu, D. Yu, and J. Wen, "Accelerated topological design of metaporous materials of broadband sound absorption performance by generative adversarial networks," Mater. Des. **207**, 109855 (2021).
[11]C. Gurbuz, F. Kronowetter, C. Dietz, M. Eser, J. Schmid, and S. Marburg, "Generative adversarial networks for the design of acoustic metamaterials," J. Acoust. Soc. Am. **149**(2), 1162–1174 (2021).
[12]P. Xu, X. Ji, M. Li, and W. Lu, "Small data machine learning in materials science," npj Comput. Mater. **9**(1), 42 (2023).
[13]J. F. Rodrigues, L. Florea, M. C. F. de Oliveira, D. Diamond, and O. N. Oliveira Jr, "Big data and machine learning for materials science," Commum. Mater. **1**(1), 12 (2021).
[14]P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. va. Hoesel, H. Schopmans, T. Sommer, and P. Friederich, "Graph neural networks for materials science and chemistry," Commun. Mater. **3**(1), 93 (2022).
[15]D. Morgan and R. Jacobs, "Opportunities and challenges for machine learning in materials science," Annu. Rev. Mater. Res. **50**(1), 71–103 (2020).
[16]J. Duan, H. Zhao, and J. Song, "Spatial domain decomposition-based physics-informed neural networks for practical acoustic propagation estimation under ocean dynamics," J. Acoust. Soc. Am. **155**(5), 3306–3321 (2024).
[17]C. Song, T. Alkhalifah, and U. B. Waheed, "Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks," Geophys. J. Int. **225**(2), 846–859 (2021).
[18]J. Yan, Y. Li, G. Yin, S. Yao, and Y. Peng, "Inverse design on customised absorption of acoustic metamaterials with high degrees of freedom by deep learning," Mech. Syst. Signal Process. **237**, 112989 (2025).
[19]H. Guo, W. Chen, Y. Wang, F. Ma, P. Sun, T. Yuan, and X. Xie, "Parametric modeling and deep learning-based forward and inverse design for acoustic metamaterial plates," Adv. Mater. Struct. **31**(30), 12986–12996 (2024).
[20]H. Baali, M. Addouche, A. Bouzerdoum, and A. Khelif, "Design of acoustic absorbing metasurfaces using a data-driven approach," Commun. Mater. **4**(1), 40 (2023).
[21]C. Qian, I. Kaminer, and H. Chen, "A guidance to intelligent metamaterials and metamaterials intelligence," Nat. Commun. **16**(1), 1154 (2025).
[22]Y. Jin, B. Wen, Z. Gu, X. Jiang, X. Shu, Z. Zeng, Y. Zhang, Z. Guo, Y. Chen, T. Zheng, Y. Yue, H. Zhang, and H. Ding, "Deep-learning-enabled MXene-based artificial throat: Toward sound detection and speech recognition," Adv. Mater. Technol. **5**(9), 2000262 (2020).
[23]M. Simonović, M. Kovandžić, I. Ćirić, and V. Nikolić, "Acoustic recognition of noise-like environmental sounds by using artificial neural network," Expert Syst. Appl. **184**, 115484 (2021).
[24]M. Bendsøe and O. Sigmund, *Topology Optimization: Theory, Method and Applications* (Springer, Berlin, Germany, 2003).
[25]M. P. Norton and D. G. Karczub, *Fundamentals of Noise and Vibration Analysis for Engineers* (Cambridge University Press, London, UK, 2003).
[26]K.-J. Bathe, *Finite Element Procedures* (Watertown, MA, 2006).
[27]M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. **378**, 686–707 (2019).