

Supporting Information

Xie et al. 10.1073/pnas.1502276112

SI Materials and Methods

Metamaterial-Based Resonances Design. The resonating dispersion of each waveguide is contributed by an array of Helmholtz resonators distributed within the fan-shaped area. Such side branch resonators are commonly used as acoustic filters (39). The resonating frequency can be estimated as $\omega_0 = c_s / \sqrt{h(h + \Delta h)}$, where c_s is the speed of sound, h is the height of the resonator, and Δh is the effective height increase due to the opening. Fig. S1A (*Inset*) shows the dispersion of the transmission of a resonator numerically retrieved from simulation.

To aid the design of the sensing system, we model the resonator using lumped elements. For a side branch resonator indexed as j in the i th waveguide, the complex transmission can be expressed as $T_{ij} = 1 - (R_{ij} / (R_{ij} + r_{ij} + j\omega L_{ij} + 1/j\omega C_{ij}))$. The i th waveguide is essentially a band-stop frequency filtering network composed of an array of resonators, as shown in Fig. S1B. The resonances are randomized and contribute to a measurement matrix that favors compressive sensing, as will be explained in *SI Text, Forward Model Derivation*, below.

Experimental Setup. The layout of the experimental setup of the system is shown in Fig. S2. The metamaterial multispeaker listener is deployed on the floor of an anechoic chamber. The sensor at the center is an omnidirectional surface mount microelectromechanical microphone (Analog Devices ADMP401). The signal acquired by the microphone is sent to a preamplifier (Stanford Research SR560) before being digitized by a data acquisition device (NI PCI-6251). Multiple speakers are used as audio sources. They are placed on the floor of the chamber and mounted toward the ceiling so that the radiation of each is symmetrical with respect to the z axis (perpendicular to the floor).

A library of 100 broadband pulses were used as words for the multispeaker listening experiment. Each word has a length of 5 ms and the adjacent words are separated by a pause of 1 s to avoid overlapping of signals of different time indices. All of the words are synthesized by superposing randomized truncated cosine waves whose center frequencies span the spectral range of interest. The time-domain expression of a synthesized word can be expressed as $s_p(t) = \sum_{\omega_q=\omega_1}^{\omega_M} a_{pq} \cos(\omega_q t + \theta_{pq})$, where a_{pq} and θ_{pq} are computer-generated pseudorandom numbers within the range $a_{pq} \in [0, 1]$ and $\theta_{pq} \in [0, 2\pi]$, respectively. Two examples are shown in Fig. 3A (*Inset*) in the main text.

SI Text

Forward Model Derivation. Assuming the resonances in each waveguide are distributed sparsely over the interested frequency range and only first-order filtering responses dominate, the overall frequency modulation of a waveguide can be approximated by the multiplication of the individual responses of the resonators: $T_i(\omega) = \prod_j T_{ij}(\omega)$. For a source located at $\vec{r}_k$, the frequency response can be derived by propagating the waveguide responses from each waveguide aperture $\vec{r}_i$ to the source location $\vec{r}_k$: $P_c(\omega, \vec{r}_k, S_0) = a(\omega) S_0(\omega) \sum_{i=1}^{36} G(\omega, \vec{r}_i, \vec{r}_k) T_i(\omega) R(\omega, \vec{r}_i, \vec{r}_k)$, where $S_0(\omega)$ is the spectrum of the audio signal from the source, $R(\omega, \vec{r}_i, \vec{r}_k)$ is the waveguide radiation pattern which is mostly determined by the shape of the waveguide aperture, and $G(\omega, \vec{r}_i, \vec{r}_k) = e^{-jk|\vec{r}_i - \vec{r}_k|} / |\vec{r}_i - \vec{r}_k|$ is the Green's function from the location $\vec{r}_i$ of the aperture of the i th waveguide to the location $\vec{r}_k$. The coefficient $a(\omega)$ includes all other factors such as sensor and speaker responses that are uniform for different source locations and audio signals.

Each column of the measurement matrix

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN} \end{bmatrix}$$

represents the discretized Fourier components of a source emitting an audio message from the predefined library from one of the possible locations in the scene. The number of columns of the matrix is $N = K \times P$, where K is the number of possible locations and P is the size of the audio library. An element in the measurement matrix can be expressed as

$$h_{mn} = P_c(\omega_m, \vec{r}_k, S_p) = a(\omega_m) S_p(\omega_m) \times \sum_{i=1}^{36} G(\omega_m, \vec{r}_i, \vec{r}_k) T_i(\omega_m) R(\omega_m, \vec{r}_i, \vec{r}_k).$$

Each row $\mathbf{H}_m$ of the measurement matrix represents a test function for the object vector at one frequency, because a measurement value in the measured data vector is sampled in the way defined by the test function as $g_m = \langle \mathbf{f}, \mathbf{H}_m \rangle$, where the angle bracket indicates the inner product. The randomization of the measurement matrix for this sensing system is contributed by the carefully designed waveguide responses $T_i(\omega)$. Discussions on the quantified characterization of the measurement matrix are detailed in *SI Text, Measurement Matrix Evaluation*, below.

Reconstruction Quality Metric. The average MSE is used as a metric to quantify the reconstruction quality. It is defined as $1/KT \sum_{k=1}^K \sum_{i=1}^T |\hat{\mathbf{f}}_k(t_i) - \mathbf{f}_{k0}(t_i)|^2$, where K is the number of all the possible source locations, T is the maximum time index, $\hat{\mathbf{f}}_k(t_i)$ is the reconstruction vector for the k th source at time index t_i , and $\mathbf{f}_{k0}(t_i)$ is the ground-truth vector for the same source at the time index. In our experiment, the nonzero values in the object vector (ground truth) are set to unity. However, nonunity relative amplitudes in the object vector in principle can still be reconstructed by the L1-norm regularization inverse algorithm (37).

Measurement Matrix Evaluation. The measurement matrix reflects the physical properties of the sampling process (such as the efficiency of phase and amplitude modulations and the spectral responses of the measurement). Here we discuss the performance of measurement matrix, including the resolution and the stability of the inverse estimation.

The resolution of the inverse problem is associated with the number of measurement M and the sampling efficiency of those measurement modes. In compressive sensing, M is less than the dimensionality of the object vector N , which means that without any prior information, the measurement cannot fully recover the scene. To overcome such limitation, enough independent sampling provided by the measurement matrix is required. A quality measurement matrix reduces overlapping of measurement modes and thus increases sampling efficiency, providing an opportunity for a reconstruction to outperform an undersampled measurement. With a reconstruction algorithm that uses proper regularization, an ill-conditioned compressive sensing might still be able to recover the scene.

Here we use the average mutual coherence to evaluate the independence of the modulation channels (36, 40). The mutual

coherence between two modulation channels $\tilde{h}_i$ and $\tilde{h}_j$ is $W_{ij} = \tilde{h}_i^* \tilde{h}_j / |\tilde{h}_i| |\tilde{h}_j|$. The average mutual coherence can be expressed as $\mu_{av} = \sum_{i \neq j} W_{ij} / (m(m-1))$, which represents the average overlap between all m modulation channels. The average mutual coherence ranges from 0 to 1, with close-to-zero values meaning highly randomized modulations whereas close-to-unity values mean highly similar modulations. The average mutual coherence for the case with metamaterials is about 0.20, whereas that without metamaterials is about 0.93. The significantly reduced mutual coherence generally indicates high reconstruction fidelity and lower error (36).

The noise sensitivity of the measurement matrix can be considered as its contribution of noise amplification during the inverse process. Singular value decomposition can be used to analyze the measurement matrix to estimate the performance of the ill-conditioned inverse problem (34). Because small singular values cause the inverse process sensitive to the noise, a stable inverse problem should reduce the number of small singular values. One way to evaluate the performance has been inspired by the covariance matrix. The variance will be amplified by the sum of inverse square of the singular values $\sum_i 1/s_i^2$. However, in an ill-posed inverse problem, such estimation will be dominated by those close-to-zero elements, which reduce the dynamic range of the evaluation. Practically, a threshold level can be set for singular value spectrum with the measured noise floor. The singular values above this threshold will be considered as effective and they dominate the sampling process. The number of effective singular values can thus provide a metric of the performance of the inverted problem.

The Spatiotemporal Degrees of Freedom of the Measurement Modes.

The quality of the measurement modes can be quantitatively described by the number of the degrees of freedom (DOF) in both spatial and spectral domains (41). The spatial DOF along a dimension x can be estimated as $X/\delta x$, where X is the total range and δx is the correlation distance along this dimension. Similarly the spectral DOF can be estimated as $\Delta\omega/\delta\omega$, where $\Delta\omega$ is the system bandwidth and $\delta\omega$ is the correlation frequency. Using the physical model described in an earlier section (*SI Text, Forward Model Derivation*), we calculated the decorrelation for both frequency and the azimuth of the measurement modes. The results are presented in Fig. S3 *A* and *B*. If we define the correlation range where the normalized correlation coefficient drops to 0.5, the azimuth correlation angle is about 8° , whereas the correlation frequency is about 80 Hz. This indicates that the number of DOF of the azimuth is about 45 and that within the spectrum of interest is about 38.

There is a fundamental tradeoff between the quality factors of the physical layer modulations, the received signal power, and the DOF. On the one hand, the quality factors of the realized system will be lower than the ideal model and will thus decrease the DOF as suggested by Lipworth et al. (42). On the other hand, the modulations offered by metamaterials are essentially band-stop filtering. The higher the quality factors, the lower the received

signals power. The interplay between these aspects needs to be tailored for further optimization of the sensing performance.

The Advantages of Using Metamaterials. Resonator-based metamaterials contribute several unique features to our sensing system: (i) the flexible material parameter dispersions allow us to design and realize the optimal measurement modes for the sensing tasks. In the demonstrated work, the randomized resonating dispersions create sufficient incoherent information channels; (ii) the metamaterials allow us to achieve required spatial complexity and frequency dependency (DOF) within much smaller size, compared with the random scatterers-based methods where the device size typically needs to be orders of magnitude larger than the wavelength; and (iii) the system performance can be estimated relatively easily with the effective parameters of the metamaterials using theoretical forward models and full-wave simulations.

Controlled Experiment Without Metamaterials. The control experiment of attempting to reconstruct three sources with only an omnidirectional microphone has an averaged MSE = 1.99, which is about 25 $\times$ that of the reconstruction with metamaterial (MSE = 0.08). Both results are obtained using the sample experimental setup and with the same regularization parameters in the reconstruction algorithm. As is obvious from Fig. S4*B*, the reconstruction without metamaterial is too poor to provide useful information about the sources. The deficient reconstruction results from the more correlated modulations (transfer functions) for the three sources when the metamaterial coating is removed, as shown in Fig. S4*A*.

Reconstruction of Two Fixed Activated Sources out of Six Possible Locations.

The task of this reconstruction is to identify the locations of two activated sources out of six possible fixed locations and at the same time segregate their overlapping signals. The reconstruction for the six locations exhibits two blocks (corresponding to the two activated sources) with dominant signal strength (shown in Fig. S5), whereas four others with negligible reconstructed signal strength are not shown. This designed task is to demonstrate a simple combination of signal segregation and localization of fixed sources. A successful recognition is counted when the content reconstructed with the highest reconstruction strength is above 50% and matches the ground-truth word for each activated source at each time index (whereas the maximum reconstruction is below 10% for each silence source). The overall recognition ratio calculated using such definition is 87.5% for this task.

Reconstruction of Randomly Emitting Sources out of Four Possible Locations.

This task is to reconstruct the signals from four possible speakers. Unlike the previous task where the activated sources are fixed, in this task two or three speakers are randomly selected from the four possibilities at each time index to emit random words. The reconstruction results are shown in Fig. S6. Using the same definition of recognition as the previous task, the overall recognition ratio is 95%.

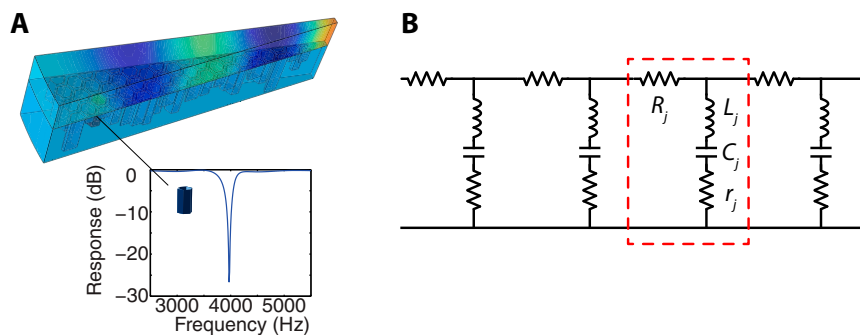


Fig. S1. (A) Fan-like waveguide with one resonator highlighted. (Inset) Simulation result of the amplitude of the transmission of the waveguide with only the contribution from the highlighted resonator. (B) Band-stop equivalent lumped elements network representation of the waveguide.

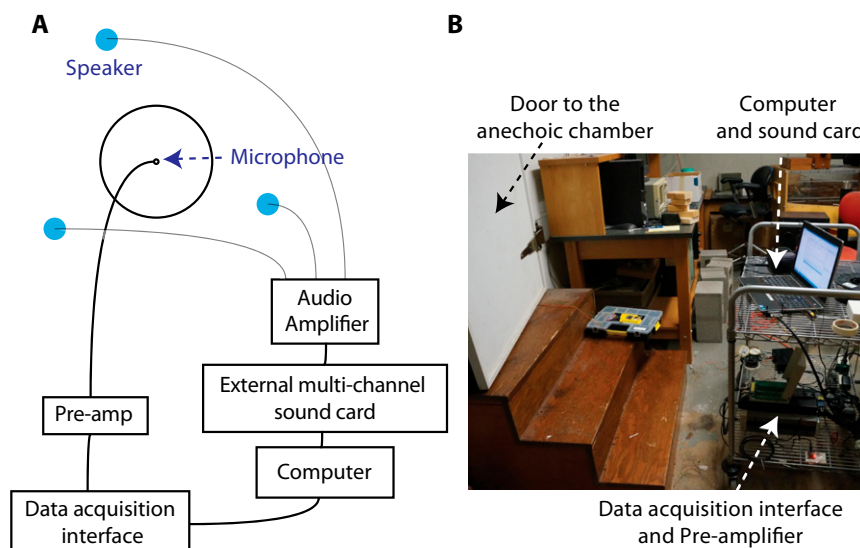


Fig. S2. (A) Schematic of the system layout. (B) Photograph of the data acquisition platform outside of the anechoic chamber.

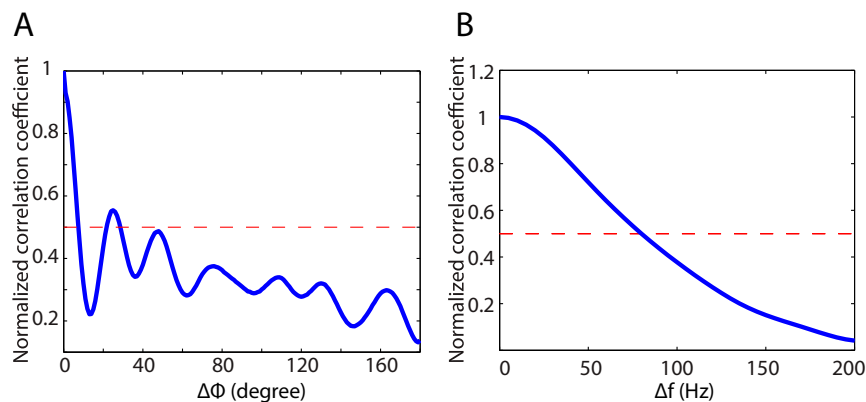


Fig. S3. Azimuthal and spectral correlation curves estimated from the forward model. (A) Normalized correlation coefficient with respect to the shifted azimuthal angle. (B) Normalized correlation coefficient with respect to the shifted frequency.

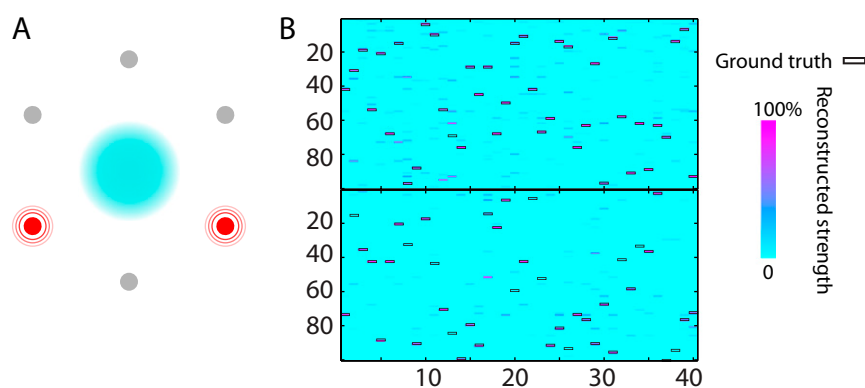


Fig. S5. Reconstruction of two active resources out of six possible locations. (A) Schematic of the reconstruction task. Two red radiating circles indicate the sound-emitting sources, whereas the gray circles represent silent sources. (B) Reconstruction of two locations with significant reconstruction strength. The overall recognition ratio for this task is 87.5% using the criterion defined in *SI Text*.

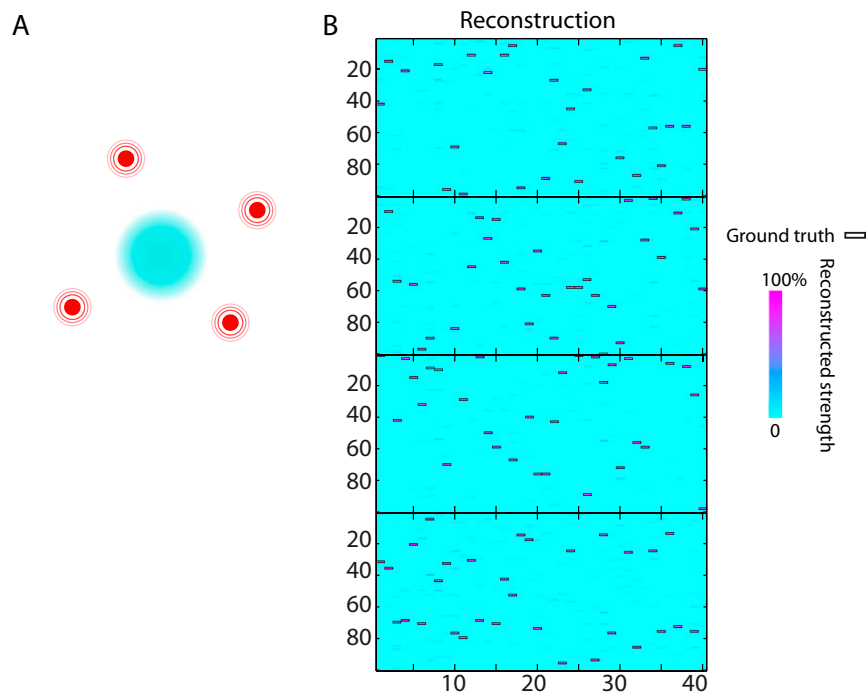


Fig. S6. Reconstruction of two or three randomly selected active sources out of four possibilities. (A) Schematic of the reconstruction task. Four sources have the same possibility of radiating source and only two or three are randomly selected to emit sound at each time index. (B) Reconstruction of the four locations. The overall recognition ratio for this task is 95% using the criterion defined in *SI Text*.