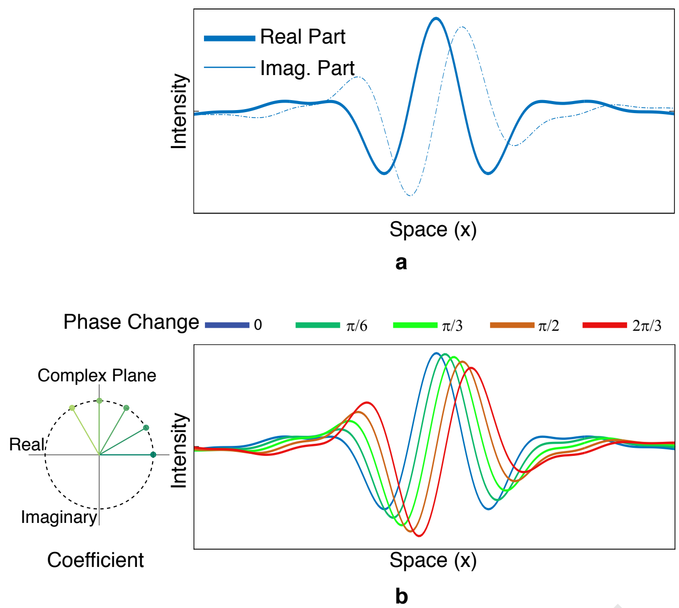


Fig. S1. Increasing the phase of complex steerable pyramid coefficients results in approximate local motion of the basis functions. **a**, A 1D slice of a complex steerable pyramid basis function. **b**, The basis function is multiplied by several complex coefficients of constant amplitude and increasing phase to produce the real part of a new basis function that is approximately translating. (Reproduced from (31).)

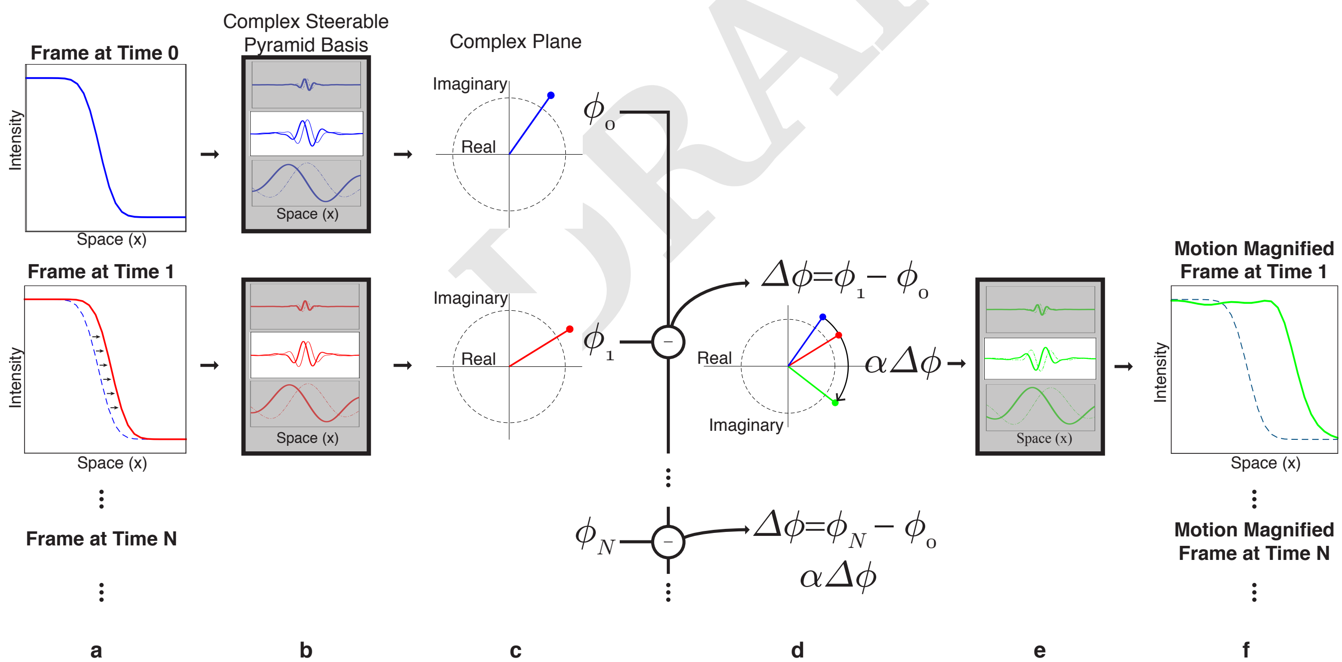


Fig. S2. A 1D example illustrating how the local phase of complex steerable pyramid coefficients is used to amplify the motion of a subtly translating step edge. **a**, Frames (two shown) from the video. **b**, Sample basis functions of the complex steerable pyramid. **c**, Coefficients (one shown per frame) of the frames in the complex steerable pyramid representation. The phases of the resulting complex coefficients are computed. **d**, The phase differences between corresponding coefficients are amplified. Only a coefficient corresponding to a single location and scale is shown; this processing is done to all coefficients. **e**, The new coefficients are used to shift the basis functions. **f**, A reconstructed video is produced by inverse transforming the complex steerable pyramid representation. The motion of the step edge is magnified. (Reproduced from (31).)

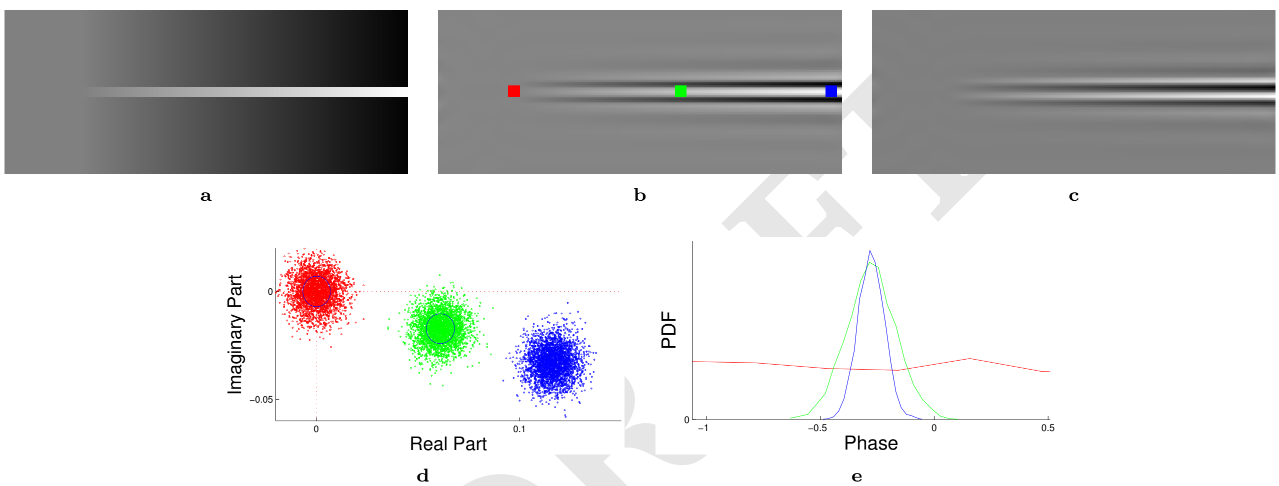


Fig. S3. Noise model of local phase. **a**, A frame from a synthetic video with noise. **b**, The real part of a single level of the complex steerable pyramid representation of **a**. **c**, The imaginary part of the same level of the complex steerable pyramid representation of **a**. **d**, A point cloud over noisy values of the real and imaginary values of the complex steerable pyramid representation at the red, green and blue points in **b**. (**e**). The corresponding histogram of phases.

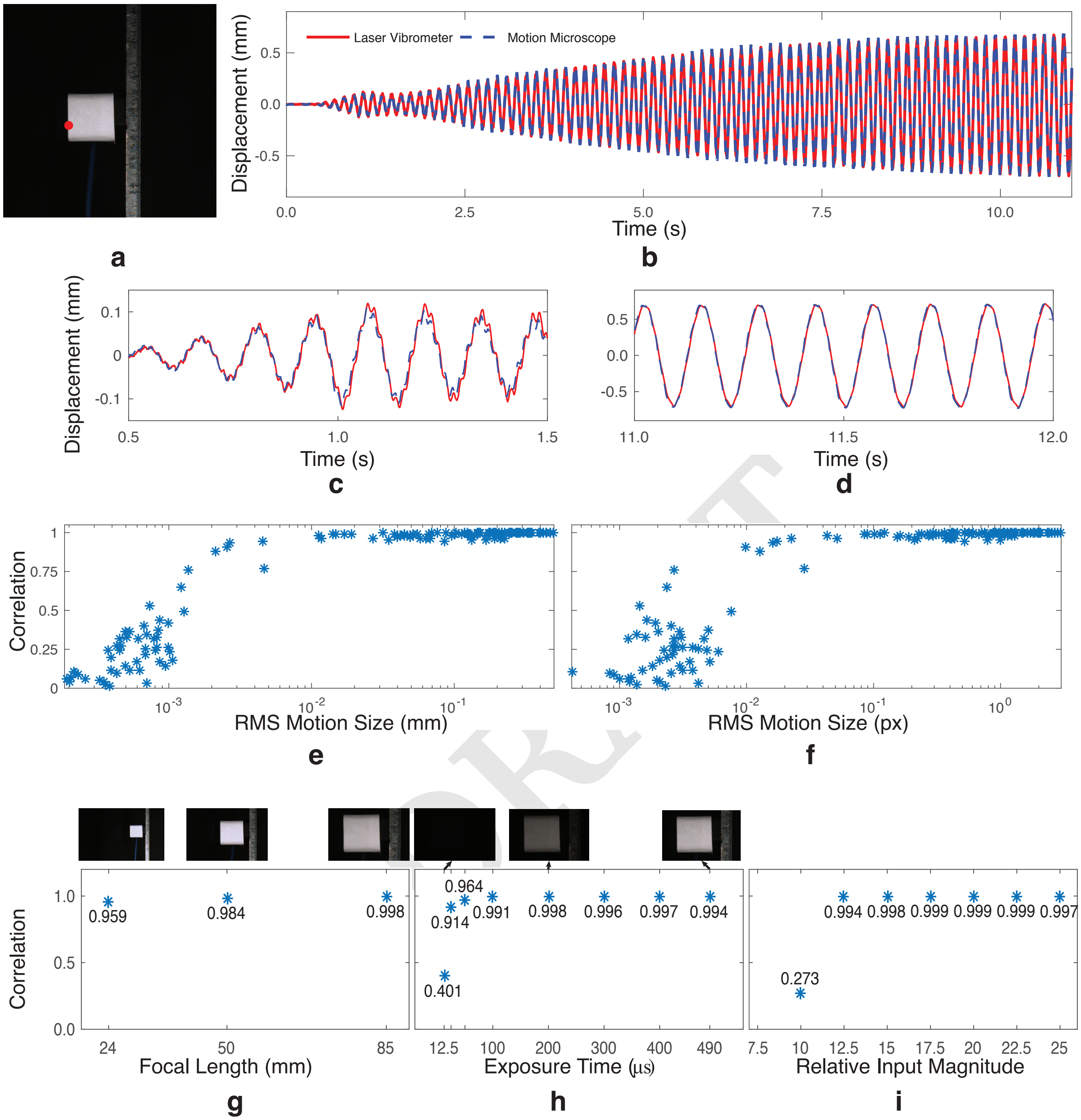


Fig. S4. A comparison of our quantitative motion estimation vs. a laser vibrometer. Several videos of a cantilevered beam excited by a shaker were taken with varying focal length, exposure times and excitation magnitude. **a**, A frame from one video. **b**, Motions from the motion microscope at the red point are compared to the integrated velocities from a laser vibrometer. **c**, **b** from 0.5-1.5s. **d**, **b** from 11-12s. **e**, The correlation between the two signals across the videos vs. the root-mean-square (RMS) motion size in millimeters measured by the laser vibrometer. **f**, The correlation between the two signals across the videos vs. the root-mean-square (RMS) motion size in pixels measured by the laser vibrometer. **g**, The correlation between the signals vs. focal length (exposure time: 490 μs, excitation magnitude: 15). **h**, Correlation vs. exposure time (focal length: 85mm, excitation magnitude: 15). Cropped frames from the corresponding videos are shown above. **i**, Correlation vs. relative excitation magnitude (focal length: 85 mm, exposure time: 490 μs). Only motions at the red point in **a** were used in our analysis.

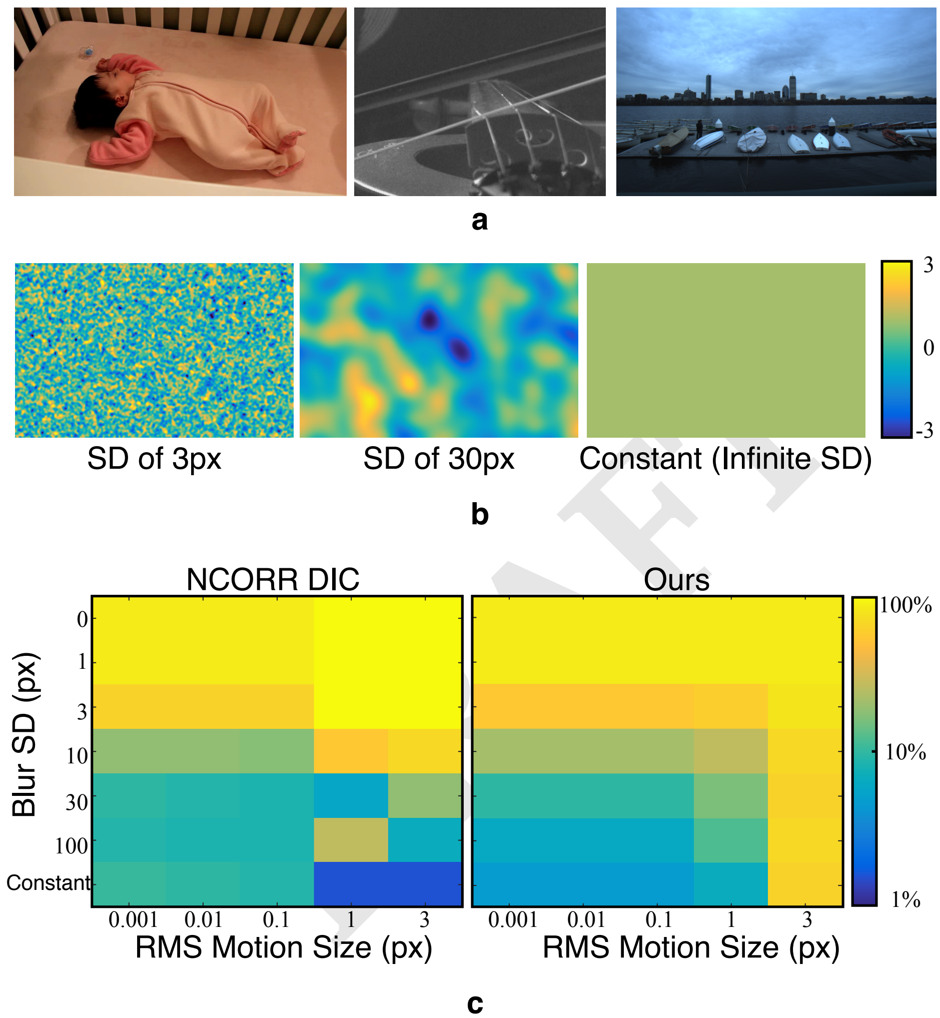


Fig. S5. An evaluation of our motion estimation method and NCORR (32) on a synthetic dataset of images. **a**, Sample frames from real videos used to create dataset. **b**, Sample of synthetic motion fields of various motion size and spatial scale used to create dataset. **c**, The motion microscope and NCORR are used to estimate the motion field and the average relative error is displayed for both methods as a function of motion size and spatial scale. Both methods are only accurate for spatially smooth motion fields. Our method is twice as accurate for spatially smooth, sub-pixel motion fields. NCORR is more accurate for larger motions.

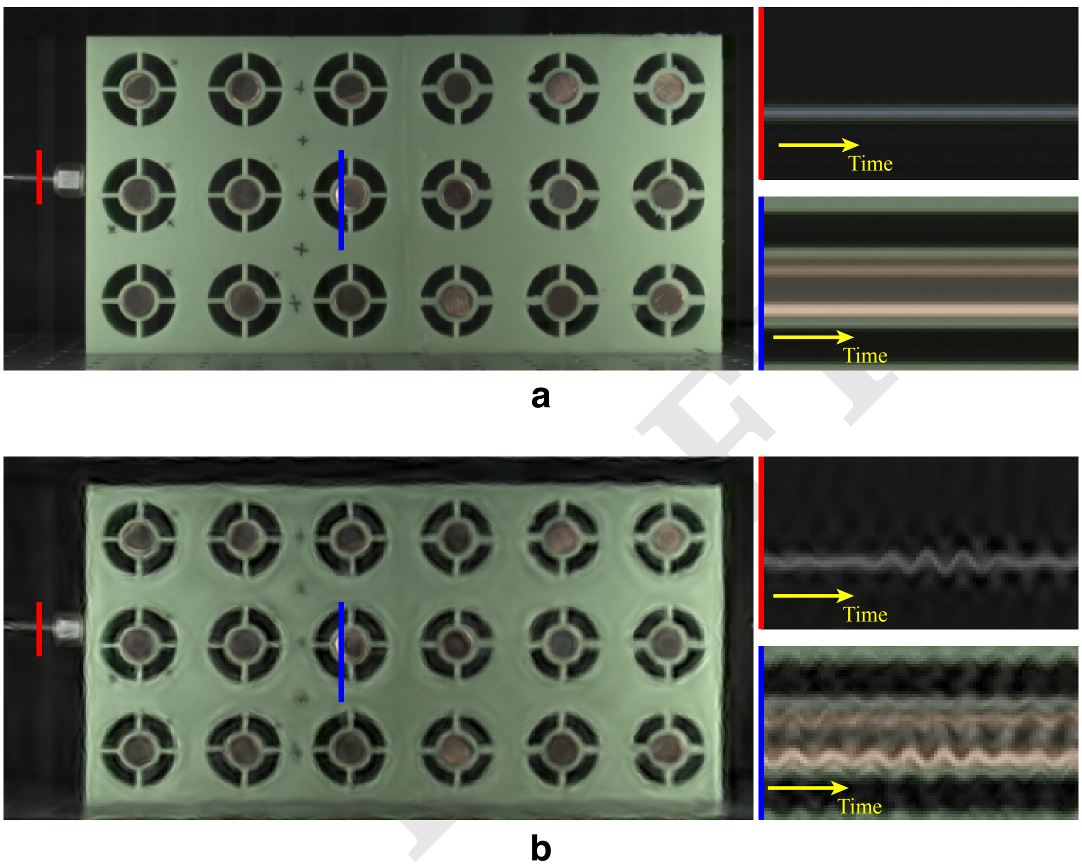


Fig. S6. Magnification of a spatial smooth and temporally filtered noise can look like real signal. **a**, Frames and time slices from a synthetic 300-frame video created by replicating a single frame 300 times and adding a different realistic noise pattern to each frame. **b**, Corresponding frame and time slices from the synthetic video motion magnified 600x in a temporal band of 40-60Hz. (b). Timeslices from the same parts of each video are shown on the right for comparison.

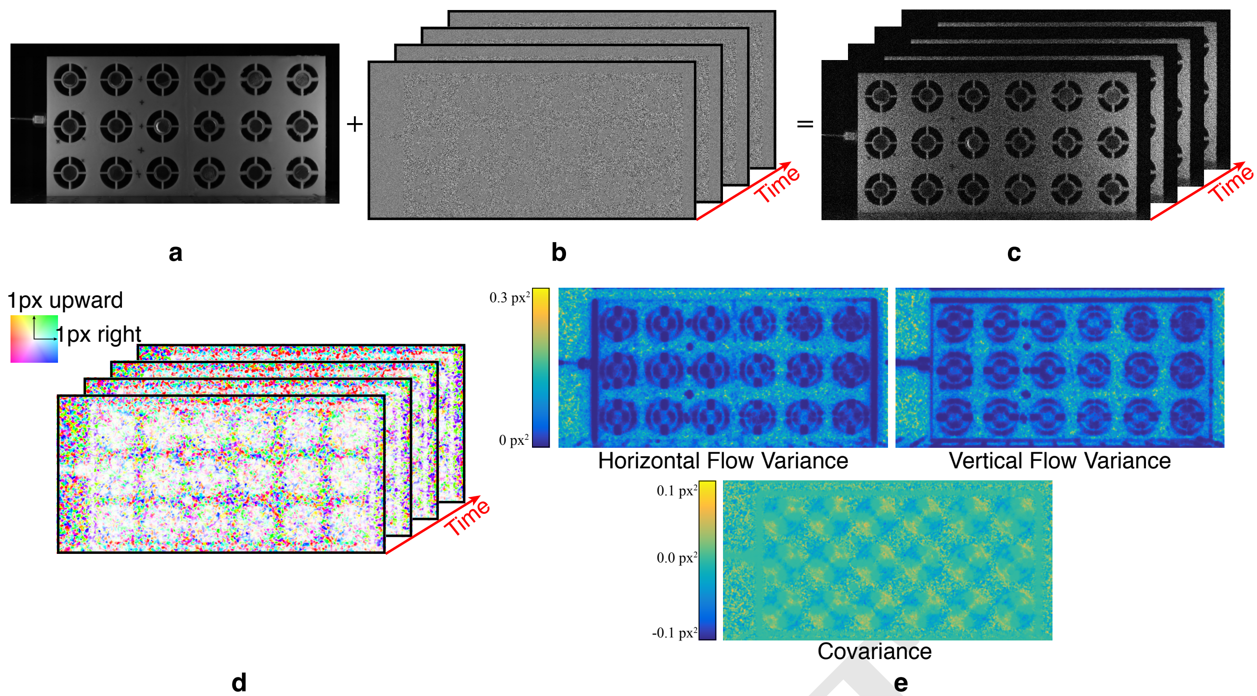


Fig. S7. Using a probabilistic simulation to compute the noise covariance of the motion estimate. **a**, A single frame from an input video, in this case of an elastic metamaterial. **b**, Simulated, but realistic noise (contrast enhanced 80x). **c**, Synthetic video with no motions consisting of the input frame replicated plus simulated noise (noise contrast-enhanced 80x). **d**, Estimated motions of this video. **e**, Sample variances and sample covariance of the vertical and horizontal components of the motion are computed to give an estimate of how much noise is in the motion estimate.

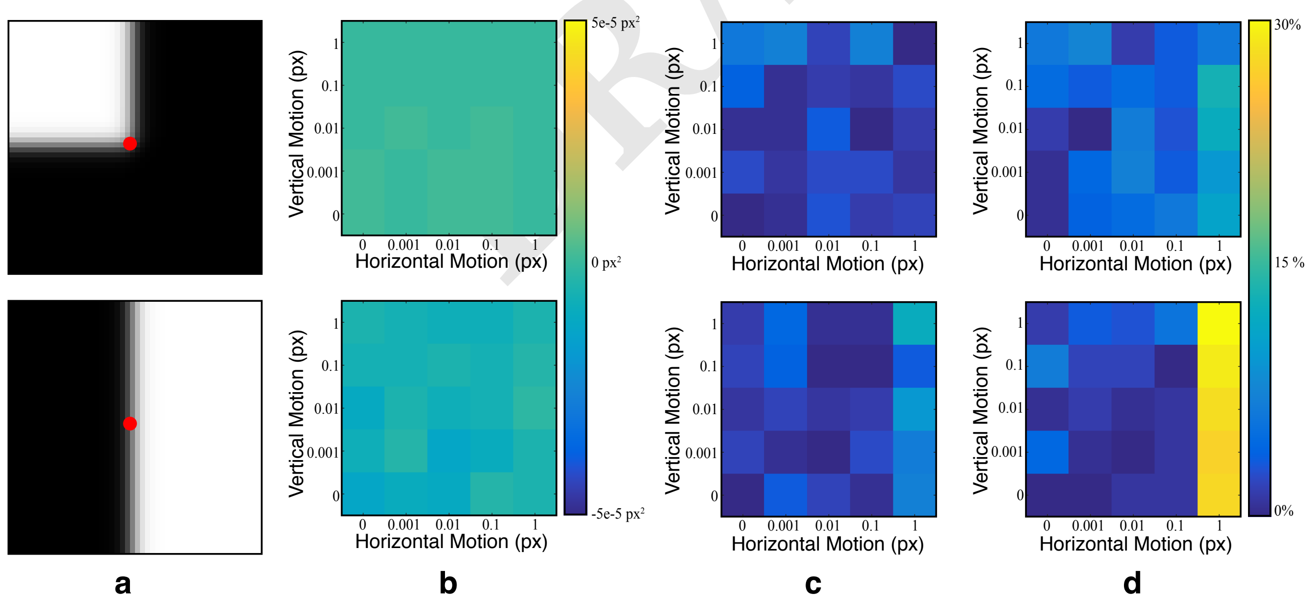


Fig. S8. Synthetic experiments showing that our noise covariance estimation, which assumes that the motions are zero, is also accurate for small non-zero motions. **a**, The motion between synthetic frames with noise and slightly translated versions (not shown) are computed over 4000 runs at the marked point in red for several different translation amounts. Each time different, but independent noise is added to the frames. **b**, The sample covariance vs. motion size. **c**, Relative error of horizontal variance vs. motion size. **d**, Relative error of vertical variance vs. motion size. **c** and **d** are on the same color scale.

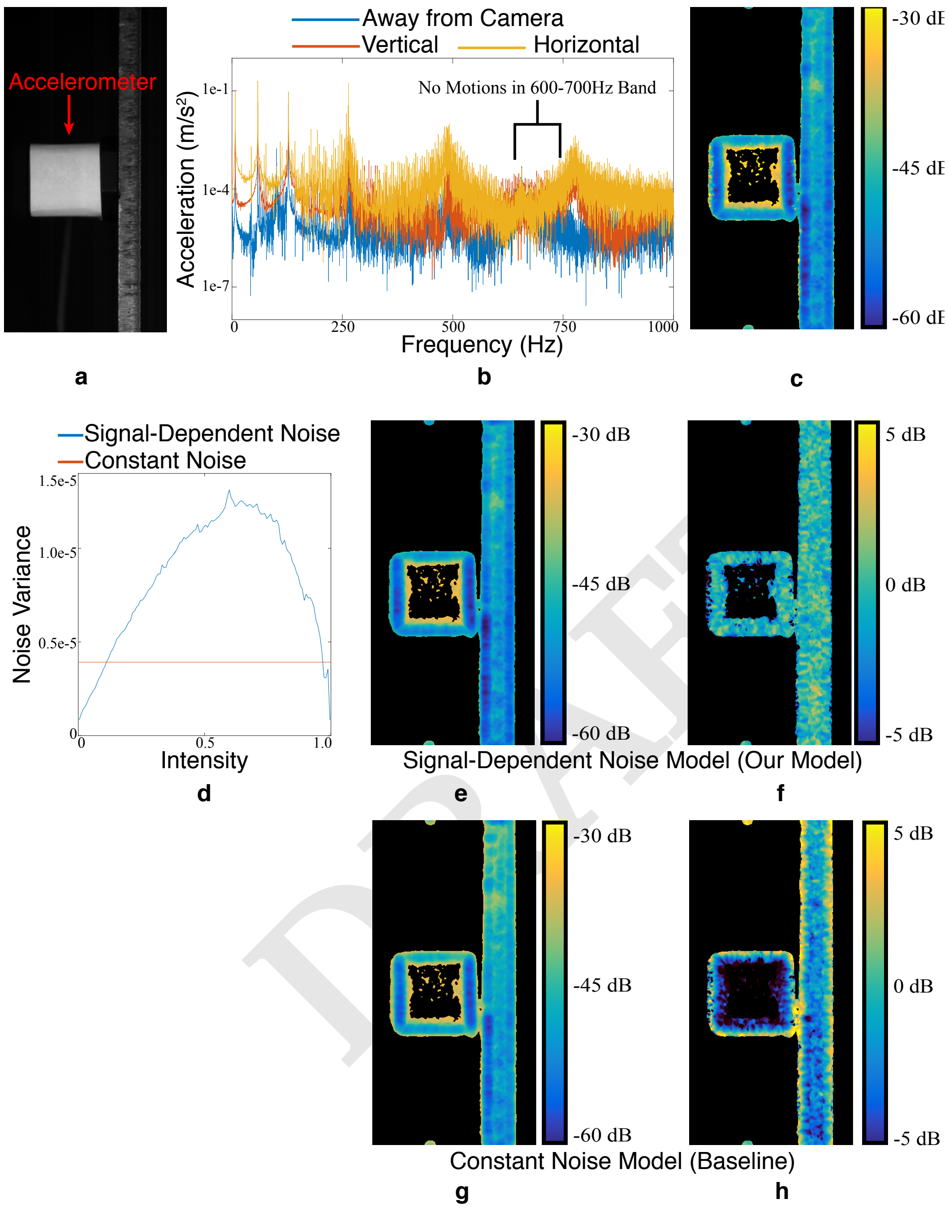


Fig. S9. Validation of our noise estimation on real data. **a**, A frame from a real video of an accelerometer attached to a beam. **b**, The accelerometer shows there are no motions in the frequency band 600 to 700 Hz. **c**, The variance of our motion estimate in the 600-700Hz band serves as a ground truth measure of noise as there are no motions. **d**, The estimated noise level vs. intensity for a signal-dependent noise model and a constant noise model. **e**, The noise estimate produced by our Monte Carlo simulation with a signal-dependent model. All variances are of the motions projected onto the direction of least variance. Textureless regions, where the motion estimation is not meaningful, have been masked out in black. **f**, Difference in decibels between ground truth and **e**. **g**, Noise estimate produced by the Monte Carlo simulation with a constant noise model. **h**, Difference in decibels between ground truth and **g**.

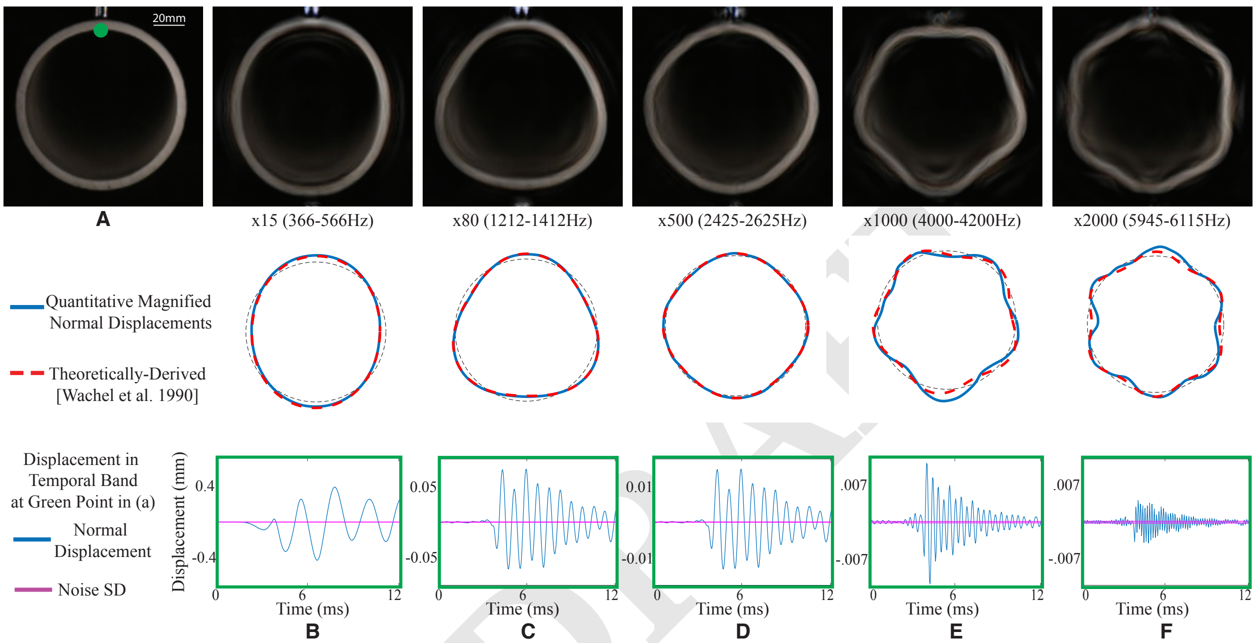


Fig. S10. The motion microscope is applied to a pipe being struck by a hammer. (**A**) A frame from the input video, recorded at 24,096 frames per second. (**B-F**) In the first row, a frame is shown from five motion magnified videos showing motions amplified in the specified frequency bands (Supplementary Video 3). In the second row, modal shapes recovered from a quantitative analysis of the motions are shown in blue. The theoretically-derived modal shapes, shown in dashed orange, are overlayed for comparison over a perfect circle in dotted black. In the bottom row, displacement vs. time and the estimated noise standard deviation is shown at the green point marked in **A**.

Supporting Information

1. Modal Shapes of a Pipe

We made a measurement of a pipe being struck by a hammer, viewed end on by a camera, to capture its radial-circumferential vibration modes. A standard 4" schedule 40 PVC pipe was recorded with a high-speed camera at 24,000 frames per second (fps), at a resolution of 192 × 192 (SI Appendix, Fig. S10a). SI Appendix, Fig. S10b-f shows frames from the motion magnified videos for different resonant frequencies showing the mode shapes, a comparison of the quantitatively measured mode shapes with the theoretically derived mode shapes, and the displacement vs. time of the specific frequency band and the estimated noise standard deviation. The tiny modal motions are seen clearly. Obtaining vibration data with traditional sensors with the same spatial density would be extremely difficult, and accelerometers placed on the pipe would alter its resonant frequencies.

This sequence also demonstrates the accuracy of our noise analysis. The noise standard deviations show that the detected motions prior to impact, when the pipe is stationary, are likely spurious.

2. Synthetic Validation

We validate the accuracy of our motion estimation on a synthetic dataset and compare its accuracy to NCORR, a digital image correlation technique (32) used by mechanical engineers (33). In this experiment, we did not employ temporal filtering.

We created a synthetic dataset of frame pairs with known ground truth motions between them. We took natural images from the frames of real videos (SI Appendix, Fig. S5a) and warped them according to known motion fields using cubic b-spline interpolation (34). Sample motions fields, shown in SI Appendix, Fig. S5b, were produced by Gaussian blurring IID Gaussian random variables. We used Gaussian blurs with standard deviations (SD), ranging from zero (no filtering) to infinite (a constant motion field). We also varied the root-mean-square (RMS) amplitude of the motion fields from 0.001px to 3px. For each set of motion field parameters, we sampled five different motion fields to produce a total of 155 motion fields with different amplitudes and spatial coherence. To test the accuracy of the algorithms rather than their sensitivity to noise, no noise was added to the image pairs.

We ran our motion estimation technique and NCORR on each image pair. We then computed the mean absolute difference between the estimated and ground truth motion fields. Then, for each set of motion field parameters, we averaged the mean absolute differences across image pairs and divided the result by the RMS motion amplitude to make the errors comparable over motion sizes. The result is the average relative error as a percentage of RMS motion amplitude (SI Appendix, Fig. S5c).

Both NCORR and our method perform best when the motions are spatially coherent (filter standard deviations greater than 10 px) with relative errors under 10%. This reflects the fact that both methods assume the motion field is spatially smooth. Across motion sizes, our method performs best for sub-pixel motions (5% relative error). This is probably because we assume that the motions are small when we linearize the phase constancy equation (Eq. 9). NCORR has twice the relative error (10%) for the same motion fields.

The relative errors reported in SI Appendix, Fig. S5c are computed over all pixels including those that are in smooth, textureless regions where it is difficult to estimate the motions. If we restrict the error metric to only take into account pixels at edges and corners, the average relative errors for small (< 1px RMS), spatially coherent (filter SD > 10px) motions drops by a factor of 2.5 for both methods.

We generated synthetic images that are slight translations of each other and added Gaussian noise to the frames (SI Appendix, Fig. S8a). For each translation amount, we compute the motion between the two frames over 4000 runs. We compute the sample covariance matrix over the runs as a measure of the ground truth noise level. We also used our noise analysis to estimate the covariance matrix at the points denoted in red.

The off-diagonal term of the covariance matrix should be zero for the synthetic frames in SI Appendix, Fig. S8a. For both examples, it is within 10^{-5}px^2 of zero for all translation amounts (SI Appendix, Fig. S8b).

The relative errors of the horizontal and vertical variances vs. translation (SI Appendix, Fig. S8c-d) are less than 5% for sub-pixels motions. This is likely due to the random nature of the simulation. For motions greater than one pixel, the covariance matrix has relative error of less than 25%.

3. Relation between Local Phase Differences and Motions

Fleet and Jepson have shown that contours of constant phase in image subbands such as those in the complex steerable pyramid approximately track the motion of objects in a video (7). We make a similar *phase constancy* assumption, in which the following equation relates the phase of the frame at time 0 to the phase of future frames:

$$\phi_{r,\theta}(x,y,0) = \phi_{r,\theta}(x-u(x,y,t), y-v(x,y,t), t), \qquad [7]$$

where $\mathbf{V}(x,y,t) := (u(x,y,t), v(x,y,t))$ is the motion we seek to compute. We Taylor-expand the right-hand side around (x,y) to get

$$\Delta\phi_{r,\theta} = \left(\frac{\partial\phi_{r,\theta}}{\partial x}, \frac{\partial\phi_{r,\theta}}{\partial y}\right)\cdot(u,v) + O(u^2,v^2), \qquad [8]$$

where $\Delta\phi_{r,\theta}(x,y,t) := \phi_{r,\theta}(x,y,t) - \phi_{r,\theta}(x,y,0)$, arguments have been suppressed and $O(u^2,v^2)$ represents higher-order terms in the Taylor expansion. Because we assume the motions are small, higher order terms are negligible and the local phase variations are approximately equal to only the linear term:

$$\Delta\phi_{r,\theta} = \left(\frac{\partial\phi_{r,\theta}}{\partial x}, \frac{\partial\phi_{r,\theta}}{\partial y}\right)\cdot(u,v). \qquad [9]$$

Fleet has shown that the spatial gradients of the local phase, $\left(\frac{\partial\phi_{r,\theta}}{\partial x}, \frac{\partial\phi_{r,\theta}}{\partial y}\right)$, are roughly constant within a subband and that they are approximately equal to the peak tuning frequency of the corresponding subband's filter (35). This frequency is a 2D vector oriented orthogonal to the direction the subband selects for, which means that the local phase changes only provide information about the motions perpendicular to this direction.

4. Low-Amplitude Coefficients have Noisy Phase

Each frame of the input video $I(x,y,t)$ is transformed to the complex steerable pyramid representation by being spatially bandpassed by a bank of quadrature pairs of filters $g_{r,\theta}$ and $h_{r,\theta}$, where r corresponds to different spatial scales of the pyramid and θ corresponds to different orientations. We use the filters of Portilla and Simoncelli, which are specified and applied in the frequency domain (26). For one such filter pair, the result is a set of complex coefficients $S_{r,\theta} + iT_{r,\theta}$ whose real and imaginary part are given by

$$S_{r,\theta} = g_{r,\theta} * I \text{ and } T_{r,\theta} = h_{r,\theta} * I \qquad [10]$$

where the convolution is applied spatially at each time instant t. This filter pair is converted to amplitude $A_{r,\theta}$ and phase $\phi_{r,\theta}$ by the operations

$$A_{r,\theta} = \sqrt{S_{r,\theta}^2 + T_{r,\theta}^2} \text{ and } \phi_{r,\theta} = \tan^{-1}(T_{r,\theta}/S_{r,\theta}) \qquad [11]$$

$g_{r,\theta}$ and $h_{r,\theta}$ are in quadrature relationship, which means that they select for the same frequencies, but are 90 degrees out of phase like sin and cos. A consequence is that they are uncorrelated and have equal root mean square (RMS) value. Complex coefficients at antipodal orientations are conjugate symmetric and contain redundant information. Therefore, we only use a half circle of orientations.

This transform has various properties that we don't use in this work such as perfect invertibility and steerability. Invertibility is used in motion magnification.

Suppose the observed video $I(x,y,t)$ is contaminated with independent and identically distributed (iid) noise $I_n(x,y,t)$ of variance σ^2:

$$I(x,y,t) = I_0(x,y,t) + I_n(x,y,t) \qquad [12]$$

where $I_0(x,y,t)$ is the underlying noiseless video. This noise causes the complex steerable pyramid coefficients to be noisy, which causes the local phase to be noisy. We show that the local phase at a point has an approximate Gaussian distribution when the amplitude is high and is approximately uniformly distributed when the amplitude is low.

The transformed representation has response

$$g_{r,\theta} * I_0 + g_{r,\theta} * I_n \text{ and } h_{r,\theta} * I_0 + h_{r,\theta} * I_n. \qquad [13]$$

The first term in each expression is the noiseless filter response, which we denote $S_{0,r,\theta} = g_{r,\theta} * I_0$ for the real part and $T_{0,r,\theta} = h_{r,\theta} * I_0$ for the imaginary part. The second term in each expression is filtered noise, which we denote as $S_{n,r,\theta}$ and $T_{n,r,\theta}$. At a single point, $S_{n,r,\theta}$ and $T_{n,r,\theta}$ are Gaussian random variables with covariance matrix equal to

$$\sigma^2 \begin{pmatrix} \sum_{x,y} g_{r,\theta}(x,y)^2 & \sum_{x,y} g_{r,\theta}(x,y) h_{r,\theta}(x,y) \\ \sum_{x,y} g_{r,\theta}(x,y) h_{r,\theta}(x,y) & \sum_{x,y} h_{r,\theta}(x,y)^2 \end{pmatrix} = \sigma^2 \sum_{x,y} g_{r,\theta}(x,y)^2 I \qquad [14]$$

where I is the identity matrix and equality follows from the fact that $g_{r,\theta}$ and $h_{r,\theta}$ are quadrature pairs.

We suppress the indices r, θ in this section for readability. From Eq. 11, the noiseless and noisy phase are given by

$$\phi_0 = \tan^{-1}(T_0/S_0) \text{ and } \phi = \tan^{-1}((T_0 + T_n)/(S_0 + S_n)). \tag{15}$$

Their difference linearized around (S_0, T_0) is

$$\tan^{-1}\left(\frac{T_0 + T_n}{S_0 + S_n}\right) - \tan^{-1}\left(\frac{T_0}{S_0}\right) = \frac{S_n S_0 - T_n T_0}{A_0^2} + O\left(\frac{S_n^2, S_n T_n, T_n^2}{A_0^4}\right). \tag{16}$$

The terms S_n^2 and T_n^2 are expected to be equal to their variance $\sigma^2 \sum g_{r,\theta}(x,y)^2$. Therefore, if $A_0^2 >> \sigma^2 \sum g_{r,\theta}(x,y)^2$, higher order terms are negligible. In this case, we see that the phase is approximately a linear combination of Gaussian random variables and is therefore Gaussian. This is illustrated empirically by local phase histograms of the green and blue points in Extended Data Fig. S3a-e.

For these high amplitude points, we compute the variance of the phase of a coefficient:

$$E\left[\left(\tan^{-1}\left(\frac{T_0 + T_n}{S_0 + S_n}\right) - \tan^{-1}\left(\frac{T_0}{S_0}\right)\right)^2\right] \tag{17}$$

$$\approx E\left[\left(\frac{T_0 S_n - S_0 T_n}{A_0^2}\right)^2\right] \tag{18}$$

$$= E\left[\frac{T_0^2 S_n^2 - 2T_0 S_0 S_n T_n + S_0^2 T_n^2}{A_0^4}\right] \tag{19}$$

$$= \frac{\sigma^2 \sum g_{r,\theta}^2 (T_0^2 + S_0^2)}{A_0^4} \tag{20}$$

$$= \frac{\sigma^2 \sum g_{r,\theta}^2}{A_0^2}. \tag{21}$$

The first approximation follows from the linearization of Eq. 16.

When the amplitude is low compared to the noise level ($A_0^2 << \sigma^2 \sum g_{r,\theta}(x,y)^2$), the linearization of Eq. 16 is not accurate. In this case, $S_0 \approx 0$ and $T_0 \approx 0$ and phase is given by

$$\tan^{-1}\left(\frac{T_n}{S_n}\right). \tag{22}$$

T_n and S_n are uncorrelated Gaussian random variables with equal variance, which means that the phase is a uniformly random number. The phase at such points contains no information and intuitively corresponds to places where there is no image content in a given pyramid level (Extended Data Fig. S3e, red point).

5. Noise Model and Creating Synthetic Video

We adopt a signal-dependent noise model, in which each pixel is contaminated with spatially independent Gaussian noise with variance $f(I)$ where I is the pixel's mean intensity (27)(36). Liu et al. (27) refer to this function f as a *noise level function* and we do the same. This reflects that sensor noise is well-modeled by the sum of zero-mean Gaussian noise sources, some of which have variances that depend on intensity (5). We show that this noise model is an improvement over a constant variance noise model in Extended Data Fig. S9.

The noise level function f is estimated from temporal variations in the input video, with observed intensities $I(x, y, t)$. Assuming that I is the sum of noiseless intensity I_0 and a zero-mean Gaussian noise term I_n with variance $f(I_0)$, the temporal variations are given by the following Taylor expansion

$$
\begin{aligned}
I(x,y,t) =& I_0(x,y,t) + I_n(x,y,t) && [23]\\
=& I_0(x-u(x,y,t), y-v(x,y,t), 0) + I_n(x,y,t) && [24]\\
\approx& I_0(x,y,0) - \frac{\partial I_0}{\partial x}u(x,y,t) - \frac{\partial I_0}{\partial y}v(x,y,t) + I_n(x,y,t). && [25]\\
& && [26]
\end{aligned}
$$

The second equality is the brightness constancy assumption of optical flow (37, 38). We exclude pixels where the the spatial gradient $\left(\frac{\partial I_0}{\partial x}, \frac{\partial I_0}{\partial y}\right)$ has high magnitude from our analysis. At the remaining pixel, temporal variations in I are mostly due to noise

$$I(x,y,t) \approx I_0(x,y,0) + I_n(x,y,t). \qquad [27]$$

At these pixels, we take the temporal variance and mean of I, which in expectation are $f(I_0)$ and I_0 respectively. To increase robustness, we divide the intensity range into 64 equally sized bins. For each bin, we take all those pixels with mean inside that bin and take the mean of the corresponding temporal variances of I to estimate the noise level function f.

With f in hand, we can take frames from existing videos and use them to create simulated videos with realistic noise, but with known, zero motion. In Extended Data Fig. S6a, we take a frame $I_0(x,y,0)$ from a video of the metamaterial, filmed with a Phantom V-10, and add noise to it via the equation

$$I_S(x,y,t) = I_0(x,y,0) + I_n(x,y,t)\sqrt{f(I_0(x,y,0))}, \qquad [28]$$

where I_n now is Gaussian noise with unit variance. We motion magnify the resulting video 600 times in a 20Hz band centered at 50Hz to show that motion magnified noise can cause spurious motions (Extended Data Fig. S6b).

We use the same simulation to create synthetic videos with which to estimate the covariance matrix of the motion vectors.

We quantify the noise in the motion vectors by estimating their covariance matrices $\Sigma_{\mathbf{V}}(x,y)$. These matrices reflect variations in the motion caused by noise. It is not usually possible to directly estimate them from the input video because both motions and noise vary across frames and the true motions are unknown. Therefore, we create a noisy, synthetic video $I_S(x,y,t)$ with known zero true motion (Eq. 28, Extended Data Fig. S7a-c).

We estimate the motions in I_S (Extended Data Fig. S7d) using our technique with spatial smoothing, but without temporal filtering, which we handle in a later step. This results in a set of 2D motion vectors $\mathbf{V}_S(x,y,t)$, in which all temporal variations in $\mathbf{V}_S$ are due to noise. The sample covariance matrix over the time dimension is

$$\Sigma_{\mathbf{V}} = \frac{1}{N-1}\sum_t \left(\mathbf{V}_S(x,y,t) - \bar{\mathbf{V}}_S(x,y)\right)\left(\mathbf{V}_S(x,y,t) - \bar{\mathbf{V}}_S(x,y)\right)^T \qquad [29]$$

where $\bar{\mathbf{V}}_S(x,y)$ is the mean over t of the motion vectors. $\Sigma_{\mathbf{V}}$ is a 2×2 symmetric matrix, defined at every pixel, with only three unique components. In Extended Data Fig. S7e, we show these components, the variances of the horizontal and vertical components of the motion and their covariance.

The motion $\mathbf{V}$ projected onto a direction vector $\mathbf{d}_\theta := (\cos(\theta), \sin(\theta))$ is $\mathbf{V} \cdot \mathbf{d}_\theta$ and has variance $\sigma_V^2(\theta) = \mathbf{d}_\theta^T \Sigma_{\mathbf{V}} \mathbf{d}_\theta$. Of particular interest is the direction θ of least variance that minimizes $\sigma_V^2(\theta)$. In the case of an edge in the image, the direction of least variance is usually normal to the edge.

6. Analytic Justification of Noise Analysis

We analyze only the case when the amplitudes at a pixel in all subbands are large ($A_{r,\theta} >> \sigma^2 \sum g^2$) because the local phases have a Gaussian distribution in this case. Such points intuitively correspond to places where there is image content in at least two directions. In this case, we show that the sample covariance matrix computed using a simulated video with *no* motions is accurate for videos with sub-pixel *small* motions.

We reproduce the linearization of phase constancy equation (Eq. 9) with noise terms added to the phase variations (n_t) and phase gradient (n_x, n_y):

$$\Delta\phi_{r,\theta} + n_t = (u, v) \cdot \left(\frac{\partial \phi_{r,\theta}}{\partial x} + n_x, \frac{\partial \phi_{r,\theta}}{\partial y} + n_y \right). \qquad [30]$$

The total noise term in this equation is $n_t + un_x + vn_y$. The noise terms n_t, n_x and n_y are of the same order of magnitude. Since u and v are much less than 1px, the predominant source of noise is from n_t and the effects of n_x and n_y are negligible and we can ignore them, allowing us to write the noisy version of the equation as

$$\Delta\phi_{r,\theta} + n_t = (u, v) \cdot \left(\frac{\partial \phi_{r,\theta}}{\partial x}, \frac{\partial \phi_{r,\theta}}{\partial y} \right). \qquad [31]$$

The motion estimate $\mathbf{V}$ is the solution to a weighted least squares problem, $\mathbf{V} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y}$ (Eq. 3). To simplify notation, let $\mathbf{B} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$, the parts of the equation that don't depend on time. Then, the flow estimate is

$$\mathbf{V} = \mathbf{B}\mathbf{Y}. \qquad [32]$$

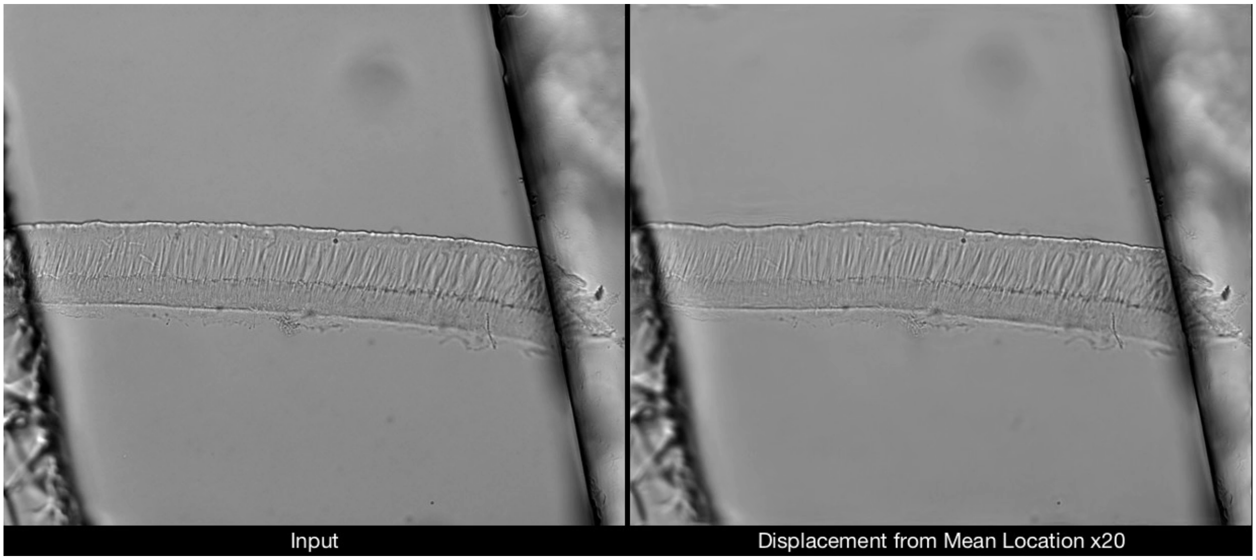
where the elements of $\mathbf{Y}$ are the local phase variations over time. $\mathbf{X}$ and $\mathbf{W}$ contains the spatial gradients of phase and amplitude respectively. We have demonstrated that $\mathbf{X}$ is close to noiseless (Eq. 31) and our assumption about the amplitudes being large means $\mathbf{W}$ is also approximately noiseless, which means that $\mathbf{B}$ is noiseless.

We split $\mathbf{Y}$ into the sum of its mean $\mathbf{Y}_0$ and variance, a multivariate Gaussian random variable, denoted as $\mathbf{Y}_n$, that has zero-mean and variance that depends only on image noise and local image content. Then, the flow estimate is

$$\mathbf{V} = \underbrace{\mathbf{B}\mathbf{Y}_0}_{\text{True flow}} + \underbrace{\mathbf{B}\mathbf{Y}_n}_{\text{Noise Term (Covariance Matrix)}} \qquad [33]$$

The noise term doesn't depend on the value of the true flow $\mathbf{BY}_0$. Therefore, the estimated covariance matrix is valid even when the motions are non-zero, but small.

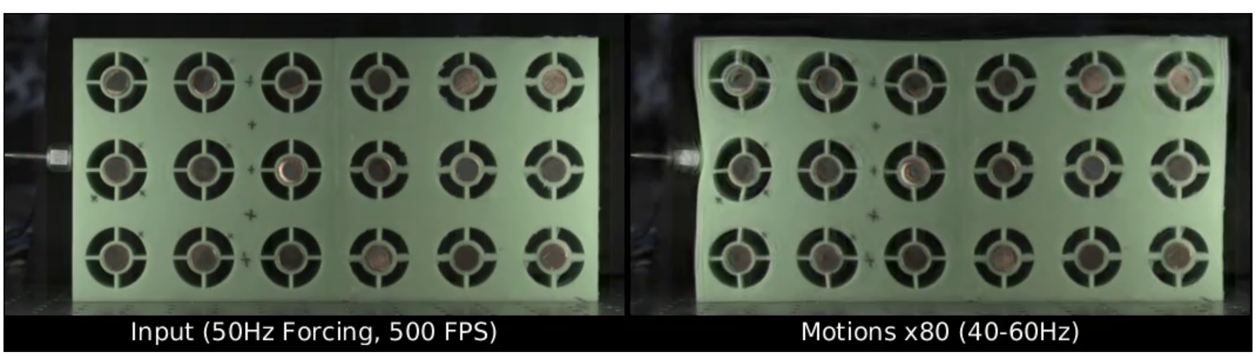
Movie Legends.



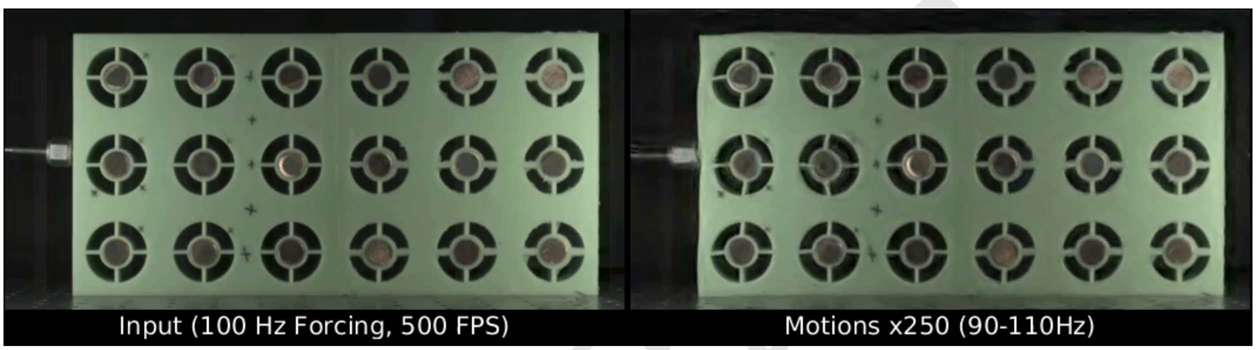
Movie S1. Traveling waves of the tectorial membrane revealed. The displacement from mean location of the membrane in the input video on the left was amplified by twenty times to produce the motion magnified video shown on the right. The original video consists of eight frames. The included video repeats these eight frames ten times for 80 frames and plays the result at 10 frames per second.

Movie S2. The input bridge video is concatenated with two motion magnified videos revealing different modal shapes of the bridge. Motions within a 1.6-1.8 Hz frequency band are amplified 400 times to produce the video on the right, in which the first bending mode is revealed. Motions within a 2.4-2.7 Hz frequency band are amplified 250 times to produce the video on the bottom, in which the first torsional mode is revealed. The impact of the central span (not shown in the video) occurs approximately five seconds after the video's start.

Movie S3. The input pipe video concatenated with five motion magnified videos revealing different modal shapes of the pipe. The sub-videos were scaled up using bicubic interpolation by a factor of 50%. The video was recorded at 24,096 FPS and played back at 30 FPS.



Movie S4. A probe vibrates the metamaterial at 50Hz. The input high speed video is shown on the left. Motions within a 40-60 Hz frequency band are amplified 80 times to produce the video on the right, in which the propagation of the vibrations is revealed. The video was recorded at 500 FPS and is played back at 30 FPS.



Movie S5. A probe vibrates the metamaterial at 100 Hz. The successful attenuation of vibrations are revealed in the motion magnified video on the right, in which motions in a frequency band of 90-110 Hz are amplified 250 Hz. The high-speed input video of the metamaterial is shown on the left. It was recorded at 500 FPS and is played back at 30 FPS.