Research Article | Vol. 14, No. 11 / 1 Nov 2023 / Biomedical Optics Express 5539
Biomedical Optics EXPRESS

Compressed sensing on displacement signals measured with optical coherence tomography

Brian L. Frost,[1,*] Nikola P. Janjušević,[2] C. Elliott Strimbu,[3] and Christine P. Hendon[1]

[1]*Department of Electrical Engineering, Columbia University, 500 W. 120th St., Mudd 1310, New York, NY 10027, USA*
[2]*New York University, Tandon School of Engineering, Electrical and Computer Engineering, 370 Jay St, Brooklyn, NY 11201, USA*
[3]*Columbia University, Department of Otolaryngology, 630 West 168th Street, New York, NY 10032, USA*
[*]*b.frost@columbia.edu*

Abstract: Optical coherence tomography (OCT) is capable of angstrom-scale vibrometry of particular interest to researchers of auditory mechanics. We develop a method for compressed sensing vibrometry using OCT that significantly reduces acquisition time for dense motion maps. Our method, based on total generalized variation with uniform subsampling, can reduce the number of samples needed to measure motion maps by a factor of ten with less than 5% normalized mean square error when tested on a diverse set of *in vivo* measurements from the gerbil cochlea. This opens up the possibility for more complex *in vivo* experiments for cochlear mechanics.

1. Introduction

Spectral domain optical coherence tomography (OCT) is a volumetric imaging modality with broad clinical and research applications [1]. OCT images are built of A-Scans – one-dimensional reflectivity maps along the device's optic axis. Formation of an image or volume requires sweeping the beam, acquiring A-Scans on a grid of points. As a result, large volumes built of many A-Scans take a substantial amount of time to acquire, and high SNR can only be achieved with averaging that further increases acquisition time. This leaves the imaging vulnerable to issues such as sample drift, or issues related to non-stationarity [2,3].

Compressed sensing imaging (CSI) has been developed to combat the acquisition time problem for images and volumes for many modalities by reducing the number of measurements required to achieve acceptable resolution [4–6]. Although CSI was originally formulated for acquisition in a transform domain incoherent to the imaging domain (e.g. Gaussian random matrices), "compressed sensing" has since been used more generally as a term encompassing sparsity-based signal recovery [7]. In this context, CSI has seen success in 3D volumetric OCT imaging [8,9] by employing either random or uniform image-domain subsampling schemes.

In addition to imaging, spectral domain OCT is also capable of vibrometry through spectral domain phase microscopy (SDPM) [10]. The A-Scan is generated by a Fourier transform of raw photodetector data, and its magnitude is proportional to the sample reflectivity (this is what is used for imaging), while the *phase* of the A-Scan is proportional to sub-pixel displacements. Taking many A-Scans at a single location over a period of time and tracking the phase, one can measure displacement responses with resolution on the order of an angstrom. This makes OCT particularly well-suited for the study of auditory mechanics, wherein structures in the inner ear all move at angstrom-to-nanometer scales to produce the sensation of hearing [11–17]. The application of compressed sensing to OCT-based displacement measurements has not yet been considered.

#503168 https://doi.org/10.1364/BOE.503168
Journal © 2023 Received 11 Aug 2023; revised 21 Sep 2023; accepted 22 Sep 2023; published 2 Oct 2023

Biomedical Optics EXPRESS

OCT vibrometry is significantly more time-consuming than imaging for two reasons: 1) measuring a displacement response requires acquiring many A-Scans at a single position over a period of time, and 2) displacements measured along an A-Scan are one-dimensional projections of the true three-dimensional motion onto the optical axis [10,14,18]. This second point means that complete description of three-dimensional motion can only be achieved if motion is measured at the same point at three or more angles, and then reconstructed from the resultant one-dimensional measurements [17,19,20]. This emphasizes the importance of developing compressed sensing vibrometry (CSVi).

Such a development is also timely; studies of recent interest in the field of cochlear mechanics use OCT to measure displacement responses at many locations in the cochlea over the course of a single experiment [14,18,19,21,22]. In *in vivo* experiments, time is restricted by the need to maintain animal anesthesia. This heavily limits the resolution and scale with which displacement maps can be acquired. CSVi could significantly reduce the amount of time required for such experiments, allowing for high resolution motion map acquisition, and opening the door to more complex experiments. For example, over-constrained 3-D motion reconstruction [19] or motion reconstruction in time-sensitive perturbation studies [13,21] are currently intractable, but could conceivably be performed in a reasonable amount of time using CSVi.

Recent years have seen a rise in deep-learning based image restoration and compressed sensing with state-of-the-art results [23,24]. These methods by and large rely on vast amounts of training data, with corresponding ground-truth acquisitions, and are often unable to perform reconstruction on samples outside of their training distribution (e.g. a change in instrument or subject) [25]. Optimization-based reconstruction methods are thereby favorable to learning based methods for the present application, as large *in vivo* OCT displacement map datasets are currently unfeasible to acquire. Optimization-based techniques remain especially popular in medical imaging contexts, wherein Total Generalized Variation regularization (see Section 4) and its derivatives comprise many state-of-the-art non-learned algorithms [26–28].

We provide a theoretical development of CSVi on OCT displacement scans, and provide evidence of its use in cochlear mechanics. We find that our method, based on a uniform subsampling pattern and total generalized variation optimization, can reduce acquisition time of dense areal motion maps in the cochlea by a factor of ten with less than 5% error. In addition, we argue that our method can be interpreted as a simultaneous denoiser by penalizing aphysical high-frequency components of displacement maps. This method is robust to stimulus frequency and level, as well as beam axis orientation with respect to the sample anatomy.

2. Background

2.1. Time-dependent OCT signals

Figure 1 shows the fundamental objects of interest for this study – OCT scans of the cochlea along with lineal and areal motion measurements. Panel **A** shows an anatomical drawing of a cross-section of the gerbil organ of Corti complex (OCC), the organ in the cochlea responsible for the transduction of fluid pressure stimuli into neurotransmitter.

The building block of OCT signals is the *A-Scan* – a complex-valued one-dimensional scan along the optic axis $\hat{\mathbf{z}}$ pointing into the sample. The magnitude of the A-Scan is related to the reflectivity of the sample at each depth, while the phase contains information about sub-pixel displacements. The nature of OCT is such that the values at all z-positions in an A-Scan are recorded simultaneously.

One can form a *B-Scan* by taking multiple A-Scans along a line segment. The magnitude of a B-Scan is a 2-D image, where pixel values encode sample reflectivity. Panel **B** in Fig. 1 shows a sample B-Scan of the OCC, along with a copy of this B-Scan with overlain markers of the anatomical structures derived from known anatomy (panel **A**). Shown also is the magnitude of a sample A-Scan, which makes up one vertical line in the B-Scan marked by a white dashed line.

Biomedical Optics EXPRESS

The axis along which A-Scans are swept to form a B-Scan is called $\hat{\mathbf{y}}$, so that a B-Scan occupies the *yz*-plane. A *C-Scan*, or volumetric scan, is formed by taking parallel B-Scans along a third orthogonal axis, $\hat{\mathbf{x}}$. The coordinates $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are referred to as *optical coordinates*, and they change dependent on the position and orientation of the scanner. We write positions in optical coordinates as $\mathbf{p}^{\theta}$, where θ is an index that describes the orientation of the scanner.

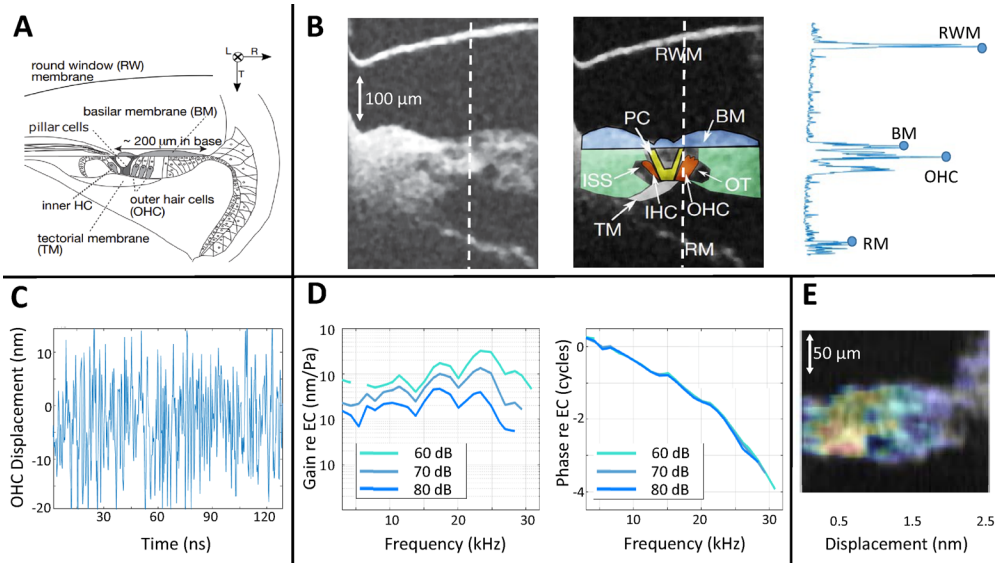

Fig. 1. Examples of OCT A-Scans, B-Scans and displacement measurements in the gerbil organ of Corti complex (OCC). **A** – Anatomical drawing of a cross-section of the OCC in gerbil, with important structures labelled. Axes at the top-right denote the longitudinal, radial and transverse axes of the cochlea. **B** – B-Scan of the gerbil cochlea taken *in vivo* through the round window membrane (RWM), 430 μm × 300 μm. The unlabeled B-Scan is recorded, and points are related to structures according to known anatomy (see panel **A**). The labeled B-Scan shows the RWM, basilar membrane (BM), outer hair cells (OHC), inner hair cells (IHC), pillar cells (PC), inner sulcus space (ISS), outer tunnel (OT), tectorial membrane (TM) and Reissner's membrane (RM). Each vertical line in the B-Scan is an A-Scan magnitude. A white dashed line denotes the position of a sample A-Scan, shown to the right. **C** – OCT-measured time-domain displacement at the OHC in response to a Zwuis stimulus at 80 dB SPL, *in vivo*. **D** – Frequency-domain gain re ear canal (EC) pressure at the OHC *in vivo*, in response to Zwuis stimuli at 60, 70 and 80 dB SPL. Only data that are at least 2.5 standard deviations above the noise level are shown. **E** – An areal map of the 25 kHz component of the OHC displacement magnitude *in vivo*, in response to an 80 dB SPL Zwuis stimulus. Scan is 200 μm × 200 μm.

Previous CSI work has focused on C-Scan *magnitudes* [8,9]. CSVi, on the other hand, concerns the phase. In SD-OCT these features are acquired simultaneously, but phase information can only be interpreted through acquisition of time-series of A-scans.

An *M-Scan* is formed by repeatedly taking A-Scans at a single location (x, y) over a period of time. We write the continuous M-Scan in polar form as

$$m(\mathbf{p}^{\theta}, t) = a(\mathbf{p}^{\theta}, t)e^{j\phi(\mathbf{p}^{\theta}, t)}, \tag{1}$$

where a and ϕ are the real magnitudes and phases of the M-Scan as a function of position and time. In practice, an M-Scan would be a sampled version of this signal.

The theory of SDPM gives a formula for the average sub-pixel displacement in terms of the phase of the M-Scan [10]. So long as the structures within a pixel are moving with displacements

Biomedical Optics EXPRESS

smaller than the pixel size (this assumption is always valid in cochlear mechanics, where displacements are usually on the order of nanometers), the displacements of the structures along the beam axis are given by

$$\delta^\theta(\mathbf{p}^\theta, t) = \frac{\lambda_c \phi(\mathbf{p}^\theta, t)}{4\pi n}, \tag{2}$$

where λ_c is the center wavelength of the OCT device and n is the refractive index of the sample. Fig. 1 panel **C** shows an example of such a displacement measurement.

Mathematically, δ^θ is a scalar field mapping points in 3-D space and time to 1-D real-valued displacements:

$$\delta^\theta : \mathbb{R}^4 \to \mathbb{R}. \tag{3}$$

This displacement is a 1-D projection of the true motion of the structure onto the optical $\hat{\mathbf{z}}$-axis. Generally speaking, a structure at a single point in space will have a 3-D displacement – that is, displacement is completely described by a vector field mapping points in space and time to a 3-D displacement vector:

$$\mathbf{d} : \mathbb{R}^4 \to \mathbb{R}^3. \tag{4}$$

It is prescient to discuss this object in terms of a global coordinate system rather than optical coordinates, as we will later consider the structure as being measured from more than one angle. Defining a general spatial coordinate $\mathbf{r} \in \mathbb{R}^3$, we can write the true displacement as

$$\mathbf{d}(\mathbf{r}, t) \in \mathbb{R}^3. \tag{5}$$

The measured displacement and true displacement are related by a projection map $\mathbf{P}^\theta$ that projects 3-vectors onto the beam axis. Similarly, the optical and global coordinates at orientation k are related by a change of basis matrix $\mathbf{R}^\theta$ and a translation $\mathbf{t}^\theta$ (as any two stationary coordinate systems in $\mathbb{R}^3$ are related). We write

$$\delta^\theta(\mathbf{R}^\theta[\mathbf{p}^\theta] + \mathbf{t}^\theta, t) = \mathbf{P}^\theta\left[\mathbf{d}(\mathbf{r}, t)\right] \in \mathbb{R}. \tag{6}$$

2.2. Frequency-dependent OCT signals

More often considered are the frequency domain versions of these displacement signals, which offer more clear information about the behavior of a system such as the cochlea. We define the temporal Fourier transforms of the true and measured displacement signals as

$$\mathcal{F}_t\{\mathbf{d}(\mathbf{r}, t)\} = \mathbf{D}(\mathbf{r}, f) \in \mathbb{C}^3. \tag{7}$$

$$\mathcal{F}_t\{\delta^\theta(\mathbf{p}^\theta, t)\} = \Delta^\theta(\mathbf{p}^\theta, f) \in \mathbb{C}, \tag{8}$$

Under an assumption of stationarity, these signals describe a pattern of simultaneous motion within the sampled structure. This stationarity assumption is generally justified so long as no sample deterioration or drift has occurred over the sampling period. This further motivates the developing of CSVi, as accelerated sampling time better ensures the accuracy of this assumption.

The frequency axis is discretized by the character of the sound stimulus. Displacements within the cochlea are usually measured in response to either a sweep of pure tone stimuli, or a multitone "Zwuis" stimulus [29]. In either case, a finite number of frequency components are presented, and we observe the displacement responses only at those stimulus frequencies.

For our specific application, we must also note that the cochlea is a nonlinear system – displacement responses differ in character dependent on the stimulus magnitude. Magnitude is referred to as sound pressure level (SPL). Its units, dB SPL, are defined as

$$SPL = 20 \log_{10} \frac{P}{P_0} \text{ dB SPL},$$

where P is the pressure and $P_0 = 20$ μPa is approximately the threshold of human hearing.

Biomedical Optics EXPRESS

Figure 1 panel **D** shows an OCT-measured frequency-domain response of the outer hair cells (OHC) to a Zwuis stimulus at 60, 70 and 80 dB SPL. The data are plotted as gains with respect to the pressure at the ear canal (EC), which is where the input stimulus is applied – that is, this resembles the "transfer function" of the system (of course, nonlinear systems do not actually have well-defined transfer functions). As the gain varies as a function of SPL, it is clear that the system is nonlinear.

The signal we wish to reconstruct is **D**. This must be done by measuring from $\Theta \geq 3$ orientations [17,19,20]. At each $\theta = 1, 2, \ldots \Theta$, we define $\hat{\Delta}^\theta$ to be the sampled version of the continuous complex scalar field Δ^θ. Supposing that at each orientation we sample at F frequencies at S SPLs along an $M \times N$ grid, and supposing that A-Scans contain L pixels along the z axis, the sampled signal is

$$\hat{\Delta}^\theta \in \mathbb{C}^{M \times N \times L \times F \times S}. \tag{9}$$

It should be noted that the process described in the present work is equally valid for areal scans (2-D scans where $N = 1$). While we will develop the most general 3-D case, the focus of the results will be on areal scans, which have been the subject of recent inquiry in cochlear mechanics [14,21]. Figure 1 panel **E** shows an example of the magnitude of such an areal motion map from the gerbil cochlea at 25 kHz and 80 dB SPL.

2.3. Dimensions of sampling and required generality

Equation (9) shows that there are *five* axes on which we sample during cochlear mechanics experiments – the x, y and z spatial coordinates, frequency and SPL. Our goal is to reduce the number of samples taken at experiment time, but it is not necessary to subsample along all of these axes.

As for frequency, all F frequency components are presented simultaneously if Zwuis multitone stimuli are used [29]. This means that sampling in frequency is not a time-limiting factor, so subsampling in frequency is unnecessary.

It is not standard in cochlear mechanics to present data as a function of SPL – generally, frequency-domain data is presented for a select hand-full of SPLs. SPL sampling is already relatively sparse, so its subsampling need not be considered.

As for the z axis, the nature of OCT is that the displacements at all z positions at a single (x, y) location are recorded simultaneously. This means that z sampling does not impact recording time, so we need not consider reducing the number of samples in z.

This leaves the x and y spatial domains in which to subsample, i.e. we look to compress the grid along which the M-Scans are taken. This means reducing the values of M and N in Eq. (9). At each SPL-orientation-frequency permutation, this is analogous to CSI for a complex volume.

Although orientation, frequency and SPL do not directly impact our subsampling scheme, they are still critical to consider in the validation of the CSVi method. Motion maps differ significantly between stimulus frequency, SPL and viewing angle [14,19,21,22]. Our method for CSVi must be sufficiently general to perform well for all frequency-orientation-SPL combinations.

3. Algorithms

We frame the compressed sensing vibrometry (CSVi) problem as complex-valued signal recovery problem. For notation, we consider our images in vectorized form, i.e. an $N_1 \times N_2$ complex-valued displacement map is a vector $\boldsymbol{x} \in \boldsymbol{C}^N$, with $N_1 \times N_2 = N$. Our observations are modeled as $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{\nu} \in \boldsymbol{C}^N$, where $\boldsymbol{M} = \mathbf{diag}(\boldsymbol{m}) \in \{0, 1\}^{N \times N}$ is a mask operator, $\boldsymbol{m}$ is a boolean vector which denotes if a pixel in the displacement map was collected, and $\boldsymbol{\nu}$ is unknown measurement error. Note that for CSVi $\boldsymbol{m}$ is a boolean image with columns removed as we consider a subsampling of A-scans only.

In the following subsections, we present signal models and associated algorithms for the recovery of displacement map $\boldsymbol{x}$ from a masked observation $\boldsymbol{y}$ via use of sparsity-based priors/regularizers.

Biomedical Optics EXPRESS

We choose to use priors that have seen success in reconstruction problems involving anatomical images. While we are not reconstructing the anatomy but rather the displacement map, these maps are constrained by the morphology of the OCC. Moreover, observed and expected smoothness of these maps suggests that gradient- and wavelet-based sparsity are reasonable priors [14,21].

The following algorithms perform optimization of real-valued objective functions of complex variables by considering the optimization jointly over real and imaginary parts of their argument. For neatness of presentation, this joint optimization is written with complex arithmetic. For example, the simple linear program over two variables

$$\min_{\boldsymbol{x}_1 \in \mathbb{R}^N,\, \boldsymbol{x}_2 \in \mathbb{R}^N} \boldsymbol{c}_1^T \boldsymbol{x}_1 + \boldsymbol{c}_2^T \boldsymbol{x}_2$$

may be written equivalently as

$$\min_{\boldsymbol{x} \in \mathbb{C}^N} \operatorname{Re}\{\boldsymbol{c}^H \boldsymbol{x}\}$$

where $\boldsymbol{x} = \boldsymbol{x}_1 + j\boldsymbol{x}_2$, $\boldsymbol{c} = \boldsymbol{c}_1 + j\boldsymbol{c}_2$, $\cdot^H$ denotes the conjugate transpose of a complex vector, and $\mathbf{R}e\cdot$ denotes the real part. Further details on optimization over complex variables can be found in [30].

3.1. Iterative shrinkage thresholding algorithm

In this section, we consider the signal prior of sparsity in a wavelet domain by performing a change of variables. Let $\boldsymbol{x} = \boldsymbol{W}\boldsymbol{z}$, where $\boldsymbol{W}$ is an (orthogonal) wavelet basis. This basis may be understood as representing a signal in terms of a single low-pass subband at a coarse image resolution, and several directional band-pass subbands at different resolutions. We then seek the wavelet coefficients $\boldsymbol{z}$ by using the sparsity-promoting ℓ_1 norm as a regularizer, often referred to as the Basis Pursuit DeNoising problem (BPDN),

$$\underset{\boldsymbol{z} \in \mathbb{C}^N}{\text{minimize}} \ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{W}\boldsymbol{z}\|_2^2 + \|\boldsymbol{\lambda} \circ \boldsymbol{z}\|_1. \tag{10}$$

Here, $\boldsymbol{\lambda}$ is a vector of hyperparamters for determining tradeoff between data-fidelity and sparsity in each subband of the wavelet transform. The element of $\boldsymbol{\lambda}$ corresponding to the low-frequency component of the wavelet coefficients is set to zero as there is no reasonable assumption of sparsity.

A popular algorithm for solving BPDN is proximal gradient descent (PGD), which tackles objectives of the form $f(\boldsymbol{x}) + g(\boldsymbol{x})$ with f smooth and g possibly non-smooth, involving the proximal operator g, defined as,

$$\mathbf{prox}_{\tau g}(\boldsymbol{v}) := \underset{\boldsymbol{x}}{\operatorname{argmin}}\ \tau g(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{v}\|_2^2, \tag{11}$$

where τ is some positive constant. PGD is then defined by the iteration $\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\eta g}(\boldsymbol{x}^{(k)} - \eta \nabla f(\boldsymbol{x}^{(k)}))$, with step-size parameter η. Convergence is guaranteed for $\eta \in (0, 2/L]$, where L is the Lipschitz-constant of ∇f [31].

When $g = \cdot_1$, the proximal operator has a closed form solution known as element-wise Soft-Thresholding,

$$\mathbf{ST}_\tau(z) := \operatorname{sign}(z)\,(|z| - \tau)_+\,, \tag{12}$$

where for complex $z = |z|e^{j\boldsymbol{\phi}}$, $\operatorname{sign}(z) = e^{j\boldsymbol{\phi}}$, and $(\cdot)_+$ denotes projection onto the positive orthant [32]. PGD applied to (10) is therefore known as the iterative shrinkage thresholding algorithm (ISTA) [33]. Algorithm 1 details the use of ISTA with a wavelet sparsity prior to estimate a displacement map $\hat{\boldsymbol{x}}$ from masked observation $\boldsymbol{y}$.

Biomedical Optics EXPRESS

Algorithm 1: Iterative Shrinkage Thresholding Algorithm (ISTA) for (10)

Input: masked observation $\boldsymbol{y}$, mask $\boldsymbol{m}$, trade-off parameter $\boldsymbol{\lambda}$;
Let: $z^{(0)} = \mathbf{0}$;
for $k = 0, 1, \ldots, \infty$ **do**
  $z^{(k+1)} := \mathbf{ST}_{2\boldsymbol{\lambda}}(z^{(k)} - 2\boldsymbol{W}^T\boldsymbol{M}(\boldsymbol{W}z^{(k)} - \boldsymbol{y}))$;
Output: $\hat{\boldsymbol{x}} = \boldsymbol{W}z^{(\infty)}$;

For our ISTA implementation, we chose to use a three-level Daubechies-7 wavelet transform. This choice was motivated by assessment of the sparsity of our data in various wavelet domains, as well as evaluation of the performance of ISTA when using 2-, 3- and 4-level wavelet transforms. More details can be found in the Supplemental Material.

3.2. Total variation

An alternative signal model is that of Total Variation (TV) regularization, in which the spatial gradient of the displacement map (in real and imaginary parts) is assumed to be sparse. We express this formulation as,

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2 + \|\boldsymbol{D}\boldsymbol{x}\|_{1,2} \tag{13}$$

where $\boldsymbol{D} = \begin{bmatrix}\boldsymbol{D}_v \\ \boldsymbol{D}_h\end{bmatrix}$ is a first order approximation of the spatial gradient, i.e.,

$$(\boldsymbol{D}_v\boldsymbol{x})[i,j] = \boldsymbol{x}[i,j] - \boldsymbol{x}[i-1,j] \tag{14}$$

$$(\boldsymbol{D}_h\boldsymbol{x})[i,j] = \boldsymbol{x}[i,j] - \boldsymbol{x}[i,j-1]. \tag{15}$$

The $\ell_{1,2}$ norm, $\|\boldsymbol{D}\boldsymbol{x}\|_{1,2} := \sum_{i,j} \left\|\begin{bmatrix}(\boldsymbol{D}_v\boldsymbol{x})[i,j] \\ (\boldsymbol{D}_h\boldsymbol{x})[i,j]\end{bmatrix}\right\|_2$, imposes an isotropic regularization by encouraging joint-sparsity in horizontal and vertical directions.

A efficient method for solving (13) can be found in Primal Dual Splitting (PDS, also known as the Chambolle-Pock Algorithm or the Primal Dual Hybrid Gradient Method) [34]. PDS tackles problems of the form

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} f(\boldsymbol{x}) + g(\boldsymbol{D}\boldsymbol{x}) \tag{16}$$

by forming the saddle-point problem,

$$\min_{\boldsymbol{x}} \min_{z} f(\boldsymbol{x}) + \mathrm{Re}\{z^H\boldsymbol{D}\boldsymbol{x}\} - g^*(z) \tag{17}$$

where g^* is the convex-conjugate of convex functional g, i.e. they are related by the identity $g(\boldsymbol{y}) = \sup_z \mathrm{Re}\boldsymbol{y}^H z - g^*(z)$. Consider, for the spatial gradient operator $\boldsymbol{D}$, $z = \begin{bmatrix}z_v \\ z_h\end{bmatrix}$. When g is a norm, g^* is the indicator function of the dual-norm ball. For $g = \lambda\|\cdot\|_{1,2}$,

$$g^*(z) = i_{\lambda\boldsymbol{B}_{\infty,2}}(z) = \begin{cases} 0, & \boldsymbol{m}ax_{i,j}\left\|\begin{bmatrix}z_v[i,j] \\ z_h[i,j]\end{bmatrix}\right\|_2 \le \lambda \\ \infty, & \text{otherwise}\end{cases}, \tag{18}$$

i.e. the indicator function of the $\ell_{\infty,2}$-norm ball with radius λ [34].

Biomedical Optics EXPRESS

The PDS algorithm can then be written as taking alternating proximal gradient descent and proximal gradient ascent steps on the primal and dual variables, $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively,

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\tau f}\left(\boldsymbol{x}^{(k)} - \tau \boldsymbol{D}^T z)\right) \tag{19}$$

$$\tilde{\boldsymbol{x}}^{(k+1)} := \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \tag{20}$$

$$z^{(k+1)} := \mathbf{prox}_{\sigma g^*}\left(z^{(k)} + \sigma \boldsymbol{D}\tilde{\boldsymbol{x}}^{(k+1)})\right) \tag{21}$$

where $\theta \in [0, 1]$ is an extrapolation-step hyperparamter, required to be non-zero for convergence when f is non-smooth, and τ, σ are step-size hyperparameters which satisfy $\tau\sigma\boldsymbol{D}_2 \leq 1$.

Returning to the CSVi TV regularization problem, our saddle-point problem is,

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \max_{z\in\mathbb{C}^{2N}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2 + \mathrm{Re}\{z^H \boldsymbol{D}\boldsymbol{x}\} - i_{\lambda\mathcal{B}_{2,\infty}}(z). \tag{22}$$

The proximal operator of the data-fidelity term can be found in closed form as,

$$\mathbf{prox}_{\tau f}(\boldsymbol{v}) := \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{C}^N} \frac{\tau}{2}\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{v}\|_2^2 = \frac{\tau\boldsymbol{y} + \boldsymbol{v}}{1 + \tau\boldsymbol{m}}. \tag{23}$$

The proximal operator of the indicator function of a convex set $\mathbb{C}$ is given by projection onto said set, $\Pi_{\mathbb{C}}$. Together, PDS for the TV-regularized CSVi problem is given in Algorithm 2.

Algorithm 2: Primal-Dual Splitting (PDS) for (22)

Input: masked observation $\boldsymbol{y}$, mask $\boldsymbol{m}$, trade-off parameter λ ;
Let: $z^{(0)} = \mathbf{0}, \boldsymbol{x}^{(0)} = \boldsymbol{y}, \tau = \sigma = \frac{1}{\sqrt{8}}$;
for $k = 0, 1, \ldots, \infty$ **do**
  $\boldsymbol{x}^{(k+1)} := \frac{\boldsymbol{x}^{(k)} + \tau(\boldsymbol{y} - \boldsymbol{D}^T z^{(k)})}{1+\tau\boldsymbol{m}}$;
  $\tilde{\boldsymbol{x}}^{(k+1)} := \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$;
  $z^{(k+1)} := \Pi_{\lambda\boldsymbol{B}_{\infty,2}}\left(z^{(k)} + \sigma\boldsymbol{D}\tilde{\boldsymbol{x}}^{(k+1)}\right)$;
Output: $\hat{\boldsymbol{x}} = \boldsymbol{x}^{(\infty)}$;

4. Total generalized variation

TV regularization assumes a piecewise constant signal model, which is ill-suited for describing the complex internal motions of the organ of Corti. A more sophisticated signal prior known as Total Generalized Variation (TGV) [34], in which the sparse spatial gradient prior is proposed in a hierarchical manner, offers a greater ability to model cochlear displacement. In the following, we consider the second order TGV prior (TGV-2). We express this formulation as,

$$\min_{\boldsymbol{x}\in\mathbb{C}^N,\, \boldsymbol{v}\in\mathbb{C}^{2N}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2 + \lambda_1\|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{v}\|_{1,2} + \lambda_0\|\boldsymbol{K}\boldsymbol{v}\|_{1,2} \tag{24}$$

where $\boldsymbol{K}$ is the spatial gradient operator applied to each channel of $\boldsymbol{v}$ independently, i.e.

$$\boldsymbol{K}\boldsymbol{v} = \begin{bmatrix} \boldsymbol{D} & 0 \\ 0 & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_v \\ \boldsymbol{v}_h \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_v\boldsymbol{v}_v \\ \boldsymbol{D}_h\boldsymbol{v}_v \\ \boldsymbol{D}_v\boldsymbol{v}_h \\ \boldsymbol{D}_h\boldsymbol{v}_h \end{bmatrix}. \tag{25}$$

The ratio of λ_0 to λ_1 play a role in balancing the importance placed on first and second spatial derivatives of the displacement map in the objective function, respectively.

Biomedical Optics EXPRESS

A saddle point problem amenable to PDS can be formed as

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathbb{C}^N,\,\boldsymbol{v}\in\mathbb{C}^{2N}}\ \max_{\boldsymbol{z}\in\mathbb{C}^{2N},\,\boldsymbol{q}\in\mathbb{C}^{4N}}\ & \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{M}\boldsymbol{x}\|_2^2+\mathrm{Re}\{\boldsymbol{z}^H(\boldsymbol{D}\boldsymbol{x}-\boldsymbol{v})\}\\ &+\mathrm{Re}\{\boldsymbol{q}^H\boldsymbol{K}\boldsymbol{v}\}-i_{\lambda_1\mathcal{B}_{\infty,2}}(\boldsymbol{z})-i_{\lambda_0\mathcal{B}_{\infty,2}}(\boldsymbol{q}).\end{aligned}\tag{26}$$

The corresponding PDS algorithm for CSVi with a TGV-2 prior (26) is given in Algorithm 3. This algorithm is derived in an identical manner to the TV PDS algorithm (Alg. 2), wherein proximal gradient descent is performed on the primal variables $(\boldsymbol{x},\boldsymbol{v})$ with dual variables $(\boldsymbol{z},\boldsymbol{q})$ fixed, then proximal gradient ascent is performed on the duals with primals fixed.

Algorithm 3: Primal-Dual Splitting (PDS) for (26)

Input: masked observation $\boldsymbol{y}$, mask $\boldsymbol{m}$, trade-off parameters $\lambda_1,\ \lambda_0$;
Let: $\boldsymbol{z}^{(0)}=\boldsymbol{v}^{(0)}=\boldsymbol{0}$, $\boldsymbol{u}^{(0)}=\boldsymbol{0}$, $\boldsymbol{x}^{(0)}=\boldsymbol{y}$, $\tau=\sigma=\frac{1}{\sqrt{12}}$;
for $k=0,1,\ldots,\infty$ **do**
  $\boldsymbol{x}^{(k+1)}:=\frac{\boldsymbol{x}^{(k)}+\tau(\boldsymbol{y}-\boldsymbol{D}^T\boldsymbol{z}^{(k)})}{1+\tau\boldsymbol{m}}$;
  $\boldsymbol{v}^{(k+1)}:=\boldsymbol{v}^{(k)}-\tau\left(\boldsymbol{K}^T\boldsymbol{q}^{(k)}-\boldsymbol{z}^{(k)}\right)$;
  $\tilde{\boldsymbol{x}}^{(k+1)}:=\boldsymbol{x}^{(k+1)}+\theta(\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)})$;
  $\tilde{\boldsymbol{v}}^{(k+1)}:=\boldsymbol{v}^{(k+1)}+\theta(\boldsymbol{v}^{(k+1)}-\boldsymbol{v}^{(k)})$;
  $\boldsymbol{z}^{(k+1)}:=\Pi_{\lambda_1\boldsymbol{B}_{\infty,2}}\left(\boldsymbol{z}^{(k)}+\sigma\left(\boldsymbol{D}\tilde{\boldsymbol{x}}^{(k+1)}-\tilde{\boldsymbol{v}}^{(k+1)}\right)\right)$;
  $\boldsymbol{q}^{(k+1)}:=\Pi_{\lambda_0\boldsymbol{B}_{\infty,2}}\left(\boldsymbol{q}^{(k)}+\sigma\boldsymbol{K}\tilde{\boldsymbol{v}}^{(k+1)}\right)$;
Output: $\hat{\boldsymbol{x}}=\boldsymbol{x}^{(\infty)}$;

The presented algorithms were implemented in the Julia programming language and run on an Intel i5-8250U 3.4 GHz CPU with 8 threads. For a single reconstruction, ISTA and TV algorithms complete in roughly 45 seconds, and TGV completes in roughly 90 seconds.

5. Methods

5.1. Experimental preparation

The experiments were approved by the Columbia University Institutional Animal Care and Use Committee. Young adult gerbils were anesthetized, their scalps were removed and their heads were attached to a two-axis goniometer. Measurements were taken at the base of the cochlea through the left ear's round window membrane using a Thorlabs Telesto 311C spectral domain OCT device, equipped with an LSM-03 lens. The device has a 1300 nm center wavelength, an axial resolution of 8 μm and a lateral resolution of 10 μm. Acquisition and initial processing of the OCT data was performed using the Thorlabs SpectralRadar software development kit. The temporal sampling rate of M-Scan recordings is approximately 100 kHz.

The cochlea was stimulated via one-second Zwuis multi-tone stimuli [29], generated by a Tucker Davis Technologies system and delivered closed-field to the animal's EC using a plastic tube. The EC pressure, used as a reference for gain and phase responses, was measured using a Sokolich ultrasonic microphone with its probe tube stationed 1-2 mm from the tympanic membrane. Further details of the animal preparation and acoustical system can be found in *Strimbu et al., 2020* [13].

The *in vivo* data set consists of 275 areal motion maps from three animals taken at very distinct orientations relative to the cochlea's anatomy, together representing the three most common orientations considered in cochlear mechanics experiments. They are:

Biomedical Optics EXPRESS

$\theta = 1$: Data taken near the 24 kHz region taken along the longitudinal axis of the cochlea through the OHC region (less technically, along the cochlea's spiral orthogonal to the B-Scan in Fig. 1 **B**). The cochlea was stimulated with Zwuis stimuli containing 25 frequency components between 2 kHz and 30 kHz at 60, 70 and 80 dB SPL. That is, there are 75 maps from this animal. Each map is 270 μm $\times$ 300 μm, containing 100 rows and 200 columns.

$\theta = 2$: Data taken near the 24 kHz region along the radial axis at an oblong angle with respect to the transverse direction of the cochlea (less technically, a skewed version of the anatomical cross-section in Fig. 1 **A**). The cochlea was stimulated with Zwuis stimuli containing 25 frequency components between 2 kHz and 30 kHz at 50, 60, 70 and 80 dB SPL. That is, there are 100 maps from this animal. Each map is 270 μm $\times$ 300 μm, containing 100 rows and 200 columns.

$\theta = 3$: Data taken near the 40 kHz region in a transverse-radial plane (less technically, the exact cross-section represented in Fig. 1 **A**). The cochlea was stimulated with Zwuis stimuli containing 25 frequency components between 2 kHz and 50 kHz at 50, 60, 70 and 80 dB SPL. That is, there are 100 maps from this animal. Each map is 270 μm $\times$ 330 μm, containing 100 rows and 330 columns.

5.2. Pre-processing

Data were pre-processed on an M-Scan by M-Scan basis, as neighboring M-Scan information is not available in a sparsely sampled signal (i.e. areal maps are pre-processed column-by-column). The noise floors of the M-Scans were estimated, and M-Scan points were replaced with 0 if they were less than 2.5 standard deviations above the noise floor. The M-Scans were then median filtered (three-pixel kernel, where the median is selected by magnitude) in the z dimension to remove outliers.

5.3. Evaluation

We look to compare the success of TV, ISTA and TGV using either uniform or random subsampling. We consider three subsampling rates, where the number of samples is reduced by a factor of $P = 2$, 5 or 10. For random subsampling, we average over ten realizations of the sampling mask.

We begin by considering the performance of these six methods on a test set of 20 maps chosen from the full dataset at random. We then measure the performance of the best tested method on the entire dataset of 275 maps.

The reconstructions are compared on two metrics: normalized mean square error (NMSE) between the complex reconstruction and densely sampled signals, and structural similarity index (SSIM) between the magnitudes of the reconstruction and the densely sampled signals. For each tested displacement map, reconstruction method, subsampling paradigm and subsampling rate, a grid search was used to determine the parameters which give the optimal NMSE, and these values were used to compute the reported NMSE and SSIM.

5.4. Visualization of areal motion maps

The complex-valued areal motion maps in Figs. 3 and 5 are represented using cyclic colormaps with varying saturation. Colorwheels are present in the first panel of both figures.

The hue, representing the phase re EC pressure at each pixel, varies moving counter-clockwise around the colorwheel in the standard manner. That is, the hue at the right-hand horizontal represents a point moving in-phase with EC pressure, while a point at the top vertical represents a point +0.25 cycles out of phase with EC pressure.

The saturation represents gain magnitude at each pixel normalized to the maximum. It varies linearly from the center (black, no motion) to the outer edge of the colorwheel (most saturated, maximum gain magnitude across the areal map).

Biomedical Optics EXPRESS

6. Results

6.1. Comparison of methods

We first compared TV, ISTA and TGV over a 20-map test set as described in Sec 5.3. The mean NMSEs and SSIMs, along with the respective standard deviations, across this test set are shown in Fig. 2.

The results show that across this test set, uniform sampling consistently provides better results than random sampling on both the NMSE and SSIM metrics for all tested methods and subsampling rates. Across the methods, TGV performs better than TV and ISTA. Within one standard deviation of the mean for this test set at $P = 10$, uniform TGV gives an NMSE of less than 3.5% and the SSIM exceeds 0.95.

This is also true qualitatively, as shown in the representative example of Fig. 3 ($P = 10$). The top row shows results for uniform subsampling. ISTA grossly reconstructs the signal well at its lowest-frequency (in a spatial sense) regions, but performs poorly near rapid changes between low- and high-amplitude positions with varying phases.

Physiologically, the bright region on the left-hand-side is at the junction between OHCs and Deiters cells sometimes referred to as the hotspot [14]. It is one of the regions of most interest in these motion maps, so poor performance here is not acceptable.

The TV reconstruction, which uses a piecewise-constant prior, is not as smooth as the densely sampled signal. The TGV signal, on the other hand, performs well at all positions in the OCC. The NMSE between this reconstruction and the densely sampled signal is below 1%.

As for the random sampling pattern, all reconstructions fail to resemble the densely sampled signal well. Random sampling patterns may include large breadths of columns that are not sampled, which we refer to as *bands*. Bands are results of the structured nature of OCT subsampling, where samples are removed column-by-column rather than pixel-by-pixel.

The reconstruction performs most poorly at these bands. ISTA reduces the signal value to zero at the centers of these bands, while TV performs nearly constant interpolation across them. TGV still outperforms ISTA and TV, but still shows significant artifacts at the largest bands. This further motivates the choice of TGV with uniform subsampling as the method of further evaluation.

6.2. Uniform TGV performance

We have shown that on the test set, TGV with uniform subsampling outperforms TV and ISTA, as well as any method with random subsampling. We now evaluate this method on all 275 maps within our data set.

Figure 4 shows NMSE and SSIM values for uniform TGV with three subsampling rates. Average values and standard deviations are computed and presented according to animal. For all animals, at one standard deviation above the mean, NMSE does not exceed 4.1% and SSIM exceeds 0.9 at even $P = 10$. At $P = 5$, NMSE does not exceed 1%. This shows quantitatively that uniform TGV performs well with generality in frequency, SPL and orientation.

To show that this is qualitatively true as well, we show one map from each animal in Fig. 5 with a subsampling rate of $P = 10$. The example maps are chosen to be at different amplitudes and frequencies so that the generality of the method can be assessed. For all three example maps, TGV performs qualitatively well. Reconstructions are slightly smoother than the original signal, but still maintain the same phase shifts and hotspot behavior as in the original map. Some amount of smoothing may be more physical than the densely sampled motion maps measured by OCT, which suffer from issues such as shot noise, signal competition and blurring by an anisotropic point spread function [1,35]. As such, the NMSE is an over-estimate for the error between our signal and the ground truth, as the signal to which we are referencing is inherently noisy. Further interpretation of CSVi as a denoiser is discussed in Sec 7.1.

Biomedical Optics EXPRESS

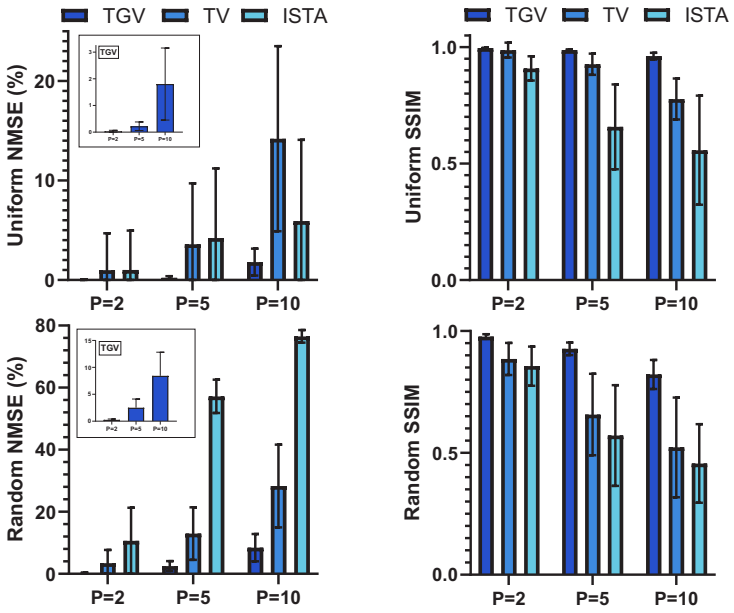

Fig. 2. Normalized mean square error and magnitude SSIM for the three tested methods: TGV, TV and ISTA. Values displayed are sample means over a test set (N = 20), and error bars represent one sample standard deviation from the mean. Results are compared for both uniform and random subsampling by a factor of P = 2, 5 and 10. The inset in the NMSE panels shows the results for TGV on a smaller set of axes, as the error is far lower than those achieved using TV and ISTA. TGV with uniform subsampling is seen to significantly outperform all other methods – using this method at P = 10, the mean NMSE is less than 2%.

7. Discussion and conclusions

We have developed a theoretical formulation for compressed sensing for OCT-measured displacement measurements. We have tested six possible methods for CSVi on *in vivo* cochlear mechanics motion maps, and have found that TGV with uniform subsampling is capable of reconstructing densely sampled maps from only 10% of the samples with less than 5% error. Our use of uniform A-scan subsampling is in contrast to Lebed *et al.* [8], which uses random subsampling of A-scans in their OCT volume sparse recovery. We note that, at high subsampling ratios, Lebed *et al.*'s reconstructions contain similar banding artifacts to those shown in Fig. 3 H, I, J. Our results suggest that uniform subsampling may also improve 3D volume recovery.

The method is also shown to function well at various beam axis orientations, stimulus frequencies and stimulus amplitudes, showing that it is sufficiently general to be of use for any experimental setup. This is the first example of compressed sensing being applied to the phase of the OCT signal.

7.1. Denoising interpretation of CSVi

The objective functions chosen do not force the recovered signal to be equal to the observation at any point, instead rewarding nearness to the observation in an ℓ_2 sense. TGV will also discourage high-frequency phenomena, as they will incur large second derivatives. Our method can be interpreted as simultaneously performing compressed sensing and denoising.

Reduction of phase noise in OCT is challenging *in vivo*. In theory, one could reduce noise by averaging recordings over a longer period of time, but this incurs two issues in practice: 1)

Biomedical Optics EXPRESS

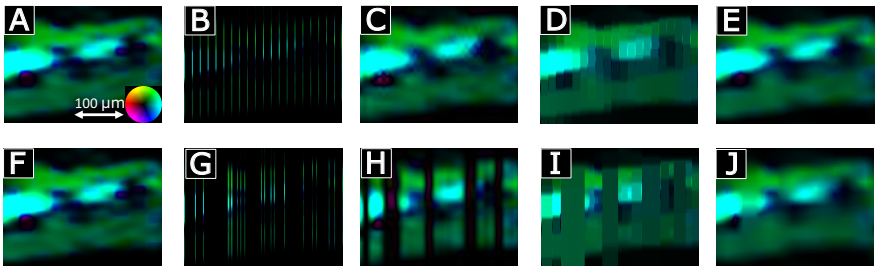

Fig. 3. Example reconstructions using 10% of the M-Scans from the gerbil OCC. Colormap is shown in **A**, with hue representing phase re EC and saturation representing gain normalized to the maximum (further described in Sec 5.4). Sample: $\theta = 2$, 80 dB SPL, 10 kHz component. Maps are 100 rows by 200 columns, or 270 μm × 300 μm. Data are further described in Sec 5.1. Top Row: Results for uniform subsampling. **A** – densely sampled motion map; **B** – map from **A** subsampled uniformly by a factor of 10; **C** – dense map reconstructed using ISTA (2.38% NMSE); **D** – dense map reconstructed using TV (5.84% NMSE); **E** – dense map reconstructed using TGV (0.95% NMSE). Bottom Row: Results for random subsampling. **F** – densely sampled motion map(identical to **A**); **G** – map from **F** subsampled randomly by a factor of 10; **H** – dense map reconstructed using ISTA (45.55% NMSE); **I** – dense map reconstructed using TV (35.59% NMSE); **J** – dense map reconstructed using TGV (10.79% NMSE). Both qualitatively and quantitatively, TGV with uniform subsampling is seen to outperform the other methods on this sample.

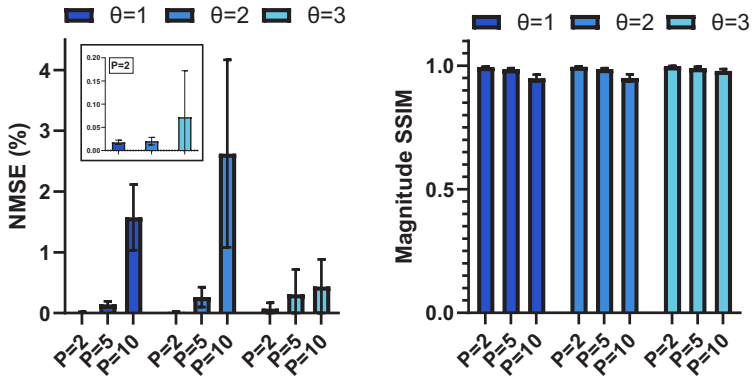

Fig. 4. NMSE and magnitude SSIM for TGV with uniform subsampling by factors of $P = 2$, 5 and 10 across the full dataset ($N = 275$), organized by orientation. At $\theta = 1$, $N = 75$; at $\theta = 2$, $N = 100$; at $\theta = 3$, $N = 100$. Data are further described in Sec 5.1. Values displayed are sample means over the set at each orientation, and error bars represent one sample standard deviation from the mean. Inset shows the NMSE for subsampling by a factor of 2, as it is much smaller than that achieved when subsampling by factors of 5 or 10.

acquisition time is a limiting factor in *in vivo* experiments, and 2) sample drift worsens over the acquisition period [2,3]. Other noise sources, such as signal competition, are not time-dependent and also cannot be solved by averaging [35]. CSVi not only aids in reduction of experiment time, but may aid in the reduction of noise as well. Although it is difficult to quantify noise when our "ground truth" displacement maps are inherently noisy, this denoising feature can be seen qualitatively in the more smooth (and thereby more likely physical) features apparent in recovered signals of Fig. 5.

Biomedical Optics EXPRESS

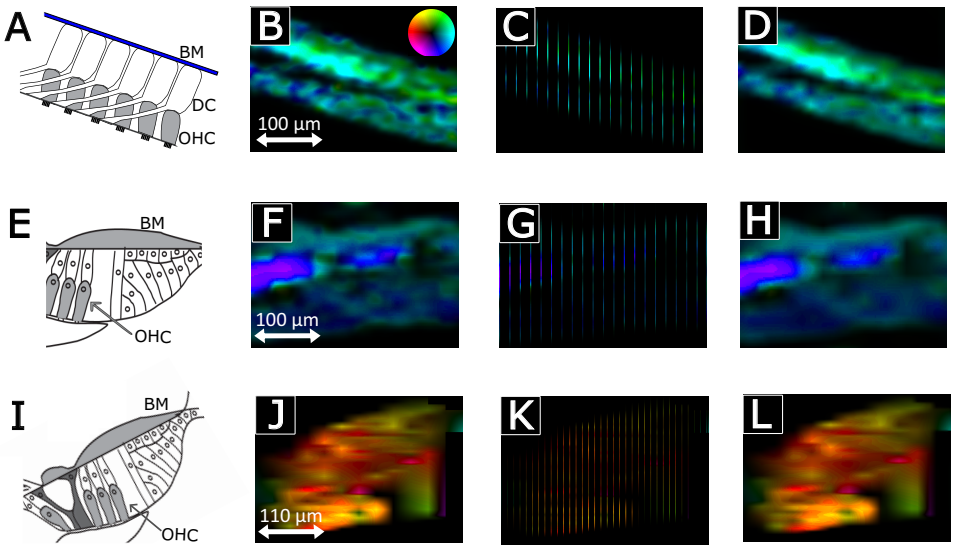

Fig. 5. Representative sample of three reconstructions made using TGV with uniform subsampling. Colormap is shown in **B**, with hue representing phase re EC and saturation representing gain normalized by maxima. **A** – Cartoon of approximate anatomy at $\theta = 1$ with basilar membrane (BM), Deiters cells (DC) and outer hair cells (OHCs) labeled. **B** – Densely sampled motion map for $\theta = 1$, 60 dB SPL, 15 kHz. Map is 100 rows by 200 columns, and 270 μm $\times$ 300 μm. **C** – Map in **B** uniformly subsampled by a factor of 10. **D** – Dense map reconstructed from the subsampled map in **B**. **E**-**H** – Same as **A**-**D**, but with $\theta = 2$, 80 dB SPL, 9 kHz. Maps are 100 rows by 200 columns, and 270 μm $\times$ 300 μm. **I**-**L** – Same as **A**-**D**, but with $\theta = 3$, 70 dB SPL, 25 kHz. Maps are 100 rows by 300 columns, and 270 μm $\times$ 330 μm. All data are further described in Sec 5.1.

Many important features, such as the important “hotspot" delineation (the difference between the higher-intensity motion at the OHC and the lower-intensity motion at the supporting cells and BM) appear high in spatial frequency [14,21]. The sharpness of this delineation is unlikely to be physical, as the OHCs are not mechanically isolated from the rest of the OCC (Fig. 1). It is more likely that this apparent sharpness is a result of noise, and is exaggerated by insufficient sampling. TGV optimization penalizes these aphysical high-frequency components.

It is useful to compare our uniform sampling method to the classical sampling theory of Nyquist. Within this framework, a bandlimited signal can be faithfully represented by fewer samples so long as the sampling frequency is greater than twice the bandwidth. High-frequency features will lead to aliasing when upsampling and interpolating discrete-space signals, and as high-frequency noise is unavoidable some noise aliasing will always be present. No denoising will take place and phantom edge phenomena will occur.

7.2. Experimental applications

CSVi is most useful in that it enables various experiments that would be otherwise intractable. OCT experiments for cochlear mechanics broadly fall into two categories – 1) comparison of the motion in healthy cochlea at different anatomical positions within the OCC, and 2) perturbation studies in which the OCC is modified by either a genetic mutation or administration of a drug (e.g. furosemide or salicylate) [11,13,14,18,21,36,37]. The former have been enhanced in recent years by the development of methods to probe 2-D or 3-D motion both quantitatively [17,19,20]

Biomedical Optics EXPRESS

and qualitatively [14,22]. All such methods require achieving measurement of a sample from multiple angles, which is extremely time-consuming.

This has made it challenging, if not impossible, perform recovery/perturbation studies while probing either 2-D or 3-D motion. For example, recovery from furosemide injection on timescales of 10 minutes are of interest [13], but dense reconstruction of 2-D motion currently takes at least 40 minutes (20 minutes at each measurement angle) [19]. A factor of 10 reduction in sample acquisition, as offered by the presented CSVi method, would accelerate experiments sufficiently for 2-D effects of such perturbations to be probed.

Even in healthy cochlea, densely sampled 3-D motion maps have not yet been achieved *in vivo*. Even methods which could theoretically reconstruct such measurements are far too time-limited. Any 3-D reconstruction would require at least three measurements from distinct angles, and the noise level is strongly multiplied by the fact that large angular differences are challenging to achieve. For this reason, even 2-D reconstructions have been limited to responses to 80 dB SPL stimuli for which the SNR is sufficiently high. Only over-constrained reconstructions involving measurements at many angles could achieve 2-D or 3-D reconstructions at lower SPL.

Densely sampled *in vivo* 3-D motion at many SPLs would offer the most complete description possible of the healthy cochlea. If acceleration by a factor of 10 is achieved, 20-minute motion measurements at each angle can be reduced to a 2-minute acquisition time.

Funding. National Institutes of Health (R03 EB032097); National Institute on Deafness and Other Communication Disorders (F31 DC020621-02, R01 DC015363).

Acknowledgments. We would like to acknowledge Elizabeth S. Olson, who provided resources for the experiments and helped in proofreading the manuscript.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. J. A. Izatt and M. A. Choma, *Theory of Optical Coherence Tomography* (Springer Berlin Heidelberg, 2008), pp. 47–72.
2. J. P. Kolb, W. Draxinger, and J. Klee, *et al.*, "Correction: Live video rate volumetric oct imaging of the retina with multi-mhz a-scan rates," PLoS One **14**(3), e0213144 (2019).
3. Y. Chen, Y.-J. Hong, S. Maxita, and Y. Yasuno, "Three-dimensional eye motion correction by lissajous scan optical coherence tomography," Biomed. Opt. Express **8**(3), 1783–1802 (2017).
4. H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "K-t focuss: A general compressed sensing framework for high resolution dynamic mri," Magn. Reson. Med. **61**(1), 103–116 (2009).
5. D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," Proc. Natl. Acad. Sci. U.S.A. **100**(5), 2197–2202 (2003).
6. M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," IEEE Trans. Signal Process. **59**(9), 4053–4085 (2011).
7. G. Kutyniok, "Theory and applications of compressed sensing," GAMM-Mitteilungen **36**, 79–101 (2013).
8. E. Lebed, P. J. Mackenzie, M. V. Sarunic, and M. F. Beg, "Rapid volumetric oct image acquisition using compressive sampling," Opt. Express **18**(20), 21003–21012 (2010).
9. J. P. McLean and C. P. Hendon, "3-d compressed sensing optical coherence tomography using predictive coding," Biomed. Opt. Express **12**(4), 2531–2549 (2021).
10. M. A. Choma, A. K. Ellerbee, C. Yang, T. L. Creazzo, and J. A. Izatt, "Spectral-domain phase microscopy," Opt. Lett. **30**(10), 1162–1164 (2005).
11. E. Fallah, C. E. Strimbu, and E. S. Olson, "Nonlinearity and amplification in cochlear responses to single and multi-tone stimuli," Hearing Res. **377**, 271–281 (2019).
12. F. Chen, D. Zha, A. Fridberger, J. Zheng, N. Choudhury, S. L. Jacques, R. K. Wang, X. Shi, and A. L. Nuttall, "A differentially amplified motion in the ear for near-threshold sound detection," Nat. Neurosci. **14**(6), 770–774 (2011).
13. C. E. Strimbu, Y. Wang, and E. S. Olson, "Manipulation of the endocochlear potential reveals two distinct types of cochlear nonlinearity," Biophys. J. **119**(10), 2087–2101 (2020).
14. N. P. Cooper, A. Vavakou, and M. Van Der Heijden, "Vibration hotspots reveal longitudinal funneling of sound-evoked motion in the mammalian cochlea," Nat. Commun. **9**(1), 3054 (2018).

Biomedical Optics EXPRESS

15. W. Dong, A. Xia, P. D. Raphael, S. Puria, B. E. Applegate, and J. S. Oghalai, "Organ of corti vibration within the intact gerbil cochlea measured by volumetric optical coherence tomography and vibrometry," J. Neurophysiol. **120**(6), 2847–2857 (2018).
16. S. S. Gao, R. Wang, P. D. Raphael, Y. Moayedi, A. K. Groves, J. Zuo, B. E. Applegate, and J. S. Oghalai, "Vibration of the organ of corti within the cochlear apex in mice," J. Neurophysiol. **112**(5), 1192–1204 (2014).
17. H. Y. Lee, P. D. Raphael, A. Xia, J. Kim, N. Grillet, B. E. Applegate, A. K. E. Bowden, and J. S. Oghalai, "Two-dimensional cochlear micromechanics measured in vivo demonstrate radial tuning within the mouse organ of corti," J. Neurosci. **36**(31), 8160–8173 (2016).
18. B. L. Frost, C. E. Strimbu, and E. S. Olson, "Using volumetric optical coherence tomography to achieve spatially resolved organ of corti vibration measurements," J. Acoust. Soc. Am. **151**(2), 1115–1124 (2022).
19. B. Frost, C. E. Strimbu, and E. S. Olson, "Reconstruction of transverse-longitudinal vibrations in the organ of corti complex via optical coherence tomography," J. Acoust. Soc. Am. **153**(2), 1347–1360 (2023).
20. W. Kim, D. Liu, S. Kim, K. Ratnayake, F. Macias-Escriva, S. Mattison, J. S. Oghalai, and B. E. Applegate, "Vector of motion measurements in the living cochlea using a 3d oct vibrometry system," Biomed. Opt. Express **13**(4), 2542–2553 (2022).
21. C. E. Strimbu and E. S. Olson, "Salicylate-induced changes in organ of corti vibrations," Hearing Res. **423**, 108389 (2022).
22. S. Meenderink and W. Dong, "Organ of corti vibrations are dominated by longitudinal motion in vivo," Commun. Biol. **5**(1), 1285 (2022).
23. N. Janjuševic, A. Khalilian-Gourtani, and Y. Wang, "CDLNet: Noise-adaptive convolutional dictionary learning network for blind denoising and demosaicing," IEEE Open J. Signal Process. **3**, 196–211 (2022).
24. A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson, "End-to-end variational networks for accelerated mri reconstruction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, eds. (Springer International Publishing, Cham, 2020), pp. 64–73.
25. D. Gilton, G. Ongie, and R. Willett, "Model adaptation for inverse problems in imaging," IEEE Trans. Comput. Imaging **7**, 661–674 (2021).
26. J. Duan, W. Lu, C. Tench, I. Gottlob, F. Proudlock, N. N. Samani, and L. Bai, "Denoising optical coherence tomography using second order total generalized variation decomposition," Biomed. Signal Process. Control **24**, 120–127 (2016).
27. R. Huber, G. Haberfehlner, M. Holler, G. Kothleitner, and K. Bredies, "Total generalized variation regularization for multi-modal electron tomography," Nanoscale **11**(12), 5617–5632 (2019).
28. D. Wang, D. S. Smith, and X. Yang, "Dynamic mr image reconstruction based on total generalized variation and low-rank decomposition," Magn. resonance in medicine **83**, 2064–2076 (2020).
29. C. P. Versteegh and M. van der Heijden, "Basilar membrane responses to tones and tone complexes: nonlinear effects of stimulus intensity," J. Assoc. for Res. Otolaryngol. **13**(6), 785–798 (2012).
30. S. O. Haykin, *Adaptive Filter Theory* (Pearson, Upper Saddle River, NJ, 2013), 5th ed.
31. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. **2**(1), 183–202 (2009).
32. S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Birkhäuser Basel, 2013).
33. I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Comm. Pure Appl. Math. **57**, 1413–1457 (2004).
34. A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," Acta Numer. **25**, 161–319 (2016).
35. N. C. Lin, C. P. Hendon, and E. S. Olson, "Signal competition in optical coherence tomography and its relevance for cochlear vibrometry," The J. Acoust. Soc. Am. **141**(1), 395–405 (2017).
36. J. B. Dewey, A. Altoe, C. A. Shera, B. E. Applegate, and J. S. Oghalai, "Cochlear outer hair cell electromotility enhances organ of corti motion on a cycle-by-cycle basis at high frequencies in vivo," Proc. Natl. Acad. Sci. U.S.A. **118**(43), e2025206118 (2021).
37. N. H. Cho and S. Puria, "Motion of the cochlear reticular lamina implies that it is not a stiff plate," Sci. Rep. **12**(1), 18715 (2022).