Optical Coherence Tomography Techniques for Contextualizing and Reconstructing Displacement Responses in the Mammalian Cochlea

Brian L. Frost

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2024

Abstract

Optical Coherence Tomography Techniques for Contextualizing and Reconstructing Displacement Responses in the Mammalian Cochlea

Brian L. Frost

Spectral domain optical coherence tomography (OCT) is a powerful tool for measuring nanometer-scale displacement responses in the cochlea, as it is capable of volumetric imaging and vibrometry at a depth into a sample. The past decade has seen a wealth of OCT-measured displacement data from structures within the organ of Corti complex (OCC) that had previously been impossible to measure *in vivo*. These data have revealed surprising features of active intra-OCC motion but have not yet led to a complete understanding of cochlear amplification, the means by which active processes enhance the tuning and gain of the cochlear displacement responses in a level-dependent manner.

Certain technical challenges arise from the properties of OCT imaging and vibrometry that obscure the interpretation of intra-OCC displacement measurements. In particular, OCT-measured responses are dependent on the orientation of the system's beam axis. The beam axis is generally chosen based on experimental convenience, and has no inherent relevance to the anatomy of the cochlea.

This introduces two problems: 1) OCT-acquired images of the cochlea may be taken at skewed angles relative to the cochlea's naturally endowed anatomical coordinates, and 2) OCT-measured displacement responses are projections of a three-dimensional motion onto the beam axis. This thesis concerns the quantification of these effects on intra-OCC displacement measurements, as

well as the development of methods to overcome these complications *in vivo*. In doing so, previously reported data that appear to disagree can be synthesized.

I present a method by which the skew of OCT images relative to cochlear anatomy can be quantified, relating the OCT system's optical coordinates to the cochlear anatomy. With this method, I have shown that OCT images resembling familiar anatomical drawings of longitudinal cross-sections often capture a completely different anatomical slice of the cochlea. This leads to large quantitative shifts in phase responses when measuring displacements along a single beam axis, as opposed to what one would measure if s/he were measuring along an anatomically relevant axis. I have also provided a method by which to account for this phenomenon to capture structures related in some desired anatomical fashion.

I then turn to the issue of projection of the three-dimensional cochlear motion onto the OCT beam axis. I have provided a method for reconstructing two- and three-dimensional displacement responses in the relevant anatomical directions by acquiring displacement measurements at multiple locations within the cochlea. In doing so, I have revealed that previously unexplained disagreements between measurements in different experimental preparations can be explained by competing components of motion being projected onto the single axis. I have also shown that motion at the junction between the outer hair cells and Deiters cells follows a lineal pattern, as opposed to non-degenerate elliptical patterns that would be expected of fluid motion in this region. This method requires the acquisition of data at many points within the OCC, making it significantly time-consuming. This makes it vulnerable to sample drift and deterioration, and reduces experimental yield. Certain applications of the method – such as reconstructing displacement maps over a dense volume – are thereby intractable. To address this problem, I have developed a compressed sensing method for vibrometry (CSVi). CSVi is a classical optimization method based on a total generalized variation signal prior, which is shown to out-perform methods using total variation and wavelet domain sparsity priors. I have also found that uniform sub-sampling schema offered significant performance benefits over random sub-sampling schema. I found that this CSVi method can reconstruct densely sampled displacement maps in the

cochlea *in vivo* with less than 5% normalized mean square error, using only 10% of samples. While these methods offer new insight into interpretation of OCT displacement measurements, there is still a challenge in measuring the motion of the stereocilia of the hair cells. The stereocilia are too small to be imaged using OCT, and the proxy measurement of differential motion of the reticular lamina and tectorial membrane (between which the stereocilia lie) is not yet achievable in the gerbil base. Stereocilia motion is related to the transduction current through the hair cells, which is critical to understanding cochlear function. These currents lead to neurotransmitter release and active electromotile responses believed to be responsible for cochlear amplification. I present a model for studying another proxy measurement of the stereocilia motion – the voltage in the cochlea's scala tympani, or cochlear microphonic (CM). This model of CM reveals that to match experimental data 1) stereocilia motion must be more sharply tuned than measured intra-OCC displacement responses, 2) the displacement-current gain of the mechano-electric transducer channels *in vivo* must be larger than what is measured *in vitro* by a factor of ~ 6, and 3) the hair cells at more basal locations of the cochlea must be compromised. These predictions offer insight into aspects of cochlear mechanics that are not easily probed using OCT.

Table of Contents

List of Figures

Acknowledgements

I would like to thank Drs. Elizabeth Olson and Christine Hendon for offering superb mentorship. Their knowledge, passion and openness have been invaluable to my growth as a scientist.

I would like to thank Dr. Stanislav Mintchev of the Cooper Union for training me in mathematics, and investing so much time and effort in my pursuit of academic research.

I would like to thank Dr. Clark Elliott Strimbu for the major role he has played in every experiment I've conducted in my time at Columbia. He has also been an excellent source of knowledge, and a pleasure to work alongside. I would also like to thank Dr. Elika Fallah and Lauren Chiriboga for being excellent labmates and friends.

I would like to thank scientists in the field of cochlear mechanics who have helped to edit and assess my work, and have taught me very much about the inner ear: Dr. Marcel van der Heijden, Dr. Jong-Hoon Nam, Dr. Karl Grosh, Dr. Sunil Puria and Dr. Tianying Ren.

I would like to thank Nikola Janjušević for pen-and-paper brainstorming sessions over weekly breakfasts, collaboration on the compressed sensing work, and mutual maintenance of sanity.

I would like to thank my parents, Kimberly Frost and Diana Butera, my brother Lance Weisberg, and my grandmothers, Claudia Sancimino and Diane Poland, for their unconditional love.

I would like to thank my partner Tosha Diamond-Huey, for sticking by my side for the last three incredible years.

I would finally like to thank my most beloved friends Michael Cohen, Mark Pirouet, Romaniya Voloshchuk, Hugo Ball, Karol Wadolowski, Dallas Skau, Shannon Jordan, David Benanti, Megan Noga, Dylan Phelan and Jaka Wescott for their continuous and immeasurable support.

Dedication

To Jerry C. LaPlante (1936-2020), incredible inventor and grandfather who sparked my love for science and engineering.

Chapter 1: Principles of Spectral-Domain Optical Coherence Tomography

Optical coherence tomography (OCT) fills a strong niche in medical imaging – it is capable of volumetric imaging several millimeters into a sample with typical resolutions on the scale of several microns in all three dimensions [1]. Additionally, it is capable of measuring sub-pixel displacements at the angstrom scale through spectral domain phase microscopy (SDPM) [2]. These features make OCT particularly well-suited for the study of cochlear mechanics, which concerns the complex sound-evoked vibration patterns of the organ of Corti complex (OCC) that are in part responsible for the sensation of hearing [3, 4, 5, 6, 7, 8].

OCT systems are generally classified into three categories: time-domain OCT, spectral-domain OCT and swept-source OCT. All data presented in this thesis was recorded using a ThorLabs Telesto 320 spectral-domain OCT system at Columbia University's Fowler Memorial Laboratory (details in App A.2). As such, all future references to OCT refer to spectral-domain OCT in particular. In this chapter, I will discuss the mechanisms by which OCT can be used to acquire images, volumes and time-domain displacement signals.

1.1 Architecture

Fig 1.1 shows a diagram of the canonical spectral-domain OCT device architecture, which resembles a Michelson-Morley interferometer. Light from a broadband near-infrared source is split into two paths – the *sample beam* that travels towards the sample to be imaged, and the *reference beam* that travels towards a reference mirror. Before reaching the sample, the sample beam is redirected by a two-axis galvo mirror which precisely controls the angle of the beam. The sample beam reflects off the sample at various depths, and returns to the system to interfere with the reference beam.

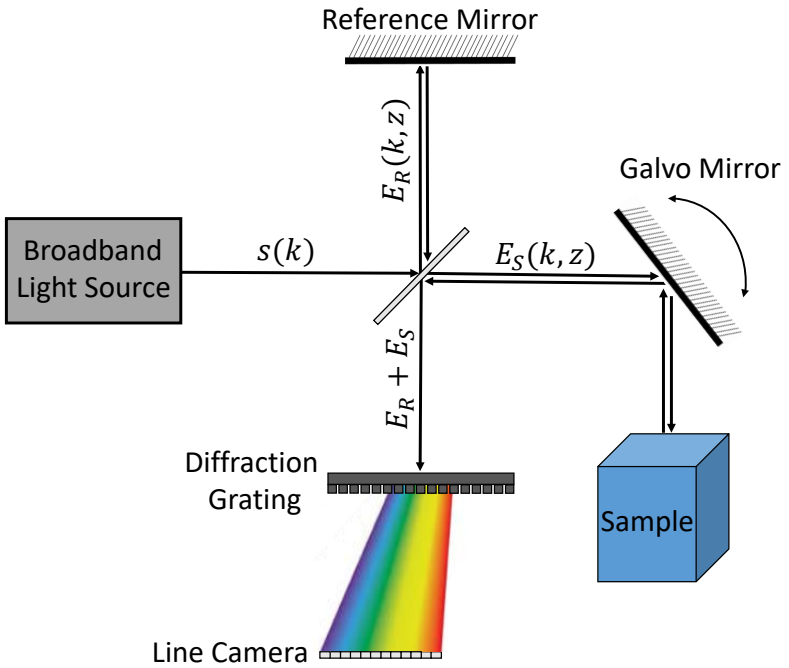

Figure 1.1: Cartoon illustration of the architecture of a spectral domain OCT device, resembling a Michelson-Morley interfermoeter. A near-infrared source with wavenumber-domain (k-domain) spectrum $s(k)$ illuminates a beam splitter (the central white rectangle), creating the reference beam E_R and the sample beam E_S. Optical path length from the beam splitter is z. The galvo mirror directs the sample beam. The reflected beams interfere and are passed through a diffraction grating, after which their intensity is measured by a line camera.

The interfering sample and reference beams pass through a diffraction grating, which spatially separates the light by wavelength. This illuminates a line camera for which each pixel corresponds to a different wavelength interval. This wavelength-domain intensity at the line camera is referred to as *the OCT signal*.

1.2 Formation of an A-Scan

The OCT signal is an interference pattern in the wavelength domain that contains the information required to generate a map of the reflectivity as a function of depth into the sample. This reflectivity map, known as an *A-Scan*, is the fundamental unit from which OCT images and displacement measurements are built. In this section, I will discuss the formation of an A-Scan from the OCT signal.

1.2.1 The OCT Signal in the Wavenumber Domain

For reasons that will soon become evident, I will represent quantities as functions of wavenumber k rather than wavelength λ, related by $k = 2\pi/\lambda$. I will also ignore noise phenomena for the time being.

Light is modeled as a plane wave traveling in the z direction, and is represented through its electric field. That is, each beam has the form

$$E(k, z, t) = A(k)e^{-j(kz-\omega t)},$$

where A [N/C] is the magnitude spectrum, z denotes the optical path length (OPL) traveled from the beam-splitter at $z = 0$, and ω is the temporal radian frequency. As time-dependence will be identical across all beams, it will be left implicit for the remainder of this section.

The light source has electric field spectrum $s(k)$ [N/C]. The sample and reference beams are given by

$$E_S(k, z) = s(k)a_S e^{-jkz}, \quad E_R = s(k)a_R e^{-jkz}$$

respectively. Here, a_S and a_R (both unitless) represent the proportion of light intensity split to their respective paths, satisfying $a_S^2 + a_R^2 = 1$.

The reference beam arrives at the reference mirror after traveling a distance z_R, reflects, and then arrives back at the beam splitter to interfere with the sample beam. That is, it interferes after traveling a distance $2z_R$. This is both the physical path length and OPL as the reference beam is assumed to pass through air with index of refraction 1.

Ignoring the distance from this point to the spectrometer (which is identical between the two beams), the detected reference beam is thereby

$$E_R^D(k) = s(k)a_R r_R e^{-jk2z_R}, \tag{1.1}$$

where r_R (unitless) is the reflectivity constant of the reference mirror. Ideally, $r_R \approx 1$.

The sample beam reflects off of the sample at various depths. The sample is modeled as a sequence of M point-reflectors, each at physical depth z_m into the sample with reflectivity r_m (unitless) for $m = 1, 2, \ldots, M$. The total sample reflectivity function is

$$r_S(z) = \sum_{m=1}^{M} r_m \delta(z - z_m) \tag{1.2}$$

where δ is the Kroenecker delta function. The portion of the sample beam that reflects off of reflector m travels a total OPL of $2nz_m$ before arriving back at the beam splitter to interfere with the reference beam, where n (unitless) is the index of refraction of the medium.

We make the assumption that $r_m \ll 1$, so that almost all of the energy incident on the first reflector is also incident at the M^{th} reflector. We also assume that no scattering occurs, and that no internal reflection occurs. The detected sample beam is then

$$E_S^D(k) = s(k)a_S \sum_{m=1}^{M} r_m e^{-jk2nz_m}. \tag{1.3}$$

The right-hand side of the above equation can be recognized as the convolution of the plane wave

with $r_S(2nz)$.

The photodetector converts time-averaged incident power to current. The irradiance of the superposition of these two beams is

$$\frac{1}{2}c\epsilon_0|E_R^D + E_S^D|^2 \quad [\mathrm{W/m^2}],$$

where c [m/s] is the speed of light and ϵ_0 [$\mathrm{C^2/Nm^2}$] is the permittivity of free space. The photodetector has some responsivity $\mathcal{R}$ [A/W] and some pixel area $\mathcal{A}$ [$\mathrm{m^2}$]. Defining

$$\rho = c\epsilon_0\mathcal{R}\mathcal{A},$$

the OCT signal, I^D [A], is

$$I^D(k) = \frac{\rho}{2}\left\langle|E_R^D(k) + E_S^D(k)|^2\right\rangle = \frac{\rho}{2}\left\langle\left(E_R^D(k) + E_S^D(k)\right)\left(E_R^D(k) + E_S^D(k)\right)^*\right\rangle, \tag{1.4}$$

where $< \cdot >$ denotes integration over the photodetector's integration time, and $\cdot^*$ denotes complex conjugation.

Defining the spectrum intensity $S(k) =< |s(k)e^{j\omega t}|^2 >$, I can write the rather large expression for measured intensity:

$$I^D(k) = \frac{\rho}{2}S(k)\left(a_R r_R e^{-jk2z_R} + a_S\sum_{m=1}^{M} r_m e^{-jk2nz_m}\right)\left(a_R r_R e^{jk2z_R} + a_S\sum_{m=1}^{M} r_m e^{jk2nz_m}\right) \tag{1.5}$$

$$= \frac{\rho}{2}S(k)\left(a_R^2 r_R^2 + a_S^2\sum_{m=1}^{M} r_m^2\right) + \tag{1.6}$$

$$+ \frac{\rho}{2}S(k)a_R a_S r_R\sum_{m=1}^{M} r_m\left(e^{j2k(nz_m - z_R)} + e^{j2k(z_R - nz_m)}\right) + \tag{1.7}$$

$$+ \frac{\rho}{2}S(k)\sum_{l\neq m} r_l r_m\left(e^{j2kn(z_l - z_m)} + e^{j2kn(z_m - z_l)}\right). \tag{1.8}$$

Three components are apparent here – the z-independent *DC term* in line 1.6 is a constant scaling

of the spectrum $S(k)$, the *cross-correlation* terms in line 1.7 represent interference between the reference beam and the components of the sample beam reflected from each reflector, and the *autocorrelation terms* in line 1.8 represent pair-wise interference between components of the sample beam reflected at distinct reflectors.

Earlier I assumed that $r_R \approx 1$ and $r_m \ll 1$ for $m = 1, 2, \ldots, M$. Considering any pair of reflectors l and m, these assumptions give $r_l r_m \ll r_R r_m$. This means that the autocorrelation terms are small compared to the DC and cross-correlation terms, and will be ignored for the remainder of the derivation.

Removing the autocorrelation terms and using Euler's formula,

$$I^D(k) \approx GS(k) + HS(k) \sum_{m=1}^{M} r_m \cos[2k(nz_m - z_R)], \tag{1.9}$$

$$G = \frac{\rho}{2}\left(a_R^2 r_R^2 + a_S^2 \sum_{m=1}^{M} r_m^2\right), \quad H = \rho a_R a_S r_R.$$

This is the compact form of the OCT signal in the wavenumber domain. Note that by the assumptions concerning reflectances, $G > H$.

1.2.2 The A-Scan

From Eqn 1.9 it can be seen that the OCT signal encodes the depth of each reflector by frequency. The natural next step is to consider the Fourier transform of this signal, where the Fourier partner of k is denoted as ζ [m]. Defining $\mathcal{F}\{I^D\} = \tilde{A}$ and $\mathcal{F}\{S\} = B$ (referred to as the background), the Fourier transform is

$$\begin{aligned} \tilde{A}(\zeta) &= GB(\zeta) + \frac{H}{2} B(\zeta) * \sum_{m=1}^{M} r_m \big(\delta[\zeta - 2(nz_m - z_R)] + \delta[\zeta - 2(z_R - nz_m)]\big) \\ &= GB(\zeta) + \frac{H}{2} \sum_{m=1}^{M} r_m \big(B[\zeta - 2(nz_m - z_R)] + B[\zeta - 2(z_R - nz_m)]\big) \end{aligned} \tag{1.10}$$

where $*$ denotes convolution and δ denotes either the Dirac or Kroenecker delta function (in the idealized continuous case or the practical discrete case, respectively).

This signal is conjugate-symmetric in ζ (as it is the Fourier transform of a real signal) and contains two components – a scaled copy of the spectrum centered about $\zeta = 0$, and scaled copies of the background centered at different ζ positions corresponding to the depths of the reflectors. This is nearly the desired object – a spatial reflectivity profile achieved by taking a Fourier transform of the OCT signal.

This concept is illustrated for a simple two-point-reflector sample in Fig 1.2. The nearer reflector (blue) produces a lower-frequency wavenumber-domain interference pattern, while the further reflector (orange) produces a higher-frequency pattern. These add with the DC component to generate the OCT signal in panel **B**. Its Fourier transform magnitude in **C** shows broad peaks at the positions of the two reflectors symmetrically distributed about the origin.

However, there are two apparent problems – 1) G is large compared to H, so the background spectrum centered at 0 may dwarf nearby signals corresponding to reflectors (see the size of the DC peak in Fig 1.2 **C**), and 2) each reflector is smoothed in the ζ domain by the background, so the resolution of the system is limited by the properties of the broadband light source (see the breadth of the peaks in Fig 1.2 **C**). Point (2) corresponds to the interpretation of $B(\zeta)$ as a point-spread function for the OCT system. For these reasons, $\tilde{A}$ is not used as the A-Scan.

Instead, we measure the spectrum $S(k)$ by blocking the sample beam. The measured $S(k)$ can be used in two ways. First, by making an estimate for G (this process usually involves observing the average value of the OCT signal), one can *subtract* $GS(k)$ from the OCT signal $I^D(k)$. This removes the DC term.

Then, one can *divide* this signal by S(k). In the spatial domain, this is equivalent to *deconvolving* with $B(\zeta)$. This process is effectively identical to the standard optical procedure of deblurring by deconvolving with the point-spread function. After performing this subtraction and division in

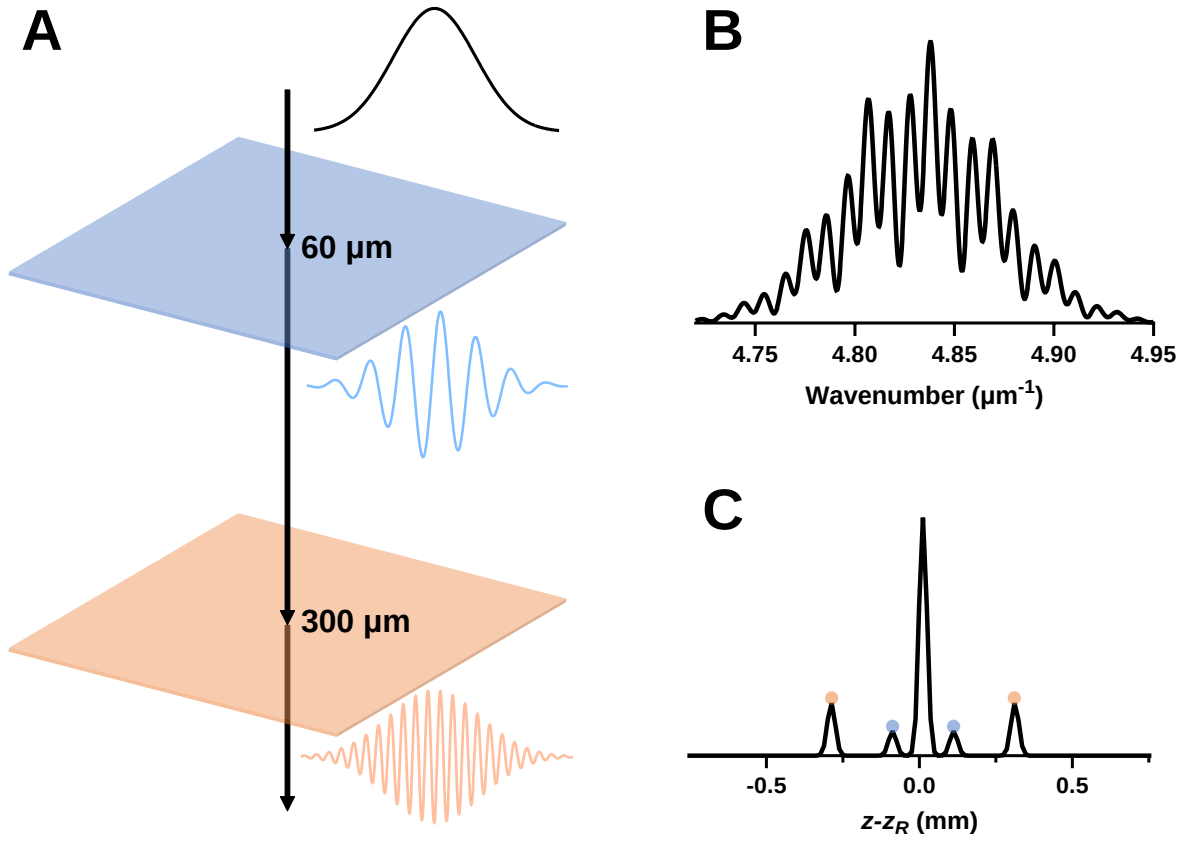

Figure 1.2: Simulated example illustrating the acquisition of $\tilde{A}$ for a simple sample containing two point-reflectors. **A** – Arrow represents the sample beam travelling through a sample consisting of two reflective planes. The uppermost black curve represents the wavenumber-domain spectrum of the incident light (the background). Blue and orange curves correspond to the wavenumber-domain interference patterns induced by the two point-reflectors. Overlain distances represent values of $z - z_R$ at the point-reflectors. **B** – Wavenumber-domain OCT signal, which is the sum of the black, blue and orange curves in **A**. **C** – Magnitude of the Fourier transform of the OCT signal in **B**, plotted against $z - z_R$. A large DC component comes from the background spectrum (black curve in **A**), while the marked blue and orange peaks correspond to the respectively colored reflectors.

the wavenumber domain, the A-Scan in the ζ domain is given by

$$\begin{aligned} A(\zeta) &= \mathcal{F}\left\{\frac{GS(k) + HS(k)\sum_{m=1}^{M} r_m \cos[2k(nz_m - z_R)] - GS(k)}{S(k)}\right\}(\zeta) \\ &= \frac{H}{2}\sum_{m=1}^{M} r_m\big(\delta[\zeta - 2(nz_m - z_R)] + \delta[\zeta - 2(z_R - nz_m)]\big), \end{aligned} \tag{1.11}$$

Considering only positive ζ, assuming $nz_m > z_R$ for all reflectors, and letting $z = (\zeta + 2z_R)/2n$, we arrive at the formula for the ideal A-Scan in physical distance into the sample:

$$A(z) = \frac{H}{2}\sum_{m=1}^{M} r_m \delta[2n(z - z_m)]. \tag{1.12}$$

Normalizing the z axis by $2n$, we get the physical distance between the reflectors and the beam splitter.

1.3 B-Scans, C-Scans and M-Scans

The modulus of the A-Scan is a single-channel one-dimensional “image" along the beam’s axis, z, into the sample. To form a 2-D image, the sample beam can be swept along an axis y using the galvo mirror. At each point along the y axis, an A-Scan can be recorded to form what is known as a *B-Scan*. The modulus of a B-Scan is proportional to the reflectivity in a subset of the yz-plane, forming a grayscale image in the standard sense.

Naturally, the beam can also be swept along a third perpendicular axis x so that the corresponding unit vectors $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ form a right-handed coordinate system. Recording A-Scans while sweeping the beam across a rectangular area in the xy-plane, a volumetric reflectivity map known as a *C-Scan* (or equivalently *volume scan*) is formed.

Until now, time-dependence of the A-Scan has been left implicit. In the case of a stationary sample, the A-Scan is time-independent. However, in a non-stationary sample, the A-Scan may change significantly as a function of time. We refer to a series of A-Scans recorded at the same xy-location over a period of time as an *M-Scan*.

In the idealized case, the z axis is continuous. This means that motion is encoded simply by letting a reflector point z_m vary as a function of time. In practice, the spectrometer has only a finite number of pixels so that z is discretely sampled with pixel size Δz. The treatment of motion then falls into two categories – 1) when the displacements of reflectors varies on a scale larger than Δz, and 2) the case of *sub-pixel* displacements, where displacement magnitudes are smaller than Δz.

Displacements of type (1) are sufficiently large to be seen in the modulus of an M-Scan. This means that such displacements can be analyzed through standard video processing methods on the real M-Scan modulus signal.

Displacements of type (2) will be invisible in the modulus, as the reflectivity at each pixel does not change as a function of time. This case requires special treatment specific to the discretization of z-space. As the sound-evoked displacements in the OCC are on the order of nanometers – far less than the several-micron scale of the z-resolution of OCT – this case is of particular interest for the study of cochlear mechanics.

1.4 Spectral Domain Phase Microscopy

Consider a reflector with refelctivity r_0 and z-direction displacement $z_0 + d(t)$, where z_0 is the z-position of the nearest pixel boundary and $0 < d(t) < \Delta z$ (i.e. the reflector lies within a single pixel for all time). The OPL traveled by the portion of the sample beam reflected by this reflector is $2n(z_0 + d(t))$. The sample beam at the detector is

$$E_S^D(k,t) = s(k)a_S r_0 e^{-jk2n(z_0+d(t))}. \tag{1.13}$$

To determine the effect of this displacement on the OCT signal, one can associate this OPL with a single z_m in Eqn 1.9. Subtracting the background and dividing by S, as in Eqn 1.11, the ζ-domain A-Scan over time, or M-Scan, will be

$$M(\zeta,t) = \mathcal{F}_k\{Hr_0\cos[2k(nz_0 - z_R) + 2knd(t)]\}(\zeta,t), \tag{1.14}$$

where the subscript k denotes that the Fourier transform is being taken with respect to wavenumber. As the time-dependent term is itself multiplied by k, its impact to the equation would modulate the detected reflector position – however, we know that is not a detectable phenomenon. We can circumnavigate this issue through a simple approximation.

Writing the central wavelength of the light source as λ_0, we can approximate the k term multiplying $d(t)$ as a constant, $2\pi/\lambda_0$. That is, we can write

$$M(\zeta, t) \approx \mathcal{F}_k\left\{Hr_0 \cos\left[2k(nz_0 - z_R) + \frac{4\pi n d(t)}{\lambda_0}\right]\right\}(\zeta, t). \tag{1.15}$$

As the Fourier transform is being taken with respect to wavenumber, the time-dependent term acts as a k-independent phasing term. Considering only positive ζ, the Fourier transform is

$$M(\zeta, t) = \frac{H}{2} r_0 \delta[\zeta - 2(nz_0 - z_R)] e^{j\frac{4\pi n d(t)}{\lambda_0}}. \tag{1.16}$$

Under the same coordinate transformation as before, I have

$$M(z, t) = \frac{H}{2} r_0 \delta[2n(z - z_0)] e^{j\frac{4\pi n d(t)}{\lambda_0}}. \tag{1.17}$$

Referring to the phase of the M-Scan as $\angle M = \phi$, the phase of the measurement at z_0 is $\phi(z_0, t) = 4\pi n d(t)/\lambda_0$. This allows one to recover the sub-pixel displacement signal from the M-Scan as

$$d(t) = \frac{\lambda_0 \phi(z_0, t)}{4\pi n}. \tag{1.18}$$

This method is known as *spectral domain phase microscopy* (SDPM), and facilitates measurement of sub-pixel displacements with incredible angstrom-scale resolution [2]. It should be noted that this method only picks up the one-dimensional motion of the reflector along the z axis.

1.5 Resolution and Field of View

The resolution of the OCT system is determined by the central wavelength (λ_0) and bandwidth ($\Delta\lambda$) of the light source, as well as the numerical aperture of the objective lens (NA) [1]. The lateral (x and y) resolution is derived from the Rayleigh criterion, given in terms of the numerical aperture and central wavelength as

$$\delta x = \delta y \approx 0.37 \frac{\lambda_0}{NA}. \tag{1.19}$$

It should be noted that this formula is only valid at the focal plane of the system, so the lateral resolution is distinct at different z-locations in an OCT scan. This means that the objective lens also determines the axial (z) field of view.

The axial resolution of the system is determined by the coherence length, and can be shown to be

$$\delta z = \frac{2 \ln 2}{n\pi} \frac{\lambda_0^2}{\Delta\lambda}. \tag{1.20}$$

There exists a fundamental tradeoff between resolution and penetration depth. In particular, a larger central wavelength yields a poorer resolution, as can be seen in Eqn 1.20, but also tends to yield a larger penetration depth due to the phenomenon of Rayleigh scattering [1].

Axial field of view is also limited by *occlusion*. The A-Scan formula is valid under the assumption that the reflectivities within the sample are small. Of course, if the reflectivities were too small, almost no signal would be picked up by the device. On the other hand, if the reflectivities are too large, very little light is transmitted deeper into the sample. This is important for *in vivo* imaging of the cochlea, as the bony walls of the cochlea obscure views due to their high reflectivity. In this case, the practical axial field of view cannot extend beneath this opaque structure.

1.6 Practical Implementation

Certain details of the OCT data processing are omitted in the discussion above. In particular, the spectrometer data recorded in the wavelength domain must be interpolated to the wavenumber domain, and the signal must be windowed before the Fourier transform is applied.

At experiment time, data acquisition is controlled through a C++ program that uses ThorLabs' SpectralRadar SDK. This program uses built-in processing functions to perform the background acquisition, subtraction and normalization, as well as the interpolation, windowing and Fourier transform steps. The output of this program is a complex-valued time series of A-Scans. All other processing steps are performed after experimentation in MATLAB.

Data presented in this thesis were recorded using a ThorLabs Telesto 320 system. Details of this system are presented in App A.2.

Chapter 2: Introduction to Cochlear Mechanics

The sensation of hearing in mammals begins with a pressure wave entering the ear canal (EC). EC pressure generates a displacement at the tympanic membrane, which in turn pushes on the ossicles of the middle ear. The last of these bones, the stapes, pushes on the oval window (OW) of the cochlea and modulates the pressure of the fluid within [9].

The cochlea is then responsible for transforming this fluid pressure into a neuronal stimulus, and it does so in a fascinating manner. In this chapter, I will discuss the anatomy of the cochlea and properties of its sound-evoked displacement responses.

2.1 Anatomy

The bony exterior of the cochlea is in the shape of a snail's shell, with the tip of this shell being referred to as the *apex*, the broad end as the *base*, and the central axis of the spiral as the modiolus. Fig 2.1 **A** shows a cartoon of the cochlear spiral with the base on the top-left and the apex at the center. Fig 2.1 **B** shows a "midmodiolar" cross-section of the guinea pig cochlea, with the modiolus at the center, the base at the bottom and the apex at the top [10].

Three fluid ducts, or scalae, fill the cavity (see Fig 2.1 **B**). The scala vestibuli (SV) contains sodium-rich perilymph, and abuts the oval window (OW) at the base where the stapes footplate supplies the cochlea's input force. Opposite SV is the scala tympani (ST), also containing perilymph, which abuts the round window membrane (RWM) at the base [10]. SV and ST are connected at the most apical position in the cochlea known as the helicotrema. As fluid is pushed in and out at the OW, the perilymph volume is conserved by an opposite displacement at the RWM (illustrated by the dotted arrows in Fig 2.1 **A**)[12, 13].

Between SV and ST lies scala media (SM), filled with potassium-rich endolymph. SV and SM

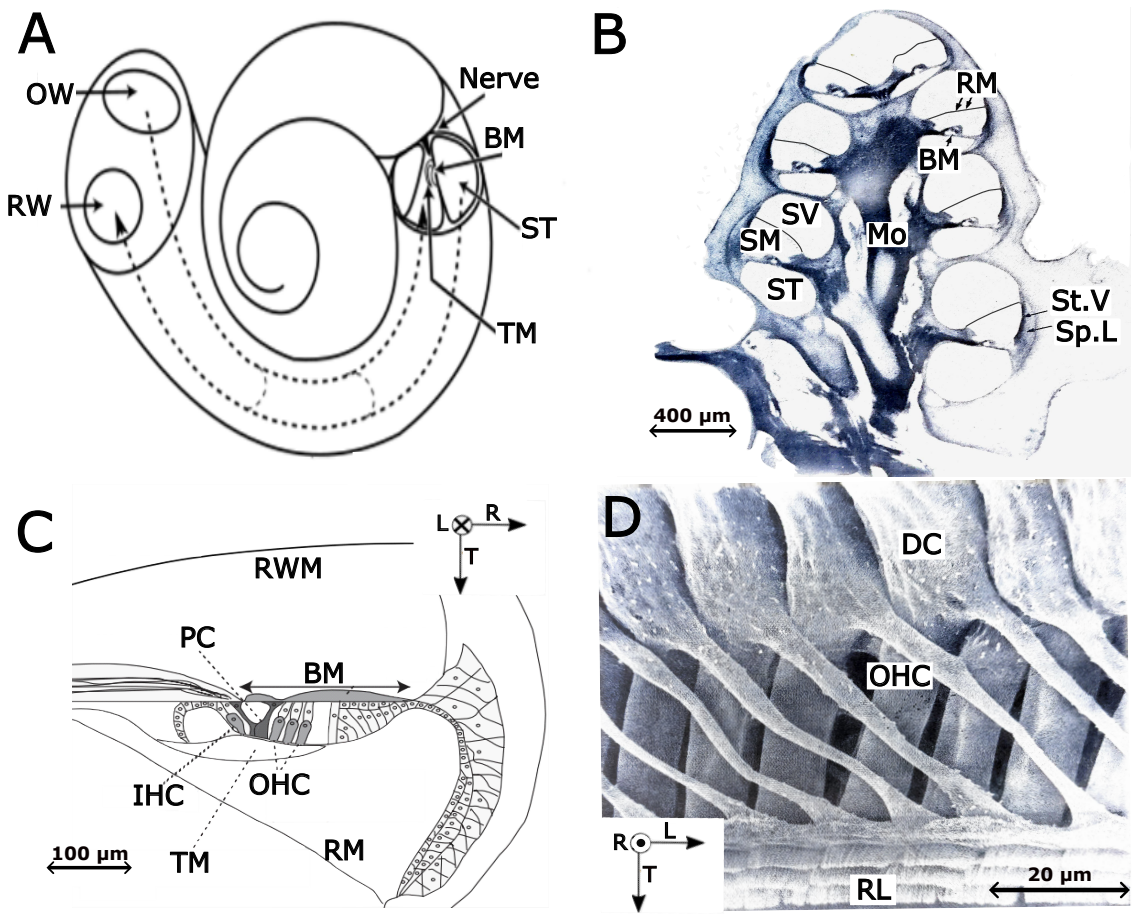

Figure 2.1: Four images of the cochlea at different scales and perspectives. **A** – Cartoon of the cochlear spiral showing the gross shape and the direction of fluid pressure procession. The stapes of the middle ear pushes on the oval window (OW), and fluid volume is conserved by an opposite displacement at the round window (RW). A cross-section is shown revealing the three fluid compartments. The basilar membrane (BM), scala tympani (ST), tectorial membrane (TM) and auditory nerve are labeled, and shown in more detail in the other panels. **B** – Midmodiolar cross-section of the guinea pig cochlea from Slepecky, with the base at the bottom and the apex on top [10]. Labels: scala tympani (ST), scala media (SM), scala vestibuli (SV), modiolus (Mo), basilar membrane (BM), Reissner's membrane (RM), stria vascularis (St.V), spiral ligament (Sp.L). **C** – Cartoon longitudinal cross-section of the organ of Corti complex at the base. The anatomical coordinate axes at this position, longitudinal (L), radial (R) and transverse (T), are shown at the top-right. Labels: round window membrane (RWM), pillar cells (PC), basilar membrane (BM), inner hair cells (IHC), outer hair cells (OHC), tectorial membrane (TM), Reissner's membrane (RM). **D** – Scanning electron micrograph of a radial cross-section of the outer hair cells (OHCs) and Deiters cells (DC) in mole rat, from Raphael *et al* [11]. This cross-section is an orthogonal plane to the cartoon in **C**, as can be seen from the coordinate axes in the bottom-left. The phalangeal processes can be seen proceeding from the OHC-DC junction to the reticular lamina (RL) in the base-to-apex direction. The stereocilia of the OHCs can be seen on the RL, as the TM has been removed in the imaged preparation.

are separated by Reissner's membrane (RM) which is only two cells thick. It is thus believed to be present only to separate endolymph from perilymph, having no mechanical function [14, 10]. Due to the ionic compositons of the fluids in the cochlea, endolymph has a $\sim 80-90$ mV potential relative to perilymph [9]. This is known as the endocochlear potential (EP). The stria vascularis (St.V) lies on the outer wall of the cochlea, and acts as an ion pump responsible for maintaining ion homeostasis [10]. When the animal dies, the St.V ceases to function and the EP disappears [15, 9]. The EP is analogous to a battery that allows for active processes to occur in a healthy cochlea.

SM and ST are separated by the organ of Corti complex (OCC) – a diverse collection of cells and membranes ultimately responsible for the transduction of fluid pressure into neurotransmitter response. The OCC also spirals around the cochlea, having a cross-sectional symmetry along the spiral (i.e. the arrangement of cells at each position along the spiral is nearly identical to that at any other position), although physical properties and relative sizes vary. This symmetry can be seen in Fig 2.1 **B**.

The portion of the OCC nearest to ST is the basilar membrane (BM) – a compliant and permeable membrane composed of collagen fibers spanning its width [10]. BM width tends to increase from base to apex in most species, while its stiffness tends to decrease in the same direction [16, 10, 17].

The cochlea is naturally endowed with an anatomical coordinate system [18]. The *longitudinal* direction points around the spiral from base to apex. At each longitudinal position, the direction from ST to SM normal to the BM is called the *transverse* direction. Within a longitudinal-transverse cross-section, the direction parallel to the BM pointing from the modiolus to the outer wall of the cochlea is called the *radial* direction. Positions closer to the modiolus are referred to as *medial* while positions closer to the outer wall are called *lateral*. Fig 2.1 **C** and **D** show these coordinate axes at distinct cross-sections.

The longitudinal, radial and transverse unit vectors $(\hat{\mathbf{l}}, \hat{\mathbf{r}}, \hat{\mathbf{t}})$ form a right-handed coordinate system at each point in space. Relative to Cartesian space, this system varies depending on position within the cochlea, so this defines a coordinate *frame* which I call the *anatomical coordinate frame*.

Fig 2.1 **C** shows a cartoon of a cross-section of the OCC with certain key structures labeled. The inner and outer pillar cells (PC) are stiff structural elements, together with the BM defining a triangular fluid space known as the tunnel of Corti. Fluid spaces within the OCC contain cortilymph, which is nearly identical to perilymph in ionic composition [10].

Medial of the inner pillar cells are the inner hair cells (IHC), which are topped with hair-like *stereocilia*. Stereocilia are made up mostly of tightly packed actin filaments, and are connected in bundles by small filaments known as tip links [19]. The stereocilia are embedded (or nearly embedded) in the collagenous, gel-like tectorial membrane (TM) [20].

Lateral of the outer pillar cells are the outer hair cells (OHC), which come in rows of three per longitudinal cross-section. These cells are also topped with stereocilia embedded in TM. The cell walls of the OHC contain the motor protein prestin, which expands and contracts according to the transmembrane potential [21, 22].

Fig 2.1 **D** shows a radial cross-section through the OHCs. The stiff surface of the OCC near the TM from which the sterocilia extrude is known as the reticular lamina (RL). The RL is an actin-packed plate with tight “ion-proof" junctions, so it is in part responsible for maintaining the ionic difference between endolymph and perilymph [10].

Between the OHCs and BM are the Deiters cells (DC) which are tightly packed in a lattice along the BM [23]. They connect to the BM by the rope-like Deiters stalk, and they connect to the RL by stiff phalangeal processes (PP) which slant towards the apex [23]. Through the Deiters stalk and the PP, the DCs lie in tension between BM and RL. Motion at the junction between the OHCs and DCs (OHC-DC) is a central consideration of this thesis.

2.2 Physiology

2.2.1 Tonotopy and the Traveling Wave

The displacement responses of all structures within the OCC are characterized by *tonotopy* – a place-frequency mapping where each single-tone stimulus causes a maximum vibration at a specific longitudinal position (i.e. the frequency has a “best place," or BP) [9, 24]. Reciprocally, each

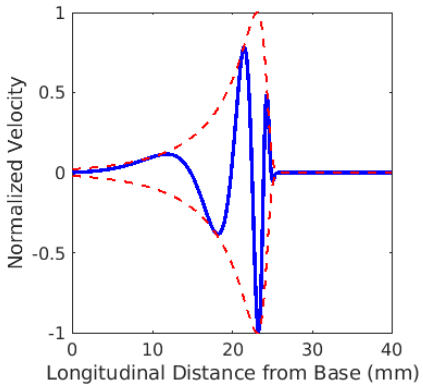

Figure 2.2: Snapshot of the traveling wave in response to a 1 kHz tone at a single instance in time. The blue curve is transverse velocity of the BM normalized by the response's maximum value over the length of the cochlea. Dashed red curves indicate the envelope of the response, or the maximum magnitude that the wave will reach across time. Features to note are the peaking of the response at a best place and the spatial variation of the wavelength (or reciprocally, the wavenumber). A Wentzel-Kramers-Brillouin (WKB) approximate solution to a box model of the cochlea was used to generate these curves (see App B for more details).

longitudinal position has a stimulus frequency at which its displacement response is maximized (best frequency, BF or characteristic frequency, CF).

Positions at the apex, where structures are generally more compliant, are tuned to lower frequencies. Basal positions, where structures are generally stiffer, are tuned to higher frequencies [25, 9]. This is due to the presence of a *traveling wave* sustained by the OCC, which can be thought of as originating at the base and transporting energy along the cochlea's length [9]. As the OCC impedance varies along its length, so too does the degree to which the OCC resonates in response to the stimulus. In response to a pure tone, the magnitude will be largest near where the impedance yields a resonant frequency equal to the stimulus frequency [12, 13, 26, 27]. Apical to this region, energy is rapidly lost [12, 13].

Fig 2.2 shows a simulated snapshot of the traveling wave at one instance in time across the longitudinal length of the cochlea in response to a pure tone stimulus. The velocity at the base is small, and grows in magnitude until reaching the BP for this frequency after which the magnitude rapidly falls off. Notably, due to the variation in impedance along the length of the cochlea, the wavelength of the traveling wave also varies in space – it is longer at the base, and gets shorter

as the traveling wave approaches its best place. Curves were generated using a two-dimensional Wentzel-Kramers-Brillouin (WKB) approximate solution to a box model of the cochlea [28, 29, 30, 31, 27, 26], described further in App B.

Fig 2.3 shows both reciprocal forms of tonotopy. Panel **A** shows the modeled transverse velocity amplitudes as a function of longitudinal position (i.e. position along the spiral) in response to tones at 1.0 kHz and 3.2 kHz. The magnitude responses (normalized to their maximum values) show that each stimulus frequency causes a maximum displacement at a specific location, with the response falling off rapidly in space apical to this position. The lower-frequency tone causes a maximal response at a more apical position, while the higher-frequency tone causes a maximal response at a more basal position.

Panel **C** shows the reciprocal phenomenon. Observing modeled frequency responses at two distinct longitudinal locations shows that the more apical position is tuned to a lower frequency than the more basal position.

The phase responses of panels **B** and **D** are also of fundamental importance, and can be explained through the physics of the traveling wave. The traveling wave is usually modeled as satisfying the wave equation with an impedance-dependent wavenumber [32, 24, 26]. That is, the wavenumber of the traveling wave also varies in space with its magnitude increasing from the base to the BP (this can be seen in Fig 2.2, recalling that wavenumber is the inverse of wavelength). By definition,

$$k = \frac{\partial \phi}{\partial x},$$

where k is the wavenumber, ϕ is the phase and x is distance along the cochlea. The increase in the magnitude of k corresponds to phase varying more quickly from base to apex. An apical-traveling wave (that is, a wave traveling in the positive x direction) has a negative wavenumber, so the phase *decreases* from base to apex, and more quickly as the BP is approached. The phase responses of panels **B** and **D** show this characteristic *phase accumulation*. The flattening of the phase response past this point is due to the wavenumber becoming very small above the best place[1] [35, 27].

[1]There is some technicality here – this formulation assumes that the traveling wave is a true single-mode wave,

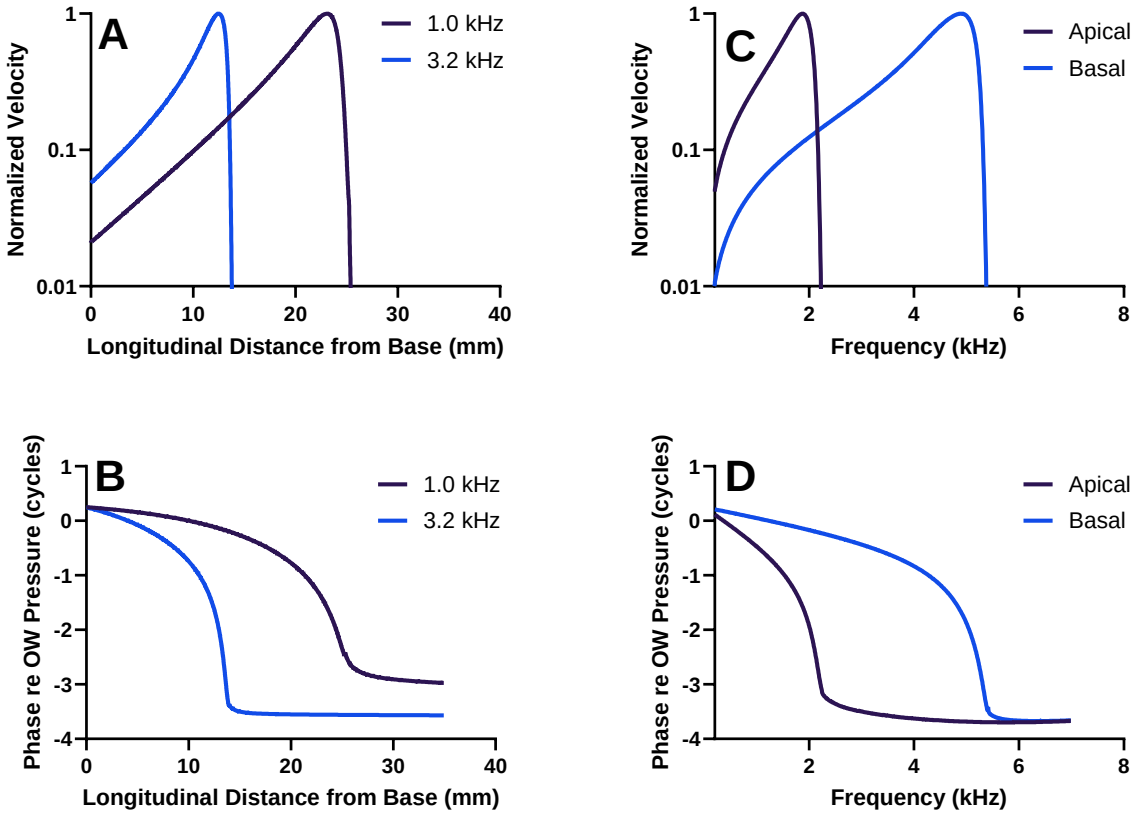

Figure 2.3: Normalized velocity responses showing the two reciprocal tonotopy phenomena apparent in cochlear mechanics. A Wentzel-Kramers-Brillouin (WKB) approximate solution to a box model of the cochlea was used to generate these curves (see App B for more details). **A**, **B** – Velocity magnitude and phase responses along the longitudinal axis in response to single-tone stimuli at 1.0 kHz and 3.2 kHz. Responses show that frequencies are tuned to specific locations. Values are normalized so that the maximum velocity is 1. **C**, **D** – Frequency responses at two longitudinal positions, showing that each position is tuned to a specific frequency. The more apical position is 18 mm from the base and the more basal position is 9 mm from the base.

2.2.2 Nonlinearity and Zwuis Stimuli

Tonotopy is present even in unhealthy or dead preparations [9], but there are a number of other interesting features that arise from active processes in healthy cochleae *in vivo*. The presence of EP allows a current to pass through the hair cells [36]. At the tips of the stereocilia are mechanically gated ion channels [37, 38, 10]. When the stereocilia are mechanically stimulated, they modulate the current due to the opening and closing of these channels [38], creating what is known as the mechano-electric transduction (MET) current. The channel open probability is a statistical phenomenon, and follows a Boltzmann distribution. That is, the current-displacement relationship is well-modeled by a Boltzmann function with an operating point determined by the resting position of the stereocilia [37]. This means that the current-displacement relationship is approximately linear for small-magnitude displacements, but the current saturates for high-magnitude displacements when nearly all channels are fully open or fully closed. Considering the current-displacement gain, there is a *compressive nonlinearity* – while the current magnitude monotonically increases as displacement magnitude increases, the gain monotonically *decreases* as the current saturates.

This is illustrated in the modeled response in Fig 2.4, where samples of a Boltzmann function, an input stereocilia deflection waveform and the output MET current are shown. In panel **B**, the Boltzmann curve can be seen as approximately linear between its minimum and maximum values, but it saturates for large stereocilia displacements (asymmetrically, due to a potentially shifted operating point). This is clear in panel **D**, where high-magnitude values of the input stimulus lead to saturation in the MET current.

Mechanical stimulation of the stereocilia occurs *in vivo* due to shear motion of the tectorial membrane. The resulting MET current through the IHCs controls the release of neurotransmitter responsible for the sensation of hearing. The current through the OHCs modulates the transmembrane potential, causing the expansion and contraction of prestin [22, 21]. This phenomenon is termed *electromotility*.

The displacement responses are larger in magnitude and more sharply tonotopically tuned in

having a well-defined single wavenumber. Discussion of alternate approaches can be found in [33, 34, 26].

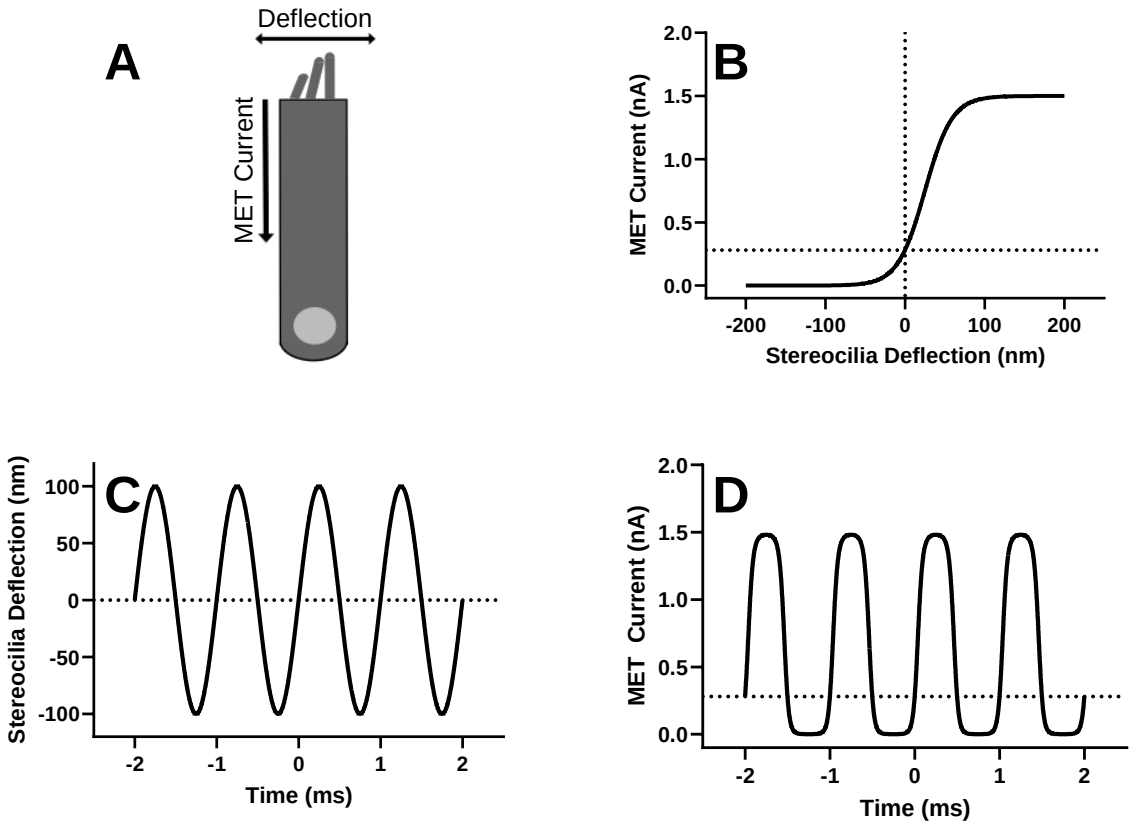

Figure 2.4: Illustration of compressive nonlinearity in mechano-electric transduction (MET) using a Boltzmann function model. **A** – A cartoon of an OHC. Deflecition of the stereocilia results in a variation in the current through the cell. **B** – A graph of a sample Boltzmann function relating stereocilia deflection and MET current. The maximum current, linear range and operating point have been set to 1.5 nA, 75 nm and -25 nm respectively. Parameters were chosen based on *in vitro* recordings of OHC MET current [37]. Dotted lines indicate the operating point. **C** – A sample time-domain stereocilia deflection signal, acting as an input to the MET channel. Its frequency is 1 kHz and its peak-to-peak amplitude is 200 nm. **D** – MET current output for the input in **C**, according to the Boltzmann relationship in **B**. The dotted line indicates the MET current when the displacement is 0. Note that the signal saturates at both the minimum and maximum current values, but more-so at the minimum due to the negative operating point.

living cochleae than in dead cochleae [39], so some collection of active processes known as the *cochlear amplifier* must be responsible for this phenomenon. The nature of this cochlear amplifier is the subject of some controversy, but it is generally agreed that the presence of prestin in the OHCs plays some role. This is evinced by the fact that blocking the current through the OHCs significantly reduces cochlear displacement responses [40, 41]. The leading theory is that electromotile motions of the OHCs produce forces that provide power to the BM and RL, amplifying and further tuning responses on a cycle-by-cycle basis [42, 43].

As the stereocilia current exhibits compressive nonlinearity, so too does the electromotile response of the OHCs [6, 41, 42]. This means that the amplification provided by electromotility is compressively nonlinear, and thereby other structures within the OCC, such as BM, will also exhibit compressive nonlinearity in their displacement responses [6, 40, 39]. In an unhealthy or dead cochlea, the cochlea will behave linearly as no such active process takes place.

Due to this nonlinearity, OCC displacement gain in a healthy cochlea will depend on the magnitude of the stimulus, or sound pressure level (SPL). This is most commonly measured in dB SPL, defined on a logarithmic scale as

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

where P is pressure in Pa and $P_0 = 20$ μPa – approximately the lowest pressure magnitude at which humans can hear.

Fig. 2.5 shows *in vivo* displacement measurements from the gerbil base illustrating compressively nonlinear SPL-dependence. Panel **A** shows displacement magnitudes measured from the BM in response to stimuli at 60, 70 and 80 dB SPL. At all SPLs, frequency tuning is present, and increasing SPL increases displacement magnitude at most frequencies. Panel **B** shows the gain, derived by dividing the displacement responses by the input pressure at EC. It can be seen that higher SPLs result in *lower* gain near the BF, where displacements are maximal. That is, there is compressive nonlinearity in the BM displacement response. Panels **C** and **D** show displacement and gain in the same preparation at OHC-DC, where compressive nonlinearity is again present. Unlike BM, this nonlinearity can be seen across the whole frequency range rather than just near

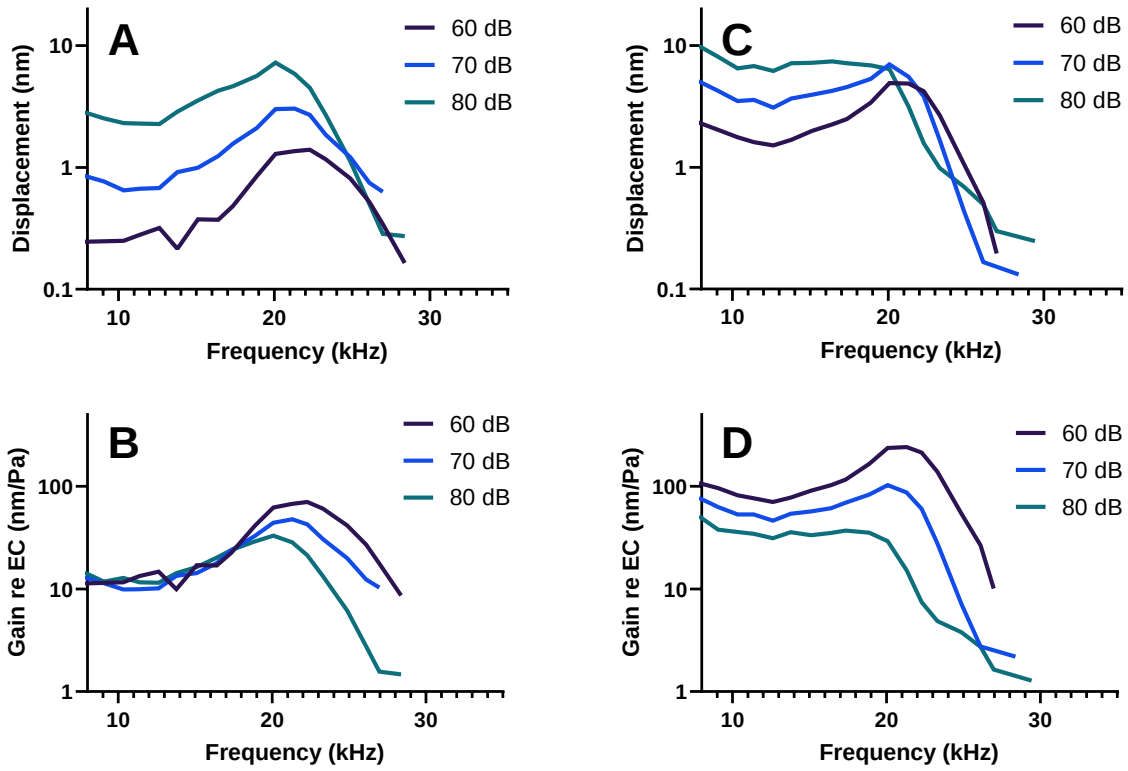

Figure 2.5: Displacement and gain magnitude responses from the gerbil base *in vivo* (Ge988), illustrating the manifestation of compressive nonlinearity. **A**, **B** – Displacement and gain at the BM in response to a multitone Zwuis stimulus containing 25 frequency components at 60, 70 and 80 dB SPL. **C**, **D** – The same, but at the OHC-DC junction.

the BF.

This nonlinearity in the gain is also called *scaling nonlinearity* to distinguish it from a second form of nonlinear behavior apparent in the cochlea – *distortion*. In response to a single tone, the nonlinear form of the current-displacement relationship will induce harmonics in the current, i.e. there will be measurable responses at integer multiples of the stimulus frequency. In response to a superposition of two tones at frequencies f_1 and f_2, *distortion products* appear [44]. These are responses at superimposed integer multiples of the stimulus frequencies, such as a response at $2f_1 - f_2$. Distortion products are a sign of a healthy cochlea, as they are a result of active nonlinearity.

The health of a cochlea can thereby be assessed using what are known as distortion product otoacoustic emissions (DPOAEs), which are a measurable pressure response coming *out* of the cochlea and into the ear canal in response to two-tone stimuli [45]. DPOAEs are recorded throughout our experiments to ensure that the animal's hearing has not deteriorated.

However useful, distortion also disallows the measurement of a broadband frequency response by the presentation of a superposition of uniformly spaced tones[2]. Tone complexes known as *Zwuis stimuli* have been developed to avoid the impact of distortion [46]. In short, Zwuis stimuli are superpositions of tones that are spaced in such a way that distortion products generated by any subset of the tones do not overlap with the stimulus frequencies. This means that the frequency responses can be efficiently measured using a single multitone stimulus rather than by separately measuring the response to single tones. All displacement data presented in this thesis were recorded using Zwuis stimuli.

Because we have used Zwuis stimuli, we do not consider distortion except for when using DPOAEs to assess cochlear health. As such, when I use term "nonlinearity," I am generally referring to the compressive scaling nonlinearity present in the SPL-dependent displacement responses.

[2]The frequency response is not defined for non-linear systems in general, however we can define a pseudo-frequency response by observing the response to individual tonal stimuli at a single input SPL. This is what I mean when I say "frequency response."

2.2.3 Displacement Characteristics within the Organ of Corti Complex

Until recently, only the motion of the BM could be measured *in vivo*, as it is the first surface visible to an interferometer when looking through the RWM or a hole drilled in the cochlea (called a cochleostomy) [39]. The motion of the BM at a single longitudinal position is tuned to its BF due to tonotopy. For stimulus frequencies below about 0.7BF (the *sub-BF region*), the displacement of the BM grows linearly with pressure with approximately the same magnitude as if the cochlea were passive. However, for stimulus frequencies between about 0.7BF and 1.2BF (the *BF region*), the displacement responses of the BM exhibit compressive nonlinearity. This can be seen in Fig 2.5 **A** and **B**.

With the advent of OCT, experimenters can now measure displacements inside of the OCC, and particular attention has been paid to the motion of OHCs. The OHCs exhibit a tuned frequency response as well, but their properties are quite different from those of the BM. For example, OHC-DC appears to move with a higher magnitude than the rest of the OCC at most stimulus frequencies, so the region is often referred to as the *hotspot* [6, 40, 47].

Of particular interest is the behavior of OHC-DC in the sub-BF region. Displacements in this region are larger in healthy cochleae than in damaged or dead cochleae, which is intuitive as electromotility should boost the region's displacement response [40, 6, 47, 48]. However, in this region the BM motion is nearly identical between healthy and damaged or dead cochleae [39]. Similarly, OHC-DC motion exhibits compressive nonlinearity across the entire frequency range in response to Zwuis stimuli [40, 6, 47, 49]. This contrasts with BM motion which is linear sub-BF. That is, the BM displacements appear mechanically isolated from this nonlinearity until about 0.7BF. This difference is clear in Fig 2.5.

One phenomenon of fundamental importance to this thesis is that OCT-measured OHC-DC phase responses vary significantly between experimental setups – in particular, when responses are measured at different longitudinal locations or with different beam axis orientations relative to the sample anatomy. To understand the mechanics of hearing, this relative phasing is critical. For example, if one interprets the OHC-DC as providing mechanical power to the BM on a cycle-by-

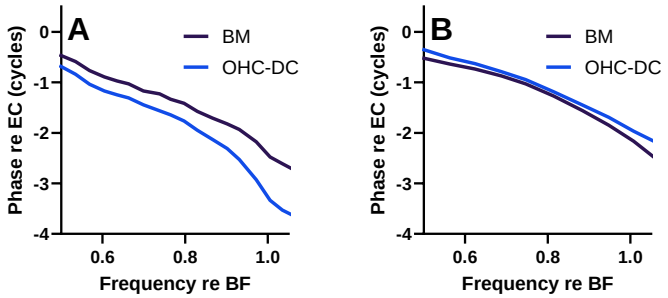

Figure 2.6: Phase responses from BM and OHC-DC in the gerbil base *in vivo*. **A** – Responses from Ge961, measured near the 50 kHz region. OHC-DC lags BM across frequency, with this lag increasing with frequency. **B** – Responses from Ge967, measured near the 26 kHz region. OHC-DC leads BM across frequency, with this lead varying non-monotonically with frequency.

cycle basis, the relative phase of these two structures could make the difference between supplying energy to the traveling wave and removing energy from the traveling wave [50, 43, 51].

Fig 2.6 gives an example of this phase response discrepancy. In both panels, BM and OHC-DC phase responses from the gerbil base *in vivo* are shown. Data in panel **A** are from a further basal position (BF of about 50 kHz) than those in panel **B** (BF of about 26 kHz). In the first dataset, OHC-DC *lags* BM across frequency, while in the second dataset OHC-DC *leads* BM across frequency.

How can we interpret and synthesize the available wealth of OCT data if they disagree on this critical point, and how can we develop experiments that provide interpretable data within the context of this apparent ambiguity? These are central questions of this thesis.

Chapter 3: Resolving Anatomical Discrepancies in Uniaxial Recordings

3.1 Limits of Uniaxial Recordings

As OCT records displacement signals at all positions along its optic axis simultaneously, a single M-Scan can be used to measure BM motion and OHC-DC motion at once. These *uniaxial* measurements are efficient, and they are a common method for reporting OCT-measured displacements in the cochlear mechanics field. This has been valuable in determining qualitative features of motion in the OHC-DC region, which were not measurable before the advent of OCT.

For example, uniaxial measurements show that OHC-DC moves more in healthy cochleae than in damaged or dead cochleae, even sub-BF where this is not true of BM [6, 41, 40, 48]. They have also shown that OHC-DC exhibits tonotopy, and moves with a larger magnitude than BM [6, 5]. This is especially true sub-BF where OHC-DC uniquely exhibits scaling nonlinearity in response to Zwuis stimuli (see Fig 2.5, as well as [3, 47, 6]).

However, uniaxial recordings have also provided significantly disparate results when compared across experimental setups, both in qualitative and quantitative ways. In particular, reported phases of intra-OCC structures relative to BM appear to differ characteristically between experiments. For example, while Dewey *et al.* found that RL lags BM across frequency in the mouse apex, Ren and He found that RL leads BM at low frequencies and lags it at high frequencies in the gerbil base [42, 52, 53]. Experiments from our lab in the gerbil base show that OHC-DC leads BM across frequency near the 25 kHz region, but other experiments from our lab and Cho *et al.* find that OHC-DC lags BM across frequency near the 50 kHz region (see Fig 2.6, as well as [5, 6, 48, 47]).

These differences could be entirely due to species or frequency location. Such explanations would have extreme implications about the very nature of cochlear mechanics, as they would imply that relative phasing can vary characteristically and still produce the sensation of hearing.

This would upend the theory of cycle-by-cycle amplification, where relative phase would be a controlling factor of power transfer from OHC to BM [42, 50, 43].

It is more likely that these discrepancies arise (at least in large part) from a limitation of uniaxial recordings: the optic axis of the OCT device differs with respect to anatomical coordinates between experiments. The optic axis is often chosen based on experimental convenience, and not with the aim of achieving a measurement in a specific anatomical direction. This has two significant consequences.

The first is that the relative positions of the structures being measured along an M-Scan are known in *optical* coordinates, but may not be known in *anatomical* coordinates. The former are simply a feature of the experimental setup, while the latter are of physiological importance.

The second is that OCT-measured displacements are one-dimensional projections of three-dimensional motion onto the optic axis (Sec 1.4). Motion along the z axis has no inherent physiological significance, as it is a weighted superposition of the physiologically meaningful components of motion. These features deeply impact the interpretation of OCT-measured displacements in the cochlea, and could be largely responsible for the major characteristic phase discrepancies described above.

The projection issue has been discussed by cochlear mechanics researchers since 2018, although largely in a qualitative or theoretical sense [6, 54]. In this chapter, I discuss a method that I have developed for the quantitative characterization of these issues inherent to uniaxial recordings. This is achieved by determining a local relationship between optical and anatomical coordinates. The method can be used to determine anatomical distances between positions in a single uniaxial recording, which allows for these distances to be reported and accounted for. It can also be used to determine the projection map of three-dimensional motion onto the measurement axis. This method was presented in *Frost et al., 2022*, and was featured on the cover of the February 2022 issue of the *Journal of the Acoustical Society of America* [18].

This chapter includes displacement data and B-Scans recorded from the gerbil base *in vivo*. The details of the animal preparation are described in App A.1, and the details of the technology

employed in these experiments are described in App A.2.

3.2 Illustration of Skew

The distinction between optical and anatomical coordinates is not immediately apparent when taking uniaxial recordings, as B-Scans taken at large angles relative to the BM normal may still resemble an anatomical cross-section of the cochlea. This is illustrated in Fig 3.1, where an anatomical drawing of a radial-transverse cross-section is compared to a B-Scan taken through the gerbil RWM.

Panel **B** shows an anatomical drawing of an rt-cross-section, and panel **C** shows a B-Scan (an optical yz-cross-section) taken through the gerbil RWM. Because OCT is label-free, we have to use known anatomy to distinguish structures. Using similarity between the B-Scan and the anatomical drawing, we can recognize the anatomy as shown in panel **D**. At experiment time, we generally orient ourselves in the cochlea by finding such a B-Scan that resembles an anatomical drawing.

This observed similarity between the anatomical drawing and the recorded B-Scan might lead one to believe that the B-Scan captures an rt-cross-section, i.e. $\hat{\mathbf{z}} = \hat{\mathbf{t}}$, but this is not actually the case. This is illustrated in Fig 3.2, where again a B-Scan is compared to an anatomical rt-cross-section. When we observe an orthogonal cross-section to this B-Scan (the xz-cross-section in **D**) it becomes clear that the B-Scan in **B** is *not* in the rt-plane.

Comparing the xz-cross-section in **D** with the anatomical drawing in **C**, we can see that the optical z axis has large components in both the longitudinal and transverse directions. This distinction between optical and anatomical coordinates has been termed *skew* [6, 18], and requires the analysis of a volume scan to characterize. While skew was considered in at least one early publication on OCT-measured displacements in the cochlea [6], its effects have often been ignored or have gone unreported.

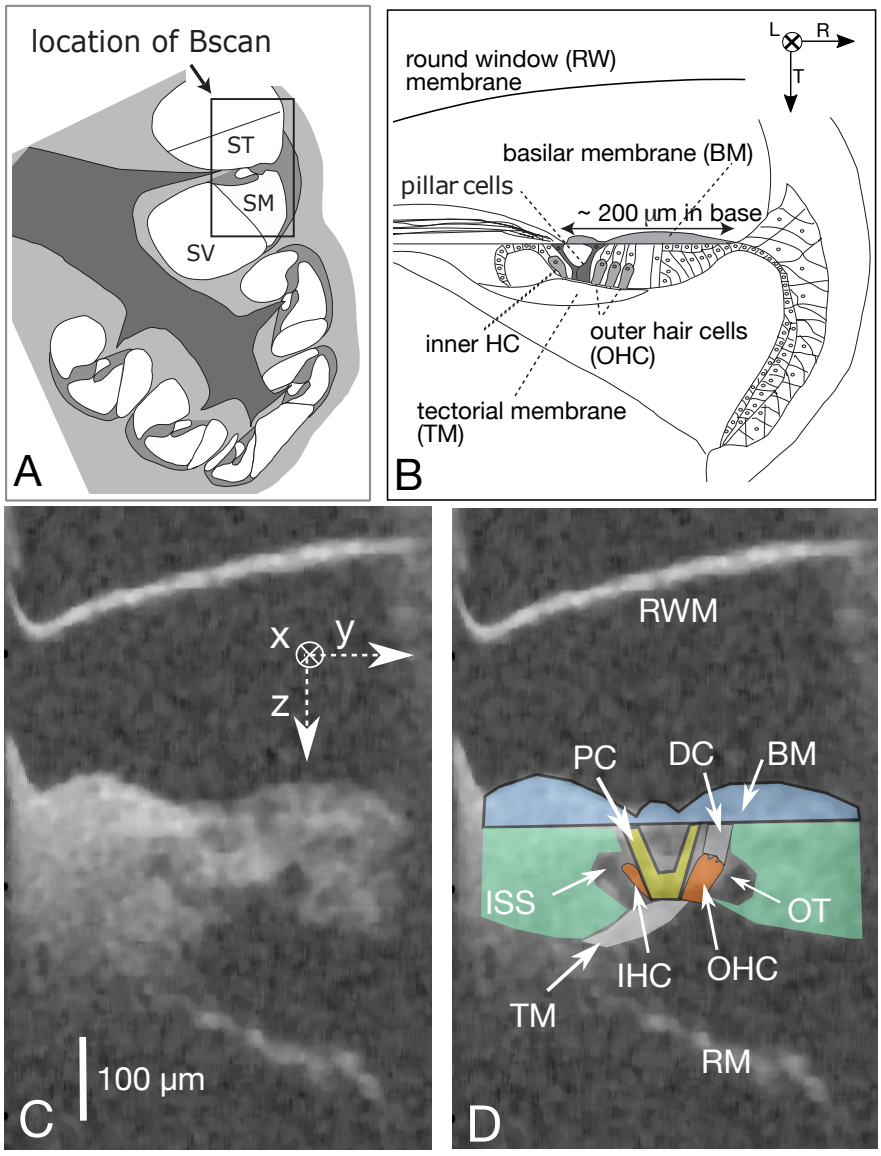

Figure 3.1: Anatomical and optical cross-sections in the gerbil base (~25 kHz region), resembling one another. **A** – Cartoon of the gerbil cochlea with the base at the top and apex at the bottom. B-Scans are taken through the RWM, capturing a basal cross-section of the OCC. **B** – Anatomical drawing of an *rt*-cross-section of the gerbil OCC with key anatomical structures labeled. The drawing is oriented as we would expect the OCC to appear in a B-Scan according to panel **A**. **C** – A B-Scan taken through the gerbil RWM as shown in the cartoon of panel **A**. B-Scans are inherently label-free, so anatomical structures are not immediately recognizable. **D** – The B-Scan from **C** with key anatomical structures labeled and color-coded. These labels are determined by comparison of the B-Scan in **C** with the anatomical drawing in **B**. ST = Scala tympani, SM = Scala media, SV = Scala vestibuli, HC = Hair cell, PC = Pillar cells, DC = Deiters cells, ISS = Inner sulcus space, OT = outer tunnel, RM = Reissner's Membrane. This figure is adapted from Fig 1 of *Frost et al., 2023* [49].

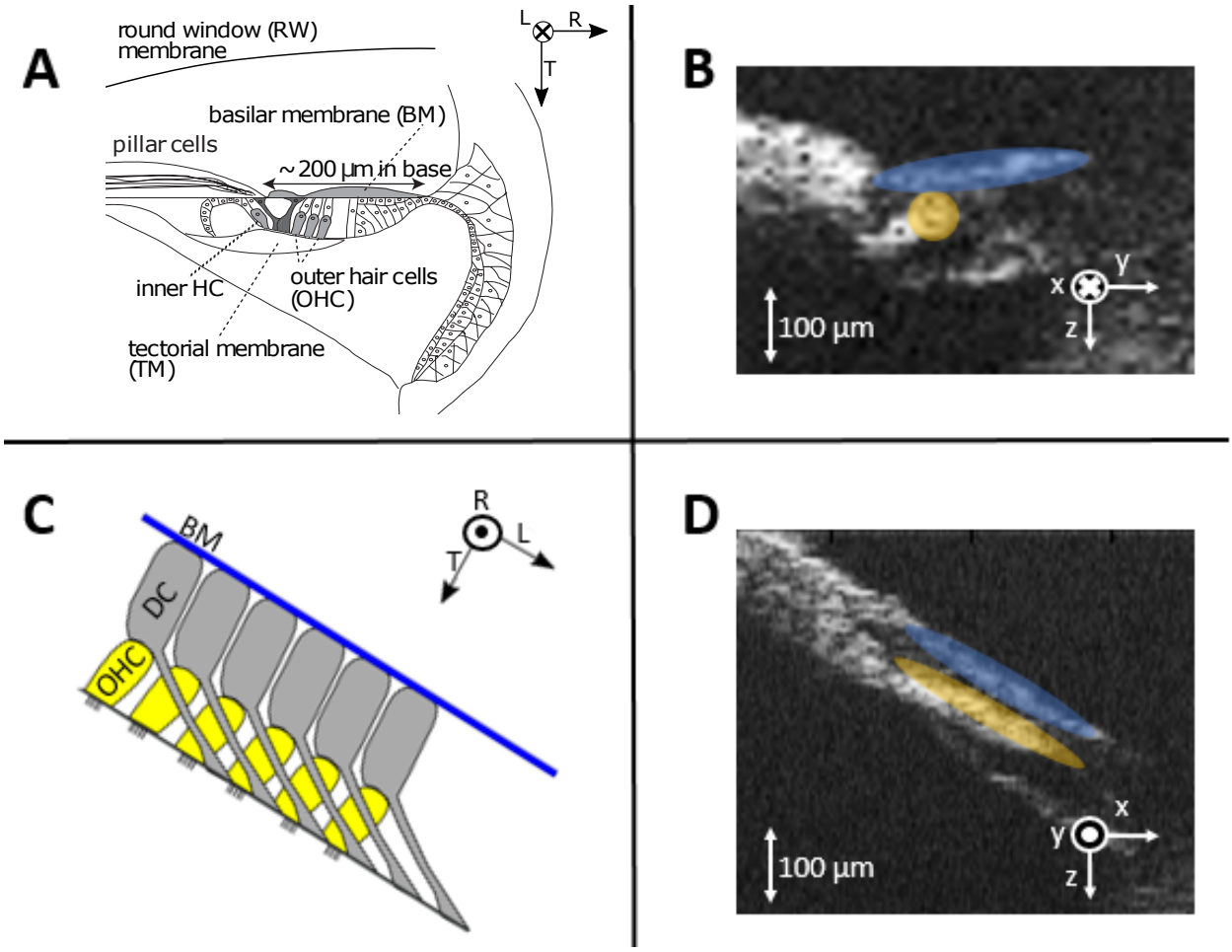

Figure 3.2: Two orthogonal views from an OCT volume taken through the gerbil RWM, revealing the difference between anatomical and optical coordinates. **A** – Cartoon of an anatomical rt-cross-section from Fig 3.1. **B** – Optical yz-cross-section with BM (blue) and OHC-DC (yellow) marked according to the known anatomy of **A**. **C** – Cartoon of an anatomical lt-cross-section through the OHC and DC, tilted. **D** – Optical xz-cross-section from the same volume as **B**, with BM and OHC labeled according to the anatomical cartoon in **C**. Together, these B-Scans show that the optical coordinate axes are distinct from the anatomical coordinate axes.

3.3 Challenges in Interpreting Skewed Motion Measurements

Before introducing the mathematical framework within which skew is quantified, some qualitative impacts can be noted. Consider an M-Scan that measures both the BM and OHC-DC in the preparation displayed in Fig 3.2. The distance between measured BM and OHC in the z direction can be measured simply by counting the number of pixels between the locations. However because the z axis has components in both t and l, as we move along the z axis we are moving not only towards SM into the OCC (i.e. in the transverse direction), but also towards the apex (i.e. in the longitudinal direction). Following a single column in Fig 3.2 **D** illustrates this feature of the measurement.

As described in Sec 2.2.1, a difference in longitudinal location corresponds to a shift in BF due to tonotopy. This leads to characteristically different behavior of both the magnitude and phase responses, as more apical positions are tuned to lower frequencies (see Fig 2.3). That is, when we compare the responses of different structures in a uniaxial recording, motion characteristics will differ because of both physiological distinctions between the structures *and* a tonotopic shift. We would prefer to isolate physiological differences, as tonotopic differences are already well-understood [9, 25].

Moreover, our measured displacements are projections onto the z axis. This means that both longitudinal and transverse components will superimpose, potentially interfering in complex manners. Longitudinal, radial and transverse motion are all likely important for the functioning of the healthy cochlea. *Longitudinal* motion is significant in fluid, and it is expected to be the primary direction of energy flow in the cochlea at lower frequencies [12, 13]. Moreover, feedforward models have been developed that suggest that the active forces of more basal OHCs may amplify responses at more apical positions [55, 56].

MET current is modulated by *radial* motion of the stereocilia, determined by differential radial motion between RL and TM [37, 38]. This controls both neural responses and the electromotile motion of the OHCs [21]. The OHCs themselves also have a significant radial tilt [10].

The BM is expected to move mostly in the *transverse* direction [6, 49], and the OHC electromotile motion has a significant transverse components as well. Many models consider only this component of motion in the OCC, and certain cycle-by-cycle amplification theories suggest that this is the primary direction of power transfer from OHC to BM [32, 26, 43, 42, 50].

What physiological significance, then, does z-direction motion have? If one cannot isolate any of the three physiologically meaningful components of motion with a uniaxial measurement, this question is not easily answered.

3.4 Methods

I have developed a method to quantify the distinction between anatomical and optical coordinates within uniaxial motion measurements. The method, which relies on a locally static approximation of anatomical coordinates, can be used to account for skew and allow for experimenters to measure at locations within the same tonotopic cross-section. This allows isolation of phenomena due to physiological differences between structures and those caused by tonotopic shifts along a single M-Scan. It also allows for experimenters to quantify and report the projection map of anatomical components of motion onto the optic axis, which adds important context to uniaxial displacement measurements[1].

3.4.1 Locally Planar Model

Relating optical and anatomical coordinates is nontrivial, because the mapping between them changes as a function of space. Anatomical coordinates are defined relative to the BM as it spirals around the cochlea. At any point, a smooth surface such as the BM can be locally modeled as a plane – specifically, by its tangent plane at that point. The quality of this model degrades as we move further from the point of approximation, and the region over which the planar approximation holds well is determined by the curvature of the BM. This is shown in Fig 3.3, which compares the optical and anatomical coordinate systems with an approximating plane.

[1] To supplement the text, a video detailing application of the method was used in a ThorLabs seminar and can be found here: https://www.youtube.com/watch?v=zEwOOnm-T3Y

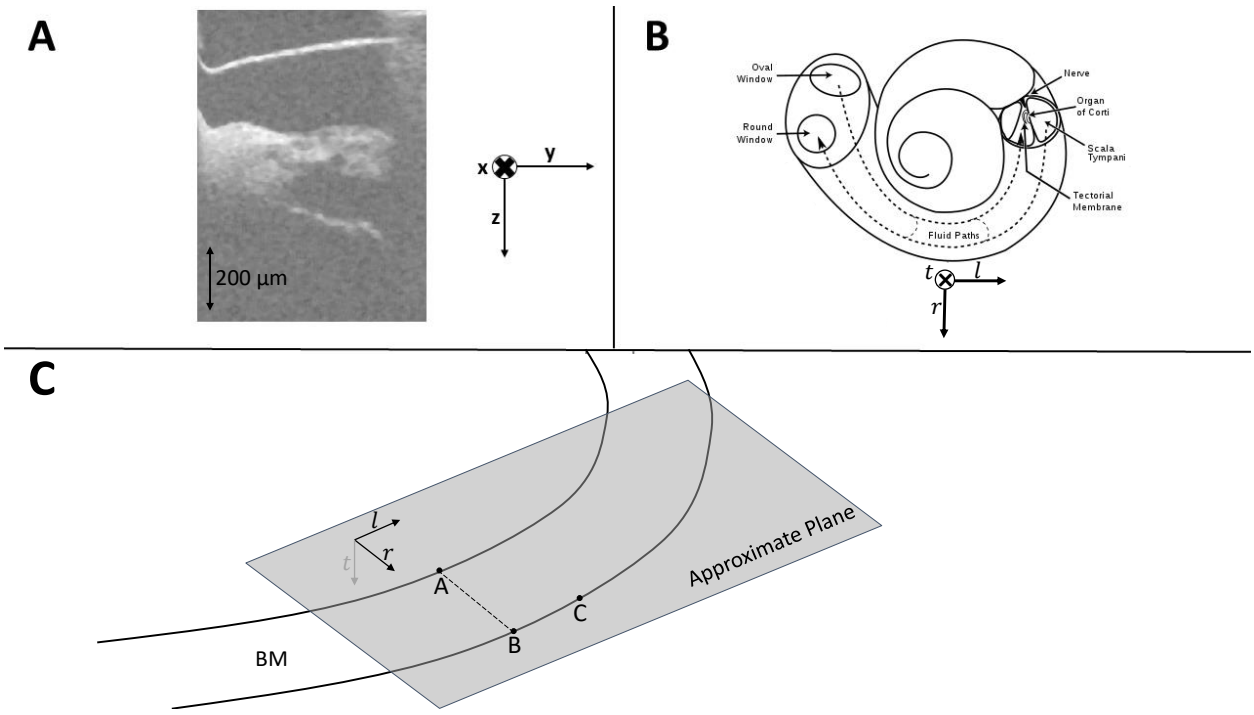

Figure 3.3: Illustration of optical, anatomical and approximate anatomical coordinates. **A** – B-Scan taken through the RWM of gerbil (same as in Fig 3.1) with optical coordinates labeled. **B** – Cartoon of the cochlea with the anatomical coordinate frame displayed as a single location. The frame spirals about the cochlea. **C** – Cartoon of a small region of the BM as a flat, spiraling sheet. The approximate plane is shown in gray, with the approximate anatomical coordinates labeled. These coordinates are static relative to Cartesian space. It can be seen that the planar approximation only appears reasonable over a small region of the BM, degrading as one moves further along the BM spiral. The dashed line indicates where the orienting B-Scan from **A** would lie. The labeled points **A**, **B** and **C** are used to determine the approximating plane according to the process described in Sec 3.4.2. This figure is adapted from Fig 2 of *Frost et al., 2022* [18].

The BM is flat, and the longitudinal radius of curvature is maximal at the base where the spiral is at its widest [10]. This suggests that in the region at which we are measuring, the BM may be well-modeled by a plane over a sufficient range to accurately determine anatomical distances of structures along M-Scans.

To quantify this, I observed an OCT volume and manually selected ten points on the medial edge of the BM, spanning 440 μm longitudinally. The true longitudinal direction traces a curve passing through these points, and the BM is only well-approximated by a plane where this curve is well-approximated by a line. A best-fit line and a correlation matrix measuring the nearness of this line to the curve was determined over a 440 μm range, as well as a 220 μm subset of this range.

Fig 3.4 shows the projections of these points and the best-fit lines onto the optical coordinate planes. Quantitatively, the R^2 values exceed 0.94 for the full 440 μm range and 0.98 over the 220 μm range. Qualitatively, the best-fit line over 220 μm seems to be an excellent approximation for the longitudinal direction. So long as we do not significantly over-step a 220 μm longitudinal range from the point at which the tangent plane is found, we are well-justified in using this approximation. As such, for the remainder of this chapter l, r and t will refer to the *approximate* anatomical coordinate axes, which are static relative to optical coordinates.

3.4.2 Determining the Approximate Anatomical Plane

Having justified the planar approximation, we now look to determine the approximate anatomical plane in terms of optical coordinates. A plane is determined by any three non-colinear points **A**, **B** and **C**. We can find the plane by selecting any three non-colinear points on the BM within a volume. I have developed a specific method for selecting these three points that also determines the anatomical l, r and t axes in optical coordinates.

As discussed above, we generally begin experiments by finding a B-Scan such as that shown in Fig 3.1 **C** so that we can orient ourselves relative to the anatomy of the OCC. We call this the *orienting B-Scan*. This method operates on a volume scan in which the orienting B-Scan is the center yz-plane.

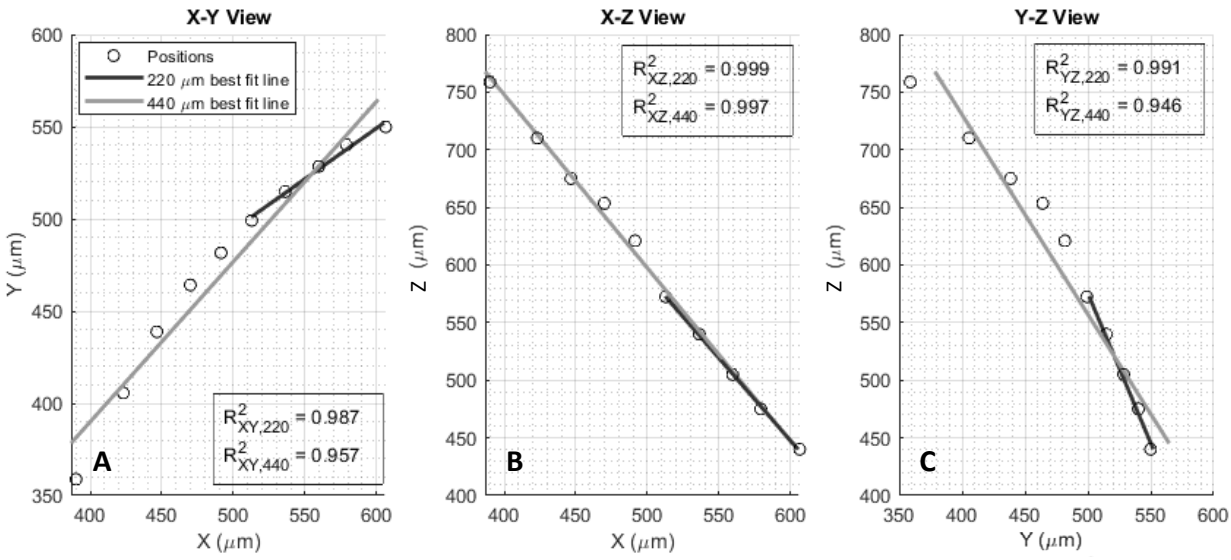

Figure 3.4: Analysis of the validity of the planar approximation in the gerbil base. The ten black circles indicate points selected along the medial edge of the spiraling BM over a 440 μm longitudinal span. Each panel shows the projection of the selected points onto a different optical plane, along with a projection of the best-fit line. Black lines show the planar approximation over half of the span (220 μm), while gray lines show the planar approximation over the full span. R^2 values are shown in each panel for both ranges. This figure is adapted from Fig 9 of *Frost et al., 2023* [49].

The left-hand B-Scan in Fig 3.5 shows an example of an orienting B-Scan, in which the medial and lateral ends of the BM are perceptible. We select our first two points, **A** and **B**, at these ends respectively. This defines a line segment that well-approximates the BM in this cross-section, and lies within the approximate plane. These points and the corresponding line segment are shown in the left-hand B-Scan of Fig 3.5.

We then consider a distinct parallel B-Scan within the same volume – i.e. another yz-cross-section at a different x position. We choose a cross-section that is apical of the orienting B-Scan for reasons that will soon become apparent. The third point **C** is picked at the most lateral position on the BM within this cross-section. This third point is guaranteed to be non-colinear with **A** and **B**, as it lies in a different optical cross-section. This means that these three selected points determine an approximate plane. An example of a selection for **C** is shown in the right-hand B-Scan of Fig 3.5, along with the projection of the resultant approximating plane onto this B-Scan.

Due to how we have selected these three points, we can determine the three anatomical unit vectors, $\hat{\mathbf{l}}$, $\hat{\mathbf{r}}$ and $\hat{\mathbf{t}}$. First, points **B** and **C** both lie on the most lateral edge of the BM, and thus they have the same radial and transverse coordinates. They are separated only longitudinally, and by construction **C** is apical of **B**. This gives the longitudinal unit vector:

$$\hat{\mathbf{l}} = \frac{\mathbf{C} - \mathbf{B}}{||\mathbf{C} - \mathbf{B}||}. \tag{3.1}$$

The transverse vector is defined as being normal to BM, pointing from ST to SM. As we have three non-colinear points in the plane, we can write the normal vector as a cross-product (being careful to choose the correct order so that $\hat{\mathbf{t}}$ points towards SM):

$$\hat{\mathbf{t}} = \frac{(\mathbf{A} - \mathbf{B}) \times (\mathbf{C} - \mathbf{B})}{||(\mathbf{A} - \mathbf{B}) \times (\mathbf{C} - \mathbf{B})||}. \tag{3.2}$$

The final coordinate vector, $\hat{\mathbf{r}}$, can then be found using that fact that $\{\hat{\mathbf{l}}, \hat{\mathbf{r}}, \hat{\mathbf{t}}\}$ is a right-handed coordinate system. That is,

$$\hat{\mathbf{r}} = \hat{\mathbf{t}} \times \hat{\mathbf{l}}. \tag{3.3}$$

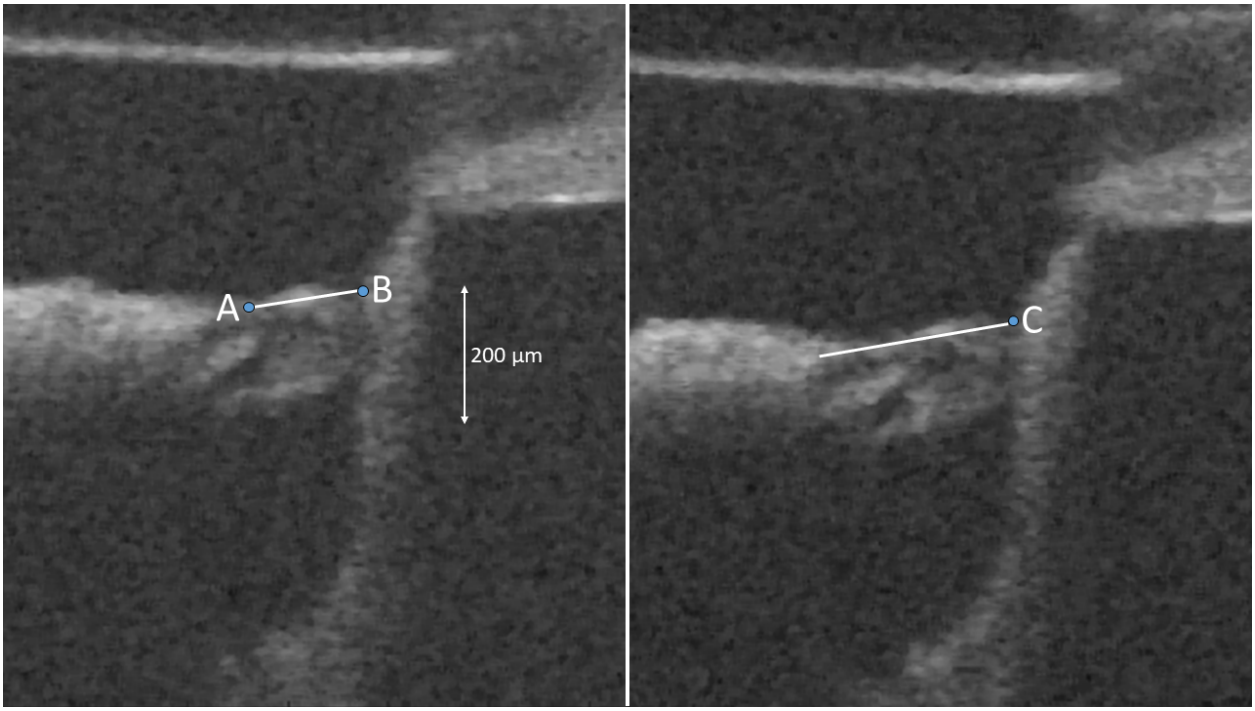

Figure 3.5: Example of the plane approximation process using a volume scan from Ge900 (25 kHz region, taken through RWM *in vivo*). Two B-Scans from a single volume scan, 20 μm apart. Points **A** and **B** are chosen in the first B-Scan (left). This determines the line segment that approximates the BM in this cross-section, shown to the user for verification before moving to the next step. In the second B-Scan (right), **C** is chosen, completely defining the plane. The projection of that plane onto the second B-Scan is shown to the user for verification before the values describing the plane are saved. This figure is adapted from Fig 5 of *Frost et al., 2022* [18].

This method requires minimal user input and only two B-Scans to perform. Further implementation details are presented in Sec 3.4.6. It should be noted that the distance between the two selected B-Scans is important. If the B-Scans are too close, small local variations in the OCC shape or user input error may make for a poor approximation of the BM over a large range. On the other hand, if the B-Scans are too far apart, the planar approximation may break down in accordance with the analysis in Fig 3.4. Generally an x distance of about 20 μm is used, and robustness of the resulting components can be assessed by considering the effect of performing the process with several choices of inter-B-scan distance up to about 200 μm.

3.4.3 Determining the Coordinate Transformation

For a single orientation of the OCT system, we write a position $\mathbf{p}$ in the optical basis, $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$, as

$$\mathbf{p} = p_x\hat{\mathbf{x}} + p_y\hat{\mathbf{y}} + p_z\hat{\mathbf{z}} = \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix}_{op} . \tag{3.4}$$

This same position can be written in the anatomical basis as

$$\mathbf{p} = p_l\hat{\mathbf{l}} + p_r\hat{\mathbf{r}} + p_t\hat{\mathbf{t}} = \begin{pmatrix} p_l \\ p_r \\ p_t \end{pmatrix}_{anat} . \tag{3.5}$$

With the planar approximation giving us static anatomical coordinates, the mapping relating these optical and anatomical components is uniquely defined by a rotation and translation. In the present context, the definition of an origin is arbitrary as we are only concerned with relative distances between points in the OCC. As such, we can assume the origin for both coordinate systems is the same, and say that optical and anatomical coordinates are related only by a rotation.

The rotation can be written as an orthogonal 3×3 matrix, which we can find by setting up a simple system of linear equations. Above, we found anatomical basis vectors $\hat{\mathbf{l}}$, $\hat{\mathbf{r}}$ and $\hat{\mathbf{t}}$ in *optical* coordinates. We also know the forms of these vectors in anatomical coordinates, as they are the standard ordered basis elements. For example,

$$\hat{\mathbf{l}} = \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix}_{op} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}_{anat} . \tag{3.6}$$

To determine the matrix U that maps anatomical to optical coordinates, we can consider the following system of three equations:

$$U \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \hat{\mathbf{l}}_{op}, \quad U \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \hat{\mathbf{r}}_{op}, \quad U \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \hat{\mathbf{t}}_{op}. \tag{3.7}$$

This gives the form of U simply as the matrix whose columns are the anatomical basis vectors written in optical coordinates:

$$U = \begin{pmatrix} l_x & r_x & t_x \\ l_y & r_y & t_y \\ l_z & r_z & t_z \end{pmatrix}. \tag{3.8}$$

The elements of this matrix were directly computed in the planar approximation process. The inverse, which maps optical coordinates to anatomical coordinates, is simply the transpose U^T (as U is orthogonal).

3.4.4 Form of the Projection

The motion measured in an M-Scan is a projection of the anatomical motion onto the beam axis, z. The form of this projection is determined by the $\hat{\mathbf{z}}$ unit vector's components in anatomical coordinates. The matrix U^T maps optical to anatomical coordinates, and $\hat{\mathbf{z}}_{op}$ is the elementary basis vector $(0\ 0\ 1)^T$. This gives

$$\hat{\mathbf{z}}_{anat} = U^T \hat{\mathbf{z}}_{op} = \begin{pmatrix} l_z \\ r_z \\ t_z \end{pmatrix}. \tag{3.9}$$

A three-dimensional displacement at position $\mathbf{p}$ written in anatomical coordinates, $\mathbf{d}(\mathbf{p}) = (d_l(\mathbf{p})\ d_r(\mathbf{p})\ d_t(\mathbf{p}))^T$, is measured by the OCT system as a projection onto the $\hat{\mathbf{z}}$ vector. We can

write the measured one-dimensional displacement δ as the inner product

$$\delta(\mathbf{p}) = \hat{\mathbf{z}}_{anat} \cdot \mathbf{d}(\mathbf{p}) = d_l(\mathbf{p}) l_z + d_r(\mathbf{p}) r_z + d_t(\mathbf{p}) t_z. \tag{3.10}$$

With the planar approximation method, we can find l_z, r_z and t_z, and thereby the form of this projection. However as the projection is a 3-to-1 map, we cannot recover $\mathbf{d}$ or even any of its three components from a measured δ.

Still, this is a very important object to understand and report as it offers context as to which components of motion are being most significantly represented in a uniaxial measurement. For example, we computed these components in Ge900 (Fig 3.5) where measurements where taken near the 25 kHz region through the RWM. In this sample, which is representative of the angle at which we often measure in this region, we found $l_z = 0.8$, $r_z = -0.1$ and $t_z = 0.6$.

That is, when measuring from this angle, the most represented anatomical component of motion is longitudinal, with significant contribution from the transverse motion component as well. Radial motion contributes far less, as the optical axis is nearly perpendicular to the radial axis.

3.4.5 Applications of Exploring Anatomical Space in Optical Coordinates

In a uniaxial scan, one can quantify the impact of skew and report the longitudinal or radial distance between measured structures using the coordinate relationship derived above. For example, one can determine how far apical measured OHC-DC lies from measured BM. This would be performed by first considering the measured point at the BM as the origin in both anatomical and optical coordinates. The other points measured along the M-Scan can all be written as $\Delta z \hat{\mathbf{z}}_{op}$, where Δz is distance of the structure of interest from the BM along the optic axis.

In anatomical coordinates, we can write these same positions as $\Delta z \hat{\mathbf{z}}_{anat}$, giving the longitudinal, radial and transverse distance between these points and the BM. These value can then be reported to provide important context to measured displacements.

Reciprocally, one might be interested in measuring two structures related in a known *anatom-*

ical manner. To do so, one would need to position the scanner so that these two points could be measured. Consider first an M-Scan where we measure the BM at one point. We wish also to measure OHC-DC in the same tonotopic cross-section, i.e. related only by a transverse translation. Assuming OHC-DC is about 45 μm transverse from the BM (the origin in this context), this is written optically as $45\,U\hat{\mathbf{t}}\ \mu$m. A second M-Scan can then be taken so that this point is captured. Data from these positions would be more intuitively compared than measurements from different tonotopic cross-sections, simplifying physiological analysis.

3.4.6 Implementation

A simple graphical program was developed for the experimenter to select the three non-colinear points defining the approximating plane. This consists of 1) choosing a spacing between the B-Scans at which the points are selected, 2) selecting points **A** and **B** in the first B-Scan and 3) selecting point **C** in the second B-Scan. An example of this program in operation is shown in Fig 3.5.

The left panel shows the B-Scan in which the user is first asked to select points **A** and **B**. The program provides some feedback to ensure that the user has selected reasonable points, showing the line segment that s/he has chosen to represent the BM in this cross-section. This line segment is the projection of the soon-to-be-determined approximating plane. If the user does not care for his/her choice, the program will restart and ask for new selections.

The right panel shows the second B-Scan, shown after the experimenter has verified the choice of line segment in the first B-Scan. This B-Scan is some amount of slices away from the first, with the precise difference specified by the user. After selecting **C** the plane is determined, and a line segment is drawn showing the projection of this plane onto this second B-Scan. Once again, the user has the choice to restart if this segment poorly represents the BM here.

Once the plane has been verified the program saves the points selected and the change of basis matrix U. Now skew correction is possible, and the projection map for the motion has been determined. Other experiments, such as ones in which we look to measure structures in the

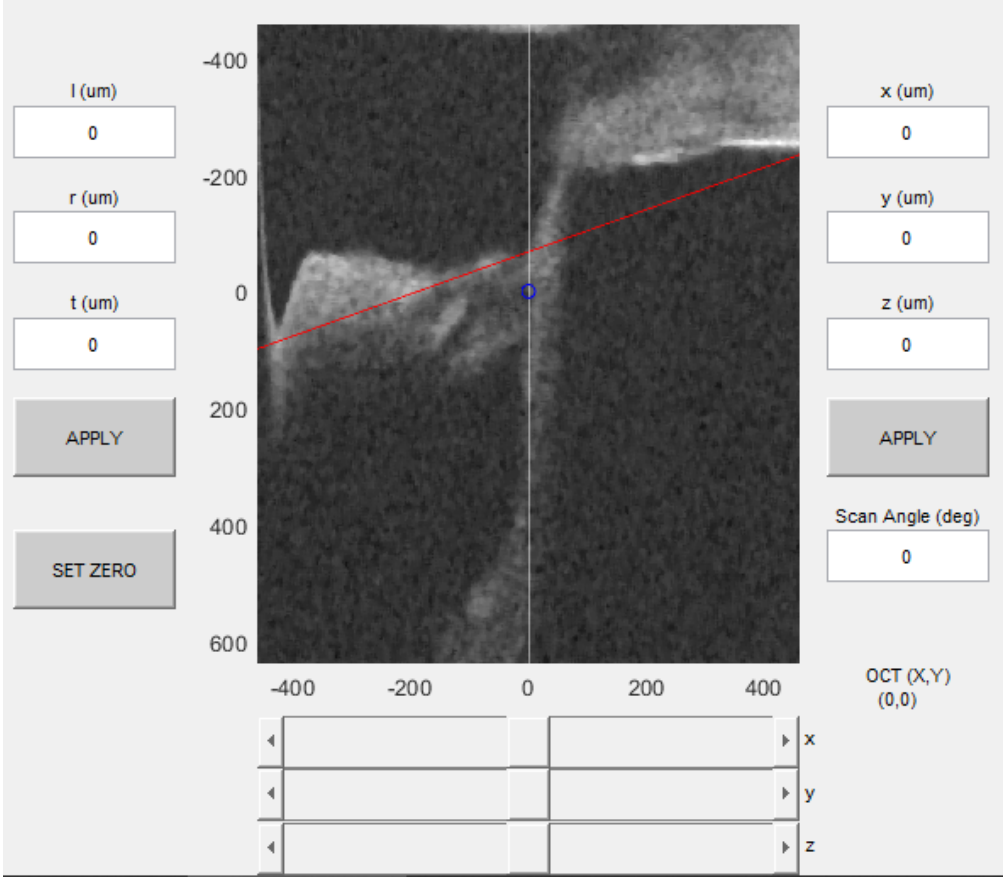

Figure 3.6: A still of the orienting GUI used to explore volumes in both optical and anatomical coordinates. Shown is the orienting B-Scan from experiment 900. The coordinate values of the blue point's location are shown in the anatomical (left) and optical (right) systems, and the A-Scan in which this point lies is shown in white. The red line is the planar approximation of the BM projected onto the displayed B-Scan. This figure is adapted from Fig 3 of *Frost et al., 2022* [18].

same longitudinal cross-section, are made possible by the ability to explore anatomical and optical coordinates simultaneously.

In practice, this is challenging without visual aid as the space is three-dimensional. I developed a graphical user interface (GUI) using MATLAB that allows the user to orient her/himself within the volume in either coordinate system. This is known as the *orienting GUI*. A still of this program is presented in Fig 3.6, where the center B-Scan of the volume in question is shown. The red line indicates the projection of the approximating plane onto this B-Scan.

The program works by moving about the blue circle using either the sliders on the bottom or manual input of coordinates in either coordinate system. As the x coordinate varies, so too does

the displayed B-Scan and the red line, which corresponds to the projection of the approximating plane.

Moving the point in either system updates both sets of displayed coordinate values. We also have the option to set the anatomical origin to be at any point (with the “SET ZERO" button) so that the coordinates are more easily interpreted.

The last elements in the GUI, on the bottom-right, are the scan angle and the OCT coordinates X and Y. This third coordinate system has gone unmentioned until now, but is experimentally vital. The optical coordinates discussed up to now are defined relative to a volume scan, but when we use ThorLabs software to acquire an M-Scan these coordinates are not used.

Instead, there is a *global coordinate system* irrespective of volume which acts as the input to ThorImage software. If the volume is taken at an angle relative to this global coordinate system, we have to reverse this rotation in computing the global coordinate value that corresponds to the volume-specific optical coordinates.

3.5 Results

3.5.1 Quantification of Skew in the Gerbil Base

First we used the planar approximation to determine by how much structures measured in a uniaxial scan differ in longitudinal location. We consider a single example taken through the RWM in the 25 kHz region of the gerbil cochlea. To achieve measurements in this region, the measurement axis must have a longitudinal component in the towards-apex direction (see Fig 3.2). This can yield longitudinal distances on the order of the sizes of structures in the cochlea.

Fig 3.7 shows an example of positions measured along a single M-Scan alongside a cartoon representing the anatomical structures. Longitudinal distances of measured structures from the BM were computed using our method, and are labeled for each point. In this example, a measured point within OHC lies 46 μm apical of the measured BM point within the same scan. The diameter of an OHC is 10 μm and adjacent OHCs are nearly in contact, so this distance is significant on the scale of the cochlea’s anatomy. The change in frequency location will have an impact on the measured

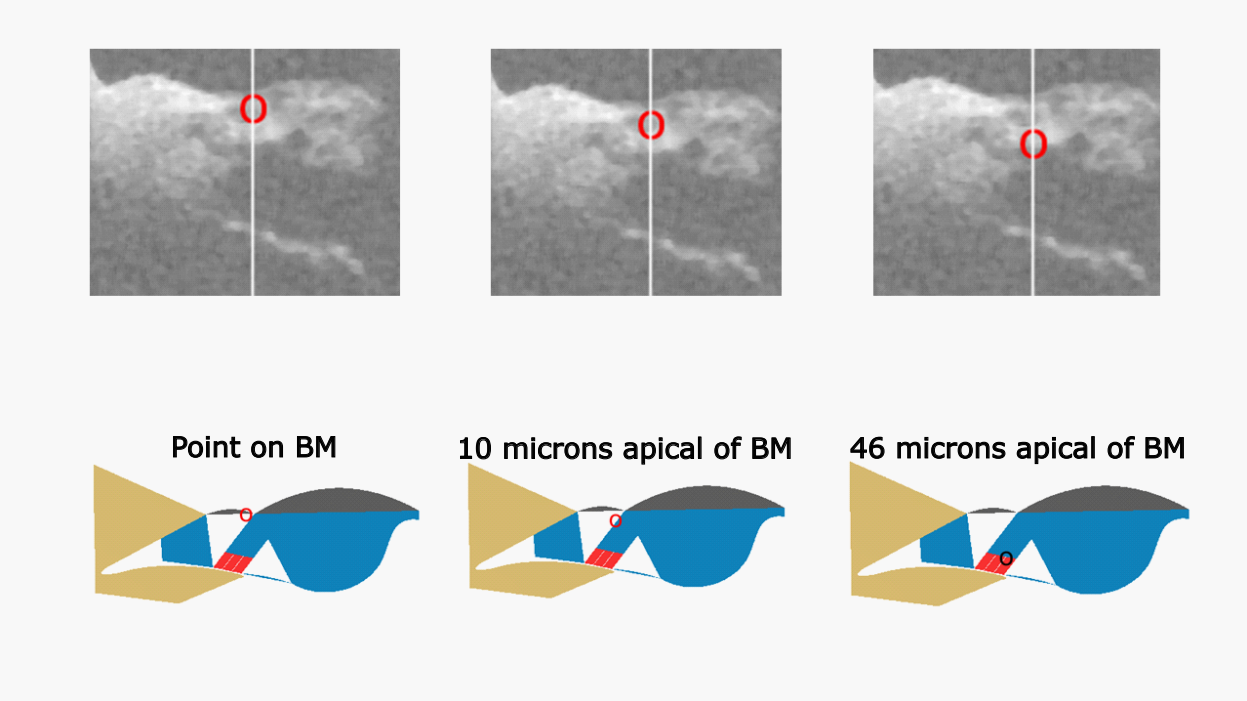

Figure 3.7: Illustration of longitudinal skew in Ge900. The panels in the top row shows the same B-Scan and A-Scan (white line), with three measured points isolated. The first is on the BM, to which other measurements are referenced. Cartoons in the bottom row show the anatomical structures at which the measured points are expected to lie, and are labeled by their longitudinal distance from BM. In particular, measured OHC-DC lie 46 μm apical of BM.

point's sound-evoked displacement responses due to tonotopy (Sec 2.2.1). This is explored in the following subsection.

3.5.2 Skew Correction at Experiment Time

This large skew can be accounted for at experiment time using the orienting GUI. An example of this is shown in Fig 3.8. The process of skew correction begins with taking a single M-Scan containing the structures of interest. In our case, these are BM and OHC-DC. The position of this M-Scan is shown as the white line in Fig 3.8 **A**, where I have set the anatomical origin to be the measured point at the BM. That means that any point such that $t = 0$ lies precisely on the BM.

I want to measure OHC-DC and BM in the same cross-section by accounting for skew. First, the skew is quantified by moving down along the z axis (that is, within the same M-Scan) to the measured OHC-DC position. This can be seen in Fig 3.8 **B**. Observing the anatomical coordinates

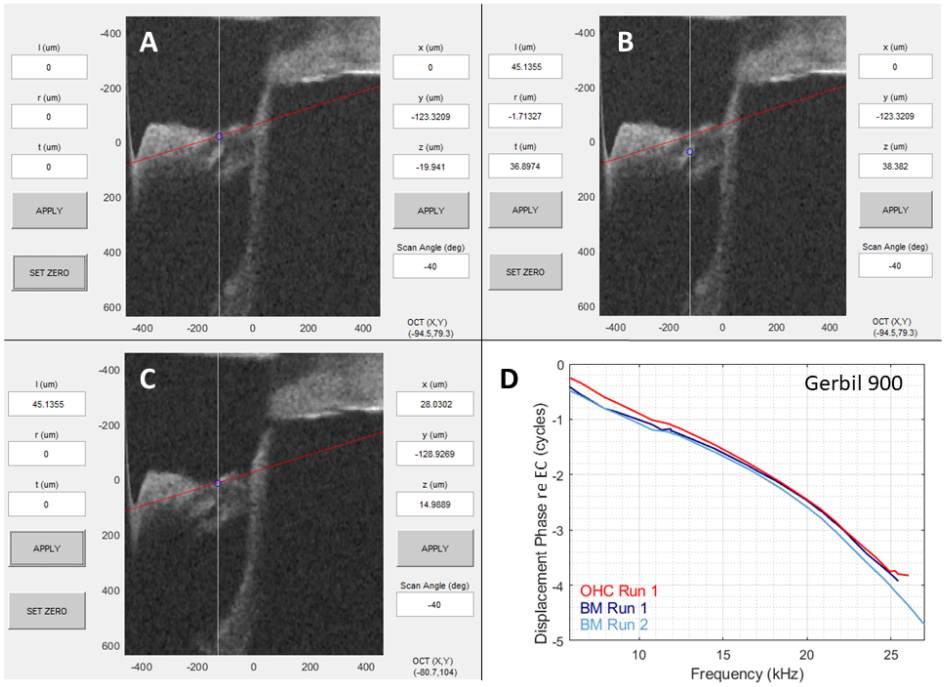

Figure 3.8: Use of the orienting GUI to measure BM and OHC in the same anatomical cross-section. **A** – To begin, we select an A-Scan containing BM and OHC. The zero point is set to be on the BM along this A-Scan, as shown here. **B** – We move the z slider so that the blue point is on the OHCs and the A-Scan has not changed. Only the z optical coordinate changes, whereas all three anatomical coordinates have changed. The l value indicates that the OHCs are about 45 μm apical of the BM in this A-Scan. **C** – We find the measurement location necessary to measure BM motion in the same cross-section as OHC from the previous A-Scan by moving to the point with the same l position but with $r = t = 0$. The OCT (X, Y) coordinates on the bottom right are the output we use to direct the OCT scanner to the desired measurement location. **D** – We display the measured displacement phase with respect to ear canal (EC) at the OHC and BM in the first A-Scan, as well as the BM in the second A-Scan. The OHC from run 1 and BM from run 2 are in the same anatomical cross-section. Data taken at 80 dB SPL. This figure is adapted from Fig 6 of *Frost et al., 2022* [18].

of this position, it differs substantially from BM in both the longitudinal and transverse directions. In particular, it lies about 45 μm apical of measured BM.

To measure a point on the BM in this same cross-section, we want to maintain this same value of l. We know that positions with $t = 0$ lie on the BM, thanks to our choice of origin, so we can simply set $t = 0$ without changing l to find a BM point in the same longitudinal cross-section as the measured OHC-DC. This is shown in Fig 3.8 **C**. The OCT coordinates for this position can now be used to record a second M-Scan which passes through this point on the BM. We now have three points of interest to observe – BM and OHC-DC in the same optical cross-section (panels **A** and **B**) and BM in the same anatomical cross-section as OHC-DC (panel **C**).

Fig 3.8 **D** shows the phase responses to 80 dB Zwuis stimuli at these three positions. I have decided to show phase but not magnitude as a 45 μm apical difference does not show a shift in the magnitude peak within our frequency resolution [25]. The phase is more sensitive to longitudinal shift due to its rapid phase accumulation near the BF (see Sec 2.2.1), so differences in BF can be observed more clearly through the phase response [49].

The red curve shows OHC-DC phase from the first measurement position (panel **B**), the dark blue curve shows BM phase from the first measurement position (**A**) and the light blue curve shows BM phase from the second measurement position (**C**). Comparing within one M-Scan amounts to comparing the red and dark blue curves, while comparing within one anatomical cross-section amounts to comparing the red and light blue curves.

Observing data within a single M-Scan, as they are often presented, one would see that OHC-DC leads BM at low frequencies, and then is in-phase with BM at frequencies near and above the BF (here about 22 kHz). On the other hand, observing measurements within a single anatomical cross-section, one would see that OHC-DC actually leads BM across frequency. This lead is approximately a quarter of a cycle near and above the BF.

Considering these data in terms of the cycle-by-cycle theory of cochlear amplification accents how significant this difference is. If OHC-DC are providing a force to BM, relative phasing between the two could make the difference between the OHC-DC providing or removing power from

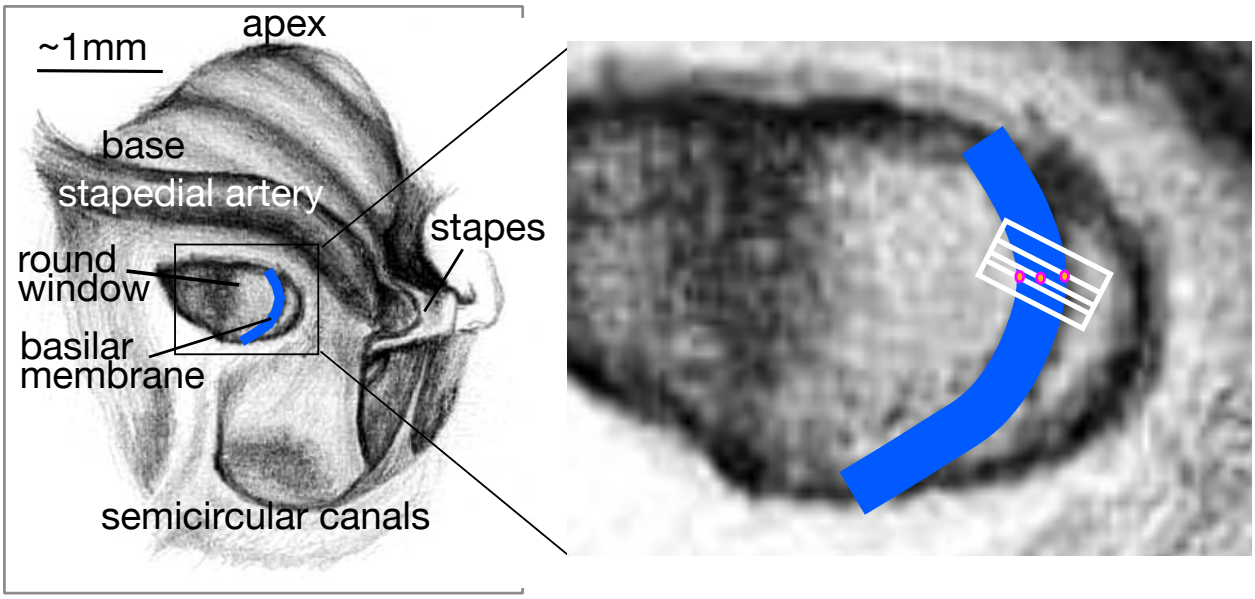

Figure 3.9: Illustration of the method used to test the operation of the orienting GUI against known physiology. The blue band in each view indicates the BM. The magenta dots in the expanded view on the right indicate points that span the BM radially. The white box is the xy-plane of a volume scan, with the interior white lines indicating the y-axis of the B-Scans in which the magenta points lie. The orienting GUI identifies the locations of the magenta points, all of which lie in different B-Scans. This figure is adapted from Fig 7 of *Frost et al., 2022* [18].

the BM [50, 43]. Looking only within an optical cross-section, the two structures appear to move in phase, which suggests no transfer of power. On the other hand, OHC-DC appear to lead BM by a quarter of a cycle within an anatomical cross-section, which suggests power transfer.

3.6 Validation

3.6.1 Radial Variation of Basilar Membrane Displacement

To validate the program's functionality, we replicated the results of an experiment from Cooper [57]. He found that the phase response of BM does not vary across its radial width. With our program, we can measure various points on the BM within the same longitudinal cross-section differing only in the radial direction. If our program did not function as intended, there would be some longitudinal shift between measured BM points which would cause a perceptible phase shift (see Figs 2.3 **D**, 3.8 **D**).

Fig 3.9 shows a cartoon of this experimental procedure. We take a volume scan through the RWM containing the BM, represented by the blue band. The B-Scans which comprise this volume are not necessarily parallel to the radial axis of the cochlea, as shown by the white line segments. We used the orienting GUI to select points on the BM within the volume that differ only radially in anatomical coordinates, even though they may differ in all three optical directions. We measured displacement responses to 80 dB SPL Zwuis stimuli.

The results of this experiment at two different head angles are shown in Fig 3.10. In the first cross-section (panels **A**, **B** and **C**), seven positions across a 60 μm radial span of the BM are measured. The data show no visible difference in the phase response over this region. Similarly, in the second cross-section (panels **D**, **E** and **F**) we measure at 4 points along an 80 μm span. Again, there is no visible difference in the phase response. These results validate our method, as they match precisely what was found by Cooper [57].

3.6.2 Sensitivity to Point Selection

It is important that small variations in point selection do not significantly change the output of the process described above, so that two different users observing the same volume arrive at the same planar approximation. To evaluate the error that could result from point selection, I have considered a single volume scan from Ge903 shown in Fig 3.11.

I chose two B-Scans spaced apart by 20 μm and considered the difference between planar approximations that may be achieved by perturbing one of the point selections, leaving the other two points fixed. In panel **A**, I considered all possible point selections within a 40 μm $\times$ 40 μm box centered at a reasonable selection for point **A**. Similarly, in panel **B**, I considered the same for selections of point **B**.

The OCT system's resolution is $\sim$8 μm in the lateral direction and $\sim$4 μm in the axial direction, so this box is far larger than the system's resolution. Points selected at far edges of this box can be seen qualitatively to be very poor guesses, and are only included to test how robust the method is to extreme error.

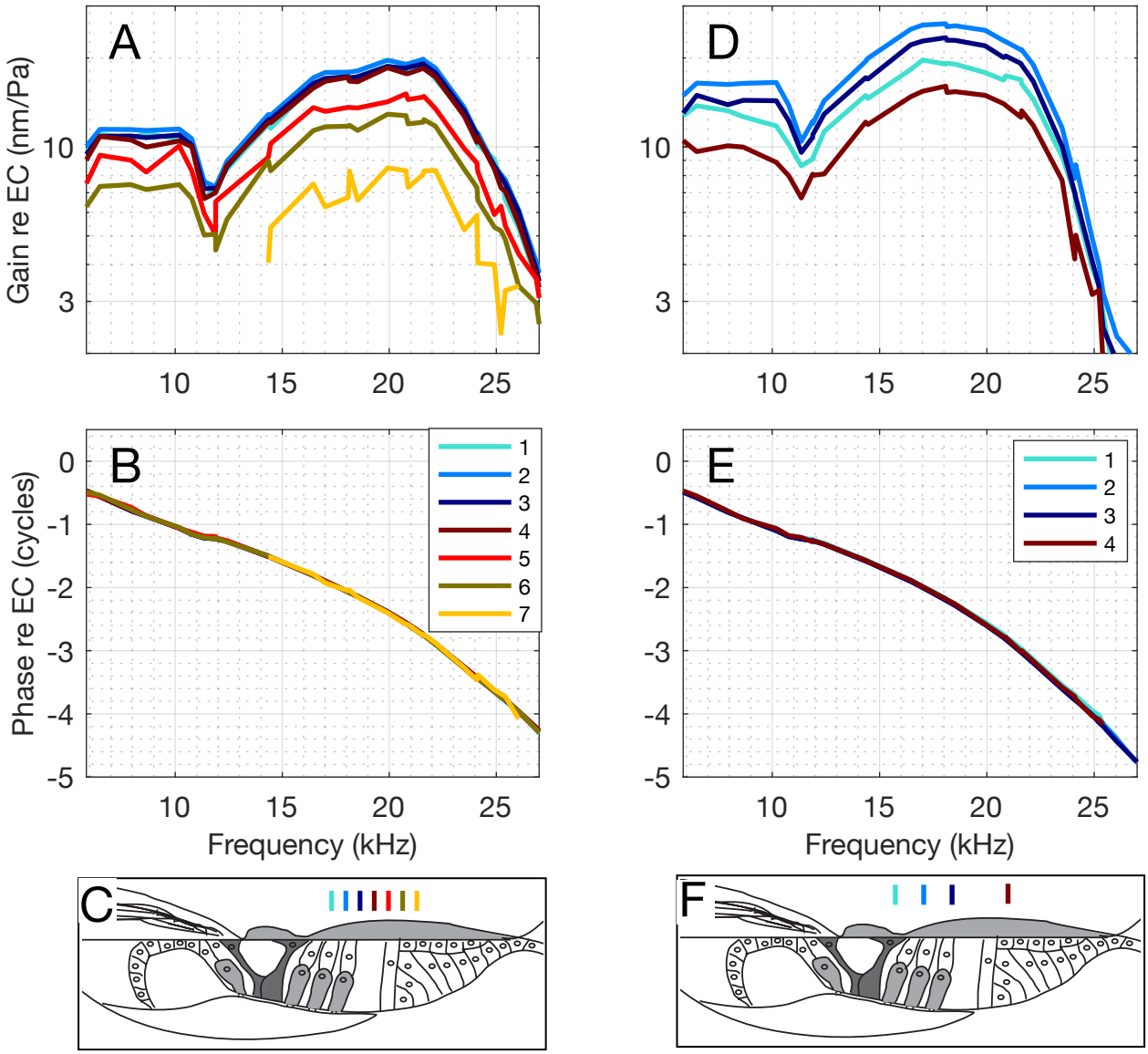

Figure 3.10: BM displacement data from Ge900 at two different head angles provide evidence that our method correctly identified the anatomical radial axis. **A** and **B** – BM displacement amplitude and phase taken along a single anatomical cross-section at seven radial locations spaced 10 μm apart medial (aqua) to lateral (yellow), with locations approximated in **C**. **D** and **E** – BM displacement amplitude and phase taken in a different anatomical cross-section at radial locations spaced 20 or 40 μm apart medial (aqua) to lateral (brown), with locations approximated in **F**. Data taken at 80 dB SPL. Figure adapted from Fig 8 of *Frost et al., 2022* [18].

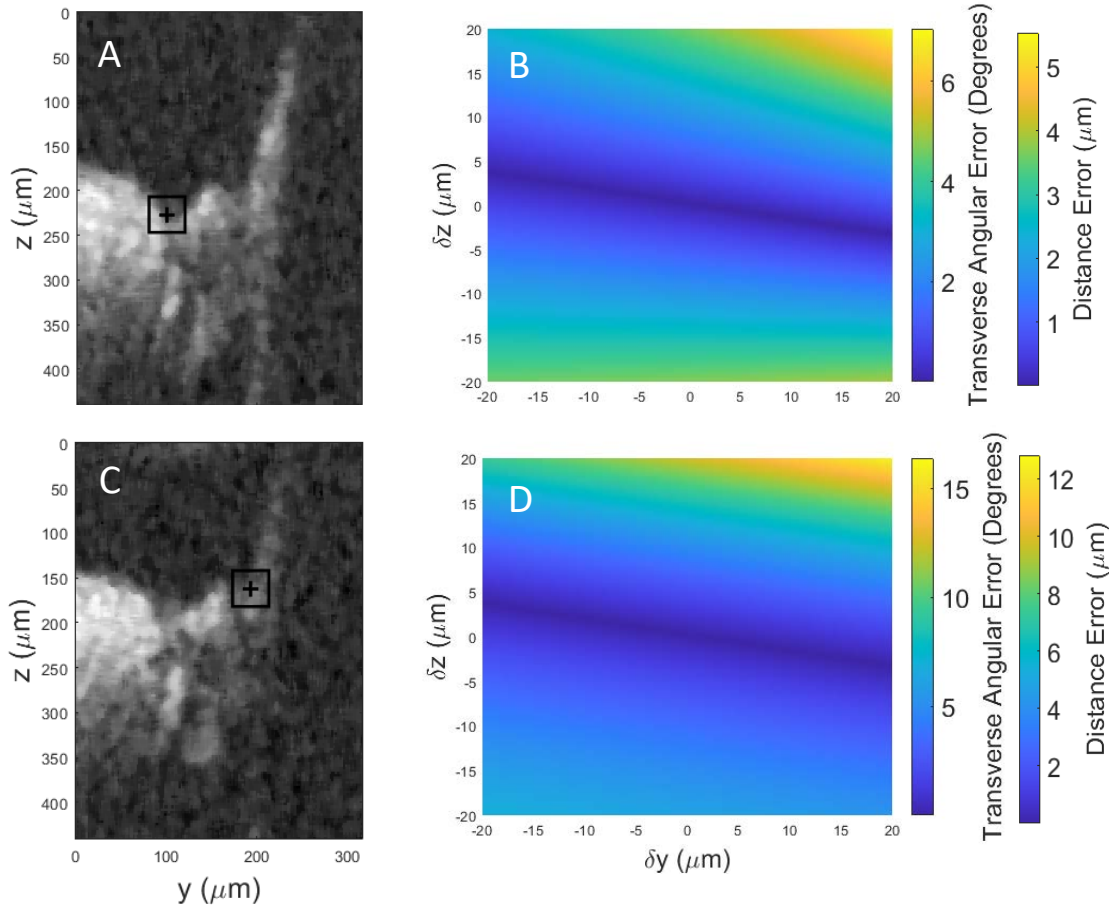

Figure 3.11: Angular and distance error induced by perturbation in selected points, as defined in Eqns 3.11 and Eqn 3.12. **A** – B-Scan with a 40 μm $\times$ 40 μm box centered around a selected point **A**, representing a range of possible selected points for which error will be computed. **B** – error induced by selecting points within this box, holding points **B** and **C** fixed. **C, D** – Same as **A** and **B**, except varying point **B** while holding **A** and **C** fixed. Fig 4 from *Frost et al., 2022* [18].

Panels **B** and **D** show the error (relative to selecting the center point) induced by selecting a different point within the box. Error is evaluated based on two metrics. The first is absolute transverse angular error $\Delta\theta$, which gives the unsigned angular difference between the normal vectors of the plane with that which would have been derived from selecting the center point. If $\hat{\mathbf{t}}_0$ is the transverse vector (in optical coordinates) computed using the center point, and $\hat{\mathbf{t}}_1$ is that computed using a perturbed point, this angular error is computed by the dot product rule as

$$\Delta\theta = |\arccos(\hat{\mathbf{t}}_0 \cdot \hat{\mathbf{t}}_1)|. \tag{3.11}$$

The second metric is referred to as distance error, and is motivated by the application of computing skew within the cochlea. We start by considering a hypothetical point at OHC-DC, **o**, which we know from anatomy lies approximately 45 μm from the BM in the transverse direction. An important use of the planar approximation would be to find a point on the BM, **b**, which differs from **o** only by a transverse translation (i.e. it is at the exact same tonotopic position *and* radial position). This is the application studied in Sec 3.5.2.

For the two computed transverse vectors, these points would be computed as $\mathbf{b}_0 = \mathbf{b} - 45\,\hat{\mathbf{t}}_0$, $\mathbf{b}_1 = \mathbf{b} - 45\,\hat{\mathbf{t}}_1$. The distance error, Δb, is the distance between these two computed points. By the law of cosines, this is

$$\Delta b = 45\sqrt{2 - 2\cos\Delta\theta}. \tag{3.12}$$

The error heat maps in Fig 3.11 **B** and **D** show that a selection that differs by 20 μm in both directions could yield a distance error of up to 12 μm. This is an extreme example, however, as this amount of point selection error is not likely. Within a 20 μm $\times$ 20 μm box, still larger than the OCT system's resolution, distance error does not exceed 5 μm (itself smaller than the OCT system's lateral resolution). This indicates that the method is reasonably robust to human error in point selection.

3.7 Discussion and Conclusions

The method presented in this chapter is valuable for the analysis and presentation of OCT-measured cochlear displacement responses in a physiologically interpretable manner. We showed that at common measurement angles, optical cross-sections and anatomical cross-sections can vary significantly despite their misleading visual similarity.

This skew between the two coordinate systems leads to large qualitative and quantitative differences in the phase responses of structures measured along a single M-Scan due to tonotopic shift. As relative phase is key to understanding power transfer that is theorized to occur between OHC and BM, this is an important consideration [50, 42, 43]. This is especially true when considering models of the OCC which either hypothesize or predict such relative phase responses.

Also important is the nature of the projection of motion onto the optic axis. Motion components in all three anatomical directions are likely to be of physiological significance, and their relationships with one another are not yet understood (further discussion of this issue is present in the following chapter). When we measure a displacement response within the OCC, we are measuring a superposition of these components which may interfere with one another in complex ways. While it does not entirely resolve this problem, presenting the form of this projection as determined by the planar approximation can offer important context.

For example, in the experiments presented above, the longitudinal component was weighted significantly more than the other two components of motion. If the three components of motion are similar in magnitude, we could say that we are presenting measurements that resemble longitudinal displacements. However, this would not be true if the other two components dwarf the longitudinal component in displacement magnitude.

This projection map is also useful in comparing data sets between experiments. Differences between experiments may be due to frequency location, health of the cochlea or species difference, but they may also well be due to the components of motion that most contribute to the measurement. Recent experiments from the Fowler lab have focused on achieving measurements from the

50 kHz region of the gerbil OCC. These results have shown characteristically different phase responses from those seen at the 25 kHz region (Fig 2.6). Applications of the planar approximation program to volumes from these experiments show that the measurements near the 50 kHz region are near-purely transverse, as opposed to the 25 kHz region where our measurements are largely longitudinal. This could be the reason for the phasing difference.

Since its publication, the program has seen consistent use in our lab. In particular, it has been used in a recent paper, *Frames and hotspots in cochlear mechanics (2023)*, which highlights how regions of the cochlea are impacted by the sub-BF activity exhibited by the OHCs. This work discusses large troughs in OHC-DC magnitude responses and corresponding phase jumps near the BF, in light of the angle at which the measurement was made [47]. Another application of the program has been to use knowledge of the projection map at various angles to reconstruct 2-D motion in the OCC. This is the subject of the following chapter.

Chapter 4: Reconstructing Anatomical Components of Motion in the Organ of Corti Complex

4.1 Introduction

4.1.1 Challenges of Measuring Individual Motion Components

In Chapter 3, I discussed limitations of OCT vibrometry introduced by the uniaxial nature of M-Scans. In particular, I discussed how OCT-measured displacements are one-dimensional projections of a three-dimensional motion. This implies that the measured signals may not resemble any of the three physiologically relevant directions of motion. In Sec 3.3, I described the physiological importance that each anatomical component of motion is theorized to have, motivating the need for studies that measure individual anatomical components of motion.

One solution to this problem would be to target viewing angles that can achieve measurements at precisely longitudinal, radial or transverse angles. For example, studies in the hook region of the gerbil cochlea have achieved near-entirely transverse measurement angles (i.e. $\hat{\mathbf{t}} = \hat{\mathbf{z}}$) through the RWM [48, 47]. This can be validated by using the program described in Chapter 3.

However, the anatomy of the gerbil cochlea makes it challenging (if not impossible) to measure at purely radial or longitudinal angles through the RWM. While this may be possible in other animal preparations, for example in mouse where measurements can be made through the cochlea's bony wall, this is not generally true across species and frequency locations.

In this chapter, I will discuss a method that I have developed for reconstructing two- and three-dimensional displacement responses using multiple uniaxial OCT measurements taken at different orientations relative to the sample anatomy. This can be done without directly targeting measurement orientations with single anatomical components. To do so requires two steps: 1) registration

of measured structures between measurement angles, and 2) mathematical reconstruction of the two- or three-dimensional displacement from the registered measurement locations.

4.1.2 Challenges of Registration in Cochlear Mechanics

OCT is inherently label-free, as scans are one-channel maps of reflectivity. This contrasts with fluorescence methods, such as confocal microscopy, in which specific tissues can be localized by staining with fluorescent dyes. Without clear delineation between tissues in an OCT scan, volume registration is a challenging task.

Success has been seen in registering translated OCT volumes using classical methods such as the Scale-Invariant Feature Transform (SIFT) algorithm [58]. However, these methods do not function well in the *in vivo* cochlear setting. Methods using SIFT have relied on applying the algorithm to *en face* images or 2-D slices (B-Scans) within a volume. This is useful in translational registration, but precludes rotational differences between volumes.

Other volume-based registration methods exist that do account for three-dimensional rotations, such as the Kabsch algorithm for point clouds [59, 60, 61, 62], or deep learning methods [63, 64]. These methods perform well on synthetically rotated OCT volumes, but they perform poorly in registering OCT volumes in the cochlear setting[1]. This is due to *reflection*, *occlusion* and *refraction*.

Reflection is an unavoidable property of OCT imaging, wherein objects too close to the objective lens of the system appear as reflected copies in the image or volume. This issue often arises with the RWM through which we measure displacements in the gerbil base. In some cases, the RWM appears “right-side-up" above the OCC in a B-Scan. However, if we wish to lower the lens so that the OCC is nearer to the objective, the RWM will exit the view and a reflection of the RWM will appear “upside-down" in the B-Scan. As the lens approaches the sample further, the reflected

[1] I have tested two such point cloud methods – The iterative Kabsch algorithm, and a convolutional neural network followed by a stage of iterative Kabsch (feature-accelerated iterative Kabsch, or FAIK). These methods performed relatively well on synthetic data, but have not seen success in *in vivo* cochlear mechancis experiments. As such, they do not appear in any peer-reviewed publications, but my discussions of their performance can be found at the following two locations: `https://brian-frost-laplante.github.io/blog/Optimal_rotation/`, `https://github.com/Brian-Frost-LaPlante/reports/blob/main/VOCT.pdf`.

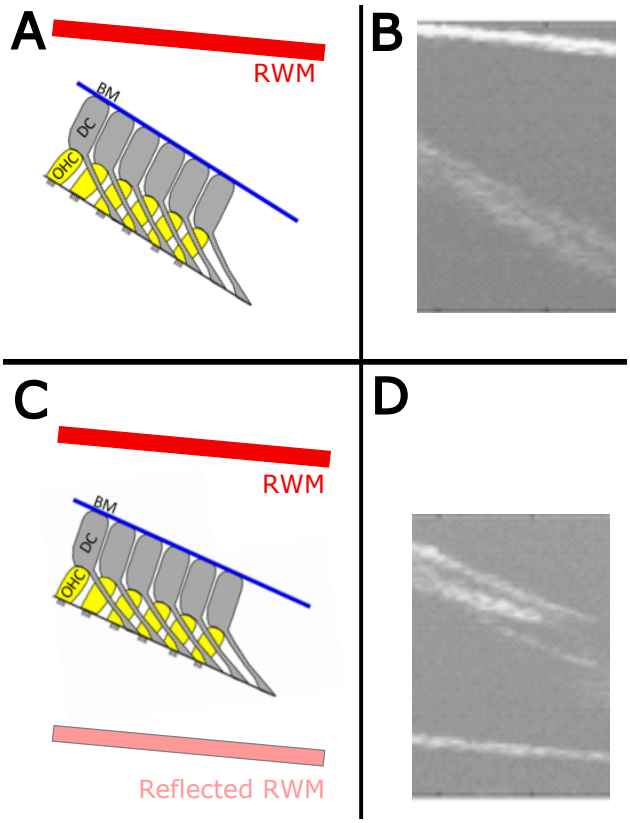

Figure 4.1: Sample B-Scans from Ge995 displaying the reflection property of OCT images. **A-B** – Anatomical cartoon and corresponding B-Scan showing the RWM “right-side-up", appearing nearer to the scanner than the OCC. **C-D** – Similar to **A-B**, except that the scanning head has been zoomed towards the sample. The RWM now appears reflected as if it were further from the scanner than the OCC.

RWM moves further down in the B-Scan, sometimes even below the OCC. This phenomenon is shown in Fig 4.1. While this can be avoided by ensuring that the objective is sufficiently far from the sample, this could move the OCC out of the focus of the system and give sub-optimal image brightness and displacement signal-to-noise ratio (SNR). Reflection is thereby unpredictable across viewing angles, and will throw off volume-based registration methods.

Occlusion is a feature particular to cochlear displacement measurements taken through the RWM. The bone surrounding the RWM is opaque, so the OCT beam can only penetrate far enough to view the OCC in the region visible through the RWM. This means that the round window is quite literally a *window* that controls the field of view of the scanner. As the orientation of the head relative to the scanner changes, so too does the region visible through the window. That is, the bone

occludes different portions of the OCC at different angles. This is similarly true for measurements made through cochleostomies, where the hole drilled in the cochlea acts as the window.

Refraction occurs because the RWM acts as an air-fluid interface. This changes relative angles between structures within the fluid (e.g. the OCC) and structures outside of the fluid (e.g. the bone surrounding the RWM). This is also true of measurements made through cochleostomies, as an air-fluid interface is present along the beam path.

One solution that accounts for these problems is to use multiple OCT systems very precisely oriented so that the relationship between the images taken by the three systems is known [65]. This method is expensive and requires a large amount of physical space. It also requires extreme precision in accounting for differing angles of refraction at the air-fluid interfaces between systems, and the fact that all three systems see different portions of the OCC due to the windowing phenomenon.

The method presented in this chapter avoids these significant roadblocks by using cochlear physiology to register points, as opposed to using complex hardware or volume processing methods. In short, the method uses the tonotopic phase response of the BM, which is expected to be identical across all orientations, to register positions between measurements taken from multiple measurement angles. This method does not require any *a priori* knowledge about structure orientation, only requires a single OCT system, and is sufficiently general to be used in various experimental preparations. The *in vivo* gerbil experiments were conducted as described in App A.1. The method presented in this chapter appeared in the *Journal of the Acoustical Society of America* in 2023 [49].

4.2 Theory of Reconstruction

I will begin with a discussion of the mathematics involved in reconstructing motion given measurements taken at multiple viewing angles. This assumes that measured structures have already been registered between viewing angles, our method for which is described in Sec 4.3.

4.2.1 The General Case

The steady-state motion of the cochlea in the frequency domain is a 3-D complex vector field as a function of 3-D space and frequency. We consider the motion at a single 3-D position $\mathbf{p}$, frequency f and SPL S as

$$\mathbf{d}(\mathbf{p}, f, S) = d_l(\mathbf{p}, f, S)\hat{\mathbf{l}} + d_r(\mathbf{p}, f, S)\hat{\mathbf{r}} + d_t(\mathbf{p}, f, S)\hat{\mathbf{t}} \in \mathbb{C}^3,$$

where $\hat{\mathbf{l}}$, $\hat{\mathbf{r}}$ and $\hat{\mathbf{t}}$ are the longitudinal, radial and transverse unit vectors and d_l, d_r and d_t are the components of motion in each of these directions. For notational purposes, I will keep position-, frequency- and SPL-dependence implicit for the remainder of this section.

Let $\hat{\mathbf{z}}_m$ with $m = 1, 2, \ldots, M$ be the optical axes at M distinct measurement orientations. They can be written in anatomical coordinates as

$$\hat{\mathbf{z}}_m = l_m\hat{\mathbf{l}} + r_m\hat{\mathbf{r}} + t_m\hat{\mathbf{t}}, \quad m = 1, 2, \ldots, M.$$

The displacement δ_m measured at orientation m is a complex number defined by the projection of $\mathbf{d}$ onto $\hat{\mathbf{z}}_m$:

$$\delta_m = \begin{pmatrix} l_m & r_m & t_m \end{pmatrix} \begin{pmatrix} d_l \\ d_r \\ d_t \end{pmatrix} \in \mathbb{C}, \quad m = 1, 2, \ldots, M. \tag{4.1}$$

Eqn 4.1 can be re-written as a system of M equations in matrix form as

$$\begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{pmatrix} = \begin{pmatrix} l_1 & r_1 & t_1 \\ l_2 & r_2 & t_2 \\ \vdots & \vdots & \vdots \\ l_M & r_M & t_M \end{pmatrix} \begin{pmatrix} d_l \\ d_r \\ d_t \end{pmatrix} \in \mathbb{C}^M. \tag{4.2}$$

In practice, we have the uniaxial displacement measurements δ_m and we can find the components l_m, r_m and t_m using the program described in Chapter 3. The unknown object to reconstruct is the 3-D vector $\mathbf{d}$. Assuming the M orientations are not colinear, this problem is *underdetermined*

if $M < 3$, *determined* if $M = 3$ and *overdetermined* if $M > 3$.

This means the system may have no unique solution or no solution at all. The “best fit" (least-squares) solution for $\mathbf{d}$ given our measurements is given by

$$\mathbf{d} = \begin{pmatrix} l_1 & r_1 & t_1 \\ l_2 & r_2 & t_2 \\ \vdots & \vdots & \vdots \\ l_M & r_M & t_M \end{pmatrix}^{\dagger} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{pmatrix}, \tag{4.3}$$

where $\cdot^{\dagger}$ denotes the Moore-Penrose pseudoinverse[2].

4.2.2 The Two-Dimensional Case

Eqn 4.3 is the most general form of the equation by which anatomical motion components can be reconstructed from multiple uniaxial measurements. Intuitively, the most accurate reconstructions could be obtained by measuring from many distinct orientations ($M > 3$), but this is practically challenging.

A more modest and tractable problem would involve the case of reconstructing two components of motion from only two measurements ($M = 2$), first approached in the radial-transverse case by Lee *et al.* [66]. For the remainder of this chapter, I will discuss only the longitudinal and transverse components of motion, ignoring radial motion entirely. This can be justified by ensuring that our measurement axes contain no radial component ($r_1 = r_2 = 0$).

In this case the motion is simply

$$\mathbf{d} = d_l\hat{\mathbf{l}} + d_t\hat{\mathbf{t}} \in \mathbb{C}^2, \tag{4.4}$$

and the measurement axes are

$$\hat{\mathbf{z}}_i = l_i\hat{\mathbf{l}} + t_i\hat{\mathbf{t}}, \quad i = 1, 2. \tag{4.5}$$

Finally, the system of equations can be simplified to

[2]The Moore-Pensrose pseudo-inverse is a generalization of the matrix inverse to all matrices. In a system $Ax = b$ (may be unsolvable), it is the matrix that minimizes $||A^{\dagger}b - b||_2$.

$$\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix} = \begin{pmatrix}l_1 & t_1\\ l_2 & t_2\end{pmatrix}\begin{pmatrix}d_l\\ d_t\end{pmatrix} \in \mathbb{C}^2. \tag{4.6}$$

This system is invertible if the 2×2 matrix is full rank. The rows of this matrix are the measurement axes, so it is full rank if and only if the measurement axes are linearly independent. As we are considering a set of only two vectors, the rows are linearly independent so long as the measurement directions are not parallel or antiparallel – an easily met criterion that amounts to precisely the statement that we cannot reconstruct 2-D motion with measurements along only one axis.

The measurement axes are unit vectors, so this matrix is orthogonal. That means its inverse is simply its transpose, and we can write the 2-D reconstruction formula as

$$\begin{pmatrix}d_l\\ d_t\end{pmatrix} = \begin{pmatrix}l_1 & l_2\\ t_1 & t_2\end{pmatrix}\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix}. \tag{4.7}$$

4.2.3 Error in the Two-Dimensional Case

While we can ensure the matrix is invertible by taking measurements from any two distinct angles, it is intuitively clear that measurements taken only, say, 1° apart from one another cannot be used to accurately reconstruct the displacement in a noisy environment. Instead, one would imagine that an accurate reconstruction would require a large angular difference between measurement orientations.

As of yet, I have not considered the impact of noise on the measured signal. Modeling noise as an additive process $\mathcal{N} \in \mathbb{C}^2$, I can write the noisy reconstruction formula as

$$\begin{pmatrix}d_l\\ d_t\end{pmatrix} = \begin{pmatrix}l_1 & l_2\\ t_1 & t_2\end{pmatrix}\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix} + \begin{pmatrix}l_1 & l_2\\ t_1 & t_2\end{pmatrix}\mathcal{N}. \tag{4.8}$$

The SNR of each measurement is determined by $\mathcal{N}$, but this noise is modified by the reconstruction matrix. In particular, the linear noise power is approximately multiplied by the *condition number* of the matrix, κ, defined by

$$\kappa = \frac{|\sigma_{max}|}{|\sigma_{min}|}, \tag{4.9}$$

where σ_{max} and σ_{min} are the maximum- and minimum-magnitude singular values of the reconstruction matrix. We can use the formulae for the singular values of a 2×2 matrix to derive a relationship between noise amplification and the angular difference between measurement axes.

Using the element names of the relevant matrix, the singular values are

$$\sigma_{max} = \sqrt{\frac{s_1 + s_2}{2}}, \tag{4.10}$$

$$\sigma_{min} = \sqrt{\frac{s_1 - s_2}{2}}, \tag{4.11}$$

where

$$s_1 = l_1^2 + t_1^2 + l_2^2 + t_2^2, \tag{4.12}$$

$$s_2 = \sqrt{(l_1^2 + t_1^2 - l_2^2 - t_2^2) + 4(l_1 l_2 + t_1 t_2)^2}. \tag{4.13}$$

Because the measurement axes are unit vectors, we can write $l_i^2 + t_i^2 = 1$ for $i = 1, 2$. Now we have

$$s_1 = 2, \quad s_2 = 2(l_1 l_2 + t_1 t_2).$$

Defining $\nu = l_1 l_2 + t_1 t_2$, I can write the condition number as

$$\kappa = \sqrt{\frac{1 + \nu}{1 - \nu}}, \tag{4.14}$$

a monotonically increasing function of ν. With a bit of manipulation, this formula can be written in terms of the angular difference between measurement axes. To do so, recognize $\nu = \hat{\mathbf{z}}_1 \cdot \hat{\mathbf{z}}_2$, or

$$\nu = ||\hat{\mathbf{z}}_1||\,||\hat{\mathbf{z}}_2|| \cos\theta = \cos\theta, \tag{4.15}$$

where θ is the angular difference between the two vectors (I have used the fact that the measurement

axes are unit vectors). A trigonometric identity gives a concise formula:

$$\begin{aligned}\kappa &= \sqrt{\frac{1+\cos\theta}{1-\cos\theta}} \\ &= \sqrt{\frac{1}{\tan^2\frac{\theta}{2}}} \\ &= \frac{1}{\left|\tan\frac{\theta}{2}\right|},\ -90^\circ \le \theta \le 90^\circ.\end{aligned} \tag{4.16}$$

This quantifies the intuition regarding the relationship between noise power and angular difference. The condition number approaches infinity as the angular difference approaches 0, and approaches a minimum at 1 as the angular difference approaches 90°. While 90° would be optimal, the maximum achievable angular difference is restricted by the experimental preparation. In particular, in measurements taken through the gerbil round window where refraction and occlusion are present, 15° is often the largest achievable angular difference.

The noise floor of our uniaxial displacement measurements is approximately 0.1 nm. For a vector of two independent measurements δ_1 and δ_2, the noise floor would thereby be about $\sqrt{0.1^2+0.1^2} \approx 0.14$ nm. The noise increase induced by reconstruction using measurements taken 15° apart (a κ value of 7.6 according to Eqn 4.16) would yield a noise floor of ~ 1 nm.

4.3 Tonotopic Registration

With Eqn 4.7, I can reconstruct 2-D motion in the OCC so long as I can 1) determine the measurement axes in anatomical coordinates (l_i and t_i), and 2) register a point of interest between two orientations so that the *same structure* is measured at both angles. In this work, our structure of choice was OHC-DC.

The method presented here performs both of these tasks without any *a priori* knowledge of precise structure locations or beam angles, using only a single OCT system. The process below is followed at both measurement angles $i = 1, 2$. Each step described below is graphically represented in a panel in Figs 4.2 and 4.3.

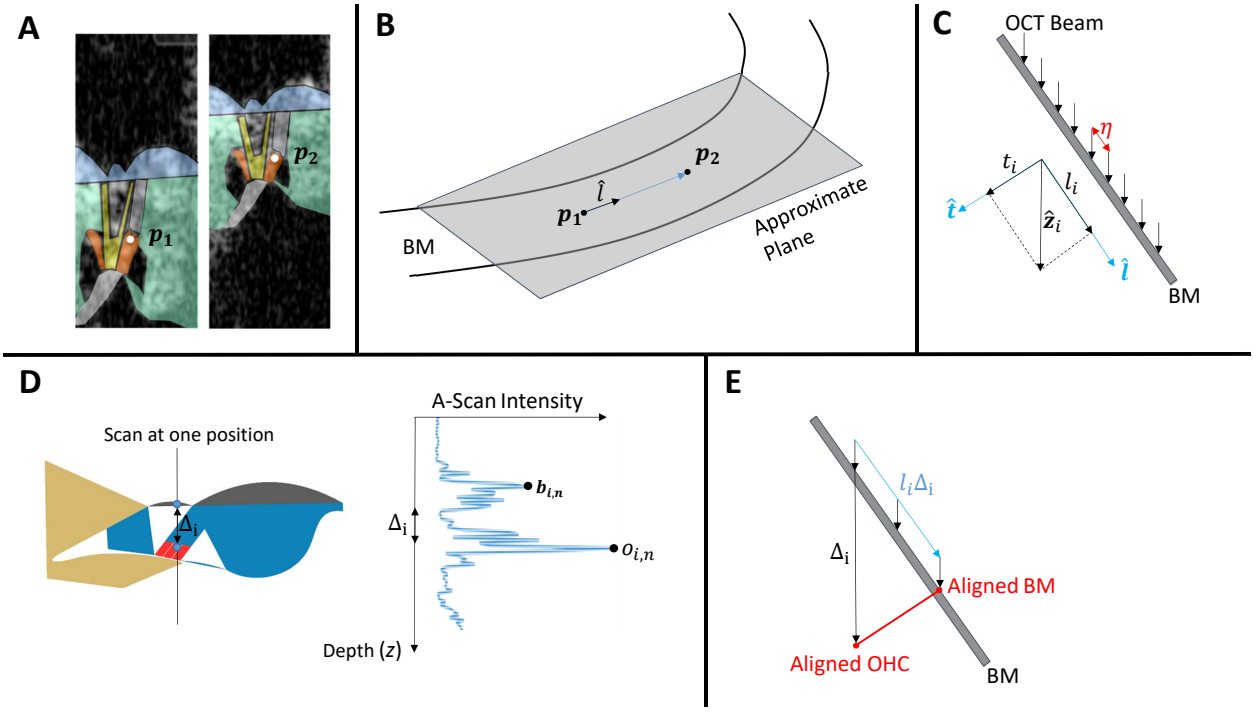

Figure 4.2: Graphical representation of the steps followed in the experimental method employed for two-dimensional reconstruction. **A** – Labeled B-Scans from a single volume, with two points $\mathbf{p}_1$ and $\mathbf{p}_2$ at the same anatomical structure (OHC-DC) marked in two different cross-sections. **B** – The BM approximated as a plane, in which the longitudinal direction connects any two points at the same anatomical structure between cross-sections. **C** – Cartoon of the BM with many measurements taken η apart longitudinally; anatomical axes with the measurement axis represented in longitudinal and transverse components. **D** – Cartoon of OCC with points labeled at BM and OHC-DC, along with an A-Scan with these same points labeled; Δ_i is the axial distance between OHC-DC and BM. **E** – Cartoon of the BM with OHC-DC and BM in the same anatomical cross-section but different A-Scans aligned by the known longitudinal component of the measurement axis. This figure is adapted from Fig 3 of *Frost et al., 2023* [49, 67].

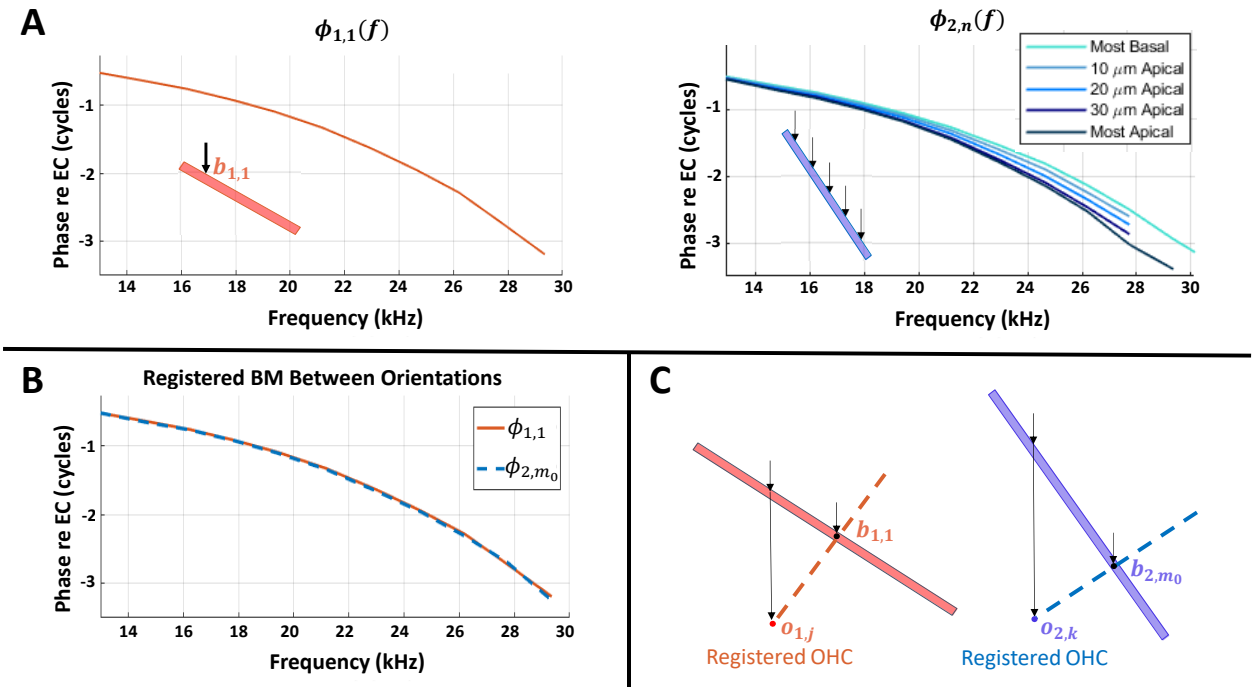

Figure 4.3: Graphical representation of the method by which cross-sections and OHC-DC points are registered between orientations post-experiment. The pink and blue rectangles represent the approximately linear BM segments observed at the two orientations. **A** – On the left, the first measured BM point's phase response at 80 dB SPL, $\phi_{1,1}(f)$; on the right, a set of BM responses at 80 dB SPL taken 10 μm apart longitudinally at the second orientation, $\phi_{2,n}(f)$ for $n = 1, 2, 3, 4, 5$. **B** – $\phi_{1,1}(f)$ and the nearest phase response at orientation 2, ϕ_{2,m_0}. The two phase responses are nearly identical, showing that the BM points measured at $b_{1,1}$ and b_{2,m_0} are in the same anatomical cross-section. **C** – Cartoons of the BM at both orientations with registered cross-sections shown by dotted lines. OHCs aligned to these BM points, $\mathbf{o}_{1,j}$ and $\mathbf{o}_{2,k}$, determined as in Fig 4.2 **E**, are thereby also in the registered anatomical cross-section. All four of these measured points lie in the same anatomical cross-section, and the OHCs are thereby registered to one another as marked. This figure is adapted from Fig 4 of *Frost et al., 2023* [49].

Determining $\hat{\mathbf{l}}$

Before taking any measurements, we find the longitudinal unit vector $\hat{\mathbf{l}}$ in optical coordinates. This is performed using a simplified version of the process described in Chapter 3. Using the planar approximation, the longitudinal axis is approximated as a line. We find this line by picking two points at each orientation, $\mathbf{p}_{i,1}$ and $\mathbf{p}_{i,2}$, at the same radial-transverse position in two different cross-sections of a volume scan. For example, in Fig 4.2 **A** I have selected the two points to be at OHC-DC.

By definition, two distinct points at the same radial-transverse position differ only along the longitudinal axis. Under the planar approximation, I can find the longitudinal vector at orientation $i = 1, 2$ simply as the normalized vector difference

$$\hat{\mathbf{l}}_{i,op} = \frac{\mathbf{p}_{i,2} - \mathbf{p}_{i,1}}{||\mathbf{p}_{i,2} - \mathbf{p}_{i,1}||}. \tag{4.17}$$

Determining l_i and t_i

We need the anatomical l and t components of the optic axis $\hat{\mathbf{z}}_i$ to form the reconstruction matrix of Eqn 4.7. We can find l_i from the unit vector $\hat{\mathbf{l}}_i$ found in Eqn 4.17 simply by observing the z-component of this vector in optical coordinates. This is the dot product

$$l_i = \hat{\mathbf{z}}_{i,op} \cdot \hat{\mathbf{l}}_{i,op} = (0\ 0\ 1)\hat{\mathbf{l}}_{i,op}. \tag{4.18}$$

The transverse component of the measurement vector can then be found by recognizing that $\hat{\mathbf{z}}_{i,anat} = (l_i\ t_i)^T$ is a unit vector. This gives t_i as

$$t_i = \sqrt{1 - l_i^2}. \tag{4.19}$$

This is displayed in the bottom-left of Fig 4.2 **C**.

Taking measurements along the longitudinal axis

At each orientation, we measure displacement at OHC-DC at many longitudinally spaced M-Scans. This is necessary for the registration method, to be described shortly. Let the first of these M-Scans at orientation i contain $\mathbf{p}_{i,1}$, which is within OHC-DC by construction. We measure at N longitudinally spaced M-Scans with spacing η, within the n^{th} scan capturing OHC-DC at location

$$\mathbf{o}_{i,n} = \mathbf{p}_{i,1} + (n-1)\eta\hat{\mathbf{l}}_i,\ n = 1, \ldots, N, \tag{4.20}$$

where again $i = 1, 2$ is the index of the scanner orientation. This process is displayed in Fig 4.2 **C**.

This describes the entire experimental process – find the longitudinal vector and take measurements across this axis at two orientations. After the experiment has ended, we turn to registering OHC-DC points so that the mathematical reconstruction can be applied.

Skew Correction

Every M-Scan capturing OHC-DC will also capture BM as a consequence of the anatomy of the cochlea. At each orientation i and scan n, I will denote measured BM and OHC-DC as $\mathbf{b}_{i,n}$ and $\mathbf{o}_{i,n}$, respectively. In Chapter 3, I discussed the concept of *skew* – the fact that although $\mathbf{b}_{i,n}$ and $\mathbf{o}_{i,n}$ are in the same optical cross-section, they do not necessarily lie in the same anatomical cross-section. To perform our registration method, we need to account for skew. The method of skew correction from Chapter 3 can be used here, although the present case – the 2-D setting with no radial component – is sufficiently simple that the orienting GUI is not necessary.

The $\hat{\mathbf{z}}_i$-direction distance between BM and OHC-DC in each M-Scan is $\Delta_i = (\mathbf{o}_{i,n} - \mathbf{b}_{i,n}) \cdot \hat{\mathbf{z}}_i$. The corresponding longitudinal distance is thereby $l_i\Delta_i$. This is displayed in Fig 4.2 **D** and **E**. Under the planar approximation, this value is independent of the cross-section n (although it does depend on orientation i).

The first OHC-DC position $\mathbf{o}_{i,1}$ is in the same anatomical cross-section as a point on the BM at $\mathbf{b}_{i,1} + l_i\Delta_i\hat{\mathbf{l}}$. In Chapter 3, we would perform skew correction at experiment time by finding optical

coordinates at which to measure the motion at this position. However, this is not necessary in this case – we can perform skew correction *post hoc*.

This is because we have taken many measurements of BM along the longitudinal axis. If the longitudinal spacing η is small, there exists an integer m such that $l_i\Delta_i \approx m\eta$. That is, $\mathbf{o}_{i,1}$ is approximately in the same anatomical cross-section as $\mathbf{b}_{i,m+1}$. We call these points *aligned*. This *post hoc* skew correction process is shown in Fig 4.2 **E**.

As a result of the planar approximation, this first alignment results in many more aligned pairs of OHC-DC and BM. For integer k, $\mathbf{o}_{i,k}$ and $\mathbf{b}_{i,m+k}$ will be aligned so long as $m + k \leq N$.

BM-Based Cross-Section Registration

The ultimate goal is to register OHC-DC positions between orientations. However, as we have performed skew correction, we can equivalently register BM positions between orientations. From this, OHC-DC registration is a natural consequence, as we know which OHC-DC positions are in the same cross-sections as these registered BM positions from the alignment procedure described above. That is, OHC-DC registration can be re-framed as a problem of BM registration.

To register BM points between orientations, we use the phase of the traveling wave. We assume that the BM moves entirely transversely, so that the phase response of BM motion will be the same at every viewing angle. I will denote the phase response at $\mathbf{b}_{i,n}$ as $\phi_{i,n}(f)$, where f is the stimulus frequency. Considering the phase response at orientation 1 and position 1, $\phi_{1,1}(f)$, we find the position m_0 at orientation 2 where the phase response best matches $\phi_{1,1}$ according to minimum ℓ_2 (in the discrete case, standard Euclidean) distance. That is,

$$m_0 = \underset{m}{\operatorname{argmin}} ||\phi_{1,1}(f) - \phi_{2,m}(f)||_2. \tag{4.21}$$

Then $\mathbf{b}_{1,1}$ and $\mathbf{b}_{2,m_0}$ are said to be registered.

Because measured positions are spaced longitudinally by η at both orientations, $\mathbf{b}_{1,2}$ and $\mathbf{b}_{2,m_0+1}$ are also registered, and so on. This yields registered BM points across some longitudinal range.

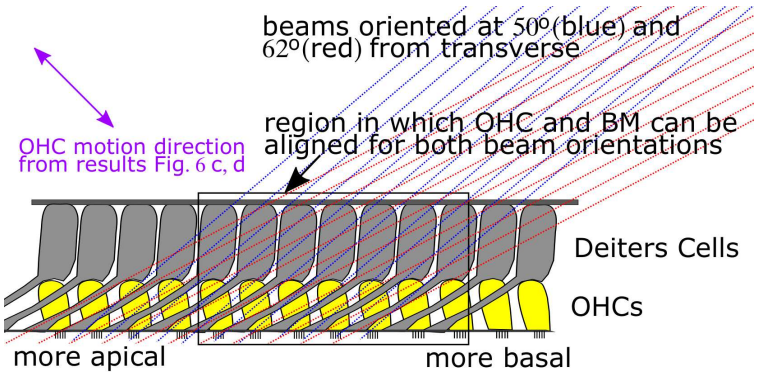

Figure 4.4: Cartoon of a longitudinal-transverse cross-section of the OCC containing the DCs and OHCs. Measurement axes at two orientations that are achievable through the gerbil RW are shown. Measured OHC and BM positions along one measurement axis will lie at different longitudinal locations. Eleven measurements at each angle are taken 10 μm apart longitudinally. The boxed region shows where we can align measured BM positions with measured OHC-DC positions – this corresponds to about six distinct OHC-DC positions being registered. The arrow in the top-left shows the direction of motion in which the OHC-DC region was found to move in the reconstruction of Fig 4.5. This figure is adapted from Fig 5 of *Frost et al., 2023* [49].

Fig 4.3 **A** illustrates the process, showing the phase response at $i = n = 1$, and the phase responses across longitudinal position at orientation $i = 2$. We compare the phase response $\phi_{1,1}(f)$ to all of the phase responses at orientation 2 and find the best match, $\phi_{2,m_0}(f)$. This match is shown in Fig 4.3 **B**, wherein the phase responses are nearly identical. This means that the cross-sections containing these two BM points are identical.

Registering OHC-DC points between orientations

Finally we are in a position to register OHC-DC points, combining the results of the skew correction and BM registration steps. Consider registered BM points $\mathbf{b}_{1,n}$ and $\mathbf{b}_{2,m}$. In performing skew correction, these BM points were aligned to OHC-DC points in the same anatomical cross-section. If $\mathbf{b}_{1,n}$ is aligned to OHC-DC point $\mathbf{o}_{1,j}$ and $\mathbf{b}_{2,m}$ is aligned to $\mathbf{o}_{2,k}$, then *all four* of these measured points are in the same anatomical cross-section. This means that these two OHC-DC points are registered. We can repeat this process for all registered BM points with aligned OHC-DC points. This is graphically represented for a single position in Fig 4.3 **C**, and across the entire measurement range in Fig 4.4.

4.4 Measured Longitudinal-Transverse Reconstructions

Having found the longitudinal and transverse components of the two measurement axes, and registered OHC-DC points, the mathematical reconstruction can be performed. In this section I will present reconstructions of longitudinal and transverse components of motion at the OHC-DC measured using this method. I will begin by walking through a single example of a reconstruction following the method described above to elucidate the process. I will then show two other examples of longitudinal-transverse reconstructions.

4.4.1 Explanatory Example

The following exposition is adapted from *Frost et al., 2023* [49]. Fig 4.5 shows a reconstruction using responses to an 80 dB SPL stimulus. In this example, from the first angle we consider data from the OHC-DC region measured at the fourth measurement location: $\mathbf{o}_{1,4}$. Skew correction is performed to align BM and OHC within the same anatomical cross-section: Using the value of l_1, the longitudinal distance between BM and OHC-DC locations along measurement axis 1 was about 40 μm (Fig 4.2 **C-E**). The longitudinal spacing (η in Fig 4.2 **C**) was 10 μm, thus $\mathbf{o}_{1,4}$ is in the same anatomical cross-section as $\mathbf{b}_{1,8}$.

The registration of BM points between orientations used least squares matching of $\mathbf{b_{1,1}}$ and $\mathbf{b_{2,m}}$ phase responses for $m = 1, \ldots, 11$, as shown in Fig 4.3. These data, as well as those of Fig 4.3 are both from Ge967 where we found $\mathbf{b_{1,1}}$ to be registered to $\mathbf{b_{2,2}}$. This means that $\mathbf{b_{1,8}}$ is also registered to $\mathbf{b_{2,9}}$. Finally, using the value of l_2, the longitudinal distance between BM and OHC-DC locations was about 30 μm, and thus $\mathbf{o}_{2,6}$ is in the same anatomical cross-section as $\mathbf{b}_{2,9}$.

Now the registration is complete: $\mathbf{b_{1,8}}$, $\mathbf{o_{1,4}}$, $\mathbf{b_{2,9}}$ and $\mathbf{o_{2,6}}$ are all in the same anatomical cross-section. Fig 4.5 **A** and **B** show the displacement frequency responses measured at these OHC-DC points at the two angles. These complex data correspond to δ_1, δ_2 in Eqn 4.7. The displacement phases at the two registered BM locations are also shown and are nearly identical, indicating successful registration. Moreover, BM amplitudes should be parallel (offset from each other ver-

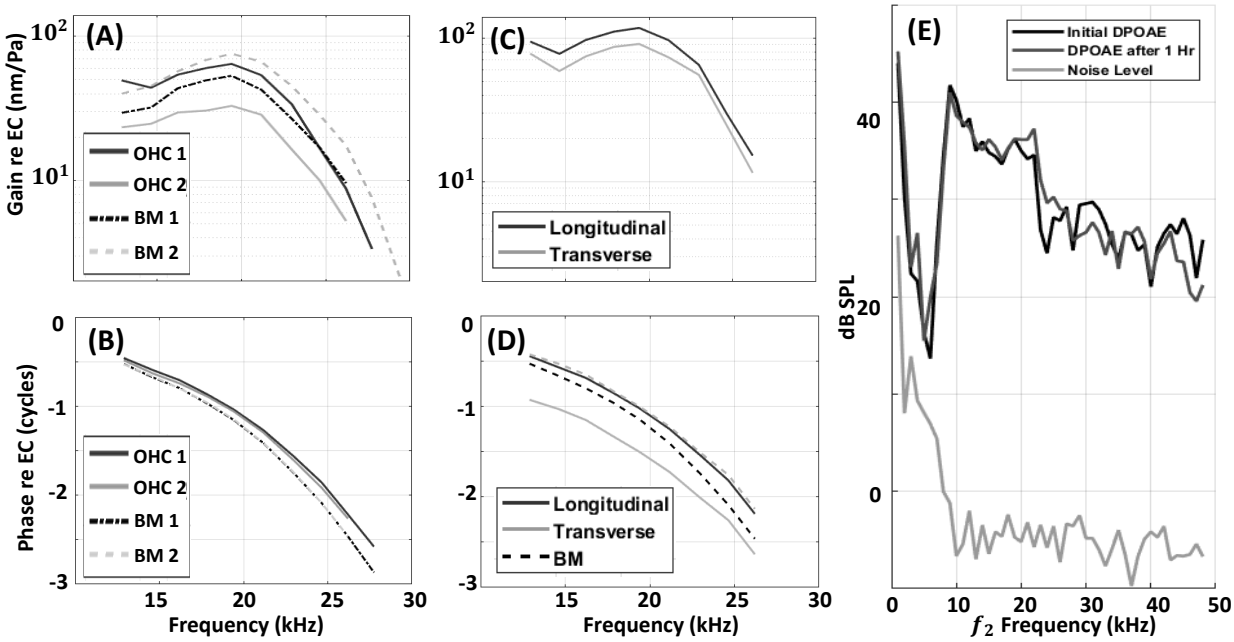

Figure 4.5: An example of responses and corresponding reconstruction from Ge967 at registered OHC-DC positions performed using responses to an 80 dB 15-frequency, 1 sec Zwuis stimulus. **A, B** – Magnitude and phase responses at aligned OHC-DC positions taken at two orientations – viewing angles 1 and 2 make 64° and 50° angles with the BM normal, respectively. BM responses in the registered cross-section measured at both orientations are also shown. Note that the BM phase responses are nearly indistinguishable between orientations, indicating that the cross-section is truly registered. Similarly, the BM magnitudes are parallel (offset vertically), and differ by a geometric factor determined by the ratio of the measurement angles' cosines. **C, D** – Reconstructed longitudinal and transverse magnitude and phase responses at the OHC-DC, generated by application of Eq 4.7 to the data in panels **A** and **B**. For reference, we also show the BM phase response as a dashed black line, and the dashed gray line shows the reconstructed transverse phase shifted vertically by 0.5 cycles. **E** – DPOAE magnitudes in response to 70 dB SPL two-tone stimuli measured 20 minutes prior to the displacement measurements at each orientation. These two DPOAE measurements were taken one hour apart. This figure is adapted from Fig 6 of *Frost et al., 2023* [49].

tically) and differ by a geometric factor: $\cos\theta_2/\cos\theta_1 \approx 1.47$. We can see that the BM responses are quite nearly parallel, and we compute that they differ in size by a factor of, on average, 1.43.

Finally, we perform the reconstruction. The components l_1, t_1, l_2 and t_2 form the reconstruction matrix (Eqn 4.7), which we apply to the measured displacements δ_1 and δ_2. The reconstructed longitudinal and transverse OHC-DC frequency responses are shown in Fig 4.5 **C** and **D**.

4.4.2 Longitudinal and Transverse Reconstructions at 70 and 80 dB SPL

Longitudinal and transverse gain and phase in response to 70 and 80 dB SPL Zwuis stimuli at two positions, alongside BM data from the same cross-section, are shown in Fig 4.6. The reconstructed transverse and longitudinal motion are out of phase by approximately half of a cycle across most of the frequency range. This is made clear with the light gray dashed line in Fig. 4.5 **D**, which shows transverse phase shifted upwards by half of a cycle. This indicates non-elliptical (specifically lineal) motion along an axis that lies between the positive longitudinal and negative transverse directions. This direction of motion is indicated in the purple arrow in Fig. 4.4.

The amplitude of the measured motion was significantly larger at orientation one (the more longitudinal measurement angle) than at orientation two. This δ_1 and δ_2 combination could result from the longitudinal component being much larger than the transverse component. However, the angular difference of 14° is too small to account for the size of the loss in amplitude at the second orientation. The actual explanation for the smaller amplitude at the second orientation is that the second angle of observation is more perpendicular to the direction of overall motion. This reasoning also explains why the reconstructed transverse and longitudinal responses are larger than the amplitude of the motion measured at either orientation. This sort of “cancellation" effect will be strongest when the motion is nearly lineal (not elliptical) because elliptical motion will never be perpendicular to the observation direction. Finally, the observation that the OHC-DC phase responses measured at the two orientations (as in Fig 4.5 **B**) are nearly equal is consistent with nearly lineal motion.

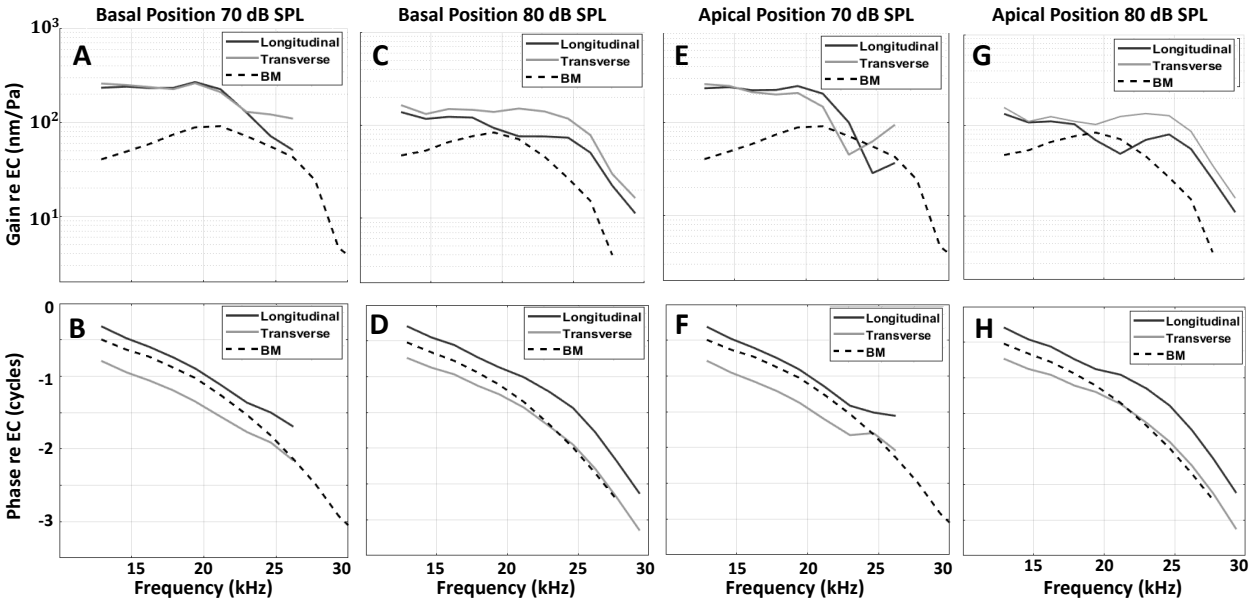

Figure 4.6: Transverse and longitudinal gain and phase responses reconstructed at the OHC-DC in response to Zwuis stimuli at 70 and 80 dB SPL at two distinct longitudinal locations. The measurement positions are spaced apart by 60 μm longitudinally. Light gray curves are reconstructed transverse OHC-DC responses, dark gray curves are reconstructed longitudinal OHC-DC responses and black dashed curves are BM responses in this same cross-section. This figure is adapted from Fig 8 of *Frost et al., 2023* [49].

4.4.3 Validation of Assumptions

As noted previously, the method operates under the assumptions that 1) the longitudinal axis of the cochlea can be approximated as a line in the region of interest, 2) the radial component of motion measurement is negligible - i.e., the A-scan axis is perpendicular to the radial axis, 3) the motion of the BM is near-entirely transverse, and 4) the imaging and condition of the cochlea is stable over the course of the experiment.

1) The approximation of the longitudinal axis as a line in the gerbil base was investigated in Chapter 3, and the results are shown in Fig 3.4. This shows that over a 200 μm longitudinal range, a line is a good approximation for the longitudinal axis.

2) The radial component of motion can be determined using the program described in Chapter 3. At experiment time, we ensure that r_1 and r_2 are negligible by computing $\hat{\mathbf{z}}$ in anatomical coordinates. As an example, at orientation 1 in the experiment of Fig 4.6, the measurement axis is given by $\hat{\mathbf{z}}_1 = (l_1\ r_1\ t_1) \approx (0.90\ 0.02\ 0.43)$. The radial component contributes ~20 times less than transverse motion, and ~40 times less than longitudinal motion to the total measured motion at this angle. Our assumption of negligible radial motion in the measurement is valid as long as the radial motion at the structure of interest is not an order of magnitude larger than the other components of motion, which appears unlikely from reconstructed radial motion *ex vivo* in response to mechanical and electrical stimuli [68].

3) It has been argued that BM motion is near-purely transverse [6]. The anatomy of the BM does not facilitate motion in the radial or longitudinal direction – The BM comprises densely packed radial fibers, clamped at both radial edges forming a contiguous unit with the spiral lamina and spiral ligament [10]. Moreover, as longitudinal shearing fluid motions are expected to be antisymmetric and similar in size, there is not expected to be any net shear on the BM to drive longitudinal motion [13, 12].

4) It is critical to ensure that the results are not impacted by instability of either imaging or the health of the cochlea. The most common cause of image instability is the accumulation of fluid on the RWM, which impacts the measurement location due to refraction. Fluid accumulation was

alleviated by placing an absorbant cotton wick near the RW opening. In the experiment detailed above, after motion was measured at each longitudinal location, we checked that the image had not shifted more than 8 μm (the lateral resolution of our system) from its original state in any of the three optical directions.

The condition of the cochlea was assessed by periodically measuring DPOAEs (Sec 2.2.2). Over the course of the experiment, we did not see significant degradation in these responses. Sample DPOAEs from Ge967 are shown in Fig 4.5 **E**.

4.5 Discussion

4.5.1 Comparison to Uniaxial Measurements

The reconstructions show that longitudinal and transverse OHC-DC phase responses differ by about 180° across frequency in the ~26 kHz region of the gerbil cochlea. This significant difference could explain the variation in phase responses seen between data acquired at different measurement angles (see Fig 2.6). To investigate this possibility, we considered two previously acquired uniaxial phase responses from gerbil OHC-DC.

In Ge976, data was measured at the ~24 kHz place. The measurement angle made an 80° angle with the BM normal and had very little radial component. These data are thereby dominated by longitudinal motion. In Ge961, data was measured at the ~50 kHz place. In this case, the measurement axis made an angle with the BM normal of less than 10°. These data are thereby dominated by transverse motion.

In Fig 4.7, we compare the reconstructions with these nearly-longitudinal and nearly-transverse data. Panels **A-D** show the uniaxially acquired and reconstructed longitudinal responses at 70 and 80 dB SPL. OHC-DC is seen to lead BM across frequency in uniaxial and reconstructed responses. At 80 dB SPL, it can bee seen that this lead grows with frequency in both the uniaxial and the reconstructed response.

As for transverse motion (panels **E-H**), uniaxial and reconstructed data show that OHC-DC generally lags BM across frequency. This pattern breaks near BF at 80 dB SPL where in both

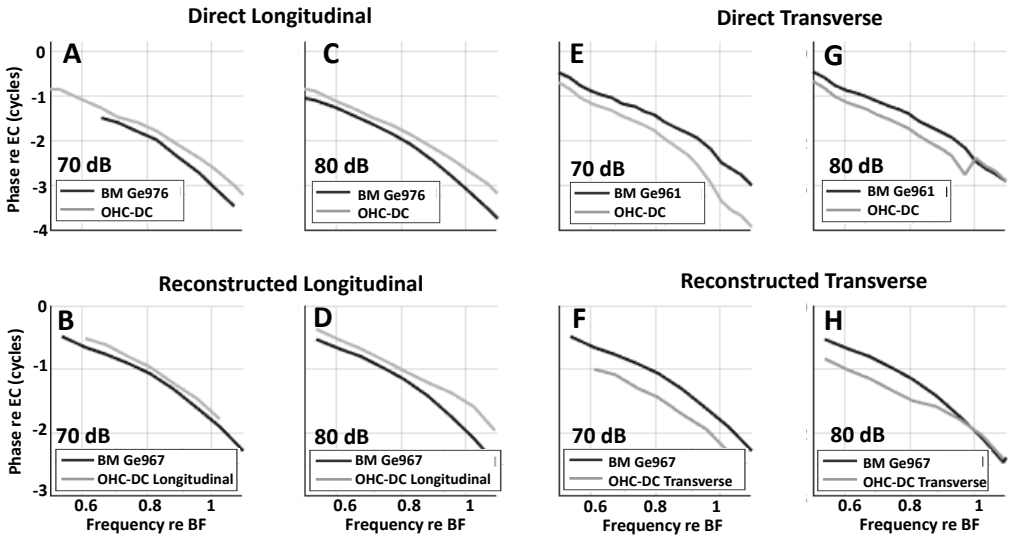

Figure 4.7: Phase responses to 70 and 80 dB SPL Zwuis stimuli at OHC-DC (light gray) and BM (black). Comparison of phase reconstructed by our method to uniaxial measurements with strong longitudinal or strong transverse components. **A**, **C** – (Ge976) Uniaxially measured phase responses dominated by longitudinal motion in the gerbil base, BF≈ 24 kHz. Measurement axis made an $\sim$ 80° angle with the BM normal. **B**, **D** – (Ge967) Reconstructed longitudinal phase responses, BF≈ 26 kHz. **E**, **G** – (Ge961) Uniaxially measured phase responses dominated by transverse motion in the gerbil base, BF≈ 50 kHz. Measurement axis made a $<$ 10° angle with the BM normal. **F**, **H** – (Ge967) Reconstructed transverse phase responses from the same position as **B** and **C**, BF≈ 26 kHz. This figure is adapted from Fig 10 of *Frost et al., 2023* [49].

uniaxial and reconstructed measurements, transverse OHC-DC phase undergoes a “phase lift" to become in-phase with BM motion. This phase lift is considered briefly in Sec 4.5.4, and in detail by *Strimbu et al* [47].

These qualitative similarities lend credence to the idea that differences between the uniaxial phase response measurements are caused by viewing angle, and indicate that our reconstruction method has succeeded. This comparison also resolves the mystery posed in Sec 2.2.3 – accounting for the viewing angle can synthesize data that appear to be in stark disagreement.

4.5.2 Physiological Implications of Broadband Nonlinearity

Comparing responses to 70 and 80 dB SPL stimuli, the reconstructions show the presence of sub-BF nonlinearity at OHC-DC in both transverse and longitudinal components of motion. The two components of motion are comparable in magnitude to one another, and significant relative to the magnitude of BM displacement. The fact that these significant active motions do not alter BM linearity sub-BF is striking.

Considering first the transverse component, the only structures between OHC-DC and the BM are the bodies of the Deiters cells [10]. If one is to accept the common notion that OHCs provide power directly to the BM in the BF region on a cycle-by-cycle basis [42, 43], the Deiters cells must be playing some regulatory role in determining at which frequencies this power transfer can occur [68]. Mechanically, the Deiters cells and BM are coupled via the rope-like Deiters stalk [23]. The transverse compression of the Deiters stalk is the difference between BM and transverse OHC displacements. This produces a nonlinear Hooke force on the BM across the frequency range.

Measurements of the pressure near the BM have shown that the force from fluid on the BM is linear sub-BF [69], so that there are both linear and nonlinear forces acting on the BM at any given frequency. Intuitively, the stimulus frequency and the stiffness of the Deiters stalk will largely determine which of these two forces is dominant, and thereby where the BM response might appear linear and where it might appear nonlinear. This simple interpretation indicates that the Deiters stalk’s stiffness may be a controlling factor for the mechanical isolation of BM from

active OHC-DC motion in the sub-BF region. This can explain the lack of sub-BF nonlinear motion at BM considering only transverse motion. In a pure-transverse framework, large active OHC-DC displacements sub-BF play no mechanical role whatsoever.

The presence of longitudinal sub-BF nonlinearity is also intriguing. As the BM is not believed to move with a significant longitudinal component, this motion cannot directly transfer power to the BM within the same longitudinal cross-section. However, the OHCs and Deiters cells do have longitudinal structure – the OHCs are tilted towards the base and the phallangeal processes of the Deiter cells are tilted towards the apex (see Fig 2.1, as well as [11, 10, 55]).

Large sub-BF OHC-DC responses do not appear to impact BM in the same cross-section, but the BM *is* moving in an amplified nonlinear fashion at more apical positions near the BP for the stimulus frequency. As opposed to the purely transverse power transfer method outlined above, the cochlea may make use of large, active longitudinal OHC-DC motion in the base to amplify motion at more apical positions.

The feedforward mechanism cannot work via direct action on BM, as the BM does not move nonlinearly sub-BF and thereby is unlikely to deliver power nonlinearly towards the apex. An alternative mechanism theorized by Guinan suggests instead that energy can be transmitted apically by modulating the cross-sectional area of the fluid spaces within the OCC [56]. This model also suggests a phasing of 2-D OHC-DC motion similar to what is found in the above data, discussed in the following subsection.

4.5.3 Lineal Motion

A striking feature of the presented reconstructions is that transverse and longitudinal components of OHC-DC motion appear to be about a half-cycle out of phase. Specifically, towards-base motion is nearly in-phase with towards-SM motion.

Cooper *et al.* discuss the fact that structures will move in longitudinal-transverse ellipses with frequency-dependent radii and aspect ratios, so that measured phase and magnitude responses will change significantly with viewing angle [6]. If this ellipse followed the expected behavior of fluid

in the scalae, this angle-dependence would be quite severe [13, 12]. However, what we find is *lineal* motion (a line is a degenerate ellipse with aspect ratio 0), showing that OHC-DC does not behave like fluid. The direction of this line is shown in Fig 4.4.

Non-elliptical motion at OHC-DC is sensible given the anatomy of the cochlea. The Deiters cells are connected to the BM by the Deiters stalk and to the reticular lamina by the phallangeal processes [23]. That is, these cells are in tension with two membranes which do not move in fluid-like ellipses.

The particular phase relationship measured here is also predicted by Guinan, and illustrated nicely in his Figs 3 and 4 [56]. As the OHCs contract, OHC-DC moves towards SM, i.e. in the positive transverse direction. This simultaneously pulls the DC towards the base, i.e. in the negative longitudinal direction.

Lineal motion is also mathematically convenient – it allows us to think of OHC-DC as having "a phase" that is not measurement-angle-dependent up to a half-cycle shift. It also explains why OHC-DC measured magnitude responses are often significantly smaller than either component of motion, as the projections of the components onto the beam axis directly subtract from one another.

4.5.4 Potential for Troughs and Phase Lifts in Uniaxial Measurements

Certain oddities occur in some OCT-measured displacement responses to high-SPL stimuli. In particular, OHC-DC measurements taken with significant components in both transverse and longitudinal directions will sometimes exhibit a phase lift and corresponding amplitude trough below the BF. An example of this phenomenon is shown in Fig 4.8.

In response to an 80 dB SPL stimulus, the amplitude exhibits a clear trough near 20 kHz. When this trough reaches its local minimum, the phase response discontinuously lifts by about half of a cycle. In this section, I will provide a mathematical explanation of this phenomenon in terms of the cancellation of longitudinal and transverse components of motion.

If the beam axis makes an angle θ with the BM normal, the measured motion at the OHC-DC

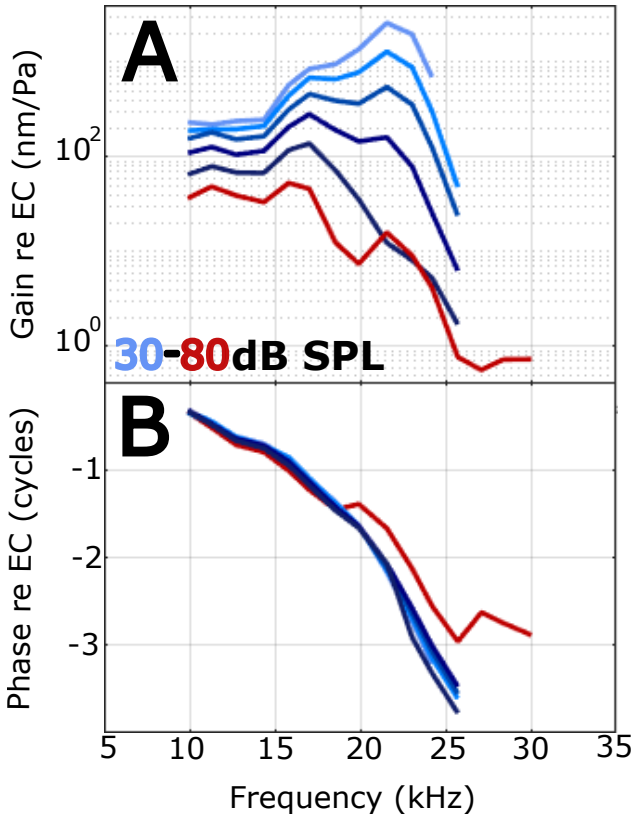

Figure 4.8: Gain (**A**) and phase (**B**) responses from Ge959, taken at a measurement angle containing significant transverse and longitudinal components. Of particular note are the amplitude trough and discontinuous phase lift seen in the 80 dB SPL response near 20 kHz.

is

$$\delta^o = d_l^o \sin\theta + d_t^o \cos\theta, \tag{4.22}$$

where $\mathbf{d}^o = (d_l^o \quad d_t^o)^T$ is the true OHC-DC motion written in its longitudinal and transverse components.

Assuming the two components are precisely a half-cycle out of phase, we can write the following formulae:

$$|\delta^o| = \left||d_l^o|\sin\theta - |d_t^o|\cos\theta\right|, \tag{4.23}$$

$$\angle\delta^o = \begin{cases} \angle d_l^o, & |d_l^o|\sin\theta > |d_t^o|\cos\theta \\ \angle d_t^o, & |d_l^o|\sin\theta < |d_t^o|\cos\theta \end{cases}. \tag{4.24}$$

Of particular interest is the winner-takes-all phase relationship of Eqn 4.24, which exhibits a discontinuity at whatever value of θ leads to the measurement axis being perpendicular to the line of motion. I will call this angle $\theta_\perp$. At $\theta_\perp$, the magnitude of the measured motion will drop to 0 according to Eqn 4.23. This means that a phase discontinuity would always correspond to a magnitude null.

Our reconstructions show that at high SPL, the magnitudes of the two motion components are comparable, but not equal, across frequency. That means this sign flip occurs near $\theta_\perp \approx 45°$, but the precise value of the normal is frequency-dependent, i.e. $\theta_\perp = \theta_\perp(f)$.

With a measurement angle θ that gives substantial weight to both transverse and longitudinal components, there may be a frequency f at which $\theta_\perp(f) = \theta$. At such a frequency, measured transverse and longitudinal components switch in dominance and a phase shift/magnitude null should occur according to Eqns 4.23 and 4.24. This may be the cause of the behavior seen in Fig 4.8 near 20 kHz.

If the response follows Eqn 4.24, the phase shift will be discontinuous, as is seen in Fig 4.8. However, the magnitude response in the data is still continuous, smoothly approaching and recovering from a local trough near the frequency at which the phase jumps. This is also expected according to Eqn 4.23, indicating that this model (motivated by our reconstructions) explains these

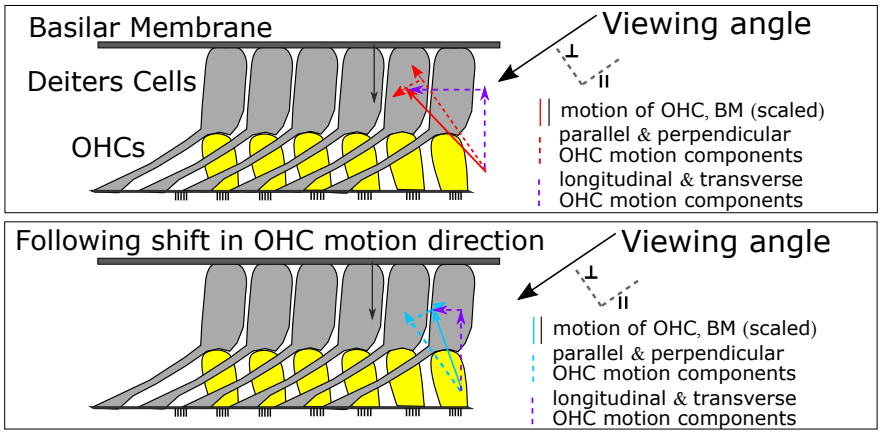

Figure 4.9: Cartoon useful for visualizing the magnitude notch and phase discontinuity, provided by Dr. Elizabeth Olson. The OHC-DC motion line is shown (solid red top, solid light blue bottom), as well as its longitudinal-transverse components (dashed dark blue) and its components parallel and perpendicular to the measurement axis (dashed red top, dashed light blue bottom). The component parallel to the measurement axis is what is measured. In the top panel, longitudinal motion is of similar magnitude to transverse motion and our measurement axis is nearly perpendicular – only a small positive motion component is present along the beam axis. Without changing the measurement angle, a change in the longitudinal component magnitude yields a parallel motion in the opposite direction. If this shift occurs continuously in frequency, the beam axis will be perpendicular to the motion line at some point, at which phase will exhibit a discontinuous jump and magnitude will fall to 0.

oddities of the data well. A cartoon visualization of this dominance switch is shown in Fig 4.9, in which two examples of OHC displacement responses are shown with the same viewing angle.

Magnitude troughs and phase notches are also seen in high-SPL responses measured at near-purely transverse measurement angles (for example, see the phase lift in Fig 4.7), but these must be different in kind from the component cancellation phenomena described above. It has been hypothesized that measured OHC-DC phase can be thought of as having two modes – one that is due to active internal motion and a second that is due to the gross simultaneous motion of the entire OCC (called “frame" motion) [47]. The active internal motion will dominate at low frequencies and SPLs, but the gross motion of the OCC (including the BM) will overpower this motion at high SPLs near BF where the active processes have saturated. The tradeoff in dominance of the internal and frame motion components is believed to lead to this other sort of trough. This is covered in detail by Strimbu *et al* [47].

As a final note, measurements taken with the beam axis pointed towards the base have a predictable behavior within this framework. In this case, positive transverse motion and *negative* longitudinal motion project positively onto the beam axis (that is, towards-base motion appears positive in the projection). Mathematically, this corresponds to a negative θ in Eqns 4.23 and 4.24.

As negative longitudinal and positive transverse motion are nearly in-phase, a notch due to a hand-off in dominance between longitudinal and transverse motion would not be expected to appear. Still, troughs may appear due to other phenomena such as the BM motion dominating internal OHC motion [47].

Moreover, measurements of OHC-DC made with the beam axis pointed towards the base would be expected to see a phase lag re BM, in line with positive transverse and negative longitudinal motion. This is precisely what is seen in previous experiments where such an angle has been achieved [6, 54].

4.5.5 Challenges

The method presented in this chapter is limited by requiring measurements at sufficiently different angles so that noise power will not become too large to reconstruct statistically significant data. Achievement of different angles can be challenging in measurements taken through the gerbil RWM due to occlusion. Anecdotally, it appears unlikely that an angular difference greater than 20° could be achieved when measuring at the ~25 kHz region in gerbil.

This increases noise power to the point that reconstructed data are only statistically significant in response to stimuli at levels greater than 70 dB SPL. This issue could be abated in part by performing an overdetermined reconstruction (Eqn 4.2), taking measurements from more than two angles. Unfortunately, this would significantly increase acquisition time.

The measurements in the example experiment were taken over about 2 hours. Each response to a one-second Zwuis stimulus takes 10 seconds to acquire, process and transfer to a hard drive, with data transfer from RAM to the hard drive being the temporal bottleneck for acquisition of a single M-Scan. Additional time must be spent ensuring the stability of the imaging and measuring

DPOAEs. Over the course of 15 minutes, sample drift is likely [70, 71]. It is often safer to pause after each M-Scan is recorded to ensure that drift has not occurred.

Rotating the preparation is also time-consuming, as we must ensure that 1) a substantial angular difference is achieved, 2) the measurements at the second angle will cover the same frequency locations as those measured at the first angle, and 3) the measurement axis contains no substantial radial component (necessary only for 2-D reconstruction, not required for 3-D reconstruction).

To achieve reconstructions of 3-D motion, or overdetermined reconstructions that allow for measurements at lower SPLs, this process would have to be significantly accelerated so that measurements at $M \geq 3$ orientations could be acquired. It would be useful, for example, if we could reduce the number of samples we would need to take at each angle to reconstruct the signal. This would reduce not only the acquisition time, but the vulnerability of the measurements to sample drift and deterioration. In the next chapter, I discuss a method by which we can reduce the number of samples required to measure dense patterns of motion across a longitudinal span, allowing for significant acceleration of data acquisition in reconstruction experiments.

Chapter 5: Compressed Sensing Vibrometry

In Chapter 4, I discussed the use of OCT to reconstruct maps of three-dimensional motion in the OCC. Many aspects of these reconstructions are limited by acquisition time, which is particularly restrictive in *in vivo* experiments. One manner by which to confront this problem is *compressed sensing*. Compressed sensing refers to a class of methods that allow for the reconstruction of signals from a sparsely sampled subset of the signal [72, 73, 74]. An example of this process is shown in Fig 5.1, where a natural image has been subsampled by a factor of 2 and then reconstructed using two methods discussed in this chapter.

Compressed sensing has seen success in medical imaging using a variety of modalities, including OCT [76, 77, 78]. OCT compressed sensing has focused on volumetric scans, composed of real-valued A-Scan magnitudes. Compressed sensing on the phase of OCT signals, encoding sub-pixel displacement [2], has not been studied previously.

Along with Nikola P. Janjušević (New York University Tandon School of Engineering), I have developed a method for compressed sensing vibrometry (CSVi) that can accelerate the acquisition of OCT displacement maps by a factor of at least ten with extremely high fidelity (normalized mean squared error less than 5%). While the method is developed with cochlear mechanics in mind, it can be used more generally for other OCT displacement measurements. This work was published in *Biomedical Optics Express* in 2023 [79].

Much of the compressed sensing work in recent years relies on neural networks, capable of achieving state-of-the-art results [80, 81]. However, neural network methods require large amounts of training data and often fail to reconstruct results that differ significantly from the training set. This is restrictive for CSVi, as we would like it to work across a range of species, stimuli and viewing angles.

The method we have used is based instead on convex optimization with total generalized vari-

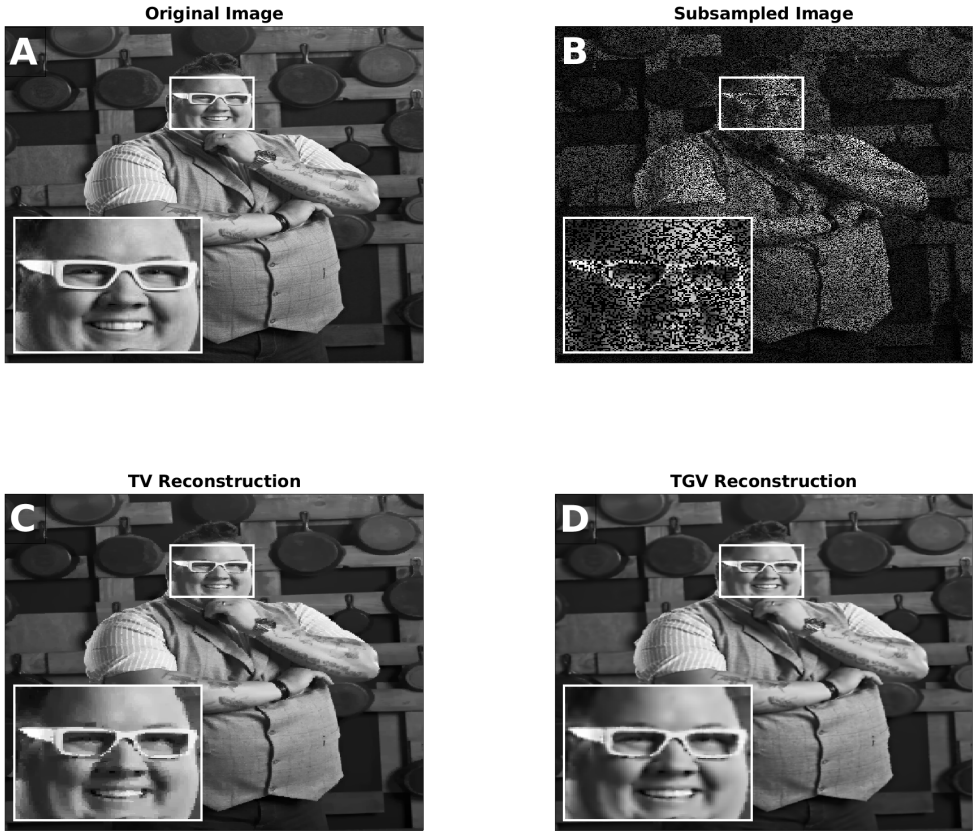

Figure 5.1: Illustration of compressed sensing for image reconstruction. Losses are reported in normalized mean square error (NMSE) between the reconstruction and the original image. **A** – Image of Graham Elliot, America's youngest four-star chef (**source**: IMDb). Inset on the bottom-left shows a zoom-in on his face, with his glasses being the feature of particular note. **B** – Subsampled version of the image in **A**, where a ramdomly chosen 50% of the original image coefficients have been set to 0. **C** – Image generated from the subsampled image in **B** using the total variation (TV) method discussed in this chapter (as well as in *Chambolle and Pock, 2016* [75]). Gross features look to be reconstructed well, while the inset shows weaknesses of the reconstruction for certain fine features. Normalized mean square error (NMSE): 1.6%. **D** – Same as **C**, except that the total generalized variation (TGV) method discussed in this chapter (as well as in *Chambolle and Pock, 2016* [75]) has been used for the image reconstruction. NMSE: 0.6%.

ation (TGV) regularization, using uniform subsampling [75]. We have tested CSVi on a diverse set of motion maps from the gerbil base *in vivo*. We have compared it to various other strategies and have found that this method performs best across all tested viewing angles, stimulus levels and stimulus frequencies. CSVi can also be interpreted as a simultaneous denoiser, as it penalizes aphysical high-spatial-frequency features.

I will first discuss the mathematical framework of compressed sensing in which our methods for CSVi are developed. Detailed derivations and algorithms are provided in Appendix C. Discussion of the application to *in vivo* cochlear displacement maps and results of the study begins in Sec 5.2.

5.1 The Compressed Sensing Framework

5.1.1 Motivation

To reconstruct dense 2-D or 3-D maps of displacement in the cochlea requires many measurements to be taken from multiple angles [66, 49]. Ideally, these measurements would also be taken at multiple SPLs and within a large spatial volume to achieve the most information possible about the character of motion in the OCC. As discussed in Chapter 4, noise levels are multiplied by the backprojection matrix used to reconstruct 2-D or 3-D maps. To achieve measurements at lower SPLs, the SNR would have to be boosted by taking measurements from more than three orientations and computing an over-constrained reconstruction. This multiplies the measurement time scale to beyond what is reasonable for an *in vivo* experiment. Even dense 2-D reconstructions at high SPLs are time-consuming, extended by the need to ensure that the animal is sufficiently anesthetized and the sample has not drifted [70, 71]. This renders the method intractable for perturbation studies where the transient changes in OCC displacement are observed after delivery of a drug (e.g. furosemide or salicylate) [41, 40].

Compressed sensing is thereby an attractive pursuit – if each set of motion measurements requires fewer samples, then 1) measurements can be achieved at more viewing angles in a more reasonable amount of time, opening the door to over-determined 3-D reconstruction experiments, and 2) reconstructions can be made quickly enough to allow measurement of 2-D or 3-D motion

in perturbation studies.

5.1.2 Form of the Data

In the most general sense, the motion in the cochlea is a frequency- and SPL-dependent 3-D vector field. Each individual displacement measurement is a 1-D projection of this vector field, and thereby the measurements are dependent on 1) 3-D position, 2) frequency, 3) SPL and 4) orientation of the beam axis. I denote this discrete frequency-domain signal by

$$\hat{\Delta}^{\theta} \in \mathbb{C}^{M \times N \times L \times F \times S}, \tag{5.1}$$

where $\theta = 1, 2, \ldots \Theta$ is an index for the orientation, M, N and L are the number of samples in the x, y and z directions, respectively, F is the number of presented frequencies and S is the number of SPLs.

For compressed sensing, we look to reduce the number of samples (i.e. *subsample*) along these axes. We need not subsample in frequency, as all frequencies are presented simultaneously in a Zwuis stimulus [46]. Similarly, all z positions are recorded at once by the nature of OCT, so we need not subsample in z [1]. The number of SPLs presented is usually relatively small, and we generally look at qualitative differences between responses to different SPLs (we never plot responses as a function of SPL) so there is no need to reduce S.

This leaves the x and y dimensions as candidates for subsampling. For the sake of simplicity and availability of data, we do not consider volumes but instead areal motion measurements (i.e. $M = 1$), so the goal is to then to reduce the number of columns in the measured areal map and recover the more densely sampled signal. If this is successful, then the same method can naturally be applied slice-by-slice to volumetric scans.

Fig 5.2 illustrates this idea. A densely sampled map is shown in panel **B**, while two examples of subsampled motion maps are shown in **C** and **D**. The subsampling occurs column-by-column. Two aspects of the data should be noted before beginning the development of the algorithms – 1) The

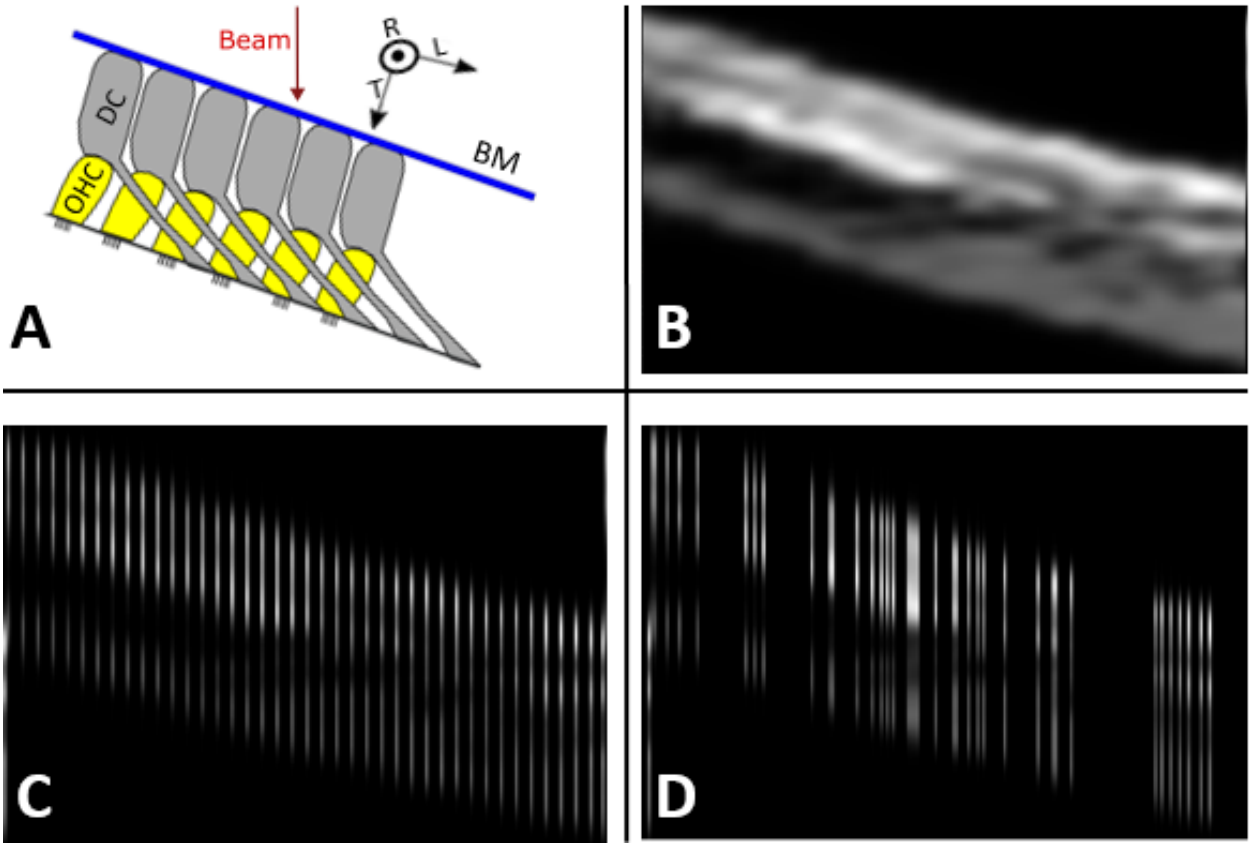

Figure 5.2: Sample vibration map with examples of subsampling patterns. **A** – Anatomical cartoon of the cochlea corresponding to the cross-section within which we measured motion. The measurement axis has strong transverse and longitudinal components, and the beam is swept across the longitudinal axis of the cochlea. **B** – Displacement magnitude map of Gerbil 988, 20 kHz component of the response to an 80 dB Zwuis stimulus. Colormap: 0 nm (black) to 6 nm (white). **C** – The same map as in **B** containing only 20% of the A-Scans, with subsampling occurring uniformly (i.e. gaps between samples are equal in width). **D** – The same map as in **B** containing 20% of the samples, with the subsampling pattern determined randomly (i.e. the width of each gap between samples is random).

data will be handled one SPL, frequency and orientation at a time, so that areal displacement maps are analogous to images, however 2) the data are complex-valued while most literature discussing compressed sensing and optimization concern real-valued signals.

5.1.3 Observation Model and the Form of the Objective Functions

CSVi is considered as a signal reconstruction problem of complex-valued images, i.e. complex arrays of size $N_1 \times N_2$, with $N_1, N_2 \in \mathbb{N}$. For notational purposes, it is simpler to consider this signal as a vector in $\mathbb{C}^N$ where $N = N_1 N_2$.

Consider some ground truth displacement signal $\mathbf{x} \in \mathbb{C}^N$. Our observation $\mathbf{y} \in \mathbb{C}^N$ is related to this signal by a *subsampling mask*, $\mathbf{m} \in \{0, 1\}^N$. The mask is a Boolean vector which denotes if a pixel at index n is measured ($\mathbf{m}_n = 1$) or not ($\mathbf{m}_n = 0$).

The *sensing map* $\mathbf{M} = \mathbf{diag}(\mathbf{m})$, encodes this masking as a matrix. Intuitively, this is an identity matrix with rows zeroed out according to the mask, so as to delete some samples while leaving others untouched. Assuming an additive noise process $\mathcal{N}$,

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathcal{N}. \tag{5.2}$$

The goal is to recover the ground truth signal $\mathbf{x}$ from the measured signal $\mathbf{y}$, but even in a noise-free environment there are infinitely many solutions to Eqn 5.2 as $\mathbf{M}$ is rank-deficient. Instead, we set up an optimization problem by seeking a solution $\hat{\mathbf{x}}$ such that $\mathbf{y} \approx \mathbf{M}\hat{\mathbf{x}}$ and that $\hat{\mathbf{x}}$ satisfies some properties expected of the underlying signal. A familiar analogue is Nyquist sampling theory, which uses the limited bandwidth of a continuous-time signal to represent it completely with a discrete-time signal.

An *objective function* is a functional $F : \mathbb{C}^N \to \mathbb{R} \cup \{\infty, \infty\}$ that encodes penalties for an optimization problem – it is the object to minimize. The objective function for the compressed sensing problem will be of the form

$$F(\mathbf{x}) = \frac{1}{2}||\mathbf{M}\mathbf{x} - \mathbf{y}||_2^2 + g(\mathbf{x}), \tag{5.3}$$

where g is called a *regularizer* [72]. The first summand encodes *data fidelity*, i.e. how near the estimated signal is to the measured signal after application of the sensing map. The regularizer encodes some feature of the signal that should be penalized according to a *signal prior* (i.e. an assumption about the ground truth signal). Examples are included in the following subsections.

Ideally, the objective function will have a unique global minimum. The solution to the problem will then be well-defined:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{C}^N}{\operatorname{argmin}} F(\mathbf{x}). \tag{5.4}$$

5.1.4 Sparsity in a Wavelet Domain

A common signal prior is *sparsity* in a domain [82]. A vector is sparse if most of its components (with respect to its representation in some basis or dictionary) are 0. This is usually written in terms of the ℓ_0 *norm*:

$$||\mathbf{x}||_0 = \sum_{n=1}^{N} [1 - \delta(\mathbf{x}_n)], \tag{5.5}$$

where $\mathbf{x} \in \mathbb{C}^N$ and δ is the Kroenecker delta function. The functional $||\cdot||_0$ gives the number of non-zero components in $\mathbf{x}$, and will be small if the signal is sparse.

The ℓ_0 norm is not actually a norm (it does not satisfy absolute homogeneity) and is also discontinuous. Instead, the better-behaved ℓ_1 *norm* is more often used as a metric for sparsity:

$$||\mathbf{x}||_1 = \sum_{n=1}^{N} |\mathbf{x}_n|. \tag{5.6}$$

The ℓ_1 norm is continuously differentiable everywhere except 0, and defines a proper norm. It is also convex, having a unique global minimum with no other local minima. It can be seen as a "next best thing" to the ℓ_0 norm intuitively, but mathematical backing of this fact is outside of the scope of this thesis (see instead *Ramirez et al., 2013* [83]). As such, ℓ_1 *regularizers* are common in sparsity-promoting optimization.

Suppose a signal is expected to be sparse in a domain, such as the Fourier domain or a wavelet domain, and the transform to that domain is represented by the invertible linear map $\mathbf{\Psi} : \mathbb{C}^N \rightarrow$

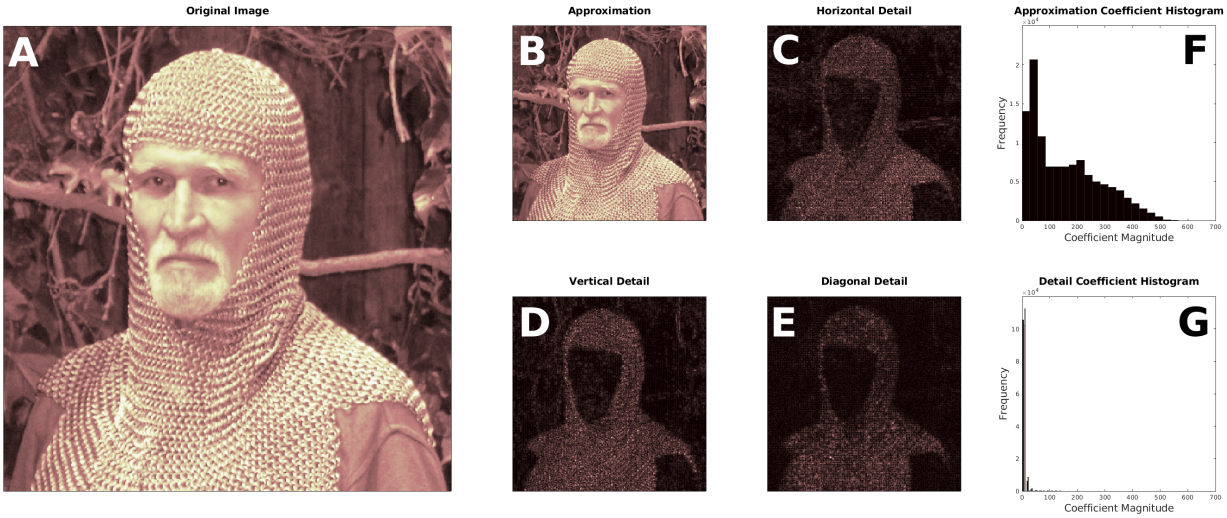

Figure 5.3: Illustration of sparsity in a wavelet domain. **A** – One-channel image of Jerry C. LaPlante. **B**-**E** – A 1-level Daubechies-7 discrete wavelet transform (DWT) of the image in **A**; Approximation coefficients, horizontal detail coefficients, vertical detail coefficients and diagonal detail coefficients. **F** – Histogram of the approximation coefficient magnitudes. **G** – Histogram of the detail coefficient magnitudes (all three sets combined), showing that they are sparse as most of the coefficients have near-zero magnitude.

$\mathbb{C}^N$. In our case, $\mathbf{\Psi}$ will be a discrete wavelet transform (DWT), and we will reformulate our optimization problem in terms of the variable $\mathbf{z} = \mathbf{\Psi x} \in \mathbb{C}^N$.

It is unreasonable to simply use $||\mathbf{z}||_1$ as a regularizer. To see this, recall the interpretation of a 1-level DWT as a set of approximation and detail coefficients [84]. In a 1-level DWT, an original image is broken down into it's approximation coefficients (interpreted as "low-low" coefficients in a 2-D filter bank), and its horizontal, vertical and diagonal detail coefficients (high-low, low-high and high-high).

The approximation coefficients form an image that is a quarter of the size of the original image, and resembles its gross/smooth features. This is generally not a sparse signal. The details, on the other hand, mark sharp transitions, and tend to be sparse for natural images. In a 2-level DWT, the approximation coefficients are broken down further into their approximation and detail coefficients, and this is repeated L times for an L-level DWT [84].

Fig 5.3 shows a sample 1-level Daubechies-7 (db7) DWT of a natural image. The approximation coefficients in panel **B** are not sparse, as can be seen by the high frequency of high-magnitude

coefficients in the histogram in panel **F**. The detail coefficients are high-magnitude only at sharp transitions in their respective directions (horizontal, vertical or diagonal), and low-magnitude elsewhere. The histogram of these coefficients in panel **G** shows that the detail coefficients are sparse, with the vast majority of coefficients being nearly 0.

As such, the better regularizer involves a selective map, $\mathcal{D}$, which analyzes only the detail coefficients [82]:

$$[\mathcal{D}\mathbf{z}]_n = \begin{cases} 0, & \mathbf{z}_n \text{ is an approximation coefficient} \\ \mathbf{z}_n, & \mathbf{z}_n \text{ is a detail coefficient} \end{cases} \tag{5.7}$$

The wavelet-domain sparsity signal prior corresponds to the detail coefficients of $\mathbf{z} = \mathbf{\Psi x}$ being sparse. The objective function in this case would be

$$F(\mathbf{z}) = \frac{1}{2}||\mathbf{M\Psi^{-1}z} - \mathbf{y}||_2^2 + \lambda||\mathcal{D}\mathbf{z}||_1. \tag{5.8}$$

The algorithm for optimizing this objective function, known as the iterative shrinkage thresholding algorithm (ISTA) [85], is derived in Appendix C and shown in Alg 1. After finding the optimal value $\hat{\mathbf{z}}$, the optimal displacement map can be found by $\hat{\mathbf{x}} = \mathbf{\Psi}^{-1}\hat{\mathbf{z}}$

5.1.5 Total Variation

Another common regularizer is *total variation* (TV) [75]. This corresponds to the assumption that the signal is approximately piecewise constant – a crude signal prior that is intuitively most accurate for cartoon images. A piecewise constant signal in continuous space would have a sparse gradient. In the discrete case presented here, the gradient is replaced by finite difference operators [75, 86].

The right-handed finite difference operators are maps from $\mathbb{C}^{N_1 \times N_2}$ to $\mathbb{C}^{N_1 \times N_2}$ defined as

$$(\partial_1^+ \mathbf{x})_{i,j} = \begin{cases} \mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}, & i < N_1 \\ 0, & i = N_1 \end{cases}$$
$$(\partial_2^+ \mathbf{x})_{i,j} = \begin{cases} \mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}, & j < N_2 \\ 0, & j = N_2 \end{cases} \tag{5.9}$$

The *discrete gradient* $\mathbf{D} : \mathbb{C}^{N_1 \times N_2} \to \mathbb{C}^{N_1 \times N_2 \times 2}$ is defined in terms of these operators as

$$(\mathbf{D}\mathbf{x})_{i,j} = \begin{pmatrix} (\partial_1^+ \mathbf{x})_{i,j} \\ (\partial_2^+ \mathbf{x})_{i,j} \end{pmatrix}. \tag{5.10}$$

If an image were piecewise constant, the discrete gradient would be zero throughout the bulk of the image and nonzero only at edges. In other words, the prior is that the discrete gradient is a sparse signal. To arrive at a regularizer, we define the norm on $\mathbb{C}^{N_1 \times N_2 \times 2}$ as a pixel-wise ℓ_2 norm across the third dimension[1], and then the ℓ_1 norm of the resultant signal [75]. That is,

$$||\mathbf{D}\mathbf{x}||_{2,1} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sqrt{|(\mathbf{D}\mathbf{x})_{i,j,1}|^2 + |(\mathbf{D}\mathbf{x})_{i,j,2}|^2}. \tag{5.11}$$

The TV objective function will be

$$F(\mathbf{x}) = \frac{1}{2}||\mathbf{M}\mathbf{x} - \mathbf{y}||_2^2 + \lambda ||\mathbf{D}\mathbf{x}||_{2,1}. \tag{5.12}$$

The algorithm for optimizing this objective function, based on the Chambolle-Pock algorithm for primal-dual splitting [75], is derived in Appendix C and shown in Alg 2.

5.1.6 Total Generalized Variation

Total generalized variation (TGV) regularization considers a relaxation of the TV assumption – it assumes the signal is approximately piecewise *linear* rather than piecewise constant [75, 86]. In

[1] A pixel-wise ℓ_1 norm can also be used here. The difference is that ℓ_1 gives an *anisotropic* penalty with grid bias, while ℓ_2 gives an *isotropic* penalty [75]. Our data does not have any grid structure, so we use ℓ_2.

the continuous case, the TGV assumption corresponds to the Hessian of the signal being sparse. In the discrete case, this is a second finite difference. The operator $\mathbf{D}$ maps from $\mathbb{C}^{N_1 \times N_2}$ to $\mathbb{C}^{N_1 \times N_2 \times 2}$. Thus, the next finite difference operation will be a map $\mathbf{K} : \mathbb{C}^{N_1 \times N_2 \times 2} \rightarrow \mathbb{C}^{N_1 \times N_2 \times 2 \times 2}$.

The map $\mathbf{K}$ is defined in terms of the right-handed finite difference operators of Eqn 5.9. Consider $\mathbf{v} \in \mathbb{C}^{N_1 \times N_2 \times 2}$, where $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{C}^{N_1 \times N_2}$ are its first and second component images along the third axis. We define $\mathbf{Kv}$ as

$$\mathbf{Kv} = \begin{pmatrix} \partial_1^+ \mathbf{v}_1 & \frac{1}{2}(\partial_2^+ \mathbf{v}_1 + \partial_1^+ \mathbf{v}_2) \\ \frac{1}{2}(\partial_1^+ \mathbf{v}_2 + \partial_2^+ \mathbf{v}_1) & \partial_2^+ \mathbf{v}_2 \end{pmatrix}. \tag{5.13}$$

The TGV objective function is defined as

$$F(\mathbf{x}, \mathbf{v}) = \frac{1}{2}||\mathbf{Mx} - \mathbf{y}||_2^2 + \lambda_1 ||\mathbf{Dx} - \mathbf{v}||_{2,1} + \lambda_0 ||\mathbf{Kv}||_{2,1}, \tag{5.14}$$

where the $\ell_{2,1}$ norm of $\mathbf{Kv}$ is defined by

$$||\mathbf{Kv}||_{2,1} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sqrt{\sum_{k=1}^{2} \sum_{l=1}^{2} |(\mathbf{Kv})_{k,l}|^2}. \tag{5.15}$$

Once again, this is the ℓ_1 norm of the pixel-wise ℓ_2 norm of the signal.

This objective function involves the introduction of an intermediate variable $\mathbf{v} \in \mathbb{C}^{N_1 \times N_2 \times 2}$, which should remain close to the discrete gradient $\mathbf{Dx}$ for the second term in the objective function to be minimized. This means that the $\mathbf{Kv}$ term resembles a discrete Hessian of $\mathbf{x}$. As this objective function is smallest for signals with sparse second derivatives, it is minimized by an approximately piecewise linear signal.

The algorithm for optimizing this objective function, based on the Chambolle-Pock algorithm for primal-dual splitting [75], is derived in Appendix C and shown in Alg 3.

5.2 Methods

5.2.1 Data Set Details

With the theory of CSVi developed, I now turn to the application of CSVi on a diverse set of *in vivo* motion maps from the gerbil base. Data were collected as described in App A.1. The data set consists of 275 areal motion maps from three gerbils *in vivo* taken at distinct orientations relative to the cochlea's anatomy, together representing the three most common orientations considered in our lab's cochlear mechanics experiments. They are:

$\theta = 1$: Data taken near the 24 kHz region along the longitudinal axis of the cochlea through the OHC-DC region. 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, with a measurement axis making an $\sim 60^\circ$ angle with the transverse direction. 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. 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.

5.2.2 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 was determined by magnitude) in the z dimension to remove outliers.

5.2.3 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.

For the ISTA method, we must choose a wavelet basis and a number of levels by which to decompose our signal. We chose a 3-level Daubechies-7 wavelet transform, for reasons explained in App D. 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.

5.2.4 Implementation Details

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. For each tested displacement map, reconstruction method, subsampling paradigm and subsampling rate, a grid search was used to determine the parameters (λ, λ_0 and λ_1) that gave the optimal NMSE, and these values were used to compute the reported NMSE and SSIM.

5.2.5 Visualization of Areal Motion Maps

The complex-valued areal motion maps in Figs 5.5 and 5.7 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 displacement gain across the map. 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).

5.3 Results

5.3.1 Comparison of Methods

We first compared the six methods – ISTA (wavelet-domain sparsity prior), TV and TGV with uniform and random subsampling – on a test set containing 20 areal maps. The results with $P = 2$, 5 and 10 are shown in Fig 5.4.

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 5.5 ($P = 10$). 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. This corresponds to the hotspot which is of critical interest, so poor performance here is not acceptable [47, 6].

The TV reconstruction, which uses a piecewise-constant prior, is not as smooth as the densely sampled signal. The TGV reconstruction, 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

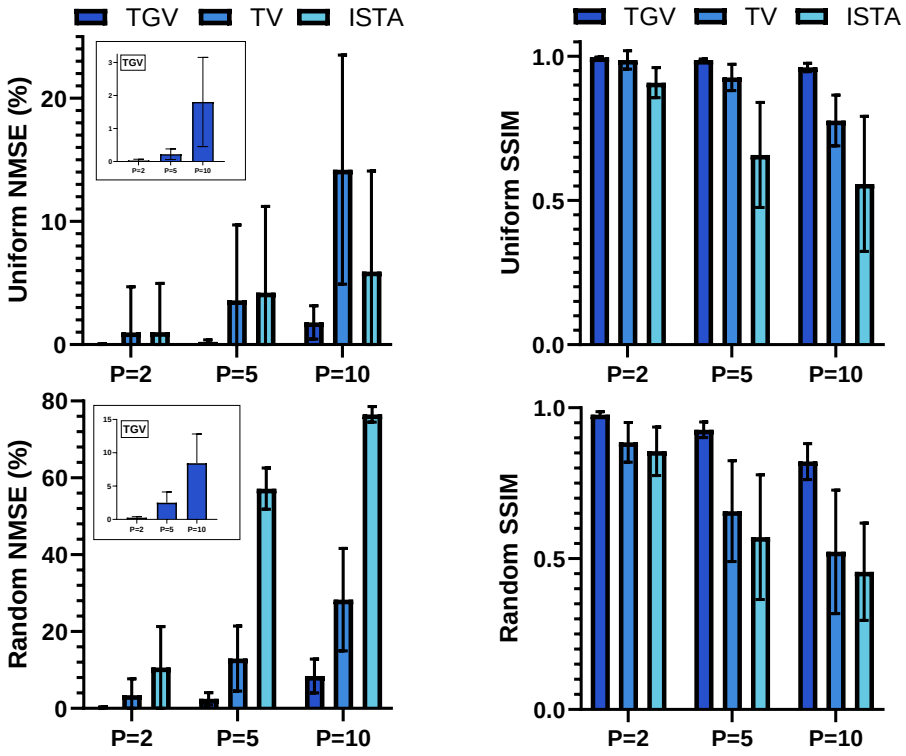

Figure 5.4: Normalized mean square error and magnitude SSIM for the three tested methods: TGV, TV and ISTA. Values displayed are sample means over a test set ($N = 20$), and error bars represent one sample standard deviation from the mean. Results are compared for both uniform and random subsampling 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%.

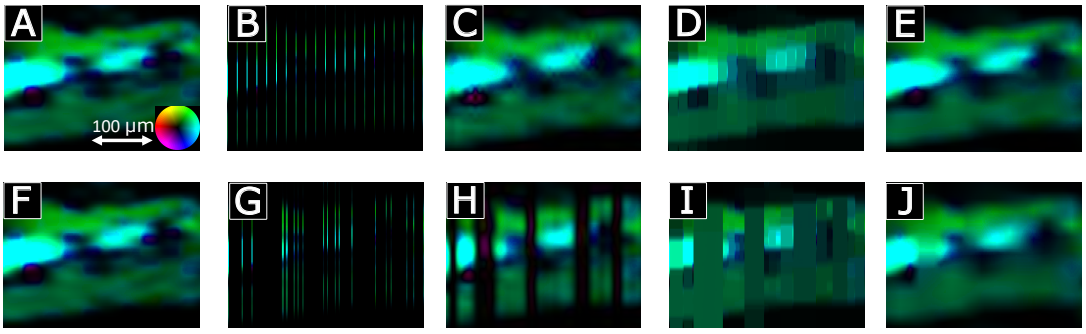

Figure 5.5: Example reconstructions using 10% of the M-Scans from the gerbil OCC. Colormap is shown in **A**, with hue representing phase re EC and saturation representing gain normalized to the maximum (further described in Sec 5.2.5). Sample: θ = 2, 80 dB SPL, 10 kHz component. Maps are 100 rows by 200 columns, or 270 μm $\times$ 300 μm. Data are further described in Sec 5.2.1. 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.

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 again outperforms ISTA and TV, but similarly shows significant artifacts at the largest bands. This is qualitatively similar to some of the artifacts seen in Lebed *et al.*'s work in compressed sensing for volumetric imaging [78]. Along with the results of Fig 5.4, these qualitative features motivate the choice of TGV with uniform subsampling as the method of further evaluation.

5.3.2 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.

Fig 5.6 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 $P = 10$. 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 5.7 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 noise, signal competition [87] and blurring [1].

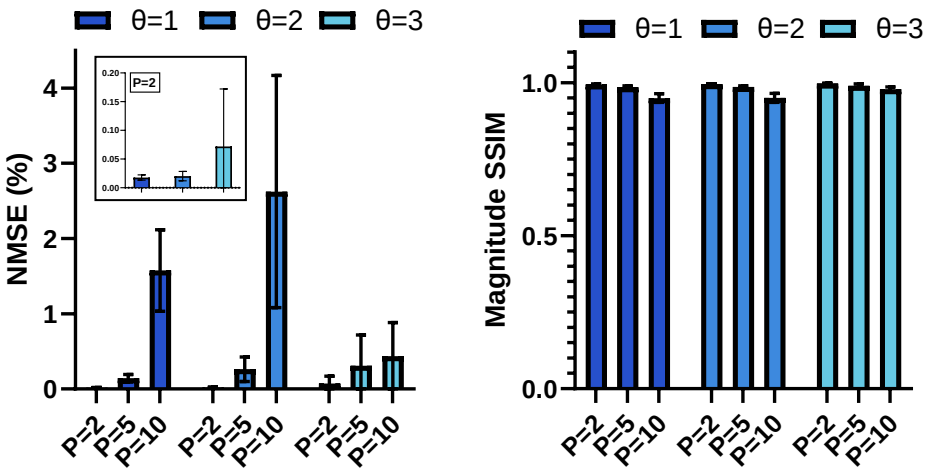

Figure 5.6: NMSE and magnitude SSIM for TGV with uniform subsampling by factors of $P = 2$, 5 and 10 across the full dataset ($N = 275$), organized by orientation. At $\theta = 1$, $N = 75$; at $\theta = 2$, $N = 100$; at $\theta = 3$, $N = 100$. Values displayed are sample means over the set at each 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.

5.4 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% NMSE. The method also appears to function well across beam axis orientations, stimulus frequencies and SPLs 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.

5.4.1 Superiority of Uniform Subsampling

Figs 5.4 and 5.5 show quantitative and qualitative superiority of uniform subsampling as opposed to random subsampling for all three tested CSVi algorithms. This was not an expected result, as random subsampling is often preferred in imaging applications [84, 76, 72, 75]. It is likely that

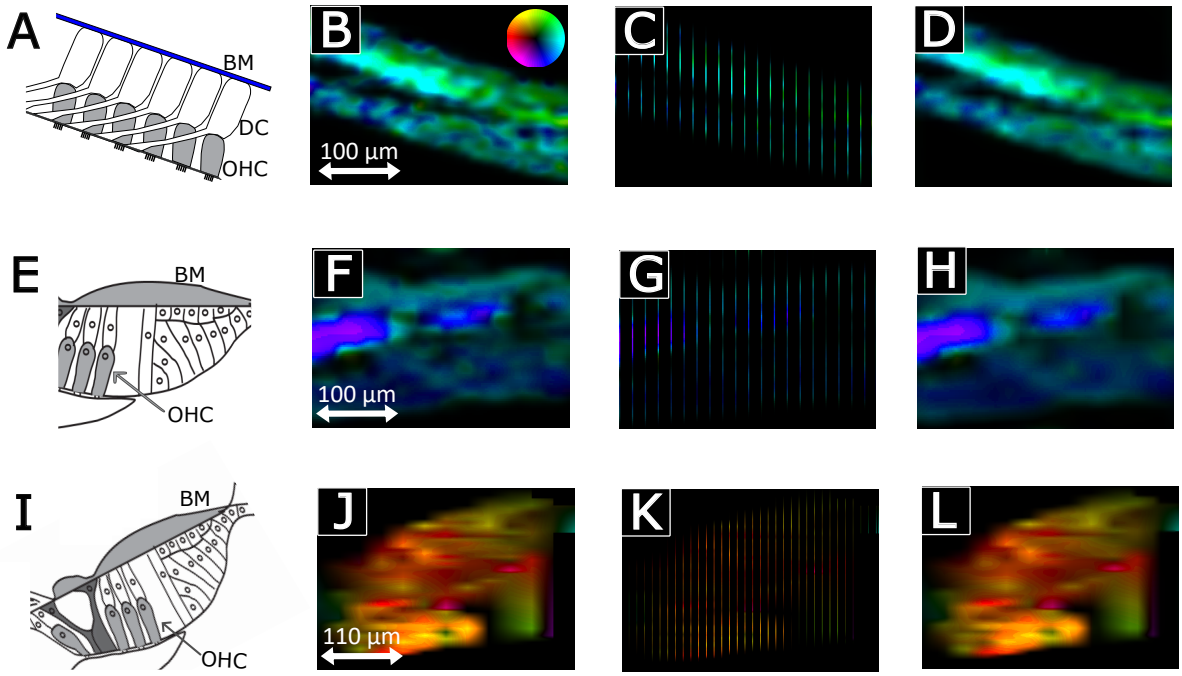

Figure 5.7: Representative sample of three reconstructions made using TGV with uniform subsampling. Colormap is shown in **B**, with hue representing phase re EC and saturation representing gain normalized by maxima. **A** – Cartoon of approximate anatomy at $\theta = 1$ with basilar membrane (BM), Deiters 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..

the poor performance of CSVi under random subsampling schema is due to the structured nature of OCT subsampling as being column-by-column rather than pixel-by-pixel. This creates the bands and corresponding banding artifacts seen in Fig 5.5.

While this work focuses on the acquisition of displacement maps, it is reasonable to infer that this result would also apply to OCT imaging. Comparative studies of this type for compressed sensing in OCT imaging have not been performed, but comparison of published results using random [78] and uniform [77] subsampling shows qualitative superiority of reconstructed images when using a uniform scheme. In particular, qualitative analysis of the reconstructions of Lebed *et al.* reveals similar banding artifacts to those seen in our CSVi results [78]. However, these studies are not directly comparable as they do not rely on the same algorithms for image reconstruction. Future study of this phenomenon in OCT imaging would be revealing and practically valuable.

5.4.2 Interpretation as a Denoiser

In addition to reducing acquisition time, CSVi can also be interpreted as a simultaneous denoiser, as can be seen from the TGV objective function of Eqn 5.14. TGV optimization does not force equality of the reconstructed maps to the measurements at sample positions. While it rewards ℓ_2 *nearness* to these measurements, it also penalizes high second derivatives that may appear in the data due to noise. This will have a smoothing effect on the entire reconstruction map, distinguishing the method from interpolation or Nyquist-style upsampling. Upsampling also incurs aliasing of both high-frequency signal components and noise which are not apparent in CSVi reconstructions.

In theory, one could also reduce noise by averaging displacement recordings over a longer period of time, but in practice this is not desirable. For one, this naturally increases experiment time. Additionally, sample drift worsens over the acquisition period, leading to diminishing returns or even *worse* results when averaging is performed over longer periods of time [70, 71]. There are also noise sources that are time-independent, such as signal competition [87].

CSVi's behavior as a denoiser is challenging to quantify – our measured displacement maps to which we are comparing are inherently noisy, so we do not have access to "ground truth."

However, we know that cochlear displacement maps are constrained by the physical morphology of the cochlea, so that smooth features are more likely to be physical than sharp transitions. Fig 5.7 shows that reconstructed signals are smoother than measurements without deeply affecting qualitative character and while maintaining quantitative similarity. This smoothing can be made more extreme by changing λ_1 and λ_0 in the TGV objective function.

5.4.3 Experimental Applications

In Chapters 3 and 4, I described the importance of measuring from multiple locations when evaluating displacements within the OCC. Valid registration of structures between two measurement angles, for example, requires dense sampling to ensure that the same position is measured at each orientation.

This is a time-limiting factor in performing experiments wherein 2-D or 3-D motion data are reconstructed. As described in the Chapter 4, this generally limits us to measurements from only two distinct orientations in a single experiment. This means that 1) 3-D motion cannot be reconstructed, and 2) the SNR of the motion data is significantly reduced relative to what could be achieved if more motion measurements were made. This has limited our results to only 2-D reconstructions at 70 and 80 dB SPL where the SNR is highest.

CSVi enables registration with far sparser sampling at each orientation, as only 20 measurements could provide motion data at 200 intermediate positions along a 270 μm span. This significantly reduces the measurement time at each angle, which could allow for the achievement of measurements from 3 or more orientations over a reasonable period of time.

This is necessary for 3-D reconstruction experiments, or overdetermined reconstruction where motion is reconstructed from measurements at more than 3 distinct orientations (offering a higher SNR). With this, CSVi could help to achieve the most complete description of cochlear micromechanics – 3-D motion data across a dense volume at various frequencies and SPLs.

Another possible application of CSVi is to perturbation studies, where a drug such as salicylate or furosemide is administered [41, 40]. Displacement responses are then monitored as their

behavior changes as a function of time. Due to the time-sensitivity of these experiments, it is not possible to monitor the response to perturbation over a densely sampled area or volume. With CSVi, however, it would be reasonable to achieve dense areal maps on a small enough time scale to monitor induced changes of displacement responses at multiple structural regions. CSVi may even allow for 2-D or 3-D displacement responses to be monitored in such perturbation studies.

Chapter 6: Model of the Cochlear Microphonic Exploring Tuning and Magnitude of Hair Cell Transduction Current

6.1 Introduction

6.1.1 Challenges in Measuring Stereocilia Motion Using OCT

The previous chapters have discussed limitations of OCT in measuring the complex 3-D vibrations of the cochlea, as well as means by which to overcome these limitations. One goal of cochlear vibrometry is the understanding of motion at the stereocilia, which are ultimately responsible for sensation (IHCs) or generation of the electromotility-inducing transmembrane potential (OHCs).

Stereocilia are embedded (or nearly embedded) in the TM, and anatomy suggests that they are stimulated by differential *radial* motion between the RL and TM [20]. Vibration measurements at the RL and TM have been measured in many different *in vivo* preparations and at many different orientations relative to the OCT beam axis. For example, transverse motion at the RL in the gerbil base has been published by multiple groups [48, 52, 47].

Achieving a purely radial measurement angle is not reasonable in most animal preparations, especially at the gerbil base where the opaque bone would block a radially-aimed beam. Instead, a method of reconstruction such as that described in Chapter 5 would be necessary to extract radial motion from multiple measurement angles. This has been attempted in mouse by Lee *et al.*, who were the first to develop a similar method for transverse-radial reconstruction [66]. However, their results are questionable – for example, they measure radial motion of the BM on the same order of magnitude as transverse BM motion, which is unlikely given the anatomy of the cochlea [6, 10]. Such oddities could be due to errors in registration, failure to account for skew, or failure to account for the contribution of longitudinal motion to the uniaxial measurements.

Other transverse-radial reconstructions have been recorded in gerbil *ex vivo* by Zhou *et al.*, who stimulated the cochlea both electrically and mechanically in a microchamber [68]. They find possible mechanical mechanisms for the transformation of transverse BM motion to radial motion at the RL. Their methods offer insight into important features such as mechanical impedance of the Deiters cells, but it is not clear to what extent their results describe *in vivo* behavior where electrical and mechanical stimulation are coupled.

In Chapter 4, I focused only on longitudinal and transverse motion, explicitly ensuring that no radial motion component was measured. The method could be used in a similar fashion to reconstruct radial motion components. However, it has proven very challenging to measure from the TM in the gerbil base. B-Scans made in the gerbil base *in vivo* using OCT devices with both 1300 nm and 880 nm central wavelengths show very little reflection at the TM [48, 47, 5]. This could be due to both its relative transparency and the fact that it lies past the entire rest of the OCC along the beam axis, so that some of the beam has been reflected prior to reaching the TM. This also induces more scattering. All of these factors reduce the brightness of the TM in the image.

Pixel brightness is related to the SNR of displacement measurements, so it is rare that statistically significant motion can be measured from the TM in these preparations[1]. Reconstruction would then multiply the noise power in these measurements so that reconstructed radial TM motion is even less likely to be statistically significant (see Sec 4.2.3). As a result, the radial differential motion that results in stereocilia stimulation at the gerbil base has not yet been measured *in vivo*.

6.1.2 The Cochlear Microphonic

Without a method for measuring differential motion of RL and TM, either directly or via reconstruction, a proxy metric for stereocilia motion must be studied instead. One common metric is the *cochlear microphonic* (CM), which is the voltage within ST in response to a sound stimulus. This voltage is modulated by the transduction current, itself generated by the stereocilia displacement. While all hair cells produce transducer current in response to stereocilia displacement, the

[1]This is not the case in guinea pig or mouse, but reconstruction studies in guinea pig have not yet been attempted and the mouse results of Lee *et al.* are questionable [66].

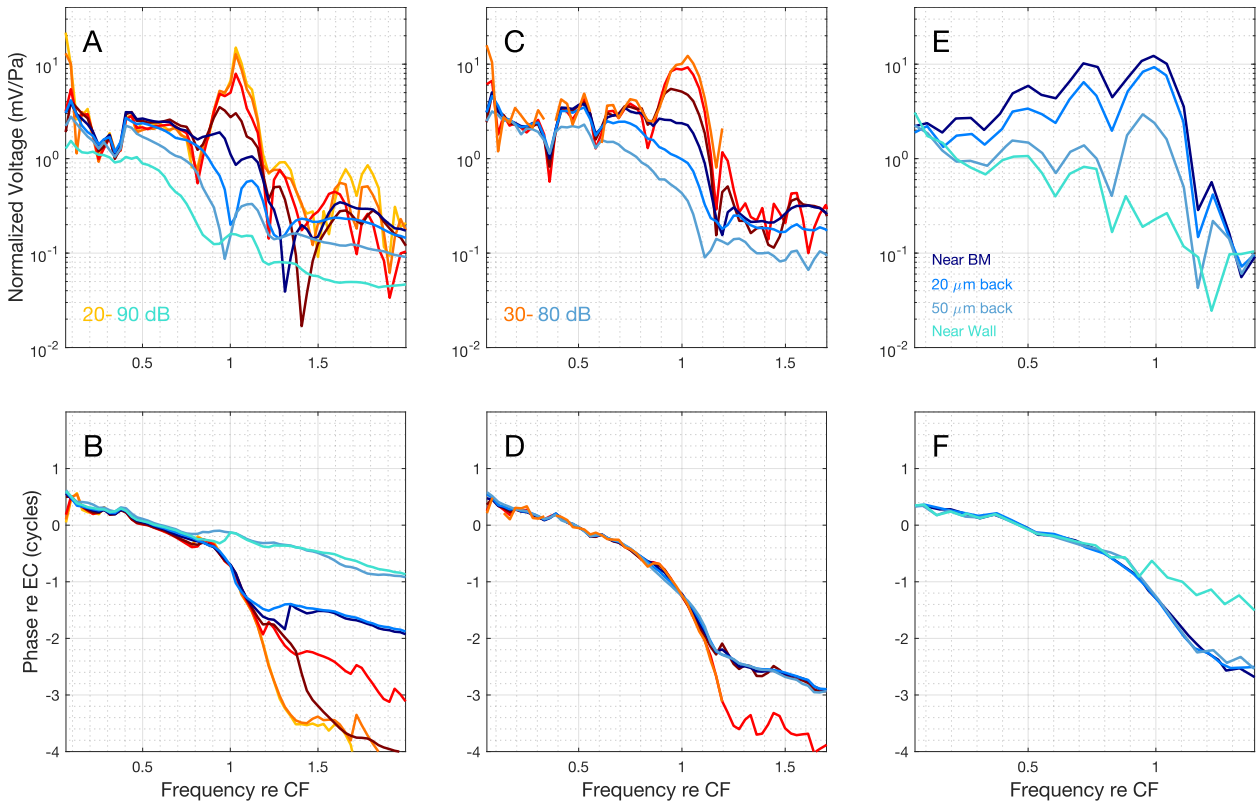

Figure 6.1: Cochlear microphonic from experimental datasets. **A** and **B** – Set 1, Ge712 [3] amplitude and phase of CM measured close to (~ 20 μm from) the BM at the 16 kHz characteristic frequency (CF) location. SPL 20-90 dB in 10 dB intervals. **C** and **D** – Set 2, Ge693 [88] amplitude and phase of CM measured close to (~ 20 μm from) the BM at the 18 kHz CF location. SPL 30-80 dB in 10 dB intervals. **E** and **F** – Ge693, amplitude and phase of CM at various distances from the BM in scala tympani at the 18 kHz CF location, 45 dB SPL. Fig 3 of *Frost and Olson, 2021* [89].

OHCs contribute significantly more to the CM than the IHCs [36]. Examples of CM responses from gerbil are shown in Fig 6.1.

In response to a single tone, stereocilia at all longitudinal positions are stimulated simultaneously with the maximum displacement occurring at the tone's best place. The CM is thereby a superposition of contributions from every hair cell, with the size of each cell's contribution to the CM depending not only on the magnitude of the response at that cell but also its distance from the measuring electrode. This can be seen in Fig 6.1 **E** and **F**, where CM responses to the same stimulus measured at four distinct locations in ST are shown.

If the electrode could be stationed within a single cell this would allow the most direct proxy measurement of stereocilia displacement. This voltage would be tuned much as the mechanical

response is. Unfortunately, such a measurement is not possible without damaging the cochlea.

The next best thing is to station the electrode very close to the BM. Local cells thereby contribute much more significantly to the voltage response at that position, so that the measured response is tuned very similarly to the mechanical response. CM near the BM is referred to as *local cochlear microphonic* (LCM), and is a more direct proxy for stereocilia displacement at a single longitudinal position.

LCM responses from two animals are shown in Fig 6.1 **A**-**D**, showing that the voltage exhibits tuning, phase accumulation and compressive non-linearity similar to BM displacement responses. However, the LCM is still affected by non-local voltage contributors as the BM is still ~80 μm transverse of the hair cells. This can be seen in the data as large troughs appear in the LCM, likely due to destructive interference between contributions from hair cells at different tonotopic locations.

6.1.3 Interpreting and Modeling the Cochlear Microphonic

CM can offer significant qualitative insight into displacement responses at the stereocilia. For example, Fallah *et al.* showed that LCM in guinea pig exhibits the same nonlinearity at frequencies below the characteristic frequency (CF)[2] in response to Zwuis stimuli that can be seen in displacement responses at OHC-DC[3, 6, 5]. This is particularly interesting because measured RL and TM displacements do not exhibit this sub-BF nonlinearity [47, 42].

Unfortunately the CM cannot offer complete information about the current at the stereocilia, as this poses an underdetermined inverse problem – we measure a single frequency response in ST and want to use it to study a distributed current source with distinct frequency responses at many locations. This means that certain features of CM are better studied using models, which can approach this inverse problem from the forward direction: assuming a reasonable current response at the hair cells, can a model produce CM responses that resemble data?

[2] In this chapter, I use characteristic frequency (CF) as opposed to best frequency (BF) to describe the displacement-maximizing frequency at a tonotopic position. These terms are generally interchangeable, but I use CF here to be consistent with the work presented in *Frost and Olson, 2021* [89].

In this chapter, I present a finite element model (FEM) of the electrical properties of the cochlea used to study CM. The FEM itself is quite simple, amounting to the solution of Ohm's law and the continuity equation on a set of nested half-cylindrical shells, stimulated by a linear current source. This simplicity is a strength of the model, as the only significant degree of freedom is the form of the current source. We assume a form of the current source based on BM displacement responses from gerbil measured *in vivo*, and determine features of the current source that must differ from BM motion to generate a CM that matches experimental data.

Our study has three major results – 1) Stereocilia displacement must be more sharply tuned than BM displacement for the model to produce the CM measured *in vivo*; 2) OHC mechano-electric transduction channels must generate about 200 pA per nm of BM displacement for the modeled response to match the magnitude of measured CM, which is about 6 times larger than current/stereocilia displacement relations measured *in vitro*; 3) the model matches data best when current from basal regions is reduced, suggesting that basal damage was present in the experiments to which our model results are compared. Much of the work presented here was published in the *Biophysical Journal* in 2021 [89].

6.2 Features of Cochlear Microphonic Data

Before presenting the model, I will discuss some features of the experimental data that were used to guide and evaluate the results of the study. Fig 6.1 shows CM data from two gerbils – Set 1 from [3] in panels **A** and **B**, and Set 2 from [88] in panels **C** and **D**. Panels **E** and **F** show CM at several distances from the BM in the same animal as Set 2.

The LCM responses show certain characteristics that are similar to BM displacement data at lower SPLs. For example, at low SPL they are tuned to a CF at which the response peaks. They also exhibit the phase accumulation characteristic of the traveling wave. On the other hand, CM is quite flat in magnitude sub-CF at low SPLs and appears to decrease monotonically at high SPLs (or when measured far from the BM), as opposed to BM displacement which monotonically increases sub-CF [39]. LCM amplitude at lower SPLs also experiences two robust notches – one below the

CF and one above the CF – that are not present in BM displacement data.

One final feature of the phase response can only be revealed by comparing CF-matched BM data and LCM data from similar preparations [3, 88, 90]. These data show that LCM phase leads BM displacement phase by approximately a quarter of a cycle for frequencies higher than about 0.7 times the CF.

These features are all sensitive to both SPL and the distance from the BM at which the response was measured. For example, in Fig 6.1 **A** and **C**, the peak can be seen to disappear as SPL is increased. In BM displacement data, the peak does become less pronounced at higher SPL, but does not dissapear entirely as it does in LCM data [39]. In panel **E**, the peak can also be seen to disappear as the electrode is moved further from the BM.

Phase accumulation is also sensitive to SPL and measurement position, as the response accumulates fewer cycles at higher SPLs and further from the BM. The phase plateaus in all measured responses, but this response occurs after fewer cycles have accumulated at higher SPLs or when measured further from the BM. These plateaus are always offset by an integer multiple of a cycle (i.e. they are in-phase). The phase is not flat at the plateaus but has a small negative slope, which can largely be attributed to the middle ear delay (about 25 μs).

Transducer current at high SPL will be greatly influenced by MET channel saturation, which is not implemented in our model. Instead, we assume that the current is linearly related to displacement. This means that our results will only be valid at lower SPL, so we restrict our study to SPL no greater than 50 dB.

I will encapsulate the above description of low-SPL response characteristics in a list of six qualitative properties that we will use to explore the FEM-predicted CM:

1. A prominent CF peak at measurement locations near the BM

2. Presence of sub- and supra-CF notches in the magnitude response

3. Loss of peak definition as measurements are taken further from the BM

4. Relatively flat amplitude response in the sub-CF region

5. Traveling wave phase accumulation through several cycles for measurements within 50 μm of the BM, lost further from the BM

6. A phase lead of the LCM with respect to BM motion at frequencies above about 0.7 times CF

Property (1) evinces tuning of the current source, which is the input of our model. Our ability to match the tuning of CM will reveal properties of the tuning of MET current responses. Property (2) may be related to the impact of non-local current, as the notches are likely caused by destructive interference between out-of-phase current responses. Along with properties (3), (4) and (5), this will inform the locality of our model determined by simulated physical properties.

The phase shift of property (6) is not expected to occur within our model, as we will model MET current as being proportional to BM displacement. The quarter-cycle shift is believed to be due to a phasing property of the current source itself, perhaps generated by a TM resonance [88]. If the shift were to appear in our modeled CM responses, this would cast doubt on the TM resonance theory.

6.3 Methods

6.3.1 Basics of the Model

The FEM is implemented in COMSOL Multiphysics as a set of nested half-cylindrical shells, as seen in Fig 6.2. This represents half of the gerbil cochlea, uncoiled with a length of 10 mm. The central half-cylinder is the organ of Corti (OC), 80 μm in radius. It is nested in the 10 μm-thick BM, which is itself nested in the 520 μm-thick ST. Lastly, this is nested in the 100 μm-thick outer wall. Sizes were determined according to known anatomy of the gerbil cochlea [17], and approximated to fit the uniform half-cylindrical shape. In particular, we used geometric values similar to those found 2.5 mm from the RW, where the data of Fig 6.1 were measured. A more realistic geometry is discussed in App E.1.1, but the character of the results was not changed by implementation of a more complex geometry.

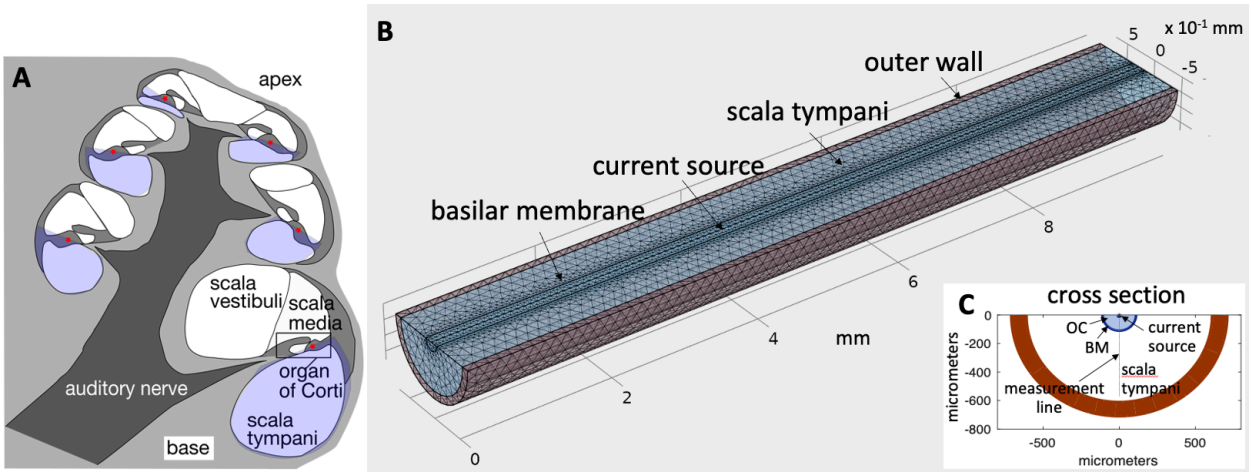

Figure 6.2: **A** – Cross-section of the gerbil cochlea, with the spiraling ST marked in blue. The red star represents the spiraling current source. **B** – Geometry of the model as it appears in the COMSOL Multiphysics user interface, representing an uncoiled version of the blue region in **A**. The outer wall is distinct from the larger fluid space, and the approximate position of the BM is marked by a half-cylindrical surface. The line current source can be seen on the flat surface. **C** shows a cross-section 2.5 mm from the base, and the vertical line from source to wall is where simulated voltages are recorded. OC = organ of Corti, BM = basilar membrane. Fig 1 of *Frost and Olson, 2021* [89].

MET current is modeled as a linear current source density running down the central axis of the rectangular boundary, denoted $A(x, f, S)$ [A/m], where x [m] is longitudinal distance from RW, f [Hz] is stimulus frequency and S [dB SPL] is stimulus magnitude. We observe the simulated voltage response along the line segment perpendicular to the rectangular boundary, running from the source to the outer wall at $x = 2.5$ mm (Fig 6.2 **C**). This corresponds to the CM at the 19.5 kHz region, similar to where the data in Fig 6.1 were collected.

6.3.2 The Boundary Value Problem and Physical Parameters

Within the interior, the the charge density continuity equation is satisfied:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0, \tag{6.1}$$

where ρ [C/m^3] is charge density and J [A/m^2] is current density. We also have Ohm's law:

$$\mathbf{J} = \sigma \mathbf{E}, \tag{6.2}$$

where σ [S/m] is the conductivity of the material and $\mathbf{E}$ [V/m] is the electric field. We are interested in finding the voltage V, which is defined in terms of electric field as

$$\mathbf{E} = -\nabla V. \tag{6.3}$$

The outer curved wall of the cylinder is considered to be fixed at ground ($V = 0$). The three other boundaries (the rectangular and semicircular boundaries) are assumed to be electrically insulating. Mathematically, this is to say that at these boundaries $\mathbf{J} \cdot \mathbf{n} = 0$ for any normal vector $\mathbf{n}$. This is justified by the fact that no current flows through the oval window or round window, as well as the presence of tight junctions at the RL isolating SM and ST [91, 92]. The RL condition also justifies our modeling only the ST, as opposed to modeling all three fluid chambers of the cochlea.

The conductivities are the only electrical parameters of the model. The OC, BM and ST are all

modeled as having the conductivity of saline (σ = 1.45 mS/cm) making them electrically identical. In this sense, the model can be seen as a single half-cylindrical fluid space (OC, BM and ST) nested in the outer wall. The outer wall conductivity σ_W is not likely to be similar to that of saline. Instead, we write $\sigma_W = \sigma/K$, where $K \in \mathbb{R}^+$ is a unitless scaling parameter. It turns out that tuning K amounts to tuning the locality of our model, and we explore the impact of varying K in Sec 6.7.1. Aside from the form of the current source A, this is the only variable parameter in the model.

6.3.3 Implementation

Equations 6.1, 6.2 and 6.3 were solved subject to the specified boundary conditions by using the finite element method based in the COMSOL Multiphysics software package. A tetrahedral mesh with elements no smaller than 40 μm and no larger than 55 μm on each side was used. This was sufficient to achieve convergence of the simulation for frequencies as high as 25 kHz. The use of this relatively coarse mesh is possible because the details of the OC and BM have not been included in the model geometry. A mesh twice as fine (in terms of minimum side length) was tested for a single input and no characteristic differences were seen in the model output. The current source is defined with a linear resolution of 11.1 μm.

All simulations were performed on a PC running the 64-bit Windows 10 Enterprise operating system, equipped with 32 GB of RAM and an Intel Xeon W-2133 CPU (3.6 GHz clock rate). We observe the model output at 32 frequencies between 1 kHz and 25 kHz. On this system, a 32-frequency sweep takes 34 minutes and 6 seconds for a single SPL.

6.3.4 Current Source

We initially model A as being proportional to BM displacement, denoted $z(x, f, S)$. We write $A = Cz$, where C [A/m^2] is a scaling constant denoting the sensitivity of the MET channels. Our model is linear, so C can be tuned *post hoc* to match the magnitude of CM data, but it is still useful to have an initial guess for the value of C. We based this guess on *in vitro* data, which suggests that

the stereocilia displacement-current gain for a single OHC is approximately 33 pA/nm [93]. There are three OHCs in each longitudinal cross-section, and these rows of OHCs are spaced about 10 μm apart along x, so that a there are about 300 OHCs per mm [94]. We have 300 OHCs/mm $\times$ 33 pA/nm per OHC, giving a first estimate of $C = 10$ kA/m^2.

Scaling Symmetry

BM motion data is usually recorded at one position $x = x_0$ at various stimulus frequencies and SPLs, i.e. we have $z(x_0, f, S)$ but need $z(x, f, S)$ to derive the current source. To overcome this problem, we use the principle of *scaling symmetry* [95, 96], which says that displacement responses in the cochlea depend only on the ratio of the stimulus frequency and the CF rather than on absolute position.

Under scaling symmetry, for a given location x_0, frequency f_0 and sound pressure level S_0, $z(x, f_0, S_0)$ can be found in terms of $z(x_0, f, S_0)$ using the tonotopic map $CF(x)$ [Hz], which maps each location x to the characteristic frequency at that location. Specifically,

$$z(x, f_0, S_0) = z\left(x_0, \frac{f_0}{CF(x)} CF(x_0), S_0\right). \tag{6.4}$$

Müller measured the tonotopic map in gerbil and fit it to an exponential function:

$$CF(x) = 0.398 \times \left(10^{0.2(11.1-x)} - 0.631\right), \tag{6.5}$$

where x has units mm and $CF(x)$ has units kHz [97]. This tonotopic map is valid for frequencies higher than 2 kHz.

BM Displacement Data

We use BM displacement data based on measurements of Ren *et al.* [98] with CF 15.5 kHz. We chose these data because they were taken over a wide frequency range, facilitating the use of scaling symmetry. The data shown in Fig. 6.3 were smoothed and the phase responses, which were

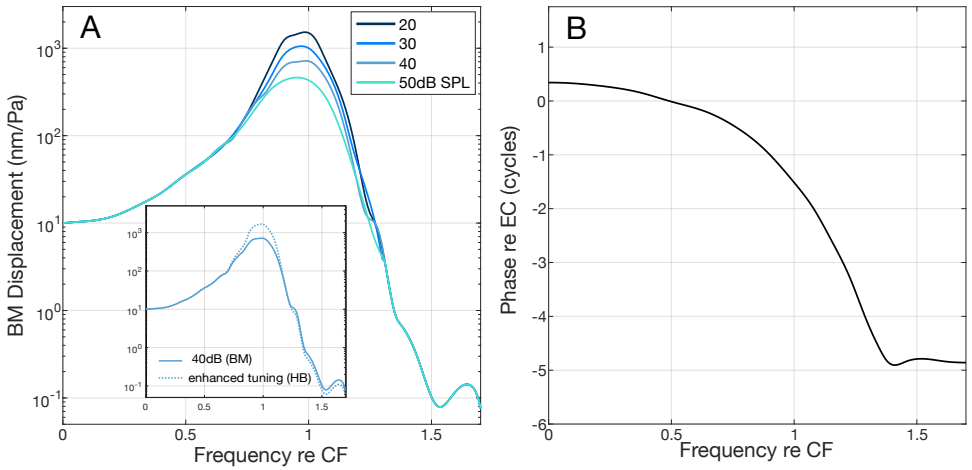

Figure 6.3: The current source is initially assumed to be proportional to BM displacement shown here. **A** Amplitude and **B** phase of BM displacement, based on gerbil data with CF 15.5 kHz [98], at sound pressure levels 20-50 dB SPL. The phase was nearly independent of SPL and the small variations were not included. Phase is shown referenced to the ear canal pressure. The data are plotted versus frequency/CF. Inset in **A** shows enhanced tuning of hair bundle (HB) displacement input used in Sec 6.5. Fig 2 of *Frost and Olson, 2021* [89].

only mildly SPL-dependent, were taken to be independent of SPL. To explore the impact of tuning at the stereocilia on the simulated CM, we also tested the modeled response to a more sharply tuned current source as shown in the inset of panel **A**. Using scaling symmetry, we generate line current source values proportional to this BM displacement data at 32 frequencies between 1 kHz and 25 kHz.

Phase References

There is a subtle but important technical issue regarding the phase of our current stimulus. The BM displacement data of Fig 6.3 and CM data of Fig 6.1 reference phase to ear canal pressure. Between EC and the stapes, there is a delay incurred by the middle ear, T, measured in gerbil to be $T \approx 25$ μs [99]. Scaling symmetry assumes that all delays are intracochlear, and thereby cannot be

applied directly to data referenced to EC pressure. We can transform ear canal phase ϕ_E [cycles] to stapes phase ϕ_S [cycles] simply subtracting this middle ear delay (it appears as addition because I am subtracting delay, a negative number):

$$\phi_s = \phi_E + fT. \tag{6.6}$$

This allows us to use scaling symmetry by deriving the input A according to stapes-referenced phase. When we observe the output of our model, we want to compare it to the data of Fig 6.1 which is referenced to EC pressure. This is easily done by adding back in the middle ear delay.

6.4 Results for Input Based on BM Displacement

6.4.1 Modeled Voltage Response

As described above, we first use a line-current source that is proportional to the BM displacement data in Fig 6.3 after application of scaling symmetry. Fig 6.4 shows the predicted voltage along the radial line at the 19.5 kHz location (see measurement line in inset in Fig 6.2) for 20, 30, 40 and 50 dB SPL. The predicted voltage at four points along this line are shown – on the line source and at distances 55, 110, 160 and 410 μm from the line source. The phase of the BM displacement used to generate the current stimulus is included for reference. For these simulations, we used $K = 50$ – this choice is discussed in Sec 6.7.2.

Fig 6.4 **A** and **B** show the predicted voltage at the position of the line source. A strong peak is present in all curves in panel **A**. Panels **C** and **D** show responses 55 μm from the line source, a distance that is not experimentally accessible but is useful for illustration. The magnitude of the peak is reduced, a supra-CF notch has developed and the phase accumulates 3 cycles before plateauing.

The predicted amplitude of the CM 110 μm from the line current source (Fig. 6.4 **E**), which approximately corresponds to the LCM measurement location from Fig 6.1 **A**-**D**, contains substantial CF peaks at 20 and 30 dB SPL, with notches skirting the peaks. At 40 and 50 dB SPL the peak is

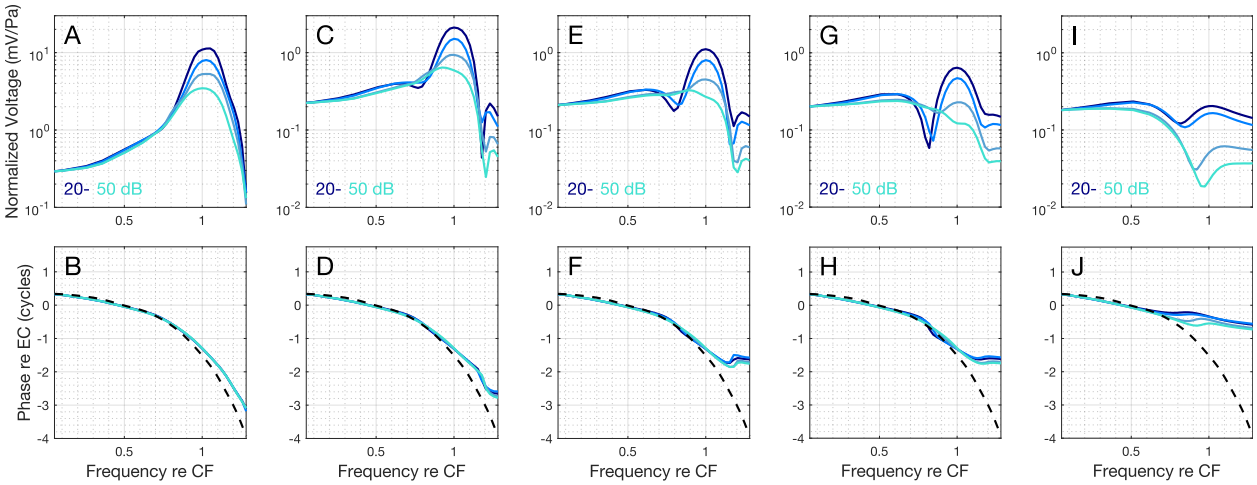

Figure 6.4: CM prediction under the assumption that current is proportional to BM displacement. $K = 50$, channel sensitivity = 33 pA/nm (starting value). Predictions are shown at five locations along the line segment 2.5 mm from the base of the cochlea (see Fig. 6.2 **C**). Magnitude and phase: **A** and **B** – at the position of the line current source; **C** and **D** – 55 μm from the source; **E** and **F** – 110 μm from the source; **G** and **H** – 160 μm from the source; **I** and **J** – 410 μm from the source. The dashed lines in the lower panels are the phase of the input (BM displacement) used to generate the current stimulus. Fig 4 of *Frost and Olson, 2021* [89].

substantially washed out. This trend continues in Fig 6.4 **G**, 160 μm from the line current source. The predicted LCM phase accumulation at the BM is 2 cycles and the phase contains ripples where the amplitude notches occur.

The predicted LCM voltages at a point 110 μm from the line current source are compared to the experimental data of Set 1 in Fig 6.5 and of Set 2 in Fig 6.6. In these figures the model's current source strength, which was reasoned from *in vitro* data to take a value of 33 pA/nm, has been multiplied by factors of approximately 6 (to 200 pA/nm, Set 1) and 8 to (260 pA/nm, Set 2) in order to match the experimental LCM amplitudes in the sub-CF band.

6.4.2 Comparison to Experimental Data

Sec 6.2 includes a list of qualitative properties of the experimental data used to explore the FEM predictions. Property (1) was the presence of a CF peak in LCM for stimuli below 60 dB SPL. In Fig 6.4 **E** and **F** (corresponding to measurements made 110 μm from the line source), the FEM predicts a significant peak at 20 and 30 dB SPL, but a peak is barely present at 40 dB SPL

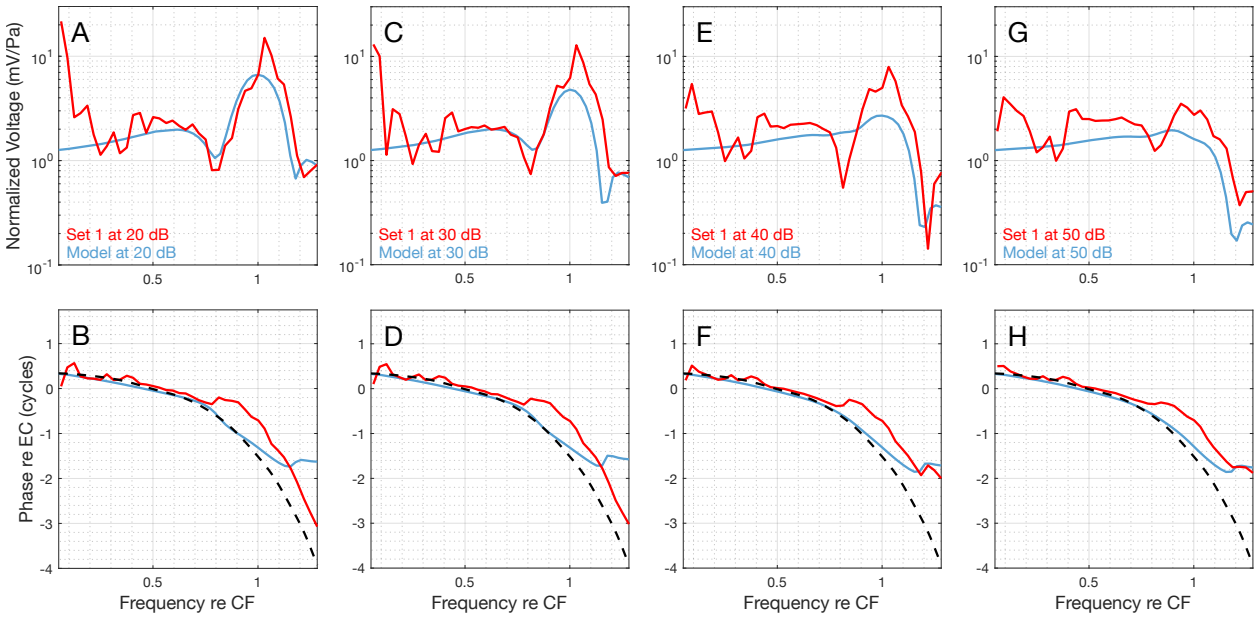

Figure 6.5: Model LCM predictions 110 μm from the line current source ($\sim$ 20 μm from the BM) compared to experimental Set 1. Results (magnitude and phase) are shown at 20 (**A** and **B**), 30 (**C** and **D**), 40(**E** and **F**) and 50 dB SPL (**G** and **H**). CM predictions are based on the assumption that current is proportional to BM motion. $K = 50$, channel sensitivity is adjusted from starting value of 33 pA/nm to 200 pA/nm to align with the experimental result. The phase of the current stimulus is shown as a dashed line in the lower panels. Fig 5 of *Frost and Olson, 2021* [89].

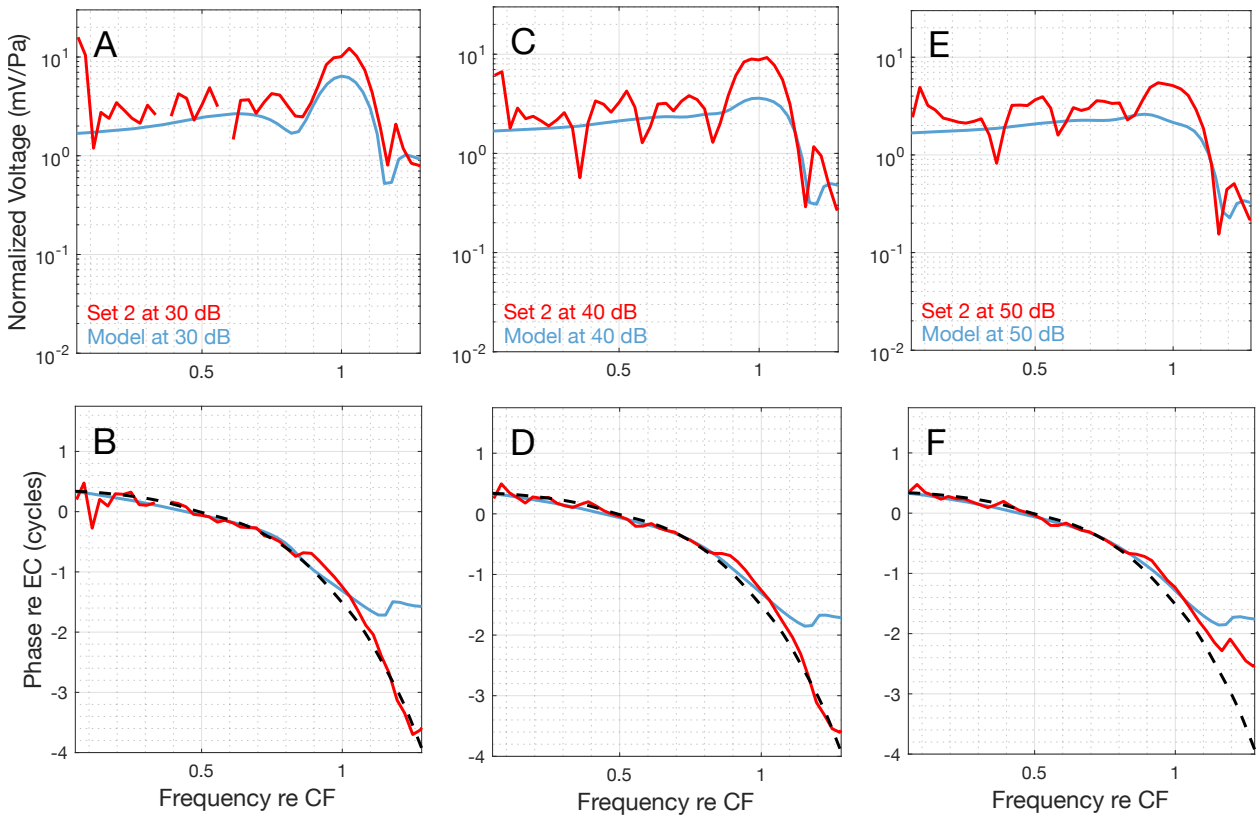

Figure 6.6: Model LCM predictions 110 μm from the line current source ($\sim$ 20 μm from the BM) compared to experimental Set 2. Results (magnitude and phase) are shown at 30 (**A** and **B**), 40 (**C** and **D**) and 50 dB SPL (**E** and **F**). CM predictions are based on the assumption that current is proportional to BM motion. $K = 50$, channel sensitivity is adjusted from starting value of 33 pA/nm to 260 pA/nm to align with the experimental result. The phase of the current stimulus is shown as a dashed line in the lower panels. Fig 6 of *Frost and Olson, 2021* [89].

and the 50 dB SPL peak is washed out. Direct comparison to the experimental data in Sets 1 and 2 (Figs. 6.5 and 6.6 respectively) shows that the experimental peak at the CF is under-predicted. In particular the peak is washed out in the model results at 40 dB SPL, where it is still strong in the experimental LCM. The inability of our model to predict the degree of tuning in the LCM using the BM-proportional line-current source supports the idea that HB motion is more sharply tuned than BM motion. This is discussed further in Sec 6.5.

Property (2) was the presence of sub- and supra- CF notches. In both predicted and experimentally measured CM voltage, the sub-CF notches near the BM only appear when there is a strong peak. The presence of the peak is poorly predicted by the model at 40 and 50 dB and the sub-CF notches are not predicted at these SPLs either. This argues that the sub-CF notch is the result of cancellation involving currents from a defined peak apical to the measurement location. The supra-CF notch is predicted relatively accurately at all SPLs.

Properties (3) and (5) – loss of peak definition and phase accumulation as we move further from the BM – are apparent in our predicted CM voltages in Fig 6.4. Finally, property (4) – the flatness of the sub-CF band – is met accurately at all SPLs in both sets. This is controlled in part by the value of K, which has been chosen to best match data in the sub-CF region (see Sec 6.7.2).

Property (6), the presence of a phase shift and lead of LCM with respect to BM motion, was not predicted by the FEM model. The predicted phase does not undergo the lift and pronounced lead apparent in the experimental data, especially clear in Fig 6.6. This experimental feature likely requires a shift in the phase of the current source as discussed in E.2.

6.4.3 Premature Plateau

The predicted LCM phase begins to plateau at a lower frequency than the measured LCM phase. It is tempting to say this plateau is due to the *fast mode* response. The fast mode is a fluid pressure wave that is in-phase with stapes displacement – as the fluid in the scalae is incompressible, compression at the stapes creates a so-called fast wave that travels at the speed of sound in water towards the round window, which then would expand opposite to the stapes to maintain fluid

volume.

Experimentally measured intracochlear displacements, LCM and pressure often feature a phase plateau above the BF [100]. The phase plateau in pressure is readily attributed to the fast mode [24], and the fast mode pressure can also produce a fast mode of BM displacement (visible as the plateau in amplitude and phase in Fig 6.3). This could similarly lead to a fast-mode-induced phase plateau in LCM.

However, in the results of Fig 6.4, the predicted LCM phase plateaued at frequencies where the input phase was still accumulating. It is likely that the LCM plateau in the FEM arises from non-local current from the CF peak rather than from the fast mode. In that case the flat LCM phase observed in the model is akin to the nearly flat phase of a *wave-fixed DPOAE*. A wave-fixed DPOAE arises from the CF peak place, where the phase is nearly invariant, and then travels or projects to the measurement location [101]. With this, the basis for the LCM phase plateau is understood, but the deviation between measurements and model predictions remains unexplained.

The difference could be due to our model being unrealistically non-local, meaning that the parameter K has been set too high. Another reasonable explanation is that the measurements are taken from gerbils in which the basal region was compromised due to experimental exposure. In that case, in the experiments, the basal CF peaks would be attenuated. The local responses with traveling-wave phase accumulation would then remain dominant through higher frequencies. The known fragility of the basal region makes this explanation credible [102]. I will explore these two possibilities in Sec 6.8, but first return to the question of the insufficient tuning in the predicted LCM peak.

6.5 Results for Input with Enhanced Tuning

As discussed in Sec 6.4.2, the use of a line-current source based on BM displacement was insufficient to reproduce the tuning seen in the experimentally measured LCM. Guided by our findings, and encouraged by previous modeling results [88], we expect the line-current source to be more sharply tuned than BM displacement. We generated a second line-current source with

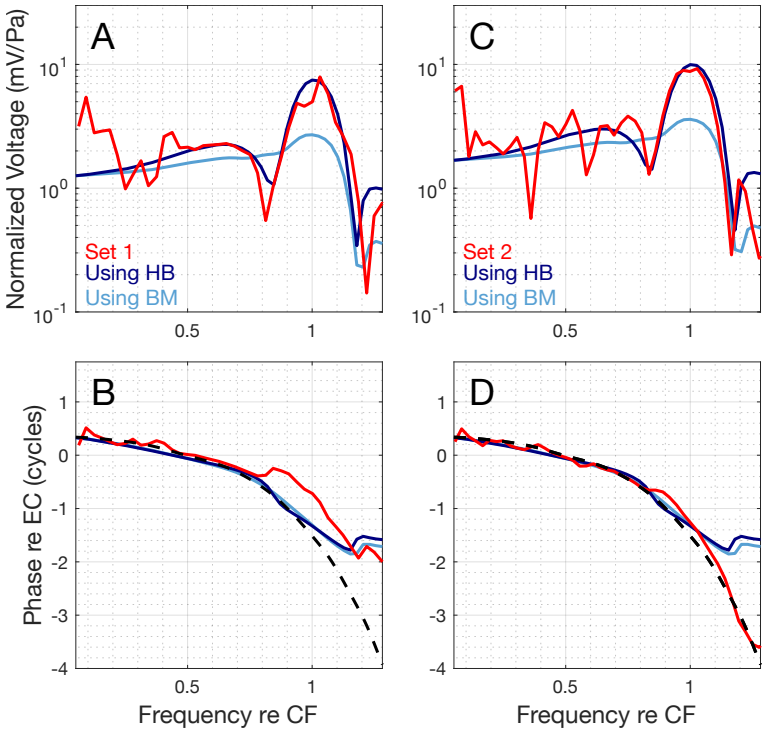

Figure 6.7: Model LCM predictions 110 μm from the line current source compared to experimental data. Current source is based on the enhanced tuning of HB motion. Comparisons are made at 40 dB SPL. **A** and **B** – Set 1 comparison. $K = 50$ and channel sensitivity adjusted from starting value to to 200 pA/nm (same as Fig. 6.5). **C** and **D** – Set 2 comparison, $K = 50$ and channel sensitivity adjusted to 260 pA/nm (same as Fig. 6.6). The phase of the current stimulus is shown as a dashed line in each phase plot. Fig 7 of *Frost and Olson, 2021* [89].

heightened tuning compared to BM motion, with a modest gain factor of ~2 at the peak. The magnitude of this source is labeled “HB" in the inset of Fig 6.3 **A**, and the phase is identical to that of the BM-proportional source.

The predicted LCM voltages at a point 110 μm from the current source using both the BM-based and HB-based current sources are compared to the data of Sets 1 and 2 in Fig 6.7. With the more tuned current source, the LCM peak is more prominent while the phase is affected very little. Just as in Fig. 6.4, the current source strength was 200 pA/nm (Set 1) and 260 pA/nm (Set 2). This value was chosen to match the sub-CF amplitude, which was not substantially affected by the enhanced tuning.

6.5.1 Comparison to Experimental Data

Whereas the predictions using the BM-proportional stimulus underestimated peaks at 20-30 SPL and showed no peak at 40 dB, the more highly tuned stimulus provides accurate predictions of both the presence and shape of the 40 dB SPL peak. In addition, the use of the more tuned current source better predicts the presence and magnitude of the sub-CF notch, satisfying properties (1) and (2).

The distance-dependent properties (3) and (5), as well as the sub-CF flatness property (4), were met similarly with and without the application of additional tuning. Regarding the phase shift of property (6), the predicted phase is barely affected by the additional tuning. This is not surprising, as we have not altered the phase of the current source. Importantly, HB-based responses also exhibit the premature phase plateau seen for BM-based phase responses.

6.6 Quantitative Difference in Sensitivity

Under the original assumption that the OHC line-current source is proportional to BM displacement, and that the OHC sensitivity is 33 pA/nm, the model-predicted voltage magnitudes were a factor of six (Set 1) or eight (Set 2) times smaller in the sub-CF region than measured LCM voltages. At the peak, this factor was even larger.

Using the more sharply tuned current source (Fig 6.7), the model predicted the tuning of CM correctly and increased the peak value. However, the magnitude of the sub-CF band was not significantly affected by the enhanced tuning. Thus, including the HB tuning did not change the finding that the OHC transducer channel sensitivity should be 6-8 times larger than our assumed sensitivity (33 pA/nm), or ~ 200-260 pA/nm.

This could mean either that a) channel sensitivity actually is 6-8 times larger than our starting value of 33 pA/nm, or b) the HB displacement is 6-8 times larger than BM displacement in the sub-CF band (and still more tuned than BM motion in the peak). We consider each of these in turn.

Our estimate of the transducer sensitivity was based on *in vitro* experimental data taken in a

gerbil hemicochlea [93]. Those experiments found that basal OHCs produced current with maximum gain of 50 pA/nm. We picked a modest starting sensitivity value of 33 pA/nm considering that all OHCs are likely not operating at maximum gain. However, these *in vitro* data may underestimate sensitivity as a result of temperature, ion concentration and holding potential differences between *in vitro* and *in vivo* conditions, as explored by Kennedy *et al* [103]. This exploration was in mouse, but the relative value shifts under changes in experimental conditions are still of interest.

For example, the holding potential in the hemicochlea experiments from which we derive our initial sensitivity estimate was -70 mV [93]. Kennedy *et al.* show decreasing the holding potential from -84 mV to -134 mV can increase the gain by about a factor of two in *in vitro* mouse OHCs. This more negative holding potential better represents *in vivo* conditions, where both the endocochlear potential and the intracellular potential are present. Kennedy *et al.* also found that changes in calcium concentration can increase OHC sensitivity by up to a factor of two.

He *et al.* [93] found that decreasing holding potential to -140 mV in the hemicochlea experiments nearly doubled the current response for a single given input stimulus. They also found that a decrease in calcium concentration can increase current by a factor of three for a given motion stimulus. Thus *in vitro* versus *in vivo* differences in cell voltage and calcium concentration could be a source of the discrepancy between our model predictions and measured LCM. This argues that the predicted $\sim$ 200 pA/nm value for OHC sensitivity is accurate.

Alternatively considering possibility (b), OCT-measured displacement responses in mice, indicate that HB radial motion motion is smaller than BM motion at frequencies lower than the CF [66]. Recent transverse RL motion measurements from gerbil also indicate that RL displacement is less than twice as large as BM displacement [47] (although it should be noted that transverse RL motion does not directly stimulate stereocilia). It is thereby most likely that the quantitative discrepancy is a result of underestimation of the sensitivity, but future radial RL and TM measurements may adjust this interpretation.

6.7 Outer Wall Conductivity

6.7.1 Effects of Varying the Wall Conductivity Ratio

In cable models, the locality of the CM measurement along the cable – the degree to which the CM represents current from local OHCs – is determined by a space constant λ [m], a parameter that is proportional to the ratio of radial to longitudinal resistances. An increase in λ corresponds to a less local measurement, i.e. it corresponds to more distant current sources having a larger impact on the measured voltage. In our FEM, the unitless parameter K modifies the locality of measurement by controlling the conductance of the outer wall relative to the fluid.

Previous FEMs have similarly found the conductivity ratio to be an important controlling factor of current flow through the cochlea [104]. An intuitive connection exists between K and the λ of cable models. In a cable model, the longitudinal resistance is entirely due to perilymph resistance, and the radial resistance is due to both perilymph and wall resistance. In our FEM, higher values of K correspond to higher radial resistance relative to longitudinal resistance, leading to more contribution to the measured CM from non-local sources. On the other hand, lower K values correspond to lower radial resistance relative to longitudinal resistance, and will yield CM measurements that more closely resemble the local current source.

Fig 6.8 shows the effect of varying the locality constant using the initial model, with current source proportional to the BM displacement of Fig 6.3. Responses are shown 110 μm from the line source (approximately at the BM, panels **A** and **B**) and 210 μm from the line source (panels **C** and **D**). These results confirm that K functions as a locality-controlling parameter, with lower values of K (more conductive outer wall) corresponding to LCM measurements that more closely resemble the local current stimulus.

The peak is less pronounced for high K, where the contribution from more distant current sources is larger. The peak shape was not changed substantially by reducing K below 50. Thus, the finding that current proportional to BM motion is not sufficiently sharply tuned to predict the tuning of measured CM is robust to changes in K. Because the peak amplitude did not change

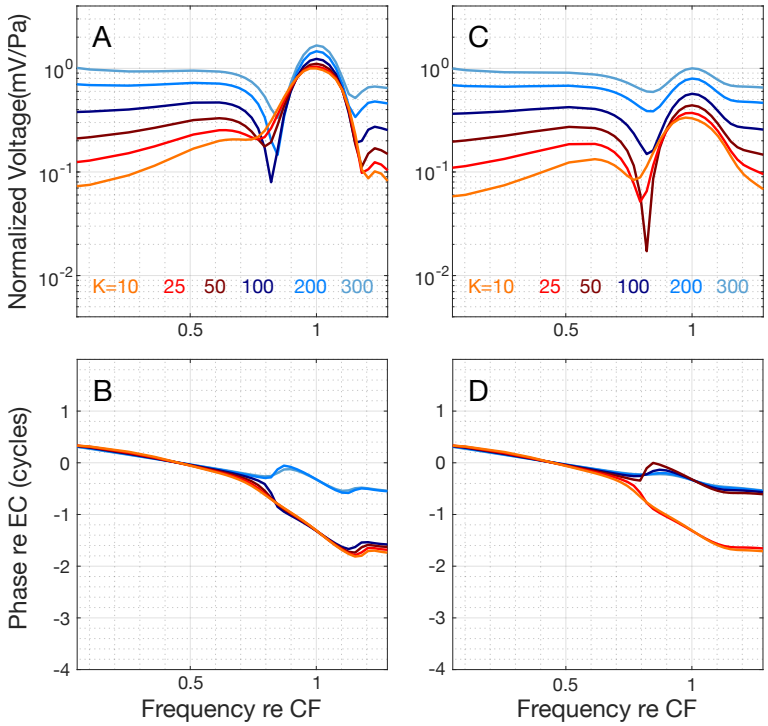

Figure 6.8: Effect of variations in lateral wall conductivity, $\sigma_W = \sigma/K$, where σ is the conductivity of the ST saline solution. Predicted CM at 20 dB SPL with K=10, 25, 50, 100, 150 and 300. **A** and **B** – 110 μm from the line current source; **C** and **D** – 210 μm from the line current source. Channel sensitivity is set to 33 pA/nm. Source is proportional to BM data if Fig 6.3. Fig 8 of *Frost and Olson, 2021* [89].

significantly as K was varied, our findings relating to transducer sensitivity are robust to choice of K as well.

6.7.2 Choice of the Outer Wall Conductivity Ratio

Responses near the BM using lower values of K exhibit an upward slope in the sub-CF, whereas the response in this band is flat for higher K. A slight upward slope is seen in the sub-CF region of the experimental data of Sets 1 and 2, best matching the slope of the $K = 50$ response. This leads to the choice of $K = 50$ for our simulations.

A K value of 50 corresponds to a wall conductivity value of $\sigma/K = 14.5/50$ mS/cm = 0.29 mS/cm, corresponding to a resistivity of $\sim 3500\ \Omega\,\text{cm}$. This is greater than the resistivity of muscle but less than that of bone, and similar to the resistivity of fat [105]. This seems reasonable, considering the tight junctions present in the tissue of the cochlear wall for maintaining the different fluid and electrical compartments of the cochlea. The wall resistivity we use is also in reasonable agreement with a previous estimate of $\sim 1000\ \Omega\,\text{cm}$ [106].

6.7.3 Relationship to Phase Accumulation

At lower SPLs, the phase in experimental LCM data traverses multiple cycles before leveling off. On the other hand, our modeled responses travel through only about two cycles before leveling off at all SPL. This difference indicates that in the experiment, local current is more dominant over remote current than it is in the FEM model. This suggests that we have made the outer wall resistance too large, which has the effect of emphasizing non-local current sources.

In Fig 6.8, it is shown that higher K values predict a lack of phase accumulation, but lower K values still did not produce the several cycles of phase accumulation observed in the experimental data of Fig 6.1. Even making $K = 1$ did not produce the experimentally measured phase accumulation.

In a cable model the response is made increasingly local by reducing the space constant, λ, and can be made completely local by setting $\lambda = 0$. In the more realistic FEM, the spread of current

cannot be reduced without limit. Therefore, other properties must be considered to explain the experimentally measured degree of locality in evidence in the several cycles of phase accumulation.

6.8 Causes of the Phase Plateau and Notch Behavior

6.8.1 Expected Relationship to the Phase Plateau and Notches

We explored a second potential explanation for the phase plateau – that the cochlear responses were less robust in the region basal to the measurement location. If this were the case, then the peak responses at supra-CF frequencies would be attenuated, leading to less contribution from basal non-local sources. The phase may then behave more similar to the local response, which continues to accumulate rather than plateauing. The base of the gerbil cochlea is known to be fragile [102], so such an explanation is within reason.

Damage of this sort would likely have an impact on other aspects of the response. For example, we hypothesized that the notches seen in the FEM's magnitude response are due to phase cancellation between current components local to the measurement location and peak (high-amplitude) non-local current sources. If this is the case, reduction of current from different regions in the cochlea would not only impact the phase plateau but also these notches.

6.8.2 Effects of Damage

To test these hypotheses, we considered three manipulations – 1) current apical to the 18 kHz location is set to zero, 2) current basal to the 21 kHz location is set to zero, and 3) current basal to the 21 kHz region is reduced by a factor of two. These frequencies are chosen to be symmetric about and near to the 19.5 kHz CF of the measurement location. While the case of reduced current is physically reasonable, the nulled apex and base cases are simply theoretical to observe the impact they have on notches in the voltage response.

Using a current source proportional to BM displacement at 20 dB SPL, we observed the model predictions and compare them to that of the intact simulations[3]. The predictions are shown in Fig

[3]To relate this to the HB-based tuning of Fig 6.7, note that the BM-based current source at 20 dB SPL is similar to

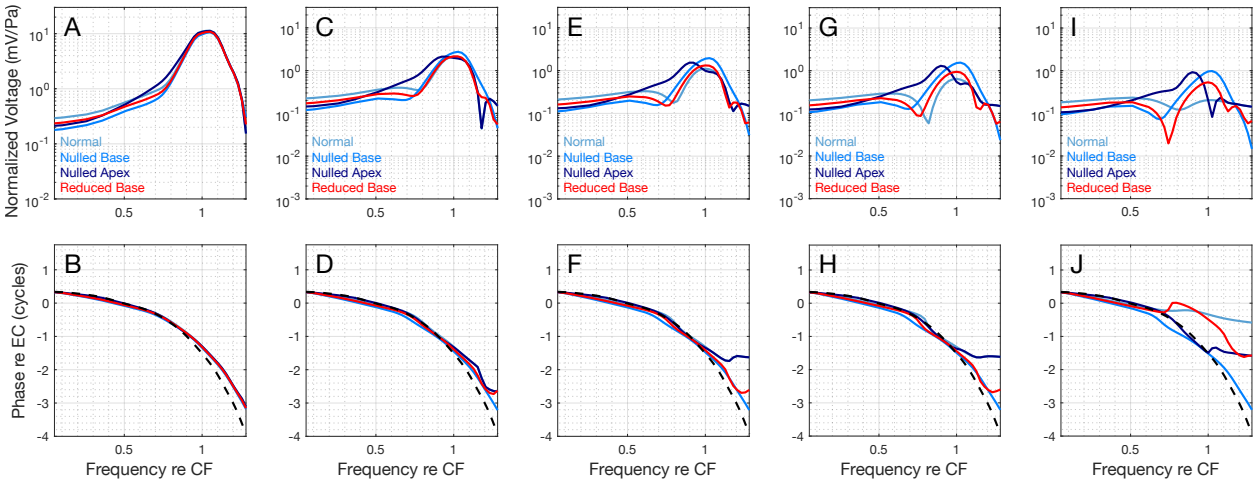

Figure 6.9: Effect of nulling or reducing the current source basal or apical of the measurement location (19.5 kHz CF place). Panel sets show magnitude and phase at various distances from the current source . The original current source, based on BM displacement tuning, and the original channel sensitivity, 33 pA/nm, are used. In the nulled-base case, the current source from the base to the 21 kHz place is set to 0. In the reduced-base case the current source from the base to the 21 kHz place is reduced by half. In the nulled-apex case, current from the apex to the 18 kHz place is set to 0. SPL = 20 dB SPL, $K = 50$. **A** and **B** – CM predictions at the position of the line-current source. **C** and **D** – 55 μm from the line current source. **E** and **F** – 110 μm from the line current source. **G** and **H** – 160 μm from the line current source. **I** and **J** – 410μm from the line current source. The phase of the current stimulus is shown as a dashed line in the lower panels. Fig 9 of *Frost and Olson, 2021* [89].

6.9.

When current basal to the measurement location was nulled, the supra-CF notch disappeared at all locations. The phase mimicked local BM phase more closely, accumulating many cycles monotonically. This is consistent with our hypothesis that both the premature phase plateau and supra-CF notch are due to the interference of current from basal OHCs. In addition, the sub-CF notch became less pronounced (Fig 6.9 **G**).

When the current apical to the measurement location was nulled, at locations 55 and 110 μm from the current source the supra-CF notch appeared at the same frequency and size as in the intact response, as did the premature phase plateau. In this case, the sub-CF notch disappeared. Thus, in the FEM, the sub-CF notch requires both basal and apical current to fully form.

Considering the less extreme case in which basal current contributions are halved, both the sub-

the HB-based source at 40 dB SPL.

CF and supra-CF notch remain. As in the case where the basal current is entirely nulled, we see more phase accumulation than in the original study. However, as opposed to the nearly distance-independent phase accumulation apparent in the nulled-base case, a distance-dependent plateau develops in the reduced-base case. This resembles what is seen in experimental CM measurements more accurately than the predictions from both the nulled-base case and the intact case.

6.8.3 Relationships between Notch Frequencies

To more concretely explore the notches in our model output, we consider the phase of the current source. Fig 6.10 shows the amplitude and phase of the BM motion data used to generate our current source, with the values at frequencies where notches occur highlighted with orange dots. The phases at the two notch frequencies are separated by almost exactly 2 complete cycles. The x-axis is labeled as frequency, but due to tonotopy this axis could also be interpreted as longitudinal location, basal (left) to apical (right).

Next we note the size of the response and the phase behavior where the response is destructively interfering with a notch position (differing by an odd number of half cycles). There are five such locations within the peak region, marked with numbered blue dots. Destructively interfering component (1) is at a point where the phase is changing slowly relative to all other interfering components. Thereby there are many current sources near this frequency that have a destructively interfering phase relative to the two notch frequencies, generating a high-magnitude distributed interfering component.

Consider the supra-CF notch. When we null current basal to the location of measurement, we null the three largest destructively interfering current components (labeled 1, 2, 3). The remaining destructively interfering components (4, 5) are both low in amplitude and at frequencies where phase is changing rapidly. Thereby, we expect to see the supra-CF notch disappear when basal current is nulled. In the case where apical current is nulled, we retain components 1, 2 and 3, and expect to retain the supra-CF notch. This expected behavior is observed in Fig 6.9.

Now consider the sub-CF notch. When we null current basal to the location of measurement,

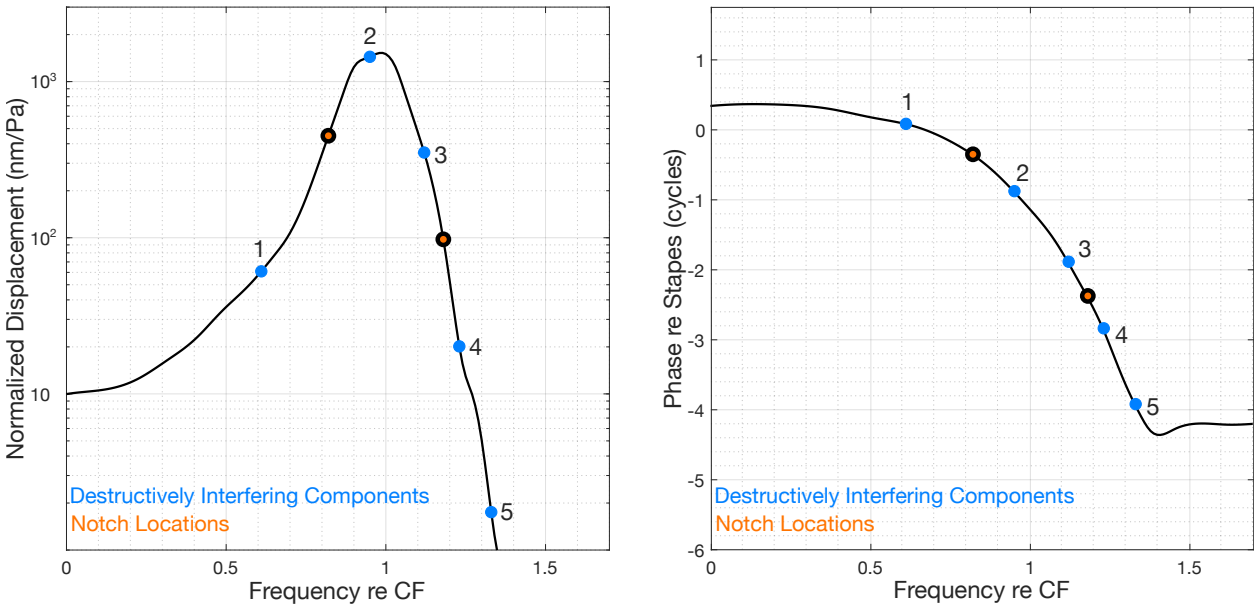

Figure 6.10: Exploration of the basis of prominent notches. Amplitude and phase of the basilar membrane displacement data used to generate the model input. As in Fig. 6.3, except here the reference is stapes (so that all phase variation occurs within the cochlea). Highlighted in orange are the values of the amplitude and phase at the frequencies where notches appear in our model predictions, ~ 0.8CF and 1.2CF. These correspond to phases of 0.36 and 2.36 cycles. Highlighted in blue are values corresponding to frequencies where the phases are half of a cycle off from the phase at the notch frequencies. Current components at the frequencies in blue will interfere destructively with those in orange. Fig 10 of *Frost and Olson, 2021* [89].

we have eliminated the destructively interfering component (1) where phase is varying most slowly. However, we maintain the other four destructively interfering components, some of which are at high amplitudes. Eliminating component 1 might decrease the prominence of the notch, but not eliminate it entirely. If instead we null apical current, we lose four destructively interfering components (2-5), including the two at the highest amplitudes, so it is likely that the sub-CF notch will be greatly reduced in prominence if not eliminated. This is what is seen in Fig 6.9, in which the sub-CF notch is lowered in prominence by nulling basal current, and nearly eliminated by nulling apical current.

To summarize, the sub-CF notch seen in the model results is due to interference between basal, local and apical components, whereas the supra-CF notch (and subsequent phase level-off) is due mostly to interference between local and basal components. The reduced phase accumulation in the intact modeled responses compared to experimental data is reasonably due to a lesser contribution of basal OHCs to the ST current.

Phase cancellation is one source of notches in the CM, but notches could also be present in the current source itself. These notches would not be captured by these results, but CM predictions where notches are present in the current source are briefly considered in App E.2.

6.8.4 Furosemide Recovery Results Indicating Effects of Basal Damage

In the experimental results reported in [15], LCM was measured before and after intravenous furosemide was administered to reduce endocochlear potential. A previously unexplained finding from that study is replotted in Fig 6.11, which shows LCM responses close to the BM before and after recovery from furosemide. The phase accumulation post-furosemide was greater than pre-furosemide, and a supra-CF notch was reduced. We explored this finding with the FEM.

We first hypothesized that the changes post-furosemide were due to the outer wall becoming more leaky, as has been observed [107], which would be equivalent to a reduction in K. However, as shown above, reducing K did not produce a substantial increase in phase accumulation.

Another possible explanation was that following furosemide, the more basal regions did not

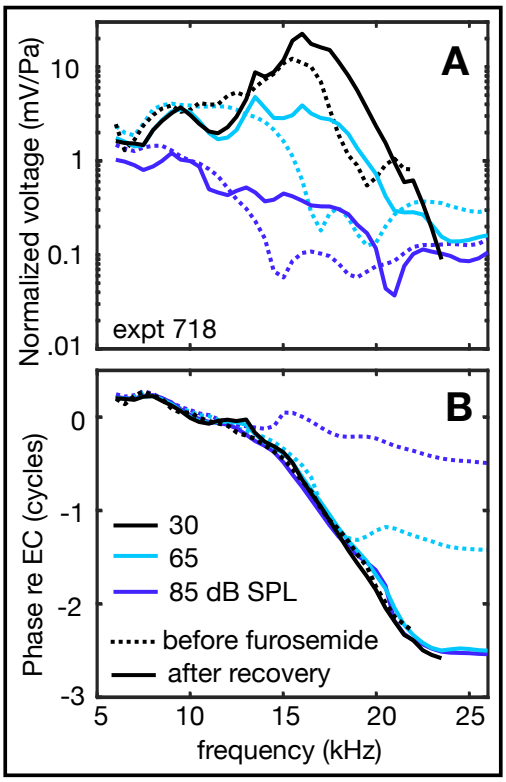

Figure 6.11: Experimental data related to model prediction with basal current nulled or reduced. CM measured close to the BM, before administering furosemide and after recovery from furosemide (3.5 hrs later). **A** – Normalized voltage amplitude. **B** – Phase relative to ear canal. Note that traveling wave phase accumulation is present even at high SPLs after recovery, indicating that the response was more local after recovering from furosemide than it was before administration of furosemide. A reasonable explanation is that the more basal region had not recovered fully, reducing the interference of non-local basal current. Fig 11 of *Frost and Olson, 2021* [89].

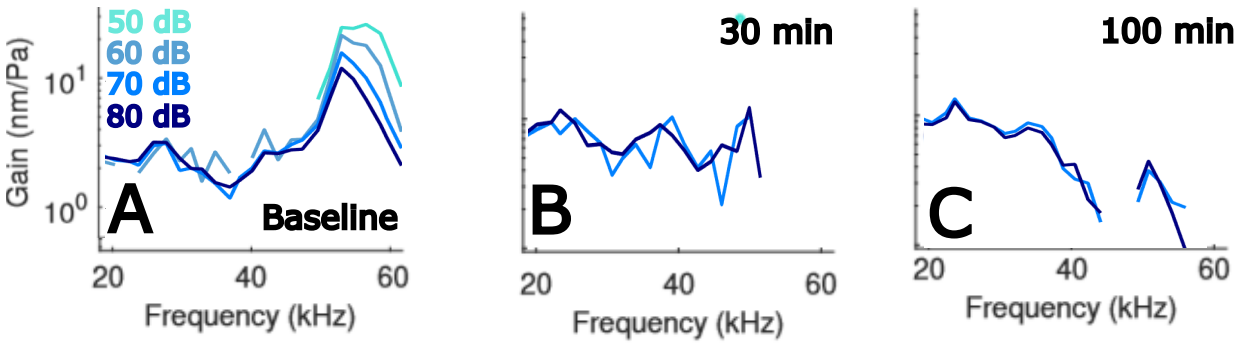

Figure 6.12: Experimental data related to model prediction with basal current nulled or reduced. OCT-measured displacement gain at the BM in the hook region of the gerbil cochlea (Ge971) before and after administration of furosemide. **A** – Baseline measurement made before administration of furosemide, showing the nonlinearity in the BF region characteristic of a healthy cochlea. **B,C** – Displacement gain measured in the same region 30 and 100 minutes after administration of furosemide, respectively. Furosemide has abolished the nonlinearity in the BF region, and this nonlinearity does not recover in any sense over the course of the experiment. This indicates that the base does not recover, as predicted by DPOAE measurements in [41].

recovery fully, which would serve to partially null the basal responses. Based on the FEM results, in which basal nulling allowed local currents to remain dominant at frequencies above the peak, this is the more likely explanation. The explanation is supported by the observation that the recovery of high frequency DPOAEs was typically incomplete following recovery from intravenous furosemide, although OCT-measured displacement responses in the ~25 kHz region do recover [15, 41].

Recent experiments from our group have also supported the damaged-base hypothesis. Dr. C. Elliott Strimbu used OCT to measure displacements in the ~50 kHz region of the gerbil cochlea before and after administration of furosemide. Displacement data from such an experiment are shown in Fig 6.12. These data show that just as high-frequency DPOAEs do not recover after administration of furosemide, high-frequency-region displacement responses also do not recover. This contrasts with OCT-measured displacement responses at somewhat lower frequency regions which did recover [41]. These data support the prediction of our model that basal damage leads to the phase accumulation characteristics of measured CM.

6.9 Conclusions

Through the use of an FEM for the electrical properties of the cochlea, we predicted the shape, size and spatial variation of the CM in order to explore the relationships between BM motion and HB motion, and between HB motion and OHC current. Comparing the model output to experimental CM data, we found that OHC current (and thus HB displacement) must be more sharply tuned than BM responses to produce the measured CM. We also found that transduction sensitivity in the gerbil base is $\sim$ 200 pA/nm, approximately 6 times larger than what *in vitro* experiments have measured.

Finally, in order to predict the measured phase accumulation through several cycles, we found that we needed to reduce the amplitude of the basal current - a reduction by half was reasonably successful. This reduction might be due to the fragile cochlear base, and may not be present in a completely healthy cochlea [15]. This progression of the model's current source is illustrated in Fig. 6.13, and these predicted current source properties are the primary findings of this study.

The present model could be advanced to address additional questions. The experimentally measured shift of the LCM phase relative to BM displacement, occurring at a frequency ~ 0.7 times the CF, was not predicted by the present FEM and is thought to be key to cochlear amplification [88, 90]. Exploring this further with the FEM will require a more complex current source, based on hair bundle motion predicted by micromechanical models and/or emerging experimental data. A prototype of such a study is presented in App E.2.

The present model could also be modified to address round window CM [108]. Of particular interest are the spectral ripple patterns observed in these measurements, where the CM magnitude response exhibits many notches and peaks over a broad spectrum. To adapt our model to address CM measured at the round window, the voltage would be recorded at the semicircular surface representing the base of the cochlea, and a model of the electrical properties of the round window membrane would be implemented. Finally, incorporation of HB current saturation into our model would allow it to probe this key and incompletely characterized aspect of cochlear operation [109].

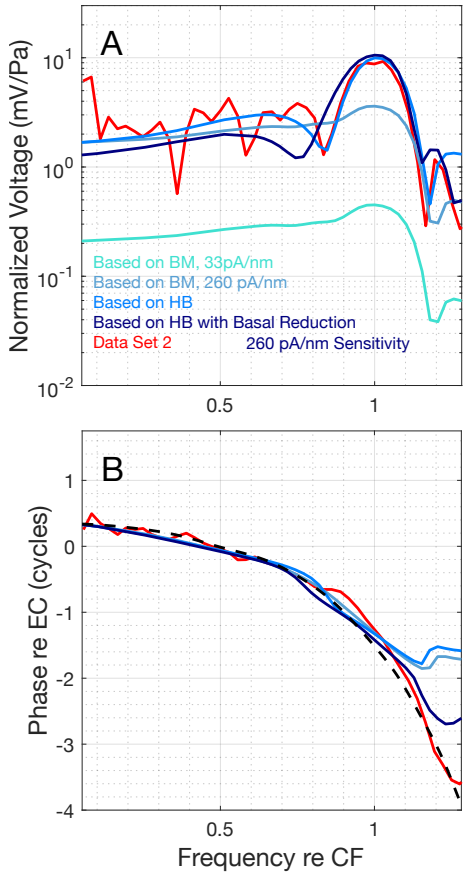

Figure 6.13: Experimental data Set 2 (40 dB SPL) and LCM predictions as the current source representation progressed. The current source was initially represented in the model as proportional to BM displacement and transducer gain estimated from *in vitro* findings (cyan), then with increased transducer sensitivity (light blue), then enhanced tuning (medium blue) and finally reduction of the current sources basal to the measurement location (dark blue), which produced a more realistic phase accumulation. The phase of the current source was not changed and is included in a dashed line. Fig 12 of *Frost and Olson, 2021* [89].

Conclusion

In this thesis I have presented limitations of OCT as a one-dimensional vibrometer and the means by which one may quantify or overcome them. I have shown that skew can lead to easily misinterepreted differences between responses of structures measured in uniaxial scans due to large tonotopic distances. I have also shown that the projection of longitudinal and transverse motion onto the OCT beam axis may be responsible for significant differences between measured displacement responses in different preparations.

The methods I have developed involve no *a priori* knowledge of the beam axis' orientation relative to the sample, and operate on well-justified assumptions such as the local planarity of the BM. Through these methods I have provided means of synthesis between disparate data, registration of volumetric measurements in the cochlea and measurements of two- or three-dimensional motion using OCT.

I have also developed CSVi, a method by which to accelerate such OCT measurements. CSVi could allow for faster acquisition of two- and three-dimensional reconstruction, opening the door for future studies of anatomical components of motion at lower SPLs with higher SNR. It could even be used to achieve reconstructions of anatomical components of motion in perturbation studies.

Lastly, I have presented an FEM of the electrical properties of the cochlea that addresses features of cochlear mechanics that cannot yet be studied using OCT. In particular, it offers insight into the motion at the OHC stereocilia that produce active electromotile responses, likely critical for cochlear amplification.

While this thesis is centered largely on methodology, I have also presented several preliminary physiological findings. For example, the skew compensation discussed in Chapter 3 has revealed that near the BF, OHC-DC motion is in fact 90° out of phase with BM motion within a tonotopic cross-section (along a mostly longitudinal measurement axis), whereas previous measurements may have led one to believe that these structures move in phase [5, 3, 41].

Preliminary two-dimensional reconstruction results show that at high SPL, longitudinal and transverse motion at OHC-DC are nearly 180° out of phase with one another across frequency. This implies a lineal motion pattern, distinguishing this motion from the non-degenerate elliptical fluid motion near the OCC [12, 13]. The reconstructions reveal that the longitudinal component of motion is similar in magnitude to the transverse component. The two are also tuned similarly, and both exhibit sub-BF nonlinearity. This lends credence to certain feedforward models of the cochlea [56, 55].

The FEM has predicted that the base of the gerbil cochlea must be compromised to produce experimentally measured CM responses in the gerbil base. This prediction was, at the time of publication of these results, in line with known vulnerability of the gerbil base [102]. It had also been implied by CM responses in gerbil after the application of furosemide, where DPOAEs did not recover basal to the CM measurement region [15]. This finding has since been further justified by basal OCT measurements after application of furosemide (see Fig 6.12).

The FEM has also predicted a larger MET transduction gain than what is measured *in vitro*. While this prediction has not directly been tested *in vivo*, it is in line with the factors of reduction that may occur in *in vitro* experiments due to variations in temperature, ion concentration and holding potential [103].

Finally, it has predicted sharper tuning at the stereocilia than at the BM. This finding offers a hypothesis for experimentalists to assess, potentially using OCT in preparations where radial reconstruction of differential RL and TM motion is tractable.

These results are preliminary, and the natural next step is for experimentalists to use the methods presented in this thesis to discover robust features of anatomical components of motion. All

three components are likely to be of importance for the healthy functioning of the OCC, so the mystery of cochlear amplification may only be solved by an exploration of three-dimensional motion.

References

[1] J. A. Izatt and M. A. Choma, "Theory of optical coherence tomography," in *Optical Coherence Tomography: Technology and Applications*, W. Drexler and J. G. Fujimoto, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 47–72, ISBN: 978-3-540-77550-8.

[2] M. A. Choma, A. K. Ellerbee, C. Yang, T. L. Creazzo, and J. A. Izatt, "Spectral-domain phase microscopy," *Opt. Lett.*, vol. 30, no. 10, pp. 1162–1164, 2005.

[3] E. Fallah, C. E. Strimbu, and E. S. Olson, "Nonlinearity and amplification in cochlear responses to single and multi-tone stimuli," *Hearing Research*, vol. 377, 271–281, 2019.

[4] F. Chen *et al.*, "A differentially amplified motion in the ear for near-threshold sound detection," *Nature Neuroscience*, vol. 14, no. 6, 770–774, 2011.

[5] C. E. Strimbu, N. C. Lin, and E. S. Olson, "Optical coherence tomography reveals complex motion between the basilar membrane and organ of corti in the gerbil cochlea," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, 1898–1898, 2018.

[6] N. P. Cooper, A. Vavakou, and M. Van Der Heijden, "Vibration hotspots reveal longitudinal funneling of sound-evoked motion in the mammalian cochlea," *Nature Communications*, vol. 9, no. 1, p. 3054, 2018.

[7] 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," *Journal of Neurophysiology*, vol. 120, 2018.

[8] S. S. Gao *et al.*, "Vibration of the organ of corti within the cochlear apex in mice," *Journal of Neurophysiology*, vol. 112, 2014.

[9] G. Von Békésy and E. G. Wever, *Exeperiments in hearing*. McGraw-Hill, 1960.

[10] P. Dallos, A. N. Popper, R. R. Fay, and N. B. Slepecky, "Structure of the mammalian cochlea," in *The Cochlea*. Springer, 1996, 44–129.

[11] Y. Raphael, M. Lenoir, R. Wroblewski, and R. Pujol, "The sensory epithelium and its innervation in the mole rat cochlea," *The Journal of Comparative Neurology*, vol. 314, no. 2, 367–382, 1991.

[12] J. Lighthill, "Energy flow in the cochlea," *Journal of Fluid Mechanics*, vol. 106, pp. 149–213, 1981.

[13] J. Lighthill, *Waves in Fluids*. Cambridge University Press, 1978.

[14] A. J. Duvall and V. T. Rhodes, "Reissner's membrane: An ultrastructural study," *Archives of Otolaryngology - Head and Neck Surgery*, vol. 86, no. 2, 143–151, 1967.

[15] Y. Wang, E. Fallah, and E. S. Olson, "Adaptation of cochlear amplification to low endocochlear potential," *Biophysical Journal*, vol. 116, no. 9, 1769–1786, 2019.

[16] B. A. Bohne and C. D. Carr, "Morphometric analysis of hair cells in the chinchilla cochlea," *The Journal of the Acoustical Society of America*, vol. 77, no. 1, 153–158, 1985.

[17] W. Plassmann, W. Peetz, and M. Schmidt, "The cochlea in gerbilline rodents," *Brain, Behavior and Evolution*, vol. 30, no. 1-2, 82–102, 1987.

[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," *Journal of the Acoustical Society of America*, vol. 151, pp. 1115–1125, 2022.

[19] J. Pickles, S. Comis, and M. Osborne, "Cross-links between stereocilia in the guinea pig organ of corti, and their possible relation to sensory transduction," *Hearing Research*, vol. 15, no. 2, 103–112, 1984.

[20] P. Hakizimana and A. Fridberger, "Inner hair cell stereocilia are embedded in the tectorial membrane," *Nature Communications*, vol. 12, no. 1, 2021.

[21] W. E. Brownell, C. R. Bader, D. Bertrand, and Y. de Ribaupierre, "Evoked mechanical responses of isolated cochlear outer hair cells," *Science*, vol. 227, no. 4683, 194–196, 1985.

[22] J. Zheng, W. Shen, D. Z. He, K. B. Long, L. D. Madison, and P. Dallos, "Prestin is the motor protein of cochlear outer hair cells," *Nature*, vol. 405, no. 6783, 149–155, 2000.

[23] A. Parsa, P. Webster, and F. Kalinec, "Deiters cells tread a narrow path – the deiters cells-basilar membrane junction," *Hearing Research*, vol. 290, pp. 13–20, 2012.

[24] L. C. Peterson and B. P. Bogert, "A dynamical theory of the cochlea," *The Journal of the Acoustical Society of America*, vol. 22, no. 1, 84–84, 1950.

[25] D. D. Greenwood, "A cochlear frequency-position function for several species – 29 years later," *Journal of the Acoustical Society of America*, vol. 87, pp. 2592–2605, 1990.

[26] B. L. Frost, "Foundations of the wkb approximation for models of cochlear mechanics in 1- and 2-d," *In Revisions with The Journal of the Acoustical Society of America, available at https://arxiv.org/abs/2307.14353*, 2023.

[27] M. A. Viergever, *Mechanics of the inner ear: A mathematical approach*. Technische Hogeschool, 1980.

[28] G. Green, "On the motion of waves in a variable canal of small depth and width," *Transactions of the Cambridge Philosophical Society*, vol. 6, pp. 457–462, 1837.

[29] A. Breton, *Nadja*. Librarie Gallimard, 1928.

[30] J. Liouville, "Extrait d'un mémoire sur le développement des fonctions en séries dont les différents termes sont assujettis à satisfaire à une méme équation différentielle linéaire, contenant un paramètre variable," *Journal de Mathématiques Pures et Appliquées*, vol. 2, 16–35, 1837.

[31] G. Zweig, R. Lipes, and J. R. Pierce, "The cochlear compromise," *The Journal of the Acoustical Society of America*, vol. 59, no. 4, 975–982, 1976.

[32] C. R. Steele and L. A. Taber, "Comparison of wkb and finite difference calculations for a two-dimensional cochlear model," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, 1001–1006, 1979.

[33] L. Watts, "The mode-coupling liouville–green approximation for a two-dimensional cochlear model," *The Journal of the Acoustical Society of America*, vol. 108, no. 5, 2266–2271, 2000.

[34] S. J. Elliott, G. Ni, B. R. Mace, and B. Lineton, "A wave finite element analysis of the passive cochlea," *The Journal of the Acoustical Society of America*, vol. 133, no. 3, 1535–1545, 2013.

[35] E. de Boer and M. Viergever, "Validity of the liouville-green (or wkb) method for cochlear mechanics," *Hearing Research*, vol. 8, no. 2, 131–155, 1982.

[36] P. Dallos and M. Cheatham, "Production of cochlear potentials by inner and outer hair cells," *The Journal of the Acoustical Society of America*, vol. 60, pp. 510–512, 1976.

[37] R. Fettiplace, "Hair cell transduction, tuning, and synaptic transmission in the mammalian cochlea," *Comprehensive Physiology*, 1197–1227, 2017.

[38] A. J. Hudspeth and D. P. Corey, "Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli.," *Proceedings of the National Academy of Sciences*, vol. 74, no. 6, 2407–2411, 1977.

[39] L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," *Physiological Reviews*, vol. 81, no. 3, 1305–1352, 2001.

[40] C. E. Strimbu and E. S. Olson, "Salicylate-induced changes in organ of corti vibrations," *Hearing Research*, vol. 423, p. 108 389, 2022.

[41] C. E. Strimbu, Y. Wang, and E. S. Olson, "Manipulation of the endocochlear potential reveals two distinct types of cochlear nonlinearity," *Biophysical Journal*, vol. 119, no. 10, 2087–2101, 2020.

[42] 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," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, p. 37 625, 2021.

[43] R. D. Rabbitt, "Analysis of outer hair cell electromechanics reveals power delivery at the upper-frequency limits of hearing," *Journal of the Royal Society Interface*, vol. 19, no. 191, 2022.

[44] L. Robles, M. A. Ruggero, and N. C. Rich, "Two-tone distortion in the basilar membrane of the cochlea," *Nature*, vol. 349, no. 6308, 413–414, 1991.

[45] P. Bonfils and P. Avan, "Distortion-product otoacoustic emissions: Values for clinical use," *Archives of Otolaryngology - Head and Neck Surgery*, vol. 118, no. 10, 1069–1076, 1992.

[46] C. P. Versteegh and M. van der Heijden, "Basilar membrane responses to tones and tone complexes: Nonlinear effects of stimulus intensity.," *Journal of the Association for Research in Otolaryngology*, vol. 13, pp. 785–798, 6 2012.

[47] C. E. Strimbu, L. A. Chiriboga, B. L. Frost, E. Fallah, and E. S. Olson, "A frame and a hotspot in cochlear mechanics," *bioRxiv*, 2023. eprint: `https://www.biorxiv.org/content/early/2023/06/30/2023.06.29.547111.full.pdf`.

[48] N. H. Cho and S. Puria, "Motion of the cochlear reticular lamina implies that it is not a stiff plate," *Scientific Reports*, vol. 12, no. 1, p. 18 715, 2022.

[49] B. Frost, C. E. Strimbu, and E. S. Olson, "Reconstruction of transverse-longitudinal vibrations in the organ of corti complex via optical coherence tomography," *Journal of the Acoustical Society of America*, vol. 153, pp. 1347–1360, 2023.

[50] A. Altoè, J. B. Dewey, K. K. Charaziak, J. S. Oghalai, and C. A. Shera, "Overturning the mechanisms of cochlear amplification via area deformations of the organ of corti," *The Journal of the Acoustical Society of America*, vol. 152, no. 4, 2227–2239, 2022.

[51] R. Caro, *The power broker: Robert Moses and the Fall of New York*. Vintage Books, 1975.

[52] W. He, D. Kemp, and T. Ren, "Timing of the reticular lamina and basilar membrane vibration in living gerbil cochleae," *eLife*, vol. 7, 2018.

[53] T. Ren, W. He, and D. Kemp, "Reticular lamina and basilar membrane vibrations in living mouse cochleae," *Proceedings of the National Academy of Sciences*, vol. 113, no. 35, 9910–9915, 2016.

[54] S. Meenderink and W. Dong, "Organ of corti vibrations are dominated by longitudinal motion in vivo," 2022.

[55] Y.-J. Yoon, C. R. Steele, and S. Puria, "Feed-forward and feed-backward amplification model from cochlear cytoarchitecture: An interspecies comparison," *Biophysical Journal*, vol. 100, no. 1, pp. 1–10, 2011.

[56] J. J. Guinan Jr., "Cochlear amplification in the short-wave region by outer hair cells changing organ-of-corti area to amplify the fluid traveling wave," *Hearing Research*, vol. 426, p. 108 641, 2022.

[57] N. P. Cooper, "Radial variation in the vibrations of the cochlear partition," in *Recent Developments in Auditory Mechanics*, Singapore: World Scientific, 2000, pp. 109–115.

[58] Y. Gan, W. Yao, K. M. Myers, and C. P. Hendon, "An automated 3d registration method for optical coherence tomography volumes," *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.

[59] B. K. Horn, H. M. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America A*, vol. 5, no. 7, p. 1127, 1988.

[60] R. Benjemaa and F. Schmitt, "A solution for the registration of multiple 3d point sets using unit quaternions," *Lecture Notes in Computer Science*, 34–50, 1998.

[61] O. Faugeras and M. Hebert, "The representation, recognition, and locating of 3-d objects," *The International Journal of Robotics Research*, vol. 5, no. 3, 27–52, 1986.

[62] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, 239–256, 1992.

[63] Y. Zhang, W. Zhang, and J. Li, "Partial-to-partial point cloud registration by rotation invariant features and spatial geometric consistency," *Remote Sensing*, vol. 15, no. 12, p. 3054, 2023.

[64] R. Zhou, X. Li, and W. Jiang, "Scanet: A spatial and channel attention based network for partial-to-partial point cloud registration," *Pattern Recognition Letters*, vol. 151, 120–126, 2021.

[65] W. Kim *et al.*, "Vector of motion measurements in the living cochlea using a 3d oct vibrometry system," *Biomedical Optics Express*, vol. 13, pp. 2542–2553, 2022.

[66] H. Y. Lee *et al.*, "Two-dimensional cochlear micromechanics measured in vivo demonstrate radial tuning within the mouse organ of corti," *Journal of Neuroscience*, vol. 36, no. 31, 8160–8173, 2016.

[67] B. Frost, C. E. Strimbu, and E. S. Olson, "Erratum: Reconstruction of transverse-longitudinal vibrations in the organ of corti complex via optical coherence tomography," *Journal of the Acoustical Society of America*, vol. 153, pp. 2537–2538, 2023.

[68] W. Zhou, T. Jabeen, S. Sabha, J. Becker, and J.-H. Nam, "Deiters cells act as mechanical equalizers for outer hair cells," *The Journal of Neuroscience*, 2022.

[69] E. S. Olson, "Intracochlear pressure measurements related to cochlear tuning," *The Journal of the Acoustical Society of America*, vol. 110, no. 1, 349–367, 2001.

[70] J. P. Kolb *et al.*, "Correction: Live video rate volumetric oct imaging of the retina with multi-mhz a-scan rates," *PLoS One*, vol. 14, no. 7, 2019.

[71] Y Chen, Y.-J. Hong, S. Maxita, and Y. Yasuno, "Three-dimensional eye motion correction by lissajous scan optical coherence tomography," *Biomedical Optics Express*, vol. 8, no. 3, pp. 1783–1802, 2017.

[72] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[73] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013, ISBN: 0817649476.

[74] G. Kutyniok, "Theory and applications of compressed sensing," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 79–101, 2013.

[75] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, 161–319, 2016.

[76] 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.*, vol. 61, no. 1, pp. 103–116, 2009.

[77] J. P. McLean and C. P. Hendon, "3-d compressed sensing optical coherence tomography using predictive coding," *Biomedical Optics Express*, vol. 12, no. 4, pp. 2531–2549, 2021.

[78] E. Lebed, P. J. Mackenzie, M. V. Sarunic, and M. F. Beg, "Rapid volumetric oct image acquisition using compressive sampling," *Opt. Express*, vol. 18, no. 20, pp. 21 003–21 012, 2010.

[79] B. L. Frost, N. Janjušević, C. E. Strimbu, and C. Hendon, "Compressed sensing of displacement signals measured with optical coherence tomography," *Biomedical Optics Express*, vol. 14, no. 11, pp. 5539–5554, 2023.

[80] A. Sriram *et al.*, "End-to-end variational networks for accelerated mri reconstruction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel *et al.*, Eds., Cham: Springer International Publishing, 2020, pp. 64–73, ISBN: 978-3-030-59713-9.

[81] N. Janjušević, A. Khalilian-Gourtani, and Y. Wang, "CDLNet: Noise-adaptive convolutional dictionary learning network for blind denoising and demosaicing," *IEEE Open Journal of Signal Processing*, vol. 3, pp. 196–211, 2022.

[82] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *PNAS*, vol. 100, no. 5, pp. 2197–2202, 2003.

[83] C. Ramirez, V. Kreinovich, and M. Argaez, "Why l1 is a good approximation to l0: A geometric explanation," *Journal of Uncertain Systems*, vol. 7, 2013.

[84] S. G. Mallat, *A wavelet tour of signal processing: The sparse way*. Elsevier, 2009.

[85] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

[86] Y. Lv, "Parameter selection and solution algorithm for tgv-based image restoration model," *SN Applied Sciences*, vol. 1, no. 7, 2019.

[87] N. C. Lin, C. P. Hendon, and E. S. Olson, "Signal competition in optical coherence tomography and its relevance for cochlear vibrometry," *The Journal of the Acoustical Society of America*, vol. 141, no. 1, 395–405, 2017.

[88] A. Nankali, Y. Wang, C. E. Strimbu, E. S. Olson, and K. Grosh, "A role for tectorial membrane mechanics in activating the cochlear amplifier," *Scientific Reports*, vol. 10, p. 17 620, 2020.

[89] B. Frost and E. S. Olson, "Model of cochlear microphonic explores the tuning and magnitude of hair cell transduction current," *Biophysical Journal*, vol. 120, no. 17, 3550–3565, 2021.

[90] W. Dong and E. S. Olson, "Detection of cochlear amplification and its activation," *Biophysical Journal*, vol. 105, no. 4, 1067–1078, 2013.

[91] R. L. Gulley and T. S. Reese, "Intercellular junctions in the reticular lamina of the organ of corti," *Journal of Neurocytology*, vol. 5, no. 5, 617–618, 1976.

[92] F Nin *et al.*, "Computer modeling defines the system driving a constant current crucial for homeostasis in the mammalian cochlea by integrating unique ion transports," *Systems Biology and Applications*, vol. 24, 1–10, 2017.

[93] D. Z. Z. He, S. Jia, and P. Dallos, "Mechanoelectrical transduction of adult outer hair cells studied in a gerbil hemicochlea," *Nature*, vol. 429, 766–770, 2004.

[94] M. Risoud, J. Sircoglou, G. Dedieu, M. Tardivel, C. Vincent, and N.-X. Bonne, "Imaging and cell count in cleared intact cochlea in the mongolian gerbil using laser scanning confocal microscopy," *European Annals of Otorhinolaryngology, Head and Neck Diseases*, vol. 134, no. 4, 221–224, 2017.

[95] G. Zweig, "Finding the impedance of the organ of corti," *The Journal of the Acoustical Society of America*, vol. 89, pp. 1229–1254, 1991.

[96] G. Zweig, "Basilar membrane motion," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 40, 619–633, 1976.

[97] M. Müller, "The cochlear place-frequency map of the adult and developing mongolian gerbil," *Hearing Research*, vol. 94, no. 1-2, 148–156, 1996.

[98] T. Ren, W. He, and E. Porsov, "Localization of the cochlear amplifier in living sensitive ears," *PLoS ONE*, vol. 6, no. 4, e20149:1–8, 2011.

[99] E. S. Olson, "Observing middle and inner ear mechanics with novel intracochlear pressure sensors," *The Journal of the Acoustical Society of America*, vol. 103, 3445 – 3463, 1988.

[100] E. S. Olson, "Fast waves, slow waves and cochlear excitation," *The Journal of the Acoustical Society of America*, vol. 133, p. 050 134, 2013.

[101] C. A. Shera and J. J. Guinan, "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian oaes," *The Journal of the Acoustical Society of America*, vol. 105, no. 2, 782–798, 1999.

[102] E. H. Overstreet, A. N. Temchin, and M. A. Ruggero, "Basilar membrane vibrations near the round window of the gerbil cochlea," *Journal of the Association for Research in Otolaryngology*, vol. 3, no. 3, 351–361, 2002.

[103] H. J. Kennedy, M. G. Evans, A. C. Crawford, and R. Fettiplace, "Fast adaptation of mechanoelectrical transducer channels in mammalian cochlear hair cells," *Nature Neuroscience*, vol. 6, no. 8, 832–836, 2003.

[104] W. Nogueira, D. Schurzig, A. Büchner, R. T. Penninger, and W. Würfel, "Validation of a cochlear implant patient-specific model of the voltage distribution in a clinical setting," *Frontiers in Bioengineering and Biotechnology*, vol. 4, p. 84, 2016.

[105] T. J. Faes, H. A. van der Meij, H. C. de Munck, and R. M. Heethaar, "The electric resistivity of human tissues (100 hz-10 mhz): A meta-analysis of review studies," *Physiological Measurement*, vol. 20, no. 4, R1–10, 1999.

[106] G. A. Misrahy, K. M. Hildreth, E. W. Shinabarger, and W. J. Gannon, "Electrical properties of wall of endolymphatic space of the cochlea (guinea pig)," *American Journal of Physiology-Legacy Content*, vol. 194, no. 2, 396–402, 1958.

[107] P. V. Pierre, J. E. Rasmussen, S. N. Aski, P. Damberg, and G. Laurell, "High-dose furosemide enhances the magnetic resonance signal of systemic gadolinium in the mammalian cochlea," *Otology and Neurotology*, vol. 41, no. 4, 545–553, 2020.

[108] K. K. Charaziak, J. H. Siegel, and C. A. Shera, "Spectral ripples in round-window cochlear microphonics: Evidence for multiple generation mechanisms," *The Journal of the Association for Research in Otolaryngology*, vol. 19, no. 4, 401–419, 2018.

[109] W Zhou and J.-H. Nam, "Probing hair cell's mechanotransduction using two-tone suppression measurements," *Scientific Reports*, vol. 9, p. 4626, 2019.

[110] H. Duifhuis, "Cochlear macromechanics," *Mechanics of Hearing Chapter 6*, 189–211, 1988.

[111] S. L. Johnson, M. Beurg, W. Marcotti, and R. Fettiplace, "Prestin-driven cochlear amplification is not limited by the outer hair cell membrane time constant," *Neuron*, vol. 70, no. 6, 1143–1154, 2011.

[112] C. A. Shera and N. P. Cooper, "Basilar-membrane interference patterns from multiple internal reflection of cochlear traveling waves," *Journal of the Acoustical Society of America*, vol. 133, pp. 2224–2239, 2013.

Appendix A: Experimental Details

A.1 Animal Preparation

Experimental protocols were approved by the Columbia University Institutional Animal Care and Use Committee. Experiments were conducted by Dr. C. Elliott Strimbu and I in an acoustical isolation booth at the Fowler Laboratory at Columbia University's Irving Medical Center, equipped with a floating optics table.

Displacement data presented in this report were acquired *in vivo* from young adult gerbils of either sex. Animals were anesthetized with ketamine (40 mg per kg of body mass) and sodium pentabarbital (PB, 40 mg per kg of body mass) via intraperitoneal (IP) injection. The gerbil's scalp was removed and the head was attached to a two-axis goniometer using dental cement. The animals were tracheotomized to facilitate breathing. The left pinna and the majority of the cartilageneous ear canal were resected.

A narrow opening in the bulla was made by dissecting the tissue and muscle around the left temporal bone and using forceps to chip the bone. The bulla was attached to the goniometer using dental cement.

Throughout the surgery and experiment, body temperature was monitored using a rectal thermometer. A servocontrolled heating blanket was used to maintain a body temperature of 38° C, and additional heating was supplied to the animal's head using a disposable hand warmer.

Throughout the experiment, supplemental PB doses were administered to maintain depth of anesthesia, evaluated by assessing the animal's response to a light toe pinch. Every 6 hours, IP injections of the analgesic buprenorphine (0.2 mg per kg of body mass) were administered.

Two data sets presented in this study involve administration of furosemide (Figs 6.11, Fig 6.12). Furosemide doses (100 mg per kg of body mass) were administered by injection into the left leg's

femoral vein.

Some data presented in this report are borrowed from previous works using different experimental protocols than those described here. We have borrowed cochlear microphonic data from Fallah *et al.*, Nankali *et al.* and Wang *et al.* for comparison to our model presented in Chapter 6 [3, 88, 15]. We have also borrowed displacement data from Ren *et al.* for use as the input to our finite element model, shown in Fig 6.3 [98].

A.2 Data Acquisition

A Tucker-Davis Technologies System (TDT) was used to generate sound stimuli and record pressure waveforms. The system consists of a PA5 Programmable Attenuator, an RP2.1 Enhanced Real-time Processor, an MS2 Monitor Speaker, an RX6 Multifunction Processor and an HB7 Headphone Driver. In most presented experimental data, digital signals were recorded using a 97656.25 Hz sampling frequency. Three data sets presented in this report – displacement phase from Fig 2.6 **A**, data from set $\theta = 3$ used for compressed sensing vibrometry evaluation in Sec 5.2.1, and displacement data of Fig 6.12 – were recorded using a 130208.3333 Hz sampling frequency.

Additional circuits were used to transform the trigger signal for the OCT system from that generated by the TDT to one suited to the OCT system's external trigger input. The details of this circuit are described in Dr. Nathan C. Lin's doctoral dissertation. The circuit was designed and implemented for the lower sampling frequency by Dr. Nathan C. Lin, and a similar circuit was designed and implemented for the higher sampling frequency by Dr. C. Elliott Strimbu.

Stimuli were delivered closed-field to the animal's ear canal using a RadioShack speaker and a plastic tube. The pressure at the ear canal (used as a reference for displacement measurements) was recorded using a Sokolich ultrasonic microphone with a probe tubed stationed 1-2 mm from the ear drum.

Displacement data were collected using a Thorlabs Telesto 320 OCT device equipped with an LSM03 objective lens. The system has a central wavelength of 1300 nm, a nominal axial resolution of 4.2 μm and an imaging depth of 2.6 mm in water. The lens has an effective focal length of 36

mm and a numerical aperture of approximately 0.6, giving a lateral resolution of about 8 μm.

A.3 Software Details

Signal generation, pressure acquisition and triggering of the OCT system were controlled by a MATLAB program. This program interfaced with the TDT through the zBUSmon software. OCT volumes were acquired through the ThorLabs ThorImage software. This software was also used to observe real-time images of the sample to determine measurement location and assess sample drift.

Data acquisition was controlled using a C++ program which made use of the ThorLabs SpectralRadar software development kit (SDK). The SDK possesses functions for acquiring raw photodetector data according to trigger signals generated by the TDT, as well as for processing this data and saving it as a time-series of complex-valued A-Scans (i.e. as M-Scans). These M-Scans were then processed alongside recorded ear canal pressure signals using various MATLAB programs to extract the frequency responses and noise levels of the displacement signals.

Compressed sensing vibrometry algorithms described in Chapter 5 were performed in the Julia programming language. Visualization of displacement maps was performed in MATLAB.

The finite element model described in Chapter 6 was implemented in COMSOL Multiphysics on various machines at Columbia University's Engineering Terrace.

Appendix B: Wentzel-Kramers-Brillouin Method

The curves in Fig 2.1 were generated using a two-dimensional WKB solution for a passive, linear box model of the cochlea [28, 30, 31, 32, 110, 27, 26]. In short, this is an approximate solution to the Laplace equation in a rectangular space with a wave displacement input at one boundary and a spatially varying impedance (representing the OCC) at another.

The modeled transverse velocity of the OCC at longitudinal position x (with $x = 0$ at the most basal position) in response to a stimulus at radian frequency ω is given by

$$\dot{u}(x,t) = \frac{P_{OW} e^{j\omega t - jx\int_0^x k(x,\omega)\,dx}}{Z(x,\omega)} \frac{k_0 h \cosh\left(k(x,\omega)h\right)}{\cosh\left(k(x,\omega)h\right)\tanh\left(k_0 h\right)} \times \times \sqrt{\frac{\tanh\left(k_0 h\right) + k_0 h\,\mathrm{sech}^2(k_0 h)}{\tanh\left(k(x,\omega)h\right) + k(x,\omega)h\,\mathrm{sech}^2(k(x,\omega)h)}}, \tag{B.1}$$

$$k(x,\omega)\tanh\left(k(x,\omega)h\right) = \frac{-2j\omega\rho}{Z(x,\omega)}, \tag{B.2}$$

where $k_0 = k(0,\omega)$. P_{OW} is the input pressure magnitude at the oval window, h is the scala height and ρ is the fluid density. To generate Fig 2.3, I used $P_{OW} = 1$ μPa, $\rho = 1$ kg/m^3 and $h = 1$ mm. The chosen input pressure is irrelevant in the presented curves as responses are normalized.

I simulated the system over a 35 mm longitudinal length with 5000 evenly spaced points, and for 1000 evenly spaced frequencies between 100 Hz and 7 kHz. The wavenumber k varies in space according to the transcendental dispersion relation in Eqn B.2. Z is the mechanical impedance of the OCC, which is lumped together as a point-impedance. It is generally written in spring-mass-damper form as

$$Z(x,\omega) = \frac{S(x)}{j\omega} + R(x) + j\omega M(x), \tag{B.3}$$

where S is stiffness, R is resistance and M is mass. Fig 2.3 was generated using the same parameters as Steele and Taber [32]:

$$R(x) = 2\ \mu\text{Ns/mm}^3,$$
$$M(x) = 1.5\ \text{mg/mm}^2,$$
$$S(x) = 10e^{-0.2x}\ \text{N/mm}^3.$$

Resistance and mass are constant, while stiffness decreases exponentially from base to apex.

The dispersion relation in Eqn B.2 is transcendental and cannot be solved analytically. In fact, at any given x and ω it has infinitely many solutions. Several methods have been developed to solve this equation for the physically relevant value of k, and I have used that presented by Viergever [27]. More details of the WKB solution's derivation and implementation can be found in my tutorial article, *Foundations of the WKB Approximation for Models of Cochlear Mechanics in 1- and 2-D* [26].

Appendix C: Derivations of Algorithms for Compressed Sensing Vibrometry

In Chapter 5, I compare the performance of CSVi using three objective functions – one promoting wavelet-domain sparsity (Eqn 5.8), one promoting sparse gradients (Eqn 5.12) and one promoting sparse Hessians (Eqn 5.14). In this appendix, I will derive the algorithms used to minimize these objective functions through the theory of convex optimization. Certain results will go underived for the sake of brevity – for a self-contained treatment, see [75].

C.1 Convex Optimization

C.1.1 Convex Functions and Conjugates

Consider a continuous functional $f : U \to \mathbb{R}$ where U is an open subset of some real topological vector space X. The function f is *convex* as a function if its epigraph is convex as a set. Conceptually, this means that the graph of the function is shaped like a cup. Mathematically, f is convex if for every $x, y \in U$ and $t \in [0, 1]$ we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

In other words, any straight line drawn from one point on the graph to another lies above the function's graph.

Convex functions are very useful for optimization problems as they have unique global minima without any other local minima. Classical optimization methods tend to involve choosing a convex objective function, and using an iterative algorithm to find the global minimum of this function [75].

In Sec 5.1.4, I noted that the ℓ_1 norm is commonly used as a regularizer for sparsity promotion,

as opposed to the ℓ_0 norm. It should now be clear that the ℓ_1 norm is convex while the ℓ_0 norm is not. The ℓ_2 norm is also convex, and all of the objective functions of Eqns 5.8, 5.12 and 5.14 can be seen to be convex.

A useful computational tool is the Legendre-Fenchel conjugate, or *convex conjugate*. For any function $f : U \rightarrow \mathbb{R}$, the convex conjugate f^* is given by

$$f^*(z) = \sup_{x \in U} \{< z, x > -f(x)\}. \tag{C.1}$$

This means $f^* : X^* \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$, where X^* is the dual space of X (i.e. X^* is the space of linear functionals from X to $\mathbb{R}$). A useful property is that if f is convex and continuous, the *biconjugate* $f^{**} = f$.

Another useful property is scaling – if $\lambda \in \mathbb{R}$, then it can be seen from Eqn C.1 that

$$(\lambda f)^*(z) = f^*(z/\lambda). \tag{C.2}$$

C.1.2 Primal-Dual Saddle Point Problems

The primal-dual saddle point method is a way to rewrite a *minimization* problem on X as a *saddle point* problem on X and X^* [75]. Consider an optimization problem

$$\min_{x \in X} \{f(Ax) + g(x)\},$$

where $f : P \rightarrow \mathbb{R}$ and $g : X \rightarrow \mathbb{R}$ are both continuous convex functionals, and $A : X \rightarrow P$ is a bounded linear operator (where P is also a real normed vector space). Consider f^{**}:

$$f^{**}(Ax) = \sup_{p \in P} \{< p, Ax > -f^*(p)\},$$

by the definition of the convex conjugate. As f is continuous and convex, $f^{**} = f$. I can write the original optimization problem as

$$\min_{x \in X} \sup_{p \in P} \{< p, Ax > -f^*(p) + g(x)\}. \tag{C.3}$$

This is a saddle point problem, minimizing a primal variable x and maximizing a dual variable p. As g is p-independent, the maximization is occurring for function $< p, Ax > -f^*(p)$, which is actually *concave* (i.e. its negative is convex). Our algorithms for TV and TGV optimization use this saddle point formulation.

C.1.3 Dual Norms

Norms on vector space X can be associated with norms on the dual space X^*, referred to as *dual norms*. These objects are useful in writing the convex conjugates of norms, required for primal-dual splitting. If $||\cdot||$ is a norm on a real vector space X, its dual norm $||\cdot||^\delta$ is defined on the dual space X^* as

$$||z||^\delta = \sup_{x \in X,\ ||x|| \leq 1} \{| < z, x > |\}. \tag{C.4}$$

Given Eqns 5.8, 5.12 and 5.14, I will need the dual norms of the ℓ_1 and ℓ_2 norms to compute convex conjugates for my objective functions. It can be shown from the definition above [75] that

$$||z||_1^\delta = ||z||_\infty = \max_{m=1,2,...,N} |z_m|, \tag{C.5}$$

$$||z||_2^\delta = ||z||_2. \tag{C.6}$$

Also necessary are the dual norms of the $\ell_{2,1}$ norms of Eqns 5.12 and 5.14. As these norms are pixel-wise ℓ_2 norms followed by ℓ_1 norms, we have

$$||v||_{2,1}^\delta = ||v||_{2,\infty}, \tag{C.7}$$

or the ℓ_∞ norm of the pixel-wise ℓ_2 norm for v in either of $\mathbb{R}^{M\times N\times 2}$ or $\mathbb{R}^{M\times N\times 2\times 2}$.

We will now use these dual norms to represent the convex conjugates of norms. To begin, we define the *indicator function* – if C is a closed, nonempty convex set, then the indicator function on C is

$$i_C(x) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases} \tag{C.8}$$

For real $\lambda > 0$, the *λ-ball* for a norm $||\cdot||$ is defined as the set containing all vectors with norm less than or equal to λ:

$$\mathcal{B}_{\lambda||\cdot||} = \{x \in X \quad \text{s.t. } ||x|| \leq \lambda\}. \tag{C.9}$$

It can be shown that the convex conjugate of a norm is the indicator function for the unit ball of its dual norm [75]:

$$||\cdot||^* = i_{\mathcal{B}_{||\cdot||^\delta}}. \tag{C.10}$$

C.1.4 Proximal Gradient Descent and Ascent

To minimize the objective functions in question, we will use methods based on the *proximal operator*. For a functional $f : U \to \mathbb{R}$, the proximal operator with real parameter $\lambda > 0$ is a function from U to U defined as

$$\mathbf{prox}_{\lambda f}(v) = \underset{x\in U}{\operatorname{argmin}} \left(\frac{1}{2\lambda}||x - v||_2^2 + f(x) \right). \tag{C.11}$$

The proximal operator can be thought of as a generalization of projection onto convex sets (POCS). To see this, consider the indicator function of a non-empty, closed convex set $C \subset U$ (see Eqn C.8). The proximal operator of this indicator function is

$$\begin{aligned} \mathbf{prox}_{\lambda i_C}(v) &= \underset{x\in U}{\operatorname{argmin}} \begin{cases} \frac{1}{2\lambda}||x - v||_2^2, & x \in C \\ \infty, & x \notin C \end{cases} \\ &= \underset{x\in C}{\operatorname{argmin}} ||x - v||_2^2. \end{aligned} \tag{C.12}$$

This is precisely the projection of v onto C. POCS plays a major role in the development of our algorithms.

A common optimization method for differentiable functions is gradient descent, where convergence is achieved by "stepping down" towards the minimum of the objective function. As the ℓ_1 norm is not differentiable, our algorithms will instead converge to the minimum through repeated application of the proximal operator.

This is justified by the properties of contractive mappings. A function $g : U \to U$ is *Lipschitz continuous* if there exists real $L > 0$ such that

$$||g(x) - g(y)||_2 \leq L||x - y||_2, \tag{C.13}$$

where L is called the *Lipschitz constant*. If g is Lipschitz continuous with $L < 1$, it is called a *contractive mapping*, as application of g shrinks the ℓ_2 distance between points.

Consider a contractive mapping g and a point $x^{(0)} \in U$, and define a sequence

$$\{x^{(k)}\}_{k=0,1,\ldots}, \quad x^{(k+1)} = g(x^{(k)}). \tag{C.14}$$

It is guaranteed that the limit of this sequence exists, and is the unique fixed point of g. This is called the Banach fixed point theorem.

Under weak assumptions, the proximal operator can be shown to be contractive [75]. This allows development of a fixed point algorithm to determine the minimizer of a convex function. Consider the optimization problem

$$\underset{x \in U}{\operatorname{argmin}}\, f(x) + g(x),$$

where f and g are convex functionals on U, f is smooth with Lipschitz-continuous gradient, and

g may be non-smooth. For initial value $x^{(0)} \in U$, the *proximal gradient descent* step is

$$x^{(k+1)} = \mathbf{prox}_{\eta g}(x^{(k)} - \eta \nabla f(x^{(k)})), \quad k = 0, 1, \ldots. \tag{C.15}$$

This resembles repeated application of the proximal operator of g to the standard gradient descent step of f with fixed step size η. The step size satisfies $\eta \leq 1/L$, where L is the Lipschitz constant of ∇f. It can be shown that the minimizer of $f + g$ is a fixed point of this operation[1], and the iterative algorithm in Eqn C.15 is guaranteed to converge to $\hat{x}$.

It should be noted that this method can easily be adapted to handle *concave* optimization. To do so, assuming f is concave, the gradient descent step within the proximal operator is simply replaced with a gradient *ascent* step

$$y^{(k+1)} = \mathbf{prox}_{\eta g}(y^{(k)} + \eta \nabla f(y^{(k)})), \quad k = 0, 1, \ldots. \tag{C.16}$$

Together, this facilitates the Chambolle-Pock algorithm, or a proximal gradient algorithm for primal-dual splitting [75]. This involves tackling the saddle point problem of Eqn C.3 by at each k performing a descent in x holding p constant, then an ascent in p holding x constant.

The *primal step* for x in Eqn C.3 can be written in the form of Eqn C.15 by recognizing that $\nabla f^*(p) = 0$ for constant p, and using the known gradient of the inner product:

$$x^{(k+1)} = \mathbf{prox}_{\sigma g}(x^{(k)} - \sigma A^* p^{(k)}). \tag{C.17}$$

The *dual step* for p is similarly found in the form of Eqn C.16 as

$$p^{(k+1)} = \mathbf{prox}_{\tau f^*}(p^{(k)} + \tau A x^{(k+1)}). \tag{C.18}$$

[1]The proof of this relies on the fact that the proximal operator is the resolvent of the subdifferential, a generalization of the gradient for non-differentiable convex functionals [75].

C.2 Algorithms

Using the methods of convex optimization developed above, we can derive the algorithms used for compressed sensing vibrometry. I will cover the algorithms for optimizing the objective functions promoting wavelet-domain sparsity, TV and TGV regularization in turn. In these problems, I will consider $U = X$ to be the whole vector space. I will also assume that maximum or minimum values are achieved on the set, so that supremum and infimum operations may be replaced by maximum and minimum operations.

C.2.1 Handling Complex Data

The theoretical development of convex optimization presented above concerned functions with domains in a real vector space. However, our objective functions are defined over complex vector spaces. This must be accounted for in the development of our optimization methods.

A vector $\mathbf{x} \in \mathbb{C}^N$ can be written as $\mathbf{x}_1 + j\mathbf{x}_2$ where $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$. The optimization over a complex vector space can then be seen as the joint optimization of two vectors over a real vector space. To avoid the explosion of notation, I will continue to write in terms of complex arithmetic as an efficient way to represent this joint optimization. The distinction is mostly important when writing inner products, which vary between real and complex vector spaces.

For example, consider also a vector $\mathbf{c} = \mathbf{c}_1 + j\mathbf{c}_2$, $\mathbf{c}_1, \mathbf{c_2} \in \mathbb{R}^N$. A joint optimization of $< \mathbf{x}, \mathbf{c} >$ for $\mathbf{x}$ would resemble

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

In terms of complex arithmetic, I can write this as

$$\min_{\mathbf{x} \in \mathbb{C}^N} \mathrm{Re}\{\mathbf{c}^H \mathbf{x}\},$$

where $\cdot^H$ denotes the conjugate transpose, and $\mathrm{Re}\{\cdot\}$ denotes the real part.

C.2.2 ISTA: Wavelet-Domain Sparsity

Assuming sparsity in a wavelet domain with wavelet transform $\mathbf{\Psi}$, we find a minimum for F in Eqn 5.8 using the iterative shrinkage thresholding algorithm (ISTA) [85]. This is related to a proximal gradient method (Eqn C.15) with $f(\mathbf{z}) = \frac{1}{2}||\mathbf{M\Psi^{-1}z} - \mathbf{y}||_2^2$ and $g(\mathbf{z}) = \lambda||\mathcal{D}\mathbf{z}||_1$. It can be shown that the proximal operator for the ℓ_1 norm is the *soft-thresholding operation*, $\mathcal{S}$, defined by

$$\left[\mathbf{prox}_{\eta\lambda||\cdot||_1}(\mathbf{z})\right]_n = \mathcal{S}_{\eta\lambda}(\mathbf{z}_n) = \text{sign}(\mathbf{z}_n)\max\{|\mathbf{z}_n| - \eta\lambda, 0\}. \tag{C.19}$$

It is important to note that in the complex case,

$$\text{sign}(\mathbf{z}_n) = e^{j\angle\mathbf{z}_n}.$$

The proximal operator of g is defined in terms of the soft-thresholding operator using a vector parameter $\mathbf{\Lambda}$, which is 0 at the indices of approximation coefficients and λ at the indices of detail coefficients. That is,

$$\mathbf{\Lambda}_n = \lambda\frac{|(\mathcal{D}\mathbf{z})_n|}{|\mathbf{z}_n|},\ n = 1, 2, \ldots, N.$$

Then we can write the proximal operator of g as

$$\left[\mathbf{prox}_{\eta\lambda||\mathcal{D}\cdot||_1}(\mathbf{z})\right]_n = \left[\mathcal{S}_{\eta\mathbf{\Lambda}}(\mathbf{z})\right]_n = \mathcal{S}_{\eta\mathbf{\Lambda}_n}(\mathbf{z}_n),\ n = 1, 2, \ldots, N. \tag{C.20}$$

To find the proximal gradient descent step for ISTA, we also need ∇f, which is

$$\nabla f(\mathbf{z}) = (\mathbf{M\Psi^{-1}})^*(\mathbf{M\Psi^{-1}z} - \mathbf{y}).$$

where $\cdot^*$ denotes the adjoint. The wavelet transform is orthogonal, so $(\mathbf{\Psi^{-1}})^* = \mathbf{\Psi}$. $\mathbf{M}$ is real and diagonal, so $\mathbf{M}^* = \mathbf{M}$. Moreover, double-application of an identical mask is equivalent to applying the mask once: $\mathbf{MM} = \mathbf{M}$. Lastly, $\mathbf{My} = \mathbf{y}$, as $\mathbf{y}$ is already subsampled. The gradient can be written

more simply as

$$\nabla f(\mathbf{z}) = \mathbf{\Psi M \Psi^{-1} z} - \mathbf{\Psi y}. \tag{C.21}$$

Plugging in to Eqn C.15, we can write the step for the ISTA algorithm as

$$\mathbf{z}^{(k+1)} = \mathcal{S}_{\eta\mathbf{\Lambda}}\left[\mathbf{z}^{(k)} - \eta\left(\mathbf{\Psi M \Psi^{-1}} \mathbf{z}^{(k)} - \mathbf{\Psi y}\right)\right], \quad k = 0, 1, \ldots. \tag{C.22}$$

This step is repeated until the algorithm converges according to some criterion. As $\mathbf{z}$ was the wavelet-domain variable, the solution in the original domain is then $\mathbf{\Psi^{-1} z}$. This is displayed in the pseudocode in Alg 1 (here $\eta = 2$ is used as it can be shown to be the optimal step size for this problem [85]).

Algorithm 1: Iterative Shrinkage Thresholding Algorithm (ISTA)

Input: masked observation $\mathbf{y}$, sensing map $\mathbf{M}$, trade-off parameter vector $\mathbf{\Lambda}$;
Let: $\mathbf{z}^{(0)} = \mathbf{0}$;
for $k = 0, 1, \ldots, \infty$ **do**
    $\mathbf{z}^{(k+1)} := \mathcal{S}_{2\mathbf{\Lambda}}(\mathbf{z}^{(k)} - 2\mathbf{\Psi M}(\mathbf{\Psi^{-1}} \mathbf{z}^{(k)} - \mathbf{y}))$;
Output: $\hat{\mathbf{x}} = \mathbf{\Psi^{-1}} \mathbf{z}^{(\infty)}$;

C.2.3 Total Variation

We now consider the TV objective function of Eqn 5.12, with $X = \mathbb{R}^{M \times N}$ and $P = \mathbb{R}^{M \times N \times 2}$. We will perform primal-dual splitting to formulate a saddle point problem (Sec C.1.2) using $f(\mathbf{x}) = \lambda ||\mathbf{x}||_{2,1}$, $A = \mathbf{D}$, and $g(\mathbf{x}) = \frac{1}{2}||\mathbf{Mx} - \mathbf{y}||_2^2$. From the scaling property of the convex conjugate (Eqn C.2) and the conjugate of the norm (Sec C.1.3), we have

$$f^*(\mathbf{p}) = i_{||\mathbf{p}||_{2,\infty} \leq \lambda}(\mathbf{p}), \tag{C.23}$$

the indicator function on the $\ell_{2,\infty}$ ball of radius λ. The saddle point problem is then

$$\min_{\mathbf{x} \in X} \max_{\mathbf{p} \in P} \mathrm{Re}\{\mathbf{p}^* \mathbf{D x}\} + \frac{1}{2}||\mathbf{Mx} - \mathbf{y}||_2^2 - i_{||\mathbf{p}||_{2,\infty} \leq \lambda}(\mathbf{p}).$$

The primal step is found by substitution into Eqn C.17:

$$\mathbf{x}^{(k+1)} = \underset{\mathbf{x}\in X}{\operatorname{argmin}} \frac{1}{2\sigma}||\mathbf{x} - \mathbf{x}^{(k)} + \sigma\mathbf{D}^*\mathbf{p}^{(k)}||_2^2 + \frac{1}{2}||\mathbf{M}\mathbf{x} - \mathbf{y}||_2^2, \tag{C.24}$$

where σ is the primal step size. As this function is differentiable, the optimality condition is simply where the $\mathbf{x}$-derivative is 0:

$$\frac{1}{\sigma}(\mathbf{x} - \mathbf{x}^{(k)} + \sigma\mathbf{D}^*\mathbf{p}^{(k)}) + \mathbf{M}^*(\mathbf{M}\mathbf{x} - \mathbf{y}) = 0, \tag{C.25}$$

As in ISTA, $\mathbf{M}$ is diagonal so $\mathbf{M}^* = \mathbf{M}$, and as $\mathbf{y}$ is already subsampled we have $\mathbf{M}^*(\mathbf{M}\mathbf{x} - \mathbf{y}) = \mathbf{M}\mathbf{x} - \mathbf{y}$. Solving for $\mathbf{x}$ in Eqn C.25, we get the primal step

$$\mathbf{x}^{(k+1)} = (I + \sigma\mathbf{M})^{-1}(\mathbf{x}^{(k)} + \sigma(\mathbf{y} - \mathbf{D}^*\mathbf{p}^{(k)})). \tag{C.26}$$

The matrix $I + \sigma\mathbf{M}$ is diagonal, so multiplication by its inverse is equivalent to element-wise division by the diagonal. I will abuse notation and write this as an element-wise quotient:

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{x}^{(k)} + \sigma(\mathbf{y} - \mathbf{D}^*\mathbf{p}^{(k)})}{\mathbf{1} + \sigma\mathbf{m}}. \tag{C.27}$$

The dual step can be found by plugging into Eqn C.18:

$$\mathbf{p}^{(k+1)} = \underset{\mathbf{p}\in P}{\operatorname{argmin}} \frac{1}{2\tau}||\mathbf{p} - \mathbf{p}^{(k)} - \tau\mathbf{D}\mathbf{x}^{(k+1)}||_2^2 + i_{||\mathbf{p}||_{2,\infty}\leq\lambda}(\mathbf{p}). \tag{C.28}$$

Outside of the λ-ball, the indicator evaluates to ∞ – that means that the minimizer must lie inside of the λ-ball, where the indicator evaluates to 0. This simplifies the proximal operator to

$$\mathbf{p}^{(k+1)} = \underset{\mathbf{p}\in P}{\operatorname{argmin}} \frac{1}{2\tau}||\mathbf{p} - \mathbf{p}^{(k)} - \tau\mathbf{D}\mathbf{x}^{(k+1)}||_2^2, \quad ||\mathbf{p}||_{2,\infty} \leq \lambda. \tag{C.29}$$

This is solved by a projection of $\mathbf{p}^{(k)} + \tau\mathbf{D}\mathbf{x}^{(k+1)}$ onto the λ-ball. This can be represented by the

element-wise quotient

$$\mathbf{p}^{(k+1)} = \frac{\mathbf{p}^{(k)} + \tau \mathbf{D}\mathbf{x}^{(k+1)}}{\max\{1, |\mathbf{p}^{(k)} + \tau \mathbf{D}\mathbf{x}^{(k+1)}|/\lambda\}}, \tag{C.30}$$

where $|\cdot|$ denotes pixel-wise 2-norm $(|\mathbf{w}|)_{m,n,s} = \sqrt{(\mathbf{w}_{m,n,1})^2 + (\mathbf{w}_{m,n,2})^2}$.

These two updates allow us to write the full primal-dual TV algorithm (using $\tau = \sigma = 1/\sqrt{8}$ as it can be shown to be optimal [75]). This is displayed in pseudocode in Alg 2.

Algorithm 2: Primal-Dual Splitting for TV

Input: masked observation $\mathbf{y}$, mask $\mathbf{m}$, trade-off parameter λ ;
Let: $\mathbf{p}^{(0)} = \mathbf{0}$, $\mathbf{x}^{(0)} = \mathbf{y}$, $\tau = \sigma = \frac{1}{\sqrt{8}}$;
for $k = 0, 1, \ldots, \infty$ **do**
  $\mathbf{x}^{(k+1)} := \frac{\mathbf{x}^{(k)} + \sigma(\mathbf{y} - \mathbf{D}^*\mathbf{p}^{(k)})}{\mathbf{1} + \sigma\mathbf{m}}$;
  $\mathbf{p}^{(k+1)} := \frac{\mathbf{p}^{(k)} + \tau\mathbf{D}\mathbf{x}^{(k+1)}}{\max\{1, |\mathbf{p}^{(k)} + \tau\mathbf{D}\mathbf{x}^{(k+1)}|/\lambda\}}$;
Output: $\hat{\mathbf{x}} = \mathbf{x}^{(\infty)}$;

To elegantly write $\mathbf{D}^*$, we need to define the left-handed fundamental finite difference operators [86]:

$$\begin{aligned}
(\partial_1^- u)_{i,j} &= \begin{cases} \mathbf{p}_{i,j} - \mathbf{p}_{i-1,j}, & 1 < i < M \\ \mathbf{p}_{i,j}, & i = 1 \\ -\mathbf{p}_{i-1,j}, & i = M \end{cases} \\
(\partial_2^- \mathbf{p})_{i,j} &= \begin{cases} \mathbf{p}_{i,j} - \mathbf{p}_{i,j-1}, & 1 < j < N \\ \mathbf{p}_{i,j}, & j = 1 \\ -\mathbf{p}_{i,j-1}, & j = N \end{cases}
\end{aligned} \tag{C.31}$$

In terms of these operators,

$$D^*\mathbf{p} = -\partial_1^- \mathbf{p}_1 - \partial_2^- \mathbf{p}_2. \tag{C.32}$$

C.2.4 Total Generalized Variation

Many of the derivations for the TV primal-dual splitting algorithm extend to the minimization of the TGV objective function in Eqn 5.14. However, because there are two primal variables – $\mathbf{x}$ and $\mathbf{v}$ – we need to introduce two dual variables – $\mathbf{p}$ and $\mathbf{q}$. This can be performed in the same

manner by forming the saddle problem in $\mathbf{x}$ holding $\mathbf{v}$ constant, then in $\mathbf{v}$ holding $\mathbf{x}$ constant. The resultant optimization problem is

$$\min_{\mathbf{x},\mathbf{v}} \max_{\mathbf{p},\mathbf{q}} \mathrm{Re}\{\mathbf{p}^*(\mathbf{D}\mathbf{x}-\mathbf{v})\} + \mathrm{Re}\{\mathbf{q}^*\mathbf{K}\mathbf{v}\} + \frac{1}{2}||\mathbf{M}\mathbf{x}-\mathbf{y}||_2^2 - i_{||\mathbf{p}||_{2,\infty}\leq\lambda_1}(\mathbf{p}) - i_{||\mathbf{q}||_{2,\infty}\leq\lambda_0}(\mathbf{q}). \quad \text{(C.33)}$$

Here, $\mathbf{x} \in \mathbb{C}^{N_1\times N_2}$, $\mathbf{v}, \mathbf{p} \in \mathbb{C}^{N_1\times N_2\times 2}$, and $\mathbf{q} \in \mathbb{C}^{N_1\times N_2\times 2\times 2}$.

The primal step for $\mathbf{x}$, holding $\mathbf{v}$, $\mathbf{p}$ and $\mathbf{q}$ constant, is identical to the primal step from TV:

$$\mathbf{x}^{(k+1)} = (I+\sigma\mathbf{M})^{-1}(\mathbf{x}^{(k)} + \sigma(\mathbf{y} - \mathbf{D}^*\mathbf{p}^{(k)})). \quad \text{(C.34)}$$

As for $\mathbf{v}$, the descent step is

$$\mathbf{v}^{(k+1)} = \operatorname*{argmin}_{\mathbf{v}} \frac{1}{2\sigma}||\mathbf{v}-\mathbf{v}^{(k)}-\sigma\mathbf{p}^{(k)}||_2^2 + \mathrm{Re}\{(\mathbf{q}^{(k)})^*\mathbf{K}\mathbf{v}\}. \quad \text{(C.35)}$$

As the function is differentiable, the optimality condition is where the $\mathbf{v}$-derivative is 0:

$$\frac{1}{\sigma}(\mathbf{v}-\mathbf{v}^{(k)}-\sigma\mathbf{p}^{(k)}) + \mathbf{K}^*\mathbf{q}^{(k)} = 0.$$

Solving for $\mathbf{v}$, we get the second primal step:

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \sigma(\mathbf{p}^{(k)} - \mathbf{K}^*\mathbf{q}^{(k)}). \quad \text{(C.36)}$$

Next, the dual step for $\mathbf{p}$ is

$$\mathbf{p}^{(k+1)} = \operatorname*{argmin}_{\mathbf{p}} \frac{1}{2\tau}||\mathbf{p}-\mathbf{p}^{(k)}-\tau(\mathbf{D}\mathbf{x}^{(k+1)}-\mathbf{v}^{(k+1)})||_2^2 + i_{||\mathbf{p}||_{2,\infty}\leq\lambda_1}(\mathbf{p}). \quad \text{(C.37)}$$

This is nearly identical to the TV case but the $\mathbf{D}\mathbf{x}$ term is replaced by $\mathbf{D}\mathbf{x}-\mathbf{v}$. Its solution is

similarly a projection onto the λ-ball:

$$\mathbf{p}^{(k+1)} = \frac{\mathbf{p}^{(k)} + \tau(\mathbf{D}\mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)})}{\max\{1, |\mathbf{p}^{(k)} + \tau(\mathbf{D}\mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)})|/\lambda_1\}}. \tag{C.38}$$

Finally, the dual step for $\mathbf{q}$ is

$$\mathbf{q}^{(k+1)} = \underset{\mathbf{q}}{\operatorname{argmin}} \frac{1}{2\tau}||\mathbf{q} - \mathbf{q}^{(k)} - \tau\mathbf{K}\mathbf{v}^{(k+1)}||_2^2 + i_{||\mathbf{q}||_{2,\infty} \leq \lambda_0}(\mathbf{p}). \tag{C.39}$$

This is characteristically similar to Eqn C.29, replacing $\mathbf{p}$ with $\mathbf{q}$, $\mathbf{D}$ with $\mathbf{K}$ and λ with λ_0. Direct substitution gives the final step:

$$\mathbf{q}^{(k+1)} = \frac{\mathbf{q}^{(k)} + \tau\mathbf{K}\mathbf{v}^{(k+1)}}{\max\{1, |\mathbf{q}^{(k)} + \tau\mathbf{K}\mathbf{v}^{(k+1)}|/\lambda_0\}}. \tag{C.40}$$

where again $|\cdot|$ is a pixel-wise 2-norm.

These four steps allow us to write the full primal-dual TGV algorithm (with $\sigma = \tau = 1/\sqrt{12}$ which can be shown to be optimal [75, 86]). This is displayed in pseudocode in Alg 3.

Algorithm 3: Primal-Dual Splitting for TGV

Input: masked observation $\mathbf{y}$, mask $\mathbf{m}$, trade-off parameters $\lambda_1,\ \lambda_0$;
τ **Let:** $\mathbf{q}^{(0)} = \mathbf{v}^{(0)} = \mathbf{0}$, $\mathbf{p}^{(0)} = \mathbf{0}$, $\mathbf{x}^{(0)} = \mathbf{y}$, $\tau = \sigma = \frac{1}{\sqrt{12}}$;
for $k = 0, 1, \ldots, \infty$ **do**
    $\mathbf{x}^{(k+1)} := \frac{x^{(k)} + \sigma(\mathbf{y} - \mathbf{D}^*\mathbf{p}^{(k)})}{\mathbf{1}+\sigma\mathbf{m}}$;
    $\mathbf{v}^{(k+1)} := \mathbf{v}^{(k)} - \sigma\left(\mathbf{K}^*\mathbf{q}^{(k)} - \mathbf{p}^{(k)}\right)$;
    $\mathbf{p}^{(k+1)} := \frac{\mathbf{p}^{(k)} + \tau(\mathbf{D}\mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)})}{\max\{1, |\mathbf{p}^{(k)} + \tau(\mathbf{D}\mathbf{x}^{(k+1)} - \mathbf{v}^{(k+1)})|/\lambda_1\}}.$;
    $\mathbf{q}^{(k+1)} := \frac{\mathbf{q}^{(k)} + \tau\mathbf{K}\mathbf{v}^{(k+1)}}{\max\{1, |\mathbf{q}^{(k)} + \tau\mathbf{K}\mathbf{v}^{(k+1)}|/\lambda_0\}}.$;
Output: $\hat{\mathbf{x}} = \mathbf{x}^{(\infty)}$;

The adjoint of the second finite difference operator, $\mathbf{K}^*$, can be written elegantly using the fundamental finite difference operators [86] as

$$\mathbf{K}^*\mathbf{q} = -\begin{pmatrix} \partial_1^-\mathbf{q}_{1,1} + \partial_2^-\mathbf{q}_{1,2} \\ \partial_1^-\mathbf{q}_{2,1} + \partial_2^-\mathbf{q}_{2,2} \end{pmatrix}. \tag{C.41}$$

Appendix D: Wavelet-Domain Sparsity of Areal Displacement Maps

The results presented for CSVi (Chapter 5) using ISTA used a three-level Daubechies-7 (db-7) wavelet transform. We chose this wavelet basis by assessing the sparsity of our signals in several wavelet bases. We then determined the number of levels to use by comparing NMSE results for reconstruction using two-, three- and four-level wavelet transforms.

D.1 Sparsity in Wavelet Domains

To assess sparsity, we considered the reconstruction of the original areal motion map after hard-thresholding the smallest wavelet coefficients. In particular, we considered the keep probability of coefficients necessary to achieve an NMSE of less than 1% between the original map and the map reconstructed from the thresholded signal.

Figure D.1 shows this keep probability for Daubechies 1-20 wavelet transforms. Compression rates of 42 maps are shown, taken from two angles, seven stimulus frequencies and three SPL.

This shows that across viewing angle, frequency and amplitude the signal is sparse in Daubechies 4-20 wavelet domains. Motion maps can be represented faithfully for all maps tested using less than 5% of wavelet coefficients in each of these domains. This makes a strong case for global sparsity in motion within the cochlea in these wavelet domains, including db-7.

D.2 Difference between Levels

We chose to use a three-level wavelet transform, motivated by the NMSE results of comparing two-, three- and four-level wavelet transforms on the test set ($N = 20$). Fig D.2 shows that at $P = 10$, the three-level wavelet transform performs best of the tested methods.

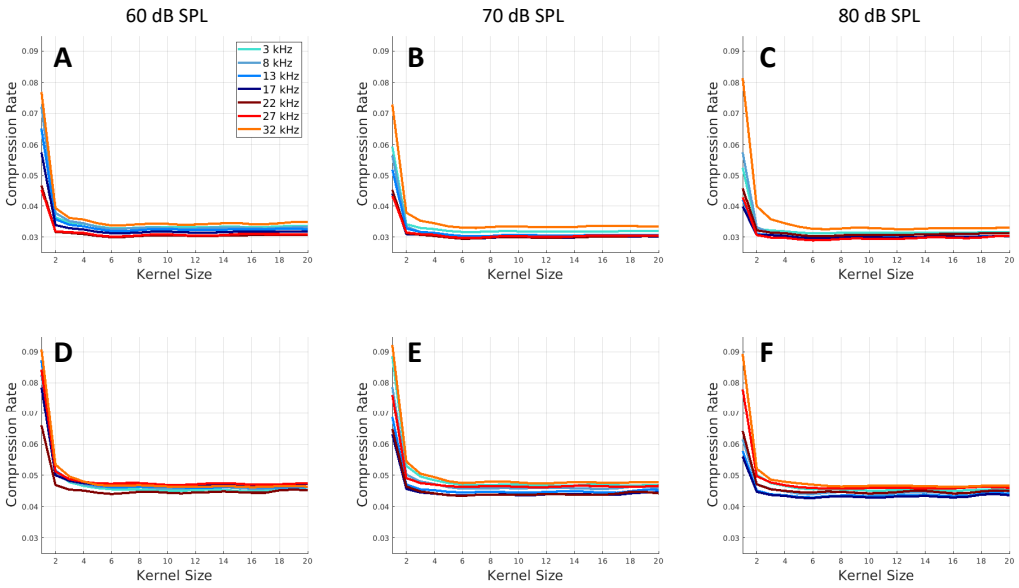

Figure D.1: Compression rates of *in vivo* areal cochlear motion patterns at three stimulus levels, seven stimulus frequencies and two viewing angles. Curves show the proportion of wavelet coefficients needed to achieve 1% NMSE from the original motion map using Daubechies 1-20 wavelet transforms. **A-C**: Compression rates for $\theta = 1$ in response to 60, 70 and 80 dB SPL stimuli, respectively . **D-F**: Same as **A-C** but for motion maps acquired at $\theta = 2$.

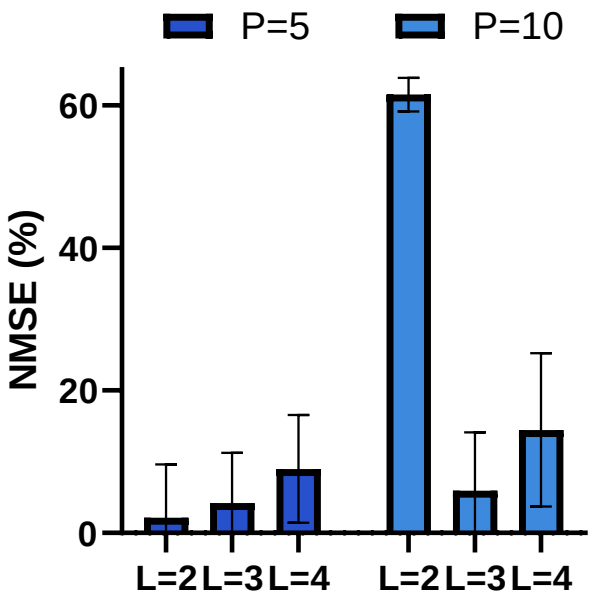

Figure D.2: NMSE for ISTA on the test set ($N = 20$) using two-, three- and four-level (L) Daubechies-7 wavelet transforms at $P = 5$ and $P = 10$.

Appendix E: Variations of the Finite Element Model

This Appendix is adapted from the supplemental material of *Frost and Olson, 2021* [89].

E.1 Longitudinal Variations

The model presented in Chapter 6 uses ST radius and OHC transducer gains that are constant along the longitudinal axis of the cochlea. In the actual gerbil cochlea both ST radius and transducer gain vary along the length of the cochlea. To justify these simplifying assumptions, we consider the effects of varying a) ST radius, and b) transducer gain along the length of the cochlea.

E.1.1 Effect of Longitudinal Variation in Scala Tympani Radius

The half-cylinder geometry (Fig 6.2) models the ST as a 520 μm wide half-cylindrical shell. This is based on the measured size of ST 2.5 mm from the round window, where the output of our model is recorded. The cross-sectional area of ST in gerbil varies non-monotonically, first increasing to a maximum about 2.5 mm from the round window, then decreasing until a point about 5 mm from the round window, after which point the cross-sectional area remains approximately constant [17].

We consider a revised model geometry in which the ST area varies in accordance with these measurements, presented in Fig E.1. The ST is modeled by three conjoined solids – 1) a truncated cone with initial radius 120 μm at the base and final radius 520 μm at the point 2.5 mm from the base; 2) a truncated cone with initial radius 520 μm at the point 2.5 mm from the base and final radius 120 μm at the point 5 mm from the base; 3) a half cylinder with radius 120 μm spanning the apical half of the model. The wall is still a 100 μm wide shell

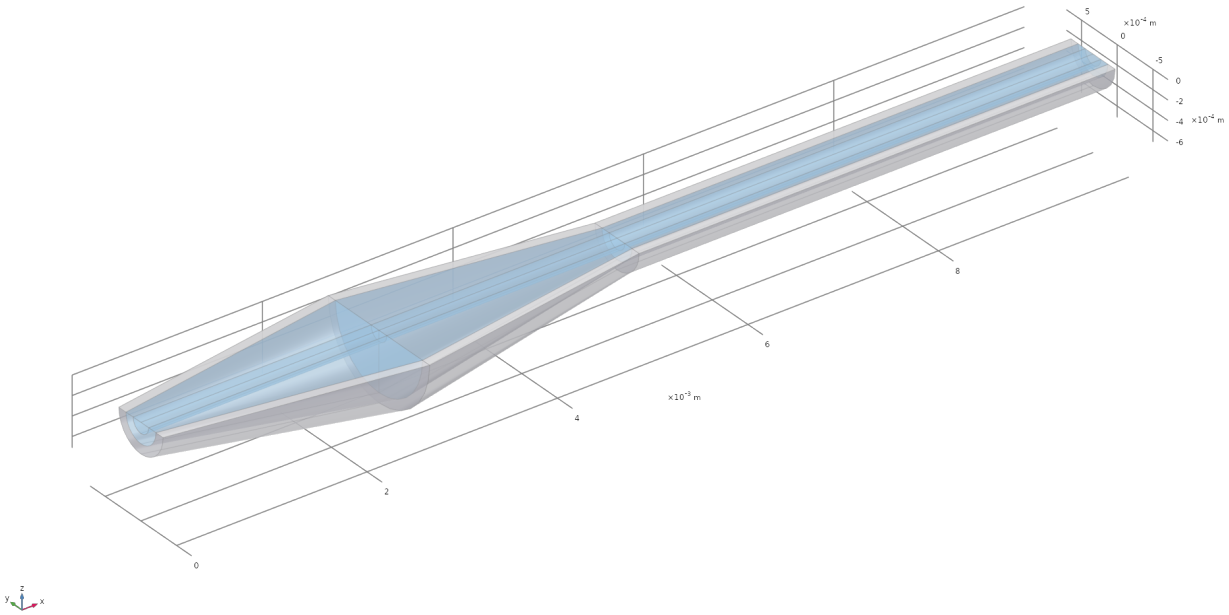

Figure E.1: Model geometry wherein ST radius varies longitudinally approximately according to the known anatomy of gerbil, as it appears in the COMSOL Multiphysics user interface. The fluid space (OC, BM and ST) is shown in light blue, while the wall is shown in gray.

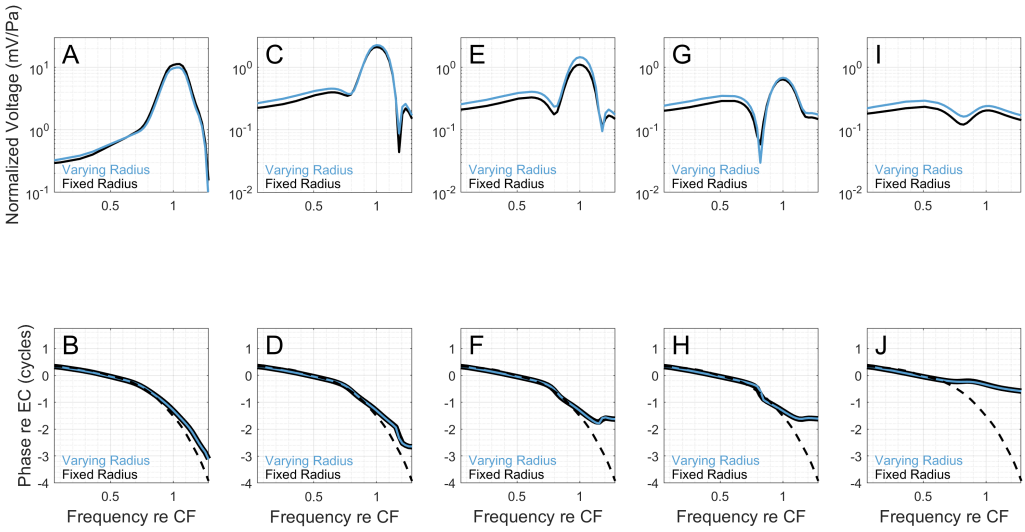

Figure E.2: Effect of more accurate geometry. CM prediction using the current source that is based on BM displacement at five locations along the line segment 2.5 mm from the base of the cochlea (19.5 kHz place), using two model geometries: that from Fig E.1 in which the radius varies, and that from Fig 1 in the main text in which the radius is constant. The presented "fixed radius" results are the predictions from Fig 6.4 SPL = 20 dB SPL, $K = 50$, channel sensitivity = 33 pA/nm. Magnitude and phase at the following distances from the current source: **A** and **B** – at the position of the line current source; **C** and **D** – 55 μm; **E** and **F** – 110 μm; **G** and **H** – 160 μm; **I** and **J** – 410 μm. The phase of the current source is shown as a dashed line in the lower panels.

The model output using this geometry along with the corresponding output using the half-cylinder geometry of Fig 6.2 is shown in Fig E.2. The presence of notches, the sharpness of tuning and the plateauing of the phase are not qualitatively affected by this change in model geometry. The amplitudes differ at some locations by a factor that is never larger than 1.5 times. The phase is essentially indistinguishable between the two model outputs. The lack of substantial difference, both qualitative and quantitative, between the model outputs using these two geometries justifies the use of the simpler geometry.

E.1.2 Effect of Longitudinal Variation of Transducer Gain

Experiments on OHCs *in vitro* have shown that the mechanoelectrical transducer gain varies across the length of the cochlea [111]. While these data are sparse in gerbil, Johnson *et al.* [111]

combined gerbil and rat hair cell data to suggest a logarithmic relationship between the OHC transducer gain and the characteristic frequency corresponding to that OHC's location in the cochlea. In the basal region of the gerbil cochlea, the frequency is related to the longitudinal location along the cochlea by an exponential relationship [97], which combined with the proposed logarithmic relationship above suggests a *linear* relationship between transducer gain and longitudinal location along the cochlea in the base.

The frequency-saturating current curve of Johnson *et al.* shows a change from about 1 nA to 3 nA between the 350 Hz and 10 kHz locations. The locations corresponding to those frequencies according to Müller are about 10.2 mm and 4 mm respectively, yielding a slope of approximately -0.3 (pA/nm)/mm. We choose an even steeper slope of $m = -0.5$ to test the robustness of our model to a varying transducer gain. So that our results between studies are comparable, we choose the gain at the 2.5 mm place (where data is recorded) to be 33 pA/nm (the starting channel sensitivity used in our study). This gives both a point and a slope for the displacement-gain linear relationship, fully defining the gain-location relationship as

$$G = 33.3 - 0.5(x - 2.5), \tag{E.1}$$

where G is the gain in pA/nm and x is the distance from the base of the cochlea in mm. The model output using this alternative stimulus is hardly different from our fixed-gain model output. Across all stimulus frequencies and all measurement locations, the maximum absolute difference between the model output using varying transducer gain and the model output using fixed transducer gain was typically too small to be noticeable in our plots, and was at most 8% of the CM magnitude, at a notch frequency. This justifies using the simpler fixed-gain model.

E.2 Effects of Notches in the Current Source

The notches in model outputs in the main paper are due to destructive interference between local and non-local current components. Notches in the LCM data could also be due to notches

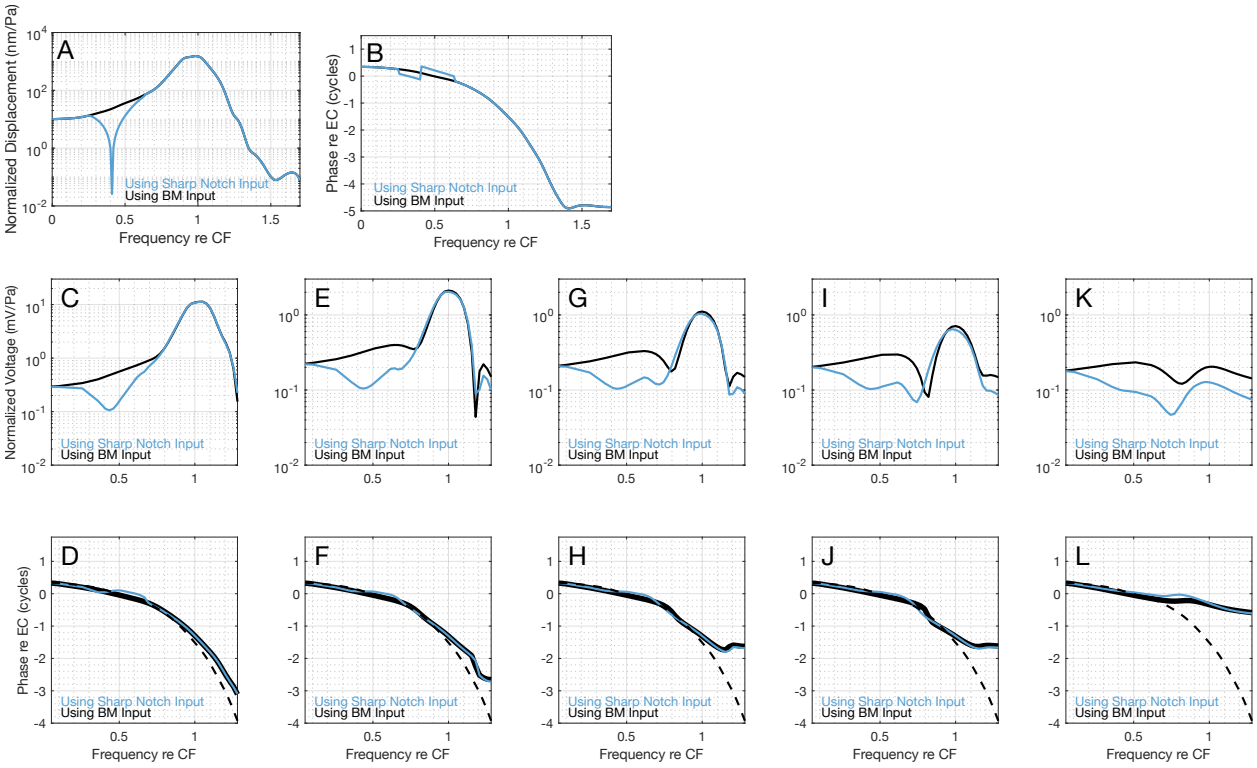

Figure E.3: Effect of a sharp notch in the current source. The notch is added to the original current source based on BM displacement, channel sensitivity = 33 pA/nm, $K = 50$, sound level = 20 dB SPL. **A** and **B** – magnitude and phase of the inputs with and without the notch. **C** through **L** show CM magnitude and phase at various distances from the current source. **C** and **D** – at the position of the line-current source; **E** and **F** – 55 μm; **G** and **H** – 110 μm; **I** and **J** – 160 μm; **K** and **L** – 410 μm. The dashed lines in the lower panels are the phase of the BM displacement used to generate the current stimulus without the notch (as in **B**).

present in the source, which could arise for several reasons, for example mechanical standing waves [112] and TM resonance [88]. We briefly explore the effects that notches in the current line source have on the FEM model output; this exploration could be expanded in future work.

Fig E.3 **A** and **B** show an alternate input generated by adding a sharp notch to the current source at 0.4CF, as well as a corresponding phase variation. The outputs of the model using this input are shown in panels **C-L**. The notch penetrates the scala as a rippled reduction without a sharp tip. The size of the reduction is related to the wider portion of the current line notch. The phase also develops mild ripples.

Another possibility is a relatively broad notch, of the type suggested by Nankali et al, as a result of TM resonance [88]. We use an input inspired by Nankali et al, as displayed in Fig E.4

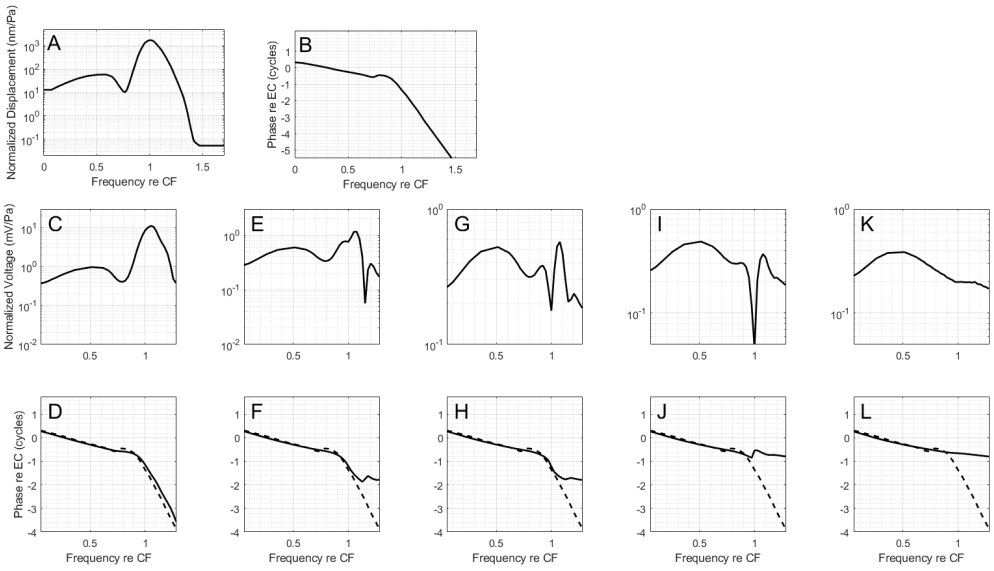

Figure E.4: CM prediction where the current source possesses a broad notch and a phase ripple, inspired by the resonant TM model of Nankali et al, Fig 5 [88]. $K = 50$ and channel sensitivity = 33 pA/nm. **A** and **B** – magnitude and phase of the current source. **C** through **L** show CM magnitude and phase at various distances from the current source. **C** and **D** – at the line-current source; **E** and **F** – 55 μm; **G** and **H** – 110 μm; **I** and **J** – 160 μm; **K** and **L** – 410 μm. The dashed lines in the lower panels are the phase of the BM displacement used to generate the current stimulus (as in **B**).

A and **B**, in which there is a broad sub-BF notch and a phase that undergoes a ripple at ~0.8 CF and accumulate rapidly at frequencies beyond the ripple. The model predictions with this current source are shown in Fig E.4 **C-L**. The broad notch penetrates the fluid with reduction, and the phase ripple penetrates the fluid and is similar to the phase of the current source until it levels off at a plateau. The degree to which the phase follows the current source phase is similar to that in the main results. A sharp notch close to the CF joins what remains of the broad notch at a distance 110 μm from the current source (Fig E.4 **G**); this sharp notch is due to phase cancellation.

In summary, the FEM predicts notches in the LCM both due to phase cancellation or due to the presence of notches in the current source, but source notches are broadened, and thus the sharp notches observed in experimental LCM measurements are most likely due to phase cancellation. On the other hand, phase shifts that are observed in LCM (for example, the shifts relative to BM displacement that were observed in [90]) are likely to be present in the current source.

Publications and Conference Proceedings

The following is a list of publications and conference abstracts I have co-authored in my time at Columbia University. Included also are manuscripts that have been submitted to journals, but have not yet been published.

Peer-Reviewed Publications

B. L. Frost and E. S. Olson, "Model of cochlear microphonic explores the tuning and magnitude of hair cell transduction current," *Biophysical Journal*, vol. 120, no. 17, 3550–3565, 2021.

B. L. Frost, C. E. Strimbu, and E. S. Olson, "Using volumetric optical coherence tomography to achieve spatially resolved organ of corti vibration measurements," *Journal of the Acoustical Society of America*, vol. 151, pp. 1115–1125, 2022.

B. L. Frost, C. E. Strimbu, and E. S. Olson, "Reconstruction of transverse-longitudinal vibrations in the organ of corti complex via optical coherence tomography," *Journal of the Acoustical Society of America*, vol. 153, pp. 1347–1360, 2023.

B. L. Frost, N. P. Janjušević, C. E. Strimbu, and C. P. Hendon, "Compressed Sensing on Displacement Signals Measured with Optical Coherence Tomography," *Biomedical Optics Express*, vol. 14, pp. 5539-5554, 2023.

Conference Abstracts

B. L. Frost and E. S. Olson, "Cochlear Microphonic Measurements Indicate Sharper Tuning at Stereocilia Than at Basilar Membrane," *Abstracts of the 44th midwinter research meeting, Associ-*

ation for Research in Otolaryngology, 2021.

B. L. Frost, C. E. Strimbu and E. S. Olson, “Using Volumetric Optical Coherence Tomography to Achieve Spatially Resolved Organ of Corti Vibration Measurements," *Abstracts of the 45th mid-winter research meeting, Association for Research in Otolaryngology*, 2022.

B. L. Frost, C. E. Strimbu and E. S. Olson, “Transverse-Longitudinal Structure Registration and Vibration Measurement via Optical Coherence Tomography," *Correspondences of the Mechanics of Hearing 2022 Meeting*, 2022.

L. A. Chiriboga, **B. L. Frost**, C. E. Strimbu and E. S. Olson, “Designing a Coupled Common-Mode OCT Probe with a Voltage Electrode for Simultaneous Intracochlear Motion and Voltage Measurements in Guinea Pig," *Abstracts of the 46th midwinter research meeting, Association for Research in Otolaryngology*, 2023.

Manuscripts In Review

C. E. Strimbu, L. A. Chiriboga, **B. L. Frost**, E. Fallah, and E. S. Olson, “A frame and a hotspot in cochlear mechanics," submitted to *Hearing Research*. Available on bioRxiv.

B. L. Frost, “Foundations of the WKB Approximation for Models of Cochlear Mechanics in 1- and 2-D," 2023. Submitted to *Journal of the Acoustical Society of America*. Available at arXiv: 2307.14353.