Measuring Nanometer, Three-Dimensional Motions with Light Microscopy

by

Charles Quentin Davis

Submitted to the Department of
Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1997

Author . . . . . .
Department of Electrical Engineering and Computer Science
May 23, 1997

Certified by . . . . . .
Dennis M. Freeman
Assistant Professor of Electrical Engineering
Thesis Supervisor

Accepted by . . . . . .
Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

Measuring Nanometer, Three-Dimensional Motions with Light Microscopy

by

Charles Quentin Davis

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 1997, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

Computational motion analysis of images from a light microscope is used to explore mechanics in the microscopic worlds of cells and man-made micromachines. Images of a moving target are taken using a strobe light and a CCD camera attached to a microscope. Temporal sequences of stop-action images are recorded at multiple planes of focus to characterize three-dimensional motions.

We demonstrate that motion resolution is limited by the motion detection algorithm. The largest source of error in gradient-based and matching algorithms is statistical bias. We present a new algorithm with errors that are typically smaller than 0.02 pixels. With this algorithm and our video microscopy system, we can measure motions of submicrometer targets with nanometer accuracy.

The system has been used to study both hearing structures and man-made micromachines. Analysis of electrically-induced motions of a microfabricated accelerometer and a microfabricated angular velocity sensor allow visualization and quantification of complex modes of motion that limit the performance of these sensors. Pilot studies demonstrate that the system can also measure sound-induced motions of the human stapes, where it can resolve complex modes of motion thought to be important for loud sounds and for certain middle-ear pathologies. The system has also been used to measure mechanical deformations in the mouse tectorial membrane, a gelatinous inner-ear structure that has previously been a difficult target of study.

Our major application has been the study of sound-induced motions of sensory cells and accessory structures in the ear of a lizard. Our results, while still preliminary, are the first direct measurements of sound-induced motions of hair bundles and their overlying tectorial membrane. They indicate that the tectorial membrane does not move as a rigid body but has significant shear throughout its thickness. Therefore, the relative motion between the tectorial membrane and the reticular lamina is not necessarily indicative of hair bundle rotation. Shear through the tectorial membrane has not been previously reported. These results demonstrate that our system can measure cochlear micromechanical processes that have previously eluded direct experimental study.

Thesis Supervisor: Dennis M. Freeman
Title: Assistant Professor of Electrical Engineering

Acknowledgments

First and foremost, I would like to thank my advisor and friend, Prof. Denny Freeman. I met Denny as a Junior seeking research experience. Seven years later, he is still my teacher, mentor, and friend. Even as an undergraduate, he gave me very challenging projects to work on. I thank him for "saving" those projects for me to work on again as a graduate student. I thank him for his faith that I could design the motion measurement system that formed the basis of my thesis, and then for his help in actually doing it. The food parties that we had with Farzad and Kristin were great, and my tongue still hasn't recovered from Ben Van Roy's Thai soup. I hope our paths continue to cross.

I would also like to thank my thesis committee: Profs. Tom Weiss and Berthold Horn. Their suggestions at committee meetings and on the thesis document greatly improved the thesis. I would also like to thank Tom for his help with the lizard preparation.

The other students in the lab really made my graduate years fun and interesting. AJ, thanks for helping me with the lizard experiments, proofreading my thesis, and trying to teach me the forehand frisbee toss. You were a great office-mate despite your loud typing. Cam, thanks for teaching me how to machine plexiglass and all about low-fat food. I didn't know styrofoam could taste that bad. Zoher, I hope some of your business savvy rubbed off on me. Laura, thanks for letting me be your advisor and for being so much fun to work with.

To all the Blasches—Erik, Ian, Kyle, Bruce, Barb, and especially AT—thanks for your support even if Barb did beat me. Erik, thanks for all the late-night phone calls to keep our spirits up!

Dad, thanks for all your encouragements, including getting me to come to MIT. Mom, thanks for believing that I would actually finish. Eric, thanks for toughening me up for my many late nights at MIT. Finally, I thank my wife, Nicole, whose support during my graduate career was invaluable. Thanks also for that "extra" incentive to graduate!

Contents

Chapter 1

Introduction

The focus of many scientific disciplines has shifted from the macroscopic to the microscopic domain. For example, the thrust of biological research has changed from the organismic to cellular and subcellular levels. Electrical structures are now built largely from micron-scale transistors. Mechanical structures seem to be following suit with the development of microelectromechanical systems (MEMS). All of these fields rely on techniques for measuring physical processes at these microscopic scales. In this thesis I describe a new technique for quantifying motions of microscopic structures, and demonstrate example applications in studying the mechanisms of hearing and the behavior of MEMS.

Measuring motions at these small scales is difficult. The structures themselves are small. Important mechanical structures in the ear include millimeter-sized bones and micrometer-sized cellular structures. Features in MEMS devices span the millimeter to micrometer range. Motions of these structures may be only a fraction of their size — from nanometers to micrometers. The diminutive size of biological and man-made micromachines enables them to have large accelerations with small forces and thereby move at high frequencies. Bats can hear sounds at frequencies exceeding 100 kHz and some MEMS devices also operate at those frequencies.

We have combined video microscopy, stroboscopic illumination, and computer vision to measure the three-dimensional motions of these microscopic structures with nanometer precision. By exploiting stroboscopic illumination and the microscope's

ability to image different structures at different planes of focus, we can obtain a sequence of stop-action, three-dimensional images of moving structures. Computer vision algorithms are then applied to the images to estimate motion. This system, which we call computer microvision, has many useful features. Since motion is estimated directly from video images, the structure whose motion is estimated is unambiguous. In addition, three-dimensional motions of all visible structures can be estimated simultaneously. Furthermore, motion measurements are made in a non-contact fashion—the only part of the measurement apparatus that touches the moving structures is light.

This thesis is organized as follows. Chapter 2, "Computer Microvision," presents a overview of the computer microvision system. Chapters 3-6 are work done on motion analysis. Chapters 7-9 highlight the applications of computer microvision. Chapters 2-5 and 7 are organized as manuscripts to appear in peer-reviewed journals. Different backgrounds are assumed in each of these chapters, owing to differences in the audiences reached by the different journals.

Chapter 3, "Equivalence of Subpixel Motion Estimators Based on Optical Flow and Block Matching," demonstrates the mathematical equivalence of these two broad classes of motion estimators, and discusses useful insights to be gleaned from each representation. This paper is in print (Davis et al., 1995). Chapter 4, "Statistics of Subpixel Registration Algorithms based on Piecewise Linear Interpolators," and Chapter 5, "Nanometer Registration with Light Microscopy: Characterization of Performance Limitations," present theoretical and technical limitations of the computer microvision system. Chapter 6, "Image Registration" further analyzes the motion detection algorithms developed in Chapter 4, and extends those methods for three-dimensional motions.

Chapter 7, "Using video microscopy to characterize micromechanics of biological and man-made micromachines," demonstrates two example applications of the computer microvision system. This paper is in print (Freeman and Davis, 1996). Chapter 8, "Direct Observations of Inner-ear Mechanics," includes an extensive description of the apparatus to measure sound-induced motions results of sound-induced motions, and work done to assess the viability of the in vitro preparation of the lizard cochlea.

Chapter 9, "Other Applications of Computer Microvision," demonstrates that the computer microvision system has broad application. Motions of a isolated tectorial membrane (used to study its mechanical properties), an in situ human stapes (used to study middle ear mechanics), and a microfabricated gyroscope (used to study its dynamic properties) are briefly examined.

Chapter 2

Computer Microvision

2.1 Abstract

Computational motion analysis of images from a light microscope provides new opportunities to explore mechanics in the microscopic worlds of cells and man-made micromachines. We have developed a system that can measure motions of submicrometer targets with nanometer accuracy. Motion measurements of a microfabricated accelerometer demonstrate the system's use as a diagnostic instrument to support the design and manufacture of microelectromechanical systems. The system has also been used to observe sound-induced motions of sensory cells and accessory structures in the intact cochlea. These measurements reveal dynamical relations that may well contribute to the ear's acute sensitivity to sound.

2.2 Computer microvision

We have combined computer vision with light microscopy to create a system that we call *computer microvision.* This system enables us to visualize motions of structures with micrometer dimensions over times from microseconds to days. Images of a moving target are taken using a microscope (Zeiss Axioplan, Thornwood, NY) with a strobe light (Chadwick-Helmuth xenon flashtube, El Monte, CA or an LED) and a CCD camera (Photometrics 200 series with a Thompson 7883 CCD, Tucson, AZ).

Temporal sequences of stop-action images are recorded at multiple planes of focus to characterize three-dimensional motions.

The sequences of recorded images can be viewed at playback speeds chosen to facilitate human interpretation of motion. In addition, quantitative motion estimates can be obtained directly from the images using algorithms from computer vision (Horn, 1986). The quantitative motion estimates can also be used to amplify or suppress specific modes of motion in the images. Viewing the modified images can reveal modes of motion that are not obvious in the originals.

The computer microvision system can achieve nanometer *motion* resolution even though the image resolution is much worse. Image resolution is limited by the optics (Hopkins and Barham, 1950) to dimensions on the order of the wavelength of light (550 nm for green light). However, one can detect motions of a blurred image that are significantly smaller. Fundamental limitations on motion detection are determined by the contrast of the target, by the resolution and precision of the optics and video imager, and by the quantum nature of light.

Motion resolution is also limited by the computer vision algorithm. We have developed algorithms to estimate displacements between two images taken at different times. Generally displacements are much smaller than the pixel spacing of the camera.[1] We identify regions of interest that contain a single moving structure and determine the best estimate of rigid-body displacement between successive images in a least-squares sense (Horn and Weldon, Jr., 1988). For subpixel displacements, this gradient-based algorithm is equivalent to finding the shift that minimizes the sum of squared differences between shifted versions of linearly interpolated images (Davis et al., 1995). Linear interpolation introduces errors that cause a target-dependent bias in the estimates. This bias is significantly larger than the variability of motion estimates resulting from quantum nature of light and therefore is the limiting factor in motion detection accuracy.

To reduce the bias,[2] we shift the second image one pixel in the direction that is

[1] Larger displacements can be accurately estimated using an iterative "pyramid" procedure (Burt, 1981) which reduces suprapixel motion estimation to subpixel estimation.

[2] For simplicity, the bias-reduction method is described for one-dimensional images, but has been

opposite to its estimated displacement, repeat the algorithm to determine a second estimate. Shifting the second image by one pixel can be done exactly and efficiently, and the second estimate provides additional information. By dividing the first estimate by the difference between the two estimates, which would be 1 if the algorithm were unbiased, bias that is directly proportional to the displacement is eliminated. This bias-reduction technique improves the accuracy of the original algorithm by a factor of 4–10, depending on the images.

We have tested our system using targets whose motions could also be determined with an independent method. Our noise floor is typically less than 4 nm — i.e., <1% of the wavelength of the light that generates the images. The corresponding image displacement is 1/100 pixel.

2.3 Cochlear micromechanics.

We have applied computer microvision to study the micromechanics of the cochlea[3], a biological micromachine that is sensitive to eardrum displacements as small as 100 picometers (von Békésy, 1960; Fay and Popper, 1994). Although the cochlea has been the target of many investigations, the mechanisms underlying its remarkable sensitivity remain controversial.

Sensory cells in the cochlea are called hair cells and are organized along with other cells and accessory structures to form a resonant partition in the saline-filled inner ear. Hair cells are specialized for mechanical-to-electrochemical transduction by a microscopic bundle of mechanically-sensitive sensory hairs (Hudspeth, 1983; Hudspeth, 1989), that project into saline and are covered by a gelatinous structure called the tectorial membrane. Relative motions between hair cells and the overlying tectorial membrane cause hair bundle rotations that modulate the ionic conductance of stress-gated ion channels in the hair-cell membrane. The mechanical stress is believed to be generated by microscopic filaments called tip-links (Pickles and Corey,

extended and implemented for two and three dimensional motions.

[3]Preliminary results have been described elsewhere (Davis and Freeman, 1995).

1992; Hudspeth and Gillespie, 1994), which connect the tips of adjacent sensory hairs.

We have developed an in vitro preparation for studying the sound-induced motions of cochlear hair bundles.[4] The cochlea of an alligator lizard is mounted in an experimental chamber so that it can be observed with a microscope while being stimulated by an underwater pressure transducer that simulates physiological stimulation. Key structures including hair bundles and the tectorial membrane are directly imaged to determine relations between the motions of these structures during acoustical stimulation — relations that have not previously been measured.

Typical results are shown in Figure 2-1. Displacements near the base of this hair bundle are larger than those near the tip, i.e., the hair bundle is rotating about its point of attachment to the hair cell. Displacements of the tectorial membrane are generally smaller than those of either the base or tip of the hair bundle. Furthermore, as the distance from the base of the hair bundle increases, the motion shows increasing phase lag. By subtracting the displacement waveform of the tip of the hair bundle from the base, one can determine that the displacement of the tip of the hair bundle relative to the base is 0.3 μm peak-to-peak, corresponding to a hair bundle rotation of 3° peak-to-peak. Motions of five other hair bundles in this preparation were analyzed and the results were qualitatively similar to those shown in Figure 2-1. Motions orthogonal to the plane of focus were generally smaller (0.1 μm peak-to-peak) than motions in the plane, and are consistent with in vivo measurements (Peake and Ling, 1980).

These data can also be used to determine displacements of individual sensory hairs. Stimulation of the cochlea by a sinusoidal pressure generates nearly sinusoidal motions of all the sensory hairs (Figure 2-2, middle left). To determine if there is relative motion between the sensory hairs, we applied our motion detection algorithm to estimate the motion of the right edge of the bundle and shifted the images to compensate for that motion. The resulting images show relative motions between sensory hairs (Figure 2-2, bottom left). These results suggest tip-links that connect

[4]Surgical procedures are described elsewhere (Freeman et al., 1993). The care and use of animals have been approved by the Massachusetts Institute of Technology Committee on Animal Care.

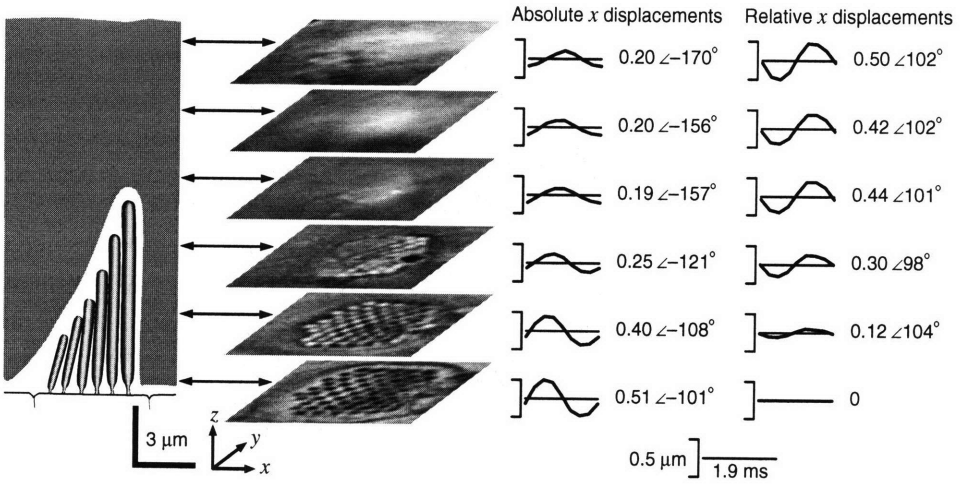


Figure 2-1: Sound-induced motions of a hair bundle and overlying tectorial membrane. The double-headed arrows associate a sequence of images from our system (3 μm spacing) with the corresponding planes of section in the schematic drawing, which shows 6 of the sensory hairs in the bundle of one cell and the overlying tectorial membrane (gray). Similar images were obtained at 8 phases of the 513 Hz stimulus. The intensity (115 dB SPL in the fluid adjacent to the basilar membrane) corresponds to approximately 89 dB SPL at the eardrum. The leftmost waveforms show displacements in the x direction during one cycle of the stimulus. The associated numbers are the peak-to-peak magnitude (in μm) and angle (in degrees) of the fundamental component of the displacement. The rightmost waveforms show differences between the left waveforms and the bottom left waveform.

the tips of neighboring stereocilia are stretched not only because the distance between insertion points increases when the sensory hairs are rotated by equal angles,[5] but also because the sensory hairs separate as they rotate. Motions of individual sensory hairs have been measured elsewhere (MacDonald and Corey, 1996; Duncan et al., 1995), but only for hair bundles that were stimulated with a water jet; the membranes through which natural stimuli are delivered had been removed.

These results are the first direct measurements of sound-induced motions of hair bundles with their overlying tectorial membrane, and the first measurements of sound-induced motions of individual stereocilia. From similar measurements we hope to better understand these micromechanical relations that contribute to the ear's acute sensitivity to sound.

[5]If d is the distance between the bases of two parallel sensory hairs (Figure 2-1) whose lengths differ by Δ, then the square of the distance between their tips is $\Delta^2 + d^2 + 2d\Delta \sin\theta$ where θ is the angular rotation from the vertical. Thus, rotations of the hair bundle lead to tip-link stretch.

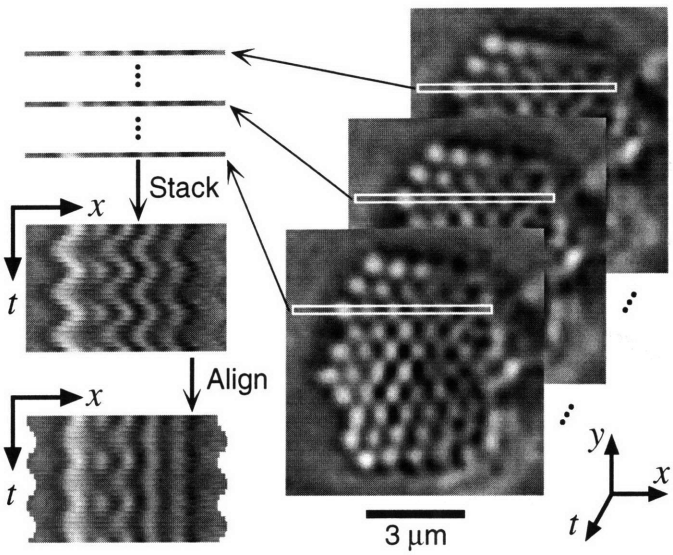


Figure 2-2: Motions of sensory hairs within a bundle. Highlighted regions of images (right panel, which is the second lowest in Figure 2-1) from eight phases of the stimulus period were stacked, with regions from earlier phases on top of regions from later phases. The resulting montage is repeated three times (middle left) to illustrate the motions of individual sensory hairs. The images in the lower left panel were first shifted to cancel the motion of the right part of the hair bundle.

2.4 Micromechanics of MEMS.

We have applied computer microvision to study the motions of man-made microelectromechanical systems (MEMS).[6] These systems can be mass-produced and have evolved during the past 15 years from scientific curiosities to commercial products, holding promise to revolutionize the fields of sensor and actuator design.[7]

Figures 2-3 and 2-4 illustrate results for a microfabricated accelerometer.[8] A shuttle and associated combs are suspended above the substrate via a folded-beam cantilever spring system. Accelerations of the substrate cause relative motions that can be sensed as a change in capacitance between the interdigitated teeth of combs attached to the shuttle and substrate. The shuttle can also be moved by applying an electrical potential across the interdigitated teeth.

We applied a sinusoidal voltage to the combs and measured the horizontal displacement of the shuttle with our computer microvision system. The resultant frequency response is well approximated by a second-order, mass-spring system with a resonant frequency of 8.86 kHz and a quality of tuning of 13.6. The standard deviation of

[6]Preliminary results have been described elsewhere (Freeman and Davis, 1996).

[7]For example, a micromachined accelerometer (Analog Devices ADXL50, Norwood, MA, http://www.analog.com) is now used as the sensor for airbag deployment on vehicles made by General Motors, Honda, and others.

[8]This accelerometer was fabricated using surface micromachining techniques at MCNC (Research Triangle Park, NC). See http://www.mcnc.org/memshome.html for fabrication information.

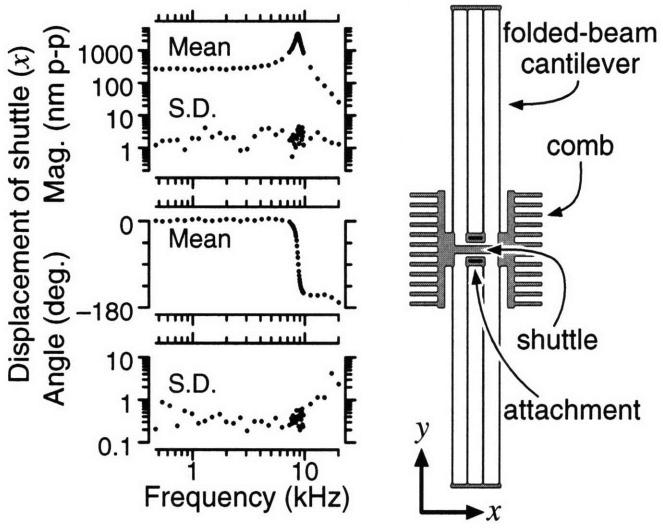


Figure 2-3: Frequency response of a microfabricated accelerometer. The right panel shows a schematic diagram of the moving parts of the accelerometer and the two attachment points to the substrate. The left panels show the fundamental component of the shuttle's horizontal displacement as a function of frequency. At each frequency, 5 trials[9] were performed to estimate the mean and standard deviation (S.D.). The stimulus was a differential electrical signal to the two comb drives ($62 \pm 2\cos(2\pi ft)$ volts), with the voltage referred to the substrate.

repeated measurements is less than 5 nm in magnitude and the signal to noise ratio is nearly 60 dB at resonance. The standard deviation of repeated phase measurements is less than 1 degree except when the magnitude is less than 50 nm.

The computer microvision system can also measure motions of the accelerometer that are orthogonal to the plane of focus. In this device, out-of-plane motions represent failure modes that limit performance. Figure 2-4 shows that the motion of teeth on opposing combs are nearly out-of-phase with one another and that the motion near the center of the shuttle is small. Thus, the shuttle is rocking about a point near its center. Repeatability of this measurement is within 4 nm in magnitude and a few degrees in phase.

These results demonstrate the use of computer microvision as a diagnostic instrument for MEMS. MEMS are fabricated using VLSI-like methodologies similar to those used to fabricate electronic devices. Unlike electronics, however, simple methods for testing and characterizing internal failure modes of micromechanical devices do not exist. This limitation has important consequences for designers and fabricators of MEMS. For the designer, it can result in large numbers of costly prototypes.

[9]Images at 8 evenly-spaced phases of the sinusoidal stimulus were taken. The computer vision algorithm was used to estimate the displacement between successive images to reconstruct the time waveform, from which the magnitude and angle were computed.

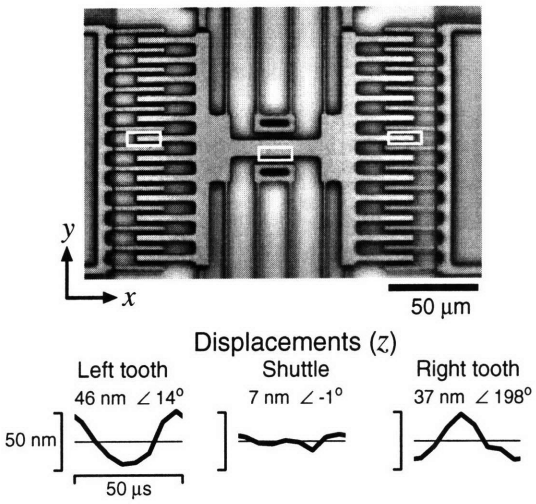


Figure 2-4: Out-of-plane motions of a microfabricated accelerometer. The top image shows three regions of interest on the accelerometer: a moving tooth on the left and right comb drives and the center portion of the shuttle. The bottom panels shows estimates of the out-of-plane displacement for a differential electrical stimulus ($62 \pm 60\cos(2\pi\, 2 \times 10^4\, t)$ volts). To obtain these waveforms, data from 30 focal planes (separated by 0.545 μm) were taken at 8 evenly-spaced phases of the stimulus cycle.

For the fabricator, yields are reduced because common techniques such as laser trimming cannot be applied. Diagnostic instruments based on computer microvision hold promise to facilitate the design and manufacture of MEMS in the same way that signal analyzers facilitate the design and manufacture of electronics.

2.5 Comparison to other methods.

During the past decade, there has been enormous progress in the development of scanned-probe microscopy techniques (e.g., Wickramasinghe, 1991) that (1) do not require special preparations to view the target and (2) have at least an order of magnitude better resolution than widefield optical microscopy. One could imagine that an increase in image resolution would increase motion resolution (e.g., Weiss et al., 1993), making a computer microvision built around a scanned-probe technique superior to a system built around a widefield optical microscope as was done here.

Unfortunately, scanned-probe techniques have other limitations that preclude their use in hearing: (1) they have extremely small working distances ($\leq$ 10 nm) and (2) they only measure surface structures. In Figure 2-1 we examined motions that were 15,000 nm below the surface of the cochlea. These limitations also preclude the use of scanned-probe microscopy in many MEMS applications. However, a computer microvision system built around a scanned-probe microscopy could dramatically increase both image and motion resolution for the class of MEMS applications

with a single moving surface, such as the accelerometer in Figure 2-3.

Computer microvision offers a number of advantages over other motion measurement systems for either biological and man-made micromachines. For example, the target whose motion is being measured is unambiguous in computer microvision, because the motions are determined directly from video images. This contrasts with other motion measurement systems (e.g., utilizing a capacitive probe, the Mössbauer effect, laser interferometry) in which the target must be inferred from observations using a separate imaging system.

Unlike other methods that estimate one component of the motion of one target point (e.g., utilizing a capacitive probe, the Mössbauer effect, laser interferometry), computer microvision allows one to estimate the 3D displacement vectors. The cost of this flexibility is the time required to digitize and store all of the images that are needed. A typical data set consists of 104 images (13 planes times 8 phases) each with 576 $\times$ 382 pixels. The total data collection time is typically 5 minutes. This amount of time is large compared to the time needed to determine the motion of a single structure using other measurement methods. However, from the resulting 104 images, one can estimate the 3D displacement waveform for *every* resolvable structure in the field of view. Thus the total measurement time is in fact extraordinarily short compared to the time it would take to make motion measurements of each structure using another method.

Another advantage of determining motions directly from video images is that these same images can be viewed by humans. Such viewing provides important checks that the quantitative results make sense. After determining the motions of an image sequence using computer microvision, we routinely shift each image to compensate for the motion. The resulting sequence of shifted images should appear to be stationary. Deviations are readily detected, and such deviations have often uncovered complex modes of motions that were not readily apparent from the quantitative results. This occurred, for example, when we first analyzed the hair bundle motions in Figure 2-1. Shifting the images by the displacements obtained using computer microvision did not stabilize the image as we had anticipated. Rather, it revealed motions between

the sensory hairs (Figure 2-2) that could have gone undetected using other motion measurement systems.

Chapter 3

Equivalence of Subpixel Motion Estimators Based on Optical Flow and Block Matching[1]

3.1 Abstract

We compare two classes of algorithms for estimating subpixel, rigid-body translation between two images. One class is based on optical flow. Optical flow algorithms determine translations between images from estimates of spatial and temporal derivatives of brightness. The other class is based on block matching. Block matching algorithms determine translations between images by minimizing the difference between shifted (warped) versions of the original images.

We show that these two classes of algorithms are equivalent for subpixel displacements. Specifically, we show that all block matching algorithms that use bilinear interpolation can be recast into equivalent optical flow formulations, and that all algorithms based on optical flow using first-order derivative estimators can be recast into equivalent block matching formulations.

[1]This chapter is reprinted with permission from Davis et al. (1995).

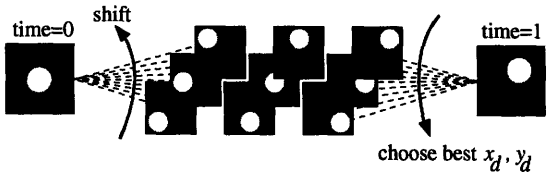


Figure 3-1: Block matching. Translation is determined by comparing shifted (warped) versions of the original images and choosing the shift that gives the closest match.

3.2 Keywords

motion detection, image registration, optical flow, block matching, computer vision, subpixel

3.3 Introduction

This study was motivated by our work in hearing. We use a strobe light, microscope, and CCD camera to take stop-action images of sound-induced motions of inner-ear structures. Our video images result from orthographic projection of objects in the microscope's focal plane. Thus rigid-body translations of inner ear structures result in rigid-body translations of the image. Even after magnification by a light microscope, sound-induced motions are much smaller than the pixel spacing of a modern video camera. Thus, we are interested in algorithms for estimating subpixel, rigid-body displacements between images.

Two popular classes of rigid-body motion estimation algorithms are optical flow (Horn and Schunck, 1981; Aggarwal and Nandhakumar, 1988) and block matching (Netravali and Haskell, 1988), both of which have been applied to estimate subpixel displacements (Tian and Huhns, 1986; Giorda and Racciu, 1975). As shown in Figures 3-1 and 3-2, the ideas behind these two families of algorithms appear to be completely different. After more carefully defining these families of algorithms, we show by example how to convert an algorithm in one family into an equivalent algorithm in the other. We then show that the conversion process is general. Thus, these families of algorithms are equivalent for subpixel displacements.

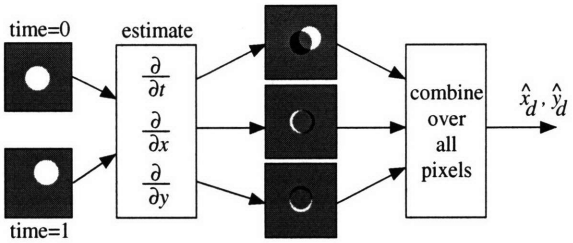


Figure 3-2: Optical flow. Temporal and spatial gradients of brightness are used to estimate translation in a single pass.

3.4 Definitions

3.4.1 Optical flow

Motion estimators based on optical flow share the constant brightness assumption (Horn and Schunck, 1981), which states that the brightness E of an image remains constant with time:

$$E(x, y, t) = E(x + x_d, y + y_d, t + t_d)$$

where x and y represent spatial coordinates and t represents time. The right hand side of the equation can be written as Taylor's series expanded about the point (x, y, t). Then, after eliminating $E(x, y, t)$ from both sides, the constant brightness equation can be written as

$$0 = \frac{\partial E(x, y, t)}{\partial x} x_d + \frac{\partial E(x, y, t)}{\partial y} y_d + \frac{\partial E(x, y, t)}{\partial t} t_d + \cdots$$

where the ellipses represent higher order terms. We can simplify this equation by using the camera's frame time as our unit of time so that $t_d = 1$ frame. Then

$$E_x x_d + E_y y_d + E_t + \cdots = 0 \tag{3.1}$$

where E_x, E_y, and E_t represent the partial derivatives of the brightness function E.

Equation 3.1 can be made discrete by changing the functions with continuous domain to functions with discrete domain (Horn and Schunck, 1981). Let $G(i, j, k)$ represent the gray (brightness) value from the camera, where i and j are the pixel coordinates and k is the picture number. The transformation from E to G is generally

nontrivial, and details are given elsewhere (Janesick et al., 1987). For simplicity (and without lack of generality) we choose our coordinate system so that $G(i,j,k) = E(i,j,k)$ for all i,j,k.

The discrete form of Equation 3.1 can be written as

$$G_x x_d + G_y y_d + G_t + \cdots = 0 \tag{3.2}$$

where G_x, G_y, and G_t are linear combinations of images and their integer shifts that approximate the partial derivatives of the brightness function E. For example, if first-order backward differences are used to approximate the spatial derivatives then $G_x(i,j,k) = G(i,j,k) - G(i-1,j,k)$, where we assume that space is measured in units of pixels. Throughout this paper, we consider only the case of rigid body translation. For this case, the displacements x_d and y_d are constants. Thus Equation 3.2 represents a system of equations—one equation for each cluster of pixels for which G_x, G_y, and G_t can be calculated—with just two unknowns, x_d and y_d. This generally overdetermined system of equations can be solved in the least squares sense by choosing $\widehat{x_d}$ and $\widehat{y_d}$ equal to the values of x_d and y_d that minimize the sum of the squares of the left-hand side of Equation 3.2:

$$(\widehat{x_d}, \widehat{y_d}) = \arg\min_{x'_d, y'_d} \sum_{i,j} (G_x x'_d + G_y y'_d + G_t + \cdots)^2 \tag{3.3}$$

where the sum is over all relevant clusters of pixels.

3.4.2 Block matching

Motion estimates based on block matching are based on the assumption that two consecutive images $E(x,y,0)$ and $E(x,y,1)$ are related by a simple shift

$$E(x,y,1) = E(x - x_d, y - y_d, 0).$$

The displacements x_d and y_d are estimated by minimizing a measure of the magnitude of the difference between shifted versions of the two images,

$$(\widehat{x_d}, \widehat{y_d}) =$$

$$\arg\min_{x'_d, y'_d} \sum_{i,j} \left| E(i + x_1, j + y_1, 1) - E(i + x_0, j + y_0, 0) \right|^q \tag{3.4}$$

where $x'_d = x_1 - x_0$ and $y'_d = y_1 - y_0$. The discrete form of this equation is simple when the trial displacements x'_d and y'_d are integer multiples of the pixel spacing. Then we can take $x_0 = y_0 = 0$ and express the minimization in terms of the discrete brightness values $G(i, j, k)$,

$$(\widehat{x_d}, \widehat{y_d}) =$$

$$\arg\min_{x'_d, y'_d \in \{0, \pm 1, \cdots\}} \sum_{i,j} \left| G(i + x'_d, j + y'_d, 1) - G(i, j, 0) \right|^q.$$

However, we are interested in applications in which the displacements x_d and y_d are subpixel. Thus to apply block matching, we must specify an interpolating function $F(x, y, k)$ to estimate the brightness between points sampled by the camera. One common interpolating function is the bilinear interpolator (Parker et al., 1983; Girod, 1993), in which $E(x + x'_d, y + y'_d, k)$ is approximated by

$$F(i + x'_d, j + y'_d, k) = (1 - x'_d)(1 - y'_d) G(i, j, k) +$$

$$x'_d (1 - y'_d) G(i + 1, j, k) + (1 - x'_d) y'_d G(i, j + 1, k) +$$

$$x'_d y'_d G(i + 1, j + 1, k) \tag{3.5}$$

where we have assumed $0 \leq x'_d < 1$ and $0 \leq y'_d < 1$.

3.5 Examples

In this section, we show that several common implementations of block matching have equivalent optical flow realizations. The first three are one dimensional; the last is two dimensional.

3.5.1 Example 1

Consider a block matching algorithm in which the displacement x_d is determined by minimizing the sum of the squared differences between the first image and a shifted (warped) second image:

$$\widehat{x_d} = \arg\min_{x'_d} \sum_i \left(E(i + x'_d, 1) - E(i, 0)\right)^2 .$$

After computing the shifts with linear interpolation (Equation 3.5) the minimization problem becomes

$$\widehat{x_d} =$$

$$\arg\min_{x'_d} \sum_i \left((1 - x'_d)G(i, 1) + x'_d G(i + 1, 1) - G(i, 0)\right)^2$$

$$= \arg\min_{x'_d} \sum_i \left(G_t + G_x x'_d\right)^2 \tag{3.6}$$

for $0 \leq x'_d \leq 1$, where

$$\begin{aligned} G_x &= G(i + 1, 1) - G(i, 1) \\ G_t &= G(i, 1) - G(i, 0). \end{aligned} \tag{3.7}$$

Notice that the form of Equation 3.6 is identical to that of Equation 3.3. It follows that the block matching technique of warping the second image to match the first using linear interpolation is equivalent to an optical flow method in which derivatives are approximated by first differences given in Equations 3.7 and illustrated in Figure 3-3.

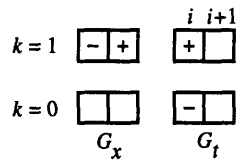


Figure 3-3: Derivative estimators for Example 1. The left panel illustrates the spatial derivative estimator, a forward difference operator that is applied to the image at time $k = 1$. The right panel illustrates the temporal derivative estimator, a forward difference operator that is applied to pixel i in the two images.

3.5.2 Example 2

Consider a block matching algorithm in which the displacement x_d is determined by minimizing the sum of the squared differences between two images, each shifted halfway:

$$\widehat{x_d} = \arg\min_{x'_d} \sum_i \left(E(i + \frac{x'_d}{2}, 1) - E(i - \frac{x'_d}{2}, 0)\right)^2.$$

After using linear interpolation (Equation 3.5), the minimization problem becomes

$$\begin{aligned}\widehat{x_d} = \arg\min_{x'_d} \sum_i \Big[&(1 - \frac{x'_d}{2})G(i,1) + \frac{x'_d}{2}G(i+1,1) \\ &- \Big((1 - \frac{x'_d}{2})G(i,0) + \frac{x'_d}{2}G(i-1,0)\Big)\Big]^2 \\ = \arg\min_{x'_d} &\sum (G_t + G_x x'_d)^2 \end{aligned} \tag{3.8}$$

for $0 \leq x'_d \leq 2$, where

$$G_x = \tfrac{1}{2}[G(i+1,1) - G(i,1) + G(i,0) - G(i-1,0)]$$

$$G_t = G(i,1) - G(i,0). \tag{3.9}$$

As in the previous example, the form of Equation 3.8 is identical to that of Equation 3.3. It follows that this block matching technique is also equivalent to an optical flow method. However, the derivative estimators, which are illustrated in Figure 3-4 are different from those for Example 1 (Figure 3-3).

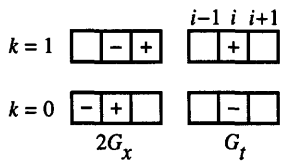


Figure 3-4: Derivative estimators for Example 2. The + and − symbols represent +1 and −1; therefore the right panel represents G_t and the left represents $2G_x$.

3.5.3 Example 3

Here we estimate the displacement x_d by minimizing the sum of the squares of a more complicated function:

$$\widehat{x_d} = \arg\min_{x'_d} \sum_i \left(E(i + \frac{1}{2} + \frac{x'_d}{2}, 1) - E(i + \frac{1}{2} - \frac{x'_d}{2}, 0) \right)^2.$$

After using linear interpolation (Equation 3.5), the minimization problem becomes

$$\widehat{x_d} = \arg\min_{x'_d} \sum_i \frac{1}{4} \Big[(1 - x'_d) G(i,1) + (1 + x'_d) G(i+1,1)$$

$$- ((1 + x'_d) G(i,0) + (1 - x'_d) G(i+1,0)) \Big]^2$$

$$= \arg\min_{x'_d} \sum (G_t + G_x x'_d)^2$$

for $-1 \leq x'_d \leq 1$, where

$$G_x = \tfrac{1}{2} \left(G(i+1,1) - G(i,1) + G(i+1,0) - G(i,0) \right)$$
$$G_t = \tfrac{1}{2} \left(G(i,1) - G(i,0) + G(i+1,1) - G(i+1,0) \right).$$

These derivative estimators, which are illustrated in Figure 3-5 are the one-dimensional versions of Horn and Schunck's (1981) derivatives for optical flow. Consequently, the standard optical flow technique is equivalent to the block matching algorithm described in this section.

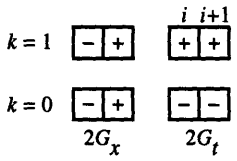


Figure 3-5: Derivative estimators for Example 3.

3.5.4 Example 4

Consider the following two-dimensional example.

$$(\widehat{x_d}, \widehat{y_d}) = \arg\min_{x'_d, y'_d} \sum_{i,j} \Big(E(i + \frac{1}{2} + \frac{x'_d}{2}, j + \frac{1}{2} + \frac{y'_d}{2}, 1)$$

$$- E(i + \frac{1}{2} - \frac{x'_d}{2}, j + \frac{1}{2} - \frac{y'_d}{2}, 0) \Big)^2$$

After using bilinear interpolation (Equation 3.5), the minimization problem is

$$(\widehat{x_d}, \widehat{y_d}) = \arg\min_{x'_d, y'_d} \sum_{i,j} \frac{1}{16} \Big((1 - x'_d)(1 - y'_d) G(i, j, 1)$$

$$+ (1 + x'_d)(1 - y'_d) G(i + 1, j, 1) + (1 - x'_d)(1 + y'_d) G(i, j + 1, 1)$$

$$+ (1 + x'_d)(1 + y'_d) G(i + 1, j + 1, 1) - (1 + x'_d)(1 + y'_d) G(i, j, 0)$$

$$- (1 - x'_d)(1 + y'_d) G(i + 1, j, 0) - (1 + x'_d)(1 - y'_d) G(i, j + 1, 0)$$

$$- (1 - x'_d)(1 - y'_d) G(i + 1, j + 1, 0) \Big)^2$$

$$= \arg\min_{x'_d, y'_d} \sum (G_t + G_x x'_d + G_y y'_d + G_c x'_d y'_d)^2$$

for $-1 \leq x'_d, y'_d \leq 1$. G_x, G_y, G_c and G_t are illustrated in Figure 3-6. G_x, G_y, and G_t are estimators for gradients in x, y, and t respectively, and are equivalent to the derivative estimators of Horn and Schunck (1981). G_c, which results from $x'_d y'_d$ terms in the bilinear interpolation (Equation 3.5), represents a cross-derivative term that is not typically included in optical flow formulations.

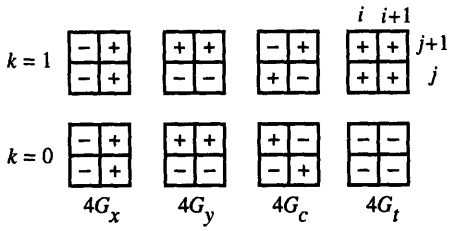


Figure 3-6: Derivative estimators for Example 4. G_x, G_y, and G_t are equivalent to the derivative estimators of Horn and Schunck(Horn and Schunck, 1981). G_c represents a cross-derivative term that is not typically included in optical flow formulations.

3.6 Equivalence

In the previous section, we showed examples of equivalent block matching and optical flow algorithms for estimating subpixel motions. In this section, we show that the equivalence is general.

Theorem 1 *All two-dimensional, block-matching algorithms that use bilinear interpolation and a mean-square error measure can be cast into an equivalent optical-flow representation containing both first-order and cross-derivative estimators.*

Proof: Substitute the bilinear approximations given in Equation 3.5 into Equation 3.4, replace x_1 with $x_0 + x'_d$ and y_1 with $y_0 + y'_d$, and set $q = 2$. The resulting equation can be re-written as

$$(\widehat{x_d}, \widehat{y_d}) = \arg\min_{x'_d, y'_d} \sum_{i,j} (\alpha + \beta x'_d + \gamma y'_d + \delta x'_d y'_d)^2$$

where neither α, β, γ, nor δ depend on x'_d or y'_d. Setting $G_t = \alpha, G_x = \beta, G_y = \gamma$, and $G_c = \delta$ completes the proof.

Theorem 2 *All two-dimensional, optical-flow algorithms using only first-order derivative estimators and the first-order cross-term can be cast into equivalent block-matching algorithms.*

Proof: We prove this theorem by constructing an interpolating function F that when substituted into Equation 3.4 with $q = 2$ gives rise to the same minimization

problem that was given in Equation 3.3. This construction is not unique; we give two possible constructions below.

Construction 1: Separate the terms in Equation 3.3 that depend on first image from those that depend on the second image by defining

$$\begin{aligned} G_x^1(i,j) &= 2G_x(i,j,1) & G_x^0(i,j) &= 2G_x(i,j,0) \\ G_y^1(i,j) &= 2G_y(i,j,1) & G_y^0(i,j) &= 2G_y(i,j,0) \\ G_c^1(i,j) &= 4G_c(i,j,1) & G_c^0(i,j) &= -4G_c(i,j,0) \\ G_t^1(i,j) &= G_t(i,j,1) & G_t^0(i,j) &= -G_t(i,j,0). \end{aligned}$$

Substitute

$$\begin{aligned} G_x &= (G_x^1 + G_x^0)/2 & G_c &= (G_c^1 - G_c^0)/4 \\ G_y &= (G_y^1 + G_y^0)/2 & G_t &= G_t^1 - G_t^0 \end{aligned}$$

into Equation 3.3. Equation 3.3 is now equivalent to Equation 3.4 with $q = 2$ if $E(x + x_1, y + y_1, 1)$ and $E(x + x_0, y + y_0, 0)$ are approximated using the following interpolating functions:

$$F(i + x_1, j + y_1, 1) = G_x^1 x_1 + G_y^1 y_1 + G_c^1 x_1 y_1 + G_t^1$$

$$F(i + x_0, j + y_0, 0) = G_x^0 x_0 + G_y^0 y_0 + G_c^0 x_0 y_0 + G_t^0$$

where $x_1 = x'_d/2 = -x_0$ and $y_1 = y'_d/2 = -y_0$. These interpolation equations (and therefore the constructed block matching algorithm) are valid for $0 \leq x'_d, y'_d \leq 2$.

Construction 2: Separate the terms in Equation 3.3 that depend on first image from those that depend on the second image by defining

$$
\begin{aligned}
G_x^1(i,j) &= 2G_x(i,j,1) - 2G_c(i,j,1) \\
G_y^1(i,j) &= 2G_y(i,j,1) - 2G_c(i,j,1) \\
G_c^1(i,j) &= 4G_c(i,j,1) \\
G_t^1(i,j) &= G_t(i,j,1) - G_x(i,j,1) - G_y(i,j,1) + G_c(i,j,1) \\
G_x^0(i,j) &= 2G_x(i,j,0) + 2G_c(i,j,0) \\
G_y^0(i,j) &= 2G_y(i,j,0) + 2G_c(i,j,0) \\
G_c^0(i,j) &= -4G_c(i,j,0) \\
G_t^0(i,j) &= -G_t(i,j,0) - G_x(i,j,0) - G_y(i,j,0) - G_c(i,j,0).
\end{aligned}
$$

Substitute

$$
\begin{aligned}
G_x &= (G_x^1 + G_x^0)/2 + (G_c^1 + G_c^0)/4 \\
G_y &= (G_y^1 + G_y^0)/2 + (G_c^1 + G_c^0)/4 \\
G_c &= (G_c^1 - G_c^0)/4 \\
G_t &= G_t^1 - G_t^0 + (G_x^1 - G_x^0)/2 + (G_y^1 - G_y^0)/2 + (G_c^1 - G_c^0)/4
\end{aligned}
$$

into Equation 3.3. Equation 3.3 is now equivalent to Equation 3.4 with $q = 2$ if $E(x + x_1, y + y_1, 1)$ and $E(x + x_0, y + y_0, 0)$ are approximated using the following interpolating functions:

$$F(i + x_1, j + y_1, 1) = G_x^1 x_1 + G_y^1 y_1 + G_c^1 x_1 y_1 + G_t^1$$

$$F(i + x_0, j + y_0, 0) = G_x^0 x_0 + G_y^0 y_0 + G_c^0 x_0 y_0 + G_t^0$$

where $x_1 = (1 + x'_d)/2$, $x_0 = (1 - x'_d)/2$, $y_1 = (1 + y'_d)/2$, and $y_0 = (1 - y'_d)/2$. These interpolation equations (and therefore the constructed block matching algorithm) are valid for $-1 \leq x'_d, y'_d \leq 1$.

3.6.1 Example 2, revisited

In Example 2 we showed that a block matching algorithm in which linear interpolation is used to shift each of two images by the same amount in opposite directions is

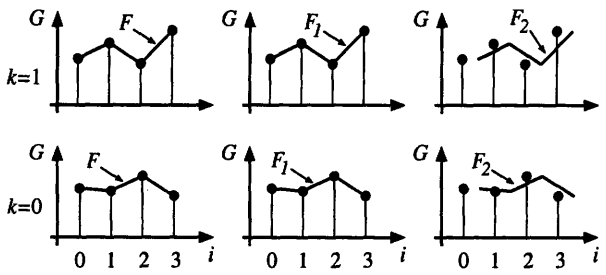


Figure 3-7: Interpolating functions for Example 2. The dots in each panel illustrate brightness $G(i,k)$ as a function of pixel i. The top row shows brightnesses at time index $k=1$; the bottom row, at time index $k=0$. The left column shows the linear interpolating function F upon which Equation 3.8 is based. The center column shows the interpolating function F_1 resulting from Construction 1 (Equation 3.10). The right column shows the interpolating function F_2 resulting from Construction 2 (Equation 3.11).

equivalent to an optical flow algorithm with the derivative estimators shown in Figure 3-4. In this section, we start with that optical flow algorithm and use Theorem 2 to generate two equivalent block matching algorithms.

Applying Construction 1 to Equation 3.9 yields the interpolation equations

$$F_1(i+x'_d/2,1) = (G(i+1,1)-G(i,1))x'_d/2+G(i,1)$$

$$F_1(i-x'_d/2,0) = (G(i-1,0)-G(i,0))x'_d/2+G(i,0). \tag{3.10}$$

Alternatively, we could apply Construction 2 to obtain

$$F_2(i+1/2+x'_d/2,1)=(G(i+1,1)-G(i,1))x'_d/2+G(i,1)$$

$$F_2(i+1/2-x'_d/2,0)=(G(i-1,0)-G(i,0))x'_d/2+G(i,0). \tag{3.11}$$

These interpolators are illustrated in Figure 3-7 along with the original linear interpolating function upon which Equation 3.8 was based. In this example, Construction 1 leads to the same linear interpolating function that was used to formulate the example. However Construction 2 leads to an unusual interpolating function: F_2 is in fact a shifted version of of F_1.

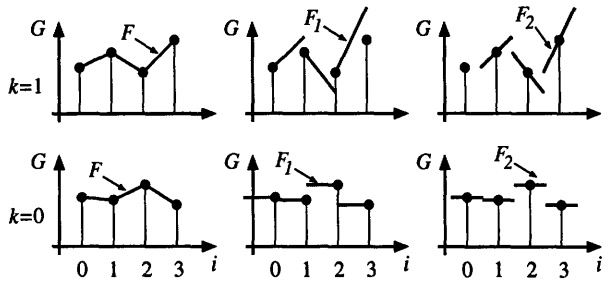


Figure 3-8: Interpolating functions for Example 1. The panels are analogous to those in Figure 3-7.

3.6.2 Example 1, revisited

We now apply the constructions to the optical flow algorithm developed in Example 1. The derivative estimators are shown in Figure 3-3, and application of Construction 1 to Equation 3.7 yields the interpolation equations

$$F_1(i + x'_d/2, 1) = 2(G(i+1,1) - G(i,1))x'_d/2 + G(i,1)$$

$$F_1(i - x'_d/2, 0) = G(i,0).$$

Alternatively, we could apply Construction 2 to obtain

$$F_2(i + 1/2 + x'_d/2, 1) = 2(G(i+1,1) - G(i,1))x'_d/2 + G(i,1)$$

$$F_2(i + 1/2 - x'_d/2, 0) = G(i,0).$$

These interpolators are illustrated in Figure 3-8 along with the original linear interpolating function upon which Equation 3.6 was based. As in the previous example, the two constructions lead to interpolating functions that are related by a shift of one half the pixel spacing. However, in this example, neither construction leads to the original linear interpolating function.

3.7 Extensions and limitations

3.7.1 Other measures of image difference

The block matching algorithms we have considered compare shifted images in a magnitude-squared sense (i.e., $q = 2$ in Equation 3.4). However other types of comparisons, such as mean absolute difference (Gharavi and Mills, 1990; Komarek and Pirsch, 1989), are often used. These other methods also produce an equivalent optical flow algorithm by modifying the exponent in Equation 3.3.

3.7.2 Other interpolators

Block matching algorithms based on higher order interpolators (Parker et al., 1983), such as bicubic splines, have equivalent optical flow algorithms which incorporate higher order derivative estimators.

3.7.3 Relaxing the rigid body constraint

One method (Bergen et al., 1992) for estimating non-rigid body motion is to (1) partition the images into subregions, (2) estimate "local" motions within the subregions using a rigid body assumption, and (3) combine the local estimates to form a "global" motion estimate that is not a simple translation. Our results still apply for the "local" estimates in such a formulation.

3.7.4 Equivalence using only first-order optical flow

Theorem 1 generally produces an optical flow algorithm with a cross term. Therefore it is not generally possible to cast an arbitrary block matching algorithm into the strictly first-order optical flow framework used by Horn and Schunck (Horn and Schunck, 1981). However, block matching based on linear (not bilinear) interpolation will always fall into Horn and Schunck's first-order optical flow framework.

3.7.5 Casting into block matching with bilinear interpolation

The interpolators for the block matching algorithms produced by the Constructions for Theorem 2 are in general not bilinear. However, our four examples can indeed be placed into a bilinear interpolation block matching framework since the derivative estimators at each time are at most two pixels wide. It is easy to show that wider derivative estimators cannot be cast into a block matching algorithm that uses bilinear interpolation.

3.8 Discussion

Block matching and optical flow algorithms appear to be completely different (Figures 3-1 and 3-2), and the accuracy of subpixel estimates based on the two methods have been reported to be different (Tian and Huhns, 1986). However, we have shown that if you choose corresponding algorithms from the two families, the estimators are identical.

We suspect that there is a very basic reason for this equivalence. Both algorithms must base subpixel estimates on an interpolation approximation to sample points in the discrete image. As long as the order of approximation used by block matching (e.g., bilinear interpolation) is equal to that used by optical flow, (e.g., through the bilinear term in the Taylor's series) it is not surprising that the two families of algorithms are equivalent.

Since block matching and optical flow methods generate numerically equivalent estimates for subpixel displacements, other features may suggest choosing one over the other. For example, the implementation of the optical flow algorithm may require significantly fewer operations than an "equivalent" block matching algorithm. On the other hand, block matching will tend to have better performance in terms of estimation accuracy over suprapixel displacements.

3.9 Acknowledgments

We would like to thank Jungwoo Lee at David Sarnoff Research Center for suggesting the use of block matching for subpixel motion estimation. This work was supported by NIH and NSF.

Chapter 4

Statistics of Subpixel Registration Algorithms Based on Piecewise Linear Interpolators

4.1 Abstract

Analysis of subpixel registration algorithms based on piecewise linear interpolators (e.g., spatio-temporal gradients or block-matching) shows that registration errors contain not only a random component from noise in the images but also a systematic bias component. The bias component is present even if the images are noise-free and its magnitude exceeds that of the random component for almost all imaging conditions.

Bias depends on both image content and on the displacement between the images. For small displacements, bias increases nearly linearly with displacement. Based on this observation, we have developed a closed-form method to reduce the bias by approximately a factor of four. This improvement is large compared to the improvement that can be obtained using averaging and/or temporal oversampling.

4.2 Keywords

registration, subpixel, statistics, bias, piecewise linear interpolation

4.3 Introduction

In this paper, we examine statistical properties of registration algorithms based on first-order spatio-temporal gradients (Horn and Schunck, 1981; Horn and Weldon, Jr., 1988; Horn, 1986; Tian and Huhns, 1986; Aggarwal and Nandhakumar, 1988). Unlike edge detection methods (Hachincha and Simon, 1988) or point-correspondence methods (Roach and Aggarwal, 1980; Weng et al., 1993), algorithms based on spatio-temporal gradients can be used to register arbitrary images: no prior knowledge of the target is necessary. Furthermore, these algorithms take advantage of information from all parts of the image (rather than just at edges or other features), which is particularly important for the registration of noisy images. For subpixel registration, these gradient-based algorithms are equivalent (Davis et al., 1995) to finding the shift that minimizes the sum of squared differences between shifted versions of linearly interpolated images (Netravali and Haskell, 1988; Giorda and Racciu, 1975). Thus results in this paper apply equally to block-matching algorithms that use piecewise-linear interpolation to achieve subpixel accuracy.

We show that registration errors consist of a random component that results from image noise and a generally much larger systematic bias component that is intrinsic to the registration algorithm. We show that bias results from systematic errors in using a first-order Taylor's expansion of the image as a function of displacement. Using a first-order Taylor's expansion of the registration estimate as a function of displacement, we develop a closed-form method of reducing bias. When the new algorithm is applied to video images from a light microscope, registration errors are typically less than 1/200 pixel, without averaging.

These results have potential applications in diverse registration tasks, ranging from lithographic mask alignment (Douglas, 1993) to satellite imagery (Pritt, 1994).

We have also applied these methods to analyze motions with nanometer precision. Stroboscopic illumination is used to obtain a sequence of stop-action video images, and the displacements between successive images are determined. Results have provided the first direct measurements of sound-induced motions of sensory receptor cells and their overlying tectorial membrane in the inner ear (Davis and Freeman, 1995). Similar techniques have also been used to measure motions of man-made microelectromechanical systems (MEMS) (Freeman and Davis, 1996) and could provide the basis for a "motion probe-station" for design and fabrication of MEMS.

In this paper, we present the registration algorithm based on first-order spatiotemporal gradients and derive our bias-reducing method. We derive closed-form expressions for the performance of these two algorithms for a family of sinusoidal images. We show computer simulations for four different images, and characterize the effect of imaging noise on registration estimates. In the companion paper (Davis and Freeman, 1997a), we demonstrate that our simulations are realistic by measuring the motion of a calibrated displacement source with nanometer accuracy.

4.4 Registration algorithms

We consider two registration algorithms. The first is gradient based (Horn and Schunck, 1981; Horn and Weldon, Jr., 1988). The second compensates for bias in the first by repeated (four times) application of the first.

4.4.1 Gradient-based algorithm

Let $A(x, y)$ and $B(x, y)$ represent misaligned images of the same target. We assume that the misalignment is purely translational, so that

$$A(x - \frac{d_x}{2}, y - \frac{d_y}{2}) = B(x + \frac{d_x}{2}, y + \frac{d_y}{2})$$

where (d_x, d_y) represents the displacement between the two images. If the displacement is small, we can expand the offset images in Taylor's series and retain just the

linear terms so that

$$A(x,y) - \frac{d_x}{2}\frac{\partial A}{\partial x} - \frac{d_y}{2}\frac{\partial A}{\partial y} = B(x,y) + \frac{d_x}{2}\frac{\partial B}{\partial x} + \frac{d_y}{2}\frac{\partial B}{\partial y}. \tag{4.1}$$

The images $A(x,y)$ and $B(x,y)$ are sampled by the camera to give arrays of pixel values $A[i,j]$ and $B[i,j]$. We approximate the images and their partial derivatives at points midway between pixels. For example, $A\left((i+\frac{1}{2})\Delta, (j+\frac{1}{2})\Delta\right)$ is approximated by $\overline{A}[i,j]$ where

$$\overline{A}[i,j] = (A[i,j] + A[i+1,j] + A[i,j+1] + A[i+1,j+1])/4$$

and Δ represents the pixel spacing. The partial derivatives $\partial A/\partial x$ and $\partial A/\partial x$ at $((i+\frac{1}{2})\Delta, (j+\frac{1}{2})\Delta)$ are approximated by

$$\begin{aligned} \overline{A}_x[i,j] &= (A[i+1,j] - A[i,j] + A[i+1,j+1] - A[i,j+1])/2\Delta \\ \overline{A}_y[i,j] &= (A[i,j+1] - A[i,j] + A[i+1,j+1] - A[i+1,j])/2\Delta. \end{aligned} \tag{4.2}$$

Substituting these approximations into the Taylor's expansion yields

$$\begin{aligned} \overline{A}[i,j] - \overline{B}[i,j] = \quad & d_x(\overline{A}_x[i,j] + \overline{B}_x[i,j])/2 \\ & + d_y(\overline{A}_y[i,j] + \overline{B}_y[i,j])/2 \end{aligned}$$

which can be regarded as an array i,j of equations with two unknowns, d_x and d_y. Because data from a camera are corrupted by noise, the constraint equations are generally inconsistent. Such equations can be solved using least squares (Horn and Weldon, Jr., 1988), and the solution can be expressed in closed form:

$$\begin{bmatrix} \widehat{d}_x \\ \widehat{d}_y \end{bmatrix} = -\begin{bmatrix} \sum\sum G_x G_x & \sum\sum G_x G_y \\ \sum\sum G_x G_y & \sum\sum G_y G_y \end{bmatrix}^{-1} \begin{bmatrix} \sum\sum G_x G_t \\ \sum\sum G_y G_t \end{bmatrix} \tag{4.3}$$

where the sums are over i and j and

$$\begin{aligned} G_x[i,j] &= \left(\overline{A}_x[i,j] + \overline{B}_x[i,j]\right)/2 \\ G_y[i,j] &= \left(\overline{A}_y[i,j] + \overline{B}_y[i,j]\right)/2 \\ G_t[i,j] &= \overline{B}[i,j] - \overline{A}[i,j]. \end{aligned} \tag{4.4}$$

Hereafter, we refer to these equations as the "gradient-based algorithm" and $(\widehat{d}_x, \widehat{d}_y)$ as the "gradient-based estimate" of displacement.

4.4.2 Linear bias compensation

Gradient-based estimates have systematic errors (Section 4.6). The relation between gradient-based estimates $(\widehat{d}_x,\ \widehat{d}_y)$ and the unknown displacement $(d_x,\ d_y)$ can be approximated by a simple linear model,

$$\begin{aligned} \widehat{d}_x &= \alpha_1 d_x + \beta_1 d_y \\ \widehat{d}_y &= \alpha_2 d_x + \beta_2 d_y \end{aligned} \tag{4.5}$$

that can be used to predict d_x and d_y from the biased estimates if the four parameters α_1, α_2, β_1, and β_2 were known. We have developed a method to estimate those parameters directly from the two images being registered. The method exploits integer-pixel shifts of the images, which are computationally simple and introduce no shifting errors. First, we apply the gradient-based algorithm to determine an estimate $(\widehat{d}_x, \widehat{d}_y)$ of the displacement between the first and second images. Then we use $\widehat{d}_x$ and $\widehat{d}_y$ to define three additional registration problems: (1) If $\widehat{d}_x > 0$ let $s_x = +1$, otherwise let $s_x = -1$. Shift the first image s_x pixels in the $+x$ direction and use the gradient-based algorithm to make a second estimate $(\widehat{d}^x_x, \widehat{d}^x_y)$, which will satisfy

$$\begin{aligned} \widehat{d}^x_x &= \alpha_1(d_x - s_x) + \beta_1 d_y \\ \widehat{d}^x_y &= \alpha_2(d_x - s_x) + \beta_2 d_y. \end{aligned} \tag{4.6}$$

(2) Analogously, use $\widehat{d}_y$ to define a shift s_y in the $+y$ direction and use the gradient-based algorithm to make a third estimate $(\widehat{d}_x^y, \widehat{d}_y^y)$ satisfying

$$\begin{aligned} \widehat{d}_x^y &= \alpha_1 d_x + \beta_1 (d_y - s_y) \\ \widehat{d}_y^y &= \alpha_2 d_x + \beta_2 (d_y - s_y) \end{aligned} \tag{4.7}$$

(3) Apply the gradient-based algorithm after shifting the first image s_x in the $+x$ direction *and* s_y in the $+y$ direction to obtain

$$\begin{aligned} \widehat{d}_x^{xy} &= \alpha_1 (d_x - s_x) + \beta_1 (d_y - s_y) \\ \widehat{d}_y^{xy} &= \alpha_2 (d_x - s_x) + \beta_2 (d_y - s_y). \end{aligned} \tag{4.8}$$

If the original motion was subpixel, then (ignoring errors due to noise) all three of the additional problems involve estimating subpixel motions.

Equations 4.5-4.8 represent eight equations with six unknowns: four unknown model parameters, α_1, α_2, β_1, and β_2; and two unknown displacements, d_x and d_y. Taking pairwise differences (i.e., Equations 4.5 $-$ Equations 4.6 then Equations 4.5 $-$ Equations 4.7 then Equations 4.5 $-$ Equations 4.8) it follows that

$$\begin{aligned} \widehat{d}_x &= \widehat{d}_x^x + s_x\alpha_1 = \widehat{d}_x^y + s_y\beta_1 = \widehat{d}_x^{xy} + s_x\alpha_1 + s_y\beta_1 \\ \widehat{d}_y &= \widehat{d}_y^x + s_x\alpha_2 = \widehat{d}_y^y + s_y\beta_2 = \widehat{d}_y^{xy} + s_x\alpha_2 + s_y\beta_2. \end{aligned}$$

Any three of the four registration problems can be used to determine the model parameters as well as the displacements d_x and d_y. We use the three with the smallest displacement magnitude, i.e., $(\widehat{d}_x)^2 + (\widehat{d}_y)^2$, $(\widehat{d}_x^x)^2 + (\widehat{d}_y^x)^2$, $(\widehat{d}_x^y)^2 + (\widehat{d}_y^y)^2$, and $(\widehat{d}_x^{xy})^2 + (\widehat{d}_y^{xy})^2$. We refer to estimates $\bar{d}_x$ and $\bar{d}_y$ based on this algorithm as estimates with "linear bias compensation".

4.5 Methods

The subpixel registration algorithms were tested by applying them to estimate the displacements between a test image and a computer-generated shift of the same image.

Computer-generated noise processes were applied to each image to simulate the effects of imaging noise.

4.5.1 Images

We analyzed registration performance for pairs of four test images, and a family of sinusoidal images. The first pair was a small dark spot on a bright background, which was implemented as a radially-symmetric raised-cosine function (Hanning window),

$$A[i,j] = \begin{cases} E_0(1 - \frac{1+\cos 2\pi r_1/5}{3.8}) & \text{for } |r_1| = \sqrt{(i\Delta - c_x)^2 + (j\Delta - c_y)^2} < 2.5 \\ E_0 & \text{otherwise} \end{cases}$$

$$B[i,j] = \begin{cases} E_0(1 - \frac{1+\cos 2\pi r_2/5}{3.8}) & \text{for } |r_2| = \sqrt{(i\Delta - c_x - d_x)^2 + (j\Delta - c_y - d_y)^2} < 2.5 \\ E_0 & \text{otherwise} \end{cases}$$

where r represents the distance from the center of the simulated bead (c_x, c_y) to the image point $[i,j]$ and (d_x, d_y) is the displacement of the scene between the images. This "simulated bead" image is easily parameterized, and therefore results obtained with this image are easily reproduced. To avoid artifacts that could result if the simulated beads were always aligned with the pixel grid, the bead's center point in the first image is chosen randomly within the central 14×14 pixel portion of the simulated image. The center point need not fall on pixel centers and is represented in floating-point. The center point for the second simulated bead is offset from that of the first by the desired test shift. To characterize the effect of signal energy on registration, tests are also done using images with multiple simulated beads. For the first image in each pair, seven beads are placed randomly in the central 14×14 pixel portion of the image, subject to the constraint that none of the beads touch one another. For the second image, each center is shifted by the same test displacement. The signal energy in these pictures is 7 times or 8.5 dB greater than that in the single bead images.

The second and third image pairs are representative of our application of subpixel registration to determining motions of sensory receptor cells in the inner ear and of

microelectromechanical systems (MEMS) using video microscopy. The fourth image pair is a synthetic aperture radar (SAR) image of the Galapagos Islands, taken on board the space shuttle.[1] Shifts between measured images were simulated by (1) computing the 256×256-point Fourier transform of the original 256×256 pixel image; (2) multiplying by the pure-phase transfer function

$$H(k_x, k_y) = e^{j2\pi(x_d k_x + y_d k_y)/256}$$

where x_d and y_d represent the desired displacements in the x and y directions respectively; (3) inverse transforming the product; and (4) cropping the resulting image to be 32×32 pixels. Cropping the large FFT reduced the Gibbs effects created by discontinuities at the image's boundary.

The family of sinusoidal image pairs is

$$\begin{aligned} A[i,j] &= E_0 + E_1 \cos \Omega i \\ B[i,j] &= E_0 + E_1 \cos \Omega (i - d_x) \end{aligned} \tag{4.9}$$

where d_x is the displacement of the scene in the second image with respect to the first, Ω is the spatial frequency of the sinusoids, and $E_0 \geq E_1$ since the brightness cannot be negative.

4.5.2 Simulated noises

All simulated images include simulations of fixed-pattern noise, shot noise, and quantization errors in the A/D converter. Each original image and shifted image is represented by an array of high-precision floating-point numbers that represent the average number of electrons for each pixel (not necessarily an integer) in the absence of fixed-pattern or shot noise. Fixed-pattern noise is simulated by multiplying the pixel values by random gain factors taken from a Gaussian distribution (Press et al., 1992). The mean of the distribution is set to one and the standard deviation is used as a pa-

[1] Image (P-43899) courtesy of the Jet Propulsion Laboratory, Pasadena, CA. http://www.jpl.nasa.gov/archive/mpe2.html.

rameter to control the amount of fixed-pattern noise. Shot noise is simulated using a Poisson pseudo-random number generator (Press et al., 1992) at each pixel, converting the average number of electrons to an integer that represents the number of electrons in one randomly chosen instance. The number of electrons is then divided by 46.1 — the average number of electrons required to increment the A/D output by one (Photometrics, 1993) — and truncated to obtain an integer representing the output of the A/D.

Simulations with different fixed-pattern noise levels are accomplished by changing the standard deviation of the Gaussian noise process. Simulations with different shot-noise levels are accomplished by changing the amplitude scaling that represents the lighting condition. For ease of comparison, both fixed-pattern and shot noises are expressed in dB. The dB value reported for fixed-pattern noise represents the ratio of the standard deviation of the Gaussian distribution to its mean. The dB value reported for shot noise represents the ratio of the standard deviation of the Poisson distribution to its mean. If the reported dB values for fixed-pattern and shot noises are equal, then the energy in the noises are also equal.

All simulations were repeated 100 times using 100 image pairs with the same fixed-pattern noise but with independent pseudo-random shot noises and quantization errors. For each pair, consisting of an original and a shifted image, shift estimates were generated. Bias was estimated as the sample mean of the difference between the estimated and actual shift. The standard deviation was estimated by taking the square root of the sample variance.

4.6 Bias in gradient-based estimates

4.6.1 Sinusoidal images

We now derive a closed-form expression for the gradient-based estimate as a function of the known displacement d_x and of the spatial frequency Ω for the sinusoidal images. Substituting Equations 4.9 into Equations 4.4 results in a set of equations that (after

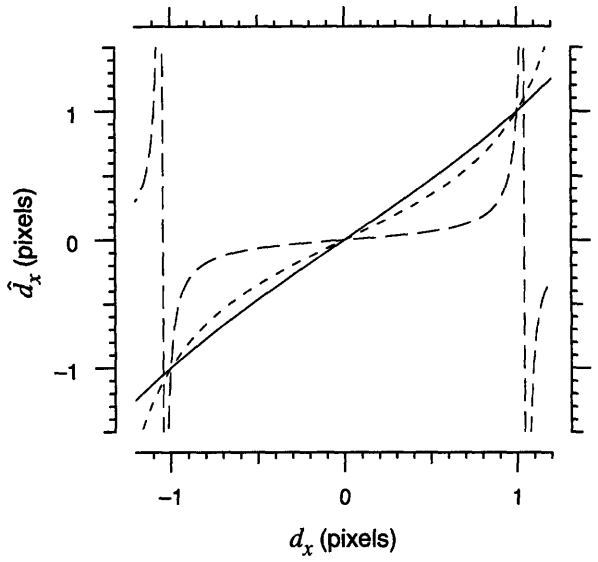


Figure 4-1: Registration estimates for a sinusoidal image displaced in one dimension. The solid, dotted, and dashed curves illustrate gradient-based estimates of displacement (Equation 4.3) for sinusoidal images with frequencies Ω equal to 1, 2, and 3, respectively. The Nyquist frequency is $\Omega = \pi$. The curves are computed from Equation 4.10.

some trigonometric simplification) can be written as

$$\begin{aligned} G_i[i,j] &= -2E_1 \sin\tfrac{\Omega}{2} \sin\tfrac{\Omega(2i+1-d_x)}{2} \cos\tfrac{\Omega d_x}{2} \\ G_j[i,j] &= 0 \\ G_k[i,j] &= 2E_1 \sin\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+1-d_x)}{2} \cos\tfrac{\Omega}{2}. \end{aligned}$$

In this special case in which $G_j[i,j] = 0$, Equation 4.3 can be simplified to give

$$\hat{d}_x = -\frac{\sum\sum G_i G_k}{\sum\sum G_i G_i}$$

so

$$\hat{d}_x = \frac{\tan(\Omega d_x/2)}{\tan\Omega/2}. \tag{4.10}$$

Note that for this special case of one sinusoid in the image, the estimate is independent of the summation region, as long as the summation region has at least one point where $G_i[i,j] \neq 0$.

Plots of the relation between the estimate $\hat{d}_x$ and the imposed displacement d_x are illustrated in Figure 4-1 for images with three different spatial frequencies. The estimated displacement is exactly equal to the imposed displacement when the imposed displacement is equal to -1, 0, and 1 pixel for all spatial frequencies. For frequencies $\Omega < 1$, $\hat{d}_x \approx d_x$. However, there is increasing difference between the estimate and

imposed displacement for higher frequencies. Gradient-based estimates are systematically too small for subpixel displacements, too large for displacements between 1 and π/Ω (half the wavelength), and are periodic with a period equal to the wavelength of the sinusoid.

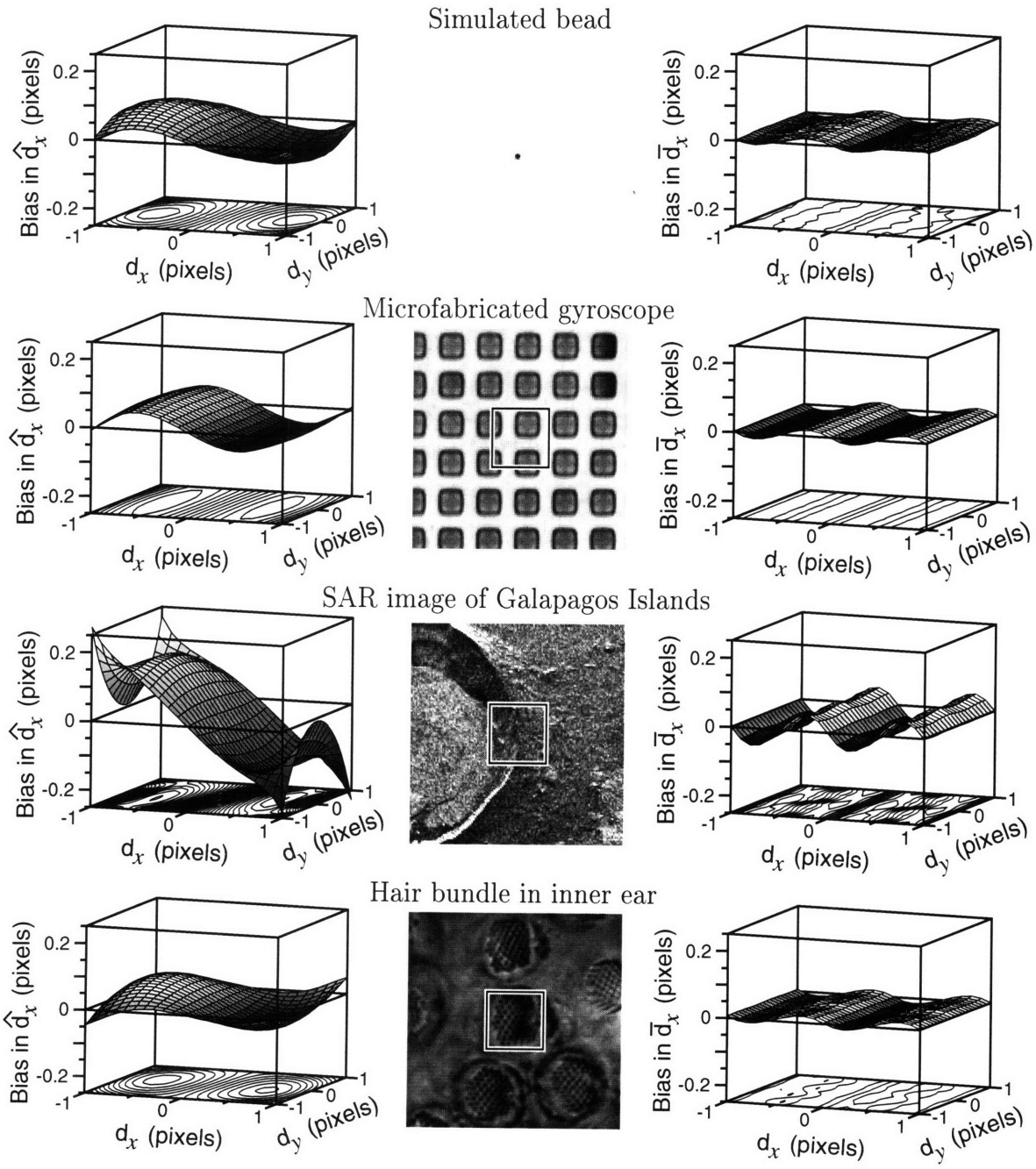
We refer to the difference between the estimated displacement and the imposed displacement $(\widehat{d}_x - d_x)$ as bias. As is evident from Figure 4-1, bias depends not only on the imposed displacement, but also on the image: bias functions for images with different spatial frequencies differ.

4.6.2 Two-dimensional images

Table 4.1 shows the two dimensional bias function for simulated motions of a simulated bead. For $d_x = 0$, the bias in the estimate of the x component of displacement $(\widehat{d}_x)$ is zero independent of d_y. For most subpixel displacements, the bias is positive for negative d_x and negative for positive d_x. Thus the gradient-based algorithm typically underestimates the imposed subpixel displacements—a trend that was also seen for sinusoidal images (Figure 4-1). For small values of $|d_x|$ and $|d_y|$, the bias in the estimate of the x component of displacement is a nearly linear function of d_x and changes little with d_y. However, for large displacements d_x and d_y, the bias depends on both d_x and d_y, and the relation between bias and displacement is quite nonlinear.

The top bias function illustrated in Table 4.1 results from images of a very small target (bead diameter of 5 pixels in a 32×32 image) containing a relatively large fraction of energy at high spatial frequencies. One might expect that bias functions for images with more energy in low spatial frequencies will tend to be smoother functions of d_x and d_y. We used computer-simulated motions (see Methods) of three measured images to test this hypothesis (Table 4.1). Results indicate that the two images taken with our video microscopy system (the gyroscope and hair bundle) have bias functions that are similar to that of the bead, and the measured SAR image has significantly more bias. Nevertheless, all the bias functions are nearly linear functions of the displacement (d_x, d_y) for small displacements, which is the hypothesis for linear bias compensation.

Table 4.1: Bias in estimates of displacement without (*Left*) and with (*Right*) linear bias compensation for (1) simulated motions of a simulated bead and (2) simulated motions of a measured image of a gyroscope, SAR, and hair bundle. For the measured images, only the center 32×32 pixels (enclosed by the box) were used—the larger region is shown for clarity. Displacements were simulated for 21×21 values of (d_x, d_y). Bias was calculated as the difference between the average estimate of the x component of displacement ($\widehat{d_x}$ or $\bar{d}_x$) and the imposed displacement d_x. The surface represents the bias as a function of the imposed displacements d_x and d_y. The contour lines in the lower part of the figure connect points in the d_x,d_y plane that have bias equal to an integral multiple of 0.01. Shot noise was -50 dB and fixed-pattern noise was -50 dB.



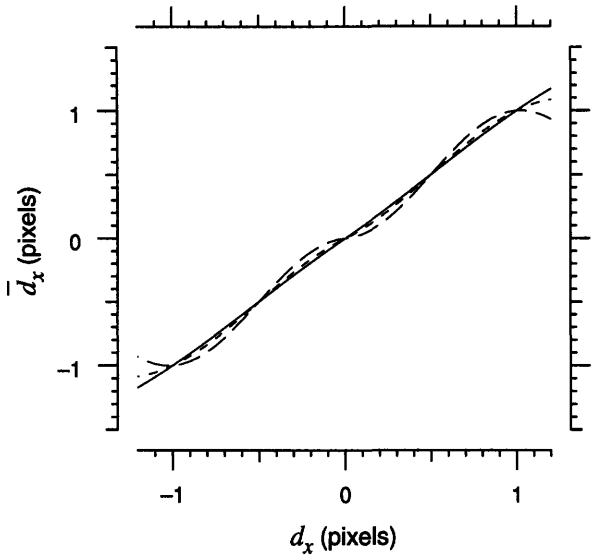


Figure 4-2: Motion estimates with linear bias compensation for a sinusoidal image. The solid, dotted, and dashed curves illustrate estimates of displacement with linear bias compensation for sinusoidal images with frequencies Ω equal to 1, 2, and 3, respectively. The curves are computed from Equation 4.11.

4.7 Bias after linear bias compensation

4.7.1 Sinusoidal images

One can derive a closed-form expression for the estimate with linear bias compensation ($\bar{d}_x$) when the image consists of a single sinusoid. An expression for $\hat{d}_x$ is given in Equation 4.10. If $d_x > 0$, then $\hat{d}_x > 0$ and therefore

$$\hat{d}_x^x = \frac{\tan \Omega(d_x - 1)/2}{\tan \Omega/2}.$$

The values of $\hat{d}_x$ and $\hat{d}_x^x$ determine α_1,

$$\alpha_1 = \frac{\tan \Omega d_x/2 - \tan \Omega(d_x - 1)/2}{\tan \Omega/2}.$$

Since $\alpha_2 = \beta_1 = \beta_2 = 0$, it follows that

$$\bar{d}_x = \frac{\tan \Omega d_x/2}{\tan \Omega d_x/2 - \tan \Omega(d_x - 1)/2} \tag{4.11}$$

for $d_x > 0$. Values of $\bar{d}_x$ for $d_x < 0$ follow from similar reasoning. Results for positive and negative values of d_x are plotted in Figure 4-2.

Figure 4-1 and Equation 4.10 both show that the gradient-based estimate $\hat{d}_x$ is

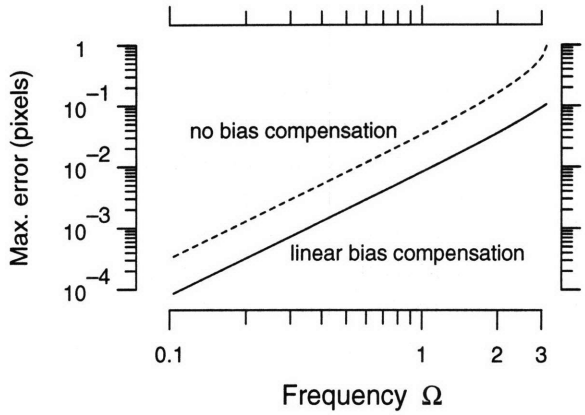


Figure 4-3: Maximum errors in estimates of the displacements of a sinusoidal image with and without linear bias compensation. The dashed line illustrates the maximum subpixel bias in gradient-based estimates as a function of the spatial frequency Ω of the image. This maximum subpixel bias is taken as the biggest difference between $\hat{d}_x$ and d_x (Equation 4.10) for any $|d_x| < 1$. The solid line represents the analogous relation after linear bias compensation (i.e., the biggest difference between $\bar{d}_x$ and d_x in Equation 4.11).

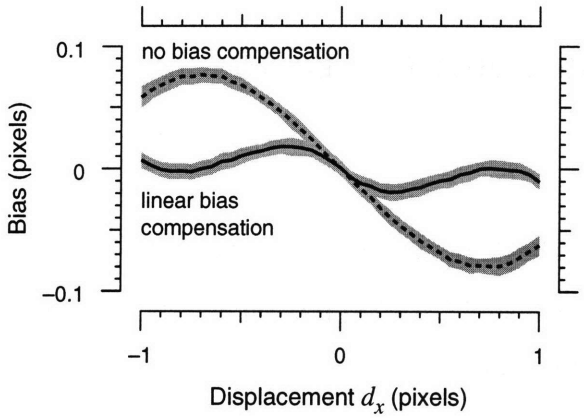


Figure 4-4: Bias in motion estimates: simulated motions of a simulated bead, default condition. Displacements are simulated at 41 values of d_x. Shot and fixed-pattern noise are both −50 dB. The dashed and solid lines indicate the bias of the two algorithms. The shaded regions around the lines are ±1 standard deviation.

equal to the imposed displacement d_x when $d_x = -1$, 0, and 1. Figure 4-2 shows that estimates with linear bias compensation $\bar{d}_x$ are equal to d_x at two additional values of d_x near ± 0.5. Furthermore, the error in the estimate with compensation for bias is generally smaller than the error in the estimate without bias compensation. A more quantitative characterization of error in the estimate is given in Figure 4-3, which shows the relation between the maximum error that results for subpixel displacements ($|d_x| < 1$) as a function of spatial frequency. Linear bias compensation reduces the maximum subpixel bias by more than a factor of four. The improvement is even greater for spatial frequencies near the Nyquist frequency ($\Omega \approx \pi$).

4.7.2 Two-dimensional images

The right column in Table 4.1 shows two-dimensional bias functions for the linear bias corrected algorithm. The algorithm had very little bias near $0, \pm 1$, and ± 0.5. The maximum subpixel error for the four images dropped by a factor of 4.2 for the bead, 3.4 for the gyroscope, 5.8 for the SAR image, and 3.5 for the hair bundle (average: 4.2). Furthermore, as seen in the plots and the contour lines, the dependence on orthogonal motions has been greatly reduced. Linear bias compensation therefore can more independently measure displacement in the two directions. This reduction enables more accurate measurements of displacements in the presence of confounding orthogonal displacements.

4.8 Statistics of motion estimates

As shown in Table 4.1, the two dimensional bias function for the four images are qualitatively similar. Registration bias for the simulated bead is in the middle of the four. Because the simulated bead image (1) generates representative results, (2) is easily parameterized and therefore easily reproduced, and (3) simplifies changing the signal energy (adding more beads), we focus on this image to characterized registration performance as a function of signal energy, shot noise, and fixed-pattern noise. To explore this multi-dimensional parameter space, we vary one parameter at a time from a default condition that is representative of images from our video-microscope system: a 32×32 pixel image of a single simulated bead with fixed-pattern and shot noise at -50 dB. These noise levels were taken to approximate typical levels in high resolution systems. Shot noise in CCD cameras is determined by the ratio of the standard deviation to the mean number of electrons per pixel (Janesick et al., 1987) and is typically on the order of -50 dB. Fixed pattern noise for the Sony XC-77 camera has been estimated (Healey and Kondepudy, 1994) to be -49 dB.

Results for the default case are shown in Figure 4-4. The bias in both algorithms is larger than the variability at almost all imposed displacements. As with the sinusoidal image, results for both algorithms shown in Figure 4-4 have no bias at $d_x = 0$;

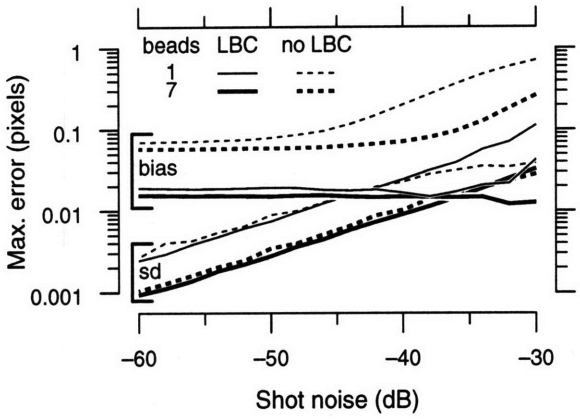


Figure 4-5: Effect of shot noise on statistics of motion estimates. Motion estimates were simulated as described in Figure 4-4 for images with 16 different levels of shot noise. We report the energy in the shot noise in dB relative to the mean energy in the image; therefore, the range of the abscissa corresponds to a 60 dB range in mean image brightness. Lines in this figure show the maximum bias and maximum standard deviations (sd) across imposed displacement d_x as a function of shot noise energy. Statistics are shown for four conditions as indicated in the legend: images with one and seven beads, with and without linear bias compensation (LBC).

however, neither algorithm reduces the bias to zero at $d_x = \pm 1$. Simulations without noise (not shown) show that the bias in both algorithms is zero for $d_x = \pm 1$ when the noise eliminated — i.e., the non-zero bias at $d_x = \pm 1$ in Figure 4-4 is a direct consequence of noise. Thus, image noise gives rise not only to variability in motion estimates but also to bias. This phenomenon is explored more fully in subsequent sections. Finally, notice that the bias in estimates with linear bias compensation is smaller than that without linear bias compensation for all imposed displacements. Linear compensation for bias reduces the maximum bias by a factor of approximately four.

4.8.1 Effect of shot noise

Figure 4-5 illustrates the effect of shot noise on simulated motions of a simulated bead. For all simulated conditions, increasing the shot noise increases the variability of the motion estimates. For shot noises smaller than -40 dB, the standard deviation for the motion estimates depends linearly on shot noise amplitude. Increasing the signal energy in the image by 8.5 dB (by changing the number of simulated beads from 1 to

7) decreases the standard deviation for the motion estimates by approximately 8 dB. These data are consistent with the idea that the standard deviation of the motion estimates is inversely proportional to the signal-to-noise ratio for the image.

When shot noise is small (< -50 dB), linear bias compensation reduces the bias by nearly a factor of four. Increasing the shot noise energy tends to increase the bias in motion estimates without bias compensation. This trend is reduced by linear bias compensation. Except for the lowest signal-to-noise conditions simulated (one bead, shot noise > -35 dB), the bias that remains after linear bias compensation is independent of the energy in the shot noise. The result is that linear bias compensation can reduce bias by more than a factor of 15 when shot noise is high (> -35 dB).

To understand the importance of these results for motion estimation algorithms, one can compare the relative magnitudes of the bias and standard deviations of the algorithms. For all energies of shot noise that were studied, the bias in estimates without bias compensation is significantly larger than the standard deviation. Thus, the systematic bias built into the gradient-based algorithm is greater than the variability in the estimates caused by noise *even at high noise levels.* This relation does not generally hold for estimates with linear bias compensation. The much smaller bias that remains after linear bias compensation is smaller than the standard deviation for noisy images (shot noise > -42 dB for one bead images; > -37 dB for seven bead images). However, for shot noise that is typical for our video microscopy system (-50 dB), bias is the most significant part of the total error in motion estimates (i.e., bias + standard deviation) even after linear bias compensation.

4.8.2 Effect of fixed-pattern (background) noise

Figure 4-6 illustrates effects of fixed-pattern noise on motion estimates — effects that are quite different from those of shot noise. The standard deviation of motion estimates tends to increase with increasing shot noise (Figure 4-5). In contrast, the variability tends to *decrease* with increasing fixed-pattern noise (Figure 4-6). More noise leads to estimates with less variability but bigger errors: bias increases very rapidly as fixed-pattern noise increases above -40 dB. Generally, high levels of fixed-

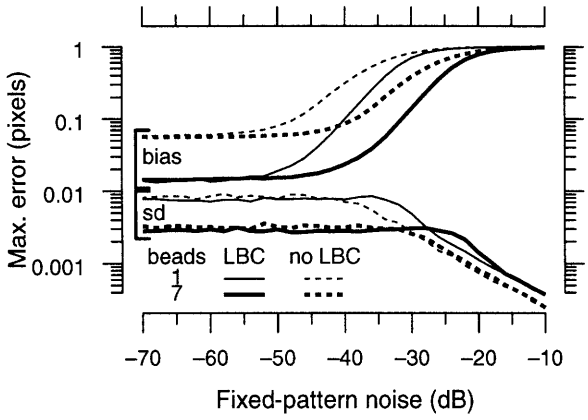


Figure 4-6: Effect of fixed-pattern noise on statistics of motion estimates. Motion estimates were simulated as described in Figure 4-4 for images with 31 different levels of fixed-pattern noise. We report the energy in the fixed-pattern noise in dB relative to the mean energy in the image. Statistics are illustrated as in Figure 4-5.

pattern noise cause motion estimators to *more consistently get the wrong answer.*

Small amounts of fixed-pattern noise (< -50 dB for one bead images, < -42 dB for seven bead images) have little effect on the statistics of motion estimates. Large amounts of fixed-pattern noise (> -30 dB for one bead images, > -22 dB for seven bead images) cause both algorithms to estimate 0 for all displacements: the maximum subpixel bias approaches one pixel. For small amounts of fixed-pattern noise, linear bias compensation reduced the bias by about a factor of four without making any significant change in standard deviation. The bias in estimates with and without linear bias compensation begin to increase at nearly the same levels of fixed-pattern noise: -50 dB for one bead images and -42 dB for seven bead images.

The energies in the one and seven bead images are -31.3 dB and -22.8 dB relative to the average energies in their backgrounds. It is interesting that the maximum errors for linear bias compensation in Figure 4-6 are equal to 0.5 pixels when the energy in the fixed-pattern noise is -32 dB for the one bead image and when the energy is -24 dB for the seven bead image. The algorithm with linear bias compensation weights the two competing motions (that of the bead and that of the background) equally when their energies are equal.

4.9 Discussion

We have developed two approaches for studying motion estimation from video images: (1) theoretical analysis of sinusoidal images and (2) computer simulated motion

of both synthetic images and a stationary image from a video camera.[2] Using these methods, we have shown that gradient-based motion estimates are intrinsically biased. The bias is image-dependent: for some images the bias is a few percent of the spacing between pixels, for other images the bias is as large as the displacements being estimated. For sinusoidal images, bias is largest for images with the highest spatial frequencies (Figures 4-1, 4-2, and 4-3). In many applications (including ours which uses a scientific-grade CCD camera with a 12-bit digitizing system), bias is by far the largest component of the error in motion estimates — much greater than the random errors caused by noise in the measurement system.

We have proposed a linearized model for the bias in estimates based on optical flow, and we have used the model to generate a new motion detection algorithm with linear bias compensation. The algorithm uses data contained in just two video images: no prior knowledge of the image is required. Prior knowledge of the target could be used to estimate the parameters of the linear model; however, prior knowledge is not the only source of information. In principle, the parameters could be determined by simulating shifts of a single image. This was in fact the method used to generate the lower bias functions in Table 4.1. However, software simulation of arbitrary shifts is computationally intensive. The method takes advantage of the accuracy and computational simplicity of shifting images by an integral number of pixels. The new algorithm reduces the maximum subpixel bias by $3.4\times$ – $15\times$, with a typical reduction near factor of four.

We have characterized the effects of several important experimental conditions, discussed below.

4.9.1 Effect of shot noise

Increasing shot noise increases the bias in estimates without linear bias compensation (Figure 4-5). We interpret this result as follows. Suppose that shot noise results in the spurious increase in the brightness of a pixel in the first of the two images used

[2] A third approach, using video images of a scene whose motion is independently known, is discussed in the companion paper (Davis and Freeman, 1997a).

to estimate motion. The four sets of derivative estimates (Equations 4.2) which use the noisy pixel[3] will tend to give contradictory conclusions in the Taylor's expansion (Equation 4.3). For example, assume that without the noise the spatial derivatives (and therefore the temporal derivatives) near the given pixel are zero. Then the estimates from pixels above, below, to the left, and to the right of the noisy pixel will result in contradictory conclusions — that the motion is downward, upward, to the right, and to the left, respectively. These errors tend to cancel each other. However, each pixel contributes to the least-mean-square registration solution (Equation 4.3) in proportion to the magnitude of the local change in brightness (Tanner and Mead, 1986). If the noise is large, the pixels near the noisy one contribute proportionately. Therefore, noise tends to bias estimates based on optical flow toward zero (Figure 4-5). Increasing the signal energy reduces the effect of shot noise by decreasing the relative importance of noisy contributions to the least-mean-square problem.

The bias in estimates with linear bias compensation is nearly independent of noise level (Figure 4-5). This follows from the previous argument by realizing that the algorithm for linear bias compensation includes the effects of estimating displacements that are both positive and negative. If the original estimate of motion is positive, then the images are shifted so that the second estimate is negative and vice-versa. Although each estimate is biased toward zero, the signs of the biases in the two estimates are always opposite. The oppositely signed biases tend to cancel with linear bias compensation.

4.9.2 Effect of fixed-pattern noise

Increasing fixed pattern noise increases the bias in estimates with and without bias compensation (Figure 4-6). For images containing one (seven) bead and fixed-pattern noises larger than -30 dB (-22 dB), both motion detection algorithms fail. If the image were empty, fixed pattern noise would produce images that appear to contain a stationary random pattern. When the image contains a moving target, motion of

[3]For example, if the noisy pixel was $[i', j']$ then $\overline{A}_x[i', j'], \overline{A}_x[i'-1, j'], \overline{A}_x[i', j'-1]$, and $\overline{A}_x[i'-1, j'-1]$ use the noisy pixel.

the target competes with the stationary, fixed-pattern noise. The important statistics for understanding fixed-pattern noise are the energy in the signal and the energy in the noise; both algorithms estimate that the displacement is about half the imposed displacement when the energy in the signal and noise are equal.

4.9.3 Other error reduction techniques

We found that bias in estimates of rigid body motion could generally be reduced by more than a factor of four by applying the linear bias compensation algorithm. Should one regard this improvement as significant, or are there other (possibly simpler) ways to achieve similar improvement? The computation required for the linear bias compensation algorithm is approximately four times that required without bias compensation. We discuss several alternative algorithms with similar computational cost.

If four image pairs could be obtained, one might consider averaging as a method to improve motion estimates. Averaging the images will decrease the shot noise of the images and therefore improve estimates without linear bias compensation by shifting the horizontal axis in Figure 4-5 to the left by 6 dB. Nevertheless, the improvement using linear bias compensation is greater. In fact, for shot noise levels lower than -33 dB, the improvement using linear bias compensation is greater than an *arbitrarily large* reduction in shot noise. Image averaging is therefore not an effective use of computational resources. Alternatively, we could average motion estimates instead of averaging the images. Since the motion estimation process is nonlinear, results for these two averaging schemes will not necessarily be the same. However, this averaging strategy is also ineffective. Since the bias is large compared to the variability of estimates without bias compensation (Figure 4-5), the total error is not significantly reduced by averaging.

If the additional images can be obtained at a faster rate, one might be able to construct better estimates by taking advantage of the fact that displacements between the images are smaller. However, the results in Figure 4-4 suggest that temporal oversampling is also ineffective. First consider the case that the total displacement

between the first and last image is less than 1/3 pixel. The bias in estimates without bias compensation is nearly a linear function of the displacement of the target for displacements less than 1/3 pixel. Therefore, the bias for smaller displacements is proportionally smaller. However, if an estimate of the total displacement is formed by summing the estimates of displacements between intermediate images, then the total bias is unchanged. Furthermore, the random component of the error in this summed estimate will be higher than that based on just the first and last image. The case in which the total displacement between the first and last image is greater than 1/3 pixel is even worse. If the original distance was greater than 1/3 pixel (and less than one), the sum of the four smaller biases will be greater than the original bias. We conclude that simple algorithms based on temporal oversampling are ineffective at reducing bias.

Alternatively, one could change the original registration algorithm by increasing the number of terms in the Taylor expansion Equation 4.1. Unfortunately, formulations with the higher-order terms cannot be solved in closed form. Furthermore, multiple minima may exist[4] for subpixel translations which forces one to use exhaustive search to find the global minimum. For 1/100th pixel accuracy, exhaustive search requires 10^4 separate runs of the algorithm, in comparison to the four runs required by linear bias compensation.

To summarize, the linear bias compensation algorithm's factor of four reduction in bias is a significant improvement which cannot be obtained using simple averaging and/or oversampling techniques, nor easily obtained by changing the underlying algorithm.

4.9.4 Extensions to supra-pixel motion estimation

Linear bias compensation can easily be extended to supra-pixel displacements through a pyramid scheme (Burt, 1984). One can low-pass filter and downsample the images

[4]For example, the algorithm resulting from keeping the bilinear term in the Taylor's expansion (Equation 4.1) has two minima near $(d_x, d_y) = (0.5, -0.8)$ and $(-0.7, -0.5)$ for the problem $A[i,j] = \sin 2i \sin 2j$, $B[i,j] = \sin(2i - 0.7)\sin(2j + 0.9)$.

to obtain a subpixel problem. Using this estimate, the pictures can be shifted so that less downsampling is required to maintain a subpixel estimation problem. The shifting and downsampling can be repeated until no downsampling is needed to obtain a subpixel motion estimation problem. This method is computationally expensive and in practice our supra-pixel problems are only a few pixels. Consequently, we use the gradient-based algorithm on the original images and shift the pictures one pixel in directions indicated by the algorithm. We repeat until the estimates change sign between steps—indicating the displacement between the images is now subpixel. We then apply linear bias compensation to this subpixel problem.

Chapter 5

Nanometer Registration with Light Microscopy: Characterization of Performance Limitations

5.1 Keywords

registration, motion detection, bias compensation, video microscopy

5.2 Introduction

Fundamental limitations on image resolution and registration are different. Image resolution is limited by the optics (Hopkins and Barham, 1950). To a first approximation, images are blurred by convolution with a point-spread function (Goodman, 1968; Born and Wolf, 1975) whose width is on the order of the wavelength of light (typically 550 nm). However, blurred images can be aligned to within a small fraction of the blurring width (Douglas, 1993; Tian and Huhns, 1986). The fundamental limitations on registration are determined by the contrast of the target, by the resolution and precision of the optics and video imager, and by the quantum nature of light.

Precision is also limited by mechanical vibrations. These vibrations can cause

relative motion between the camera and the target. This relative motion causes errors in registration estimates. We show that this effect can be attenuated to less than 1 nm through the use of a vibration-isolation table and long (> 200 ms) integration times.

Precision is also limited by the registration algorithm. In this paper, we examine the two registration algorithms discussed in the companion paper (Davis and Freeman, 1997b). Using a target with calibrated displacements, we show that registration errors consist of a random component that results from image noise and a generally much larger systematic component—bias—that is intrinsic to the registration algorithm. When the new algorithm is applied to video images from a light microscope, registration errors are typically less than 5 nm without averaging — i.e., <1% of the wavelength of the light that generates the images.

These results have potential applications in diverse registration tasks, ranging from lithographic mask alignment (Douglas, 1993; Grenon et al., 1995; Inderhees and Kiefer, 1995) to construction of Gbyte memory chips (Behringer et al., 1991). We have also applied these methods to analyze motions with nanometer precision. Stroboscopic illumination is used to obtain a sequence of stop-action video images, and the displacements between successive images are determined. Results have provided the first direct measurements of sound-induced motions of sensory receptor cells and their overlying tectorial membrane in the inner ear (Davis and Freeman, 1995). Similar techniques have also been used to measure motions of man-made microelectromechanical systems (MEMS) (Freeman and Davis, 1996) and could provide the basis for a "motion probe-station" for design and fabrication of MEMS.

5.3 Methods

This section describes our hardware for obtaining images of microscopic targets and our method for reducing fixed-pattern noise. The section also describes our two test targets.

5.3.1 Video images

Images from a light microscope (Zeiss Axioplan, Thornwood, NY) are projected onto a scientific grade CCD camera (Photometrics 200 series with a Thompson 7883 CCD, Tucson, AZ) and digitized with 12-bit resolution. Under most lighting conditions, the predominate sources of image noise are shot noise and fixed-pattern noise. Shot noise results from the quantum nature of light. The arrival of photons at a given pixel location—and therefore the number of hole-electron pairs generated at that location—can be modeled as a Poisson process (Janesick et al., 1987). Typical images from our system have approximately 10^5 electrons per pixel (Photometrics, 1993). The standard deviation of a Poisson random variable is equal to the square root of the mean. Therefore, the ratio of the standard deviation to the mean number of electrons collected is typically about $10^{2.5}/10^5 = 10^{-2.5}$ which is about -50 dB.

Fixed-pattern noise results from optical imperfections or from spatial variations in the sensitivity of the video imager (Janesick et al., 1987). To reduce the effect of fixed-pattern noise, we use a two-point correction technique (Hiraoka et al., 1987; Healey and Kondepudy, 1994). After the microscope has been adjusted for Koehler Illumination (Inoué, 1986), the focal plane of the microscope is moved so that the target is out of focus. One hundred images are acquired and averaged (to reduce shot noise); the resulting "bright" image gives an estimate of fixed-pattern noise at typical illumination levels. Next, one hundred additional images are acquired and averaged (to reduce dark and read noise (Janesick et al., 1987)) with the microscope light off. This "dark" image gives an estimate of fixed-pattern offsets in the absence of light. Subsequent images are then corrected for pixel-to-pixel variations in sensitivity and offset by calculating a corrected image G, where

$$G[i,j] = \frac{M[i,j] - D[i,j]}{B[i,j] - D[i,j]}.$$

$M[i,j]$ represents the gray value generated by the $[i,j]^{\text{th}}$ pixel in the measured image, and $B[i,j]$ and $D[i,j]$ represent corresponding gray values in the bright and dark images, respectively. For targets reported in this study, the ratio of the energy in

fixed-pattern noise to the average energy in the uncorrected image was typically -45 dB. After two-point correction, this ratio was typically -56 dB.

5.3.2 Target with known position

To test our registration methods, we have constructed a target whose position is calibratible. The target is a TiO_2 particle with a diameter of 0.3 μm (Polysciences, Warrington PA), which we refer to as a bead. The bead is embedded in mounting medium (Mount Quick, Electron Microscopy Sciences, Fort Washington, PA) between a pair of cover slips.

The cover slips are attached to an aluminum mounting block via a piezoelectric stack, so that motion of the stack results in motion of the target relative to the microscope. The distance the target moves is measured using a fiber-optic detector (Angstrom Resolver, Opto Acoustics Sensors, Raleigh, NC) that uses an optic lever technique (Cook and Hamm, 1979) to sense the distance between the tip of its fiber-optic probe and a reflecting surface attached to the moving face of the piezoelectric crystal. The relation between the fiber-optic detector and displacement is linear over a 50 μmrange—much greater than the maximal displacement of the piezoelectric stack. With the detector, we determined that the relation between our electrical drive to the stack ($< \pm 125$ V) and displacement is very linear (correlation coefficient of 0.9999992 between the piezoelectric drive and the detector output for 14 evenly-spaced drive voltages). Unfortunately, the detector's sensitivity (μm/volt) was not known and very difficult to calibrate (different calibration methods yielded different results). Therefore, we did a 1 parameter fit between the simulations and the measurements for the two registration algorithms at the 26 displacements examined. Thus our results had 1 degree of freedom for 52 data points. The fitted sensitivity fell near the mean of the sensitivity factors derived from the different calibration methods.

5.3.3 Motion detection

Pairs of 32×32 images are acquired — the first with the target in one fixed position, and the second after the target has been displaced by the piezoelectric stack. The light integration time was 62 ms. Registration algorithms are used to estimate the known displacement of the stack. Tests consist of one hundred independent trials. Each trial was characterized by an error equal to the difference between the displacement estimate and the displacement measured with the optic lever device. Bias was estimated as the sample mean of the errors and the standard deviation was taken as the square root of the sample variance of the errors.

The performance of two motion detection algorithms, described in (Davis and Freeman, 1997b) are compared to each other and to simulations of the above experiment. The simulations included a mathematical model of the target, fixed-pattern noise, shot noise, and quantization errors. See the companion paper (Davis and Freeman, 1997b) for details on the simulations.

5.3.4 Stationary high-contrast target

To determine the effect of mechanical vibrations on registration estimates, we examined a stationary, very high-contrast target. The target was part of a glass micrometer which had sharp contrast in both directions. A high-contrast target was used to minimize the effects of shot/fixed-pattern noise on registration performance. The target was imaged for a variable amount of time while the total light, and thus shot noise, remained constant. One hundred image pairs (size: 32×32 pixels) were acquired, two point corrected, and analyzed for motion. Standard deviations of registration estimates were computed.

Our video microscopy system rests on a vibration isolation table. This passive air table is activated through pressurization with nitrogen. We determined the effect of the air table on the variability of registration estimates by activating and deactivating the table.

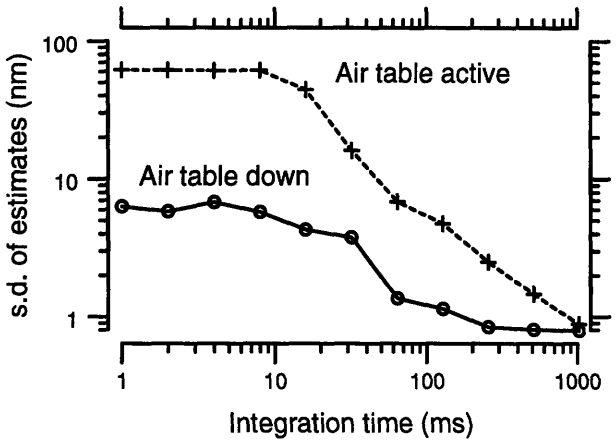


Figure 5-1: Effect of integration time on registration estimates on a high-contrast, stationary target. The standard deviation of registration methods, as a function of the time span of illumination, is plotted for two air table configurations. The total amount of light to the camera is kept constant. Both algorithms yielded the same results.

5.4 Results

Images of a high-contrast, stationary target were analyzed for motion as a function of the light integration time (Figure 5-1). The total amount of light was kept constant. For integration times less than 10 ms, the vibration isolation table reduced registration variability by 20 dB. For integration times greater than 10 ms, the variability decreased with increasing integration time. At a one second integration time, the standard deviation of registration estimates were 0.8–0.9 nm, nearly a 20 dB improvement when using the air table and a 40 dB improvement without the air table. For integration times less than one second, the vibration isolation table improved results; however, the amount of improvement decreases with increasing integration time.

We now compare results of the computer simulated bead from (Davis and Freeman, 1997b) to measurements of a 0.3 μm diameter bead whose motion is independently known (Figure 5-2). The results of the experiment and the simulations are similar. The simulated and measured biases in motion estimates with and without linear bias compensation are similar in both shape and peak magnitude. In the experiment, the largest bias without bias compensation was 15.6 nm (0.068 pixels). The largest bias with linear bias compensation was 4.5 nm (0.020 pixels) — a factor of 3.4 smaller. This factor is slightly smaller than the 4.4× improvement in the simulations. The biases from the simulation generally fall well within a standard deviation of the biases from the measurements, but typically outside one standard error. The largest standard

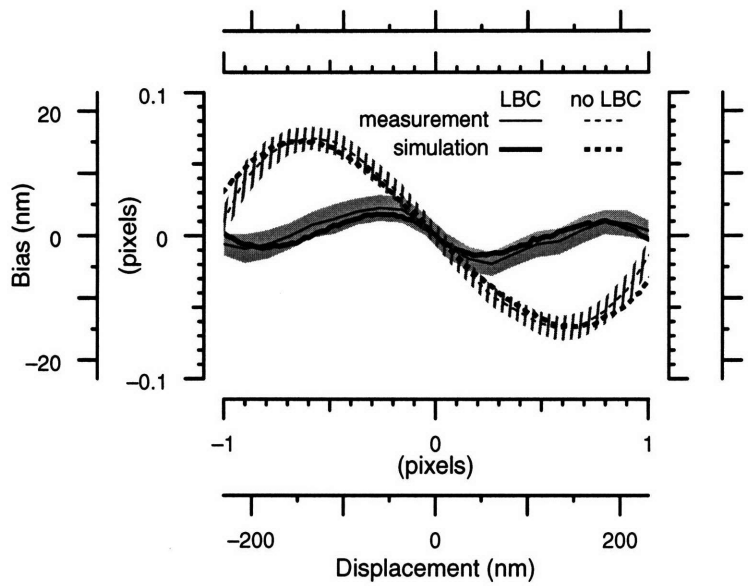


Figure 5-2: Comparison of simulations and measurements of statistics of registration estimates. Calibrated motions of a single TiO_2 bead were estimated from video images obtained from 100 independent trials at each of 20 different displacements. The thin lines and shaded regions represent the bias and ± 1 standard deviation for measurements. The thick lines represent the bias for simulations. The standard deviations of the simulations are not shown. Noise levels in the simulations were taken to match those in the experiment: shot noise was -50 dB and fixed-pattern noise was -55 dB.

deviation of motion estimates obtained from the experiments are 2.4 nm (0.010 pixels) with LBC and 2.8 nm (0.012 pixels) without. These standard deviations are 36%/27% larger than the simulations; the simulations did not include effects of mechanical vibrations.

5.5 Discussion

5.5.1 Mechanical vibrations

Eliminating mechanical vibrations is critical to obtaining nanometer accuracy registration (Figure 5-1). There is a trade-off between how long one is willing to wait for an image (integration time) and precision of the estimate. The vibration-isolation table helps in that regard—reducing the time to sub-nanometer precision by a factor of four. With a more sophisticated isolation table (e.g., an actively stabilized one), we expect the need for long integration times to be reduced.

5.5.2 Simulation versus measurements

The bias function that was measured experimentally has almost the same shape and magnitude as the bias from the simulations. This result suggests that our simulations have captured many of the features found in real-world experiments.

Furthermore, linear bias compensation reduced experimentally determined bias by a factor of 3.4 — nearly the same result as predicted by the computer simulations. These results suggest there is utility in using computer simulation for analyzing and improving motion detection algorithms.

5.5.3 Importance for hearing research

We have demonstrated a method by which video imaging can be used to reliably measure the subpixel motion of not only simple targets but also of complex targets. Specifically, we have shown that our method can be applied to the measurement of the small motions of inner-ear structures that result from acoustical stimulation. These results provide an important new methodology — video imaging — for studying inner-ear motion. Preliminary results (Davis and Freeman, 1995) demonstrate that this is the first methodology which allows direct and concurrent measurement of the motion of all inner-ear structures visible with light microscopy — not only the motions of macroscopic structures but also motions of single sensory receptor cells (typical dimensions of $\sim$ 10 μm) and even motions of single sensory hairs (typical dimensions of $\sim$ 1 μm).

Chapter 6

Image Registration

6.1 Introduction

In this chapter, we further analyze and extend the 2D algorithms developed earlier.

First, we describe performance results from registering sinusoidal images. To estimate subpixel motions, registration algorithms require interpolating functions, or equivalently derivative estimators (Chapter 3). This approximation causes systematic errors in registration estimates. We describe results which indicate that when using a first-order Taylor's expansion of the constant brightness equation (i.e., using linear interpolating functions), the derivative estimators that we use (developed by Horn and Schunck, 1981) have the least subpixel bias for images that vary sinusoidally in space. In particular, two spatial derivative estimators of infinite support perform worse.

Second, we describe an extension of the 2D algorithms to estimating 3 translations for rigid body motion from two volumes of data. These algorithms are used in other chapters to estimate the motion of microscopic structures imaged with our computer microvision system.

Finally, we describe a method to estimate all six degrees of freedom for rigid body motion from two volumes of data. This algorithm has potential application in our computer microvision system, but has not been implemented.

6.2 Sine wave analysis of subpixel registration algorithms

The registration problem—estimating the motion between two images—is examined for six algorithms when estimating the displacement between a one dimensional sine wave and its subpixel translations. The algorithms can be considered to be gradient based (derived from the constant brightness assumption) or block-matching based (minimizing the sum of the squared differences between shifts of the pictures) (Davis et al., 1995). All the algorithms had bias: displacement estimates generally did not equal the displacement. Closed form solutions are found for five out of six basic algorithms. The derivative estimators of Horn and Schunck (1981) have the smallest errors, including two algorithms whose spatial derivatives had infinite support.

6.2.1 Introduction

Image registration is the process of aligning two images by finding the translation, rotation, scaling, or other transformation between them so that they are most alike. Typical applications include medical imaging, satellite imaging, semiconductor manufacturing (e.g., Pritt, 1994). Registration can also be used as a general-purpose motion detection algorithm by using it between sequential images (e.g., Burt, 1984; Bergen et al., 1992; Davis and Freeman, 1995).

Systematic registration errors—bias—are typically the largest source of error in subpixel registrations (Chapter 4). Further, for motion detection applications, bias error accumulates at every time step, generating very poor estimates after many frames have been analyzed.

We examine bias in algorithms for the case of one dimensional sinusoidal images undergoing translation. The algorithms considered are one-dimensional and use differing linear interpolation schemes. Several of the algorithms considered can be interpreted as the one-dimensional simplification of existing linear or bilinear algorithms from the literature (Horn and Schunck, 1981; Burt, 1984; Tian and

Huhns, 1986; Jähne, 1990). Other algorithms considered have a heuristic argument for good performance.

6.2.2 Methods

Algorithm

The continuous-domain constant brightness assumption (Horn and Schunck, 1981) states that the brightness E of an image remains constant with time:

$$E(x,t) = E(x + x_d, t + t_d)$$

where x represents our 1D spatial coordinate and t represents time. The right hand side of the equation can be written as Taylor's series expanded about the point (x,t). Then, after eliminating $E(x,t)$ from both sides, the constant brightness equation can be written as

$$0 = \frac{\partial E(x,t)}{\partial x} x_d + \frac{\partial E(x,t)}{\partial t} t_d + \cdots$$

where the ellipses represent higher order terms. We can simplify this equation by using the camera's frame time as our unit of time so that $t_d = 1$ frame. Then

$$E_x x_d + E_t + \cdots = 0$$

where E_x, and E_t represent the partial derivatives of the brightness function E. The discrete-domain analog of the equation is

$$G_x d_x + G_t + \cdots = 0$$

where $G_x[i,k]$ and $G_t[i,k]$ are linear combinations of images and their integer shifts that approximate the partial derivatives of the brightness function E. We assume that a patch of the image is undergoing a rigid-body translation, and least squares is used to combine information across the patch to estimate the translation: choose $\hat{d}_x$

equal to the values of d_x that minimizes

$$\sum_i (G_x d_x + G_t + \cdots)^2$$

where i ranges over the patch. We call this range the summation region.

We consider only the linear term in the Taylor expansion. Thus, our estimate of the x translation is

$$\hat{d}_x = -\frac{\sum_i G_t G_x}{\sum_i G_x G_x}.$$

The algorithms studied here differ by their derivative approximations. Note that by equivalence (Davis et al., 1995), this class of algorithms can be re-cast as block matching methods which minimize the sum of squared differences between shifted versions of the original images, where the shifting is done using linear interpolation.

Sinusoidal images

Consider a brightness field that (1) varies sinusoidally in the x direction, and (2) moves everywhere with velocity v in the x direction:

$$E(x,t) = E_0 + E_1 \cos \Omega(x + \phi - vt)$$

where $E_0 \geq E_1$ since the brightness cannot be negative. Suppose that a camera samples this brightness field uniformly in space and time with unit inter-pixel distance and unit frame time. Then two consecutive images obtained at time indices 0 and 1 equal

$$\begin{aligned} G[i,j,0] &= E_0 + E_1 \cos \Omega(i + \phi) \\ G[i,j,1] &= E_0 + E_1 \cos \Omega(i + \phi - d_x) \end{aligned} \tag{6.1}$$

where d_x is the displacement of the scene in the second image with respect to the first. We apply registration algorithms to these sinusoidal images to generate estimates of displacement as a function of the known displacement d_x, spatial frequency Ω, and phase ϕ of the sinusoid. The Nyquist frequency of the sinusoids is $\Omega = \pi$.

6.2.3 Results

In this section we describe the performance of six different registration algorithms, from the most accurate to the least. The algorithms all use the linear term in the one dimensional constant brightness equation. The algorithms differ in how they estimate the space (G_x) and time (G_t) derivatives.

Algorithm 1

Horn and Schunck (1981) suggested that when approximating a set of partial derivatives, the approximations should estimate the derivatives at the same point. They suggested using an average of several first differences to approximate the partial derivatives midway between the sample points in space and time. For one dimensional images, these approximations reduce to

$$\begin{aligned} G_t &= \tfrac{1}{2}\textstyle\sum_{i'=i}^{i+1} G[i',1] - G[i',0] \sim E_t \\ G_x &= \tfrac{1}{2}\textstyle\sum_{k'=0}^{1} G[i+1,k'] - G[i,k'] \sim E_x. \end{aligned} \tag{6.2}$$

These derivative operators are shown graphically in Figure 6-1. The spatial and temporal operators best estimate the derivatives at the same point: $(i + 1/2, 1/2)$.

Substituting Equations 6.1 into Equations 6.2 results in a set of equations that (after some trigonometric simplification) can be written as

$$G_t = 2E_1 \sin\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+2\phi+1-d_x)}{2} \cos\tfrac{\Omega}{2}$$

$$G_x = -2E_1 \sin\tfrac{\Omega}{2} \sin\tfrac{\Omega(2i+2\phi+1-d_x)}{2} \cos\tfrac{\Omega d_x}{2}$$

so that

$$\hat{d_x} = -\frac{\sum_i G_t G_x}{\sum_i G_x G_x} = \frac{\tan(\Omega d_x/2)}{\tan\Omega/2}. \tag{6.3}$$

Note that after factoring the derivative approximations, the only term that depends on the summation region is the same for the two derivative approximations. Conse-

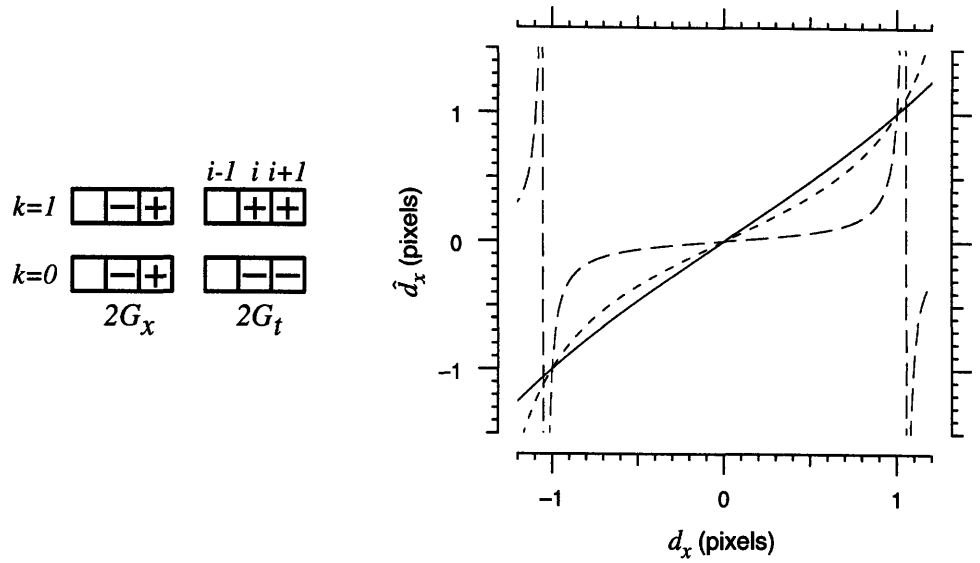


Figure 6-1: Algorithm 1's computational molecules (left) and displacement estimates (right) for a 1D sinusoidal image. The left computational molecule illustrates the spatial derivative estimator, the average of forward difference operators applied to the image at times $k = 0$ and $k = 1$. The right computational molecule illustrates the temporal derivative estimator, the average of forward difference operator applied to pixels i and $i+1$ in the two images. The solid, dotted, and dashed curves in the plot (right) illustrate estimates of displacement based on Algorithm 1 for sinusoidal images with frequencies Ω equal to 1, 2, and 3, respectively. The curves are computed from Equation 6.3.

quently, all dependence on the summation region cancels in the registration estimate.

Plots of the relation between the estimate $\hat{d}_x$ and the imposed displacement d_x are illustrated in Figure 6-1 for images with three different spatial frequencies. The estimated displacement is exactly equal to the imposed displacement when the imposed displacement is equal to -1, 0, and 1 pixel for all spatial frequencies. Estimates are odd symmetric with respect to the displacement, systematically too small for subpixel displacements, too large for displacements between 1 and π/Ω (half the wavelength), and are periodic with a period equal to the wavelength of the sinusoid. The difference between the estimate and imposed displacement approaches 0 for low spatial frequencies (Figure 6-7 and Figure 6-8) and approaches the displacement for higher frequencies (Equation 6.3).

Algorithm 2

Instead of testing spatial derivative estimators with a support of 4,5,6, etc., we will test infinite support estimators by using the partial derivative of E in each image:

$$G_x[i,k] = \frac{1}{2}(E_x(i,1) + E_x(i,0))$$

In practice we cannot use this estimator, but we can with our mathematically-specified example. Since we only have two samples in the time dimension, we use first differences:

$$G_t[i,k] = G[i,1] - G[i',0] \sim E_t$$

where we have selected approximations that best estimate derivatives at the same point: $(i, 1/2)$.

Substituting Equations 6.1 into these derivative estimators results in a set of equations that can be written as

$$G_t = 2E_1 \sin\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+2\phi-d_x)}{2} G_x = -\Omega E_1 \cos\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+2\phi-d_x)}{2}$$

so that

$$\hat{d_x} = -\frac{\sum_i G_t G_x}{\sum_i G_x G_x} = \frac{2}{\Omega}\tan\frac{\Omega d_x}{2}. \tag{6.4}$$

Note that, similar to Algorithm 1, all the terms that depend on the summation region cancel, leaving the estimate independent of the summation region.

This estimator (Figure 6-2) is odd symmetric, correctly estimates 0 displacement, and overestimates non-zero displacements. It has the distinction of being the only algorithm examined whose slope at $d = 0$ is correct. Thus, for very small motions, $|d| \ll 1$, this algorithm is superior to the other 8. The price paid for this excellent small displacement performance is poor estimates for $d = 1$: For all frequencies, its maximum subpixel error is at least 2.5 times worse than Algorithm I (Figure 6-7). Further, its maximum subpixel error is unbounded as the frequency approaches the Nyquist rate (π). For frequencies below half the Nyquist rate, the high-quality, small-

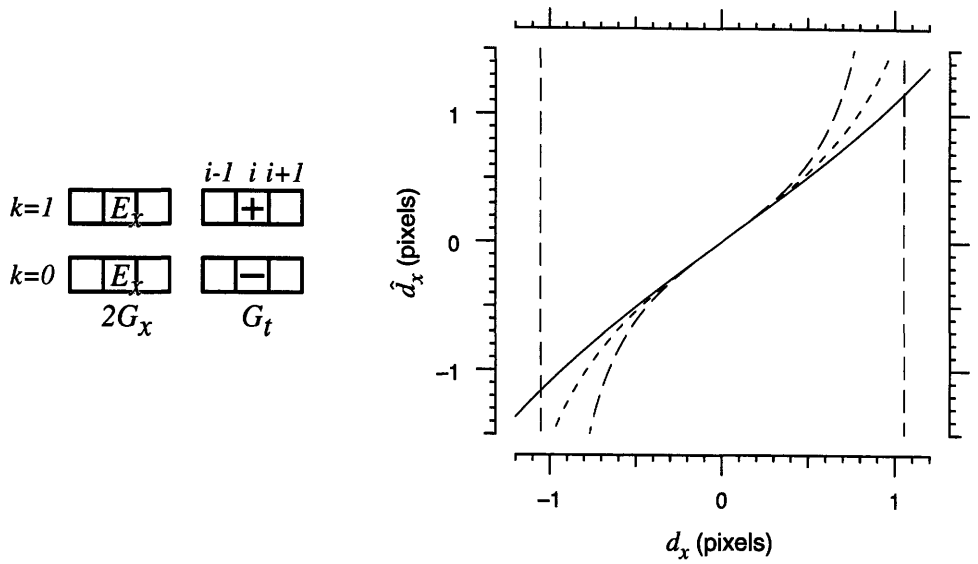


Figure 6-2: Algorithm 2's computational molecules (left) and displacement estimates (right) for a 1D sinusoidal image. The left computational molecule illustrates the spatial derivative estimator, the average of the exact derivative of the sinusoid at times $k = 0$ and $k = 1$. The right computational molecule illustrates the temporal derivative estimator, the forward difference operator applied to pixel i in the two images. The solid, dotted, and dashed curves in the plot (right) illustrate estimates of displacement based on Algorithm 2 for sinusoidal images with frequencies Ω equal to 1, 2, and 3, respectively. The curves are computed from Equation 6.4.

displacement estimates and the low-quality, large-displacement estimates cancel: This algorithm has nearly the same error, averaged over all subpixel displacements, as Algorithm I.

Algorithm 3

One possible explanation for the worse performance of Algorithm 2 when compared to Algorithm 1 is the location of the derivative estimators. To test this hypothesis, we use the same temporal derivative approximation as Algorithm 1:

$$G_t = \frac{1}{2}\sum_{i'=i}^{i+1} G[i',1] - G[i',0] \sim E_t.$$

We use the average of two spatial partial derivatives :

$$G_x[i,k] = \frac{1}{2}(E_x(i+\frac{1}{2},1) + E_x(i+\frac{1}{2},0))$$

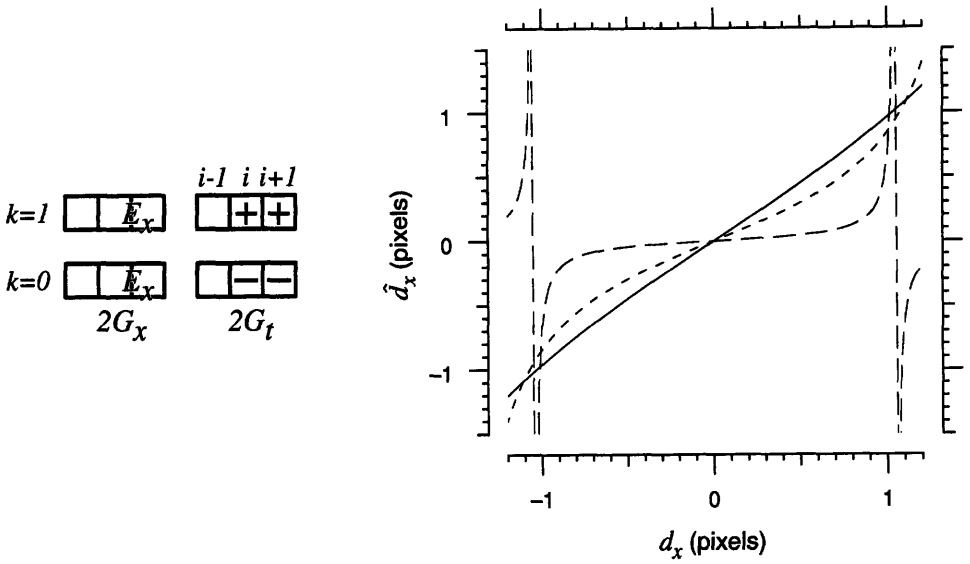


Figure 6-3: Algorithm 3's computational molecules (left) and displacement estimates (right) for a 1D sinusoidal image. The curves are computed from Equation 6.5.

in order to estimate both the space and time derivatives at $(i + 1/2, 1/2)$. We can evaluate this infinite-support derivate estimator for our mathematically specified sinusoidal images.

Substituting Equations 6.1 into these derivative estimators results in

$$G_t = 2E_1 \sin \tfrac{\Omega d_x}{2} \sin \tfrac{\Omega(2i+2\phi+1-d_x)}{2} \cos \tfrac{\Omega}{2}$$

$$G_x = -\Omega E_1 \cos \tfrac{\Omega d_x}{2} \sin \tfrac{\Omega(2i+2\phi+1-d_x)}{2}$$

so that

$$\hat{d}_x = -\frac{\sum_i G_t G_x}{\sum_i G_x G_x} = \frac{2}{\Omega} \tan \frac{\Omega d_x}{2} \cos \frac{\Omega}{2}. \tag{6.5}$$

Note that, similar to Algorithms 1 and 2, all the terms that depend on the summation region cancel, leaving the estimate independent of the summation region.

This algorithm (Figure 6-3) is odd symmetric, correctly estimates 0 displacement, and underestimates non-zero displacements. This algorithm and Algorithm 2 have the same solution to sinusoidal estimation, except for the $\cos \Omega/2$ term. This term has the effect of controlling the maximum subpixel error as the frequency approaches the Nyquist rate. In fact, its maximum subpixel error is always smaller than Algorithm

2. Nevertheless, it has a larger average error (Figure 6-8) than Algorithm 2 except for very high frequencies ($\Omega >\approx 2.5$).

This algorithm is qualitatively very similar to Algorithm 1, but quantitatively worse than Algorithm 1 in every respect. It can be easily shown from the equations of their estimates that for all subpixel displacements and all frequencies greater than 0, the error in Algorithm 1 is less than or equal to the error in this algorithm (Algorithm 3). The inequality is strict for nonzero displacements.

Algorithm 4

Another realizable variation (Jähne, 1990) of Algorithm 2 is to use an average of two three-point derivative estimators for the spatial derivative approximation:

$$G_t = G[i,1] - G[i,0] \sim E_t$$

$$G_x = \tfrac{1}{4}\left(G[i+1,1] - G[i-1,1] \; + \; G[i+1,0] - G[i-1,0]\right) \sim E_x.$$

This third-order derivative estimator could potentially generate better estimates than second-order estimators. The odd symmetry of derivative estimators require that the center point of the estimator be zero. Thus, the three-point derivative estimator has only two non-zero terms. The derivative estimators are co-localized at $(i, 1/2)$.

Substituting Equations 6.1 into these derivative estimators results in

$$G_t = 2E_1 \sin\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+2\phi-d_x)}{2}$$

$$G_x = -E_1 \sin\Omega \cos\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+2\phi-d_x)}{2}$$

so that

$$\hat{d}_x = -\frac{\sum_i G_t G_x}{\sum_i G_x G_x} = \frac{2\tan\frac{\Omega d_x}{2}}{\sin\Omega} \tag{6.6}$$

Note that, similar to Algorithms 1, 2, 3 and 4, all the terms that depend on the summation region cancel, leaving the estimate independent of the summation region.

This algorithm (Figure 6-4) is odd symmetric, correctly estimates 0 displacement,

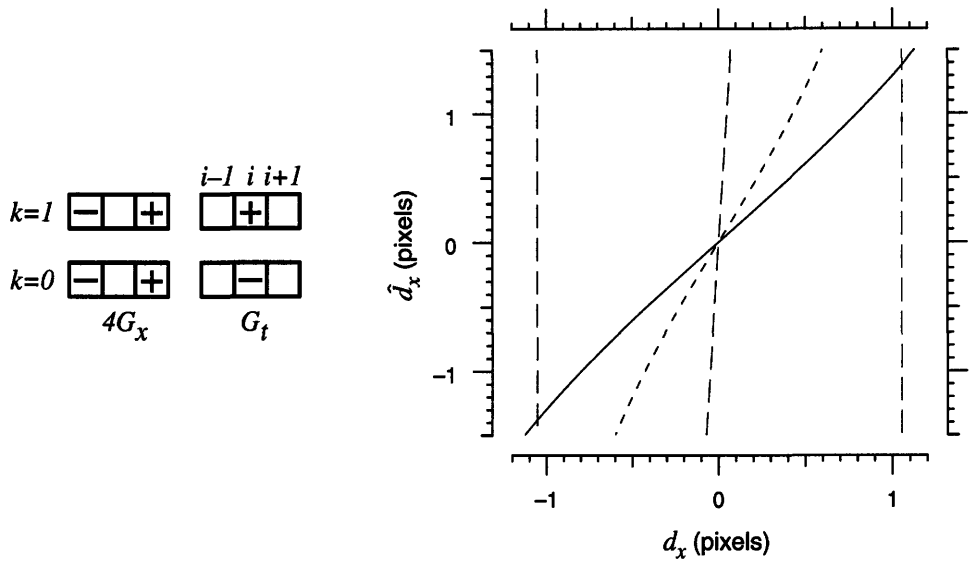


Figure 6-4: Algorithm 4's computational molecules (left) and displacement estimates (right) for a 1D sinusoidal image. The curves are computed from Equation 6.6.

and overestimates non-zero displacements. It is a poor performer: nearly a order of magnitude worse than Algorithm I in both average (Figure 6-8) and maximum (Figure 6-7) error for all frequencies.

Algorithm 5

This algorithm uses one first difference in space and time to estimate the derivatives. The estimators are not co-localized: the space derivative is best estimated at $(i + 1/2, 1)$ and the time at $(i, 1/2)$. This algorithm is equivalent (Davis et al., 1995) to a block matching algorithm in which the displacement x_d is determined by minimizing the sum of the squared differences between the first image and a shifted second image (Burt, 1984; Bergen et al., 1992; Tian and Huhns, 1986):

$$\widehat{x_d} = \arg\min_{x'_d} \sum_i \left(G[i + x'_d, 1] - G[i, 0]\right)^2$$

where G at non-integer coordinates is approximated by linear interpolation.

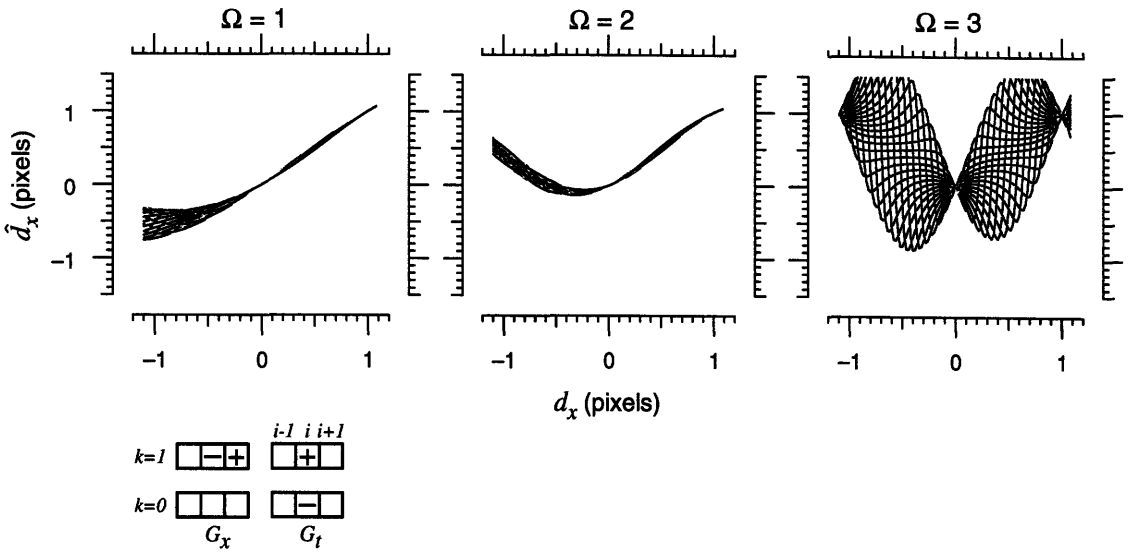


Figure 6-5: Algorithm 5's computational molecules (bottom) and displacement estimates (top) for three frequencies Ω of a 1D sinusoidal image. The displacement estimates were calculated using a summation region of five pixels. Estimates from images with 17 evenly-spaced phases of the sinusoid's phase ϕ are plotted for each frequency. The curves are computed from Equation 6.7.

The derivative estimators are

$$G_t = G[i,1] - G[i,0] \sim E_t$$
$$G_x = G[i+1,1] - G[i,0] \sim E_x.$$

Substituting Equations 6.1 into these derivative estimators results in

$$G_t = 2E_1 \sin\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+2\phi-d_x)}{2}$$
$$G_x = -2E_1 \sin\tfrac{\Omega}{2} \sin\tfrac{\Omega(2i+2\phi-2d_x+1)}{2}$$

so that

$$\hat{d}_x = -\frac{\sum_i G_t G_x}{\sum_i G_x G_x} = \frac{\sin\frac{\Omega d_x}{2} \sum \sin\frac{\Omega(2i+2\phi-d_x)}{2} \sin\frac{\Omega(2i+2\phi-2d_x+1)}{2}}{\sin\frac{\Omega}{2} \sum \sin^2\frac{\Omega(2i+2\phi-2d_x+1)}{2}} \tag{6.7}$$

Unlike the previous estimators, this algorithm depends on the summation region. Denote the number of pixels in the summation region N_s. $N_s = 5$ is common in practice (Burt, 1984; Bergen et al., 1992); consequently, we analyze performance for this case.

The derivative approximations are not odd-symmetric (about i), and neither is the estimate. As shown in Figure 6-5, this algorithm correctly estimates displacements of 0 and 1. At all other displacements, the estimate depends on the phase of the sinusoid. Estimates of negative displacements are very poor for all frequencies — estimates may even have the wrong sign for $\Omega \geq 1.43$. Estimates of positive displacements are poor at high spatial frequencies, but become very good (comparable to algorithm 1) for frequencies below half the Nyquist rate. Its maximum (Figure 6-7) and average (Figure 6-8) errors for all subpixel displacements do not go to 0 as the frequency goes to 0, and is greater than Algorithm I for all frequencies.

Algorithm 6

One possible fix to Algorithm 5 is treat both images equally. Instead of minimizing the sum of the squared differences between the first image and a shifted second image

$$\widehat{x_d} = \arg\min_{x'_d} \sum_i \left(G[i + x'_d, 1] - G[i, 0]\right)^2$$

shift both images half way:

$$\widehat{x_d} = \arg\min_{x'_d} \sum_i \left(G[i + \frac{x'_d}{2}, 1] - G[i - \frac{x'_d}{2}, 0]\right)^2$$

G at non-integer coordinates is approximated by linear interpolation.

The equivalent (Davis et al., 1995) derivative estimators use the average of two first differences to estimate the spatial derivatives, while keeping the simple first difference for the time derivative:

$$\begin{aligned} G_t &= G[i, 1] - G[i, 0] \sim E_t \\ G_x &= \tfrac{1}{2}\left(G[i + 1, 1] - G[i, 1] \; + \; G[i, 0] - G[i - 1, 0]\right) \sim E_x. \end{aligned}$$

The offset first differences in the space derivative approximation and the time derivative approximation best approximate the derivatives at the same point as Algorithm 2: $(i, 1/2)$.

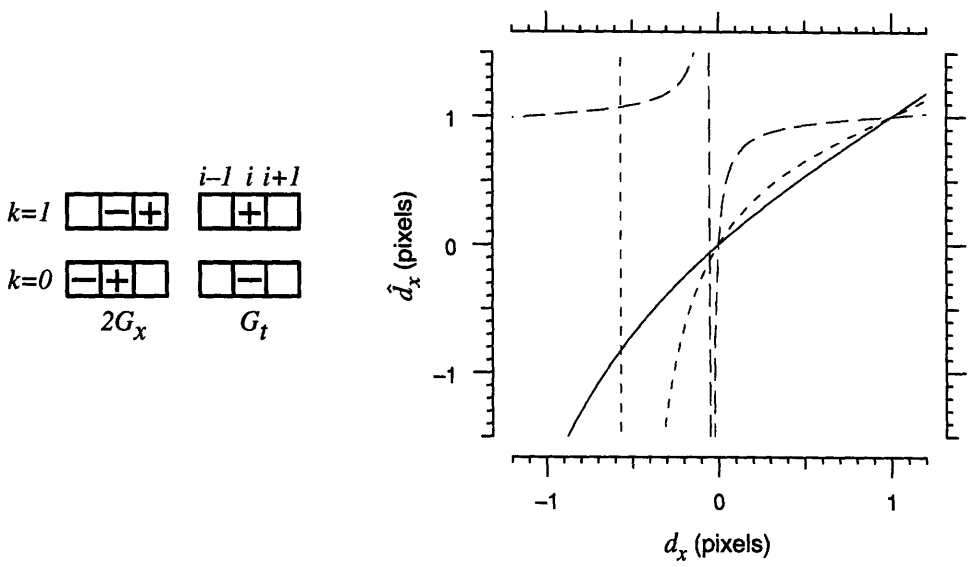


Figure 6-6: Algorithm 6's computational molecules (left) and displacement estimates (right) for a 1D sinusoidal image. The curves are computed from Equation 6.8.

Substituting Equations 6.1 into these derivative estimators results in a set of equations that can be written as

$$G_t = 2E_1 \sin\tfrac{\Omega d_x}{2} \sin\tfrac{\Omega(2i+2\phi-d_x)}{2}$$
$$G_x = -2E_1 \sin\tfrac{\Omega}{2} \cos\tfrac{\Omega(d_x-1)}{2} \sin\tfrac{\Omega(2i+2\phi-d_x)}{2}$$

so that

$$\hat{d}_x = -\frac{\sum_i G_t G_x}{\sum_i G_x G_x} = \frac{\sin\frac{\Omega d_x}{2}}{\sin\frac{\Omega}{2}\cos\frac{\Omega(d_x-1)}{2}} \tag{6.8}$$

Note that, in contrast to Algorithm 5, all the terms that depend on the summation region cancel, leaving the estimate independent of the summation region.

This algorithm correctly estimates displacements of 0 and 1, overestimates positive displacements and very poorly estimates negative displacements. Unlike all previous Algorithms 1–4, this algorithm is not odd symmetric (Figure 6-6). While this algorithm has very small error (matching Algorithm 1) for positive displacements, its poor negative-displacement makes it worse than Algorithms 1-4 for subpixel displacements in terms of maximal (Figure 6-7) and average (Figure 6-8) error. In contrast to Algorithm 5, this algorithm's bias goes to zero as the spatial frequency goes to zero. At higher spatial frequencies, this algorithm performs worse than Algorithm 5.

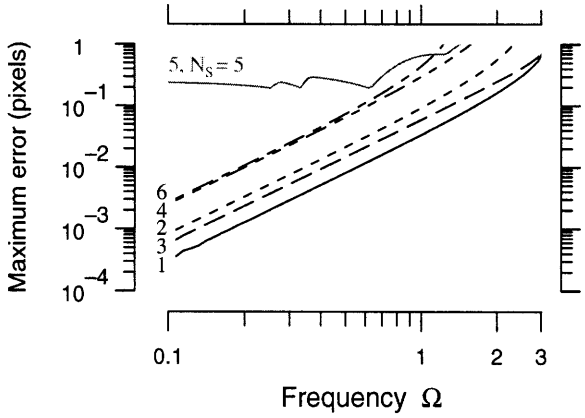


Figure 6-7: Maximum subpixel errors in estimates of the displacements of a sinusoidal image as a function of the spatial frequency Ω of the image. This maximum subpixel bias is taken as the biggest difference between $\hat{d}_x$ and d_x for any $|d_x| < 1$. For Algorithms 1, 2, 3, 4, and 6, the maximum error was found analytically (black lines). For Algorithm 5, the maximum for a given frequency was approximated by the maximum error from estimates with 1,000 evenly-spaced phases ($0 \leq \phi < 2\pi$) at 100 evenly-spaced subpixel displacements (gray line).

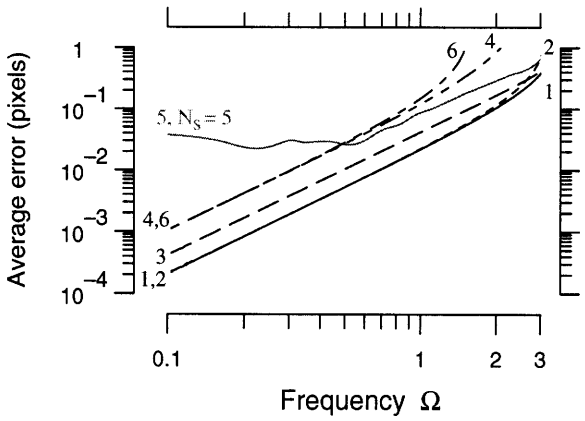
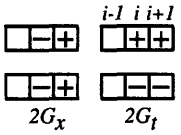
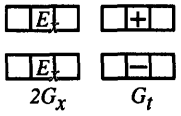
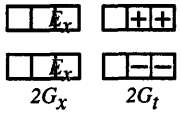
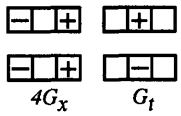
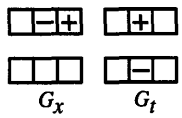
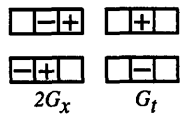


Figure 6-8: Average absolute subpixel errors in estimates of the displacements of a sinusoidal image as a function of the spatial frequency Ω of the images. For Algorithms 1. 2, 3, 4, and 6, the average absolute error was found analytically by taking one half of the integral of the absolute value of the estimate minus the displacement, as the displacement ranged from -1 to 1 (black lines). For Algorithm 5, the average absolute error for a given frequency was approximated by the average absolute error from estimates with 1,000 evenly-spaced phases ($0 \leq \phi < 2\pi$) at 100 evenly-spaced subpixel displacements (gray line).

A fix that one might consider is to use a combination of this algorithm and its mirror (generated by flipping the computational molecules about $i = 0$). The first step is to get Algorithm 4's estimate. If the estimate is positive, use that estimate; if negative, use its mirror estimate. This fix would work if Algorithm 4's estimate always has the correct sign; however, it does not. If the spatial frequency is greater than half the Nyquist rate ($\Omega > \pi/2$), then the estimate is singular at $d_x = 1 - \pi/\Omega$. For more negative displacements, the sign of the estimate is wrong. Thus, Algorithm 4's poor negative displacement performance cannot be eliminated with the above algorithm.

Table 6.1: Summary of algorithms and their performance for 1D sinusoidal images. The left column describes the derivative estimators in terms of computational molecules. The right column shows that algorithm's estimate of displacement between the sinusoidal images (Equation 6.1).

	Computational Molecules	Estimate of displacement d_x
1.		$\frac{\tan\frac{\Omega d_x}{2}}{\tan\frac{\Omega}{2}}$
2.		$\frac{2}{\Omega}\tan\frac{\Omega d_x}{2}$
3.		$\frac{2}{\Omega}\tan\frac{\Omega d_x}{2}\cos\frac{\Omega}{2}$
4.		$\frac{2\tan\frac{\Omega d_x}{2}}{\sin\Omega}$
5.		$\frac{\sin\frac{\Omega d_x}{2}\sum\sin\frac{\Omega(2i+2\phi-d_x)}{2}\sin\frac{\Omega(2i+2\phi-2d_x+1)}{2}}{\sin\frac{\Omega}{2}\sum\sin^2\frac{\Omega(2i+2\phi-2d_x+1)}{2}}$
6.		$\frac{\sin\frac{\Omega d_x}{2}}{\sin\frac{\Omega}{2}\cos\frac{\Omega(d_x-1)}{2}}$

The algorithms and their estimates are summarized in Table 6.1.

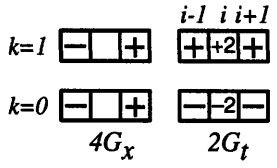


Figure 6-9: Computational molecule resulting from a 2 pixel moving average performed on the images before applying Algorithm I. This algorithm has the same estimate as Algorithm I.

6.2.4 Generalizations

Composition rules for equivalent algorithms

Since our test image contains only one frequency, linear filtering merely changes the DC value (E_0), and the amplitude (E_1) and phase (ϕ) of the images. None of the estimates from the 6 algorithms depend on E_0 or E_1. Only Algorithm 5 depends on ϕ; changing ϕ may improve or degrade performance for a specific image, but not general performance parameters like the maximum (Figure 6-7) or average (Figure 6-8) error.

Thus many algorithms have different computational molecules but the same estimate for sinusoidal images. For example, when the filter $h[i,k] = \frac{1}{2}(G[i-1,k]+G[i,k])$ is combined with Algorithm I, we get the computational molecule illustrated in Figure 6-9, but the same estimate as Algorithm I.

One dimensional motion of 2D images

We show that under fairly broad conditions, the results for Algorithms 1, 2, 3, 4, and 6 are unaffected by allowing our image to vary in the y direction. That is, when the image is sinusoidal in the x direction and displaced only in the x direction, the x estimate is given in Table 6.1 and the y estimate is 0. The proof uses the two-dimensional constant brightness equation taken out to the linear terms. In this case, the estimates are

$$\hat{d}_x = \frac{\sum G_x G_y \sum G_y G_t - \sum G_y^2 \sum G_x G_t}{\sum G_x G_y \sum G_x G_y - \sum G_x^2 \sum G_y^2}$$

$$\hat{d}_y = \frac{-\sum G_x^2 \sum G_y G_t + \sum G_x G_y \sum G_x G_t}{\sum G_x G_y \sum G_x G_y - \sum G_x^2 \sum G_y^2}.$$

Let $\alpha_x \sum \sigma_x = \sum G_x$, where α_x is the product of all the terms which can be factored out of the summation and σ_x contains all the terms left inside the summation. Make similar definitions for the other summations.

Theorem 3 *For a given estimation algorithm and pair of 1D images, if* $\sigma_x = \sigma_t$, *then* $\hat{d}_x = \alpha_t/\alpha_x$.

Proof: The 1D estimates can be re-written as

$$\hat{d}_x = -\frac{\sum G_t G_x}{\sum G_x G_x} = \frac{\alpha_i \alpha_k \sum \sigma_x \sigma_t}{\alpha_x^2 \sum \sigma_i^2}$$

Substituting σ_x for σ_t and simplifying gives the desired results.

This factorization is how the estimates were derived in the Results section. The next theorem states that this factorization is unaffected by letting the image vary in the y direction, as long as that variation affects the part of G_x and G_t that depend on the summation region in the same manner.

Theorem 4 *For a given estimation algorithm and pair of 2D images, if* $\sigma_x = \sigma_t$, *then (1)* $\hat{d}_x = \alpha_t/\alpha_x$ *and (2)* $\hat{d}_y = 0$.

Proof: The 2D estimates can be re-written as

$$\hat{d}_x = \frac{\alpha_x \alpha_y^2 \alpha_t \left(\sum \sigma_x \sigma_y \sum \sigma_y \sigma_t - \sum \sigma_y^2 \sum \sigma_x \sigma_t\right)}{\alpha_x^2 \alpha_y^2 \left(\sum \sigma_x \sigma_y \sum \sigma_x \sigma_y - \sum \sigma_x^2 \sum \sigma_y^2\right)}$$

$$\hat{d}_y = \frac{\alpha_x^2 \alpha_y \alpha_t \left(\sum \sigma_x \sigma_y \sum \sigma_x \sigma_t - \sum \sigma_x^2 \sum \sigma_y \sigma_t\right)}{\alpha_x^2 \alpha_y^2 \left(\sum \sigma_x \sigma_y \sum \sigma_x \sigma_y - \sum \sigma_x^2 \sum \sigma_y^2\right)}.$$

Substituting σ_x for σ_t and simplifying gives the desired results.

The premises for Theorem 4 are true, for example, for the following generalization to the images in Equation 4.9:

$$G[i, j, 0] = E_0 + E_1 f(j) \cos \Omega(i + \phi)$$
$$G[i, j, 1] = E_0 + E_1 f(j) \cos \Omega(i + \phi - d_x)$$

where $f(j)$ is an arbitrary function subject to the constraint that brightness cannot be negative. If we use the same derivative estimators as in Table 6.1 at location j, then $\sum G_x = \alpha_x \sum f(j) \sigma_x$ and $\sum G_t = \alpha_k \sum f(j) \sigma_t$. Thus if the 1D algorithm was independent of the summation region (i.e., $\sigma_x = \sigma_t$), then the 2D algorithm is also.

Similarly, if we average the 1D estimators at the same n points in j and if the 1D algorithm was independent of the summation region then Theorem 2 still applies and we get an x estimate that is independent of the variations in y.

A specific example of the 2D generalization is

$$G[i,j,0] = E_0 + E_1 \cos \Omega_1 j \cos \Omega i$$
$$G[i,j,1] = E_0 + E_1 \cos \Omega_1 j \cos \Omega(i + \phi - d_x)$$

which is a two-dimensional sinusoid pattern previously used, for example, by Barron et al. (1994). In that paper, they allowed the image to translate in both directions, but only reported results for two velocity/frequency pairs. Our results characterize algorithms in an infinite, three parameter space (x and y spatial frequencies and x displacement).

6.3 Three-dimensional, pure translation image registration

In this section we briefly describe the extension of the two algorithms described in Chapter 4 to three dimensions.

6.3.1 Gradient-based method

Following Section 4.4.1, let $A(x, y, z)$ and $B(x, y, z)$ represent misaligned 3D images of the same target. We assume that the misalignment is purely translational, so that

$$A(x - \frac{d_x}{2}, y - \frac{d_y}{2}, z - \frac{d_z}{2}) = B(x + \frac{d_x}{2}, y + \frac{d_y}{2}, z + \frac{d_z}{2})$$

where (d_x, d_y, d_z) represents the displacement between the two 3D images. If the displacement is small, we can expand the offset images in Taylor's series and retain just the linear terms so that

$$A(x, y, z) - \frac{d_x}{2}\frac{\partial A}{\partial x} - \frac{d_y}{2}\frac{\partial A}{\partial y} - \frac{d_z}{2}\frac{\partial A}{\partial z} = B(x, y, z) + \frac{d_x}{2}\frac{\partial B}{\partial x} + \frac{d_y}{2}\frac{\partial B}{\partial y} + \frac{d_z}{2}\frac{\partial B}{\partial z}.$$

The 3D images $A(x, y, z)$ and $B(x, y, z)$ are sampled by the camera to give arrays of pixel values $A[i, j, k]$ and $B[i, j, k]$. We approximate the images and their partial derivatives at points midway between pixels in order to obtain consistent estimates. For example, $A\left((i + \frac{1}{2})\Delta, (j + \frac{1}{2})\Delta, (k + \frac{1}{2})\Delta\right)$ is approximated by $\overline{A}[i, j, k]$ where

$$\overline{A}[i, j, k] = (A[i, j, k] + A[i + 1, j, k] + A[i, j + 1, k] + A[i + 1, j + 1, k] +$$
$$A[i, j, k + 1] + A[i + 1, j, k + 1] + A[i, j + 1, k + 1] + A[i + 1, j + 1, k + 1])/8$$

and $\partial A/\partial x$ at $((i + \frac{1}{2})\Delta, (j + \frac{1}{2})\Delta, (z + \frac{1}{2})\Delta)$ is approximated by

$$\overline{A}_x[i, j, k] = (A[i + 1, j, k] - A[i, j, k] + A[i + 1, j + 1, k] - A[i, j + 1, k] +$$
$$A[i + 1, j, k + 1] - A[i, j, k + 1] + A[i + 1, j + 1, k + 1] - A[i, j + 1, +1k])/4\Delta$$

where Δ represents the pixel spacing. Substituting these approximations into the Taylor's expansion yields

$$\begin{aligned}\overline{A}[i,j,k] - \overline{B}[i,j,k] = \;& d_x(\overline{A}_x[i,j,k] + \overline{B}_x[i,j,k])/2 \\ +\;& d_y(\overline{A}_y[i,j,k] + \overline{B}_y[i,j,k])/2 \\ +\;& d_z(\overline{A}_z[i,j,k] + \overline{B}_z[i,j,k])/2.\end{aligned}$$

Substituting

$$\begin{aligned}G_x(i,j,k) = \;& \left(\overline{A}_x[i,j,k] + \overline{B}_x[i,j,k]\right)/2 \\ G_y[i,j,k] = \;& \left(\overline{A}_y[i,j,k] + \overline{B}_y[i,j,k]\right)/2 \\ G_z[i,j,k] = \;& \left(\overline{A}_z[i,j,k] + \overline{B}_z[i,j,k]\right)/2 \\ G_t[i,j,k] = \;& \overline{B}[i,j,k] - \overline{A}[i,j,k].\end{aligned}$$

we obtain

$$G_t[i,j,k] = d_x G_x[i,j,k] + d_y G_y[i,j,k] + d_z G_z[i,j,k]$$

which can be regarded as an array i,j,k of equations with three unknowns, d_x, d_y, and d_z. Because data from a camera are corrupted by noise, the constraint equations are generally inconsistent. Such equations can be solved using least squares and the solution can be expressed in closed form:

$$\begin{bmatrix} \widehat{x_d} \\ \widehat{y_d} \\ \widehat{z_d} \end{bmatrix} = - \begin{bmatrix} \sum\sum\sum G_x G_x & \sum\sum\sum G_x G_y & \sum\sum\sum G_x G_z \\ \sum\sum\sum G_y G_x & \sum\sum\sum G_y G_y & \sum\sum\sum G_y G_z \\ \sum\sum\sum G_z G_x & \sum\sum\sum G_z G_y & \sum\sum\sum G_z G_z \end{bmatrix}^{-1} \begin{bmatrix} \sum\sum\sum G_x G_t \\ \sum\sum\sum G_y G_t \\ \sum\sum\sum G_z G_t \end{bmatrix}$$

where the sums are over i, j, and k. We refer to these equations as the "3D gradient-based algorithm" and $(\widehat{x_d}, \widehat{y_d}, \widehat{z_d})$ as the "3D gradient-based estimate" of displacement.

6.3.2 Linear bias compensation

Following Section 4.4.2, we expect that the 3D gradient-based algorithm has systematic errors. The relation between 3D gradient-based estimates $(\widehat{x_d}, \widehat{y_d}, \widehat{z_d})$ and the unknown displacement (d_x, d_y, d_z) can be approximated by a simple linear model,

$$\begin{bmatrix} \widehat{x_d} \\ \widehat{y_d} \\ \widehat{z_d} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} = \mathbf{P} \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix}.$$

If the 3D gradient-based algorithm were unbiased, the parameter matrix $\mathbf{P}$ would equal the identity matrix. In general, however, the algorithm is biased and one must estimate $\mathbf{P}$ in order to improve the estimate.

Just as in the 2D case, we re-run the 3D gradient-based algorithm after shifting the second image by known integer amounts to obtain the information needed to estimate $\mathbf{P}$. All 8 corners of the shift cube containing the displacement (d_x, d_y, d_z) are performed to determine the shift at which the estimated displacement is smallest. The estimate from that corner is combined with the estimates from the corners that share an edge with the closest corner (see Figure 6-10) are combined to estimate $\mathbf{P}$. This problem has 12 non-linear equations and 12 unknowns (9 model parameters and the 3 displacements). Nevertheless, by taking pairwise differences between these equations, they are easily solved. For example, p_{12}, which represents the dependence on y of the x estimate, equals the difference between x estimates at the two shifts that share the y edge.

All of the 3D results in other chapters use this 3D linear bias compensated algorithm.

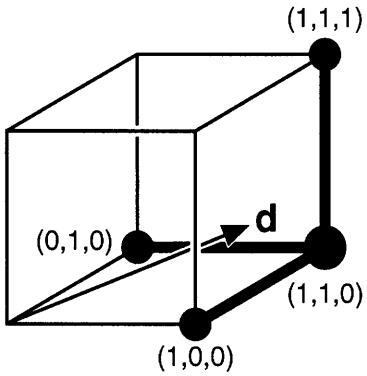


Figure 6-10: Selection of shifts for 3D linear bias correction. The **d** represents the 3D subpixel displacement to be estimated. The arrow indicates the distance that the 3D gradient-based algorithm has to estimate without shifting the second 3D image. In this example, the estimated distance is smallest if the second 3D image is shifted +1 pixel in the x and y directions. The 3D linear bias correction algorithm uses the gradient-based algorithm's estimates from the bolded corners to determine the coefficients in the parameter matrix **P**.

6.4 Extensions to six degrees of freedom

In this section we describe an algorithm for estimating general 3D rigid motion, which includes both translations and rotations. Unlike the derivations for estimating pure translations, we derive this algorithm in the continuous domain for both time and space. This simplification enables us to estimate velocities rather than displacements. Consequently, the discrete implementation of this algorithm requires the rotations and translations to be small. How small "small" is has not been investigated. This method is adapted from one described in Horn and Weldon, Jr. (1988).

Assuming that the brightness of the image is not affected by motion of the specimen (Horn and Schunck, 1981), the expansion of the total derivative of image brightness $\mathrm{E}(x, y, z, t)$ is

$$\frac{\partial \mathrm{E}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \mathrm{E}}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \mathrm{E}}{\partial z}\frac{\partial z}{\partial t} + \frac{\partial \mathrm{E}}{\partial t} = 0$$

By denoting $(\partial \mathrm{E}/\partial x, \partial \mathrm{E}/\partial y, \partial \mathrm{E}/\partial z)^T$ as $\mathbf{E}_r$, $(\partial \mathrm{E}/\partial t)$ as E_t, and $(\partial x/\partial t, \partial y/\partial t, \partial z/\partial t)^T$ as $\mathbf{r}_t$, the constant brightness assumption can be re-written as

$$\mathbf{E}_r \cdot \mathbf{r}_t + \mathrm{E}_t = 0.$$

$\mathbf{r}_t$ represents the rigid-body velocity of the imaged target, which is composed of a translational $\mathbf{v}$ and rotational $\boldsymbol{\omega}$ component:

$$\mathbf{r}_t = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{r}.$$

Combining these last two equations gives

$$\mathbf{E}_r \cdot (\mathbf{v} + \boldsymbol{\omega} \times \mathbf{r}) + \mathrm{E}_t = 0 \tag{6.9}$$

which in the absence of noise is valid for all (x, y, z) at each time t.

In the presence of noise, Equation 6.9 represents an over-determined system equations for the unknown translational $\mathbf{v}$ and rotational $\boldsymbol{\omega}$ velocities. This system can

be solved in a least-squares sense by minimizing the sum of the squares of errors:

$$\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\omega} \end{bmatrix} = \underset{\mathbf{v}',\boldsymbol{\omega}'}{\arg\min} \int\int\int [\mathbf{E}_r \cdot \mathbf{v}' + \mathbf{E}_r \cdot (\boldsymbol{\omega}' \times \mathbf{r}) + \mathrm{E}_t]^2 \, dx \, dy \, dz.$$

Using the triple-scalar product identity, $\mathbf{E}_r \cdot (\boldsymbol{\omega} \times \mathbf{r}) = (\mathbf{r} \times \mathbf{E}_r) \cdot \boldsymbol{\omega}$, we obtain

$$\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\omega} \end{bmatrix} = \underset{\mathbf{v}',\boldsymbol{\omega}'}{\arg\min} \int\int\int [\mathbf{E}_r \cdot \mathbf{v}' + (\mathbf{r} \times \mathbf{E}_r) \cdot \boldsymbol{\omega}' + \mathrm{E}_t]^2 \, dx \, dy \, dz.$$

Differentiating with respect to $\mathbf{v}$ and $\boldsymbol{\omega}$ and setting equal to zero leads to a pair of vector equations:

$$\left(\int\int\int \mathbf{E}_r {\mathbf{E}_r}^T\right) \mathbf{v} + \left(\int\int\int \mathbf{E}_r (\mathbf{r} \times \mathbf{E}_r)^T\right) \boldsymbol{\omega} = -\int\int\int \mathbf{E}_r \mathrm{E}_t$$

$$\left(\int\int\int (\mathbf{r} \times \mathbf{E}_r) {\mathbf{E}_r}^T\right) \mathbf{v} + \left(\int\int\int (\mathbf{r} \times \mathbf{E}_r)(\mathbf{r} \times \mathbf{E}_r)^T\right) \boldsymbol{\omega} = -\int\int\int (\mathbf{r} \times \mathbf{E}_r) \mathrm{E}_t$$

where the differential $dx \, dy \, dz$ has been omitted for conciseness, and $\mathbf{A}\mathbf{B}^T$ represents the outer product of the two vectors $\mathbf{A}$ and $\mathbf{B}$.

Assuming the symmetric, 6x6 matrix is invertible, the optimal solution in the least-squares sense is

$$\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} \int\int\int \mathbf{E}_r {\mathbf{E}_r}^T & \int\int\int \mathbf{E}_r (\mathbf{r} \times \mathbf{E}_r)^T \\ \int\int\int (\mathbf{r} \times \mathbf{E}_r) {\mathbf{E}_r}^T & \int\int\int (\mathbf{r} \times \mathbf{E}_r)(\mathbf{r} \times \mathbf{E}_r)^T \end{bmatrix}^{-1} \begin{bmatrix} -\int\int\int \mathbf{E}_r \mathrm{E}_t \\ -\int\int\int (\mathbf{r} \times \mathbf{E}_r) \mathrm{E}_t \end{bmatrix} \tag{6.10}$$

The conversion of the above continuous representation into a discrete representation is a direct extension of methods discussed in Horn and Schunck (1981). The partial derivatives of the image brightness must be estimated from the measured brightness $G[i, j, k, l]$. As demonstrated earlier in this chapter, the estimates for the four partial derivatives should be co-localized for best performance. We estimate the

derivatives at $(i+1/2, j+1/2, k+1/2, 1/2)$:

$$G_x = \frac{1}{8}\sum_{j'=j}^{j+1}\sum_{k'=k}^{k+1}\sum_{t'=0}^{1} G[i+1, j', k', t'] - G[i, j', k', t'] \sim \mathrm{E}_x$$

$$G_y = \frac{1}{8}\sum_{i'=i}^{i+1}\sum_{k'=k}^{k+1}\sum_{t'=0}^{1} G[i', j+1, k', t'] - G[i', j, k', t'] \sim \mathrm{E}_y$$

$$G_z = \frac{1}{8}\sum_{i'=i}^{i+1}\sum_{j'=j}^{j+1}\sum_{t'=0}^{1} G[i', j', k+1, t'] - G[i', j', k, t'] \sim \mathrm{E}_z$$

$$G_t = \frac{1}{8}\sum_{i'=i}^{i+1}\sum_{j'=j}^{j+1}\sum_{k'=k}^{k+1} G[i', j', k', 1] - G[i', j', k', 0] \sim \mathrm{E}_t$$

The rigid body motion can now be solved by discretizing Equation 6.10 and using the above derivative approximations. At least six linear independent equations are needed to solve for six unknowns in the least-squares sense. We expect in practice that many more equations are necessary to provide robustness.

The units of time in this derivation are in terms of the camera's frame time. Thus, the velocities estimated have units of pixels/frame or radians/frame, which is numerically equivalent to the translation and rotation between the two image volumes. Since the derivation used velocity, and velocity is the limit of a displacement or rotation per length of time as the length of time approaches 0, this method implicitly assumes that the translation and rotation between the two images is small. For arbitrary translations, the images are still related by linear functions. In contrast, for arbitrary rotations, the images are related by trigonometric functions which have to be linearized by a small angle approximation to obtain the above solution. Furthermore, rotations are order-dependent except for small angles.

Chapter 7

Using Video Microscopy to Characterize Micromechanics of Biological and Man-made Micromachines[1]

7.1 Abstract

Microscopic mechanics at the scale of individual sensory receptor cells play an important signal processing role in the inner ear. However, there are few direct measurements — in large part because the measurements are technically challenging. The important structures are small (micrometers) and the motions are even smaller (nanometers) as well as fast (audio frequencies from 20 Hz to 20 kHz). Furthermore, to relate the motions of different structures, simultaneous measurements from many targets are important.

We have developed a new method for visualizing and measuring the sound-induced motions of inner-ear structures. We use (1) a light microscope to magnify the image of the inner ear and to project the image onto a CCD camera, (2) optical sectioning

[1]This chapter is based on Freeman and Davis (1996).

to obtain sequences of images that characterize the entire 3D structure, and (3) stroboscopic illumination to create slow motion video sequences in response to audio-frequency stimuli. Even after optical magnification, sound-induced displacements of inner-ear structures are small compared to the pixel spacing of modern CCD cameras. Nevertheless, subpixel motions modulate the brightness field. We have modified robot vision algorithms to determine motion from changes in the brightness field.

Although this system was designed to study mechanics of the inner ear, it has broader application. We illustrate possible applications in MEMS with measurements of the three-dimensional motions of the teeth in a microfabricated comb drive.

7.2 Background

During the past 15 years, microelectromechanical systems have evolved from scientific curiosities to viable commercial products, holding promise for revolutionizing the fields of sensor and actuator design. Enormous progress in molecular biology and cell physiology during those same 15 years has enabled scientists to further understand the pervasiveness and elegance of design of biological microelectromechanical systems. Our research group is studying one such system — the cochlea, which is the part of the inner ear that responds to sound.

The cochlea is a remarkable signal processing unit. It has a single mechanical input: the piston-like motion of the stapes, which is the last in the chain of middle ear bones that move when sound strikes the eardrum. Therefore, one could view the cochlea as a simple displacement sensor — albeit, one with extraordinary sensitivity (von Békésy, 1960; Fay and Popper, 1994). Using neural signals generated by the cochlea, humans can reliably detect sounds that cause motions of the stapes on the order of picometers — i.e., on the order of the diameter of a hydrogen atom. However, the cochlea is more than a simple displacement sensor (Pickles, 1988). Its neural outputs manifest considerable signal processing. The cochlea performs a high-quality frequency analysis of low-level sounds: $Q_{10dB} > 600$ has been recorded for motions in

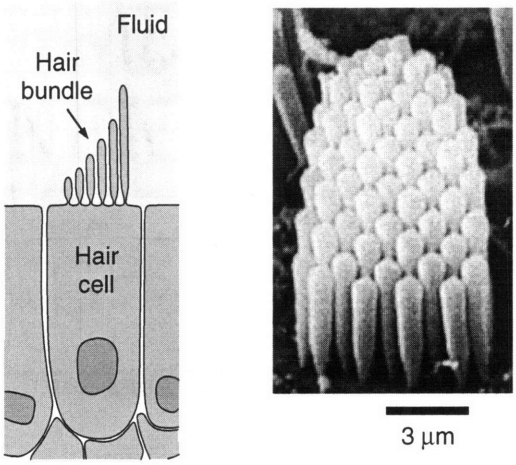


Figure 7-1: Structure of a hair cell. The left panel shows a schematic diagram of a hair cell and its mechanically sensitive hair bundle. The scanning electron micrograph in the right panel shows a surface view of a hair bundle in the cochlea of the lizard viewed in the direction of the arrow. Each bundle contains 50 to 100 sensory hairs.

the saline-filled cochlea of the bat (Kössl and Russell, 1995). Furthermore, cochlear signal processing is profoundly nonlinear, even at the threshold of hearing. One important nonlinearity is compression. Humans detect sounds differing in intensity over a 120 decibel range. This range of intensities is compressed as it is encoded for neurons with considerably smaller (20-50 dB) dynamic ranges. Our group is interested in understanding the micromechanical mechanisms underlying these signal processing capabilities of the cochlea.

Cochlear mechanics are organized hierarchically. Sensory reception in the cochlea is mediated by hair cells (Hudspeth, 1983; Hudspeth, 1989), which are specialized for mechanoreception by a microscopic bundle of mechanically sensitive hairs (Figure 7-1). Rotations of a hair bundle cause changes in the intracellular potential of the hair cell, thereby stimulating the release of a neurotransmitter that causes action potentials of cochlear neurons.

The mechanisms by which motions of hair bundles change intracellular potential involve more microscopic mechanical events. Like all cells (Weiss, 1996b; Weiss, 1996a), hair cells are surrounded by a cell membrane. The cell membrane consists of a lipid bilayer that acts as an electrical and chemical barrier between the inside

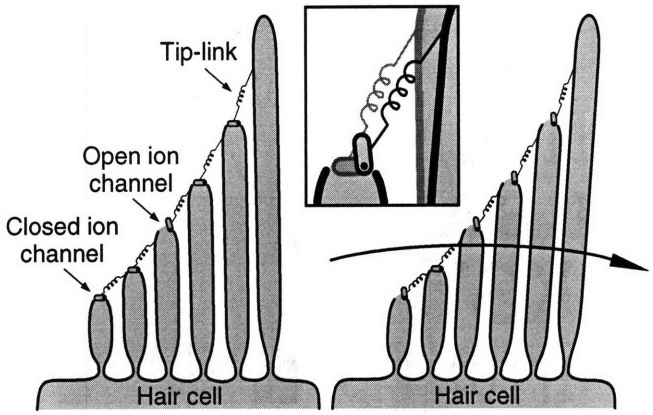


Figure 7-2: Schematic representation of stress-gated ion channels in hair cells. Rotation of the hair bundle (right) changes the mechanical stress in tip-links, which are represented here as springs connecting adjacent sensory hairs. The change in tip-link stress modulates the ionic conductivity of ion channels, schematized here as trap-doors. The inset shows a magnified view near one tip-link. The gray and black images were taken from the left and right panels, respectively.

and outside of the cell. In addition, the cell membrane contains integral membrane proteins. One class of membrane proteins, the ion channels, act as pores through which ionic currents can pass between the inside and outside of the cell. Hair cells contain ion channels whose state of ionic conductivity is modulated by mechanical stress. The mechanical stress that acts on these stress-gated ion channels is believed to be generated by filaments called tip-links (Pickles and Corey, 1992; Hudspeth and Gillespie, 1994), which connect the tips of adjacent sensory hairs (Figure 7-2).

Fluid dynamics plays a key role in cochlear micromechanics. Hair cells do not operate in isolation but are organized along with other cells and tissues to form a sensory epithelium (Figure 7-3). Displacements of the stapes cause pressures in the fluids that act to deform the sensory epithelium. As the tissues deform, shearing motions between the hair cells and overlying tectorial membrane cause rotations of hair bundles. Forces of fluid origin play a key role in determining motions of this coupled, hydromechanical system. Interaction of mechanical properties of the epithelium with inertial fluid forces accounts for much of the frequency selectivity exhibited in motion of the epithelium (Neely, 1981; Siebert, 1974; Viergever and Diependaal, 1986). Fluid forces play a key role in determining the motions of hair bundles (Freeman and

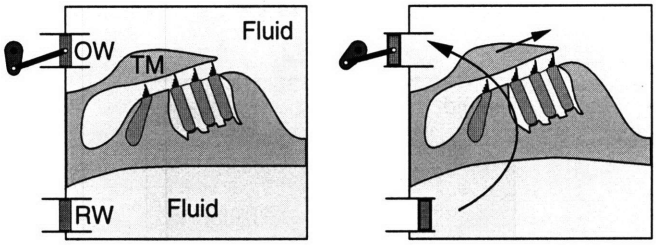


Figure 7-3: Schematic cross-section of the cochlea. The sensory epithelium and overlying tectorial membrane (TM) separate two fluid-filled compartments. One is coupled to the oval window (OW), into which the piston-like foot of the stapes fits; the other is connected to the round window (RW), which is covered by a thin membrane. The right panel illustrates deformations caused by motions of the stapes.

Weiss, 1988; Dallos et al., 1972) and contribute to a micromechanical resonance of some hair bundles of lizards (Freeman and Weiss, 1990).

Cochlear structures are extensively cross-coupled. Although organized hierarchically, the levels are not independent. For example, the state (open or closed) of stress-gated ion channels (nanometer scale) has been shown to affect the stiffness of the hair bundle (micrometer scale) (Howard and Hudspeth, 1988). Motions of hair bundles are coupled both mechanically and hydrodynamically to one another and to the sensory epithelium (millimeter scale). Through the surrounding fluids, hair cells can also interact chemically and electrically.

In humans, the cochlear micromachine consists of approximately 15,000 hair cells, each with more than 50 sensory hairs and associated tip-links in a saline-filled environment. The challenge is to understand how these parts — more than a million in each ear — interact to form a system with the remarkable signal processing capabilities of the ear.

7.3 Methods

We have developed a video system to measure sound-induced motions of cochlear structures in an in vitro preparation of the alligator lizard cochlea (Davis and Freeman, 1995). The cochlea is mounted in an experimental chamber so that it can be viewed with a microscope while simultaneously being stimulated by an underwater

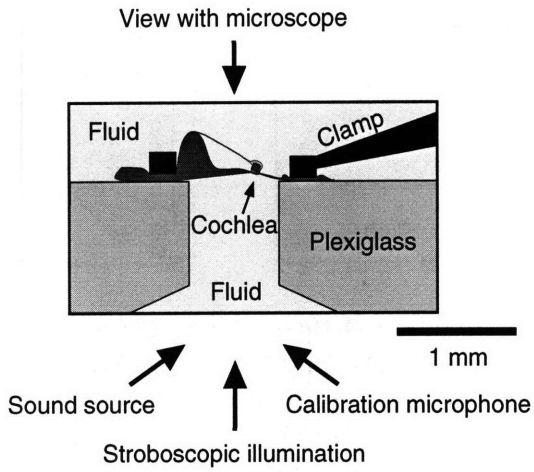


Figure 7-4: Cross-sectional view of the experimental chamber. The cochlea is clamped so that it seals a circular opening between two fluid-filled compartments. The cochlea is stimulated by a piezoelectric disk that generates hydrodynamic pressures in the bottom compartment; the top compartment is vented to the atmosphere. Sound pressure is calibrated using an underwater microphone located just under the cochlea.

pressure source (Figure 7-4). The chamber is mounted on the stage of a light microscope (Zeiss Axioplan, New York, NY) that projects magnified images of cochlear structures onto a scientific grade CCD camera (Photometrics 200 series with a Thompson 7883 CCD, Tucson, AZ). The camera has 576 $\times$ 384 pixels with 23 μm spacing. Total magnification between the target and camera can be adjusted from 5 to 600.

Optical sectioning. Many biological tissues, including hair cells, are nearly transparent. Images can be obtained from any part of such a tissue by altering the microscope's focus control. Thus images from a series of focal depths can be combined into a three-dimensional image that characterizes all of the structures in the cochlea, including hair cells, hair bundles, and the tectorial membrane. Optical sectioning can also be used to generate three-dimensional images of opaque targets. However, only unoccluded surface structures are visible. To automate the imaging of three-dimensional structure, a computer-controlled stepper motor is used to adjust the focus control of the microscope with a resolution of 1/11 μm per step.

Stroboscopic illumination. The important motions of cochlear structures are at audio frequencies, i.e., from 20 Hz to 20 kHz. These frequencies exceed the maximum

rates of scientific-grade (12-bit) video imagers (it takes 3 seconds for our system to digitize a full frame image and save it to the disk). Therefore, stroboscopic illumination is used to slow the apparent rate of motion. We typically use a gas discharge strobe lamp (Chadwick-Helmuth Co. Inc., El Monte, CA) which generates high intensity illumination. However, the light from a gas discharge strobe is not spatially uniform and pulse-to-pulse variations can cause motion artifacts. Therefore, light from the strobe is coupled to the microscope through a fiber optic scrambler (Technical Video Ltd., Woods Hole, MA). The scrambler also isolates our system from the considerable electromagnetic interference that is generated by the strobe. The duration of the strobe pulse (8 μs) limits the maximum frequency of motion that our system can measure. To relax this constraint, we have recently developed a stroboscopic light source using an LED (AND Division of Purdy Electronics Corp., Sunnyvale, CA). Although they generate less intense light, LEDs radiate much less electromagnetic interference and are orders of magnitude faster.

Subpixel motion estimates from video images. Fundamental limitations on image resolution and on motion detection from images are different. Image resolution is limited by the optics (Hopkins and Barham, 1950). To a first approximation, images are blurred by convolution with a point-spread function (Goodman, 1968; Born and Wolf, 1975), which limits resolution to distances on the order of the wavelength of light (typically 550 nm). However, one can detect motions of a blurred image that are significantly smaller. The fundamental limitations on motion detection are determined by the quantum nature of light (shot noise) and by the video imager.

Subpixel displacements can be estimated from changes in brightness (Figure 7-5) using methods originally developed for robot vision (Horn, 1986). Generally we identify regions of interest that contain a single moving structure. Images at successive times are then analyzed to determine the best rigid-body translation in a least-squares sense (Horn and Weldon, Jr., 1988).

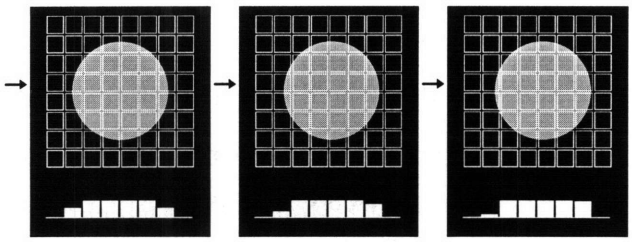


Figure 7-5: Variations in brightness caused by subpixel motions. The array of squares in each panel represents the pixels of a video camera; the gray disk represents the image of a simple scene; and the bar plots represent the brightness of the row of pixels indicated by the arrow. The three panels illustrate rightward motion of the disk. Although the displacements are subpixel, information about the motion is available from changes in the brightness of the pixels near the edges.

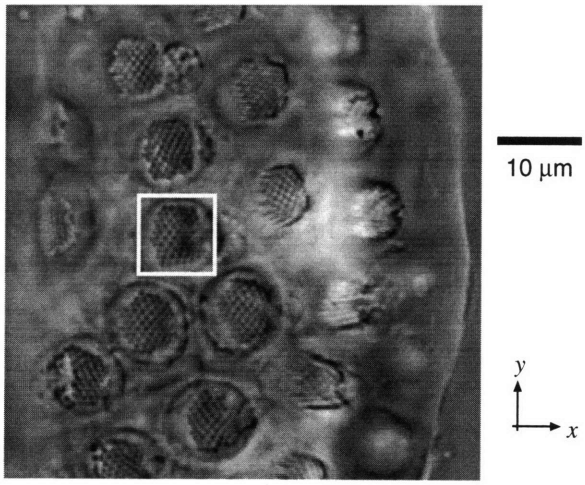


Figure 7-6: Cochlear structures in the alligator lizard. This figure shows a portion (240 × 240 pixels) of a video image of the lizard cochlea. The white square surrounds one hair bundle; several other hair bundles are also visible. The faint vertical structure at the right edge of the image is the edge of the tectorial membrane (see Figure 7-3). Total optical magnification from specimen to video camera was 100×.

7.4 Results

7.4.1 Sound-induced motions of cochlear structures

Figure 7-6 shows the cochlea of an alligator lizard as imaged with our system. The hair bundles of hair cells and the overlying tectorial membrane are clearly visible. Generally 30 to 60 hair cells are contained in the field of view. Approximately 20 can be seen in Figure 7-6, which shows only a portion of the camera's field of view.

Because the cochlea is nearly transparent, one can view structures both above and

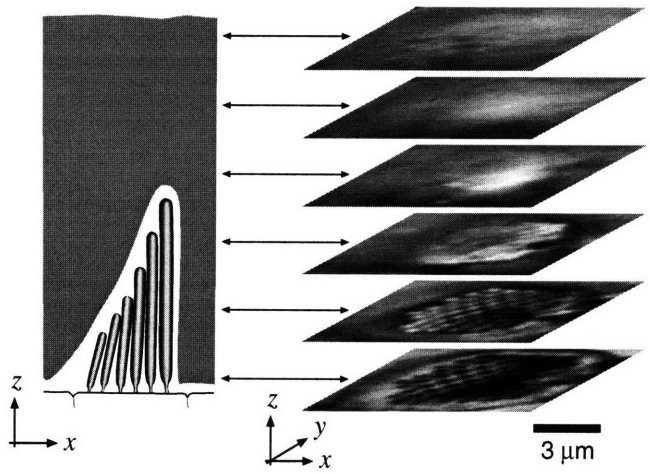


Figure 7-7: Three-dimensional image of a hair bundle. The bottom 2D image shows a magnified and tilted version of the highlighted part of Figure 7-6. The other images are from a sequence of planes, each 3 μm above the previous one. The double-headed arrows associate each image with the corresponding plane of section in the schematic drawing on the left. The scale bar applies to both the x and z directions; y is foreshortened.

below those shown in Figure 7-6 by changing the plane of focus. In this experiment, images were acquired at 13 planes of focus separated by 3 μm. In general, one can obtain sufficiently many images to determine the three-dimensional structure of any visible structure in the cochlea. Figure 7-7 illustrates how the sequence of two-dimensional images can be combined to obtain a three-dimensional image of a hair bundle and overlying tectorial membrane.

Three-dimensional images of the cochlea were obtained at 8 phases of the sinusoidal stimulus. The resulting data set consisted of 104 images of 576 $\times$ 382 pixels (13 planes times 8 phases) and total data collection time was 5 minutes. This amount of time is large compared to the time needed to determine the motion of a single structure using other measurement methods, such as laser interferometry. However, from the resulting 104 images, one can estimate the three-dimensional displacement waveform for *every* resolvable structure in the field of view. Thus the total measurement time is in fact extraordinarily short compared to the time it would take to make motion measurements of each structure using another method. We illustrate the utility of this feature by showing how two separate analyses of the same data set may be used to address two issues at different spatial scales — the relation between hair

bundle motion and that of the overlying tectorial membrane, and relations among the motions of neighboring sensory hairs within a hair bundle.

To determine motions of the hair bundle and overlying tectorial membrane, we analyzed motions of the structures in the 6 images of Figure 7-7. Displacements near the base of the hair bundle were larger than those near the tip (Figure 7-8), which demonstrates that the hair bundle is rotating about its point of attachment to the hair cell. Displacements of the tectorial membrane were generally smaller than those of either the base or tip of the hair bundle. Furthermore, as distance from the base of the hair bundle increases, the motion shows increasing phase lag.

To get a quantitative estimate of hair bundle rotation, each displacement waveform was subtracted from the displacement waveform for the base. The results show that the displacement of the tip of the hair bundle relative to the base is 0.3 μm peak-to-peak. Since the planes through the tip and base are separated by 6 μm, hair bundle rotation is approximately 3° peak-to-peak. Motions of five other hair bundles in this data set were analyzed and the results were qualitatively similar to those shown in Figure 7-8.

In all of our 12 experiments to date, displacements of the tips and bases of hair bundles have been larger than those of the overlying tectorial membrane for frequencies throughout the lizard's auditory range. This result is consistent with the idea that inertial forces arising from the mass of the tectorial membrane and surrounding fluid tend to resist motion of the tectorial membrane (Mammano and Nobili, 1993). This result is not consistent with other theories that suggest that the tectorial membrane is a resonant structure (Allen, 1980). However, our results should be regarded as preliminary. Control experiments must be performed to assure that the cochlea is not damaged during experimentation. Regardless of the ultimate resolution of this particular issue, the important point is that video microscopy can be used to make measurements to directly test theoretical conceptions.

The same data set used in the previous analysis can also be used to study mechanical events at the level of individual sensory hairs. The sinusoidal stimulus generates nearly sinusoidal motions of all the sensory hairs (top right panel of Figure 7-9). To

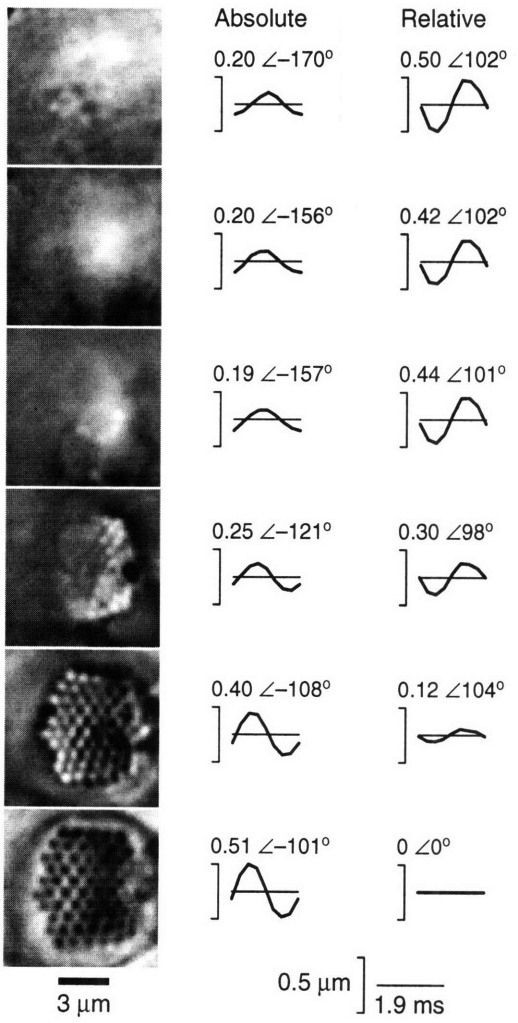


Figure 7-8: Sound-induced motions at six planes of focus through a hair bundle and overlying tectorial membrane. Images (left column) corresponding to those illustrated in Figure 7-7 were obtained at 8 phases of the stimulus period and were analyzed to quantify motion at each plane of section. The waveforms in the center column show average displacements in the x direction during one cycle of the stimulus. The associated numbers are the peak-to-peak magnitude (in μm) and angle (in degrees) of the fundamental component of the displacement. The waveforms in the right column show differences between the waveforms in the center column and the bottom waveform in the center column. The intensity of the 513 Hz stimulus (115 dB SPL in the fluid adjacent to the basilar membrane) corresponds to approximately 89 dB SPL at the eardrum.

determine if there is relative motion between the sensory hairs, we applied our motion detection algorithm to estimate the motion of the right edge of the bundle and shifted the images to compensate for that motion. The resulting images show relative motions between sensory hairs (bottom right panel of Figure 7-9).

Although tip-links are too small to be resolved in a light microscope, the results in Figure 7-9 show that we can directly measure the motions of the sensory hairs to which the tip-links are attached. The relative motions of the tips of the sensory hairs give direct estimates of tip-link stretch, which is the physiologically important mechanical input for transduction mechanisms at the nanometer scale. Although preliminary, our results suggest that tip-links are stretched not only because the distance between insertion points increases when the hairs are rotated by equal angles (as illustrated in Figure 7-2), but also because the angular displacement of shorter hairs is greater

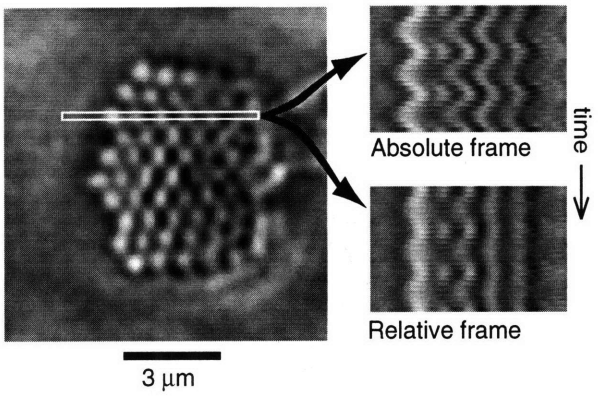


Figure 7-9: Motions of sensory hairs within a bundle. The left panel is an enlarged image of the hair bundle in Figure 7-7, at a plane of section near the center of the bundle (second from bottom). The highlighted region was extracted from each of eight images acquired during the stimulus period and positioned so that regions from earlier phases are on top of regions from later phases. The resulting montage, which illustrates motions during one period, is repeated three times to generate the images in the right panels. The upper and lower right panels are similar, except that the images used to generate the lower panel were first shifted to cancel the motion of the right part of the hair bundle.

than that of longer hairs. Motions of individual sensory hairs have been measured elsewhere (Duncan et al., 1995; MacDonald and Corey, 1996) but only for hair bundles that were stimulated with a water jet; the membranes through which natural stimuli are delivered had been removed.

7.4.2 Three-dimensional motions of a comb drive

Although originally developed to measure motions of cochlear structures, our video system has broader application. We have performed several pilot experiments to determine if it can be used to measure the motions of man-made micromechanical devices.

Figure 7-10 shows an image of a microfabricated accelerometer obtained using our system. The comb drive was driven with an 8.8 Hz sinusoidal voltage, 44 volts peak-to-peak with an 80 volt DC offset. Images were obtained at 45 planes of focus with 0.45 μm spacing to generate a three-dimensional image. Three-dimensional images taken at 8 phases of the sinusoidal stimulus were analyzed to determine the motions of the stationary and fixed teeth highlighted in Figure 7-10. Displacements of the moving tooth were approximately 2 μm peak-to-peak in the horizontal direction (left

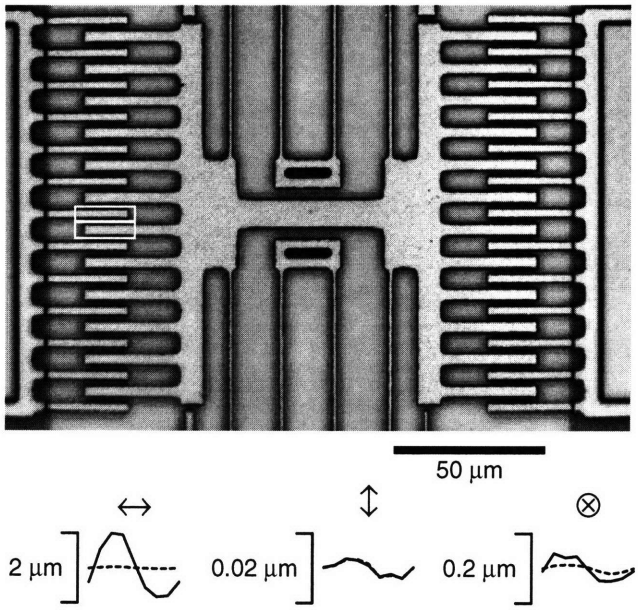


Figure 7-10: Motions of teeth in a comb drive. The field of view of this image (64× magnification) of a microfabricated accelerometer is focused on the central shuttle and associated comb drives. Most of the folded-beam cantilever that supports the shuttle lies above and below the field of view; only a small portion of it is visible near the center of the image. The portions of the images in the white boxes were analyzed to determine the motions of a stationary (top) and moving (bottom) tooth of the comb drive, and results are shown in the three plots as dashed and solid curves, respectively. The left, center, and right panels show average displacements in the horizontal, vertical, and out-of-plane directions, respectively.

plot, solid line). These displacements are significantly larger than the corresponding displacements of the fixed tooth, which serve as control measurements to assess the stability of the measurement system. Vertical motions (center plot) were about 6 nm peak-to-peak for both the fixed and moving teeth, which is near the noise floor of our measurement system (without averaging).

Our system can also measure out-of-plane motions. Although one cannot use optical sectioning to "see through" opaque structures, three-dimensional images of opaque structures still provide information about three-dimensional structure and motion, as illustrated in Figure 7-11. Image brightness is greatest when the tooth is "in focus" and decreases as the plane of focus is moved higher or lower. Our motion detection algorithm takes advantage of these changes in brightness to align three-dimensional images of the same tooth taken at different phases of the sinusoidal stimulus. Results indicate that the out-of-plane displacement of the moving tooth was

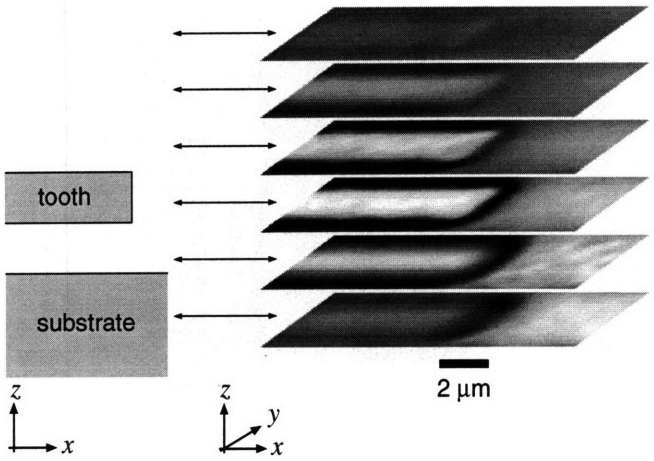


Figure 7-11: Three-dimensional image of a tooth in a comb drive. This figure illustrates 6 images of the stationary tooth highlighted in Figure 7-10. The images are at different planes of focus, separated by 2.27 μm. The double-headed arrows associate each image with the corresponding plane of section in the schematic drawing on the left. The scale bar applies to both the x and z directions; y is foreshortened.

80 nm peak-to-peak. The out-of-plane displacement of the stationary tooth was much smaller (26 nm peak-to-peak).

7.5 Discussion

We have developed a video system that can be used to characterize the motions of both biological and man-made micromechanical systems. Using stroboscopic illumination, we observed motions over the entire audio range of frequencies, and can in principle observe motions at much higher frequencies. Using the optical sectioning property of a light microscope, we measured three-dimensional motions of both transparent (biological) and opaque (silicon) structures. Quantitative results were obtained directly from the video images, based on motion detection algorithms originally developed for robot vision.

7.5.1 Implications for hearing research

Using the video system, we have obtained the first direct experimental observations of the relation between sound-induced motions of the tectorial membrane and hair bundle displacement. We have also obtained the first direct experimental observations of sound-induced motions of sensory hairs within a bundle. Although these results must be regarded as preliminary until control experiments have tested the viability of our preparations, they demonstrate how video microscopy can be used to probe important cochlear mechanisms at a variety of spatial scales: from ion channels (nanometers) to hair bundles (micrometers) to the entire cochlea (millimeters). Video microscopy has opened the experimental door to understanding key relations in cochlear mechanics that have previously been addressed only theoretically.

7.5.2 Implications for MEMS

Tools for characterizing the performance of man-made micromachines are primitive by comparison to those available for electronic devices. For many micromachines, tools for assessing internal failure modes do not exist, and only the terminal behaviors can be tested. The inability to characterize internal failure modes has important consequences for both the designer and fabricator. For the designer, it can result in large numbers of prototypes. For the fabricator, it means reduced yields because common techniques such as laser trimming cannot be applied. As the complexity of micromachines increases and as their production becomes a commercial endeavor, developing better tools for characterizing the performance of micromachines will become crucial. Results in this paper demonstrate the potential use of video methods for characterizing the performance of MEMS devices.

7.5.3 Comparison to other motion measurement systems

A number of systems have been developed to measure micromechanical motions. Our method is different from and has several advantages over previous methods, including the Mössbauer method (Johnstone and Boyle, 1967), photodiode method (Crawford

and Fettiplace, 1985), capacitive probe (Wilson and Johnstone, 1975), and laser interferometry (Khanna and Leonard, 1982; Ruggero and Rich, 1991). (1) Our motion estimates are derived directly from sequences of images. Thus all of our quantitative results can be checked qualitatively by simply viewing the slow-motion video sequences. For example, to test that the displacement estimates in Figure 7-8 were qualitatively correct, each image was shifted by the estimated displacement in a direction to compensate for the motion. The resulting sequence of images, which should be stationary, provided a test that motions were estimated properly. (2) In the video method, the identity of the target whose motion is measured is unambiguous since motion is determined directly from images of the target. All the other methods listed above require additional observations to determine which motion is being measured. (3) Unlike any of the other methods listed above, the video method allows simultaneous measurement of motions of many structures in the same specimen. This is a significant advantage when determining interrelations among motions of different structures. (4) The video method provides estimates of all three components of rigid body translation for every structure in the image. The other methods listed above provide an estimate of just one component of motion. (5) In contrast to some methods (e.g., the Mössbauer method, some laser interferometric methods (Cooper and Rhode, 1992) and some photodiode methods (Crawford and Fettiplace, 1985)) our method does not require any mechanical contact with the moving specimen. Mechanical contact has been shown to affect motions of cochlear structures (Khanna and Leonard, 1986), and it seems inescapable that similar degradations would occur in high-performance man-made micromachines.

The video method has one important disadvantage: it is particularly sensitive to vibrations of the microscope and camera. To reduce vibrations, we use a pneumatic vibration isolation table, which very effectively reduces vibrations at high frequencies but has little effect for low frequencies (<10 Hz). Although low frequency vibrations contaminate the output signals of all motion detectors, the contamination can generally be filtered out — provided that the frequencies of interest are sufficiently different from the contamination frequencies. However, in the video method, images are ac-

quired at a low rate, typically 0.3 Hz. Low frequency vibrations of the microscope or camera cause displacements of the images that cannot be distinguished from motions near the frequency of interest. These vibrations determine the noise floor of our measurements, which under optimal conditions (e.g., high contrast target, no nearby structures with confounding motions) is approximately 5 nm without averaging — i.e., 1% of the wavelength of the light that generates the images. With averaging we have reduced the noise to less than 1 nm. We are currently working on other methods to reduce this noise floor.

7.6 Acknowledgments

The authors thank Thomas F. Weiss for his contributions throughout this project, RuthAnne Eatock who prepared the scanning electron micrograph shown in Figure 7-1, and A. J. Aranyosi for helpful comments on the manuscript. This work was supported by grants from the NSF and NIH (R01 DC00238 and P01 DC00119).

Chapter 8

Direct Observations of Inner-ear Mechanics

8.1 Introduction

All vertebrates transduce the mechanical energy in sound into an electrical representation at sensory cells called hair cells (e.g., Howard et al., 1988). Hair cells have a mechanically sensitive tuft of sensory hairs (hair bundles) that protrude into the fluids in the inner ear. In most cochleae including all mammals, a gelatinous membrane called the tectorial membrane (TM) surrounds the hair bundles. Because of the close proximity to the transduction apparatus (the hair bundles), the TM is thought to play a critical role in cochlear function (e.g., Davis, 1958; Dallos et al., 1972; Allen, 1980; Freeman and Weiss, 1988; Mammano and Nobili, 1993).

Although the tectorial membrane plays a prominent role in theories of cochlear function, we are the only group to have published measurements of sound-induced motions of both the TM and underlying hair bundles (Davis and Freeman, 1995; Freeman and Davis, 1996). The paucity of direct measurements is due to the difficulties in measuring sound-induced motions: the motions that result from moderate sounds are smaller than 1 μm and the tectorial membrane is almost transparent.

This chapter describes in detail our work done to measure sound-induced motions of an in vitro preparation of the alligator lizard cochlea, including motions of the

tectorial membrane (TM) and underlying hair bundles.

The lizard was chosen as the experimental animal for many reasons. First, the lizard cochlea is planar (Mulroy, 1974) and therefore simple to image with a microscope. Second, the lizard cochlea has two distinct anatomical regions—one with and one without a TM. Thus, insight into the functional significance of the TM can be obtained by comparing responses from the two regions. Third, there exists a large body of information about cochlear function both proximal and distal to the micromechanics in the lizard cochlea. This information along with measurements of the sound-induced cochlear motions can be used to understand the entire peripheral auditory system. Fourth, the lizard shares non-linear properties such as two tone suppression (Holton and Weiss, 1978; Holton, 1980) and a compressive non-linearity (Holton and Weiss, 1983) with mammalian cochleae. Consequently, understanding the lizard cochlea may provide insight into mammalian hearing as well.

An in vitro preparation was chosen primarily for imaging considerations. In situ, one has difficultly imaging the basilar papilla and cannot image hair bundles or the TM. In contrast, in vitro, individual sensory hairs of hair bundles and the TM can be resolved.

This work included four major areas: (1) designing the apparatus for holding, stimulating, and measuring motions of in vitro preparation, (2) developing the in vitro preparation, (3) measuring sound-induced motions, and (4) developing measurement protocols for addressing important issues in cochlear physiology. As soon as the apparatus and in vitro preparation were sufficiently advanced to make sound-induced motion measurements, work in the four areas proceeded in parallel. Parallel development enabled us to identify the most significant technical hurdles more quickly. However, most of our sound-induced motion measurements were not done with our most recent experimental apparatus. Furthermore, we have only very recently been able to measure an in vitro electrical response from the mechanoelectric transduction of sound. The methods for maintaining an electrical response in vitro are still being worked out; consequently, none of the motion measurements were done on cochleae that could transduce sound. Nevertheless, the motion measurements performed are

the first report on sound-induced motions of the tectorial membrane and underlying hair bundles, and they demonstrate how these methods can be used to address important issues in cochlear mechanics.

With motion measurements of the tectorial membrane (TM) one could determine if the TM moves as a rigid body at auditory frequencies. The TM has been modeled both as a rigid body (e.g., Davis, 1958; Freeman and Weiss, 1988) and as a compliant structure (e.g., Allen, 1980).

With additional measurements of the reticular lamina, one could determine the relation between TM and reticular lamina (RL) motion. In particular, by comparing the relative motion magnitudes, one could determine if the TM is resonant as hypothesized by Allen (1980). While the ratio of the motions of the TM to the RL is useful for determining which structure moves the most, that ratio is not the most significant for understanding the functional role of the TM. Relative motions of the tectorial membrane and reticular lamina is thought to play a key role in determining the angular displacement of hair bundles. To assess the functional role of the TM, one could compute the difference between the radial displacements of the TM and RL. To account for the variability in the magnitude of the motion in different preparations and at different frequencies, each difference between displacements of the TM and RL is normalized by the corresponding RL displacement to yield the micromechanical transfer function used in previous theoretical studies (Allen, 1980; Freeman and Weiss, 1988). With this transfer function, one could determine whether the frequency selectivity found on the auditory nerve (Weiss et al., 1976) is the same as the frequency selectivity of the relative motion between the TM and RL.

A widely held belief in cochlear mechanics is that the rotation of hair bundles that are in contact with the tectorial membrane is proportional to the relative motion between the TM and RL. For example, Davis (1958) hypothesized that all hair bundles are deflected by differential displacement between the TM and RL. When later evidence suggested that mammalian inner hair cells are not attached to the TM, Dallos et al. (1972) modified Davis's (1958) model for inner hair cells, but restated the claim that outer hair cells are deflected by differential displacement between

the TM and RL. This belief is also held in more modern theories of mechanics (e.g., Allen, 1980; Freeman and Weiss, 1988; Neely, 1993). Experimentally, Hemmert et al. (1995) measured motions of the TM and basilar membrane, but their method precluded measurement hair bundle or reticular lamina motion.[1] Consequently, in order to relate their measurements to hair cell stimulation, they also have to assume some relation between TM motion and hair bundle rotation.

With additional measurements of the tips of the hair bundles underlying the TM, one could determine the relation between motions of the TM, RL, and tips of the hair bundles. By taking the ratio between radial displacements of the TM and tips of the hair bundle, one could determine if the they are in fact the same. Furthermore, one could also directly compute the relation between hair bundle rotation and TM/RL shear by dividing the difference between the displacement of the tips and the RL by the difference between the displacement of the TM and the RL. With anatomical knowledge that the TM covers the tips of the hair bundle, these ratios can also determined whether the TM shears throughout its thickness, which is a mode of motion not previously reported.

8.2 Methods

8.2.1 Motion-measurement apparatus

The small displacements that we wish to measure place special constraints on the video-imaging system. Even after magnification by a light microscope, the displacements generate only small intensity differences in the individual pixels of the CCD camera. These small intensity differences, and therefore motion detection algorithms based on those differences are acutely sensitive to imaging degradations such as those caused by shot noise, dark noise, read noise, charge transfer efficiency, and inter-pixel gain variations (See Janesick et al. (1987) for details on CCD cameras). Accordingly,

[1] They placed high-contrast marker beads on the TM and basilar membrane and used both laser doppler and a photodiode pair to estimate the beads' motion. Beads could not be placed on the reticular lamina or hair bundles; therefore, their motion could not be measured.

we use a scientific-grade CCD camera (Photometrics 200 series, 12 bit dynamic range, Photometrics AZ) to improve the system's motion resolution. Since the frame rate of scientific-grade CCD cameras is very slow compared to audio frequencies, the inner ear must be stroboscopically illuminated to slow its apparent motion.

To prevent spatial aliasing by the CCD camera, the image of the specimen must be bandlimited by a spatial low pass filter. Since all microscopes have finite-sized apertures, they all bandlimit the images they transmit. Accordingly, we use microscope optics (Zeiss Axioplan with a 40x, 0.75 NA water immersion objective) that (1) have the greatest resolution (highest spatial cutoff frequency) and (2) have enough magnification to prevent spatial aliasing.

The signal for the acoustic drive and the strobe light were generated by a two-channel, 12 bit D/A converter (DT 1497, Data Translations). The acoustic drive was temporally low passed filtered with a passive 9 kHz reconstruction filter (J1202, TTE, Inc. Los Angeles, CA), amplified by a 40 dB amplifier, and powered a piezoelectric disk attached to the experimental chamber. The piezoelectric disk converted the electrical energy into acoustic energy in the experimental chamber's saline. The other channel of the D/A converter is used to trigger the strobe light at known phases of the stimulus frequency, so that stop-action pictures can be taken.

With the above hardware, stroboscopic pictures can be taken with an average rate of one picture every three seconds. We will now describe some aspects of the motion measurement system in more detail.

Illumination

Stroboscopic illumination is used to slow the apparent motion of the cochlea. We use two different sources for stroboscopic illumination. The first is a gas-discharge tube (Model 8440, Chadwick-Helmuth, CA), and the second is an LED[2].

Three types of instabilities are common in strobe lights: temporal, intensity, and

[2]Typically we use a 18 candela, 590 nm, non-diffused LED with a 4° viewing angle (AND190HYP, AND division of Purdy Electronics, CA) or a 1 candela diffused red LED with a 60° viewing angle (CMD53SRD/G, Chicago Miniature Lamp, Buffalo Grove, IL).

spatial instabilities. Temporal instabilities (variations in the time between the trigger and the flash) cause the specimen to be illuminated at the wrong time, leading to incorrect motion measurements. Compensating for temporal instabilities is extremely difficult; accordingly, both of the devices that we use have very small temporal instabilities ($\ll 1$ μs). The gas-discharge tube followed by a broad-band green filter has an instantaneous brightness about 4,000 times that of the LED; however, the strobe light has much greater intensity and spatial instabilities.

Gas-discharge strobe. The intensity instabilities (variations in the brightness of individual flashes) for the strobe light are typically 4.5%. Since small motions do not change the entire picture's average brightness, the average brightness of all the pictures can be normalized to have the same value. Thus, problems resulting from intensity instabilities are largely eliminated and typically do not cause errors in motion estimation.

Spatial instabilities, caused by wandering of the strobe light's arc, result in a nonuniformly-illuminated specimen and the structure of the nonuniformity changes with every picture. Since subpixel motion detection algorithms interpret changes in intensity as motion; spatial instabilities must be minimized. The spatial instabilities are minimized through the use of a fiber optic scrambler (Technical Video, Ltd., Woods Hole, MA).

If more than one strobe flash per picture is desired (e.g., to increase the light level, or to average out substrate vibrations, or to average out intensity instabilities), the strobe light can flash up to 500 Hz. However, above 70 Hz the total amount of light received no longer increases linearly with the number of flashes. Below a 70 Hz flash rate, the coefficient of variation (standard deviation divided by the mean) for the average picture brightness is 4.5% for 1 flash, $< 3.8\%$ for 2 flashes, and $< 1.5\%$ for 25 flashes per picture.

The gas-discharge strobe has a pulse-width (full width at half the maximum light output) of 12–23 μs, depending on the firing rate.[3] Thus the sampling theorem implies

[3]Above 140 flashes per second the pulse width is 12 μs; below 70, 23 μs. We typically use the

that the gas-discharge strobe cannot be used to sample signals above 41–22 kHz. We typically acquire images at 8 phases of motion; therefore, the strobe light starts to limit the independence of the 8 phases above 10–5 kHz (depending on the firing rate).

LED. The LED is placed in a tungsten lamp house at the same position as a tungsten bulb. The LED is driven by a computer-controlled D/A converter via a voltage-to-current converter. The intensity of the emitted light from an LED is nearly linear with respect to the current through the device, and has a response time in the 10s of nanoseconds. Because LED's are so fast, the shortest light pulse width is limited in practice by the electronics driving the LED to roughly 1 μs. The lens on the LED makes the illumination almost critical rather than Kölher (Inoué, 1986). The LED's die is visible in the specimen plane; a low-angle diffuser is used to make the light more uniform. LED's with integrated diffusers can also be used, and preliminary results indicate that they yield more consistent motion estimates along the optic axis. Nevertheless, the only diffused LED that is bright enough for motion measurements emits red light. Red light reduces the microscope's resolution (Born and Wolf, 1975). The apparent size of the LED at the specimen plane of the microscope depends on the magnification used and the viewing angle of the LED. Typical viewing angles used are 4–15°; smaller viewing angles increase the brightness but decrease the illuminated field of view.

Unlike the gas-discharge strobe, the LED must be flashed many times in order to generate sufficient light. Exposure times on the order of 62 ms are needed at 100× magnification for the lizard preparation. In order to achieve this amount of light, the LED is pulsed on with typically a 1/8th duty cycle for 0.5 seconds (See Figure 8-1). Pulsing the LED on and off over 0.5 seconds has the advantage of averaging out the substrate vibrations (Figure 5-1).

The LED has negligible spatial, temporal, and intensity instabilities, making it the strobed light source of choice. On the other hand, in high light-level applications (e.g., very high magnifications or polarized microscopy), the gas-discharge strobe can

strobe at the lowest firing rate setting.

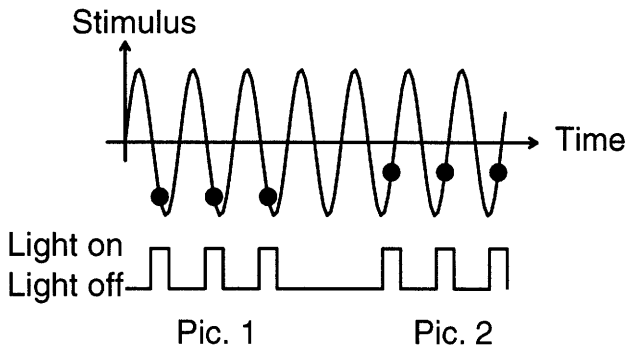


Figure 8-1: Light integration. The top curve shows a time waveform of the electrical stimulus to the specimen. The bottom curve shows the command voltage (that controls the current) to the LED. The LED is turned on multiple times at the same stimulus phase before the camera reads out the image.

generate 30 times ($4000 \times 14 \cdot 10^{-6} \times 70/.125$) more light per second than the brightest LED.

Magnification

We analyze the effects due to diffraction to find the minimum microscope magnification consistent with the sampling theorem. To the extent that other optical degradations cause more blurring than the diffraction limit, even smaller magnifications may be acceptable.

The amplitude point spread function for the in-focus plane $h(x, y)$ is (Born and Wolf, 1975)

$$h(x, y) = I_o \frac{\lambda^2 q^2}{\pi a^2 r} J_1\left(\frac{a 2\pi r}{\lambda q}\right)$$

where I_o is the average light intensity, a is the aperture radius in the objective, q is the objective's focal length, and λ is the wavelength of the light. For the Zeiss 40×, 1.95 mm, water immersion objective, the aperture radius is 3.1 mm and focal length is 4.11 mm (Hohman, 1997). Using $\alpha = \lambda q/a$, the Fourier transform of the amplitude PSF $H(f_x, f_y)$ can be determined:

$$\begin{aligned} h(x, y) = \left(I_o \frac{\lambda^3 q^3}{\pi^2 a^3}\right) \frac{1}{\alpha^2 (r/\alpha)} J_1\left(\frac{2\pi r}{\alpha}\right) &\overset{\mathcal{F}}{\leftrightarrow} \\ H(f_x, f_y) = \left(I_o \frac{\lambda^3 q^3}{\pi^2 a^3}\right) \operatorname{circ}(\alpha f) & \end{aligned} \tag{8.1}$$

where

$$\operatorname{circ}(r) = \begin{cases} 1 & r \leq 1 \\ 0 & r > 1 \end{cases}$$

and $\overset{\mathcal{F}}{\leftrightarrow}$ indicates Fourier transform pairs.

The output from the camera is a function of the light's intensity as opposed to the light's amplitude. Thus, the amplitude PSF has to be squared to get a physically meaningful result; the intensity PSF $k(x, y)$:

$$k(x, y) = h^2(x, y) \overset{\mathcal{F}}{\leftrightarrow} K(f_x, f_y) = H(f_x, f_y) * H(f_x, f_y)$$

where $*$ denotes convolution.

The amplitude of the intensity PSF $k(x, y)$ monotonically decreases with distance from the origin. Because $H(f_x, f_y)$ is non-zero for $\alpha f < 1$, $K(f_x, f_y)$ is non-zero for $\alpha f < 2$. Thus the intensity PSF $K(f_x, f_y)$ is bandlimited to

$$f = \sqrt{f_x^2 + f_y^2} < \frac{2}{\alpha} = \frac{2a}{\lambda q}.$$

For the Zeiss 40× objective, the intensity PSF is bandlimited to 2.74 μm^{-1}, assuming that $\lambda = 550\text{nm}$. The Nyquist sampling frequency f_s is twice the highest frequency in the signal or 5.49 μm^{-1}. This sampling frequency corresponds to a sampling period T_s of $1/f_s = 0.18$ μm/pixel.

Since the sampling period of a camera is the ratio of the pixel spacing (23 μm for our Photometrics camera) to the magnification, the lowest magnification required not to alias the signal is (23 μm/pixel)/(0.18 μm/pixel) = 128×.

We now compare this theoretical prediction with measurements. Figure 8-2 shows the intensity PSF measured at two magnifications. The 400× PSF indicates that the Fourier transform is non-zero (i.e., above the noise floor which has been reduced 10-fold by averaging 100 images) for $f < \pm 0.3\pi$. Therefore, a magnification of 400× oversamples the specimen by a factor of 3.33; the Nyquist magnification is 120×. Thus, the theoretical and experimental Nyquist magnifications agree to within 6%. The data from 100× magnification illustrates that while this magnification aliases a little, the power aliased is negligible. Since 100× provides a much larger field of view than our next possible magnification (160×), we use 100×.

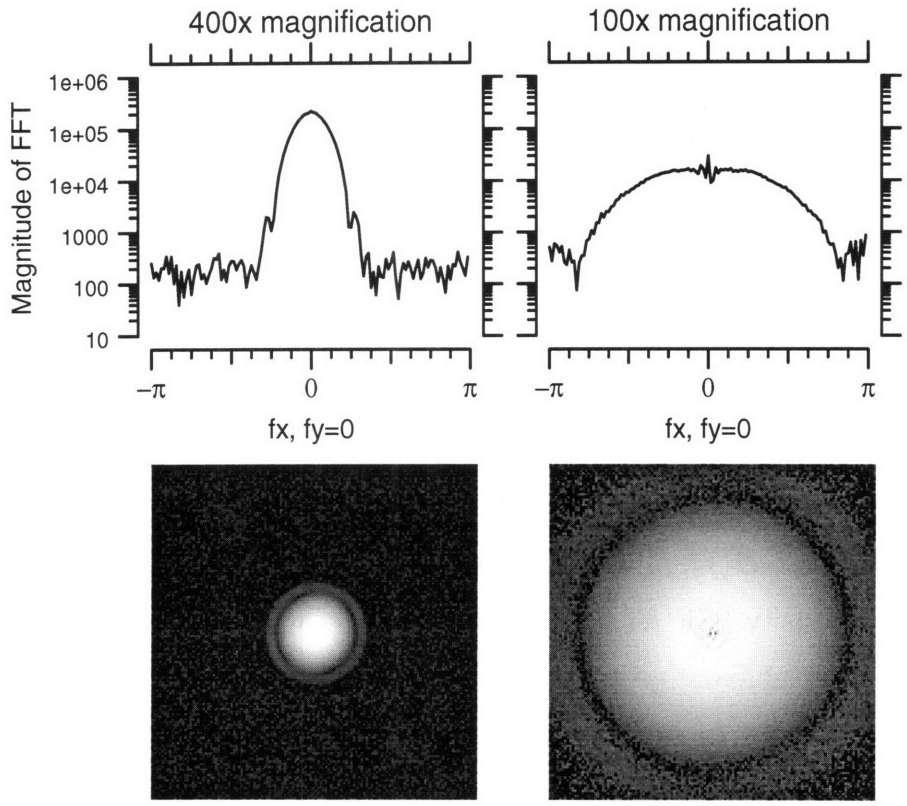


Figure 8-2: Magnitude spectra of the microscope's intensity point spread function. The left panels show the magnitude of the Fourier transform of the image (averaged 100 times to reduce shot noise) of a 0.3 μm TiO_2 bead at 400× magnification. The right panels are data from the same bead taken at 100× magnification. The bottom images show the two-dimensional FFT magnitude, with 0 frequency at the center of the image. The top plots show the FFT magnitude as a function of the horizontal spatial frequency f_x, with the vertical spatial frequency f_y set to 0.

Motion detection

We characterize the magnitude and phase of radial and longitudinal displacements of cochlear structures from sequences of video images taken at different phases of the stimulus (Figure 8-3). Rectangular regions of interest are manually drawn to enclose an object of interest (e.g., a hair bundle). Displacements between the object in successive images are estimated using the linear bias corrected gradient-based algorithm discussed in Chapter 4.

We characterize transverse motions as well as radial and longitudinal motions by collecting the above images at multiple planes of focus. Thus, we acquire a temporal sequence of 3D images. We estimate the three rigid-body translations between the

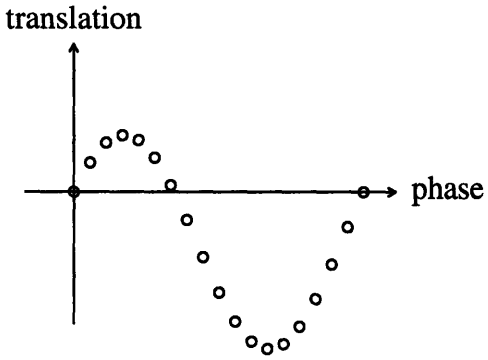


Figure 8-3: Illustration of the method for estimating motion from video images. Data are taken with the strobe synchronized to n different phases of the stimulus. The phases are evenly spaced in 360 degrees. If the estimated displacement between the first two phases is $d_{1,2}$, between the second and third is $d_{2,3}$, etc., then the data points plotted are (1,0), (2,$d_{1,2}$), (3,$d_{1,2}$+$d_{2,3}$), etc. Thus the symbols map out one period of the motion, with the caveat that the DC component is arbitrary. For clarity in the plots, the first point (1,0) is repeated as the last point (n+1,0).

images using the linear bias correction method discussed in Chapter 6.

8.2.2 Preparation

Adult alligator lizards (*Gerrhonotus multicarinatus*), 13-75 g body weight, were anesthetized by cooling to 4–8° C for 60 minutes and then sacrificed by decapitation followed by pithing the brain. A dorsal approach was used—first removing the skin, muscle, and triangular bone to expose the skull. A razor blade was used to break through the skull into a semi-circular canal. From this starting hole, the bone over the sacculus was carefully chipped away. In earlier experiments the sacculus was removed by washing it away with saline; in later experiments, by pulling the endolymphatic membrane apart with two forceps and then gently suctioning the sacculus (in pieces) out through 0.8 mm ID glass tube connected to a peristaltic pump. After removal of the sacculus, the cochlear duct was visible in its bony recess. The eighth nerve was cut by a razor blade fragment that had been sharpened to a point. In all but the last few experiments, cochlear duct was removed by using fine forceps to manually grasp the auditory nerve and to lift the duct out of the head. In the latest experiments, the duct was extracted from the head using locking forceps mounted on a micromanipulator. The forceps were opened remotely through a hydraulic mechanism to reduce operator-induced vibrations of the forceps. The cochlear duct was lifted from the head by the auditory nerve at a rate of approximately 100 μm/sec. In earlier experiments the cochlear duct was extracted from the head into air, and then pushed through an

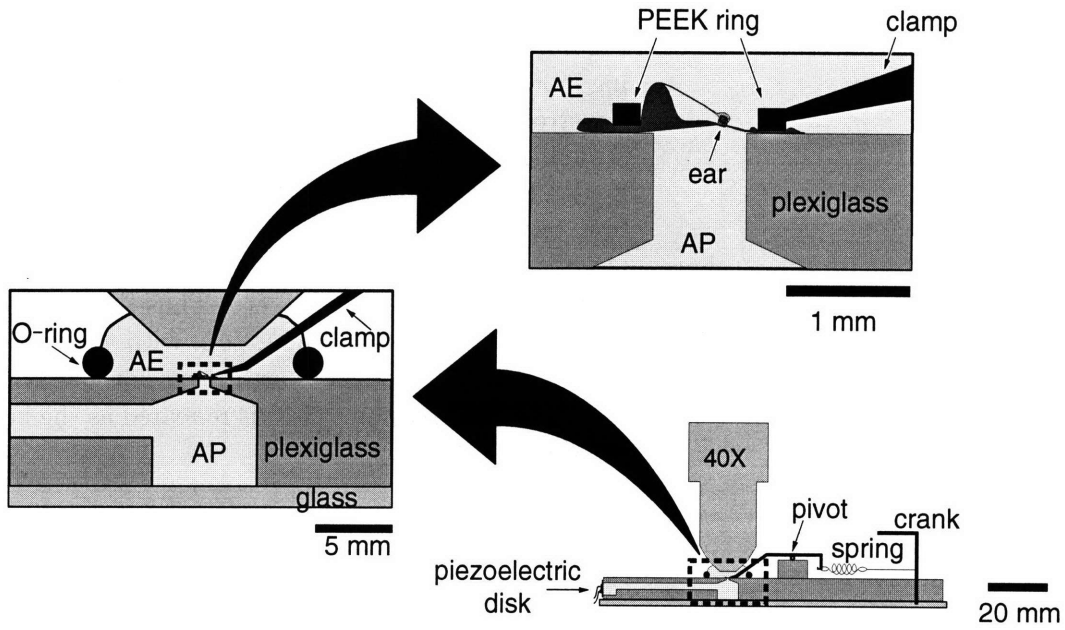


Figure 8-4: Experimental chamber for measuring motions of inner ear structures: three views with different scales. The sensory receptor organ is positioned over a hole in plexiglass (top panel) and held in place by a PEEK ring concentric with the plexiglass hole (shown in cross section). Force on the ring, provided through a spring-loaded beam (bottom panel), maintains a chemical and hydrodynamic seal between artificial endolymph (AE) above the organ and artificial perilymph (AP) below. Hair cells and the TM can be viewed through the 40X microscope objective while a piezoelectric disk generates sound pressure in AP.

air/water interface into the experimental chamber. In later experiments, the entire lizard head was submerged in saline so that the duct could be extracted and manipulated without going through an air/water interface. After extraction, the vestibular membrane was sometimes opened before and sometimes after positioning over a hole in the experimental chamber that separated two fluid-filled regions (Figure 8-4).

The care and use of animals have been approved by the Massachusetts Institute of Technology Committee on Animal Care.

8.2.3 Experimental chamber

Clamp

In the experimental chamber, the cochlear duct is held over a 0.74 mm diameter hole in a plexiglass sheet (Acrylite FF acrylic, Cryo Industries, Rockway, NJ) by a 0.76 mm diameter PEEK (polyetheretherketone) ring. The size of the hole and

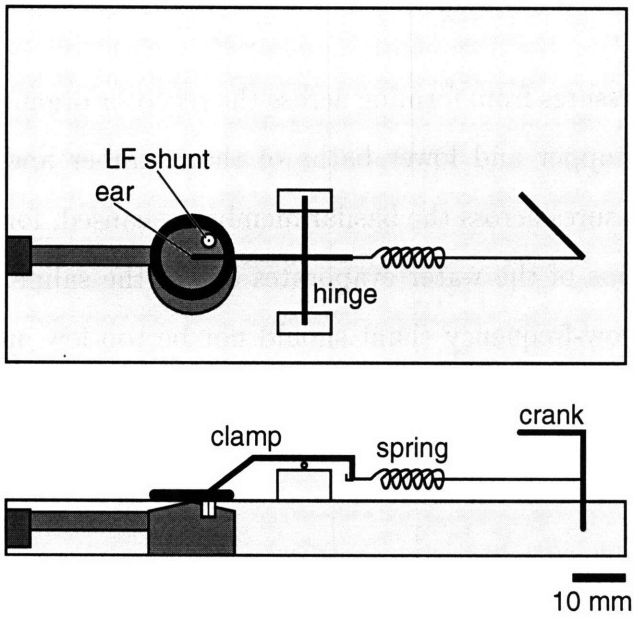


Figure 8-5: Experimental chamber for measuring motions of inner ear structures, highlighting the clamping apparatus and the low-frequency shunt (LF shunt). Rotating the crank pulls on the spring which causes the clamp to pivot and hold the ear over a small hole (See Figure 8-4). A hole in the plexiglass connects the upper and lower baths and provides a low-frequency hydrodynamic shunt.

ring were chosen to simulate the bony ring that supports the duct in vivo. Plastic was chosen so that upon accidental penetration of the epithelium, the cochlear duct would not be electrically shorted. Electrical shorting would shunt the gross electrical response of the cochlea. The ring is approximately 100 μm thick and was cut with a razor blade from 30 mil ID PEEK tubing (McMaster-Carr, NJ). The ring is glued with a plastic adhesive (Plastic Welder, Devcon Corp., Danvers, MA) to a stainless steel clamp.

The clamp is glued to a hinge 20 mm away and 6 mm above the chamber hole (Figure 8-5). The clamp's path through space is therefore circular. Ideally, the clamp would come straight down on the cochlear duct since any lateral motions will pull on the tissues. The ratio of the vertical motion to the horizontal motion toward the spring is approximately 3.4. Making the clamp longer or the hinge shorter would increase this ratio, but would also increase unintended motion in the orthogonal direction. The hinge, made from concentric stainless steel tubes, is designed to minimize rotations orthogonal to the intended rotations. These rotations cause undesirable translations of the clamp that increase with the clamp's length. Decreasing the height of the hinge would have the undesirable effect of reducing the force applied by the spring to clamp the ear.

Low-frequency shunt

In vivo the helicotrema prevents DC pressures from forming across the receptor organ. A hole in the plexiglass connects the upper and lower baths of the chamber and prevents the developments of static pressures across the basilar membrane caused, for example, by osmotic pressure when some of the water evaporates out of the saline. The hydrodynamic impedance of this low-frequency shunt should not be too low or else the cochlea will not be stimulated hydrodynamically. Furthermore, the electrical impedance of this shunt should not be too low or else the cochlea's electrical response will be short circuited.

The low-frequency shunt is constructed using a glass tube 3 mm long with an ID of 0.1 mm. The hydrodynamic resistance through a tube is

$$R_h = \frac{8\mu l}{\pi r^4}$$

where μ is the viscosity of the fluid (approximately that of water), l is the length of the tube, and r is the radius. Thus $R_h = 1.1 \times 10^{12}$ Ns/m^5. The hydrodynamic inertance through a tube is

$$L = \frac{\rho l}{\pi r^2}$$

where ρ is the density of the fluid (approximately that of water). Thus $L = 1.0 \times 10^9$ Ns/m^5. The hydrodynamic impedance $Z = R_h + Lj\omega$ is compared in Figure 8-6 to a model of the impedance of the basilar membrane in vivo (Rosowski et al., 1985). This glass tube provides a low-frequency shunt while attenuating the magnitude of the volume velocity of the fluid to the basilar membrane by no more than 8 dB.

The electrical impedance of the low-frequency shunt must be sufficiently large so that any electrical response from the cochlea is not attenuated significantly. The electrical resistance of the saline in the glass tube is

$$R_e = \frac{l}{\pi r^2 \Lambda_{\mathrm{NaCl}} [\mathrm{NaCl}]}$$

where Λ_{NaCl} is the equivalent ionic conductivity for NaCl (the predominate species

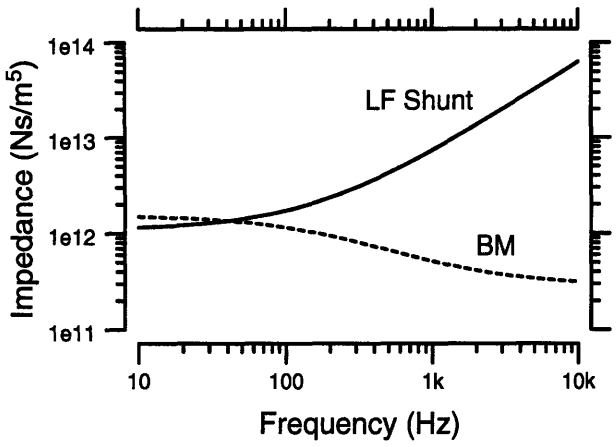


Figure 8-6: Impedance of a model of the basilar membrane (BM) in vivo (dashed curve) and the low-frequency shunt used in the experimental chamber (solid curve). The low-frequency shunt has a larger impedance above 40 Hz.

in perilymph), and [NaCl] is the concentration of NaCl. At a concentration of 175 mM (175 mol/m^3), $\Lambda_{\mathrm{NaCl}} = 0.0116$ m$^2/\Omega$ mol (Weast, 1987, page D-167). Thus $R_e = 190$ kΩ. The calculation for 175 mM KCl ($\Lambda_{\mathrm{NaCl}} = 0.0139$ m$^2/\Omega$ mol), the predominate species in endolymph, is $R_e = 160$ kΩ. In comparison, we estimate the resistance of the cochlear duct (R_{cd}) by multiplying the specific resistance of the sensory epithelium of the frog sacculus, 50 Ω cm^2 (Corey and Hudspeth, 1983), by the area of the chamber's hole for the cochlear duct, $\pi(.074\ \mathrm{cm})^2$ to obtain $R_{cd} = 3$ kΩ $\ll R_e$. Since the length of the glass tube is much greater than the thickness of the basilar membrane, the electrical capacitance of the cochlear duct is much greater than that of the low-frequency shunt. Consequently, the electrical impedance of the low-frequency shunt is much greater than the cochlear duct for all frequencies.

The low-frequency shunt provides a DC hydrodynamic path across the cochlea, while not significantly affecting the AC hydrodynamic path or electrical impedance across the cochlea.

Stimulation

A piezoelectric disk morph, obtained from a Panasonic EFR series 40 kHz ultrasonic transducer and waterproofed with Krulex urethane glue, seals the left-most opening (Figure 8-8) in the bottom fluid-filled region and provides the hydrodynamic stimulation to the cochlear duct. In later experiments the piezoelectric disk was electrically shielded with copper foil (See Electrical Measurements section for details.).

A pressure transducer (EPX, Entran Devices Inc., Farfield, NJ) monitors the fluid

pressure under the duct in order to calibrate the sound source.

The chamber's hydrodynamic stimulator can produce at least 100 dB SPL pressures over the frequency range 0.02–20 kHz. The stimulator's power spectrum is not constant across frequency; consequently, the drive voltage to the piezoelectric disk must change when collecting data for a constant-amplitude, swept-frequency experiment. On the other hand, the piezoelectric disk is fairly linear with respect to both amplitude and frequency distortion products. For example, upon driving the piezoelectric disk to generate two 114 dB SPL hydrodynamic tones, the largest distortion product (difference tones) was 42 dB down from the drive tones. The largest cubic distortion product was 45 dB down.

Perfusion

The requirement of perfusing both apical and basal sides of the cochlear duct complicates the design and construction of the chamber. Perfusion is needed to provide the cells with nutrients and a constant chemical environment.

Perfusion requirements are similar to those for the low-frequency shunt: (1) minimize the DC pressure across the duct when perfusing and (2) minimize shorting the duct either electrically or hydrodynamically. Our implementation is shown in Figure 8-7.

Saline influx to the basal surface is through a 1.5 m, 0.15 mm ID PEEK tube. This influx tube's hydrodynamic impedance is more than 40 dB larger than that of the model of the basilar membrane (Rosowski et al., 1985). Perfusion efflux is through a very large inside diameter (6.4 mm) tube in order to keep the fluid beneath the basilar membrane near atmospheric pressure. The efflux from the apical and basal surfaces is drawn by a peristaltic pump from the same collecting o-ring, ensuring that the hydrostatic head on both sides is the same. The collecting o-ring is distant from both surfaces of the cochlear duct to reduce fluid mixing near the duct. Apical influx is accomplished through a glass tube placed on the cochlear duct's o-ring. Apical efflux is through a covered slot connecting the two o-rings. The lower perfusion efflux path can be sealed with a sliding vent to enable hydrodynamic stimulation and electrical

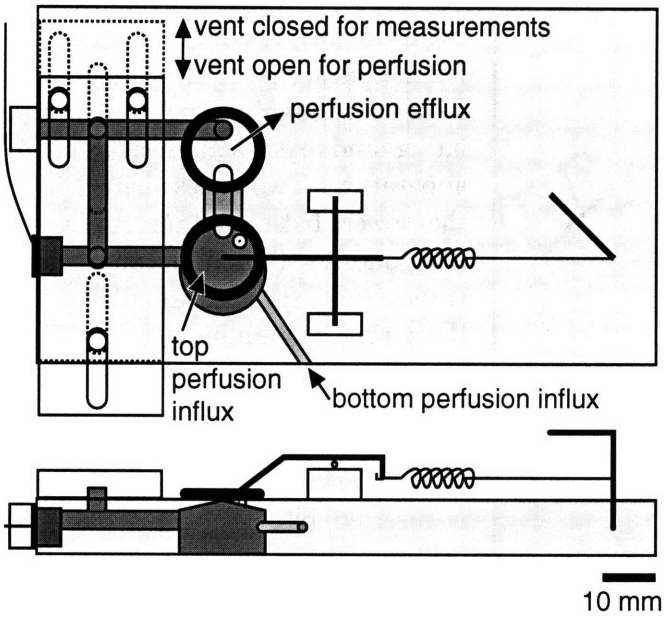


Figure 8-7: Experimental chamber for measuring motions of inner ear structures, highlighting the perfusion apparatus. A sliding plexiglass (Acrylite FF acrylic, Cryo Industries, Rockway, NJ) top piece is held by three screws with spring washers (not shown). The top piece can either vent (solid lines) the lower chamber to the efflux o-ring (perfusion out) or can hydrodynamically seal (dotted lines) the lower bath from the top bath. A partially covered slot connects the top bath to the efflux o-ring.

isolation. The vent is held in place by 3 screws with spring washers. A thin layer of Vaseline is put on the vent to improve the low frequency seal. A sliding vent is used as a switch because it displaces very little volume to open and close. Any fluid volume displaced by the switch would cause a large increase in static pressure on the basal side of the cochlear duct. This pressure increase would tension or even tear the basilar membrane.

The chemical isolation between the top and bottom baths in the chamber was tested by perfusing saline with red food coloring in the bottom bath and blue food coloring in the top bath. The dyes stained the cochlear duct, but did not pass through. In particular, the red dye perfused in the bottom bath stained the region around the papilla and the limbic bulge. The blue dye stained the part of the cochlear duct outside of the clamp and the limbic bulge (which made the limbic bulge look purple).

Electrical Measurements

The cochlea transduces a mechanical input signal into an electrical output signal; therefore, measuring the electrical response to sound is a common and natural way to assess the viability of cochlear preparations. Assessing the viability of each preparation is critical in biological experiments, because viability may affect the motions of inner ear structures. For example, the motion of the basilar membrane in mam-

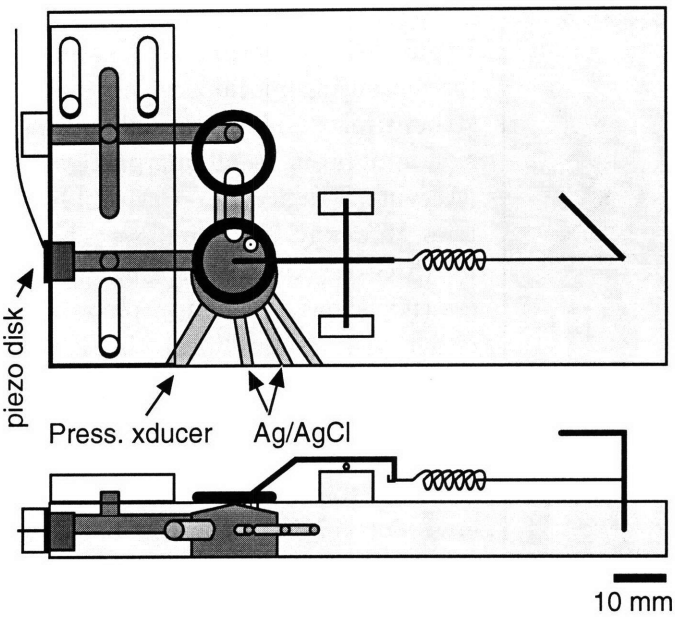


Figure 8-8: Experimental chamber for measuring motions of inner ear structures, highlighting the stimulating and electrical measurement apparatus. Two holes drilled into the lower bath contain Ag/AgCl electrodes. Two additional electrodes (not shown) are placed in the o-ring containing the ear. The four electrodes are used to measure the ear's electrical response to sound stimulation provided by the piezoelectric disk. A pressure transducer is inserted into the lower bath to calibrate the sound pressure of the acoustic stimulus.

mals is nonlinearly related to the sound pressure when the electrical responses of the cochlea are normal, but becomes linear as the electrical response deteriorates (Dallos et al., 1996). Consequently, we designed the chamber to be able to measure electrical responses. In addition to measuring the audio-frequency electrical response, our electronics duplicate the DC electrical potential (called the endolymphatic potential or EP) that exists across the apical and basal sides of the cochlea in vivo. Our implementation is shown in Figure 8-8.

Two Ag/AgCl electrodes (E205, Warner Instruments, CT) are placed in both the top and bottom baths. A four-electrode voltage clamp configuration is used (Corey and Hudspeth, 1983). The electric potential is sensed with one of the electrodes in each pair and the other is used to drive current. The bottom bath is kept at ground, while the top bath is kept at 16 mV, the EP for the alligator lizard (Weiss, Altmann and Mulroy, 1978). The current needed to keep the top bath at 16 mV is converted into voltage, amplified, and low-pass filtered with a 2nd order Butterworth at 20 kHz to reduce high-frequency noise. The circuit has two outputs: a 10^6 V/A DC-coupled output and a 10^8 V/A AC-coupled output. The source resistance of the voltage clamp circuit (excluding the Ag/AgCl electrodes) is 0.1 Ω at DC. The DC-coupled output enables measurement of the cochlear duct's impedance. The impedance is typically 4-5 kΩ, which is comparable to the estimate of 3 kΩ made above.

The AC-coupled output enables measurement of the cochlear microphonic current. Since the cochlea responds at the same frequency as the stimulus, the limiting factor in making AC measurements is cross-talk with the piezoelectric stimulator[4]. To reduce cross-talk, (1) the ground side of the piezoelectric disk faces the chamber; (2) the drive to the piezoelectric disk is coaxial cable; (3) the first 6 mm of the liquid path in the chamber is an insulating oil rather than saline (flexible 20 μm thick plastic wrap separates the oil and saline regions); and (4) the entire piezoelectric disk is covered with 25 μm thick copper foil which is soldered to the ground on the coaxial cable. To measure the cross-talk, a piece of flexible 20 μm thick plastic (Saran Wrap, Dow Chemical Co., Midland, MI or Glad sandwich bag, First Brands Corp., Danbury, CT) was placed where the cochlear duct should go. With only methods (1) and (2), 7 volt rms 500 Hz tone on the piezo (equivalent 104 dB SPL at the eardrum) produced 400 pA response. Frog sacculus hair cells have a maximal response near 170 pA (Assad et al., 1991). Consequently, the cross-talk level with methods (1) and (2) were a significant fraction of the best response we could expect with the lizard (10 hair cells times 170 pA is only a factor of 4 larger than the cross-talk). With all four the cross-talk reducing methods, the cross-talk was reduced to $\sim$10 pA — equal to the current through only 10 transduction channels (Holton and Hudspeth, 1986). The microphonic measuring equipment is sensitive enough to measure 6% of the maximum current through a single hair bundle.

8.2.4 In situ microphonic measurements

We measured the microphonic in situ, in order to assess the viability of the cochlea during surgery, The measurement methods are comparable to Kaplan et al. (1983), however the electrode placements were different because we used a dorsal approach while Kaplan et al. (1983) used a ventral approach. The reference electrode was placed in muscle either anterolateral to the cochlea or along the midline in the neck. The positive electrode was placed as close to the cochlea as possible. After removing

[4]The cochlea also responds at harmonics of the stimulus (usually with a smaller amplitude) making amplifier noise another limitation.

the dorsal portion of the skull, the positive electrode was placed on the sacculus. After removing the sacculus, the positive electrode was placed within 0.5 mm of the cochlear duct. The electrical signal was amplified by an 80 dB instrumentation amplifier, low-pass filtered with a 3 pole Butterworth at 20 kHz, and sampled by an A/D converter (DSP-16, Ariel Corporation, NJ) at 50 kHz using the SYSid software package (version 4.18, AT&T Bell Labs, NJ and Ariel Corporation, NJ).

Acoustic clicks were presented with a microphone system (ER-10B, Etymotic Research, IL) that was sealed to the external ear of the lizard with Vaseline. A computer-generated click was sent through a unity-gain power amplifier (D-75, Crown International, Inc., Elkhart, IN) to the microphone system's speaker. Clicks were presented at a rate of 98 per second, which was chosen to minimize the time to collect averaged responses (typically 2,000 averages), while not degrading the microphonic potential's amplitude. Kaplan et al. (1983) determined that the amplitude of the cochlear microphonic potential is constant with respect to click rates out to 100 Hz. The neural response amplitude is down by a factor of 2 at 100 Hz from its value at low click repetition rates; however, this reduction is of little concern since we are more interested in the microphonic component.

The microphonic potential was measured by playing both condensing and rarefacting acoustic clicks to the ear and recording the difference between the two averaged response waveforms. The contralateral columella was cut to prevent the undisturbed contralateral cochlea from responding to the ipsilateral acoustic stimulation. In some experiments the measurements were done in vivo. The lizard was anesthetized with urethane (1.4% of its body weight). Although severed blood vessels were closed with hemostats, blood loss was a problem. In other experiments, the lizards were decapitated and pithed, following the procedure outlined in Section 8.2.2. These experiments were done with A. J. Aranyosi.

8.2.5 Glass probe stimulation

A mechanical method of stimulating the cochlea was developed in order to assess the viability of the preparation at surgical stages at which acoustic stimulation was

impossible. A glass pipette was pulled (Vertical pipette puller model 700C, David Kopf Instruments, Inc., Tujunga, CA) from a glass capillary tube (Model GC200-15, 2mm OD, Warner Instrument Corp., CT). The pipette was fire polished to a sealed tip diameter of 100 μm. The pipette was attached to a piezoelectric bimorph that was mounted on a 3-axis micromanipulator (Rucker and Kolls, Inc., Milpitas, CA). Motions of the pipette on the order of 25 μm p-p were observed under stroboscopic illumination at 300 Hz and 10 V p-p, when the pipette was in air. At 1 kHz, the amplitude dropped to 8 μm p-p.

8.2.6 Intracellular recordings

Intracellular potential recordings were performed with the assistance of Pankaj Oberoi and A. J. Aranyosi at the Eaton Peabody Laboratory of Auditory Physiology in a sound-proof, electrically shielded chamber (Ver et al., 1975). The cochlear duct was extracted and clamped into the sound stimulating chamber. Microelectrodes were pulled (Sutter Instrument Model P80/PC Flaming Brown) from glass pipettes (1.2 mm OD; 0.6 ID; omega dot) with approximately 0.2 μm tip diameter (30 MΩ typical electrode resistance) and filled with 2 M KCl and methyl blue dye to improve visualization. A Ag/AgCl wire was inserted for electrical contact. The electrode was placed on a micromanipulator that advances the electrode hydraulically from outside the chamber. Recording electronics (Axoprobe 1a, Axon Instruments) measured the potential difference between the microelectrode and a Ag/AgCl reference electrode in the top bath. The cochlea was stimulated with sound via the piezoelectric disk built into the sound stimulating chamber.

8.2.7 Solutions

The following were the three most commonly used solutions.

- Artificial perilymph (171 mM NaCl, 2 mM KCl, 2 mM $CaCl_2$, 3 mM D-glucose, 5 mM HEPES buffer) (Freeman et al., 1993). Osmolarity: 330 mosmoles.

- L-15 Medium (Leibovitz) with 20 mM HEPES buffer and 5 mM D-glucose. Osmolarity: 350 mosmoles.

- 90% L-15 Medium (Leibovitz) and 10% chicken serum, with 20 mM HEPES buffer and 5 mM D-glucose. (Freeman et al., 1993) Osmolarity: 360 mosmoles.

All solutions were adjusted to a pH of 7.3 ±0.05 by adding NaOH.

The L-15 solutions have amino acids and vitamins, which increase the osmolarity and also provide many of the nutrients that the cochlear duct may need. L-15/chicken serum also provides proteins that the duct may need; however, those proteins greatly increase the number of bubbles in solution (similar to whipped egg whites)—attenuating the hydrodynamic stimulus in the experimental chamber. Most of the sound-induced motion studies and in vitro microphonic studies used artificial perilymph. L-15/chicken serum was used in the intracellular recording experiments. L-15 was used in most of the in situ microphonic studies. While a systematic comparison of the effect of the different solutions on the cochlea was not performed, no effects were noticed.

8.3 Results

8.3.1 Microphonic

In vitro microphonic measurements

After clamping the cochlear duct into the experimental chamber, we were able to observe sound induced motions and to measure the cochlea's electrical response.

In seven preparations, the voltage across the cochlea was measured while keeping the current through the cochlea at 0 amps. In all seven, the potentials measured—< 0.2 μV @ 350 Hz, and < 2 μV @ 3 kHz—were comparable to the potentials measured when the cochlea was replaced with saran wrap. These potentials are hard to compare with existing in vivo measurements because the electrical connections of the in vitro system to the recording electrodes differs from that in vivo. Nevertheless,

the in vivo cochlear microphonic measured on the round window at 320 Hz is 20 dB larger than what we measured (Kaplan et al., 1983). In vivo microphonics measured in scala media at 500 Hz are 40 dB larger than what we measured at 350 Hz (Weiss, Altmann and Mulroy, 1978). Therefore, if the cochleae were transducing sound, we expect that the responses should have been well above our cross-talk levels measured in the absence of a cochlea.

In 14 preparations, we measured the current through the cochlea while keeping the voltage across the basilar papilla constant. The DC voltage was fixed at 16.5 mV, which is the same as the voltage measured in vivo (Weiss, Altmann and Mulroy, 1978). In this voltage clamp mode, the source resistance is less than 2 Ω below 1 kHz; as long as the impedance of the duct is much greater 2 Ω, impedance changes due to surgical methods should have minimal effects on the measured currents. The median impedance was 4.5 kΩ, with an interquartile range of 3.5–5.7 kΩ and a range of 1.3–72 kΩ. Thus, comparisons with other studies are easier to interpret. Nevertheless, transduction currents have not been measured in the lizard; so we have to compare to other reptiles. In the chick, maximal cochlear hair cell transduction currents have been reported to be 114 pA (Kimitsuki and Ohmori, 1992) and later 50 pA (Kimitsuki et al., 1996). In the turtle, maximal cochlear hair cell transduction currents have been reported to be 200–400 pA (Crawford et al., 1989) and later 140 pA (Crawford et al., 1991). Before shielding the piezoelectric disk stimulator and using oil next to the piezoelectric disk, the median microphonic current at 330 Hz was 200 pA (n=6). After shielding, the median dropped to 38 pA (n=8). In both cases, the current was comparable to the cross-talk measured under similar conditions. Thus, not even one hair bundle was transducing in the 8 preparations after the improved shielding, and no more than one was transducing before the improved shielding.

We now explore the four hypotheses of why we were unable to measure a microphonic in vitro.

Hypothesis I: Bad electronics

We tested our voltage clamp electronics for measuring the cochlear microphonic to verify that if there were a microphonic we could measure it. In addition to standard tests such as replacing the cochlea by a resistor, we tested the circuit under conditions more similar to the cochlea by measuring the electrical response of the frog sacculus.[5] The sacculus was stimulated by a vibrating glass probe (Corey and Hudspeth, 1983) at 95 Hz. With Corey and Hudspeth's (1983) original circuit, we measured a 2.5 nA fundamental and a 9 nA second harmonic component of the electrical response. The stimulating probe had to then be raised to switch measuring circuits. Upon lowering the probe, our circuit measured a 1 nA fundamental and a 6 nA second harmonic. Both sets of measurements are comparable to those in Corey and Hudspeth (1983); the difference in the two circuits' responses is most likely due to a slightly different stimulating condition.

Hypothesis II: Microphonic is electrically shorted

We performed intracellular recordings[6] on 4 in vitro preparations in order to determine if the hair cells were responding electrically to sound but the extracellular response was being short-circuited.

In all 4 preparations, we were unable to measure an electrical response from cells in the free-standing region of the basliar papilla.[7] The sensitivity of the measurement apparatus was 0.2 mV peak-to-peak. In comparison, the intracellular response to tone bursts measured in vivo was 20 mV peak-to-peak (Holton and Weiss, 1983) at comparable sound pressure levels.

[5]This experiment was done with the assistance of Rick MacDonald and A. J. Aranyosi.
[6]These experiments were done with the assistance of Pankaj Oberoi and A. J. Aranyosi.
[7]Resting potentials are described later and are shown in Figure 8-10.

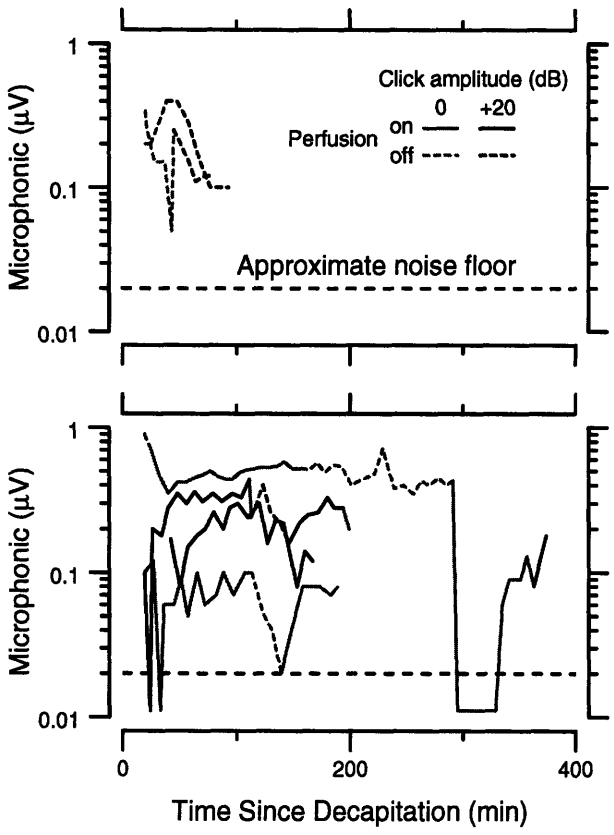


Figure 8-9: In situ cochlear microphonic potential vs. time. Each line represents the peak-to-peak microphonic potential from one preparation as a function of time. The top plot shows data from lizard cochleae which were never perfused with saline. The bottom plot shows data from lizard cochleae which were perfused in some part of the experiment. The horizontal dashed line represents the approximate noise floor for our measurements that averaged 8,000 click responses; most of the data was taken with 2,000 averages.

Hypothesis III: Cochlea is damaged from an inappropriate chemical environment

We tested the hypothesis that our saline solutions could be compromising the cochlear microphonic response by (1) performing the surgery up to the point where we could control the chemical environment around the cochlea, and (2) testing the effect of our saline by measuring changes in the in situ cochlear microphonic. The microphonic was elicited by sound presented at the eardrum, and measured by a gross electrode placed within 0.5 mm of the cochlea.

Figure 8-9 shows results from two experiments where saline was not perfused near the cochlea except to keep the recording electrodes wet. The microphonic potential dropped with time and increased when the acoustic click was louder. The basic trend in the data suggests that the microphonic potential drops roughly at a rate of 20 dB per 30 min. without perfusion.

Figure 8-9 also shows results from four experiments where we did perfuse saline (oxygenated L-15). Perfusion was accomplished by mounting a glass capillary on a

micromanipulator and lowering it within 0.5 mm of the cochlear duct. Saline was suctioned from this capillary using a peristaltic pump while fresh saline was perfused onto the head posterior to the saccular cavity with a syringe pump. The microphonic typically dropped 6-10 dB in the first 15 minutes, and then stabilized for up to several hours. In two of the three experiments where perfusion was stopped, the microphonic potential dropped at roughly the same rate as when the cochlea was never perfused—20 dB per 30 min. Upon restarting perfusion, the microphonic potential increased at about the same rate.

Hypothesis IV: Cochlea is damaged from mechanical trauma

Several strategies were taken to determine (1) if the cochlea was being mechanically damaged, (2) when the damage was happening, and (3) how to reduce the damage.

Intracellular recordings. We performed intracellular recordings[8] on 4 in vitro preparations in order to measure the resting potential of the cells in the cochlear duct. Figure 8-10 shows the resting potentials in the four preparations. In 3 out of the 4 preparations, the median DC membrane potentials for cells on the basilar papilla were -64, -77, and -70 mV. These potentials are comparable to those measured in vivo for hair cells (median of -73 mV) and 25 mV too positive for supporting cells (Weiss, Peake, Ling and Holton, 1978). Thus, our recordings were either from hair cells that were not transducing (none of the cells had a measurable response to sound) but could maintain a resting potential or in supporting cells with depolarized resting potentials. In the fourth preparation, membrane potentials near the basilar papilla had a median of -15 mV, but cells on the limbic bulge had a median resting potential of -71 mV.

In vivo microphonic measurements during surgery. Microphonic measurements were taken at different stages of the surgery in 6 in vivo preparations of anesthetized lizards. These experiments were very difficult to do without causing severe

[8]These experiments were done with the assistance of Pankaj Oberoi and A. J. Aranyosi.

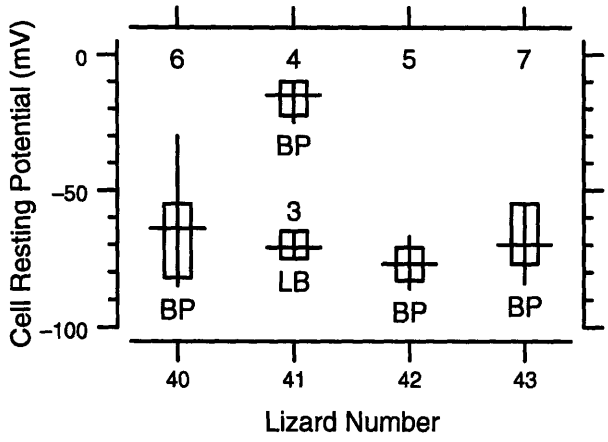


Figure 8-10: Intracellular resting potentials from cells in 4 in vitro preparations. The box plots represent the median, interquartile range, and range of the cells measured in each preparation. The number of cells in each box plot is indicated above the plot. BP denotes cells measured in the basliar papilla; LB, in the limbic bulge.

blood loss from a large vessel beneath the triangular bone[9] which has to be cut to widely access the sacculus dorsally. In 4 of the 6 preparations, we were able to measure both a neural response and a microphonic. These responses were maintained until exposing the sacculus, during which cutting the triangular bone was necessary. The other 2 failed from instrumentation problems or overdosing the lizard with anesthetic.

In a seventh anesthetized lizard experiment, a baseline of microphonic/neural recordings were made, and then the lizard was decapitated. Results show that the neural response lasts for 10–15 minutes after decapitation, and the microphonic lasts over 2 hours. Since the cochlear duct can be removed in under 30 minutes after decapitation, we decided to repeat this type of microphonic surgery experiments after decapitation. These experiments are vastly simpler and faster to perform, and more closely mimic the surgery used in motion experiments.

In situ microphonic measurements during surgery. Microphonic measurements were taken at different stages of the surgery in 11 preparations[10] (See Figure 8-11). In these preparations, the lizards were decapitated, pithed, and their cochlear ducts were exposed following the procedure outlined in Section 8.2.2. Quantitative interpretation of the microphonic across different stages of the surgery is difficult because the hydrodynamic environment in which the cochlea rests changes at the various

[9]Figure 1 in (Mulroy, 1974) shows the triangular bone unlabeled.

[10]These experiments were done with the assistance of A. J. Aranyosi.

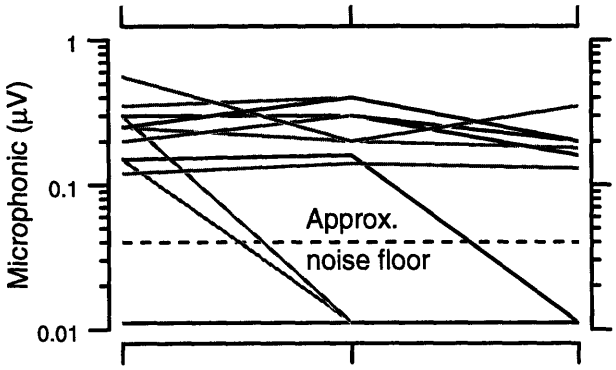


Figure 8-11: Magnitude of the cochlear microphonic as a function of the stage in the surgery for 11 preparations. Each line represents microphonic measurements for a single preparation at three surgical stages. Two levels of acoustic clicks were used: black line segments represent responses from clicks that were 20 dB louder than the gray line segments.

surgical stages. Consequently, we examine the binary decision: does a measurable cochlear microphonic response exist.

In 7 out of the 11 preparations, we could still measure a cochlear microphonic after cutting the VIIIth nerve—the last step before removal of the cochlear duct. In 3 out of the remaining 4, we were able to measure a microphonic at some stage and then the microphonic dropped below our noise floor. In 2 out of the 11, we had to increase the amplitude of the acoustic click by 20 dB in order to measure a microphonic. In 1 case, we did not increase the click amplitude when the microphonic potential dropped below our noise floor.

Unlike the microphonic, the neural response to acoustic clicks was typically absent. Occasionally, a neural response was present in the first measurement (15–20 minutes post decapitation), but the amplitude of the neural dropped quickly with time.

In vitro microphonic measurement. In a very recent experiment, the cochlear duct was extracted from the head using locking forceps mounted on a micromanipulator[11]; previous extraction were done manually. The cochlear duct was held 1 cm above the head, which was submerged in saline (L-15). The cochlear duct was shaken with the glass probe stimulator to elicit an electrical response.

Results from this experiment are shown in Figure 8-12. The cochlear microphonic responded non-linearly with stimulus amplitude after removal from the head. Fifty

[11]This experiment was done with the assistance of A. J. Aranyosi.

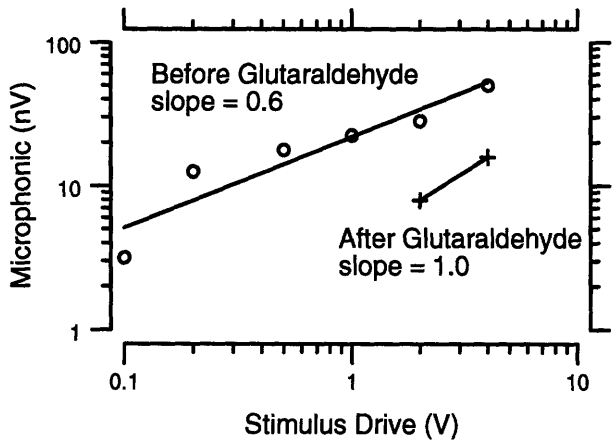


Figure 8-12: Magnitude of the cochlear microphonic as a function of the 2 kHz electrical drive to the glass-probe stimulator. The circles represent the magnitude of the cochlear microphonic after the cochlear duct was removed from the head and before application of glutaraldehyde. The line represents a least-squares fit to the data and has a slope of 0.6 on a log-log scale. The pluses represent data after the application of glutaraldehyde. The least-squares fit has a slope of 1.0 indicating a linear dependence on the stimulus drive.

microliters of a 50% glutaraldehyde solution was perfused near the duct. After eight minutes fresh saline was perfused near the duct so that the electrical environment near the recording electrode would be the same as prior to perfusion. The magnitude of the electrical response dropped by 12 dB and became linear—indicating that we reached the crosstalk level of our experimental apparatus.

8.3.2 Preliminary results of lizard ear motions

In this section we describe measurements of micromechanics in the tectorial region of the alligator lizard cochlea. A microphonic was not measured in any of the preparations, and for that reason these results must be considered preliminary. Nevertheless, these results demonstrate how this method can be used to address important issues in cochlear mechanics. Specifically, we examine (1) three-dimensional motions of the whole basilar papilla, (2) the relation between the motion of the tectorial membrane and the reticular lamina, and (3) the relation between the motion of the tectorial membrane and the tips of the hair bundles.

Gross motions of the sensory receptor organ

In order to observe audio frequency motions of the *in vitro* preparation, a strobe light was used to slow the apparent motion. When the preparation is stimulated, the entire

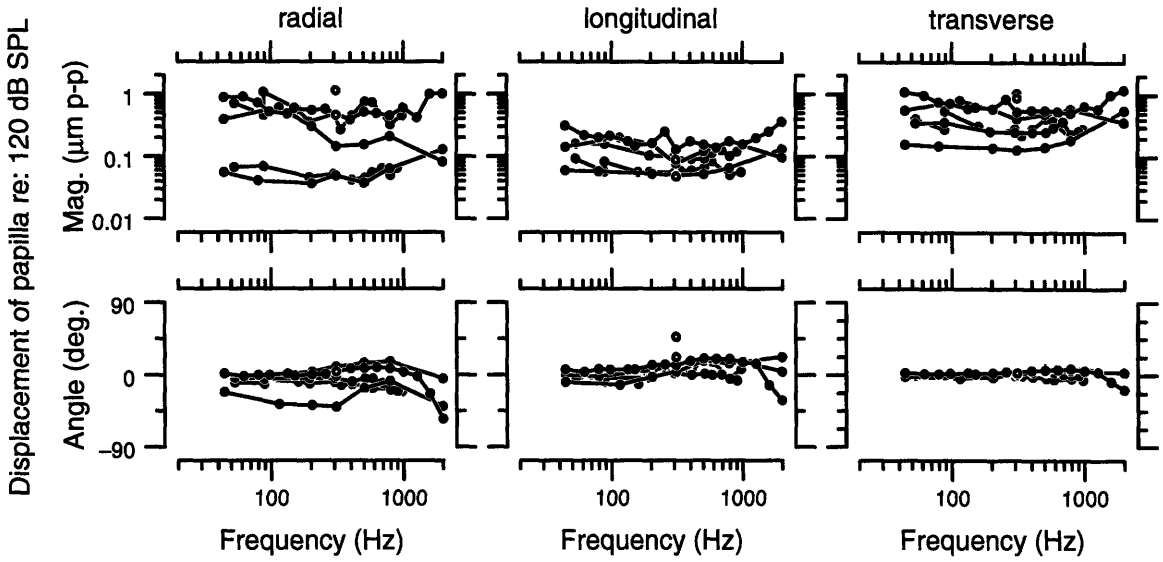


Figure 8-13: Frequency response of the displacement of the tectorial region of the basilar papilla for seven preparations. Data points for an individual experiment are represented as circles connected by a line. The three orthogonal directions are defined in Figure 8-16. The displacement magnitudes have been normalized to 120 dB SPL on the scala tympani side of the basilar membrane. The average sound pressure used in the experiments was 119 dB SPL, range 105–140 dB SPL. The angle is relative to the pressure on the scala tympani side of the basilar membrane.

receptor organ moves roughly as a rigid body—pivoting about an axis parallel to its longitudinal axis. Motions were clearly visible for sound pressures on the order of 100 dB SPL and for frequencies from 20 Hz to 5 kHz.

3D frequency response of the basilar papilla

In seven preparations, three-dimensional, sound-induced motions were examined at frequencies between 40 Hz and 2 kHz in the tectorial region of the basilar papilla. Displacements were analyzed in a $40 \times 64 \times 24$ μm region (175 pixels $\times$ 280 pixels $\times$ 13 planes) near the center of the TM region of the basilar papilla.[12] Figure 8-13 shows the motion of this region of the basilar papilla as a function of frequency. To first order, the basilar papilla moves in phase with the sound pressure and the magnitude of motion is independent of frequency.

The magnitude of the motion in the radial direction varied by 30 dB across prepa-

[12]In one preparation, only 9 planes where used, spanning 16 μm in depth. In another, 13 planes were used, but they spanned 36 μm.

rations. Variability in the displacement magnitude of the other directions were 20 dB. Radial motions were larger than longitudinal motions in six of the seven preparations. The average ratio of radial to longitudinal motions across experiments was 14 dB with a range of −8 to 20 dB. The ratio of the magnitude of transverse motions to radial motions had a mean of 3 dB, and ranged from −6 dB to 14 dB.

Figures 8-14 and 8-15 are images taken near the beginning of the experiment of the preparations used in the frequency response plots. Many hair bundles can be seen as well as the right (abneural) edge of the tectorial membrane. Circular cell herniations (blebs) can be seen in 6 of the 7 preparations.

3D cochlear motions

To characterize three-dimensional, sound-induced motions in one alligator lizard cochlea, a time sequence of 8 3D-images were analyzed. The cochlea was stimulated with a 513 Hz, 92 dB SPL (referred to the eardrum or 118 dB SPL under the basilar membrane) tone. Three-dimensional motion was estimated at a regular array of regions of interest (See Figure 8-16). Estimates were binned in the longitudinal direction and plotted as a function of the radial distance from the neural edge of the basilar papilla. Motions in the radial direction were largest, between 400 and 500 nm peak-to-peak. Longitudinal motions, which have been shown to excite hair cells minimally (Shotwell et al., 1981), were an order of magnitude smaller. Transverse motions increased abneurally from 100 nm to 330 nm.

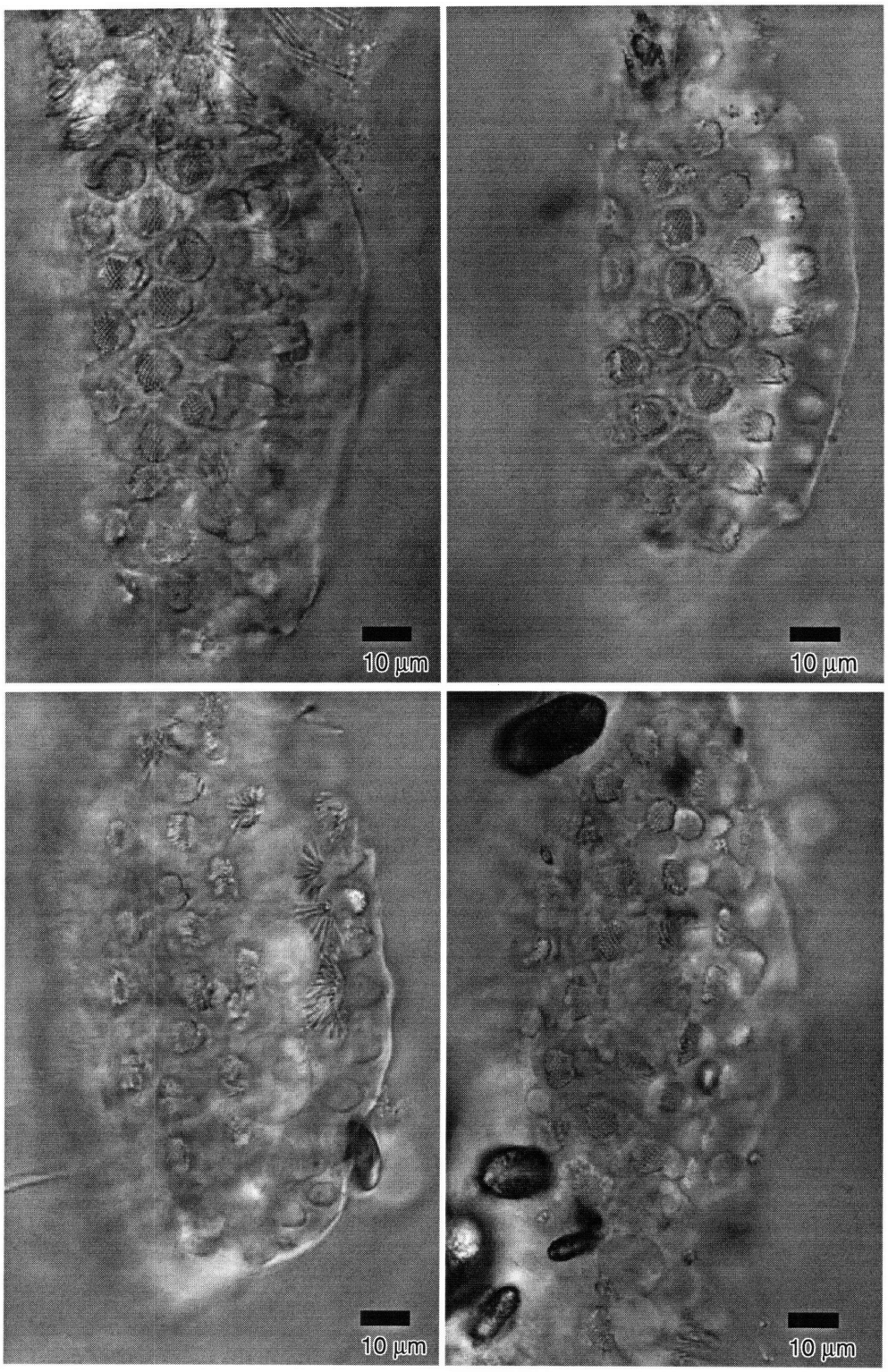


Figure 8-14: Images of the tectorial region of the basilar papilla at a focal plane near the base of the hair bundles for 4 preparations. Five large otoconia can be seen in the lower right preparation; one is visible in the lower left preparation.

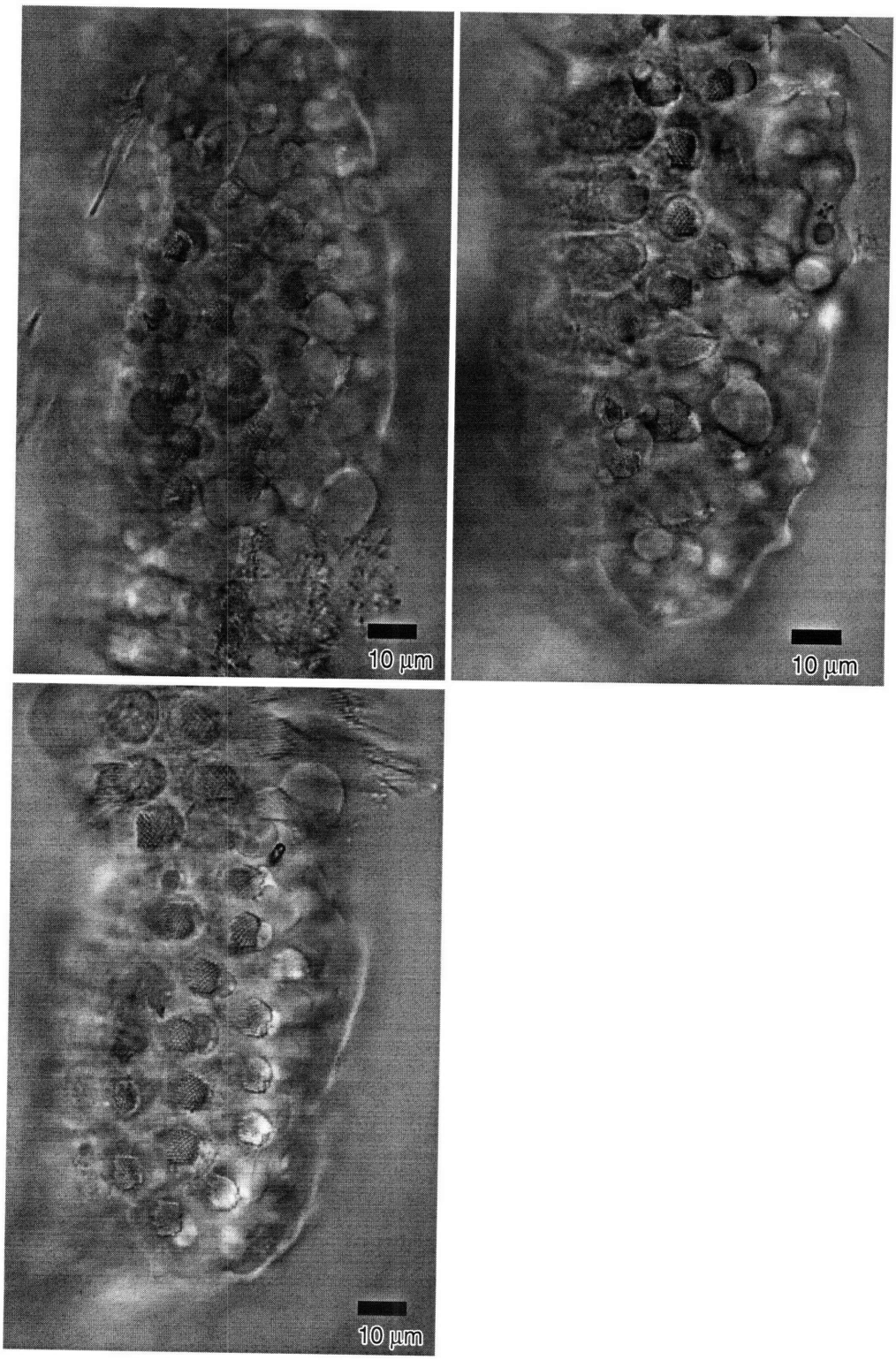


Figure 8-15: Images of the tectorial region of the basilar papilla at a focal plane near the base of the hair bundles for 3 preparations.

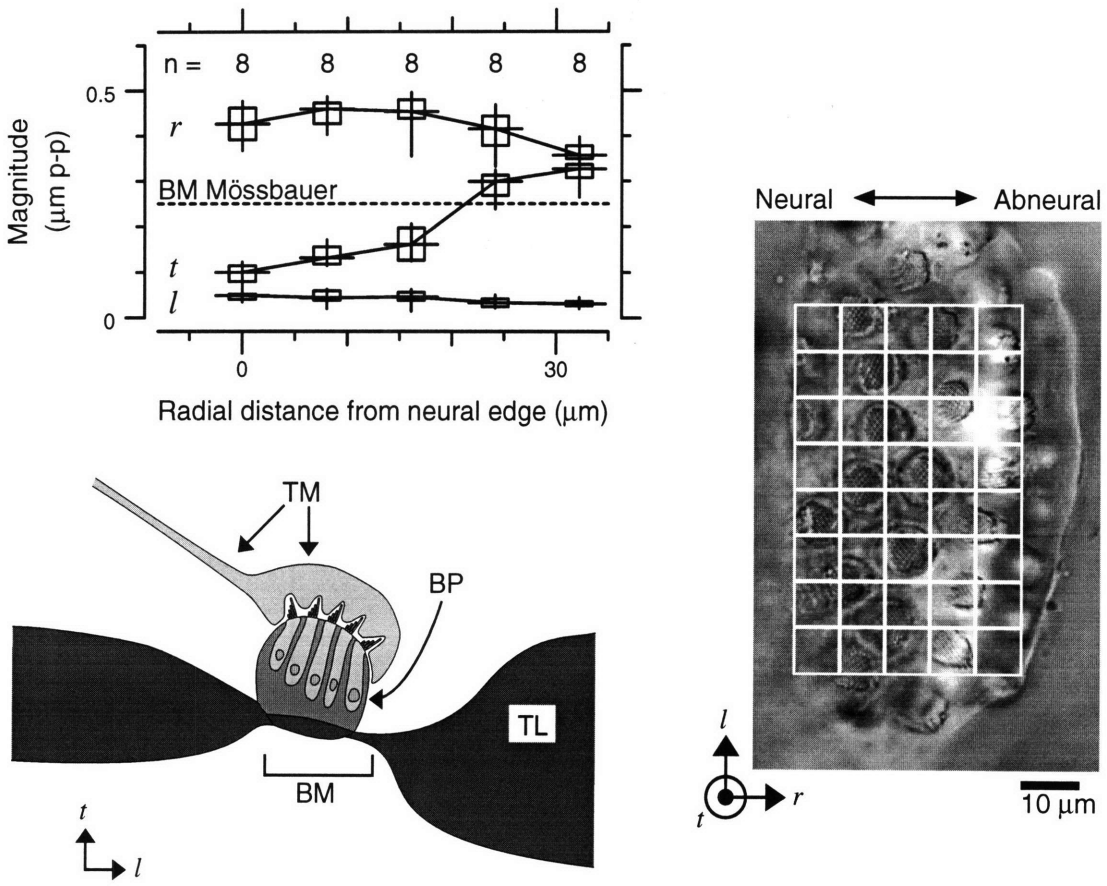


Figure 8-16: Three-dimensional motions of the alligator lizard cochlea. The right image shows the cochlea and the regular array of 40 regions of interest in which motion was estimated. The magnitude of the motion in the radial, longitudinal, and transverse directions are plotted on the left. The horizontal dashed line shows the measured motion of the basilar membrane in the transverse direction obtained with the Mössbauer technique (Peake and Ling, 1980). The schematic cross-section in the lower left shows the basliar papilla (BP) with a cluster of 5 hair cells resting on the basilar membrane. The hair cells' bundles protrude into the overlying tectorial membrane (TM). The basilar membrane (BM) is the thinnest (and therefore most compliant) just to the right of the basilar papilla (Mulroy, 1974). The triangular limbus (TL) is a supporting structure abneural to the basliar papilla.

Comparison of motions of the TM and reticular lamina

The tectorial membrane (TM) of the alligator lizard is almost completely transparent when viewed with a light microscope. The most apparent structure is on its top surface (top from the view in Figure 8-16). Circular structures slightly larger than hair cells, fine filamentous structure, granular structures, and the radial and longitudinal boundaries of the TM were the only structures in the TM with sufficient contrast to estimate motion. To compare motions of the TM with those of the underlying reticular lamina (RL), we identified a region of the TM in seven[13] preparations that could be readily tracked. Regions used 1 plane of focus and typically were 9×9 μm in area (38×38 pixels). We analyzed sound-induced radial displacements of these regions of the TM as well as the region of the RL directly underlying these TM regions.

Figure 8-17 shows the ratio of the radial motion of the tectorial membrane and reticular lamina as a function of frequency at constant sound pressure (mean 119 dB SPL underneath the basilar membrane, range 105–140 dB SPL). The sound pressure varied by less than 2 dB in each experiment, and had a range across experiments of 35 dB. The magnitude of this ratio is always less than 2 and is less than 1 in 120 out of the 130 measurements. Thus, the tectorial membrane moves less than the reticular lamina. Within each experiments, there was a slight trend for the tectorial membrane to move less at high frequencies than at low frequencies. The relative phase between the tectorial membrane and the reticular lamina is clustered around zero (indicating that they are moving in phase), and $-180°$, (indicating that they are moving exactly out of phase). In one experiment repeated trials gave results which differed by 180°.

Figure 8-18 plots the micromechanical transfer function: the ratio of the difference

[13]In one other preparation, motions parallel to the excitatory (radial) direction were smaller than motions along the length (longitudinal) direction. This motion is inconsistent with the other seven preparations. Additionally, anatomical evidence (Mulroy, 1974) suggests that the most compliant part of the cochlea would cause the basilar papilla to move radially the most, and the hair cells are sensitive to those radial motions (Mulroy, 1974; Shotwell et al., 1981). We believe that the motion in this preparation was distorted by a modification of the basilar membrane's compliance when clamping the cochlear duct in the experimental chamber. Therefore, we excluded that experiment from this micromechanical analysis.

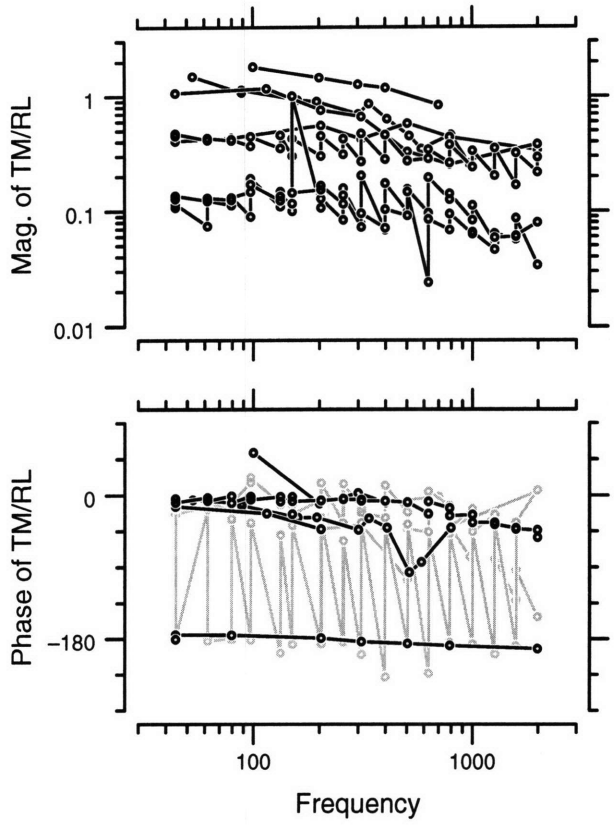


Figure 8-17: Ratio of tectorial membrane displacement to the reticular lamina displacement for seven preparations. Displacements were measured at the fundamental frequency of the acoustic stimulus. Data points for an individual experiment are represented as circles and connected by lines. Motions of the tectorial membrane were determined in regions of interest that were hand-selected for good contrast. Reticular lamina motions were determined directly below the region selected for the tectorial membrane. Phase estimates are indicated in black when the numerator and denominator of the ratio were both greater than 100 nm. These estimates are more reliable than the phase estimates indicated in gray, for which either the numerator or denominator was less than 100 nm.

between the radial motion of the TM and RL to the radial motion of the RL. In all 7 preparations, the magnitude of the micromechanical transfer function was within 14 dB of unity for all frequencies. The micromechanical transfer functions for individual experiments were roughly constant with frequency. In 4 of the 7 preparations, the phase was near $-180°$ for all frequencies. In the other 3 preparations, the phase was near $0°$ for low frequencies and approached $-180°$ near 800 Hz.

Comparison of TM and hair bundle tip motion

Figure 8-19 shows the ratio of the radial motion of the TM to the tips of a nearby hair bundle. This ratio varied by 30 dB across preparations, but was fairly constant across frequency for six of the seven preparations. At high frequencies, the TM moved at least 2 times more than the tip of the hair bundle in five of the seven preparations. At low frequencies, the TM moved at least 2 times more than the tip of the hair bundle in four of the seven preparations. The other preparations had ratios near 1. Four of the five preparations whose phase could be accurately estimated (both the

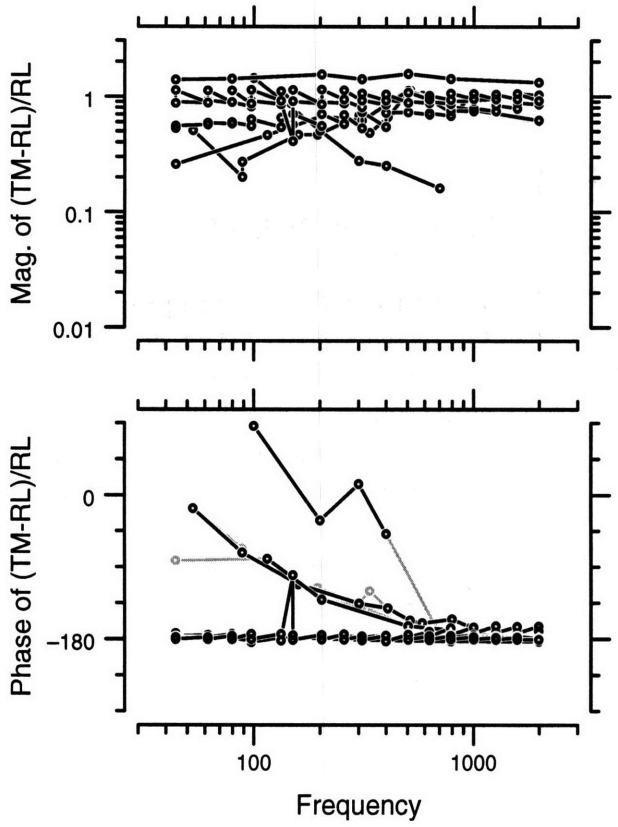


Figure 8-18: Ratio of the shear (the difference between tectorial membrane and reticular lamina motions) to the reticular lamina motion for seven preparations. The curves are derived from data as in Figure 8-17.

numerator and the denominator of the ratio exceeding 100 nm), had phases near 0° (indicating that the TM and tips of the hair bundles were moving in phase), while the other had phases −180°, (indicating that the TM and tips of the hair bundles were moving exactly out of phase).

Figure 8-20 shows the ratio between the differential radial motion of the hair bundles tips and RL to the differential radial motion of the TM and RL. This ratio was less than unity 126 out of the 130 measurements, which indicate that hair bundle shear is not the smaller than the TM/RL shear. The phase of the ratio is near zero.

8.4 Discussion

8.4.1 Microphonic

Measuring the cochlear microphonic is a commonly used method to assess the viability of the cochlea, since a critical step in cochlear function is mechano-to-electric transduction. In 21 preparations, we were unable to measure an electrical response

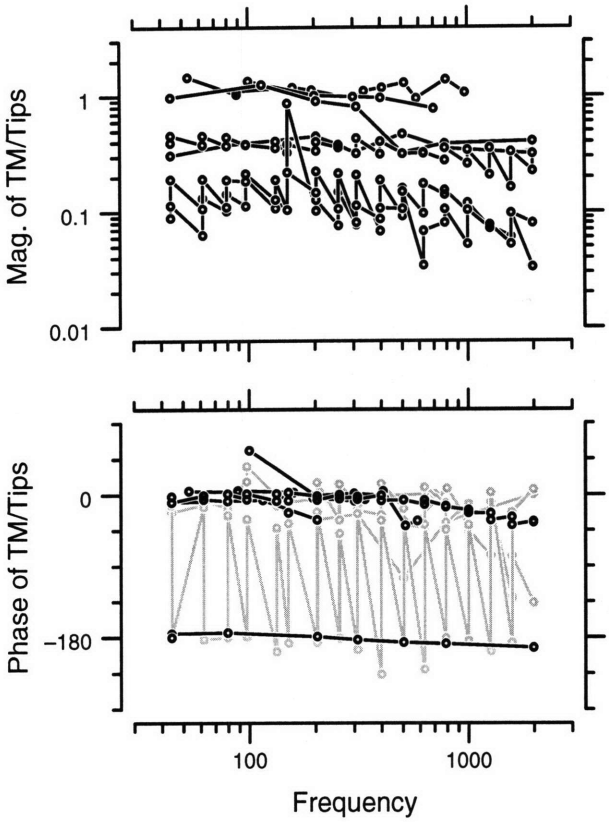


Figure 8-19: Ratio of tectorial membrane displacement to that of the tips of a hair bundle for seven preparations. The displacement of the tectorial membrane is the same as Figure 8-17. The displacement of the tips of a hair bundle was determined in a region of interest enclosing the closest hair bundle to the region of interest used in the tectorial membrane calculation (typically within 10 μm). Shading in the phase data is analogous to that in Figure 8-17.

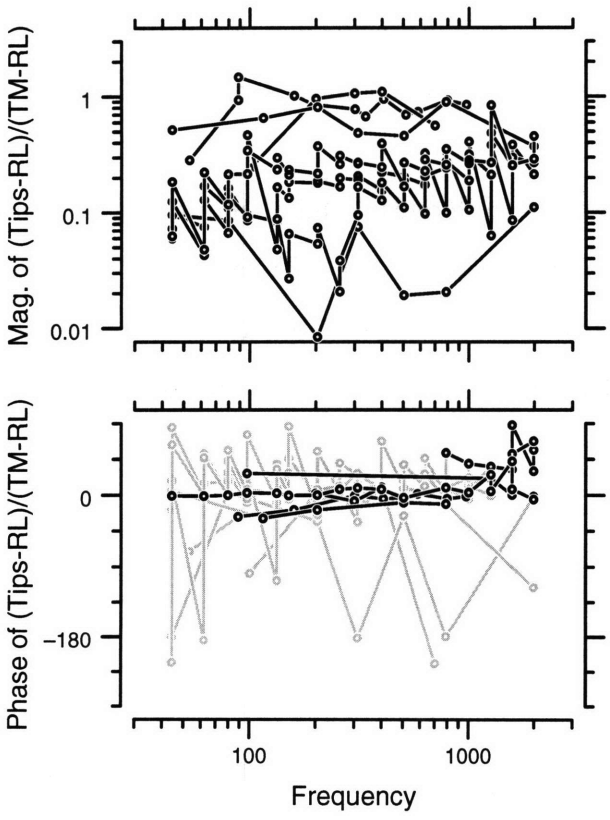


Figure 8-20: Ratio of the TM shear (the difference between tectorial membrane and reticular lamina displacements) to the hair bundle shear (the difference between hair bundle tips and reticular lamina displacement) for seven preparations. Displacements of the tectorial membrane and the tips of hair bundles are the same as in Figure 8-19. Displacements of the reticular lamina were determined separately for the TM and tips, directly below the respective region for the TM or tips. The curves are derived from data as in Figure 8-17.

to sound. We considered four hypotheses to explain why we were unable to measure a microphonic in vitro.

The first hypothesis was that the cochlea was transducing sound, but the electronics were unable to measure it. To demonstrate that this hypothesis was incorrect, we placed a bullfrog saccular epithelium in the chamber and were able to measure the typical $\sim$ 10 nA transduction current (Corey and Hudspeth, 1983).

The second hypothesis was that the cochlea was transducing sound, but the response was being short-circuited. We demonstrated that this hypothesis was incorrect by measuring intracellular potentials in the cochlea. We did not see a response to sound.

The third hypothesis was that the cochlea had lost its ability to transduce sound due to an inappropriate chemical environment. We demonstrated that this hypothesis was invalid by perfusing saline across the cochlea in situ and maintaining a microphonic for several hours. Furthermore, the cochlear microphonic decreased when perfusion stopped and increased upon re-starting the perfusion.

The fourth hypothesis was the cochlea had lost its ability to transduce sound due to damage during surgery. From intracellular recordings on one preparation, we found resting potentials on the limbic bulge to be 56 mV more negative than cells in the basilar papilla. Since the limbic bulge is not as mechanically sensitive as the basilar papilla (Peake and Ling, 1980), this evidence supports the surgical damage hypothesis. From cochlear microphonic measurements in situ, we found that we can typically perform all the surgical steps including cutting the VIIIth nerve while maintaining the cochlear microphonic. Since we could not measure a sound-induced or a mechanical displacement induced (using the glass probe stimulator) cochlear microphonic for the 21 preparations in the experimental chamber, this evidence supports the hypothesis that damage is occurring after cutting the nerve and before measuring the microphonic in the chamber. We have performed one experiment in which the cochlear duct was removed from the head using hydraulically-controlled forceps mounted on a micromanipulator. After lifting the cochlear duct 1 cm above the head, we were able to measure an electrical response to mechanical displacement (induced with a glass

probe stimulator). The electric response was non-linear (compressive) before application of glutaraldehyde. After application of glutaraldehyde, the potential dropped 12 dB and became linear. We conclude that the potential measured before application of glutaraldehyde was an electrical response from the cochlear duct — a response that we were unable to measure previously. The cochlear duct contains both the basilar papilla, which responds to sound, and the lagena, which is an otolith organ. The role of the lagena has not been determined, and the lagena may respond to the glass probe stimulator. Consequently, we cannot conclude that the response from the cochlear duct was from the basilar papilla. Nevertheless, since we were unable to measure any response before using the hydraulically-controlled forceps mounted on a micromanipulator, we conclude that previous extraction techinques were damaging the cochlea. The slow extraction technique holds promise for future in vitro experiments. Our hypothesis for the mechanism of our success was that the slow extraction generates forces at only very low frequencies, to which the hair bundles can adapt.

Our results from testing Hypothesis III (inappropriate chemical environment) also provides useful information about both how to conduct the surgery and experiments. Specifically, our results indicate that the microphonic decays/recovers at a rate of 20 dB per 30 minutes when perfusion is stopped/started. Furthermore, the cochlear microphonic can recover from 30 minute periods with no perfusion. Consequently, during surgeries to remove the cochlear duct, the surgeon has at least 30 minutes to expose the cochlear duct and start perfusion. Typically only 15 minutes are needed to reach the duct; consequently, the surgeon should feel no need to rush through the surgery. Taking extra care not to mechanically disturb the duct is perhaps more important than speed. Furthermore, allowing the duct 30 minutes to "recover" from the surgery should make the cochlea have larger electrical responses during the experiment. During motion measurements, the decay rate implies that the cochlea should be perfused at least every 30 minutes. Since (1) the flow of saline (perfusion) causes the basilar papilla to move and (2) basolateral perfusion requires the chamber's vent to be open, perfusion is turned off during motion measurements. Motion measurements take three seconds per picture or 20 seconds per 2D frequency/amplitude point

(assuming eight evenly-spaced phases are taken); perfusion requirements dictate that no more than 75 2D frequency/amplitude points are taken between perfusions. If 3D motions are desired, then one is limited to about six points (assuming 13 planes of section) between perfusions.

8.4.2 Motions of the sensory receptor organ

None of the motion measurements were done on cochleae that could transduce sound. Nevertheless, these motion measurements are the first reported sound-induced motions of the tectorial membrane and underlying hair bundles. Where there is overlap, our results agree with previously reported in vitro and in vivo measurements. Furthermore, they demonstrate how computer microvision can be used to address important issues in cochlear mechanics.

Comparison with other measurements

Our results on the gross motion of the basilar papilla are consistent with all previously reported data (both in vitro and in vivo).

There are two reports of sound-induced motions of the basilar papilla (Holton and Hudspeth, 1983; Frishkopf and DeRosier, 1983). Their thrust was measuring motions in the free-standing region of the basilar papilla, while our thrust was measuring motions in the tectorial region. However, they also reported on the gross motion of the basilar papilla to which we can compare. A cochlear microphonic was not measured in either of the previous in vitro reports.

Hydrodynamic stimulation causes a rocking motion of the basilar papilla about an axis parallel to its length (Figure 8-16), which is consistent with previous reports (Holton and Hudspeth, 1983; Frishkopf and DeRosier, 1983). This observation is easy to understand in terms of the anatomy. The basilar papilla is asymmetrically located on the thin basilar membrane. With the assumptions that the basilar membrane has spatially uniform stiffness and that the stiffness of the basilar membrane predominates that of other structures, it follows that the abneural side of the basilar membrane will

strech more than the neural side. This asymmetry would cause the basilar papilla to rock back-and-forth as observed.

We found radial displacements of the basilar papilla on the order of 0.5 μm in response to sound pressures on the order of 120 dB SPL in the fluid. This finding is consistent with those reported by Frishkopf and DeRosier (1983). Holton and Hudspeth (1983) did not record the sound pressure in their chamber, thus a direct comparison cannot be made with their results.

The transverse component of the 3D frequency response of the tectorial region of the basilar papilla at constant sound pressure is, to first order, constant in magnitude and has zero phase (Figure 8-13). This result is consistent with the basilar membrane and extrastapes having similar tuning curves (Weiss, Peake, Ling and Holton, 1978; Peake and Ling, 1980). Furthermore, the transverse motions of the basliar papilla that we measured, in vitro and without a cochlear microphonic, span the same range of magnitudes as those measured with the Mössbauer technique in vivo and with a cochlear microphonic (Peake and Ling, 1980).

8.4.3 Implications for the tectorial membrane

Unlike the gross motions of the basilar papilla, our results are the first measurements of sound-induced motions of the tectorial membrane and underlying hair bundle. Consequently, comparisons with others results cannot be made.

The motion of the TM was generally less than that of the RL (Figure 8-17) and decreased slightly with frequency. Thus, these results indicate that the TM's mass becomes mechanically important at high frequencies. If the TM were stationary, then the micromechanical transfer function (Figure 8-18) would be equal to -1, which is consistent with four of the seven preparations. In the other three preparations, the phase was near zero at low frequencies and decreased to $-180°$ with increasing frequency. These results are characteristic of the response of a highly damped second order resonant system. Thus an alternative explanation of the 180 phase lag is that it could result as the high frequency asymptote of a second order resonant system. These results indicate that damping should figure prominently in mechanical models

of the TM.

Figures 8-19 and 8-20 demonstrate that the top of the tectorial membrane and the tips of the hair bundles move differently. Since the tectorial membrane covers the hair bundles (Mulroy, 1974), these results indicate that the TM shears through its thickness. Thus, the TM cannot be treated as a rigid body.

8.4.4 Implications for frequency selectivity

The macromechanical motions of the basilar papilla (Figure 8-16) do not contribute to the frequency selectivity seen on the auditory nerve (Weiss et al., 1976; Weiss, Peake, Ling and Holton, 1978; Holton, 1980). A "second filter" is needed to account for the sharp frequency selectivity of the auditory nerve in comparison to the macromechanical motion of the basilar papilla. However, neither the micromechanical transfer function (Figure 8-18) nor the transfer function relating TM/RL shear to hair bundle rotation (Figure 8-20) is sharply tuned. Thus, these results indicate that the second filter is not mechanical at the level of hair bundle rotation in the TM region.

Chapter 9

Other Applications of Computer Microvision

In the previous chapters, we have shown that computer microvision can be used to solve problems in inner-ear mechanics and to analyze motions of a microfabricated accelerometer. In this chapter we demonstrate that computer microvision has more general applications. In particular, we demonstrate that computer microvision can be used to measure motions of (1) the fibrillar structure in the mouse tectorial membrane, (2) the human stapes, and (3) a microfabricated angular velocity sensor.

9.1 Mouse tectorial membrane

The tectorial membrane (TM) is strategically located in mammalian cochleae on top of the hair bundles—the mechano-to-electric transduction apparatus. Consequently, the TM is assumed to have great importance in cochlear function (See references in Hawkins et al., 1996). The mechanical properties of the TM are critical to cochlear models, but have not been extensively studied experimentally.

Abnet and Freeman (1997) have developed an isolated preparation of the mouse tectorial membrane to study its mechanical properties. The TM is isolated, glued to a glass slide, and 20 μm magnetizable bead is attached to its surface. Sinusoidal forces are generated using electromagnets. The displacement of the bead and of the

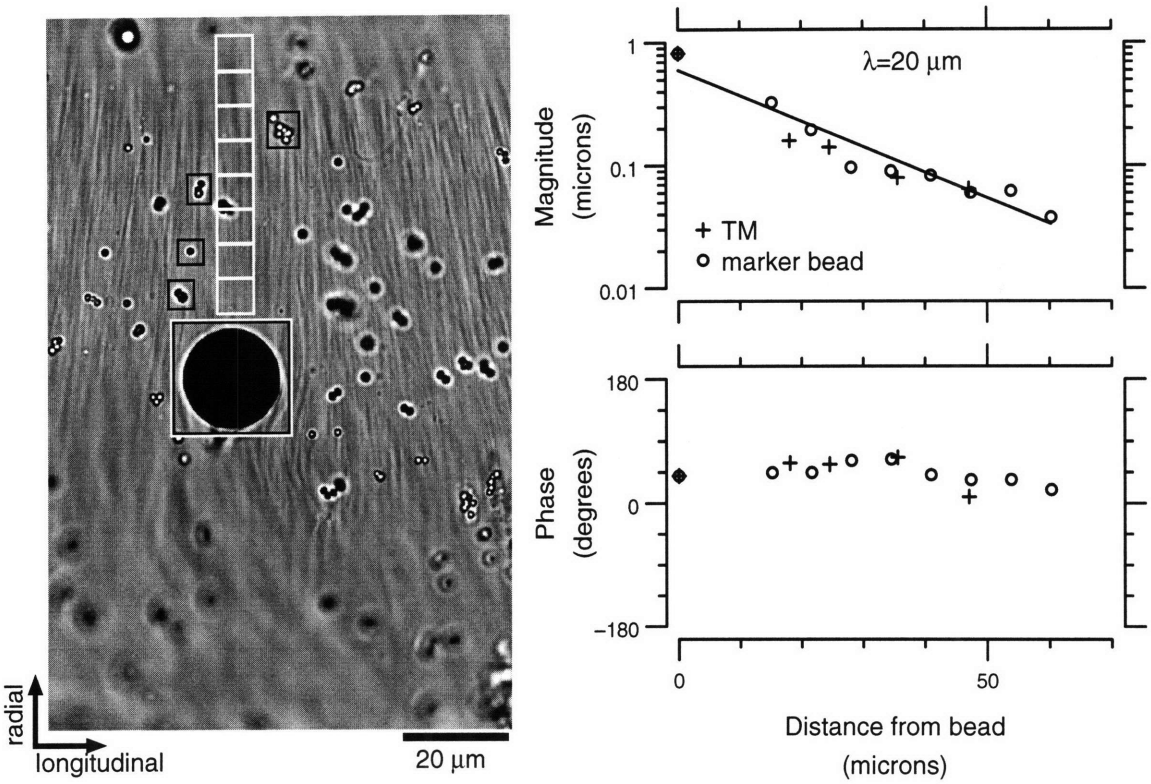


Figure 9-1: Mechanical properties of the mouse tectorial membrane. The image (left) shows the isolated mouse tectorial membrane. The radial fibrillar structure of the TM is easily seen. Regions of interest (boxes) that were used to determine the deformation pattern in the TM as a function of radial distance from the bead are shown on the TM. Two sets of regions for each direction were made; a regularly-spaced grid along the TM (white boxes) and regions surrounding non-magnetic marker beads (black boxes). The large (20 μm diameter) magnetizable bead is surrounded by a black box and a white box. The plots (right) show the magnitude and phase of the fundamental component of longitudinal motion as a function of the distance from the magnetizable bead. The magnetizable bead was driven with a 10 Hz, 0.8 μN longitudinal force.

TM are determined with computer microvision.

Motion measurements were taken in two series of regions of interest, one containing and the other lacking high-contrast, non-magnetic marker beads (Figure 9-1). The data indicate that there is not a significant difference in the motion estimates between regions of interest that contain or lack marker beads. Consequently, marker beads do not have to be used to characterize the motion of the TM.

Computer microvision provides significant advantages over previous methods (von Békésy, 1960; Zwislocki and Cefaratti, 1989) for studying mechanical properties of the TM. For example, the tectorial membrane can be studied without chemical con-

trast agents. The point impedance of the TM can be measured, which describes the mechanical impedance of the TM seen by individual hair bundles. The space constant (λ) can also be determined from the same images, which describes the spatial extent of coupling in the TM (e.g., Figure 9-1). In general, the 3D strain field throughout the TM can be measured, which is important for characterizing (1) mechanical loads on hair bundles, (2) interactions between hair bundles, and (3) the functional role of the TM's fibrillar structure. Finally, measurement of displacements as small as a few nanometers enables study of the mechanical role of the TM at normal hearing levels.

9.2 Human Stapes

Measurements of stapes motion from a fresh human cadaver were taken with the computer microvision system using polarized, reflected light microscopy[1]. The ossicles and tympanic membrane were left intact. Sound was delivered to the tympanic membrane. The stapes was imaged through a hole drilled into the middle ear space.

A 3D frequency response of the stapes at a constant, uncalibrated SPL near 115 dB is shown in Figure 9-2. Variability is typically a decade below the mean magnitude and around 10° of phase. The dependence of magnitude on frequency is similar to that of the cat (Guinan and Peake, 1967). The similar phase of the motions in all three directions indicate that this section of the stapes is moving in a straight line.

[1]These experiments were done with Susan Voss.

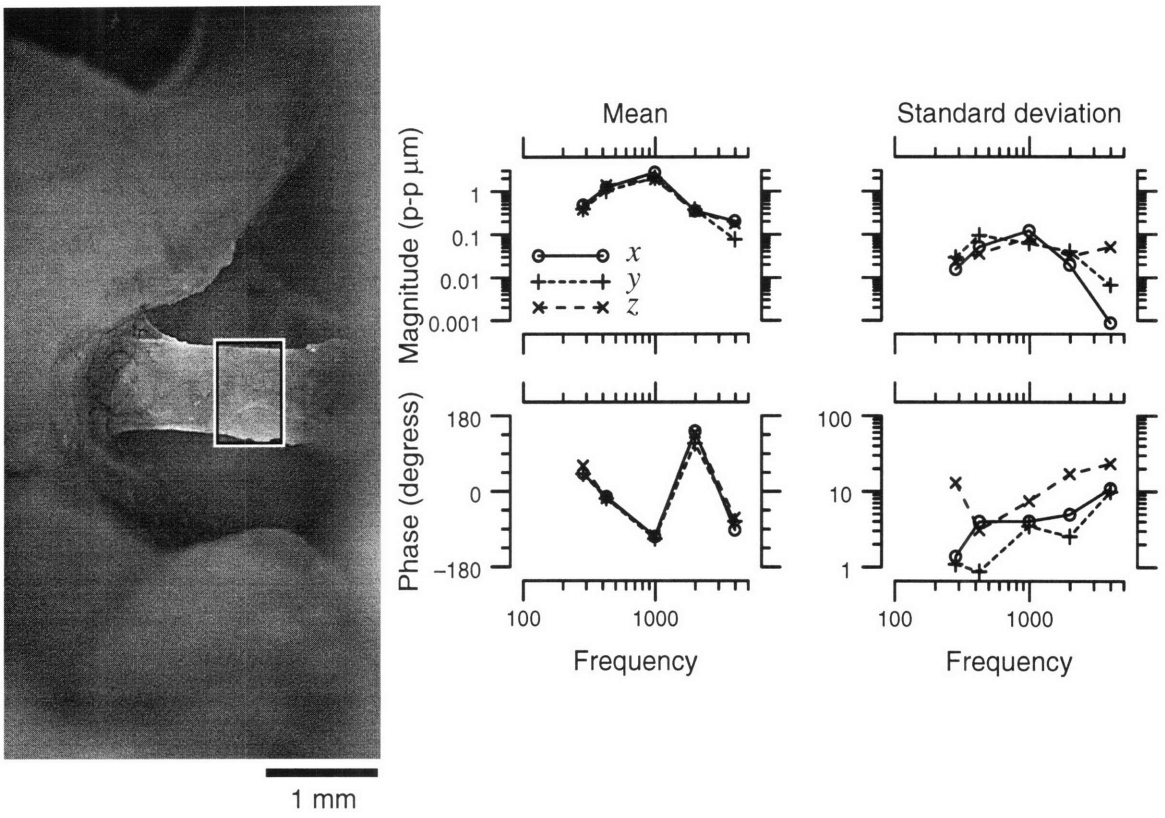


Figure 9-2: Human stapes 3D frequency response. The image (left panel) shows the round window (upper left) and the stapes (center). The rectangular region of interest highlights a section of the posterior crus of the stapes near its joint with the incus. The anterior crus can be faintly seen in the background. The right panels show the 3D frequency response for the motion of posterior crus inside the highlighted region. A higher magnification (20× 0.4 NA objective) was used to collect motion data, which gave a field of view equivalent to the indicated region. That region was split lengthwise into four, so that separate motion estimates could be made to calculate standard deviations of the measurements. The data used to estimate the motion consisted of images taken at 8 time points for each of 40 planes of focus (separated by 5 μm). In the image, the $+x$ direction is to the right; $+y$, up; and $+z$, out of the page.

Computer microvision provides significant advantages over other methods of measuring middle ear mechanics (Guinan and Peake, 1967; Gyo et al., 1987). Motion can be measured without contrast aids (e.g., mirrors, silver particles, or steel spheres), the placement of which could load or damage the ossicle. Each measurement provides 3D motion estimates. Three-dimensional motion estimates from at least two positions (and the assumption that the bone is rigid), can be used to calculate five degree of freedom motion for the ossicle[2]. This description could then determine, for example, (1) the extent of flexing in ossicular joints, (2) the extent to which stapes motion is piston-like, and (3) the ratio of stapes motion to malleus motion (lever arm ratio).

Middle ear mechanics provides a number of challenges to computer microvision, the primary challenge being that the motions are thousands of times smaller than the structures. At magnifications/numerical apertures sufficient for motion estimation, only a fraction of an ossicle is visible. Thus, one has the additional problem of determining orientation and position changes between measurements on different parts of the ossicle. This challenge is not unique to computer microvision, and should be easier to solve using a video method (like computer microvision) than other methods of motion detection (such as laser doppler). With video methods, low-magnification images can be taken to help register the higher magnification motion images.

9.3 Microfabricated angular velocity sensor

Microelectromechanical systems (MEMS) hold great commercial promise for revolutionizing the fields of sensor and actuator design (O'Connor, 1992; Gabriel, 1995). MEMS devices can be mass-produced, and their small size opens up a host of new applications. We examined a microfabricated accelerometer in previous chapters; here we examine a microfabricated angular velocity sensor called a tuning-fork gyroscope.[3]

The angular velocity sensor has two shuttles which are driven in the plane of the

[2]The sixth degree of freedom, rotation about the line connecting the regions of interest, can be determined from motion estimates in three non-collinear regions of interest.

[3]This work was done with Laura Johnson. For more details, see Johnson (1997). The gyroscope was designed and manufactured by C.S. Draper Laboratories (Cambridge, MA).

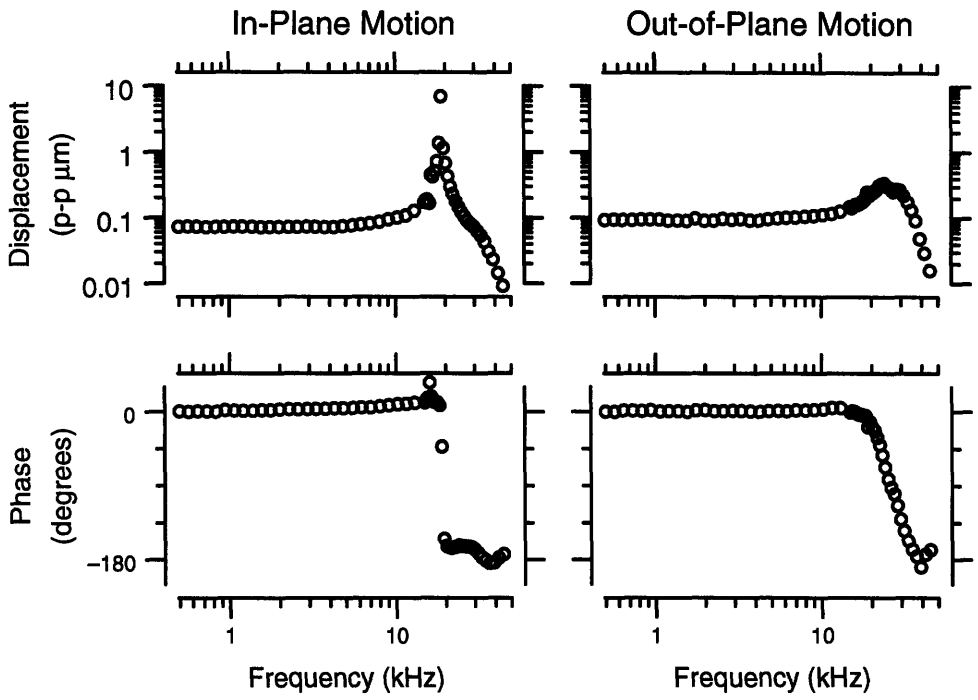


Figure 9-3: Frequency response of a small section of one shuttle of microfabricated angular velocity sensor in the driven (in-plane) and unintended (out-of-plane) directions. For 56 frequencies, images at 8 evenly-spaced phases of the stimulus were taken at 30 planes of section spanning 16 μm. The magnitude (top) and phase (bottom) of the fundamental component of motion is plotted.

device. The Coriolis force causes deflections of the shuttle's motion in proportion to the imposed angular velocity. These deflections change the distance between the shuttles and the device's substrate, which is measured as a change in capacitance and interpreted as an angular velocity.

Figure 9-3 demonstrates that we can obtain a frequency response of the shuttle's motion. The in-plane motion frequency response demonstrates that computer microvision can measure a 60 dB range of motion amplitudes. Even though no angular velocity was imposed on the device, we measured out-of-plane motions well above our noise floor. These out-of-plane motions limit the device's ability to measure angular velocities.

The computer microvision system can easily measure the resonant frequencies for in-plane and out-of-plane motions, which are different. Furthermore, the out-of-plane resonance consists of three sub-peaks. Computer microvision can resolve those peaks as different modes of motion. The lowest-frequency peak coincides with the in-plane

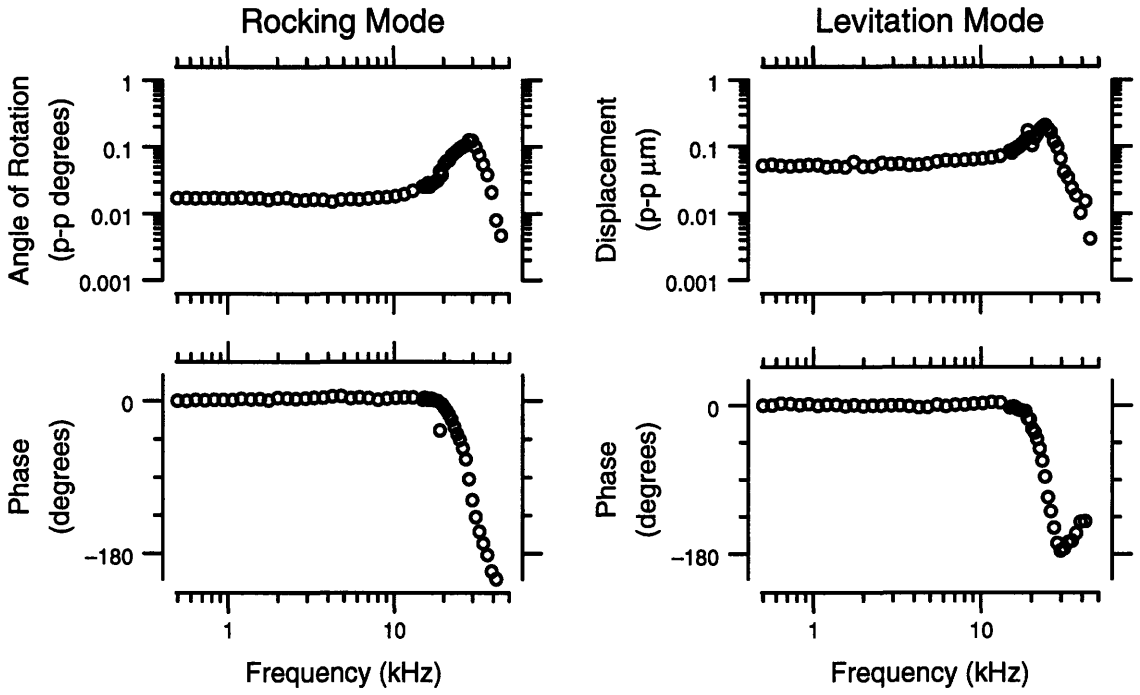


Figure 9-4: Frequency response of a the rocking and levitation out-of-plane modes of motion of a microfabricated angular velocity sensor. For 56 frequencies, images at 8 evenly-spaced phases of the stimulus were taken at 30 planes of section spanning 16 μm. Three-dimensional motion estimates from an array of 12 regions of interest spanning the shuttle's length were combined to estimate the modes of motion. For the rocking mode, an angle waveform (relative to the substrate) was computed from the slope of the linear least-squares fit to the array of displacement waveforms. The magnitude and phase of the fundamental component of this angle waveform is plotted for each frequency. For the levitation mode, a levitation waveform was computed as the average across the regions of the out-of-plane displacement not accounted for by the rotation. Plotted are magnitude and phase of the fundamental component of this levitation waveform. The center of rotation was chosen to be in the center of the shuttle.

resonance. The other two peaks result from (1) the entire shuttle "levitating" above the substrate and (2) the shuttle "rocking" about an axis that is parallel to the substrate and orthogonal to the in-plane motion.

Figure 9-4 demonstrates the frequency response of the levitation and rocking modes of the shuttle's motion. The data illustrate that computer microvision can measure rotations approaching one millidegree. Further, one can see that the resonant frequencies of the two modes are different, and their combination accounts for the higher frequency peaks in the out-of-plane frequency response shown in Figure 9-3. The device has other modes of motion, which for example caused the spurious

positive phase in the in-plane frequency response (Johnson, 1997). By combining three-dimensional motion estimates across different regions of the shuttle, computer microvision can characterize *all* the modes of motion. Knowledge of the modes is critical for designers to be able to fabricate better devices. Just as electronic engineers have electronic probe stations to test their devices, computer microvision holds promise to become a "motion probe station" for MEMS designers.

9.4 Conclusions

Computer microvision combines video microscopy with computer vision to estimate nanometer, three-dimensional motions. We have measured motions of structures whose size spanned 68 dB (0.3 μm to 800 μm), whose motions spanned 66 dB (5 nm to 10,000 nm), and at frequencies spanning 94 dB (1 Hz to 50 kHz). In principle, computer microvision can measure motions of structures over an even larger range of size, motions, and frequencies.

The system provides non-contact motion measurements without the need for artificial contrast enhancements. The target whose motion is estimated is unambiguous, since estimates are determined from images of the target. Furthermore, the images can be analyzed at multiple scales to estimate gross motions and fine details.

We have used this system to make important contributions in hearing—by analyzing sound-induced motions of the alligator lizard papilla, sound-induced motions of the human stapes, and the mechanical deformation pattern in isolated tectorial membranes—as well in the study of micromachines—by analyzing electrically-induced motions of a microfabricated accelerometer and a microfabricated angular velocity sensor.

We feel that computer microvision is a very general-purpose tool whose range of applications is only now being discovered.

Bibliography

Abnet, C. C. and Freeman, D. M. (1997). Mechanical properties of the isolated mouse tectorial membrane, *Abstracts of the Twentieth Midwinter Research Meeting*, Association for Research in Otolaryngology, St. Petersburg Beach, Florida.

Aggarwal, J. and Nandhakumar, N. (1988). On the computation of motion from sequences of images—a review, *Proc. IEEE* **76**(8): 917–934.

Allen, J. B. (1980). Cochlear micromechanics – a physical model of transduction, *J. Acoust. Soc. Am.* **68**: 1660–1670.

Assad, J. A., Shepard, M. G. and Corey, D. P. (1991). Tip-link integrity and mechanical transduction in vertebrate hair cells, *Neuron* **7**: 985–994.

Barron, J., Fleet, D. and Beauchemin, S. (1994). Performance of optical flow techniques, *IJCV* **12**(1): 43–77.

Behringer, U., Haug, W., Meissner, K., Ziemlich, W., Bohlen, H., Bayer, T., Kulcke, W., Rothuizen, H., Sasso, G. and Vettiger, P. (1991). The electron beam proximity printing lithography, a candidate for the 0.35 and 0.25 micron chip generations, *Microelectronic Engineering* **13**: 361–364.

Bergen, J. et al. (1992). Hierarchical model-based motion estimation, *Proceedings of the Second European Conference on Computer Vision*, Santa Margherita Ligure, Italy.

Born, M. and Wolf, E. (1975). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Pergamon Press, New York.

Burt, P. J. (1981). Fast filter transfroms for image processing, *Computer Graphics and Image Processing* **16**: 20–51.

Burt, P. J. (1984). *Multiresolution image processing and analysis*, Springer-Verlag, chapter The pyramid as a structure for efficient computation.

Cook, R. O. and Hamm, C. W. (1979). Fiber optic lever transducer, *App. Optics.* **18**: 3230–3241.

Cooper, N. and Rhode, W. (1992). Basilar membrane mechanics in the hook region of the cat and guinea pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts, *Hearing Res.* **63**: 163–190.

Corey, D. P. and Hudspeth, A. J. (1983). Analysis of the microphonic potential of the bullfrog's sacculus, *J. Neurosci.* **3**(5): 942–961.

Crawford, A. C., Evans, M. G. and Fettiplace, R. (1989). Activation and adaptation of transducer currents in turtle hair cells, *J. Physiol.* **419**: 405–434.

Crawford, A. C., Evans, M. G. and Fettiplace, R. (1991). The actions of calcium on the mechano-electrical transducer current of turtle hair cells, *J. Physiol.* **434**: 369–398.

Crawford, A. C. and Fettiplace, R. (1985). The mechanical properties of ciliary bundles of turtle cochlear hair cells, *J. Physiol.* **364**: 359–379.

Dallos, P., Billone, M. C., Durrant, J. D., Wang, C. Y. and Raynor, S. (1972). Cochlear inner and outer hair cells: Functional differences, *Science* **177**: 356–358.

Dallos, P., Popper, A. N. and Fay, R. R. (eds) (1996). *The Cochlea*, Springer-Verlag, New York.

Davis, C. Q. and Freeman, D. M. (1995). Direct observations of sound-induced motions of the reticular lamina, tectorial membrane, hair bundles, and individual

stereocilia, *Abstracts of the Eighteenth Midwinter Research Meeting*, Association for Research in Otolaryngology, St. Petersburg Beach, Florida.

Davis, C. Q. and Freeman, D. M. (1997a). Nanometer registration with light microscopy: characterization of performance limitations, *Optical Engineering* . In preparation.

Davis, C. Q. and Freeman, D. M. (1997b). Statistics of subpixel registration algorithms based on piecewise linear interpolators, *Optical Engineering* . In preparation.

Davis, C. Q., Karu, Z. Z. and Freeman, D. M. (1995). Equivalence of subpixel motion estimators based on optical flow and block matching, *IEEE International Symposium for Computer Vision*, Coral Gables, Florida, pp. 7–12.

Davis, H. (1958). A mechano-electrical theory of cochlear action, *Ann. Otol., Rhinol. and Laryngol.* **67**: 789–801.

Douglas, S. C. (1993). A frequency-domain subpixel position estimation algorithm for overlay measurement, *SPIE* **1926**: 402–411.

Duncan, R. K., Hernandez, H. N. and Saunders, J. C. (1995). Relative stereocilia motion of chick cochlear hair cells during high-frequency water-jet stimulation, *Aud. Neurosci.* **1**: 321–329.

Fay, R. R. and Popper, A. N. (eds) (1994). *Comparative Hearing: Mammals* in *Springer Handbook of Auditory Research*, Springer-Verlag, New York.

Freeman, D. M. and Davis, C. Q. (1996). Using video microscopy to characterize micromechanics of biological and man-made micromachines, *Technical Digest of the Solid-State Sensor and Actuator Workshop*, Transducers Research Foundation, Inc., Hilton Head Island, SC.

Freeman, D. M., Hendrix, D. H., Shah, D., Fan, L. F. and Weiss, T. F. (1993). Effect of lymph composition on an *in vitro* preparation of the alligator lizard cochlea, *Hearing Res.* **65**: 83–98.

Freeman, D. M. and Weiss, T. F. (1988). The role of fluid inertia in mechanical stimulation of hair cells, *Hearing Res.* **35**: 201–208.

Freeman, D. M. and Weiss, T. F. (1990). Hydrodynamic analysis of a two-dimensional model for micromechanical resonance of free-standing hair bundles, *Hearing Res.* **48**: 37–68.

Frishkopf, L. S. and DeRosier, D. J. (1983). Mechanical tuning of free-standing stereociliary bundles and frequency analysis in the alligator lizard cochlea, *Hearing Res.* **12**: 393–404.

Gabriel, K. J. (1995). Engineering microscopic machines, *Scientific American* **273**: 150–153.

Gharavi, H. and Mills, M. (1990). Blockmatching motion estimation algorithms - new results, *IEEE Transactions on Circuits and Systems* **37**(5): 649–651.

Giorda, F. and Racciu, A. (1975). Bandwidth reduction of video signals via shift vector transmission, *IEEE Transactions on Communications* **23**: 1002–1004.

Girod, B. (1993). Motion-compensating prediction with fractional-pel accuracy, *IEEE Transactions on Communications* **41**(4).

Goodman, J. W. (1968). *Introduction to Fourier Optics*, McGraw-Hill Book Company, New York.

Grenon, B., Defibaugh, D., Sprout, D. and Taft, C. (1995). Manufacturing performance of the ALTA 3000 mask laser writer, *15th Annual Symposium on Photomask Technology and Management*, Vol. 2621, SPIE, pp. 14–18.

Guinan, J. J. and Peake, W. T. (1967). Middle-ear characteristics of anesthetized cats, *J. Acoust. Soc. Am.* **41**: 1237–1261.

Gyo, K., Aritomo, H. and Goode, R. L. (1987). Measurement of the ossicular vibration ratio in human temporal bones by use of a video measuring system, *Acta Otolarygol* **103**: 87–95.

Hachincha, A. and Simon, S. (1988). Subpixel edge detection for precise measurements by a vision system, *Proc. SPIE, Industrial Inspection* **1010**: 148–157.

Hawkins, H. L., McMullen, T. A., Popper, A. N. and Fay, R. R. (eds) (1996). *Auditory Computation*, Springer-Verlag, New York.

Healey, G. E. and Kondepudy, R. (1994). Radiometric CCD camera calibration and noise estimation, *IEEE Trans. Pattern Anal. Machine Intell.* **16**: 267–276.

Hemmert, W., Zenner, H.-P. and Gummer, A. (1995). Three-dimensional vibration measurements on different locations of the cochlear partition, *Abstracts of the Eighteenth Midwinter Research Meeting*, Association for Research in Otolaryngology, St. Petersburg Beach, Florida.

Hiraoka, Y., Sedat, J. W. and Agard, D. A. (1987). The use of charge-coupled device for quantitative optical microscopy of biological structures, *Science* **238**: 36–41.

Hohman, B. (1997). Product Manager, Carl Zeiss, Inc., Personal Communication.

Holton, T. (1980). Relations between frequency selectivity and two-tone rate suppression in lizard cochlear-nerve fibers, *Hearing Res.* **2**: 21–38.

Holton, T. and Hudspeth, A. J. (1983). A micromechanical contribution to cochlear tuning and tonotopic organization, *Science* **222**: 508–510.

Holton, T. and Hudspeth, A. J. (1986). The transduction channel of hair cells from the bull-frog characterized by noise analysis, *J. Physiol.* **375**: 195–227.

Holton, T. and Weiss, T. F. (1978). Two-tone rate suppression in lizard cochlear nerve fibers, relation to receptor organ morphology, *Brain Res.* **159**: 219–222.

Holton, T. and Weiss, T. F. (1983). Receptor potentials of lizard cochlear hair cells with free-standing stereocilia in response to tones, *J. Physiol.* **345**: 205–240.

Hopkins, H. H. and Barham, P. M. (1950). The influence of the condenser on microscopic resolution, *Proc. Phys. Soc.* **63**: 737–744.

Horn, B. K. P. (1986). *Robot Vision*, MIT Press, Cambridge, MA.

Horn, B. K. P. and Weldon, Jr., E. (1988). Direct methods for recovering motion, *Internatl. J. of Computer Vision* **2**: 51–76.

Horn, B. K. and Schunck, B. G. (1981). Determining optical flow, *Artificial Intelligence* **17**: 185–203.

Howard, J. and Hudspeth, A. J. (1988). Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the bullfrog's saccular hair cell, *Neuron* **1**: 189–199.

Howard, J., Roberts, W. M. and Hudspeth, A. J. (1988). Mechanoelectrical transduction by hair cells, *Ann. Rev. Biophys. Chem.* **17**: 99–124.

Hudspeth, A. J. (1983). The hair cells of the inner ear, *Sci. Am.* **248**: 54–64.

Hudspeth, A. J. (1989). How the ear's works work, *Nature* **341**: 397–404.

Hudspeth, A. J. and Gillespie, P. G. (1994). Pulling strings to tune transduction: adaptation by hair cells, *Neuron* **12**: 1–9.

Inderhees, G. and Kiefer, B. (1995). Integration of an advanced laser writer into a manufacturing environment, *15th Annual Symposium on Photomask Technology and Management*, Vol. 2621, SPIE, pp. 36–49.

Inoué, S. (1986). *Video Microscopy*, Plenum Press.

Jähne, B. (1990). Motion determination in space-time images, *ECCV* pp. 161–173.

Janesick, J. R., Elliot, T., Collins, S., Blouke, M. M. and Freeman, J. (1987). Scientific charge-coupled devices, *Optical Engineering* **26**(8): 692–714.

Johnson, L. K. (1997). *Using computer microvision to characterize the motions of a microfabricated gyroscope*, Master's thesis, MIT.

Johnstone, B. and Boyle, A. (1967). Basilar membrane vibration examined with the Mössbauer technique, *Science* **158**: 389–390.

Kaplan, M. S., Szaro, B. G. and Weiss, T. F. (1983). Components of cochlear electric responses in the alligator lizard, *Hearing Res.* **12**: 323–351.

Khanna, S. M. and Leonard, D. G. B. (1982). Basilar membrane tuning in the cat cochlea, *Science* **215**: 305–306.

Khanna, S. M. and Leonard, D. G. B. (1986). Measurement of basilar membrane vibrations and evaluation of the cochlear condition, *Hearing Res.* **23**: 37–53.

Kimitsuki, T., Nakagawa, T., Hisashi, K., Komune, S. and Komiyama, S. (1996). Gadolinium blocks mechano-electric transducer current in chick cochlear hair cells, *Hearing Res.* **101**: 75–80.

Kimitsuki, T. and Ohmori, H. (1992). The effect of caged calcium release on the adpatation of the transduction current in chick hair cells, *J. Physiol.* **458**: 27–40.

Komarek, T. and Pirsch, P. (1989). Array architectures for block matching algorithms, *IEEE Transactions on Circuits and Systems* **36**(10): 1301–1308.

Kössl, M. and Russell, I. J. (1995). Basilar membrane displacement in the cochlea of the mustached bat, *Abstracts of the Eighteenth Midwinter Research Meeting*, Association for Research in Otolaryngology, St. Petersburg Beach, Florida.

MacDonald, R. B. and Corey, D. P. (1996). Stereocilia bundles of the bullfrog sacculus hair cells do not splay in response to stimulation, *Abstracts of the Nineteenth Midwinter Research Meeting*, Association for Research in Otolaryngology, St. Petersburg Beach, Florida.

Mammano, F. and Nobili, R. (1993). Biophysics of the cochlea: Linear approximation, *J. Acoust. Soc. Am.* **93**: 3320–3332.

Mulroy, M. J. (1974). Cochlear anatomy of the alligator lizard, *Brain Behav. Evol.* **10**: 69–87.

Neely, S. T. (1981). Finite difference solution of a two-dimensional mathematical model of the cochlea, *J. Acoust. Soc. Am.* **69**(5): 1386–1393.

Neely, S. T. (1993). A model of cochlear mechanics with outer hair cell motility, *J. Acoust. Soc. Am.* **94**(1): 137–146.

Netravali, A. N. and Haskell, B. G. (1988). *Digital pictures: representation and compression*, Plenum Press, New York.

O'Connor, L. (1992). MEMS: Microelectromechanical systems, *Mechanical Engineering* **114**(2): 40–47.

Parker, J., Kenyon, R. and Troxel, D. (1983). Comparison of interpolating methods for image resampling, *IEEE Transactions on medical imaging* **2**(1): 31–39.

Peake, W. T. and Ling, A. L. (1980). Basilar-membrane motion in the alligator lizard: Its relation to tonotopic organization and frequency selectivity, *J. Acoust. Soc. Am.* **67**: 1736–1745.

Photometrics (1993). *Final test report, series 200 camera system*, Photometrics, Ltd., Tucson, Arizona.

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing*, second edn, Academic Press, San Diego, CA.

Pickles, J. O. and Corey, D. P. (1992). Mechanoelectric transduction by hair cells, *Trends in Neurosci.* **15**: 254–259.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vettering, W. T. (1992). *Numerical Recipes in C*, second edn, Cambridge University Press, Cambridge.

Pritt, M. D. (1994). Automated subpixel image registration of remotely sensed imagery, *IBM J. Res. Develop.* **38**(2): 157–166.

Roach, J. and Aggarwal, J. (1980). Determining the movement of objects from a sequence of images, *IEEE Trans. PAMI* **2**(6): 554–562.

Rosowski, J. J., Peake, W. T., Lynch, T. J., Leong, R. and Weiss, T. F. (1985). A model for signal transmission in an ear having hair cells with free-standing stereocilia: II. macromechanical stage, *Hearing Res.* **20**: 139–155.

Ruggero, M. A. and Rich, N. C. (1991). Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration, *Hearing Res.* **51**: 215–230.

Shotwell, S. L., Jacobs, R. and Hudspeth, A. J. (1981). Directional sensitivity of individual vertebrate hair cells to controlled deflection of their hair bundles, *Ann NY Acad. Sci.* **374**: 1–10.

Siebert, W. M. (1974). Ranke revisited – a simple short-wave cochlear model, *J. Acoust. Soc. Am.* **56**: 594–600.

Tanner, J. and Mead, C. (1986). An integrated analog optical motion sensor, *VLSI Signal Processing II: Proceedings of the IEEE Acoustics, Speech, and Signal Processing Society Workshop on VLSI Signal Processing*, pp. 50–67.

Tian, Q. and Huhns, M. N. (1986). Algorithms for subpixel registration, *Computer Vision, Graphics, and Image Processing* **35**: 220–233.

Ver, I. L., Brown, R. M. and Kiang, N. Y. S. (1975). Low-noise chambers for auditory research, *Journal of the Acoustical Society America* **58**(2): 392–398.

Viergever, M. and Diependaal, R. (1986). Quantitative validation of cochlear models using the Liouville-Green approximation, *Hearing Res.* **21**: 1–15.

von Békésy, G. (1960). *Experiments in Hearing*, McGraw-Hill, New York.

Weast, R. C. (ed.) (1987). *CRC Handbook of Chemistry and Physics*, 68th edn, CRC Press.

Weiss, S., Ogletree, D. F., Botkin, D., Salmeron, M. and Chemla, D. S. (1993). Ultrafast scanning probe microscopy, *Appl. Phys. Lett.* **63**(18): 2567–2569.

Weiss, T. F. (1996a). *Cellular Biophysics: Electrical Properties*, MIT Press, Cambridge, MA.

Weiss, T. F. (1996b). *Cellular Biophysics: Transport*, MIT Press, Cambridge, MA.

Weiss, T. F., Altmann, D. W. and Mulroy, M. J. (1978). Endolymphatic and intracellular resting potential in the alligator lizard cochlea, *Pflugers Arch.* **373**: 77–84.

Weiss, T. F., Mulroy, M. J., Turner, R. G. and Pike, C. L. (1976). Tuning of single fibers in the cochlear nerve of the alligator lizard: Relation to receptor organ morphology, *Brain Res.* **115**: 71–90.

Weiss, T. F., Peake, W. T., Ling, A. and Holton, T. (1978). Which structures determine frequency selectivity and tonotopic organization of vertebrate cochlear nerve fibers? evidence from the alligator lizard, *in* R. Naunton and C. Fernandez (eds), *Evoked Electrical Activity in the Auditory Nervous System*, Academic Press, New York, pp. 91–112.

Weng, J., Ahuja, N. and Huang, T. (1993). Optimal motion and structure estimation, *IEEE Trans. PAMI* **15**(9): 864–884.

Wickramasinghe, H. K. (ed.) (1991). *AIP Conference Proceedings 241: Scanned Probe Microscopy*, American Institute of Physics, New York.

Wilson, J. P. and Johnstone, J. R. (1975). Basilar membrane and middle-ear vibration in guinea pig measured by capacitive probe, *J. Acoust. Soc. Am.* **57**: 705–723.

Zwislocki, J. J. and Cefaratti, L. K. (1989). Tectorial membrane II: Stiffness measurements *in vivo*, *Hearing Res.* **42**: 211–228.