Rapid confocal imaging for measuring sound-induced motion of the hearing organ in the apical region

Stefan Jacob
Igor Tomo
Anders Fridberger
Jacques Boutet de Monvel
Mats Ulfendahl
Karolinska Institutet
Center for Hearing and Communication Research
Departments of Otolaryngology and Clinical Neuroscience
M1 Karolinska Universitetssjukhuset
SE-171 76 Stockholm, Sweden
E-mail: anders.fridberger@ki.se

Abstract. We describe a novel confocal image acquisition system capable of measuring the sound-evoked motion of the organ of Corti. The hearing organ is imaged with a standard laser scanning confocal microscope during sound stimulation. The exact temporal relation between each image pixel and the sound stimulus is quantified. The motion of the structures under study is obtained by fitting a Fourier series to the time dimension of a continuous sequence of acquired images. Previous versions of this acquisition system used a simple search to find pixels with similar phase values. The Fourier series approach permits substantially faster image acquisition with reduced noise levels and improved motion estimation. The system is validated by imaging various vibrating samples attached to a feedback-controlled piezoelectric translator. When using a rigid sample attached to the translator, the system is capable of measuring motion with peak-to-peak amplitudes smaller than 50 nm with an error below 20% at frequencies between 50 and 600 Hz. Examples of image sequences from the inner ear are given, along with detailed performance characteristics of the method. © *2007 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.2718568]

Keywords: hearing; hair cells; confocal microscopy; rapid imaging; optical flow; wavelets.

Paper 06252SSR received Sep. 15, 2006; revised manuscript received Dec. 20, 2006; accepted for publication Dec. 21, 2006; published online Mar. 29, 2007.

1 Introduction

Hearing relies on the transformation of basilar membrane vibration into deflection of stereocilia atop the sensory hair cells. Although substantial data have accumulated about the vibration of the basilar membrane in the base of the cochlea[1] and about reticular lamina motion in the apex,[2] we have very limited knowledge about the vibration of parts bridging these two crucial anatomical structures. How do the outer hair cells move in relation to the basilar membrane? How do forces produced by outer hair cells affect this motion pattern? How is the motion of the organ of Corti transformed into stereocilia deflection? These are crucial questions if we want to understand the operation of the hearing organ. The standard technique for making inner ear vibration measurements is laser interferometry. Although interferometers are usually exquisitely sensitive, only one or a few points can be measured at a time. Thus, it is difficult to quantify the motion of a 3-D object such as the organ of Corti. New experimental methods are therefore needed to answer the aforementioned questions.

Laser scanning confocal microscopy (LSM) enables visualization of the cellular structures within thick tissue samples and has been successfully used for imaging studies of the inner ear.[3,4] However, one of the limitations of confocal microscopy is the relatively long acquisition time, owing to the inherent point-to-point scanning mode of the microscope. Consequently, it takes around 1 s to acquire a single standard size confocal image (512×512 pixels). This is too slow to capture the motion of the organ of Corti, as it occurs during

sound stimulation. In this work, we describe a modified confocal acquisition system that permits imaging during sound stimulation. Using this system, high-resolution images can be acquired from structures vibrating at several hundred hertz. The accuracy of the system was confirmed by calibration measurements using a computer-controlled, piezoelectrically driven translator. The currently described system is the latest setup of an acquisition system that has been under constant development since 2001.[5,6] The new developments described in this work make image acquisition substantially faster and increase the resolution of the motion by half an order of magnitude. The noise level is also reduced, thereby improving both the accuracy of the motion estimates and the sensitivity of the technique. Therefore we are able to characterize the motion of the organ at significantly lower sound pressure levels than before.[5,6]

2 Methods

2.1 Preparation and Sound Stimulation

Adult guinea pigs of both genders weighing between 250 to 500 g were used in accordance with the rules approved by the local ethics committee (permit N311/03). Detailed descriptions of the preparation have been published previously.[7,8] Only a brief description is given here. The excised temporal bone was sealed to a holder and immersed in tissue culture medium [minimum essential medium (MEM)
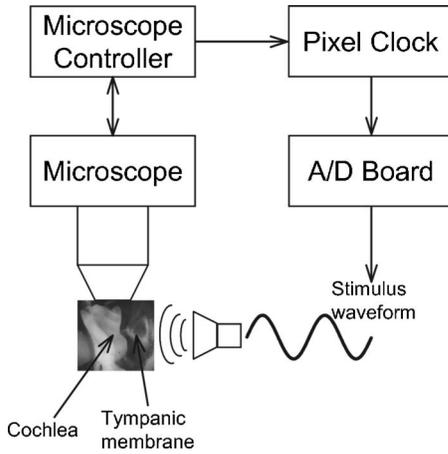
Fig. 1 Schematic overview of the acquisition system. The basic principle is to use the pixel clock of the LSM as a time base for the generation of the sound stimulus. This ensures synchronization between image acquisition and stimulus generation. The phase of the sound stimulus for each pixel is then saved for processing.

with Earle's salts, Gibco, Palsley, UK]. The external auditory canal was attached in the holder to face a loudspeaker, thus enabling sound stimulation. To gain access for imaging as well as to establish a perfusion system, a portion of the bony shell covering the apical part of the cochlea was carefully removed. A perilymph-like oxygenated solution (MEM) was used to perfuse scala media via an opening in the basal turn of the cochlea. The fluorescent dye calcein acetoxymethyl ester (Molecular Probes, Leiden, the Netherlands, 4-μg/ml concentration) was also delivered by perfusion system to stain Reissner's membrane and the supporting cells of the organ of Corti. The medium was then exchanged and 12.5 μg/ml of the dye RH 795 was added to the perfusion system. This styryl, potentiometric dye labels preferentially cell membranes of the sensory cells and nerve fibers. The overall condition of the preparation was routinely assessed using a low magnification lens. In the case of structural damage, swollen sensory cells, or other abnormalities, the preparation was discarded. For imaging we used a Carl Zeiss LSM510 laser scanning confocal microscope equipped with a 40×, NA 0.8 water immersion lens. Pixel sizes were in the range 50 to 100 nm. Due to anatomical constraints, a perfect profile view of the organ of Corti cannot be acquired. However, if the preparation is rotated with respect to the optical axis of the microscope, images that are quite close to true cross sections can be obtained. In this study, sound stimulation was applied in a range of frequencies between 100 and 600 Hz. The frequencies, at which the imaged section responded most, ranged between 150 and 225 Hz. Stimulus levels were between 88 and 110 dB sound pressure level (SPL). Due to the immersion of the preparation, the middle ear is fluid filled, which, along with the opening of the apical turn, reduces the effective stimulus level by 20 to 30 dB.[9] The data presented in this text were not corrected for this attenuation.

2.2 Image Sampling and Noise-Free Phase Extraction

A schematic overview of the acquisition system is shown in Fig. 1. Note that the LSM controller outputs a timing signal

for each pixel acquired. This pixel clock is used as the time base for the stimulus generation by an A/D board (PCI-6115, National Instruments, Austin, Texas). Thus, when a new pixel is acquired by the LSM, a new sample of the stimulus is written to the output of the A/D board. In the previous version,[6] the sound was generated by a function generator and sampled. The phase of the sound was estimated and pixels sorted to have approximately the same temporal relation to the stimulus. With this old method, noise in the sound channel influenced the recorded phase, and therefore the reconstructed images did not have an absolutely fixed temporal relation to the stimulus. These problems yielded artifacts, which limited the resolution of the technique. In the present setup, the discrete-time stimulus is generated by software (LabView 7.1, National Instruments, Austin, Texas), and the phase argument of each sound sample (one sample corresponds to one pixel) is saved and used for Fourier series extraction. This new image sampling technique and the noise-free phase extraction yield less artifacts in the reconstructed images due to low-pass filtering of the time series and an absolutely fixed temporal relation between images and stimulus.

2.3 Fourier Series Extraction

During experiments, a fixed number of images (N_1) are acquired without an interframe delay. At the same time, the phase array of the stimulus is saved with dimensions equal to the image stack. Because the pixel value (I_m) and the phase (P_m) of each pixel are known, a Fourier series may be used to reconstruct an image sequence where all pixels in a given frame have an identical temporal relation to the sound stimulus. To illustrate this, imagine a sample of a tiny bright structure surrounded by darker ones, which results in one bright pixel in the image. If the sample moves periodically in the image plane, the intensity will drift out of the bright pixel into a darker one and back again, according to the motion direction. Plotting the intensity of the bright pixel as a function of the motion phase will yield a periodic behavior, which can be analyzed by a Fourier series approach.

The Fourier series coefficients are computed for each pixel location along the time dimension, which maps to the phase domain.

$$a_0 = \frac{2}{N_1}\sum_{m=1}^{N_1} I_m; \quad a_n = \frac{2}{N_1}\sum_{m=1}^{N_1} \cos[nP_m]I_m;$$

$$b_n = \frac{2}{N_1}\sum_{m=1}^{N_1} \sin[nP_m]I_m. \tag{1}$$

We use only a few coefficients for reconstruction, namely the constant and the sine and cosine ones up to harmonics of order α. A reconstruction (I) from the coefficients of each pixel location is performed at the desired phase values (p).

$$I[p] = \frac{1}{2}a_0 + \sum_{n=1}^{\alpha} a_n \cos[np] + \sum_{n=1}^{\alpha} b_n \sin[np]. \tag{2}$$

The number of reconstructed images does not depend on the number of images acquired, but α does! To ensure stability, the parameter α should not exceed half the number of frames in the acquisition. This also means that the reconstructed im-

ages are low-pass filtered, which reduces noise significantly.

2.4 Optical Flow Calculations

A wavelet-based differential optical flow algorithm was used to quantify the motion seen in the image sequences. A detailed description of this algorithm and a thorough performance characterization can be found in Ref. 10. In general, direct comparisons show that the performance of this optical flow algorithm is similar to the standard Lucas-Kanade method[11] when noise-free image sequences are used. Under more realistic conditions, where noise is added to the image sequence, the wavelet-based algorithm has a higher density of correct motion estimates.

2.5 Calibration Experiments

To test the acquisition system described, rigid autofluorescent samples were attached to a vibrating piezoelectric translator (P841.10, Physik Instrumente, Karlsruhe, Germany). The motion amplitude of the piezoelectric translator was measured by a built-in strain gauge sensor, which was calibrated by an interferometer. Due to mechanical limitations of the microscope setup, the translator had to be attached to one end of a 10-cm-long rod, which was mounted at the opposite end. The sample itself was glued to the head of a 2-cm-long screw, and the screw was mounted in the piezo head. The calibration experiments are influenced by the mechanical mount, which could not be used during the interferometric measurements. This may influence the magnitude and frequency responses, especially in the direction perpendicular to the one of the applied motion. By aligning the piezoelectric translator to one axis of the image, it is possible to separate the parallel and the perpendicular component of the motion relative to the translator elongation axis.

The samples were imaged with the same objective lens used during real experiments. However, it is not possible to use water immersion during these calibration experiments; the effective numerical aperture of the lens is therefore decreased and the image quality worse than that found during real experiments. In this regard, the calibration experiments described next represent worst-case estimates.

3 Results

The Fourier series extraction is implemented as custom Matlab scripts. The input to the algorithm is the image sequence generated by the confocal microscope, and the matching array of stimulus phases. To cover a phase of 2π for each pixel, with equally spaced points, sound stimulus frequencies that introduce a phase offset of $2\pi/N_1$ between each image were used (N_1 is the number of images in the stack). A reconstructed image stack is obtained by Fourier series extraction. In the ear, the motion is periodic when pure-tone stimulation is used. It proved to be sufficient to use only a few harmonics for the reconstruction. This results in improved images, because the reconstructed sequence is low-pass filtered along the time dimension. A smoothing of the images is performed by convolution with a Gaussian kernel to remove localized reconstruction artifacts.
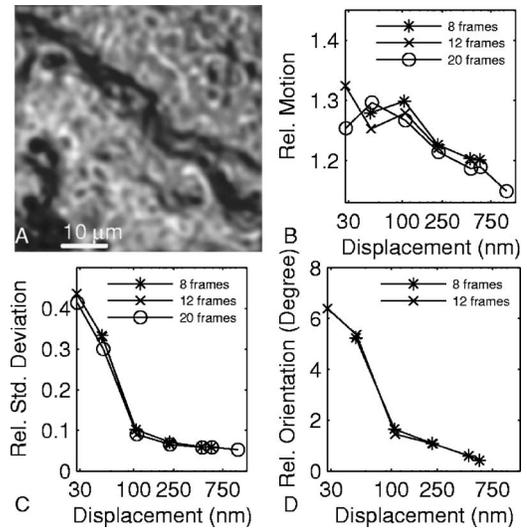
Fig. 2 (A) Micrograph showing an autofluorescent piece of plastic used for initial calibration (pixel size 112 nm). (B) Motion measured from the reconstructed image stack by the optical flow algorithm is compared to the real, physical motion of the piezoelectric translator. Physical motion was measured with a built-in strain gauge sensor. (C) Relative standard deviation of the estimated motion distribution as a function of the applied motion. (D) Angular errors as a function of the applied motion.

3.1 Artificial Target Tests

Figure 2(A) shows an image of a test sample, consisting of a piece of autofluorescent plastic attached to the piezo stack. Note that the scale of the structural details of the target matches approximately the one typically found in biological samples. In the first set of experiments, the piezoelectric translator was vibrating at 200 Hz. Figure 2(B) shows the relative motion amplitudes parallel to the piezo-elongation direction. A systematic overestimation of motion amplitudes occurs when displacements in the order of the pixel size (112 nm) or larger are to be measured. Compared to the physical motion, the measured motion is about 25% too large for motion amplitudes close to the pixel size and drops to 14% for motion amplitudes of roughly eight times the pixel size. This behavior is relatively independent of the number of images in the acquisition. This is not so surprising, since the signal-to-noise ratio in the images is very good (pixel time 6.4 μs). We also tried to use just four frames in the acquisition, which failed in most of the cases, because the noise influence increases dramatically when fewer frames are acquired. The dependency of the errors on the number of acquired frames becomes more serious at a faster scan speed (pixel time 2.5 μs), because the signal-to-noise ratio is much worse. Therefore scanning faster to reduce the laser exposure time of the biological sample is yielding larger errors. To get the same errors with faster scan speeds, one needs to scan more frames. Hence the laser exposure time is almost the same as with slower scan speeds.

The accuracy depends to some extent on the number of harmonics included in the reconstruction. Although similar values are obtained for small displacements, results are improved at large displacements if two harmonics are used in the reconstruction. Adding more harmonics than two did not ap-
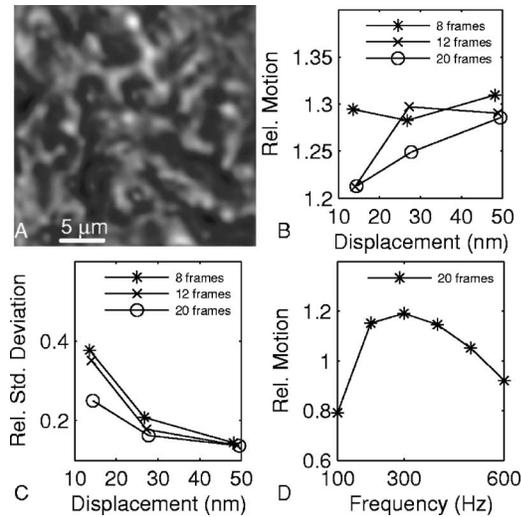
Fig. 3 (A) Calibration sample at higher zoom than Fig. 2 (pixel size 56 nm). (B) The mean motion estimate relative to the physical applied motion. (C) Standard deviation of the estimated motion distribution relative to the applied motion. (D) Relative motion estimated as a function of the applied frequency.

preciably alter the accuracy (data not shown). The width of the motion estimate distribution for one acquisition is described by the standard deviation relative to the applied motion [see Fig. 2(C)]. In essence, this number describes the variability inherent to one motion trajectory. It is smaller than 10% when motions larger than the pixel size are applied. The situation changes when motions smaller than the pixel size are applied. Figure 2(B) shows an overestimation of this motion of roughly 25 to 30%; note that this part of the graph is very noisy. The reason for this can be seen in Fig. 2(C), where the relative standard deviation increases drastically (up to 40%) for motions smaller than the pixel size.

The motion estimate is limited by the noise in the acquired images, giving rise to a virtual motion. To access these limitations, we acquired a time series of images when the sample was not moving. The estimated motion amplitudes of these acquisitions was always smaller than 10 nm, with standard deviation of less than 6 nm (pixel size 112 nm) and 5 nm with standard deviation of less than 3 nm (pixel size 56 nm). This is comparable to the one measured in a real temporal bone experiment, where the estimation of the virtual motion is below 10 nm.

By measuring the angle of the major axis of the motion trajectories, it is possible to assess the angular error of the motion estimate. For motions larger than the pixel size, the angular error is below 2 deg. It increases for motions smaller than the pixel size, but remains smaller than 10 deg. The angular error is not substantially affected by the amount of acquired frames. The error of the motion estimation depends to some extent on the structure of the sample, especially on the characteristic length scale over which contrast occurs. If this length scale is smaller than the pixel size, the pixel size is the limiting factor.

Figure 3(A) shows the same target at higher zoom (pixel size 56 nm). The translator was vibrating at 200 Hz. We applied motion on the order of the pixel size and smaller. Figure

3(B) shows the relative motion amplitudes. The motion is overestimated by 20 to 30%. The relative standard deviation of the estimation, shown in Fig. 3(C), is again very small for motions on the order of the pixel size, and increases dramatically as the applied motion becomes smaller. Figure 3(C) also shows that the amount of acquired frames has an influence on the width of the estimate distribution. For displacements larger than half the pixel size, there is almost no difference between 8, 12, and 20 frames scanned. But at motions much smaller than half the pixel size, the width of the estimate distribution for 20 frames scanned is only 2/3 of the one with 8 or 12 frames scanned. This means that at these small displacements, the noise in the image acquisition limits the information content of the motion in the images. Therefore it is necessary to scan more frames to reduce the uncertainty in the motion estimation. The frequency dependence of the motion estimation [Fig. 3(D)] shows a resonance-like behavior, with a broad maximum around 300 Hz and a dropoff to each side. This behavior has been observed with all test samples and also at different scan speeds (data not shown). Therefore it may reflect the mechanical properties of the sample mount (see calibration experiments in Sec. 2.5).

Thus it is concluded that this acquisition system performs very well. Optimum performance is achieved for motions about the size of a pixel, and it would therefore make sense to adjust the pixel size of the microscope according to the expected size of the motion. Two harmonics should be included in the calculations. The number of acquired frames has a relatively small influence. Note that these minor estimation errors are not caused by the optical flow algorithm, but rather by the acquisition system itself. When using image sequences with computer-generated translations, motion was accurately measured, even for very small displacements and even if the image sequence was substantially degraded by noise (data not shown).

3.2 Cochlear Motion Data

We have also obtained vibration data in several temporal bone preparations. An example is shown in Fig. 4(A). The image of these two outer hair cells was obtained *in situ* during simultaneous sound stimulation at the characteristic frequency 225 Hz, at a stimulus level of 97-dB SPL (the effective stimulus level is reduced by immersion of the preparation and by the opening of the apical turn). The asterisk in the image marks a point on the reticular lamina, the motion of which is shown in Figs. 4(B)-(D). The three motion trajectories display the motion for three different frequencies (at the characteristic frequency, below, and above it). Changing the stimulus frequency altered the overall displacement amplitude, but the shape and orientation of the trajectories remained similar across frequencies. The main axis of the trajectories approximately coincided with the long axis of the outer hair cells. This vibration component dominated the response; components in other directions were generally quite small.

In Fig. 5, a tuning curve obtained from this preparation is plotted. For comparison, we also include a previously published tuning curve, which was measured with a laser interferometer.[7] The overall shape of the curves is similar, as are the response amplitudes. In both cases, a notch is seen on the high-frequency side of the curve, although the notch ap-
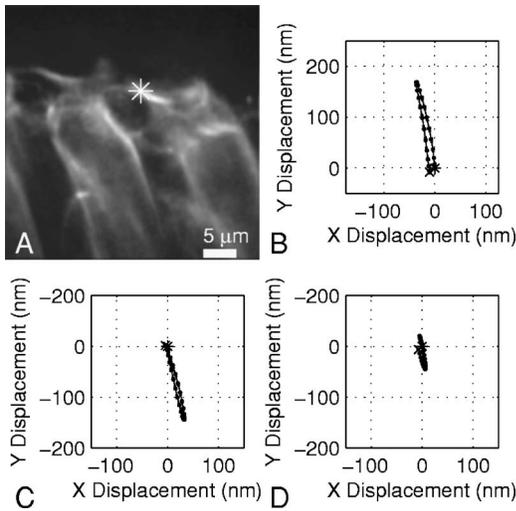
Fig. 4 (A) Two outer hair cells from the second and third row. The image was obtained during sound stimulation at 225 Hz. (B), (C), and (D) Motion trajectories at the characteristic frequency, at 150 and at 400 Hz, respectively.

pears more pronounced in the recently measured dataset. Although both curves were obtained in isolated preparations, they are quite similar to those seen in living, anesthetized animals.[12]

4 Discussion

We describe a novel acquisition system for obtaining confocal image sequences of rapidly moving objects. The strength of the technique is that it is possible to acquire high-quality image sequences showing the relative motion of different structures in the organ of Corti. The previously concealed internal mechanics of the hearing organ thereby become experimentally accessible. This is an important step forward. The technique is not limited to motion measurements. Using suffi-

ciently fast indicators, it would be eminently feasible to measure rapid alterations in ion concentrations or even membrane potentials. Although presently available membrane potential indicators suffer from lack of sensitivity, new indicators with improved response properties are becoming available.[13]
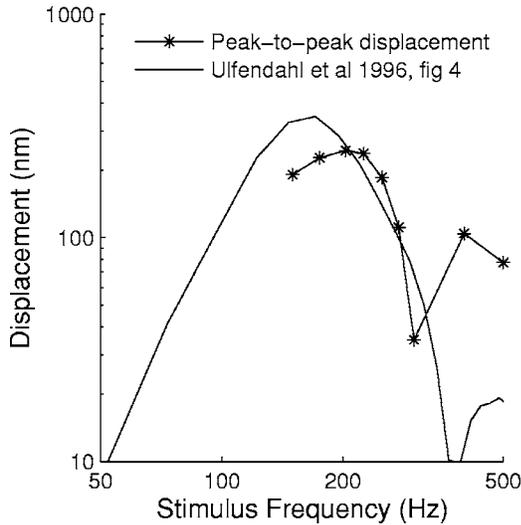
The results of the calibration measurements presented here have practical implications for motion measurements. For optimal results, the pixel size should be tailored to the motion to be measured. If this basic requirement is fulfilled, the accuracy of the technique is very high. Fortunately, the pixel size of most modern confocal microscopes can be varied over a large range (in our case from approximately 20 to 600 nm using a 40× water immersion lens). However, even if the motion would be smaller than the pixel size, a 20% amplitude estimation error is still not disastrous. Frequently used biophysical techniques have systematic measurement errors that fall within this range. For example, in patch-clamp recordings, it is rarely possible to compensate more than 80 to 90% of the series resistance, leading to a 10 to 20% estimation error when measuring membrane currents.

The Fourier series approach used here has several appealing properties. As compared to our previous acquisition system, which was based on searching the acquired image stack for pixels with similar phase, the Fourier technique leads to a substantially decreased acquisition time. The total acquisition time for one time series is around 40 s, a value that permits data to be acquired at multiple stimulus frequencies within a reasonable time frame. When stimulating the ear with pure tones, the resulting motion is nearly sinusoidal and it is therefore sufficient to use only a few harmonics in the reconstruction. This leads to much decreased noise levels and substantially improved image quality. Any nonperiodic motion of the sample (e.g., drift) will result in artifacts in the reconstructed image series. Hence the condition of the sample and the mechanical stable mount are very important.

It is important to realize that the characteristics of the object under study also influences measurement results. This is easily appreciated for the extreme case when the "image" consists of only a homogeneous field with identical pixel values. In such a case, no optical flow algorithm can detect motion, even if motion is present. For more realistic objects, the size of the smallest features of the image is likely to influence the minimum motion amplitude that can be reliably quantified. Thus, if we want to extend the measurement range to amplitudes substantially lower than the minimum values used here, it may be necessary to circumvent the limits imposed by diffraction. Such microscope systems are being developed,[14] and may prove to be useful in this context.

Our system is not the first to acquire images of the moving hearing organ. Using artificial stimulation and a sectioned preparation, Hu, Evans, and Dallos[15] acquired images that showed some aspects of hearing organ motion, albeit at a very low stimulus frequency. By combining video microscopy with stroboscopic illumination, the frequency range of such a measurement can be substantially enhanced.[16,17] The primary advantage of the present technique is that the inherent optical sectioning of the confocal microscope makes it possible to perform experiments on intact preparations while using the natural stimulus—sound delivered to the tympanic membrane. No physical sectioning needs to be done. Compared to video

Fig. 5 A tuning curve measured from the same preparation as in Fig. 4 is plotted together with a displacement tuning curve measured at 93 dB SPL (Ref. 7).

microscopy, confocal microscopy results in a large improvement of image quality and resolution. This improvement makes it possible to accurately measure the small displacements typically found in the ear.

Acknowledgments

References

1. L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," *Physiol. Rev.* **81**, 1305–1352 (2001).
2. M. Ulfendahl, "Mechanical responses of the mammalian cochlea," *Prog. Neurobiol.* **53**, 331–380 (1997).
3. Å. Flock, B. Flock, A. Fridberger, E. Scarfone, and M. Ulfendahl, "Supporting cells contribute to control of hearing sensitivity," *J. Neurosci.* **19**, 4498–4507 (1999).
4. A. Fridberger, J. Boutet de Monvel, and M. Ulfendahl, "Internal shearing within the hearing organ evoked by basilar membrane motion," *J. Neurosci.* **22**, 9850–9857 (2002).
5. A. Fridberger and J. Boutet de Monvel, "Sound-induced differential motion within the hearing organ," *Nat. Neurosci.* **6**, 446–448 (2003).
6. A. Fridberger, I. Tomo, M. Ulfendahl, and J. Boutet de Monvel, "Imaging hair cell transduction at the speed of sound: dynamic behavior of mammalian stereocilia," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1918–1923 (2006).
7. M. Ulfendahl, S. M. Khanna, A. Fridberger, Å. Flock, B. Flock, and W. Jäger, "Mechanical response characteristics of the hearing organ in the low-frequency regions of the cochlea," *J. Neurosci.* **76**, 3850–3862 (1996).
8. M. Ulfendahl, E. Scarfone, Å. Flock, S. Le Calvez, and P. Conradi, "Perilymphatic fluid compartments and intercellular spaces of the inner ear and the organ of Corti," *Neuroimage* **12**, 307–313 (2000).
9. R. Franke, A. Dancer, S. M. Khanna, and M. Ulfendahl, "Intracochlear and extracochlear sound pressure measurements in the temporal bone preparation of the guinea-pig," *Acustica* **76**, 173–182 (1992).
10. A. Fridberger, J. Widengren, and J. Boutet de Monvel, "Measuring hearing organ vibration patterns with confocal microscopy and optical flow," *Biophys. J.* **86**, 535–543 (2004).
11. B. Lucas and T. Kanade, "An iterative image registration technique with an application in stereo vision," in *Proc. 7th Intl. Joint Conf. Artif. Intell.*, pp. 674–679, Morgan Kaufman, New York (1981).
12. C. Zinn, H. Maier, H. Zenner, and A. W. Gummer, "Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the in-vivo guinea-pig cochlea," *Hear. Res.* **142**, 159–183 (2000).
13. A. Grinvald and R. Hildesheim, "VSDI: a new era in functional imaging of cortical dynamics," *Nat. Rev. Neurosci.* **5**, 874–885 (2004).
14. G. Donnert, J. Keller, R. Medda, M. A. Andrei, S. O. Rizzoli, R. Lührmann, R. Jahn, C. Eggeling, and S. W. Hell, "Macromolecular-scale resolution in biological fluorescence microscopy," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11440–11445 (2006).
15. X. Hu, B. N. Evans, and P. Dallos, "Direct visualization of organ of Corti kinematics in a hemicochlea," *J. Neurophysiol.* **82**, 2798–2807 (1999).
16. H. Cai, C. P. Richter, and R. S. Chadwick, "Motion analysis in the hemicochlea," *Biophys. J.* **85**, 1929–1937 (2003).
17. A. J. Aranyosi and D. M. Freeman, "Sound-induced motions of individual cochlear hair bundles," *Biophys. J.* **87**, 3536–3546 (2004).