Research Article | Vol. 10, No. 7 | 1 Jul 2019 | BIOMEDICAL OPTICS EXPRESS 3317
Biomedical Optics EXPRESS

In vivo tomographic visualization of intracochlear vibration using a supercontinuum multifrequency-swept optical coherence microscope

Samuel Choi,[1,2,*] Fumiaki Nin,[2,3,4] Takeru Ota,[2,3] Kouhei Sato,[1] Shogo Muramatsu,[1,2] and Hiroshi Hibino[2,3,4]

[1]*Niigata University, Department of Electrical and Electronics Engineering, 8050 Ikarashi-2, Niigata 950-2181, Japan*
[2]*AMED-CREST, AMED, Japan*
[3]*Niigata University, School of Medicine, Department of Molecular Physiology, 757 Ichibancho, Asahimachi, Niigata 951-8510, Japan*
[4]*Niigata University, Center for Transdisciplinary Research, 8050 Ikarashi-2, Niigata 950-2181, Japan*
**schoi@eng.niigata-u.ac.jp*

Abstract: This study combined a previously developed optical system with two additional key elements: a supercontinuum light source characterized by high output power and an analytical technique that effectively extracts interference signals required for improving the detection limit of vibration amplitude. Our system visualized 3D tomographic images and nanometer scale vibrations in the cochlear sensory epithelium of a live guinea pig. The transverse- and axial-depth resolution was 3.6 and 2.7 μm, respectively. After exposure to acoustic stimuli of 21–25 kHz at a sound pressure level of 70–85 dB, spatial amplitude and phase distributions were quantified on a targeted surface, whose area was 522 × 522 μm^2.

1. Introduction

The cochlea of the inner ear transduces sound energy, which is a form of mechanical energy, into electrical signals, which are essential for a neurotransmitter release. This process is triggered by nanoscale vibrations induced in the cochlear sensory epithelium, which contains a layer of sensory hair cells and the basilar membrane (BM), i.e., the underlying extracellular matrix. The vibrations in the BM are controlled by the active motion of hair cells [1–4]. Although this arrangement is thought to critically contribute to the high sensitivity and sharp tuning of hearing, the *in vivo* behavior of each layer as well as the correlation of the dynamics among multiple layers remain unclear.

To address these issues, various optical measurement systems have been developed. Laser Doppler vibrometer (LDV) techniques can detect epithelial vibrations in the picometer range [5–10]. For example, in the cochlea, traveling waves elicited by sound stimulation propagate from the base to apex on the BM thus forming a spatial amplitude distribution. To determine the vibration distribution caused by the motion of a traveling wave, asynchronous measurement has been conducted by means of beam scans in an LDV system [11]. Nonetheless, such methods are inherently unable to simultaneously determine the vibration distribution on a targeted surface and extract tomographic information from a sample. Therefore, they are not applicable to the analysis of each layer in the sensory epithelium. In this regard, stroboscopic detection schemes may be useful for detecting the vibrations [12–15]; however, systems based on such methods are usually too complicated to be integrated into a microscope for *in vivo* measurement.

To analyze the motions inside the tissue and overcome the above-mentioned shortcomings, we recently proposed multifrequency-swept optical coherence microscopic

#365730 https://doi.org/10.1364/BOE.10.003317
Journal © 2019 Received 23 Apr 2019; revised 5 Jun 2019; accepted 7 Jun 2019; published 13 Jun 2019

Biomedical Optics EXPRESS

vibrometry (MS-OCMV), which is a combination of wide-field heterodyne interferometric vibrometry (WHIV) [16] and multifrequency-swept interferometry [17,18]. This approach can successfully record not only 3D volumetric tomography but also wide-field vibrations on a surface inside a biological tissue [18]. Nevertheless, it cannot accurately measure nanoscale vibrations in the sensory epithelium for the following two reasons. First, the power of the installed superluminescent diode (SLD) is too low to achieve adequate signal reflection from the tissue (the power of the light applied to the sample: 0.9 mW). Second, the method for analysis of the interference signals is unsuitable because one of the frequency components required for quantifying vibration amplitude overlapped with direct current (DC) component at frequency of 0 Hz.

To overcome these problems, we employ a supercontinuum (SC) light source, which provides more powerful irradiation than an SLD does in the present study. Furthermore, to effectively extract the interference signals that represent vibrations of an object in the nanometer range, we modulate the motion of the reference mirror in the WHIV system.

As for newly devised technologies, in recent years, optical coherence tomography (OCT) was widely used and can be regarded as a competitor of our optical system. Among such methods, spectral-domain OCT and swept-source OCT combined with Doppler techniques have been applied to the measurement of inner-ear vibration [19–21]. The scanning in these two technologies is oriented differently. Doppler types of OCT can immediately determine a cross-sectional distribution of vibration in the depth direction along an "a-scan" line. On the other hand, our system can reduce the lag for the lateral beam scan by performing the *en face* measurement using a CMOS camera and can immediately capture the lateral vibration on a surface. Nonetheless, it requires multifrequency sweeping for the cross-sectional scan in the axial depth direction, as is the case for time domain OCT. Therefore, Doppler types of OCT are specialized for cross-sectional imaging, whereas our technique is useful for *en face* imaging of a laterally spread vibration distribution on an internal surface such as the BM in the cochlear sensory epithelium.

When we limit the application to *en face* vibration measurement in a sensory epithelium with low reflectance (0.02%–0.06%) [22], Doppler types of OCT require repeated A-scans to average the data and reduce the noise floor [23]. In addition, the M-scan mode is often used for monitoring temporal changes in the vibrations. Thus, these two configurations result in an asynchronous B-scan with a lag, which makes simultaneous measurement of wide-range motions difficult in a live biological tissue.

To overcome this difficulty, we attempted *in vivo en face* vibration analysis of a sensory epithelium by MS-OCMV. The improvements enable us to quantitatively visualize the wide-field vibrations on the surface of a desired depth position in the sensory epithelium of a live animal. Moreover, the motion and a three-dimensional (3D) volumetric image of the tissue can be captured without averaging the data.

2. Methods

2.1. Instrumentation

The setup of the improved MS-OCMV is shown in Fig. 1(A). This system consists of a multifrequency generation unit, microscopic interferometer, and detection unit.

The multifrequency generation unit consists of an SC light source (SuperK EXR-4; NKT Photonics, Denmark), a Fabry–Pérot filter (FPF), and an optical bandpass filter (OBF). For practical use, we extracted a wavelength band ranging from 600 to 980 nm by means of the OBF. The light beam characterized by discrete multifrequency components was acquired by transmitting the collimated and filtered SC light through the FPF. Figure 1(B) depicts a comparison between the spectrum of the multifrequency light from the SC and that of an SLD (T-850-HPI, Superlum, Ireland), which we used in the original system [17]. These data were obtained through the FPF. The bandwidth of the SC was similar to that of the SLD (~200 nm); nevertheless, maximum irradiation of the sample surface in the case of the former light

Biomedical Optics EXPRESS

source was improved to 37 mW, whereas maximum irradiation in the case of the latter light source was 0.9 mW.

The FPF consists of two partially reflecting mirrors with reflectivity 0.8, each of which is attached to a different piezoelectric actuator (PA) (MOB-A or MD-140L; MESS-TEC, Japan). Cavity length, d, determines the interval frequency (i.e., the free spectral range), $\Delta\nu$, via the relation $\Delta\nu = c/(2d)$, where c is the light speed in air. The linewidth of the longitudinal mode is determined by the finesse value of the plates, which was ~14 in our implementation. In our experiments, d was set to ~20 mm; $\Delta\nu$ was estimated to be 7.5 GHz. Thus, considering the finesse, multifrequency components were produced with estimated linewidth of 535.7 MHz (1.2 pm in terms of wavelength). The cavity length was varied by means of the two aforementioned PAs to perform an axial depth scan by multifrequency-swept interferometry [18]. The PAs were driven by a ramp signal from a function generator (WW5064; Tabor Electronics, Israel). The signal was magnified by a high-power amplifier (TZ-0.5P; Matsusada Precision, Japan), and the total stroke of the PAs was 980 μm.

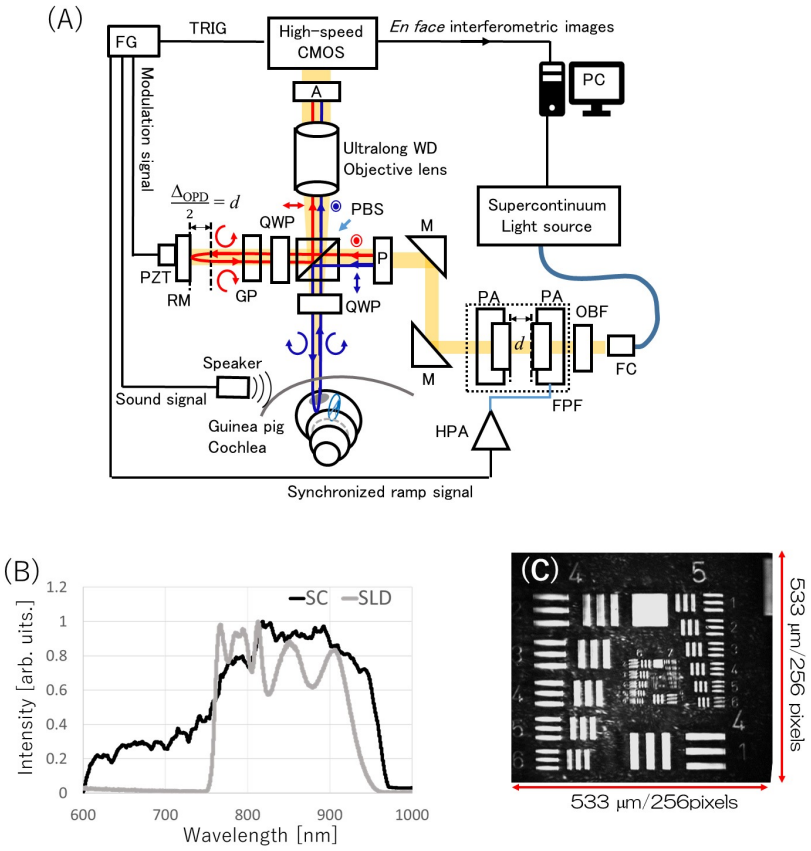

Fig. 1. Instrumentation of the improved MS-OCMV system. (A) The schematic of the setup. A: analyzer; FC: fiber collimator; FG: function generator; FPF: Fabry–Pérot filter; GP: glass plate for dispersion compensation; HPA: high-power amplifier; M: mirror; OBF: optical bandpass filter; P: polarizer; PA: piezoelectric actuator; PBS: polarized beam splitter; QWP: quarter wave plate; PZT: piezoelectric transducer; RM: reference mirror; TRIG: trigger. (B) Spectra of SC and SLD light sources. Note that the signal intensity in either case was normalized. (C) A micrograph of a test target. Our system distinguished 144 line pairs/mm (Group 7, Element 2).

The generated multifrequency light entered the microscopic interferometer that was subjected to full-field tomographic and vibration measurements. The incident light was split

Biomedical Optics EXPRESS

using a polarization beam splitter (PBS) to irradiate the sample and reference mirror surfaces. The polarizer was combined with the PBS to adjust the branching ratio between the two split beams. These beams were polarized orthogonally to each other. In each arm, a quarter wave plate (QWP) was inserted to avoid unexpected reflections from the optical elements. The beam that passed through the QWP was incident toward either the reference mirror or the sample. The reflected beam again passed through the QWP and was directed to the PBS. The polarization of the reflected beam was rotated by 90° as compared to the polarization of the incident beam. Consequently, the beam in the reference arm was polarized orthogonally to the beam in the sample arm; these two beams were recombined in the PBS and entered the objective lens. In this process, polarizations of stray light beams, which were scattered from other optical elements located between the QWPs and PBS or polarizer, were not rotated; therefore, they could not enter the objective lens through the PBS. Moreover, the system was equipped with a rotatable linear polarizer as an analyzer to attain optimal interference contrast by controlling the polarization extinction ratio between the two beams from reference and sample arms.

The reference mirror is a well-polished glass plate with a reflectance of approximately 4%. The reference path length is modulated by a piezoelectric transducer (PZT) attached to the mirror; this mechanism plays a key role in the operation of the improved WHIV (see the next subsection).

An *en face* interferometric image of the sample surface was captured by means of an inverted microscope having an objective lens characterized by an ultralong working distance of 205 mm (UWZ200; Union Optics, Japan). This profile provided us with sufficient space for laying an anesthetized guinea pig under the lens. Optical magnification varied from 0.7 to 9.8. The depth of focus and numerical aperture at maximum magnification were 62 μm and 0.093, respectively. This information is available in the manufacturer's instructions (URL: http://www.union.co.jp/en/union_uwz.php). Imaging resolution was estimated to be 3.6 μm via microscopic examination of a test target (USAF Resolving Power Test Target 1951) as shown in Fig. 1(C). In this figure, we confirmed that 1 pixel of the CMOS camera corresponds to approximately 2.1 μm at maximum magnification.

The axial depth scan can be operated without changing the optical path difference (OPD), although this process requires vertical motion of the microscopic interferometer in conventional time domain full-field OCT systems.

The principle of multifrequency-swept interferometry has been established in our earlier study [18]. As mentioned above, the interval of the spectral multifrequency components, $\Delta\nu$, is correlated with the cavity length, d [because $\Delta\nu = c/(2d)$]. The interference signal of the multifrequency light manifested repetitive fringe peaks (i.e., high-order interference) with a constant interval of $\Delta_{OPD} = c/\Delta\nu = 2d$. Thus, the OPD that yields the first-order interference peak is $2d$. During our measurement, the optical path length of the object arm was set to be identical to focal length, whereas the length of the reference arm was set to approximately 22 mm (i.e., half of Δ_{OPD}) longer than that of the other arm. Owing to this arrangement, the first-order interference peak overlapped with the focal plane. When d was varied by controlling the FPF, Δ_{OPD} changed, resulting in operation of the axial depth scanning. Note that for this mode, the motion of the reference mirror is unnecessary. The maximum scanning range of the first-order interference peak is equivalent to the maximum stroke of the FPF.

Depth imaging of 3D OCT and vibration analysis with the WHIV technique, which are described in detail in the next subsection, are separate procedures. Because d of the FPF corresponds to Δ_{OPD} of the interferometer, the position on the Z-axis where interference occurs changes according to the cavity length of the FPF. In this manner, we can determine a depth position of interest for the *en face* vibration measurement and obtain the vibration parameters on the X-Y plane at the depth position.

To rapidly capture *en face* interferometric images, a high-speed CMOS camera (FASTCAM Mini AX200; Photron, Japan; 1024 × 1024 pixels; pixel size, 20 × 20 μm; bit

Biomedical Optics EXPRESS

depth, 12; frame rate, 2000 fps) was employed as a detector. The sensitivity of the sensor is ISO 40,000 as per ISO standard. Full-well capacity is 16,000 e^-, and the sensor dynamic range is 54.8 dB. The captured images were transferred to a PC, which also controlled the CMOS camera, SC light source, and function generator. The trigger for initiating a recording was generated by the function generator. In our system, two types of acquisition were performed: volume scans (three spatial dimensions) and vibrometry (two spatial dimensions + one temporal dimension). In both cases, the acquisition speed of the 3D volume data was 2 Gvoxels/s. Transfer of one volume data set of 1.024 GB took approximately 1 min. The maximum data size was 1024 pixels on the X-axis, 1024 pixels on the Y-axis, and 8000 frames along the Z-axis for a typical volumetric data set and a similar approach for the vibrometry data sets. All the captured images were processed and analyzed in the MATLAB software.

2.2. Improvement of the WHIV technique

Our previous WHIV technique requires Fourier domain analysis using two different frequency components to quantify the amplitude and phase of the sample's vibration [16–18]. These two components are called a "zeroth-order signal" at a frequency of 0 Hz and a "first-order signal," which corresponds to a difference frequency (i.e., beat frequency) between the frequencies of sample and reference vibrations. In the original method, the DC component overlapped and added linearly to the zeroth-order signal because the frequency of the first-order signal was 0 Hz. Therefore, a possible disadvantage of the original method is that the DC component interferes with proper extraction of the zeroth-order component required for estimation of the vibration amplitude (see Subsection 4.2). This problem is prominent when the measurement targets are biological samples that yield weak interference signals, such as the cochlear sensory epithelium.

To overcome this drawback, we added low-frequency DC offset modulation to the modulation provided for the reference mirror. Figure 2 illustrates a comparison of the spectral signal obtained by the improved method with the signal obtained by the original technique. In the former case, the zeroth-order frequency component was clearly separated from the bias noise owing to the offset modulation.

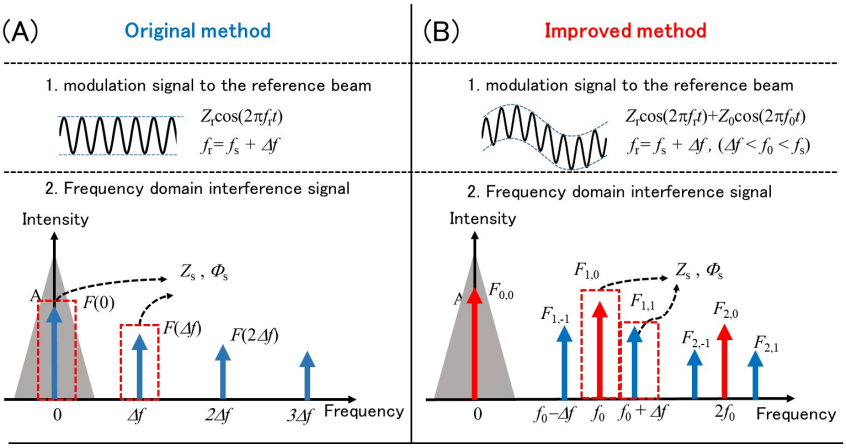

Fig. 2. Modification of the WHIV technique. In the original vibrometric method (A) [16], simple sinusoidal modulation illustrated in the upper panel was applied to the reference mirror. In this condition, the zeroth component $F(0)$ overlapped with the DC component, A (lower panel). By contrast, in the improved method (B), sinusoidal DC offset modulation was added to the reference modulation (upper panel). This approach completely separates the zeroth component from bias noise at a frequency of f_0 Hz, i.e., $F_{1,0}$ (lower panel). From frequency components $F_{1,0}$ and $F_{1,1}$, vibration amplitude Z_s and phase Φ_s can be properly estimated without disturbance by the DC component (see Subsection 4.2).

Biomedical Optics EXPRESS

The details of the improvement are as follows. Suppose that the sample is stimulated by a pure tone sound at a frequency of f_s and the reference mirror is sinusoidally vibrated at a slightly different frequency, $f_r = f_s + \Delta f$. Moreover, offset modulation with a low frequency, f_0, is added to the motion of the reference mirror. The effect of these three factors on the temporal interference signal is expressed as

$$I_{xy}(t) = A_{xy} + B_{xy}\cos[Z_s\cos(2\pi f_s t + F_s) + Z_r\cos(2\pi f_r t + F_r) + Z_0\cos(2\pi f_0 t + F_0) + a_{xy}], \quad (1)$$

where A_{xy} and B_{xy} denote DC component that do not contribute to the interference and interferometric amplitude, respectively. α_{xy} is the spatial interferometric phase distribution at depth position d', and Z_s, Z_r, and Z_0 are the spatial amplitude distributions of the vibrating sample, reference mirror, and offset modulation, respectively. In addition, Φ_s, Φ_r, and Φ_0 are the initial phase distributions of the acoustic stimulus, reference mirror vibration, and offset modulation, respectively. In general, the frame rate of a standard CMOS camera is much lower than that of conventional photodetectors whose sampling rate ranges from a few hundred kilohertz to several tens of gigahertz. This configuration averages and reduces relatively high-frequency components of the temporal interference signal [see Eq. (1)] but affects low-frequency components negligibly. Therefore, the components Δf and f_0 that are sufficiently lower than the frame rate of the camera are retained in the frequency domain. After application of the Jacobi–Anger expansion [24] and exclusion of the terms associated with higher frequencies, Eq. (1) can be rewritten as follows:

$$\begin{aligned} I_{xy}(t) = {} & A_{xy} + B_{xy}\cos\alpha_{xy}\sum_{m=1}^{M}\sum_{n=1}^{N}(-1)^{m+n}J_{2m}(Z_0)J_n(Z_s)J_n(Z_r)\cos\{2\pi(2mf_0 \pm n\Delta f)t + 2m\Phi_0 \pm n\Phi\} \\ & + B_{xy}\sin\alpha_{xy}\sum_{m=1}^{M}\sum_{n=1}^{N}(-1)^{m+n+1}J_{2m-1}(Z_0)J_n(Z_s)J_n(Z_r)\cos[2\pi\{(2m-1)f_0 \pm n\Delta f\}t + (2m-1)\Phi_0 \pm n\Phi] \end{aligned} \quad (2)$$

where m and n are positive integers that are indices associated with the harmonics f_0 and Δf, and J_k denotes the kth order Bessel function of the first kind. M and N are limits of harmonic orders, which are determined by the exposure time and frame rate of the camera. Φ denotes a relative phase, which is expressed as $\Phi = \Phi_r - \Phi_s$. The absolute phase value of Φ is relatively changed depending on the timing of triggering to start capturing. In our current system, however, the trigger and signals for modulations were not synchronized. Thus, absolute phase value Φ was changed randomly with each measurement. The absolute value of Φ can be changed arbitrarily by adjusting the start point of the recorded signal during this data processing. Thus, in this case, relative spatial phase differences become more important. Therefore, for convenience, we redefine Φ_s as a resulting phase $\Phi_s = \Phi$ including the reference phase Φ_r.

In the frequency domain, the signal represented by Eq. (2) is composed of a carrier frequency of f_0 and neighboring sidebands with a frequency spacing of Δf (Fig. 2(B)). After the Fourier transform presented in Eq. (2), complex amplitudes of high-order components defined as $F_{m,n}$ associated with harmonic frequency $mf_0 + n\Delta f$ can be denoted as

$$\begin{aligned} F_{2m-1,n} &= (-1)^{m+n+1}B\sin\alpha J_n(Z_s)J_n(Z_r)J_{2m-1}(Z_0)\exp i[n\Phi_s + (2m-1)\Phi_0], \\ F_{2m,n} &= (-1)^{m+n}B\cos\alpha J_n(Z_s)J_n(Z_r)J_{2m}(Z_0)\exp i[n\Phi_s + 2m\Phi_0]. \end{aligned} \quad (3)$$

For estimating Z_s and Φ_s, frequency components $F_{1,0}$ and $F_{1,1}$ are extracted from the observed frequency components. Here, an intensity ratio, r_{01}, is defined as $r_{01} = |F_{1,0}|/|F_{1,1}|$. It can also be described as $|J_0(Z_s)J_0(Z_r)|/|J_1(Z_s)J_1(Z_r)|$. In this context, we can derive an evaluation function, $\varepsilon(z) = \{r_{01} - |J_0(z)J_0(Z_r)|/|J_1(z)J_1(Z_r)|\}^2$. Amplitude distribution $Z_s(x, y)$ can be obtained with such z that this value minimizes $\varepsilon(z)$ at each x-y coordinate. We developed a MATLAB code that evaluated $\varepsilon(z)$ and solved for z by comparing the measured value of r_{01} with a precomputed data set of $|J_0(z)J_0(Z_r)|/|J_1(z)J_1(Z_r)|$ as the function of z varied from 0.000 rad to

Biomedical Optics EXPRESS

2.404 rad in increments of 0.001 rad. Therefore, the accuracy of this solving procedure was 0.001 rad. The identified value of z can be estimated to be Z_s within the constraint condition of $z \le 2.404$ rad. Preferred values of Z_r and Z_0 are approximately 1.5 and 2.0 rad, respectively. These parameter settings provide an ideal condition. Spatial phase distribution Φ_s is calculated as

$$\Phi_s = \tan^{-1}\left[\frac{\mathrm{Im}(-F_{1,1}/F_{1,0})}{\mathrm{Re}(-F_{1,1}/F_{1,0})}\right]. \tag{4}$$

Furthermore, interference phase α is obtained using $F_{1,0}$ and $F_{2,0}$:

$$\alpha = \tan^{-1}\left[\frac{|F_{2,0}|}{J_2(Z_0)}\Big/\frac{|F_{1,0}|}{J_1(Z_0)}\right]. \tag{5}$$

For this purpose, Z_r and Z_0 should be determined before measurement; the two parameters can be calibrated arbitrarily via the function generator and assumed to be constant. For the calibration, these parameters were measured in advance using a conventional LDV. This approach is similar to the sinusoidal phase modulation technique [25] except for one characteristic: the analysis described in this study involves the beat signals resulting from the three different modulations mentioned above.

Note that the modified WHIV technique can determine all the vibration parameters two-dimensionally without lateral scanning. On the other hand, this method has the following disadvantage. When the values of α are in the vicinity of integer multiples of π rad, the intensities of frequency components of $F_{1,1}$ and $F_{1,0}$ are hardly detectable due to the $\sin(\alpha)$ dependence for odd terms in Eq. (2) and (3). Therefore, in this so-called "unmeasurable area," the amplitude and phase values cannot be obtained accurately (see Subsection 3.2). Therefore, we filtered out the unmeasurable area in the process of determination of the parameters necessary for characterization of the sample's vibration.

2.3. Animal preparation

In vivo wide-field tomographic and vibration measurements were performed on the cochlear sensory epithelium of a guinea pig as a sample, as follows.

First, a guinea pig was deeply anesthetized with intraperitoneal injection of urethane (1.5 g/kg). The toe pinch, corneal reflexes, and respiratory rate were examined to evaluate the depth of anesthesia. When anesthesia was insufficient, urethane (0.3 g/kg) was additionally injected into the animals. After tracheotomy, which was conducted for the maintenance of spontaneous breathing, the animals were paralyzed by intravenous injection of vecuronium bromide (3 mg/kg) (Vecuronium for intravenous injection; Fuji Pharma, Japan) [26]. Subsequently, the animal was artificially ventilated with room air using a respirator (SN-408-7; Shinano Manufacturing, Japan) [27]. We stopped the ventilation during the measurements for 4 s to prevent the motion artifact.

A fenestra was surgically opened on the lateral site of the bulla in order to shine the SC light on the sensory epithelium in the basal turn through the transparent round window. Basically with this hole the measurements can be carried out. In our preparations, an additional hole was made on the anterior portion of the bulla. This arrangement allowed us to directly confirm the position of the beam spot during the recording and thereby significantly improved the efficiency of the experiments.

Biomedical Optics EXPRESS

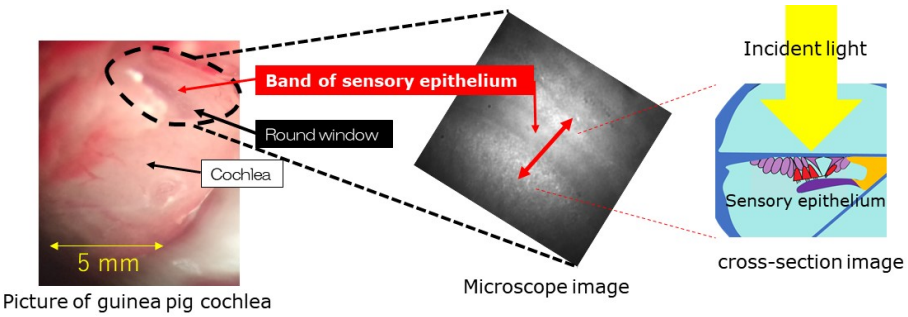

Fig. 3. Sample preparation. The cochlea of a live guinea pig was surgically exposed (left panel) and subjected to the experiment. Light was applied to the sensory epithelium (middle panel) through a transparent round window (the dotted circle in the left-hand panel). A schematic image of irradiation of the epithelium is provided in the right-hand panel.

Then, the animal's head was fixed on an acrylic plate (50 × 20 × 5 mm). The plate was tightly connected to an articulating base stage (SL20/M; Thorlabs, USA). In this process, the angle and position of the cochlea were manually controlled to enable the laser beam to irradiate the sensory epithelium through the round window membrane as perpendicularly as possible. This method does not require a hole to be artificially made in the cochlear bony wall for the irradiation and is hence noninvasive (Fig. 3).

The irradiation time during the measurements was several minutes in total. In spite of high irradiation power (37 mW), the tissue is likely to be damaged only minimally because the body fluid in the cochlea dissipates the heat generated by the irradiation. Alternatively, because the area irradiated by the light beam is relatively wide, the energy actually received by the tissue might be weaker than expected. To preserve an animal's condition as much as possible, all the procedures in the experiment were completed within 4 hours.

The experimental protocol was in compliance with federal guidelines for the care and handling of small rodents and was approved by the Institutional Animal Care and Use Committee of Niigata University [28].

3. Results

3.1. Validation of OCT imaging

We initially evaluated the axial resolution and sensitivity of the imaging by the improved MS-OCMV. Figure 4(A) depicts a first-order interference fringe detected by a pixel of the CMOS camera when a planar mirror was illuminated in the OCT system. The frames of interference images were captured as a time series and were used to reconstruct 3D volumetric data. As shown in Fig. 4(A), the fringe was obtained after removing the DC component from the raw data. This signal was processed with the Hilbert transform [29] and an adequate bandpass filter to reduce noise and extract the envelope. The signal process was carried out along the z-axis (axial depth direction) at each x and y coordinate. As depicted in Fig. 4(B), to properly apply the Hilbert transform, at least eight frames are necessary to construct one period of the fringe. Finally, a cross-sectional distribution in the depth direction was obtained as an "A-line" by squaring the extracted envelope as presented in Fig. 4(C). This process was carried out at all x-y coordinates simultaneously via the 3D fast Fourier transform (FFT) in the Matlab software. The axial resolution determined via full width at half maximum (FWHM) of the A-line was approximately 2.7 μm.

Biomedical Optics EXPRESS

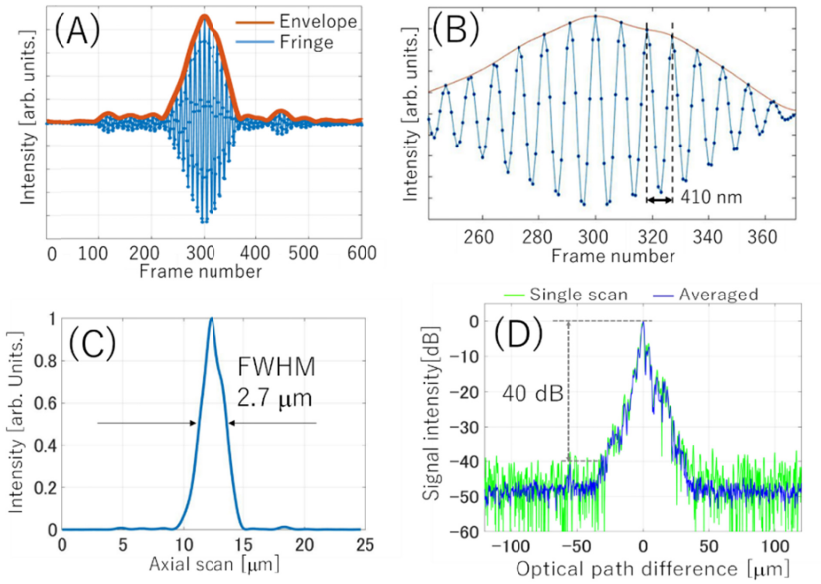

Fig. 4. The first-order interference peak obtained by the axial depth scan. (A) Interference fringe with its envelope after reduction of the DC component from the raw signal. (B) An enlarged plot near the interference fringe peak. (C) A typical A-line obtained from the envelope of the interference fringe. FWHM: full width at half maximum. (D) The logarithmic scale of (C). The green and blue plots in (D) denote the intensity profiles obtained by a single scan and by averaging the data from the neighboring 100 pixels, respectively.

In Fig. 4(D), the interference signal is shown on the logarithmic scale, indicating that sensitivity was 40 dB in terms of the signal-to-noise ratio (SNR). This value is lower than that of standard OCT systems, because in our system, full-well capacity is low and the resolution of the analogue-to-digital converter is poor (12 bits). Nevertheless, when the interference signals collected from the neighboring area of 10 × 10 pixels were averaged, the noise floor decreased by approximately 10 dB (Fig. 4(D)). Therefore, similar procedures, such as camera binning, may be effective in improving the sensitivity of our OCT system. Note that ripples remain visible around the main peak (OPD range: −25 to 25 μm).

3.2. A pilot experiment with the improved WHIV technique

Next, we performed a proof-of-concept experiment using the WHIV technique that was improved as described in Subsection 2.2. The comparison between the original and the improved WHIV is discussed in Subsection 4.2. The target sample was a thick planar mirror that was vibrated at a frequency (f_s) of 26.000 kHz by an attached PZT. During the measurement, the reference mirror was sinusoidally modulated at f_r = 26.080 kHz (Δf = 80 kHz) with the addition of offset modulation having a frequency of f_0 = 170 Hz. Based on these conditions, in theory, the CMOS camera can detect two heterodyne signals: one at a frequency of $f = \Delta f + f_0$ = 250 Hz (1/8 of the frame rate) and the other at the frequency of the offset modulation (170 Hz). Four-thousand frames of interference images were captured during 2 s. The measured planar area was 522 × 522 μm with a resolution of 256 × 256 pixels. The optical magnification of the objective lens was set to 9.8. Figure 5(A) shows microscopic *en face* raw images of the mirror surface 0.42, 1.00, and 1.53 s from the outset of the measurement. As mentioned in Subsection 2.2, it is difficult to accurately measure vibration parameters when the interference phase is approximately 0 rad or integer multiples of π rad. To characterize this so-called unmeasurable area, spatial interference fringe patterns were induced via tilting of the reference mirror.

Biomedical Optics EXPRESS

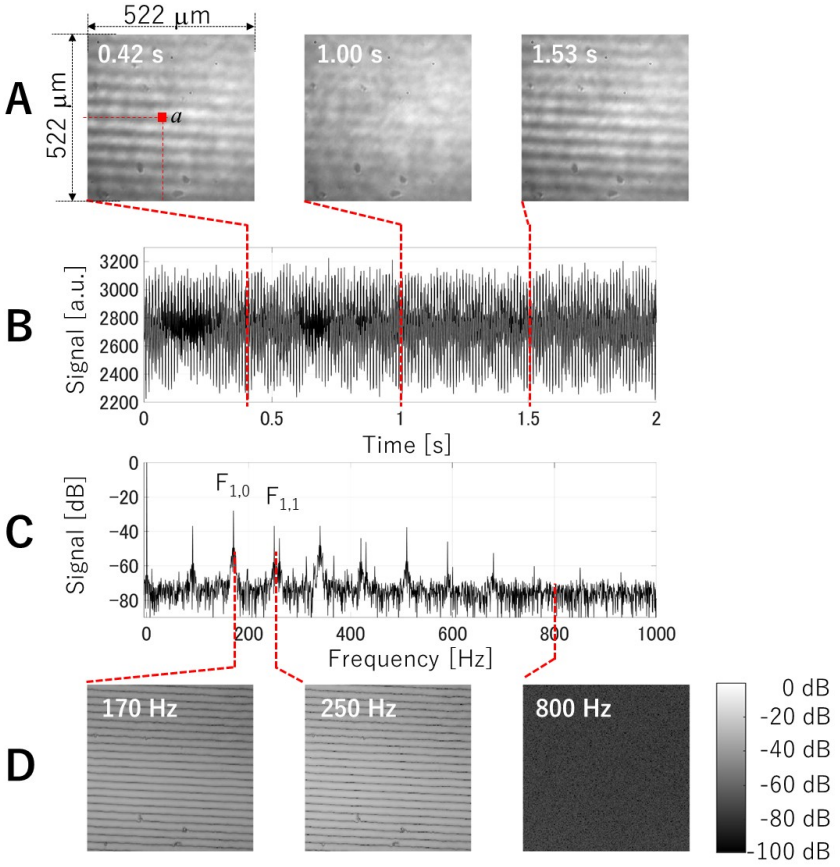

Fig. 5. Heterodyne signals detected by the WHIV technique with a vibrated mirror. (A) Microscopic *en face* interference raw images of the sample's surface at time points 0.42, 1.00, and 1.53 s. (B) A typical temporal heterodyne interferogram obtained in a point region indicated by *a* in (A). (C) Frequency domain signals obtained by FFT. (D) A 2D distribution of the frequency components $|F_{0,1}|$ (170 Hz) and $|F_{1,1}|$ (250 Hz) and a noise component (800 Hz).

Figure 5(B) illustrates the temporal change in the heterodyne signals at one point on the mirror (see dot *a* in Fig. 5(A)). In this experiment, the PZT was stimulated with an alternating current (AC) voltage of 5 V. From the recorded data, the frequency-domain signals were obtained by fast FFT, as shown in Fig. 5(C). We detected two components, i.e., $F_{1,1}$ and $F_{1,0}$, at 250 and 170 Hz, respectively. These results are consistent with the theoretical observations mentioned above. We further analyzed all the data points obtained on the surface of the sample with FFT (Fig. 5(A)). Then, from the signals that ranged from 0 to 1 kHz, the $F_{1,1}$ and $F_{1,0}$ components were extracted; they are visualized two-dimensionally in Fig. 5(D). Note that these two series of data were subjected to detection of the amplitude and phase distribution of the vibrations in the sample. As expected, little spatial information was available from the background noise observed at 800 Hz as shown in Fig. 5(D).

Biomedical Optics EXPRESS

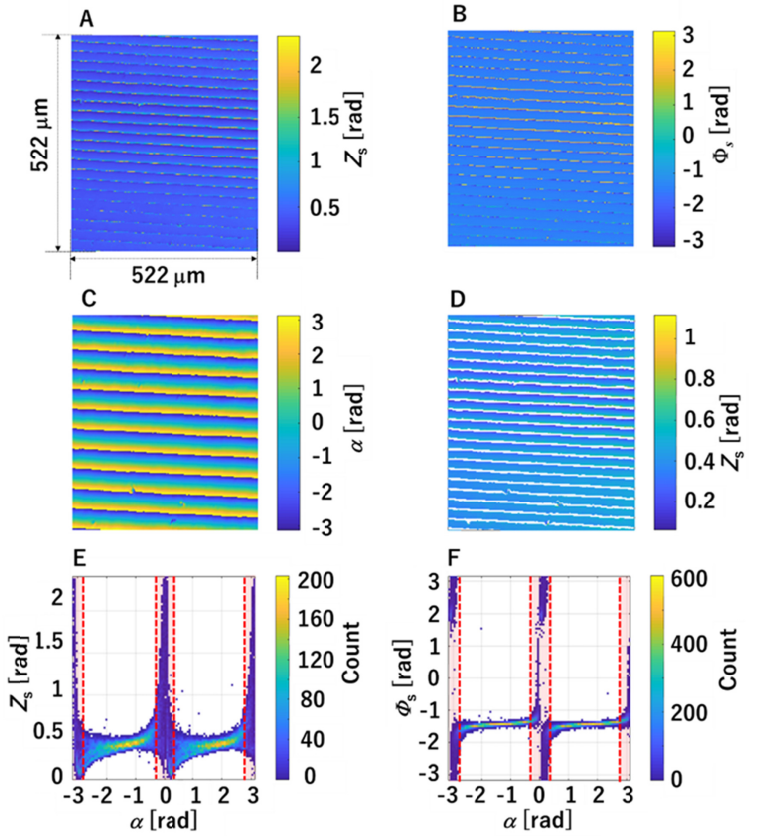

Fig. 6. Analysis of the vibrations measured in the mirror by the WHIV technique. For this assay, the data described in Fig. 5 were used. 2D distributions of vibration amplitude Z_s, vibration phase Φ_s, and interference phase α are displayed in panels (A), (B), and (C), respectively. The signals of the unmeasurable area were removed from the data, and the distributions of Z_s are reproduced in (D). The pixels were profiled in accordance with the Z_s and α values, and they are plotted in the histogram in (E). Similarly, the pixels of phase Φ_s values are plotted in the histogram in (F). The unmeasurable areas are indicated by hatching in panels (E) and (F).

These frequency components were analyzed by the estimation method described in Subsection 2.2. Figure 6(A–C) presents the distributions of vibration amplitude Z_s, phase Φ_s, and interference phase α. The pixels were profiled in accordance with the Z_s and α values; they are plotted in the histogram given in Fig. 6(E). It is expected that the Z_s values are constant regardless of the α values because the surface of the tested mirror should move uniformly under any conditions. Nevertheless, we found that the Z_s values markedly varied when the α values were integer multiples of π. At $|\alpha| < 0.1\pi$ or $|\alpha| > 0.9\pi$, the dispersion of Z_s was the largest; furthermore, regarding the vibration phase, the values were significantly varied in this area as well (Fig. 6(F)). Hence, we defined this region as the unmeasurable area. This area is indicated by hatching with vertical dotted lines in Fig. 6(E) and (F). The signals of this area were discarded, and the spatial distribution of Z_s was reconstructed as shown in Fig. 6(D).

We next evaluated the measurement accuracy of the improved WHIV technique. The values of the vibration amplitude obtained by this technique were compared to those of a conventional LDV, which can target only one point in a sample. The mirror was sinusoidally moved via application of an AC voltage of 0.2, 0.5, 1.0, 2,0, 3.0, 4.0, or 5 V to the PZT. As

Biomedical Optics EXPRESS

shown in Fig. 7(A), the average amplitude detected by the WHIV technique increased linearly with the strength of the stimulus for the PZT, and the values were in agreement with those acquired by the LDV. The average measurement error between the two methods was less than ~0.6 nm. Nonuniformity of the amplitude distribution was evaluated as a standard deviation denoted as error bars in Fig. 7(A). The FFT analysis indicated that when the voltage was 0.2 V, the peak of $F_{1,1}$ was not clear-cut; in this case, the noise floor (approximately −70 dB relative to the signal intensity at 0 Hz) exceeded the intensity of the signal (Fig. 7(B)). This limit of detection corresponded to ~1.1 nm in the vibration measurement.

To assess the measurement error, standard deviations (SDs) of the vibration amplitude and the phase were examined. Figure 7(C) shows the changes in the SDs of amplitude σ_Z and phase σ_Φ as a function of measured Z_s. These SDs represent the spatial fluctuations in the surface distribution for each parameter Z_s and Φ_s. Note that here, σ_Φ was defined by the following equation, considering that the average of the angles is expressed as addition of vectors [30]:

$$\sigma_\Phi = \sqrt{-2\ln\left|\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\exp\{i\Phi_s(m,n)\}\right|}, \tag{6}$$

where M and N are the numbers of pixels represented along the x-axis and y-axis, respectively.

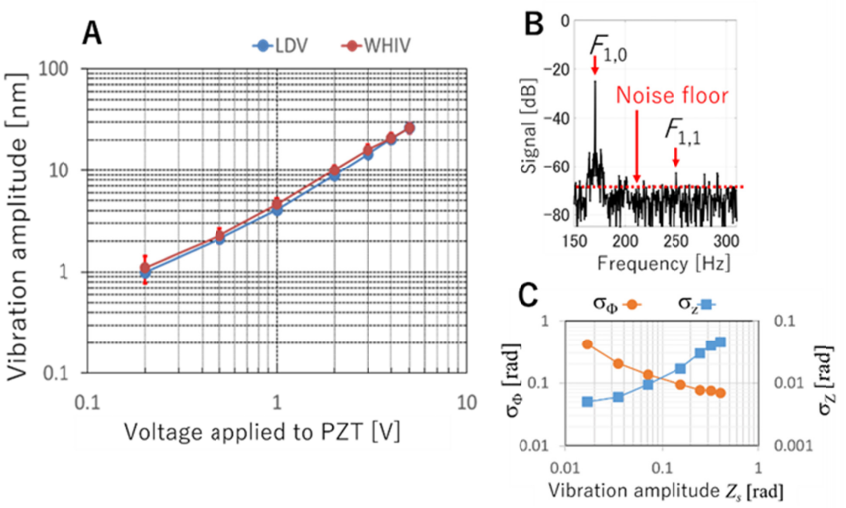

Fig. 7. Performance evaluation of the improved WHIV technique. (A) Comparison of measurement accuracy between the WHIV technique and a conventional LDV. For the two methods, a planar mirror vibrated at a frequency of 26 kHz served as a sample. Nonuniformity of the amplitude distribution can be evaluated by the standard deviation indicated as error bars. (B) The limit of detection of the vibration amplitude in the WHIV technique. The mirror was stimulated with 0.2 V. The obtained data were transformed to frequency domain heterodyne signals, and components $F_{1,0}$ and $F_{1,1}$ are presented in the panel. The noise floor is indicated by a dotted line. Refer to the text for the determination of the measurement threshold. (C) Changes in the SDs of vibration amplitude σ_Z and phase deviation σ_Φ with respect to Z_s.

As shown in Fig. 7(C), σ_Z increased in proportion to the measured Z_s. It remained almost constant at approximately 11% of the Z_s value in the range of 0.03 to 0.3 rad (0.5–5 nm). Nevertheless, the deviation increased to approximately 33% at an amplitude of 1.1 nm (Z_s = 0.03 rad) when the SNR of $|F_{1,1}|$ reached the limit indicated by Fig. 7(B). The deviation of phase σ_Φ changed in inverse proportion to Z_s. As a result, because the increase in the vibration amplitude made the SNR of $|F_{1,1}|$ higher, the SD of the phase tended to decrease. Therefore,

Biomedical Optics EXPRESS

σ_Φ can be a measure of accuracy in the detected signal. On the other hand, σ_Z was proportional to Z_s, indicating that the rate of measurement error was constant irrespective of the amplitude value in our system.

3.3. In vivo wide-field vibration analysis of the sensory epithelium

Using the MS-OCMV with the improved WHIV technique, we examined the cochlear sensory epithelium of a live guinea pig. First, to perform tomography of the tissue, we carried out an *en face* OCT measurement in a planar area of 522 × 522 μm at a resolution of 850 × 850 pixels. A total of 8000 *en face* images were acquired at a frame rate of 2000 fps by scanning within approximately 560 μm in the axial direction. Because of the high output power of the SC light source, the data were obtained in a single scan. The time of acquisition of these volumetric data (i.e., total scanning time) was 4 s. Optical magnification of the objective lens was set to 9.8, resulting in numerical aperture of 0.0093 and the depth of focus of 92 μm.

From the obtained data, we selected an area of 256 × 256 pixels for subsequent analyses. This component was processed as described in Subsection 3.1. After that, 3D volumetric images were reconstructed Fig. 8. (A), (B) shows the 3D volumetric images of the sensory epithelium including the neighboring bony component from different viewpoints. The contrast of the images was controlled by discarding low-intensity data below a threshold manually determined for each 3D data series. The dynamic range of the visualization was approximately 11 dB.

Figure 8(C) illustrates a remeasured result from the sensory epithelium in the portion enclosed by the dotted line in Fig. 8(B). In this measurement, 4000 *en-face* images were acquired by scanning within approximately 350 μm in the axial direction. Acquisition time was 2s. Figure 8(D–F) depicts the *X*-*Z* cross-sectional images sliced at *Y* axes corresponding to dotted lines 1, 2, and 3, respectively, in Fig. 8(C). The outline of the cross-sectional images was similar to a well-known view of the guinea pig sensory epithelium displayed in Fig. 8(G) [31]. In particular, the sensory epithelium has multiple regions that lack cells (e.g., the tunnel of Corti and inner sulcus); they are hallmarks for identifying such components of the epithelium as the BM, reticular lamina (RL), and tectorial membrane (TM). Therefore, in the images in Fig. 8(D–F), we could roughly detect the structures of BM, RL, and TM.

We next intended to visualize and quantify the vibrations of the sensory epithelium in the wide-field mode. Axial OCT scanning was performed near the BM, and this position was chosen as a target. During the measurement with the improved WHIV technique, the animal was exposed to a pure tone sound of 21, 22, 23, 24, or 25 kHz through a Y-shaped waveguide that was connected to the exit of a speaker (EC1; Tucker-Davis Technologies, FL, USA). The intensities of the acoustic stimuli were monitored by an ultrasonic microphone inserted into one output port of the waveguide. The other output port was tightly inserted into the left external ear canal of the animal. All the parameters for operating the system and the algorithm for analyzing the data were the same as those used in the performance evaluation experiment with the mirror, except for the following characteristic. In the animal experiment, the data throughout the image (256 × 256 pixels) were analyzed by FFT, and the 2D distribution and intensity of the $F_{1,1}$ component were monitored by the computer. This preliminary analysis indicated that the target region in the sensory epithelium responded most markedly to 23 kHz among the five frequencies that we tested (see the results described later).

Biomedical Optics EXPRESS

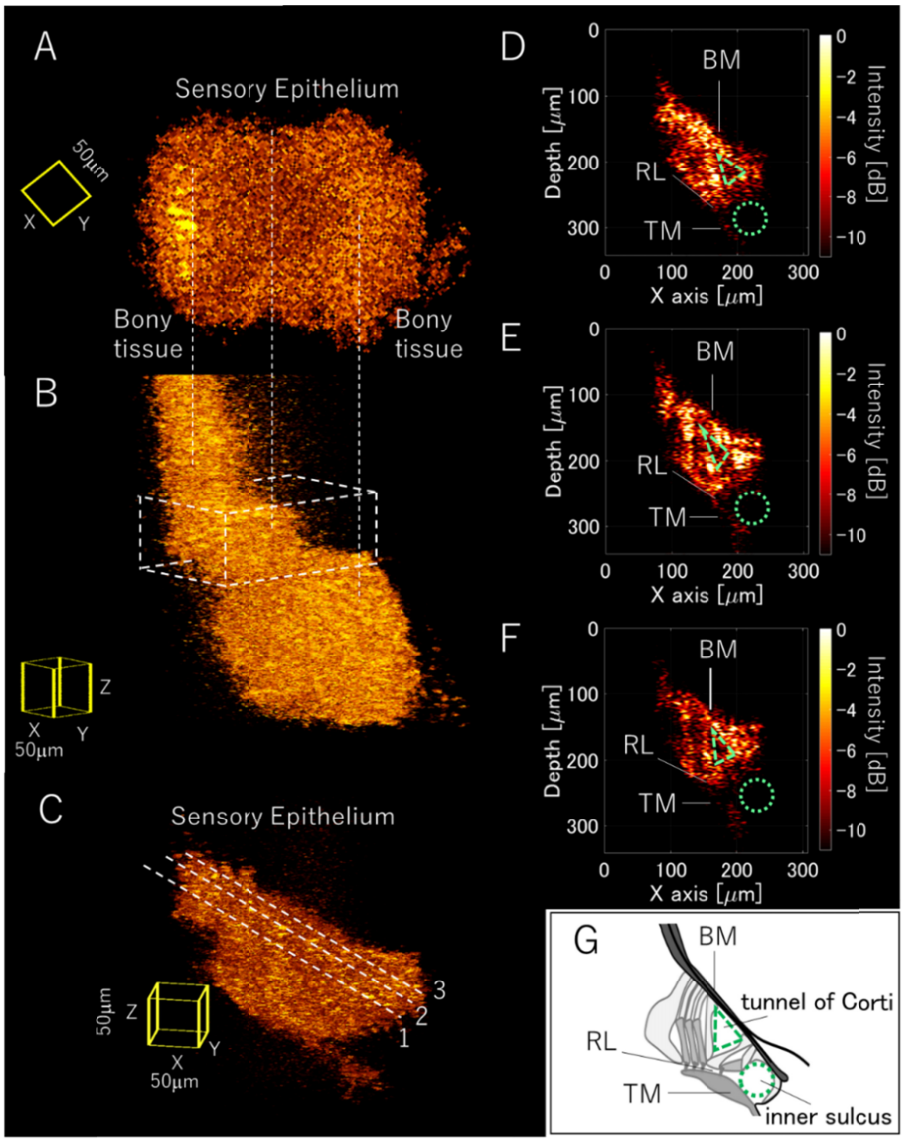

Fig. 8. Results of 3D volumetric imaging. (A) The 3D volumetric image of the sensory epithelium located between neighboring bony tissues. (B) A recalibrated image of the sensory epithelium from different viewpoints (see Visualization 1). (C) The remeasurement result from the portion enclosed by the dotted line in (B). The length of each side of the yellow parallelepiped is 50 μm. (D–F) *X-Z* cross-sectional images along lines 1, 2, and 3, respectively, indicated in (C). (G) Schematic illustration of the cross-sectional view of the sensory epithelium. BM, RL, and TM denote the basilar membrane, reticular lamina, and tectorial membrane, respectively. The inner sulcus and tunnel of Corti are marked by a dotted circle and triangle, respectively, which are also overlaid in D–F.

Furthermore, to analyze the area of the sensory epithelium as exclusively as possible, we extracted a region of interest (ROI) from the acquired *en face* image by a masking procedure. Figure 9 shows an example with a stimulus of 23 kHz. First, using the FFT data on the BM vibrations in each pixel, we averaged the intensity at frequencies of 0.9 to 1 kHz to determine a threshold for the masking procedure, because in this range, significant heterodyne signals were negligible (Fig. 5(C)). The mean + 2SD was defined as the threshold. Second, the pixel was configured to assume the value of "1" or "white" when the absolute value of the peak of

Biomedical Optics EXPRESS

$F_{1,1}$ ($|F_{1,1}|$) exceeded the threshold; otherwise, the configuration was set to "0" or "black." This process was applied to all the pixels at the same time, and the result was transformed into a black-and-white 2D map. This image is referred to as "mask 1." Third, on the basis of the interference phase α distribution, a pixel that manifested itself as an unmeasurable area was set to "0" or "black," whereas a measurable pixel was set to "1" or "white" (Fig. 6(D)). This digitization was carried out for all the pixels simultaneously, thus affording a 2D image called "mask 2." The final step was the "AND operation," in which a pixel with the value of "1" in both masks was redefined as "1" or "white," and data smoothing was performed by means of a median filter. These procedures produced the "conclusive mask" (Fig. 9).

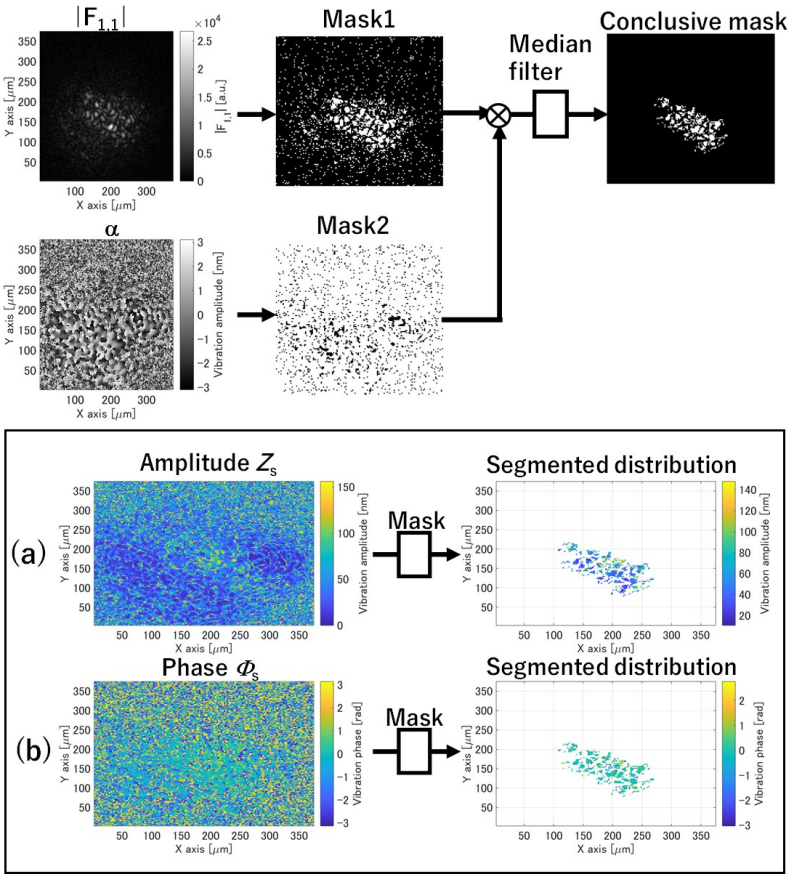

Fig. 9. The masking procedure. In mask 1, when the absolute value of the peak of $F_{1,1}$ in a pixel exceeded the threshold determined as described in the text, the pixel value was transformed to "1" or "white." Mask 2 served to filter out the unmeasurable area according to the value of interference phase α. In this configuration, the measurable pixel was referred to as "1" or "white." Masks 1 and 2 were merged by the "AND operation," in which the pixel with the value of "1" in both masks was set to "1" or "white." Finally, to the conclusive mask, the median filter was applied for data smoothing. For the detailed procedure, refer to the text. As indicated in the boxed panel, the distributions of (a) Z_s and (b) Φ_s were segmented through the conclusive mask. The ROI of the sensory epithelium in the wide-field mode was determined by this procedure.

Biomedical Optics EXPRESS

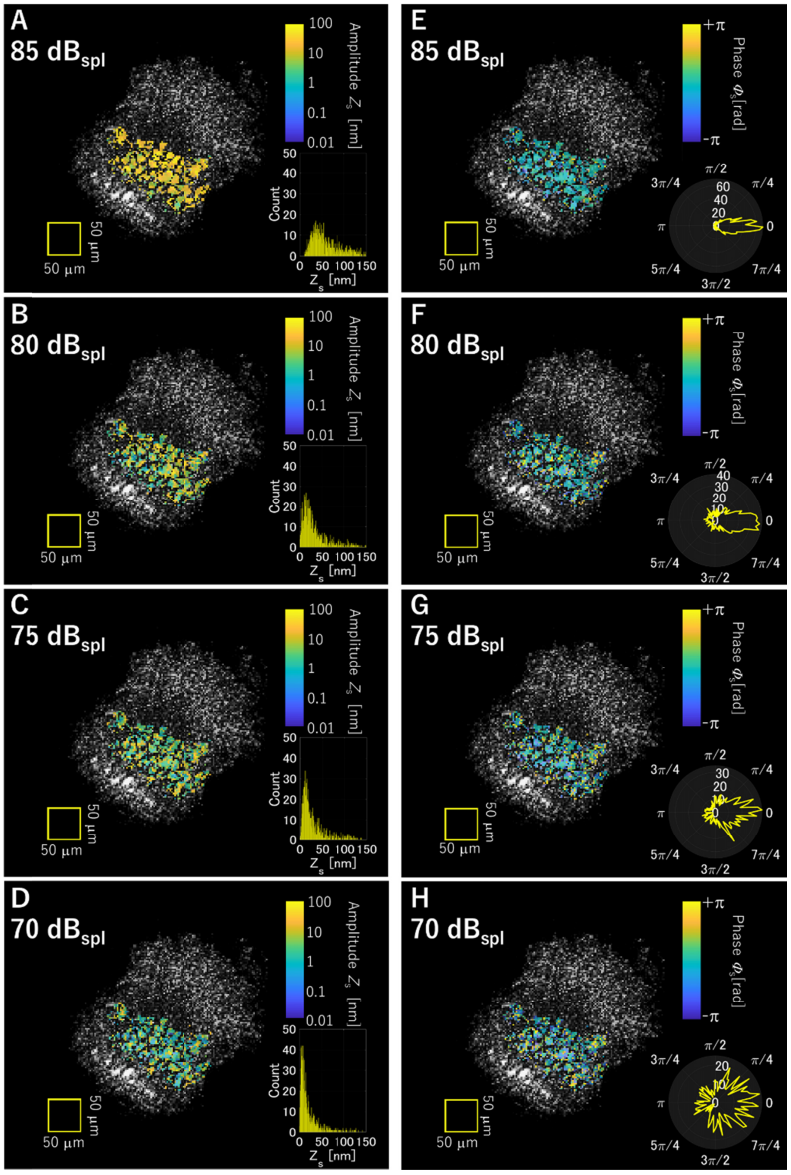

Fig. 10. Two-dimensional visualization of vibrations in the cochlear sensory epithelium of a live guinea pig. Vibration amplitude Z_s (A–D) and phase Φ_s (E–H) in the ROI on the epithelium (see text) are denoted by the color bar shown in the upper right part of each panel and mapped on a slice section of the 3D volumetric image obtained in Fig. 8(A). In the measurements, the animal was exposed to acoustic stimuli at various SPLs: 85 dB [(A) and (E)], 80 dB [(B) and (F)], 75 dB [(C) and (G)], or 70 dB [(D) and (H)]. The insets in the lower right corners of panels (A–D) indicate the amplitude histograms, whereas those in (E–H) indicate phase histograms. By means of these parameters and the volumetric data, the pattern of the vibration can be visualized schematically (see Visualization 2).

Figure 10 presents the properties of the BM vibrations in the ROI when the sound pressure levels (SPLs) of the acoustic stimuli applied to the animal were 70, 75, 80, or 85 dB (frequency: 23 kHz). In this 2D map, the values of Z_s, and Φ_s in each pixel were converted into the amplitude and phase histograms. Phase values Φ_s were normalized to the averaged angular phase of each distribution. Overall, the amplitude increased with the strength of the acoustic stimuli. Furthermore, when the animal was exposed to sounds at 70 and 75 dB SPLs,

Biomedical Optics EXPRESS

the phase distribution was markedly different. This parameter was relatively homogenous with the stimuli of 80 and 85 dB SPLs. Therefore, the measurement is likely to become more reliable as the response of the BM increases.

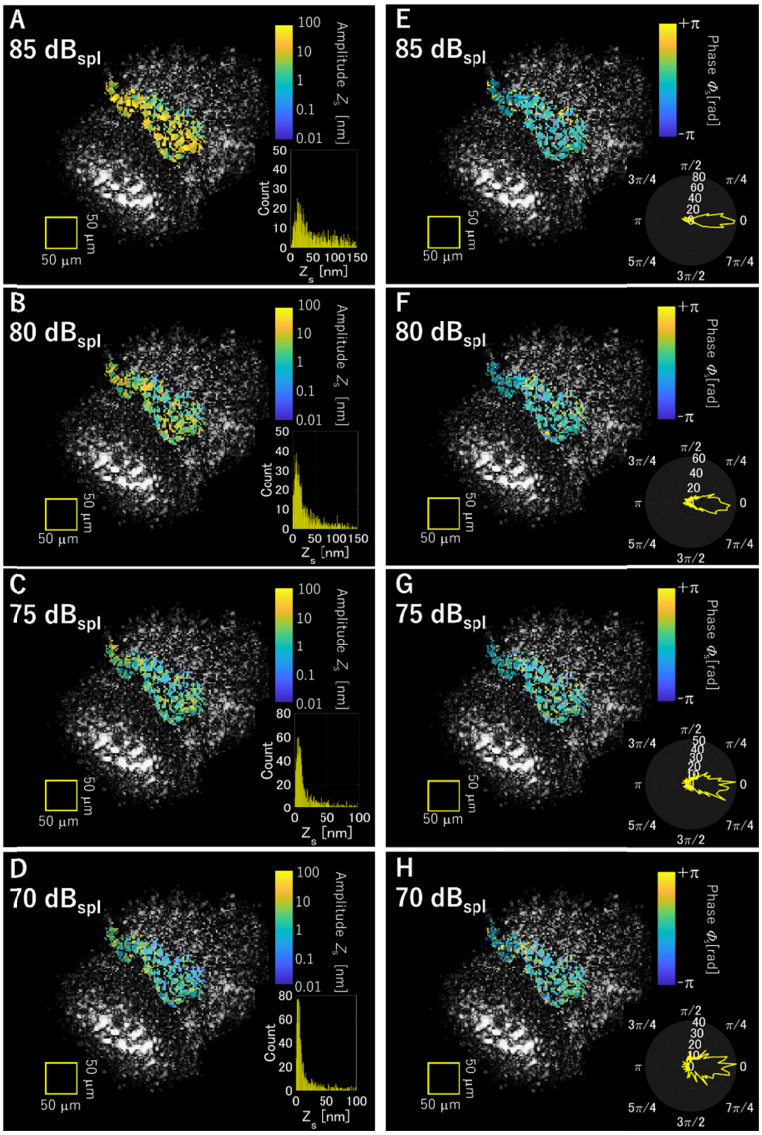

Fig. 11. Two-dimensional visualization of vibrations in the sensory epithelium of the guinea pig postmortem. The tested animal, experimental conditions, data analyses, and displayed parameters are the same as those in Fig. 10.

Approximately 1 h after the measurement shown in Fig. 10, the guinea pig was euthanized. The ROI was again determined by the procedure described in Fig. 9, and the vibrations of the BM were recorded via the protocol depicted in Fig. 11. For each type of stimulus, the vibration amplitude quantified in the animal postmortem was likely to be less than that in the live animal (Fig. 11(A–D); see also Fig. 10(A–D)). This observation may result from dysfunction of hair cells' active processes that can amplify the motion of the BM [32]. In addition, for a stimulus of any intensity, little dispersion of the phase distribution was detected throughout the ROI in the animal postmortem (Fig. 11(E–H)), although the spatial

Biomedical Optics EXPRESS

deviation of this parameter was apparent under the control conditions (Fig. 10(E–H)). This difference seems to stem from the lack of noise sources, such as breathing and blood flow, after death.

Fig 12(A) presents a plot of the average values and standard deviations calculated from the vibration amplitude recorded for all the pixels included in the ROI (approximately 2000 pixels) versus different sound intensities. Note that the number of pixels differed among the trials. Overall, the values for the live guinea pig (control) exceeded those for the postmortem animal, as mentioned above. A comparison of these two series of data revealed that when the stimuli were relatively weak (70 and 75 dB), the response increased under the control conditions [2,31]. This nonlinear amplification supports the idea that the cochlea (including the sensory epithelium) was damaged only minimally during the measurement.

Fig 12(B) illustrates the measurements at 21, 22, 23, 24, and 25 kHz (85 dB SPL). For each type of stimulus, the ROI was determined. The averaged values of the amplitude were obtained with the procedure utilized in Fig. 12(A). The other displayed parameters are the deviations of phase values (σ_Φ). At the stimuli of 23 kHz, the amplitude was maximal and the deviation was minimal. Because the lower σ_Φ value means a higher SNR of the detected signal as mentioned in Subsection 3.2, we inferred that the characteristic frequency at the point we examined was 23 kHz. This observation is in agreement with the aforementioned preliminary finding that the characteristic frequency of the epithelium targeted by the laser was 23 kHz.

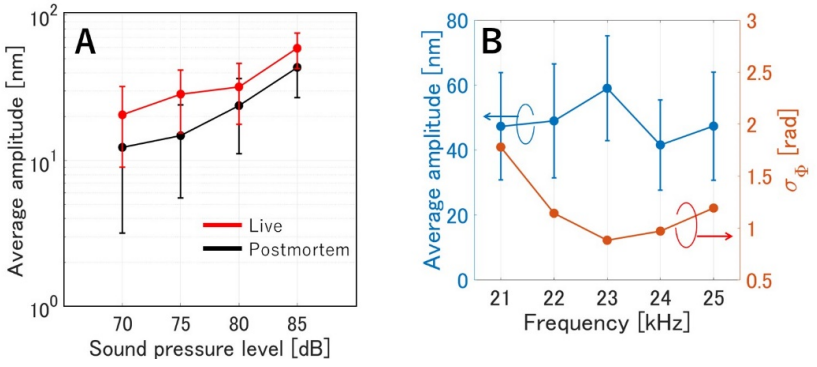

Fig. 12. Epithelial vibrations measured with the improved WHIV technique. (A) Vibration amplitude. The data were collected from the ROI (~2000 pixels) of the sensory epithelium in the control (red curve; live guinea pig) and postmortem (black curve). The animal was exposed to stimuli of 23 kHz at various SPLs. The average values and standard deviations (SD) are plotted. (B) The tuning curve of the vibration amplitude (blue curve) and phase (orange curve; σ_Φ). In this assay series, the live animal was acoustically stimulated with different frequencies (21–25 kHz; 85 dB SPL). The averages and SD of the data in the ROI individually determined for each stimulus are shown. The phase values were obtained with Eq. (6) (see text).

Theoretical models and experiments on the base of the cochlea corroborate that the BM shows wave propagation behavior with the phase gradient along the cochlea from the base to the apex in the region of the characteristic frequency [2]. In the region of the sensory epithelium approximately 2–3 mm in length observed through the window, the characteristic frequency is known to be in the range 28–32 kHz [33] along the cochlear spiral on the basal side. The other apical side can be predicted to be 19–22 kHz according to Ref [32]. Furthermore, the experimental results mentioned above suggest that the characteristic frequency varies in the range ~21–25 kHz in the region we examined. Especially significant signals were detected at 23 kHz (Fig. 12(B)). Therefore, it is expected that the phase difference caused by the traveling wave that is scale invariance can be detected in this frequency range.

Biomedical Optics EXPRESS

To confirm this theoretical prediction, we analyzed the phase data using the same spatial distributions as in the above experiment (Fig. 12 (B)). Figure 13 shows the distributions obtained at different frequencies and their averaged phase gradients on the BM. Because of the limitation on sensitivity in this system, we could examine only the data obtained at >80 dB SPL. Overall, more reliable results were obtained from a postmortem animal than from a live one. The reason is that noise was deceased because movement causing the artifacts was reduced in the postmortem animal.

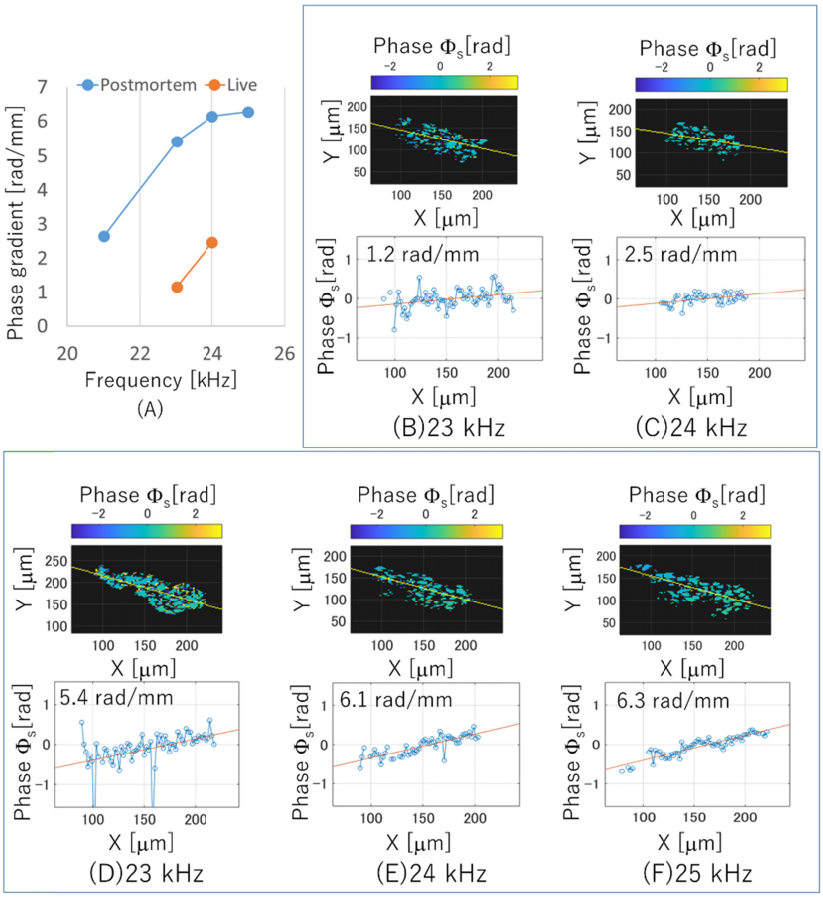

Fig. 13. Analysis of the phase gradient caused by a travelling wave on the BM. (A) Variation of the phase gradient plotted in terms of radian/mm as a function of sound stimulation frequency. The gradient values were obtained from the spatial phase distributions of vibration stimulated by a pure tone sound at 85 dB SPL. (B, C) Spatial phase distributions and their phase gradients on the BM of a live guinea pig at a frequency of 23 and 24 kHz, respectively. The phase gradients were obtained by calculating the average phase value along the centroid line (yellow lines on each distribution). (D–F) Spatial phase distributions and their phase gradient corresponding to frequencies of 23, 24, and 25 kHz, respectively. The tested animal, experimental conditions, and data analysis are the same as those in panels (B) and (C).

In the analysis, first, the centroid line crossing the longitudinal direction of the sensory epithelium was determined for quantitative evaluation, as reference baseline representing the direction of the membrane (yellow lines on the distributions in Fig. 13). This line was calculated by linear approximation using the centroid y-axis coordinates with respect to the x-axis. Then, the phase values aligned on a vertical line normal to the direction were averaged. Thereby, the averaged values were plotted along the centroid line (Fig. 13(B)). The averaged

Biomedical Optics EXPRESS

result was removed where the number of valid points was less than 2.5% of the pixel number in the vertical line (for example, most of data located out of the ROI). Finally, from the averaged phase plot, the tilt of an approximate straight line obtained via polynomial regression was estimated as the phase gradient. In this analysis, the standard error of the approximation for the line was ~0.05 rad on average.

As a result, the gradient of the phase varied between approximately 1 and 7 rad/mm. The phase changes were observed at 23, 24, and 25 kHz oscillations of the postmortem animal. In the case of the live animal, comparatively small phase gradients were observed. The result at 23 kHz, which was considered the characteristic frequency, was estimated to be 1.15 and 5.39 rad/mm respectively in the live and postmortem animal. The largest gradient was estimated to be 6.27 rad/mm at 25 kHz oscillation postmortem. Reasonable values of phase gradients were obtained in this analysis as compared with the result obtained in Ref [34]. We confirmed the phase changes predicted theoretically by analyzing selected distributions with relatively less noise. Further improvements are needed for more sensitive and accurate measurements in a live tissue.

4. Discussion

MS-OCMV has been developed for *en face* measurement of biological vibrations by the WHIV technique. In comparison with Doppler SD-OCT, however, disadvantages and problems are yet to be improved. Generally, standard Doppler SD-OCT is superior to our method in terms of real-time performance and sensitivity (SNR of 90–100 dB and picometer accuracy for vibrometry). Our system might be superior in terms of the speed of capturing 3D volumetric data with spatial simultaneous *en face* detection using a high-speed CMOS camera. SD-OCT and a conventional LDV requires a photodetector or a line sensor with high sampling frequency to avoid aliasing. In our technology, in principle, there is no limitation on the vibration frequency because the heterodyne signal sufficiently detectable by a CMOS camera produced by two sinusoidal phase modulations is utilized to analyze the vibration.

Nevertheless, for practical applications, high-speed cameras are preferred, to avoid the effects of low-frequency noise mentioned above. In addition, sensitivity of the high-speed CMOS cameras is generally lower than that of standard photodetectors owing to the lower full-well capacity (16,000 e^-). Besides, it is known that SD-OCT methods based on Fourier spectroscopy guarantee a higher SNR than qualitative time domain methods [35,36]. A representative Doppler SD-OCT system (e.g., Ganymede SD-OCT, Thorlabs, USA) provides sensitivity of ~100 dB for OCT imaging. On the other hand, OCT sensitivity of an MS-OCMV system were estimated to be approximately 40 dB, respectively. Further improvement of the system is needed for practical *in vivo* measurements. In this section, we discuss current issues from this perspective to clarify the limitations and prospects.

4.1. The noise level of OCT

The original MS-OCMV, which was equipped with an SLD light source (center wavelength: 820 nm), could not clearly detect either the 3D volumetric image or the sound-induced vibrations in the sensory epithelium. In this series of experiments, the light source typically administered an optical power of 0.06 μW to each pixel of the CMOS camera during a usual exposure time of 0.5 ms. Because of the low reflectance of the sensory epithelium, it is estimated that a pixel of the camera with a quantum efficiency of 25% received only 54 e^- from the tissue. In such a condition, the noise of the signal can be determined as

$$noise = \sqrt{N_{td}^{\ 2} + N_s^{\ 2}}\ , \tag{7}$$

where N_{td} and N_s denote temporal dark noise (i.e., read noise) and shot noise, respectively. In accordance with the native profile of the image sensor [37], temporal dark noise (N_{td}) was 29 e^-. Shot noise (N_s), calculated by the square root of the number of photon fluxes on a one-

Biomedical Optics EXPRESS

pixel surface, was approximately 15 e^-. Consequently, the noise calculated by Eq. (7) was 33 e^-. These data indicate a low SNR of approximately 4.3 dB, which is a major reason for the unsuccessful analysis of the epithelium by the previous system. On the other hand, in the present study, the SLD light source was replaced with the SC light source. This state of affairs increased optical power approximately 41-fold; therefore, in accordance with Eq. (7), the SNR was increased to 27.4 dB with the same CMOS camera.

When a mirror was examined with the improved MS-OCMV, ripples were detected around the main peak in the interference signals (Fig. 4(D)). Such a disturbance may affect sensitivity and quality of the imaging. The interference signals are closely associated with the power spectrum of the light source. The ripples were caused by the rectangular shape of the power spectrum. To resolve this problem, an optical filter that can reshape the edge of the spectral envelope should be applied to the multifrequency generation unit that contains the light source.

The actual axial resolution depicted in Fig. 4(C) deteriorates more than the ideal axial resolution, probably because of the aberration of the microscope, dispersion between the reference and object paths of the interferometer, and the wavelength characteristic of the CMOS camera. The spectral-amplitude distribution of the SC light source also influences the shape of the coherence function (e.g., large ripples and double peaks). Besides, absorption and wavelength dispersion in the sample could influence the axial resolution. Further improvement is necessary in future experiments, e.g., a dispersion compensation method for low-coherence interferometry [38].

4.2. Comparison between the original and improved WHIV techniques

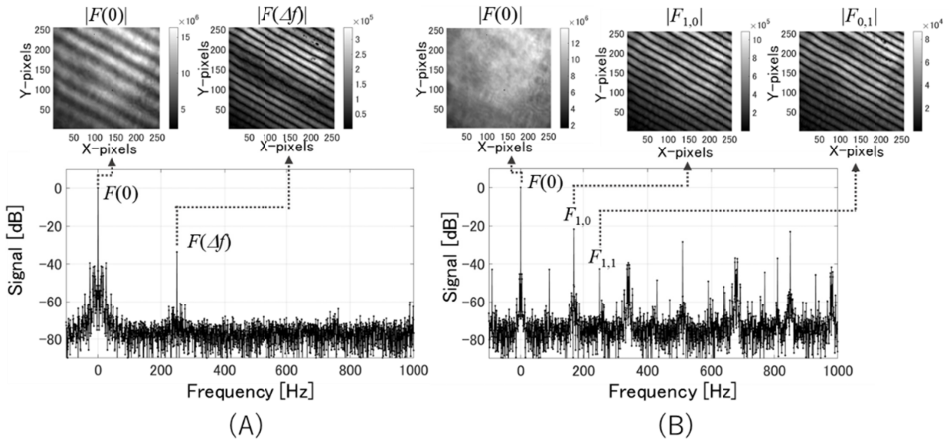

Fig. 14. Typical heterodyne signals in the frequency domain detected by (A) the original and (B) improved WHIV techniques with a vibrated mirror. In the original methods, 2D distributions of the frequency components $|F(0)|$ (0 Hz) and $|F(\Delta f)|$ (250 Hz) were obtained. The distribution of $|F(0)|$ shows an interference fringe pattern, which means that the DC component and first-order signal merged in the distribution. On the other hand, in the improved technique, the distributions of the DC component of $|F(0)|$ (0 Hz) and high-order frequency components of $|F_{0,1}|$ (170 Hz) and $|F_{1,1}|$ (250 Hz) were obtained separately.

We compared the improved WHIV with the original method to demonstrate the validity of this improvement. Figure 14 shows typical frequency domain signals obtained by these methods. The reference mirror was vibrated by a pure-tone sinusoidal signal and by the DC offset modulated signal for the improved and original WHIV, respectively. The sample mirror was vibrated by a sinusoidal signal with a frequency of f_s = 23.000 kHz in both cases. The reference frequencies for the improved method, f_r, and f_0 were 23.080 kHz and 170 Hz, respectively. For the original method, f_r was set to 23.250 kHz. Thus, Δf was 250 Hz. The

Biomedical Optics EXPRESS

amplitude voltages applied to the PZTs for Z_s, Z_r, and Z_0 were set to 1.0, 1.0, and 5.0 V, respectively. The *en face* images consisted of 256 × 256 pixels were extracted to display the distribution set of $F(0)$ and $F(\Delta f)$ or of $F_{0,0}$, $F_{0,1}$, and $F_{1,1}$ in the original method and improved method, respectively, as illustrated in Fig. 14.

The intensity of the frequency component $F(0)$ in the original method can be written as [16]

$$\left|F(0)\right| = \left| A_{xy} + B_{xy}\cos(a_{xy})J_0(Z_s)J_0(Z_r)\right|. \tag{8}$$

Vibration amplitude Z_s can be obtained by extracting the second term from Eq. (8) and calculating the intensity ratio to first-order component $F(\Delta f)$ via a process similar to the one described in Subsection 2.2. Nevertheless, it is difficult to eliminate DC component A_{xy} because these two terms are linearly added and merged into one signal in the frequency domain. Thus, initially, the original technique [16,18] employed 2D image processing based on the traditional Fourier transform method [39] focusing on the spatial interference fringe pattern appearing in the obtained *en face* distribution (Fig. 14(A)). Nevertheless, this arrangement rather resulted in diminished accuracy for quantifying the vibration amplitude especially when the analyzed surfaces were complicated as in biological tissues. In Ref [17], a substitute method utilizing a second-order component instead was adopted. Nonetheless, the intensity of the substitute signal is inherently weaker than that of the zeroth- and first- order signals. This problem is prominent when we analyze biological samples such as the cochlear sensory epithelium that yield a weak interference signal.

To solve these problems, in the improved method, the zeroth-order component can be clearly separated from the DC component by means of a third signal (DC offset modulation). As shown in Fig. 14(B), obviously $F(0)$ contained only DC term A_{xy}. The zeroth-order component was obtained as $F_{1,0}$ independently. The principle that can be discerned in the frequency domain is that the zeroth-order component was shifted by offset modulation frequency f_0 from 0 Hz, and that first-order component $F_{1,1}$ was regarded as a sideband around $F_{1,0}$. This feature enables accurate extraction of this essential signal for the vibration amplitude without interference from the DC component.

An advantage of this improvement is that the method can also estimate the interference phase α. In addition, the vibration amplitude can be estimated in the "unmeasurable area" using the pair of second-order signals, $F_{2,0}$ and $F_{2,1}$, alternatively. A disadvantage of this improvement is that in the comparison of the peak intensities between $|F(\Delta f)|$ and $|F_{1,1}|$, the SNR in the improved method deteriorated by approximately 10 dB as shown in Fig. 14.

The effect of low-frequency noise was also confirmed in Fig. 14(A). The low-frequency noise was distributed up to about 50 Hz (Fig. 14 (A)), and when WHIV was used, it appeared as sidebands of each frequency component. This low-frequency noise was mainly due to mechanical disturbances of the interferometer. In the original method, modulation frequency Δf is set higher to prevent the influence of low-frequency noise. Even in the improved method, it was necessary to guarantee a sufficient frequency spacing between the longitudinal modes to separate them from the sidebands of noise components. Therefore, in our system, by introducing a high-speed CMOS camera with 2000 fps or more, accurate measurement could be achieved that was less susceptible to the low-frequency noise.

4.3. The performance limit of the improved WHIV technique

In general, verification of the limit of detection is crucial for characterization of any analytical instrument or method. As for the improved WHIV technique, here, we chose an *in silico* approach to evaluate the error in the measurement of vibration amplitude. According to Eq. (1), in the simulation, we set several parameters to reconstitute the interferometric heterodyne signals in the time domain as follows: Z_r = 1.8 rad, Z_0 = 1.8 rad, f_s = 23 kHz, f_r = 23080 Hz, and f_0 = 170 Hz. In addition, we reproduced the noise floor of −70 dB as shown in Fig. 7(B)

Biomedical Optics EXPRESS

by additionally providing the simulated heterodyne signals with uniformly distributed random noise that ranged within 6.25% of interference intensity. Furthermore, interference phase α was configured to be $\pi/2$ such that both $|F_{1,1}|$ and $|F_{1,0}|$ were maximal.

Four series of heterodyne signals were generated at a rate of 1 MHz; each series consisted of 10^6 points (1 s), 2×10^6 points (2 s), 4×10^6 points (4 s), or 8×10^6 points (8 s). In the experiments, we captured interference images with the CMOS camera at 2000 fps. According to this sampling period (0.5 ms), in each simulated signal series, sets of 500 points were taken in order and were averaged. Consequently, the number of sampling points (N) were 2000, 4000, 8000, or 16000. Note that in the *in vivo* measurements, biological specimens were examined for 2 s; such experimental conditions could be mimicked with 4000 sampling points in the simulation.

Next, WHIV was applied to these signals. As described above, with a certain input vibration amplitude value (Z_s), different resulting values (ζ_s) were calculated owing to random noise. The relative error between the two is defined as $100|Z_s\text{-}\zeta_s|/Z_s$. The relation between this value and Z_s is shown in Fig. 15(A). Furthermore, as presented in Fig. 15(B), the SNR of $|F_{1,1}|$ is obtained with reference to the noise floor and plotted as a function of Z_s.

In each series of sampling points, as Z_s increased, the relative error decreased (Fig. 15(A)) and the SNR increased (Fig. 15(B)). Under the same conditions as in the experiment (at N = 4000), the relative error exceeded 25% at $Z_s \leq 1$ nm (Fig. 15(A)). In this case, the SNR was approximately 3 dB; in other words, the signal intensity was approximately twice that of the noise floor (Fig. 15(B)). This result is consistent with the limit of detection in the measurement of vibration amplitude, i.e., ~1 nm.

In addition, an increase in N reduced the relative error and elevated the SNR (Fig. 15(A), (B)). This observation strongly indicates that to improve the sensitivity and accuracy of the measurement, the duration of data acquisition should be extended. Nevertheless, these settings are accompanied by an increase in motion artifacts in the case of live animals and should hence be carefully applied to *in vivo* assays in future studies.

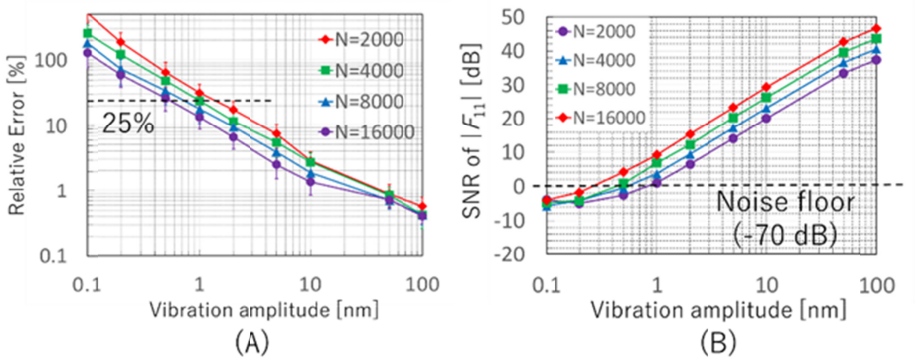

Fig. 15. *In silico* analysis for evaluation of the improved WHIV technique. Four series of sampling points (2000, 4000, 8000, and 16000), which represent measurements during 1, 2, 4, and 8 s, respectively, were computationally prepared as described in the text. As a function of vibration amplitude the relative error (A) and SNR of the $|F_{1,1}|$ component (B) were plotted. Refer to the text for details. In (A), the calculation was carried out 100 times, and the averaged values with standard deviations are presented (error bars).

Note that this analysis of the error implies the ideal case when the planar mirror vibration was examined with the noise floor of −70 dB. Given that sensitivity depends on the optical SNR, it should be considered with respect to actual cochlear vibration measurement for estimating the measurement error and sensitivity. In the *in vivo* experiment, the noise floor of $|F_{1,1}|$ was approximately −55 dB with the sound pressure level of 70 dB for stimulation. Therefore, we could estimate the detectable minimum amplitude of the sensory epithelium

Biomedical Optics EXPRESS

and found it to be roughly 5 nm (Fig. 15(B)). Further studies are needed to elucidate the details of the actual sensitivity of *in vivo* measurement.

4.4. Motion artifacts

A motion artifact generally leads to a loss of vibration data and deterioration of the SNR ratio. If the fluctuation of the OPD is limited to a phase change of approximately several hundred nanometers, it is possible to remove its influence because the fluctuation is detected as low-frequency noise of the signal. Further changes lead to alteration of the measurement area. Therefore, it is desirable that the OPD fluctuate within a few micrometers, which is the coherence length during acquisition of the data. To prevent the motion artifact, we controlled the movement due to breathing of the animal by artificial ventilation via a respirator during the measurements as mentioned in Subsection 2.3.

Another solution is to reduce measurement time by increasing the frame rate of the CMOS camera. The advantage of this improvement method is that it can reduce the accumulation effect of the detector and improve time resolution to capture the movement of the animal. Accumulation during exposure time in a pixel of the image sensor may cause blurring of the detected signal because it is very sensitive to the phase change due to the path length variation. Therefore, the accumulation effect of the detector worsens contrast in the interference signal.

We investigated the influence on the contrast reduction exerted by accumulation time and scan speed in a previous study [17]. For instance, according to the simulation, it is possible to improve the degradation of interference contrast from approximately 78% to 94% when the frame rate increases from 2000 to 4000 fps if the frequency of heterodyne signal $F_{1,1}$ is 250 Hz. Nevertheless, application of this improvement requires careful consideration of the overall-light-intensity reduction due to the exposure time shortening. Given that intensity of the accumulated interference signal is also proportional to exposure time (e.g., see Eq. (6) in Ref [17].), acquisition at 4000 fps, for instance, causes deterioration of the SNR by approximately 3 dB as compared to the results of this study. As demonstrated in Fig. 15, deterioration of the SNR increases the error of the measured amplitude value. Therefore, there is a trade-off between amelioration of a motion artifact due to acceleration of the frame rate and an increase in measurement error due to shorter exposure time.

5. Conclusion

In this study, we drastically modified our previously developed MS-OCMV system for the acquisition of 3D tomographic images and for analysis of a wide-field vibration distribution of objects. The present system was equipped with an SC light source to enhance irradiation Furthermore, in the WHIV technique, we added offset modulation to the reference signal to accomplish wide-field measurement of ultrafast vibrations in a biological sample. The performance of the improved MS-OCMV system for 3D volumetric imaging is characterized by a transverse resolution of 3.6 μm and an axial depth resolution of 2.7 μm. The vibration amplitude detectable with this system was estimated to be ~1.1 nm, and measurement accuracy is similar to that of conventional LDVs. These settings enabled us to detect the structure and motion in an acoustically stimulated cochlear sensory epithelium of a live guinea pig, even though this tissue has an extremely low reflectance rate. With sounds of different intensities or frequencies, the spatial distribution of the vibration amplitude and phase was quantified and mapped onto a 3D volumetric image. The profile changed when the animal was euthanized. The proposed technique can provide a platform for effective analysis of the cochlea as well as other organs, thereby contributing to advances in life sciences.

Funding

AMED-CREST, AMED (JP18gm0810004); JSPS KAKENHI (16H03164); and the Joint Research Program of the Biosignal Research Center, Kobe University (281003).

Biomedical Optics EXPRESS

Acknowledgments

We would like to thank Prof. Takamasa Suzuki, Niigata University, for his technical expertise in optical interferometry and valuable support throughout our research. We would like to thank Editage (www.editage.jp) for English language editing.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. A. J. Hudspeth, "Integrating the active process of hair cells with cochlear function," Nat. Rev. Neurosci. **15**(9), 600–614 (2014).
2. L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," Physiol. Rev. **81**(3), 1305–1352 (2001).
3. P. K. Legan, V. A. Lukashkina, R. J. Goodyear, M. Kössi, I. J. Russell, and G. P. Richardson, "A targeted deletion in α-tectorin reveals that the tectorial membrane is required for the gain and timing of cochlear feedback," Neuron **28**(1), 273–285 (2000).
4. M. M. Mellado Lagarde, M. Drexl, V. A. Lukashkina, A. N. Lukashkin, and I. J. Russell, "Outer hair cell somatic, not hair bundle, motility is the basis of the cochlear amplifier," Nat. Neurosci. **11**(7), 746–748 (2008).
5. S. M. Khanna, "Homodyne interferometer for basilar membrane measurements," Hear. Res. **23**(1), 9–26 (1986).
6. M. Ulfendahl, S. M. Khanna, and C. Heneghan, "Shearing motion in the hearing organ measured by confocal laser heterodyne interferometry," Neuroreport **6**(8), 1157–1160 (1995).
7. M. Ulfendahl, S. M. Khanna, and A. Flock, "The vibration pattern of the hearing organ in the waltzing guinea-pig measured using laser heterodyne interferometry," Neuroscience **72**(1), 199–212 (1996).
8. A. L. Nuttall and D. F. Dolan, "Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig," J. Acoust. Soc. Am. **99**(3), 1556–1565 (1996).
9. N. P. Cooper, "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. **509**(Pt 1), 277–288 (1998).
10. M. A. Ruggero, N. C. Rich, A. Recio, S. S. Narayan, and L. Robles, "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**(4), 2151–2163 (1997).
11. J. A. N. Fisher, F. Nin, T. Reichenbach, R. C. Uthaiah, and A. J. Hudspeth, "The spatial pattern of cochlear amplification," Neuron **76**(5), 989–997 (2012).
12. P. Picart, J. Leval, J. C. Pascal, J. P. Boileau, M. Grill, J. M. Breteau, B. Gautier, and S. Gillet, "2D full field vibration analysis with multiplexed digital holograms," Opt. Express **13**(22), 8882–8892 (2005).
13. I. Shavrin, L. Lipiäinen, K. Kokkonen, S. Novotny, M. Kaivola, and H. Ludvigsen, "Stroboscopic white-light interferometry of vibrating microstructures," Opt. Express **21**(14), 16901–16907 (2013).
14. M. Khaleghi, C. Furlong, M. Ravicz, J. T. Cheng, and J. J. Rosowski, "Three-dimensional vibrometry of the human eardrum with stroboscopic lensless digital holography," J. Biomed. Opt. **20**(5), 051028 (2015).
15. S. Sato, T. Kurihara, and S. Ando, "Real-time vibration amplitude and phase imaging with heterodyne interferometry and correlation image sensor," Proc. SPIE **7063**,70630I (11 August 2008)
16. S. Choi, Y. Maruyama, T. Suzuki, F. Nin, H. Hibino, and O. Sasaki, "Wide-field heterodyne interferometric vibrometry for two-dimensional surface vibration measurement," Opt. Commun. **356**, 343–349 (2015).
17. S. Choi, K. Sato, T. Ota, F. Nin, S. Muramatsu, and H. Hibino, "Multifrequency-swept optical coherence microscopy for highspeed full-field tomographic vibrometry in biological tissues," Biomed. Opt. Express **8**(2), 608–621 (2017).
18. S. Choi, T. Watanabe, T. Suzuki, F. Nin, H. Hibino, and O. Sasaki, "Multifrequency swept common-path en-face OCT for wide-field measurement of interior surface vibrations in thick biological tissues," Opt. Express **23**(16), 21078–21089 (2015).
19. S. S. Gao, P. D. Raphael, R. Wang, J. Park, A. Xia, B. E. Applegate, and J. S. Oghalai, "In vivo vibrometry inside the apex of the mouse cochlea using spectral domain optical coherence tomography," Biomed. Opt. Express **4**(2), 230–240 (2013).
20. S. S. Gao, R. Wang, P. D. Raphael, Y. Moayedi, A. K. Groves, J. Zuo, B. E. Applegate, and J. S. Oghalai, "Vibration of the organ of Corti within the cochlear apex in mice," J. Neurophysiol. **112**(5), 1192–1204 (2014).
21. H. Y. Lee, P. D. Raphael, J. Park, A. K. Ellerbee, B. E. Applegate, and J. S. Oghalai, "Noninvasive in vivo imaging reveals differences between tectorial membrane and basilar membrane traveling waves in the mouse cochlea," Proc. Natl. Acad. Sci. U.S.A. **112**(10), 3128–3133 (2015).
22. S. M. Khanna, J. F. Willemin, and M. Ulfendahl, "Measurement of optical reflectivity in cells of the inner ear," Acta Otolaryngol. Suppl. **467**(sup467 s467), 69–75 (1989).
23. S. M. Khanna, C. J. Koester, J.-F. Willemin, R. Daendliker, and H. Rosskothen, "Noninvasive optical system for the study of the function of inner ear in living animals," Proc. SPIE **2732**, 64–81 (1996).
24. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (Dover Publications, 1965) chap. 9.
25. O. Sasaki and H. Okazaki, "Sinusoidal phase modulating interferometry for surface profile measurement," Appl. Opt. **25**(18), 3137–3140 (1986).

Biomedical Optics EXPRESS

26. H. Hibino, Y. Horio, A. Inanobe, K. Doi, M. Ito, M. Yamada, T. Gotow, Y. Uchiyama, M. Kawamura, T. Kubo, and Y. Kurachi, "An ATP-dependent inwardly rectifying potassium channel, KAB-2 (Kir4. 1), in cochlear stria vascularis of inner ear: its specific subcellular localization and correlation with the formation of endocochlear potential," J. Neurosci. **17**(12), 4711–4721 (1997).
27. G. Ogata, Y. Ishii, K. Asai, Y. Sano, F. Nin, T. Yoshida, T. Higuchi, S. Sawamura, T. Ota, K. Hori, K. Maeda, S. Komune, K. Doi, M. Takai, I. Findlay, H. Kusuhara, Y. Einaga, and H. Hibino, "A microsensing system for the in vivo real-time detection of local drug kinetics," Nat. Biomed. Eng. **1**(8), 654–666 (2017).
28. https://www.niigata-u.ac.jp/contribution/research/policy/animal-experiment/
29. J. Na, W. J. Choi, E. S. Choi, S. Y. Ryu, and B. H. Lee, "Image restoration method based on Hilbert transform for full-field optical coherence tomography," Appl. Opt. **47**(3), 459–466 (2008).
30. https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Circular_Data_Analysis.pdf.
31. F. Chen, D. Zha, A. Fridberger, J. Zheng, N. Choudhury, S. L. Jacques, R. K. Wang, X. Shi, and A. L. Nuttall, "A differentially amplified motion in the ear for near-threshold sound detection," Nat. Neurosci. **14**(6), 770–774 (2011).
32. D. Zha, F. Chen, S. Ramamoorthy, A. Fridberger, N. Choudhury, S. L. Jacques, R. K. Wang, and A. L. Nuttall, "In vivo outer hair cell length changes expose the active process in the cochlea," PLoS One **7**(4), e32757 (2012).
33. N. P. Cooper and W. S. Rhode, "Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts," Hear. Res. **63**(1-2), 163–190 (1992).
34. F. Nin, T. Reichenbach, J. A. N. Fisher, and A. J. Hudspeth, "Contribution of active hair-bundle motility to nonlinear amplification in the mammalian cochlea," Proc. Natl. Acad. Sci. U.S.A. **109**(51), 21076–21080 (2012).
35. R. Leitgeb, C. Hitzenberger, and A. Fercher, "Performance of fourier domain vs. time domain optical coherence tomography," Opt. Express **11**(8), 889–894 (2003).
36. J. F. de Boer, B. Cense, B. H. Park, M. C. Pierce, G. J. Tearney, and B. E. Bouma, "Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography," Opt. Lett. **28**(21), 2067–2069 (2003).
37. http://www.mra.pt/repositorio/bf89/pdf/9859/2/fastcam-mini-ax200-tech-datasheet.pdf
38. S. Luo, T. Suzuki, O. Sasaki, S. Choi, Z. Chen, and J. Pu, "Signal correction by detection of scanning position in a white-light interferometer for exact surface profile measurement," Appl. Opt. **58**(13), 3548–3554 (2019).
39. M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," J. Opt. Soc. Am. **72**(1), 156–160 (1982).