# Long-range, wide-field swept-source optical coherence tomography with GPU accelerated digital lock-in Doppler vibrography for real-time, *in vivo* middle ear diagnostics

**DAN MACDOUGALL,[1,*] JOSHUA FARRELL,[1] JEREMY BROWN,[1,2] MANOHAR BANCE,[1,3] AND ROBERT ADAMSON[1,2]**

[1]*School of Biomedical Engineering, Dalhousie University, 5981 University Ave, Halifax NS, B3H4R2, Canada*

[2]*Department of Electrical and Computer Engineering, Dalhousie University, 1459 Oxford St, Halifax NS, B3H4R2, Canada*

[3]*Department of Surgery, Dalhousie University, 1276 South Park St, Halifax NS, B3H2Y9, Canada*

*\*dan.macdougall@dal.ca*

**Abstract:** We present the design, implementation and validation of a swept-source optical coherence tomography (OCT) system for real-time imaging of the human middle ear in live patients. Our system consists of a highly phase-stable Vernier-tuned distributed Bragg-reflector laser along with a real-time processing engine implemented on a graphics processing unit. We use the system to demonstrate, for the first time in live subjects, real-time Doppler measurements of middle ear vibration in response to sound, video rate 2D B-mode imaging of the middle ear and 3D volumetric B-mode imaging. All measurements were performed non-invasively through the intact tympanic membrane demonstrating that the technology is readily translatable to the clinic.

## References and links

1.  K. J. Cruickshanks, T. S. Tweed, T. L. Wiley, B. E. K. Klein, R. Klein, R. Chappell, D. M. Nondahl, and D. S. Dalton, "The 5-year incidence and progression of hearing loss: the epidemiology of hearing loss study," Archives of Otolaryngology–Head & Neck Surgery **129**, 1041–1046 (2003).
2.  J. A. Brown, Z. Torbatian, R. B. Adamson, R. Van Wijhe, R. J. Pennings, G. R. Lockwood, and M. L. Bance, "High-frequency ex vivo ultrasound imaging of the auditory system," Ultrasound in medicine & biology **35**, 1899–1907 (2009).
3.  J. J. Rosowski, R. P. Mehta, and S. N. Merchant, "Diagnostic utility of laser-Doppler vibrometry in conductive hearing loss with normal tympanic membrane," Otol Neurotol **24**, 165–175 (2003).
4.  S. E. Voss and J. B. Allen, "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384 (1994).
5.  M. P. Feeney, B. Stover, D. H. Keefe, A. C. Garinis, J. E. Day, and N. Seixas, "Sources of variability in wideband energy reflectance measurements in adults," Journal of the American Academy of Audiology **25**, 449–461 (2014).
6.  W. Chien, J. J. Rosowski, M. E. Ravicz, S. D. Rauch, J. Smullen, and S. N. Merchant, "Measurements of stapes velocity in live human ears," Hearing Research **249**, 54–61 (2009).
7.  H. H. Nakajima, M. E. Ravicz, S. N. Merchant, W. T. Peake, and J. J. Rosowski, "Experimental ossicular fixations and the middle ear's response to sound: evidence for a flexible ossicular chain," Hearing Research **204**, 60–77 (2005).
8.  C. Pitris, K. T. Saunders, J. G. Fujimoto, and M. E. Brezinski, "High-resolution imaging of the middle ear with optical coherence tomography: a feasibility study," Arch Otolaryngol **127**, 637–642 (2001).
9.  Z. Hubler, N. Shemonski, R. Shelton, G. Monroy, R. Nolan, and S. Boppart, "Real-time automated thickness measurement of the in vivo human tympanic membrane using optical coherence tomography," Quant Imaging Med Surg. **5**, 69–77 (2015).
10. H. R. Djalilian, J. Ridgway, M. Tam, A. Sepehr, Z. Chen, and B. J. Wong, "Imaging the human tympanic membrane using optical coherence tomography in vivo," Otol Neurotol **29**, 1091–1094 (2008).
11. M. Pawlowski, S. Shrestha, J. Park, B. Applegate, J. Oghalai, and T. TS, "Miniature, minimally invasive, tunable endoscope for investigation of the middle ear," Biomed. Opt. Express **6**, 2246–2257 (2015).

12. S. S. Hong and D. M. Freeman, "Doppler optical coherence microscopy for studies of cochlear mechanics," J.Biomed. Opt. **11**, 054014 (2006).

13. H. M. Subhash, A. Nguyen-Huynh, R. K. Wang, S. L. Jacques, N. Choudhury, and A. L. Nuttall, "Feasibility of spectral-domain phase-sensitive optical coherence tomography for middle ear vibrometry," J.Biomed. Opt. **17**, 0605051–0605053 (2012).

14. E. W. Chang, J. T. Cheng, C. Röösli, J. B. Kobler, J. J. Rosowski, and S. H. Yun, "Simultaneous 3d imaging of sound-induced motions of the tympanic membrane and middle ear ossicles," Hearing Res. **304**, 49–56 (2013).

15. J. Park, E. Carbajal, X. Chen, J. Oghalai, and B. Applegate, "Phase-sensitive optical coherence tomography using a vernier-tuned distributed bragg reflector swept laser in the mouse middle ear," Opt. Lett. **39**, 6233–6236 (2014).

16. J. Park, J. Cheng, D. Ferguson, G. Maguluri, E. Chang, C. Clancy, D. Lee, and N. Iftimia, "Investigation of middle ear anatomy and function with combined video otoscopy-phase sensitive oct," Biomed. Opt. Express **7**, 238–250 (2016).

17. D. MacDougall, J. Rainsbury, J. Brown, M. Bance, and R. Adamson, "Optical coherence tomography system requirements for clinical diagnostic middle ear imaging," J.Biomed. Opt. **20**, 56008 (2015).

18. M. Bonesi, M. Minneman, J. Ensher, B. Zabihian, H. Sattmann, P. Boschert, E. Hoover, R. Leitgeb, M. Crawford, and W. Drexler, "Akinetic all-semiconductor programmable swept-source at 1550 nm and 1310 nm with centimeters coherence length," Opt. Express **22**, 2632–2655 (2014).

19. Y. Jian, K. Wong, and M. V. Sarunic, "Graphics processing unit accelerated optical coherence tomography processing at megahertz axial scan rate and high resolution video rate volumetric rendering," J.Biomed. Opt. **18**, 026002 (2013).

20. K. Zhang and J. U. Kang, "Graphics Processing Unit-Based Ultrahigh Speed Real-Time Fourier Domain Optical Coherence Tomography," IEEE J. Sel. Top. Quantum Electron. **18**, 1270–1279 (2012).

21. D. Hillmann, T. Bonin, C. Lurhs, G. Franke, M. Hagen-Eggert, P. Koch, and G. Huttermann, "Common approach for compensation of axial motion artefacts in swept-source oct and dispersion in fourier-domain oct," Opt. Express **20**, 6761–6776 (2021).

22. R. Z. Gan, M. W. Wood, and K. J. Dormer, "Human middle ear transfer function measured by double laser interferometry system," Otol Neurotol **25**, 423–435 (2004).

23. W. F. Decraemer, O. de La Rochefoucauld, W. R. J. Funnell, and E. S. Olson, "Three-dimensional vibration of the malleus and incus in the living gerbil," Journal of the Association for Research in Otolaryngology: JARO **15**, 483–510 (2014).

24. M. S. M. G. Vlaming and L. Feenstra, "Studies on the mechanics of the normal human middle ear," Clinical Otolaryngology & Allied Sciences **11**, 353–363 (2009).

25. W. Drexler, M. Liu, A. Kumar, T. Kamali, A. Unterhuber, and R. A. Leitgeb, "Optical coherence tomography today: speed, contrast, and multimodality," J.Biomed. Opt. **19**, 071412 (2014).

26. W. Choi, B. Potsaid, V. Jayaraman, B. Baumann, I. Grulkowski, J. J. Liu, C. D. Lu, A. E. Cable, D. Huang, J. S. Duker, and J. G. Fujimoto, "Phase-sensitive swept-source optical coherence tomography imaging of the human retina with a vertical cavity surface-emitting laser light source," Optics Letters **38**, 338 (2013). 00041.

## 1.    Introduction

The middle ear is located at the end of the S-shaped outer ear canal, medial to the eardrum or tympanic membrane (TM), and houses three bones called the ossicles (malleus, incus and stapes) that act to efficiently conduct acoustic vibrations from the TM to the cochlea. Pathologies of the middle ear such as cholesteatoma, otitis media, otosclerosis and ossicular erosions lead to conductive hearing losses, a condition that affects hundreds of millions worldwide [1]. Conventional non-invasive imaging modalities such as computed tomography and magnetic resonance imaging lack the resolution to adequately resolve the fine anatomical structures of the middle ear [2]. Optical microscopy cannot image past the translucent tympanic membrane and so exploratory surgery is often needed to diagnose middle ear disorders with certainty [3]. In addition to imaging techniques, there are a number of functional diagnostic tools such as tympanometry and wideband immitance testing [4] that aim to measure the mechanical response of the middle ear by applying sound with a known pressure and observing the reflectance or motion of the ear in response. Unfortunately, these techniques have proven to be of very limited diagnostic value for most conditions [5] because the measurements are dominated by the mechanical properties of the tympanic membrane which are usually not well correlated to hearing loss. In cadaveric studies, laser Doppler vibrometry (LDV) has been widely used to obtain accurate, spatially resolved vibration measurements of the middle ear response to sound. However, LDV lacks depth resolution and so can only accurately measure vibration if the vibrating structure is the only strong reflector upon which the laser is incident. This requires that the specimen be surgically

prepared so as to expose the structure of interest in order to obtain a vibration measurement. While LDV has seen limited deployment as a tool for *in vivo* functional diagnostics [3], it can only measure structures exposed at the surface of the eardrum (i.e. the umbo) or during surgery when the middle ear is exposed [6]. The vibration of these surface structures is not tightly correlated to the vibration of deeper structures due to the compliance of the ossicular chain [7], and so provides less diagnostically useful information than would be obtained from a direct vibration measurement of deeper structures [3].

Optical coherence tomography is a rapidly evolving imaging modality that uses optical interferometry to obtain depth-resolved images of tissue. Feasibility studies in cadaveric temporal bones have demonstrated the potential of OCT as a tool for imaging the middle ear [8] and there has been some clinical investigation of its use in diagnosing certain pathologies of the middle ear through imaging of the TM [9, 10], and for investigating middle and inner ear physiology in animal models [11, 12]. To date, however, in vivo OCT systems have had very limited depth range of a few millimeters and have not been able to image deep middle ear structures such as the stapes and cochlear promontory.

A key recent development has been the application of phase-resolved OCT Doppler vibrography (OCT-DV) to study the vibration of middle ear structures in response to sound and the demonstration that OCT-DV may be able to diagnose certain functional pathologies that are not evident in purely anatomical images [13–16]. To date, however, these techniques have not been demonstrated in live subjects, either animal or human. OCT-DV offers a nearly ideal tool for functional imaging of the middle ear as it provides a non-invasive means of obtaining spatially-resolved vibrational responses. Such diagnostic information has the potential for greatly improved sensitivity and specificity as compared to current approaches since most middle ear disorders create fixations or discontinuities at specific locations within the ossicular chain.

Translating OCT middle ear imaging to the clinic involves a number of challenges [17]. Given the natural curvature and oblique orientation of the human TM relative to the ear canal, and allowing for normal anatomical variability between patients, a practical system must be capable of imaging over a depth range exceeding 10 mm in order to capture the entire middle ear; a range that has not been achievable with spectral-domain systems used in earlier studies [13, 16]. In addition, the optical losses incurred when imaging through the intact TM result in a 15-20 dB reduction in SNR relative to imaging without the TM [17], so high sensitivity and dynamic range are required. Ear canals are curved and so obtaining a clear line of sight into the middle ear can be challenging in certain individuals and requires careful optical design. Furthermore, the signal processing required to perform imaging and Doppler processing in real time represents a significant computational challenge. Other authors have noted the lack of a sufficiently fast Doppler processing engine as a key barrier to clinical deployment of the technology [16].

The present study describes an OCT imaging system capable of non-invasive, real-time 2D, 3D and Doppler mode imaging of the middle ear in live, awake subjects. Our system generates B-mode images at a nominal frame-rate of 20 frames per second (FPS), and can acquire Doppler lines at 0.2 lines per second at a safe sound stimulus presentation level of $< 100 \text{dB}_{\text{SPL}}$ for 5 seconds. Real-time Doppler processing is performed using a digital lock-in detection technique implemented on a graphics processing unit (GPU) in parallel with standard B-mode image processing. The system has been deployed clinically and we present imaging and Doppler data obtained with the system in live subjects including a patient with an ossicular prosthesis. This is, to our knowledge, the first report of OCT imaging of the entire middle ear volume and of middle ear OCT-DV in live subjects.

## 2. Methods

### 2.1. OCT engine design

Our imaging system, shown diagrammatically in Figure 1, is built around a Vernier-tuned distributed Bragg-reflector (VT-DBR) akinetic swept laser (Insight Photonics Solutions, Model: SLE-101, central wavelength $\lambda_0$ = 1550 nm, tuning bandwidth $\Delta\lambda$ = 40 nm, nominal repetition rate $f_s$ = 100 kHz, nominal power $P_0$ = 20 mW). This laser's fast sweeping and extremely long coherence length $l_c > 200$mm makes imaging of the full depth of the human middle ear possible in real-time. The laser's wavelength repeatability (RMS wavelength jitter < 20 pm) and the resulting low interferometric phase noise [18] that it provides enables phase-sensitive OCT-DV measurements with vibrational sensitivity near the shot-noise detection limit [15]. Because of these properties, akinetic swept lasers are very well suited to OCT imaging of the human middle ear and to performing functional vibrography in response to a sound stimulus.
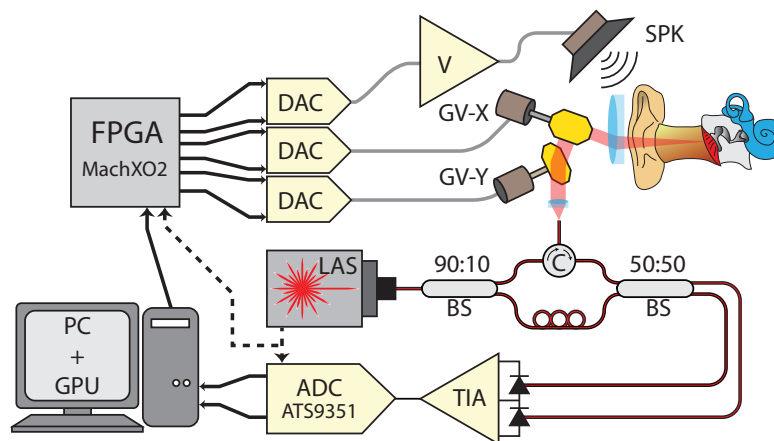


Fig. 1. Diagram of the SS-OCT system designed for real-time human middle ear imaging *in vivo*. FPGA: field programmable gate array, DAC: digital-to-analog converter, ADC: analog-to-digital converter, LAS: wavelength swept laser, TIA: trans-impedance amplifier, V: voltage amplifier, SPK: speaker, GV: galvanometer mirror, BS: fiber beam splitter, PC: personal computer, GPU: graphics processing unit.

A Mach-Zehnder fiber interferometer was designed for use with this laser and a Thorlabs PDB-470C balanced detector. 90% of the laser output is directed to the sample and a circulator directs the collected backscatter for 50:50 interference with the remaining 10% in the reference fiber. The measured reference arm power is 300$\mu$W at each detector element, which is adequate for shot noise to dominate over receiver noise. With the laser configured to output 13mW, after optical losses a total of 4.2 mW of power is incident on the eardrum. The interference signal is sampled at 12 bits and 400 MSPS on a high-speed data acquisition card (ATS9351, Alazar Technologies), with the acquisition card clocked and triggered directly from the laser's sample-clock and sweep-start signals.

### 2.2. Synchronization of the akinetic VT-DBR laser

The SLE-101 operates in a free-running mode and cannot be triggered externally. During sweeping operation, the laser generates light of stepwise-linear increasing optical frequency in time, but containing a number of brief mode-hopping events during which the laser behaves unpredictably. Since the timing of these events is highly repeatable from sweep to sweep and the

affected time-points can be identified during calibration, the corresponding invalid data samples can be removed from the collected interferograms and the piece-wise segments of the desired linear sweep can be stitched together [18]. After self-calibration, the VT-DBR laser reports the locations of the erroneous data in the sweep over a serial communication interface, and since the stitching produces an effectively linear optical frequency sweep, re-sampling of the measured signal is not required.

In order to maintain phase locking between the laser frequency-sweeping, the beam-scanning optics and the generation of an acoustic tone stimulus, we generate drive signals for the s-peaker and scanning mirrors using a set of digital-to-analog converters controlled by a field-programmable-gate-array (FPGA, MachXO2 by Lattice Semiconductor) that is clocked directly by the sweep-start signal output of the laser. This approach provides system-wide synchrony and avoids phase errors associated with the invalid, mode-hopping sample points. The laser's 100kHz sweep clock is sufficiently fast for adequate sampling of acoustic tones at any frequency in the human range of audibility (20Hz to 20kHz).

Because of the presence of invalid points in the VT-DBR laser sweep, more samples must be collected than will be used to form an A-line. As a result, the effective sweep rate is reduced by the fraction of invalid samples in a particular calibration run. Because the laser sweeping controls all timing in the system, the line-rate, frame-rate and acoustic stimulus frequency are all reduced by this same factor. This effect increases with sweep-rate as larger fractions of the sweep are spent in invalid states. In the two SLE-101 units we have used in this system, we observed typical frame-rate reductions of 11% and 18% at full speed. Additionally, in order to guarantee reliable data acquisition by the ATS-9351, 256 sweep points are ignored at the end of each sweep. While this has no effect on line or frame-rate, it reduces the number of useful points in the measured spectral interferograms, and represents a larger fraction of the total sweep at higher sweep-rates.

For simplicity in what follows we ignore these effects and give the nominal values of sweep-rate, line-rate and frame-rate that would be obtained if all the samples were valid. In reality, though, these parameters all change from one calibration run to the next and are always less than their nominal values.

## 2.3. Extraction of Doppler vibrographic information

In SS-OCT of biological tissue, the measured optical phase $\psi(x, y)$ at any location $(x, y)$ in the B-mode image is random owing to the random micro-structure of the sample. However, when the sample is set into periodic motion at an acoustic frequency $f_a$, the phase at each location oscillates periodically about the value it has when the sample is stationary. Neglecting noise, the optical phase measured at time $t$ at a spatial location $(x, y)$ of the image is given by:

$$\psi(x, y, t, f_a) = \frac{4\pi A(x, y, f_a)}{\lambda_0} \sin\left[2\pi f_a t + \phi(x, y, f_a)\right] + \psi_0(x, y) \tag{1}$$

where $A(x, y, f_a)$ is the displacement amplitude of acoustic vibration at frequency $f_a$, $\phi(x, y, f_a)$ is the acoustic phase shift of the reflector at $f_a$, and $\psi_0(x, y)$ is the optical phase that would be measured with the sample at rest.

If it can be assumed that motion associated with the acoustic stimulus occurs on time scales much longer than the sweep time of the laser, then all structures are approximately stationary during each sweep. Under this assumption, time $t$ in Equation 1 can be discretized into the number $i$ of sweeps that have elapsed since $t = 0$. We index each B-mode image frame by $w$ and each of the $X$ A-lines in the image within each B-mode frame by $x$. Each A-line contains $Y$ pixels, indexed by $y$, and is generated from $N$ laser sweeps, indexed by $n$. The index number of the current laser sweep $i$ is then given by

$$i = (n + xN + wXN) \tag{2}$$

The amount of acoustic phase in radians accumulated during a single laser sweep is

$$\Delta\phi = 2\pi f_a T_s \tag{3}$$

where $T_s = 1/f_s$ is the laser sweep period. The total acoustic phase accumulated since the measurement began at time $t = 0$ and $(w, x, n) = (0, 0, 0)$ is therefore:

$$\phi_i = 2\pi f_a i T_s \tag{4}$$

$$\phi_{wxn} = 2\pi f_a \left(n + xN + wXN\right) T_s \tag{5}$$

The measured optical phase, *i.e.* the phase of the discrete Fourier transform (DFT) of the spectral interferogram of a vibrating reflector located at pixel $y$ of A-line $x$ is therefore given by:

$$\psi_{wxyn} = \frac{4\pi A_{xy}(f_a)}{\lambda_0} \sin\left[2\pi\frac{f_a}{f_s}(n + xN + wXN) + \phi_{xy}(f_a)\right] + \psi_{0,xy} \tag{6}$$

$A_{xy}(f_a)$ and $\phi_{xy}(f_a)$ represent the 2-D magnitude and phase of the steady-state vibrational response of the pixel. Both are of diagnostic significance for middle ear structures and can be extracted from the measured optical phase data by cross-correlating $\psi_{wxyn}$ with a phasor oscillating at the acoustic frequency $f_a$, which can be written in terms of the indices $w$, $x$ and $n$ as $e^{j\phi_{wxn}} = e^{j2\pi(n+xN+wXN)f_aT_s}$. If a total of $W$ image frames contribute to the extracted vibration level, our estimation can be thought of as the computation of the coefficient of the $W \times N$-pt DFT of $\psi_{wxyn}$ at the frequency corresponding to the applied stimulus tone. By separating the Fourier sum into two sums over $w$ and $n$, the estimate of $A_{xy}(f_a)$ and $\phi_{xy}(f_a)$ can be written in the following form:

$$A_{xy}(f_a)e^{j\phi_{xy}(f_a)} = \frac{\lambda_0}{4\pi}\frac{1}{NW}\sum_{w=0}^{W-1}\sum_{n=0}^{N-1}e^{-j2\pi(n+xN+wXN)f_aT_s}\psi_{wxyn} \tag{7}$$

Equation 7 is well suited for evaluation in parallel on a GPU since the sums over $n$ can be carried out in parallel. This digital lock-in approach can be used for arbitrarily long continuous averaging of the OCT-DV data in real-time, providing a narrow effective detection bandwidth. Because phase-resolved Doppler measurements use the same interferogram data as is used to generate B-mode images, the processing approach described here is amenable to real-time, simultaneous OCT-DV and B-mode imaging. The approach is also readily extensible to three dimensions for volumetric OCT-DV imaging.

## 2.4. *Real-time accelerated signal processing on a graphics processing unit*

GPU acceleration has previously been used to process OCT B-mode data [19, 20], but to our knowledge this represents its first use for real-time OCT-DV processing. We implemented the processing in NVidia's parallel programming platform, the Compute Unified Device Architecture (CUDA) which allows arrays of thousands of threads of execution (called a *grid* in CUDA) to execute simultaneously, providing a massive speed-up for appropriately formulated, highly parallel algorithms. In general, processing throughput with CUDA can be limited either by the number of operations being carried out or by the memory bandwidth of the device, *i.e.* the device's ability to read and write data to and from its own internal memory. Execution speed-up is achieved by identifying and leveraging the parallelism in the problem, by minimization of redundancy, and through carefully orchestrated memory access, *e.g.* coalescing logically adjacent threads in the grid to access physically adjacent memory locations simultaneously. We now describe the logical implementation of our CUDA OCT-DV processing kernels.

### 2.4.1.   Pre-discrete-Fourier-transform processing

Our processing pipeline incorporates a normalization stage that can be used to flatten systematic amplitude ripple due to wavelength-dependent transmission of optical components by scaling the collected interferograms by a pre-measured amplitude profile. This normalization can be helpful for suppressing unwanted broadening of the axial point-spread function (PSF) due to amplitude errors in the interferogram. Previously demonstrated numerical dispersion compensation capabilities [21] have also been incorporated that apply a complex multiplication vector of phase correction factors to the interferogram. This phase vector can be made to conjugate unwanted wavelength-dependent phase accumulations between the two arms of the interferometer that arise due to dispersion imbalance. This normalization can prevent unwanted broadening of the PSF due to phase errors in the collected interferogram. Following these normalizations of amplitude and phase, the resulting data is then multiplied by a Hamming window function prior to calculating the DFT in order to control dynamic range and PSF width in the image. Since the normalization steps (ripple-rejection, dispersion-compensation, windowing) are all multiplicative in nature, by associativity, they can be pre-multiplied into a single complex normalization vector and stored on the GPU.

These processing steps can be applied efficiently by grouping them into a single kernel of execution to minimize memory accesses and by structuring the CUDA grid of threads to maximize coalesced memory access. Since the VT-DBR's raw interferograms must first be stitched to eliminate invalid data in a way that depends on the laser's most recent calibration, the unpredictable pattern of memory accesses prevents perfectly coalesced reading of the raw interference pattern into the grid. However, since invalid data tends to be sparsely distributed with each sweep consisting of long sections of valid points interspersed with short blocks of invalid ones, sufficient data locality remains to get significant speed improvement through coalescing.

In CUDA we implement a 1D grid of threads where each thread corresponds to a single valid point in the laser sweep, as shown diagrammatically in Figure 2A. When operating at maximum speed, our system acquires interferograms containing 4000 valid samples, and so uses a grid of 4000 threads to execute the kernel. With the sweeps being sampled at 400MHz, this corresponds to a nominal sweep rate of $f_s$ = 100kHz. Refering to Equation 7, each of the 4000 threads loops across and operates on $N$ laser sweeps to collectively produce a single A-line in the final image, *i.e.* data is transferred and pre-processed on the GPU in single-A-line-sized groups. Writing of the result back to global memory is also coalesced and efficient.

This approach is compatible with using CUDA streams for concurrent data transfer and execution for pre-processing of many A-lines simultaneously. However, the Microsoft Windows drivers for our GPU (NVidia GTX660) lacked the ability to disable batched dispatching of asynchronous CUDA operations making it impossible to take full advantage of CUDA streams. Use of a newer GPU or the NVidia Linux driver would lift this limitation.

Following pre-processing, an optimized single-precision, complex-to-complex Fast-Fourier-Transform (FFT) is executed using the CUDA library, cuFFT.

### 2.4.2.   Post-discrete-Fourier-transform processing

Following the FFT, we simultaneously perform sweep averaging for SNR improvement in B-mode imaging and extraction of OCT-DV information through digital lock-in detection. Throughput is maximized by minimizing and coalescing global memory accesses, and by grouping B-mode averaging and Doppler processing into a single kernel of execution. This is achieved by creating a 2D grid of CUDA threads executed in a single processing stream where each thread corresponds to a single pixel in the final OCT image, as shown in Figure 2B. Simultaneous transfer and execution can also be done using a separate processing stream for each A-line and a thread of execution for each pixel in the final image.

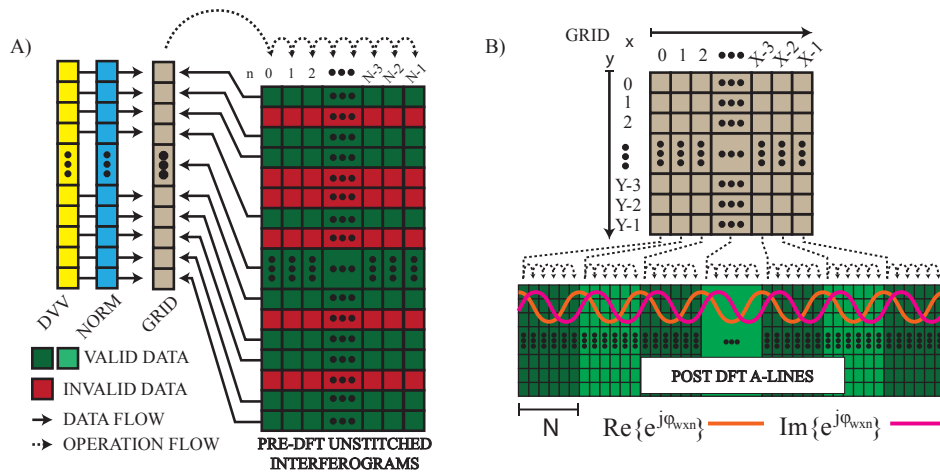A key feature of this approach is that it enables efficient calculation of Equation 7. After

Fig. 2. Conceptual diagram describing the memory and grid structure of the CUDA streams used for A) pre- and B) post-DFT processing the raw unstitched interferometric data. DVV: data valid vector describing the pattern of invalid data to be ignored in the spectral interferograms, NORM: complex normalization vector used for application of ripple-rejection, dispersion compensation and windowing prior to discrete Fourier-transformation.

single-precision floating point calculation of the phase angles $\psi_{wxyn}$ and application of a non-iterative phase unwrapping algorithm, the phasor $e^{j\phi_{wxn}}$ can be generated inside each thread from the grid parameters $(w, x, n)$. This way, $A_{xy}(f_a)e^{j\phi_{xy}(f_a)}$ can be estimated for all A-lines simultaneously. In a pattern similar to that used during pre-processing, each thread loops across $N$ sweeps and accumulates the B-mode average and OCT-DV results into efficiently accessed registers, and only the final results are written to global memory before transfer to the CPU.

The selection of the parameter $N$, *i.e.* the number of laser sweeps contributing to each A-line for a particular imaging configuration is an important one. Obviously it sets the frame rate according to $FPS = f_s/NX$, but its impact on phase calculations is less apparent. Phase unwrapping is an inherently serial operation since every value's required phase correction term depends on all previous values and their associated phase correction terms, making it a classically difficult problem to solve efficiently on a GPU. In the current system, this problem is avoided by always choosing the period of the acoustic stimulus to be $N$ sweeps in duration. By doing so, each serial thread need only unwrap the small subset of phases that it processes without additional memory accesses. This ensures that whole acoustic cycles are always accurately phase-unwrapped. Any spurious jumps that remain between acoustic periods have no impact on the OCT-DV output. This also implies that the effective A-line rate during OCT-DV measurements, $f_s/N$, cannot be made shorter in duration than a single period of the acoustic frequency of interest. However, any acoustic frequency that produces an integer multiple of whole acoustic cycles in $N$ sweeps can readily be accommodated with this approach. The set of allowable acoustic frequencies are defined by $f_a = f_s m/N$ where $m$ is an integer. Finer frequency resolution could potentially be obtained by using a clock multiplier to clock the FPGA at a frequency of $kf_s$, which would then allow frequencies of $f_a = f_s mk/N$ to be used. We typically perform OCT-DV measurements at octave separated frequencies below $f_a = 10$kHz (N=320, 160, 80, 40, 20, 10 produces nominal acoustic frequencies of 312.5Hz, 625Hz, 1.25kHz, 2.5kHz, 5kHz and 10kHz when $f_s = 100$kHz).

It is the combination of the system's synchrony, the careful placement of measurement data into memory on the GPU, and the appropriate structuring of the CUDA threads to have correspondence between grid structure and the indices $(w, x, y, n)$ that makes real-time B-mode and OCT-DV processing calculations fast and efficient. The architecture allows us to perform continuous real-time processing of B-mode and OCT-DV data with 4000-point sweeps at the laser's maximum sweep rate of $f_s = 100$kHz and 100% duty on a commercial-grade GPU.

The accuracy of the system was determined by imaging the vibration of a B&K 8001 impedance head/accelerometer mounted on a B&K 4810 minishaker. The vibration amplitude measured by OCT-DV agreed with the accelerometer reading to within 0.3dB over an acoustic frequency range from 250Hz to 8kHz. When the phase of vibration of the minishaker is tuned over $360°$, the detected phase of vibration is accurately tracked by OCT-DV and agrees to within $0.5°$.

### 2.5.  Patient-compatible scanning optics

*In vivo* imaging in human middle ears requires accommodation of the natural curvature of the external ear canal, which tends to limit line-of-sight to the eardrum from outside the head. While a pure endoscopic solution could provide adequate field-of-view (FOV) at the eardrum, endoscopes require sterilization between patients and may suffer from motion artifacts if they are handheld – a particularly serious concern for Doppler imaging. In contrast, a mounted surgical microscope solution can provide excellent imaging stability, but has a poor FOV.

We have addressed these tradeoffs by adopting a hybrid approach to the system's optical design. By imaging down the ear canal through standard 4 mm disposable plastic specula with a custom microscope mounted on the articulating arm of a Zeiss OPMI3 surgical microscope, we benefit from hands-free stability while avoiding the sterility problem. Our scanning microscope, depicted in Figure 3, consists of a pair of orthogonal galvanometer mirror scanners and a three-lens system of off-the-shelf achromatic doublets from Thorlabs. Lenses L1 (AC254-050) and L2 (AC508-100) operate roughly as a pupil relay, and L3 (AC254-060) as an objective. Adjustment of element separation D2 provides a means to vary the depth of best-focus within the sample. Adjustment of element separation D3 provides a means to adjust the maximum fan angle of the scan pattern. A wide FOV capturing the entire eardrum ($\approx 10mm \times 10mm$) is achieved by placing a pupil image (*i.e.* an image of the deflecting mirrors) near the speculum tip. The numerical aperture (NA) of the system is made small ($\approx 0.022$) in order to extend depth-of-field along the middle ear's axial length. At this NA and wavelength, diffraction-limited lateral resolution is acceptable for middle ear imaging at $\frac{\lambda}{2nNA} \approx 35\mu$m, which is an order of magnitude better than CT or MRI.

A custom clamp was machined to allow changing of the speculum between patients. The sound stimulus was presented to the ear through a tube speaker (Etymotic Research, ER3A) and fed into the airspace within the speculum. The sound pressure level was monitored using a calibrated tube microphone (Etymotic Research, ER7C-Series B).

## 3.   Results

### 3.1.   Ex vivo imaging

A cadaveric human temporal bone, obtained fresh-frozen, was prepared for OCT-DV imaging to demonstrate the system's potential for applications in fundamental research in hearing physiology. We partially drilled away the ear canal to facilitate alignment of the system for imaging. Our 40nm optical bandwidth around 1550nm in air results in an axial pixel-length of $\approx 30\mu$m, thus requiring only 10mm/$30\mu$m $\approx 330$ post-DFT samples in order to produce images of the desired 10mm in length. In configuring the system for imaging a trade-off must be made between A-line density, Doppler sensitivity at each A-line, and total measurement time. For this experiment,
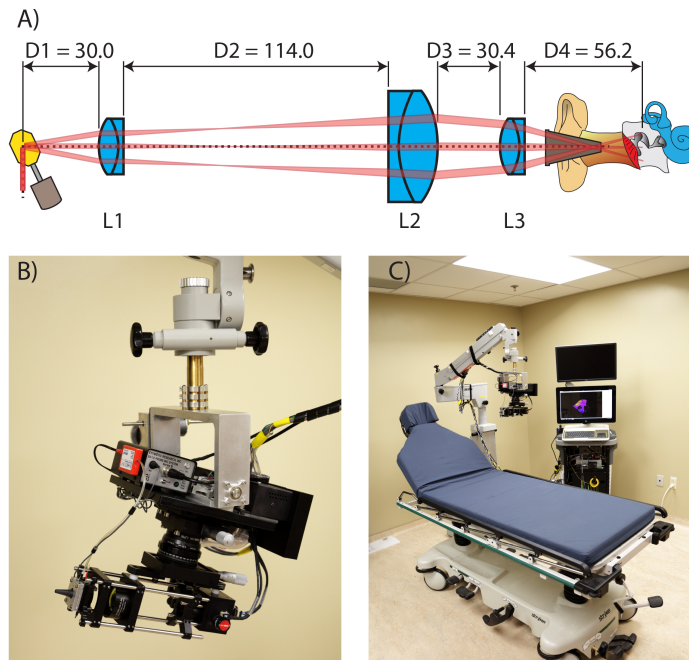
Fig. 3. A) Optical layout B) Closeup of the middle ear OCT scanning microscope used for imaging down the ear canal through a 4 mm otoscopic speculum and mounted to surgical microscope arm. C) Complete in-clinic, real-time imaging system mounted to an articulating arm. Units are in mm.

we configured the system for $128 \times 330$ pixel 2D B-mode imaging at 20 FPS ($X = 128$ in Equation 7), for $128 \times 128 \times 330$ voxel 3D B-mode imaging at $\approx 0.15$ volumes per second (VPS), and for excitation and detection of vibrations at 515Hz (nominal $f_a = 625$Hz, $N = 160$ in Equation 7). We then performed 2D OCT-DV imaging by applying digital lock-in detection over 128 acoustic cycles at each A-line to acquire images like those shown in Figure 4A and 4B, showing the cross-sectional functional response of the middle ear measured transtympanically with our system. These 2D images were acquired over $\approx 33s$. Similarly, a full 3D functional OCT-DV volume was collected and processed over a total acquisition time of 70 minutes, which is simply the cumulative duration of 128 acoustic cycles over the $128 \times 128 = 16,384$ A-lines in the volume. Figures 4C and 4D show a full 3D render of the excited response before and after digital removal of the TM, respectively (see Visualization 1). The measured vibration levels are consistent with those previously reported in LDV studies of temporal bones [22]. Using the same dataset, we generated an animation of the complete vibration response of the ear with peak-to-peak displacements magnified by $4,000\times$ and slowed in time by $1,000\times$ shown in Visualization 2. The relative phases of motion of all structures in the volume are preserved and accurately depicted. Since we can only measure the velocity component along the optical axis of the system the motion is only shown along this direction.

### 3.2. In vivo imaging

Figure 5A shows the wide-field, long-range view of the middle ear as viewed through the TM from the intact ear canal, *in vivo*. The images of Figure 5 were collected with the system configured for $256 \times 330$ pixel 2D imaging at 20 FPS and $256 \times 256 \times 330$ voxel 3D imaging at $\approx 0.08$ VPS ($\approx 12.5s$ per volume). Visualization 3 demonstrates the ability to observe the
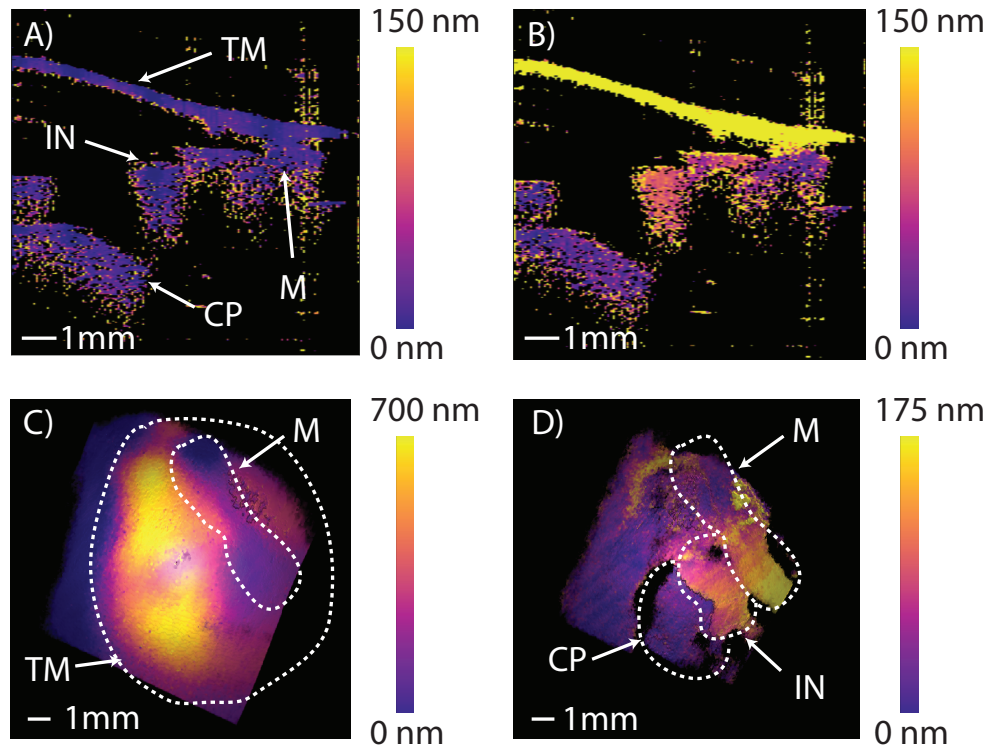
Fig. 4. *Ex vivo*, Doppler-vibrographic response of a cadaveric right ear at an acoustic frequency of 515Hz showing a 2D view of the ear's measured peak-to-peak vibrational response A) without acoustic stimulus, and B) with stimulus applied at 100dB$_{SPL}$. C) and D) show the color-mapped vibrational response to stimulus in 3D before and after digital removal of the TM (see Visualization 1). Visualization 2 uses the same data to animate the middle ear's vibrational response. Tympanic membrane (TM), malleus (M), incus (IN), cochlear promontory (CP).

real-time macroscopic dynamics of the middle ear in 2D using video-rate B-mode OCT imaging. Between times 0:16s and 0:21s the outward bulge of the TM due to pressurization of the middle ear cavity by a Valsalva maneuver (pinching of the nose and popping of the ears) is clearly seen, as well as the subsequent depressurization that occurs when the subject swallows. 3D volume renders of the entire live middle ear can be seen in Figures 5B, 5C, and Visualization 4.

Another particularly promising application of our system's wide-field, long-range anatomical imaging capabilities lies in post-operative tracking of middle ear prosthetics. Prosthetic ossicles, which are surgically implanted to restore function in ears afflicted by various forms of conductive hearing loss, can migrate over time and eventually fail. Figure 6 and Visualization 5 show anatomical images of a patient's middle ear acquired during a post-operative check-up visit following a stapedotomy surgery to treat otosclerosis. This is a procedure where the stapes, which has been fixed in place by plaque formation over the footplate, is excised and replaced in function by a piston-like prosthetic that is crimped onto the incus and fed into the inner ear, as depicted in Figure 6A. Figure 6B shows a 3D render of the patient's ear with the crimp of the prosthetic appearing as a prominent bulge wrapped around the long process of the incus.
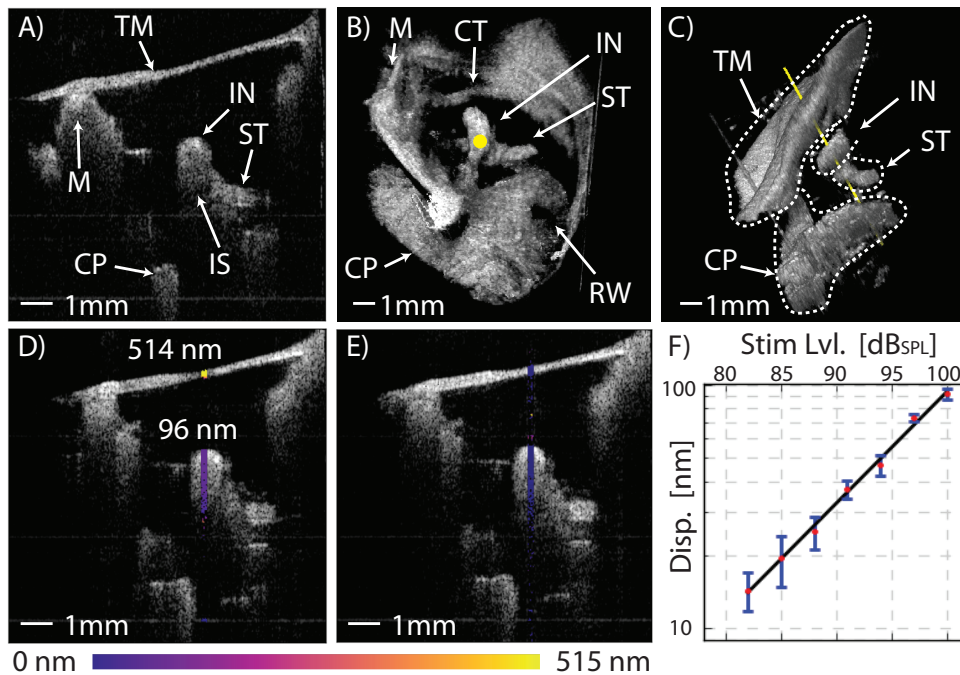
Fig. 5. *In vivo*, real-time functional imaging of a normal left ear's response at 1030Hz. A) Shows a $1 \times 1 \text{cm}^2$ 2D cross-section of the middle ear in the transverse plane. Visualization 3 shows the macroscopic changes to the ear anatomy during a Valsalva maneuver. B) Shows a $1 \times 1 \times 1 \text{cm}^3$ 3D volume render of the middle ear as seen from the perspective of the ear canal with the TM digitally removed (see Visualization 4), and C) from an inferior-posterior perspective with the TM in-place showing the axis of Doppler measurement along the yellow line passing through the incus at the stapedius tendon. Functional measurements of the TM and incus' peak-to-peak vibrational response at 1kHz are shown in D) with a $100\text{dB}_{\text{SPL}}$ tone applied to the ear and E) without stimulus. F) A plot of displacement response versus sound pressure level showing excellent linearity from $80\text{dB}_{\text{SPL}}$ to $100\text{dB}_{\text{SPL}}$. Error bars represent $\pm$ one standard deviation of the response over the pixels along the axial length of the incus. Tympanic membrane (TM), malleus (M), incus (IN), incudo-stapedial joint (IS), stapedius tendon (ST), chorda tympani nerve (CT), cochlear promontory (CP), round-window niche (RW).

Figures 6C and 6D show 2D transverse cross-sections through the patient's incus with the bone and titanium prosthetic having starkly different appearances. In this patient the implant was well fixed and had not migrated.

When a live subject is imaged with OCT-DV, motion generated due to heartbeat, breathing and muscle movement induce phase disturbances in the collected interferograms that can dwarf the physiological vibrations of interest by several orders of magnitude. While these unwanted macroscopic motions are slow and well-separated in frequency from the acoustic response, when combined with the random micro-structure of the biological tissues, they produce phase noise that extends into the acoustic frequency band, significantly degrading Doppler phase sensitivity as compared to cadaver experiments.

Fortunately, because the ossicles move as rigid bodies at acoustic frequencies [23], it is possible to apply spatial averaging through the depth of the ossicle to significantly reduce the

motion-induced noise without compromising imaging speed. Segmentation of the ossicles to implement spatial averaging is straightforward owing to the high contrast between the ossicles and the surrounding air. Typically, with no acoustic stimulus applied to the ear, we are able to consistently average down to an apparent vibration < 5nm in 5s of temporal averaging in the kHz range, which we take as our effective noise floor in vivo. This may be compared against the 1nm noise floor we achieve with just 50ms of averaging in the kHz range when imaging stationary objects in the same setting.

The procedure we adopted for imaging live subjects is as follows. Using real-time B-mode imaging, we align the microscope in the patient's ear canal to obtain a preliminary assessment of the anatomy. We then capture a 3D volumetric scan of the ear and render it onscreen with the eardrum digitally removed to unveil the locations of the ossicles, nerves and tendons in the middle ear cavity. Based on the anatomical landmarks, individual lines are selected within the 3D volume for performing OCT-DV measurements. We collect Doppler data for 5s along each selected line, obtaining the vibrational response for every pixel along the line. The resulting vibration magnitude is displayed on the screen as a co-registered colour overlay on the B-mode image to provide immediate feedback to the clinician. If results appear satisfactory, spatial averaging can be applied in post-processing to improve the SNR of the displacement measurements for each ossicle.

This procedure was applied to obtain the live patient Doppler data shown in Figure 5. With the system configured to excite and detect vibrations at 1030Hz (nominal $f_a = 1250$Hz, $N = 80$ in Equation 7) we measured the displacement of the structures lying along the yellow line shown in the 3D volume renders in Figures 5B and 5C. The colour-mapped Doppler results with and without a $100$dB$_{SPL}$ stimulus applied are shown in Figures 5D and 5E. The measurement was repeated at seven sound pressure levels between $80$dB$_{SPL}$ and $100$dB$_{SPL}$ to produce the plot in Figure 5F. The slope of the data shows good agreement with the 20dB/decade reference line on the plot indicating that the response of the middle ear is linear over the measured range of amplitudes, in agreement with measurements performed in cadavers [24].

## 4. Discussion

Because in our implementation, Doppler sensitivity is limited primarily by patient motion, the system could be improved by mitigating or eliminating its effects. Gating of the heartbeat and/or breathing are both possible, but will add additional complicating steps to the imaging process. Improved vibration isolation, while technically feasible, is difficult to implement in a clinical setting. Alternatively, if the laser sweep rate could be increased, lines could be collected faster which would reduce the effect of patient motion. Both VT-DBR [25] and MEMS-VCSEL [26] lasers offer phase stability high enough to allow Doppler imaging in architectures that are compatible with very fast line rates approaching 1 MHz. The GPU-based Doppler processing architecture presented here is readily scalable to higher speeds once faster phase-stable swept sources become available. Currently-available high-end GPUs could handle processing rates an order of magnitude higher than those needed for our current system.

## 5. Conclusions

We have demonstrated a SS-OCT system for use in clinical middle ear imaging in humans, and its performance is summarized in Table 1. The use of an akinetic VT-DBR swept laser allows for imaging of the full depth of the human middle ear volume and for OCT-DV mapping of the ear's dynamic response to acoustic tones. The presented architecture incorporates scalable GPU acceleration of the OCT-DV calculations, which we have shown achieves sufficient throughput for continuous 100% duty processing at full-speed sweeping with the SLE-101 VT-DBR swept-laser on a consumer-grade nVidia GTX660 GPU. The system was validated through imaging trials in a cadaveric middle ear specimen and in live subjects.
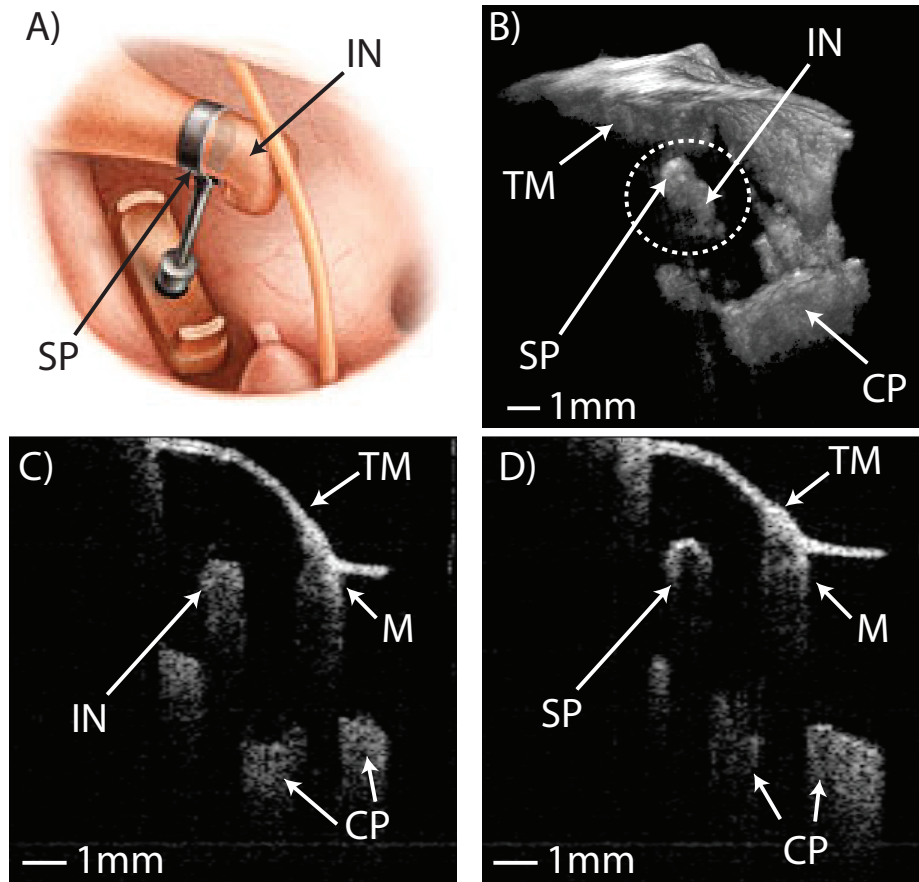
Fig. 6. A) Illustration of the placement of a stapes piston prosthesis during stapedotomy surgery (taken from "Practical Otology for the Otolaryngologist" by Seilesh Babu, 2013) B) *In vivo* 3D render of patient's middle ear containing a stapes piston prosthethis, clearly showing the crimp of the piston around the long process of the incus (see Visualization 5). C) and D) show real-time 2D B-mode images comparing the appearance of normal incus bone, characterized by gradual signal drop-off and multiple scattering, to the appearance of the titanium prosthesis, characterized by a single strong surface reflection. Tympanic membrane (TM), incus (IN), stapes piston prosthesis (SP), cochlear promontory (CP), malleus (M)

Table 1. Summary of current system capabilities

| Max. A-line rate[a] | 100kHz |
|---|---|
| Optical Bandwidth | 40nm@1550nm |
| 2D Frame Rate[a] | 20FPS |
| 3D Volume Rate[a,b] | 0.15VPS |
| Imaging Depth in Air | 10mm |
| Field of View | $10 \times 10$mm |
| Axial and Lateral Resolution | $< 40\mu$m |
| Vibrographic Noise Floor[c] | $< 1$nm |

[a] Based on nominal laser sweeping speed
[b] With 128 volume slices
[c] At 4kHz with $50ms$ of averaging