Miniaturized endoscopic 2D US transducer for volumetric ultrasound imaging of the auditory system

Lucas Lavenir, João Cavalcanti Santos, Nabil Zemiti, Akil Kaderbay, Frederic Venail, Philippe Poignet

Towards Minimally Invasive Volumetric Ultrasound Imaging of the Auditory System

Lucas Lavenir, João C. Santos, Nabil Zemiti, Akil Kaderbay, Frédéric Venail and Philippe Poignet

***Abstract*— In this paper, we focus on the carrying out and validation of minimally invasive three-dimensional (3D) ultrasound (US) imaging of the auditory system, which is based on a new miniaturized endoscopic 2D US transducer. This unique probe consists of a 18 MHz 24 elements curved array transducer with a distal diameter of 4 mm so it can be inserted into the external auditory canal. Typical acquisition is achieved by rotating such a transducer around its own axis using a robotic platform. Reconstruction of a US volume from the set of acquired B-scans during the rotation is then performed using scan-conversion. The accuracy of the reconstruction procedure is evaluated using a dedicated phantom that includes a set of wires as reference geometry. Twelve acquisitions obtained from different probe poses are compared to a micro-computed tomographic model of the phantom, leading to a maximum error of 0.20 mm. Additionally, acquisitions with a cadaveric head highlight the clinical applicability of this set up. Structures of the auditory system such as the ossicles and the round window can be identified from the obtained 3D volumes. These results confirm that our technique enables the accurate imaging of the middle and inner ears without having to deteriorate the surrounding bone. Since US is a real-time, wide available and non-ionizing imaging modality, our acquisition setup could facilitate the minimally invasive diagnosis and surgical navigation for otology in a fast, cost-effective and safe way.**

***Index Terms*— Auditory system, ultrasound imaging, volumetric reconstruction.**

I. Introduction

HEARING impairment is the third main cause of disability worldwide. In its last world report on hearing released in 2021, the World Health Organization (WHO) has estimated the prevalence of hearing loss to more than 1.5 billion people [1]. Since this number is expected to rise over 2.5 billion by 2050, hearing impairment has been elevated to a major public health issue. In parallel with the prevention of hearing loss, WHO consequently promotes the development of low-cost technologies for facilitating diagnosis and treatment, especially in low- and middle-income countries [2].

Overall, hearing loss can be categorized as either conductive or sensorineural [3]. The first one refers to pathologies that hinder sound transmission through the eardrum or the ossicular chain to the inner ear while the latter is typically responsible for permanent hearing impairment and results from a dysfunction of the cochlear sensory cells or the hearing nerve.

Despite the increased use of magnetic resonant imaging and computed tomography (CT) for the study of the temporal bone [4], such modalities lack resolution for accurate identification of middle and inner ear pathologies. In some conductive hearing loss cases, the accurate diagnosis can only be made during a surgical exploratory tympanotomy [5], which allows for a direct visualization of the tympanic cavity and the auditory ossicles [6]. To overcome those spatial resolution issues, conebeam CT (CBCT) has emerged in the ear, nose, and throat (ENT) area, providing an isotropic spatial resolution up to 0.150 mm, and is now commonly used for postoperative verification of cochlear implantation [7]. However, the high investment cost of the apparatus, the radiation hazards and the inability to provide functional information significantly limit the application of CBCT as a cost-effective modality for early diagnosis. From this perspective, optical coherence tomography (OCT) appears as a promising alternative. OCT is a safe modality based on white light interferometry that provides real-time rendering with a spatial resolution of a few tens of microns. Recently, it has gained interest for the imaging of both the middle [8] and inner ears [9], but its limited penetration depth in scattering media impedes its application for intraoperative imaging of the auditory system [10].

Although also limited in depth, ultrasound (US) imaging allows to image further than OCT and was even shown to evaluate the functionality of the ossicles through the eardrum [11]. Besides, depending on the applied frequency, US imaging is capable to achieve sufficient spatial resolution to visualize anatomical landmarks of the ear that range from the ossicular chain [11], [12] to cochlear structures such as the basilar membrane [13], [14]. Combined with the low-cost, ease of use and safety of the corresponding imaging systems, these features make US an appropriate modality for initial contact with the patient. Yet, the large housing of commercially available US transducers makes the use of US for the imaging of the ear inapplicable in clinical practice since the auditory system is

This work was supported by the French "Fondation pour l'Audition" (FPA RD-2017-4), the French Occitanie Region and European Council (RR19093FF FEDER found) within the Investissements d'Avenir Program (Labex CAMI, ANR-11-LABX0004, the Equipex ROBOTEX, ANR-10-EQPX-44-01). (Correspondence author: Nabil Zemiti.)

Lucas Lavenir, João C. Santos, Nabil Zemiti and Philippe Poignet are with the LIRMM, University of Montpellier, CNRS, Montpellier, France (e-mail: lucas.lavenir@lirmm.fr; joao.cavalcanti-santos@lirmm.fr; nabil.zemiti@lirmm.fr; philippe.poignet@lirmm.fr).

Frédéric Venail is with the INM, INSERM U105, Montpellier, France (e-mail: frederic.venail@inserm.fr). Frédéric Venail has been awarded of the Early Scientific Career FPA award.

Akil Kaderbay and Frédéric Venail are with the University Hospital Gui de Chauliac, University of Montpellier, Montpellier, France (e-mail: a-kaderbay@chu-montpellier.fr; f-venail@chu-montpellier.fr).

deeply anchored in the temporal bone. In fact, studies that reported *ex vivo* imaging of the middle ear actually required prior removal of the external auditory canal (EAC) [11], [12].

In this paper, we introduce a new miniaturized endoscopic 2D US transducer whose dimensions are compatible with the EAC. Since the anatomy of the middle and inner ears is one of the most complex of the human body, conventional 2D US imaging may actually result in a mental workload for the medical practitioner who must picture the organ of interest from various planar images. For instance, the operator in [11] had to visualize cross-sections from different angles to construct a mental representation of the imaged structures. A proper solution then is to resort to three-dimensional (3D) US imaging, which provides a better understanding of the anatomy and facilitates the decision-making process. Various technologies based on 2D transducers are reported in the literature, where the main dichotomy lies in the use of position sensing. Sensorless techniques, such as mosaicing, rely on the sole intrinsic contents of the acquired images to deduce their global alignement. Best results, however, come with a high computational cost due to the use of specific similarity measures [15]. On the contrary, sensor-based techniques consist in reconstructing a volume based on the relative pose of collected data. In freehand US, the localization of each B-scan is achieved using a tracking device, which leads to inaccuracies and a high computational cost in reconstruction due to the sparsity of acquired data [16]. In scan-conversion US, frame acquisitions are conversely performed using a motor stage that includes an encoder. By directly refering to the geometry of the probe, this technique thus benefits from a high degree of precision as well as a short reconstruction time, both of which are of interest for diagnosis purposes [17]–[19]. Synthetic aperture has also been studied for 3D US using mechanically translated [20] or rotated [21] 2D transducers since it increases both the resolution and acquisition rate. Yet, this method requires to access RF signals from the individual transducer elements, which is difficult to fulfill on commercial devices and may necessitate dedicated equipment [22]. In this paper, we present a new imaging platform that takes advantages of our newly miniaturized endoscopic 2D US transducer to provide minimally invasive volumetric US imaging of the auditory system using scan-conversion. The actuation of the probe is based on an existing robotic system called *RobOtol* that is dedicated to middle ear surgery. As for robot-assisted 3D US imaging [23], [24], the use of such a system enables both the precise and repetable positioning of the probe, which is of particular interest for the insertion along the EAC.

Since it is meant for medical purposes, we consider of primary importance to validate both quantitatively and qualitatively the proposed acquisition setup. As a result, this study additionally incorporates a two-fold validation stage. Firstly, we focus on assessing the correctness of the reconstructed volume's geometry compared to the reality of acquisition using a dedicated phantom. Beyond the lack of a clear methodology in the literature, a significant difficulty for quantifying the reconstruction error is due to the unusual small dimensions of the imaging probe and, thereby, of the corresponding acquisition window. To adress this point, a new evaluation protocol is presented in this work as well. It consists in the automatic detection and registration of pairs of lines of a custom 3D-printed cross-wire phantom from US and micro-CT acquisitions, the latter being the reference model. The accuracy of the reconstruction procedure is then computed as the mean distance between the registered sets of lines. Secondly, we demonstrate the clinical potential of our system by performing *ex vivo* acquisitions with a complete anatomical context on a cadaveric head.

The contributions of this paper lie in both the carrying out and validation of minimally invasive 3D US imaging of the auditory system using a miniaturized endoscopic 2D US transducer whose dimensions are compatible with the EAC. To our knowledge, this is the first time that 3D US imaging of the auditory system is reported without prior deterioration of the EAC. Although our imaging platform has already been introduced in [25], this previous study only addressed the issue of calibrating such a miniaturized US transducer. In this paper, we focus on a detailed description of the system and its applicability for medical purposes with a special attention on the validation aspect.

II. Materials and methods

A. Imaging platform

The imaging platform is based on a custom designed 2D US transducer (Fig. 1), whose fabrication was subcontracted to Vermon, a world leader in composite piezoelectric transducers. The dimensions of the distal part were optimized to enable adequate access to the middle ear by raising a tympanomeatal flap. This procedure, which is routinely performed for exploratory tympanotomy [5], [6], consists in elevating the annulus in order to expose the tympanic cavity. To preserve the functionality of the ossicular chain while running surgical instruments and endoscopes into the EAC, the maximum allowable diameter was set to 4 mm [26]. Thereby, the manufactured probe consists in a 4 mm diameter 24 elements curved array transducer with a field of view of 47.55° and a center frequency of 18 MHz $\pm$ 10%. The center frequency value was motivated by the fact that a compromise between compactness and spatial resolution had necessarily to be done. From that perspective, a center frequency of 18 MHz allows to operate at frequencies that provide a spatial resolution of a few tens of microns while remaining close to the frequency range of the 5-14 MHz US transducer we used in a previous study [27] to prove the feasability of using US for imaging an *ex vivo* human cochlea. The probe is also compatible with the SonixTouch® Ultrasound System (BK Medical formerly Ultrasonix, Peabody, United States) that provides real-time rendering by online conversion of the acquired raw data into 2D B-mode images.

Regarding the actuation of the US transducer, we used a *RobOtol* system (Collin Medical, Bagneux, France), a commercial 6 degrees of freedom robotic arm that was specifically designed for middle ear surgery [28]. As depicted in Fig. 2, the probe was inserted into the optical system holder and secured with an additional mechanical assembly which was fabricated by means of additive manufacturing with polyactic acid. Since

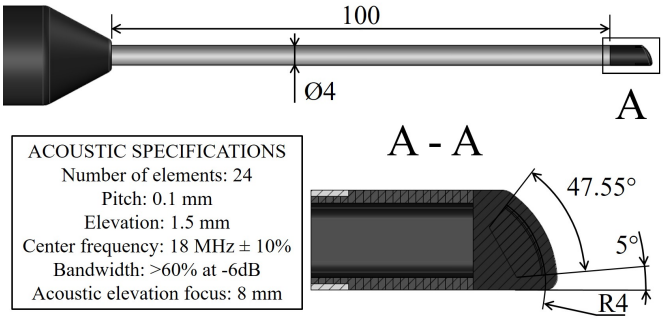

Fig. 1: Technical specifications of the miniaturized endoscopic 2D US transducer. Distances are expressed in mm.

RobOtol is a teleoperated system, a master joystick allowed the operator to adapt the pose of the probe tip in order to adjust the imaging plane. Once the pose was estimated to be satisfying, the transducer was rotated around the distal axis of the arm to collect a set of B-scans. The rotational speed and the frame rate were respectively set to 1°/s and 10 images per second, leading to an angular resolution of 0.1°. The frequency of the probe was also set to 20 MHz to achieve high spatial resolution. All the acquired 2D frames were transferred to a workstation using Ethernet through the OpenIGTLink protocol and then stacked in a DICOM file for further processing. In the following sections, we refer to the stack of acquired B-scans as the pre-scan converted image I_{pre} and the corresponding reconstructed volume as the post-scan converted image I_{post}.

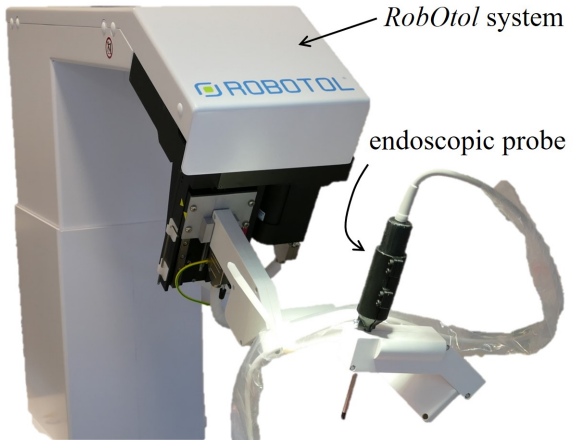

Fig. 2: Imaging platform with the endoscopic probe mounted onto the end-effector of a *RobOtol* system.

B. Reconstruction procedure

Scan-conversion (SC) consists in resampling the collected B-scans on a regular Cartesian grid using coordinate system transformation derived from the geometry of the probe. In summary, the reconstruction procedure can be summarized in three steps [17]. Firstly, for each sample of the volume to be reconstructed, one must compute the corresponding point in the transducer coordinate system thanks to *backward mapping*. Secondly, since the resulting point may fall at non-grid location, *interpolation* is then necessary to estimate its associated gray value from the pre-scan converted data. Finally, the interpolated intensity value is *assigned* to the studied sample.

Despite a clear outline of the procedure, SC lacks methodology and remains cumbersome to perform. Reported implementations are actually transducer specific [19] as, for instance, there is no systematic and automatic way to derive the coordinate system transformation regardless of the used probe. Thanks to our expertise as roboticists, we propose in the following sections an adequate method for implementing SC that is not only meant to be straightforward for the non-specialist reader but also can be extended to any transducer geometry.

Forward mapping: The objective of *forward mapping* is to allow for the conversion of the transducer coordinates (x_I, y_I, φ) into the world coordinates (x, y, z). x_I and y_I denote the coordinates of a point in an acquired B-scan relative to the US transducer coordinate frame $\mathcal{F}_4$ and φ the rotation angle of the transducer while the 3-tuple (x, y, z) corresponds to the coordinates of the same point expressed in the world coordinate frame $\mathcal{F}_0$ (Fig. 3). $\mathcal{F}_0$ is chosen so that the y-axis is collinear to the rotation axis of the probe pointing forward, whereas the x-axis passes through the transducer array focus (TAF) when $\varphi = 0$. In order to derive the (4×4) homogeneous transformation matrix ${}^0\mathbf{T}_4$ that defines $\mathcal{F}_4$ relative to $\mathcal{F}_0$, we introduce intermediary frames as depicted in Fig. 3. As $\mathcal{F}_0$ is subjected to four consecutive transformations, ${}^0\mathbf{T}_4$ can be deduced from the postmultiplication

$$
\begin{aligned}
{}^0\mathbf{T}_4 &= \overbrace{{}^0\mathbf{T}_1}^{\mathbf{Rot}_y(\varphi)}\ \overbrace{{}^1\mathbf{T}_2}^{\mathbf{Tr}_x(-r)}\ \overbrace{{}^2\mathbf{T}_3}^{\mathbf{Rot}_z(-\theta_0)}\ \overbrace{{}^3\mathbf{T}_4}^{\mathbf{Tr}_y(\Delta y)\,\mathbf{Tr}_x(-\Delta x)} \\
&= \begin{pmatrix} \mathbf{R}_y(\varphi).\mathbf{R}_z(-\theta_0) & \begin{matrix} -c_\varphi.\left[\Delta x.c_{\theta_0} - \Delta y.s_{\theta_0} + r\right] \\ \Delta x.s_{\theta_0} + \Delta y.c_{\theta_0} \\ s_\varphi.\left[\Delta x.c_{\theta_0} - \Delta y.s_{\theta_0} + r\right] \end{matrix} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix}
\end{aligned} \tag{1}
$$

with $c_\alpha = \cos(\alpha)$ and $s_\alpha = \sin(\alpha)$. r is the radius of the endoscopic part of the US probe ($r = 4$ mm) and θ_0 the elevation angle up to half of the field of view ($\theta_0 = 0.50$ rad). Δx and Δy correspond to the offsets in the x and y directions, respectively, between the origins of the frames $\mathcal{F}_3$ and $\mathcal{F}_4$ and depend on both the spatial spacing and the extent of the acquired B-scans. $\mathbf{Rot}_y$ and $\mathbf{Tr}_y$ denote respectively the (4×4) transformation matrices of pure rotation and translation about the y-axis and similarly for other axes whereas $\mathbf{R}_y$ and $\mathbf{R}_z$ represent the (3×3) rotation matrices about the y- and z-axes.

Following (1), it is possible to derive the $\mathcal{F}_0$ relative world coordinates ${}^0\mathbf{p} = (x, y, z, 1)$ from the $\mathcal{F}_4$ relative US transducer coordinates ${}^4\mathbf{p} = (x_I, y_I, 0, 1)$ thanks to the relation

$$
{}^0\mathbf{p} = {}^0\mathbf{T}_4.{}^4\mathbf{p}, \tag{2}
$$

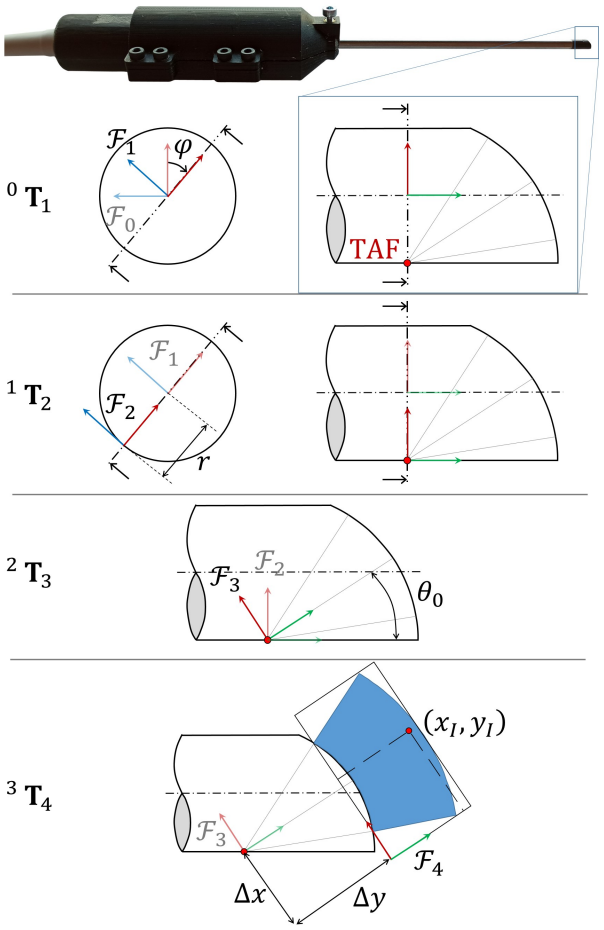

Fig. 3: Successive transformations corresponding to *forward mapping*.

which, after simplification, yields the forward mapping system

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c_\varphi[(x_I - \Delta x).c_{\theta_0} + (y_I + \Delta y).s_{\theta_0} - r] \\ -(x_I - \Delta x).s_{\theta_0} + (y_I + \Delta y).c_{\theta_0} \\ -s_\varphi[(x_I - \Delta x).c_{\theta_0} + (y_I + \Delta y).s_{\theta_0} - r] \end{pmatrix}. \quad (3)$$

This step is required to generate the sampling domain Ω_{post} of the volume to be reconstructed, which is based on the physical extent of the pre-scan converted image and the spacing that we want to achieve.

Backward mapping: On the contrary, *backward mapping* denotes the relation that allows to pass from world to transducer coordinates. To derive the inverse relationship, we first determined the rotational angle φ from (3) using the x and z coordinates:

$$\varphi = -\tan^{-1}\left(\frac{z}{x}\right) \quad (4)$$

After finding φ, we determined the x_I and y_I coordinates in the corresponding B-scan by solving the linear system formed by the first and third lines of (3), which yields:

$$\begin{pmatrix} x_I \\ y_I \end{pmatrix} = \begin{pmatrix} (\frac{x}{c_\varphi} + r).c_{\theta_0} - y.s_{\theta_0} + \Delta x \\ (\frac{x}{c_\varphi} + r).s_{\theta_0} + y.c_{\theta_0} - \Delta y \end{pmatrix}. \quad (5)$$

Implementation: According to (4) and (5), each sample $\mathbf{x}_{post}$ that belongs to the regular grid Ω_{post} of the 3D image $I_{post} : \Omega_{post} \rightarrow [0, 255]$ to be reconstructed can then be mapped to its corresponding location $\mathbf{x}_{pre}$ in the stack of the acquired B-scans $I_{pre} : \Omega_{pre} \rightarrow [0, 255]$. For a given voxel that may not fall at grid locations, the estimation of the associated gray is performed using trilinear interpolation, which provides a good compromise between reconstruction quality and computational time [29]. From an algorithmic perspective, the SC procedure can be described by the following *pseudocode*:

Algorithm Scan-conversion

Require: $\Omega_{post}, \Omega_{pre}, I_{pre}$
  for all $\mathbf{x}_{post} = (x, y, z) \in \Omega_{post}$ **do**
    compute $\mathbf{x}_{pre} = (x_I, y_I, \varphi)$ using (4) and (5)
    if $\mathbf{x}_{pre} \in \Omega_{pre}$ **then**
      Compute the voxel intensity $I_{post}(\mathbf{x}_{post})$ from I_{pre} evaluated at the eight corner points surrounding $\mathbf{x}_{pre}$ using trilinear interpolation
    else
      $I_{post}(\mathbf{x}_{post}) = 0$
    end if
  end for
  return I_{post}

By working with a stack of already converted B-scans instead of directly using the raw radio frequency signals collected by the transducer, this implementation of SC allows for a shorter reconstruction time at the expense of an additional interpolation step [18].

The complete reconstruction procedure was implemented using the C++ Insight Tool Kit (ITK) library and run on an Intel® Xeon^{TM} i7-8750H 2.20GHz processor.

C. Evaluation of system performances

Resolution measurement: The spatial resolution that is achievable by a 2D transducer is defined as the axial, or lateral, full width at half maximum (FWHM). This corresponds to the -6 dB width, along either the axial or lateral direction respectively, of a punctual object that is imaged at the optimal focus. In order to replicate such a punctual information, we designed a single 0.18 mm thick wire phantom (Fig. 4a) that is meant to be imaged perpendicularly to the acquisition plane of the transducer. Thus, the obtained B-scan contains a shape that comes closest to a point and from which we can compute the FWHM. Since the objective is to characterize the FWHM as a function of the distance from the probe, the phantom was placed on a positionning Z stage (Fig. 4c), which allows sub-mm control over the height of the linear platform.

For any given cross-section of the wire, the associated FWHM values (axial and lateral) were computed following the measurement protocol defined in [30]. First, we estimated the position of the maximal intensity pixel and placed a rectangular measurement box around it. Then, the gray level profiles centered on the previous pixel position and bounded by the box were retrieved along the axial (y-axis) and lateral

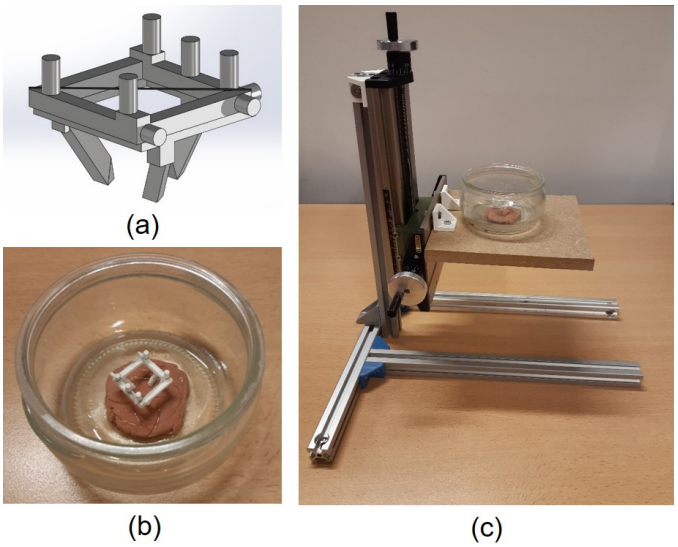

Fig. 4: Experimental setup for the measurement of the full width at half maximum. (a) CAD view of the 0.18 mm thick wire phantom. (b) Top view of the phantom after being secured in a recipient with some clay. (c) Positionning Z stage for controlling the height of the phantom.

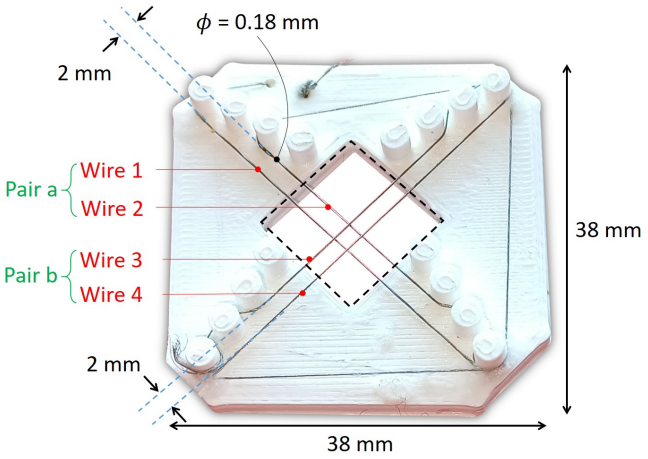

Fig. 5: Top view of the 3D printed phantom used for evaluating the accuracy of the reconstruction procedure.

(x-axis) directions and recalculated to dB scale. Finally, a parabola was fitted to each profile using least squares based on the gray level points above -20 dB with respect to the 0 dB level maximum, from which we deduced the FWHM as the -6 dB width. For each position of the phantom, a rotational sweep of 4° was collected and used for computing the FWHM values so that to quantify the precision of the resolution measurement.

Correctness of reconstructed volume's geometry measurement: Beyond assessing the resolution of the 2D B-mode images, this paper also aims to quantitatively evaluate the correctness of volumetric reconstruction since this new imaging platform is meant for medical purposes. To the best of the authors' knowledge, there is no agreed standard in the literature, and existing works on SC actually rely only on visual inspection for validation [17]–[19]. The only paper that presents such a quantitative test is a freehand US related work [31]. The proposed protocol consists in measuring to what extend the volumetric reconstruction is capable of reproducing the geometry of a well-known phantom. This evaluation test, though, induces a bias in the reported accuracy since measurements needs to be performed manually in the reconstructed volumes. In that sense, this technique would benefit from being automatized. Besides, a stumbling block for extending the test presented in [31] to our need arises from the fact that the phantom geometry must fit into the unusual small acquisition window of our setup.

To adress the abovementioned issues, we propose in this section a novel operator-independent protocol for evaluating the correctness of the reconstructed volume's geometry, which is based on a dedicated phantom that fits our acquisition window. Ideally, such a phantom should include point-like reflectors distributed in space as the reference geometry. Yet, its design remains a challenging task and even more so because of the unusual small dimensions of the acquisition window. Consequently, lines were chosen as the reference geometry as they both provides redudant information in 3D and are easy to build. The phantom consists of a cross-wire structure including two pairs of 2-mm-distant parallel wires that were mounted in orthogonal directions (Fig. 5). The plastic support was fabricated with 3D printing and metal wiring was used to enhance the constrast between the wires and the background. In addition, a reference volume of the studied phantom was required in order to account for both the 3D printing and wiring positioning errors that may occur during the fabrication. This would then serve as a ground truth in our evaluation protocol. Accordingly, we resorted to micro-CT that not only can provide high resolution images but also benefits from a high signal-to-noise ratio. Thus, the reference volume was acquired with an EasyTom 150® scanner (RX Solutions, Chavanod, France), which resulted in a volume of 389 $\times$ 1897 $\times$ 1243 voxels with an isotropic spacing of 29.35 µm.

For any given US volume, our evaluation protocol is organized as follows. First, each of the four 0.18 mm thick wires imaged in both US and micro-CT volumes was *modelled* by a line segment, which we represented by the 3-tuple $\mathcal{S}_i = \{\mathbf{c}_i, \mathbf{e}_i, l_i\}$ with $i \in \{1, 2, 3, 4\}$. $\mathbf{c}_i$ denotes the position vector of the center of the segment line, $\mathbf{e}_i$ the unit vector along its direction and l_i its length. Following the resulting sets of line segments $\mathcal{S}_{US} = \left\{ \mathcal{S}_i^{US} \right\}$ and $\mathcal{S}_{\mu CT} = \left\{ \mathcal{S}_i^{\mu CT} \right\}$, we then computed the transformation that allows to *match line segments* of both modalities. Finally, we estimated the *accuracy* of volumetric reconstruction by computing the mean distance from the registered US set of lines to the micro-CT one. To our knowledge, no similar quantitative and automatic evaluation protocol for volumetric US imaging has been reported in the literature.

The evaluation protocol was implemented using the same C++ library as for the reconstruction procedure.

1) Line segment fitting: For both the US and micro-CT modalities, the modelling of the set of wires was performed in two stages. First, each wire was isolated separately using a cylinder defined by two selected endpoints to provide a set of voxels $\mathcal{V}$ with coordinates $\mathcal{X} \subset \mathbb{R}^3$ to work on and then modelled by a line corresponding to the subset

$$\mathcal{L} = \{\mathbf{x} \in \mathbb{R}^3 \mid \exists \lambda \in \mathbb{R},\ \mathbf{x} = \mathbf{p} + \lambda \mathbf{e}\}, \tag{6}$$

where $\mathbf{p}$ is the position vector of a point belonging to the line and $\mathbf{e}$ the unit vector along its direction. Finally, according to the estimated parameters, we deduced the end points of the studied set of voxels $\mathbf{p}_{min} = \mathbf{p} + \lambda_{min}\mathbf{e}$ and $\mathbf{p}_{max} = \mathbf{p} + \lambda_{max}\mathbf{e}$ such as

$$\begin{cases} \lambda_{min} & = \min_{\mathbf{x}\in\mathcal{X}} (\mathbf{x}-\mathbf{p})^\intercal \mathbf{e} \\ \lambda_{max} & = \max_{\mathbf{x}\in\mathcal{X}} (\mathbf{x}-\mathbf{p})^\intercal \mathbf{e} \end{cases}, \tag{7}$$

from which we derived the center of the corresponding line segment $\mathbf{c} = \frac{\mathbf{p}_{max}+\mathbf{p}_{min}}{2}$ and its length $l = \|\mathbf{p}_{max}-\mathbf{p}_{min}\|$.

For the micro-CT acquisition, the wiring part was extracted from the volume using simple thresholding due to the high density of metal compared to plastic. Then, the best line model was derived by principal component analysis (PCA) [32].

Regarding the US volumetric reconstruction, PCA was not applicable since US imaging suffers from various artifacts such as speckle and distortion that may hinder the line modelling step. Consequently, we applied a modified version of the procedure described in [33], in which a random sample consensus (RANSAC) is used to robustly localize a thin surgical tool in US volumes. The RANSAC approach is preferred to using Hough transform techniques as it does not suffer from a high computational effort in the 3D case. First, in each RANSAC iteration, two control points $\mathbf{p}_A$ and $\mathbf{p}_B$ are randomly selected from $\mathcal{X}_{US}$ to fit a line so that $\mathbf{p} = \frac{\mathbf{p}_B-\mathbf{p}_A}{2}$ and $\mathbf{e} = \frac{\mathbf{p}_B-\mathbf{p}_A}{\|\mathbf{p}_B-\mathbf{p}_A\|}$ in (6). From the estimated line $\mathcal{L}_{US}$, each voxel of coordinates $\mathbf{x} \in \mathcal{X}_{US}$ is then classified as either belonging to the model ($m = 1$) or not ($m = 0$) following the classification function

$$m(\mathbf{x}, \mathcal{L}_{US}) = \begin{cases} 1, \text{ if } d(\mathbf{x}, \mathcal{L}_{US}) \leq r_{wire} \\ 0, \text{ otherwise} \end{cases}, \tag{8}$$

where $d(\mathbf{x}, \mathcal{L}_{US})$ represents the orthogonal distance of the point $\mathbf{x}$ to the line $\mathcal{L}_{US}$ and r_{wire} the radius of the studied wire ($r_{wire} = 0.09$ mm). Thanks to (8), it is then possible to define the set of inliers

$$\mathcal{X}_{inl} = \{\mathbf{x} \in \mathcal{X}_{US} \mid m(\mathbf{x}, \mathcal{L}_{US}) = 1\}, \tag{9}$$

which is involved in the selection of the best model across the iterations. However, contrary to the method described in [33], whose voting process is based on the usual number of estimated inliers, we decided to weight the vote of each inlier voxel by their corresponding intensity value. Accordingly, this RANSAC procedure no longer consists in maximizing $|\mathcal{X}_{inl}|$ but the cost function

$$C(\mathcal{X}_{inl}) = \sum_{\mathbf{x}\in\mathcal{X}_{inl}} I_{post}(\mathbf{x}), \tag{10}$$

which relies on the fact that the intensity of the wire's voxels is maximal at its center. Unlike Uhervčík *et al.* in [33], using this specific cost function allows to avoid prior segmentation to extract the set of voxels $\mathcal{V}_{US}$ and, thereby, eliminates the need for an additional tuning stage of the threshold that may be volume specific.

2) Model matching: Regarding the matching of the resulting sets of line segments, we used the approach presented in [34]. For conciseness purposes, we only discuss here the outline of the technique we implemented and refer readers to the original paper for further details.

The proposed algorithm in [34] takes as input two line sets: one referred as the model set $\mathcal{S}_M$ and the other as the image set $\mathcal{S}_I$. The output corresponds to the rigid transformation ${}^M\mathbf{T}_I$ that allows to align $\mathcal{S}_I$ with $\mathcal{S}_M$ such that

$${}^M\mathbf{T}_I : \mathcal{S}_I = \left\{ \left\{\mathbf{R}\mathbf{c}_i^I + \mathbf{t}, \mathbf{R}\mathbf{e}_i^I, l_i\right\} \right\}, \tag{11}$$

where $\mathbf{R}$ and $\mathbf{t}$ are the (3×3) rotation matrix and the translation vector associated to ${}^M\mathbf{T}_I$, respectively.

Depending on whether the model and/or the image line sets consist of sets of finite line segments or infinitely long lines, the authors distinguish several cases that lead to different derivations of the matching transformation. Since we previously estimated the length of each US line segment, the **F**inite **M**odel **F**inite **I**mage (FMFI) algorithm, when model and image line sets both consist of finite line segments, was the most relevant for our application. Yet, according to the authors' instructions, the FMFI algorithm performs best when initialized with the **I**nfinite **M**odel **I**nfinite **I**mage approach (IMII), which applies when model and image both consist of infinite lines. Consequently, we proceeded in two steps by first estimating an initial guess of ${}^M\mathbf{T}_I$ using the IMII algorithm and then refining it using the FMFI algorithm. In the next step, we denote by $\mathcal{S}_{mUS}$ the matched set of line segments resulting from operating the transformation ${}^M\mathbf{T}_I$ on $\mathcal{S}_{US}$.

3) Accuracy evaluation: When it comes to quantifying the spatial consistency between volumes after multimodal matching, the target registration error (TRE) is considered as the most relevant evaluation tool in the literature. Still, this metric does not apply for our phantom since it requires detected features to be points and is no longer applicable on line segments. As a result, we introduce in this section a new metric that is meant to quantify the accuracy (ACC) of volumetric reconstruction by measuring the difference between the reconstructed US and the micro-CT volumes based on the matched sets of line segments.

Since two matched lines may not be strictly parallel with each other, we defined the relative distance D_i between them as the mean squared integral of the distance

$$d_i(\lambda) = \| \left(\mathbf{p}_i^{mUS} - \mathbf{c}_i^{\mu CT}\right) \times \mathbf{e}_i^{\mu CT} \| \tag{12}$$

from the parameterized point $\mathbf{p}_i^{mUS}(\lambda) = \mathbf{c}_i^{mUS} + \lambda \mathbf{e}_i^{mUS}$ to the line segment $\mathcal{S}_i^{\mu CT}$ with respect to the variable λ, which

yields

$$D_i = \sqrt{\frac{1}{\lambda_{i,2} - \lambda_{i,1}} \int_{\lambda_{i,1}}^{\lambda_{i,2}} d_i^2(\lambda) d\lambda}. \quad (13)$$

The values $\lambda_{i,1}$ and $\lambda_{i,2}$ of the interval for integration were defined such that the projections of both points $\mathbf{p}_i^{mUS}(\lambda_{i,1})$ and $\mathbf{p}_i^{mUS}(\lambda_{i,2})$ on the line segment $\mathcal{S}_i^{\mu CT}$ exactly coincide with the endpoints $\mathbf{c}_i^{\mu CT} \pm \frac{l_i}{2}\mathbf{e}_i^{\mu CT}$.

Thereby, (13) allows us to obtain the ACC as the mean value of the relative distances D_i from $\mathcal{S}_i^{mUS}$ to $\mathcal{S}_i^{\mu CT}$ line segments:

$$ACC = \frac{1}{4}\sum_{i=1}^{4} D_i \quad (14)$$

Although the proposed evaluation protocol only applies to wire phantoms, it allows to validate the constructability of more complicated structures than just line segments. Indeed, in case of an error in the reconstruction of either the orientations or the interdistances between the wires, the reconstruction procedure would worsen the line segment fitting, which would then impact the model matching and finally result in a degraded accuracy value. Consequently, the metric defined in (14) enables to quantify to what extent the reconstruction procedure allows to preserve both the orientations and distance, which holds true for any structure.

D. Evaluation of clinical potential

To assess the clinical potential of the proposed setup, we carried out additional acquisitions on the left ear of a cadaveric head to account for a representative intraoperative context. Four volumetric acquisitions requiring prior anatomical preparation were actually performed with the help of an ENT surgeon.

First acquisition: The first acquisition was conducted on a left ear featuring an intact tympanic membrane (TM). A posterior incision was performed to elevate the pinna. This was required to insert the probe into the EAC without damage, since the cadaveric rigidity of the pinna prevents the normal stretching of the cartilaginous structures. This provided a visual feedback on the positioning of the tip of the transducer. Finally, a transtympanic injection of saline solution was carried out to allow for acoustic coupling, and the acquisition was launched.

Second acquisition: For this acquisition, the TM was entirely removed using tympanotomy. In the absence of this opaque membrane, it was then possible to adjust more properly the pose of the tip and thus to focus on middle ear structures. Again, a transtympanic injection of saline solution was performed before starting the rotational sweep.

Third and fourth acquisitions: Afterwards, we concentrated on investigating the potential of volumetric US imaging applied to cochlear implants by placing a Neuro ZTI CLASSIC electrode array (Oticon Medical, Vallauris, France) in a representative anatomical context. To that end, a mastoidectomy with posterior tympanotomy was performed on the same ear to enable the access to the tympanic cavity as it is routinely done in cochlear implantation, and the implant was inserted through the round window into the scala tympani. Following the implantation, two volumetric reconstructions (third and fourth acquisitions) were performed using the same sweep settings, imaging parameters and pose of the probe: one with the electrode array in place and the other after its removal. This step enabled to extract the implant from the acquired data by subtracting the corresponding reconstructed volumes. Volumetric reconstruction of the same implant outside any anatomical context in a water tank was also performed for comparison purposes.

The handling and imaging of the cadaveric tissues were conducted with the approval of the local ethics committee of the Faculty of Medicine of the University of Montpellier (FRANCE).

III. Results

A. Resolution measurement

Fig. 6 shows the FWHM values (axial and lateral) obtained from the single wire phantom images as functions of the axial distance from the probe. Overall, a total of 80 acquisitions were performed with an incremental step of 0.2 mm in the axial direction between two successive rotational sweeps, leading to a study range of 16 mm. For each distance, the axial and lateral FWHM are plotted with a central point representing the mean value computed over the rotational sweep while whiskers indicate the corresponding standard deviation. The axial distance from the probe was automatically measured from the acquired images in order to be operator-independent. According to the figure, the FWHM have different behaviors along the two directions. The lateral FWHM (blue) tends to degrade linearly with the distance from the probe, ranging from 0.23 mm in the near-field to 1.04 mm in the far-field, whereas the axial FWHM (red) does not vary significantly with a mean value of 0.16 ± 0.01 mm.

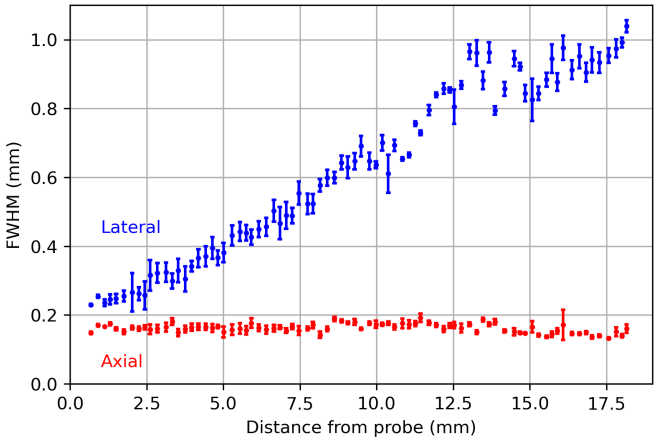

Fig. 6: Measurements of the axial and lateral resolutions as the full-width at half maximum (FWHM) of the single wire phantom at various distances from the probe.

B. Correctness of reconstructed volume's geometry measurement

Three-dimensional renderings of the volumetric reconstructions of the custom 3D printed cross-wire phantom using micro-CT and US are shown in Fig. 7. As mentioned in section II-C, the density of metal wiring provides a high contrast between the background and wires. Yet, the US acquisition window remains limited since the proposed setup only allowed to image a small subpart of the phantom compared to micro-CT (Fig. 7b).

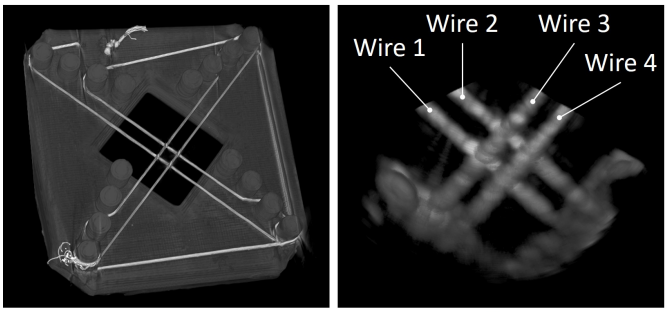

Fig. 7: 3D renderings of the phantom with (a) micro-CT and (b) volumetric US.

Qualitatively, Fig. 7 indicates that the acquisition setup allows to preserve the orientations since the rendered wires 1 and 2 (*resp.* 3 and 4) look parallel with each other, which is consistent with the designed phantom. Although the pair of parallel wires a (wires 1 and 2) may also look perpendicular to the pair b (wires 3 and 4), orthogonality is more difficult to assess from visual inspection as it depends on the rendered pose.

Regarding the evaluation protocol, Fig. 8 illustrates the processing steps we described in section II-C. On the micro-CT side, the reference volume was first thresholded in order to extract the wiring part from the rest of the image (Fig. 8a). In addition, the segmented wires were cropped according to the central square (dotted diamond in Fig. 5) so that to provide similar extent line segments than the acquired with volumetric US. Then, line fitting was performed on the corresponding subpart using PCA (Fig. 8b). On the US side, the result obtained from line fitting performed on the reconstructed US volume using RANSAC is depicted in Fig. 8c. Finally, Fig. 8d presents a rendering of the two matched sets of line segments resulting from applying successively the IMII and FMFI algorithms.

Overall, a total of 12 US reconstructed volumes were used for the quantitative evaluation with varying orientations and depths of the probe with respect to the phantom. In order to study the influence of the incidence of the transducer relative to the phantom on the ACC, the spherical coordinates of the transducer array focus (TAF in Fig. 3) were derived from each reconstructed volume and reported in Table I. To that end, we adopted the following convention (Fig. 9): the radial distance r from the intersection point between the wires 1

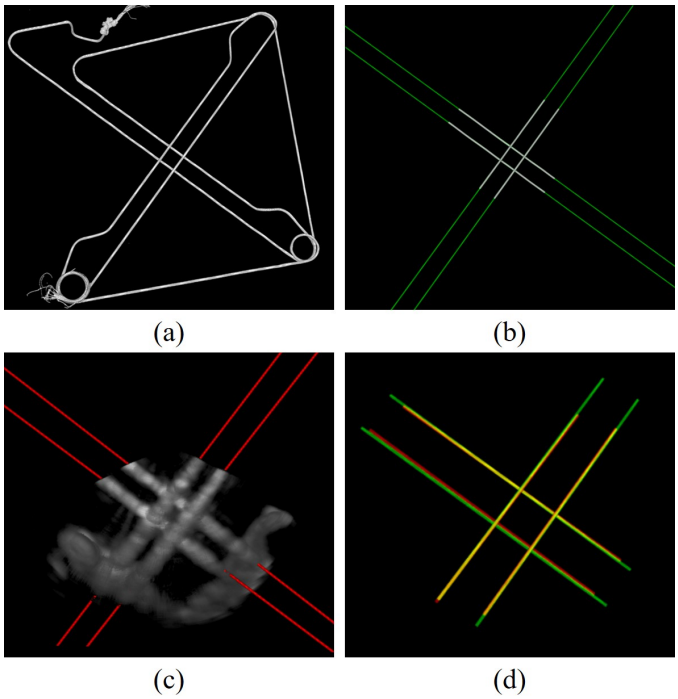

Fig. 8: Processing steps of the evaluation protocol. (a) Segmentation of the wiring part in the micro-CT volume after thresholding and (b) line segment modelling (green) on the cropped part using PCA. (c) Line segment modelling (red) in the US volume using RANSAC. (d) Matching of both micro-CT (green) and US (red) sets of line segments.

and 4; the elevation angle θ measured from a reference plane; the azimuthal angle φ between the position vector projected onto the reference plane and the z-axis. To provide a broad picture of the range of adopted configurations, the extreme poses of the tip of the transducer are also depicted in Fig. 10.

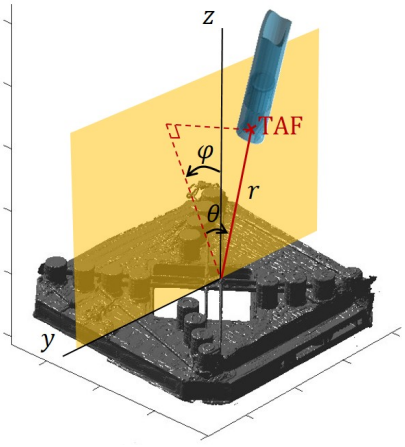

Fig. 9: Spherical coordinates (r, θ, φ) used to determine the incidence of the US tranducer relative to the phantom: radial distance r, elevation angle θ and azimuthal angle φ.

For each acquisition, the phantom was placed in a container

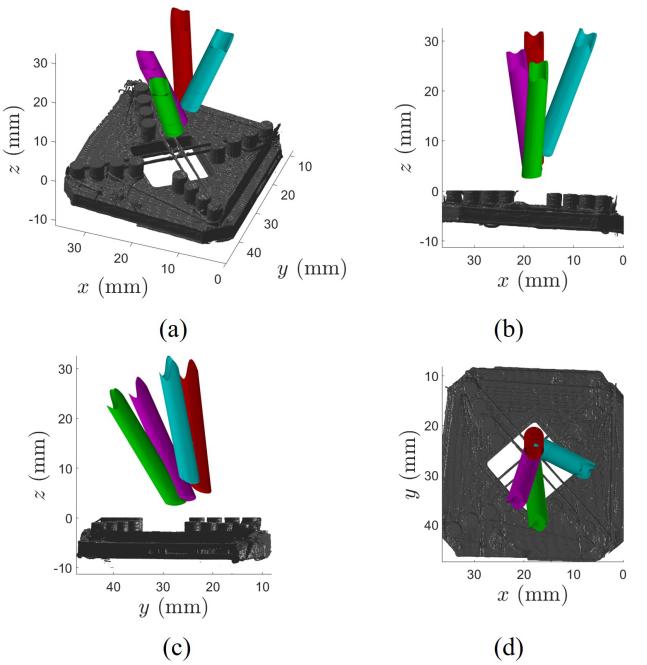

Fig. 10: Renderings of the extreme poses of the tip of the probe (colors) with regard to the phantom for the complete set of scans. (a) Isometric view. (b) Front view. (c) Right view. (d) Top view. Red, green, cyan and magenta colors correspond to acquisitions 1, 9, 11 and 12, respectively.

filled with water to allow for acoustic coupling. The rotational sweep was set to $\Delta\varphi = 120°$ to enable a complete scan of the wires. Regarding the imaging parameters, the frequency was set to 20 MHz, the B-mode gain to 30% and the imaging depth to 1.6 cm. Thereby, each reconstructed volume resulted in a 3D image of 609 × 645 × 722 voxels with a spatial spacing of 31 µm.

The calculations of the ACC on the complete set of scans are given in Table I. The reconstruction errors range from 0.08 to 0.43 mm with a mean value of 0.20 ± 0.09 mm. The lowest ACC was obtained on the US volume 2 while the highest one was obtained on the US volume 11. The spherical coordinates of the TAF were $(r_2, \theta_2, \varphi_2)$ = (16.62, 3.54, 1.40) and $(r_{11}, \theta_{11}, \varphi_{11})$ = (14.28, 12.70, 3.12), respectively. These results were obtained by setting the threshold value to 30 for the micro-CT while the iterative procedures were stopped after reaching a total of 10^4 iterations for the RANSAC and 10^2 iterations for the FMFI algorithm.

C. Acquisitions on the cadaveric head

The results presented in this section refer to the acquisitions described in section II-D.

First acquisition: Imaging of the auditory system with an intact TM is presented in Fig. 11. The first volumetric reconstruction resulted in an image of 586 × 658 × 625 voxels with an isotropic spacing of 28 µm. The TM with its typical cone shape is clearly visible in the cross-sectional

US volume	r (in mm)	θ (in °)	φ (in °)	ACC (in mm)
1	12.35	01.33	-09.20	0.12
2	16.62	03.54	01.40	0.08
3	10.01	04.30	02.71	0.19
4	16.30	05.18	01.96	0.22
5	10.02	01.23	02.78	0.16
6	16.49	03.92	07.41	0.20
7	10.03	01.17	12.87	0.18
8	16.70	04.11	14.43	0.15
9	09.89	-01.85	26.07	0.25
10	16.33	01.82	21.12	0.32
11	14.28	12.70	03.12	0.43
12	10.26	-03.52	11.65	0.14
Mean				0.20 ± 0.09

TABLE I: Accuracy of volumetric reconstruction for each US volume with regard to the micro-CT reference.

B-scan as well as the cochlear promontory, which results in a brighter structure due to its high density (Fig. 11a). Yet, because of bone distortion, localization of the round window (RW) remains cubersome considering only 2D images. On the contrary, 3D rendering of the reconstructed volume easily allows to identify the RW niche (RWN) by highlighting the corresponding cavity in the cochlear promontory (Fig. 11b). Parts of the malleus are also identified such as the umbo, which is located at the tip of the conic TM, and the lateral process of the malleus. In Fig. 11b, the malleus manubrium is not continuously delineated as it does not entirely lie within the acquisition window.

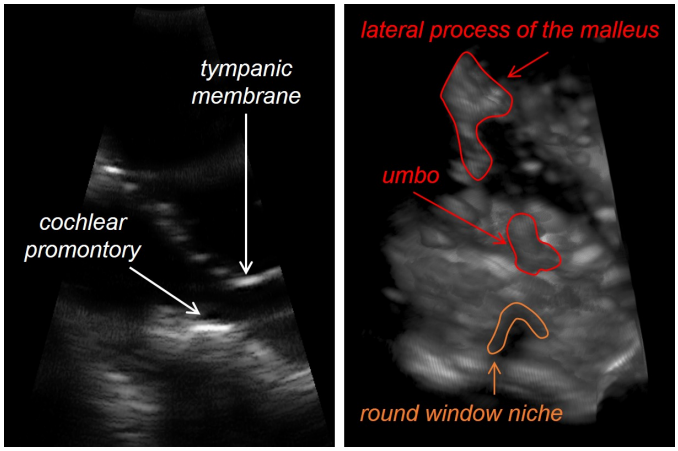

Fig. 11: US imaging of the auditory system with an intact eardrum. (a) B-scan. (b) 3D rendering.

Second acquisition: Fig. 12 presents the imaging of the auditory system after complete removal of the TM. Fig. 12a and Fig. 12b depict otomicroscopy and CBCT views of the corresponding tympanic cavity in order to facilitate the identification of the imaged structures with 3D US (Fig 12c). The second volumetric reconstruction led to an image of 558 × 609 × 599 voxels with a spatial spacing of 38 µm. Compared to Fig. 11b, additional structures of the middle ear were included within the acquisition window by adjusting

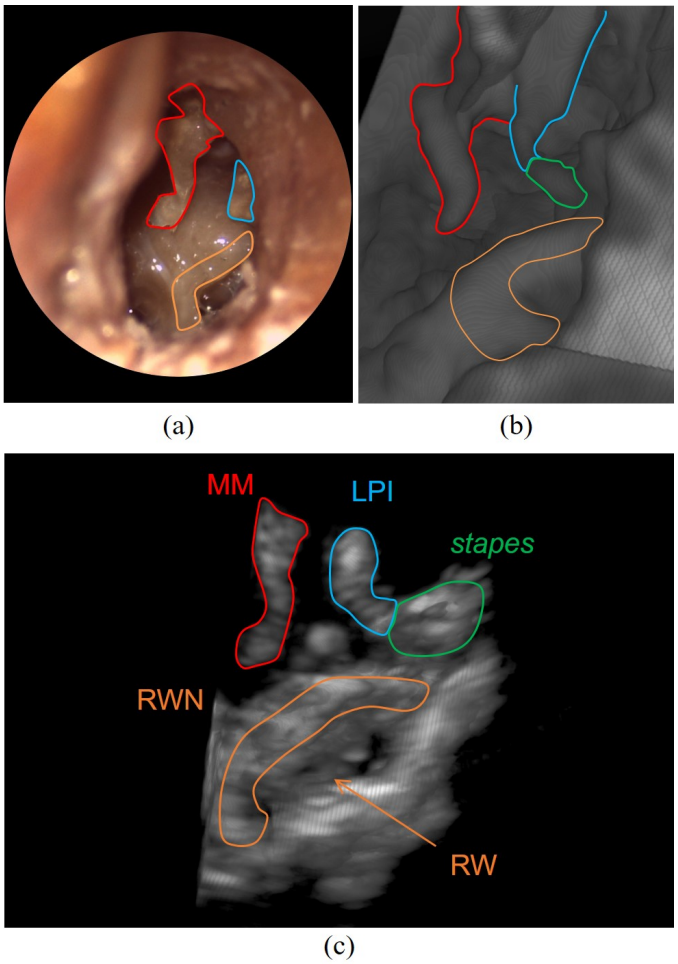

Fig. 12: Imaging of the tympanic cavity after tympanotomy with (a) otomicroscopy, (b) CBCT and (c) volumetric US highlighting the malleus manubrium (MM), the long process of the incus (LPI), the stapes, the round window niche (RWN) and the round window (RW).

properly the pose of the probe tip thanks to visual feedback. A significant part of the ossicular chain is clearly visible such as the malleus manubrium, long process of incus and stapes. Again, the cavity in the bottom part suggests the location of the RWN which overlooks the RW.

Third and fourth acquisitions: Results of the investigation on cochlear implants are presented in Fig. 13 by depicting 3D US imaging of the electrode array in a water bath (Fig. 13a) and after implantation through the RW (Fig. 13b). The third volumetric reconstruction resulted in an image of $609 \times 643 \times 724$ voxels with a spatial spacing of 34 µm. Fig. 13a shows that the used imaging parameters are adequate for the detection of the electrode array in water. Fig. 13b, however, depicts a composite rendering of the electrode array (green) and the middle ear (gray). Since no change in the sweep settings and imaging parameters was applied before and after the implantation, third and fourth volumetric reconstructions led to images of $558 \times 609 \times 599$ voxels with a spatial spacing of 38

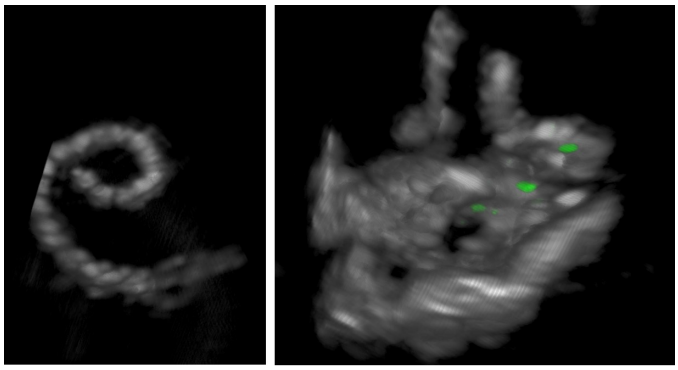

Fig. 13: 3D rendering of US acquisitions of a cochlear implant (a) placed in a water bath and (b) after insertion into the cochlea. The platinum contacts of the electrode array are depicted in green outside the cochlea.

µm. Compared to Fig. 13a, only small segments of the implant are visible in the clinical setting. This may be due to the fact that the electrode array is no longer placed perpendicularly to the probe. As a result, only the eletrodes that are localized outside the cochlea can be imaged since they are the most echogenic parts of the implant.

As a lone 2D view of a 3D rendering may not allow for a clear understanding of volumetric reconstruction, we also provide videos of the corresponding reconstructed volumes as supplementary materials.

IV. Discussion

In this paper, we presented a novel imaging platform for the 3D US imaging of the auditory system, which is based on the motorization of a miniaturized endoscopic US transducer specially developped for fitting the EAC.

Regarding the performances of the system, 2D B-mode imaging was shown to achieve a lateral resolution of 0.23 mm in the near-field that degrades as the distance increases and a stable axial resolution of 0.16 mm $\pm$ 0.01 mm. Consequently, volumetric imaging requires a tradeoff to be made between the spatial resolution to be achieved and the size of the acquisition window since the latter enlarges with the distance due to the conical shape of the collected B-scans. As regards with volumetric imaging, typical 3D scans were acquired within 2 minutes with an angular range of 120° at a rotational speed of 1°/s. We are perfectly aware that such a scanning time is not suited for clinical applications and, as a result, future works should focus on its reduction. One possibility could be to explore synthetic aperture since it allows to significantly increase the data rate and has been studied for achieving real-time 3D US imaging using 2D arrays transducers [35], [36]. Concerning the reconstruction time, it ranged from 50 to 72 s on CPU depending on the physical extent and the desired spacing. As for the rotational sweep, such a computational time is excessive for a clinical use. However, implementing the SC algorithm on GPU could be a relevant solution to achieve almost real-time reconstruction.

Concerning volumetric imaging, the correctness of the reconstruction procedure was evaluated using a novel operator-independent protocol that relies on the matching of a micro-CT volume of a reference geometry. A modified version of the RANSAC procedure described in [33] was used for the modelling of line segments in 3D US images. The rationale for modifying the original procedure was to avoid any prior segmentation of the data so that to automatize the evaluation. The calculations performed on the validation set composed of a total of 12 reconstructed US volumes (Table I) demonstrated that our setup allows to reconstruct 3D US data with a mean maximum error of 0.2 mm. It is actually crucial to understand that the reported ACC does not reflect the error of reconstruction itself but its highest bound since the incertainties of each step composing the evaluation protocol are contained in this measure. Indeed, line fitting in US still remains a challenging task that may lead to outliers due to the presence of artifacts while model matching based on the FMFI algorithm may converge to a local optimum. All of these are error prone tasks that impact the matching of the reference geometry in both modalities, even in case of a perfect reconstruction. According to Table I, the lowest and highest ACC were obtained on the volumes 2 and 11 with spherical coordinates (16.62, 3.54, 1.40) and (14.28, 12.70, 3.12), respectively. These results suggest that, above a certain value, the elevation angle tends to deteriorate the ACC and that the transducer should be positioned perpendicularly for optimal reconstruction. Measures performed on the volumes 9 and 10 tend to confirm this observation since the ACC significantly deviate from the mean value for azimuthal angles above 20°. Consequently, the more the axis of the transducer is close to the normal of the structure to be imaged, the better the correctness of the reconstructed volume. To our knowledge, this is the first time that such a quantitative and operator-independent validation protocol is reported for 3D US imaging techniques. Regarding the evaluation metric, a relevant alternative would have been to use the gold standard TRE instead of our tailored ACC by taking advantages of the intersections of lines as target points and increasing the number of wires of the phantom. In order for the TRE to account for local reconstruction errors, target points should yet be distributed in the acquisition space, which would thus require a 3D net of wires for the validation phantom. Given the unusual small dimensions of the acquisition window of our imaging platform, such a phantom would however remain difficult to build. Besides, the measurability of such a net of wires, and so target points, would also be hindered by the intrinsic features of our imaging platform considering the degraded lateral resolution in the far-field (Fig. 6) as well as the narrowed and conical acquisition window in the near-field.

The main strength of our setup lies in the minimally invasive aspect of this imaging technique. Contrary to [11] and [12], in which prior removal of the EAC was required due to the footprint of the utilized US probes, no ablation of the surrounding bone was needed in our experiments. From the clinician's viewpoint, the use of such a modality could facilitate the diagnostic process of the patient. Volumetric reconstruction of the auditory system could actually allow to assess the integrity of the ossicles and thus identify cases of CHL that may be caused by otosclerosis or chronic ear infections. Yet, since the real-time aspect of the acquired images is lost during the rotational sweep, 3D scans are currently restricted to structural information. As a result, clinical applications may also benefit from 2D B-scans that could provide functional information about the ossicular chain and enable to detect impairments at the incudomalleolar joint, which cannot be imaged (Fig. 12).

Regarding ENT surgeries, one possible application could be for ossicular chain reconstruction where functional information could allow to verify the prosthesis placement after putting the eardrum back in place. We also investigated the potential use of this modality for inner ear surgery by imaging a cochlear implant in different conditions (Fig. 13). As depicted in Fig. 13b, segmentation of the implant parts was possible by merely substracting acquired volumes before and after the insertion. Combined with a registered preoperative scan, this information could allow to maintain an appropriate axis of insertion, which was shown to be a critical factor in intracochlear trauma during cochlear implantation [37]. Additionally, 3D registration between pre- and intra-operative imaging of the tympanic cavity could enable surgical navigation, which would be of great use for the surgeon given the complexity of the ear anatomy.

Since calibration is inherent to the development of any surgical navigation system, we actually conducted in a previous paper a feasability study on the calibration of such an imaging apparatus [25]. Despite the unusual small dimensions of the acquisition window, effective calibration of our setup was shown to be possible using a classic scheme based on a phantom with the same reference geometry than the one used in section II-C. However, in the absence of a reliable ground truth, no conclusion could be drawn regarding the accuracy of the procedure.

Consequently, future works will focus on the evaluation of a calibration procedure of our acquisition setup and, if necessary, on improving its accuracy for the further development of a navigation system intended for otological surgery. In this perspective, a US / CBCT registration algorithm will also be needed to be investigated, with a special focus on achieving sub-mm accuracy and a quantitative validation using cadaver datasets.

V. Conclusion

In this article, we focused on the carrying out of minimally invasive 3D US for the imaging of the auditory system using a new miniaturized endoscopic 2D US transducer. The latter consists of a 18 MHz 24 elements curved array transducer with a distal diameter of 4 mm that can be inserted into the EAC. Volumetric acquisition was achieved by rotating this very probe around its own axis using a *RobOtol* system.

Since this acquisition setup is meant for medical purposes, we focused on evaluating the correctness of the acquired volume's geometry. To that end, we introduced a new protocol that relies on the automatic registration and comparison of US volumes to a ground truth micro-CT acquisition of a custom phantom including a set of wires as reference geometry. US

acquisitions obtained with twelve different probe poses were used to quantify the accuracy of volumetric reconstruction, which was estimated to achieve a maximum error of 0.2 mm. This value is consistent with the dimensions of the anatomical structures of the auditory system to be imaged.

To explore its clinical applicability in a representative anatomical context, our system was also tested on a cadaveric head. Typical structures such as the ossicles and the round window could be identified from the obtained US volumes. These results confirm that this technique enables the imaging of the middle and inner ears without requiring prior removal of the EAC. Such a system could then facilitate the minimaly invasive diagnosis and surgical navigation for otology in a fast, cost-effective and safe way.

Acknowledgment

The authors would like to thank Tony Blo, Frédéric Guérin and thanatopractitioner Maud Moulin (Faculty of Medecine of Montpellier) for their assistance in the evaluation of the acquisition setup on the cadaveric head and Dr Renaud Lebrun (Montpellier Rio Imaging platform and LabEx CeMEB - Institut des Sciences de l'Evolution - Université de Montpellier - CNRS UMR 5554 - Campus Triolet Place Eugène Bataillon 34095 Montpellier Cedex 5) for his contribution in the micro-CT acquisition.

References

[1] World Health Organization, "World report on hearing," 2021.
[2] A. C. Davis and H. J. Hoffman, "Hearing loss: rising prevalence and impact," *Bulletin of the World Health Organization*, vol. 97, no. 10, p. 646, 2019.
[3] L. L. Cunningham and D. L. Tucci, "Hearing loss in adults," *New England Journal of Medicine*, vol. 377, no. 25, pp. 2465–2473, 2017.
[4] A. F. Juliano *et al.*, "Imaging review of the temporal bone: part i. anatomy and inflammatory and neoplastic processes," *Radiology*, vol. 269, no. 1, pp. 17–33, 2013.
[5] G. Robertson and R. Mills, "Findings at exploratory tympanotomy for conductive hearing loss," *The Journal of laryngology and otology*, vol. 123, no. 10, p. 1087, 2009.
[6] M. M. Paparella and S. Koutroupas, "Exploratory tympanotomy revisited." *The Laryngoscope*, vol. 92, no. 5, pp. 531–534, 1982.
[7] A. Miracle and S. Mukherji, "Conebeam ct of the head and neck, part 2: clinical applications," *American Journal of Neuroradiology*, vol. 30, no. 7, pp. 1285–1292, 2009.
[8] D. MacDougall *et al.*, "Optical coherence tomography system requirements for clinical diagnostic middle ear imaging," *Journal of biomedical optics*, vol. 20, no. 5, p. 056008, 2015.
[9] S. S. Gurbani *et al.*, "Robot-assisted three-dimensional registration for cochlear implant surgery using a common-path swept-source optical coherence tomography probe," *Journal of biomedical optics*, vol. 19, no. 5, p. 057004, 2014.
[10] M. Bance *et al.*, "New imaging modalities in otology," *Advances in Hearing Rehabilitation*, vol. 81, pp. 1–13, 2018.
[11] J. W. Rainsbury *et al.*, "High frequency ex vivo ultrasound imaging of the middle ear to show simulated ossicular pathology," *Otology & Neurotology*, vol. 37, no. 5, pp. 586–592, 2016.
[12] T. G. Landry *et al.*, "Real-time imaging of in-vitro human middle ear using high frequency ultrasound," *Hearing research*, vol. 326, pp. 1–7, 2015.
[13] J. A. Brown *et al.*, "High-frequency ex vivo ultrasound imaging of the auditory system," *Ultrasound in medicine & biology*, vol. 35, no. 11, pp. 1899–1907, 2009.
[14] T. G. Landry *et al.*, "Real-time intracochlear imaging of automated cochlear implant insertions in whole decalcified cadaver cochleas using ultrasound," *Cochlear implants international*, vol. 19, no. 5, pp. 255–267, 2018.
[15] C. Wachinger *et al.*, "Three-dimensional ultrasound mosaicing," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2007, pp. 327–335.
[16] O. V. Solberg *et al.*, "Freehand 3d ultrasound reconstruction algorithms—a review," *Ultrasound in medicine & biology*, vol. 33, no. 7, pp. 991–1009, 2007.
[17] Q. Duan *et al.*, "Fast interpolation algorithms for real-time three-dimensional cardiac ultrasound," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, vol. 2. IEEE, 2003, pp. 1192–1195.
[18] B. Zhuang *et al.*, "Real-time 3-d ultrasound scan conversion using a multicore processor," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 571–574, 2009.
[19] E. R. Pospisil *et al.*, "4-d x 3-d ultrasound: real-time scan conversion, filtering, and display of displacement vectors with a motorized curvilinear transducer," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 57, no. 10, pp. 2271–2283, 2010.
[20] H. Andresen *et al.*, "Precise time-of-flight calculation for 3-d synthetic aperture focusing," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 9, pp. 1880–1887, 2009.
[21] S. I. Nikolov *et al.*, "Three-dimensional real-time synthetic aperture imaging using a rotating phased array transducer," in *2002 IEEE Ultrasonics Symposium, 2002. Proceedings.*, vol. 2. IEEE, 2002, pp. 1585–1588.
[22] J. A. Jensen *et al.*, "Ultrasound research scanner for real-time synthetic aperture data acquisition," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 52, no. 5, pp. 881–891, 2005.
[23] R. Monfaredi *et al.*, "Robot-assisted ultrasound imaging: Overview and development of a parallel telerobotic system," *Minimally Invasive Therapy & Allied Technologies*, vol. 24, no. 1, pp. 54–62, 2015.
[24] S. E. Salcudean *et al.*, "Robot-assisted medical imaging: A review," *Proceedings of the IEEE*, 2022.
[25] J. C. Santos *et al.*, "Geometric calibration of a new miniaturized endoscopic ultrasound probe," in *2021 20th International Conference on Advanced Robotics (ICAR)*. IEEE, 2021, pp. 228–233.
[26] M. Miroir *et al.*, "Design of a robotic system for minimally invasive surgery of the middle ear," in *2008 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 2008, pp. 747–752.
[27] L. Lavenir *et al.*, "Hfus imaging of the cochlea: A feasibility study for anatomical identification by registration with microct," *Annals of Biomedical Engineering*, vol. 49, no. 5, pp. 1308–1317, 2021.
[28] M. Miroir *et al.*, "Robotol: from design to evaluation of a robot for middle ear surgery," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 850–856.
[29] P. Thévenaz *et al.*, "Interpolation revisited [medical images application]," *IEEE Transactions on medical imaging*, vol. 19, no. 7, pp. 739–758, 2000.
[30] J. M. Thijssen *et al.*, "Objective performance testing and quality assurance of medical ultrasound equipment," *Ultrasound in medicine & biology*, vol. 33, no. 3, pp. 460–471, 2007.
[31] O. V. Solberg *et al.*, "3d ultrasound reconstruction algorithms from analog and digital data," *Ultrasonics*, vol. 51, no. 4, pp. 405–419, 2011.
[32] S. Wold *et al.*, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
[33] M. Uherčík *et al.*, "Model fitting using ransac for surgical tool localization in 3-d ultrasound images," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 1907–1916, 2010.
[34] B. Kamgar-Parsi, "Algorithms for matching 3d line sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 582–593, 2004.
[35] G. R. Lockwood *et al.*, "Real-time 3-d ultrasound imaging using sparse synthetic aperture beamforming," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 45, no. 4, pp. 980–988, 1998.
[36] M. Yang *et al.*, "Separable beamforming for 3-d medical ultrasound imaging," *IEEE transactions on signal processing*, vol. 63, no. 2, pp. 279–290, 2014.
[37] R. Torres *et al.*, "Cochlear implant insertion axis into the basal turn: a critical factor in electrode array translocation," *Otology & Neurotology*, vol. 39, no. 2, pp. 168–176, 2018.