

# Windowed Radon Transform for Robust Speed-of-Sound Imaging With Pulse-Echo Ultrasound

Samuel Beuret<sup>1</sup>, *Graduate Student Member, IEEE*, Baptiste Hériard-Dubreuil<sup>2</sup>, *Student Member, IEEE*, Naiara Korta Martiartu<sup>3</sup>, *Member, IEEE*, Michael Jaeger<sup>4</sup>, *Member, IEEE*, and Jean-Philippe Thiran<sup>5</sup>, *Senior Member, IEEE*

**Abstract**—In recent years, methods estimating the spatial distribution of tissue speed of sound with pulse-echo ultrasound are gaining considerable traction. They can address limitations of B-mode imaging, for instance in diagnosing fatty liver diseases. Current state-of-the-art methods relate the tissue speed of sound to local echo shifts computed between images that are beamformed using restricted transmit and receive apertures. However, the aperture limitation affects the robustness of phase-shift estimations and, consequently, the accuracy of reconstructed speed-of-sound maps. Here, we propose a method based on the Radon transform of image patches able to estimate local phase shifts from full-aperture images. We validate our technique on simulated, phantom and in-vivo data acquired on a liver and compare it with a state-of-the-art method. We show that the proposed method enhances the stability to changes of beamforming speed of sound and to a reduction of the number of insonifications. In particular, the deployment of pulse-echo speed-of-sound estimation methods onto portable ultrasound devices can be eased by the reduction of the number of insonifications allowed by the proposed method.

**Index Terms**—Pulse-echo ultrasound, Radon transform, speed-of-sound imaging, inverse problem, liver imaging.

Manuscript received 11 August 2023; revised 19 October 2023; accepted 7 December 2023. Date of publication 18 December 2023; date of current version 3 April 2024. This work was supported by the École Polytechnique Fédérale de Lausanne (EPFL). (*Corresponding author: Samuel Beuret.*)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Samuel Beuret is with the Laboratory of Signal Processing 5, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (e-mail: samuel.beuret@epfl.ch).

Baptiste Hériard-Dubreuil is with E-Scopics, 13090 Aix-en-Provence, France, and also with the Laboratory of Signal Processing 5, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (e-mail: baptiste.heriard-dubreuil@e-scopics.com).

Naiara Korta Martiartu and Michael Jaeger are with the Institute of Applied Physics, University of Bern, 3012 Bern, Switzerland (e-mail: naiara.korta@unibe.ch; michael.jaeger@unibe.ch).

Jean-Philippe Thiran is with the Laboratory of Signal Processing 5, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland, and also with the Radiology Department, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1005 Lausanne, Switzerland (e-mail: jean-philippe.thiran@epfl.ch).

Digital Object Identifier 10.1109/TMI.2023.3343918

## I. INTRODUCTION

THE biomarker underlying brightness (B)-mode pulse-echo ultrasound imaging is the tissue reflectivity function (TRF). Whereas B-mode imaging is able to indicate the presence of a wide range of conditions, it can suffer from both non-specific contrast and low sensitivity to certain disease types. In recent years, there has been a rise in methods seeking to image other biomarkers that can potentially address the shortcomings of B-mode imaging. An excellent example is the elastography techniques that reveal spatial variations of tissue stiffness [1].

Methods to image local tissue speed of sound (SoS) are also gaining considerable attention. For instance, through-transmission techniques are able to estimate the spatial distribution of SoS [2], [3], [4], [5], [6] and have led to promising results in diagnosing breast tumors [7], [8]. However, they acquire data using transducer arrays surrounding an acoustically transparent medium, limiting their application mostly to breast imaging. They also involve a dedicated heavy system that cannot leverage the advantages of low cost and portability that are generally associated with ultrasound imaging.

To circumvent this limitation, various techniques estimating SoS from pulse-echo measurements have recently been developed. Methods assuming a uniform SoS in the imaged medium are well mastered and have shown to be particularly useful for diagnosing fatty liver disease [9], [10], [11]. However, this assumption is restrictive and does not capture the actual tissue complexity. In an attempt to reconstruct the spatial distribution of tissue SoS, several techniques proposed to approximate the imaged tissues as a series of horizontal layers with uniform SoS [12], [13], [14], [15]. With regard to liver imaging, this approximation is justified for locations where the different tissues of the abdominal wall and the liver are parallel. However, it cannot be extended to a more realistic scenario with lateral variations of tissue composition. Moreover, such methods cannot image inclusions with inhomogeneous SoS, occurring typically in the case of tumors.

Tomographic pulse-echo ultrasound methods drop the assumption of lateral uniformity to fully recover the tissue SoS

distribution. They require a localized detection of aberration phase shifts—or equivalently displacements between images—and inverting a linear model that relates these measurements to tissue SoS, assuming straight rays.

Such a procedure was first proposed by Jaeger et al. [16], [17], who computed phase shifts by cross-correlating images corresponding to different transmit (Tx) angles and reconstructed the SoS image in the spatial frequency domain. Since then, several improvements have been proposed, such as a spatial model able to increase the spatial domain of the SoS reconstruction [18]. Particularly relevant to our work is the approach developed by Stähli et al. [19], [20], recently extended to curved arrays [21], [22]. Here, phase shifts are detected by jointly limiting the Tx and receive (Rx) angular apertures and comparing images with Tx/Rx angle pairs grouped around a set of common mid angles. However, the limitation of Tx and Rx angular apertures restricts the number of signals combined to reconstruct an image. The sensitivity to phase errors occurring in each signals is thus worsened. As a result, the robustness of aberration phase-shift estimation is reduced, along with the quality of SoS images reconstructed in practice. In particular, this fact increases the number of insonifications the method requires and hinders its implementation on portable devices since they have limited power, memory and data transfer bandwidth.

To circumvent this limitation, we recently investigated a method to estimate aberration phase shifts based on the Radon transform of the cross-correlation between patches extracted from different images [23]. The use of the Radon transform to estimate aberration phase shifts is supported by the change of basis from the canonical receiver basis to plane waves (PWs) basis—sometimes denoted as Radon domain—that has been investigated recently [24], [25], [26]. The rationale behind our approach is to express data in a local PW basis rather than in a global PW basis to take into account the locality of the aberration phase shifts. Based on this reasoning, we recently introduced an adaptive beamformer for ultrafast ultrasound able to correct for SoS aberrations [27].

In this article, our aim is to propose an improved version of the technique presented in [23]. In particular, we want to

- establish the mathematical formalization of the use of the windowed Radon transform for the computation of aberration phase shifts,
- demonstrate that the proposed method is able to provide meaningful results when data is acquired using both large and limited numbers of insonifications,
- quantify the robustness of the proposed method and compare it to a state-of-the-art method [19].

## II. THEORY

In this section, we detail the theoretical aspects of the proposed method. We begin by characterizing a novel complex radio-frequency (CRF) image beamforming procedure referred to as *constant-dif-angle beamforming*. Then, we show that the windowed Radon transform of images reconstructed using constant-dif-angle beamforming can be used to estimate phase aberrations caused by SoS variations in the imaged tissues.

### A. Plane-Wave Measurement Model

The theory developed in this section relies on several assumptions. First, we assume that the transducer emits a PW steered at an arbitrary Tx angle  $\theta^{\text{Tx}}$  and measures, for each time instant  $t$ , the radio-frequency (RF) signals associated with a PW steered at Rx angles  $\theta^{\text{Rx}}$ . Therefore, we can define a function  $m^{\text{PW}}(\theta^{\text{Tx}}, \theta^{\text{Rx}}, t) \in \mathbb{C}$  describing the CRF echo signals scattered within the insonified medium and measured by the transducer.

According to [28], we can model the measured echo signals expressed in a PW basis  $m^{\text{PW}}$  as

$$m^{\text{PW}}(\theta^{\text{Tx}}, \theta^{\text{Rx}}, t) = v_{\text{pe}}(t) *_t \int_{\mathbf{r}} \left[ \tilde{h}^{\text{Tx}}(\theta^{\text{Tx}}, \mathbf{r}, t) *_t \tilde{h}^{\text{Rx}}(\theta^{\text{Rx}}, \mathbf{r}, t) \right] \gamma(\mathbf{r}) d\mathbf{r}. \quad (1)$$

We denote by  $\gamma(\mathbf{r})$  the complex and continuous TRF generating the echo signals, with  $\mathbf{r} = [x, z]$  a point in the insonified medium. Functions  $\tilde{h}^{\text{Tx}}$  and  $\tilde{h}^{\text{Rx}}$  refer to the Tx and Rx spatial impulse responses (SIRs), respectively. Temporal convolution is denoted by  $*_t$ . Finally,  $v_{\text{pe}}$  stands for the pulse-echo wavelet, a term including the electromechanical impulse response of the transducer and the electric excitation waveform. It is assumed to be a complex analytic signal.

To reduce the complexity of the measurement model and ease its mathematical analysis, we follow the simplifications proposed in [29], [30], where the authors reduce the convolution with the SIRs  $\tilde{h}^{\text{Tx/Rx}}$  to a temporal shift according to the Tx and Rx propagation times and a multiplication with real functions  $h^{\text{Tx/Rx}}$ . To do so, the authors posit that the entire medium is located in the far-field of an element of the transducer. Under this assumption, we rewrite (1) as

$$m^{\text{PW}}(\theta^{\text{Tx}}, \theta^{\text{Rx}}, t) \approx \int_{\mathbf{r}} h^{\text{Tx}}(\theta^{\text{Tx}}, \mathbf{r}) h^{\text{Rx}}(\theta^{\text{Rx}}, \mathbf{r}) v_{\text{pe}}\left(t - \tau(\theta^{\text{Tx}}, \theta^{\text{Rx}}, \mathbf{r}) - \Delta\tau(\theta^{\text{Tx}}, \theta^{\text{Rx}}, \mathbf{r})\right) \gamma(\mathbf{r}) d\mathbf{r}. \quad (2)$$

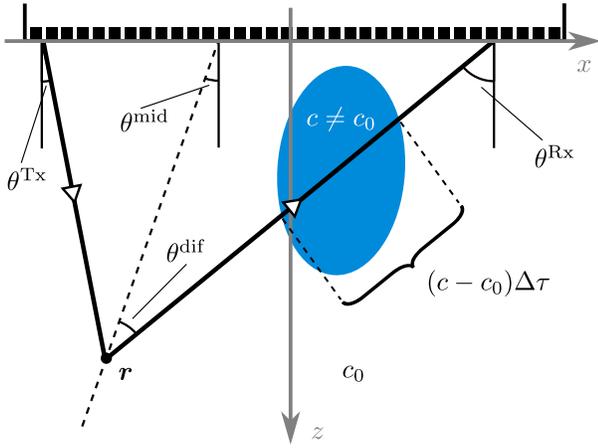
Here, the propagation times are expressed as the sum of two distinct terms. The first term  $\tau$  is the expected round-trip time associated with Tx and Rx PWs. Assuming a uniform SoS  $c_0$ , it is defined as

$$\begin{aligned} \tau(\theta^{\text{Tx}}, \theta^{\text{Rx}}, \mathbf{r}) &= \frac{1}{c_0} \langle \mathbf{u}_{\theta^{\text{Tx}}}, \mathbf{r} \rangle + \frac{1}{c_0} \langle \mathbf{u}_{\theta^{\text{Rx}}}, \mathbf{r} \rangle \\ &= \frac{2}{c_0} \cos(\theta^{\text{dif}}) \langle \mathbf{u}_{\theta^{\text{mid}}}, \mathbf{r} \rangle, \end{aligned} \quad (3)$$

with

$$\mathbf{u}_{\theta} = \begin{bmatrix} \sin(\theta) \\ \cos(\theta) \end{bmatrix}, \quad \theta^{\text{mid}} = \frac{\theta^{\text{Tx}} + \theta^{\text{Rx}}}{2}, \quad \theta^{\text{dif}} = \frac{\theta^{\text{Tx}} - \theta^{\text{Rx}}}{2}, \quad (4)$$

and where  $\langle \cdot, \cdot \rangle$  denotes the dot product. Throughout this article, we refer to the angles  $\theta^{\text{mid}}$  and  $\theta^{\text{dif}}$  as the *mid angle* and *dif angle*, respectively. The second term  $\Delta\tau$  is the aberration delay caused by deviations of local SoS from  $c_0$  along both the Tx and Rx paths. Our model of  $\Delta\tau$  will be detailed in Section III-D. A summary of the proposed configuration is depicted in Fig. 1.



**Fig. 1.** Studied configuration. A linear transducer emits a plane wave with steering angle  $\theta^{\text{Tx}}$ . The data reflected at a point  $r$  in the medium is received by the transducer with an angle  $\theta^{\text{Rx}}$ . An inclusion with a speed of sound  $c$  different from the assumed speed  $c_0$  introduces an aberration delay  $\Delta\tau$ .

The far-field hypothesis introduced in (2) is not entirely met in practice, since we do also consider echo signals scattered in the near field of the transducer's elements. Therefore, the frequency content of such signals differ from the ones predicted by (2). However, as long as their center frequencies do not deviate significantly from the center frequency of  $v_{\text{pe}}$ , the far-field hypothesis bears only a limited impact on the validity of our mathematical analysis.

### B. Constant-Dif-Angle Beamforming

We now introduce constant-dif-angle beamforming. The basic principle of the method is to reconstruct images by taking into account only signals  $m^{\text{PW}}$  with the same difference between their Tx and Rx angles. We propose to reconstruct a series of  $N^{\text{dif}}$  constant-dif-angle images  $y_k^{\text{dif}}$  corresponding to dif-angles  $\theta_k^{\text{dif}}$ , with  $k = 1, \dots, N^{\text{dif}}$ . They are defined mathematically as

$$y_k^{\text{dif}}(\mathbf{r}') = \int_{\theta^{\text{mid}}} a(\theta_k^{\text{dif}}, \theta^{\text{mid}}, \mathbf{r}') m^{\text{PW}} \left( \theta^{\text{mid}} + \theta_k^{\text{dif}}, \theta^{\text{mid}} - \theta_k^{\text{dif}}, \frac{2 \cos(\theta_k^{\text{dif}})}{c_0} \langle \mathbf{u}_{\theta^{\text{mid}}}, \mathbf{r}' \rangle \right) d\theta^{\text{mid}}, \quad (5)$$

with  $\mathbf{r}'$  denoting the position in the image. In (5), we delay the signals  $m^{\text{PW}}$  according to the propagation times defined in (3). We denote by  $a$  arbitrary apodization weights applied to each signal at each position  $\mathbf{r}'$  of the beamformed image. Finally, we assume implicitly in (5) that we have access to  $m^{\text{PW}}$  for a continuous interval of  $\theta^{\text{Tx}}$  and  $\theta^{\text{Rx}}$ .

As implied in [31], [32], DAS and coherent compounding only improve image resolution when signals corresponding to different mid angles  $\theta^{\text{mid}}$  are summed. In turn, the summation of signals corresponding to different dif angles but the same mid angle provides robustness to artifacts—for example, multiple scattering and clutter—since the signals are redundant. The suggested constant-dif-angle beamformer aims at decoupling explicitly the two effects of beamforming,

namely focusing and artifact reduction, so as to only exploit the former. We ultimately propose an alternative way to exploit the redundancy existing between constant-dif-angle images, namely to estimate aberration phase shifts rather than to reduce artifacts.

### C. Windowed Radon Transform

Given the constant-dif-angle images  $y_k^{\text{dif}}$ , we want to estimate a quantity linked to the aberration delays at an arbitrary point  $\mathbf{r}^0$  in the medium. To do so, we first propose to multiply each image  $y_k^{\text{dif}}$  by a window  $w^{\text{Rad}}$  centered in  $\mathbf{r}^0$ . The windowing allows us to restrict the image to the vicinity of  $\mathbf{r}^0$ . In a second step, we compute the Radon transform of the result. The full operation is a *windowed Radon transform*, and we define it mathematically as

$$y_k^{\text{Rad}}(\theta^{\text{Rad}}, d, \mathbf{r}^0) = \int_{\mathbf{r}'} w^{\text{Rad}}(\mathbf{r}' - \mathbf{r}^0) \delta(\langle \mathbf{u}_{\theta^{\text{Rad}}}, \mathbf{r}' - \mathbf{r}^0 \rangle - d) y_k^{\text{dif}}(\mathbf{r}') d\mathbf{r}', \quad (6)$$

where  $d$  is the offset along  $\mathbf{u}_{\theta^{\text{Rad}}}$  with respect to  $\mathbf{r}^0$ , and  $\theta^{\text{Rad}}$  is the Radon filtering angle.

We now introduce two fundamental assumptions. First, we assume that the point-spread function (PSF) of  $y_k^{\text{dif}}$  is sufficiently small with respect to the size of the window. It allows us to approximate the value of the window at the image position  $\mathbf{r}'$  with the value of the window at the medium position  $\mathbf{r}$

$$w^{\text{Rad}}(\mathbf{r}' - \mathbf{r}^0) \approx w^{\text{Rad}}(\mathbf{r} - \mathbf{r}^0). \quad (7)$$

Second, we assume that  $h^{\text{Tx}}$ ,  $h^{\text{Rx}}$ , and  $\Delta\tau$  vary slowly in the medium and can be assumed constant in the vicinity of  $\mathbf{r}^0$ —defined as  $\{\mathbf{r}' : w^{\text{Rad}}(\mathbf{r}' - \mathbf{r}^0) > 0\}$ . Likewise, we restrict the choice of  $a$  in (5) to slowly-varying functions such that it can also be assumed constant in the vicinity of  $\mathbf{r}^0$ .

To recover the aberration phase-shifts, we propose to estimate the Fourier transform  $\hat{y}_k^{\text{Rad}}$  of  $y_k^{\text{Rad}}$  with respect to  $d$ , given the two hypotheses detailed above. Considering (2) and (5) and after a series of algebraic manipulations detailed in Appendix A, we obtain

$$\hat{y}_k^{\text{Rad}}(\theta^{\text{Rad}}, \xi, \mathbf{r}^0) = \frac{2\pi c_0}{2 \cos(\theta_k^{\text{dif}}) \xi} g_k(\theta^{\text{Rad}}, \mathbf{r}^0) \hat{v}_{\text{pe}} \left( \frac{c_0 \xi}{2 \cos(\theta_k^{\text{dif}})} \right) e^{-j \frac{c_0 \xi}{2 \cos(\theta_k^{\text{dif}})} \Delta\tau (\theta_k^{\text{dif}}, \theta^{\text{Rad}}, \mathbf{r}^0)} \hat{\gamma}^w(\xi \mathbf{u}_{\theta^{\text{Rad}}}, \mathbf{r}^0), \quad (8)$$

where  $\xi$  is the frequency coordinate associated with the variable  $d$ . In (8),  $g_k$  denotes a positive function encompassing the effects of  $h^{\text{Tx/Rx}}$  and  $a$ ,  $\hat{v}_{\text{pe}}$  is the Fourier transform of  $v_{\text{pe}}$ , and  $\hat{\gamma}^w$  represents a two-dimensional complex function factoring in the reflectivity map  $\gamma$  and the window  $w$ .

Two key insights can be inferred from (8). First, the Radon angle  $\theta^{\text{Rad}}$  used in the windowed Radon transform allows us to recover the signals associated exactly with a mid angle  $\theta^{\text{mid}}$ . From now on the two angles are thus considered equivalent. Second, each signal  $y_k^{\text{Rad}}(\theta^{\text{Rad}}, \cdot, \mathbf{r}^0)$  is associated with a

specific value of the aberration delay  $\Delta\tau(\theta_k^{\text{dif}}, \theta^{\text{Rad}}, \mathbf{r}^0)$ . The value of the aberration delay is parameterized by the dif angle  $\theta_k^{\text{dif}}$  associated with the CRF image  $y_k^{\text{dif}}$ , the Radon angle  $\theta^{\text{Rad}}$  of the windowed Radon transform and the position  $\mathbf{r}^0$  of the window.

#### D. Zero-Lag Cross-Correlation

Unfortunately, the value of  $\Delta\tau$  cannot be directly deduced from  $y_k^{\text{Rad}}$  according to (8), since it is impossible to know the phase of  $\hat{\gamma}^w$  a priori. To circumvent this issue, we propose to follow [16], [19] and compute the phase of the zero-lag cross correlation between the signals associated with two different dif angles  $\theta_k$  and  $\theta_{k+1}$

$$\Delta\phi_{k \rightarrow k+1}(\theta^{\text{Rad}}, \mathbf{r}^0) = \arg \left\{ \int_d \left( y_{k+1}^{\text{Rad}} \right)^* (\theta^{\text{Rad}}, d, \mathbf{r}^0) y_k^{\text{Rad}}(\theta^{\text{Rad}}, d, \mathbf{r}^0) dd \right\}, \quad (9)$$

where  $*$  denote the complex conjugate. Upon inserting (8) into (9) expressed in the frequency domain, we obtain

$$\Delta\phi_{k \rightarrow k+1}(\theta^{\text{Rad}}, \mathbf{r}^0) = \arg \left\{ \int_{\xi} \frac{1}{\xi^2} e^{-j \frac{c_0 \xi}{2 \cos(\theta_k^{\text{dif}})} \Delta\tau(\theta_k^{\text{dif}}, \theta^{\text{Rad}}, \mathbf{r}^0)} e^{j \frac{c_0 \xi}{2 \cos(\theta_{k+1}^{\text{dif}})} \Delta\tau(\theta_{k+1}^{\text{dif}}, \theta^{\text{Rad}}, \mathbf{r}^0)} \hat{v}_{\text{pe}} \left( \frac{c_0 \xi}{2 \cos(\theta_k^{\text{dif}})} \right) \hat{v}_{\text{pe}}^* \left( \frac{c_0 \xi}{2 \cos(\theta_{k+1}^{\text{dif}})} \right) |\hat{\gamma}^w(\xi \mathbf{u}_{\theta^{\text{Rad}}, \mathbf{r}^0})|^2 d\xi \right\}. \quad (10)$$

To conclude, we first posit that  $\theta_k^{\text{dif}}$  and  $\theta_{k+1}^{\text{dif}}$  are sufficiently small such that

$$\hat{v}_{\text{pe}} \left( \frac{c_0 \xi}{2 \cos(\theta_k^{\text{dif}})} \right) \hat{v}_{\text{pe}}^* \left( \frac{c_0 \xi}{2 \cos(\theta_{k+1}^{\text{dif}})} \right) \approx \left| \hat{v}_{\text{pe}} \left( \frac{c_0 \xi}{2} \right) \right|^2. \quad (11)$$

We also assume, for now, that  $v_{\text{pe}}$  is narrow-band around its center frequency  $\omega_0$ , such that  $\xi \approx 2\omega_0/c_0$  in the second line of (10). Under the two hypothesis introduced above, the integrand in (10) becomes real and positive, allowing us to reduce the value of the aberration phase shift to

$$\Delta\phi_{k \rightarrow k+1}(\theta^{\text{Rad}}, \mathbf{r}^0) \approx \frac{\omega_0}{\cos(\theta_{k+1}^{\text{dif}})} \Delta\tau(\theta_{k+1}^{\text{dif}}, \theta^{\text{Rad}}, \mathbf{r}^0) - \frac{\omega_0}{\cos(\theta_k^{\text{dif}})} \Delta\tau(\theta_k^{\text{dif}}, \theta^{\text{Rad}}, \mathbf{r}^0). \quad (12)$$

In practice, however, the pulse-echo wavelet  $v_{\text{pe}}$  must possess a fractional bandwidth large enough to limit the size of the PSF of beamformed images—in the order of 50% to 100%. Fortunately, a simple mathematical argument shows that (12) also holds if the function multiplying the complex exponential in the integrand of (10) is real and symmetric around the center spatial frequency  $2\omega_0/c_0$ . A similar argument is set forth in previous works on local SoS estimation [16], [19] and aberration correction [33]. The symmetry property does not hold exactly in practice due to variations in  $\hat{\gamma}^w$  and the presence of  $1/\xi^2$  in (10). Therefore, the phase estimation is expected to suffer from a certain amount of noise.

According to (12), the phase shift  $\Delta\phi_{k \rightarrow k+1}$  measured between two signals  $y_{k+1}^{\text{Rad}}$  and  $y_k^{\text{Rad}}$  is eventually given

by a weighted difference of the local aberration delays  $\Delta\tau$  associated with each signal. At last, constant-dif-angle beamforming, windowed Radon transform, and zero-lag cross-correlation successfully allows us to recover a quantity that is directly linked to the local aberration delays. Importantly, we must highlight that (12) is valid as long as the weighted difference between the aberration delays is small. If the difference is too large—namely, if its magnitude is greater than  $\pi$ —, phase wrapping occurs, and we obtain inaccurate phase-shift estimates. To minimize the risk of phase wrapping, we must ensure during beamforming that adjacent dif angles  $\theta_k^{\text{dif}}$  and  $\theta_{k+1}^{\text{dif}}$  are always sufficiently close.

Following the mathematical analysis presented here, we propose the following procedure to estimate a SoS map of the insonified medium. First, we beamform constant-dif-angle CRF images and compute their windowed Radon transform. Second, we compute the zero-lag cross-correlation between the windowed Radon transform of different images to estimate a series of phase-shift maps  $\Delta\phi_{k \rightarrow k+1}$ . Third, we define a measurement model relating the phase shift maps to the aberration delays (12), which are in turn a function of the SoS map. Ultimately, we solve a regularized inverse problem to estimate the SoS map, given the phase-shift maps  $\Delta\phi_{k \rightarrow k+1}$ .

### III. METHOD

In this section, we detail the practical aspects of the proposed method, beginning with the ultrasound transducer, system, and sequence used. Then, we present our implementation of both the constant-dif-angle beamforming and the aberration phase shift estimation. Finally, we describe the inverse problem used to recover the SoS map from the phase-shift estimates.

#### A. Ultrasound System and Sequence

The experiments shown herein are performed using an L7-4 linear vascular transducer (ATL Philips, WA, USA) attached to a Vantage 64 LE (Verasonics Inc., WA, USA) research US system. The transducer consists of  $N^s = 128$  piezoelectric elements with 0.29 mm pitch and a bandwidth ranging from 4 MHz to 7 MHz, with a 4.8 MHz center frequency. We posit that the transducer is aligned with the x-axis, such that the piezoelectric elements are located at positions  $\mathbf{r}_j^s = [x_j^s, 0]^T$ ,  $j = 1, \dots, N^s$ . We acquire data by emitting sequentially  $N^{\text{Tx}} = 115$  steered PWs ranging from  $-28.5^\circ$  to  $28.5^\circ$  with a  $0.5^\circ$  angular step size. In practice, the data acquisition of a single PW requires two insonifications since we record the echo signals using two different 64 elements sub-arrays. Furthermore, the measured real RF signals are sampled at 19.2 MHz and the CRF signals are obtained by computing the Hilbert transform of RF signals.

We also test the performance of the proposed method with a reduced amount of data. Hence, we generate a second data set by downsampling the set of steering angles by a factor 10 prior to beamforming, simulating data acquisition with only 11 PWs ranging from  $-25^\circ$  to  $25^\circ$  with a  $5^\circ$  step.

We must highlight that the maximum magnitude of the steering angles is large with respect to the pitch of the transducer. Therefore, non-negligible grating-lobe artifacts appear in the beamformed images, affecting the accuracy of phase-shift measurements. However, these errors are minor compared to the benefits obtained with large steering angles in mitigating the ill-posedness of the SoS inverse problem [16], [18], [19].

## B. Beamforming

Constant-dif-angle images cannot be computed directly according to (5) for two important reasons. First, we do not measure signals associated to Rx PWs since the signals are acquired at specific sensor positions  $\mathbf{r}_j^s$ . Second, the Tx angles  $\theta^{\text{Tx}}$  are inherently discrete since they correspond to the steering angles of the PW insonifications. To circumvent these issues, we propose an alternative method to estimate constant-dif-angle images  $y_k^{\text{dif}}$ . Following [19], we first reconstruct a single image per insonification using DAS

$$y_i^{\text{Tx}}(\mathbf{r}') = \sum_{j=1}^{N^s} a^{\text{Rx}}(\mathbf{r}' - \mathbf{r}_j^s) m_{i,j} \left( \frac{1}{c_0} \langle \mathbf{u}_{\theta_i^{\text{Tx}}}, \mathbf{r}' \rangle + \frac{1}{c_0} \|\mathbf{r}' - \mathbf{r}_j^s\| \right), \quad (13)$$

where  $m_{i,j}$  denotes the continuous CRF signal associated with steering angle  $\theta_i^{\text{Tx}}$  and sensor  $\mathbf{r}_j^s$ , and  $a^{\text{Rx}}$  denotes an adpodization function. Since the signals  $m_{i,j}$  are in reality discrete, we implement (13) using cubic interpolation.

Then, we introduce a series of elementary images  $y_{i,j}^{\text{Tx,Rx}}$ , each one associated with the Tx angle  $\theta_i^{\text{Tx}}$ ,  $i = 1, \dots, N^{\text{Tx}}$  and a Rx angle  $\theta_j^{\text{Rx}}$ ,  $j = 1, \dots, N^{\text{Rx, filt}}$ . The elementary images  $y_{i,j}^{\text{Tx,Rx}}$  are obtained by filtering in the spatial frequency domain the images  $y_i^{\text{Tx}}$  according to the mid angle of  $\theta_i^{\text{Tx}}$  and  $\theta_j^{\text{Rx}}$ , namely

$$\hat{y}_{i,j}^{\text{Tx,Rx}}(\mathbf{v}') = \delta \left( \langle \mathbf{u}_{\theta^{\text{mid}}}, \mathbf{v}' \rangle \right) \hat{y}_i^{\text{Tx}}(\mathbf{v}'), \quad (14)$$

$$\theta^{\text{mid}} = \frac{\theta_i^{\text{Tx}} + \theta_j^{\text{Rx}}}{2},$$

where  $\hat{y}_{i,j}^{\text{Tx,Rx}}$ ,  $\hat{y}_i^{\text{Tx}}$  are the Fourier transform of  $y_{i,j}^{\text{Tx,Rx}}$  and  $y_i^{\text{Tx}}$  respectively,  $\mathbf{v}'$  refers to the spatial frequency associated with  $\mathbf{r}'$ , and  $\mathbf{u}_{\theta^{\text{mid}}}$  is a unit vector perpendicular to  $\mathbf{u}_{\theta^{\text{mid}}}$ . Ultimately, we estimate the dif-angle images  $y_k^{\text{dif}}$  according to

$$y_k^{\text{dif}}(\mathbf{r}) \approx \sum_{i=1}^{N^{\text{Tx}}} \sum_{j=1}^{N^{\text{Rx, filt}}} w_k^{\text{dif}} \left( \frac{\theta_i^{\text{Tx}} - \theta_j^{\text{Rx}}}{2} - \theta_k^{\text{dif}} \right) w_k^{\text{mid}} \left( \frac{\theta_i^{\text{Tx}} + \theta_j^{\text{Rx}}}{2} \right) y_{i,j}^{\text{Tx,Rx}}(\mathbf{r}), \quad (15)$$

with

$$w_k^{\text{dif}}(\theta) = \begin{cases} \cos^2 \left( \frac{\pi \theta}{2\alpha^{\text{dif}}} \right), & \text{if } -\alpha^{\text{dif}} < \theta < \alpha^{\text{dif}} \\ 0, & \text{otherwise} \end{cases}, \quad (16)$$

a Hann window of aperture  $\alpha^{\text{dif}}$ . We denote by  $w_k^{\text{mid}}$  a Tukey window of aperture  $\max_i |\theta_i^{\text{Tx}} - \theta_k^{\text{dif}}|$  and 0.25 cosine

fraction. This window discards elementary images  $\hat{y}_{i,j}^{\text{Tx,Rx}}$  corresponding to angles that are not considered in the phase estimation.

In practice, we evaluate (13) on a discrete grid of points  $\mathbf{r}'_{m,n}$ ,  $m = 1, \dots, N^{x,y}$ ,  $n = 1, \dots, N^{z,y}$ , with  $N^{x,y} = 1400$  and  $N^{z,y} = 1500$ . We define the grid spacing as  $\lambda_0/8$  along both  $x$  and  $z$  axes, such that  $\lambda_0 = c_0/f_0$ , where  $f_0$  is the center frequency of the transducer and  $c_0$  the beamforming SoS. The grid spacing along  $z$  is determined to ensure that the Nyquist criterion is fulfilled for spatial frequencies corresponding to twice the center frequency of the transducer. The grid spacing in the  $x$  direction is equal to its  $z$ -axis counterpart since an isotropic grid eases the implementation of the proposed method.

We define the apodization function  $a^{\text{Rx}}$  as a Tukey window acting on the Rx angle, with a  $60^\circ$  angular aperture and a 0.125 cosine fraction. The  $60^\circ$  aperture is chosen to discard signals with Rx angles larger than the maximum Tx angle— $28.5^\circ$  plus a  $1.5^\circ$  margin to account for the cosine fraction—since these signals do not intervene in the aberration phase shift estimation (c.f. Section III-C). Grating-lobe artifacts are thus reduced [34], since their likelihood decreases with the angular aperture considered. Respectively, a non-zero cosine fraction is necessary for function  $a^{\text{Rx}}$  to be smooth (c.f. Section II-C). The same reasoning applies for  $w_k^{\text{mid}}$  in (15). We consider  $N^{\text{Rx, filt}} = 581$  Rx angles ranging from  $-30^\circ$  to  $30^\circ$  to cover the aperture of  $a^{\text{Rx}}$ . Such a high number of Rx angles is required to cover the full spatial spectrum of  $y_k^{\text{dif}}(\mathbf{r})$  in (14). Detrimental grating-lobe-like artifact can occur if  $N^{\text{Rx, filt}}$  is too small.

In (15), we would ideally sum up only the elementary images with dif-angles exactly equal to  $\theta_k^{\text{dif}}$ . This is not possible in practice for arbitrary values of  $\theta_k^{\text{dif}}$  since  $\theta^{\text{Tx}}$  and  $\theta^{\text{Rx}}$  are discrete. Therefore, we choose to weight the different elementary images according to a Hann window. Its aperture  $\alpha^{\text{dif}}$  needs to be large enough for two reasons. First, to ensure a sufficient sampling of the frequency plane. Detrimental grating-lobe-like artifacts can appear in  $y_k^{\text{dif}}(\mathbf{r})$  if  $\alpha^{\text{dif}}$  is too small compared to the Tx angle spacing. Second, to achieve sufficient robustness to artifacts such as multiple scattering and clutter, as discussed in Section II-B. However, the accuracy of the phase shift estimation is reduced if  $\alpha^{\text{dif}}$  is too large since the hypotheses introduced in Section II-C are mitigated. We must therefore determine empirically the optimal value of  $\alpha^{\text{dif}}$ . In particular, we set  $\alpha^{\text{dif}} = 5^\circ$ . This value gave us the best trade-off in terms of SoS contrast, SoS granularity, and the noise level in phase-shift maps when 11 PWs in-vivo data are considered. Eventually, we reconstruct  $N^{\text{dif}} = 21$  constant-dif-angles images  $y_k^{\text{dif}}$  with  $\theta_k^{\text{dif}}$  ranging from  $-20^\circ$  to  $20^\circ$  with a  $2^\circ$  step in accordance with [19], [35]. We depict in Fig. 2 a visual summary of the beamforming procedure.

## C. Aberration Phase Shift Estimation

Once the constant-dif-angle images are computed, we apply (6) and (9) to estimate the local phase-shift maps  $\Delta\phi$ .

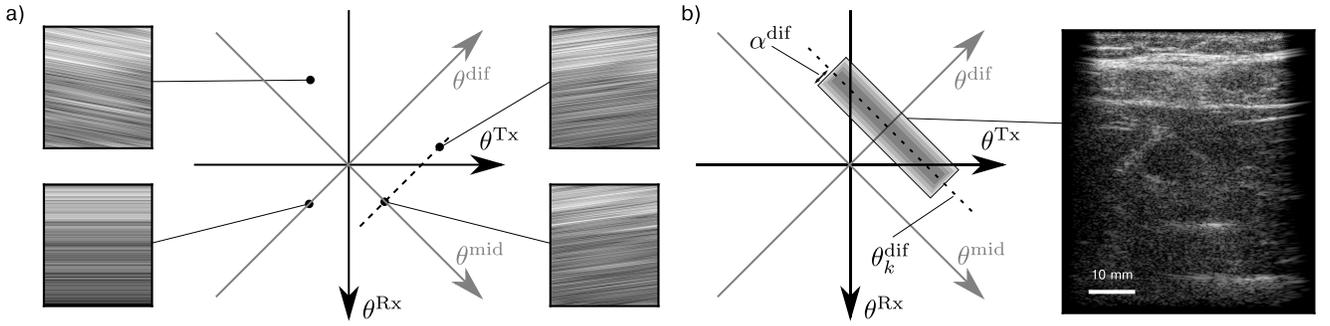


Fig. 2. Description of the beamforming process: a) Spatial filtering of beamformed single-insonification images according to a mid angle  $\theta^{mid} = (\theta^{Tx} + \theta^{Rx})/2$ . b) Reconstruction of a single constant-dif-angle image by summing images weighted by a Hann window centered at dif angle  $\theta_k^{dif}$  with aperture  $\alpha^{dif}$ .

We must highlight that hypothesis (7) introduces a limit to the resolution of  $y_k^{Rad}$  with respect to  $\theta^{Rad}$ . The limit stems from the convolution in the frequency domain between the spectra of the images and the window  $w^{Rad}$ . To maximize the bound on the resolution of  $y_k^{Rad}$  with respect to  $\theta^{Rad}$ , we choose a circular window  $w^{Rad}$  since it is characterized by a narrow main lobe. In addition, the radius of the window  $R^{Rad}$  must be sufficiently large to enforce (7) and provide robustness to artifacts, but sufficiently small to posit that  $a$ ,  $h$ , and  $\Delta\tau$  are constant within  $w^{Rad}$ . An excessive value of  $R^{Rad}$  would result in overly smoothed phase-shift maps, thus reducing the accuracy of the SoS reconstruction. Since the magnitude of the spatial variations of  $\Delta\tau$  and the amount of artifacts in the images vary significantly with the imaged medium, we select  $R^{Rad}$  using the same criteria as for  $\alpha^{dif}$  (c.f. Section III-B). In particular, we set  $R^{Rad} = 1$  mm. As we will discuss in Section VI, the mixing between different mid-angle signals introduced by the windowing may explain the robustness of the proposed method to wavefront aberrations.

In practice, we sample  $\mathbf{r}^0$  from a grid with  $N^x = 124$  and  $N^z = 149$  points along the  $x$  and  $z$  axes respectively. The grid spacing of  $\mathbf{r}^0$  is set to eight times the spacing of the beamforming grid. Moreover, we evaluate (6) and (9) at a series of  $N^{Rad} = 18$  Radon angles ranging from  $-17^\circ$  to  $17^\circ$  with a  $2^\circ$  spacing, identical to the dif-angle spacing. In addition, the displacement with respect to the center of the patch  $d$  ranges from  $-R^{Rad}$  to  $R^{Rad}$  with a discretization step equal to the grid spacing. Finally, the phase-shift maps  $\Delta\phi_{k \rightarrow k+1}$  associated with  $|\theta^{Tx}| = |\theta^{Rad} + \theta_{k/k+1}^{Rad}| > 25^\circ$  or  $|\theta^{Rx}| = |\theta^{Rad} - \theta_{k/k+1}^{dif}| > 25^\circ$  are discarded to match the Tx angle range used in the 11 PWs case.

As discussed in Section II-C, phase wrapping may occur in (9) when  $|\theta_{k+1}^{dif} - \theta_k^{dif}|$  is too large; thus, we must consider small dif-angle differences. Considering such a high number of phase-shift maps in our measurement model is however redundant and can lead to a slow and memory demanding inversion. To avoid this phenomenon, we accumulate phase shifts between  $N^{acc}$  successive dif-angle differences according to

$$\Delta\phi_{k \rightarrow (k+N^{acc})}(\theta^{Rad}, \mathbf{r}^0) = \sum_{n=1}^{N^{acc}} \Delta\phi_{(k+n-1) \rightarrow (k+n)}(\theta^{Rad}, \mathbf{r}^0), \quad (17)$$

which corresponds to the phase shift associated with a larger difference of dif angles. In our case, we accumulate  $N^{acc} = 4$  consecutive dif-angle pairs, leading to phase-shift maps associated with a dif-angle difference of  $8^\circ$ .

Ultimately, we recover a series of  $N^\phi = 56$  phase shift maps that are gathered into a vector  $\Delta\phi \in \mathbb{R}^{N^\phi N^x N^z}$ . The final phase shift maps are parametrized by a vector of mid angles  $\theta^{mid} \in \mathbb{R}^{N^\phi}$ , a vector of initial dif angles  $\theta^{dif,0} \in \mathbb{R}^{N^\phi}$ , and a vector of final dif angles  $\theta^{dif,1} \in \mathbb{R}^{N^\phi}$ . A summary of the phase shift estimation is depicted in Fig. 3, along with examples of phase shift maps generated by a  $1560 \text{ m s}^{-1}$  circular inclusion embedded within a  $1540 \text{ m s}^{-1}$  background.

#### D. Model Inversion

To reconstruct the spatial distribution of SoS, we rely on the procedure detailed in [19], [21]. In this section, we summarize its most important aspects.

First, we introduce the *slowness difference* map defined as

$$\Delta s(\mathbf{r}) = \frac{1}{c(\mathbf{r})} - \frac{1}{c_0}, \quad (18)$$

where  $c_0$  denotes the beamforming SoS. Second, we define a linear operator  $\mathcal{T}_\theta$  acting on the continuous slowness difference map

$$\mathcal{T}_\theta \{\Delta s\}(\mathbf{r}) = \int_{\zeta = \frac{-z}{\cos(\theta)}}^0 \Delta s(\mathbf{r} + \zeta \mathbf{u}_\theta) d\zeta. \quad (19)$$

The operator  $\mathcal{T}_\theta$  simulates the local aberration delay gained by a  $\theta$ -steered PW as it propagates through the medium. To define it, we use a straight-ray approximation. Importantly, this hypothesis implies that refraction effects are neglected. The operator  $\mathcal{T}_\theta$  allows us to define a second operator  $\mathcal{M}_n$  to estimate the phase-shift map defined in (12), namely

$$\mathcal{M}_n \{\Delta s\} = \frac{\omega_0}{\cos(\theta_n^{dif,1})} \left[ \mathcal{T}_{\theta_n^{mid} + \theta_n^{dif,1}} \{\Delta s\} + \mathcal{T}_{\theta_n^{mid} - \theta_n^{dif,1}} \{\Delta s\} \right] - \frac{\omega_0}{\cos(\theta_n^{dif,0})} \left[ \mathcal{T}_{\theta_n^{mid} + \theta_n^{dif,0}} \{\Delta s\} + \mathcal{T}_{\theta_n^{mid} - \theta_n^{dif,0}} \{\Delta s\} \right]. \quad (20)$$

In practice, we define multiple versions of  $\mathcal{M}_n$ , one per index  $n = 1, \dots, N^\phi$ . We discretize the resulting operators and gather them into a sparse matrix  $\mathbf{M} \in \mathbb{R}^{N^\phi N^x N^z \times N^x N^z \Delta s}$ , where  $N^x, \Delta s$ ,  $N^z, \Delta s$  denote the size of the discretization grid

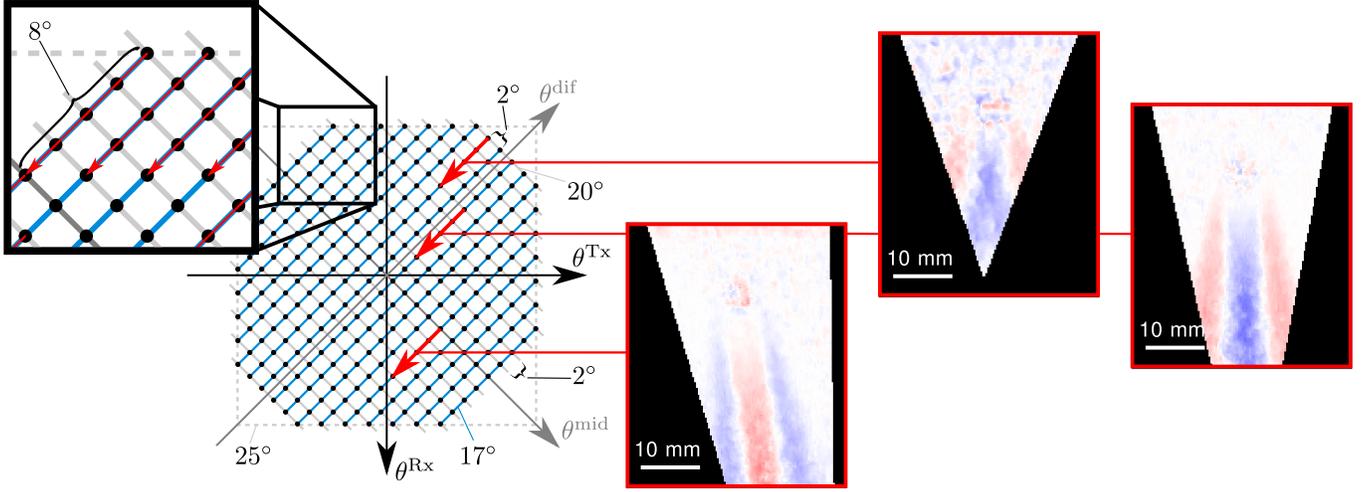


Fig. 3. Left: Configuration of the phase-shift estimation. Black dots represent signals  $y_k(\theta^{\text{Rad}}, \dots)$ . Grey lines indicate the dif angle  $\theta_k^{\text{dif}}$  used in constant-dif-angle beamforming. Blue lines indicate the values of  $\theta^{\text{Rad}}$  considered in the windowed Radon transform. Cross-correlation is performed along blue lines between adjacent signals with constant  $\theta^{\text{Rad}}$ . Red arrows represent the accumulated phase-shift maps from which the speed-of-sound map is estimated. Right: Three examples of phase-shift maps generated by a  $1560 \text{ m s}^{-1}$  inclusion embedded within a  $1540 \text{ m s}^{-1}$  background. Black areas show the masks used in the inverse problem.

of the slowness difference map. Importantly, the slowness-difference grid is not necessarily identical to the one used for the phase shift maps. In particular, we set a  $0.6 \text{ mm}$  slowness-difference grid spacing.

Due to the limited aperture of the transducer, some areas of the medium are not insonified by Tx or Rx PWs, leading to *shadows*. The phase shifts detected in these areas are therefore invalid, and we need to mask the shadows accordingly. To do so, we first define the mask for a single steered PW as

$$w_\theta^{\text{angle}}(\mathbf{r}) = \begin{cases} 1, & \text{if } x_1^s + x^{\text{mar}} < x - z \tan(\theta) < x_{N_s}^s - x^{\text{mar}} \\ 0, & \text{otherwise,} \end{cases} \quad (21)$$

where  $x^{\text{mar}}$  denotes a margin required to discard inaccurate phase-shift estimates appearing at the boundary of the PW, typically occurring due to edge waves. Then, we define a second series of masks

$$w_n = w_{\theta_{n^{\text{mid}} + \theta_n^{\text{dif},0}}^{\text{angle}}} w_{\theta_{n^{\text{mid}} - \theta_n^{\text{dif},0}}^{\text{angle}}} w_{\theta_{n^{\text{mid}} + \theta_n^{\text{dif},1}}^{\text{angle}}} w_{\theta_{n^{\text{mid}} - \theta_n^{\text{dif},1}}^{\text{angle}}}, \quad (22)$$

one per phase shift map index  $n = 1, \dots, N^\phi$ . Ultimately, all functions (22) are discretized according to the phase-shift grid and gathered into a vector  $\mathbf{w} \in \mathbb{R}^{N^\phi N^x N^z}$ . The black areas in the sample phase-shift maps of Fig. 3 illustrate masks  $w_n$ .

We estimate the discretized slowness-difference map  $\Delta \mathbf{s}$  with the following inverse problem

$$\min_{\Delta \mathbf{s}} \frac{1}{2} \|\mathbf{M} \Delta \mathbf{s} - \Delta \phi\|_{\mathbf{w}}^2 + \frac{\lambda^x}{2} \|\mathbf{D}^x \Delta \mathbf{s}\|^2 + \frac{\lambda^z}{2} \|\mathbf{D}^z \Delta \mathbf{s}\|^2. \quad (23)$$

Here,  $\mathbf{D}^x$  and  $\mathbf{D}^z$  denote the forward finite-difference derivative operators along  $x$  and  $z$  directions, respectively, whereas  $\lambda^x$  and  $\lambda^z$  encode the strength of the regularization along the two directions. We can find a closed-form solution to (23):

$$\Delta \mathbf{s} = [\mathbf{M}^T \text{Diag}(\mathbf{w}) \mathbf{M} + \lambda^x \mathbf{D}^{xT} \mathbf{D}^x + \lambda^z \mathbf{D}^{zT} \mathbf{D}^z]^{-1} \mathbf{M}^T \text{Diag}(\mathbf{w}) \Delta \phi. \quad (24)$$

TABLE I  
PARAMETERS OF THE PROPOSED METHOD

Transducer	ATL Philips L7-4
Center frequency $f_0$	4.8 MHz
Bandwidth	4 MHz - 7 MHz
Tx angles (115 PWs)	-28.5°, -28°, ..., 28.5°
Tx angles (11 PWs)	-25°, -20°, ..., 25°
Image grid spacing	$c_0/(8f_0)$
Angular width for Rx apodization	60°
Tukey cosine fraction for Rx apodization	0.125
Dif-angle aperture $\alpha^{\text{dif}}$ (16)	5°
Rx filtering angles $\theta^{\text{Rx}}$	-30°, -30° + 60°/580, ..., 30°
Beamforming dif angles $\theta^{\text{dif}}$ (15)	-20°, -18°, ..., 20°
Tukey cosine fraction for $w^{\text{mid}}$ (15)	0.25
Radon transform radius $R^{\text{Rad}}$ (6)	1 mm
Radon angles $\theta^{\text{mid}}$ (6)	-17°, -15°, ..., 17°
Phase-shift map grid spacing	$c_0/f_0$
SoS map grid spacing	0.6 mm
Regularization parameters $\lambda_x, \lambda_z$ (23)	$5 \cdot 10^5, 5 \cdot 10^5/40$
PW shadow margin $x^{\text{mar}}$ (21)	2 mm

Since the regularization parameters  $\lambda_x, \lambda_z$  and the mask  $\mathbf{w}$  are known a priori, the inverse matrix in (24) can be precomputed. Thus, the solution of (23) only requires two matrix multiplications, leading to fast inversion times.

Due to the difference in the number and nature of the phase-shift maps considered here, we cannot use the same regularization parameter than [19]. However, to provide a meaningful comparison, we fix the same ratio between  $\lambda_x$  and  $\lambda_z$  used in the reference method, namely 40. Moreover, we determine the regularization parameters such that the contrast and granularity of the SoS maps are as similar as possible to the ones achieved by the reference method. Overall, we set the regularization parameters to  $\lambda_x = 5 \cdot 10^5$  and  $\lambda_z = 5 \cdot 10^5/40$ , respectively. For clarity, Table I summarizes all the relevant parameters introduced in this section.

## IV. EXPERIMENTS

We assess the performance of the proposed method using three types of data: simulated data, data acquired on in-vitro phantoms, and in-vivo data. For the phantom and in-vivo data, we compare the proposed method with the reference method proposed in [19]. We focus on liver imaging since SoS has proven to be a relevant biomarker of fatty liver diseases thanks to its correlation with the liver fat fraction [9], [11], [22]. In this section, we detail the data acquisition, the metrics used to quantify robustness, and the implementation of the reference method. In particular, we consider the same phantom and in-vivo data used in [20].

### A. Simulated Data

First, we assess the accuracy of the proposed method with simulated data. We consider a geometric model of the ultrasound transducer and simulate, with k-Wave [36], the emission of the 115 PWs sequence and the reception of echoes by the transducer elements. We model the imaged medium to mimic a potential SoS distribution of the abdominal wall. An interweave of three layers with SoS  $1480\text{ m s}^{-1}$  and  $1580\text{ m s}^{-1}$  is placed on top of a  $1560\text{ m s}^{-1}$  layer. Moreover, we suppose uniform Gaussian density variations in the medium, such that a B-mode image of the medium figures a uniform speckle. To test the dependency of the proposed method to changes in speckle pattern, we repeat the simulation 10 times with different realizations of the density noise. We estimate the local SoS with the proposed method with two different beamforming SoS  $c_0$ ,  $1500\text{ m s}^{-1}$  and  $1540\text{ m s}^{-1}$ . We compare the results with the ground-truth SoS map. Moreover, we compare the phase-shift maps estimated by the proposed method to the result of the measurement model (20) applied to the ground truth SoS map. Results are eventually averaged over the 10 experiments.

### B. Phantom Design and In-Vivo Data

We also test the proposed method on three different gelatin phantoms—denoted as P1, P2, and P3—simulating the abdominal wall and liver. A detailed explanation regarding how the phantoms were built and how the quantitative SoS values were determined can be found in [20]. Three superficial layers mimicking respectively a subcutaneous fat layer, a triangular-shaped muscle, and a peritoneal fat layer were placed on top of a liver-mimicking layer with known SoS. The SoS in the liver-mimicking layer is different for each phantom,  $1525\text{ m s}^{-1}$  for P1,  $1555\text{ m s}^{-1}$  for P2, and  $1585\text{ m s}^{-1}$  for P3.

We also consider 11 in-vivo data sets acquired on the liver of a healthy volunteer from the same publication [20]. These data sets correspond to four different transducer positions—denoted by letters A to D—, with two to three position variations in each case—denoted by digits.

For both phantom and in-vivo data, we reconstruct SoS maps from the full (115 PWs) and reduced (11 PWs) data sets. Additionally, we take into account 4 different beamforming SoS  $c_0$ , namely  $1500\text{ m s}^{-1}$ ,  $1520\text{ m s}^{-1}$ ,  $1540\text{ m s}^{-1}$ , and  $1560\text{ m s}^{-1}$ .

### C. Robustness Metrics

We want to quantify the robustness of the proposed method and compare it to the one of the reference method. To do so, we test the variability of both methods to changes in the number of insonifications and beamforming SoS. Indeed, since the true SoS map does not depend on the number of insonifications, a SoS map reconstructed by a robust method with 11 PWs data is expected to be as close as possible to a map reconstructed with 115 PWs. Similarly, we expect a robust method to be agnostic to changes of beamforming SoS [19]. Therefore, we assess the variability to reduction of insonifications by the median absolute difference between the 115 PWs and 11 PWs SoS maps. Similarly, the variability to beamforming SoS is measured by the median of the pixel-wise standard deviation over the four different beamforming SoS tested. In both cases, we also compute the upper and lower quartiles to estimate the uncertainty of the estimators.

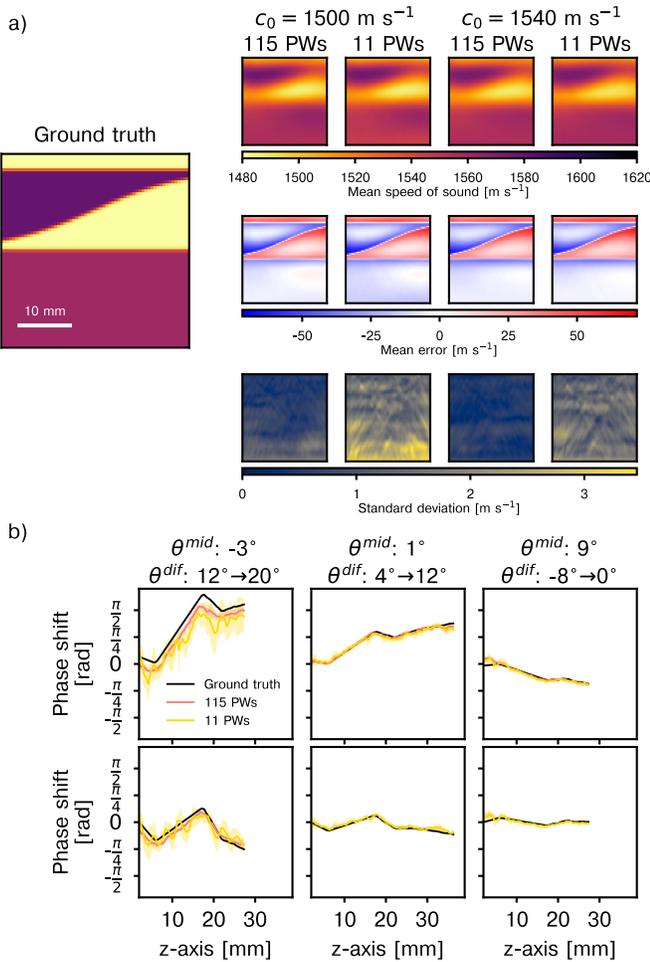
A potential target of the proposed method in a clinical context is to estimate the SoS inside the liver to quantify its fat fraction. With this purpose in mind, we also estimate the median SoS in the liver-mimicking layers and liver, for the phantom and in-vivo experiments respectively. The interquartile distance is computed as well to estimate the spread of the SoS distribution. Regarding the phantom data, we compare the values measured in the liver-mimicking layer with the ground-truth values to assess the biases of the proposed and reference methods. For in-vivo data, the true liver SoS is expected to be very similar for the different variations of a transducer position—A, B, C, and D. Similarity between the liver SoS estimates within a transducer position is therefore a sign of robustness. To quantify the similarity, we compute the standard deviation of the liver SoS within a given transducer position and average it over the four positions. Moreover, we expect the liver SoS to be mostly constant within the scale of an image. Therefore, we interpret a low interquartile distance of the liver SoS as a sign of robustness.

### D. Reference Method

The reference method was implemented as in [19], but parameters were adapted to best match the ones used in the proposed method. The tracking Tx and Rx angles range from  $-25^\circ$  to  $25^\circ$  degrees with a spacing of  $2.5^\circ$ , the Tx and Rx angular apertures are set to  $2.5^\circ$  and the angles used in the inverse problem after accumulation range from  $-25^\circ$  to  $25^\circ$  with a  $10^\circ$  step. In addition, the ratio between the regularization parameters  $\lambda_x/\lambda_z$  is kept equal to 40/1 but the magnitude of the regularization was adapted to the grid of the proposed method. For comparison with the proposed method with reduced angles, the same angle reduction was also used with the reference method.

## V. RESULTS

In this section, we present the results of the proposed method applied to simulated, phantom, and in-vivo data. In the phantom and in-vivo cases, we compare them with the results of the reference method.



**Fig. 4.** Simulation results. **a)** Left: ground truth speed-of-sound (SoS) map assumed in the simulation. Right, top row: Examples of SoS maps reconstructed from 115 and 11 PWs data, using  $1500 \text{ m s}^{-1}$  and  $1540 \text{ m s}^{-1}$  beamforming SoSs, averaged over 10 measurement realizations. Middle row: Mean error of the reconstructed SoS maps with respect to the ground-truth map. Bottom row: Standard deviation of the SoS maps over the 10 measurement realizations. **b)** Examples of phase shifts estimated by the proposed method at  $x = 0 \text{ mm}$  compared with the phase-shift measurement model applied to the ground truth SoS map. Three examples are presented using beamforming SoSs of  $1500 \text{ m s}^{-1}$  (top row) and  $1540 \text{ m s}^{-1}$  (bottom row). We depict the mean value over 10 measurement realizations and the standard deviation.

### A. Simulated Data

We display in **Fig. 4** examples of SoS maps and phase shifts at  $x = 0 \text{ mm}$  computed by the proposed method from simulated data. The mean values over the 10 realizations of the measurements are depicted, along with the standard deviations.

Regarding the mean SoS maps, we observe that their dependency to the number of PWs or to the beamforming SoS is limited. The SoS in the  $1580 \text{ m s}^{-1}$  and  $1480 \text{ m s}^{-1}$  layers are under- and overestimated, respectively. This mismatch can be attributed to the smoothing effect of the regularizer. The median SoS values in the bottom layer amount to  $1555.9 \text{ m s}^{-1}$  and  $1549.7 \text{ m s}^{-1}$  with a beamforming SoS of  $1500 \text{ m s}^{-1}$ —for 115 PWs and 11 PWS—, whereas they are  $1555.2 \text{ m s}^{-1}$  and  $1557.0 \text{ m s}^{-1}$  with a  $1540 \text{ m s}^{-1}$  beamforming SoS. We can see that the SoS biases remain limited—under  $5 \text{ m s}^{-1}$ —except with 11 PWs in the case where the SoS in the bottom layer is the furthest from the beamforming SoS.

**TABLE II**  
RELATIVE MEDIAN ABSOLUTE ERRORS OF THE PHASE SHIFTS  
ESTIMATED FROM SIMULATED DATA

	1500 <sup>a</sup>		1540 <sup>a</sup>	
	115 <sup>b</sup>	11 <sup>b</sup>	115 <sup>b</sup>	11 <sup>b</sup>
$\theta^{\text{mid}} : -3^\circ$	0.1900	0.2955	0.1720	0.2676
$\theta^{\text{dif}} : 12^\circ \rightarrow 20^\circ$	(0.0037)	(0.0087)	(0.0020)	(0.0050)
$\theta^{\text{mid}} : 1^\circ$	0.0733	0.1255	0.0784	0.1345
$\theta^{\text{dif}} : 4^\circ \rightarrow 12^\circ$	(0.0015)	(0.0030)	(0.0019)	(0.0038)
$\theta^{\text{mid}} : 9^\circ$	0.1660	0.3074	0.1500	0.2781
$\theta^{\text{dif}} : -8^\circ \rightarrow 0^\circ$	(0.0041)	(0.0105)	(0.0022)	(0.0082)
All	0.1617	0.3009	0.1594	0.2895
	(0.0008)	(0.0012)	(0.0009)	(0.0025)

<sup>a</sup>Beamforming SoS in  $\text{m s}^{-1}$     <sup>b</sup>Number of PWs

With respect to the standard deviation of the SoS estimation, we observe it is small in absolute terms. Nonetheless, a large number of insonifications reduces the dependency of the results to the measurement realization. In particular, the median standard deviation is improved by factors 2.6— with a beamforming SoS of  $1500 \text{ m s}^{-1}$ —and 2.9— with a beamforming SoS of  $1540 \text{ m s}^{-1}$ —when the number of PWs is increased from 11 PWs to 115 PWs.

We can notice in **Fig. 4b** that the examples of estimated and ground-truth phase-shifts coincide broadly. Nevertheless, a small bias is observed in the leftmost case. It is independent of the beamforming SoS but more pronounced with a reduced number of PWs. In general, the phase-shifts reconstructed with 11 PWs show a greater dependency to the signal realization, as with SoS maps.

To quantify the phase-shift errors, we provide in **Table II** the median absolute error of the estimated phase-shift maps, normalized by the median absolute values of the ground-truth phase-shift maps. We present the average over the 10 measurement realizations for the three examples depicted in **Fig. 4b** and the full set of  $N^\phi = 56$  phase-shift maps, with the standard deviation between parentheses. We observe that the errors nearly doubles when the number of PWs is reduced, and that the standard deviation increases. Moreover, the error is more important with large dif angles (first case) than with small ones (second case). The error also increases with higher mid angles (third case), even it is not clear from **Fig. 4b** due to the low magnitude of the phase shifts.

Overall, the magnitude of the phase-shift normalized error is limited, as well as its dependency to the measurement realization. Moreover, the SoS maps recovered by the proposed method are satisfactory in term of structure and amplitude, even if they are overly smoothed compared to the ground truth map due to the effect of the regularizer.

### B. Phantom Data

We display in **Fig. 5** the results of the proposed method applied to phantom data, along with the results of the reference method and B-mode images of the medium. We only depict the SoS maps reconstructed using  $1500 \text{ m s}^{-1}$  and  $1540 \text{ m s}^{-1}$  as beamforming SoS. We present the variability values in **Table III**, with the variability to changes of insonifications on

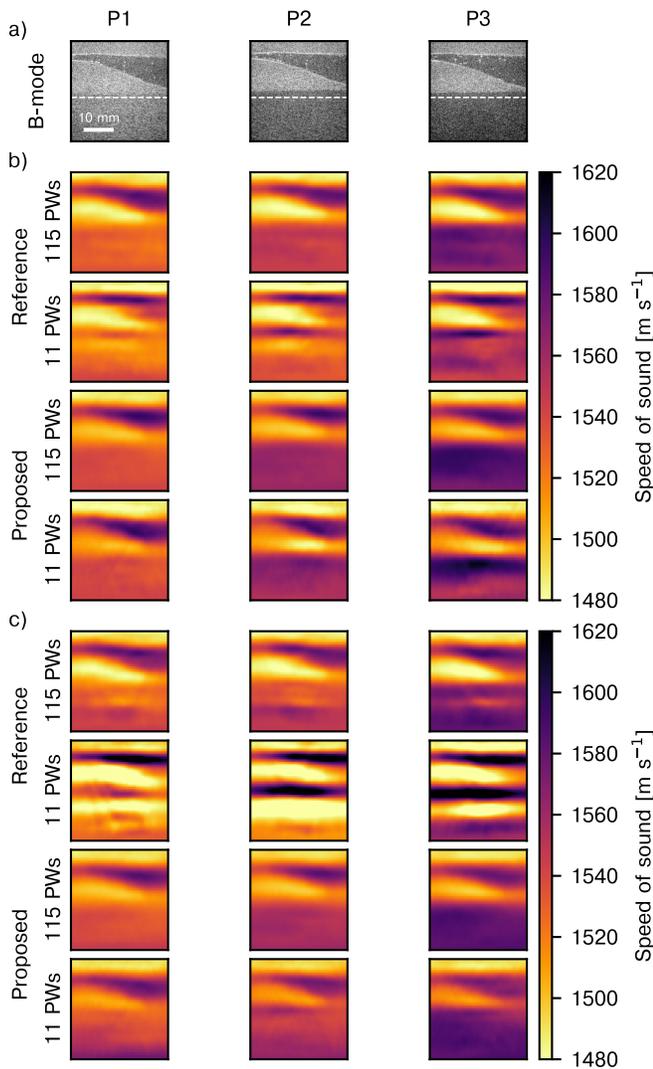


Fig. 5. Phantom results. a) B-mode images displayed with a 60 dB dynamic range, with the white dashed lines indicating the top of the liver-mimicking layers. b) Speed-of-sound (SoS) maps reconstructed using a  $1500 \text{ m s}^{-1}$  beamforming SoS, with the reference method by Stähli et al. [19] with 115 plane waves (PWs) (first row), the reference method with 11 PWs (second row), the proposed method with 115 PWs (third row), and the proposed method with 11 PWs (fourth row). c) SoS maps reconstructed using a  $1540 \text{ m s}^{-1}$  beamforming SoS.

the top—for the four beamforming SoSs—and the variability with respect to beamforming SoS on the bottom—for both the full and reduced data sets. The lower and upper quartiles are indicated in-between parenthesis, and we indicate by a bold number which of the proposed or reference method achieves the lowest variability in each case. In addition, we show the median SoS in the liver-mimicking layer in the left part of Table IV, with the interquartile distance in-between parenthesis. We indicate the top of the areas used to compute the median SoS by dashed lines in Fig. 5a.

Regarding first the results using 115 PWs, the SoS maps estimated by the two methods are qualitatively similar, as it can be seen in Fig. 5. The variability to changes of beamforming SoS—presented in the fifth row of Table III—is overall very low for both the proposed and the reference methods. Moreover, both variability values only differ by  $0.1 \text{ m s}^{-1}$ .

TABLE III  
VARIABILITY TO DATA REDUCTION (TOP ROWS) AND CHANGES OF BEAMFORMING SoS (BOTTOM ROWS) IN THE PHANTOMS

	P1	P2	P3	All	
$1500^a$	Ref.	8.0 (4.9-15.7)	13.3 (8.9-24.5)	17.9 (10.8-24.3)	13.2 (7.0-22.2)
	Pro.	<b>3.3</b> (1.5-11.0)	<b>7.4</b> (3.2-14.2)	<b>13.5</b> (6.4-21.4)	<b>7.7</b> (2.6-16.6)
$1520^a$	Ref.	16.0 (9.1-26.2)	20.3 (13.2-40.7)	19.8 (12.4-40.0)	18.0 (11.2-35.4)
	Pro.	<b>6.1</b> (2.9-11.8)	<b>3.0</b> (1.1-8.4)	<b>6.6</b> (2.8-10.4)	<b>5.2</b> (2.1-10.4)
$1540^a$	Ref.	24.3 (13.3-39.3)	37.9 (21.0-64.5)	27.6 (14.6-56.6)	30.5 (15.2-52.9)
	Pro.	<b>6.4</b> (2.5-11.4)	<b>5.5</b> (3.0-11.1)	<b>2.4</b> (1.1-7.5)	<b>4.6</b> (1.9-10.7)
$1560^a$	Ref.	39.2 (21.7-52.8)	58.1 (27.2-71.9)	49.3 (27.1-70.1)	46.6 (24.4-65.5)
	Pro.	<b>9.1</b> (4.1-17.6)	<b>6.6</b> (3.5-14.3)	<b>8.9</b> (3.8-13.6)	<b>7.8</b> (3.7-14.7)
$115^b$	Ref.	2.9 (1.4-7.7)	<b>2.5</b> (1.4-4.4)	<b>2.5</b> (1.5-5.7)	<b>2.6</b> (1.5-5.5)
	Pro.	<b>2.0</b> (1.4-2.7)	3.4 (1.7-4.6)	3.7 (2.2-5.7)	2.7 (1.6-4.5)
$11^b$	Ref.	10.2 (6.0-15.7)	16.5 (9.5-22.4)	11.8 (7.9-21.8)	12.6 (7.4-20.5)
	Pro.	<b>4.9</b> (2.5-8.4)	<b>6.0</b> (3.2-11.3)	<b>8.6</b> (5.0-13.7)	<b>6.4</b> (3.3-11.5)

<sup>a</sup>Beamforming SoS in  $\text{m s}^{-1}$  <sup>b</sup>Number of PWs

Regarding the SoS in the liver, we observe in Table IV a stronger bias in the liver-mimicking layer of P1 with the proposed method compared to the reference one. However, the biases in the liver-mimicking layers of P2 and P3 are more pronounced with the reference method.

When 11 PWs are used, Fig. 5 shows that the proposed method is more robust than the reference method against changes of beamforming SoS. For the latter, artifacts appear in the SoS map, and they are especially severe with a beamforming SoS of  $1540 \text{ m s}^{-1}$ . The improvement allowed by the proposed method is also confirmed by the quantitative measures shown in the last row of Table III. The overall variability nearly doubles with the reference method compared to the proposed one.

We observe in Table IV that a reduction of the number of PWs generally increases the bias and variance of the SoS inside the liver-mimicking layers. In the case of the proposed method, this bias is larger when the difference between the beamforming SoS and the ground-truth SoS in the liver-mimicking layer increases. However, the biases are in the large majority of cases more important with the reference method in the 11 PWs case, irrespectively of the beamforming SoS.

Finally, the overall variability to a reduction of the number of insonifications is substantially smaller with the proposed method compared to the reference method, as it can be seen in the first four rows of Table III. Regarding the proposed method, the variability to insonification reduction generally decreases the closer the beamforming SoS is to the ground-truth SoS of the liver-mimicking layer. This observation, along with the increase in bias highlighted above, show that additional care must be taken to determine the optimal

TABLE IV  
SPEED OF SOUND IN THE LIVER

		P1	P2	P3	A1	A2	A3	B1	B2	B3	C1	C2	D1	D2	D3	
Ground truth		1525	1555	1585	-	-	-	-	-	-	-	-	-	-	-	
1500 <sup>a</sup>	Ref.	115 PWs	1529.6 (6.2)	1546.5 (4.4)	1574.2 (9.1)	1535.9 (19.2)	1555.1 (15.9)	1588.3 (14.6)	1573.8 (13.1)	1558.4 (20.9)	1552.6 (19.9)	1564.1 (12.1)	1556.6 (12.2)	1572.6 (15.1)	1583.9 (7.7)	1580.0 (15.0)
		11 PWs	1522.0 (17.6)	1531.8 (16.0)	1552.9 (11.6)	1535.8 (18.0)	1532.6 (16.5)	1544.6 (14.6)	1570.3 (8.8)	1560.6 (18.0)	1546.3 (16.4)	1587.4 (23.6)	1569.6 (8.6)	1555.3 (12.9)	1566.2 (14.3)	1575.2 (14.3)
	Pro.	115 PWs	1538.2 (4.4)	1560.8 (3.1)	1585.1 (13.2)	1535.9 (32.9)	1559.8 (15.4)	1569.4 (9.7)	1577.7 (20.7)	1580.0 (22.9)	1577.0 (19.0)	1552.9 (20.1)	1559.5 (11.7)	1582.6 (17.1)	1601.1 (17.4)	1585.0 (10.5)
		11 PWs	1540.5 (5.7)	1561.6 (10.8)	1567.8 (34.0)	1544.0 (16.7)	1545.6 (11.2)	1555.4 (11.3)	1595.9 (27.1)	1588.8 (21.4)	1580.0 (14.1)	1575.5 (14.6)	1576.2 (8.4)	1588.2 (19.0)	1594.1 (18.9)	1585.3 (11.6)
1540 <sup>a</sup>	Ref.	115 PWs	1537.8 (18.4)	1547.9 (14.7)	1575.1 (15.5)	1558.3 (27.8)	1567.8 (19.7)	1596.6 (32.9)	1595.0 (21.3)	1576.0 (16.5)	1563.2 (13.3)	1598.5 (15.2)	1575.0 (11.8)	1575.0 (28.3)	1592.4 (28.0)	1596.7 (54.8)
		11 PWs	1513.2 (30.1)	1501.4 (43.8)	1557.2 (60.2)	1563.0 (37.0)	1575.9 (34.6)	1561.4 (22.5)	1591.0 (27.0)	1571.6 (18.0)	1566.5 (13.9)	1606.4 (20.8)	1588.6 (17.1)	1554.9 (32.8)	1577.6 (33.1)	1586.4 (47.0)
	Pro.	115 PWs	1537.8 (7.0)	1556.6 (5.0)	1583.3 (5.5)	1546.5 (18.2)	1562.4 (6.8)	1568.5 (8.8)	1571.1 (18.8)	1575.7 (23.9)	1567.1 (14.7)	1563.7 (16.8)	1551.8 (15.1)	1567.4 (7.3)	1595.6 (5.7)	1575.8 (10.3)
		11 PWs	1545.9 (23.0)	1553.5 (15.1)	1584.0 (6.2)	1556.8 (12.9)	1562.9 (7.6)	1565.2 (7.1)	1588.4 (14.6)	1578.4 (17.8)	1575.8 (18.8)	1577.0 (14.4)	1563.9 (12.6)	1577.7 (7.7)	1597.3 (9.0)	1586.1 (8.5)

<sup>a</sup>Beamforming SoS in m s<sup>-1</sup>

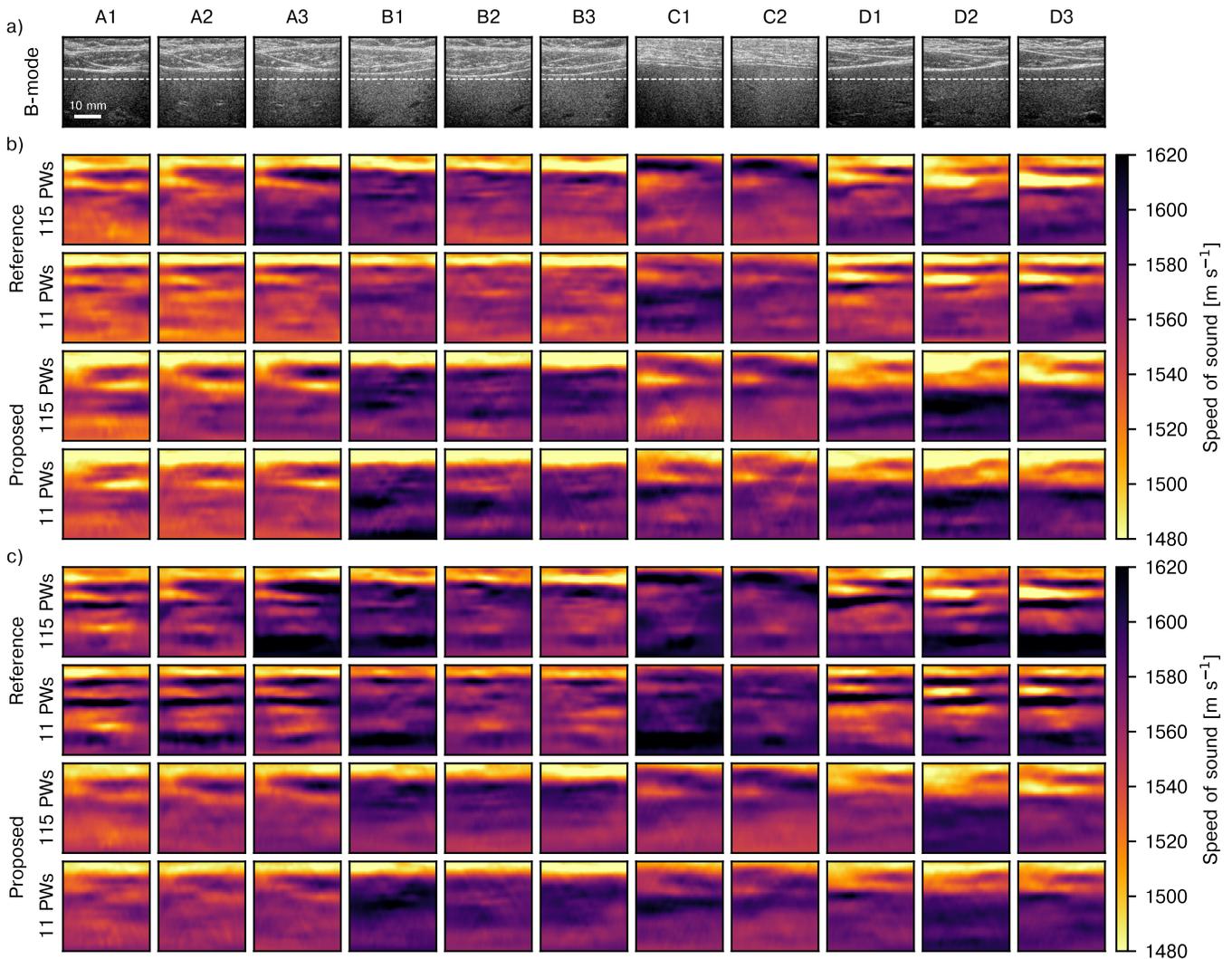


Fig. 6. In-vivo results. a) B-mode images displayed with a 60 dB dynamic range, with the white dashed line indicating the top boundary of zone used to compute liver speed of sound (SoS) b) SoS maps reconstructed using a 1500 m s<sup>-1</sup> beamforming SoS, with the reference method by Stähli et al. [19] with 115 plane waves (PWs) (first row), the reference method with 11 PWs (second row), the proposed method with 115 PWs (third row), and the proposed method with 11 PWs (fourth row). d) SoS maps reconstructed using a 1540 m s<sup>-1</sup> beamforming SoS.

TABLE V  
IN-VIVO VARIABILITY TO DATA REDUCTION (TOP ROWS) AND CHANGES OF BEAMFORMING SoS (BOTTOM ROWS)

	A1	A2	A3	B1	B2	B3	C1	C2	D1	D2	D3	All	
1500 <sup>a</sup>	Ref.	10.7 (6.1-18.4)	18.7 (10.5-26.8)	26.0 (12.6-43.7)	<b>8.8</b> (3.5-18.3)	<b>8.9</b> (3.8-18.5)	9.9 (4.4-17.9)	<b>19.6</b> (9.0-30.3)	<b>11.8</b> (7.1-17.7)	14.6 (6.8-24.2)	15.5 (8.2-24.4)	10.5 (5.5-17.9)	13.3 (6.1-23.5)
	Pro.	<b>8.7</b> (5.0-12.5)	<b>12.9</b> (6.4-20.6)	<b>11.1</b> (6.1-19.7)	9.3 (3.6-15.0)	15.2 (6.5-24.0)	<b>6.5</b> (2.9-11.5)	19.8 (11.8-29.7)	15.8 (8.2-25.0)	<b>5.8</b> (3.0-11.8)	<b>7.9</b> (4.0-12.0)	<b>6.4</b> (2.4-13.2)	<b>9.9</b> (4.6-17.9)
1520 <sup>a</sup>	Ref.	12.8 (6.8-20.9)	11.6 (5.1-22.2)	26.4 (13.4-41.1)	<b>8.1</b> (3.2-17.5)	<b>8.6</b> (4.6-15.8)	9.1 (4.6-18.8)	<b>16.8</b> (8.8-23.1)	<b>11.3</b> (6.6-16.7)	12.6 (5.7-22.4)	16.0 (10.7-24.9)	12.0 (6.6-18.8)	12.7 (6.0-22.2)
	Pro.	<b>7.3</b> (3.8-11.7)	<b>7.1</b> (3.4-13.0)	<b>7.4</b> (3.5-13.4)	10.0 (5.9-14.9)	12.5 (6.4-19.0)	<b>6.6</b> (2.9-11.5)	17.3 (8.8-24.2)	12.0 (6.5-17.6)	<b>6.0</b> (2.6-12.2)	<b>4.9</b> (2.4-9.9)	<b>9.1</b> (4.9-13.5)	<b>8.6</b> (4.0-14.7)
1540 <sup>a</sup>	Ref.	12.7 (6.0-22.1)	19.6 (9.5-28.4)	25.9 (15.0-43.8)	<b>9.6</b> (3.7-19.6)	10.8 (4.7-17.4)	9.7 (4.1-20.8)	<b>11.2</b> (6.4-18.6)	12.2 (6.6-19.8)	15.7 (7.5-27.0)	20.9 (11.4-29.3)	17.3 (10.2-23.0)	14.4 (6.8-24.6)
	Pro.	<b>9.5</b> (4.6-13.7)	<b>4.9</b> (2.4-8.8)	<b>5.8</b> (2.6-11.5)	13.0 (6.6-16.2)	<b>9.6</b> (4.8-14.7)	<b>8.3</b> (4.5-12.4)	14.7 (6.3-21.1)	<b>11.8</b> (8.0-15.6)	<b>8.5</b> (3.8-14.4)	<b>5.3</b> (2.0-12.7)	<b>11.5</b> (7.6-15.3)	<b>9.2</b> (4.2-14.7)
1560 <sup>a</sup>	Ref.	13.5 (5.7-25.4)	26.1 (11.9-42.0)	28.7 (15.5-47.6)	<b>12.7</b> (5.9-27.0)	15.1 (6.4-22.6)	11.4 (5.6-22.5)	<b>11.9</b> (5.5-18.5)	14.6 (7.5-23.5)	26.6 (12.8-39.6)	30.0 (17.1-41.1)	24.9 (13.6-39.4)	18.1 (8.3-32.7)
	Pro.	13.5 (7.6-19.5)	<b>8.7</b> (4.1-14.3)	<b>6.5</b> (3.3-10.2)	15.6 (6.9-18.5)	<b>6.8</b> (2.5-13.5)	11.4 (6.1-14.8)	13.1 (5.5-17.3)	<b>11.5</b> (8.2-15.7)	<b>10.6</b> (4.8-15.5)	<b>7.0</b> (3.2-17.8)	<b>14.0</b> (9.3-18.4)	<b>10.7</b> (4.9-16.4)
115 <sup>b</sup>	Ref.	10.8 (5.8-23.1)	9.4 (6.2-14.3)	8.9 (4.6-15.2)	8.3 (4.6-15.8)	9.1 (4.4-17.5)	8.8 (4.5-13.7)	13.6 (8.9-24.0)	7.3 (4.0-15.9)	7.4 (4.4-11.7)	9.6 (5.7-15.7)	12.8 (6.5-19.1)	9.4 (5.2-16.6)
	Pro.	<b>6.5</b> (3.5-11.4)	<b>5.9</b> (3.7-7.9)	<b>4.1</b> (2.8-6.4)	<b>4.3</b> (2.6-5.9)	<b>4.5</b> (2.9-6.4)	<b>4.5</b> (3.4-6.7)	<b>8.0</b> (4.8-10.3)	<b>3.4</b> (2.3-4.1)	<b>6.8</b> (4.4-10.1)	<b>5.6</b> (3.6-8.6)	<b>5.2</b> (3.8-8.9)	<b>4.9</b> (3.3-8.0)
11 <sup>b</sup>	Ref.	15.7 (7.8-26.3)	19.3 (13.5-34.3)	11.3 (6.7-16.9)	11.4 (5.9-21.0)	13.1 (6.7-18.6)	11.8 (6.2-19.0)	13.6 (7.7-22.5)	12.2 (5.8-19.7)	8.6 (5.2-15.4)	13.2 (9.3-19.0)	12.9 (8.4-18.0)	13.0 (7.3-20.1)
	Pro.	<b>9.8</b> (5.1-16.0)	<b>11.5</b> (7.1-16.2)	<b>7.9</b> (5.5-12.1)	<b>6.9</b> (4.5-9.2)	<b>7.7</b> (5.2-10.6)	<b>6.3</b> (4.3-9.0)	<b>8.3</b> (6.4-11.5)	<b>6.9</b> (5.0-9.4)	<b>7.8</b> (5.3-11.1)	<b>7.4</b> (4.4-12.1)	<b>7.9</b> (5.6-11.1)	<b>7.7</b> (5.3-11.4)

<sup>a</sup>Beamforming SoS in  $m s^{-1}$     <sup>b</sup>Number of PWs

beamforming SoS when a 11 PWs sequence is used by the proposed method. In contrast, we can notice in Table III that the variability of the reference method to data reduction generally increases with larger beamforming SoSs.

In conclusion, both methods show similar performance when 115 PWs are used. However, the proposed method achieves significantly better results with 11 PWs, in term of bias, variability to SoS beamforming, and variability to insonification reduction.

### C. In-Vivo Data

Fig. 6b and Fig. 6c depict the SoS maps reconstructed from the in-vivo data with  $1500 m s^{-1}$  and  $1540 m s^{-1}$  as beamforming SoSs  $c_0$ , respectively. We display in Fig. 6a the B-mode images of the medium for each data set, where a dashed line indicates the top boundary of the area used to compute the liver SoS. Table V details the quantitative variability values, in the same format as in Table III. The right part of Table IV presents the medians and interquartile distances of the SoS in the liver, for each data set. In addition, the first row of Table VI (\*) displays the standard deviation of the liver SoS over a single transducer position—i.e. A, B, C, or D—, averaged over the four positions. Finally, the second row of Table VI (\*\*) presents the median SoS interquartile distance in the liver over the 11 data sets.

Addressing first the variability to insonification reduction, we notice in the first four rows of Table V that the overall relative improvement allowed by the proposed method varies from 25% ( $c_0 = 1500 m s^{-1}$ ) to 40 % ( $c_0 = 1560 m s^{-1}$ ). Moreover, the cases where the reference method dominates the proposed one are all occurring when we image media

with simpler geometries, namely with transducer positions B and C. The SoS maps in cases A and D show stronger lateral and axial variations than cases B and C. There, the proposed method often largely dominates the reference method regarding variability to insonification reduction. The variability to beamforming SoS—presented in the last two rows of Table V—is also consistently improved by the proposed method, achieving overall relative improvements of 48% and 41% for the 115 PWs and 11 PWs cases, respectively. This improvement can be observed visually as well in Fig. 6. Artifacts similar to the ones observed in the phantom appear in the SoS maps reconstructed with the reference method using  $c_0 = 1540 m s^{-1}$ .

Regarding the SoS in the liver, the standard deviation within a transducer position is reduced with respect to the proposed method. This phenomenon occurs irrespectively of the beamforming SoS or number of PWs as it can be seen in the first row of Table VI. In contrast, the interquartile distance achieved by the proposed method—second row of Table VI—is worse than the reference method with  $c_0 = 1500 m s^{-1}$ . Nonetheless, the difference between the two methods is small and the opposite observation can be made with  $c_0 = 1540 m s^{-1}$ , with a larger difference.

Overall, all the in-vivo metrics concur with a general robustness improvement allowed by the proposed method with respect to the reference method. Interestingly, the gain in stability to insonification reduction is larger with the phantom experiments compared to the in-vivo experiments. In-vivo data are expected to suffer from more severe levels of phase noise compared to data acquired on a phantom. We can hypothesise that this phase noise has two effects. First, the larger amount of

TABLE VI

AVERAGE VARIABILITY OF THE IN-VIVO LIVER SoS FOR A TRANSDUCER POSITION (\*) AND AVERAGE INTERQUARTILE DISTANCE OF THE IN-VIVO LIVER SoS ESTIMATES (\*\*)

	1500 <sup>a</sup>				1540 <sup>a</sup>			
	Ref.		Pro.		Ref.		Pro.	
	115 <sup>b</sup>	11 <sup>b</sup>	115 <sup>b</sup>	11 <sup>b</sup>	115 <sup>b</sup>	11 <sup>b</sup>	115 <sup>b</sup>	11 <sup>b</sup>
*	9.7	8.0	6.7	3.9	12.6	9.8	7.6	5.9
**	15.1	15.1	17.9	15.8	24.5	27.6	13.3	11.9

<sup>a</sup>Beamforming SoS in m s<sup>-1</sup>    <sup>b</sup>Number of PWs

phase noise generally affect the performances of both methods. The relative improvement allowed by the proposed method with respect to the reference one is thus expected to be reduced. Second, the phase noise probably have a regularizing effect on the phase-shift estimation of the reference method, alleviating errors induced by wavefront distortion and reducing major artifacts.

## VI. DISCUSSION AND CONCLUSION

In this article, we propose a new tomographic method to reconstruct the local SoS of soft tissues from pulse-echo ultrasound measurements. We confirm the accuracy of the proposed method with simulated and in-vitro data. We also establish with in-vitro and in-vivo experiments that the proposed method is more stable to changes of beamforming SoS and to a reduction of the number of insonifications in comparison with a state-of-the-art method. Moreover, we show that the proposed method leads to more robust estimates of SoS in the liver, thus demonstrating a potential use of the method to diagnose fatty liver diseases.

A major difficulty occurring when we image a medium with spatial and lateral SoS variations—in opposition to a medium with constant or constant-by-layers SoS—stems from wavefront distortions. Such distortions mitigate the straight-ray hypothesis and introduce amplitude variations along the wavefront. Our experiments tend to characterize the proposed method as more resilient to wavefront distortions than the reference method. Indeed, the stability gap between the proposed and reference methods is more important with complex media—cases P, A, and D—than with media closer to a constant-by-layer structure—cases B and C. We believe that the windowed Radon transform explains this improvement. The reference method reduces the Tx-Rx aperture of the images, thus limiting the amount of input signals taken into account to estimate the phase-shift maps. According to the hypothesis introduced in Section II-C, the result of the windowed Radon transform should be equivalent to signals obtained by limiting the aperture of the images. However, in practice the windowing of the images prior to the Radon transform introduces a mixing of the signals associated with different angles  $\theta^{\text{Rad}}$ . We believe that it introduces a smoothing effect which leads to a more robust phase-shift estimation and mitigates the negative effects of wavefront distortions and phase noise.

However, strong outliers can still appear in the phase-shift maps. For example, phase wrapping may occur, especially if the signals in (9) are only weakly correlated due to noise or

artifacts. An  $\ell_1$  norm could therefore improve the robustness of the model inversion to outliers, as proposed in [18] with block-matching estimation of aberration delays. More generally, we believe that the most significant improvement to our method can be provided by an adaptation of the inverse problem. For instance, a total-variation (TV)-norm-based regularizer [18] could enforce piece-wise constant SoS maps. Thus, it could provide a prior that is more representative of the real SoS distribution in human soft-tissues than the proposed regularizer. In particular, a change of regularizer could alleviate the loss of SoS constraint observed in the experiments with simulated data (c.f. Section V-A). Furthermore, more elaborate regularization schemes such as a Bayesian soft-prior [20] have been proposed to factor in information from B-mode image. A neural network based inversion similar to [37] could also enforce a learned prior on the reconstructed SoS maps and improve the accuracy of reconstruction. Overall, we believe that we can close the remaining performance gap between the 115 PWs and 11 PWs cases by updating the inverse problem.

Unfortunately, our non-optimized prototype implementation of the proposed method requires 1.5 to 4 hour to run, depending on the amount of PWs considered. However, we must highlight that the constant-dif-angle beamforming and the windowed Radon transform—the two most computationally demanding steps of the proposed method—can be easily parallelized. Currently, we can neither confirm nor deny that the computation time can be brought down in the future to allow real-time imaging. Nonetheless, we believe that an optimized GPU implementation of the proposed method can provide a time-to-solution suitable for clinical practice.

The results presented in this article target liver SoS determination. Curved arrays are generally more adapted to this imaging configuration compared to linear arrays. Thus, it would be meaningful to adapt the proposed method to this type of array [21]. In principle, the application domain of the proposed method can be extended to other organs where SoS has proven to be relevant. This may include, for instance, breast imaging [38] as already hinted in our previous work showing phantom-based preliminary results [23].

However, we believe that the main practical interest of the proposed method is its applicability to portable ultrasound devices. Indeed, they are often limited in terms of power consumption, data transfer bandwidth, memory capability, and propensity to overheating. There are therefore strong incentives to reduce the number of insonifications as much as possible. Finally, our framework could, in principle, be extended to 3D to reconstruct volumetric images of tissue SoS since the large memory requirements would also benefit from the reduced number of insonifications that the proposed method may permit.

## APPENDIX A WINDOWED RADON TRANSFORM— DETAILED COMPUTATION

Taking into account (2), (5) and the two assumptions introduced Section II-C, the Fourier transform of (6) with

respect to  $d$  can be written as

$$\begin{aligned} \hat{y}_k^{\text{Rad}}(\theta^{\text{Rad}}, \xi, \mathbf{r}^0) \\ \approx \int_{\theta^{\text{mid}}} g_k(\theta^{\text{mid}}, \mathbf{r}^0) \\ \int_{\mathbf{r}} \gamma^w(\mathbf{r}, \mathbf{r}^0) \hat{f}_k(\theta^{\text{mid}}, \theta^{\text{Rad}}, \xi, \mathbf{r}, \mathbf{r}^0) d\mathbf{r} d\theta^{\text{mid}}, \end{aligned} \quad (25)$$

where  $\xi$  is the frequency associated with  $d$ , with

$$\begin{aligned} \gamma^w(\mathbf{r}, \mathbf{r}^0) &= w(\mathbf{r} - \mathbf{r}^0) \gamma(\mathbf{r}), \\ g_k(\theta^{\text{mid}}, \mathbf{r}^0) &= a(\theta_k^{\text{dif}}, \theta^{\text{mid}}, \mathbf{r}^0) \\ &h^{\text{Tx}}(\theta^{\text{mid}} + \theta_k^{\text{dif}}, \mathbf{r}^0) h^{\text{Rx}}(\theta^{\text{mid}} - \theta_k^{\text{dif}}, \mathbf{r}^0), \end{aligned} \quad (26)$$

$$(27)$$

a term encompassing magnitude effects, and

$$\begin{aligned} \hat{f}_k(\theta^{\text{mid}}, \theta^{\text{Rad}}, \xi, \mathbf{r}, \mathbf{r}^0) &= \int_{\mathbf{r}'} e^{-j\xi \langle \mathbf{u}_{\theta^{\text{Rad}}}, \mathbf{r}' - \mathbf{r}^0 \rangle} \\ v_{\text{pe}} \left( \frac{2}{c_0} \cos(\theta_k^{\text{dif}}) \langle \mathbf{u}_{\theta^{\text{mid}}}, \mathbf{r}' - \mathbf{r} \rangle - \Delta\tau(\theta_k^{\text{dif}}, \theta^{\text{mid}}, \mathbf{r}^0) \right) d\mathbf{r}'. \end{aligned} \quad (28)$$

If we express  $v_{\text{pe}}$  in function of its Fourier transform, we can rewrite (28) as

$$\begin{aligned} \hat{f}_k(\theta^{\text{mid}}, \theta^{\text{Rad}}, \xi, \mathbf{r}, \mathbf{r}^0) &= \int_{\omega} \hat{v}_{\text{pe}}(\omega) \\ &e^{-j\left[\omega \frac{2}{c_0} \cos(\theta_k^{\text{dif}}) \langle \mathbf{u}_{\theta^{\text{mid}}}, \mathbf{r} \rangle\right]} e^{j\xi \langle \mathbf{u}_{\theta^{\text{Rad}}}, \mathbf{r}^0 \rangle} e^{-j\omega \Delta\tau(\theta_k^{\text{dif}}, \theta^{\text{mid}}, \mathbf{r}^0)} \\ &\frac{1}{2\pi} \int_{\mathbf{r}'} e^{-j\left[\xi \langle \mathbf{u}_{\theta^{\text{Rad}}}, \mathbf{r} - \mathbf{r}' \rangle - \omega \frac{2}{c_0} \cos(\theta_k^{\text{dif}}) \langle \mathbf{u}_{\theta^{\text{mid}}}, \mathbf{r}' \rangle\right]} d\mathbf{r}' d\omega, \end{aligned} \quad (29)$$

with  $\omega$  the frequency associated with the time coordinate  $t$ . We can recognize in the last line of (29) the Fourier transform of a two-dimensional Dirac delta distribution. If we express the Dirac delta distribution in polar coordinates and evaluate the outer integral in (29), we obtain

$$\begin{aligned} \hat{f}_k(\theta^{\text{mid}}, \theta^{\text{Rad}}, \xi, \mathbf{r}, \mathbf{r}^0) &= \hat{v}_{\text{pe}} \left( \frac{c_0}{2 \cos(\theta_k^{\text{dif}})} \xi \right) \delta(\theta^{\text{mid}} - \theta^{\text{Rad}}) \\ &\frac{2\pi c_0}{2 \cos(\theta_k^{\text{dif}}) \xi} e^{-j\left[\xi \langle \mathbf{u}_{\theta^{\text{Rad}}}, \mathbf{r} - \mathbf{r}^0 \rangle\right]} e^{-j \frac{c_0}{2 \cos(\theta_k^{\text{dif}})} \xi \Delta\tau(\theta_k^{\text{dif}}, \theta^{\text{mid}}, \mathbf{r}^0)}. \end{aligned} \quad (30)$$

Introducing (30) back into (25) and evaluating the outer integral leads to

$$\begin{aligned} \hat{y}_k^{\text{Rad}}(\theta^{\text{Rad}}, \xi, \mathbf{r}^0) &= \frac{2\pi c_0}{2 \cos(\theta_k^{\text{dif}}) \xi} g_k(\theta^{\text{Rad}}, \mathbf{r}^0) \\ &\hat{v}_{\text{pe}} \left( \frac{c_0}{2 \cos(\theta_k^{\text{dif}})} \xi \right) e^{-j \frac{c_0}{2 \cos(\theta_k^{\text{dif}})} \xi \Delta\tau(\theta_k^{\text{dif}}, \theta^{\text{Rad}}, \mathbf{r}^0)} \hat{\gamma}^w(\xi \mathbf{u}_{\theta^{\text{Rad}}}), \end{aligned} \quad (31)$$

with

$$\hat{\gamma}^w(\mathbf{v}, \mathbf{r}^0) = \int_{\mathbf{r}} e^{-j\langle \mathbf{v}, \mathbf{r} - \mathbf{r}^0 \rangle} \gamma^w(\mathbf{r}, \mathbf{r}^0) d\mathbf{r}, \quad (32)$$

the two-dimensional Fourier transform of  $\gamma^w$ , shifted by  $\mathbf{r}^0$ .

## REFERENCES

- [1] J.-L. Gennisson, T. Defieux, M. Fink, and M. Tanter, "Ultrasound elastography: Principles and techniques," *Diagnostic Interventional Imag.*, vol. 94, no. 5, pp. 487–495, May 2013.
- [2] H. Gemmeke and N. V. Ruiters, "3D ultrasound computer tomography for medical imaging," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 580, no. 2, pp. 1057–1065, 2007.
- [3] N. Duric et al., "Breast imaging with SoftVue: Initial clinical evaluation," *Proc. SPIE*, vol. 9040, pp. 208–215, Mar. 2014.
- [4] J. Wiskin, D. T. Borup, S. A. Johnson, and M. Berggren, "Non-linear inverse scattering: High resolution quantitative breast tissue tomography," *J. Acoust. Soc. Amer.*, vol. 131, no. 5, pp. 3802–3813, May 2012.
- [5] J. Wiskin, B. Malik, D. Borup, N. Pirshafiey, and J. Klock, "Full wave 3D inverse scattering transmission ultrasound tomography in the presence of high contrast," *Sci. Rep.*, vol. 10, no. 1, p. 20166, Nov. 2020.
- [6] N. K. Martiartu, C. Boehm, and A. Fichtner, "3-D wave-equation-based finite-frequency tomography for ultrasound computed tomography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 7, pp. 1332–1343, Jul. 2020.
- [7] N. Duric et al., "Using whole breast ultrasound tomography to improve breast cancer risk assessment: A novel risk factor based on the quantitative tissue property of sound speed," *J. Clin. Med.*, vol. 9, no. 2, p. 367, Jan. 2020.
- [8] R. Natesan, J. Wiskin, S. Lee, and B. H. Malik, "Quantitative assessment of breast density: Transmission ultrasound is comparable to mammography with tomosynthesis," *Cancer Prevention Res.*, vol. 12, no. 12, pp. 871–876, Dec. 2019.
- [9] M. Imbault et al., "Robust sound speed estimation for ultrasound-based hepatic steatosis assessment," *Phys. Med. Biol.*, vol. 62, no. 9, pp. 3582–3598, May 2017.
- [10] M. Imbault et al., "Ultrasonic fat fraction quantification using in vivo adaptive sound speed estimation," *Phys. Med. Biol.*, vol. 63, no. 21, Oct. 2018, Art. no. 215013.
- [11] M. D. Burgio et al., "Ultrasonic adaptive sound speed estimation for the diagnosis and quantification of hepatic steatosis: A pilot study," *Ultraschall der Medizin-Eur. J. Ultrasound*, vol. 40, no. 6, pp. 722–733, 2019.
- [12] R. Ali, D. Hyun, and J. J. Dahl, "Application of common midpoint gathers to medical pulse-echo ultrasound for optimal coherence and improved sound speed estimation in layered media," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2020, pp. 1–4.
- [13] R. Ali and J. J. Dahl, "Distributed phase aberration correction techniques based on local sound speed estimates," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–4.
- [14] R. Ali et al., "Local sound speed estimation for pulse-echo ultrasound in layered media," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 2, pp. 500–511, Feb. 2022.
- [15] B. Hériard-Dubreuil, A. Besson, F. Wintzenrieth, C. Cohen-Bacrie, and J.-P. Thiran, "Refraction-based speed of sound estimation in layered media: An angular approach," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 70, no. 6, pp. 486–497, Jun. 2023.
- [16] M. Jaeger, G. Held, S. Peeters, S. Preisser, M. Grünig, and M. Frenz, "Computed ultrasound tomography in echo mode for imaging speed of sound using pulse-echo sonography: Proof of principle," *Ultrasound Med. Biol.*, vol. 41, no. 1, pp. 235–250, Jan. 2015.
- [17] M. Jaeger and M. Frenz, "Towards clinical computed ultrasound tomography in echo-mode: Dynamic range artefact reduction," *Ultrasonics*, vol. 62, pp. 299–304, Sep. 2015.
- [18] S. J. Sanabria, E. Ozkan, M. Rominger, and O. Goksel, "Spatial domain reconstruction for imaging speed-of-sound with pulse-echo ultrasound: Simulation and in vivo study," *Phys. Med. Biol.*, vol. 63, no. 21, Oct. 2018, Art. no. 215015.
- [19] P. Stähli, M. Kuriakose, M. Frenz, and M. Jaeger, "Improved forward model for quantitative pulse-echo speed-of-sound imaging," *Ultrasonics*, vol. 108, Dec. 2020, Art. no. 106168.
- [20] P. Stähli, M. Frenz, and M. Jaeger, "Bayesian approach for a robust speed-of-sound reconstruction using pulse-echo ultrasound," *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 457–467, Feb. 2021.
- [21] M. Jaeger et al., "Pulse-echo speed-of-sound imaging using convex probes," *Phys. Med. Biol.*, vol. 67, no. 21, Nov. 2022, Art. no. 215016.

- [22] P. Stähli, C. Becchetti, N. K. Martiartu, A. Berzigotti, M. Frenz, and M. Jaeger, "First-in-human diagnostic study of hepatic steatosis with computed ultrasound tomography in echo mode," *Commun. Med.*, vol. 3, no. 1, p. 176, 2023.
- [23] S. Beuret, B. Hériard-Dubreuil, S. Canales, and J.-P. Thiran, "Directional cross-correlation for improved aberration phase estimation in pulse-echo speed-of-sound imaging," in *Proc. IEEE Int. Ultrason. Symp.*, Sep. 2021, pp. 1–4.
- [24] G. Jansen, N. Awasthi, H.-M. Schwab, and R. Lopata, "Enhanced radon domain beamforming using deep-learning-based plane wave compounding," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2021, pp. 1–4.
- [25] H.-M. Schwab, F. van de Vosse, and R. Lopata, "An R-space theorem for plane wave ultrasound reconstruction," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2020, pp. 1–4.
- [26] H.-M. Schwab and R. Lopata, "A radon diffraction theorem for plane wave ultrasound imaging," *J. Acoust. Soc. Amer.*, vol. 153, no. 2, pp. 1015–1026, Feb. 2023.
- [27] S. Beuret and J.-P. Thiran, "Windowed radon transform and tensor rank-1 decomposition for adaptive beamforming in ultrafast ultrasound," *IEEE Trans. Med. Imag.*, Jul. 2023.
- [28] J. A. Jensen, "A model for the propagation and scattering of ultrasound in tissue," *J. Acoust. Soc. Amer.*, vol. 89, no. 1, pp. 182–190, Jan. 1991.
- [29] A. Besson et al., "Ultrafast ultrasound imaging as an inverse problem: Matrix-free sparse image reconstruction," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 3, pp. 339–355, Mar. 2018.
- [30] A. Besson et al., "A physical model of nonstationary blur in ultrasound imaging," *IEEE Trans. Comput. Imag.*, vol. 5, no. 3, pp. 381–394, Sep. 2019.
- [31] W. Lambert, J. Robin, L. A. Cobus, M. Fink, and A. Aubry, "Ultrasound matrix imaging—Part I: The focused reflection matrix, the F-factor and the role of multiple scattering," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3907–3920, Dec. 2022.
- [32] W. Lambert, L. A. Cobus, J. Robin, M. Fink, and A. Aubry, "Ultrasound matrix imaging—Part II: The distortion matrix for aberration correction over multiple isoplanatic patches," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3921–3938, Dec. 2022.
- [33] H. Bendjador, T. Deffieux, and M. Tanter, "The SVD beamformer: Physical principles and application to ultrafast adaptive ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3100–3112, Oct. 2020.
- [34] B. Denarie et al., "Coherent plane wave compounding for very high frame rate ultrasonography of rapidly moving targets," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1265–1276, Jul. 2013.
- [35] R. Rau, D. Schweizer, V. Vishnevskiy, and O. Goksel, "Speed-of-sound imaging using diverging waves," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 16, no. 7, pp. 1201–1211, Jul. 2021.
- [36] B. E. Treeby and B. T. Cox, "K-wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.*, vol. 15, no. 2, 2010, Art. no. 021314.
- [37] V. Vishnevskiy, S. J. Sanabria, and O. Goksel, "Image reconstruction via variational network for real-time hand-held sound-speed imaging," in *Machine Learning for Medical Image Reconstruction. MLMIR (Lecture Notes in Computer Science)*, vol. 11074. Berlin, Germany: Springer, 2018, pp. 120–128.
- [38] L. Ruby et al., "Breast cancer assessment with pulse-echo speed of sound ultrasound from intrinsic tissue reflections: Proof-of-concept," *Investigative Radiol.*, vol. 54, no. 7, pp. 419–427, Jul. 2019.