Refraction-Based Speed of Sound Estimation in Layered Media: an Angular Approach

Baptiste Hériard-Dubreuil, *Student Member, IEEE*, Adrien Besson *Member, IEEE*, Frédéric Wintzenrieth *Member, IEEE*, Claude Cohen-Bacrie and Jean-Philippe Thiran, *Senior Member, IEEE*

Abstract—**Speed of sound estimation in ultrasound imaging is a growing modality with several clinical applications such as hepatic steatosis stages quantification. A key challenge for clinically-relevant speed of sound estimation is to obtain repeatable values independent from superficial tissues and available in real-time. Recent works have demonstrated the feasibility to achieve quantitative estimations of the local speed of sound in layered media. However, such techniques require a high computational power and exhibit instabilities. We present a novel speed of sound estimation technique based on an angular approach of ultrasound imaging in which plane-waves are considered in transmit and in receive. This change of paradigm allows us to rely on refraction properties of plane-waves to infer local speed of sound values directly from the angular raw-data. The proposed method robustly estimates the local speed of sound with only few ultrasound emissions and with a low computational complexity which makes it compatible with real-time imaging. Simulations and *in vitro* experimental results show that the proposed method outperforms state-of-the-art approaches with biases and standard deviations lower than $10 \, \mathrm{m \, s^{-1}}$, 8 times less emissions and 1000 times lower computational time. Further *in vivo* experiments validate its performance for liver imaging.**

Index Terms— **Angular Framework, Plane-Waves, Speed of Sound Estimation, Ultrasound Imaging.**

I. INTRODUCTION

IN ultrasound imaging, speed of sound estimation techniques aim at measuring the speed at which an emitted ultrasonic wave propagates in a medium. In case of a biological medium, speed of sound relates to tissue characteristics and its estimation yields several clinical applications, from hepatic steatosis staging [1] to breast cancer detection [2]–[4]. In the case of liver steatosis, several studies have demonstrated that speed of sound values measured in the liver correlate well with histological fat score [5], and more recently to fat fraction and steatosis grade [6]–[9]. Nevertheless, it is not broadly adopted in clinical practice due to the lack of a technique that is both easy to use and reliable.

The first speed of sound estimation techniques in pulse-echo ultrasound imaging appeared forty years ago, making use of several transducer arrays [10] or probe displacements [11]–[13]. They tracked time of flight of ultrasonic beams, refraction or deformations, and yielded a good correlation with histological grades [5] but were difficult to apply clinically.

More recently, several methods involving a single ultrasonic probe have been proposed. They rely on the hypothesis of an homogeneous speed of sound in the medium. Several are based on the shape of the measured delay profiles of ultrasonic waves coming from the medium [14], while others find the speed of sound that maximizes a coherence-based criterion [15]. Such a criterion is typically based on the Van-Cittert-Zernike theorem [16]–[18]. Further methods search for the speed of sound that optimizes focusing properties [19], entropy [20], or investigate image deformations during compounding [21].

However, such techniques cannot measure the speed of sound of a given region independently from shallower tissues. They rather measure the round-trip propagation time of the ultrasound wave between the probe (positioned at the skin surface) and the region of interest, and output the corresponding average speed of sound, which is equal to the local speed of sound plus a bias caused by superficial layers. We call such methods average methods or integrated methods.

To tackle this problem, a novel framework has been introduced by Jaeger *et al.* [22], in which a tomographic inverse problem is solved in order to retrieve a map of local speed of sound from integrated measurements. Later developments have reduced artifacts, unlocked *in vivo* applications [23]–[27], generalized the framework to other physical models [28]–[31], to different array geometries e.g. convex probes [32], and have been validated clinically [9].

In the context of hepatic steatosis assessment, clinicians are interested in quantitative speed of sound values within regions of interest of the liver. While desirable in some cases, a fine-grained map of speed of sound values is not required. As a result, specific methods have appeared in the last five years, making profit of the layered geometry of the liver and overlying tissues to compensate for superficial tissues. A first class of approaches use prior knowledge of superficial layers [15] or rely on prior segmentation of such layers [33], but the need for such priors makes them too complex to be used in most cases. Others simplify the tomographic inversion approach mentioned earlier to a one dimensional scenario where the inversion is only performed in the axial direction

[34]–[37]. Nevertheless, to the best of our knowledge, all existing methods are based either on the derivation or on the compensation of average methods, leading to instabilities in practice.

We propose a new speed of sound estimation method that yields robust and accurate local measurements. This method requires only a few plane-wave emissions, is compatible with real-time imaging, and does not suffer from instabilities in deeper regions. Indeed, the proposed method is neither based on derivation nor on inversion of average estimations, but leverages refraction, a local physical phenomenon, to retrieve directly the local speed of sound. More specifically, as the local steering angle of a plane-wave is affected by speed of sound changes through refraction, we recover the local speed of sound by estimating the local steering angle.

The remainder of the paper is organized as follows. Theoretical aspects related to the proposed method are described in Section II. Experiments and results are detailed in Sections III and IV and discussed in Section V. Concluding remarks are drawn in section VI.

II. THEORY

A. Notations and Basic Considerations

We consider a linear ultrasound probe composed of M ultrasonic transducers regularly arranged along a line. We note x_j the lateral position of transducer j on this line. Each transducer, upon electrical excitation, generates an acoustic pulse $h(t)$ that is supposed to be a gaussian pulse of central pulsation ω_c and standard deviation σ:

$$h(t) = e^{-\frac{t^2}{2\sigma^2}} e^{j\omega_c t}, \qquad (1)$$

using the hilbert notation.

We transmit a planar wavefront within the medium by applying the following linear delay profile to the transducers elements:

$$\tau_j = \frac{x_j \sin \alpha}{c_{th}}, \qquad (2)$$

where c_{th} is a default speed of sound (assumed to be $1540 \, \text{m s}^{-1}$ in most cases) and α is the so-called steering angle. Under the wavefront assumption, we assume that such a planar wavefront propagates in a unique direction α.

During its propagation, the generated wave is scattered by the medium. In most biological tissues, the main scattering regime is called *speckle* and is caused by microscopic granularity of typical size significantly smaller than the wavelength of the ultrasound wave. We model the reflectivity of such tissues statistically using a density function $\chi(\boldsymbol{r})$, where $\boldsymbol{r}$ represents the position in the medium.

As the wavefront is scattered, reflected waves propagate back to the probe where they are recorded and processed. Traditionally, they are delayed and summed to focus on specific points of interest by means of a so-called beamforming algorithm [38]. In this work, we implement a different strategy since we receive plane-waves, in a similar way to the strategy used in transmit. To receive a steered plane-wave of angle β, we delay the element-raw-data $\{s_j(t)\}_{1 \leq j \leq M}$ recorded by each transducer element with a linear delay law and we

sum the delayed signals, as in the standard delay-and-sum algorithm, such that:

$$s(t) = \sum_{1 \leq j \leq M} s_j \left(t - \frac{x_j \sin \beta}{c_{th}} \right). \qquad (3)$$

As for the transmit, this reception strategy is equivalent to selecting a unique direction in the received echoes.

We now introduce two key quantities for the understanding of the proposed method: the mid-angle $\gamma = (\alpha + \beta)/2$ and the half-angle difference $\delta = (\alpha - \beta)/2$ (angles are defined using the trigonometric convention, i.e. positive in the anticlockwise direction). Such quantities are displayed in Fig. 1a.

We consider a medium Ω composed of a superposition of N horizontal layers. We note c_i the speed of sound in the i-th layer. We position our linear probe on top of this medium. Such a configuration is represented in Fig. 1b.
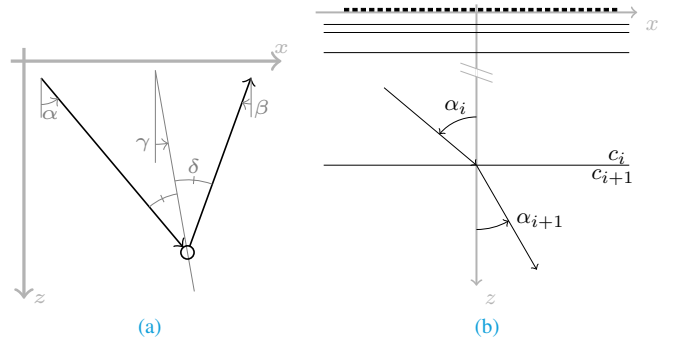

Fig. 1. (a) Definition of the mid-angle γ and of the half-angle-difference δ upon transmission of angle α and reception of angle β. (b) Example of a refraction configuration. The linear probe is represented on top, with filled square for transducers. Underneath is the imaged medium, made of several layers.

As the speed of sound in the medium may not match the default speed of sound used in (2), we denote by $-_{th}$ all the quantities based on the default speed of sound c_{th}. If the speed of sound in the shallowest layer c_1 is different from c_{th}, the steering angle in the first layer α_1 will be different from α_{th}. By geometric considerations, we obtain that α_{th}, α_1, c_{th} and c_1 are linked by the formula:

$$\frac{\sin \alpha_{th}}{c_{th}} = \frac{\sin \alpha_1}{c_1}. \qquad (4)$$

Moreover, at the interface of two different layers, it is well known that ultrasound waves undergo refraction. Their direction of propagation is directly affected by the values of speed of sound in each layer. More specifically, at the interface between layers i and $i + 1$ for $i \in [1, N - 1]$, noting α_i the steering angle in the i-th layer, Snell's law of refraction states that:

$$\frac{\sin \alpha_{i+1}}{c_{i+1}} = \frac{\sin \alpha_i}{c_i}, \quad \forall i \in [1, N - 1]. \qquad (5)$$

Using (4) and (5) and a recursion on the layers, we obtain that:

$$\frac{\sin \alpha_{th}}{c_{th}} = \frac{\sin \alpha_i}{c_i}, \quad \forall i \in [1, N]. \qquad (6)$$

Thus, since α_{th} and c_{th} are known, the problem of recovering the speed of sound in the i-th layer amounts to estimating

the steering angle α_i of the planar wavefront in the i-th layer, independently of all other layers.

Due to our specific angular transmit-receive strategy, Equation (6) is valid both in transmission and reception, and the two directions must be taken into account to estimate accurately the local speed of sound.

B. Theoretical Basis of the Proposed Speed of Sound Estimation Method

The proposed method exploits discrepancies between two angular-raw-data in order to infer local values of steering angles. To do so, it relies on remarkable tomographic properties of speckle in angular-raw-data. These properties link time scaling of angular-raw-data to local steering angles and therefore to local values of speed of sound. Thus, by measuring a scaling factor between two angular-raw-data, the proposed method directly estimates the local value of speed of sound.

Consider a layer of interest in which the speed of sound is uniform. We note this speed of sound c_r and denote by $-_r$ all associated quantities. For the sake of simplicity, we choose a time origin as well as a spatial origin located at the top of this layer, so that we only deal with the layer of interest without having to handle offset constants.

We acquire a first signal $s(t)$ from the transmission and the reception of steered plane-waves of theoretical angles α_{th} and β_{th}. Using a pulse-echo spatial impulse response model [39] where standard hypotheses are made, e.g. neglecting multiple scattering or transducers directivity, such a signal can be expressed as :

$$s(t) = \iint_{\boldsymbol{r}} \chi(\boldsymbol{r})h(t - \tau(\boldsymbol{r}))d\boldsymbol{r}, \qquad (7)$$

where h is the emitted pulse described in (1), $\boldsymbol{r}$ is the position in the medium, and $\tau(-)$ is the round trip time of flight. In the angular framework, the round trip time-of-flight τ is defined as:

$$\tau(\boldsymbol{r}) = \frac{x \sin \alpha_r + z \cos \alpha_r}{c_r} + \frac{x \sin \beta_r + z \cos \beta_r}{c_r}. \qquad (8)$$

With the trigonometric variable change of Fig. 1a applied to real angles, we introduce γ_r and δ_r the mid-angle and the half-angle-difference. We define the corresponding axial coordinate $\zeta = x \sin \gamma_r + z \cos \gamma_r$ and lateral coordinate $\xi = z \sin \gamma_r - x \cos \gamma_r$. With these new coordinates, we can write $\boldsymbol{r} = (\zeta \sin \gamma_r - \xi \cos \gamma_r, \zeta \cos \gamma_r + \xi \sin \gamma_r)$. Using a change of variable $(x, z) \to (\zeta, \xi)$ in (7), we obtain:

$$s(t) = \int_{\zeta} \left(\int_{\xi} \chi(\boldsymbol{r}) d\xi \right) h\left(t - 2\frac{\zeta \cos \delta_r}{c_r} \right) d\zeta. \qquad (9)$$

In this equation, the integral along ξ corresponds to the integral of reflectivity function along a line of angle γ_r. It can be reformulated using the Radon transform formalism as $\mathcal{R}\chi(\gamma_r, \zeta)$, the Radon transform of the reflectivity function χ along angle γ_r at depth ζ. The external integration over the variable ζ corresponds to the contribution along the direction

orthogonal to γ_r. With these new notations, Equation (9) can be simplified as follow:

$$s(t) = \int_{\zeta} \mathcal{R}\chi(\gamma_r, \zeta) h\left(t - 2\frac{\zeta \cos \delta_r}{c_r} \right) d\zeta. \qquad (10)$$

As the reflectivity function is unknown, local steering angles are well hidden in this signal. To recover their values, we compare $s(t)$ to another signal $s'(t)$ obtained from the emission and reception of a steered plane-waves of different angles α'_{th} and β'_{th}. Similarly to (10), we obtain:

$$s'(t) = \int_{\zeta} \mathcal{R}\chi(\gamma'_r, \zeta) h\left(t - 2\frac{\zeta \cos \delta'_r}{c_r} \right) d\zeta. \qquad (11)$$

The comparison between (10) and (11) reveals that if we choose the steering angles such that $\gamma_r = \gamma'_r$, the signal $s(t)$ and $s'(t)$ would be equal up to a factor and a rescaling. For instance, under the simplification that $h(t)$ is the Dirac delta function, it can be easily shown that $s(t \cos \delta_r) \propto s'(t \cos \delta'_r)$, or equivalently $s(t \cos \delta_r / \cos \delta'_r) \propto s'(t)$. Hence, providing that the scaling factor $SF = \cos \delta_r / \cos \delta'_r$ can be measured from s and s', we can retrieve the true speed of sound using (6) and a second order small angle approximation as described in Supplementary-S-I, leading to:

$$c_r \approx c_{th} \sqrt{2 \frac{1 - SF}{\delta_{th}^2 - \delta_{th}'^2}}. \qquad (12)$$

In practice, we choose our theoretical angles such that $\gamma_{th} = \gamma'_{th}$, which is equivalent to $\gamma_r = \gamma'_r$ under the small angle approximation. The use of (12) supposes that we are able to compute the scaling factor SF accurately, which is in fact a difficult task.

To do so, instead of a dirac simplification, we rely on the temporal extent and oscillation of our gaussian pulse $h(t)$. We suggest the following two-step approach. First, the signals $s(t)$ and $s'(t)$ are rescaled into $\widehat{s}(t) = s(t \cos \delta_{th})$ and $\widehat{s}'(t) = s'(t \cos \delta'_{th})$. At this point, the two signals are theoretically re-aligned and equal, but a residual scaling factor may remain due to the difference between theoretical and real angles. In a second step, we measure this residual scaling as follows:

$$RSF := \frac{\partial \angle \langle \widehat{s}(t)\widehat{s}'(t)^* \rangle}{\partial t}, \qquad (13)$$

where $-^*$ is the complex conjugation, $\langle - \rangle$ the averaging operator and $\angle -$ the phase operator. Such a scaling factor embodies different operations that can be detailed for better understanding. Intuitively, since $\widehat{s}(t)$ and $\widehat{s}'(t)$ are oscillating signals, the phase of their correlation gives us information about their relative time shift. The averaging process is used to remove randomness induced by different speckle realizations. Finally, a derivation is applied to extract the evolution of the time shift, i.e. the residual scaling factor. Using speckle properties and the small angles approximation, we can demonstrate that the residual scaling factor RSF defined above is equal to (*cf* Appendix-A):

$$RSF \approx \frac{\omega_c}{2} \left(1 - \frac{c_r^2}{c_{th}^2} \right) \left(\delta_{th}'^2 - \delta_{th}^2 \right). \qquad (14)$$

This expression directly links the residual scaling factor to the real local speed of sound value c_r. It is interesting to note that due to the derivation, this residual scaling factor is independent to any phase offset that may be introduced by unknown shallow tissues. It is therefore a local quantity. We can indeed recover the exact location that corresponds to a RSF measurement. The RSF is estimated via (13) from $\langle \widehat{s}(t)\widehat{s}'(t)^* \rangle$, which can be written as an integral of a gaussian pulse (cf Appendix-A) centered around $\zeta = \mu_{tot} \approx c_r t/2$. Thus, a RSF computed at t roughly corresponds to the line of angle γ_r with respect to the horizontal, that crosses the central vertical axis $x = 0$ at $z = c_r t/2$.

The preliminary rescaling makes the RSF different from the scaling factor SF mentioned in (12), thus leading to a different speed of sound formula. In this case, we obtain from (14) the following expression of the local speed of sound:

$$c_r = c_{th}\sqrt{1 - \frac{\frac{2}{\omega_c}RSF}{\delta_{th}'^2 - \delta_{th}^2}}. \tag{15}$$

Equation (15) is groundbreaking in two aspects. First, it enables the extraction of the local speed of sound from refraction effects in speckle. Refraction is most of the time neglected in ultrasound imaging as it is a second order phenomenon when using small angles. The precision given by the angular transmit-receive strategy allows us to leverage this effect reliably. Secondly, it links directly the local speed of sound to the two signals acquired in the region of interest, in stark contrast with state-of-the-art methods which most of the time require the analysis of a large part of the medium and the resolution of an optimization problem. The estimation process does not require knowledge of superficial layers and can be applied at any depth indifferently.

C. Proposed Speed of Sound Estimation Method

A summary of the presented method is given in Fig. 2.

First, a reasonable value of speed of sound $c_{th} = 1540\,\mathrm{m\,s^{-1}}$ is considered. Then, two angle pairs $(\alpha_{th}, \beta_{th})$ and $(\alpha_{th}', \beta_{th}')$ are chosen so that they share the same mid-angle ($\gamma_{th} = \gamma_{th}'$). In practice, we choose two pairs with not too close squared half-angle-differences ($|\delta_{th}'^2 - \delta_{th}^2| > \epsilon$, with $\epsilon = 0.015\,\mathrm{rad}^2$) in order to prevent instabilities in (15). Then, the following steps are performed.

1) Two signals $s(t)$ and $s'(t)$ are acquired by transmitting plane-waves of theoretical angles α_{th} and α_{th}' and by receiving along directions β_{th} and β_{th}' using (3).
2) Angular signals s and s' are rescaled into $\widehat{s}(t) = s(\cos\delta_{th}t)$ and $\widehat{s}'(t) = s'(\cos\delta_{th}'t)$.
3) The residual scaling factor RSF is computed from $\widehat{s}$ and $\widehat{s}'$ using (13), where the average is performed by taking the mean of the correlation signal over $1\,\mathrm{cm}$, windowed by a hanning function, and the derivation is a simple forward temporal difference.
4) The local speed of sound is deduced from RSF using (15).

This method only requires two plane-wave transmits and a few processing steps with low computation complexity.

However, to increase the robustness of the method, we decide to run it multiple times with different transmit and receive angles. Each run gives us an estimate of the local speed of sound. The different estimates are averaged to obtain a more stable result. Empirical observations led us to transmit 15 angles ranging from -14° to 14° with 2° angle steps, as it optimizes the trade-off between performances and number of emissions. From the backscattered echoes of each transmitted angle, we typically choose a hundred mid-angles ranging from -15° to 15° thus defining a hundred receive angles that we only consider if lower than 30°. We then gather these signals in pairs that share a common mid-angle with not too close half-difference angles, leading to a few thousands speed of sound estimates, each associated with a specific line of angle γ_r and depth $\zeta = c_r t/2$. To preserve the spatial equivalence between the image beamformed at c_{th} and the estimated speed of sound, we represent this line as a line of angle γ_{th} and depth $\zeta = c_{th}t/2$. To combine all the estimates, we neglect all γ_{th} (i.e. suppose all lines horizontal) and average the speed of sound values at each depth. Such mixing affect the resolution of the method, as a mid-angle of 15° can lead to a depth offset up to $5\,\mathrm{mm}$ on the edges of the region of interest. However, the average offset is smaller than the half length of the hanning window used for the correlation averaging and have therefore a limited impact in our case.

In order to limit the lateral extension of the proposed method, we use a lateral apodization in the reception step to select a region of interest two centimeters wide.

The code of the resulting method is available at `https://github.com/E-Scopics/refraction-based-sos`.

III. EXPERIMENTS

The performance of the proposed method is assessed against two state-of-the-art approaches.

The first one is a coherence maximization technique, a standard speed of sound estimation method first described in [16] and implemented for instance in [15]. We label this method *average method* as it estimates the average slowness (inverse of the speed of sound) between the probe and a given region of interest. The second one is the *dix inversion* method proposed by Ali *et al.* [37], a layered approach that deduces the local speed of sound from the output of the average method. We used the python code from the github repository provided by the authors. We tuned the regularization parameter to 10^9 to obtain a similar resolution than the proposed method while maximizing the robustness and cropped profiles on both sides to prevent border effects.

The proposed method is computed on a 15 plane-waves dataset, and the two benchmark methods are computed on a synthetic aperture dataset made of 128 single element transmits. The three methods are compared on simulations, *in vitro* and *in vivo* experiments.

A. Simulation

A five layered medium is simulated using k-Wave [40]. Ultrasonic speckle is generated with a random density distribution with small granularity. Five horizontal layers are defined
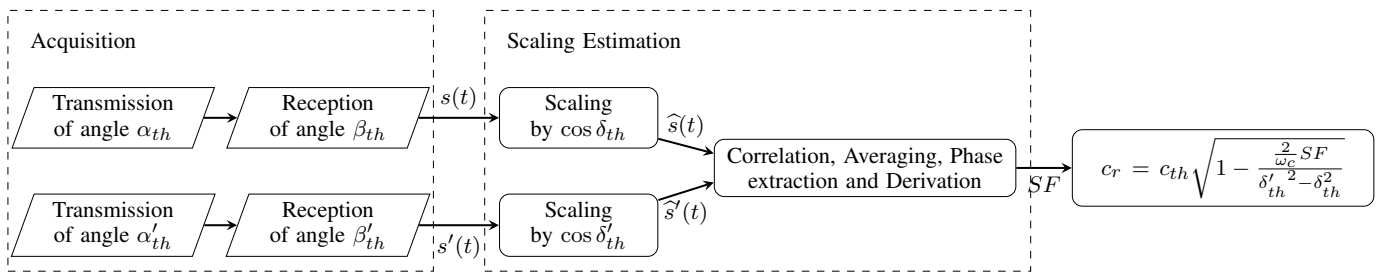
Fig. 2. Summary of the proposed method. The technique is divided in three blocks. First, the acquisition block composed of the two transmissions and receptions. Then, the residual scaling factor estimation block, were signal are scaled by a theoretical factor and compared. Finally, equation (15) is applied to deduce the local speed of sound.

with speed of sound values typically encountered *in vivo* and no attenuation. Their characteristics are reported in Table I.

TABLE I

SIMULATION LAYOUT: SPEED OF SOUND VALUES AND LAYER THICKNESSES.

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Speed of Sound ($\mathrm{m\,s^{-1}}$)	1540	1450	1580	1540	1480
Thickness (mm)	2	15	10	40	40

We simulate a 128-element linear probe of pitch $300\,\mu m$ which impulse response is approximated by a Gaussian pulse (center frequency $5\,MHz$, relative bandwidth 66%) The full synthetic aperture dataset is generated, and the plane-wave dataset is synthesized a posteriori by summing the recorded signals with linear delay laws. Radio frequency (RF) signals are demodulated at $5\,MHz$ and decimated to obtain complex in-phase quadrature (IQ) signals.

The speed of sound estimation methods are compared qualitatively by computing the speed of sound at various depths and generating a profile along depth. The quantitative comparison is performed between $30\,mm$ and $60\,mm$, which is in the fourth layer. Within this range, speed of sound is estimated every millimeter, leading to 30 estimates for each method. Means and standard deviations are computed to assess biases and variabilities.

To evaluate the techniques in realistic settings, a noiseless and a noisy configuration are considered. The noisy configuration is obtained by adding a complex gaussian noise of power equal to the plane-wave signal power to the final IQ signals.

In order to evaluate the real-time capability of the local speed of sound estimation techniques, we compare the variability for different number of transmits, ranging between 2 and 100, in the two configurations. Transmit angles (respectively elements) are chosen to be uniformly spaced within a constant range of 30° (resp. 100 elements).

B. In vitro

The three methods are compared *in vitro* on a multipurpose tissue mimicking reference phantom (Model 054GS, CIRS, USA) whose ground truth speed of sound has been measured at $1550\,\mathrm{m\,s^{-1}}$, used alone (uniform medium) or underneath a homogeneous layer of gelified water measured at $1490\,\mathrm{m\,s^{-1}}$, either $22\,mm$ thick (*bilayer 1*) or $40\,mm$ thick (*bilayer 2*).

Ultrasonic pulses are emitted at $5\,MHz$ using an E-Scopics system equipped with a 128 elements linear probe of pitch $300\,\mu m$. Demodulated raw ultrasonic signals (IQ) are recorded after low-pass filtering and decimation at $3.9\,MHz$. To compare the three methods, we use an ultrasound transmission sequence that alternatively emits single transducers and plane-wave transmissions to record both datasets.

Estimation methods are compared qualitatively by computing the speed of sound profile along depth. For quantitative results, three depth ranges are considered when possible, namely $10-30\,mm$, $30-50\,mm$ and $50-70\,mm$. Speed of sound is estimated every millimeter, and values are gathered by ranges. Means and standard deviations are computed to assess biases and variabilities.

C. In vivo

In vivo acquisitions are performed by imaging the liver of healthy volunteers (in compliance with the ethical principles of the Declaration of Helsinki 2018). Three different imaging planes (typically intercostal, subcostal and epigastric) are selected on four volunteers to study the reproducibility of liver measurements for different configurations of superficial tissues. For each imaging plane, the same acquisition setup as for *in vitro* experiments is used but repeated ten times in a $5\,s$ long time interval to assess repeatability and robustness against disturbances due to breathing, cardiac beats and probe movements.

A $20\,mm$ wide depth range is considered for each acquisition and manually placed within the liver, trying to avoid strong heterogeneities or reverberation artifacts.

Speed of sound is estimated within this range for each acquisition. Mean and standard deviation across the ten acquisitions are finally computed to assess the repeatability.

IV. RESULTS

In Fig. 3 is displayed the B-Mode image (leftmost plot) corresponding to the noiseless simulation experiment along with the speed of sound profiles (rightmost plot).

Different layers can be distinguished in the beamformed image through visual assessment as speed of sound changes imply acoustic impedance changes. On the right, the three estimated profiles are compared to the ground truth. We observe that the dix inversion method and the proposed method both resolve the different layers and estimate accurately the
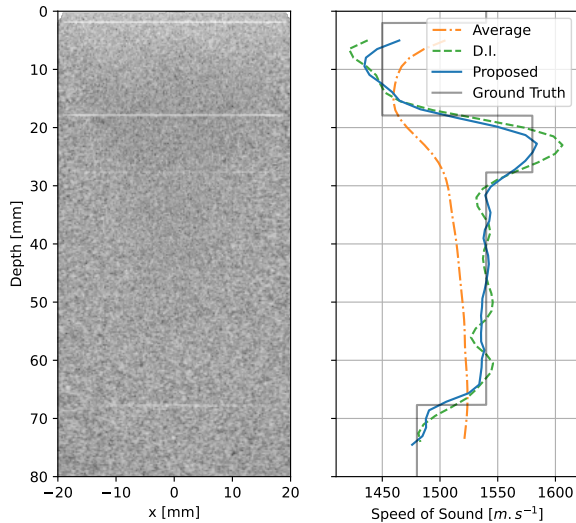
Fig. 3. Simulation results in a noiseless configuration, with the B-mode image on the left and the speed of sound profiles along depth on the right. D.I. stands for the dix inversion method. First and last **5 mm** are cropped to remove border effects of averaging and regularization.

local speed of sound. Both profiles appear smooth, with an axial resolution of approximately $10\,\mathrm{mm}$. Such characteristics are mainly driven by the averaging of the correlation for the proposed method, and by the regularization term for the dix inversion method. Furthermore, compared to the dix inversion method, the proposed method stays stable along depth and does not exhibit growing oscillations. As for the average method, it yields estimates biased by superficial layers.

The noisy configuration is displayed in Fig. 4. We observe that the average method is still biased and slightly less robust, while the dix inversion method shows strong oscillations and is no longer accurate. On the other hand, the proposed method is still local, accurate and stable.
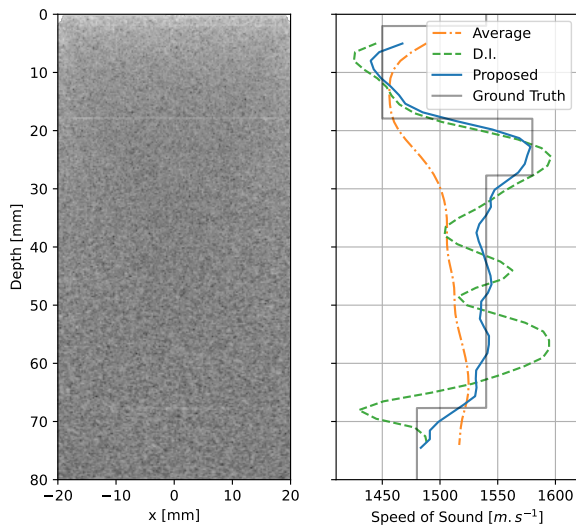

Fig. 4. Simulation results in a noisy configuration, with the B-mode image on the left and the speed of sound profiles along depth on the right. D.I. stands for the dix inversion method.

Quantitative measurements are reported in Table II. They reveal that in the noiseless case, the two local methods are

accurate within $2\,\mathrm{m\,s^{-1}}$ of the ground truth, whereas the average method is biased by more than $20\,\mathrm{m\,s^{-1}}$. The proposed method is about two times more robust than the dix inversion method as highlighted by values of standard deviation. In the noisy case, the dix inversion method performances are significantly affected, with a variability five times higher than in the noiseless scenario, whereas the proposed method is still accurate within $2\,\mathrm{m\,s^{-1}}$ of the ground truth, and exhibits only a slightly higher variability than in the noiseless case.

TABLE II
SIMULATION RESULTS

	Ground Truth	Average Method		Dix Inversion Method		Proposed Method	
		Mean	Std	Mean	Std	Mean	Std
Noiseless	1540	1515	4.7	**1538**	5.1	**1538**	**2.5**
Noisy	1540	1510	5.6	1546	25.0	**1539**	**4.0**

Mean and standard deviation of speed of sound measurements for the three methods and two noise configurations. Values are expressed in $\mathrm{m\,s^{-1}}$. Bold values indicate the best mean and standard deviation for each configuration.

Finally, the variability of the dix inversion and the proposed methods as a function of the number of transmits in the noiseless and noisy scenarios is displayed Fig. 5. We observe
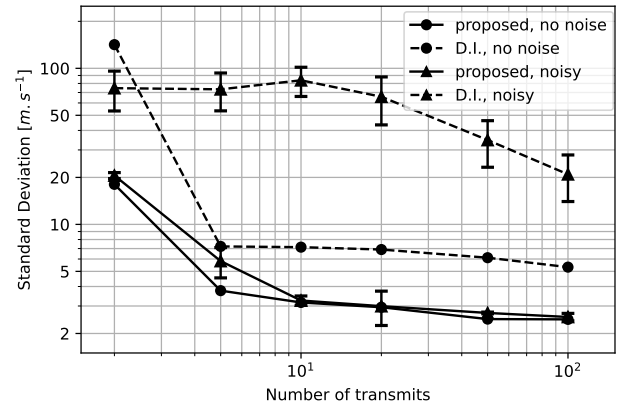

Fig. 5. Speed of sound variations in the 30-60mm depth range for different number of transmits, in noiseless and noisy configurations. Noisy configurations are repeated 3 times with independent noise to obtain 95% confidence intervals. Std values of Table II correspond to one of these 3 realizations, with 128 transmits for the dix inversion method and 15 transmits for the proposed method.

that in the noiseless case (circles), both methods could work well with 5 transmits, with a slight advantage for the proposed method. In the noisy scenario (triangles), the performances of the proposed method stay roughly the same whereas the dix inversion method shows a significant performance drop. In the end, in the noisy case, the proposed method with 5 transmits performs better than the dix inversion method with 100 transmits.

Similarly, for the *in vitro* experiment, the beamformed images are displayed on Fig. 6 alongside estimated speed of sound profiles for the three configurations. Visually, the average method is the most robust but is biased in the bilayer cases, and the dix inversion method exhibits strong oscillations near the interface and at high depths, which affects its ability to
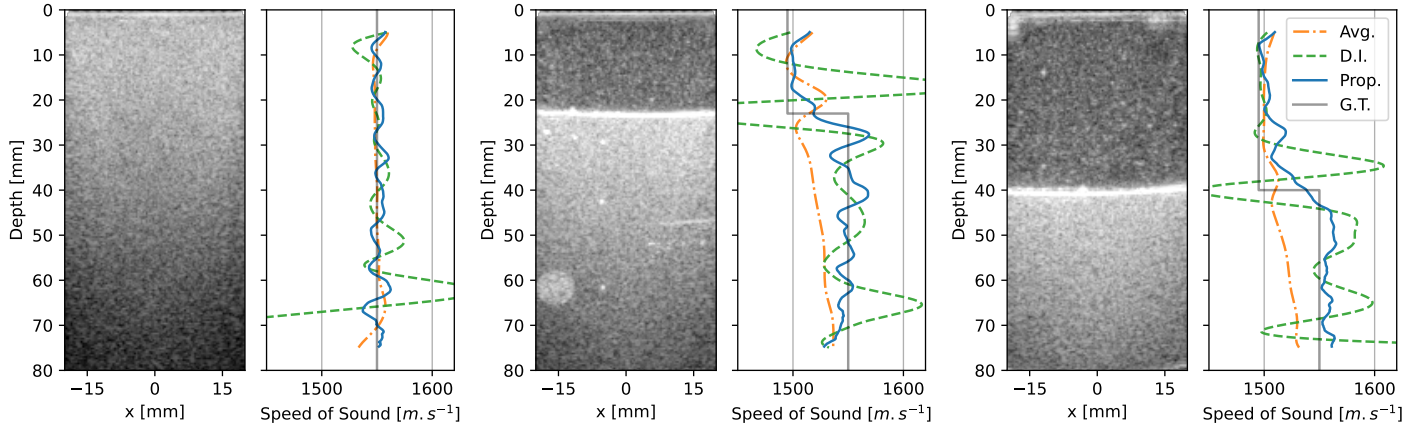
Fig. 6. *In vitro* results, with the beamformed image on the left and the compared profiles on the right, in three different configurations (from left to right, uniform medium, bilayer 1 with interface at **22 mm**, and bilayer 2 with an interface at **40 mm**). *Avg.* stands for average method, *D.I.* for the dix inversion method, *Prop.* for the proposed method and *G.T.* for the ground truth.

provide reliable local estimate of speed of sound. The proposed method seems to exhibit the best compromise as it is unbiased with reduced oscillations. Quantitative results are presented in Table III for different depth ranges. In every configuration, the

TABLE III
In Vitro RESULTS

		Average Method		Dix Inversion Method		Proposed Method	
		Mean	Std	Mean	Std	Mean	Std
Uniform Medium	10-30 mm	1547	0.6	1548	4.3	1550	3.3
	30-50 mm	1549	0.7	1553	6.3	1553	4.1
	50-70 mm	1553	2.3	1560	54.8	1549	6.8
Bilayer 1	30-50 mm	1518	4.6	1555	12.2	1549	10.0
	50-70 mm	1530	2.7	1560	29.3	1547	4.7
Bilayer 2	50-70 mm	1523	4.1	1568	18.1	1557	3.7

Mean and standard deviation of speed of sound measurements for the three methods and different regions of interest in three different configurations. Values are expressed in $\mathrm{m\,s^{-1}}$.

proposed method estimates accurately the calibrated phantom speed of sound with biases lower than $7\,\mathrm{m\,s^{-1}}$. Moreover, we notice that the proposed method is more robust than the dix inversion method, especially at high depths where the signal to noise ratio (SNR) is the lowest.

In vivo results are displayed in Fig. 7. In this case, we compare the three methods by plotting speed of sound estimates along with the corresponding variability over ten acquisitions for each situation. Regarding the box positioning, we relied on a deeper region of interest for patient 4 as b-mode images exhibit strong reverberation artifacts, and a shallower region for the third acquisition of patient 3 whose b-mode image revealed heterogeneities at 40 mm.

In vivo results, displayed in Fig. 7, show the difference between the three methods. First, the average method is stable and repeatable for each imaging plane, as shown by the low uncertainty measured for each imaging plane (visually assessed by the size of the errorbar in the rightmost plots of Fig. 7). Two estimates show higher variability, likely caused by motion between the acquisitions and closeness to the

superficial layers. However, estimated speed of sound varies across imaging planes, as confirmed by the results of the Welch t-test of independence which demonstrate significant differences (p-value <0.001) in 4 cases among 12. This lack of reproducibility can be understood as the average method is not local and thus sensitive to speed of sound variations in superficial layers. Secondly, the dix inversion method exhibits a low repeatability, with variations of several tens meters per second within the same imaging plane. This low reliability makes any further statistical study difficult. Finally, the proposed method exhibits a high repeatability for each imaging plane (low standard deviations), and a high reproducibility across imaging planes for the same volunteer (similar average values for the three imaging planes for each patient).

We can therefore conclude that from a repeatability and reproducibility point of view, the proposed method is the only one that is reliable enough for *in vivo* speed of sound quantification.

V. DISCUSSION

In this study, we present a speed of sound estimation method based on refraction. This method detects the scaling between signals caused by refraction, using tomographic and speckle properties. We established a theoretical relationship between the detected scaling factor to the actual local speed of sound through a direct formula reported in equation (15). Although the elemental algorithm only requires two signals extracted from two plane-wave transmissions, a generalized version making use of multiple signals extracted from 15 plane-wave transmissions was tested experimentally.

Simulation results prove the ability of the proposed method to estimate local speed of sound values from refraction effects in speckle. Moreover, the absence of bias experimentally validates the theoretical analysis and resulting formula (15).

The proposed technique is conceptually different from state-of-the-art local speed of sound estimation methods. Indeed, these methods traditionally measure quantities that relate to the integrated slowness, like time of flight differences or a global speed of sound, and extract the local speed of sound
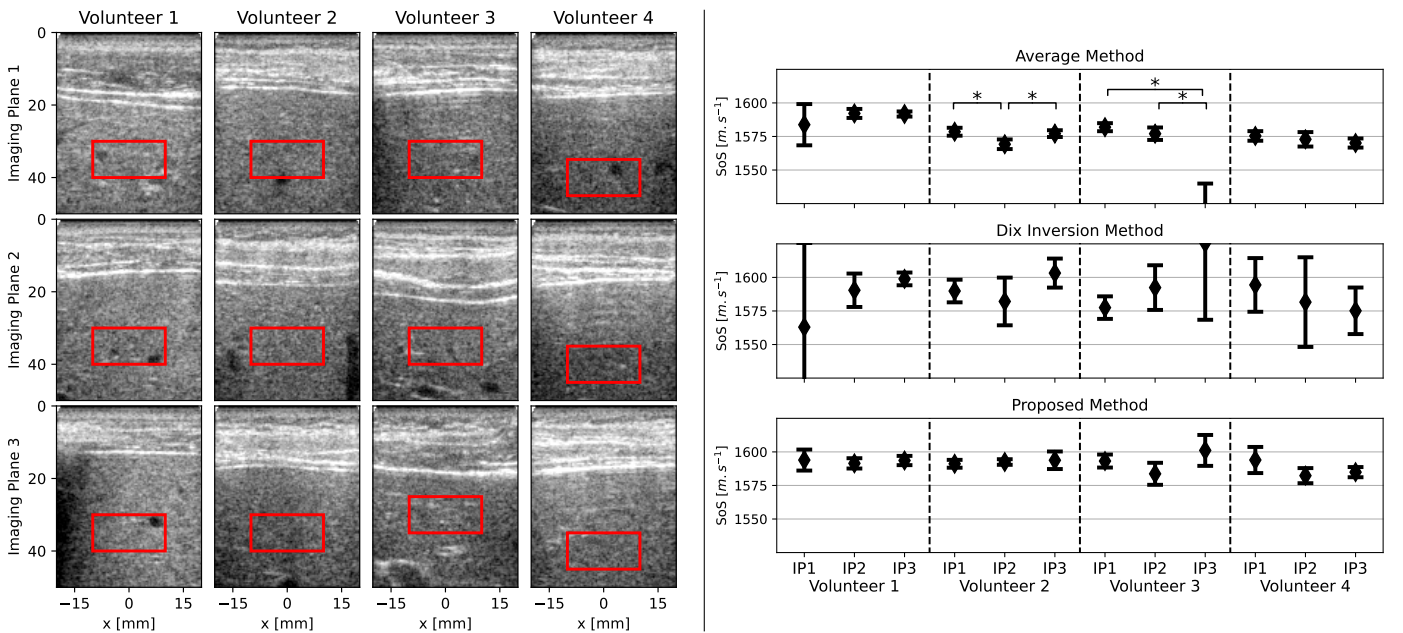
Fig. 7.   *In vivo* study. One beamformed image for each imaging plane, for each volunteer is displayed on the left, with a red box delimiting the region of interest. Speed of sound means and standard deviations over ten acquisitions are plotted via error bars on the right, for each estimation method. * represents a p-value <0.001 in Welch's t-test of independent means.

from such integral, by means of prior knowledge on superficial layers, dix inversion or tomographic inversion. Our method, on the other hand, cancels the cumulative effects with a derivation and measures refraction, a local phenomenon.

The computed ultrasound tomography in echo mode method (CUTE) described in [25] exhibits strong similarities with the proposed method. Both emit plane waves, receive angularly the echoes, map them to spatial coordinates (beamforming for CUTE and scaling for the proposed method) and extract phase differences using the common mid-angles strategy. The extracted quantity, called phase-shifts in [25], relates strongly to $\angle \langle \widehat{s}(t)\widehat{s}'(t) \rangle$ used in (13). However, the processing chain of the proposed method maximizes the angular accuracy at the expense of lateral resolution whereas CUTE maximizes the trade-off between angular accuracy and spatial resolution. Indeed, phase-shifts are used in completely different ways. CUTE leverages a ray-tracing approach to output a map of local speed of sound from the accumulated slowness embedded in phase shifts. Hence, lateral resolution is required for sufficient conditioning. On the other hand, the proposed method removes such cumulative effects by deriving phase shifts into RSF and estimates the local speed of sound at the depth of interest from refraction effects hidden in RSF. In this case, the purely angular transmission-reception strategy is not only optimal for phase shift accuracy but also magnifies refraction effects.

Benchmarks show that the proposed method is reliable *in vivo*, as opposed to the two methods we compared against. First, global methods (e.g. the average method) are not local and thus not repeatable. We observed biases of 10 to $30\,\mathrm{m\,s^{-1}}$, but we believe that it can be far higher for overweight patients that show more than $4\,\mathrm{cm}$ thick superficial layers. On the other hand, most local methods (e.g. the dix inversion method) are not robust enough in challenging situations (low SNR,

aberrations, movements), most likely due to the inversion process that transform the global speed of sound into local measurements. These issues are overcome by the proposed method. Indeed, a higher SNR than single element transmissions is allowed by plane-wave transmissions and reinforced by the remarkable coherence in speckle brought by the angular approach. Moreover, the choice to sacrifice lateral localization in order to maximize angular precision in the proposed method leads to robust and accurate phase shifts, which appear smooth with reliable derivatives.

We noticed that the proposed method is sensitive to reverberation artifacts. This aspect is under study and will be published in future works.

Initial hypotheses consisted in supposing a horizontally layered medium. Still, the limit case of slightly non-horizontal configurations can be studied theoretically and it can be demonstrated (*cf.* Supplementary-S-II) that the proposed method remains valid up to $10\,\mathrm{m\,s^{-1}}$ for interfaces angled up to $15°$. *In vivo* experiments show that the method is indeed robust against not perfectly horizontal layers and small aberrations.

On top of accuracy and stability, the main improvement compared to existing methods lies in its practical simplicity, making it a strong candidate for clinical applications. Firstly, it does not require the user to manually define layers or enter prior information. The user only needs to position a box inside the medium of interest. Secondly, the proposed method needs a low number of transmits, typically 15 plane-waves, as opposed to the hundred of emissions required by state-of-the-art techniques. This leads to a lower data rate, a simplified ultrasound sequence and a reduced acquisition duration, which makes it compatible with duplex configurations in which speed of sound estimation is mixed with B-Mode imaging. Besides, it drastically reduces the impact of probe motion and tissues

displacement that may cause instabilities to speed of sound estimation. The video, provided as supplementary material to this manuscript, displays the proposed method operating in real-time alongside B-Mode images. It illustrates the practical interest of the technique. Thirdly, its computational cost is low. As seen in Fig. 2, it only requires an interpolation and negligible operations (e.g. correlation or filtering) per pair of signals. Its computational cost is therefore significantly lower than iterative methods that optimize an index computed at multiple speed of sounds, or tomographic methods that solve an inverse problems involving large matrices. We roughly compare the computational needs between the proposed method and the dix inversion method in Appendix-B. Results are displayed Table IV. We see that the proposed method needs 1000 times

TABLE IV
COMPUTATIONAL LOAD

	D.I. Method	Proposed Method	Gain
Number of operations	10^{12}	10^9	x1000
Computational time (CPU)	>15 min	<1 s	x1000

less operations than the dix inversion method, leading to a computational time gain of three order of magnitude in CPU-based implementations. Furthermore, its highly parallelizable architecture will likely lead to a computational time of tens of milliseconds in GPU-based implementations.

As a result, several clinical applications can be considered for this method, such as steatosis assessment in the liver. Future work will consist in transposing the method to the convex configuration which is usually preferred for abdominal applications.

VI. CONCLUSION

The study of refraction effects in ultrasound imaging enables a robust and accurate local speed of sound estimation.

In summary, we rely on an angular point of view of ultrafast ultrasound imaging based on the use of steered plane-waves both in transmission and in reception. Such a strategy allows us to exploit refraction effects to robustly estimate the speed of sound from the received raw signals in a fast and efficient way, with a few transmits and a low computational cost. Experiments show that the resulting estimates are accurate and more robust than state-of-the-art methods, especially at high depths.

We obtain a local speed of sound estimation technique that is compatible with real-time imaging and seems a promising candidate for future clinical investigations.

APPENDIX

A. Residual Scaling Factor Expression in Speckle

Starting from (10) and (11), we can write the re-aligned signals:

$$\widehat{s}(t) = \int_{\zeta} \mathcal{R}\chi\left(\gamma_r, \zeta\right) h\left(t \cos\delta_{th} - 2\frac{\zeta \cos\delta_r}{c_r}\right) d\zeta, \quad \text{(A.1)}$$

and similarly for $\widehat{s}'(t)$ with $\cos\delta'_{th}$ and $\cos\delta'_r$. To simplify further calculations, we introduce intermediate functions:

$$\begin{cases} f(t,\zeta) = t \cos\delta_{th} - 2\frac{\zeta \cos\delta_r}{c_r} \\ f'(t,\zeta) = t \cos\delta'_{th} - 2\frac{\zeta \cos\delta'_r}{c_r} \end{cases} \quad \text{(A.2)}$$

From there, we obtain for the averaged correlation:

$$\langle \widehat{s}(t)\widehat{s}'(t)\rangle = \iint_{\zeta\,\zeta'} \langle \mathcal{R}\chi\left(\gamma_r, \zeta\right) \mathcal{R}\chi\left(\gamma_r, \zeta'\right)^*\rangle \cdot$$
$$h(f(t,\zeta))h^*(f'(t,\zeta'))d\zeta d\zeta'. \quad \text{(A.3)}$$

Let us first focus on the first part of this equation, i.e. on speckle properties. We have that:

$$\langle \mathcal{R}\chi\left(\gamma_r, \zeta\right) \mathcal{R}\chi\left(\gamma_r, \zeta'\right)^*\rangle = \iint_{\xi\,\xi'} \langle \chi(\boldsymbol{r})\chi(\boldsymbol{r})^*\rangle \, d\xi d\xi'. \quad \text{(A.4)}$$

Following [41], we suppose that speckle is uniform and independent in the layer of interest. Thus, if we average the correlation of different speckle realisations (in our case across depth), we obtain:

$$\langle \chi(\boldsymbol{r})\chi(\boldsymbol{r}')^*\rangle = \begin{cases} |\chi|^2 & \text{if } \boldsymbol{r} = \boldsymbol{r}', \\ 0 & \text{else.} \end{cases} \quad \text{(A.5)}$$

From there, we obtain that the integrated quantity is always 0 except when $\xi = \xi'$ and $\zeta = \zeta'$. We obtain for the general expression:

$$\langle \widehat{s}(t)\widehat{s}'(t)\rangle = |\chi|^2 l \int_{\zeta} h(f(t,\zeta))h^*(f'(t,\zeta))d\zeta, \quad \text{(A.6)}$$

with l the lateral extent of the considered medium. We obtain a simpler expression, where the two pulses are multiplied at the same depth. As the two pulses are scaled and shifted due to the unknown speed of sound, even after theoretical scaling, this remaining expression is not as simple as it seems. To obtain its analytic expression, we need to go back to the gaussian pulse expression of $h(t) = g(t)e^{j\omega_c t}$, with g a gaussian of mean 0 and standard deviation σ. We obtain:

$$\frac{\langle \widehat{s}(t)\widehat{s}'(t)\rangle}{|\chi|^2 l} = \int_{\zeta} g(f(t,\zeta))g(f'(t,\zeta))e^{j\omega_c\left(f(t,\zeta)-f'(t,\zeta)\right)}d\zeta. \quad \text{(A.7)}$$

Regarding the first part of the integral, we obtain a multiplication of two gaussian functions of variable ζ of mean and standard deviation:

$$\begin{cases} g(f(t,-)) \sim \mathcal{N}\left(\frac{c_r t}{2}\frac{\cos\delta_{th}}{\cos\delta_r}, \frac{\sigma c_r}{2\cos\delta_r}\right), \\ g(f'(t,-)) \sim \mathcal{N}\left(\frac{c_r t}{2}\frac{\cos\delta'_{th}}{\cos\delta'_r}, \frac{\sigma c_r}{2\cos\delta'_r}\right). \end{cases} \quad \text{(A.8)}$$

Besides, we know that the product of two gaussian functions is a gaussian function of mean $\mu_{tot} = \frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}$. We obtain that the global function $g_{tot}(\zeta) = g(f(t,\zeta))g(f'(t,\zeta))$ is a gaussian function of mean:

$$\mu_{tot} = \frac{c_r t}{2}\frac{\cos\delta_{th}\cos\delta_r + \cos\delta'_{th}\cos\delta'_r}{\cos^2\delta_r + \cos^2\delta'_r}. \quad \text{(A.9)}$$

On the other hand, the phase of the complex exponential in the second part of the integral is equal to:

$$\phi = \omega_c \left(t(\cos \delta_{th} - \cos \delta'_{th}) - \frac{2\zeta}{c_r}(\cos \delta_r - \cos \delta'_r) \right), \tag{A.10}$$

which is an affine function in ζ and can be written $\phi = a\zeta + b$. We obtain:

$$\frac{\langle \widehat{s}(t)\widehat{s}'(t) \rangle}{|\chi|^2 l} = \int_{\zeta} g_{tot}(\zeta) e^{j(a\zeta+b)} d\zeta. \tag{A.11}$$

From there, the symmetry around μ_{tot} shows that the phase of such an integral is equal to the value of the affine function at μ_{tot}, i.e. $\angle \langle \widehat{s}(t)\widehat{s}'(t) \rangle = a\mu_{tot} + b$.

We obtain after development and re-factorization:

$$\angle \langle \widehat{s}(t)\widehat{s}'(t) \rangle = \omega_c t \frac{\cos \delta_r + \cos \delta'_r}{\cos^2 \delta_r + \cos^2 \delta'_r} \cdot \\ (\cos \delta_{th} \cos \delta'_r - \cos \delta'_{th} \cos \delta_r). \tag{A.12}$$

Here, the local speed of sound does not appear explicitly. However, refracted angles can be linked to their theoretical value, c_r and c_{th} from refraction analysis, using (6). With a second order Taylor expansion on the angles (supposed small under the small angle approximation), the previous equation can be simplified, leading to a derivative (*cf* Supplementary-S-I):

$$\frac{\partial \angle \langle \widehat{s}(t)\widehat{s}'(t) \rangle}{\partial t} \approx \frac{\omega_c}{2} \left(1 - \frac{c_r^2}{c_{th}^2} \right) \left({\delta'_{th}}^2 - \delta_{th}^2 \right). \tag{A.13}$$

B. Computational loads

In this section, we compare roughly the computational loads of the dix inversion method and of the propose method. We note N_z the number of points considered axially (typically 100) and N_{bf} the number of operations per point needed for beamforming (typically 10 for an interpolation and an apodization).

For the dix inversion method, the main computational load comes from the beamforming at 100 different speed of sounds, along 100 points laterally, N_z points axially, 128 transmits, and 128 channels in reception. This leads to $2 \cdot 10^8 \cdot N_z N_{bf}$ operations.

For the proposed method, the first step is the computation of the angular signals. We receive approximately 100 plane-waves for each 15 transmits, from the 128 transducers signals. To do so, we perform an interpolation and an apodization for each depth, leading to a load of approximately $2 \cdot 10^5 \cdot N_z N_{bf}$. Other steps consist in a correlation and filtering for each pair of signal at each depth. With around $5 \cdot 10^3$ signals, it leads to $2 \cdot 10^5 \cdot N_z$ operations if we count 50 operations per point per pair. In total, the first step is predominant, leading to $2 \cdot 10^5 \cdot N_z N_{bf}$ operations in total.

REFERENCES

[1] J. Ormachea and K. J. Parker, "A preliminary study of liver fat quantification using reported ultrasound speed of sound and attenuation parameters," *Ultrasound Med. Biol.*, vol. 48, no. 4, pp. 675–684, Apr. 2022.

[2] M. Sak *et al.*, "Using speed of sound imaging to characterize breast density," *Phys. Med. Biol.*, vol. 43, no. 1, pp. 91–103, Aug. 2017.

[3] E. A. O'Flynn *et al.*, "Ultrasound tomography evaluation of breast density: a comparison with noncontrast magnetic resonance imaging," *Invest. Radiol.*, vol. 52, no. 6, p. 343, Jun. 2017.

[4] L. Ruby *et al.*, "Breast cancer assessment with pulse-echo speed of sound ultrasound from intrinsic tissue reflections: Proof-of-concept," *Invest. Radiol.*, vol. 54, no. 7, pp. 419–427, Jul. 2019.

[5] D. E. Robinson, J. Ophir, L. S. Wilson, and C. F. Chen, "Pulse-echo ultrasound speed measurements: Progress and prospects," *Ultrasound Med. Biol.*, vol. 17, no. 6, pp. 633–646, Jan. 1991.

[6] M. Imbault *et al.*, "Ultrasonic fat fraction quantification using in vivo adaptive sound speed estimation," *Phys. Med. Biol.*, vol. 63, no. 21, p. 215013, Oct. 2018.

[7] M. D. Burgio *et al.*, "Ultrasonic adaptive sound speed estimation for the diagnosis and quantification of hepatic steatosis: a pilot study," *Ultraschall in der Med.-Eur. J. of Ultrasound*, vol. 40, no. 06, pp. 722–733, Jun. 2019.

[8] A. V. Telichko *et al.*, "Noninvasive estimation of local speed of sound by pulse-echo ultrasound in a rat model of nonalcoholic fatty liver," *Phys. Med. Biol.*, vol. 67, no. 1, p. 015007, Jan. 2022.

[9] 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 (cute)," *Preprint*, Sept. 2022.

[10] J. Ophir, "Estimation of the speed of ultrasound propagation in biological tissues: A beam-tracking method," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 33, no. 4, pp. 359–368, Jul. 1986.

[11] J. Ophir and Y. Yazdi, "A transaxial compression technique (TACT) for localized pulse-echo estimation of sound speed in biological tissues," *Ultrason. Imaging*, vol. 12, no. 1, pp. 35–46, Jan. 1990.

[12] D. E. Robinson, F. Chen, and L. S. Wilson, "Measurement of velocity of propagation from ultrasonic pulse-echo data," *Ultrasound Med. Biol.*, vol. 8, no. 4, pp. 413–420, Jan. 1982.

[13] D. E. Robinson, C. F. Chen, and L. S. Wilson, "Image matching for pulse echo measurement of ultrasonic velocity," *Image Vis. Comput.*, vol. 1, no. 3, pp. 145–151, Aug. 1983.

[14] M. E. Anderson and G. E. Trahey, "The direct estimation of sound speed using pulse–echo ultrasound," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3099–3106, Nov. 1998.

[15] M. Imbault *et al.*, "Robust sound speed estimation for ultrasound-based hepatic steatosis assessment," *Phys. Med. Biol.*, vol. 62, no. 9, pp. 3582–3598, Apr. 2017.

[16] R. Mallart and M. Fink, "Adaptive focusing in scattering media through sound-speed inhomogeneities: The van Cittert Zernike approach and focusing criterion," *J. Acoust. Soc. Am.*, vol. 96, no. 6, pp. 3721–3732, Dec. 1994.

[17] J.-L. Robert and M. Fink, "Green's function estimation in speckle using the decomposition of the time reversal operator: Application to aberration correction in medical imaging," *J. Acoust. Soc. Am.*, vol. 123, no. 2, pp. 866–877, Feb. 2008.

[18] G. Montaldo, M. Tanter, and M. Fink, "Time reversal of speckle noise," *Phys. Rev. Lett.*, vol. 106, no. 5, p. 054301, Feb. 2011.

[19] W. Lambert, L. A. Cobus, M. Couade, M. Fink, and A. Aubry, "Reflection matrix approach for quantitative imaging of scattering media," *Phys. Rev. X*, vol. 10, no. 2, p. 021048, Jun. 2020.

[20] W. Lambert, L. A. Cobus, T. Frappart, M. Fink, and A. Aubry, "Distortion matrix approach for ultrasound imaging of random scattering media," *Proc. Natl. Acad. Sci.*, vol. 117, no. 26, pp. 14 645–14 656, Jun. 2020.

[21] J. F. Krucker, J. B. Fowlkes, and P. L. Carson, "Sound speed estimation using automatic ultrasound image registration," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 51, no. 9, pp. 1095–1106, Sep. 2004.

[22] M. Jaeger, G. Held, S. Peeters, S. Preisser, M. Grünig, and M. Frenz, "Computed ultrasound tomography in echo mode for imaging speed of sound sing pulse-echo sonography: Proof of principle," *Ultrasound Med. Biol.*, vol. 41, no. 1, pp. 235–250, Jan. 2015.

[23] 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.

[24] 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, p. 215015, Oct. 2018.

[25] P. Stähli, M. Kuriakose, M. Frenz, and M. Jaeger, "Improved forward model for quantitative pulse-echo speed-of-sound imaging," *Ultrasonics*, vol. 108, p. 106168, Dec. 2020.

[26] P. Stähli, M. Frenz, and M. Jaeger, "Bayesian approach for a robust speed-of-sound reconstruction using pulse-echo ultrasound," *IEEE Trans. Med. Imaging*, vol. 40, no. 2, pp. 457–467, Feb. 2021.

[27] 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 *2021 IEEE Int. Ultrason. Symp.*, Xi'an, China, Sep. 2021, paper 4901, pp. 1–4.

[28] M. Jakovljevic, S. Hsieh, Rehman, G. Chau Loo Kung, D. Hyun, and J. J. Dahl, "Local speed of sound estimation in tissue using pulse-echo ultrasound: Model-based approach," *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. 254–266, Jul. 2018.

[29] R. Ali, T. Brevett, D. Hyun, L. L. Brickson, and J. J. Dahl, "Distributed aberration correction techniques based on tomographic sound speed estimates," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 69, no. 5, pp. 1714–1726, May 2022.

[30] H. Bendjador, "Correction d'aberrations et quantification de vitesse du son en imagerie ultrasonore ultrarapide," Ph.D. dissertation, Dept. of Sci. Mec., Ac., Elec. and Rob., Univ. Paris Sci. et Lettres, Paris, France, 2020.

[31] R. Rau, D. Schweizer, V. Vishnevskiy, and O. Goksel, "Speed-of-sound imaging using diverging waves," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 7, pp. 1201–1211, Jul. 2021.

[32] M. Jaeger *et al.*, "Pulse-echo speed-of-sound imaging using convex probes," *Phys. Med. Biol.*, vol. 67, no. 21, p. 215016, Oct. 2022.

[33] D. D'Souza and D. Rakhmatov, "Stratified-medium sound speed profiling for CPWC ultrasound imaging," in *2020 IEEE Int. Ultrason. Symp.*, Las Vegas, NV, USA, Sep. 2020, paper 1173, pp. 1–4.

[34] R. Ali and J. J. Dahl, "Distributed phase aberration correction techniques based on local sound speed estimates," in *2018 IEEE Int. Ultrason. Symp.*, Kobe, Japan, Oct. 2018, paper 3D-4, pp. 1–4.

[35] 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 *2020 IEEE Int. Ultrason. Symp.*, Las Vegas, NV, USA, Sep. 2020, paper 1225, pp. 1–4.

[36] R. Ali *et al.*, "Sound speed estimation in layered media using the angular coherence of plane waves," in *Med. Imaging 2020 Ultrason. Imaging Tomogr.*, vol. 11319, Houston, TX, USA, Mar. 2020, paper 13/0F, pp. 1–10.

[37] 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.

[38] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 56, no. 3, pp. 489–506, Mar. 2009.

[39] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 39, no. 2, pp. 262–267, Mar. 1992.

[40] 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, p. 021314, Apr. 2010.

[41] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," *IEEE Trans. Sonics Ultrason.*, vol. 30, no. 3, pp. 156–163, May 1983.

Adrien Besson (Member, IEEE) was born in Grenoble, France, in 1989. He received the Engineering degree from Supélec, Paris, France in 2013, and a M.Sc. in electrical engineering from École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2013. After working for two years as a research and development engineer, he returned to EPFL to receive his Ph.D. degree with Signal Processing Laboratory 5 (LTS5) in 2019.

Since 2019, he has been working as an Ultrasound Engineer at E-Scopics where he has developed an innovative platform for ultrasound imaging and quantitative modalities. His research interests include the application of compressed sensing techniques and inverse problems to ultrasound imaging.

Frédéric Wintzenrieth (Member, IEEE) was born in Annecy, France, in 1987. He received an Engineering degree from École Polytechnique, Paris, France, in 2010 and a M.Phil. degree in chemical engineering from University of Cambridge, Cambridge, United Kingdom in 2011. He received his Ph.D. degree from the Université Pierre et Marie Curie, Paris, France, in 2015 for his work on light and sound scattering in foams.

He worked as an Ultrasound Engineer for Supersonic Imaging (now Hologic) until 2019 and for E-Scopics until 2022. He is now Principal Ultrasound and Signal Processing Engineer for Austral Diagnostics, where he tracks heartbeats and lung vibrations using airborne ultrasound.

Baptiste Hériard-Dubreuil (Student Member, IEEE) was born in St-Cyr-l'Ecole, France, in 1996. He received an Engineering degree from École Polytechnique, Paris, France, in 2017 and a M.Sc. degree in electrical engineering from the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2019. He is currently pursuing a Ph.D. degree with the Signal Processing Laboratory 5 (LTS5), EPFL, under the supervision of Prof. Jean-Philippe Thiran, and with E-Scopics, Aix-en-Provence, France.

His current research interests are focused on angular approaches and few plane-waves ultrasound imaging for different applications such as speed of sound estimation.

Claude Cohen-Bacrie was born in 1969. He received an Engineering degree in signal and image processing from École Nationale Supérieure de l'Électronique et de ses Applications (ENSEA), France in 1992 and a M.Sc. degree in biomedical sciences from Polytechnique Montreal in 1994. After managing ultrasound research within Philips Research, he co-founded SuperSonic Imagine (now Hologic) in 2005, and acted as executive vice president, Chief Technology Officer and Chief Operating Officer.

In 2018, he founded E-Scopics, an innovative ultrasound company, which aims at democratizing ultrasound tools to healthcare professionals. The company owns and develops proprietary technologies to allow the dematerialization of ultrasound systems. Using the flexible platform developed by this company, he supports and contributes to projects with fundamental and clinical research and laboratories to explore new applications of medical ultrasound.

Jean-Philippe Thiran (Senior Member, IEEE) was born in Namur, Belgium, in 1970. He received his M.Sc. degree in electrical engineering and his Ph.D. degree from the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1993 and 1997, respectively. He joined the École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 1998. He is currently a Full Professor at EPFL and Director of the Signal Processing Laboratory (LTS5). Since June 2021, he has been the Director of the Institute of Electrical and Micro Engineering, EPFL.

His research field is computational imaging, with applications in many domains, including medical image analysis (diffusion magnetic resonance imaging (MRI), ultrasound imaging, and digital pathology) and computer vision. He is currently a part-time Associate Professor with the Department of Radiology, University Hospital Center (CHUV) and University of Lausanne (UNIL), Lausanne. He is the author or the coauthor of one book, nine book chapters, 270 journal articles, and more than 290 peer-reviewed papers published in the proceedings of international conferences. He holds 12 international patents.

Prof. Thiran is a fellow of the European Association for Signal Processing (EURASIP). Among many other duties, he has been the General Chairperson of the 2008 European Signal Processing Conference (EUSIPCO 2008) and the Technical Co-Chair of the 2015 IEEE International Conference on Image Processing (IEEE ICIP 2015). From 2001 to 2005, he was Co-Editor-in-Chief of the Signal Processing international journal (published by Elsevier Science). He has been an Associate Editor of the IEEE Transactions on Image Processing.

Refraction-Based Speed of Sound Estimation in Layered Media: an Angular Approach - Supplementary Materials

S-I. Taylor expansions

To solve implicit equations that involve angles and trigonometric formulas, we suppose that all considered angles are small.

In practice, we approximate quantities asymptotically around 0 using Taylor expansions and more specifically the small-o notation of the Bachmann-Landau family. For two real-valued functions f and g, $f = o(g)$ means that f is asymptotically dominated by g around 0, i.e. $\lim_{x \to 0} f(x)/g(x) = 0$. In our case, we want the variable x to represent α_{th}, β_{th}, α'_{th} and β'_{th} at the same time, as all these quantities are supposed small. Mathematically, we choose $x = \max(\alpha_{th}, \beta_{th}, \alpha'_{th}, \beta'_{th})$.

The refraction analysis that led to Equation (6) showed that for any set of superficial horizontal layers, in a layer of real speed of sound c_r, the real transmitted angle α_r is equal to:

$$\alpha_r = \arcsin\left(\frac{c_r}{c_{th}} \sin \alpha_{th}\right). \tag{S-1}$$

With the small-o notation, we have:

$$\alpha_r = \arcsin\left(\frac{c_r}{c_{th}} \alpha_{th} + o(x^2)\right) \tag{S-2}$$

$$= \frac{c_r}{c_{th}} \alpha_{th} + o(x^2). \tag{S-3}$$

Due to the symmetric transmit-receive strategy, this analysis can be applied equivalently to all the considered angles.

Using the small-o properties on composition, sum and differences, we can compute the asymptotic expression of a few angular quantities summarized in Table S-I hereafter.

TABLE S-I
ASYMPTOTIC EXPRESSION OF ANGULAR QUANTITIES

Quantity	Exact Expression	Asymptotic Expression
α_r	$\arcsin\left(\frac{c_r}{c_{th}} \sin \alpha_{th}\right)$	$\frac{c_r}{c_{th}} \alpha_{th} + o(x^2)$
α'_r	$\arcsin\left(\frac{c_r}{c_{th}} \sin \alpha'_{th}\right)$	$\frac{c_r}{c_{th}} \alpha'_{th} + o(x^2)$
β_r	$\arcsin\left(\frac{c_r}{c_{th}} \sin \beta_{th}\right)$	$\frac{c_r}{c_{th}} \beta_{th} + o(x^2)$
β'_r	$\arcsin\left(\frac{c_r}{c_{th}} \sin \beta'_{th}\right)$	$\frac{c_r}{c_{th}} \beta'_{th} + o(x^2)$
γ_r	$\frac{\alpha_r + \beta_r}{2}$	$\frac{c_r}{c_{th}} \gamma_{th} + o(x^2)$
γ'_r	$\frac{\alpha'_r + \beta'_r}{2}$	$\frac{c_r}{c_{th}} \gamma'_{th} + o(x^2)$
δ_r	$\frac{\alpha_r - \beta_r}{2}$	$\frac{c_r}{c_{th}} \delta_{th} + o(x^2)$
δ'_r	$\frac{\alpha'_r - \beta'_r}{2}$	$\frac{c_r}{c_{th}} \delta'_{th} + o(x^2)$

As the cosine function can be written asymptotically as $\cos u = 1 - u^2/2 + o(u^2)$, we obtain by composition the asymptotic expression of more complex trigonometric functions summarized in Table S-II.

TABLE S-II
ASYMPTOTIC EXPRESSION OF TRIGONOMETRIC QUANTITIES

Quantity	Asymptotic Expression
$\cos \delta_{th}$	$1 - \frac{\delta_{th}^2}{2} + o(x^2)$
$\cos \delta'_{th}$	$1 - \frac{\delta'^2_{th}}{2} + o(x^2)$
$\cos \delta_r$	$1 - \frac{c_r^2}{c_{th}^2} \frac{\delta_{th}^2}{2} + o(x^2)$
$\cos \delta'_r$	$1 - \frac{c_r^2}{c_{th}^2} \frac{\delta'^2_{th}}{2} + o(x^2)$

We can now estimate the asymptotic expressions of scaling factors SF and RSF, which are ratios of the trigonometric quantities studied in Table S-II.

A. Scaling factor SF

In the case of the scaling factor SF, we have:

$$SF = \frac{\cos \delta_r}{\cos \delta'_r} = \frac{1 - \frac{c_r^2}{c_{th}^2} \frac{\delta_{th}^2}{2} + o(x^2)}{1 - \frac{c_r^2}{c_{th}^2} \frac{\delta'^2_{th}}{2} + o(x^2)}. \tag{S-4}$$

We know that the inverse function can be written as $1/(1 - u) = 1 + u + u^2 + o(u^2)$. By composition (as $\delta'_{th} \to 0$ when $x \to 0$), and using the fact that $\delta'_{th}{}^n = o(x^2)$ for $n > 2$:

$$SF = \left(1 - \frac{c_r^2}{c_{th}^2} \frac{\delta_{th}^2}{2} + o(x^2)\right)\left(1 + \frac{c_r^2}{c_{th}^2} \frac{\delta'^2_{th}}{2} + o(x^2)\right) \tag{S-5}$$

$$= 1 + \frac{c_r^2}{c_{th}^2} \frac{\delta'^2_{th}}{2} - \frac{c_r^2}{c_{th}^2} \frac{\delta_{th}^2}{2} + o(x^2) \tag{S-6}$$

$$= 1 - \frac{c_r^2}{c_{th}^2} \frac{\delta_{th}^2 - \delta'^2_{th}}{2} + o(x^2). \tag{S-7}$$

From above Equation (S-7) we deduce Equation (12) with a second order approximation (where we remove the $o(x^2)$ part).

B. Residual scaling factor RSF

By differentiating (A.12), we obtain the following equation of RSF:

$$RSF = \omega_c \frac{\cos \delta_r + \cos \delta'_r}{\cos^2 \delta_r + \cos^2 \delta'_r} \cdot \left(\cos \delta_{th} \cos \delta'_r - \cos \delta'_{th} \cos \delta_r \right). \tag{S-8}$$

By product and sum of asymptotic expressions of Table S-II, we deduce the asymptotic expressions of more complex terms summarized in Table S-III.

TABLE S-III
ASYMPTOTIC EXPRESSION OF RSF TERMS

Term	Asymptotic Expression
$\frac{\cos \delta_r + \cos \delta'_r}{2}$	$1 - \frac{1}{4} \frac{c_r^2}{c_{th}^2} \left(\delta_{th}^2 + \delta_{th}'^2 \right) + o(x^2)$
$\frac{\cos^2 \delta_r + \cos^2 \delta'_r}{2}$	$1 - \frac{1}{2} \frac{c_r^2}{c_{th}^2} \left(\delta_{th}^2 + \delta_{th}'^2 \right) + o(x^2)$
$\cos \delta_{th} \cos \delta'_r - \cos \delta'_{th} \cos \delta_r$	$\frac{1}{2} \left(\delta_{th}'^2 - \delta_{th}^2 \right) \left(1 - \frac{c_r^2}{c_{th}^2} \right) + o(x^2)$

To compute RSF, we need the ratio of the first two lines of Table S-III. By composition with $1/(1-u) = 1+u+u^2+o(u^2)$ and by product, we obtain:

$$\frac{\cos \delta_r + \cos \delta'_r}{\cos^2 \delta_r + \cos^2 \delta'_r} = 1 + \frac{1}{4} \frac{c_r^2}{c_{th}^2} \left(\delta_{th}^2 + \delta_{th}'^2 \right) + o(x^2) \tag{S-9}$$

The residual scaling factor RSF is obtained by multiplication of ω_c with (S-9) and the last line of Table S-III. We obtain:

$$RSF = \frac{\omega_c}{2} \left(\delta_{th}'^2 - \delta_{th}^2 \right) \left(1 - \frac{c_r^2}{c_{th}^2} \right) + o(x^2) \tag{S-10}$$

C. Discussion

The approximation of scaling factors is a necessary step to obtain an explicit expression of c_r. However, it may lead to small errors to the speed of sound estimation. It is to be noted that the asymptotic analysis only guarantees a convergence property. Mathematically, using the definition of convergence, it means that:

$$\forall \epsilon \in \mathbb{R}_+, \; \exists x_0 \text{ s.t.} \forall |x| \le x_0, \quad o(x^2) \le \epsilon x^2, \tag{S-11}$$

i.e. the error is bounded by a factor times x^2 if all angles are sufficiently small. In practice, we need to verify that such an error is sufficiently small.

To quantify the error in our configuration, we simulate, for a range of speed of sound values c_r, the real residual scaling factor using (A.12), and estimate from this RSF the speed of sound using the approximate formula (A.13). The values of angles are chosen according to the framework described in II-C to assess the approximation error of the final method. We obtain the approximation error displayed in Fig. S-1.

We observe that the small angle approximation leads to errors in the estimation of $\pm 6 \, \text{m s}^{-1}$ in the worst cases, i.e. a relative error lower than 0.5%. This effect can be considered as small with respect to the variability of the proposed method in real conditions. Note that, as it is deterministic and monotonous, it could be compensated by applying as a correction the opposite of the error measured in Fig. S-1.
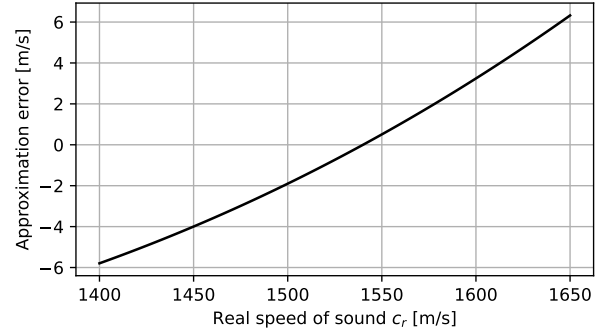
Fig. S-1. Approximation error as a function of real speed of sound c_r.

S-II. NON-HORIZONTAL LAYERS

In the case of non-horizontal layers, equation (6) obtained recursively is no longer valid. However, such a configuration can still be studied theoretically. We specifically focus on the configuration displayed Fig. S-2, where only two layers are considered with an interface of angle η with respect to the horizontal.
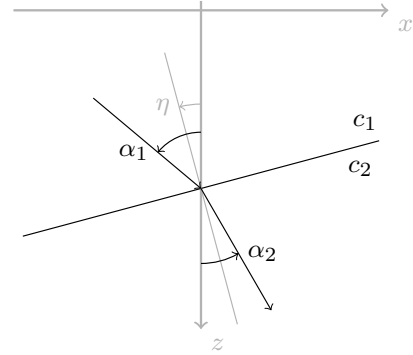
Fig. S-2. Two-layer configuration with a non horizontal interface.

The two layers are considered uniform and their speed of sounds are noted c_1 and c_2. We consider an incident wave steered with an angle α_1 with respect to the z-axis. Snell's laws of refraction give:

$$\frac{\sin (\alpha_2 - \eta)}{c_2} = \frac{\sin (\alpha_1 - \eta)}{c_1}. \tag{S-12}$$

We obtain:

$$\alpha_2 = \eta + \arcsin \left(\sin (\alpha_1 - \eta) \frac{c_2}{c_1} \right). \tag{S-13}$$

We observe that refraction effects are impacted by the interface angle. The same effect happens in reception, and considering both the emission and the reception, we can compute the mid-angle and the half-angle difference:

$$\begin{cases} \gamma_2 = \eta + \frac{\arcsin\left(\sin(\alpha_1-\eta)\frac{c_2}{c_1}\right) + \arcsin\left(\sin(\beta_1-\eta)\frac{c_2}{c_1}\right)}{2}, \\ \delta_2 = \frac{\arcsin\left(\sin(\alpha_1-\eta)\frac{c_2}{c_1}\right) - \arcsin\left(\sin(\beta_1-\eta)\frac{c_2}{c_1}\right)}{2}. \end{cases} \tag{S-14}$$

Using a second order approximation in a similar way to Supplementary-S-I, Equation (S-14) can be simplified into:

$$\begin{cases} \gamma_2 \approx \eta \left(1 - \frac{c_2}{c_1} \right) + \frac{c_2}{c_1} \gamma_1, \\ \delta_2 \approx \frac{c_2}{c_1} \delta_1. \end{cases} \quad \text{(S-15)}$$

As γ_2 only depends on γ_1 and not in α_1 or β_1 independently, we can still compare two pairs that share the same mid-angle. The half angle difference behaves approximately as if the interface was horizontal.

However, for some values of α, β and η, the small angle approximation is no longer valid for $\alpha - \eta$ and $\beta - \eta$. This would result in a decrease of the observed coherence among pairs (pairs will not share perfectly equal mid-angles) and a bias in the final estimation of c_2 (due to the half angle difference being affected by η). As for Supplementary-S-I, we can simulate the bias brought by η against different speed of sound discontinuities in the configuration considered in II-C. Such biases are displayed Fig. S-3.
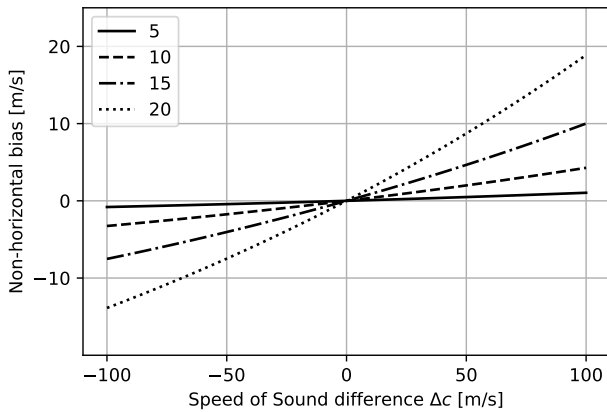
Fig. S-3. Bias caused by the non-horizontal-ness of the interface as a function of speed of sound difference $\mathbf{\Delta c}$ at the interface, for different values of the interface angle ranging from 5 to 20 degrees. Due to the symmetry of the problem, biases only depend on the absolute value of the interface angle so negative angles are not represented.

We can see that strong interfaces (with high speed of sound differences) bring biases to the estimation if they are not horizontal. For interfaces with angles of less than 15° with respect to the horizontal, such biases are lower than $10 \, \text{m s}^{-1}$ and are not significant relatively to the variability of the proposed method. Nevertheless, interfaces of higher angles could lead to a significant error in the speed of sound estimation. This is a limit of the proposed method. Nevertheless, the most biased pairs would theoretically appear as the less coherent ones and could thus be discarded.

The remaining issue is to understand how non-horizontal refraction accumulates in multilayer media. Multiple effects can arise, but we can point out the specific situation described Fig. S-4, that is likely to be seen *in vivo* (e.g. vessels or thin muscle layer), and where the angles are unaffected by a non-horizontal layer as long as it is bounded by two parallel interfaces.
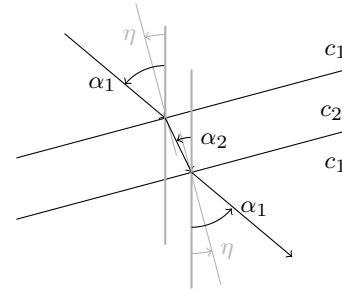
Fig. S-4. Specific configuration of a non-horizontal layer bounded by parallel interfaces, in a bigger homogeneous layer. In this case, the incident wave angle is completely unaffected by the intermediate layer.