

Robust Sonographic Muscle Quality Assessment: A Live, Accurate Tissue Speed-of-Sound Estimation
Framework

by

Di Xiao

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2024

© Di Xiao 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	ORCUN GOKSEL Associate Professor, Department of Information Technology, Uppsala University
Supervisor	ALFRED C. H. YU Professor, Department of Electrical and Computer Engineering, University of Waterloo
Internal Member	ZHOU WANG Professor, Department of Electrical and Computer Engineering, University of Waterloo
Internal Member	PATRICK MITRAN Professor, Department of Electrical and Computer Engineering, University of Waterloo
Internal-External Member	MARINA MOURTZAKIS Professor, Department of Kinesiology and Health Sciences, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The author contributions for each Chapter are described below.

Chapters 1 and 6: Di Xiao wrote these chapters.

A version of **Chapter 2**, titled “Real-time Speed-of-Sound Estimation *in vivo* via Steered Plane Wave Ultrasound” is being prepared for submission to IEEE Transactions on Biomedical Engineering, co-authored by Di Xiao, Pat de la Torre, and Alfred C. H. Yu. Di Xiao carried out the algorithmic design, implemented the algorithm for real-time operation on the research scanner, performed data analysis, and wrote the manuscript draft. Di Xiao and Pat de la Torre constructed the agar phantoms for validation. Pat de la Torre designed the mechanical components for the through-transmit device and Di Xiao designed the software for the through-transmit device. Di Xiao and Pat de la Torre provided algorithmic input. Di Xiao and Pat de la Torre collected the data for all *in vitro* and *in vivo* acquisitions. All authors will review the results and manuscript prior to submission.

A version of **Chapter 3**, titled “Sparse Matrix Beamforming: A New Look for Delay-and-Sum in Ultrasound Image Formation” is being prepared for submission to Ultrasonics, authored by Di Xiao and Alfred C. H. Yu. Di Xiao carried out the algorithmic design, programmed and provided the code samples, performed all experiments, and wrote the manuscript draft. All authors will review the results and manuscript prior to submission.

Chapter 4 comprises material from an article entitled “Minimizing Image Quality Loss After Channel Count Reduction for Plane Wave Ultrasound via Deep Learning Inference” which has been published in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. This article was co-authored by Di Xiao, William M. K. Pitman, Billy Y. S. Yiu, Adrian J. Y. Chee, and Alfred C. H. Yu. Di Xiao carried out the design of the neural network architecture, training and evaluation of the framework, analyzed the results, and wrote the manuscript draft. Billy Y. S. Yiu and Adrian J. Y. Chee supplied the data for training and validation and provided feedback on image quality metrics. Billy Y. S. Yiu and William Pitman provided input in editing the manuscript. Alfred Yu reviewed the results and edited the manuscript prior to publication.

Versions of **Chapter 5**, titled “A New Protocol for Speed-of-Sound Estimation in Large Muscles: Results from an *in vivo* Study” is being prepared for submission to Ultrasound in Medicine & Biology. This article is co-authored by Di Xiao, Pat de la Torre, Malak Saif El Nasr, Adrian J. Y. Chee, Marina

Mourtzakis, and Alfred C. H. Yu. Di Xiao designed the study, programmed the ultrasound scanner for all acquisitions, performed the data processing and statistical analysis, and wrote the manuscript drafts. Di Xiao, Pat de la Torre, Malak Saif El Nasr recruited and performed data collection on all participants. Adrian J. Y. Chee and Marina Mourtzakis provided input on the study design. All authors will review the results and manuscript prior to submission.

Abstract

Muscle quality can act as an indicator for physical health through qualitative – yet measurable – changes in muscle architecture and composition. One option for assessing muscle quality is through musculoskeletal ultrasound, which can provide metrics such as echogenicity or texture. These existing ultrasound-based metrics relating to muscle quality may depend on scanner capabilities and operator assessment of the resulting B-mode images. Tissue speed-of-sound (SoS) can robustly augment muscle quality assessment as an intrinsic property of the tissue; however, sonographic measurement of muscle SoS is frequently limited by hardware which requires bilateral access to the tissue, or which cannot perform simultaneous ultrasound imaging in real-time.

In this dissertation, I designed an ultrasound framework for live, accurate SoS estimation *in vivo*. A novel global SoS estimation algorithm with real-time potential was created using the principles of high-frame-rate ultrasound with a standard pulse-echo ultrasound probe. This algorithm was experimentally validated to be highly accurate both *in vitro* and *in vivo* with demonstrable live imaging capacity; a portable research scanner and laptop were used to realize the framework. In the process of framework design, the underlying engineering challenges of real-time data transfer rates from the probe to the system and image formation efficiency were addressed using deep learning principles and sparse matrix formulations respectively. Lastly, the novel framework was then used to conduct a human study consisting of forty volunteers. The study served to demonstrate the applicability of a live SoS system, establish a replicable experimental protocol for SoS measurement in large muscles, and investigate relationships between demographics and muscle SoS.

This dissertation research is intended to bridge the engineering innovations of ultrasound algorithm development with the clinical applications for SoS as a tissue biomarker. For the targeted muscle quality application, the results confirm the physiological relevance of muscle SoS and exhibit the utility of real time SoS estimation with simultaneous B-mode imaging. The significant relationships between muscle SoS and demographic factors support the potential for clinical translation of such a real-time system. The realization of live SoS estimation can help derive new insights between the tissue SoS and pathological conditions or imaging applications.

Acknowledgements

I would firstly like to thank Prof. Alfred Yu for the opportunity to pursue a PhD under his supervision at the University of Waterloo. I am very grateful for both his professional and personal support during these years; under his mentorship, I have grown in leaps and bounds as a researcher. He constantly provided me with opportunities to expand my capacity in research while allowing me generous space for independent thought and investigation.

I want to express my appreciation to Prof. Marina Mourtzakis for her encouragement and support – her excitement towards my research and advice for conducting human studies have allowed me to broaden the scope and horizons of this research. I would also like to thank Prof. Orcun Goksel, Prof. Zhou Wang, and Prof. Patrick Mitran for their service as members of my examination committee.

I am fortunate to have had a series of fantastic research supervisors (Alfred continues the trend); the ideas and concepts I learned under their tutelage have proved invaluable during this PhD. I am grateful to Prof. Jarvis Haupt at the University of Minnesota for his guidance; his perspectives on signal processing are thematically evident in this PhD thesis. I also am grateful to Prof. Steven Conolly and Dr. Bo Zheng at UC Berkeley for providing my first opportunity to do research in medical imaging.

I am thankful for all the current and former members of the LITMUS lab at the University of Waterloo for their camaraderie and intellectual stimulation. Specifically, I would like to acknowledge Dr. Billy Yiu and Dr. Adrian Chee for their availability, feedback, and patience. I would also like to acknowledge my colleagues William Pitman and Pat de la Torre, both of whom I initially mentored, but who very quickly became equal collaborators in research instead. I would especially like to thank Pat for believing in speed-of-sound research from the start and for participating in endless discussions both topical and irrelevant.

My heartfelt gratitude goes to my family – first and foremost my incredible wife Jenny Wang. She has been my best friend and truly a partner in this adventure (along with our dog Misha). Because of her love and kindness, I have succeeded far more than I could have dreamt of alone. I owe an incredible amount to my parents Renfu Shiao and Ying Du and my parents-in-law Hanming Wang and Yongjuan He for their ceaseless support, encouragement, and prayers. I also want to mention my lifelong friends Eric Chen and Patrick Liao who ensured that I maintained a healthy balance between work and life during these years.

Table of Contents

Examining Committee Membership.....	ii
Author’s Declaration	iii
Statement of Contributions.....	iv
Abstract	vi
Acknowledgements	vii
List of Figures	xv
List of Abbreviations.....	xxii
List of Symbols (Order of Appearance)	xxiv
Chapter 1 Introduction.....	1
1.1 Musculoskeletal Ultrasound and Tissue Speed-of-Sound	1
1.2 A Real-Time Framework for Global Speed-of-Sound Estimation	2
1.2.1 Thesis Objectives: Real-Time Imaging Framework.....	2
1.2.2 Hypothesized Solution: Plane Wave Ultrasound Imaging	2
1.3 Thesis Outline.....	3
1.4 Background Information	4
1.4.1 Medical Ultrasound Imaging.....	4
1.4.2 Speed-of-Sound of Human Tissues	5
1.4.3 What is High-Frame-Rate Ultrasound?	8
1.4.4 Ultrasound Image Formation – Delay-and-Sum Beamforming	9
1.5 Summary of Introduction	12
Chapter 2 Real-time Speed-of-sound Estimation <i>in vivo</i> via Steered Plane Wave Ultrasound.....	13
2.1 Introduction	13
2.2 Theory	14

2.2.1 Effect of Speed-of-Sound Mismatch on Point Targets for Steered Plane Wave Transmissions	15
2.2.2 Beamforming Images for Speed-of-Sound Estimation.....	16
2.2.3 Estimating the Speed-of-sound through Loss Minimization	19
2.2.4 Full Overview of SoS Estimation Framework	19
2.3 Experimental Methods.....	20
2.3.1 Generating Accurate Reference Speed-of-Sound Measurements with Water Validation ...	20
2.3.2 Construction of In Vitro Agar Phantoms.....	22
2.3.3 <i>In vitro</i> and <i>in vivo</i> Measurement and Data Collection	22
2.3.4 Beamforming Specifics	24
2.3.5 Speed-of-Sound Estimation.....	24
2.3.6 Real-Time Implementation using Sparse Matrices and GPU Acceleration on Portable Scanner	27
2.4 Results	29
2.4.1 High Accuracy of Speed-of-Sound Estimates Compared to Reference Measurements	29
2.4.2 Refraction Correction and Depth Alignment are Important for Accurate Estimates	32
2.4.3 Framework is Capable of Real-Time SoS Estimation on Portable Open Platform Ultrasound Scanner	32
2.5 Discussion	34
2.5.1 A New Algorithm and Device for Real-Time Speed-of-Sound Estimation	34
2.5.2 Importance of Live Speed-of-Sound Estimation Framework.....	35
2.5.3 Limitations and Future Work	36
2.6 Conclusion.....	37
2.7 Acknowledgements	37

Chapter 3 Sparse Matrix Beamforming: A New Look for Delay-and-Sum in Ultrasound Image

Formation	38
3.1 Introduction	38
3.2 Theory	40
3.2.1 Delay-and-Sum Beamforming as a Sparse Matrix Multiplication	40
3.2.2 Extending Sparse Matrix Beamforming for Batch Processing and Coherent Compounding	44
3.2.3 Efficient Sparse Matrix Multiplication using Compressed Sparse Row Format.....	47
3.3 Example Applications of Sparse Matrix Beamforming	47
3.3.1 Data Acquisition and System Parameters for All Example Applications	47
3.3.2 Single Plane Wave Delay-and-Sum Beamforming via Sparse Matrices.....	48
3.3.3 Batch Processing to Reduce Overhead for Sparse Matrix Beamforming.....	53
3.3.4 Coherent Compounding with Single Sparse Matrix for Plane Wave Imaging.....	55
3.3.5 Beamforming-Integrated Neural Networks through Sparse Matrix Multiplication	57
3.4 Discussion	60
3.4.1 Sparse Matrix Beamforming as an Efficient and Accessible Delay-and-Sum Beamformer	60
3.4.2 Integration for Software Frameworks via Sparse Matrix Beamforming	61
3.5 Conclusion.....	62
3.6 Acknowledgment.....	62
Chapter 4 Minimizing Image Quality Loss after Channel Count Reduction for Plane Wave Ultrasound via Deep Learning Inference	63
4.1 Introduction	63
4.2 Theory	65
4.2.1 Background Considerations.....	65
4.2.2 Overview of the Proposed Framework.....	66

4.2.3 Architecture of the Custom Encoder–Decoder Neural Network for RF Data Recovery.....	67
4.2.4 RF Data as an Image for Encoder–Decoder Input–Output.....	68
4.3 Custom Encoder-Decoder Implementation	68
4.3.1 Number of Layers, Kernels per Layer, and Kernel Sizes	68
4.3.2 Convolutional Strides	69
4.3.3 Skip Connections.....	69
4.3.4 Activation Function.....	69
4.4 Experimental Methods.....	70
4.4.1 Acquisition of Neural Network Training Data	70
4.4.2 Data Preprocessing	70
4.4.3 Neural Network Training and Evaluation	71
4.4.4 Image Quality Evaluation of the Proposed Framework	72
4.4.5 Evaluation of Image Quality After Coherent Compounding.....	73
4.4.6 Evaluation of Computational Capacity.....	73
4.5 Results	73
4.5.1 Trained CNN Encoder–Decoder Can Recover the Unacquired Channel Structure	73
4.5.2 Grating Lobe Reduction and Preservation of Resolution in Multiple Point Target Phantom Across Transmit Angles	75
4.5.3 Recovery of Carotid In Vivo Lumen-to-Tissue Contrast	78
4.5.4 Consistent Image Quality Improvement After Coherent Compounding of Single Transmit Images	79
4.5.5 Real-Time Computational Capacity	81
4.6 Discussion	81
4.6.1 Convolutional Encoder–Decoder as a New Framework for Channel Data Recovery.....	81

4.6.2 Impact of a Pure RF-to-RF Framework.....	83
4.6.3 Limitations and Perspectives on Future Work	84
4.7 Conclusion.....	85
Chapter 5 A New Protocol for Speed-of-Sound Estimation for in Large Muscles: Results From an <i>in vivo</i> Study	86
5.1 Introduction	86
5.2 Experimental Methods.....	88
5.2.1 Speed-of-Sound Estimation Framework	88
5.2.2 Imaging Hardware and Parameters.....	89
5.2.3 Recruitment of Sample and Survey Details.....	90
5.2.4 Measurement of Large Muscles (Calf, Biceps, Quadriceps) via Ultrasound	91
5.2.5 Evaluation of New Speed-of-Sound Estimation Protocol for Musculoskeletal Imaging	93
5.2.6 Statistical Analysis of Physiological Relationships for Muscle Speed-of-Sound	94
5.3 Results – Establishment of Imaging Protocol.....	95
5.3.1 Highest Inter-Sample Reliability Views of the Biceps, Calf, and Quadriceps	95
5.3.2 Highest Intra-Sample Reliability Views of the Biceps, Calf, and Quadriceps	95
5.3.3 Full Contact of Probe with Skin is Important for Estimating the Muscle Speed-of-Sound	96
5.3.4 Minimal Differences Between Dominant and Non-Dominant Appendages	97
5.3.5 Summary of Ideal Imaging Views for Speed-of-Sound Estimation.....	98
5.4 Results – Statistical Analysis of Physiological Relationships.....	98
5.4.1 High Correlation Between Body Mass Index and Muscle Speed-of-Sound for All Measured Muscles.....	98
5.4.2 Male Muscles Statistically Faster Speed-of-Sound Compared to Female Muscles	99
5.4.3 No Significant Change of Speed-of-Sound with Muscle Activation.....	100

5.4.4 For Males Only, Significant Correlation was Found Between Physical Activity and Muscle Speed-of-Sound	100
5.4.5 Estimation Accuracy of Speed-of-Sound Framework	101
5.5 Discussion	101
5.5.1 Using Real-Time Speed-of-Sound Measurements to Investigate Physiological Relationships	101
5.5.2 A New Measurement Protocol for Speed-of-Sound Characterization in Large Human Muscles.....	103
5.5.3 Potential of Muscle Speed-of-Sound to Act as a Robust Quantitative Tissue Biomarker.	103
5.6 Conclusion.....	104
5.7 Acknowledgements	104
Chapter 6 Summary and Future Directions	105
6.1 Summary of Contributions	105
6.1.1 Realization of Real-Time, Global Speed-of-Sound Estimator with Simultaneous B-Mode Imaging Capacity.....	107
6.1.2 Novel Solutions to Sub-Challenges for Real-Time High-Frame-Rate Imaging Methods.	108
6.1.3 Demonstration of Speed-of-Sound Estimation Framework <i>in vivo</i> for Assessing Muscle Quality	109
6.2 Limitations and Future Directions.....	110
6.2.1 From Global Speed-of-Sound Estimation to Local Speed-of-Sound Mapping.....	110
6.2.2 Further Investigation of Physiological Relevance of Tissue Speed-of-Sound	110
6.2.3 Conclusion.....	111
References	112
Appendix A Combining Sparse Matrix Beamforming with Linear Interpolation of RF Signals.....	133
Appendix B Explanation of Structural Redundancy of Channel-Domain Signals.....	135

Appendix C List of Video Files Referenced in Thesis.....	137
--	-----

List of Figures

Figure 1.1 High-level view of the medical ultrasound pipeline. Relation of four core thesis modules to pipeline elements is diagrammed.	3
Figure 1.2 The speed-of-sound ranges for a variety of tissues in the human body. Figure reproduced based on Figure 1.11 from (Cobbold, 2006).	6
Figure 1.3 On the left, the transmission pattern for conventional focused imaging is shown, while on the right, the broad view insonifications for high-frame-rate plane wave ultrasound is shown.	9
Figure 1.4 (a) shows the transmission path for a steered plane wave to arrive at the point (z,x) . (b) shows the receive path for the reflected echo to arrive back at the plane of the transducer.	10
Figure 2.1 B-mode images of a wire target suspended in distilled water beamformed at different SoS values, displayed at 6 dB dynamic range from the peak signal (corresponding to point-spread function full width at half maximum level). Overlaid point targets for different steering angles shown in blue (1° steering) and red (15° steering) in bottom row.	15
Figure 2.2 Differences in transmit paths for refraction corrected beamforming and standard delay-and-sum beamforming (Montaldo et al., 2009). In (a), the difference between the assumed SoS for a plane wave transmit and the actual medium SoS leads to a refraction effect for the transmit path, compared to the straight transmit path of (b)	17
Figure 2.3 Four-step overview for speed-of-sound estimation framework. Data is acquired from several steered plane waves (step 1), then beamformed for many SoS candidates while accounting for refraction and adjusting the imaging depths so that no new structures are introduced in the beamforming view (step 2). The beamformed images are then evaluated for image differences between steering angles via a loss function (step 3). Lastly, the loss function is smoothed and the loss minimum is declared to be the SoS estimate (step 4).	20
Figure 2.4 (a) Setup for measuring a reference SoS via through transmissions by attaching transducers to an electronic caliper. (b) Using the setup to measure the SoS of an agar staircase phantom.	21
Figure 2.5 (a) Setup for measuring a reference SoS via through transmissions by attaching transducers to an electronic caliper. (b) Using the setup to measure the SoS of an agar staircase phantom.	25

Figure 2.6 Speed-of-sound estimates for the collected ultrasound data as compared to the reference measurements. Two beamforming depths and two loss functions are computed for both <i>in vitro</i> agar staircase phantoms (a)-(d) and for <i>in vivo</i> human calves in two states (e)-(h). On each plot, mean signed difference (MSD) error and the maximum intra-sample variability (MIV) is reported in m/s. Values in bold represent our baseline framework.	29
Figure 2.7 Error analysis graphs comparing the absolute SoS estimation error (coefficient of variation loss) versus the measured subcutaneous fat via B-mode imaging for two beamforming depths (a), (b). Figures (c) and (d) show sample B-mode images along with the points used for manual estimation of subcutaneous fat depth – the upper boundary of the muscle fibers.	30
Figure 2.8 Speed-of-sound estimates for the collected ultrasound data for: (a) the proposed framework with coefficient of variation loss, (b) the same framework as in (a), but without refraction compensation, (c) the same framework as in (a), but without depth alignment. On each plot, mean signed difference (MSD) error and the maximum intra-sample variability (MIV) is reported in m/s. 31	
Figure 2.9 Plots comparing the SoS estimation accuracy of the full original proposed framework (a) using 31 steering angles and 71 SoS candidates with a reduced framework (b) with 10 angles and 18 SoS candidates. On each plot, mean signed difference (MSD) error and the maximum intra-sample variability (MIV) is reported in m/s.	32
Figure 2.10 Snapshots of Movies 1, 2, and 3 along with corresponding timestamps. (a) Probe is slid in the middle of bipartite phantom, showing SoS estimate is in between reference measurements. (b) Probe is placed on ex vivo porcine loin, a leaner cut of meat. (c) Probe is placed on flexed biceps muscle in vivo. (d) Probe is placed on same biceps, but muscle is relaxed.	33
Figure 3.1 Pictorial depiction of formation of sparse beamforming matrix. (a) Delay-and-sum (DAS) is the process that forms B-mode images from RF data. (b) Each beamformed pixel is the result of coherently summing appropriately delayed samples from the RF data. (c) The delay-and-sum procedure can be encoded as a sum of dot products between each channel of RF data and a sparse indicator vector. (d) Since the same RF data is used for each pixel, the dot products for each pixel can be stacked to form rows of the beamforming matrix.....	43
Figure 3.2 (a) Example structure of sparse matrix for single transmit event. Successively zoomed snippets of the sparse matrix structure are shown in dark blue and purple outlines. (b) Example block-	

sparse matrix to perform beamforming for multiple transmit events. Block-diagonal matrix for LRI beamforming and fat matrix for HRI beamforming are shown.....	45
Figure 3.3 (a) Timing for three compared beamforming implementations in Matlab. (b) Timings for same three compared beamforming implementations in Python.....	51
Figure 3.4 (a) Nvidia Profiler analysis for sparse matrix beamforming implemented via the cuSPARSE library showing performance is bound by memory. (b) Nvidia Profiler analysis for pixel-parallelized custom CUDA kernel for delay-and-sum showing performance is bound by computation	52
Figure 3.5 Timing for beamforming of focused scanline transmit data. Conventional delay-and-sum (DAS) beamforming is shown in blue, while beamforming of each successive batch of scanline data using sparse matrices is shown in green.....	54
Figure 3.6 Timing for beamforming of compounded images via sparse matrix beamforming (SMB). Looping through sparse matrices for each transmit is shown in red and is approximately 20% slower than using either the HRI matrix to perform compounding and beamforming with one matrix multiply or the LRI matrix to beamform all transmits using a single sparse matrix.....	56
Figure 3.7 (a) Simple neural network architecture to demonstrate proof-of-principle for using sparse matrix beamforming to enable beamforming-integrated neural networks. (b) DAS image from Matlab. (c) DAS image resulting from untrained CNN. (d) DAS image from trained CNN in TensorFlow. (e) DAS image from trained CNN in PyTorch. (f) Kernels from a trained CNN showing that the identity operation was correctly learned both before and after the sparse matrix beamforming layer.....	57
Figure 4.1 Diagram of overall ultrasound imaging system (solid box), with RF data formation framework shown (dashed box). The proposed framework takes in odd-indexed channels, uses a CNN to infer the even-indexed channels, and then interleaves the channels to output a full-channel RF dataset.....	66
Figure 4.2 Proposed architecture of the convolutional encoder–decoder. Odd-indexed RF channels (light blue) are input to the network and zero-padded (green) into a multiple of 8 in size. The encoding layers compress the information, while the decoding layers interpret the compressed information. The number and size of the kernels for each encoding and decoding layers are shown below the corresponding layers (dark and light gray). Two sets of skip connection concatenate the	

outputs of the encoding layers (start of arrows) to the inputs to the decoding layers (end of arrows). The last convolutional layer (dark blue) recombines the information into a single RF frame, which is cropped (orange) to ensure preservation of input and output dimensions, resulting in the inferred even-indexed RF channels (purple).....	67
Figure 4.3 Data manipulation flowchart for all acquired RF data.....	71
Figure 4.4 (a) B-mode image of human carotid in vivo and the corresponding acquired even-indexed RF data (left column) and inferred even-indexed RF data (right column). Whole RF data “images” are shown in second row, (b) and (c). Snippets corresponding to respective orange dashed boxes in (d)–(i).	74
Figure 4.5 Single plane wave transmission B-mode images (40-dB dynamic range) of point target phantom beamformed from original full-channel RF dataset (first column), CNN-inferred channel dataset (second column), linearly interpolated channel dataset (third column), and the original odd-indexed channel dataset (fourth column). Images are shown for -14.75° transmission [top row, (a)–(d)] and for -0.25° transmission [bottom row, (e)–(h)]. Yellow boxes highlight spatial aliasing regions used to calculate artifact power in Figure 4.6.....	76
Figure 4.6 Measured spatial aliasing artifact power in B-mode images across transmission angles for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and the original odd-indexed channel dataset (light blue). Thumbnails of B-mode images beamformed from CNN-inferred channels shown at bottom.....	77
Figure 4.7 (a) B-mode image of point target phantom for 0.25° steering. (b) Cross Section along green lines depicted in (a). (c) Expanded cross Section showing point targets within dashed box. Number labels in (a) correspond to point target lateral FWHM measurements in Table 4.4.....	77
Figure 4.8 Single plane wave transmission B-mode images (40-dB dynamic range) of human carotid beamformed from original full-channel RF dataset (first column), CNN-inferred channel dataset (second column), linearly interpolated channels dataset (third column), and the original odd-indexed channel dataset (fourth column). Two different carotid images are shown for 10° transmission [top row, (a)–(d)] and for -10° transmission [bottom row, (e)–(h)]. Orange box and green box show lumen and tissue regions, respectively, used to calculate lumen–tissue contrast in Figure 4.9.	78

Figure 4.9 Measured lumen–tissue contrast in B-mode images across transmission angles for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and the original odd-indexed channel dataset (light blue). Thumbnails of B-mode images beamformed from CNN-inferred channels shown at bottom.....	79
Figure 4.10 Measured spatial aliasing artifact power in B-mode images across a number of coherently compounded single-transmit images for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and original odd-indexed channel dataset (light blue). Thumbnails of compounded B-mode images beamformed from CNN-inferred channels shown at bottom.	80
Figure 4.11 Measured lumen–tissue contrast in B-mode images across a number of coherently compounded single-transmit images for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and original odd-indexed channel dataset (light blue). Thumbnails of compounded B-mode images beamformed from CNN-inferred channels shown at bottom.	81
Figure 5.1 SoS estimation is performed by beamforming data from multiple steered plane wave transmissions. 1. Data is first acquired from a set of 10 steered plane waves ranging from -15° to 15° . 2. SoS-aware beamforming is performed at a set of 18 SoS candidates by accounting for the shifts in axial position and the effect of refraction between the probe and the medium. 3. For each SoS candidate, an inter-steering loss (Eq. 2.5) is calculated. 4. The inter-steering loss values are fit to a curve and the minimum of the fitted curve is declared the SoS estimate.....	88
Figure 5.2 Equipment used for SoS measurement. (a) Portable ultrasound scanner operating in pulse-echo mode connected via Thunderbolt 3 to a laptop with a discrete GPU for live data processing. (b) Custom through-transmit SoS estimation device.	90
Figure 5.3 Demographics distribution for sampled population. (a) Sample age. (b) Sample body mass index (BMI) (c) Sample physical activity as measured by the Baecke questionnaire.....	90
Figure 5.4 Images of landmarked regions for imaging. (a) The posterior surface was marked temporarily. (b) The tangent point of the resting measuring tape on the anterior surface was marked for imaging. (c) The posterior surface of the calf was marked temporarily. (d) A mocked through-	

transmit device was used to mark both sides of the calf for imaging. (e) The quadriceps was temporarily marked above the rectus femoris muscle. 91

Figure 5.5 All imaging views used during the study for ultrasound data acquisition along with corresponding image thumbnails (a) Anterior surface of biceps in extended position. (b) Anterior surface of biceps in flexed position. (c) Anterior surface of biceps in flexed position with muscle contraction. (d) Lateral surface of calf, leg resting at 90° angle. (e) Medial surface of calf, leg resting at 90° angle. (f) Through-transmit acquisition signal from oscilloscope. (g) Cross-sectional view of rectus femoris muscle, leg extended and relaxed. (h) Cross-sectional view of rectus femoris muscle, leg extended and contracted. (i) Long-axis view of rectus femoris muscle, leg extended and relaxed. 92

Figure 5.6 Effect of applied probe pressure on an imaging view of the calf and corresponding SoS estimates (a) Minimal pressure view with an excess of ultrasound gel, leading to layer of gel between skin surface and probe. (b) View using minimal amount of pressure required for probe to be fully in contact with the skin. (c) View with additional pressure applied beyond what is required for probe to be fully in contact with the skin. 96

Figure 5.7 Box-and-whiskers plots comparing the estimated SoS values for the dominant and nondominant appendages. Boxes correspond to lower and upper quartile, while whiskers correspond to nonoutlier maximum and minimum. The p-values shown are reported based on paired t-tests. 97

Figure 5.8 Plots comparing the measured muscle SoS with the participant body mass index (BMI). Linear regression was performed for males (M, blue) and females (F, red) separately, with the lines of best fit (dashed lines) and associated Pearson correlation coefficient and p-value reported in the corners of each plot. 98

Figure 5.9 Box-and-whiskers plots comparing the estimated SoS values for males and females separated for each muscle group. Boxes correspond to lower and upper quartile, while whiskers correspond to nonoutlier maximum and minimum. The p-values shown are reported based on unpaired t-tests. 99

Figure 5.10 Box-and-whiskers plots comparing the estimated SoS values for different states of biceps flexion. Boxes correspond to lower and upper quartile, while whiskers correspond to nonoutlier maximum and minimum. The p-values shown are reported based on paired t-tests. 99

Figure 5.11 Plots comparing the measured biceps muscle SoS with the (a) participant Baecke index and the (a) participant sport index. The sport index is one of three constituents of the Baecke index. Linear regression was performed for males (M, blue) and females (F, red) separately, with the lines of best fit (dashed lines) and associated Pearson correlation coefficient and p-value reported in the corners of each plot. 100

Figure 5.12 Plots assessing the difference between the through-transmit reference SoS values and the pulse-echo estimated SoS values. (a) The reference SoS and the estimated SoS are plotted against each other and results from linear regression are shown. (b) The signed difference between the reference SoS and the estimated SoS are plotted against the participant BMI..... 101

List of Abbreviations

A-line	Amplitude Line
ASIS	Anterior Superior Iliac Spine
BMI	Body Mass Index
B-mode	Brightness Mode
CNN	Convolutional Neural Network
COO	Coordinate Format
CPU	Central Processing Unit
CS	Compressed Sensing
CSR	Compressed Sparse Row
CT	Computed Tomography
CUBDL	Challenge on Ultrasound Beamforming with Deep Learning
CUDA	Compute Unified Device Architecture
CUTE	Computed Ultrasound Tomography in Echo-mode
DAS	Delay-and-Sum
F	Female
FPGA	Field-Programmable Gate Array
FWHM	Full Width at Half Maximum
GPU	Graphics Processing Unit
HiFRUS	High Framerate Ultrasound
M	Male
MIV	Maximum Intra-Sample Variability
MRI	Magnetic Resonance Imaging

MSD	Mean Signed Difference
MSK	Musculoskeletal Ultrasound
ORE	Office of Research Ethics
PLA	Polylactic Acid
PSF	Point-Spread Function
ReLU	Rectified Linear Unit
RF	Radiofrequency
Rx	Receive
SMB	Sparse Matrix Beamforming
SoS	Speed-of-Sound
Tx	Transmit
USCT	Ultrasound Computed Tomography
VRAM	Video Random Access Memory

List of Symbols (Order of Appearance)

c	Speed-of-sound
K	Bulk modulus
ρ	Density
z	Axial position of pixel
x	Lateral position of pixel
z_e	Axial position of transducer element
x_e	Lateral position of transducer element
θ_T	Plane wave steering angle on transmit
p	Index of beamforming point
e	Index of beamforming element
$\tau(p, e)$	Time delay between point p and element e
$d_{TX}(p)$	Geometric distance for transmit to reach point p
$d_{RX}(p, e)$	Geometric distance for received echo to reach element e from point p
c_o	Speed-of-sound
$\Delta(p, e)$	Sample delay between point p and element e
f_s	Sampling frequency
$s(p, e)$	Correctly delayed signal value for point p and element e
$\mathbf{D}$	Preprocessed RF data matrix
$\alpha(p, e)$	Generalized receive apodization weight for point p and element e
$\mathbf{I}$	Image intensity value matrix
v_M	Speed-of-sound of medium
v_T	Speed-of-sound assumed on transmit

θ_M	Refracted transmit steering angle through medium
z_T	Intended beamforming depth based on transmit SoS
z_M	Corrected beamforming depth based on transmit and medium SoS
N_θ	Number of steering angles used in transmit
N_Z	Number of axial pixel positions used in beamforming
N_X	Number of lateral pixel positions used in beamforming
$\mathcal{L}$	Inter-steering loss function
N_{RF}	Number of RF samples per channel
N_C	Number of RF channels
N_S	Chapter 2: Number of SoS candidates used for beamforming
N_S	Chapter 3: Number of RF samples per channel
$\sigma_\theta(p)$	Complex standard deviation at pixel p evaluated across steering angles
$ \mu_\theta (p)$	Absolute mean at pixel p evaluated across steering angles
$R_p(\theta)$	Autocorrelation at pixel p evaluated for lag θ
$\vec{\alpha}(p, e)^T$	Transpose of indicator vector encoding apodization and sample delay
$\vec{I}$	Column-wise vectorization of 2D image intensity matrix $\mathbf{I}$
$\vec{D}$	Column-wise vectorization of 2D RF data matrix $\mathbf{D}$
N_B	Number of batches to beamform at once

Chapter 1

Introduction

1.1 Musculoskeletal Ultrasound and Tissue Speed-of-Sound

The overall objective of my thesis is the design of a real-time tissue speed-of-sound (SoS) estimation framework to provide an additional quantitative tissue biomarker for muscle quality assessment. By developing a novel pulse-echo ultrasound scheme for SoS estimation, standard ultrasound probes can be used bimodally to augment musculoskeletal (MSK) ultrasound imaging. Real-time operation will enable this framework to be used at the point-of-care, where ultrasound imaging has the most advantages as an imaging modality.

Ultrasound is a non-invasive and versatile imaging modality that can be used to visualize structures and functions of the human body based on pulse-echo acoustics in real-time (Whittingham, 2007). MSK assessment is one application of ultrasound which can provide valuable information about the morphology and physiology of muscles, tendons, ligaments, and bones (Backhaus, 2001; Czynny, 2017; Grassi et al., 2004; Peetrons, 2002). Not only is MSK ultrasound useful for identifying acute pathologies, but it can also be used as an indicator of overall health through measurement of muscle quality or body composition (Nagae et al., 2023). For example, muscle mass or thickness can act as a predictor of duration of intensive care stays or for age-related muscle decay (Paris & Mourtzakis, 2016; Strasser et al., 2013). Quantifying muscle quality has become an important area of research because of the importance of muscle as a predictor of health and the role muscle plays in health and disease (Mourtzakis et al., 2017; Nagae et al., 2023; Wolfe, 2006). Thus, researchers have investigated not only the general structure of MSK ultrasound images, but quantitative metrics based on the ultrasound images (Brandenburg et al., 2014; Harris-Love et al., 2014; Parry et al., 2015). These metrics can include measurements such as muscle thickness, echogenicity, stiffness, texture, and pennation angle (Drakonaki et al., 2012; Mourtzakis et al., 2017; Nagae et al., 2023; Nodera et al., 2018). Some parameters such as muscle thickness are directly measurable, but can be affected by applied pressure (Paris et al., 2017). Others such as echogenicity are quantitative but can depend on scanner settings and operator skills (Young et al., 2015).

A robust measurable quantity that reflects an intrinsic property of the muscle tissue is the tissue speed-of-sound (SoS). Tissue SoS has shown potential as a tissue biomarker for conditions such as osteoporosis, breast cancer, and liver steatosis (Imbault et al., 2017; Li, Duric, et al., 2009; Rossman et

al., 1989). In the context of MSK ultrasound, the tissue SoS has been shown to be related to pathologies such as edema and sarcopenia (Sanabria et al., 2019; Topchyan et al., 2006). However, measuring tissue SoS *in vivo* frequently requires niche hardware, such as an ultrasound computed tomography system (Wiskin et al., 2020), which is not widely available, or a through-transmission setup (Cummings et al., 2002), which does not have imaging potential. Moreover, these methods may not be suitable for some applications of MSK ultrasound, as they are limited by the bilateral nature of the imaging window. Therefore, a framework that can measure the SoS in real-time and with a standard ultrasound probe is critical to translate SoS insights to augment MSK ultrasound imaging.

1.2 A Real-Time Framework for Global Speed-of-Sound Estimation

1.2.1 Thesis Objectives: Real-Time Imaging Framework

Considering the challenges posed in the previous section for quantitative real-time MSK ultrasound, my thesis will meet the following specific objectives:

1. To develop and evaluate a framework for accurate SoS estimation (<5 m/s mean error) using a combination of signal processing and ultrasound imaging techniques. This objective will advance the use of tissue SoS as a robust and quantitative biomarker.
2. To implement and optimize the framework for simultaneous SoS estimation and B-mode imaging (>15 fps) using parallel computing through GPU acceleration. This objective will preserve the advantages of ultrasound as a live imaging modality and demonstrate the ability to perform point-of-care assessment of tissue SoS.
3. To validate the framework via *in vivo* studies on healthy volunteers using a point-of-care-capable ultrasound scanner. This objective will demonstrate the clinical translation potential of the framework and statistically compare the outcomes with alternative SoS methods reported in prior literature.

1.2.2 Hypothesized Solution: Plane Wave Ultrasound Imaging

I hypothesize that an accurate (<5 m/s mean error) and real-time (>15 fps) pulse-echo SoS estimation framework can be realized on a portable ultrasound scanner via steered plane wave transmissions. Our solution consists of three primary steps:

1. Transmission of steered plane waves to insonify the imaging medium. Using plane waves, the number of total transmissions required is minimized, while each transmission still provides a full view of the imaging medium.
2. SoS-aware beamforming at many SoS candidate values. By efficiently beamforming the data from all transmitted steering angles for a selection of different SoS candidates, image dissimilarity can be evaluated for all images across all SoS candidates in real-time.
3. Employment of quality metric to identify optimal beamforming SoS, which is subsequently declared the SoS estimate. The quality metric provides an estimate for image dissimilarity between steering angles and the metric should have rapid computation time.

1.3 Thesis Outline

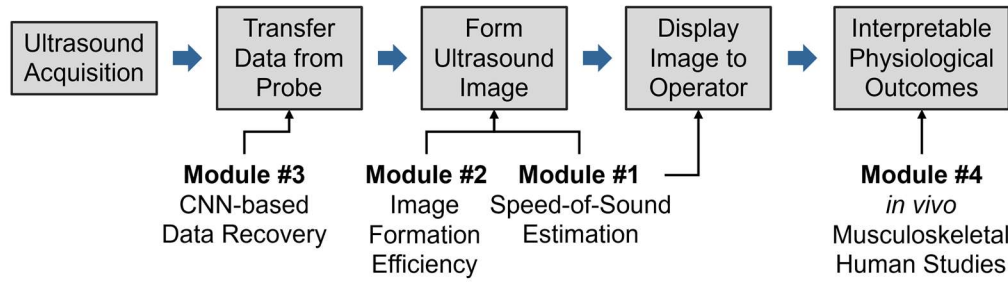


Figure 1.1 High-level view of the medical ultrasound pipeline. Relation of four core thesis modules to pipeline elements is diagrammed.

The remainder of **Chapter 1** will serve as a literature review of speed-of-sound in ultrasound and summarize ultrasound image formation principles. Figure 1.1 provides a high-level overview of the ultrasound imaging pipeline along with the relation of the four core thesis modules to the pipeline, which is covered by the next four chapters of the thesis.

Chapter 2 describes **Module #1**, the core real-time SoS algorithm and implementation on a portable scanner. By leveraging image differences between differently steered plane wave images of the same imaged medium, a novel high-accuracy global SoS estimation algorithm is developed. Through GPU-based acceleration of SoS-aware beamforming, this algorithm is implemented in real-time on a portable scanner and demonstrated *ex vivo* and *in vivo*.

Chapter 3 describes **Module #2**, which explores in detail the acceleration of delay-and-sum (DAS) beamforming for real-time operation. By reformulating delay-and-sum beamforming as a single sparse matrix multiplication, access to powerful sparse matrix libraries become available as tools for acceleration of DAS. The foundations of sparse matrix beamforming (SMB) are derived and additional means for increasing the efficiency of SMB are explored.

Chapter 4 describes **Module #3**, which addresses the data transfer sub-challenge for real-time high-frame-rate imaging. For plane wave imaging, the data transfer rate from the front-end acquisition to the imaging system can be a bottleneck for real-time implementation. A convolutional neural network (CNN) approach is proposed to reduce the data transfer without sacrificing image quality due to reduced quantity of data.

Chapter 5 describes **Module #4**, the application of the real-time framework for an *in vivo* human study to investigate physiological relationships. By using the real-time portable system developed in the previous modules, a study on human muscles is conducted using a small sample of volunteers. The study had two primary aims; firstly, to establish which imaging views are optimal for SoS estimation in large muscles (calf, quadriceps, biceps); secondly, to investigate physiological relationships between demographic factors (age, BMI, etc.) and muscle SoS.

Chapter 6 summarizes the thesis contributions, discusses the limitations of the presented work, and suggests possible future directions for continuing this research.

1.4 Background Information

1.4.1 Medical Ultrasound Imaging

Ultrasound imaging, also known as sonography, is a non-invasive medical imaging modality using high-frequency sound waves to image internal structure in the human body (Whittingham, 2007). Additionally, as a handheld and live imaging modality, ultrasound can be used dynamically at the point-of-care, rather than assessed offline after the scan has occurred (Díaz-Gómez et al., 2021; Moore & Copel, 2011). The general principle of ultrasound imaging is the transmission of acoustic pulses and the subsequent reception of the reflected echoes from structure of the image (Whittingham, 2007). By applying time-of-flight calculations to account for the acoustic transmission and reception paths, ultrasound images can be formed from the echo data (Perrot et al., 2021). The grayscale images typically associated with ultrasound are known as B-mode, or brightness mode, images (Whittingham,

2007). B-mode ultrasound has a wide range of applications in medicine; some specialties that regularly apply ultrasound include obstetrics, emergency medicine, or rheumatology (Grassi et al., 2004; Richards & McGahan, 2017; Torloni et al., 2009).

Musculoskeletal (MSK) ultrasound is a subspecialty of ultrasound focusing on imaging of muscles, connective tissues, and bones in the human body. MSK ultrasound provides detailed information about the morphology and physiology of the musculoskeletal system (Backhaus, 2001; Czyrny, 2017; Grassi et al., 2004; Peetrons, 2002). MSK ultrasound can be used to identify pathologies such as muscle or tendon tears, bursitis, and tumors or cysts among others (Backhaus, 2001; Peetrons, 2002). The dynamic nature of ultrasound allows for MSK imaging to be used interventionally for ultrasound-guided procedures; for example, to mitigate injection-related injury (Grassi et al., 2004). Ultrasound imaging has also been used for quantitative measurements of various parameters that reflect the quality and function of musculoskeletal tissues (Nagae et al., 2023). These quantitative parameters have been correlated with physiological pathologies of sarcopenia, muscle atrophy and more (S. E. Campbell et al., 2005; Minetto et al., 2016; Puthucherry et al., 2015).

1.4.2 Speed-of-Sound of Human Tissues

Ultrasound imaging strongly depends on the speed at which the acoustic pulse propagates through the imaged medium. The SoS c of a liquid is governed by the Newton-Laplace equation (Cobbold, 2006):

$$c = \sqrt{\frac{K}{\rho}} \quad (1.1)$$

where K is the bulk modulus and ρ is the density of the medium. Implicit in this equation is also a dependence on the temperature of the medium, which can nonlinearly affect both the density and the bulk modulus properties. For example, the SoS in pure water varies nonlinearly, but predictably with the temperature of the water (Bilaniuk & Wong, 1993). Given this expression for the SoS, naturally different tissues in the body will have different intrinsic SoS values. Figure 1.2 shows example SoS values for some different tissues in the body using a chart reproduced based on Figure 1.11 from (Cobbold, 2006). The human body is predominantly composed of water (~60%); thus, the SoS of soft tissue is similar to that of water (~1500 m/s). It can be observed from Figure 1.2 that the expected range of SoS in soft tissue varies from around 1350 m/s to 1700 m/s.

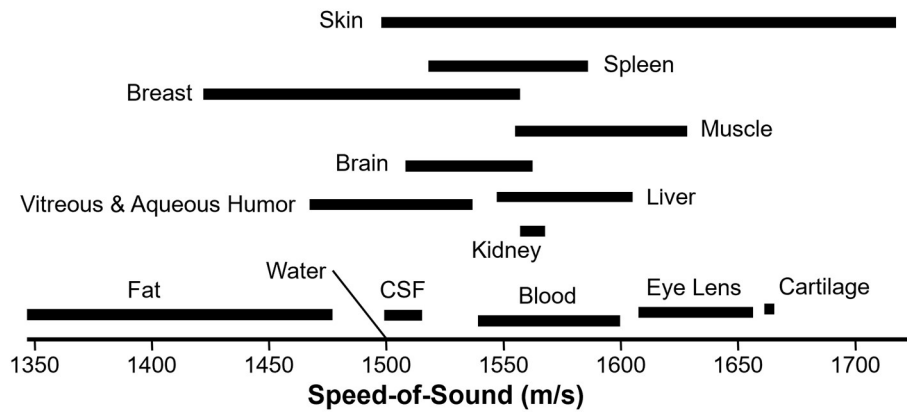


Figure 1.2 The speed-of-sound ranges for a variety of tissues in the human body. Figure reproduced based on Figure 1.11 from (Cobbold, 2006).

The SoS of human tissue could only be measured *ex vivo* until recent decades. These previous systems typically used small samples of human tissue in setups that precisely measured parameters such as the time-of-flight, sample distance, and other geometric techniques to calculate the SoS of the excised tissue (Bamber & Hill, 1979; Sehgal et al., 1986). Using similar frameworks, the speed-of-sound in excised animal tissue has also been measured (Bamber & Hill, 1979). Samples of bovine meat with more intramuscular fat have slower SoS, while freeze-thaw cycles in porcine samples did not significantly change the SoS (Koch et al., 2011a, 2011b; Park et al., 1994). These types of *ex vivo* studies increased the understanding of the relationship between pathologies and the intrinsic quantitative SoS property of tissues (Lin et al., 1987). For example, excised liver in both humans and in rats have demonstrated slower SoS if affected by steatohepatitis, or fatty liver disease (Kumagai et al., 2019; Lin et al., 1987; Sehgal et al., 1986).

In recent years, measurement of the SoS *in vivo* has grown in interest for use as a tissue biomarker. One direction for SoS measurement was the development of ultrasound computed tomography (USCT) systems. USCT, similar to X-ray computed tomography (CT), uses tomographic principles and an ultrasound ring transducer for either SoS mapping of the issue via time-of-flight tomography or attenuation mapping via backprojection of the signal intensity (Li, Duric, et al., 2009; Wiskin et al., 2020). USCT research has demonstrated clinical relevance, with some prototype systems currently available for purchase (Duric et al., 2014; Wiskin et al., 2020). Via USCT, SoS maps of human tissue can be constructed and have shown, for example, the ability to distinguish between malignant and

benign breast tumours (Li, Duric, et al., 2009). Another direction is the use of through-transmit systems to directly measure the SoS through a section of human tissue, where many clinical applications exist in bone densitometry to assess bone health or risk of osteoporosis in women (Cummings et al., 2002). These machines can measure the SoS through bone using lower frequency pulses to provide a quantitative score for bone health (Cummings et al., 2002). Bone densitometry can help identify pathologies such as osteoporosis or bone fragility (Frost et al., 2001; Gnudi et al., 2000; Litniewski et al., 2000). More recently, through-transmit systems for muscle quantification have found applications in identifying issues such as sarcopenia in elderly women, or muscle atrophy from casting (Ruby et al., 2021; Sanabria et al., 2019). Both USCT and through-transmit systems require bilateral access to the salient tissue, which can limit their applicability.

Spurred by the potential for using tissue SoS as a pathological biomarker, efforts for pulse-echo based algorithms for SoS estimation and imaging have grown in popularity. The advantages of pulse-echo SoS estimation are two-fold; firstly, the diagnostic value of using the SoS to differentiate tissue, and secondly, the ability to use the estimated SoS for aberration correction of the pulse-echo images (Ali et al., 2023). Indeed, inspired by USCT results, one of the early pulse-echo SoS mapping algorithms was denoted computed ultrasound tomography in echo-mode (CUTE) (Jaeger et al., 2015). In addition, pulse-echo SoS methods make use of the same probes as found in conventional ultrasound, which can accelerate the clinical adoption of these techniques. Many of these recent algorithms rely on broad-view unfocused transmissions such as plane wave, diverging wave, or synthetic aperture (Bezek & Goksel, 2023; Rau et al., 2021). Pulse-echo SoS estimation algorithms can loosely be classified into two categories: 1) local SoS, and 2) global SoS.

In local SoS imaging, the goal is to construct an SoS mapping of the imaged tissue. These methods will typically model the SoS on pixel-wise basis, and correspondingly calculate the propagation paths based on the SoS model. Either straight-ray paths, discretely integrated along the ray paths (Sanabria et al., 2018), or paths accounting for refraction, which can be calculated via the eikonal equation for example, are assumed to model acoustic propagation (Ali, Brevett, et al., 2022). Once the model is selected, for different transmissions some metric is calculated that depends on the SoS map. For example, CUTE uses phase differences (Jaeger et al., 2015), while other methods can use displacement differences instead (Sanabria et al., 2018). After the metric is calculated, a large optimization problem is solved to generate the SoS map (Jaeger et al., 2015). This optimization problem can be challenging

to implement for real-time operation. Another newer method uses deep learning based on simulated data, however it is difficult to translate these results for clinical use (Feigin et al., 2020).

In global pulse-echo SoS estimation, the entire imaging view is considered to form an overall estimate of the SoS. By modelling the medium via a single SoS value, assumptions can be made based expected effects of the SoS mismatch to estimate the SoS (Anderson et al., 2000); these assumptions are easier to model compared to local SoS estimation. Some methods rely on coherence patterns in raw acoustic data (Yoon et al., 2011). Others rely on post-beamformed images to evaluate a quality metric, such as image alignment through image registration or speckle brightness maximization (Anderson et al., 2000; Krucker et al., 2004). However; none of these methods have been validated *in vivo* for SoS estimation accuracy (Krucker et al., 2004). While computationally simpler than local SoS methods, global SoS estimation methods have yet demonstrate real-time operability. Thus, an accurate, real-time framework would aid in the clinical translation of SoS insights, which I design in **Chapter 2** of my thesis.

1.4.3 What is High-Frame-Rate Ultrasound?

In the past decade, acoustic pulses of unfocused planar or diverging wavefronts that insonify the entire imaging view has enabled the development of new imaging techniques. These unfocused pulses allow for orders of magnitude higher framerates compared to conventional scanline imaging, as an image frame can be generated for each pulse, instead of requiring multiple pulses for a single image frame (Tanter & Fink, 2014). This high-frame-rate, or ultrafast, ultrasound imaging paradigm can achieve imaging framerates of over 1000 frames per second (Boni et al., 2018; Tanter & Fink, 2014). High-frame-rate ultrasound has proven effective for novel imaging modes such as elastography (Montaldo et al., 2009), vector flow imaging (Jensen et al., 2016a, 2016b), and ultrasound localization microscopy (Errico et al., 2015).

Figure 1.3 illustrates the difference in transmission between scanline and plane wave imaging. On the left, multiple scanlines are shown along with the probe. Each scanline is generated by focusing a subset of the transducer elements (active transducers shown in black), and this process is repeated until the entire imaging medium has been insonified. In contrast, for plane wave imaging shown on the right of Figure 1.2, each transmission insonifies the entire imaging medium. These plane waves can also be steered (transmitted at an angle) to acquire additional data for imaging enhancements, such as coherently compounding the resulting B-mode images (Montaldo et al., 2009). These plane wave

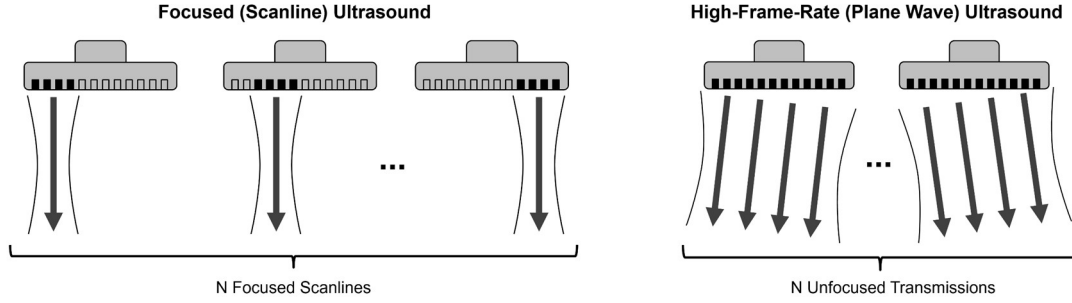


Figure 1.3 On the left, the transmission pattern for conventional focused imaging is shown, while on the right, the broad view insonifications for high-frame-rate plane wave ultrasound is shown.

transmissions are unfocused, and focusing and spatial localization occur only upon reception of the acoustic echoes, instead of during both transmit and receive. The spatial resolution of this plane wave imaging, when compounding of multiple transmits are applied, can exceed conventional scanline imaging and the high frame rates can reveal dynamic events happening on a very short temporal scale (<1 ms) (Bae et al., 2020; Tanter & Fink, 2014). A key challenge of these high-frame-rate systems is the high streaming bandwidth to transfer the raw data from the array front-end to the computed back-end (Boni et al., 2018). Upwards of 10 GB of raw data can be transmitted each second, which can be demanding for conventional hardware; thus, I propose a deep learning-based approach to reduce data transfer in **Chapter 4** of my thesis.

1.4.4 Ultrasound Image Formation – Delay-and-Sum Beamforming

Delay-and-sum (DAS) beamforming is the predominant image formation algorithm for biomedical ultrasound imaging (Perrot et al., 2021). The objective of DAS is to perform spatiotemporal localization of the received radiofrequency (RF) data after a pulse-echo event; specifically for high-frame-rate ultrasound, each transmit-receive event leads to a corresponding image frame. In DAS, each image is formed by first anticipating the time-of-flight given a hypothetical scatterer at that pixel position. By repeating this process for each possible scatterer position, corresponding to the pixels, the RF data can be appropriately delayed. Next the summation is performed for the delayed samples from each channel. This summation is performing coherently, accounting for the phase of each channel; where there is strong acoustic scattering the signals will sum more coherently, and where there is no scattering the signals will sum incoherently and result in lower magnitude. In this way, an intensity image can be formed from the received RF data.

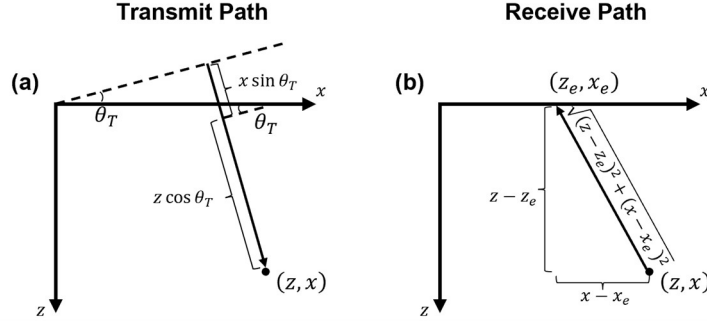


Figure 1.4 (a) shows the transmission path for a steered plane wave to arrive at the point (z, x) . (b) shows the receive path for the reflected echo to arrive back at the plane of the transducer

Formalizing the DAS algorithm for a linear probe, let an imaged pixel p be assigned a position (z, x) on a Cartesian grid. Similarly, let a transducer element e be assigned a position (z_e, x_e) on the probe face. Then, let a plane wave be transmitted with angle θ_T towards the pixel position. Figure 1.4 illustrates this example for the transmissions path and reception path of the plane wave. First, the plane wave propagates towards the pixel p . Upon encountering the hypothetical point reflector, the acoustic echoes are isotropically reflected and must travel back to the probe for detection.

To begin image formation, the total time-of-flight for each channel must be calculated by evaluating both the transmit and the receive geometry. The total time-of-flight $\tau(p, e)$ for a point and a transducer element can be calculated as:

$$\tau(p, e) = \frac{d_{TX}(p) + d_{RX}(p, e)}{c_o} \quad (1.2)$$

where $d_{TX}(p)$ represents the distance between the initial plane wave transmit, the point p , $d_{RX}(p, e)$ is the distance between point p element e , and c_o is the assumed homogenous sound speed of the medium. In general, $d_{TX}(p)$ depends on the type of transmitted pulse and $d_{RX}(p, e)$ depends on the transducer geometry. For the case of a plane wave with transmit angle θ , we can express d_{TX} and d_{RX} using geometrical arguments (Montaldo et al., 2009) as:

$$d_{TX}(p) = z \cos \theta_T + x \sin \theta_T \quad (1.3)$$

$$d_{RX}(p, e) = \sqrt{(z - z_e)^2 + (x - x_e)^2} \quad (1.4)$$

Noting that $\tau(p, e)$ lies in the time domain, it must be converted to a sample index. We can use the sampling frequency f_s to calculate this best-fitting sample $\Delta(p, e)$ via:

$$\Delta(p, e) = f_s * \tau(p, e) \quad (1.5)$$

Eq. (1.2) – (1.5) form the delay steps of the DAS beamforming algorithm. Now that the sample has been correctly delayed, the summation step can be performed. Let the preprocessed (analog-to-digital conversion, bandpass filter, analytic signal conversion) RF data matrix be represented by the $N_S \times N_C$ sized matrix $\mathbf{D}$ where N_S, N_C are the sampling depth (number of samples) and the number of channels respectively. Then, let the appropriately delayed signal (with sample delay Δ) from each channel e to form a specific pixel $p = (z, x)$ can be represented as

$$s(p, e) = \mathbf{D}(\Delta(p, e), e) \quad (1.6)$$

The summation step across channels can then be expressed as a weighted summation, which results in the intensity $I(p)$ at a given pixel:

$$I(p) = \sum_{e=1}^{N_C} \alpha(p, e) s(p, e) \quad (3.2)$$

We note that the apodizations weights $\alpha(p, e)$ correspond to applying receive apodization to the DAS procedure (Hoen, 1982) but represented in a more generalized form (one weight factor per channel per pixel). We also note that the selection of an F-number for receive beamforming can be considered a form of apodization with weights set to zero outside of the aperture (Montaldo et al., 2009; Perrot et al., 2021).

Acceleration of DAS is essential for real-time ultrasound operation. For example, general purpose GPU programming for acceleration of DAS via parallelization is one popular solution for achieving real-time plane wave and synthetic aperture ultrasound imaging (Yiu et al., 2011). An alternative to GPUs for DAS acceleration is to use field-programmable gate arrays (FPGAs) to perform image formation (Boni et al., 2017; N. A. Campbell & Brown, 2022; Kou et al., 2023). Both GPU and FPGA implementations of DAS required specialized knowledge for optimal acceleration. Software libraries with ultrasound beamforming capabilities have been made available in recent years, but may not be suitable for adaptation for more specific applications (Garcia, 2021; Hyun et al., 2019; Rodriguez-Molares et al., 2017). Thus, an efficient and accessible DAS beamformer would be beneficial for the testing of real-time ultrasound algorithms, which I investigate in **Chapter 3** of my thesis.

1.5 Summary of Introduction

The overall objective of my thesis is to design a real-time pulse-echo framework to estimate the tissue SoS. For this goal, I first introduce a novel global SoS estimation algorithm demonstrates high accuracy, then implemented this algorithm for real-time operation (**Chapter 2**). This algorithm depends on the principles of high-frame-rate imaging (Sec. 1.4.3) for insonification of the entire imaging view. While implementing the real-time framework, I address two sub-challenges for high-frame-rate imaging: acceleration of DAS beamforming (**Chapter 3**) and reduction of data transfer rates (**Chapter 4**). Lastly, I conduct an *in vivo* human study to validate the use of the proposed framework for providing an additional quantitative tissue biomarker for musculoskeletal (MSK) ultrasound (**Chapter 5**). Through this thesis, I strive to demonstrate the ability for real-time SoS estimation to find clinical applications.

Chapter 2

Real-time Speed-of-sound Estimation *in vivo* via Steered Plane Wave Ultrasound

2.1 Introduction

Pulse-echo ultrasound imaging relies on time-of-flight calculations based on the speed-of-sound (SoS) to form ultrasound images. Typically, for image formation, the SoS inside the human body is assumed to be a constant 1540 m/s/ (Whittingham, 2007). While this assumption is sufficient for forming high-quality images, it neglects to account for the varying SoS of different types of tissue (e.g., muscle vs. fat) in the human body (Cobbold, 2006; Goss et al., 1978). The tissue SoS can serve as a biomarker for tissue composition; studies have shown the tissue SoS can help to assess intramuscular fat content (Park et al., 1994; Ruby et al., 2020), detect age-related or cast-related muscle decay (Ruby et al., 2021; Sanabria et al., 2019), and differentiate between benign and malignant tumors (Li, Duric, et al., 2009). While quantitative SoS measurement can be useful, B-mode ultrasound is still ideal for live imaging of human tissue at the bedside. Thus, it would be advantageous to simultaneously connect the potential insights of quantitative SoS estimation with real-time B-mode ultrasound imaging.

Ultrasound-based SoS estimation can be accomplished via three types of methods: 1) global (or average) SoS estimation, 2) local SoS estimation to generate a pixel-wise SoS image, 3) time-of-flight ultrasound computed tomography (USCT) (Ali et al., 2023). For local SoS estimation and tomography, both methods result in a localized SoS map that can help identify tissue characteristics or identify local structures not visible in the B-mode image (Jaeger et al., 2015; Li, Duric, et al., 2009). Local SoS estimation with pulse-echo ultrasound uses computationally expensive methods that currently have limited suitability for real-time imaging, while suffering from reduced quantification accuracy *in vivo* (Jaeger et al., 2015; Sanabria et al., 2018). Compared to local SoS estimation, USCT has superior accuracy *in vivo* but requires specialized ring transducer hardware that limits its applicability outside of breast or small limb imaging (Li, Duric, et al., 2009; Wiskin et al., 2020). Additionally, there are no systems that can perform USCT in real-time (Song et al., 2022). Because of these limitations for local SoS estimation and USCT, global SoS estimation is a potential solution for accurate *in vivo* SoS measurement with real-time capabilities; the primary drawback for global SoS estimation is the lack of

localization. For roughly homogeneous media such as muscle tissue, where the SoS can be a useful biomarker (Ruby et al., 2021), this drawback is lessened.

Global SoS estimation is a well-studied area of ultrasound research; the speed-of-sound of the imaged medium is crucial for calculating the time-of-flights required for delay-and-sum image formation. Early methods for estimating the SoS could only measure the SoS using transmission methods (Greenspan & Tschiegg, 1957). These methods required two-sided access to the measured tissue and were frequently only applied *ex vivo* for human tissue characterization (Goss et al., 1978; Van Venrooij, 1971). In the recent decades, new pulse-echo SoS estimation methods that enable SoS estimation without bilateral *in vivo* were developed. These methods can be based on principles such as beam tracking (Ophir, 1986), speckle brightness maximization (Anderson et al., 2000), or image registration for steered transmits (Krucker et al., 2004). There has yet to be a demonstration of these SoS techniques applied in real-time. Another challenge of all these methods is validating that the estimation is accurate *in vivo* (Krucker et al., 2004). Thus, a new global SoS estimation method that functions in real-time while maintaining high accuracy *in vivo* can be beneficial for advancing the practical use of the tissue SoS as a tissue biomarker.

In this work, we propose a new global SoS estimation framework with demonstrable real-time capability. We hypothesize that the SoS can be estimated by finding the beamforming SoS that minimizes differences between images beamformed from steered plane wave transmissions. The images are reconstructed using modified delay-and-sum beamforming, accounting for the effects of assuming a beamforming SoS that does not match the medium SoS. We validate the accuracy of our framework through custom *in vitro* agar phantoms with varying glycerol concentrations to control the SoS. We also assess the accuracy *in vivo* on human calf muscles by comparing to reference SoS measurements acquired with through transmissions. Lastly, we demonstrate with video recordings the first device capable of concurrent B-mode imaging and SoS estimation in real-time using a portable open platform research scanner. Our work shows that the quantitative insights from real-time SoS estimation can aid in preliminary investigations for relating SoS with muscle health and function.

2.2 Theory

To estimate the SoS, we propose to leverage the information contained in steered plane wave transmits. Steered plane waves allow for insonification of the entire imaging view with a single transmit event, considerably reducing the number of transmissions required to form images from many steering angles (Montaldo et al., 2009). Each steered transmission follows a different path to the scatterers in the

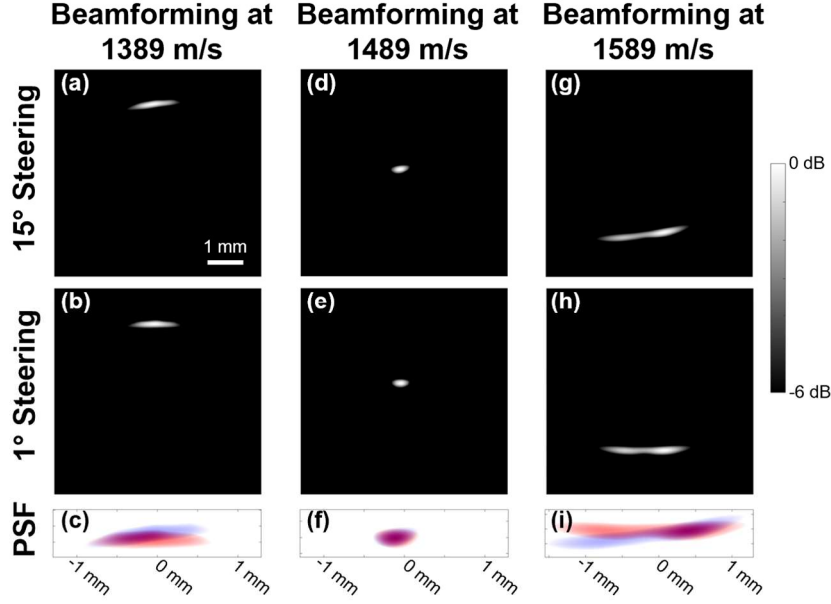


Figure 2.1 B-mode images of a wire target suspended in distilled water beamformed at different SoS values, displayed at 6 dB dynamic range from the peak signal (corresponding to point-spread function full width at half maximum level). Overlaid point targets for different steering angles shown in blue (1° steering) and red (15° steering) in bottom row.

medium, resulting in differences in the post-beamformed images. By finding the SoS that minimizes these differences, we can then generate an estimate of the average SoS in the medium. The following subsections explore this principle in greater detail.

2.2.1 Effect of Speed-of-Sound Mismatch on Point Targets for Steered Plane Wave Transmissions

When the beamforming SoS does not match the true medium SoS, there is a corresponding loss of image quality. We define this type of scenario to be a SoS mismatch. SoS mismatch results in a quantifiable loss of the imaging resolution and contrast and has previously been studied for focused transmissions (Anderson et al., 2000). Because our proposed method relies on unfocused transmissions, the effect of SoS mismatch on the point-spread function for the plane wave transmission scenario must also be examined.

For unfocused plane wave transmissions, we use a wire suspended in distilled water to study the effect of SoS mismatch on the point-spread function (PSF). Figure 2.1 shows B-mode images of a

single point target placed at approximately 25 mm depth relative to the transducer surface, suspended in distilled water at 22.3 °C which has a known SoS of 1489 m/s (Bilaniuk & Wong, 1993). Steered plane wave acquisitions were then beamformed at a lower (Figure 2.1(a)-(c)), matching (Figure 2.1(d)-(f)), and higher SoS (Figure 2.1(g)-(i)) compared to the medium for two transmit steering angles (15° and 1°). Naturally, the resolution of the point target was optimized when the beamforming SoS matches the medium SoS (Figure 2.1(d)-(f)). In contrast, the PSF spreading is apparent in SoS mismatch scenarios. The loss of resolution in scenarios with SoS mismatch is caused by summation along the incorrect hyperbolas during delay-and-sum beamforming (Perrot et al., 2021). For scenarios with SoS mismatch, not only is the resolution worsened, but also the PSFs of the point targets are positionally offset for different steering angles. The bottom row of Figure 2.1 shows the beamformed point targets overlaid and made transparent for the different SoS scenarios (blue target – 15° steering, red target – 1° steering). It can be observed that when there is SoS mismatch (Figure 2.1(c) and (i)), the non-overlapping area for the point targets is much larger compared to when there is no SoS mismatch (Figure 2.1(f)). The defocusing of the scatterers across steering angles is the main physical principle that we can leverage to estimate the SoS via plane wave ultrasound.

2.2.2 Beamforming Images for Speed-of-Sound Estimation

To leverage the previous principle, we begin by acquiring raw RF data from a range of steering angles. From the data for each steering angle, images can then be beamformed at a range of beamforming SoS values. We denote the specific SoS values used for beamforming as SoS candidates - these beamforming values will be used to test for differences between steering angles. By delay-and-sum beamforming the same set of RF data using many SoS candidates, we can examine the B-mode images to select which SoS candidate best corresponds to the medium SoS.

Our beamformer applies two key modifications to the conventional plane wave delay-and-sum method (Montaldo et al., 2009). Firstly, the effect of SoS mismatch on the steered plane waves must be accounted for. The ultrasound system assumes a SoS to determine the delays necessary for transmission at a given steering angle; however, if there is a SoS mismatch, the actual transmitted angle will differ from the intended steering angle. Secondly, when using different SoS candidates for beamforming, the apparent physical imaging view depends on the SoS – the same structures will appear at shallower or deeper imaging depths in the beamformed images for lower or higher beamforming SoS values,

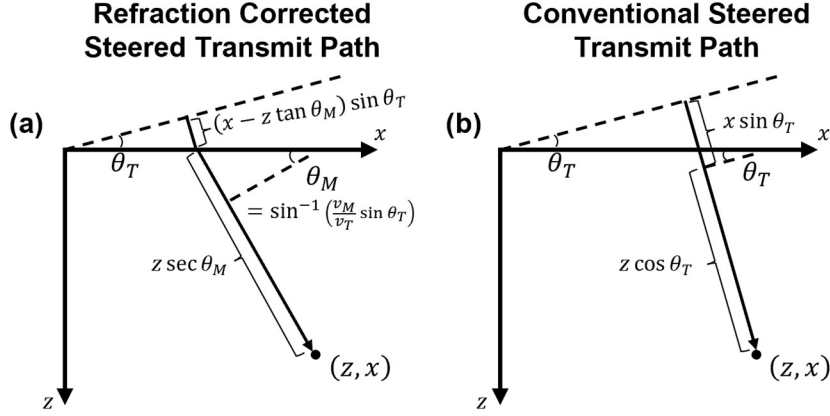


Figure 2.2 Differences in transmit paths for refraction corrected beamforming and standard delay-and-sum beamforming (Montaldo et al., 2009). In (a), the difference between the assumed SoS for a plane wave transmit and the actual medium SoS leads to a refraction effect for the transmit path, compared to the straight transmit path of (b)

respectively (Figure 2.1(a)-(h)). To best evaluate differences between post-beamformed images for each steering angle, these two effects must both be accounted for in the delay-and-sum beamforming.

Compensating for the SoS mismatch on transmit, the steering angle used for beamforming is adjusted based on the interface between the probe surface and the imaged medium. Figure 2.2 highlights the relationship between the intended steering angle and the actual steering angle when there is SoS mismatch. If the medium SoS is higher than the transmit SoS, the steering angle is effectively increased upon transmission (Figure 2.2(a)); likewise, if the medium SoS is lower than the transmit SoS, the steering angle is effectively reduced. Compensating for this effect in delay-sum-beamforming is equivalent to compensating for the effect of refraction between two media, thus we can apply Snell's law at the interface to calculate the theoretical transmitted steering angle as follows (Krucker et al., 2004):

$$v_M \sin \theta_T = v_T \sin \theta_M \quad (2.1)$$

In Eq. (2.1), v_T and θ_T are the assumed SoS and intended transmit steering angle respectively, while v_M and θ_M are the SoS of the imaged medium and the corresponding actual steering angle after refraction. In the delay calculation for beamforming at pixel position (z, x) , we follow the modified transmit for time-of-flight calculation as in Figure 2.2(a), leading to the delay equation:

$$\tau_{TX} = \frac{z \sec \theta_M}{v_M} + \frac{(x - z \tan \theta_M) \sin \theta_T}{v_T} \quad (2.2)$$

The first term in the expression corresponds to the portion of the path that is travelling through the medium, while the second term corresponds to the virtual delay for the transducer to form a steered plane wave. The transmit time-of-flight τ_{TX} of equation (2.2) contrasts with the conventional plane wave delay-and-sum beamforming time-of-flight, which at the same pixel position (z, x) is given instead by (Montaldo et al., 2009):

$$\tau_{TX} = (z \cos \theta_T + x \sin \theta_T)/v_T \quad (2.3)$$

When equation (2.2) is applied to the DAS calculations instead of equation (2.3) for each SoS candidate, we can calculate accurate time-of-flight values to beamform images for each steering angle, assuming the medium SoS is equal to the beamforming SoS.

To adjust for the effect of different beamforming SoS values on the imaging view, the axial positions of the pixels should be adjusted depending on the SoS. In Figure 2.1, the effect of different beamforming SoS values on the apparent point target position was observed. For higher beamforming SoS values (Figure 2.1(h)), the point target appears at a deeper imaging depth. For lower beamforming SoS values (Figure 2.1(b)), the point target appears at a shallower imaging depth. Without restricting the axial imaging view based on the beamforming SoS, new structures can be unintentionally introduced to or removed from the imaging view depending on the SoS candidate. To fairly compare image differences for each SoS candidate, effective imaging view should be the same, regardless of the beamforming SoS. We can align the imaging depth in the following way to ensure the same physical imaging view (though the actual positions of the pixels will not be the same): 1) set an imaging view for a specific beamforming SoS, 2) scale the imaging view for other SoS candidates by the ratio the candidate and the set SoS. If we denote the intended imaging depth at beamforming SoS v_T to be z_T , then the converted position for beamforming SoS v_M is given by the equation:

$$z_T/v_T = z_M/v_M \quad (2.4)$$

For example, if the desired axial field-of-view is 10 mm to 40 mm at 1540m/s, then the axial field-of-view for a SoS candidate of 1400 m/s would be 9.1 mm to 36.4 mm and the axial field-of-view for an SoS candidate of 1700 m/s would be 11.0 mm to 44.2 mm. In this manner, we ensure that no new structures are introduced to or removed from the beamformed images for any of the SoS candidates;

thus, any differences between images beamformed from separate steering angles must be caused by the SoS mismatch.

2.2.3 Estimating the Speed-of-sound through Loss Minimization

Applying the previous principles, images are beamformed for each SoS candidate and each steering angle. Next, we can evaluate the image differences between images from each steering angle. For each SoS candidate, we encapsulate the image differences between angles as a single loss value, from which the SoS that minimizes the loss can be identified. This SoS candidate can then be declared the SoS estimate for the imaged medium.

Formalizing the idea of minimizing the differences, let $\mathcal{L}$ be a loss function that encapsulates image differences. Also let N_θ , N_Z , and N_X respectively be the number of steering angles, axial pixel positions, and lateral pixel positions. Then the input to $\mathcal{L}$ is a $N_\theta \times N_Z \times N_X$ tensor consisting of the complex (after analytic signal conversion) pixel values for each steering angle, while the output of $\mathcal{L}$ is a single real value representative of the image differences. The loss function will be evaluated on all images beamformed using each SoS candidate, resulting in an associated real loss value for each candidate. From these values, a curve can be fitted to evaluate where the minimum loss lies – we denote this loss to be the *inter-steering loss*. We assert that the SoS that minimizes the inter-steering loss is the SoS that also minimizes differences between beamformed images for different steering angles. Based on our analysis of the effect of SoS mismatch (Sec. 2.2.2), we thus declare the SoS leading to the minimal loss to be the SoS estimate produced by our framework.

2.2.4 Full Overview of SoS Estimation Framework

A full detailed overview of the SoS estimation framework is shown in Figure 2.3. We begin by acquiring raw radiofrequency (RF) data using several steered plane waves (Figure 2.3, step 1) – let the number of transmit steering angles be N_θ . After one set of acquisitions, we then have a full $N_{RF} \times N_C \times N_\theta$ set of channel radiofrequency data, where N_{RF} is the number of RF samples per channel and N_C is the number of elements on the probe. Then, an N_S number of SoS candidates ranging from a minimum possible SoS $v_{\min}$ to a maximum possible SoS $v_{\max}$ are selected for beamforming images to calculate the inter-steering loss. For each of the N_S SoS candidates, an $N_Z \times N_X$ image is then beamformed according to Sec. 2.2.2 using that SoS candidate for each of the N_θ steering angles, leading to an $N_Z \times N_X \times N_S \times N_\theta$ data tensor (Figure 2.3, step 2). From the data tensor, the inter-steering loss

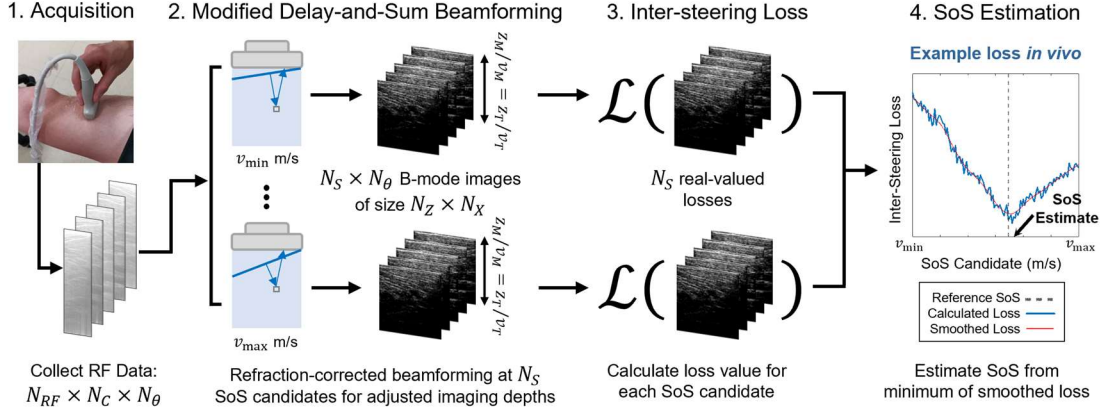


Figure 2.3 Four-step overview for speed-of-sound estimation framework. Data is acquired from several steered plane waves (step 1), then beamformed for many SoS candidates while accounting for refraction and adjusting the imaging depths so that no new structures are introduced in the beamforming view (step 2). The beamformed images are then evaluated for image differences between steering angles via a loss function (step 3). Lastly, the loss function is smoothed and the loss minimum is declared to be the SoS estimate (step 4).

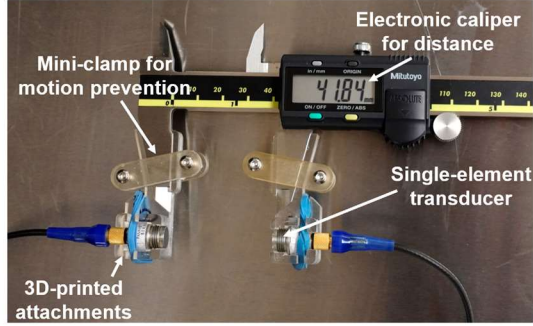
is then calculated for each SoS candidate leading to an $N_S \times 1$ vector (Figure 2.3, step 3), from which a more refined inter-steering loss can be formed via a smoothing curve fitting (Figure 2.3, step 4). After the inter-steering loss has been smoothed (Figure 2.3, step 4, red line), a minimum value can be found (Figure 2.3, step 4, black arrow); the associated SoS leading to that loss is then declared the SoS estimate.

2.3 Experimental Methods

2.3.1 Generating Accurate Reference Speed-of-Sound Measurements with Water Validation

To validate that our proposed algorithm leads to accurate SoS estimates, we measured the SoS of *in vitro* phantoms and *in vivo* human muscles with through transmissions. For reference measurements, we constructed a device (Figure 2.4(a)) to simultaneously measure the distance travelled and the time-of-flight for an ultrasound pulse, allowing for precise calculation of the SoS. For the distance measurement, an electronic caliper (ABSOLUTE; Mitutoyo; Kawasaki, Japan) with 0.01 mm resolution was used to accurately measure the distance. Two custom resin 3D-printed attachments

(a) Reference Setup



(b) Reference Measurement

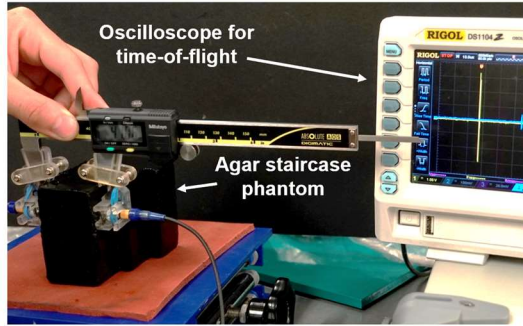


Figure 2.4 (a) Setup for measuring a reference SoS via through transmissions by attaching transducers to an electronic caliper. (b) Using the setup to measure the SoS of an agar staircase phantom.

(Figure 2.4(a)) matching the lower jaws of the caliper were then used to hold two single-element transducers (C567; Olympus; Tokyo, Japan) in alignment with the caliper. One transducer was connected to an arbitrary waveform generator (SDG2042X; Siglent; Solon, OH, USA), emitting a 3 cycle, 10 V peak-to-peak, 5 MHz sinusoidal pulse. Both transducers were also connected to an oscilloscope (DS1104Z; Rigol; Beijing, China) to measure the time-of-flight from one transducer to the other. For automated time-of-flight picking from the acquired oscilloscope data, we applied the Akaike information criterion time-of-flight picker used for USCT (Li, Huang, et al., 2009). The accuracy of our custom SoS measurement setup was validated in distilled water at 22.1 °C which has a known SoS of 1488.6 m/s (Bilaniuk & Wong, 1993); the temperature was measured with an electronic thermometer with 0.4 °C accuracy (Traceable; Webster, TX, USA). In the water, the mean device measurement of the SoS was 1488.9 m/s (SD: 0.3 m/s) with 23 independent measurements at distances ranging from 43.9 mm to 104.9 mm; this level of accuracy was within the temperature measurement

error of the validation system. The custom device was then used to take reference SoS measurements (Figure 2.4(b)) both *in vitro* and *in vivo* to validate the accuracy of the proposed pulse-echo SoS estimation method. Before each new measurement session, the reference measurement setup was calibrated and validated again in water to ensure the correctness of our SoS reference values.

2.3.2 Construction of In Vitro Agar Phantoms

The *in vitro* phantoms were constructed using a mix of agar (2% by weight), graphite (2%), water (64.6%-94%), and glycerol (2%-35.6%). The water and glycerol added up to 96% of the mixture by weight, with higher glycerol concentrations leading to corresponding higher phantom SoS.

To construct *in vitro* phantoms with a homogeneous SoS, three-step staircase molds were first 3D printed out of polylactic acid (PLA), with step heights at 4 cm, 5 cm, and 6 cm. Next, the water and agar were boiled while stirring to form the material for the phantom. After the mixture boiled, glycerol and graphite were added alongside a small amount of potassium sorbate (0.3% of the total mixture weight) as a preservative. This mixture was stirred for five minutes on heat, then transferred to a magnetic stirrer to cool until the mixture reached 50 °C. Lastly, the mixture was poured into the staircase mold and then refrigerated until they were used for the experiments. Using this procedure, we constructed fifteen total phantoms akin to Figure 2.4(b), which enabled robust assessment of the SoS *in vitro* via the reference measurement device. Before any measurements, all phantoms were removed from the refrigerator and left at room temperature overnight to ensure a consistent temperature, minimizing the effects of the SoS of the medium changing during the experiments. The measured SoS of the constructed phantoms ranged from 1508.0 m/s to 1681.9 m/s, with the standard deviation within the reference measurements for each staircase ranging from 0.5 m/s to 1.9 m/s.

2.3.3 In vitro and in vivo Measurement and Data Collection

We collected both *in vitro* and *in vivo* data to assess our proposed SoS estimation framework. First, we used B-mode ultrasound to locate and mark the position on the phantom or the muscle that we intended to measure. Next, pulse-echo plane wave RF data was acquired using the parameters of Table 2.1. Single pulse-cycle steered plane waves were transmitted at 5 MHz, with steering angles ranging from -15° to 15° in intervals of 1°. A SonixTouch research scanner (Analogic Ultrasound; Peabody, MA, USA) with the ability to save raw channel RF data was used for data collection. To establish a reference

TABLE 2.1
ULTRASOUND SCANNER DETAILS AND IMAGING PARAMETERS

Parameter	Details
Ultrasound Probe	L14-5
Number of Tx/Rx Channels	128
Array Pitch	0.3048 mm
Number of Tx Pulse Cycles	1
Transmit Frequency	5 MHz
Transmit Angle Range	-15° to 15°
Transmit SoS Assumption	1540 m/s
Sampling Rate	40 MHz
Pulse Repetition Period	150 μ s

SoS, each phantom or muscle was measured several times with the custom caliper-based setup (Sec. 2.3.1) at the marked location. Lastly, the raw RF data was transferred to a computing server for offline processing of the data with Matlab (2020b; Natick, Natick, MA, USA).

For the *in vitro* staircase phantoms, all phantoms were left outside the refrigerator overnight before any measurements were taken; this process ensured that the phantoms were at a consistent temperature. Each step of each staircase was measured five times with the custom reference setup, then all measurements were averaged to form a reference SoS for each staircase. Afterwards, the top-most step (6 cm step) of each staircase was insonified with steered plane wave transmissions as in Table 2.1 and the raw RF data was saved for later processing. The pulse-echo acquisitions were repeated a total of ten times for each phantom.

For the *in vivo* human data, we acquired data from muscles of 9 healthy volunteers (age: 27.7 ± 2.9 ; 6 males, 3 females). For each volunteer, we examined the gastrocnemius (calf) muscle with both the leg extended and the leg bent. The calf muscle on the dominant side of the body was selected for measurement. To minimize sources of error during SoS measurement, the reference measurements for each muscle were taken immediately following the pulse-echo acquisitions. Starting with the volunteer’s leg in a bent position, ten pulse-echo acquisitions (31 angles per acquisition) were performed without moving the probe position. Immediately afterwards, the muscle was measured seven times with the reference setup with slight changes in position between each measurement to cover the same field-of-view as the linear probe. The volunteer was then asked to extend the muscle and the same quantity of pulse-echo and reference data was acquired at the same location on the muscle as in the bent state. This research data acquisition protocol has been approved by the Research Ethics Board of the University of Waterloo (Protocol No. 44778)

TABLE 2.2
IMAGE BEAMFORMING PARAMETERS

Parameter	Details
F-number	1.4
Apodization	Rectangular with tapered ends
Number of Axial Pixels	128
Number of Lateral Pixels	128
Prefilter Passband	3.5 MHz – 6.5 MHz
Filter Design Method	Equiripple (40 th order)

2.3.4 Beamforming Specifics

From the RF data, we then beamformed the set of images for each SoS candidate according to our modified delay-and-sum beamforming scheme (Sec. 2.2.2). We selected SoS candidates ranging from 1400 m/s to 1750 m/s in 5 m/s intervals; this range encompasses the expected range of SoS for soft tissue in the body. To avoid near-field effects, the axial imaging view was selected to range from 10 mm to 45 mm at 1540 m/s, which was correspondingly scaled for each SoS candidate. We additionally examined using a reduced imaging view ranging from 20 mm to 45 mm at 1540 m/s to mitigate errors in estimation due to thicker regions of subcutaneous fat. For these axial ranges, refraction-corrected delay-and-sum beamforming was then performed according to the parameters summarized in Table 2.2 after the RF data was prefiltered and Hilbert transformed. From each set of steered transmit RF data (size $3120 \times 128 \times 31$), we thus formed a corresponding $128 \times 128 \times 31 \times 71$ complex data tensor consisting of beamformed, analytic images for each SoS estimate. We note that while the axial dimension has been subsampled significantly, the sampled pixels can still be compared across transmit steering angles (as in the subsequent subsection) without increasing the required image formation complexity.

2.3.5 Speed-of-Sound Estimation

Once the images for each SoS candidate for all steering angles have been beamformed, the SoS estimates were formed by minimizing the inter-steering loss (Sec 2.2.3). We examined two pixel-wise loss functions (evaluated on the $128 \times 128 \times 31 \times 71$ complex data tensor) to capture image differences between steered angles, as pixel-wise functions are amenable to GPU-acceleration for real-time operation. The first loss function is based on the coefficient of variation (CV), defined as the ratio of the standard deviation to the mean. For each pixel and each SoS candidate, we calculate the CV across the steering angles, then sum the CV over all the pixels to arrive at a single real value for each SoS

candidate. Expressed symbolically, the CV loss for a specific SoS candidate v is given by the sum over the pixels p :

$$\mathcal{L}_v = \sum_p \frac{\sigma_\theta(p)}{|\mu_\theta|(p)} \quad (2.5)$$

$\sigma_\theta(p)$ is the standard deviation across angles at pixel p and $|\mu_\theta|(p)$ is the mean of the absolute pixel magnitudes across angles. Intuitively, images that match better across angles should have lower standard deviation for each pixel. The second loss function is based on the complex autocorrelation of the pixels across angles. For each pixel and each SoS candidate, we evaluate the autocorrelation function across all steering angles, then take the negative of the absolute discrete integral of the autocorrelation at each pixel position. Expressed symbolically, the autocorrelation loss at a specific SoS candidate v is given by the sum over the pixels p :

$$\mathcal{L}_v = - \sum_p \left| \sum_\theta R_p(\theta) \right| \quad (2.6)$$

$R_p(\theta)$ is the autocorrelation function at pixel p over the steering angles evaluated at lag θ . Intuitively, the pixel-wise autocorrelation across angles should be high if the images match well across

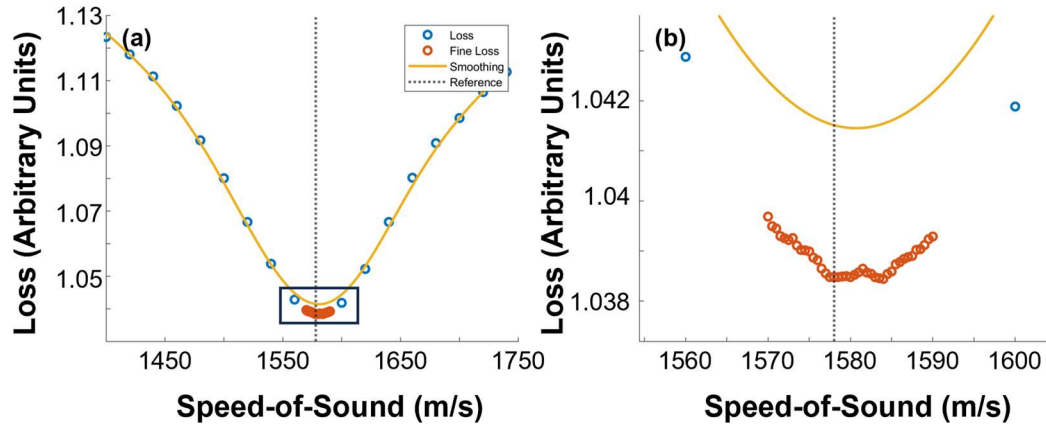


Figure 2.5 (a) Setup for measuring a reference SoS via through transmissions by attaching transducers to an electronic caliper. (b) Using the setup to measure the SoS of an agar staircase phantom.

steering angles. Both of these pixel-wise loss functions (Eq. (2.5) and Eq. (2.6)) permit fast GPU implementations.

When the loss was calculated, only the pixel positions that are fully insonified by all plane wave transmissions were used for the summations. In doing so, we avoided situations where image regions of naturally poor image quality were being compared. The loss was normalized to the number of pixels that were included in the summation. Once the loss was calculated for each SoS candidate, we then fit a normalized smoothing spline in Matlab (using the `fit` function) with a smoothing parameter of 0.99. The minimum of the smoothing spline was selected as the SoS estimate for each sample. A sample calculated loss (blue circles) along with smoothing spline fit (yellow line) for an *in vivo* calf scenario is shown in Figure 2.5. Additionally, a fine-grain loss is calculated (orange circles), demonstrating the benefits of the fitting approach, as a minimization on the fine-grain loss could lead to erroneous estimates.

To evaluate the accuracy of the SoS estimates generated by our framework, we compared the estimates with the through-transmit reference measurements. For the *in vivo* samples, we additionally analyzed the error in the SoS estimates due to the subcutaneous fat layer by measuring the thickness of subcutaneous fat and skin via B-mode imaging. We also performed paired t-tests on both the reference SoS values and the estimated values *in vivo* to assess whether there was a statistically significant difference between the SoS of calf muscle in a bent state versus an extended state (Sec. 2.3.3). Lastly, we assessed how the modified delay-and-sum beamforming calculations (Sec. 2.2.2) affected the accuracy by estimating the SoS without the two modifications (refraction correction, depth alignment) to the beamformer.

For all accuracy evaluations, we quantified the error via the mean signed difference (MSD) and standard deviation between the estimates and the reference measurements. A negative mean signed difference implies our framework underestimates the SoS, while a positive value implies our framework overestimates the SoS; values close to 0 indicate accurate estimation on average. The standard deviation of the mean signed difference indicates the amount of variation in the accuracy of our SoS estimates. We additionally report the maximum intra-sample variability (MIV) as the maximum over all the samples of the standard deviation across the 10 acquisitions for each sample. For the *in vitro* samples where the probe is fixed to a stationary stand, this value indicates the robustness to noise across multiple

acquisitions. For the *in vivo* samples where the probe is held free hand, this value additionally indicates the robustness to operator and tissue motion during the acquisition.

2.3.6 Real-Time Implementation using Sparse Matrices and GPU Acceleration on Portable Scanner

Implementing the SoS estimation procedure in real-time added additional challenges for our algorithm. Firstly, an image for every steering angle must be beamformed for each SoS candidate. Thus, the operation time scales linearly not only with pixels for the DAS beamformer, but also with both the number of steering angles and the number of SoS candidates. Secondly, the loss function evaluation must be performed in an efficient manner after all images have been beamformed. For both of these challenges, we relied on efficient GPU implementations for a real-time realization of our estimation framework.

For real-time operation, the US4R-Lite (US4US; Warsaw, Poland) portable research scanner was connected via USB-C to a GPU-enabled ThinkPad X1 Extreme Gen 4 laptop (Lenovo; Beijing, China). The laptop was running an i9-11950H CPU (Intel; Santa Clara, CA, USA) and an RTX 3080 Laptop GPU (Nvidia; Santa Clara, CA, USA) with 16 GB of VRAM. All programming was done in Python (ver. 3.8.5), with GPU acceleration provided through the CuPy (ver. 10.6.0) library and CUDA (ver. 11.2), interfacing with the US4R Python API (US4US; Warsaw, Poland) for real-time imaging. During operation, SoS estimation was applied as pass-through method; the raw channel RF data was first processed for our custom SoS estimation framework, then the same raw channel RF data was output again for image formation using the built-in US4R beamforming pipeline.

To efficiently beamform the entire data tensor, we applied sparse matrix beamforming (SMB) for GPU acceleration. SMB has demonstrated high efficiency previously for high-intensity focused ultrasound monitoring and passive acoustic mapping but has also proven to be an effective means of acceleration for delay-and-sum beamforming (Hou et al., 2014; Kamimura et al., 2021; Perrot et al., 2021). SMB enabled simple conversion of our modified beamforming algorithm (Sec. 2.2.2) for GPU processing through CuPy. By expressing DAS beamforming as a single sparse matrix multiplication with the vectorized RF data, we could concurrently beamform the images for every SoS candidate and steering angle. Compared to conventional GPU-accelerated DAS beamforming (Yiu et al., 2011), the delay calculations for SMB are instead implicitly encoded in the indices of the sparse matrix instead of being calculated on-the-fly. GPU-based sparse matrix multiplication was implemented via the

`csr_matrix` class in CuPy. For real-time implementation on our portable scanner we additionally reduced the scope of the framework in two ways: 1) reducing the number of transmit steering angles to be -15° to 15° in intervals of 3° instead of 1° , skipping the 0° transmit, 2) reducing the number of SoS candidates used for beamforming – 1400 m/s to 1740 m/s in 20 m/s intervals. We conducted additional SoS estimation experiments to ensure we minimized the loss of accuracy with the framework scope reduction.

To evaluate the inter-steering loss efficiently, we used functions from the CuPy library for GPU-acceleration. Because of the similarity in estimation accuracy from the two loss functions (Eq. (2.5) and Eq. (2.6)), we implemented the coefficient of variation loss (Eq. (2.5)) for real-time operation as it is more computationally efficient. In CuPy, the complex beamformed data was split into real and imaginary portions for computation of the mean and standard deviation for each pixel, matching the Matlab definitions of those functions for complex data. For loss smoothing, the calculated losses for each SoS candidate were transferred to the CPU in order to apply the `UnivariateSpline` function from the SciPy (ver. 1.5.2) library for smoothing (setting $s = 0$) before finding the loss minimum. This loss minimum was then selected as the SoS estimate for the corresponding compounded B-mode image frame. For smoother real-time estimation, the displayed SoS estimate was a running average of the most recent ten estimates.

Videos of live simultaneous B-mode and SoS estimation were captured for three scenarios: 1) a bipartite *in vitro* agar phantom with a slower SoS on one side and a faster SoS on the other side, 2) *ex vivo* porcine meat from the shoulder (fatty cut) and the loin (lean cut), 3) *in vivo* human biceps where the muscle is flexed and relaxed during the acquisition. For scenario 1, the SoS of the two sides were quantified by creating staircase phantoms using the same mixtures as the bipartite phantom. For scenario 2, the SoS was quantified using our custom SoS measurement device. For scenario 3, only qualitative observations were made. These three scenarios served to demonstrate the generalizability and applicability of our SoS estimation framework. Additionally, we measured the computation time during real-time operation for each step of our framework, averaging all the timings over 200 frames.

2.4 Results

2.4.1 High Accuracy of Speed-of-Sound Estimates Compared to Reference Measurements

On the acquired data (Sec. 2.3.3), the estimates generated by our proposed framework are accurate both *in vitro* and *in vivo*, regardless of the applied loss function. The top row of Figure 2.6 shows the SoS estimates compared to the reference values for the *in vitro* agar staircase phantoms for two loss functions (coefficient of variation loss – (a), (c); autocorrelation loss – (b), (d)) at two different beamforming ranges (10-45 mm, left side; 20-45 mm, right side). Substantiating the estimation accuracy for the *in vitro* cases, all framework variations have less than 2.5 m/s mean error. Examining one subplot in greater detail, in Figure 2.6(c), the mean error is -0.6 m/s, while the error standard deviation is 2.3 m/s. For all plots in Figure 2.6, the error is quantified via the mean signed difference (MSD) in the bottom right corner. Also shown in the bottom right corner, the maximum intra-sample

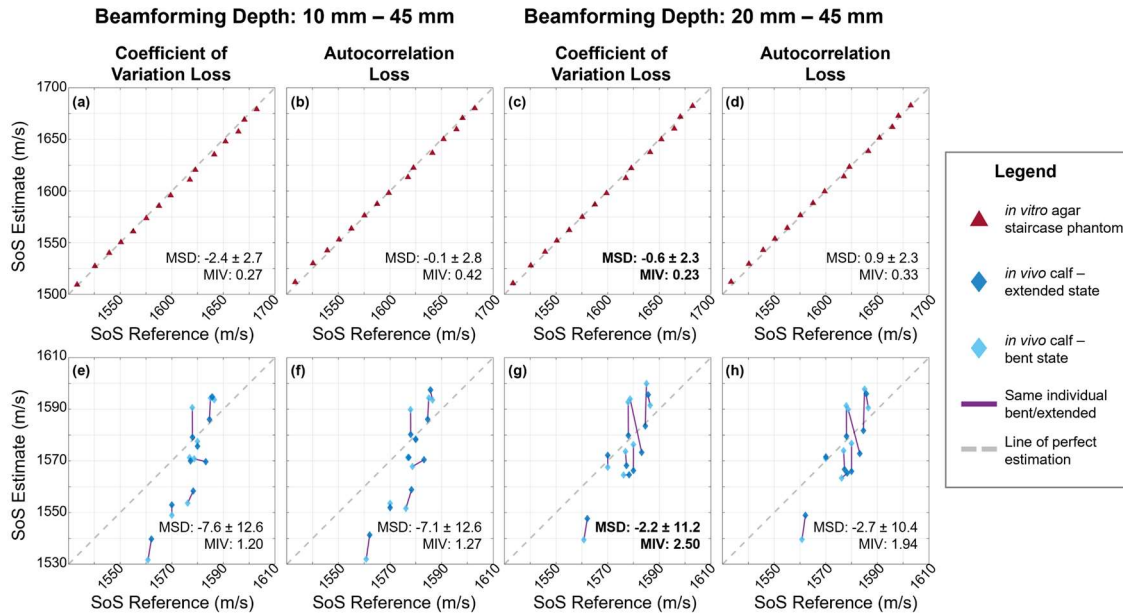


Figure 2.6 Speed-of-sound estimates for the collected ultrasound data as compared to the reference measurements. Two beamforming depths and two loss functions are computed for both *in vitro* agar staircase phantoms (a)-(d) and for *in vivo* human calves in two states (e)-(h). On each plot, mean signed difference (MSD) error and the maximum intra-sample variability (MIV) is reported in m/s. Values in bold represent our baseline framework.

variability (MIV) for all *in vitro* samples was less than 0.5 m/s in all plots. For the more challenging *in vivo* scenarios (Figure 2.6, bottom row), our framework still maintains less than 10 m/s mean error across the different framework conditions. There is minimal difference in the estimates between the two compared loss functions (Figure 2.6; (e) vs. (f), (g) vs. (h)), but a large difference in the estimates between the two beamforming depths for the same loss function (Figure 2.6; (e) vs. (g), (f) vs. (h)). It can be observed that beamforming at depths that excluded the subcutaneous fat layer (Figure 2.6, (g) and (h)) resulted in reduced SoS estimation error compared to beamforming at depths that may include the subcutaneous fat layer for specific individuals (Figure 2.6, (e) and (f)). In the *in vivo* cases, the MIV was higher than in the *in vitro* cases, but did not exceed 2.5 m/s. While the SoS has an accumulative effect in beamforming, the restriction of the considered window leads the loss value to be weighted less heavily for shallower imaging depths. We denote the framework of Figure 2.6(c) and Figure 2.6(g), using coefficient of variation loss and 20-45 mm beamforming depths, as the baseline framework.

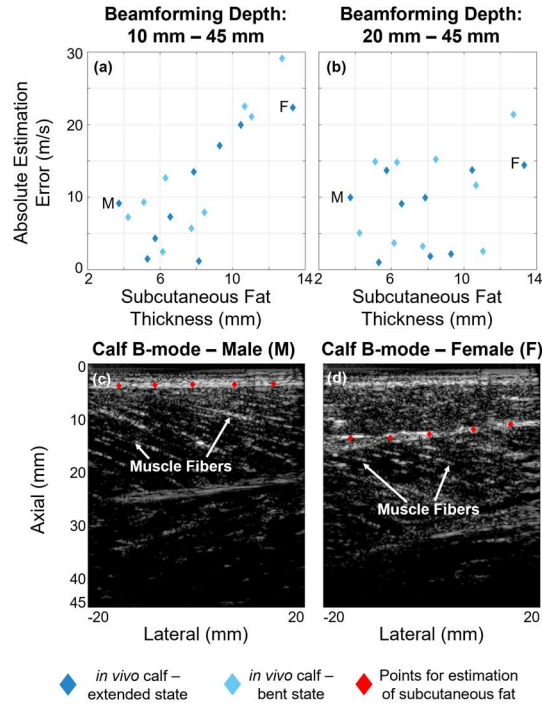


Figure 2.7 Error analysis graphs comparing the absolute SoS estimation error (coefficient of variation loss) versus the measured subcutaneous fat via B-mode imaging for two beamforming depths (a), (b). Figures (c) and (d) show sample B-mode images along with the points used for manual estimation of subcutaneous fat depth – the upper boundary of the muscle fibers.

Further investigating the effect of the subcutaneous fat on the SoS estimates, Figure 2.7 graphs the absolute estimation error compared to the measured subcutaneous fat for coefficient of variation loss for two beamforming depths. In Figure 2.7(a), beamformed at depths from 10-45 mm, a trend can be observed showing that the error increases with subcutaneous fat depth, especially after 10 mm fat depth; quantified, the mean absolute error was 11.8 m/s. In contrast, the baseline framework restricts the beamforming depths between 20-45 mm and led to an error profile less dependent on the subcutaneous fat layer with a mean absolute error of 9.3 m/s. Figure 2.7(c) and (d) show two sample B-mode images of a male subject (M) with a thin subcutaneous fat layer (including skin) and a female subject (F) with a thick subcutaneous fat layer (including skin), alongside the points used to estimate the thickness of the layer in the B-mode images (red diamonds). The points were manually selected to be on the upper boundary of the first region containing muscle fibers (white arrows on Figure 2.7(c) and (d)).

We additionally observed whether there was a difference between bent-leg and extended-leg measurements on the calf SoS (with no muscle activation). Using the baseline framework setup, a paired t-test ($n=9$) on the estimated SoS for the pairs of measurements resulted in the null hypothesis ($p=0.19$). A paired t-test on the reference SoS for the same measurements also resulted in the null hypothesis ($p=0.18$).

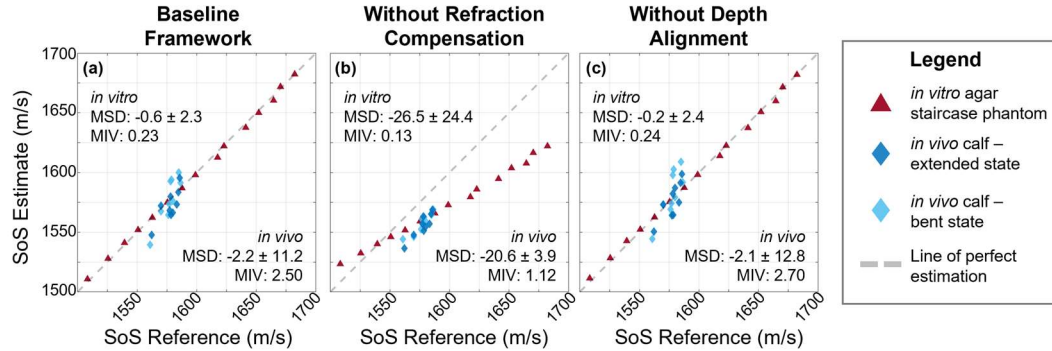


Figure 2.8 Speed-of-sound estimates for the collected ultrasound data for: (a) the proposed framework with coefficient of variation loss, (b) the same framework as in (a), but without refraction compensation, (c) the same framework as in (a), but without depth alignment. On each plot, mean signed difference (MSD) error and the maximum intra-sample variability (MIV) is reported in m/s.

2.4.2 Refraction Correction and Depth Alignment are Important for Accurate Estimates

To assess why modified delay-and-sum beamforming (Sec. 2.2.2) is important for accurate estimates, we separately omitted refraction correction and depth alignment from the baseline framework. Figure 2.8 shows the results of performing SoS estimation without the beamforming modifications. Compared to the baseline SoS estimation framework (Figure 2.8(a)), we observed that eliminating refraction compensation leads to large errors (in vitro MSD: -26.5 m/s, in vivo MSD: -20.6) on the SoS estimates overall (Fig. 7(b)), while eliminating depth alignment reduces the error standard deviation for the *in vivo* estimates only (Fig. 7(c); MSD standard deviation of 12.8 m/s compared to 11.2 m/s baseline). Our baseline framework with the beamforming modifications (Figure 2.8(a)) has the least mean error, along with the lowest error standard deviation, for both the *in vivo* and *in vitro* samples as compared to not applying beamforming modifications.

2.4.3 Framework is Capable of Real-Time SoS Estimation on Portable Open Platform Ultrasound Scanner

For real-time implementation, we decreased the scope of the estimation framework by reducing the number of transmit angles (from 31 to 10) and the number of SoS candidates (from 71 to 18) for beamforming (Sec. III-F). We observed that the reduction in framework scope did not significantly reduce the SoS estimation accuracy (Figure 2.9) compared to the baseline framework. For both the *in*

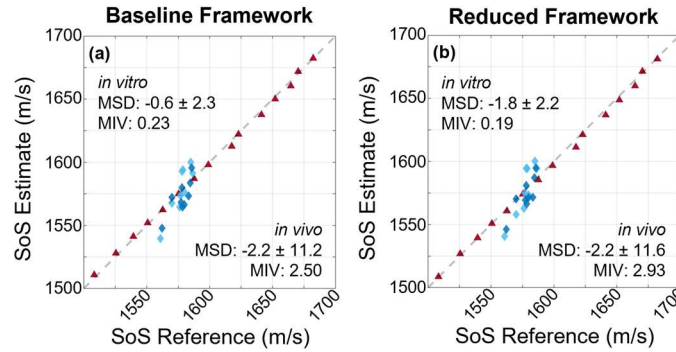


Figure 2.9 Plots comparing the SoS estimation accuracy of the full original proposed framework (a) using 31 steering angles and 71 SoS candidates with a reduced framework (b) with 10 angles and 18 SoS candidates. On each plot, mean signed difference (MSD) error and the maximum intra-sample variability (MIV) is reported in m/s.

vitro and *in vivo* samples, the reduced framework (Figure 2.9(b)) had similar accuracy as the baseline framework (Figure 2.9(a); identical to Figure 2.8(a)). Quantifying the error on the data set, the reduced framework indeed achieves only marginally lower accuracy metrics *in vivo* (MSD: $-2.2 \text{ m/s} \pm 11.6 \text{ m/s}$ vs. $-2.2 \text{ m/s} \pm 11.2 \text{ m/s}$ baseline) as the baseline framework, along with a similar marginal increase in the MIV (2.93 m/s vs 2.50 m/s).

Three videos of live SoS estimation (Appendix C) were taken to demonstrate the applicability and accuracy of our real-time framework (Sec. 2.3.6). In the first video (Movie 2.1), the probe was slid along the length of an *in vitro* bipartite phantom. We observed that when the imaged region was fully contained in only one medium, the SoS estimates closely matched the corresponding measured SoS values. Additionally, the transition of the SoS estimates as the probe was slid from one side of the phantom to the other side was smooth (as in Figure 2.10(a)). In the second video (Movie 2.2), the probe moves between a fattier sample (porcine shoulder – left side) and a leaner sample (porcine loin – right side). Qualitatively, the shoulder has a slower speed-of-sound, while the loin has a faster speed-of-sound (Figure 2.10(b)). Compared to the SoS measurements, our live estimates are within 5 m/s of the reference SoS. In the third video (Movie 2.3), the probe is held in approximately the same position on the right biceps muscle while the individual flexes and relaxes the muscle. We observed that not only

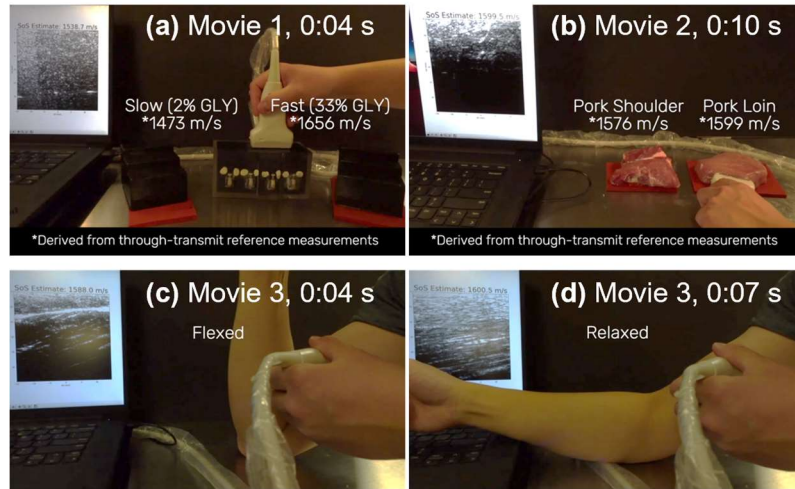


Figure 2.10 Snapshots of Movies 1, 2, and 3 along with corresponding timestamps. (a) Probe is slid in the middle of bipartite phantom, showing SoS estimate is in between reference measurements. (b) Probe is placed on ex vivo porcine loin, a leaner cut of meat. (c) Probe is placed on flexed biceps muscle *in vivo*. (d) Probe is placed on same biceps, but muscle is relaxed.

was there a qualitative difference in the SoS between the flexed (Figure 2.10(c)) and relaxed (Figure 2.10(d)) states, but the displayed estimates were also stable when the muscle was not in the process of transitioning between states. Due to the rounded shape of the muscle, through-transmit measurements were difficult to obtain.

The average framerate during real-time operation was 18.1 fps, measured over 200 frames. Further quantifying this framerate, we measured the average time taken for each section of the estimation algorithm and the built-in image formation and display pipeline. The average preprocessing (prefiltering, Hilbert transform) time for the RF data from all ten steered transmissions was 7.8 ms. The average time taken for beamforming at each SoS candidate and angle was 38.6 ms. The average time for loss calculation and minimization was 1.2 ms. Lastly, the average time taken by the built-in image formation pipeline was 6.1 ms.

2.5 Discussion

2.5.1 A New Algorithm and Device for Real-Time Speed-of-Sound Estimation

In this work, our two main contributions are: 1) a new algorithm for SoS estimation via steered plane waves, and 2) the development of a device capable of simultaneous SoS estimation and B-mode imaging in real-time.

Our new SoS estimation algorithm leverages the defocusing of scatterers across steered plane wave transmissions when the incorrect SoS is used for beamforming (Figure 2.1). By beamforming the same RF data at many SoS candidates, we can find the SoS that minimizes the differences between images for the set steering angles – the minimized SoS is declared the SoS estimate (Figure 2.3). To validate that our new framework resulted in accurate SoS estimates, a custom through-transmit device (Figure 2.4) was used to take reference SoS measurements for custom agar phantoms and for human calves. The average error of the proposed framework, as quantified by the mean signed difference, was -0.6 m/s *in vitro* (Figure 2.6(c)) and -2.2 m/s *in vivo* (Figure 2.6(g)) on the collected data sets, showing a high degree of accuracy compared to the reference measurements. The maximum intra-sample variability was 0.23 m/s *in vitro* and 2.50 m/s *in vivo*, demonstrating our framework’s consistency and robustness.

For real-time application, the proposed framework was implemented on a portable open-platform ultrasound scanner (Sec. 2.3.6). The number of steering angles and SoS candidates used for

beamforming were both reduced to accommodate the device hardware and improve the real-time framerate, without reducing the SoS estimation accuracy (Figure 2.9). Using the US4R-Lite system, we demonstrated concurrent SoS estimation and B-mode imaging at 18 fps for *in vitro* (Movie 1), *ex vivo* (Movie 2), and *in vivo* (Movie 3) imaging scenarios.

Compared to previous works for global SoS estimation, our new framework has a few advantages. Many previous works used focused transmissions for estimating the SoS (Anderson et al., 2000; Krucker et al., 2004; Ophir, 1986). Similar to our framework, one of those works also compares differences between steered transmissions by using image registration for SoS estimation (Krucker et al., 2004). By using plane wave transmissions instead, the real-time potential of the system increases significantly as more than 100 times fewer transmit events are needed to form a single SoS estimate. Requiring only a few transmit events has the added benefit of increasing the system resilience to both patient and operator motion. Despite being a global SoS estimation framework, another advantage of our system is the qualitative demonstration of robustness to inhomogeneous scenarios (Movie 1, Movie 2). When the probe was slid across the middle of the bipartite phantom (Figure 2.10(a)), the SoS estimates smoothly transitioned from one side to the other, even though the imaging view was split between two mediums. When the probe was placed on the porcine shoulder with fat clearly running across the meat sample (Movie 2), the displayed SoS was still stable and matched our reference measurements.

2.5.2 Importance of Live Speed-of-Sound Estimation Framework

The ability to perform point-of-care imaging with ultrasound is a distinct advantage over other medical imaging modalities (Wells, 2006). By performing tissue SoS estimation in real-time, we can better connect the SoS as a biomarker with the B-mode images displayed to the sonographer. For example, in Movie 3, the estimated SoS can be associated with the state of the biceps muscle as it transitions from one state to another. Recently, it has been shown that ultrasound of the biceps using shear wave elastography can help assess rigidity in Parkinson’s disease (Oppold et al., 2023); SoS estimation, like elastography, quantitatively measures a biomechanical property of the muscle tissue. Our work can aid in establishing preliminary investigations for relating SoS with muscle health and function. We envision that one possibility is for tissue speed-of-sound to serve as a one-number metric, similar to how FibroScan summarizes liver stiffness via elastography (Foucher et al., 2006).

For researchers, our work additionally contributes new ideas to the speed-of-sound literature. Firstly, we show the importance of accounting for refraction effects when modeling the effects of SoS on the beamforming paths (Figure 2.8(b)). Local SoS estimation methods can also benefit from refraction correction; potentially increasing the estimation accuracy of those techniques that assume straight-line propagation (Ali, Telichko, et al., 2022; Jaeger et al., 2015; Sanabria et al., 2018). One potential method for local SoS mapping that already applies a form of refraction correction uses the eikonal equation to calculate time-of-flights (Ali, Brevett, et al., 2022). Secondly, we show that global SoS estimation in real time is feasible, and the corresponding real-time insights can be used simultaneously with B-mode imaging. Other global SoS estimation algorithms may also be amenable to real-time implementations; we believe that live imaging will have the most impact for SoS as a biomarker. Lastly, our framework does not depend on any system hardware modifications – any system capable of plane wave transmission and raw RF data collection can use the proposed framework for SoS estimation. Because we used a pass-through method (Sec. 2.3.6), our framework can be inserted into the processing pipeline for any system capable of real-time software beamforming (Boni et al., 2018). Additionally, by passing the live SoS estimate into the beamformer for displaying the image, B-mode image quality can be improved as well, though we did not implement this feature on our system.

2.5.3 Limitations and Future Work

While we achieved our goal of real-time SoS estimation, our work can still be improved via both framework improvements and applicational studies. SoS estimation is highly sensitive to the measured delays and distances. Before each experimental session, calibration of the reference measurement device (Figure 2.4) on the order of 0.01 mm was necessary to ensure accuracy. Similarly, the ultrasound probe’s physical parameters and any delays introduced by the imaging system must be precisely characterized for accurate SoS estimation. An even more precise understanding of all the system and probe parameters would certainly improve the SoS estimation accuracy. Furthermore, the demonstrated frame rate of 18 fps can also be improved; some options include using a custom GPU implementation for the entire framework or another study to assess the minimum number of angles and SoS candidates required for accurate estimation.

Measuring the SoS *in vivo* is challenging for several reasons. For our experimental data set, the calf muscles of the healthy volunteers were almost all within the 1570 m/s to 1590 m/s range, leading to limitations for broader validation of our framework *in vivo*. Measuring the SoS at fine resolution *in vivo*

is also inherently challenging as the SoS depends on the temperature of the tissue at the time of measurement (Sehgal et al., 1986). We are planning a future study involving more muscle groups and more participants to better assess how muscle SoS relates to muscle health and other related parameters. Yet another *in vivo* challenge is how the subcutaneous fat thickness affects the accuracy of our global estimation framework (Figure 2.7). There are two primary ways in which our model is affected. The first is the accumulative effect of the SoS of the medium at shallower depths – because fat tissue has a much slower SoS compared to muscle tissue, our framework will tend towards underestimation of the global SoS. The second is the refraction modelling of the beamforming scheme – an additional refractive effect will physically occur at the fat-muscle boundary. While our real-time framework does demonstrate robustness for inhomogeneous scenarios (Movie 1, Movie 2), separating the SoS values of the fat and muscle layers would be superior for using the tissue SoS as a biomarker to quantify tissue health. Frameworks that have potential for real-time capability for performing SoS estimation layer-wise, region-wise, or locally can be extremely useful for improved quantification of tissue health via the SoS biomarker; deep learning with real-time inference is one option that could potentially be used for real-time local SoS mapping (Feigin et al., 2020). We are currently working on extending the methods from this paper for real-time layer-wise SoS estimation *in vivo* by investigating the loss metrics and estimated SoS in scenarios with two distinct layers with different SoS properties (e.g. *ex vivo* fat-muscle, agar-agar phantoms).

2.6 Conclusion

The ability to estimate tissue speed-of-sound in real-time augments the benefits of ultrasound as a point-of-care imaging modality. In this paper, we proposed a new real-time algorithm for estimating the speed-of-sound using steered plane wave transmissions and implemented it on a portable ultrasound scanner. By creating a device capable of simultaneous speed-of-sound estimation and B-mode imaging, we hope to encourage research regarding the use of the speed-of-sound as a biomarker for tissue health.

2.7 Acknowledgements

The authors would like to thank Forequarter Butcher Shop in Kitchener, ON, Canada for helpful discussions on the meat selection for experiments.

Chapter 3

Sparse Matrix Beamforming: A New Look for Delay-and-Sum in Ultrasound Image Formation

3.1 Introduction

Among medical imaging modalities, ultrasound is notable for its ability to be used at the point-of-care (Whittingham, 2007). Point-of-care ultrasonography can be used for procedural guidance, diagnostic assessment, and screening as ultrasound images are immediately available (Moore & Copel, 2011). These capabilities are facilitated by the real-time nature of ultrasound imaging, which contrasts with other imaging modalities such as CT or MRI. Using an ultrasound probe, a trained clinician directly involved in patient care can identify the most relevant tissue regions without waiting for outside consultation (Díaz-Gómez et al., 2021; Moore & Copel, 2011). Ultrasound researchers developing new imaging techniques must thus be aware of maintaining the real-time advantages of the imaging modality, which open platform ultrasound research systems can enable (Boni et al., 2018). By allowing precise control over the ultrasound transmission and reception process, programmable open platform scanners facilitate researchers' ability to expand the scope of ultrasound research as compared to clinical scanners (Boni et al., 2018).

Recent innovations in medical ultrasound depend on the use of these open platform research scanners. Many of these innovations rely on achieving very high effective imaging frame rates to advance new directions for ultrasound imaging (Tanter & Fink, 2014). The use of high-frame-rate imaging has proven effective for novel ultrasound imaging modes such as transient elastography (Montaldo et al., 2009), vector flow imaging (Jensen et al., 2016a, 2016b), and ultrasound localization microscopy (Errico et al., 2015). The high frame rates required to enable these new imaging modes are realized through broad view transmission schemes, where more of the imaged medium is insonified with each transmission compared to focused beamlines (Tanter & Fink, 2014). These schemes can include emitting acoustic plane waves or diverging waves (Montaldo et al., 2009; Tanter & Fink, 2014), simultaneously transmitting multiple focused beams as in multi-line transmission (Mallart & Fink, 1992), or firing transducer elements individually for synthetic aperture imaging (Jensen et al., 2006). Common to all the new imaging modes is the dependence on delay-and-sum (DAS) beamforming for image formation (Mallart & Fink, 1992; Montaldo et al., 2009; Perrot et al., 2021). DAS forms

ultrasound images by performing spatiotemporal localization on the RF data received by the transducer elements by coherently summing appropriately delayed data from each channel at each pixel position (Perrot et al., 2021).

To maintain the real-time advantage of ultrasound on open-platform scanners, researchers are invested in accelerating software-based DAS beamforming (Boni et al., 2018; So et al., 2011). For example, general purpose GPU programming for acceleration of DAS via parallelization is one popular solution for achieving real-time plane wave and synthetic aperture ultrasound imaging (Yiu et al., 2011). One challenge is that optimizing the code for custom CUDA kernels on the GPU is a nontrivial task and requires specific expertise (Romero-Laorden et al., 2016; Yiu et al., 2011). An alternative to GPUs for DAS acceleration is to use field-programmable gate arrays (FPGAs) to perform image formation. While FPGA-based beamformers have also demonstrated real-time capabilities for real-time image formation on open platform scanners, they demand even more expertise for optimal implementations (Boni et al., 2017; N. A. Campbell & Brown, 2022; Kou et al., 2023). Libraries with ultrasound beamforming capabilities have been made available in recent years, but may not be suitable for adaptation for more specific applications (Garcia, 2021; Hyun et al., 2019; Rodriguez-Molares et al., 2017). Thus, an easily accessible, yet still customizable software-based DAS beamformer would be beneficial for the research community for the testing of real-time ultrasound algorithms.

In this work, we pedagogically focus on the newer technique of sparse matrix beamforming (SMB) for acceleration of ultrasound image formation (Hou et al., 2014). SMB has proven to be effective not only for B-mode imaging (Perrot et al., 2021), but also for high-intensity focused ultrasound monitoring (Hou et al., 2014), passive cavitation mapping (Kamimura et al., 2021), and speed-of-sound estimation (Chapter 2). SMB relies on the fact that each beamformed pixel results from the linear combination of samples from the raw radiofrequency data; thus, the beamforming operator can be expressed as a matrix multiplication (Hou et al., 2014). While the use of SMB has been effective for specific applications, the causes and extent of the beamforming acceleration from the sparse matrix formulation have not been investigated deeply. In this paper, we review the steps for forming the sparse beamforming matrix, then show that the beamforming matrix can readily be applied in common research frameworks based on Matlab, Python, or CUDA C++. We provide code samples and additionally evaluate the degree of acceleration provided by SMB in these frameworks using a minimal working beamformer. Next, we demonstrate how SMB can be employed for different ultrasound transmit schemes such as conventional focused imaging or coherently compounded plane wave imaging (Montaldo et al., 2009). Lastly, as

SMB encodes the DAS beamforming framework into a matrix multiplication, it is also primed to support integration of DAS beamforming into the most popular machine learning frameworks of TensorFlow and PyTorch. We demonstrate the viability of using SMB to design beamforming layers embedded in neural network architectures. Overall, we believe that SMB is a tool that is well-positioned to support ultrasound researchers for testing real-time imaging innovations.

3.2 Theory

3.2.1 Delay-and-Sum Beamforming as a Sparse Matrix Multiplication

Delay-and-sum (DAS) beamforming is the predominant image formation algorithm for biomedical ultrasound imaging. While DAS is typically performed by iterating through the pixels, it can also be encoded as a single sparse matrix multiplication. Encoding DAS as a precalculated sparse matrix representing an alternative approach to DAS with potential computational benefits. DAS functions by performing spatiotemporal localization of the data received after a pulse-echo transmission (Perrot et al., 2021). In pulse-echo ultrasound, an acoustic pulse is first transmitted by the probe; the acoustic waves then propagate through the imaged medium and are partially reflected by acoustic scatterers. These reflections are then received by each transducer element on the probe and are used to form the received data matrix, denoted the raw radiofrequency (RF) data. After preprocessing the RF data, time-of-flight calculations based on the transmit and receive paths to each pixel position and each transducer element are then used to delay, then coherently sum RF data from each channel to calculate the pixel magnitudes. Repeating this procedure for each possible pixel position results in the final beamformed ultrasound image.

Formalizing the DAS procedure, let the preprocessed (analog-to-digital conversion, bandpass filter, analytic signal conversion) RF data matrix be represented by the $N_S \times N_C$ sized matrix $\mathbf{D}$ where N_S, N_C are the sampling depth (number of samples) and the number of channels respectively. Then, let the appropriately delayed signal (with sample delay Δ) from each channel e to form a specific pixel $p = (z, x)$ of an $N_Z \times N_X$ (axial pixels by lateral pixels) image be represented as

$$s(p, e) = \mathbf{D}(\Delta(p, e), e) \quad (3.1)$$

With the delay step of DAS applied, the summation step across channels can then be expressed as a weighted summation which results in the intensity $\mathbf{I}(p)$ at a given pixel:

$$\mathbf{I}(p) = \sum_{e=1}^{N_C} \alpha(p, e) s(p, e) \quad (3.2)$$

We note that the apodizations weights $\alpha(p, e)$ correspond to applying receive apodization to the DAS procedure (Hoen, 1982) but represented in a more generalized form (one weight factor per channel per pixel). We also note that the selection of an F-number for receive beamforming can be considered a form of apodization with weights set to zero outside of the aperture (Montaldo et al., 2009; Perrot et al., 2021). The typically seen grayscale B-mode ultrasound images result from log-scaling and envelope detection of the intensity matrix $\mathbf{I}$.

DAS beamforming is fundamentally an operation that maps the $N_S \times N_C$ RF data matrix $\mathbf{D}$ to the $N_Z \times N_X$ image matrix $\mathbf{I}$. In the 2D-to-2D domain, the selection of RF data varies based on the imaging depth and the speed-of-sound and is not shift invariant (Perrot et al., 2021). Changing our perspective by vectorizing both the RF data matrix and the image matrix, the DAS operation of Eq. (3.2) becomes exactly a linear map from the RF data space to the pixel intensity space (Hou et al., 2014); each beamformed pixel is the result of a linear combination of samples from the entire set of RF data. Because the relationship is linear, there exists a matrix representation of DAS beamforming where the pixels result from the multiplication of the DAS matrix with the vectorized RF data [Hou 2014]. Through Eq. (3.2) this matrix can be precisely calculated; however, the matrix (sized $N_Z N_X \times N_S N_C$) is expensive in terms of both storage and operational efficiency. For example, assume we desire to beamform a 256×256 sized image from a 3120×128 sized RF data matrix. Then a single-precision beamforming matrix would be of size 65536×399360 , requiring 104.7 GB of memory and would be correspondingly expensive to apply. Because of these considerations, expressing DAS as a simple dense matrix multiplication would be computationally intractable.

We can resolve these issues with the beamforming matrix by recognizing the inherent sparsity present in the matrix, an approach substantiated by several previous works (Besson et al., 2018; Hou et al., 2014; Ozkan et al., 2018; Perrot et al., 2021). For the purposes of pedagogy, the vector manipulations required for the construction of these matrix are reviewed. In the dense matrix scenario, many operations are wasted through zero-multiplications as each pixel only requires one sample from each channel for summation. Each row of the matrix thus has a sparsity level of just $\frac{1}{N_S}$, which for most open platform ultrasound systems results in less than 0.05% nonzero values. To construct the sparse matrix,

we re-examine Eq. (3.1) and (3.2) and express them in a form amenable to matrix conversion. Eq. (3.1), where the RF data is selected, can be expressed as a dot product of the RF data vector (from a single element e) with a naturally sparse indicator vector as follows:

$$s(p, e) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{D}(1, e) \\ \vdots \\ \mathbf{D}(\Delta(p, e) - 1, e) \\ \mathbf{D}(\Delta(p, e), e) \\ \mathbf{D}(\Delta(p, e) + 1, e) \\ \vdots \\ \mathbf{D}(N_S, e) \end{bmatrix} \quad (3.3)$$

Substituting this expression into Eq. (3.2), we obtain the sum of $N_S \times 1$ inner products:

$$\mathbf{I}(p) = \sum_{e=1}^{N_C} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha(p, e) \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{D}(1, e) \\ \vdots \\ \mathbf{D}(\Delta(p, e) - 1, e) \\ \mathbf{D}(\Delta(p, e), e) \\ \mathbf{D}(\Delta(p, e) + 1, e) \\ \vdots \\ \mathbf{D}(N_S, e) \end{bmatrix} \quad (3.4)$$

The sum of inner products can also be expressed as a single inner product; in this way, each beamformed pixel results from a single inner product of $N_S N_C \times 1$ vectors. As these inner products are all multiplied with the same RF data vector, the sparse matrix can be constructed by stacking the inner products to form the $N_Z N_X \times N_S N_C$ beamforming matrix. The full matrix expression for DAS beamforming the sized $N_Z N_X \times 1$ image vector $\vec{\mathbf{I}}$ is thus:

$$\vec{\mathbf{I}} = \begin{bmatrix} \vec{\alpha}(1,1)^T & \vec{\alpha}(1,2)^T & \cdots & \vec{\alpha}(1,N_C)^T \\ \vec{\alpha}(2,1)^T & \vec{\alpha}(2,2)^T & \cdots & \vec{\alpha}(2,N_C)^T \\ \vdots & \vdots & \ddots & \vdots \\ \vec{\alpha}(N_X N_Z, 1)^T & \vec{\alpha}(N_X N_Z, 2)^T & \cdots & \vec{\alpha}(N_X N_Z, N_C)^T \end{bmatrix} \begin{bmatrix} \mathbf{D}(1,1) \\ \vdots \\ \mathbf{D}(N_S, 1) \\ \mathbf{D}(1,2) \\ \vdots \\ \mathbf{D}(N_S, 2) \\ \vdots \\ \mathbf{D}(1, N_C) \\ \vdots \\ \mathbf{D}(N_S, N_C) \end{bmatrix} \quad (3.5)$$

where $\vec{\alpha}(p, e)$ represents the sparse indicator vector at pixel index p for RF channel index e . Expressed more succinctly, let $\mathbf{S}$ be the sparse beamforming matrix for a given transmit scheme and set

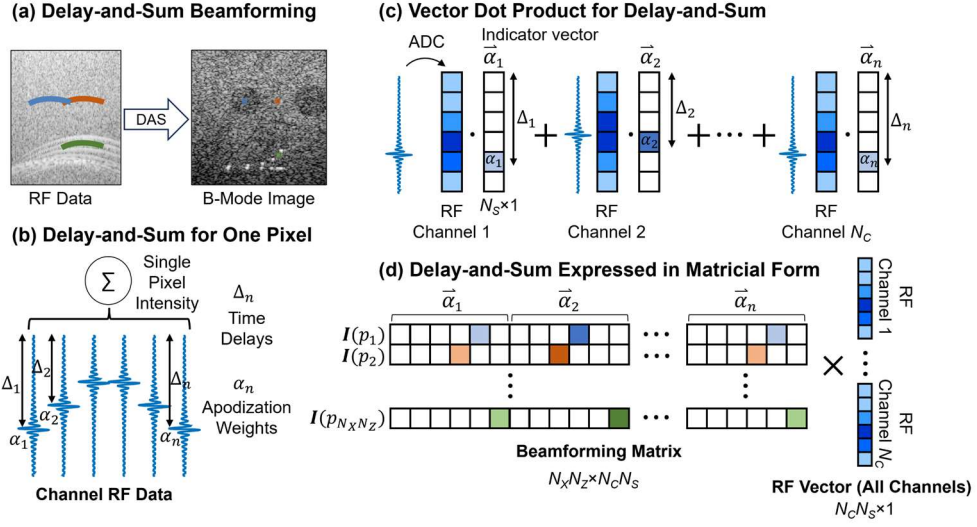


Figure 3.1 Pictorial depiction of formation of sparse beamforming matrix. (a) Delay-and-sum (DAS) is the process that forms B-mode images from RF data. (b) Each beamformed pixel is the result of coherently summing appropriately delayed samples from the RF data. (c) The delay-and-sum procedure can be encoded as a sum of dot products between each channel of RF data and a sparse indicator vector. (d) Since the same RF data is used for each pixel, the dot products for each pixel can be stacked to form rows of the beamforming matrix.

of pixel locations (left matrix in Eq. (3.5)), then the vectorized DAS image $\vec{I}$ is equal to the multiplication of $\mathbf{S}$ with the vectorized channel RF data $\vec{D}$:

$$\vec{I} = \mathbf{S} \vec{D} \quad (3.6)$$

The formation of the sparse DAS matrix is summarized in pictorial form in Figure 3.1. Figure 3.1(a) depicts that DAS transforms channel RF data in B-mode images, where each pixel is formed by summing the appropriately delayed RF data across the channels (matching colors in Figure 3.1(a)). For a single pixel (Figure 3.1(b)), the summation is performed with apodization weights α_n after the channels have been delayed by Δ_n to align coherent signals. For that pixel, the DAS procedure can be reformulated as a sum of dot products (Figure 3.1(c)), where the inner product between each channel of RF data (blue column vector) is taken with a very sparse indicator vector encoding both the appropriate delay for that pixel and the apodization weight used for summation. Because the dot products are being summed together, these indicator vectors can be combined to form a single row of

the sparse beamforming matrix (Figure 3.1(d)), corresponding to the DAS procedure for that pixel. By repeating this process for all pixels in the image, the entire beamforming matrix can be formed. Incorporation of linear interpolation for more accurate discrete RF sample selection can also be encoded using the sparse indicator vector (Yiu et al., 2011). For this case, the indicator vector $\vec{\alpha}(p, e)$ would be 2-sparse instead of 1-sparse and encode both the linear interpolation weights and the apodization weights (see Appendix A). In this way, for a single frame, the entire DAS operation can thus be encoded as a single sparse matrix multiplication with a dense RF vector.

3.2.2 Extending Sparse Matrix Beamforming for Batch Processing and Coherent Compounding

For beamforming a single frame from a single transmit event, the formulation in the previous section is appropriate; however, the construction of the sparse beamforming matrix can naturally be extended for simultaneously beamforming either multiple frames as a batch, or for changes in the transmit scheme. This extension has a few benefits. Firstly, by expressing the matrix multiplication as a batch, real-time systems can operate on a buffered structure instead of only on a frame-by-frame basis, which reduces computational overhead. Secondly, by handling changes in the transmit scheme, such as adjusting the steering angle or transmit center, real-time systems can use a single precalculated matrix for all operations, as opposed to using and storing many different matrices.

Let us first consider the case of batched transmissions. If the transmit scheme does not change frame-to-frame, as in the case of repeated single plane wave transmissions, then the sparse matrix-vector multiply operation (Eq. (3.5)) can readily be performed batchwise. Because the transmit scheme is constant, the same sparse matrix can be re-used to beamform each subsequent frame; because the beamforming remains the same, each vectorized RF frame can simply be concatenated to form a matrix. This batched RF matrix is then of size $N_S N_C \times N_B$, where N_B is the number of frames in each batch. Performing the sparse matrix-matrix multiply results in another matrix of size $N_Z N_X \times N_B$, where the columns of the matrix are the vectorized image frames. Expressed as an equation, let $\vec{I}_i, \vec{D}_i$ be the vectorized image and corresponding vectorized RF data from batch i respectively, then:

$$\begin{bmatrix} \vec{I}_1 & \vec{I}_2 & \cdots & \vec{I}_{N_B} \end{bmatrix} = \mathbf{S} \begin{bmatrix} \vec{D}_1 & \vec{D}_2 & \cdots & \vec{D}_{N_B} \end{bmatrix} \quad (3.7)$$

Correctly reshaping the resultant matrix after multiplication gives each beamformed image frame. From a linear algebra perspective, this batching process is apparent and results in the same number of

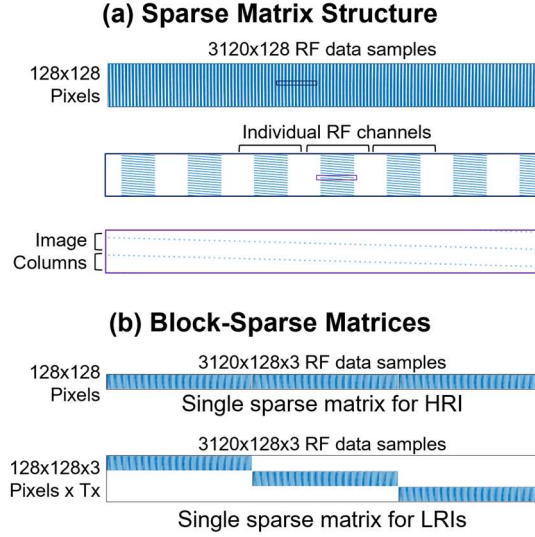


Figure 3.2 (a) Example structure of sparse matrix for single transmit event. Successively zoomed snippets of the sparse matrix structure are shown in dark blue and purple outlines. (b) Example block-sparse matrix to perform beamforming for multiple transmit events. Block-diagonal matrix for LRI beamforming and fat matrix for HRI beamforming are shown.

operations performed. However, the computational advantage to considering batched matrix-matrix operations over matrix-vector operations is an increase in algorithm efficiency (example provided in Sec. 3.3.3). By expressing the operation as a matrix-matrix multiplication, improved threading and use of memory is achieved (e.g. loading sparse matrix coefficients only once for multiple frames).

Now consider when the transmission scheme does change RF frame-to-RF frame, as in full synthetic aperture imaging or multi-angle plane wave imaging. As an example, let us examine the case of steered plane wave imaging for compounding. First, multiple steered plane waves are used to insonify the imaged medium. After reception, the corresponding received raw RF data for each plane wave is formed by the imaging system. Typically, single plane wave data are beamformed individually to form low-resolution images (LRIs), which are then compounded to form one single high-resolution image (HRI) from the whole set of steered transmissions. Using the sparse matrix formulation for DAS, we can form a sparse beamforming matrix for each transmit event. An example beamforming matrix structure is shown in Figure 3.2(a) for a single unsteered plane wave transmission. Examining the matrix structure more deeply reveals how individual RF channels and pixels are related in the rows and columns. While these individual sparse matrices can be used for beamforming each frame of RF data separately, they

can also be combined into a single sparse matrix as the entire DAS process is linear. This large sparse matrix can be formed in block-wise fashion either left-to-right (Figure 3.2(b) top) or block diagonally (Figure 3.2(b) bottom). Then, by vertically concatenating the RF vectors for each transmit event into a single long vector, multiplication of the large sparse matrix results in the HRI vector or the stacked LRI vectors respectively. Letting the number of unique transmit events be N_T , the HRI sparse beamforming matrix is thus sized $N_Z N_X \times N_S N_C N_T$, while the LRI sparse beamforming matrix is sized $N_Z N_X N_T \times N_S N_C N_T$ with block diagonal structure. The number of nonzero entries for both these matrices are identical. Expressed as an equation, let $\vec{I}_i$, $\vec{D}_i$, and $\mathbf{S}_i$ be the vectorized LRI, vectorized RF data, and sparse beamforming matrix corresponding to the i th transmit respectively. Then, the HRI beamforming matrix can be expressed as:

$$\vec{I}_1 + \vec{I}_2 + \dots + \vec{I}_{N_T} = [\mathbf{S}_1 \quad \mathbf{S}_2 \quad \dots \quad \mathbf{S}_{N_T}] \begin{bmatrix} \vec{D}_1 \\ \vec{D}_2 \\ \vdots \\ \vec{D}_{N_T} \end{bmatrix} \quad (3.8)$$

Similarly, the LRI beamforming matrix can be expressed as:

$$\begin{bmatrix} \vec{I}_1 \\ \vec{I}_2 \\ \vdots \\ \vec{I}_{N_T} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{S}_{N_T} \end{bmatrix} \begin{bmatrix} \vec{D}_1 \\ \vec{D}_2 \\ \vdots \\ \vec{D}_{N_T} \end{bmatrix} \quad (3.9)$$

Where $\mathbf{0}$ represents an $N_Z N_X \times N_S N_C$ all-zeros matrix block (example shown in Figure 3.2(b)). In these two ways, a single large, sparse matrix can represent the entire DAS beamforming procedure for a sequence of transmits, which also reduces computational overhead (example provided in Sec. 3.3.4). The computational advantage of encoding the entire beamforming operation in a single matrix lies in the threading performance. From a computing perspective, when individual matrices are used naively for beamforming, the program flow is blocked until the matrix multiplication is complete, leading to greater thread inefficiency.

These two ideas of batch processing and block-sparse matrices for transmit changes can also be applied simultaneously. If the received RF data is structured to repeat the same sequence of multiple transmits (e.g. vector Doppler imaging), then combining batch processing and block-sparse matrices can be considered to minimize processing overhead.

3.2.3 Efficient Sparse Matrix Multiplication using Compressed Sparse Row Format

There are many methods for multiplying sparse matrices with dense vectors or matrices. Among these methods, one commonly implemented, and relatively efficient, algorithm is the compressed sparse row (CSR) algorithm (Bell & Garland, 2008). The CSR format of a sparse matrix uses the column indices and values of each nonzero entry to store the matrix while compressing the row indices. In contrast, the coordinate (COO) format, which is the simplest method, will store a matrix using all the column indices, row indices, and values for each nonzero entry. Compressing the row indices has two effects – not only does it reduce storage requirements in memory, but it also allows for the CSR matrix multiply to efficiently loop through the rows of the multiplying sparse matrix (Bell & Garland, 2008). Additionally, the CSR algorithm is highly amenable to acceleration via parallelization. In the subsequent example applications, we will be either explicitly converting to the CSR format or implicitly relying on the prebuilt sparse matrix library to perform fast multiplication.

3.3 Example Applications of Sparse Matrix Beamforming

In the following subsections, we pedagogically demonstrate SMB using select example applications. The examples serve to illustrate ways in which SMB can be applied to a variety of common ultrasound research scenarios.

3.3.1 Data Acquisition and System Parameters for All Example Applications

To demonstrate all sample applications of SMB, we use data acquired from a physical scanner in the subsequent sections. The data acquisition system used was a research scanner equipped with a DAQ (SonixTouch; Analogic Ultrasound; Peabody, MA, USA) using a 40 MHz sampling rate. An L14-5 linear probe (Analogic Ultrasound; 128 Tx/Rx channels; 0.3048 mm pitch) was used for all acquisitions. All data was collected from a CIRS point target phantom (Multi-Purpose Multi-Tissue Ultrasound Phantom; Computerized Imaging Reference Systems, Inc., Norfolk, VA, USA). The beamforming procedure does not change depending on the imaged medium and SMB is numerically equivalent to delay-and-sum beamforming, so all the methods shown can also be applied to *in vivo* imaging scenarios. For the beamforming implementations, a computer system containing an Intel i7-6700K CPU (Intel, Santa Clara, CA, USA) and an Nvidia GTX 1080 GPU (Nvidia, Santa Clara, CA, USA) were used to evaluate the performance.

3.3.2 Single Plane Wave Delay-and-Sum Beamforming via Sparse Matrices

The first example demonstrates DAS beamforming on data corresponding to a single zero-degree plane wave transmission. This simple example serves two purposes; firstly, we can demonstrate how a sparse beamforming matrix is constructed based on conventional DAS code; secondly, we can evaluate and compare the image reconstruction times.

To illustrate SMB in a pedagogical way, we first developed a minimum-viable beamformer (Code Sample 3.1). The beamforming function is programmed in Matlab (ver. 2023b; MathWorks, Natick, MA, USA). This beamformer performs DAS imaging for an unsteered plane wave transmit; the minimum-viable beamformer does not consider handling sample interpolation, out-of-index errors, apodization, and assumes a linear array geometry and rectangular pixel grid. The plane wave RF data (`rf_data`) is stored as a single 2D-array (consisting of 3120 samples by 128 channels). The sampling rate (`fs`), speed-of-sound (`sos`), and receive delay (`rx_delay`; time of transmit relative to first received sample) are provided as scalar variable inputs. The transducer element positions (`pos_trans`), axial pixel positions (`pos_z`), and lateral pixel positions (`pos_x`) are provided as 1D-array inputs. The total time-of-flight (`total_time`) is calculated for each pixel, discretized into the best sample indices

Code Sample 3.1 – Minimum-Viable Delay-and-Sum Beamformer (Matlab)

```
function [img] = bmfrm_min(rf_data,fs,sos,rx_delay,pos_trans,pos_z,pos_x)
Nz = length(pos_z);
Nx = length(pos_x);
Nc = size(rf_data,2);

img = zeros(Nz,Nx);

for z = 1:Nz
    for x = 1:Nx
        total_time = (pos_z(z) + sqrt(pos_z(z)^2 + (pos_x(x) - pos_trans).^2))/sos;
        samples = round(fs*(total_time-rx_delay));
        for c = 1:Nc
            img(z,x) = img(z,x) + rf_data(samples(c),c);
        end
    end
end
```

Code Sample 3.2 – Minimum-Viable Sparse Matrix Formation (Matlab)

```
function [sp_mat] = bmfrm_smb_gen(rf_data,fs,sos,rx_delay,pos_trans,pos_z,pos_x)
Nz = length(pos_z);
Nx = length(pos_x);
Nc = size(rf_data,2);
Ns = size(rf_data,1);

s_row = zeros(1, Nz*Nx*Nc);
s_col = zeros(1, Nz*Nx*Nc);
s_val = ones(1, Nz*Nx*Nc);

count = 1;
for z = 1:Nz
    for x = 1:Nx
        total_time = (pos_z(z) + sqrt(pos_z(z)^2 + (pos_x(x) - pos_trans).^2))/sos;
        samples = round(fs*(total_time-rx_delay));
        for c = 1:Nc
            s_row(count) = z + Nz*(x-1);
            s_col(count) = samples(c) + Ns*(c-1);
            count = count + 1;
        end
    end
end
sp_mat = sparse(s_row,s_col,s_val,Nz*Nx,Ns*Nc);
```

(samples), and used for coherent summation into each image pixel ($\text{img}(z,x)$). In this way, a single beamformed image can be formed from the RF data by appropriately calling the function.

Based on this minimum-viable DAS beamformer, we can also convert the code to generate a sparse matrix instead of a beamformed image. After calculating the best sample, instead of selecting the RF data for coherent summation, we generate the row and column indices to form the sparse beamforming

Code Sample 3.3 – Sparse Matrix Beamforming (Matlab)

```
function [img] = bmfrm_smb(rf_data,sp_mat,Nz,Nx)
img_vec = sp_mat * rf_data(:);
img = reshape(img_vec,[Nz Nx]);
```

matrix. Sparse matrix construction is demonstrated through Code Sample 3.2, while applying the sparse matrix for beamforming is demonstrated through Code Sample 3.3. In Code Sample 3.2, the structure of the code is the same as in Code Sample 3.1 except for the inner-most loop. In that loop, the row and column indices are stored instead of performing the summation operation. While not pertaining to the minimum-viable beamformer where apodization is precluded, the inner loop is also where the apodization weights would be saved (as opposed to using an all-ones vector for the values at the row and column indices). We note that for apodizations where the support is limited (e.g. applying an F-number), SMB increases in efficiency due to ignoring the zero-valued weights. Once the sparse matrix is formed, it can then be used for beamforming. Code Sample 3.3 gives an example of how the sparse matrix can be applied to perform DAS; first the RF data is vectorized, then the sparse matrix multiplication is performed. After multiplying, the image remains vectorized and must then be appropriately reshaped to reform the 2D beamformed image array. This image array should exactly match the result of beamforming using Code Sample 3.1. The vectorize-multiply-unvectorize procedure is the core of SMB.

Using these three code samples as a basis, we can evaluate the difference in computation times in Matlab (ver. 2023b). Moreover, we developed similar code examples with the minimum-viable beamformer in Python (ver. 3.11.4) using NumPy (ver. 1.24.3) and SciPy (ver. 1.10.1) to show the language-independent benefits of SMB. Instead of generating the sparse matrix in Python, the sparse beamforming matrix was generated in Matlab was then saved and loaded into Python to demonstrate the flexibility of applying SMB cross-language. Timings were evaluated for image sizes ranging from 128×128 pixels to 1024×1024 pixels. For all timing experiments, each beamforming step was repeated 100 times on the same set of RF data, then averaged to find the mean beamforming time. CPU-only timings were evaluated for the minimum-viable DAS beamformer and corresponding SMB function in both Matlab and Python. Given that sparse matrix multiplication has natural GPU support in Matlab and Python, we leveraged this advantage to develop GPU-based SMB with minimal additional effort. In Matlab, we simply converted the relevant arrays into `gpuArrays`; in Python, we used the CuPy library (ver. 12.2.0) which is nearly a one-to-one conversion from the NumPy and SciPy libraries. The GPU backend for both Matlab and Python was the Nvidia CUDA framework (ver. 12.2). Thus, we were also able to evaluate the timings on GPU for SMB. In Matlab, timings were computed using the `tic-toc` functionality for CPU timings, and the `gputimeit` function for GPU timings. In Python, timings were

computed using the `time_ns` function with synchronization performed prior to measurement for GPU timings.

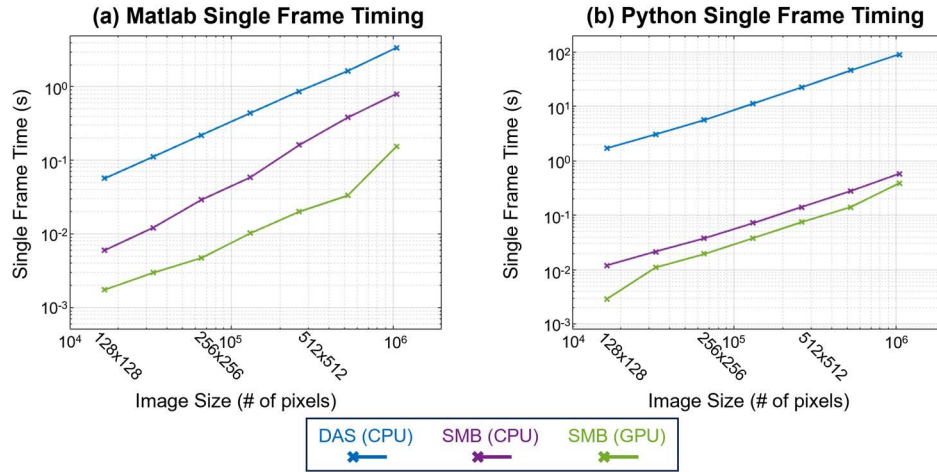


Figure 3.3 (a) Timing for three compared beamforming implementations in Matlab. (b) Timings for same three compared beamforming implementations in Python.

Figure 3.3 shows the results of these timing experiments in Matlab (Figure 3.3(a)) and Python (Figure 3.3(b)). For both log-log plots, we observed that the timing as the number of pixels increases approximately follows a linear trend for both DAS and SMB. This trend is expected for DAS as it is $O(N_Z N_X N_C)$ which is evident through the triple-nested for-loop structure in Code Sample 3.1. Similarly, the linear trend is expected for SMB given that sparse matrix multiplication scales as the number of nonzeros in the matrix; the number of nonzero values in the sparse beamforming matrix is also $O(N_Z N_X N_C)$. In both Matlab (Figure 3.3(a)) and Python (Figure 3.3(b)), CPU-based SMB (purple lines) was at least an order of magnitude faster than DAS (blue lines), with additional speedup provided by moving SMB to the GPU (green lines). As scientific programming languages typically operate much faster on vectorized operations compared to for-loops, this result is unsurprising. Examining one specific example in greater detail, for a 256×256 image in Matlab, the minimal-viable DAS beamformer requires 0.22 s per frame, while SMB performs faster on both CPU and GPU at 0.029 s per frame and 0.0047 s per frame respectively. Moreover, the sparse beamforming matrix for a 256×256 image required only 20 MB of storage and RAM. Across all pixel densities for the CPU implementations, SMB required 16% of time per frame compared to DAS in Matlab and 0.7% in Python. Overall, SMB is an accessible way to provide beamforming acceleration for ultrasound image formation.

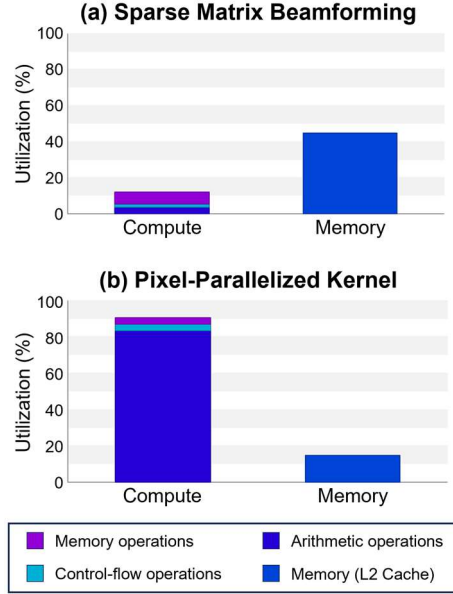


Figure 3.4 (a) Nvidia Profiler analysis for sparse matrix beamforming implemented via the cuSPARSE library showing performance is bound by memory. (b) Nvidia Profiler analysis for pixel-parallelized custom CUDA kernel for delay-and-sum showing performance is bound by computation

To further investigate the causes of acceleration provided by SMB, we designed and compared custom DAS and SMB GPU kernels developed in CUDA (ver. 10.2). For the DAS implementation, the kernel was pixel-parallelized; each thread as assigned to beamform a single pixel intensity. The DAS kernel was designed for plane wave beamforming and accounted for linear interpolation of RF samples and the steering angle. The SMB kernel performed the exact same beamforming operation (RF interpolation, steering angle), but the beamformer was encoded into a sparse matrix instead of being pixel parallelized. To perform the sparse matrix multiplication, the CUDA cuSPARSE library was used to store the sparse beamforming matrix in CSR format. Timings were evaluated for beamforming a size 512×256 image with both kernels using the Nvidia profiler across 1500 instances. For additional investigation, a second GPU – an Nvidia RTX 2080Ti (Nvidia, Santa Clara, CA, USA) – was also used to assess the influence of GPU hardware on the two processes. On the GTX 1080, the pixel parallelized DAS kernel and the SMB kernel averaged 1.69 ms and 7.25 ms per frame respectively. On the more advanced RTX 2080Ti, the same kernels averaged 1.16 ms and 0.69 ms per frame respectively. Thus, depending on the hardware used for implementation, custom CUDA kernels for DAS and SMB can

vary comparatively. Figure 3.4 shows the results of using the Nvidia Visual Profiler to analysis the causes for the discrepancy. In Figure 3.4(a), we observed that SMB was memory dominant, matching expectations as the delay computation was precalculated and offloaded to the sparse matrix coefficients. In contrast, Figure 3.4(b) shows that the pixel-parallelized DAS kernel was computation dominant and memory light; this imbalance is caused by the on-the-fly time-of-flight calculations. Assessing the GPU specifications between the GTX 1080 and RTX 2080Ti, there is greater memory capacity increase (1251 MHz memory clock, 320.3 GB/s memory bandwidth vs. 1750 MHz memory clock, 616.3 GB/s memory bandwidth) compared to computation capacity increase (8.9 TFLOPS peak performance vs. 13.5 TFLOPS peak performance), likely causing the greater improvement in efficiency for SMB. These GPU-specific results highlight the computation-memory tradeoff between DAS and SMB.

The primary advantage of applying SMB lies in the offloading of the repeated delay calculations required for DAS to the sparse matrix structure. While encoding the delays and apodization weights in a sparse matrix can be similar to precalculating a delay table for DAS (Ibrahim et al., 2015), SMB is better suited for natural acceleration of DAS through prebuilt functions. Matlab is already well-equipped for sparse matrix operations, while Python libraries such as SciPy have established sparse matrix functionality for scientific programming. Additionally sparse matrices have inherent GPU support in popular programming frameworks. In Matlab, there is the `gpuArray` object; in Python, the CuPy library; in C/C++, there is the CUDA CuSPARSE library. With the SMB formulation of DAS, the use of prebuilt libraries can aid researchers in reducing the time and effort required to accelerate their beamforming code for real-time performance.

3.3.3 Batch Processing to Reduce Overhead for Sparse Matrix Beamforming

The next example illustrates two aspects of the SMB paradigm: 1) the integration of batch processing for enhanced efficiency and reduction of overhead, 2) the application of SMB for different transmission schemes using focused imaging as a case study.

In focused imaging, many transmit/receive events are used to form a single image for display. Each event corresponds to forming a single A-line of the final beamformed image. DAS beamforming is used for forming each individual A-line from a subset of the received RF channels, the number of channels in the subset is denoted the receive aperture. Typically, the same transmit/receive scheme is used for all scanlines in the image; the scheme is shifted laterally to form successive scanlines. Because the image formation process for each A-line is the same, we can form the sparse beamforming matrix

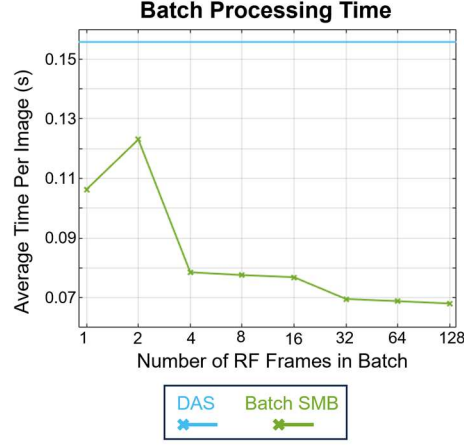


Figure 3.5 Timing for beamforming of focused scanline transmit data. Conventional delay-and-sum (DAS) beamforming is shown in blue, while beamforming of each successive batch of scanline data using sparse matrices is shown in green.

corresponding to a single scanline first, then apply SMB in a batchwise manner to form the full B-mode image with only a single matrix multiplication. To apply SMB batchwise, first assume the full set of RF data is sized $N_S \times N_C \times N_B$, where N_B is the number of batches. Then, each sparse beamforming matrix for focused imaging is of size $N_Z \times N_S N_C$, since only one A-line is formed from each set of RF data. Preparing the RF data for batched multiplication, each RF frame of the batch is vectorized separately, forming a $N_S N_C \times N_B$ matrix representation of the full batch of RF data. Multiplying this vectorized data matrix by the sparse beamforming matrix results in the full $N_Z \times N_B$ intensity image. The key to batched SMB is the vectorization of each frame to form the RF matrix, leading to SMB requiring a sparse matrix-matrix multiplication instead of a sparse matrix-vector multiplication, and thus reducing overhead (Sec. 3.2.2).

For this example, we used a linear L14-5 probe (128 elements) transmitting a total of 128 scanlines. These transmissions all use the same transmit and receive aperture of 64 elements, with the transmit center shifting by one pitch-length for each subsequent scan line. Thus, the overall received RF was of the size $3120 \times 64 \times 128$ (samples $\times$ channels $\times$ batch). We compared the timings of a minimum-viable beamformer (similar to the plane wave example but adapted for focused imaging) and the batched sparse matrix beamformer for batch sizes ranging from 1-128 using an axial pixel density of 2048,

averaging over 100 executions. All comparisons were done using CPU processing only; the timings were evaluated using Matlab's tic toc functionality.

Figure 3.5 charts the results of the timing experiments for batched SMB. The fixed amount of time (0.156 s) for DAS beamforming of the full image (light blue line) was higher compared to SMB regardless of the batch size (green line). We note that the performance gains of SMB for focused imaging compared to plane wave imaging (Sec. 3.3.2) are reduced because fewer delay calculations are required; the delays are calculated on-the-fly for one scanline, then applied to the RF data for the entire batch. For batched SMB, we observed the overall trend of decreasing image formation time with increasing batch size. This observation matches our expectations, as the operational overhead decreases as the batch size increases up to the number of frames in the batch, where each image results from a single sparse matrix-matrix multiplication. The sole outlier to this trend was for a batch size of 2 (totaling 64 matrix-matrix multiplies). Considering that batch size 1 is equivalent to looping through each frame and applying the sparse beamforming matrix (totaling 128 matrix-vector multiplies), the cause of this discrepancy may be an increase in overall overhead for using very small batch sizes. The fastest average frame formation time used a batch size of 128 (0.068 s), corresponding to performing only a single matrix-matrix multiply. Overall, batch processing was an effective technique to reduce the computational overhead of SMB.

3.3.4 Coherent Compounding with Single Sparse Matrix for Plane Wave Imaging

In the next example, we use coherent plane wave compounding to illustrate how the SMB framework can incorporate multiple transmissions into a single sparse matrix operation. Recalling Sec. 3.2.2, the sparse matrices for forming both the LRIs (Low-Resolution Images) and the HRI (High-Resolution Image) for compounding was shown to have block-sparse structure. These block sparse matrices can be formed by first forming the individual sparse matrices for each transmit independently, then combining the matrices in block-wise fashion either diagonally or left-to-right. This formulation encodes the entire coherent compounding paradigm into a single sparse matrix, which can reduce computational overhead.

For this example, we examined the coherent compounding of steered plane wave transmissions. In total, thirty transmissions were used ranging from -15° steering to 15° steering, with the angles separated by 1° increments and the exclusion of the unsteered transmission. The timing experiments were performed by gradually increasing the number of angles compounded in pairs of two. Three ways

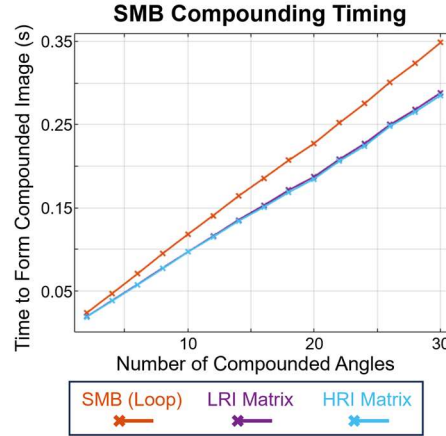


Figure 3.6 Timing for beamforming of compounded images via sparse matrix beamforming (SMB). Looping through sparse matrices for each transmit is shown in red and is approximately 20% slower than using either the HRI matrix to perform compounding and beamforming with one matrix multiply or the LRI matrix to beamform all transmits using a single sparse matrix.

of applying SMB for coherent compounding were tested: 1) looping through the individual sparse beamforming matrices for each transmit, 2) forming a single large LRI matrix, then summing after image formation, 3) forming a single large HRI matrix which directly outputted the compounded HRI. Figure 3.6 plots the results from the timing experiments. We observed a linear increase in the image formation time as the number of compounded angles increased for all methods. For all sets of angles, using a single large sparse HRI matrix (light blue line) or a single large sparse LRI matrix (purple line) for SMB was more efficient than looping through each individual sparse matrix (red line); this effect was likely caused by the threading efficiency of taking an integrated sparse matrix approach (Sec. 3.2.2). The HRI matrix was slightly more efficient (22.9% faster on average than looped SMB across angles) compared to the LRI matrix (21.5% faster on average than looped SMB across angles). The only computational difference between the two approaches is that the summation portion of DAS occurs during the matrix multiplication for the HRI matrix, while it happens after the matrix multiplication and reshaping for the LRI matrix. Overall, using block-sparse matrix structure for SMB for multiple transmissions can be an effective tool for increasing image formation efficiency.

3.3.5 Beamforming-Integrated Neural Networks through Sparse Matrix Multiplication

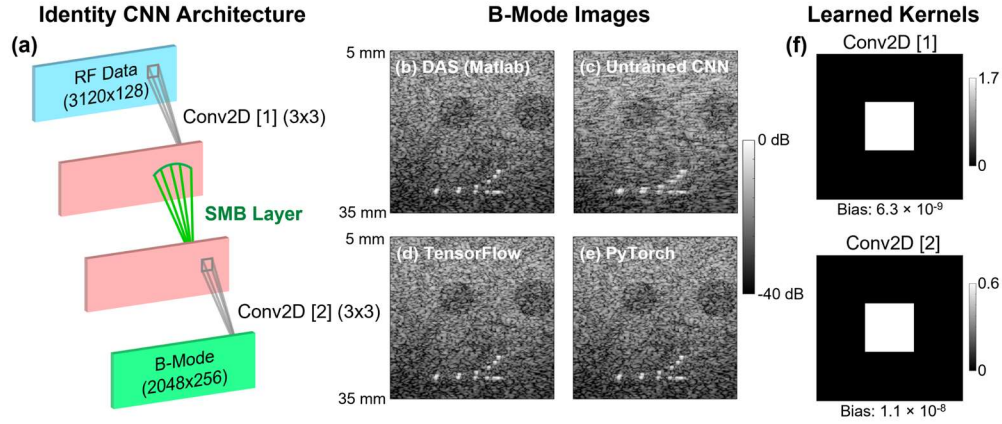


Figure 3.7 (a) Simple neural network architecture to demonstrate proof-of-principle for using sparse matrix beamforming to enable beamforming-integrated neural networks. (b) DAS image from Matlab. (c) DAS image resulting from untrained CNN. (d) DAS image from trained CNN in TensorFlow. (e) DAS image from trained CNN in PyTorch. (f) Kernels from a trained CNN showing that the identity operation was correctly learned both before and after the sparse matrix beamforming layer.

There has been recent growth of research on explicitly incorporating domain knowledge into machine learning models, defined as *informed machine learning* (von Rueden et al., 2023). Because the entire DAS beamforming operation is encoded into a single sparse matrix, SMB is primed for integrating prior knowledge of ultrasound image formation into neural network architectures. Popular machine learning frameworks such as TensorFlow and PyTorch are already equipped with specific libraries for working with sparse matrices (Abadi et al., 2016; Paszke et al., 2019). We propose two main advantages to using SMB for the implementation of beamforming-integrated neural networks. Firstly, SMB enables GPU acceleration through the built-in sparse matrix functionality, alleviating the need for custom GPU-parallelized kernels. Secondly, SMB ensures the compatibility of the beamforming step with the gradient calculations required for the training of neural network models. Harnessing these advantages, we demonstrate that beamforming-integrated neural networks can be designed with an interoperable SMB layer for fast informed machine learning.

To verify that SMB can correctly perform informed machine learning, we design and train a simple convolutional neural network with convolutional layers placed both before and after the beamforming

layer. The architecture of this neural network is shown in Figure 3.7(a). The CNN was designed to perform a trivial identity DAS task to ascertain whether the training procedure was able to correctly update the kernel weights both before and after the beamforming layer. The input to the network was real-valued channel RF data, while the output of the network was the corresponding real-valued post-beamformed intensity values after DAS. Without the introduction of the convolutional layers, the network would simply perform DAS vis SMB (example image in Figure 3.7(b)). The introduction of randomly initialized convolutional layers results in the CNN failing to correctly perform DAS as can be seen in Figure 3.7(c). Thus, if the network once again correctly applies DAS after training, we can then comment on the ability of an SMB-integrated network to learn weights both before and after the beamforming layer. In this way, we can demonstrate that SMB is an efficient and effective way to integrate beamforming into neural network architectures without the need for custom beamforming code.

To demonstrate the effectiveness of our approach in popular machine learning frameworks, we implemented the CNN architecture of Figure 3.7(a) in both Tensorflow (ver. 2.10.1) and PyTorch (ver. 2.0.0). In contrast to the previous example applications, a custom machine learning server was used to perform training and testing – this server was running an EPYC 7313P CPU (AMD, Santa Clara, CA, USA) and an RTX A4000 GPU (Nvidia, Santa Clara, CA, USA). The sparse matrix for beamforming was first generated in Matlab and the coefficients were then stored for loading into the Python frameworks. In Tensorflow, we used the `tf.sparse.SparseTensor` object alongside the `tf.sparse.sparse_dense_matmul` functions for implementing the sparse beamforming layer. The vectorize-multiply-unvectorize scheme (Sec. 3.3.2) was all encoded into a Keras Lambda layer. For timing evaluation in Tensorflow, we used the output times reported by the `model.fit` and `model.predict` functions. In PyTorch, we used the `torch.sparse_csr_tensor` object alongside `torch.matmul` for implementing the sparse matrix multiplication. The sparse beamforming matrix was incorporated as part of a `torch.nn.Module` subclass. For timing in Pytorch, we used the Pytorch profiler functionality. We used a very small set of data collected from the CIRS point target phantom to train and to test the network (training set: 276 RF and B-mode image pairs). In both frameworks, we additionally compared using batch sizes of 1, 4, and 12 for the timing evaluations. All networks were trained for 1000 epochs with no stopping criterion using the Adam optimizer with a learning rate of 0.01 and mean-squared error loss. Bandpass filtered real-valued channel RF data was used for beamforming – analytic signal conversion and log scaling was performed outside of the network for image evaluation only. We also

compared the network timings to a publicly available PyTorch beamforming layer which was provided as part of the CUBDL challenge (Hyun et al., 2021).

Figure 3.7(b)-(e) show the results of the beamforming-integrated neural network before and after training. A successfully trained network would, in theory, learn the identity kernels for both convolutional layers. A correctly beamformed image of the CIRS phantom from the test set using the Matlab-based DAS beamformer is shown in Figure 3.7(a). Applying an untrained, randomly initialized network resulted in the distorted image in Figure 3.7(b). After training, the Tensorflow network and the PyTorch network both successfully learned to correct the DAS operation to its original state, as shown in images in Figure 3.7(c) and (d) respectively. Furthermore, examining the kernels of one trained instance of the network (Figure 3.7(f); top kernel before SMB layer, bottom kernel after SMB layer) demonstrates that the network training behaved in the expected manner. In this way, we experimentally prove that using sparse matrices for beamforming-integrated neural networks leads to accurate updates of convolutional kernels both *before* and *after* the beamforming layer. We note that in this example, the DAS layer itself was held fixed – no updates were applied to the sparse beamforming matrix.

To ensure tractability of training and inference, Table 3.1 summarizes the average training and inference times for the simple networks, along with the time required to apply the sparse beamforming layer only. For both the Tensorflow and PyTorch implementations of the simple CNN architecture, the millisecond-order training timings (per batched step) for the two frameworks indicate efficiency of updating the weights around the SMB layer. The even faster inference times demonstrate the potential for real-time application of beamforming-integrated neural networks. Assessing the application of the SMB layer alone reveals that the inference time for the simple network is dominated by the beamforming; however, this discrepancy is primarily caused by the simplicity of the convolutional layers. The advantage of the sparse matrix approach can be observed upon comparison of the SMB layer in the network with an open-source beamforming layer provided with the CUBDL challenge

TABLE 3.1
TIMING EVALUATION FOR BEAMFORMING-INTEGRATED NEURAL NETWORKS

	TensorFlow			PyTorch		
Batch Size	1	4	12	1	4	12
Training Step (ms)	22	65	183	133	152	168
Inference Step (ms)	5	12	29	4	11	25
SMB Layer (ms)	4	11	25	2	10	17

(Hyun et al., 2021). We replaced the SMB layer with the CUBDL beamforming layer (`das_torch` function) in the same simple network architecture (Figure 3.7(a)) in PyTorch and evaluated the network timings for a batch size of 1. Compared to the SMB layer, the CUBDL layer was slower in training (133 ms vs 383 ms) and much slower in inference and beamforming alone (inference: 3 ms vs 380 ms, beamforming: 2 ms vs 349 ms). Lastly, to qualitatively assess the ease of implementation of a training DAS layer, we compared the lines of code required for the SMB layer versus the CUBDL layer. The proposed SMB approach requires fewer than 10 lines of code in Python (example online), while the CUBDL function needed over 100 lines of code.

Overall, this example demonstrates the ease and efficiency of using sparse matrices to embed beamforming within neural network architectures and benefit from informed machine learning.

3.4 Discussion

3.4.1 Sparse Matrix Beamforming as an Efficient and Accessible Delay-and-Sum Beamformer

Real-time imaging is a significant advantage of ultrasound as a medical imaging modality, allowing for use at the point-of-care. Open-platform ultrasound scanners can allow researchers to adapt new imaging innovations for real-time operation; however, the expertise required for efficient parallel programming on these scanners can be challenging (Boni et al., 2018). In this work, we investigated the encoding of delay-and-sum (DAS) beamforming as a sparse matrix multiplication, denoted sparse matrix beamforming (SMB). We demonstrated how to construct sparse beamforming matrices based on the principles of linearity (Figure 3.1), then extended the results for batch-wise or multi-transmit processing. This reformulation of DAS using sparse matrices enables researchers to leverage powerful sparse matrix libraries to accelerate beamforming processes.

Experimental results showed that SMB can provide both acceleration and ease-of-implementation for software-based DAS beamforming. We designed a minimum-viable beamformer to demonstrate the principles of SMB pedagogically (Code Samples 3.1, 3.2, 3.3). For CPU-based timing comparisons (Figure 3.3), SMB achieved one order of magnitude acceleration compared to conventional DAS in Matlab (Figure 3.3(a)), and two orders of magnitude more in Python (Figure 3.3(b)). Moreover, using SMB allowed for nearly one-to-one code conversion for CPU to GPU implementation, demonstrating the simplicity of application for parallel programming. Further experiments revealed the memory-

computation tradeoff (Figure 3.4) of applying SMB, since the delays were effectively precalculated and encoded in matrix coefficients. The efficiency of SMB was increased further through reduced overhead both batch-wise for a focused imaging example (Figure 3.5) and multi-transmit for a coherent compounding example (Figure 3.6). Lastly, we showed how SMB is well-positioned to support integration of DAS into common machine learning frameworks such as PyTorch and TensorFlow. By designing a simple convolutional neural network to test the proof-of-concept (Figure 3.7(a)), we observed that neural network weights correctly updated both before and after the inter-operable SMB layer (Figure 3.7(b)-(f)). The millisecond-order training and inference times (Table 3.1) showed that the sparse matrix approach was highly tractable, and the sparse matrix functionality of the two frameworks led to straightforward implementation. Overall, we demonstrated that SMB is a versatile technique for software-based DAS beamforming, offering significant advantages in terms of speed, simplicity, and compatibility with modern scientific programming libraries.

3.4.2 Integration for Software Frameworks via Sparse Matrix Beamforming

SMB is an accessible tool which can be useful for ultrasound researchers working on software-based systems. By encoding DAS as a sparse matrix multiplication, the option of applying prebuilt sparse matrix libraries for acceleration becomes available. In our work, we have shown the use of sparse matrix functionality in Matlab, Python (CuPy, TensorFlow, PyTorch), and C/C++ (CUDA, cuSPARSE). While these are the most common tools used for software-based ultrasound pipelines, sparse matrix packages also exist for less popular scientific programming languages such as R, Julia, or Fortran. One advantage of SMB for these scenarios is its portability; customized DAS schemes can be designed in a home language, such as Matlab, then saved as a sparse matrix and loaded into other languages for efficient operation. SMB can benefit not only software-based implementations, but also possibly FPGA designs, as FPGA-based sparse matrix multiplication exists (Fowers et al., 2014). A sparse matrix encoding for DAS can also enable new research directions for beamforming. For example, optimization research frequently applies matrix operations, which SMB can naturally fit into as a modular component.

While SMB can be highly efficient and accessible, specialized custom programs for ultrasound beamforming can surpass the image formation efficiency of SMB (for example, CUDA beamformer in Sec. 3.3.2). Once a library and language are selected, there is no way to further improve SMB from a sparse matrix multiplication perspective. Highly customized beamforming code can leverage efficient

GPU memory management and minimize the number of on-the-fly computations (Yiu et al., 2011). FPGA innovations for rapid encoding and loading of delays have also demonstrated extremely efficient operation (Kou et al., 2023). For SMB, the tradeoff between the flexibility and peak performance of custom DAS versus the ease-of-implementation of SMB can be the deciding factor for researchers. One possible development paradigm for SMB is to use sparse matrices to quickly assess new ideas for real-time operation, then devoting time and resources to design custom high-performance beamforming accelerants.

3.5 Conclusion

Sparse matrix beamforming (SMB) is a reformulation of delay-and-sum (DAS) beamforming as a single sparse matrix multiplication. Expressing DAS in this way enables researchers to access powerful sparse matrix libraries for acceleration of ultrasound image formation. Compared to loop-based DAS implementations, SMB is highly efficient, has low complexity, and is easily integrated into scientific programming frameworks for parallel computing or machine learning. The SMB framework is well-suited for researchers to quickly integrate DAS into real-time software-based ultrasound systems.

3.6 Acknowledgment

The authors are grateful to Dr. Billy Y. S. Yiu for his insightful perspectives on beamforming and to Pat de la Torre for other helpful discussions.

Chapter 4

Minimizing Image Quality Loss after Channel Count Reduction for Plane Wave Ultrasound via Deep Learning Inference

Reprinted under Creative Commons Attribution License (CC BY), from Xiao, D.; Pitman, W. M. K.; Yiu, B. Y. S.; Chee, A. J. Y.; Yu, A. C. H. Minimizing Image Quality Loss after Channel Count Reduction for Plane Wave Ultrasound via Deep Learning Inference, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, July 2022.

4.1 Introduction

Ultrasound imaging is well regarded as a noninvasive medical imaging modality. For current clinical scanners with array scanning capabilities, they typically form each image frame via a beamline-based pulse-echo sensing paradigm that involves sending one focused pulse firing along each of a set of beamlines in the imaging view. From each pulse-echo sensing event, one axial line of the image frame is generated by performing delay-and-sum (DAS) beamforming based on the channel-domain radio frequency (RF) data samples acquired from all array elements (Whittingham, 2007). Real-time frame rates can be achieved with this conventional ultrasound imaging paradigm. In contrast, in the past decade, unfocused pulsing of planar or diverging wavefronts, which can transiently insonify the entire imaging view, has demonstrated efficacy in achieving high frame rates of over 1000 frames per second (fps) (Boni et al., 2018; Tanter & Fink, 2014). For this high-frame-rate imaging paradigm, one image frame can be generated from the channel RF data array of the corresponding unfocused pulse-echo firing. The beamformed images obtained from unfocused pulse-echo firings with different steering angles can also be coherently compounded to improve the lateral image resolution (Montaldo et al., 2009; Yiu et al., 2011). Accordingly, researchers have successfully imaged hemodynamics in vivo on a time-resolved basis with submillisecond temporal resolution (Ketterling et al., 2017, 2017; Udesen et al., 2008). Others have also developed new applications, such as shear wave elastography and functional ultrasound imaging (Macé et al., 2011; Tanter et al., 2002).

While conceptually powerful, high-frame-rate ultrasound imaging is not without its practical implementation challenges. Its two primary hurdles are: 1) high computational capacity is required to generate images in real time (So et al., 2011) and 2) high streaming bandwidth is needed in transferring the channel RF data from the array front end to the computing back end (Boni et al., 2018). The first

hurdle can possibly be overcome by applying software beamforming methods that make use of graphics processing unit (GPU) computing devices to parallelize the DAS beamforming process over different pixel positions (Hansen et al., 2011; Yiu et al., 2011). Nevertheless, the second hurdle is more challenging to address. Specifically, upward of 10 GB of channel RF data can be transmitted every second, and few scanners are designed to accommodate this level of data streaming traffic (Boni et al., 2018). In light of this bottleneck, it is of practical interest to reduce the amount of channel RF data that needs to be streamed to the computing back end.

Several approaches have attempted to circumvent the data streaming bottleneck in high-frame-rate ultrasound imaging. One emerging direction in recent years is compressed sensing (CS), which combines sub-Nyquist sampling with assumptions of sparsity to achieve close to 90% data reduction (Chernyakova et al., 2018; Chernyakova & Eldar, 2014; Liebgott et al., 2013; Liu et al., 2017). However, these methods often require an iterative solution to optimization problems (i.e., there is no closed-form solution), so it is inherently challenging to implement them in real time even though attempts to accelerate CS have been made (Chen et al., 2021; Huijben et al., 2020). Another algorithmic solution is to perform sparse convolutional beamforming that works by using a sum coarray algorithm to improve image quality while nonuniformly reducing the total number of array elements. Nevertheless, this solution has yet to be demonstrated in real time, and its feasibility in high-frame-rate imaging via unfocused transmissions has not been shown (Cohen & Eldar, 2018).

Other approaches to reduce the data streaming traffic in ultrasound imaging involve hardware modifications, either by reducing the total number of array channels or by executing preprocessing within the front-end hardware. The use of sparse arrays is one example of such an approach. Although their use can directly reduce the data streaming traffic, sparse arrays typically have higher sidelobe levels in their field profile, and this issue is known to compromise image quality (Austeng & Holm, 2002; Diarra et al., 2013; Mitra et al., 2010; Von Ramm et al., 1976). Alternatively, the incorporation of hardware microbeamformers within the transducer front end, which execute part of the DAS beamforming process, also allows for fewer data samples to be streamed to the system back end (Larson, 1993; Leavitt et al., 2002). Yet, this approach inherently makes the system less reconfigurable, so it is ill-suited for the implementation of new channel-domain imaging methods (Boni et al., 2018). Perhaps, a less restrictive data transfer reduction solution is to add a lossless compression–decompression algorithm to the system’s data streaming operations (Lok et al., 2015). Nonetheless, the compression ratio that can be achieved tends to vary depending on the characteristics of the acquired

channel RF samples, so it is uncertain as to whether the size of the compressed data streaming traffic can always be accommodated by the system.

In this article, we present a new real-time framework for data transfer reduction when performing high-frame-rate ultrasound imaging via a software beamforming approach. Our proposed framework works by uniformly halving the channel count on receive and then reconstructing the other half of the channels using a data-driven, deep-learning approach. We hypothesize that the total number of receive channels for robust image formation can be reduced by exploiting redundancies between the pulse echoes of adjacent channels. This redundancy stems from the prevailing concept in pulse-echo sensing that different transducer channels on the same aperture would tend to receive, for a given scatterer, a similar echo signature that is merely shifted in time. Accordingly, we posit that a convolutional neural network (CNN) architecture can be trained to capture the redundant structure within the channel RF data array. The trained network can then be applied to reconstruct full-channel RF datasets from truncated ones acquired with fewer array elements used on reception. In doing so, data streaming traffic can be effectively halved with minimal loss in image quality and with bearable computational cost that does not compromise real-time feasibility, as will be shown later in this article.

4.2 Theory

4.2.1 Background Considerations

In principle, it is readily possible to reduce raw data transfer in ultrasound imaging by uniformly decimating every other channel on the array aperture during pulse-echo reception. Nevertheless, in doing so, the array pitch for beamforming would be concomitantly doubled. Using a large pitch relative to the acoustic wavelength (exceeding the spatial Nyquist limit of $\lambda/2$) results in spatial aliasing that, in turn, introduces imaging artifacts along the lateral dimension and reduces the contrast of echolucent regions (Kremkau & Taylor, 1986). Fortunately, this drawback is in theory addressable because, according to modern signal processing, even the Nyquist sampling limit can be overcome with additional knowledge of the structure of the signal (Candes et al., 2006; Mishali & Eldar, 2010). As explained in Appendix B, ultrasound RF data are indeed highly structured across adjacent channels. Accordingly, our proposed framework has been devised to implicitly take advantage of this structural redundancy via a data-driven, deep-learning approach to recover decimated channel-domain RF data and, in turn, generate ultrasound images with comparable quality as those formed using RF data acquired from the fully populated array.

4.2.2 Overview of the Proposed Framework

A convolutional encoder–decoder network is designed to use learned convolutional kernels to extract features in the RF data and to perform data recovery accordingly. This network forms the basis of our proposed pre-beamformed data formation framework as shown in Figure 4.1, together with the overall imaging system. After a transmit event, the framework (dashed box) first acquires from half of the receive channels (odd-indexed) of the ultrasound probe, then infers the unacquired channels (even-indexed) using the neural network, and finally interleaves the acquired and inferred channels to reform a full-channel set of data. The reconstructed full dataset is then passed to a beamforming module to generate an image frame.

The encoder–decoder network is chosen because this architecture is suitable for the RF data inference task based on image processing literature. In modern image processing, encoder–decoders are used for image restoration tasks such as inpainting (i.e., the recovery of lost data in images and videos (Xie et

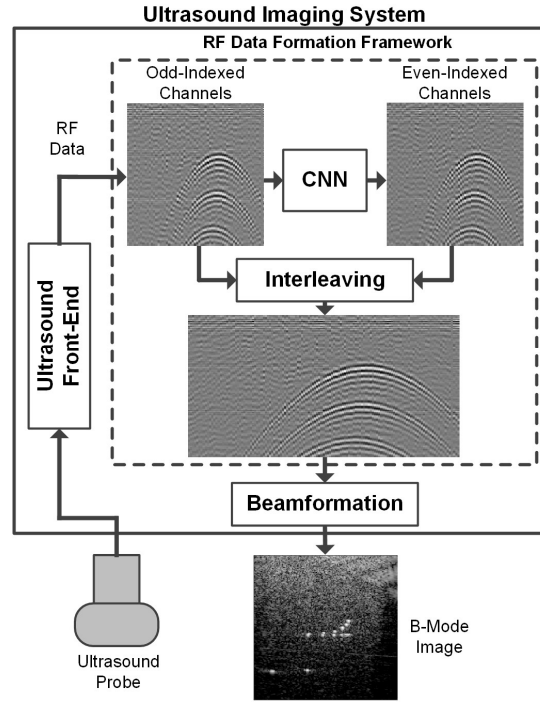


Figure 4.1 Diagram of overall ultrasound imaging system (solid box), with RF data formation framework shown (dashed box). The proposed framework takes in odd-indexed channels, uses a CNN to infer the even-indexed channels, and then interleaves the channels to output a full-channel RF dataset.

al., 2012)), which can be considered analogous to the recovery of missing RF data between channels in ultrasound. Similar to inpainting, after data-driven training, the encoding stage of the neural network learns to generate a compressed representation of the most important spatiotemporal features and the decoding stage learns to regenerate the missing portions of RF data based on those encoded features.

4.2.3 Architecture of the Custom Encoder–Decoder Neural Network for RF Data Recovery

Figure 4.2 shows the overall structure of the proposed encoder–decoder network. The network first takes in the odd-indexed RF channels as its input (light blue) and then prepares the data using zero-padding (green). Next, the prepared data enter the encoding stage of the network, where the redundant spatiotemporal features are extracted and compressed via four layers of convolutional kernels. Afterward, the compressed representation is decoded to infer the unacquired RF data via four

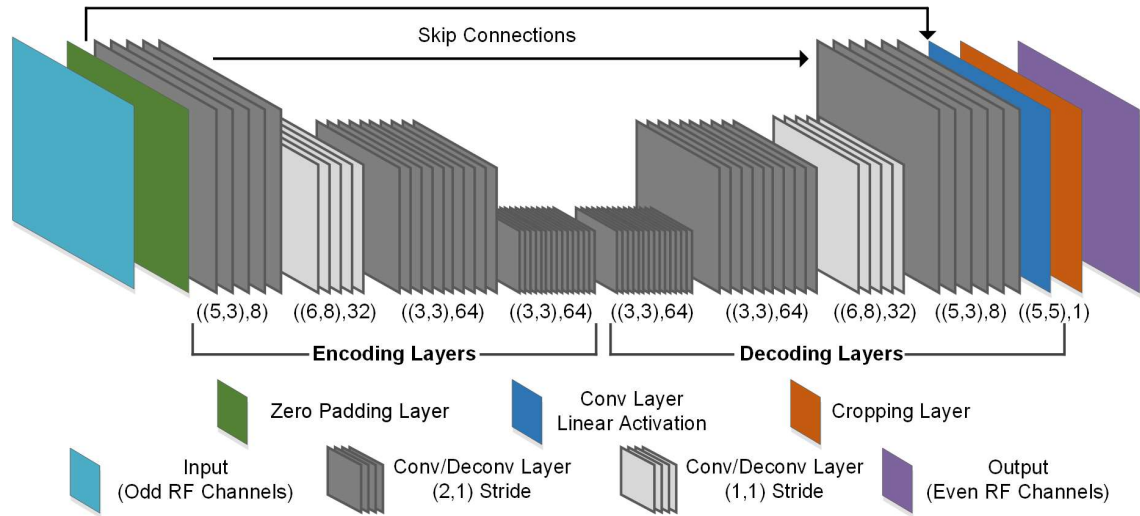


Figure 4.2 Proposed architecture of the convolutional encoder–decoder. Odd-indexed RF channels (light blue) are input to the network and zero-padded (green) into a multiple of 8 in size. The encoding layers compress the information, while the decoding layers interpret the compressed information. The number and size of the kernels for each encoding and decoding layers are shown below the corresponding layers (dark and light gray). Two sets of skip connection concatenate the outputs of the encoding layers (start of arrows) to the inputs to the decoding layers (end of arrows). The last convolutional layer (dark blue) recombines the information into a single RF frame, which is cropped (orange) to ensure preservation of input and output dimensions, resulting in the inferred even-indexed RF channels (purple).

deconvolutional layers. After the decoding stage, the resulting feature maps are combined into one using the last convolutional layer (blue). Lastly, the feature map is cropped (orange) to form the inferred even-indexed RF channel data for output (purple). In Sec. 4.2.4 and 4.3, the RF data input–output and the encoder–decoder network are discussed in detail.

4.2.4 RF Data as an Image for Encoder–Decoder Input–Output

The input and output of the convolutional encoder–decoder are formed by treating the RF data across channels as a single image. Channelwise concatenation of RF data, visualized as an image in Figure 4.1, exhibits spatiotemporal structure across both fast time and across the channels. Even in just half the channels, the hyperbolic shape that the framework leverages for RF data inference is still present in the RF data frame, as shown in top left of Figure 4.1 (under “odd-indexed channels”). Before entering the encoding–decoding stage of the network, the RF data frame is zero-padded to ensure that the appropriate sample size is used for downsampling along the depth dimension across all encoding layers. The kernels of the proposed network then capture and compress the structure in the odd-indexed channels and use that structure to infer the even-indexed channels. Finally, the RF data frame is cropped to return the output to the original size of the input RF data.

4.3 Custom Encoder-Decoder Implementation

4.3.1 Number of Layers, Kernels per Layer, and Kernel Sizes

In a CNN, the capacity to capture complex structures is governed by the number of layers, the number of kernels per layer, and the kernel sizes. Increasing these parameters allows for increased capacity to capture complex structures; however, doing so comes at the expense of increased computational cost.

To balance this tradeoff, a nine-layer encoder–decoder network was chosen. The encoder–decoder has four encoding layers to extract the redundant structure in the RF frame, four mirrored decoding layers for inference, and one recombination layer to combine the inferred features into a single RF frame. The number of kernels per layer was also chosen to balance the tradeoff between capacity and computation. The four encoding layers, respectively, have 8, 32, 64, and 64 number of kernels; this subarchitecture was mirrored in the decoding layers. The number of kernels per layer along with the convolutional stride length helps to compress the RF data into feature maps in encoding and to infer the RF data in decoding.

The convolutional kernel sizes affect the extent of the structure captured in each layer. The sizes were selected, so that the first layer captures more of the features related to the fundamental RF frequency along the depth dimension by using a long kernel of size (5,3). The second layer captures more of the salient spatiotemporal features using a wide kernel of size (6,8). The third and fourth convolutional layers, size (3,3), add depth to the network to better capture the redundant structure. The subsequent deconvolutional layers (first four layers of decoding stage in Figure 4.2) mirror the parameters of the encoding layers to form a standard encoder–decoder architecture. The convolutional layer at the end of the decoding stage (Figure 4.2, in blue) used a kernel of size (5,5) to recombine the inferred features into a single image.

4.3.2 Convolutional Strides

To compress and decompress the extracted RF data features, the proposed network used a stride length of two for all layers but the second encoding layer and its corresponding decoding layer (white layers in Figure 4.2). The stride of a convolutional layer represents sample “skipping” in each dimension (Springenberg et al., 2015). For example, a kernel with a stride length of one performs convolution at each index. A convolutional kernel with a stride length of two performs convolution at every other index, compressing the features of the RF data. Similarly, deconvolutional strides expand the features of the RF data.

4.3.3 Skip Connections

As shown in Figure 4.2, skip connections were used to connect the first and second encoding layers to the last and second last decoding layers. These connections enable extracted features from the first two layers to have additional effect in inferring the unacquired channels in the decoder stage of the network. Skip connections also stabilize the gradient descent methods used for neural network training (Drozdzal et al., 2016). A drawback of using skip connections is that additional computation time is required for inference; thus, in the proposed encoder–decoder, the skip connections are limited to the outer layers with the fewer number of kernels.

4.3.4 Activation Function

Apart from the last convolutional layer, a leaky rectified linear unit (ReLU) activation function ($\alpha = 0.01$) is used to better facilitate the training. While a standard ReLU takes a value of 0 for inputs below 0, the leaky ReLU has a very small slope, which has been shown to help in training by alleviating the

TABLE 4.1
TIMING EVALUATION FOR BEAMFORMING-INTEGRATED NEURAL NETWORKS

Parameter	Details
Ultrasound Scanner	SonixTouch
Ultrasound Probe	L14-5
Number of Tx/Rx Channels	128
Array Pitch	0.3048 mm
Number of Tx Pulse Cycles	2
Transmit Frequency	5 MHz
Transmit Angle Range	-15° to 15°
Sampling Rate	20 MHz
Imaging Depth	60 mm
Pulse Repetition Frequency	10 kHz

vanishing gradient problem (Maas et al., 2013). In contrast, the last convolutional layer uses a linear activation, since the expected even-channelled RF data outputs span both positive and negative values.

4.4 Experimental Methods

4.4.1 Acquisition of Neural Network Training Data

In order to train the CNN presented in the previous section, a total of 62 799 steered in vivo RF datasets were acquired. These datasets were captured along both the short axis and the long axis of human carotid artery from seven healthy volunteers (age: 25.9 ± 4.9) using a research imaging platform (SonixTouch; Analogic Ultrasound; Peabody, MA, USA). The research platform was programmed to transmit 31 steered plane wave pulses (2-cycle 5-MHz pulse; -15° to 15° , 1° separation; pulse-repetition frequency (PRF): 10 kHz) with an L14-5 array transducer (pitch: 0.3048 mm; Analogic Ultrasound). This research data acquisition protocol has been approved by the Clinical Research Ethics Committee of the University of Waterloo (Protocol No. 31694). The received echoes were digitized using a pre-beamformed data acquisition tool (Cheung et al., 2012) with 12-bit resolution at 20-MHz sampling rate. The acquired RF data were subsequently transferred to a computer server for off-line processing. Table 4.1 summarizes the ultrasound scanner details and imaging parameters used for data acquisition.

4.4.2 Data Preprocessing

After acquisition, the data were preprocessed to prepare the training inputs and training references in MATLAB (version 2016a; MathWorks, Natick, MA, USA) using a computer server (SYS-4028-TRT;

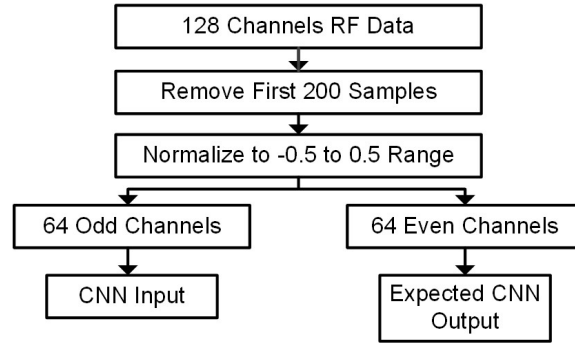


Figure 4.3 Data manipulation flowchart for all acquired RF data.

Super Micro, San Jose, CA, USA) running a Xeon E5-2620 central processing unit (Intel, Santa Clara, CA, USA) and an RTX-2080 GPU (Nvidia, Santa Clara, CA, USA). Figure 4.3 depicts the data processing pipeline that the acquired RF data follow, so that it is properly split and scaled to serve as training input and reference pairs. First, the initial 200 samples from each channel were discarded to minimize the effect of the transmit pulse on neural network training. The acquired samples were then limited to within the -0.5 to 0.5 range by dividing by 4096 (the 12-bit samples acquired from the SonixTouch are stored as integers from -2048 to 2047) as it is within the range that most CNN implementations work best. The odd-numbered channels and even-numbered channels were then split into 1300×64 sized input and reference RF frames for the convolutional encoder–decoder to learn.

4.4.3 Neural Network Training and Evaluation

The proposed encoder–decoder CNN was implemented in Python (version 3.6.7) using the TensorFlow-GPU (version 1.12.0) back end for GPU-integrated training neural network training. Keras (version 2.1.6), TensorFlow’s high-level application programming interface (API), was used to define the custom encoder–decoder architecture and to call the TensorFlow-GPU back end for GPU-accelerated training and inference. To train the implemented network, 80% of the acquired dataset was used and further split into training set (85%) and validation set (15%). The training was performed for 50 epochs by optimizing the mean-squared error (MSE) loss function using the Adam optimization algorithm (Kingma & Ba, 2017) at a learning rate of 0.001 and a batch size of 4. Subsequently, the remaining 20% of the acquired dataset was used to evaluate the performance of the network by examining the normalized root-mean-squared error (NRMSE). The NRMSE was evaluated on three data variants: the raw RF data, the amplitude of the Hilbert transformed RF data, and the phase of the Hilbert transformed RF data.

4.4.4 Image Quality Evaluation of the Proposed Framework

After training, new RF datasets were acquired and beamformed with both original and inferred data to assess the performance of the overall framework. Using the same parameters as in Table I, new datasets were acquired from the carotids of healthy subjects outside of the training set subjects. Additionally, RF dataset was acquired from a multi-point target phantom (Multi-Purpose Multi-Tissue Ultrasound Phantom; Computerized Imaging Reference Systems, Inc., Norfolk, VA, USA) using a steering angle span of -14.75° to $+14.75^\circ$ (increments of 0.5°). This phantom test case served to assess the generalizability of our network both to angles and to imaged structure that was entirely different from our training dataset. The images formed from these acquisitions served to evaluate image quality in terms of resolution, contrast, and spatial aliasing artifacts.

To analyze the efficacy of the proposed framework in recovering image quality and suppressing spatial aliasing artifacts, three derivative RF datasets were generated from the newly acquired RF data.

1. *Odd-indexed channels*: The odd-indexed channels (64 in total) of the acquired RF data for each angle.
2. *Linearly interpolated channels*: This RF dataset represents a reference data recovery scheme for the unacquired channel data. The recovery was done by linearly interpolating the neighboring odd-indexed channels to form the unacquired even-indexed channels to restore the full-channel (128) RF dataset.
3. *CNN-inferred channels*: This dataset was formed using our proposed framework. The odd-indexed channels were extracted from the original acquisition and passed into the CNN encoder–decoder to infer the even-indexed channels for full-channel (128) restoration.

B-mode images were subsequently formed from the original full-channel dataset and the three derivative RF datasets using a conventional DAS beamforming framework according to the parameters in Table 4.2. The RF data were first prefiltered with a 3–7-MHz finite-impulse response (FIR) bandpass filter applied along the fast-time dimension. The RF data were then converted to the analytic form using

TABLE 4.2
ULTRASOUND IMAGE BEAMFORMING PARAMETERS

Parameter	Details
Prefilter Passband	3-7 MHz
Filter Design Method	Equiripple (30 th Order)
Apodization	Rectangular
Focusing	Constant $F_{\text{num}} = 1.25$
Image Dimension	256 (axial) $\times$ 256 (lateral)
Lateral Resolution	0.15 mm/pixel
Axial Resolution	0.15 mm/pixel

the Hilbert transform. Afterward, DAS beamforming was performed to generate the four sets of images; a constant F-number of 1.25 and a uniform (rectangular) apodization were used.

4.4.5 Evaluation of Image Quality After Coherent Compounding

Plane wave images are often coherently compounded to improve the image quality as compared to images beamformed from a single transmission. The image quality (spatial aliasing artifacts, contrast) was also assessed after coherently compounding increasing subsets of the beamformed images from steered transmissions. The steering angles used for compounding were selected in pairs, so that each compounded image was always formed from balanced positive and negative steering angles. The single transmit images were also selected to be evenly distributed along the angle span (e.g., -8° and 8° for two compounded angles; -10° , -5° , 5° , and 10° for four compounded angles).

4.4.6 Evaluation of Computational Capacity

To assess the possibility for real-time inference using the proposed framework, the inference time for RF channel regeneration was calculated using Python’s built-in timing functionality (`time.clock`); 250 RF data inference operations were conducted and their timings were averaged.

4.5 Results

4.5.1 Trained CNN Encoder–Decoder Can Recover the Unacquired Channel Structure

The proposed CNN encoder–decoder was able to recover the structure of even-channelled RF data for the unseen test set data (i.e., not included for training and validation). As an example, Figure 4.4(a) shows an *in vivo* human carotid B-mode image (40-dB dynamic range) generated from the test set along with the corresponding reference even-channel RF frame (see Figure 4.4(b)) and the inferred RF frame (see Figure 4.4(c)) based on the odd-channelled RF input. The overall structure in the inferred RF data

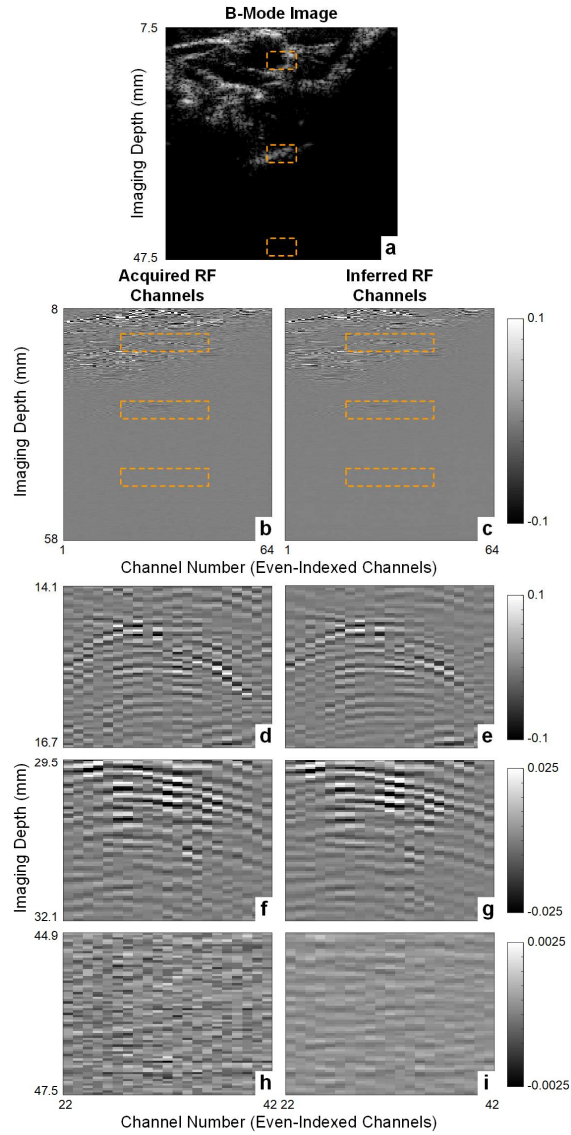


Figure 4.4 (a) B-mode image of human carotid in vivo and the corresponding acquired even-indexed RF data (left column) and inferred even-indexed RF data (right column). Whole RF data “images” are shown in second row, (b) and (c). Snippets corresponding to respective orange dashed boxes in (d)–(i).

generally matched the reference one; this observation is substantiated by the low NRMSE (0.012) between the two RF images.

To more closely examine the performance of the network at different depths, Figure 4.4(d)–(i) shows the enlarged regions of RF data at increasing depths (depicted by the corresponding dotted boxes in

TABLE 4.3
NRMSE OF INFERRED RF DATA COMPARED TO REFERENCE (FIGURE 4.4)

Region	Entire RF	Shallow Depth	Medium Depth	Deep Depth
CNN Interpolated (Int.)	0.012	0.050	0.026	0.137
CNN Int. Amplitude	0.017	0.099	0.049	0.195
CNN Int. Phase	0.206	0.165	0.092	0.239
Linear Interpolated	0.022	0.075	0.037	0.138
Linear Int. Amplitude	0.025	0.142	0.096	0.198
Linear Int. Phase	0.269	0.247	0.254	0.264

Figure 4.4(b) and (c)). As can be observed, the local structures in the inferred RF data (right column) resembled that of reference RF data (left column) for shallow and medium depths (see Figure 4.4(d)–(g)) but were dissimilar for deep imaging depths (see Figure 4.4(h) and (i)). Table 4.3 quantifies the NRMSE (normalized to RF data range) in the image for each region and compares the CNN inference approach with linear interpolation of the channels. Additionally, the Hilbert transform of the RF data is taken, so that the NRMSE of the amplitude (normalized to RF amplitude range) and phase (normalized to 2π) can be evaluated. RF data inferred with our CNN show that the CNN better matches the reference RF data in every scenario compared to linear interpolation. The higher NRMSE in deeper regions (see Figure 4.4(h) and (i)) is due to the lack of structured scattering from that region as shown in the corresponding B-mode images (bottom of Figure 4.4(a)).

4.5.2 Grating Lobe Reduction and Preservation of Resolution in Multiple Point Target Phantom Across Transmit Angles

To demonstrate that our data recovery scheme does not result in a loss of post-beamformed image quality, Figure 4.5 displays point target images generated from the full-channel and the derivative datasets (left to right, see Section IV-D) with -14.75° (top row) and -0.25° (bottom row) steering transmissions. It can be observed that the images beamformed with the proposed method (second column) contain reduced spatial aliasing artifacts (highlighted by yellow squares) from the channel reduction as compared to images from other derivate datasets (third and fourth columns). This observation holds across the two steering angles shown in Figure 4.5.

To quantify the reduction in grating lobe power across transmission angles, Figure 4.6 plots the mean grating lobe power at different transmission angles, averaged over a $4 \times 4 \text{ mm}^2$ region centered on the peak spatial aliasing artifacts on the left side for each angle (depicted as yellow squares for examples in Figure 4.5). Images from the CNN-derived dataset (orange line) show up to 10-dB improvement over the odd-indexed channel images and similar intensity ($< 2\text{-dB}$ difference) to the full-channel

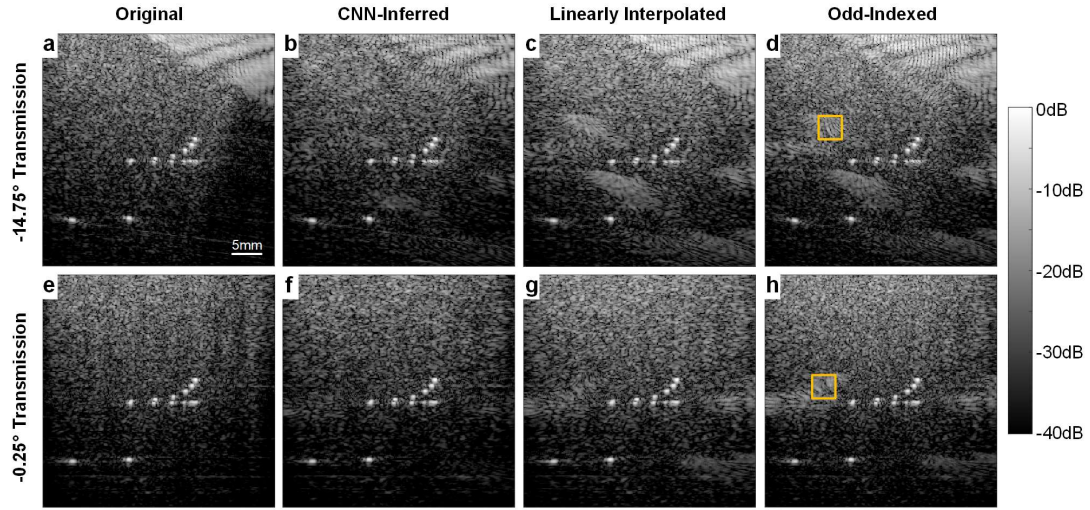


Figure 4.5 Single plane wave transmission B-mode images (40-dB dynamic range) of point target phantom beamformed from original full-channel RF dataset (first column), CNN-inferred channel dataset (second column), linearly interpolated channel dataset (third column), and the original odd-indexed channel dataset (fourth column). Images are shown for -14.75° transmission [top row, (a)-(d)] and for -0.25° transmission [bottom row, (e)-(h)]. Yellow boxes highlight spatial aliasing regions used to calculate artifact power in Figure 4.6.

images (dark blue line) across most transmission angles in the artifact region. For transmission angles above 5° , the intensity deviates from the full-channel images due to reduced insonification from plane wave steering (as shown in the bottom-right thumbnail). Although the artifact intensity of images from the linearly interpolated RF data (light purple line) demonstrates minor improvement (~ 4 dB) over the artifact intensity of images from the odd-indexed channels only (light blue line), the results are still inferior to our proposed method.

As a further evaluation of the effect of using CNN-inferred channels for image formation, a cross section [depth taken shown in Figure 4.7(a)] along the row of point targets is plotted for images generated from each compared dataset in Figure 4.7(b) for an unsteered plane wave transmission. The artifact reduction of the proposed method (orange line) can be clearly observed on the left and right sides of Figure 4.7(b). Figure 4.7(c) expands the portion of the cross section corresponding only to the wire point targets to demonstrate the lack of change in resolution. Additionally, the full-width half-maximum (FWHM; measured 6 dB below the peak for each target) for each point target labeled in Figure 4.7(a) was compared and shown in Table 4.4 to quantify the resolution difference. From the

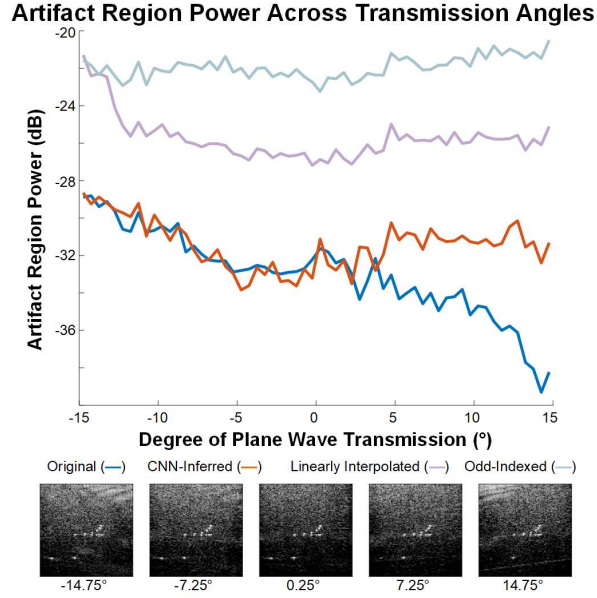


Figure 4.6 Measured spatial aliasing artifact power in B-mode images across transmission angles for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and the original odd-indexed channel dataset (light blue). Thumbnails of B-mode images beamformed from CNN-inferred channels shown at bottom.

table, it can be observed that the FWHM values across all four images for each of the three point targets are comparable, with the linear interpolated channel image showing the worst point target resolution. This shows that there is a minor loss of resolution by forming images using the proposed data inference method.

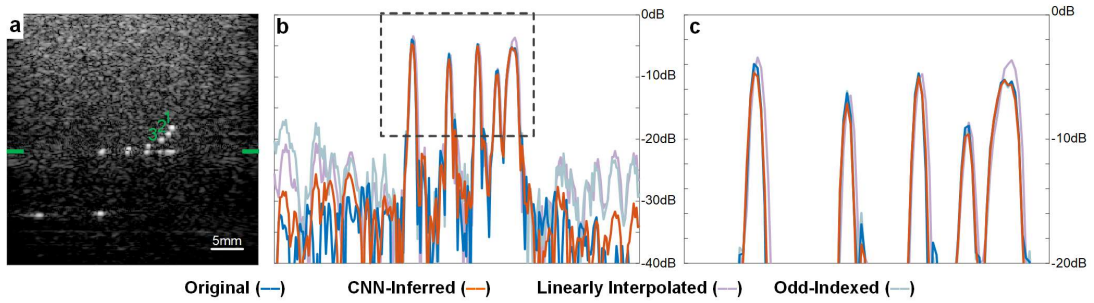


Figure 4.7 (a) B-mode image of point target phantom for 0.25° steering. (b) Cross Section along green lines depicted in (a). (c) Expanded cross Section showing point targets within dashed box. Number labels in (a) correspond to point target lateral FWHM measurements in Table 4.4.

TABLE 4.4
FULL-WIDTH HALF-MAXIMUM (MM) OF POINT TARGETS IN FIGURE 7

Point	Original	CNN-Inferred	Linearly Interpolated	Odd-Indexed
#1	0.58	0.62	0.66	0.59
#2	0.56	0.58	0.66	0.56
#3	0.58	0.62	0.69	0.59
Mean	0.57	0.60	0.67	0.58

4.5.3 Recovery of Carotid In Vivo Lumen-to-Tissue Contrast

Images beamformed from RF data inferred with our method also show improved image quality for in vivo scenarios. Specifically, it can be observed in Figure 4.8 that the images of the human carotid in vivo generated from the full-channel (first column) and the CNN-derived (second column) datasets demonstrate clear visibility of the common carotid artery lumen (center-top of each image). In contrast, in the linearly interpolated (third column) and odd-indexed (fourth column) datasets, the carotid lumen is hidden by artifacts from channel count reduction for both carotids, regardless of the transmit steering.

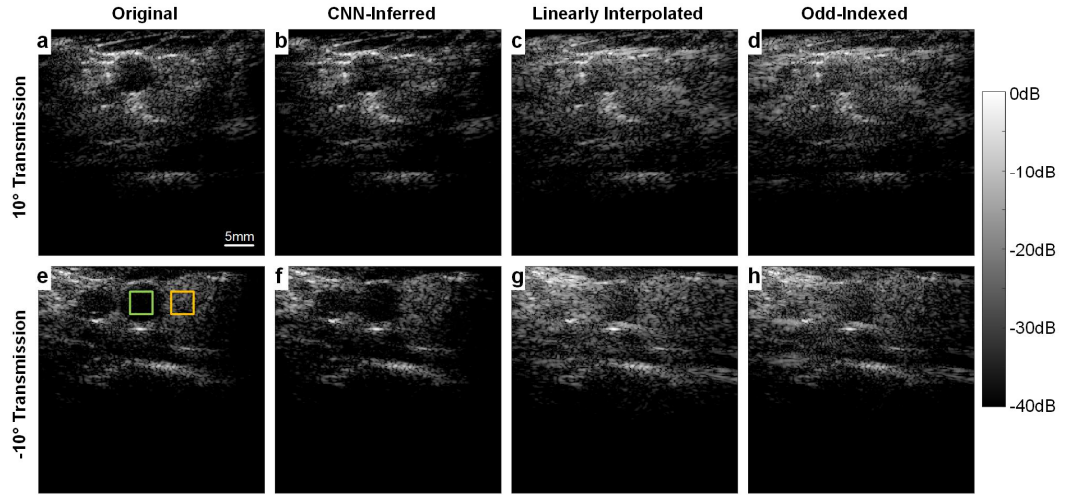


Figure 4.8 Single plane wave transmission B-mode images (40-dB dynamic range) of human carotid beamformed from original full-channel RF dataset (first column), CNN-inferred channel dataset (second column), linearly interpolated channels dataset (third column), and the original odd-indexed channel dataset (fourth column). Two different carotid images are shown for 10° transmission [top row, (a)-(d)] and for -10° transmission [bottom row, (e)-(h)]. Orange box and green box show lumen and tissue regions, respectively, used to calculate lumen-tissue contrast in Figure 4.9.

Lumen-Tissue Contrast Across Transmission Angles

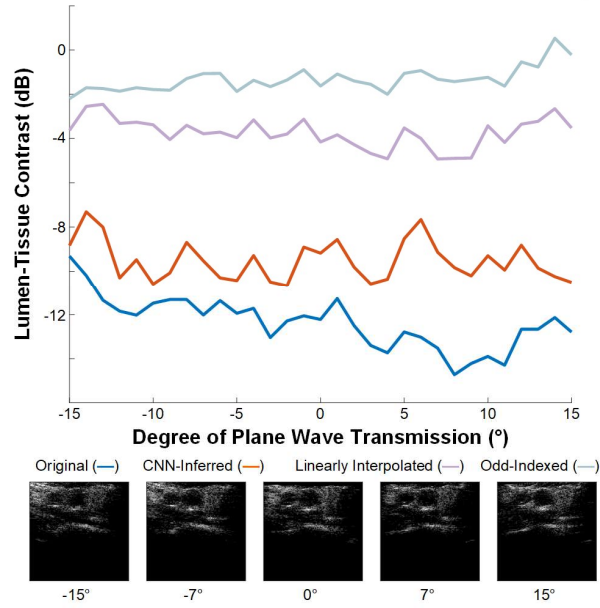


Figure 4.9 Measured lumen–tissue contrast in B-mode images across transmission angles for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and the original odd-indexed channel dataset (light blue). Thumbnails of B-mode images beamformed from CNN-inferred channels shown at bottom.

Examining the lumen-to-tissue contrast across transmission angles for each compared dataset, Figure 4.9 plots the measured contrast difference over $4 \times 4 \text{ mm}^2$ regions of lumen and tissue (yellow and green squares, respectively, in Figure 4.8(e)). On average, the lumen-to-tissue contrast of images beamformed from the CNN-derived dataset (orange line) is within 3 dB of the contrast of images beamformed from the full-channel dataset (dark blue line), while demonstrating between 8- and 10-dB contrast improvement over images beamformed from the odd-indexed channels (light blue line) alone. The lumen-to-tissue contrast of images generated from linearly interpolated data (light purple line) shows minor improvement but does not reach the contrast improvement of the proposed method.

4.5.4 Consistent Image Quality Improvement After Coherent Compounding of Single Transmit Images

B-mode images formed from steered plane wave transmissions were coherently compounded to assess the image quality improvement compared to using only a single transmission. Figure 4.10 and Figure

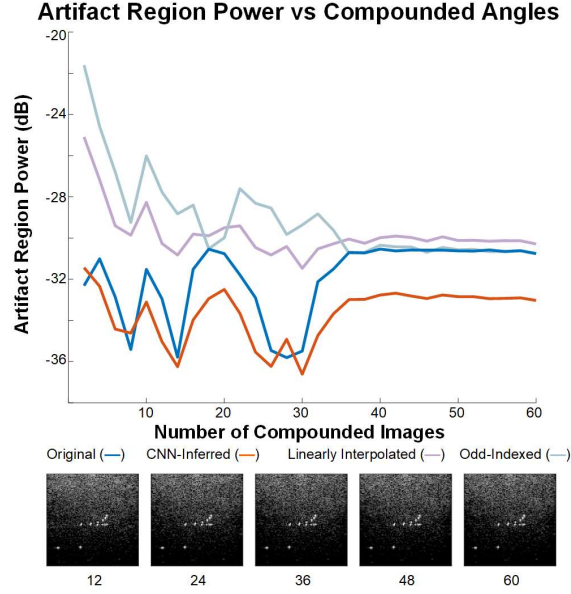


Figure 4.10 Measured spatial aliasing artifact power in B-mode images across a number of coherently compounded single-transmit images for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and original odd-indexed channel dataset (light blue). Thumbnails of compounded B-mode images beamformed from CNN-inferred channels shown at bottom.

4.11 demonstrate the effect of compounding images beamformed with CNN-inferred data using the same imaging scenarios as Figure 4.6 and Figure 4.9, respectively. Figure 4.10 shows that coherent compounding reduces the spatial aliasing artifacts even without inference (light blue line), but our proposed CNN inference scheme results in reduced artifacts with fewer angles compounded (left half of red line). The artifact power is measured via a $4 \times 4 \text{ mm}^2$ region centered on the peak spatial aliasing artifacts on the left side for each compounded image. It is notable that the images beamformed with CNN-inferred data have lower average power in the artifact region when many angles are compounded (right half of red line). This result is attributed to the lack of structured scattering (see Section 4.5.1) changing the speckle pattern. Figure 4.11 shows that coherent compounding also improves visibility of the lumen without inference (light blue line), but CNN inference (red line) results in consistently improved (6–8 dB) lumen–tissue contrast overall. The contrast was measured over the same $4 \times 4 \text{ mm}^2$ regions of lumen and tissue (yellow and green squares, respectively, in Figure 4.8(e)). In both Figure 4.10 and Figure 4.11, linear interpolation (light purple line) slightly improves the

measured metrics compared to no channel inference (light blue line), but not to the degree of the proposed method.

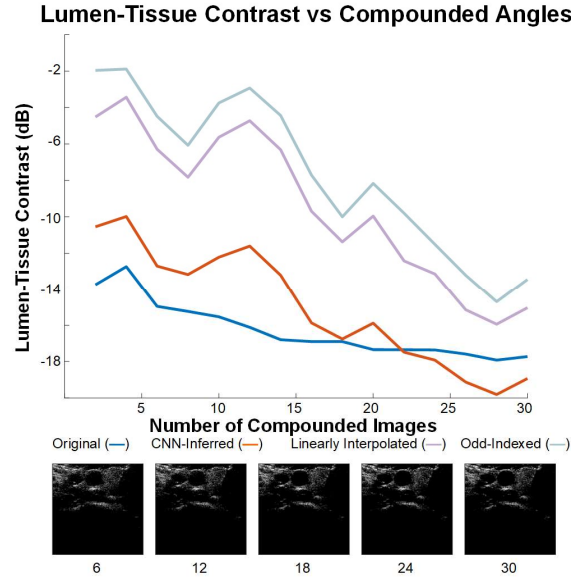


Figure 4.11 Measured lumen–tissue contrast in B-mode images across a number of coherently compounded single-transmit images for original full-channel RF dataset (dark blue), CNN-inferred channel dataset (orange), linearly interpolated channels dataset (light purple), and original odd-indexed channel dataset (light blue). Thumbnails of compounded B-mode images beamformed from CNN-inferred channels shown at bottom.

4.5.5 Real-Time Computational Capacity

Using the TensorFlow-GPU framework (version 1.12.0) running on an Intel Xeon E5-2620 CPU (Intel, Santa Clara, CA, USA) and an Nvidia RTX 2080 GPU (Nvidia, Santa Clara, CA, USA), an inference rate of 4 ms is achieved per 64 channels of RF data. This rate corresponds to a 250-fps inference throughput on our system.

4.6 Discussion

4.6.1 Convolutional Encoder–Decoder as a New Framework for Channel Data Recovery

A primary advantage of ultrasound over other medical imaging modalities, such as CT or MRI, lies in its ability to perform point-of-care imaging (Wells, 2006). Given its bedside applicability, the novel

imaging insights provided by high-frame-rate imaging (Tanter & Fink, 2014) are most effective when they are available to the sonographers in real time. However, two barriers in enabling real-time high-frame-rate imaging are: 1) the immense data transfer rate to move the data from the sampling front end to the processing back end within the imaging system and 2) the computational cost of real-time beamforming at high frame rates.

Our work addresses the data transfer challenge of real-time high-frame-rate imaging via channel count reduction without sacrificing image quality. Simply reducing the transducer receive channel count results in a loss of post-beamformed image quality due to the introduction of artifacts and the loss of contrast. To compensate for the channel reduction, a neural network was trained, using over 60 000 *in vivo* sets of plane wave data (see Table 4.1), to infer unacquired RF channel data from the acquired channel RF data. An encoder–decoder network structure (see Figure 4.2) was designed to leverage the hyperbolic structure of the acoustic echoes to regenerate data from unacquired channels. Examining only the inferred RF data (see Figure 4.4), our CNN correctly captures the expected spatiotemporal structure. After CNN inference, a set of full-channel RF data can be formed by interleaving the acquired and inferred channel data (see Figure 4.1). B-mode images beamformed from the regenerated RF data closely matched the images beamformed from the full-channel RF data in grating lobe artifact power (see Figure 4.6), point target resolution (see Figure 4.7), and lumen–tissue contrast (see Figure 4.9), even when coherent compounding was applied (see Figure 4.10 and Figure 4.11).

To minimize CNN inference time for real-time performance, only nine convolutional layers were employed, leading to an inference rate of 250 RF fps on our system when executed on a GPU. Such inference rate can support both a live imaging mode and replay mode of a real-time high-frame rate imaging system, such as the one described in (Yiu et al., 2019). To form a functional system that involves the entire signal processing chain from channel data recovery to image formation, the computational challenge of real-time high-frame-rate beamforming must also be addressed. Previous works have already demonstrated leveraged GPUs for real-time image formation (Yiu et al., 2011). Thus, we anticipate that a two-GPU system, where one GPU performs the inference before passing the full set of RF data to the second GPU for beamforming, can help realize such a system.

Our proposed data reduction and recovery framework has several advantages. First, the network demonstrated angle independence on data from steering angles ranging from -15° to $+15^\circ$. As such,

the same network can be used for inference from all the acquired angles. This capability represents an improvement over our preliminary results (Xiao et al., 2019) that only inferred unsteered plane wave data using the same number of layers. Second, the trained CNN showed high generalizability across imaging scenarios. Reconstructed images were derived from a wire phantom *in vitro* (see Figure 4.5), but the training data consisted solely of *in vivo* samples that contained no clear point targets, highlighting the generalizability of our framework. Third, the ease of training data formation is another important advantage of the proposed framework. While our method falls under the umbrella of supervised learning within the topical theme of deep learning, our training input–output pairs (odd/even indexed RF channels) are formed without any user-assisted labeling (see Figure 4.3). One consequence of this feature is that for a given ultrasound imaging system, all plane wave RF data acquired without data reduction can then form a new training pair to further improve the network.

4.6.2 Impact of a Pure RF-to-RF Framework

Conventionally, when receiving acoustic echoes, the ultrasound transducer pitch obeys the Nyquist spatial sampling criterion. Violation of the Nyquist sampling rate leads to spatial aliasing, but other works have shown that the Nyquist rate can be overcome given knowledge of signal structure (Candes et al., 2006; Mishali & Eldar, 2010). In Appendix B, we show based on physical principles that spatiotemporal structure exists in received acoustic echoes from a plane wave ultrasound transmission. Our experimental results showed that the proposed deep learning method successfully inferred the RF data structure using convolutional kernels. According to the parameters of system used for this study (see Table 4.1), the fundamental frequency was 5 MHz (wavelength of 0.308 mm at 1540-m/s sound speed), while the effective receive pitch was double of the wavelength (i.e., 0.61 mm for the pitch of odd-indexed channels). Given this discrepancy, our results suggest that the Nyquist limit for transducer element pitch is too conservative in ultrasound, given an appropriate compensation technique. This conclusion is in the vein of similar insights from works, such as (Cohen & Eldar, 2018) or (Wilcox & Zhang, 2018). Additionally, because CNNs are universal function approximators (Zhou, 2020), there may exist undiscovered alternative nonlearning methods, which can also perform RF-to-RF data regeneration. The question of how little data is too little is an unanswered one.

The value of our framework’s RF-to-RF nature is its ability to output raw RF data. This benefit can potentially enable the derivation of high-frame-rate imaging insights outside of B-mode imaging. Some high-frame-rate ultrasound methods, such as color Doppler (Demené et al., 2015), vector flow imaging

(Jensen et al., 2016b; Lenge et al., 2015), or shear wave elastography (Bercoff et al., 2004), rely on the availability of the raw RF data or the phase to implement custom signal processing techniques to extract the desired information. Our proposed framework attempts to regenerate the raw RF data, unlike other methods to reduce channel count, which focus only on improving B-mode image quality, such as (Cohen & Eldar, 2018) or (Synnevag et al., 2009). The fidelity of the reconstructed RF data (see Fig. 4) lends credence to our data reduction framework finding applications in non-B-mode ultrasound imaging and evaluating our framework’s applicability to color Doppler is an ongoing effort. The ability to reduce the data on acquisition while maintaining a full RF dataset for image formation is essential for real-time display in a clinical setting.

4.6.3 Limitations and Perspectives on Future Work

Our proposed framework demonstrated millisecond inference frame rates (Section 4.5.5). However, true real-time high-frame-rate ultrasound requires submillisecond inference to achieve 1000-fps imaging. There are several avenues to pursue to further improve the inference speed without compromising the fidelity of the regenerated RF data. Without modifying the network architecture, inference frameworks such as Nvidia’s TensorRT (Nvidia, Santa Clara, CA, USA) can perform network optimizations and have shown up to twofold speedup of inference. Apart from software acceleration, next-generation hardware, such as Google’s tensor processing unit (Google, Mountain View, CA, USA), is being produced and specifically tailored for neural network operations. These hardware advances can be integrated into an ultrasound open platform to add efficient CNN operations to the ultrasound imaging pipeline. Another possibility is to reduce the number of layers or kernels and correspondingly decrease the inference time. Nonetheless, in doing so, data recovery performance may suffer as the network learns to capture less.

High-frame-rate ultrasound imaging is also not limited to the imaging system used in this study. While this work was demonstrated using a specific probe (L14-5) and imaging system (SonixTouch), the proposed deep learning framework can be generalized to other linear probes and systems through acquisition of new data and network retraining. The existence of spatiotemporal redundancy in the RF data and the insights of applying an encoder–decoder architecture by treating the RF data as an image frame (i.e., a 2-D data array) are widely applicable. A more challenging barrier is expanding the framework beyond linear array probes as the current framework relies on applying spatially invariant convolutional kernels on raw RF data. Curvilinear probes, for example, show an increase in physical

distance in neighboring channels with increased imaging depth. One suggestion to extend our work for curvilinear arrays is through scan conversion of the raw RF data before inference. Yet, this approach may degrade the fidelity of the data through excess interpolation.

In this work, the total data transfer was halved. Further reduction may be achieved via two means. First, the receive channels can potentially be decimated three or even four times with some modifications to the neural network, as the principles of data structure redundancy (see Appendix) are still valid for those scenarios. One option is to use a branched convolutional neural network architecture, maintaining the encoder subnetwork and adding multiple decoder subnetworks for multiple degrees of channel decimation. Second, as our framework is purely RF-to-RF, it can be used in conjunction with other data reduction schemes. The proposed method does not currently require hardware modifications but could potentially be applied together with a hardware-based compression scheme (Lok et al., 2015). It remains to be investigated what the practical limit for channel reduction is for ultrasound imaging.

High-volume-rate 3-D ultrasound imaging is an area where data transfer reduction is even more necessary and is a natural application direction for our framework. Densely populated matrix probes can have upward of 1000 elements, while sparsely populated probes can suffer from low image quality (Diarra et al., 2013; Provost et al., 2014). Our work provides a possible framework for bridging the gap by acquiring from a sparse matrix probe and then regenerating the RF data to emulate a fully populated matrix probe. This strategy can be realized either by applying the 2-D convolutional network architecture presented in our work one dimension at a time or by redevelopment of a network using 3-D convolutional layers (Shan et al., 2018; Tran et al., 2015) with the possibility of using simulated data for network training.

4.7 Conclusion

Reducing the data transfer rate can alleviate one bottleneck in real-time high-frame-rate ultrasound imaging. In this article, RF data transfer was halved by acquiring data from only half the transducer channels. The subsequent loss of post-beamformed image quality due to channel reduction was mitigated in vitro and in vivo using a deep learning approach to regenerate the RF data from the unacquired channel set. This method not only helps to enable real-time high-frame-rate imaging but also has a natural extension to the 3-D ultrasound imaging domain, where the data transfer rate remains an even larger barrier.

Chapter 5

A New Protocol for Speed-of-Sound Estimation for in Large Muscles: Results From an *in vivo* Study

5.1 Introduction

Ultrasound is a non-invasive and versatile imaging modality that can be used to visualize structures and functions of the human body in real-time (Whittingham, 2007). Musculoskeletal (MSK) assessment is an application of ultrasound which can provide valuable information about the morphology and physiology of muscles, tendons, ligaments, and bones (Backhaus, 2001; Czynny, 2017; Grassi et al., 2004; Peetrons, 2002). MSK ultrasound can be used to identify pathologies such as muscle or tendon tears, bursitis, and tumors or cysts among others (Backhaus, 2001; Peetrons, 2002). Because of the dynamic nature of ultrasound, MSK imaging can also be used interventionally for ultrasound-guided procedures; for example, to mitigate injection-related injury (Grassi et al., 2004).

Not only is MSK ultrasound useful for identifying acute pathologies, but it can also be used as an indicator of overall health through quantitative measurement of muscle quality or body composition (Nagae et al., 2023). For example, muscle mass or thickness can act as a predictor of duration of intensive care stays or for age-related muscle decay (Paris & Mourtzakis, 2016; Strasser et al., 2013). Because of the importance of muscle as a predictor of health, and the role muscle plays in health and disease, quantifying muscle quality has become an important area of research (Mourtzakis et al., 2017; Nagae et al., 2023; Wolfe, 2006). Thus, efforts have been made to look not only at the general structure of MSK ultrasound images, but also to derive quantitative metrics based on the ultrasound images to assess muscle quality (Brandenburg et al., 2014; Harris-Love et al., 2014; Parry et al., 2015). These metrics can include site-specific measurements such as muscle thickness, echogenicity, stiffness, texture, and pennation angle (Drakonaki et al., 2012; Nagae et al., 2023; Nodera et al., 2018; Paris et al., 2017). These quantitative parameters have been correlated with physiological pathologies of sarcopenia, muscle atrophy and more (S. E. Campbell et al., 2005; Minetto et al., 2016; Nodera et al., 2018). Some parameters such as muscle thickness are directly measured (Paris et al., 2017) while others such as echogenicity or texture can be assessed based on B-mode pixel values (Drakonaki et al., 2012; Young et al., 2015). While quantitative, these measurements can depend on the ultrasound scanner

hardware, imaging parameters, and operator skills for assessment (Paris et al., 2017; Young et al., 2015).

Tissue speed-of-sound (SoS) can act as a robust measurable quantity that reflects an intrinsic property of the muscle tissue. Tissue SoS has shown potential as a tissue biomarker for conditions such as osteoporosis, breast cancer, and liver steatosis (Imbault et al., 2017; Li, Duric, et al., 2009; Rossman et al., 1989). In the context of MSK ultrasound specifically, it has been shown to be related to pathologies such as edema and sarcopenia (Sanabria et al., 2019; Topchyan et al., 2006). However, measuring tissue SoS in vivo can require niche hardware, such as an ultrasound computed tomography system (Wiskin et al., 2020), which is not widely available, or a through-transmission setup (Cummings et al., 2002; Topchyan et al., 2006), which does not have imaging potential. Moreover, these methods may not be suitable for some applications of MSK ultrasound, as they are limited by the bilateral nature of the imaging window. SoS frameworks that operate in pulse-echo mode with standard ultrasound probes can expand the clinical potential of SoS estimation. Some of these frameworks use local SoS imaging methods, which can produce a SoS mapping of the tissue (Ali, Brevett, et al., 2022; Jaeger et al., 2015; Sanabria et al., 2018). These methods are often computationally intensive, and thus limited in real-time capabilities. Global SoS methods can be more tractable through simpler modelling (Anderson et al., 2000; Krucker et al., 2004), however few have been demonstrated in real-time and in vivo accuracy can be limited (Krucker et al., 2004). Therefore, a framework that can measure the SoS in real-time with a standard ultrasound probe is necessary to translate SoS insights for augmenting MSK ultrasound assessment of muscle quality.

In this work, we investigate the use of a real-time global SoS estimation framework in vivo human MSK ultrasound. We designed a study and recruited 40 human participants, then measured the SoS of the biceps, quadriceps, and calf muscles. The data was collected on both the left and right appendages and landmarking based on bone markers were used to ensure consistent probe placement across the study. We also collected information on demographics such as age, sex, and physical activity. Using the acquired data, we first refined the site-specific SoS estimation protocol to identify which imaging views were best suited for robust SoS measurement. Once the appropriate imaging views were selected, we used those views to establish statistical relationships between the measured SoS and the demographic factors. Using our pulse-echo real-time framework, we discovered similar relationships between sex and body-mass index (BMI) with SoS as compared to through-transmit SoS systems (Ruby

et al., 2020; Topchyan et al., 2006). Overall, we aim to demonstrate the viability of using a real-time pulse-echo framework for live assessment of muscle quality, which essential for clinical translation.

5.2 Experimental Methods

5.2.1 Speed-of-Sound Estimation Framework

SoS estimation was performed via our previously developed real-time framework alongside a custom through-transmit device (Chapter 2). In the real-time framework, the SoS is estimated based on post-beamformed differences between unfocused plane wave transmissions. The SoS estimate is selected based on the SoS value that minimizes the differences between images from different steering angles. Figure 5.1 reviews the steps of this framework in detail. First, a batch of steered plane wave data is acquired using 5 MHz single pulse cycle transmissions at angles from -15° to 15° in 3° increments, excluding the 0° angle, for 10 total sets of raw RF (radiofrequency) data per batch (Figure 5.1, step 1). Next, each steered transmission is beamformed at SoS values ranging from 1400 to 1750 m/s, corresponding to the range of SoS values that would be expected for human muscle tissue, resulting in 180 B-mode images (Figure 5.1, step 2). Subsequently, for each SoS candidate the pixel-wise

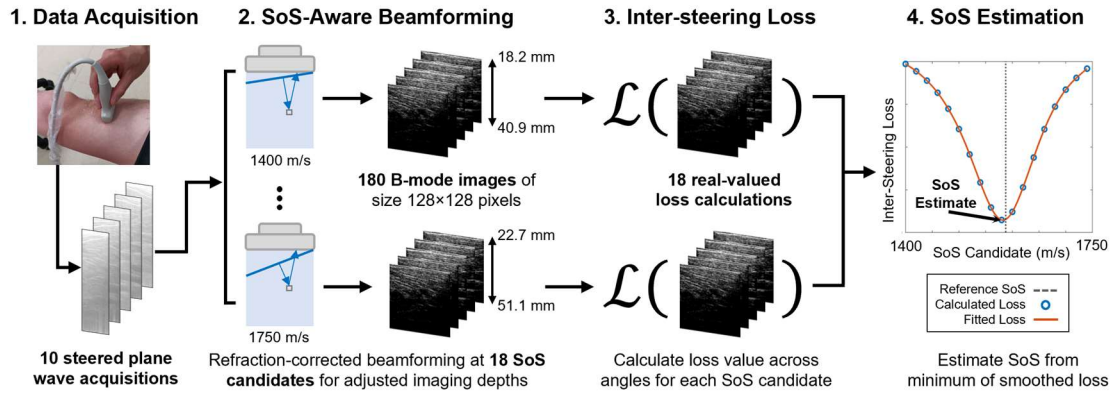


Figure 5.1 SoS estimation is performed by beamforming data from multiple steered plane wave transmissions. 1. Data is first acquired from a set of 10 steered plane waves ranging from -15° to 15° . 2. SoS-aware beamforming is performed at a set of 18 SoS candidates by accounting for the shifts in axial position and the effect of refraction between the probe and the medium. 3. For each SoS candidate, an inter-steering loss (Eq. 2.5) is calculated. 4. The inter-steering loss values are fit to a curve and the minimum of the fitted curve is declared the SoS estimate.

coefficient of variance is evaluated across the angles, then summed to form a set of 18 points representing the image differences, denoted the inter-steering loss (Figure 5.1, step 3). Lastly, a curve is fit to these points and the minimum of this fitted curve is declared the SoS estimate (Figure 5.1, step 4). This framework has demonstrated an average *in vitro* error of $-1.8 \text{ m/s} \pm 2.2 \text{ m/s}$ on homogeneous agar phantoms and an *in vivo* error of $-2.2 \text{ m/s} \pm 11.6 \text{ m/s}$ on human calf muscles (Sec. 2.4.3Chapter 2). This framework was capable of operating in real-time at 18.1 fps (Sec. 2.4.3). For through-transmission SoS measurements, we used a previously developed custom through-transmission device consisting of two single-element transducers connected to the lower jaws of a caliper (Sec. 2.3.1). An oscilloscope connected to both transducers was used to measure the time-of-flight of a 5 MHz acoustic pulse (matching the transmitted frequency of the ultrasound scanner) from one transducer to the other, with an algorithm used for automatic time-of-flight selection (Li, Huang, et al., 2009). After calibration, this device was validated in pure water at 22.1 °C, which has a known SoS value of 1488.6 m/s (Bilaniuk & Wong, 1993); for the pure water, the device reading was found to be $1488.9 \text{ m/s} \pm 0.3 \text{ m/s}$. Both ultrasound devices were approved for use on humans (University of Waterloo ORE #44778).

5.2.2 Imaging Hardware and Parameters

The pulse-echo SoS estimation framework was realized using a software-based portable research ultrasound scanner (US4R-Lite; us4us, Poland). This scanner was connected via Thunderbolt 3 to a laptop (ThinkPad X1 Extreme Gen 4; Lenovo; Beijing, China) equipped with a discrete GPU (RTX 3080 Laptop; Nvidia; Santa Clara, CA, USA), enabling real-time GPU processing of the acquired RF data (Figure 5.2(a)). For imaging, the scanner was equipped with an L14-5 linear probe which was used for all data acquisitions. The SoS framework was programmed in Python (ver. 3.8.5) using the CuPy (ver. 10.6.0) library for GPU acceleration; the SoS and B-mode images were displayed to the operator in real-time using the Matplotlib (ver. 3.3.0) library.

The through-transmission SoS measurement framework was constructed using two single-element transducers designed for non-destructive testing (C567; Olympus; Tokyo, Japan). These transducers were affixed using 3D-printed resin attachments to a high-precision electronic caliper (ABSOLUTE; Mitutoyo; Kawasaki, Japan) (Figure 5.2(b)). Transducers were connected to an oscilloscope (DS1104Z; Rigol; Beijing, China) for time-of-flight measurement; one transducer was also connected to an arbitrary waveform generator (SDG2042X; Siglent; Solon, OH, USA). Matlab (ver. 2022b) was used

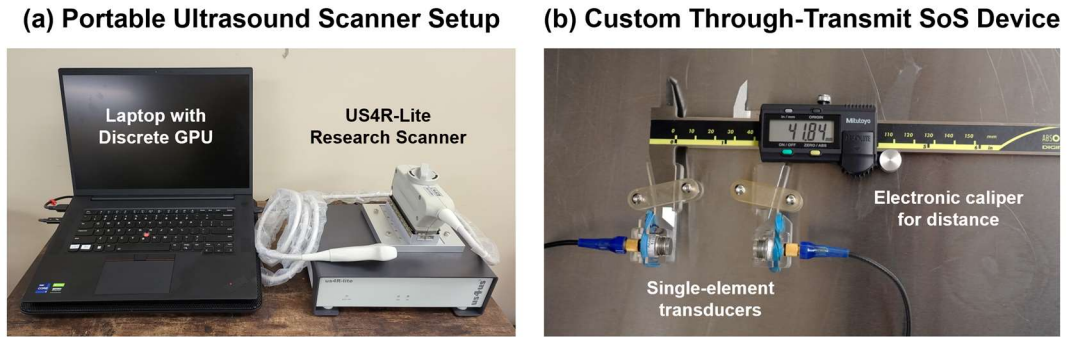


Figure 5.2 Equipment used for SoS measurement. (a) Portable ultrasound scanner operating in pulse-echo mode connected via Thunderbolt 3 to a laptop with a discrete GPU for live data processing. (b) Custom through-transmit SoS estimation device.

to process the acquired signals; the calculated SoS and the detected waveforms were displayed to the user at the time of measurement.

5.2.3 Recruitment of Sample and Survey Details

A total of forty adult volunteer participants were recruited for the purposes of this study. The volunteers were informed of their consent prior to the study (ORE #44778). The only exclusion criteria used for requirement was the requirement to be 18 years of age or older. Each participant was asked to fill out a questionnaire before any ultrasound was applied. Conditions that would acutely affect the muscles were noted on the questionnaire. Participants were asked for their age and sex. The participants were

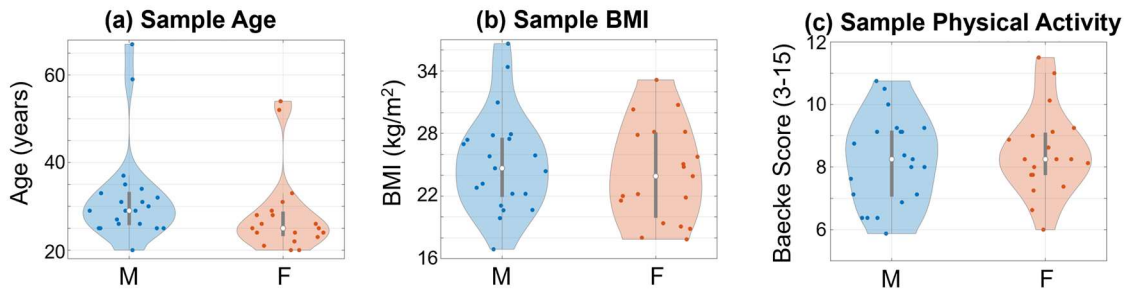


Figure 5.3 Demographics distribution for sampled population. (a) Sample age. (b) Sample body mass index (BMI) (c) Sample physical activity as measured by the Baecke questionnaire.

then asked to fill out a questionnaire regarding their physical activity profile based on a well-regarded measure from the literature (Baecke et al., 1982). The physical activity score (Baecke score, range 3-15) was recorded as three indices (work, sport, leisure) each ranging from 1-5. Lastly, the participants were asked to remove their shoes and their height and weight were measured and body mass index (BMI) was calculated. Figure 5.3 summarizes the measured demographics of the sample, separating the measurements by sex. All demographics (age (Figure 5.3(a)), BMI (Figure 5.3(b)), Baecke physical activity score (Figure 5.3(c))) were approximately Gaussian distributed, as tested by the Kolmogorov-Smirnov test of normality; the Kolmogorov-Smirnov test resulted in the null hypothesis with p-values above 0.1 for all six of the violin plots shown in Figure 5.3.

5.2.4 Measurement of Large Muscles (Calf, Biceps, Quadriceps) via Ultrasound

Ultrasound data was acquired from for every participant both the left and right appendages on the biceps, calf, and quadriceps muscles. Each appendage was first landmarked via manual palpation as

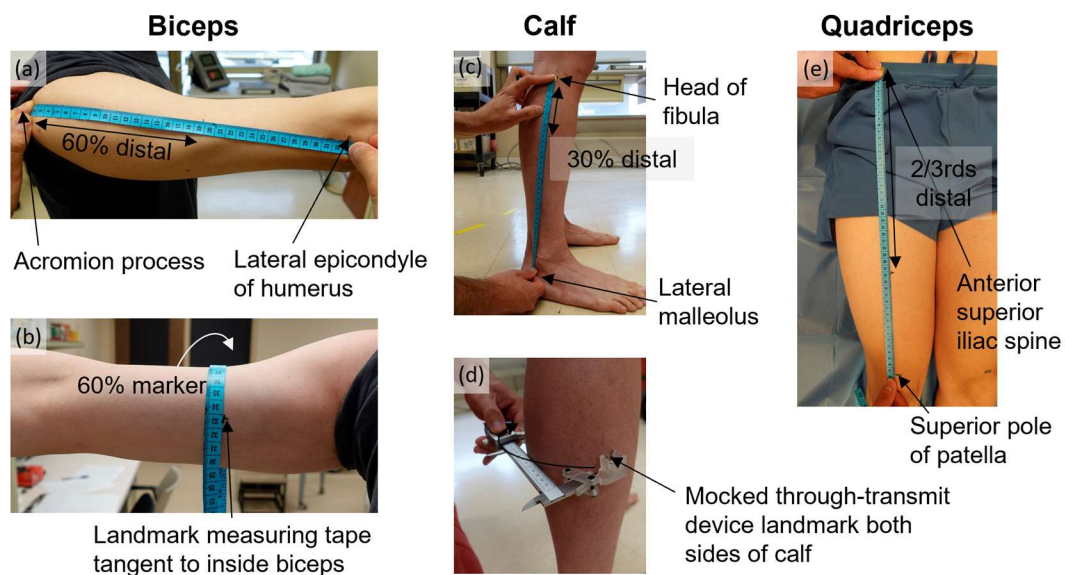


Figure 5.4 Images of landmarked regions for imaging. (a) The posterior surface was marked temporarily. (b) The tangent point of the resting measuring tape on the anterior surface was marked for imaging. (c) The posterior surface of the calf was marked temporarily. (d) A mocked through-transmit device was used to mark both sides of the calf for imaging. (e) The quadriceps was temporarily marked above the rectus femoris muscle.

shown in Figure 5.4 to ensure a consistent probe placement across all participants. The landmarking was performed according to previously reported standards for ultrasound imaging of the large muscles for quality and quantity assessment (Paris et al., 2017; Takai et al., 2014). For the biceps, the posterior surface (Figure 5.4(a)) was temporarily marked with easily removable ink based on 60% of the distance to the lateral epicondyle of the humerus from the acromial process of the scapular. The anterior surface then was marked with easily removable ink for imaging using the last point of contact of the measuring tape to the surface of the bicep (Figure 5.4(b)). For the calf, both sides of the calf were marked with easily removable ink for imaging at 30% of the distance to the lateral malleolus from the head of the fibula (Figure 5.4(c)), based on the placement of the mock through-transmit device (Figure 5.4(d)) to

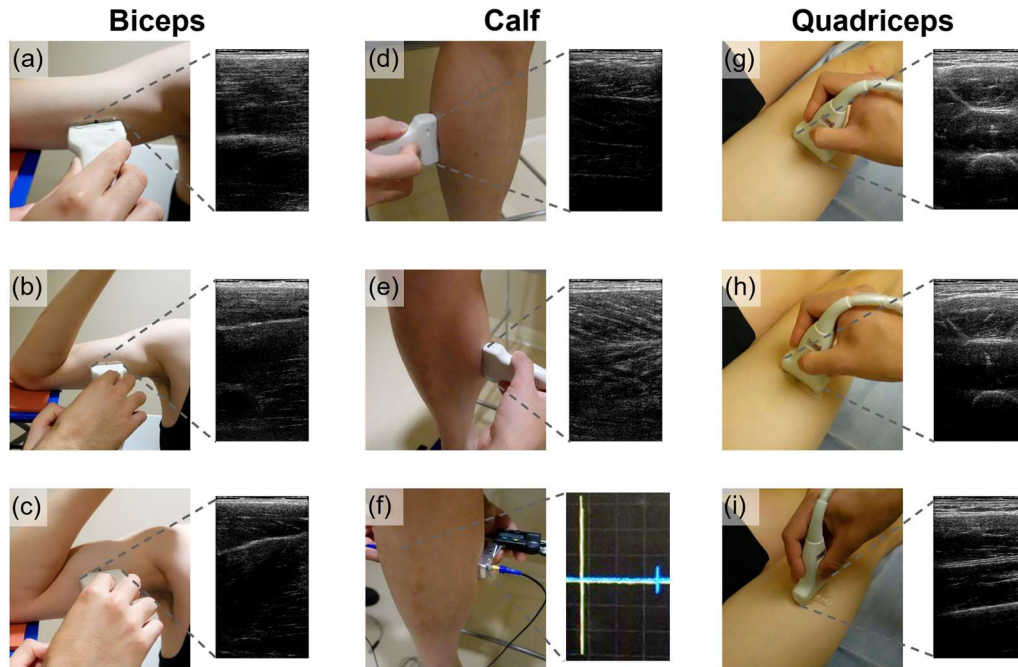


Figure 5.5 All imaging views used during the study for ultrasound data acquisition along with corresponding image thumbnails (a) Anterior surface of biceps in extended position. (b) Anterior surface of biceps in flexed position. (c) Anterior surface of biceps in flexed position with muscle contraction. (d) Lateral surface of calf, leg resting at 90° angle. (e) Medial surface of calf, leg resting at 90° angle. (f) Through-transmit acquisition signal from oscilloscope. (g) Cross-sectional view of rectus femoris muscle, leg extended and relaxed. (h) Cross-sectional view of rectus femoris muscle, leg extended and contracted. (i) Long-axis view of rectus femoris muscle, leg extended and relaxed.

ensure the same imaging view across both the ultrasound scanner and the custom device. For the quadriceps, the removable ink mark for imaging was placed two-thirds the distance from the anterior superior iliac spine (ASIS) to the superior pole of the patella (Figure 5.4(e)).

Ultrasound data was acquired based on the landmarked points for imaging, with the probe placement for different muscles shown in Figure 5.5 alongside thumbnail examples of the corresponding B-mode images. For the biceps, we acquired data from three stages of muscle flexion; with the biceps extended (Figure 5.5(a)), with biceps flexed (Figure 5.5(b)), and with the biceps flexed with contraction (Figure 5.5(c)). For the calf, the probe was placed based on both the lateral (Figure 5.5(d)) and medial (Figure 5.5(e)) sides of the calf for ultrasound acquisition. For the through-transmit reference measurements, the custom device was placed on the appropriately landmarked position (Figure 5.5(f)). For the quadriceps, all ultrasound data was acquired from the quadriceps in an extended position, centered on the rectus femoris muscle. Two states of muscle flexion were assessed; one cross-sectional view of the rectus femoris in a relaxed state (Figure 5.5(g)) and one cross-sectional view of the rectus femoris in a contracted state (Figure 5.5(h)). One additional view of the rectus femoris was acquired using a longitudinal view (Figure 5.5(i)). Each ultrasound acquisition totalled approximately one second of duration with 15 sequential data sets for the same view. The reported SoS for each muscle was the mean over the 15 acquisitions. For all views, one acquisition was acquired with minimal possible pressure and an abundance of gel (Puthuchear et al., 2015), while the other was acquired with the minimal amount of pressure required for the probe to be fully in contact with the skin. In total, the number of ultrasound data sets was 32 per volunteer (3 biceps, 2 calf, 3 quadriceps for both left and right appendages for both minimal probe pressure and full contact with skin probe pressure). For the through-transmit device, five separate readings were assessed and the evaluated SoS was the mean of all five readings.

5.2.5 Evaluation of New Speed-of-Sound Estimation Protocol for Musculoskeletal Imaging

To evaluate the ideal imaging views for global SoS estimation, we analyzed the data via four different viewpoints. First, we examined the *inter-sample reliability* by assessing which imaging views across participants had the lowest average standard deviation and coefficient of variation of the estimated SoS over the one second acquisition. The SoS was first estimated individually for all 15 sets of RF data acquired during the second-long scan duration, then those standard deviations and corresponding

coefficients of variation was assessed averaged across both the left and right appendages for each imaging view. Second, we examined the *intra-sample reliability* by examining data from a single volunteer. We repeated the entire set of 32 acquisitions five times on this single individual, then assessed which imaging views had the most standard deviation and coefficient of variation over the repeated acquisitions. Third, we assess differences in dominant versus nondominant appendages via paired t-tests applied to the estimated SoS values for the reported dominant and nondominant appendages. Fourth, we assessed whether the amount of pressure applied to the probe affected the SoS estimates from our framework. Using these four indicators, we selected imaging views for each of the three imaged muscles that demonstrated the highest consistency for further statistical analysis.

5.2.6 Statistical Analysis of Physiological Relationships for Muscle Speed-of-Sound

Using the most consistent imaging views for analysis for the three muscle groups, we examined relationships between muscle SoS and the recorded demographic factors. Most analyses were performed in sex-separated fashion; this approach was taken because females typically have a thicker subcutaneous adipose tissue layer due to sexual dimorphism (White & Tchoukalova, 2014), which can impact the SoS estimation algorithm. For BMI, physical activity, and age we performed linear regression analysis and report the Pearson correlation coefficient and the p-value. We performed an unpaired t-test to evaluate muscle SoS differences between sexes, and a paired t-test to evaluate muscle SoS differences between states of flexion, reporting the p-value for all tests. Lastly, we used the through-transmission calf SoS measurements to compare with the pulse-echo SoS estimates from the real-time framework. We perform linear regression analysis on the reference versus the estimate and on the BMI versus the signed difference between the reference and estimate. We note that the through-transmission does not act a true reference, as the transmission traverses the entirety of the calf muscle, while the pulse-echo estimate is limited to the imaging view; however, *in vivo* evaluation can provide insight into the behaviour of the SoS estimation framework. For all statistical results comparing the SoS and the demographics of the sample, analysis was performed based on the recorded live SoS estimate at the time of participation.

5.3 Results – Establishment of Imaging Protocol

5.3.1 Highest Inter-Sample Reliability Views of the Biceps, Calf, and Quadriceps

To quantify the inter-sample reliability, the amount of variation over the approximately second-long acquisition for each muscle is assessed using the standard deviation (σ) and coefficient of variation (σ/μ). The average standard deviation and average coefficient of variation for each of the muscle imaging views across all participants is shown in Table 5.1. The most reliable views were found to be the biceps in flexed position, the longitudinal view of the quadriceps in extended position with no contraction, and the medial view of the calf (most reliable views are bolded in Table 5.1). The cross-sectional views of the quadriceps had the least reliability over the second-long acquisition.

TABLE 5.1
INTER-SAMPLE RELIABILITY OF MUSCLE IMAGING VIEWS

Biceps			Calf			Quadriceps		
Imaging View	σ (m/s)	σ/μ	Imaging View	σ (m/s)	σ/μ	Imaging View	σ (m/s)	σ/μ
Extended	4.2	2.7e-3	Medial	4.8	3.1e-3	Extended, Cross-Section	10.3	6.7e-3
Flexed	2.5	1.6e-3	Lateral	6.5	4.2e-3	Contracted, Cross-Section	11.0	7.2e-3
Flexed, Contracted	3.4	2.2e-3				Extended, Longitudinal	5.0	3.2e-3

5.3.2 Highest Intra-Sample Reliability Views of the Biceps, Calf, and Quadriceps

To quantify the intra-sample reliability, the amount of variation over multiple acquisitions for the same participant is assessed, again using the standard deviation (σ) and coefficient of variation (σ/μ). The intra-sample reliability for each of the imaging views is shown in Table 5.2. The most reliable views were again found to be the biceps in flexed position, the longitudinal view of the quadriceps in extended position with no contraction, and the medial view of the calf (most reliable views are bolded in Table

TABLE 5.2
INTRA-SAMPLE RELIABILITY OF MUSCLE IMAGING VIEWS

Biceps			Calf			Quadriceps		
Imaging View	σ (m/s)	σ/μ	Imaging View	σ (m/s)	σ/μ	Imaging View	σ (m/s)	σ/μ
Extended	5.5	3.5e-3	Medial	2.0	1.5e-3	Extended, Cross-Section	7.5	4.8e-3
Flexed	3.4	2.1e-3	Lateral	5.1	3.9e-3	Contracted, Cross-Section	8.8	5.6e-3
Flexed, Contracted	3.7	2.3e-3				Extended, Longitudinal	3.9	2.5e-3

2). The cross-sectional views of the quadriceps again had the least reliability over the second-long acquisition.

5.3.3 Full Contact of Probe with Skin is Important for Estimating the Muscle Speed-of-Sound

The amount of pressure applied by the sonographer is an important factor to consider for quantification of muscle quality. Figure 5.6 shows examples of B-mode calf images taken with minimal pressure (Figure 5.6(a)), with the pressure required for the probe to be fully in contact with the skin (Figure 5.6(b)), and with strong applied pressure (Figure 5.6(c)). We observe in the minimal pressure image (Figure 5.6(a)) that the layer of ultrasound gel is visible between the probe and the skin to maintain the minimal pressure. Comparatively, when the probe is fully in contact with the skin, regardless of pressure, there is no layer of ultrasound gel; moreover, the estimated SoS values are higher in the examples with full contact compared to with minimal pressure. Over the entire sample population only the minimal pressure and full contact views were acquired for all participants. To quantify the effect of minimal pressure versus full contact pressure further, the SoS difference between minimal pressure and full contact pressure was assessed across all participants. For the three biceps views, the average difference was $20.3 \text{ m/s} \pm 11.7 \text{ m/s}$ lower in the minimal contact scenarios compared to the full contact

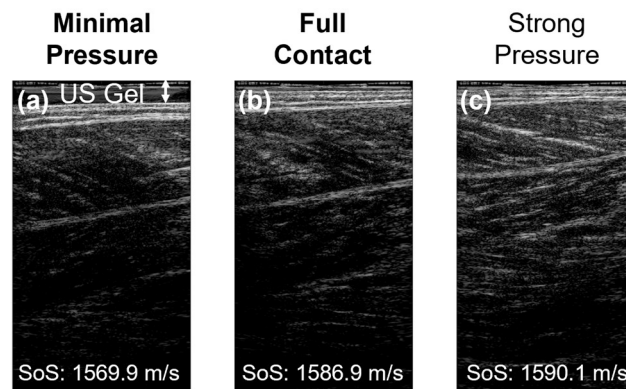


Figure 5.6 Effect of applied probe pressure on an imaging view of the calf and corresponding SoS estimates (a) Minimal pressure view with an excess of ultrasound gel, leading to layer of gel between skin surface and probe. (b) View using minimal amount of pressure required for probe to be fully in contact with the skin. (c) View with additional pressure applied beyond what is required for probe to be fully in contact with the skin.

scenarios. For the quadriceps views, the difference was $10.6 \text{ m/s} \pm 16.8 \text{ m/s}$ lower in the minimal contact scenarios compared to the full contact scenarios. For the calf views, the difference was $16.8 \text{ m/s} \pm 13.9 \text{ m/s}$ lower in the minimal contact scenarios compared to the full contact scenarios. Overall, minimal pressure resulted in over 10 m/s on average lower SoS estimation compared to full contact pressure.

5.3.4 Minimal Differences Between Dominant and Non-Dominant Appendages

SoS differences between dominant and non-dominant appendages are relevant for establishing ideal imaging views for muscle SoS estimation. To quantify the difference between the self-reported dominant and nondominant appendages, Figure 5.7 highlights the differences between the SoS values for the dominant and nondominant appendages for each sex. The SoS estimates for the boxplots are derived from the most reliable imaging views assessed in the previous subsections. Overall, there is no significant difference ($p > 0.05$) for all imaging views as assessed by paired t-tests, except for the male biceps view (Figure 5.7(a)), where there is a significant difference ($p = 3.6\text{e-}5$) between the dominant and nondominant appendages.

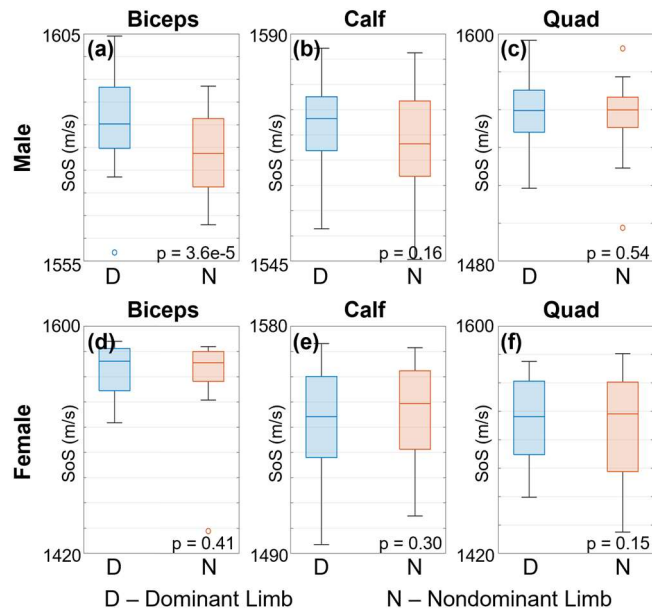


Figure 5.7 Box-and-whiskers plots comparing the estimated SoS values for the dominant and nondominant appendages. Boxes correspond to lower and upper quartile, while whiskers correspond to nonoutlier maximum and minimum. The p-values shown are reported based on paired t-tests.

5.3.5 Summary of Ideal Imaging Views for Speed-of-Sound Estimation

From these results, we identify one imaging view for further statistical analysis on the estimated SoS for each imaged muscle. For the biceps, we selected the biceps of the dominant arm, imaged in a flexed but not contracted state. For the calf, we select the calf of the dominant leg, averaging the SoS of the inside and outside views. For the quadriceps, we selected the longitudinal view of the rectus femoris in an extended and relaxed state. For all views, we used the imaging view where the probe was fully in contact with the skin. These imaging views were selected because of the relative high reliability and independence of imaging view on appendage dominance.

5.4 Results – Statistical Analysis of Physiological Relationships

5.4.1 High Correlation Between Body Mass Index and Muscle Speed-of-Sound for All Measured Muscles

The relationship between BMI and muscle SoS for the identified reliable imaging views for each muscle are shown in Figure 5.8. For the biceps (Figure 5.8(a)), calf (Figure 5.8(b)), and quadriceps (Figure 5.8(c)), there is a clear negative correlation between the BMI and the muscle SoS for both sexes; as the participant BMI increases, the muscle SoS decreases. Further quantifying this relationship, the Pearson correlation coefficient is highest for the biceps ($r = -0.75$) for males and lowest for the quadriceps muscle ($r = -0.54$). For females, the highest correlation coefficient is found in the calf ($r = -0.77$) and the lowest is the biceps ($r = -0.69$). We find a similar BMI-SoS relationship for the calf through-transmit SoS values (Figure 5.8(d)), which matches what is reported in prior through-transmit SoS literature (Ruby et al., 2020).

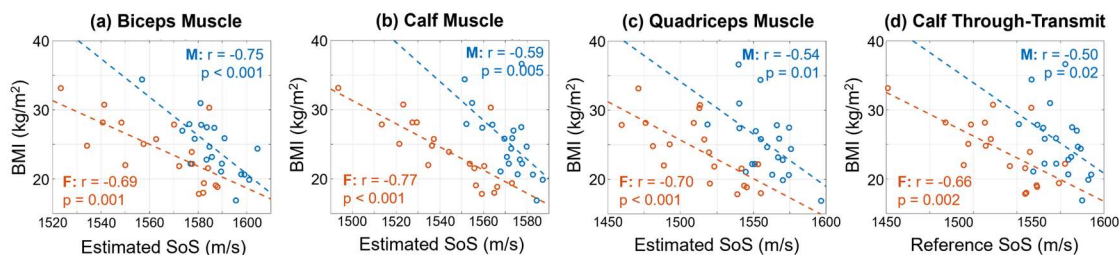


Figure 5.8 Plots comparing the measured muscle SoS with the participant body mass index (BMI). Linear regression was performed for males (M, blue) and females (F, red) separately, with the lines of best fit (dashed lines) and associated Pearson correlation coefficient and p-value reported in the corners of each plot.

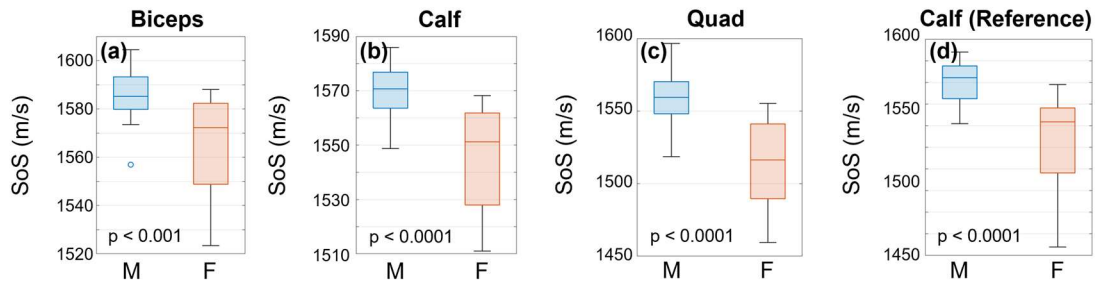


Figure 5.9 Box-and-whiskers plots comparing the estimated SoS values for males and females separated for each muscle group. Boxes correspond to lower and upper quartile, while whiskers correspond to nonoutlier maximum and minimum. The p-values shown are reported based on unpaired t-tests.

5.4.2 Male Muscles Statistically Faster Speed-of-Sound Compared to Female Muscles

Comparing muscle SoS across sexes for the different imaging views, we can see that there is an observed effect of sexual dimorphism on muscle SoS values (Figure 5.9). This is true for all three muscle views (biceps, Figure 5.9(a); calf, Figure 5.9(b); quadriceps, Figure 5.9(c)) and for the through-transmit device as well. Quantifying the difference, unpaired t-tests reveal highly significant ($p < 0.0001$) differences between male and female muscle SoS. We find a similar difference between sexes for the through-transmit calf measurements (Figure 5.9(d)), which again matches what is reported in prior through-transmit SoS literature (Topchyan et al., 2006).

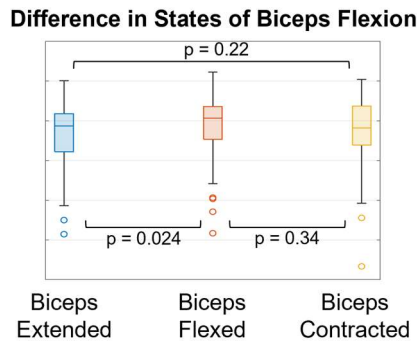


Figure 5.10 Box-and-whiskers plots comparing the estimated SoS values for different states of biceps flexion. Boxes correspond to lower and upper quartile, while whiskers correspond to nonoutlier maximum and minimum. The p-values shown are reported based on paired t-tests.

5.4.3 No Significant Change of Speed-of-Sound with Muscle Activation

The change in estimated SoS with muscle activation was assessed for the three biceps imaging views only. The restriction of statistical analysis to the biceps was due to the low reliability of the cross-sectional quadriceps imaging views (Table 5.1 and 5.2). The change in SoS associated with muscle activation is shown in the box-and-whiskers plots in Figure 5.10. We observed that there was little statistically significant difference in the muscle activation states; paired t-tests revealed that only the extended-flexed comparison (without contraction) had a significant difference ($p = 0.024$).

5.4.4 For Males Only, Significant Correlation was Found Between Physical Activity and Muscle Speed-of-Sound

The relationship between the Baecke physical activity index and muscle SoS for the flexed but uncontracted view for the biceps muscle is shown in Figure 5.11(a). For males, there is a significant correlation ($r = 0.50$, $p = 0.024$) between the estimated SoS and the Baecke index; in contrast, there was no detected correlation between the SoS and the Baecke index for females. A similar male-only detected relationship was found for the calf (M: $r = 0.45$, $p = 0.04$; F: $r = 0.09$, $p = 0.72$) and quadriceps (M: $r = 0.50$, $p = 0.02$; F: $r = 0.16$, $p = 0.52$) imaging views as well. Examining one constituent of the Baecke index in greater detail, Figure 5.11(b) shows the relationship between the muscle SoS and the sport index. Compared to the overall Baecke index, we observed that the relationship between the sport

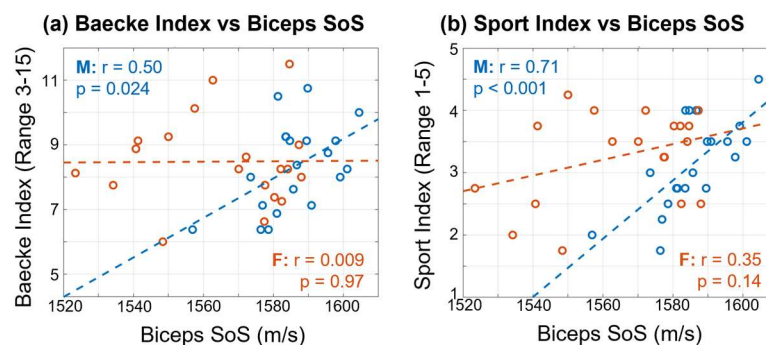


Figure 5.11 Plots comparing the measured biceps muscle SoS with the (a) participant Baecke index and the (a) participant sport index. The sport index is one of three constituents of the Baecke index. Linear regression was performed for males (M, blue) and females (F, red) separately, with the lines of best fit (dashed lines) and associated Pearson correlation coefficient and p-value reported in the corners of each plot.

index and the muscle SoS demonstrates even higher correlation coefficient ($r = 0.50$ versus $r = 0.71$) for males. The other constituent indices are less correlated compared to the sport index.

5.4.5 Estimation Accuracy of Speed-of-Sound Framework

Given the availability of the through-transmit SoS measurements alongside the pulse-echo SoS estimates, the in vivo performance of our real-time framework (Sec. 2.4.3) can be evaluated in greater detail. Figure 5.12(a) shows a comparison between the through-transmit reference SoS and the pulse-echo estimated SoS both the left and right calves. We observed that the global estimation framework consistently reported a lower SoS value compared to the through-transmit reference; however, there was a high degree of correlation (M: $r = 0.81$; F: $r = 0.88$) and similar line of best fit for both sexes. Further analyzing the discrepancy, the signed difference between the SoS estimate and reference was plotted against participant BMI in Figure 5.12(b). We observed that the difference between the through-transmit and pulse-echo SoS values was highly dependent on the participant BMI.

5.5 Discussion

5.5.1 Using Real-Time Speed-of-Sound Measurements to Investigate Physiological Relationships

Muscle quality has predictive power for assessing overall physical health through quantitative metrics such as cross-sectional area and echogeneity (Nagae et al., 2023). These quantitative measurements can

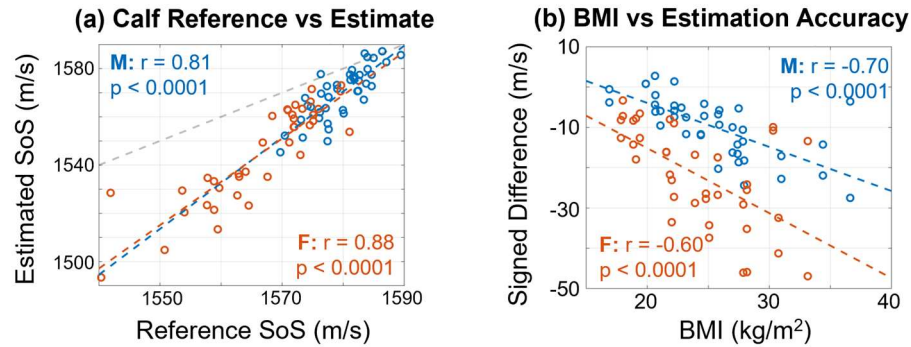


Figure 5.12 Plots assessing the difference between the through-transmit reference SoS values and the pulse-echo estimated SoS values. (a) The reference SoS and the estimated SoS are plotted against each other and results from linear regression are shown. (b) The signed difference between the reference SoS and the estimated SoS are plotted against the participant BMI.

be used to identify pathological conditions (Paris & Mourtzakis, 2016; Strasser et al., 2013); however, they can be prone to system and user error (Paris et al., 2017; Young et al., 2015). Tissue SoS is a quantitative biomarker that can help assess muscle quality as an intrinsic property of the tissue, rather than depending on the imaging system.

In our work, we demonstrated the use of a previously developed real-time pulse echo SoS estimation framework (Chapter 2) for assessing relationships between muscle SoS and demographic factors. We found statistically significant correlations between the biceps, calf, and quadriceps muscles with BMI (Figure 5.8), which is similar to what has been reported in the through-transmit SoS literature (Ruby et al., 2020). Analogously, we found significant differences between male and female muscles for all measured muscles (Figure 5.9), which again is similar to what has been reported in the through-transmit SoS literature (Topchyan et al., 2006). For the different states of biceps flexion, we found no statistically significant differences (Figure 5.10). For males only, we found that physical activity increases muscle SoS (Figure 5.11). Lastly, we showed how our framework will tend to report lower muscle SoS compared to the through-transmit reference (Figure 5.12); however, the difference is both consistent (Figure 5.12(a)) and dependent on participant BMI (Figure 5.12(b)).

There are certain limitations to the scope of the presented work. One challenge was the demographic distribution of the recruited participants; the majority of our participants were in the 20-40 year age range, which limited the analysis of SoS relationships with age. Yet another demographic challenge was the highly sedentary work profile of the participants which led to few changes in the work index component of the Baecke index (Baecke et al., 1982). Another limitation to the study is the analysis of muscle flexion state (Figure 5.10). While the consistency of probe placement was assessed, the shifting of internal structures due to the muscle flexion could lead to differences in the estimated SoS. The difference in subcutaneous adipose tissue due to sexual dimorphism also influenced the results from this study (White & Tchoukalova, 2014), which was also observed in a prior study (Sec. 2.4.1). Particularly, participants with a thicker layer of subcutaneous tissue, which is more common in females, can violate the assumption of homogeneity used for the global SoS estimation algorithm. This effect is a possible cause of the statistical analysis that were significant on males only, such as the positive correlation between SoS and physical activity (Figure 5.11). There is also clear relationship between BMI, which is far from a perfect metric, and the subcutaneous adipose tissue (Kindblom et al., 2009; Kruschitz et al., 2013); this link could explain the increasing difference between the through-transmit reference SoS measurements and the pulse-echo SoS estimates (Figure 5.12). The limitations imposed

by thicker regions of subcutaneous adipose tissue indicate the promising nature of using layer-based or regional models for SoS estimation (Hériard-Dubreuil et al., 2023; Jakovljevic et al., 2018; Stähli et al., 2021); however, the limited real-time capacity of many of these frameworks remains a constraining factor.

5.5.2 A New Measurement Protocol for Speed-of-Sound Characterization in Large Human Muscles

There has previously been no established protocol for identifying ultrasound imaging views ideal for SoS estimation in the muscles. In this work, we adopt an approach based on manual palpation of bone landmarks for identifying ultrasound imaging views previously used for muscle quality and quantity assessment (Figure 5.4) (Paris et al., 2017; Takai et al., 2014). Based on the selected imaging views, we used inter- and intra- sample reliability analysis via the standard deviation and coefficient of variation (Tables 5.1 and 5.2) to identify the most reliable (for SoS estimation) imaging views and states of muscle flexion for the biceps, calf, and quadriceps. For the biceps, the best view was of a flexed but not contracted state. For the calf the best view with the probe placed medially. For the quadriceps, the best view was the longitudinal view of the rectus femoris in an extended and relaxed state. We additionally assessed the effects of applied probe pressure. In contrast to prior measurement of muscle quality with an excess of gel and minimal pressure (Puthuchearry et al., 2015), we found that having the probe fully in contact with the skin was important to avoid underestimation of the muscle SoS; having a thick layer of ultrasound gel can reduce the homogeneity of the imaging view. Because of the demonstrated framerate of our system, which uses 15 views over a second-long acquisition, we believe the overall imaging protocol developed in this work can also be generalized and guide other researchers in ultrasound for muscle SoS.

5.5.3 Potential of Muscle Speed-of-Sound to Act as a Robust Quantitative Tissue Biomarker

While the tissue SoS has been shown to be an effective tissue biomarker (Imbault et al., 2017; Li, Duric, et al., 2009; Rossman et al., 1989), there are many barriers to its clinical adoption. The impact of our work acts to bridge the engineering developments for SoS estimation with the application of SoS estimation at the bedside. By performing all imaging with a real-time system, and all statistical analysis (Sec. 5.4) for physiological relationships on the displayed SoS values, we proved that SoS estimation is feasible at the point-of-care. This capability is essential for the translation of SoS tissue insights to

clinical application. Our study does not address any pathological conditions; however, the connection between physiology and tissue SoS can be observed in the relationships between muscle SoS and demographic factors. The correlations between SoS, BMI, and sex that were observed in the study matched similar research studies in through-transmit SoS literature (Ruby et al., 2020; Topchyan et al., 2006), indicating the prospective ability for a pulse-echo framework to replace through-transmit SoS measurement systems. Moreover, the ability to access specific muscles, such as the rectus femoris in the quadriceps, can be essential for assessing site-specific muscle quality. In this way, additional research on the relationship between the SoS, an intrinsic property of the tissue, and the muscle can be conducted. We plan to assess the relationship more deeply between subcutaneous adipose tissue, muscle, and tissue SoS in future studies. Overall, our results indicate the potential to apply pulse-echo SoS estimation for applications where bilateral access to tissue is impossible, such as the quadriceps or abdomen, and provide simultaneous imaging and SoS estimation capacity.

5.6 Conclusion

Speed-of-sound has high potential as a new, robust quantitative tissue biomarker for assessing muscle quality. Establishment of pulse-echo speed-of-sound estimation in real-time is essential for clinical viability of tissue speed-of-sound as a biomarker. In our work, we used a portable scanner equipped with a live global speed-of-sound estimation algorithm to conduct a study of volunteer participants. The physiological relationships derived between the muscle speed-of-sound and other demographic factors indicate that such a framework can indeed bridge from engineering innovations to clinical application.

5.7 Acknowledgements

The assistance of Lawrence Ju on the development of the user interface for live imaging is greatly appreciated.

Chapter 6

Summary and Future Directions

6.1 Summary of Contributions

Monitoring of muscle health is essential for overall physical health; early detection of pathological conditions such as muscle wasting can improve quality-of-life (Mourtzakis et al., 2017; Nagae et al., 2023; Wolfe, 2006). Measures of both muscle quantity and muscle quality, defined as changes in muscle architecture and composition, are essential for monitoring muscle health (Mourtzakis et al., 2017; Nagae et al., 2023). MSK ultrasound, with its ability to be used at the point-of-care, is an important tool for monitoring muscle quality through quantitative metrics such as echogenicity, texture, or stiffness that are associated with pathological conditions (Drakonaki et al., 2012; Nodera et al., 2018; Paris & Mourtzakis, 2016). Scanner settings and operator skills can affect or change the quantitative assessment of muscle quality (Young et al., 2015). Tissue speed-of-sound (SoS) is an intrinsic property of the tissue that can robustly augment muscle quality assessment; however, sonographic measurement of muscle speed-of-sound is frequently limited by the measurement hardware. Most tissue SoS measurement frameworks require bilateral access to the tissue or cannot perform simultaneous ultrasound imaging in real-time (Krucker et al., 2004; Sanabria et al., 2019; Wiskin et al., 2020).

In this dissertation, I devised an ultrasound framework for live, accurate SoS estimation *in vivo*. Using the principles of high-frame-rate ultrasound, we designed a novel global SoS estimation algorithm to operate in real-time. This algorithm was experimentally validated to have an average error less than 5 m/s error *in vitro* and less than 10 m/s *in vivo* while operating at over 15 fps using a portable scanner implementation (**Chapter 2**). The challenge of efficiently beamforming B-mode images in real-time was resolved through reformulation of delay-and-sum beamforming as a sparse matrix multiplication (**Chapter 3**). The challenge of real-time data transfer rates during high-frame-rate transmissions was addressed by designing a custom convolutional neural network architecture to reduce data transfer without compromising image quality (**Chapter 4**). Lastly, the novel framework was then applied to conduct a human study for three purposes: 1) to demonstrate the applicability of a live SoS system, 2) to establish a replicable experimental protocol for SoS measurement in large muscles, 3) to investigate relationships between demographics and muscle SoS (**Chapter 5**).

The overall impact of this dissertation is to bridge the engineering innovations of ultrasound algorithm development with the clinical applications for SoS as a tissue biomarker. Tissue SoS can be an effective indicator of pathological conditions in tissue, such as differentiating between malignant and benign tumors, assessing liver steatosis, or identifying sarcopenia in elderly women (Kumagai et al., 2019; Li, Duric, et al., 2009; Sanabria et al., 2019; Telichko et al., 2022). This dissertation work confronts two major hurdles for clinical translation of tissue SoS as a biomarker. Firstly, by introducing a new method to perform SoS estimation in pulse-echo mode (Sec. 2.2.4), we overcome the requirement for bilateral access to the imaged tissue (via modes such as USCT or through-transmit ultrasound). Secondly, by designing the method to be highly efficient (Sec. 2.3.6), we can achieve concurrent real-time SoS estimation alongside B-mode imaging. By addressing these two challenges, this work acts as a milestone towards the clinical viability of tissue SoS and can enable future studies with greater clinical potential.

From an engineering standpoint, this dissertation makes specific contributions to engineering ultrasound research. The proposed framework is one of the first demonstrated real-time capable systems for SoS estimation. Simultaneous display of the SoS alongside the B-mode image allows for use at the point-of-care, where ultrasound has the most advantages as a medical imaging modality (Nagae et al., 2023). The rigorous validation of the novel SoS estimation algorithm was crucial for the claims of accuracy. By designing a custom through-transmit device with provable accuracy in water (Sec. 2.3.1), then constructing agar phantoms and measuring the SoS (Sec. 2.3.2), we were able to confidently assess the accuracy of the proposed pulse-echo SoS estimation algorithm (Sec. 2.4.1). The realization of the algorithm on a portable scanner (Sec. 2.4.3) further advances the types of applications where tissue SoS can find clinical applications. Implementation on the portable scanner was achieved by connecting deep learning models (Chapter 4) and sparse matrix multiplication (Chapter 3) with the ultrasound image formation pipeline in ways that naturally extend even outside the SoS estimation context.

MSK ultrasound is a core clinical application of ultrasound impacted by this dissertation. From the perspective of muscle SoS, the replication of through-transmit muscle SoS studies with the proposed pulse-echo framework is an important contribution (Sec. 5.4.1, Sec. 5.4.2). These results signal that pulse-echo SoS measurements can possibly replace through-transmit measurements for muscles, negating the requirements for bilateral access. Pulse-echo SoS estimation allows for targeted investigation of a wider variety of muscles, such as abdominal muscles, where bilateral access is an impossibility. The dissertation also introduces a relationship that is new to the literature – the

relationship between physical activity and muscle SoS (Sec. 5.4.4), which was significant for males only. This effect can lead to new studies that include measuring the impact of interventions for exercise (e.g. proscribed exercise regimen). The newly established protocol demonstrates the reliability of muscle SoS as an intrinsic property of the tissue, and that the measurement is robust to small variations in operator placement and probe pressure (Sec. 5.3). Overall, the results obtained with the proposed framework exhibits that muscle SoS can be a powerful new metric for assessing muscle quality and further studies should be conducted.

6.1.1 Realization of Real-Time, Global Speed-of-Sound Estimator with Simultaneous B-Mode Imaging Capacity

In Chapter 2, the two primary research contributions were: 1) a new algorithm for SoS estimation via steered plane waves, and 2) the development of a device capable of simultaneous SoS estimation and B-mode imaging in real-time.

The SoS estimation algorithm leveraged the defocusing of scatterers across steered plane wave transmissions when the incorrect SoS is used for beamforming (Figure 2.1). To validate that the framework resulted in accurate SoS estimates, a custom through-transmit device (Figure 2.4) was used to take reference SoS measurements for custom agar phantoms and for human calves. The average error of the proposed framework, as quantified by the mean signed difference, was -0.6 m/s *in vitro* (Figure 2.6(c)) and -2.2 m/s *in vivo* (Figure 2.6(g)) on the collected data sets, confirming a high degree of accuracy compared to the reference measurements. For real-time application, the proposed framework was implemented on a portable open-platform ultrasound scanner (Sec. 2.3.6). Using the US4R-Lite system, we realized concurrent SoS estimation and B-mode imaging at 18 fps (Sec. 2.4.3).

By performing tissue SoS estimation in real-time, the SoS as a biomarker is better connected with the B-mode images displayed to the sonographer at the point-of-care. This capability leads to higher usability for human studies, which can then advance the clinical use of tissue SoS. From an algorithmic standpoint, the framework is highly amenable to customization; the core steps of SoS-aware beamforming, comparing B-mode images, then loss minimization for SoS estimation can be readily generalized. There may be benefits to incorporating alternative beamforming schemes such as minimum variance beamforming or null subtraction imaging, which depend even more on the SoS of the medium (Agarwal et al., 2019; Synnevag et al., 2009). In Sec. 2.3.5, two specific loss functions were compared, but there may also be advantages to other imaging metrics comparing images across

transmit angles such as coherence (Hériard-Dubreuil et al., 2023). Lastly, we show the importance of accounting for refraction effects when modeling the effects of SoS on the beamforming paths (Figure 2.8(b)). Local SoS estimation methods can also benefit from refraction correction compared to assuming straight-line propagation (Ali, Telichko, et al., 2022; Jaeger et al., 2015; Sanabria et al., 2018).

6.1.2 Novel Solutions to Sub-Challenges for Real-Time High-Frame-Rate Imaging Methods

In Chapter 3 and Chapter 4, my research addressed two challenges for real-time high-frame-rate imaging: 1) image formation efficiency for DAS beamforming was improved via sparse matrix beamforming, 2) radiofrequency data transfer rate from the probe to the imaging system was reduced via convolutional neural networks.

For the first challenge of image formation efficiency (Chapter 3), we reformulated DAS beamforming into a sparse matrix multiplication to leverage powerful sparse matrix libraries for acceleration. The use of SMB was demonstrated for a variety of potential use cases (single plane wave (Sec. 3.3.2), focused imaging (Sec. 3.3.3), plane wave compounding (Sec. 3.3.4), and beamforming integrated neural networks (Sec. 3.3.5)), alongside execution timings demonstrating the benefit of sparse matrix optimizations. SMB facilitated rapid GPU acceleration of the SoS estimation framework through encoding of the full beamforming structure (180 DAS images) into a single sparse matrix multiplication (Sec. 2.3.6). The sparse matrix paradigm is not only useful for SoS estimation but also for any DAS dependent process. The reformulation of DAS enabled purely software-based acceleration as opposed to leveraging hardware-based acceleration such as FPGAs. In this way, SMB is an accessible tool which can be very useful for ultrasound researchers working on software-based systems to enhance real-time operational potential of new imaging algorithms and eventually lead to improved clinical translatability.

For the second challenge of data transfer rate (Chapter 4), we acquired data from only half the transducer elements, learned an encoder-decoder CNN to infer the unacquired channels, then demonstrated that this approach led to restored image quality. The RF-to-RF network (4.3) was trained on purely *in vivo* data (Sec. 4.4.1, Sec. 4.4.3), reduced aliasing artifacts in an out-of-distribution *in vitro* point target phantom (Sec. 4.5.2), and improved carotid lumen contrast *in vivo* (Sec. 4.5.3). Moreover, aligned with the real-time theme of this thesis, we ensured that the inference times were capable of live operation (Sec. 4.5.5). While this module was not connected to the final SoS estimation framework due

to limitations of the scanner hardware, we demonstrated the feasibility of real-time data reduction. The RF-to-RF nature of the CNN permits the full data set to be used for subsequent ultrasound imaging algorithms. This capability allows for not only SoS estimation, but also alternative beamforming dependent imaging modes such as vector flow or elastography (Bercoff et al., 2004; Jensen et al., 2016b).

6.1.3 Demonstration of Speed-of-Sound Estimation Framework *in vivo* for Assessing Muscle Quality

The study performed in Chapter 5 fulfilled three primary purposes: 1) to demonstrate the applicability of a live SoS framework, 2) to establish a replicable experimental protocol for SoS measurement in large muscles, 3) to investigate relationships between demographics and muscle SoS.

Forty volunteer participants were recruited for measurement of muscle SoS using the live SoS estimation framework (Chapter 2) and to provide demographic information on age, sex, BMI, and physical activity (Sec. 5.2.3). Previously, there has been no established protocol for identifying the ultrasound imaging views ideal for SoS estimation in the muscles. By adopting an approach based on manual palpation from literature (Paris et al., 2017; Takai et al., 2014), consistent imaging views could be identified across participants. From this replicable process, statistical analysis showed the ideal views and muscle states for consistent SoS estimation (Sec. 5.3.5). Further statistical analysis on the most reliable imaging views showed significant relationships between muscle SoS and BMI (Sec. 5.4.1), sex (Sec. 5.4.2), and physical activity for males only (Sec. 5.4.4). These analyses compared favorably with similar results in the through-transmit SoS literature (Ruby et al., 2020; Topchyan et al., 2006).

This study highlighted the utility of having a real-time SoS framework; the statistical analyses were performed using the displayed and recorded SoS at the time of acquisition. By uniting the B-mode display and SoS estimation functionality, stable displayed SoS values aided in the ease of finding the correct imaging view. This feature is essential for maximizing the clinical translation potential as providing the information to the operator at bedside is core to point-of-care ultrasound. The development and standardization of the new protocol for SoS estimation in large muscles provides a foundation for further comparative studies in the future. From the perspective of muscle quality, this dissertation has demonstrated that the muscle SoS is a robustly measurable quantity that has

physiological significance. Further studies are necessary to better quantify the relationship between subcutaneous adipose tissue, muscle, and tissue SoS.

6.2 Limitations and Future Directions

6.2.1 From Global Speed-of-Sound Estimation to Local Speed-of-Sound Mapping

A natural direction for extending this dissertation work is to progress from global SoS estimation to local SoS mapping. In Sec. 2.4.1 and Sec. 5.4.5, we observed limitations of the global SoS model assumption that the imaged medium is roughly homogeneous (2.2) due to the effects of subcutaneous adipose tissue. Options for local-wise extension of SoS estimation include layer-wise modelling (Hériard-Dubreuil et al., 2023; Jakovljevic et al., 2018) or true local SoS mapping (Ali, Brevett, et al., 2022; Jaeger et al., 2015; Sanabria et al., 2018). Local SoS methods remain challenging to implement in parallel with live imaging; refinement of both the beamforming and SoS estimation procedures would facilitate the transition to real-time. Extending the work of this dissertation for layer-wise SoS analysis can also be performed through regressive modelling. Automated, real-time measurements of tissue thickness or constructing a theory-based model for materials with layer-wise structure are two possible methods of distinguishing the SoS of the adipose tissue from the underlying muscles. After selecting any of these alternatives, the statistical analysis for the study (Sec. 5.4) can be reperformed as the raw RF data remains accessible. A layer-wise extension can expand the body of applications for a live SoS estimation system to key applications where the tissue SoS has already demonstrated clinical utility such as the liver (Kumagai et al., 2019; Sehgal et al., 1986; Telichko et al., 2022).

6.2.2 Further Investigation of Physiological Relevance of Tissue Speed-of-Sound

With the availability of the real-time SoS framework, additional human studies should be conducted to explore the physiological relevance of the tissue SoS. The study performed in this thesis (Chapter 5) was limited by the demographics of the sampled population – the study was primarily conducted on young and healthy individuals who predominantly engaged in sedentary work. Studies that examine populations with more pathological conditions, such as the elderly or immobilized, could lead to new relationships between the muscle SoS and muscle quality. From a clinical perspective, monitoring of conditions such as muscle wasting through casting or sarcopenia can be more easily achieved with a real-time ultrasound system at the bedside. Advancement towards local SoS measurement is also beneficial for measuring specific muscle. Local muscle studies, such as comparing the SoS in the rectus

femoris muscle versus the vastus lateralis/medialis/intermedius muscles in the quadriceps, will provide more detailed information on muscle function. Studies using direct measurements of muscle strength instead of a self-reported physical activity questionnaire will also better quantify the relationship between physical activity and muscle SoS. These studies can then lead to interventional muscle studies both acute (e.g., measurement before and after exercise) or long-term (e.g., proscribed exercise regimen). These muscle studies can all be facilitated by the ease-of-use of the portable scanner framework described in this dissertation.

6.2.3 Conclusion

Tissue SoS has high potential as a robust quantitative tissue biomarker for assessing muscle quality. This dissertation presents one of the first live, accurate tissue SoS estimation algorithms realized on a portable ultrasound scanner concurrently with B-mode imaging. This research is essential for the clinical translation potential of using tissue SoS as a biomarker.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems* (arXiv:1603.04467). arXiv. <https://doi.org/10.48550/arXiv.1603.04467>
- Agarwal, A., Reeg, J., Podkowa, A. S., & Oelze, M. L. (2019). Improving Spatial Resolution Using Incoherent Subtraction of Receive Beams Having Different Apodizations. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 66(1), 5–17. <https://doi.org/10.1109/TUFFC.2018.2876285>
- Ali, R., Brevett, T., Hyun, D., Brickson, L. L., & Dahl, J. J. (2022). Distributed Aberration Correction Techniques Based on Tomographic Sound Speed Estimates. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(5), 1714–1726. <https://doi.org/10.1109/TUFFC.2022.3162836>
- Ali, R., Brevett, T., Zhuang, L., Bendjador, H., Podkowa, A. S., Hsieh, S. S., Simson, W., Sanabria, S. J., Herickhoff, C. D., & Dahl, J. J. (2023). Aberration correction in diagnostic ultrasound: A review of the prior field and current directions. *Zeitschrift Für Medizinische Physik*, 33(3), 267–291. <https://doi.org/10.1016/j.zemedi.2023.01.003>
- Ali, R., Telichko, A. V., Wang, H., Sukumar, U. K., Vilches-Moure, J. G., Paulmurugan, R., & Dahl, J. J. (2022). Local Sound Speed Estimation for Pulse-Echo Ultrasound in Layered Media. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(2), 500–511. <https://doi.org/10.1109/TUFFC.2021.3124479>

- Anderson, M. E., McKeag, M. S., & Trahey, G. E. (2000). The impact of sound speed errors on medical ultrasound imaging. *The Journal of the Acoustical Society of America*, 107(6), 3540–3548. <https://doi.org/10.1121/1.429422>
- Austeng, A., & Holm, S. (2002). Sparse 2-D arrays for 3-D phased array imaging—Design methods. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 49(8), 1073–1086. <https://doi.org/10.1109/TUFFC.2002.1026019>
- Backhaus, M. (2001). Guidelines for musculoskeletal ultrasound in rheumatology. *Annals of the Rheumatic Diseases*, 60(7), 641–649. <https://doi.org/10.1136/ard.60.7.641>
- Bae, S., Jang, J., Choi, M. H., & Song, T.-K. (2020). In Vivo Evaluation of Plane Wave Imaging for Abdominal Ultrasonography. *Sensors*, 20(19), Article 19. <https://doi.org/10.3390/s20195675>
- Baecke, J., Burema, J., & Frijters, J. (1982). A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *The American Journal of Clinical Nutrition*, 36(5), 936–942. <https://doi.org/10.1093/ajcn/36.5.936>
- Bamber, J. C., & Hill, C. R. (1979). Ultrasonic attenuation and propagation speed in mammalian tissues as a function of temperature. *Ultrasound in Medicine & Biology*, 5(2), 149–157. [https://doi.org/10.1016/0301-5629\(79\)90083-8](https://doi.org/10.1016/0301-5629(79)90083-8)
- Bell, N., & Garland, M. (2008). *Efficient sparse matrix-vector multiplication on CUDA*. Nvidia Technical Report NVR-2008-004, Nvidia Corporation.
- Bercoff, J., Tanter, M., & Fink, M. (2004). Supersonic shear imaging: A new technique for soft tissue elasticity mapping. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 51(4), 396–409. <https://doi.org/10.1109/TUFFC.2004.1295425>
- Besson, A., Perdios, D., Martinez, F., Chen, Z., Carrillo, R. E., Arditi, M., Wiaux, Y., & Thiran, J.-P. (2018). Ultrafast Ultrasound Imaging as an Inverse Problem: Matrix-Free Sparse Image

- Reconstruction. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(3), 339–355. <https://doi.org/10.1109/TUFFC.2017.2768583>
- Bezek, C. D., & Goksel, O. (2023). Analytical estimation of beamforming speed-of-sound using transmission geometry. *Ultrasonics*, 134, 107069. <https://doi.org/10.1016/j.ultras.2023.107069>
- Bilaniuk, N., & Wong, G. S. K. (1993). Speed of sound in pure water as a function of temperature. *The Journal of the Acoustical Society of America*, 93(3), 1609–1612. <https://doi.org/10.1121/1.406819>
- Boni, E., Bassi, L., Dallai, A., Meacci, V., Ramalli, A., Scaringella, M., Guidi, F., Ricci, S., & Tortoli, P. (2017). Architecture of an Ultrasound System for Continuous Real-Time High Frame Rate Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 64(9), 1276–1284. <https://doi.org/10.1109/TUFFC.2017.2727980>
- Boni, E., Yu, A. C. H., Freear, S., Jensen, J. A., & Tortoli, P. (2018). Ultrasound Open Platforms for Next-Generation Imaging Technique Development. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(7), 1078–1092. <https://doi.org/10.1109/TUFFC.2018.2844560>
- Brandenburg, J. E., Eby, S. F., Song, P., Zhao, H., Brault, J. S., Chen, S., & An, K.-N. (2014). Ultrasound Elastography: The New Frontier in Direct Measurement of Muscle Stiffness. *Archives of Physical Medicine and Rehabilitation*, 95(11), 2207–2219. <https://doi.org/10.1016/j.apmr.2014.07.007>
- Campbell, N. A., & Brown, J. A. (2022). A Real-Time Dual-Mode High-Frequency Beamformer for Ultrafast and Focused Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(4), 1268–1276. <https://doi.org/10.1109/TUFFC.2022.3151218>

- Campbell, S. E., Adler, R., & Sofka, C. M. (2005). Ultrasound of Muscle Abnormalities. *Ultrasound Quarterly*, 21(2), 87. <https://doi.org/10.1097/01.ruq.0000164504.30592.9f>
- Candes, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2), 489–509. <https://doi.org/10.1109/TIT.2005.862083>
- Chen, Y., Liu, J., Luo, X., & Luo, J. (2021). ApodNet: Learning for High Frame Rate Synthetic Transmit Aperture Ultrasound Imaging. *IEEE Transactions on Medical Imaging*, 40(11), 3190–3204. <https://doi.org/10.1109/TMI.2021.3084821>
- Chernyakova, T., Cohen, R., Mulayoff, R., Sde-Chen, Y., Fraschini, C., Bercoff, J., & Eldar, Y. C. (2018). Fourier-Domain Beamforming and Structure-Based Reconstruction for Plane-Wave Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(10), 1810–1821. <https://doi.org/10.1109/TUFFC.2018.2856301>
- Chernyakova, T., & Eldar, Y. C. (2014). Fourier-domain beamforming: The path to compressed ultrasound imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 61(8), 1252–1267. <https://doi.org/10.1109/TUFFC.2014.3032>
- Cheung, C. C. P., Yu, A. C. H., Salimi, N., Yiu, B. Y. S., Tsang, I. K. H., Kerby, B., Azar, R. Z., & Dickie, K. (2012). Multi-channel pre-beamformed data acquisition system for research on advanced ultrasound imaging methods. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 59(2), 243–253. <https://doi.org/10.1109/TUFFC.2012.2184>
- Cobbold, R. S. (2006). *Foundations of biomedical ultrasound*. Oxford university press.
<https://books.google.com/books?hl=en&lr=&id=Hwb5D60vb5IC&oi=fnd&pg=PA3&dq=biomedical+ultrasound+cobbold&ots=dWL9oGdSHA&sig=I9AojfC9nHtg6zz24xw58odJYEE>

- Cohen, R., & Eldar, Y. C. (2018). Sparse Convolutional Beamforming for Ultrasound Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(12), 2390–2406.
<https://doi.org/10.1109/TUFFC.2018.2874256>
- Cummings, S. R., Bates, D., & Black, D. M. (2002). Clinical Use of Bone Densitometry. *JAMA*, 288(15), 1889–1897. <https://doi.org/10.1001/jama.288.15.1889>
- Czyrny, Z. (2017). Standards for musculoskeletal ultrasound. *Journal of Ultrasonography*, 17(70), 182–187. <https://doi.org/10.15557/JoU.2017.0027>
- Demené, C., Deffieux, T., Pernot, M., Osmanski, B.-F., Biran, V., Gennisson, J.-L., Sieu, L.-A., Bergel, A., Franqui, S., Correas, J.-M., Cohen, I., Baud, O., & Tanter, M. (2015). Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity. *IEEE Transactions on Medical Imaging*, 34(11), 2271–2285.
<https://doi.org/10.1109/TMI.2015.2428634>
- Diarra, B., Robini, M., Tortoli, P., Cachard, C., & Liebgott, H. (2013). Design of Optimal 2-D Nongrid Sparse Arrays for Medical Ultrasound. *IEEE Transactions on Biomedical Engineering*, 60(11), 3093–3102. <https://doi.org/10.1109/TBME.2013.2267742>
- Díaz-Gómez, J. L., Mayo, P. H., & Koenig, S. J. (2021). Point-of-Care Ultrasonography. *New England Journal of Medicine*, 385(17), 1593–1602. <https://doi.org/10.1056/NEJMr1916062>
- Drakonaki, E. E., Allen, G. M., & Wilson, D. J. (2012). Ultrasound elastography for musculoskeletal applications. *The British Journal of Radiology*, 85(1019), 1435–1445.
<https://doi.org/10.1259/bjr/93042867>
- Drozdal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2016). The Importance of Skip Connections in Biomedical Image Segmentation. In G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, & J. Cornebise (Eds.), *Deep Learning and Data Labeling for Medical*

- Applications* (pp. 179–187). Springer International Publishing. https://doi.org/10.1007/978-3-319-46976-8_19
- Duric, N., Littrup, P., Roy, O., Li, C., Schmidt, S., Cheng, X., & Janer, R. (2014). Clinical breast imaging with ultrasound tomography: A description of the SoftVue system. *The Journal of the Acoustical Society of America*, 135(4_Supplement), 2155. <https://doi.org/10.1121/1.4876990>
- Errico, C., Pierre, J., Pezet, S., Desailly, Y., Lenkei, Z., Couture, O., & Tanter, M. (2015). Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. *Nature*, 527(7579), Article 7579. <https://doi.org/10.1038/nature16066>
- Feigin, M., Freedman, D., & Anthony, B. W. (2020). A Deep Learning Framework for Single-Sided Sound Speed Inversion in Medical Ultrasound. *IEEE Transactions on Biomedical Engineering*, 67(4), 1142–1151. <https://doi.org/10.1109/TBME.2019.2931195>
- Foucher, J., Chanteloup, E., Vergniol, J., Castéra, L., Bail, B. L., Adhoute, X., Bertet, J., Couzigou, P., & Lédinthen, V. de. (2006). Diagnosis of cirrhosis by transient elastography (FibroScan): A prospective study. *Gut*, 55(3), 403–408. <https://doi.org/10.1136/gut.2005.069153>
- Fowers, J., Ovtcharov, K., Strauss, K., Chung, E. S., & Stitt, G. (2014). A High Memory Bandwidth FPGA Accelerator for Sparse Matrix-Vector Multiplication. *2014 IEEE 22nd Annual International Symposium on Field-Programmable Custom Computing Machines*, 36–43. <https://doi.org/10.1109/FCCM.2014.23>
- Frost, M. L., Blake, G. M., & Fogelman, I. (2001). Quantitative Ultrasound and Bone Mineral Density Are Equally Strongly Associated with Risk Factors for Osteoporosis. *Journal of Bone and Mineral Research*, 16(2), 406–416. <https://doi.org/10.1359/jbmr.2001.16.2.406>

- Garcia, D. (2021). Make the most of MUST, an open-source Matlab UltraSound Toolbox. *2021 IEEE International Ultrasonics Symposium (IUS)*, 1–4.
<https://doi.org/10.1109/IUS52206.2021.9593605>
- Gnudi, S., Ripamonti, C., & Malavolta, N. (2000). Quantitative Ultrasound and Bone Densitometry to Evaluate the Risk of Nonspine Fractures: A Prospective Study. *Osteoporosis International*, *11*(6), 518–523. <https://doi.org/10.1007/s001980070095>
- Goss, S. A., Johnston, R. L., & Dunn, F. (1978). Comprehensive compilation of empirical ultrasonic properties of mammalian tissues. *The Journal of the Acoustical Society of America*, *64*(2), 423–457. <https://doi.org/10.1121/1.382016>
- Grassi, W., Filippucci, E., & Busilacchi, P. (2004). Musculoskeletal ultrasound. *Best Practice & Research Clinical Rheumatology*, *18*(6), 813–826. <https://doi.org/10.1016/j.berh.2004.05.001>
- Greenspan, M., & Tschiegg, C. E. (1957). Speed of sound in water by a direct method. *Journal of Research of the National Bureau of Standards*, *59*(4), 249–254.
- Hansen, J. M., Schaa, D., & Jensen, J. A. (2011). Synthetic aperture beamformation using the GPU. *2011 IEEE International Ultrasonics Symposium*, 373–376.
<https://doi.org/10.1109/ULTSYM.2011.0089>
- Harris-Love, M. O., Monfaredi, R., Ismail, C., Blackman, M. R., & Cleary, K. (2014). Quantitative Ultrasound: Measurement Considerations for the Assessment of Muscular Dystrophy and Sarcopenia. *Frontiers in Aging Neuroscience*, *6*.
<https://www.frontiersin.org/articles/10.3389/fnagi.2014.00172>
- Hériard-Dubreuil, B., Besson, A., Wintzenrieth, F., Cohen-Bacrie, C., & Thiran, J.-P. (2023). Refraction-Based Speed of Sound Estimation in Layered Media: An Angular Approach. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *70*(6), 486–497.
<https://doi.org/10.1109/TUFFC.2023.3261541>

- Hoen, P. (1982). Aperture apodization to reduce the off-axis intensity of the pulsed-mode directivity function of linear arrays. *Ultrasonics*, 20, 231–236.
- Hou, G. Y., Provost, J., Grondin, J., Wang, S., Marquet, F., Bunting, E., & Konofagou, E. E. (2014). Sparse Matrix Beamforming and Image Reconstruction for 2-D HIFU Monitoring Using Harmonic Motion Imaging for Focused Ultrasound (HMIFU) With In Vitro Validation. *IEEE Transactions on Medical Imaging*, 33(11), 2107–2117.
<https://doi.org/10.1109/TMI.2014.2332184>
- Huijben, I. A. M., Veeling, B. S., Janse, K., Mischi, M., & van Sloun, R. J. G. (2020). Learning Sub-Sampling and Signal Recovery With Applications in Ultrasound Imaging. *IEEE Transactions on Medical Imaging*, 39(12), 3955–3966. <https://doi.org/10.1109/TMI.2020.3008501>
- Hyun, D., Li, Y. L., Steinberg, I., Jakovljevic, M., Klap, T., & Dahl, J. J. (2019). An Open Source GPU-Based Beamformer for Real-Time Ultrasound Imaging and Applications. *2019 IEEE International Ultrasonics Symposium (IUS)*, 20–23.
<https://doi.org/10.1109/ULTSYM.2019.8926193>
- Hyun, D., Wiacek, A., Goudarzi, S., Rothlübbers, S., Asif, A., Eickel, K., Eldar, Y. C., Huang, J., Mischi, M., Rivaz, H., Sinden, D., van Sloun, R. J. G., Strohm, H., & Bell, M. A. L. (2021). Deep Learning for Ultrasound Image Formation: CUBDL Evaluation Framework and Open Datasets. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(12), 3466–3483. <https://doi.org/10.1109/TUFFC.2021.3094849>
- Ibrahim, A., Hager, P., Bartolini, A., Angiolini, F., Arditi, M., Benini, L., & De Micheli, G. (2015). Tackling the bottleneck of delay tables in 3D ultrasound imaging. *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 1683–1688.
<https://doi.org/10.7873/DATE.2015.1151>

- Imbault, M., Faccinnetto, A., Osmanski, B.-F., Tissier, A., Deffieux, T., Gennisson, J.-L., Vilgrain, V., & Tanter, M. (2017). Robust sound speed estimation for ultrasound-based hepatic steatosis assessment. *Physics in Medicine & Biology*, 62(9), 3582. <https://doi.org/10.1088/1361-6560/aa6226>
- Jaeger, M., Held, G., Peeters, S., Preisser, S., Grünig, M., & Frenz, M. (2015). Computed Ultrasound Tomography in Echo Mode for Imaging Speed of Sound Using Pulse-Echo Sonography: Proof of Principle. *Ultrasound in Medicine & Biology*, 41(1), 235–250. <https://doi.org/10.1016/j.ultrasmedbio.2014.05.019>
- Jakovljevic, M., Hsieh, S., Ali, R., Chau Loo Kung, G., Hyun, D., & Dahl, J. J. (2018). Local speed of sound estimation in tissue using pulse-echo ultrasound: Model-based approach. *The Journal of the Acoustical Society of America*, 144(1), 254–266. <https://doi.org/10.1121/1.5043402>
- Jensen, J. A., Nikolov, S. I., Gammelmark, K. L., & Pedersen, M. H. (2006). Synthetic aperture ultrasound imaging. *Ultrasonics*, 44, e5–e15. <https://doi.org/10.1016/j.ultras.2006.07.017>
- Jensen, J. A., Nikolov, S. I., Yu, A. C. H., & Garcia, D. (2016a). Ultrasound Vector Flow Imaging—Part I: Sequential Systems. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63(11), 1704–1721. <https://doi.org/10.1109/TUFFC.2016.2600763>
- Jensen, J. A., Nikolov, S. I., Yu, A. C. H., & Garcia, D. (2016b). Ultrasound Vector Flow Imaging—Part II: Parallel Systems. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63(11), 1722–1732. <https://doi.org/10.1109/TUFFC.2016.2598180>
- Kamimura, H. A. S., Wu, S.-Y., Grondin, J., Ji, R., Aurup, C., Zheng, W., Heidmann, M., Pouliopoulos, A. N., & Konofagou, E. E. (2021). Real-Time Passive Acoustic Mapping Using Sparse Matrix Multiplication. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(1), 164–177. <https://doi.org/10.1109/TUFFC.2020.3001848>

- Ketterling, J. A., Aristizábal, O., Yiu, B. Y. S., Turnbull, D. H., Phoon, C. K. L., Yu, A. C. H., & Silverman, R. H. (2017). High-speed, high-frequency ultrasound, in utero vector-flow imaging of mouse embryos. *Scientific Reports*, 7(1), Article 1.
<https://doi.org/10.1038/s41598-017-16933-x>
- Kindblom, J. M., Lorentzon, M., Hellqvist, Å., Lönn, L., Brandberg, J., Nilsson, S., Norjavaara, E., & Ohlsson, C. (2009). BMI Changes During Childhood and Adolescence as Predictors of Amount of Adult Subcutaneous and Visceral Adipose Tissue in Men: The GOOD Study. *Diabetes*, 58(4), 867–874. <https://doi.org/10.2337/db08-0606>
- Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (arXiv:1412.6980). arXiv. <https://doi.org/10.48550/arXiv.1412.6980>
- Koch, T., Lakshmanan, S., Brand, S., Wicke, M., Raum, K., & Mörlein, D. (2011a). Ultrasound velocity and attenuation of porcine soft tissues with respect to structure and composition: I. Muscle. *Meat Science*, 88(1), 51–58. <https://doi.org/10.1016/j.meatsci.2010.12.002>
- Koch, T., Lakshmanan, S., Brand, S., Wicke, M., Raum, K., & Mörlein, D. (2011b). Ultrasound velocity and attenuation of porcine soft tissues with respect to structure and composition: II. Skin and backfat. *Meat Science*, 88(1), 67–74. <https://doi.org/10.1016/j.meatsci.2010.12.004>
- Kou, Z., You, Q., Kim, J., Dong, Z., Lowerison, M. R., Sekaran, N. V. C., Llano, D. A., Song, P., & Oelze, M. L. (2023). High-Level Synthesis Design of Scalable Ultrafast Ultrasound Beamformer With Single FPGA. *IEEE Transactions on Biomedical Circuits and Systems*, 17(3), 446–457. <https://doi.org/10.1109/TBCAS.2023.3267614>
- Kremkau, F. W., & Taylor, K. J. (1986). Artifacts in ultrasound imaging. *Journal of Ultrasound in Medicine*, 5(4), 227–237. <https://doi.org/10.7863/jum.1986.5.4.227>

- Krucker, J. F., Fowlkes, J. B., & Carson, P. L. (2004). Sound speed estimation using automatic ultrasound image registration. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 51(9), 1095–1106. <https://doi.org/10.1109/TUFFC.2004.1334842>
- Kruschitz, R., Wallner-Liebmann, S. J., Hamlin, M. J., Moser, M., Ludvik, B., Schnedl, W. J., & Tafeit, E. (2013). Detecting Body Fat—A Weighty Problem BMI versus Subcutaneous Fat Patterns in Athletes and Non-Athletes. *PLOS ONE*, 8(8), e72002. <https://doi.org/10.1371/journal.pone.0072002>
- Kumagai, H., Taniguchi, N., Yokoyama, K., Katsuyama, K., Yamamoto, H., Hara, S., Hirota, N., Itoh, K., & Yamagata, T. (2019). The Speed of Sound in Rat Liver With Steatohepatitis: Ex Vivo Analysis Using Two Types of Ultrasound Systems. *Ultrasound in Medicine & Biology*, 45(8), 2258–2265. <https://doi.org/10.1016/j.ultrasmedbio.2019.05.004>
- Larson, J. D. I. (1993). *2-d phased array ultrasound imaging system with distributed phasing* (United States Patent US5229933A). <https://patents.google.com/patent/US5229933A/en>
- Leavitt, S. C., Savord, B. J., Herrick, B. M., Fallon, J. R., & Mooney, M. (2002). *Sub-beamforming apparatus and method for a portable ultrasound imaging system* (United States Patent US6491634B1). <https://patents.google.com/patent/US6491634B1/en>
- Lenge, M., Ramalli, A., Tortoli, P., Cachard, C., & Liebgott, H. (2015). Plane-wave transverse oscillation for high-frame-rate 2-D vector flow imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 62(12), 2126–2137. <https://doi.org/10.1109/TUFFC.2015.007320>
- Li, C., Duric, N., Littrup, P., & Huang, L. (2009). In vivo Breast Sound-Speed Imaging with Ultrasound Tomography. *Ultrasound in Medicine & Biology*, 35(10), 1615–1628. <https://doi.org/10.1016/j.ultrasmedbio.2009.05.011>

- Li, C., Huang, L., Duric, N., Zhang, H., & Rowe, C. (2009). An improved automatic time-of-flight picker for medical ultrasound tomography. *Ultrasonics*, 49(1), 61–72.
<https://doi.org/10.1016/j.ultras.2008.05.005>
- Lieb Gott, H., Prost, R., & Friboulet, D. (2013). Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing. *Ultrasonics*, 53(2), 525–533.
<https://doi.org/10.1016/j.ultras.2012.09.008>
- Lin, T., Ophir, J., & Potter, G. (1987). Correlations of sound speed with tissue constituents in normal and diffuse liver disease. *Ultrasonic Imaging*, 9(1), 29–40. [https://doi.org/10.1016/0161-7346\(87\)90016-2](https://doi.org/10.1016/0161-7346(87)90016-2)
- Litniewski, J., Nowicki, A., & Sawicki, A. (2000). Detection of bone disease with ultrasound—Comparison with bone densitometry. *Ultrasonics*, 38(1), 693–697.
[https://doi.org/10.1016/S0041-624X\(99\)00191-2](https://doi.org/10.1016/S0041-624X(99)00191-2)
- Liu, J., He, Q., & Luo, J. (2017). A Compressed Sensing Strategy for Synthetic Transmit Aperture Ultrasound Imaging. *IEEE Transactions on Medical Imaging*, 36(4), 878–891.
<https://doi.org/10.1109/TMI.2016.2644654>
- Lok, U.-W., Fan, G.-W., & Li, P.-C. (2015). Lossless Data Compression for Improving the Performance of a GPU-Based Beamformer. *Ultrasonic Imaging*, 37(2), 135–151.
<https://doi.org/10.1177/0161734614547280>
- Maas, A. L., Hannun, A. Y., Ng, A. Y., & others. (2013). Rectifier nonlinearities improve neural network acoustic models. *Proc. Icml*, 30(1), 3.
- Macé, E., Montaldo, G., Cohen, I., Baulac, M., Fink, M., & Tanter, M. (2011). Functional ultrasound imaging of the brain. *Nature Methods*, 8(8), Article 8. <https://doi.org/10.1038/nmeth.1641>

- Mallart, R., & Fink, M. (1992). Improved imaging rate through simultaneous transmission of several ultrasound beams. *New Developments in Ultrasonic Transducers and Transducer Systems*, 1733, 120–130. <https://doi.org/10.1117/12.130591>
- Minetto, M. A., Caresio, C., Menapace, T., Hajdarevic, A., Marchini, A., Molinari, F., & Maffiuletti, N. A. (2016). Ultrasound-Based Detection of Low Muscle Mass for Diagnosis of Sarcopenia in Older Adults. *PM&R*, 8(5), 453–462. <https://doi.org/10.1016/j.pmrj.2015.09.014>
- Mishali, M., & Eldar, Y. C. (2010). From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals. *IEEE Journal of Selected Topics in Signal Processing*, 4(2), 375–391. <https://doi.org/10.1109/JSTSP.2010.2042414>
- Mitra, S. K., Mondal, K., Tchobanou, M. K., & Dolecek, G. J. (2010). General polynomial factorization-based design of sparse periodic linear arrays. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 57(9), 1952–1966. <https://doi.org/10.1109/TUFFC.2010.1643>
- Montaldo, G., Tanter, M., Bercoff, J., Benech, N., & Fink, M. (2009). Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 56(3), 489–506. <https://doi.org/10.1109/TUFFC.2009.1067>
- Moore, C. L., & Copel, J. A. (2011). Point-of-Care Ultrasonography. *New England Journal of Medicine*, 364(8), 749–757. <https://doi.org/10.1056/NEJMra0909487>
- Mourtzakis, M., Parry, S., Connolly, B., & Puthuchear, Z. (2017). Skeletal Muscle Ultrasound in Critical Care: A Tool in Need of Translation. *Annals of the American Thoracic Society*, 14(10), 1495–1503. <https://doi.org/10.1513/AnnalsATS.201612-967PS>

- Nagae, M., Umegaki, H., Yoshiko, A., & Fujita, K. (2023). Muscle ultrasound and its application to point-of-care ultrasonography: A narrative review. *Annals of Medicine*, 55(1), 190–197. <https://doi.org/10.1080/07853890.2022.2157871>
- Nodera, H., Sogawa, K., Takamatsu, N., Mori, A., Yamazaki, H., Izumi, Y., & Kaji, R. (2018). Age-dependent texture features in skeletal muscle ultrasonography. *The Journal of Medical Investigation*, 65(3.4), 274–279. <https://doi.org/10.2152/jmi.65.274>
- Ophir, J. (1986). Estimation of the Speed of Ultrasound Propagation in Biological Tissues: A Beam-Tracking Method. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 33(4), 359–368. <https://doi.org/10.1109/T-UFFC.1986.26843>
- Oppold, J., Breu, M.-S., Gharabaghi, A., Grimm, A., Del Grosso, N. A., Hormozi, M., Kleiser, B., Klocke, P., Kronlage, C., Weiß, D., & Marquetand, J. (2023). Ultrasound of the Biceps Muscle in Idiopathic Parkinson’s Disease with Deep Brain Stimulation: Rigidity Can Be Quantified by Shear Wave Elastography. *Diagnostics*, 13(2), Article 2. <https://doi.org/10.3390/diagnostics13020213>
- Ozkan, E., Vishnevsky, V., & Goksel, O. (2018). Inverse Problem of Ultrasound Beamforming With Sparsity Constraints and Regularization. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(3), 356–365. <https://doi.org/10.1109/TUFFC.2017.2757880>
- Paris, M. T., Lafleur, B., Dubin, J. A., & Mourtzakis, M. (2017). Development of a bedside viable ultrasound protocol to quantify appendicular lean tissue mass. *Journal of Cachexia, Sarcopenia and Muscle*, 8(5), 713–726. <https://doi.org/10.1002/jcsm.12213>
- Paris, M. T., & Mourtzakis, M. (2016). Assessment of skeletal muscle mass in critically ill patients: Considerations for the utility of computed tomography imaging and ultrasonography. *Current Opinion in Clinical Nutrition and Metabolic Care*, 19(2), 125–130. <https://doi.org/10.1097/MCO.0000000000000259>

- Park, B., Whittaker, A. D., Miller, R. K., & Hale, D. S. (1994). Predicting intramuscular fat in beef longissimus muscle from speed of sound. *Journal of Animal Science*, 72(1), 109–116.
<https://doi.org/10.2527/1994.721109x>
- Parry, S. M., El-Ansary, D., Cartwright, M. S., Sarwal, A., Berney, S., Koopman, R., Annoni, R., Puthuchear, Z., Gordon, I. R., Morris, P. E., & Denehy, L. (2015). Ultrasonography in the intensive care setting can be used to detect changes in the quality and quantity of muscle and is related to muscle strength and function. *Journal of Critical Care*, 30(5), 1151.e9-1151.e14.
<https://doi.org/10.1016/j.jcrc.2015.05.024>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32.
https://proceedings.neurips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html
- Peetrons, P. (2002). Ultrasound of muscles. *European Radiology*, 12(1), 35–43.
<https://doi.org/10.1007/s00330-001-1164-6>
- Perrot, V., Polichetti, M., Varray, F., & Garcia, D. (2021). So you think you can DAS? A viewpoint on delay-and-sum beamforming. *Ultrasonics*, 111, 106309.
<https://doi.org/10.1016/j.ultras.2020.106309>
- Provost, J., Papadacci, C., Arango, J. E., Imbault, M., Fink, M., Gennisson, J.-L., Tanter, M., & Pernot, M. (2014). 3D ultrafast ultrasound imaging in vivo. *Physics in Medicine & Biology*, 59(19), L1. <https://doi.org/10.1088/0031-9155/59/19/L1>

- Puthucheary, Z. A., Phadke, R., Rawal, J., McPhail, M. J. W., Sidhu, P. S., Rowlerson, A., Moxham, J., Harridge, S., Hart, N., & Montgomery, H. E. (2015). Qualitative Ultrasound in Acute Critical Illness Muscle Wasting. *Critical Care Medicine*, 43(8), 1603.
<https://doi.org/10.1097/CCM.0000000000001016>
- Rau, R., Schweizer, D., Vishnevskiy, V., & Goksel, O. (2021). Speed-of-sound imaging using diverging waves. *International Journal of Computer Assisted Radiology and Surgery*, 16(7), 1201–1211. <https://doi.org/10.1007/s11548-021-02426-w>
- Richards, J. R., & McGahan, J. P. (2017). Focused Assessment with Sonography in Trauma (FAST) in 2017: What Radiologists Can Learn. *Radiology*, 283(1), 30–48.
<https://doi.org/10.1148/radiol.2017160107>
- Rodriguez-Molares, A., Rindal, O. M. H., Bernard, O., Nair, A., Lediju Bell, M. A., Liebgott, H., Austeng, A., & LØvstakken, L. (2017). The UltraSound ToolBox. *2017 IEEE International Ultrasonics Symposium (IUS)*, 1–4. <https://doi.org/10.1109/ULTSYM.2017.8092389>
- Romero-Laorden, D., Villazón-Terrazas, J., Martínez-Graullera, O., Ibáñez, A., Parrilla, M., & Peñas, M. S. (2016). Analysis of Parallel Computing Strategies to Accelerate Ultrasound Imaging Processes. *IEEE Transactions on Parallel and Distributed Systems*, 27(12), 3429–3440.
<https://doi.org/10.1109/TPDS.2016.2544312>
- Rossman, P., Zagzebski, J., Mesina, C., Sorenson, J., & Mazess, R. (1989). Comparison of speed of sound and ultrasound attenuation in the os calcis to bone density of the radius, femur and lumbar spine. *Clinical Physics and Physiological Measurement*, 10(4), 353.
<https://doi.org/10.1088/0143-0815/10/4/007>
- Ruby, L., Kunut, A., Nakhostin, D. N., Huber, F. A., Finkenstaedt, T., Frauenfelder, T., Sanabria, S. J., & Rominger, M. B. (2020). Speed of sound ultrasound: Comparison with proton density

- fat fraction assessed with Dixon MRI for fat content quantification of the lower extremity. *European Radiology*, 30(10), 5272–5280. <https://doi.org/10.1007/s00330-020-06885-8>
- Ruby, L., Sanabria, S. J., Martini, K., Frauenfelder, T., Jukema, G. N., Goksel, O., & Rominger, M. B. (2021). Quantification of immobilization-induced changes in human calf muscle using speed-of-sound ultrasound: An observational pilot study. *Medicine*, 100(11), e23576. <https://doi.org/10.1097/MD.00000000000023576>
- Sanabria, S. J., Martini, K., Freystätter, G., Ruby, L., Goksel, O., Frauenfelder, T., & Rominger, M. B. (2019). Speed of sound ultrasound: A pilot study on a novel technique to identify sarcopenia in seniors. *European Radiology*, 29(1), 3–12. <https://doi.org/10.1007/s00330-018-5742-2>
- Sanabria, S. J., Ozkan, E., Rominger, M., & Goksel, O. (2018). Spatial domain reconstruction for imaging speed-of-sound with pulse-echo ultrasound: Simulation and in vivo study. *Physics in Medicine & Biology*, 63(21), 215015. <https://doi.org/10.1088/1361-6560/aae2fb>
- Sehgal, C. M., Brown, G. M., Bahn, R. C., & Greenleaf, J. F. (1986). Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers. *Ultrasound in Medicine & Biology*, 12(11), 865–874. [https://doi.org/10.1016/0301-5629\(86\)90004-9](https://doi.org/10.1016/0301-5629(86)90004-9)
- Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M. K., Sun, L., Cong, W., & Wang, G. (2018). 3-D Convolutional Encoder-Decoder Network for Low-Dose CT via Transfer Learning From a 2-D Trained Network. *IEEE Transactions on Medical Imaging*, 37(6), 1522–1534. <https://doi.org/10.1109/TMI.2018.2832217>
- So, H., Chen, J., Yiu, B., & Yu, A. (2011). Medical Ultrasound Imaging: To GPU or Not to GPU? *IEEE Micro*, 31(5), 54–65. <https://doi.org/10.1109/MM.2011.65>
- Song, J., Zhang, Q., Zhou, L., Quan, Z., Wang, S., Liu, Z., Fang, X., Wu, Y., Yang, Q., Yin, H., Ding, M., & Yuchi, M. (2022). Design and Implementation of a Modular and Scalable

- Research Platform for Ultrasound Computed Tomography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(1), 62–72.
<https://doi.org/10.1109/TUFFC.2021.3105691>
- Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2015). *Striving for Simplicity: The All Convolutional Net* (arXiv:1412.6806). arXiv. <https://doi.org/10.48550/arXiv.1412.6806>
- Stähli, P., Frenz, M., & Jaeger, M. (2021). Bayesian Approach for a Robust Speed-of-Sound Reconstruction Using Pulse-Echo Ultrasound. *IEEE Transactions on Medical Imaging*, 40(2), 457–467. <https://doi.org/10.1109/TMI.2020.3029286>
- Strasser, E. M., Draskovits, T., Prashak, M., Quittan, M., & Graf, A. (2013). Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *AGE*, 35(6), 2377–2388. <https://doi.org/10.1007/s11357-013-9517-z>
- Synnevag, J., Austeng, A., & Holm, S. (2009). Benefits of minimum-variance beamforming in medical ultrasound imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 56(9), 1868–1879. <https://doi.org/10.1109/TUFFC.2009.1263>
- Takai, Y., Ohta, M., Akagi, R., Kato, E., Wakahara, T., Kawakami, Y., Fukunaga, T., & Kanehisa, H. (2014). Applicability of ultrasound muscle thickness measurements for predicting fat-free mass in elderly population. *The Journal of Nutrition, Health & Aging*, 18(6), 579–585.
<https://doi.org/10.1007/s12603-013-0419-7>
- Tanter, M., Bercoff, J., Sandrin, L., & Fink, M. (2002). Ultrafast compound imaging for 2-D motion vector estimation: Application to transient elastography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 49(10), 1363–1374.
<https://doi.org/10.1109/TUFFC.2002.1041078>

- Tanter, M., & Fink, M. (2014). Ultrafast imaging in biomedical ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 61(1), 102–119.
<https://doi.org/10.1109/TUFFC.2014.2882>
- Telichko, A. V., Ali, R., Brevett, T., Wang, H., Vilches-Moure, J. G., Kumar, S. U., Paulmurugan, R., & Dahl, J. J. (2022). Noninvasive estimation of local speed of sound by pulse-echo ultrasound in a rat model of nonalcoholic fatty liver. *Physics in Medicine & Biology*, 67(1), 015007. <https://doi.org/10.1088/1361-6560/ac4562>
- Topchyan, A., Tatarinov, A., Sarvazyan, N., & Sarvazyan, A. (2006). Ultrasound velocity in human muscle in vivo: Perspective for edema studies. *Ultrasonics*, 44(3), 259–264.
<https://doi.org/10.1016/j.ultras.2006.02.003>
- Torloni, M. R., Vedmedovska, N., Merialdi, M., Betrán, A. P., Allen, T., González, R., & Platt, L. D. (2009). Safety of ultrasonography in pregnancy: WHO systematic review of the literature and meta-analysis. *Ultrasound in Obstetrics & Gynecology*, 33(5), 599–608.
<https://doi.org/10.1002/uog.6328>
- Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*, 4489–4497.
- Udesen, J., Gran, F., Hansen, K. L., Jensen, J. A., Thomsen, C., & Nielsen, M. B. (2008). High frame-rate blood vector velocity imaging using plane waves: Simulations and preliminary experiments. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 55(8), 1729–1743. <https://doi.org/10.1109/TUFFC.2008.858>
- Van Venrooij, G. E. P. M. (1971). Measurement of ultrasound velocity in human tissue. *Ultrasonics*, 9(4), 240–242. [https://doi.org/10.1016/0041-624X\(71\)90392-1](https://doi.org/10.1016/0041-624X(71)90392-1)

- Von Ramm, O. T., Smith, S. W., & Thurstone, F. L. (1976). Gray scale imaging with complex TGC and transducer arrays. *Application of Optical Instrumentation in Medicine IV*, 70, 266–270. <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/0070/0000/Gray-Scale-Imaging-With-Complex-TGC-And-Transducer-Arrays/10.1117/12.954609.short>
- von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C., & Schuecker, J. (2023). Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 614–633. <https://doi.org/10.1109/TKDE.2021.3079836>
- Wells, P. N. T. (2006). Ultrasound imaging. *Physics in Medicine & Biology*, 51(13), R83. <https://doi.org/10.1088/0031-9155/51/13/R06>
- White, U. A., & Tchoukalova, Y. D. (2014). Sex dimorphism and depot differences in adipose tissue function. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(3), 377–392. <https://doi.org/10.1016/j.bbadis.2013.05.006>
- Whittingham, T. A. (2007). Medical diagnostic applications and sources. *Progress in Biophysics and Molecular Biology*, 93(1–3), 84–110. <https://doi.org/10.1016/j.pbiomolbio.2006.07.004>
- Wilcox, P. D., & Zhang, J. (2018). Quantification of the Effect of Array Element Pitch on Imaging Performance. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(4), 600–616. <https://doi.org/10.1109/TUFFC.2018.2794627>
- Wiskin, J., Malik, B., Borup, D., Pirshafiey, N., & Klock, J. (2020). Full wave 3D inverse scattering transmission ultrasound tomography in the presence of high contrast. *Scientific Reports*, 10(1), Article 1. <https://doi.org/10.1038/s41598-020-76754-3>
- Wolfe, R. R. (2006). The underappreciated role of muscle in health and disease. *The American Journal of Clinical Nutrition*, 84(3), 475–482. <https://doi.org/10.1093/ajcn/84.3.475>

- Xiao, D., Yiu, B. Y. S., Chee, A. J. Y., & Yu, A. C. H. (2019). Channel Count Reduction for Plane Wave Ultrasound Through Convolutional Neural Network Interpolation. In F. Karray, A. Campilho, & A. Yu (Eds.), *Image Analysis and Recognition* (pp. 442–451). Springer International Publishing. https://doi.org/10.1007/978-3-030-27272-2_39
- Xie, J., Xu, L., & Chen, E. (2012). Image Denoising and Inpainting with Deep Neural Networks. *Advances in Neural Information Processing Systems*, 25. <https://proceedings.neurips.cc/paper/2012/hash/6cdd60ea0045eb7a6ec44c54d29ed402-Abstract.html>
- Yiu, B. Y. S., Tsang, I. K. H., & Yu, A. C. H. (2011). GPU-based beamformer: Fast realization of plane wave compounding and synthetic aperture imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 58(8), 1698–1705. <https://doi.org/10.1109/TUFFC.2011.1999>
- Yiu, B. Y. S., Walczak, M., Lewandowski, M., & Yu, A. C. H. (2019). Live Ultrasound Color-Encoded Speckle Imaging Platform for Real-Time Complex Flow Visualization In Vivo. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 66(4), 656–668. <https://doi.org/10.1109/TUFFC.2019.2892731>
- Yoon, C., Lee, Y., Chang, J. H., Song, T., & Yoo, Y. (2011). In vitro estimation of mean sound speed based on minimum average phase variance in medical ultrasound imaging. *Ultrasonics*, 51(7), 795–802. <https://doi.org/10.1016/j.ultras.2011.03.007>
- Young, H.-J., Jenkins, N. T., Zhao, Q., & McCully, K. K. (2015). Measurement of intramuscular fat by muscle echo intensity. *Muscle & Nerve*, 52(6), 963–971. <https://doi.org/10.1002/mus.24656>
- Zhou, D.-X. (2020). Universality of deep convolutional neural networks. *Applied and Computational Harmonic Analysis*, 48(2), 787–794. <https://doi.org/10.1016/j.acha.2019.06.004>

Appendix A

Combining Sparse Matrix Beamforming with Linear Interpolation of RF Signals

In Sec. 3.2.1, we showed how the sparse beamforming matrix could be constructed using the delay-and-sum procedure. In that scenario, an assumption was made that the delayed sample $\Delta(p, e)$ represented the optimally selected signal index for beamforming. With high probability, the geometrically calculated delay values will fall in between discrete sample steps, thus linear interpolation is often applied to derive a better sample value (Yiu et al., 2011). Sparse matrix beamforming is well-equipped to handle linear interpolation schemes, as the linearity of beamforming with respect to the RF data is maintained. The cost of applying linear interpolation is an increase in the nonzero coefficients and accompanying computation efficiency.

Formalizing the incorporation of linear interpolation, we begin from the combination of Eq. (3.1) and Eq. (3.2). Assuming the ideal sample $\Delta(p, e)$ is non-integer, then the value falls between two integers at distance $(1 - \lambda)$ and λ respectively, where λ is some value between 0 and 1. Applying a minor overload of notation, linear interpolation of the ideal RF sample would result in a pixel value equal to:

$$I(p) = \sum_{e=1}^{N_C} \alpha(p, e)(1 - \lambda)\mathbf{D}(\lfloor \Delta(p, e) \rfloor, e) + \lambda\alpha(p, e)\mathbf{D}(\lfloor \Delta(p, e) \rfloor + 1, e) \quad (\text{A1})$$

It is evident in Eq. (A1) that linearity is maintained with the application of linear interpolation. We also observe the distribution of the apodization weights $\alpha(p, e)$ into the terms in the sum. From Eq. (A1), we can follow the steps of Sec. 3.2.1 until we arrive at the updated form of Eq. (3.4), which is expressed as:

$$I(p) = \sum_{e=1}^{N_C} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ (1 - \lambda)\alpha(p, e) \\ \lambda\alpha(p, e) \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{D}(1, e) \\ \vdots \\ \mathbf{D}(\lfloor \Delta(p, e) \rfloor - 1, e) \\ \mathbf{D}(\lfloor \Delta(p, e) \rfloor, e) \\ \mathbf{D}(\lfloor \Delta(p, e) \rfloor + 1, e) \\ \mathbf{D}(\lfloor \Delta(p, e) \rfloor + 2, e) \\ \vdots \\ \mathbf{D}(N_S, e) \end{bmatrix} \quad (\text{A2})$$

Eq. (A2) once again re-expresses the delay-and-sum operation as the sum of inner products of a sparse indicator vector with the corresponding channel of RF data. The coefficients for interpolation and the apodization weights have been consolidated into a single weight value. From the indicator vectors for each channel, the sparse beamforming matrix can once again be constructed, albeit with reduced sparsity. In this way, SMB has naturally been extended to handle linear interpolation. The same argument can be applied for linear-type interpolation schemes that require even more samples, but SMB is limited in its application to more adaptive schemes.

Appendix B

Explanation of Structural Redundancy of Channel-Domain Signals

To explain the origin of the redundant structure present in the pulse-echo signals received over different array channels, let us consider the basic case of pulse-echo sensing for a point target. Figure B1 illustrates a setup of five channels and a reflecting point source on a 2-D plane. Each receiving element has coordinates $(0, x_0)$, $(0, x_1)$, $(0, x_2)$, $(0, x_3)$, and $(0, x_4)$, respectively, and the point target has coordinates (z_p, x_p) . Assuming an ideal plane wave transmission at time $t_0 = 0$ and a constant speed of sound c_0 , each transducer element (indexed by e) would receive the backscattered signal from the point source at the following time:

$$t_e = \frac{z_p + \sqrt{z_p^2 + (x_p - x_e)^2}}{c_0} \quad (\text{B1})$$

Rearranging the terms to remove the radical and forming the equation of a conic section, we arrive at:

$$\frac{c_0^2}{z_p^2} \left(t_e - \frac{z_p}{c_0} \right)^2 - \frac{1}{z_p^2} (x_p - x_e)^2 = 1 \quad (\text{B2})$$

In this scenario, the signal received by each element is identical apart from the time of reception. Because of the isotropic reflection of the point source, the delays form a hyperbola in relation to the spacing between elements as in Eq. (B2). This delay calculation is the fundamental principle behind DAS beamforming, but this spatial structure can also be leveraged in a novel way for data reduction.

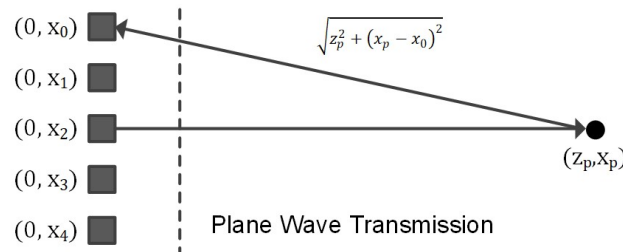


Figure B1 Illustration of a five-element transducer transmitting an unsteered plane wave pulse (dashed line); elements (denoted with squares) and scattering point source (circle) are projected onto a 2-D plane.

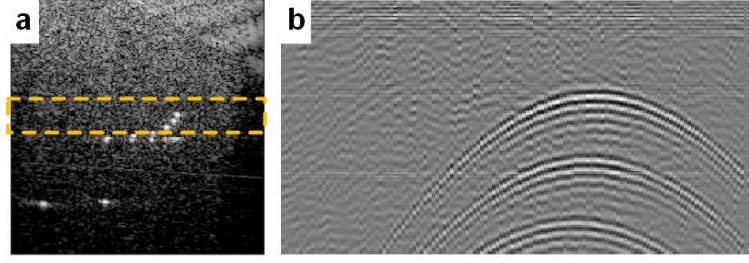


Figure B1 (a) B-mode image of a multi-point target phantom and (b) corresponding 128 channel by 200 sample snippet of raw RF data highlighted in dashed yellow box.

From algebra and geometry, given only two pairs of values (t_e, x_e) and knowing the delays form a hyperbolic shape, the equation becomes fully determined and can be obtained by solving a system of equations with two variables. Therefore, in the single point-source, five-transducer element example without noise or other adverse effects, three of the five delayed signals can be considered to carry redundant information and can be perfectly reconstructed from the other two. Extrapolating from this example to a fully populated linear array, we argue that there exist many such spatiotemporal redundancies between fast-time sets of RF data collected from neighboring channels. Each channel receives time-shifted version of echo pulses from the same scattering sources in the insonified medium.

Let us now extend the explanation to the case of a multi-point target phantom. Figure B1 shows a snippet of the spatiotemporal structure of the RF data for such a phantom (the same one reported in Section 4.4.4; data were acquired using the same imaging parameters as those reported in Table 4.1). For each point target in the phantom, three of which are highlighted in yellow in Fig. B1(a), the hyperbolic shape of the corresponding echoes is shown in Fig. B1(b). While the hyperbolic relationship in (B2) still exists, the signals can suffer from the nonidealities of attenuation and transducer angle dependence, as depicted by the reduction in signal strength away from the center of the hyperbolic structure, and potential superposition from other scatterers. Additionally, the signals from human tissue rarely manifest as those from true echogenic point sources. These reasons prohibit the exact use of (B2) to recover the unacquired channels in a typical imaging scenario. Instead, we take the data-driven approach presented in this work to learn an optimal method for leveraging the transducer channel redundancy to recover unacquired channels.

Appendix C

List of Video Files Referenced in Thesis

Movie2.1.mp4: A video of the SoS framework operating in real-time, applied to *in vitro* agar phantoms. Shown in the movie is a camera recording of real-time operation on the right, with the screen capture for clearer visibility of the SoS estimate and B-mode image on the left. In the camera recording, there are three phantoms: 1) an agar staircase labeled “Slow (2% GLY) 1473 m/s” on the left, 2) an agar bilateral phantom with wire targets in the center, 3) an agar staircase labeled “Fast (33% GLY) 1656 m/s” on the right. The probe starts on the left staircase phantom, moves across the bilateral phantom, then ends on the right staircase phantom. The real-time SoS estimates on both the staircase phantoms closely match the labelled SoS values derived from through-transmit reference measurements. As the probe slides across the middle bilateral phantom, the SoS smoothly transitions from the values for the slower phantom to the values for the faster phantom.

Movie2.2.mp4: A video of the SoS framework operating in real-time, applied to *ex vivo* porcine samples. Shown in the movie is a camera recording of real-time operation on the right, with the screen capture for clearer visibility of the SoS estimate and B-mode image on the left. In the camera recording, there are two porcine samples: a sample labeled “Pork Shoulder 1576 m/s” on the left, a sample labeled “Pork Loin 1599 m/s” on the right. The probe starts on the left samples, is moved and placed on the right sample, then moves back to the left sample and finally ends on the right sample. The real-time SoS estimates on both porcine samples closely match the labelled SoS values derived from through-transmit reference measurements.

Movie2.3.mp4: A video of the SoS framework operating in real-time, applied to *in vivo* human biceps. Shown in the movie is a camera recording of real-time operation on the right, with the screen capture for clearer visibility of the SoS estimate and B-mode image on the left. In the camera recording, the focus is on the individual’s biceps. The probe is placed on the medial side of the biceps and held in position as the individual relaxes and flexes the biceps muscle. The real-time SoS estimates change as the biceps is flexed or relaxed, ranging from 1600 m/s (relaxed) to 1583 m/s (flexed).