

# Numerical Simulations for Characterization of Missing Articulatory Information in Ultrasound Imaging for Speech

Sarah R. Li,<sup>1</sup> Suzanne Boyce, PhD, CCC-SLP,<sup>2</sup> & T. Douglas Mast, PhD<sup>1</sup>

<sup>1</sup>Biomedical Engineering; <sup>2</sup>Communication Sciences and Disorders | University of Cincinnati

## Introduction

- Due to the air-tissue interface at the tongue surface, submental (below the chin) **B-mode ultrasound images provide tongue shape information** useful for understanding typical and disordered tongue articulation during speech production (e.g., for speech therapy [1]).
  - Speech acoustics analyzes the location and extent of constrictions in the vocal tract (e.g., between the tongue and palate) [2], so **interest is the geometric shape of the air-tongue interface**, often shown as a tongue surface contour.
- However, the anterior **tongue tip may be obscured by shadowing** from the sublingual air space and/or mandible bone [3].
  - Depends on an individual speaker's anatomy and speech movement patterns
  - The amount of anterior tongue missing from the image has not been investigated thoroughly.
- Magnetic resonance images (MRI) show the entire vocal tract [2] and can thus be used to understand missing information.
- To avoid ambiguity of tongue shapes, **numerical simulations of acoustic wave propagation with the k-Wave toolbox** [4] can generate B-mode ultrasound images with known tongue shapes segmented from MRI.
  - Simulated images can help understanding of tongue tip visibility (see **5aSC37**).

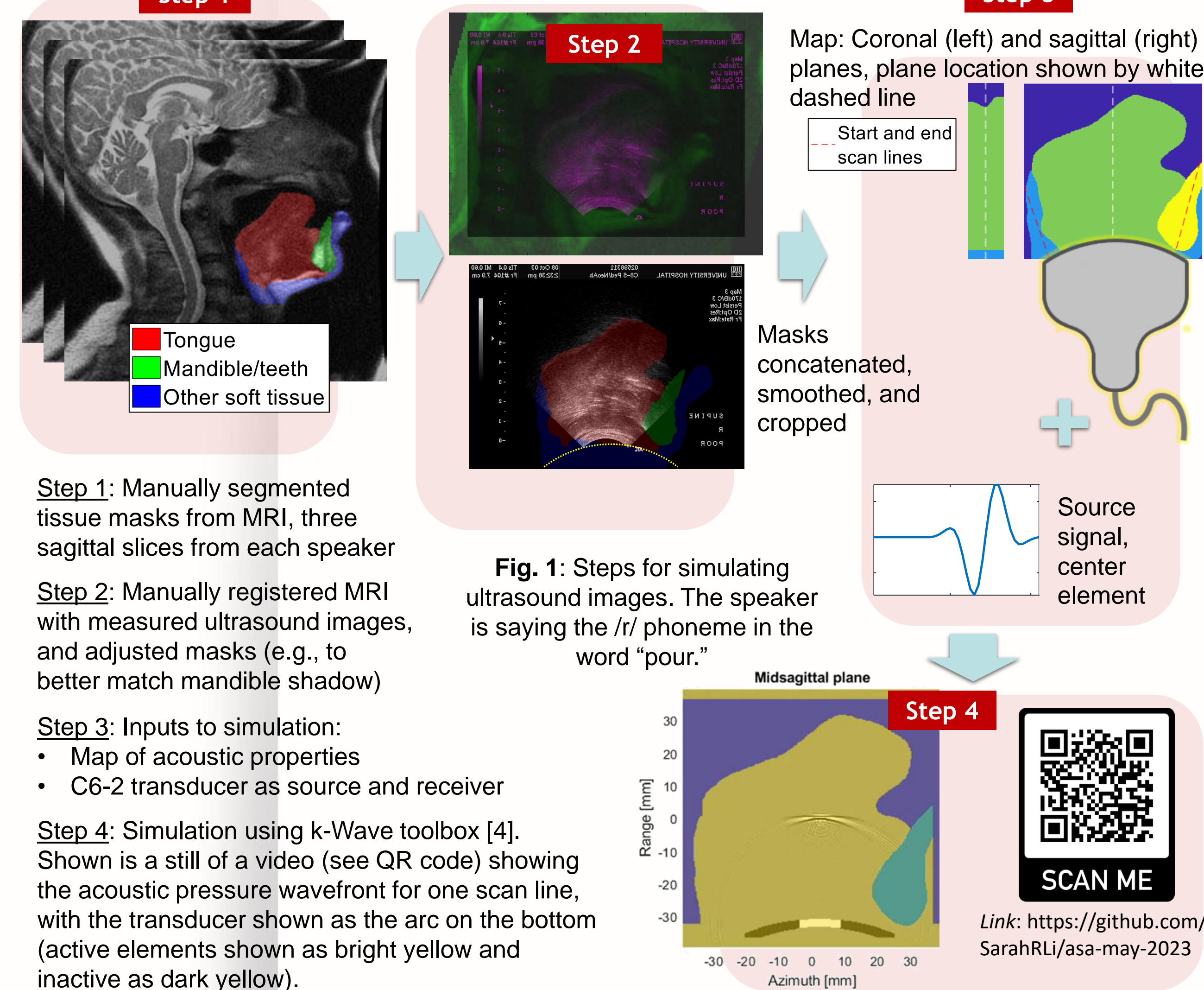
## Hypothesis

Numerical simulations using the k-Wave toolbox [4] and segmented tissue maps from MRI can efficiently generate midsagittal B-mode ultrasound images:

- Replicate features of interest (e.g., shadowing, reverberation artifacts) in measured B-mode images
- Secondary goal: determine whether information on tongue tip can be recovered from reverberation artifacts

## Methods: Data

### Data simulation workflow



## Datasets used

- Speakers:** 23 speakers producing the American English /r/ phoneme
- Stimuli and image parameters:**
  - MRI (for tongue tissue maps):** in supine position; midline sagittal slice with two parasagittal slices, 240x240 mm<sup>2</sup> field-of-view, 3-5 mm slice width, resolution of 0.938-1 mm per pixel
  - Measured ultrasound (for validating simulated ultrasound):** midsagittal image in the supine or upright position
    - Varying imaging parameters: curvilinear probes C5-2, C7-4, C8-5 on HDI 5000 machine or C6-2 (center frequency 4 MHz) on Siemens Acuson X300

## General parameters

- A linear, fluid (compressional-wave) simulation of acoustic wave propagation in 3D with attenuation was performed in k-Wave.

<sup>a</sup> Based on attenuation power law exponent,  $\gamma$  of 1.032 (see section "Post-process filtering" on choice of 4 MHz)  
<sup>b</sup> Tongue tissue additionally has scatterers added from a random distribution.  
<sup>c</sup> See section "Air-tissue interface"  
<sup>d</sup> See section "Replicating the bone shadow"

Table 1: Transducer model

Transmit/receive apodization	Hanning
Active elements	27 of 192
Element pitch (mm)	0.385
Element length (cm)	1.1
Azimuthal focal depth (cm)	4

Table 2: Final simulation parameters.

Grid step size, dx (mm)	0.154
Points per wavelength (PPW, for soft tissue at 4MHz)	2.6
Courant-Friedrichs-Lewy (CFL) number	0.26

Table 3: Final acoustic properties for media used in grid, from [5, 6].

Medium	Speed of sound (m/s)	Density (kg/m <sup>3</sup> )	Attenuation (Np/m) at 4 MHz <sup>a</sup>
Soft tissue (including tongue muscle)	1588.4 <sup>b</sup>	1090 <sup>b</sup>	12.13
Air (at 35°C)	351.9	90 <sup>c</sup>	12.13
Bone and teeth	925 <sup>d</sup>	1090 <sup>d</sup>	1444

## Air-tissue interface

Adjustments to spatial and time steps were made as long as most of the acoustic wave was reflected at the air-tissue interface, demonstrated by reflection amplitude values close to 1 in Table 4.

- Lowered PPW:** Based on soft tissue (PPW for air: 0.55)
- Final model:** Avoided computational instability at higher CFL numbers
  - Instability possibly associated with large air/tissue impedance mismatch
  - Increasing the air density allowed stability at higher CFL (larger time steps).

Table 4: Models for measuring reflection amplitudes at tissue-air interface, from a linear 1D simulation without attenuation

Model	PPW	Density of air (kg/m <sup>3</sup> )	CFL	Reflection amplitude
Realistic	500	1.15	0.03	0.9995
Lowered PPW	2.5	1.15	0.03	0.8538
Final model	2.5	90	0.3	0.9556

## Replicating the bone shadow

- Unrealistic reflections from mandible resulted from realistic compressional-wave bone properties, likely from neglect of mode conversion to shear waves. Elastic model [7] is computationally expensive.

- Matching soft tissue:** To reduce reflection, sound speed and density in bone was matched to soft tissue. Bone absorption coefficient was increased to attenuate greater energy transmitted into bone.
  - Reflections still high, likely due to dispersion effects

- Final model:** Bayesian optimization used to find sound speed minimizing reflections

<sup>1</sup> Required a CFL of 0.075

Table 5: Models adjusting bone properties

Model (fluid)	Speed of sound (m/s)	Density (kg/m <sup>3</sup> )	Attenuation (Np/m) at 4 MHz <sup>a</sup>
(A) Realistic <sup>1</sup>	3198	1990	170
(B) Matching soft tissue	1588.4	1090	722
(C) Final model	925	1090	1444

Artifact: reflection from mandible

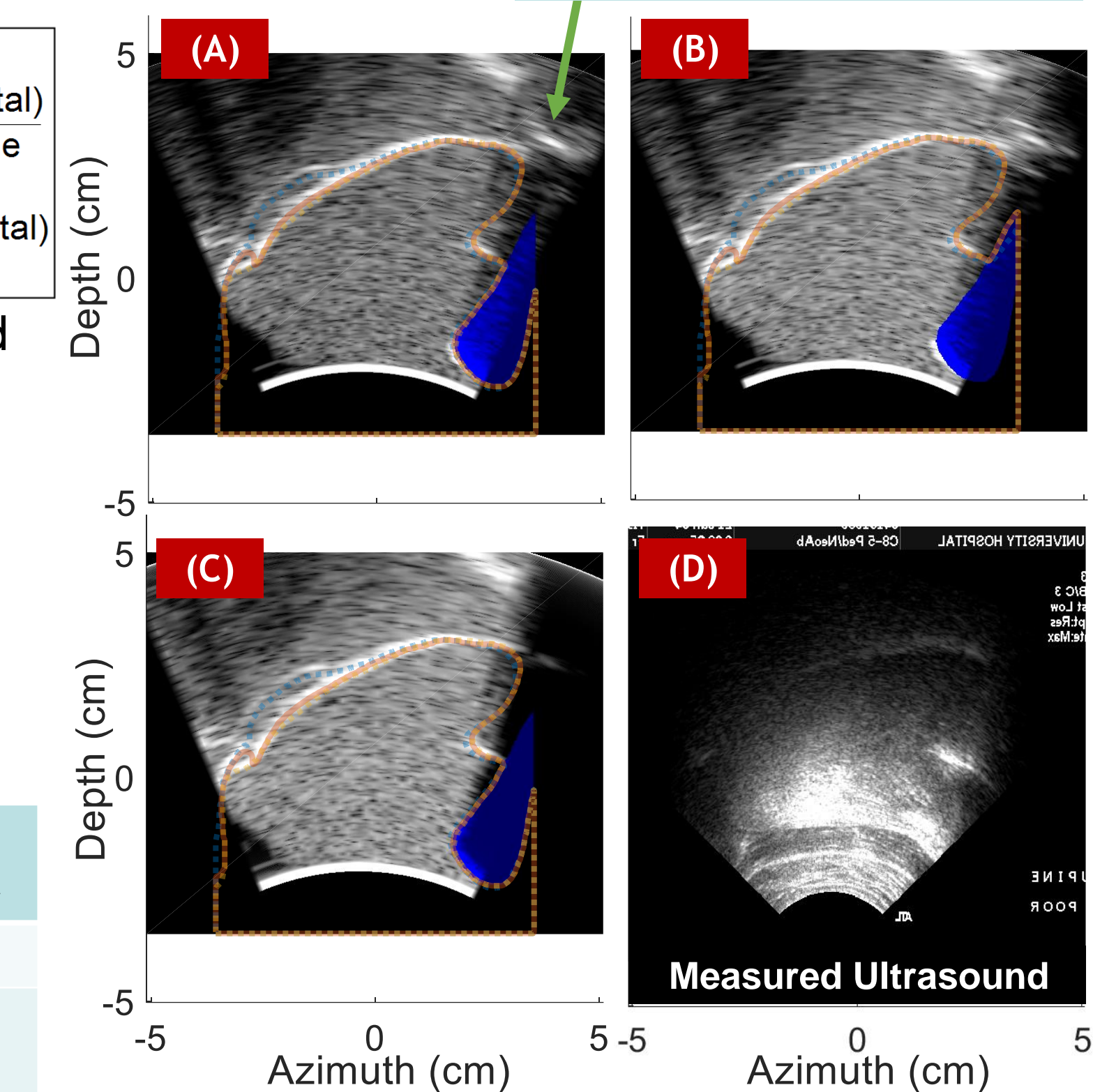


Fig. 2: Ultrasound images for testing bone properties; simulated images (A, B, C) filtered with center at 3 MHz

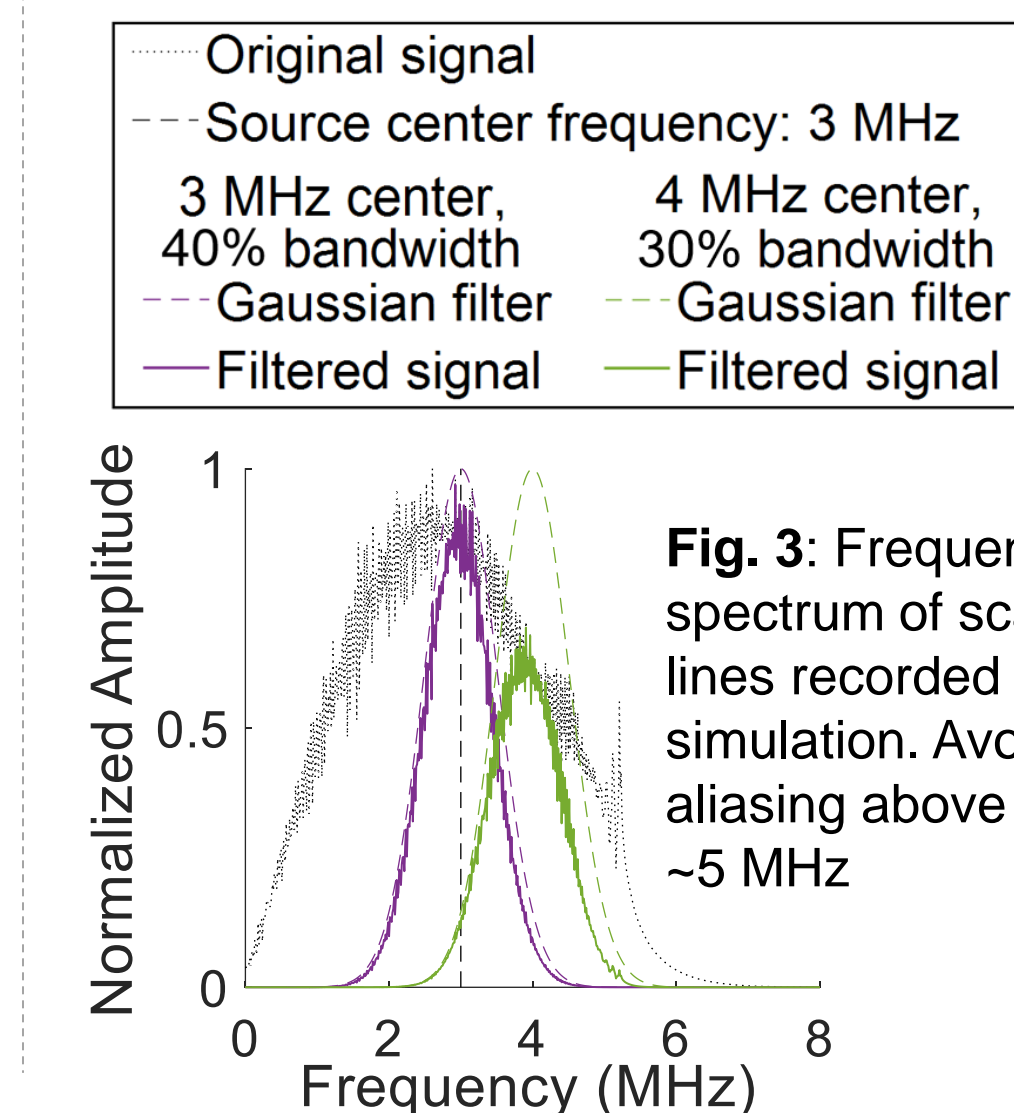


Fig. 3: Frequency spectrum of scan lines recorded in simulation. Avoids aliasing above ~5 MHz

## Post-process filtering

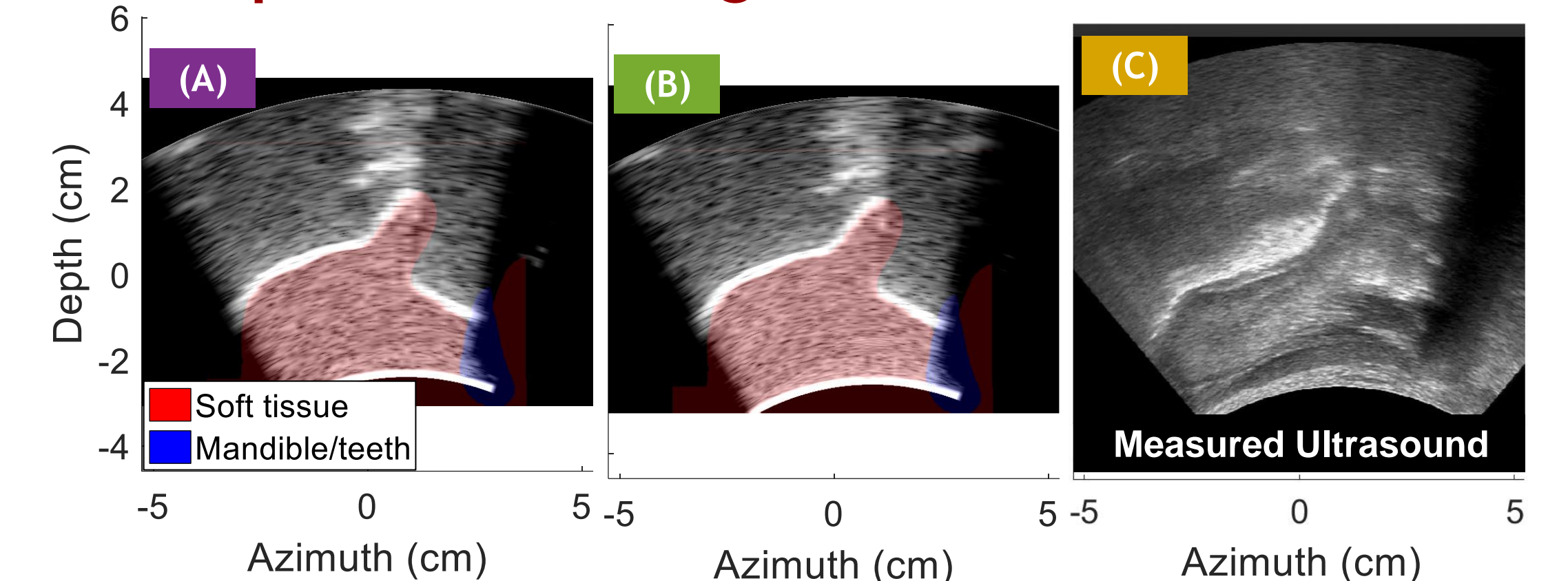


Fig. 4: Simulated images from bandpass-filtered (Fig. 3) scan lines, Gaussian centered at (A) 3 MHz and (B) 4 MHz. Final image used (B) was more similar to (C), measured ultrasound

## Results and discussion

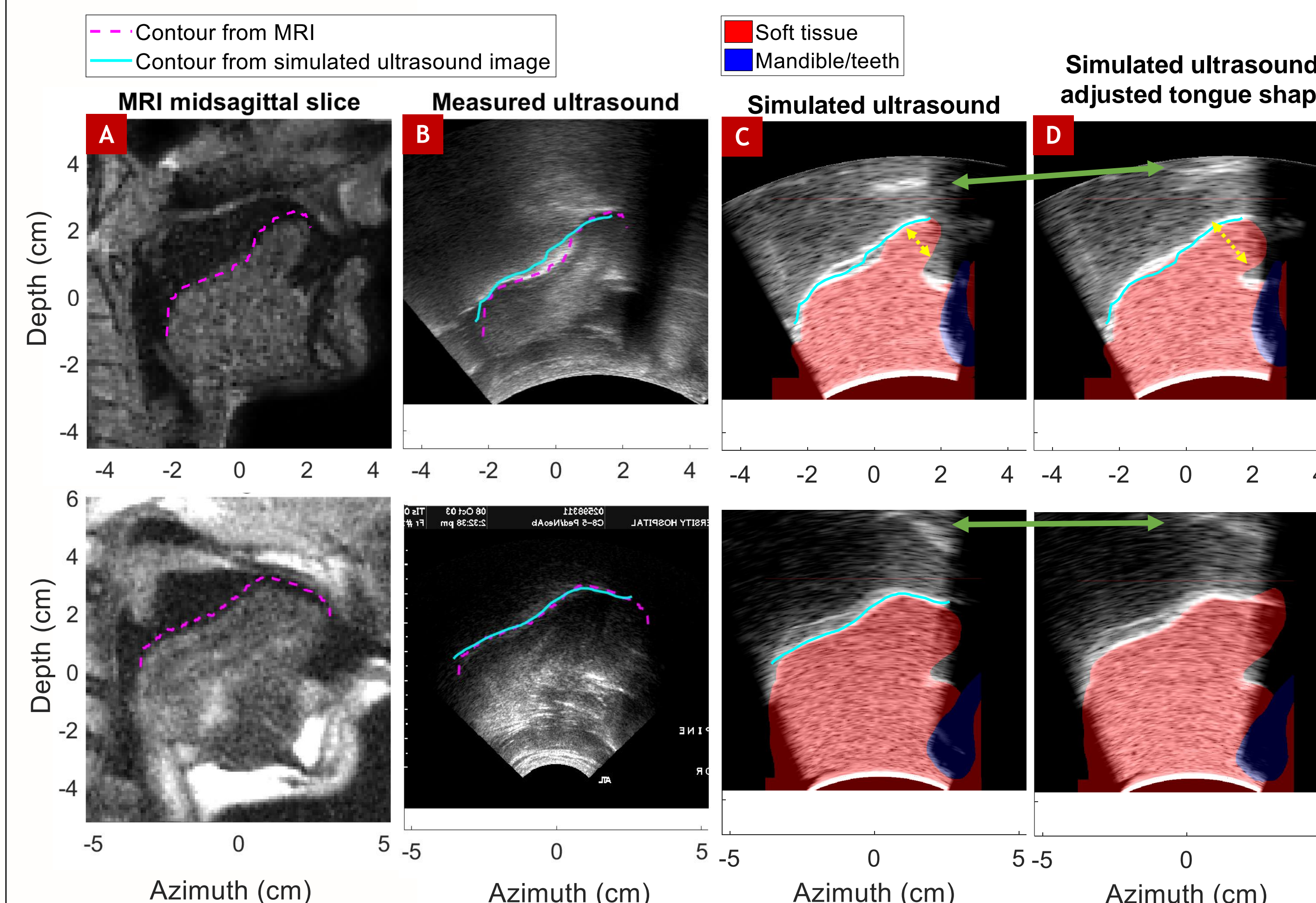


Fig. 5: Example images and contours. The simulated ultrasound images (C & D) are additionally tinted to show soft tissue as red and mandible/teeth as blue in the tissue maps used, with D showing that a different simulation with an altered tongue tip shape changes the reverberation artifact (green arrows indicating artifact).

- Simulated ultrasound images can replicate shadowing and artifacts in measured ultrasound images.
- Adjustments to the anterior tongue shape (Fig. 5D) changed the reverberation artifact (green arrows), demonstrating that the artifact provides tongue tip information.
- Tongue tip thickness (yellow dotted line in Fig. 5C/D, row 1) is apparent.
  - Similar to distance between artifact and tongue surface
- Tongue tip position within the shadow may not be apparent.
  - In Fig. 5D, row 2, little indication that the adjusted tongue tip was curled upwards compared to 5C
- Some artifacts not seen in measured images; tongue musculature may increase scattering, decreasing energy returned to transducer compared to simulations.
- Still, may be enough to categorize tongue shapes (see **5aSC37**): less tongue tip thickness implies a retroflex /r/ vs. a bunched /r/ shape.

## Computational efficiency

<sup>1</sup> Perfectly matched layer (PML) for absorbing boundaries

Table 6: Total computation times; maps are cropped to tongue size.

	Smallest tongue / map	Largest tongue / map
Computation time: NVIDIA A100 (40 GB) GPU	53 s × 35 scan lines = 31 min	141 s × 51 scan lines = 2 hr
Computation time: NVIDIA RTX2080 Super (8 GB) GPU	91 s × 35 scan lines = 53 min	489 s × 51 scan lines = 7 hr
Number of grid points (excluding PML <sup>1</sup> )	388 × 392 × 58	602 × 627 × 58

## Future work

- Adjust bone parameters to better match measured reflection
  - To understand visibility of tongue tip for the /r/ phoneme
- Decrease PML size, which was large (varied for each map, mean of 36 grid points) to decrease computation time
- Use simulated maps from 3D MRI to understand double-edge artifact (i.e., groove in tongue causes two apparent contours)

## Acknowledgements

- NIH/NIDCD grants F31 DC020672, R01 DC05250, 8 UL1 TR000077-05, R01 DC013668, and R01 DC017301
- Additional mentors Dr. Steven Lulich and Dr. Shrikanth Narayanan for discussions on research project, as well as Dr. Maureen Stone for discussion on initial project idea
- Siemens Medical Solutions (lent Acuson X300 ultrasound scanner)

## References

- J. L. Preston et al., "Ultrasound Images of the Tongue: A Tutorial for Assessment and Remediation of Speech Sound Errors," *J Vis Exp*, no. 119, p. e55123, Jan. 2017, doi: 10.3791/55123.
- X. Zhou, C. Y. Espy-Wilson, S. Boyce, M. Tiede, C. Holland, and A. Choe, "A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/," *J Acoust Soc. Am*, vol. 123, no. 6, p. 17, Jul. 2008, doi: 10.1121/1.2902168.
- M. Stone, "A guide to analysing tongue motion from ultrasound images," *Clin Linguist Phon*, vol. 19, no. 6-7, pp. 455-501, Jan. 2005, doi: 10.1080/02699200500113558.
- B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J Biomed Opt*, vol. 15, no. 2, p. 021314, 2010, doi: 10.1117/1.3360308.
- T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*, in Biomedical Engineering. Elsevier Science, 2013.
- P. A. Hasgall et al., "IT'IS Database for thermal and electromagnetic parameters of biological tissues," Version 4.1, Feb. 2022, doi: 10.13099/VIP21000-04-1. itis.swiss/database.
- B. E. Treeby, J. Jaros, D. Rohrbach, and B. T. Cox, "Modelling elastic wave propagation using the k-Wave MATLAB Toolbox," in *2014 IEEE International Ultrasonics Symposium*, Chicago, IL, USA: IEEE, Sep. 2014, pp. 146-149, doi: 10.1109/ULTSYM.2014.0037.