

Comparison of Deep Learning and Signal Processing Methods for Removing a Ringing Artifact from Ultrasound Signals

Yana Sosnovskaya
dept. of ECE
University of Washington
Seattle, WA, USA
ysos@uw.edu

Eli Shlizerman
dept. of Appl. Math.
University of Washington
Seattle, WA, USA
shlizee@uw.edu

Blake Hannaford
dept. of ECE
University of Washington
Seattle, WA, USA
blake@uw.edu

Mika N. Sinanan
dept. of Gen. Surg.
University of Washington
Seattle, WA, USA
mssurg@uw.edu

Abstract—Minimally Invasive Surgeries can benefit from having miniaturized sensors on surgical graspers to provide additional information to the surgeons. In this work, a 6mm ultrasound transducer was added to a surgical grasper, intended to measure acoustic properties of the tissue. However, the ultrasound sensor has a ringing artifact arising from the decaying oscillation of its piezo element, and at short travel distances, the artifact blends with the acoustic echo. Without a method to remove the artifact from the blended signal, this makes it impossible to measure one of the main characteristics of an ultrasound waveform – Time of Flight. In this paper, six filtering methods to clear the artifact from the ultrasound waveform were compared: Bandpass filter, Adaptive Least Mean Squares (LMS) filter, Spectrum Suppression (SPS), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). Following each filtering method, four time of flight extraction methods were compared: Magnitude Threshold, Envelope Peak Detection, Cross-correlation and Short-time Fourier Transform (STFT). The RNN with Cross-correlation method pair was shown to be optimal for this task, performing with the root mean square error of 3.6 %.

Index Terms—ultrasound sensor, surgical grasper, noise removal algorithms, signal processing, deep learning

I. INTRODUCTION

Minimally Invasive Surgery (MIS) is now standard in modern medicine and involves operating through small incisions using laparoscopic graspers for manipulation, and endoscopic cameras for visual feedback. Advantages of MIS include faster recovery, less blood loss, and a lower risk of complications. Alongside its benefits, MIS also brings new challenges, such as the lack of tactile feedback for surgeons. In open surgery, surgeons can palpate the tissue to gain information about its non-visible structure, such as tumors and blood vessels [1].

Part of our approach to providing the tactile sensing and diagnostic capability for surgical graspers is the addition of a miniaturized Steminc SMD063T07R111 ultrasound transducer to the tip of a Surgical Babcock Grasper, intended to mechanically interface with the tissue and measure its acoustic properties.

This material is based upon work supported by the National Science Foundation under Grant No. 2036255.

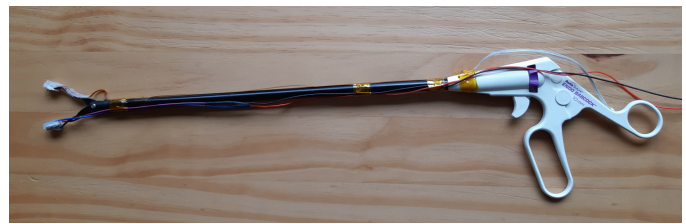
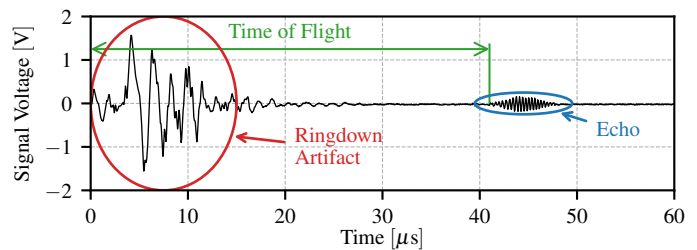
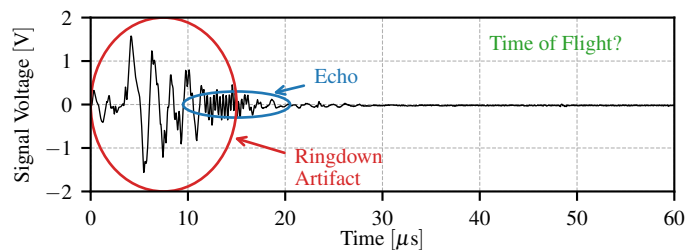


Fig. 1. Surgical grasper with an attached ultrasound sensor on the tip.



(a) Long distance ultrasound signal with artifact.



(b) Short distance ultrasound signal with artifact.

Fig. 2. Raw ultrasound signals with ringdown artifacts.

Ultrasound is widely used for diagnostic, prognostic and treatment applications, such as abnormality detection, tissue characterization, high-intensity ultrasound ablation therapy and others. An ultrasound transducer is a piezo crystal that has a resonant oscillation frequency. All piezo crystals can operate in both directions: transmitting and receiving sound waves. If a time-varying potential difference is applied across the electrodes, the piezo element will start oscillating to produce

a sound wave. Conversely, if a sound pressure is applied to a piezo crystal, it will generate a voltage across its leads. When an ultrasound pulse travels through the tissue, it undergoes continuous modifications, which depend on the characteristics of the sound waves as well as tissue properties. Several characteristics of these sound waves are particularly important for tissue characterization: time of flight (TOF), which depends on the velocity of sound, and attenuation [2].

Miniaturized ultrasound transducers have not been widely used in surgical graspers. In one application, a transmitter-receiver transducer pair on Martin forceps was used to detect soft tissue cancers, by evaluating the echo attenuation [3]. Our use differs by using a 1D ultrasound transducer in bidirectional A-mode, which switches between transmitter and receiver functionalities. Operation in bidirectional mode saves space and gives an opportunity to add other sensors on the surgical grasper's jaws by using only one transducer instead of two. However, it is also a drawback: the system records exponentially decaying "ringing" (aka, the "ringdown artifact") of the transducer after excitation. When a surgical grasper's jaws operate on a tissue, the distance between them can be under 1 cm. As seen in Fig. 2, at these short distances, the ringdown artifact blends with the received echo, making it impossible to identify the TOF without a method to remove the artifact.

Short-distance ultrasound, in which the echo can blend with the ringdown, is applicable in many fields. These include the estimation of the degenerative loss of skeletal muscles [4], obstacle avoidance and mapping ultrasound sensors for underwater robots [5], thickness and defect detection applications in non-destructive testing [6], and other applications. All of these applications fundamentally suffer from the ringdown artifact when an A-mode bidirectional ultrasound transducer is used. In spite of all the above-mentioned applications, limited published literature exists on ringdown artifact removal from ultrasound signals, and there does not appear to be applicable work on AI-based noise removal algorithms for 1D ultrasound waveforms.

One application of a ringdown artifact removal from an ultrasound signal is a gastrointestinal capsule with a 1D ultrasound for microanatomical diagnostics [7]. The capsule uses a ringdown-compensating filter by subtracting a moving average of the ringdown portion of the signal from the combined signal. This approach can help at longer TOFs, but not at short ones: at short TOFs, the ringdown blends with the echo, and the moving average starts including the echo as well.

Besides 1D applications like the surgical grasper or the diagnostic capsule above, ringdown artifact removal is also needed for short-range ultrasound imaging, in which the ringdown artifact adds a blind spot or dead space. A patent by Barlow, et al. claims a method for such removal by obtaining multiple "reference scans" (scans taken in an echo-free space), averaging them in a specific way, and subtracting them from the combined scan [8]. This method could potentially be extended to a single bidirectional transducer, but its main limitation (specified by the inventors) comes from "ringdown drift": the

changing of the ringdown of a sensor over time. Applying such a method to a transducer of a surgical grasper would therefore require frequent (and potentially lengthy, depending on the number of reference scans required) recalibration, which is not feasible during a surgery.

The need to remove ringdown from short-range ultrasound signals also arises in ultrasonic thickness measurement systems with bidirectional transducers. Most such systems use multiple transducers, but some, such as the Elcometer Ultrasonic Precision Thickness Gauges, use single bidirectional transducers. These systems are able to measure short-range echoes by placing a "delay line" material between the transducer and the material being inspected, which artificially elongates TOF, thus giving the ringdown time to decay prior to receiving the echo. The signal is also filtered through frequency selective filters, which further helps remove the ringdown, and also Gaussian noise. However, this delay line is also at least 9 mm thick, which makes it prohibitively bulky for a surgical grasper.

The goal of this paper is to apply, evaluate and compare several signal processing and AI-based methods for clearing the corrupted ultrasound echo waveforms to accurately estimate the TOF, using the algorithm illustrated in Fig. 3.

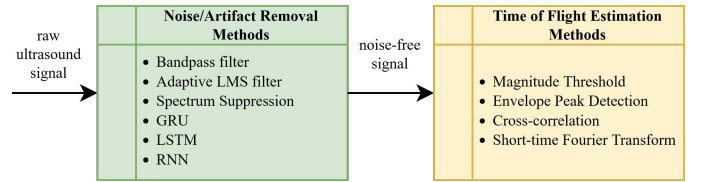


Fig. 3. Architecture for method comparison in two-stage TOF estimation.

II. MATERIALS AND METHODS

A. Denoising Algorithms Description

Fig. 3 illustrates the two-stage process with which the ringdown artifact is removed from the ultrasound signal, and TOF is then estimated from the filtered signal. Six noise removal algorithms for the first stage are compared: three traditional, and three deep learning ones for comparison. For traditional algorithms, the Bandpass filter, the Adaptive LMS filter, and the Spectrum Suppression method (SPS) were chosen. These traditional algorithms were chosen because these algorithms are used for denoising conventional audio signals, speech, and Natural Language Processing.

The Bandpass filter was designed with passband frequencies of 2.5 MHz to 3.5 MHz to bracket it over the 3 MHz resonant frequency. The Adaptive LMS filter uses the target signal to optimize the coefficients for the filter. From preliminary attempts, the LMS filter order of 13 with a step size of 0.004 was selected for use on all signals of interest. Initial coefficients were based on the coefficients of the above Bandpass filter with the Hamming window.

The SPS method used here was adapted from a paper by Boll [9]. The method relies on a separate noise signal (i.e., the ringdown artifact measured from a free transducer) and the

blended signal. The signals are transformed to the frequency domain via Fast Fourier Transform (FFT), and in the frequency domain, the noise is subtracted from the blended signal. An inverse FFT is then used to obtain the denoised signal in the time domain. Boll followed this sequence with a lowpass filter; in this paper, the above Bandpass filter was used on the denoised signal instead.

For deep learning algorithms, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) were chosen for comparison, because these algorithms work with time series data, and have previously been used for denoising speech signals. Of the three, RNN is the simplest and slowest to train one, LSTM has the most representational power and is generally cheaper to train than standard RNNs, and GRU is a streamlined, cheaper to train version of LSTM. Chollet presents the theory of the three methods [10]. In this work, the tanh activation function was used for all three deep learning networks, with the Adam optimizer with the learning rate of 5×10^{-4} and 330 training epochs. The Euclidean norm of the error between the input and target signals was used as the loss function. All trained networks had two layers with 512 and 256 units. All three networks used a bidirectional architecture.

B. Time of Flight Estimation Methods

In the second stage of the two-stage TOF estimation process in Fig. 3 the actual TOF estimation algorithm is applied, producing the TOF from the cleared signal. TOF is the time required for a sound wave to travel through a medium. TOF depends on the distance traveled, and the medium's properties, including density, compressibility, and rigidity, which are temperature-dependent.

Four TOF estimation methods are utilized in this work.

Magnitude Threshold Method, used widely in hardware setups for automatic detection of TOF. When the amplitude of the received signal reaches a preset threshold, it records the time. The threshold should bypass the maximum noise level expected. This method is sensitive to noise fluctuations and signal decay with distance traveled, but is the most straightforward to implement [11].

Envelope Peak Detection consists of approximating the echo signal as an analytical signal with complex components, from which a Hilbert envelope is computed. Then, peaks are found in the signal's envelope. The TOF is then detected either by locating the peak with the largest amplitude. Such envelope signal is more robust to fluctuations of the echo's peaks, as well as a variable white noise level [12].

Cross-correlation Method (CC), in its simplest form, consists of cross-correlating the transmitted and received signals, and estimating the TOF as the maximum in the correlation function [13].

Short-time Fourier Transform Method (STFT) for TOF estimation uses the time-frequency representation of ultrasound signals. STFT is obtained with a sliding window over the time domain signal, and taking the Fourier transform of the

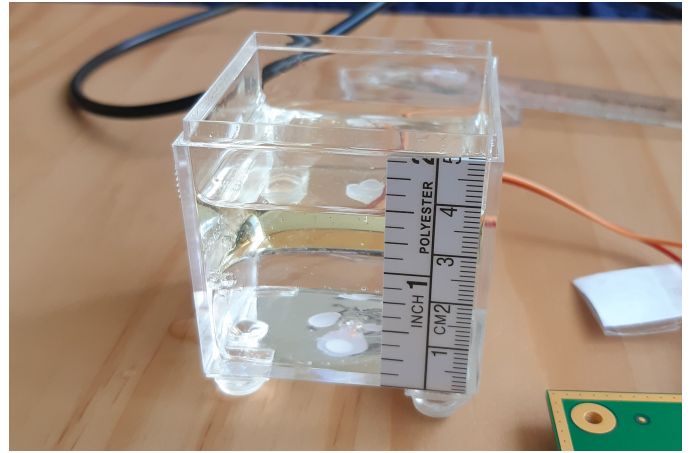


Fig. 4. Acrylic container with an attached ultrasound transducer at the bottom.

window. Then, only windows with the frequency of interest are considered for the estimation of TOF [14].

C. Data Collection and Preprocessing

Apart from the sensorized surgical grasper (Fig. 1), another experimental setup (Fig. 4) was developed for ultrasound data collection, consisting of an acrylic container with an attached ruler to measure the level of liquid in the container – the one-way distance traveled. The bottom acrylic thickness of the container is 2 mm. The same SMD063T07R111 transducer as the one used on the surgical grasper was glued to the bottom of the acrylic container. Similarly, the same data acquisition setup was used: TI TDC1000 to drive the transducer at 3 MHz with a pulsed square waveform with 6 impulses, Siglent SDS 1104X-E oscilloscope to provide the external clock and acquire the signal. The data was recorded by the oscilloscope in *csv* format, combined and aligned in MATLAB. Then, the dataset was downsampled by a factor of 26 to a new sampling rate of 19.23 MHz, leading to 1279 samples per waveform.

In order to use the dataset for training in deep learning models, it is required to have target signals to compare with and to adjust weights accordingly. When generating target signals, at distances of 2 cm and longer, the ringdown and the echo were sufficiently far apart, which made it possible to isolate the echo by zeroing the ringdown time segment. The waveforms obtained at shorter distances were not used as target signals for the training dataset, due to the difficulty of extracting the echo from such signals. The training dataset contains 993 waveform pairs (consisting of a raw signal, and a target signal with only the echo), with an approximately equal number of waveforms per distance. A separate test dataset was collected at 9 distances from 0.5 cm to 4.0 cm, using liquid water as the medium. This testing dataset has 270 waveforms, 30 waveforms per distance. Both datasets were saved in *mat* format for further processing in Google Colaboratory. Both the downsampled and original resolution datasets have been published on IEEE DataPort [15].

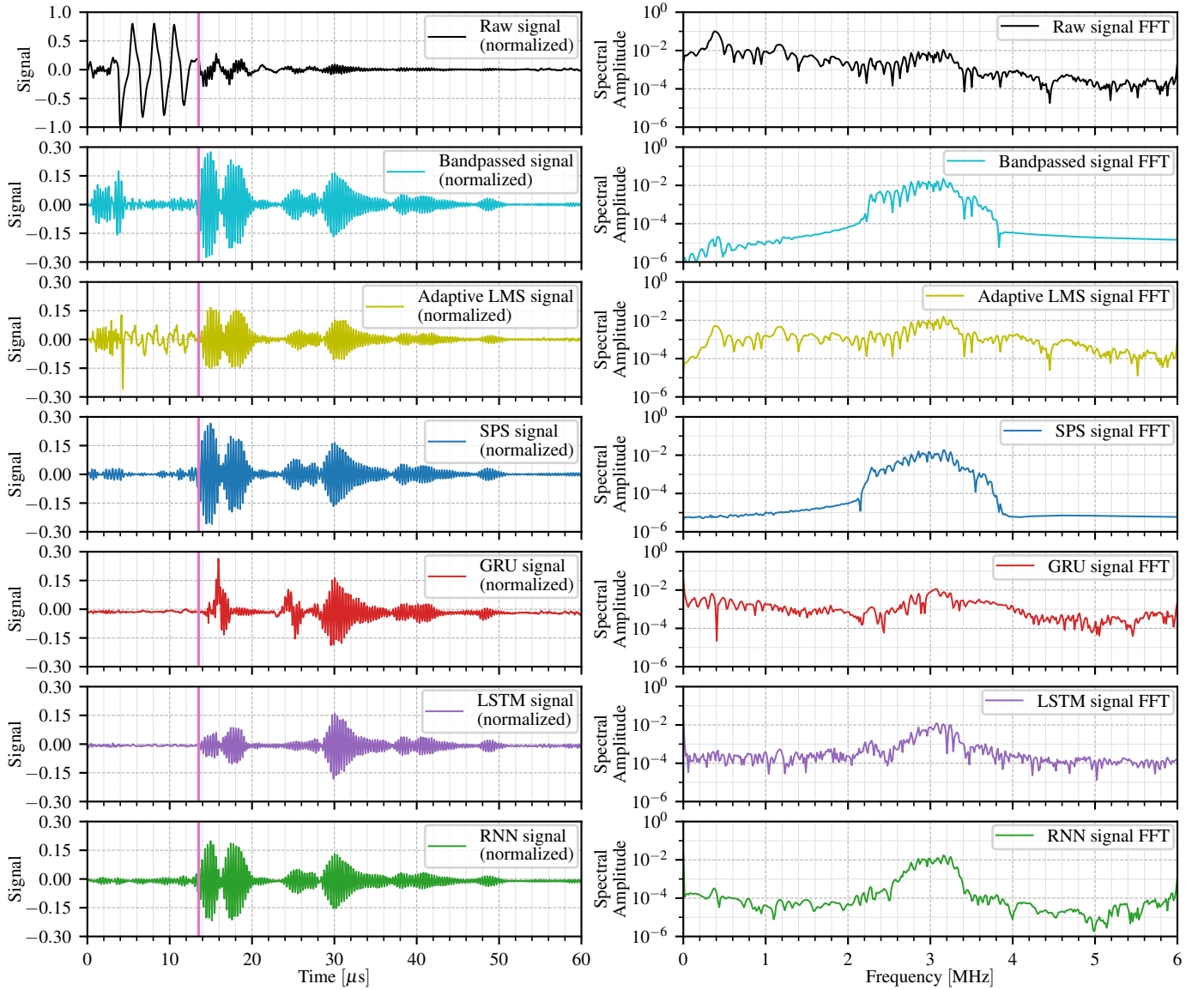


Fig. 5. Denoising results for the Bandpass filter, the Adaptive LMS filter, Spectrum Suppression, GRU, LSTM and RNN. Left: the results of the denoising methods in time domain. Vertical pink line shows the true TOF with the travel distance of 1 cm. Right: the results of the denoising methods in the frequency domain.

III. RESULTS

A. Noise Removal Qualitative Comparison

Each method's results for artifact removal from a single raw ultrasound signal at 1 cm flight distance (here and below, this refers to the distance traveled until reflection and return to the transducer) are presented on Fig. 5. At this distance, the echo is capable of traveling back and forth multiple times, without decaying enough to become immeasurable; this multi-echo phenomenon can be observed in all filtered signal plots, at approximately 13.5 μ s, 27.0 μ s and 40.5 μ s. The training dataset did not have any data with multi-echo.

The upper left plot of Fig. 5 shows the raw signal, corrupted by the ringdown artifact; the echo's amplitude is smaller than the ringdown artifact's by about a factor of 3. At a shorter

flight distance, this corruption would become more severe, thus making it even harder to extract the time of flight from the blended signal. Because both the ringdown and the echo are physically generated by the oscillating transducer, part of the ringdown's frequency spectrum is similar to the echo's, and due to the short TOF, they blend with each other in time domain as well. For this reason, the Bandpass filter still retains a lot of noise, particularly in the first 5 μ s of the waveform. The adaptive LMS filter exhibits the same behavior. The SPS method, however, despite its reliance on the frequency domain, cleaned the signal much better, with only negligible noise remaining in the first 5 μ s. The performance of deep learning methods was more varied. As Fig. 5 shows, at this short distance, GRU significantly distorted the signal, although the ringdown was removed completely. LSTM showed some

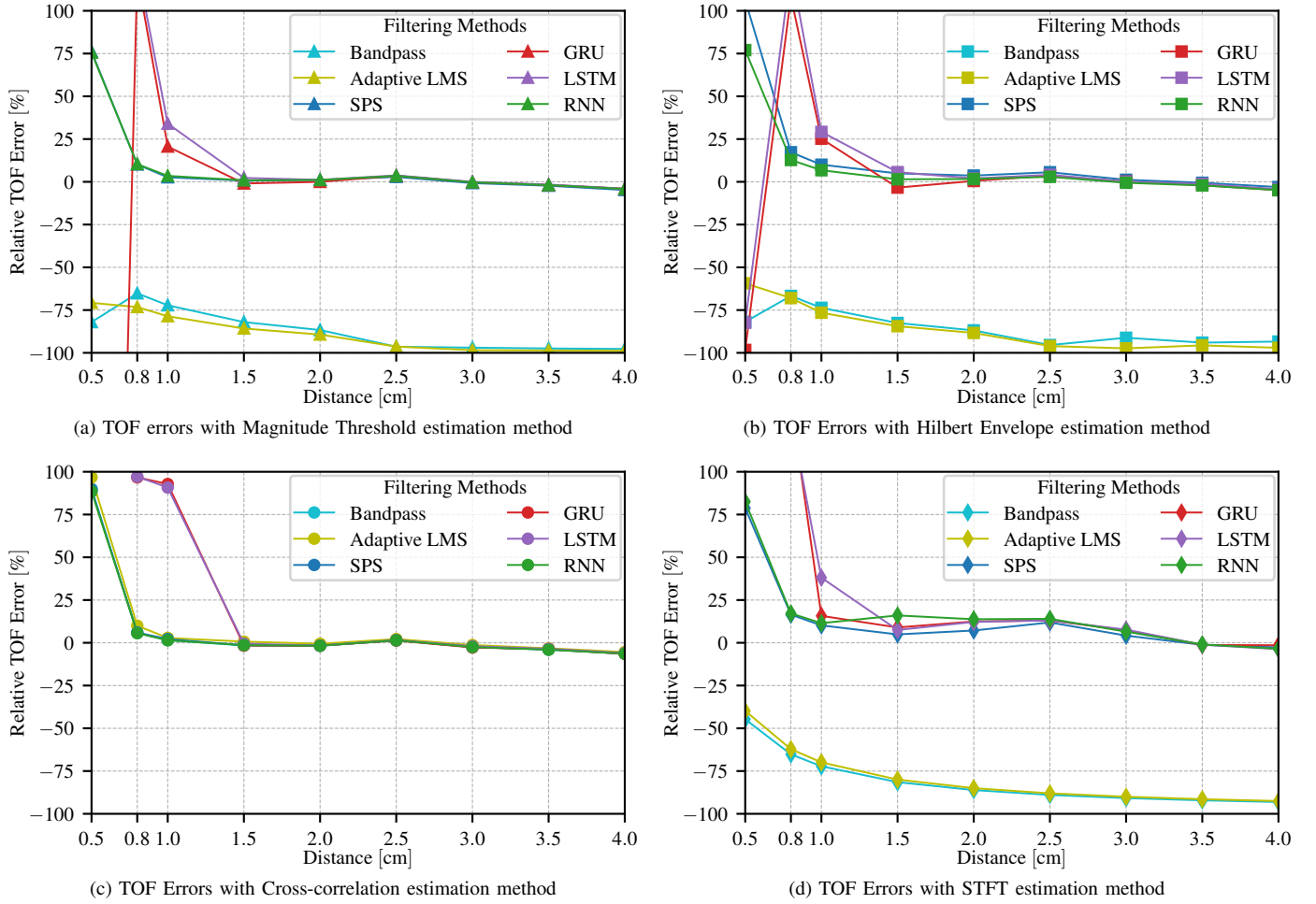


Fig. 6. TOF mean percentage errors by different filtering and TOF estimation method. Errors over 100 % indicate method failure, and are cropped out.

distortion as well, although echoes remained visually recognizable. RNN performed the best among the three deep learning methods.

B. Time of Flight Comparison

Following the denoising via the six methods discussed in Section II-A, TOF was estimated via the four methods discussed in Section II-B. We evaluate the true TOFs based on the velocity of sound in water and the known distance. Mean relative errors in TOFs, as estimated by each method pair for each distance measured, are presented on Fig. 6.

As we can see from Fig. 6, all pairs of methods fail at the distance of 0.5 cm, which indicates the system's limitation. It's observable that deep learning denoising methods LSTM and GRU with all TOF estimation methods started to fail at distances lower than 1.5 cm. Distances lower than 2 cm were never shown during the training of deep learning denoising methods, and LSTM and GRU fail to generalize the solution for short-distance echo waveforms. It may be possible to use them with additional training data on short distances. On the other hand, standard RNN showed remarkably good results

in spite of not having short-distance echo waveforms shown during the training.

The Bandpass and the Adaptive LMS filters had relative errors of more than 75 % paired with three of the four methods for TOF estimation at all distances. This is likely because as Fig. 5 illustrates, after these filters are applied, the remaining artifact is larger in amplitude than the echo, which causes threshold-based methods to fail. However, the Bandpass and the Adaptive LMS filters performed well with the CC TOF estimation method, because it does not depend on thresholding, but on signal correlation. CC method is also immune to white noise more than other methods.

Root Mean Square Error (RMSE) over all valid distances (i.e., 0.8 cm to 4.0 cm) was calculated for all 24 pairs of methods. 6 pairs demonstrated an RMSE of less than 5 %:

- 1) SPS with Magnitude Threshold: RMSE of 4.31 %.
- 2) RNN with Magnitude Threshold: RMSE of 4.39 %.
- 3) Bandpass with CC: RMSE of 3.62 %.
- 4) Adaptive LMS with CC: RMSE of 4.40 %.
- 5) SPS with CC: RMSE of 3.70 %.
- 6) RNN with CC: RMSE of 3.60 %.

These methods' RMSEs for each distance, and their stan-

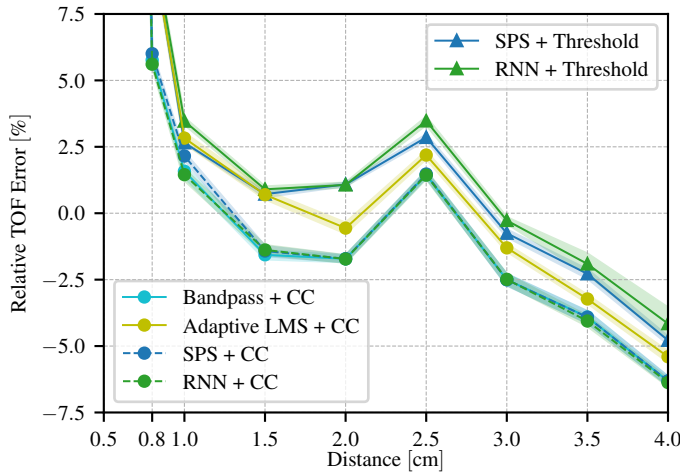


Fig. 7. Relative TOF errors for best method pairs across all distances.

standard deviations over all samples taken at each distance are illustrated on Fig. 7. Based on their RMSEs, three pairs performed particularly well: Bandpass with CC, SPS with CC and RNN with CC. RNN with CC has the lowest RMSE error. The error variances are also small, which confirms that because the noise removal algorithms used in this work do not rely on subtracting a particular ringdown signal to remove it, they are not sensitive to ringdown drift.

All 24 method pairs compared are usable in real-time, each providing a TOF result in under 15 ms. The bulk of this time is taken by the denoising stage.

IV. DISCUSSION

The RNN denoising method, combined with the CC TOF Estimation method, appear to be best suited for ringdown artifact removal, followed by TOF extraction. In literature, LSTM and GRU are typically considered superior methods to RNN due to the vanishing and exploding gradient problems [10]. At longer TOFs, with the echo and the ringdown artifact completely separated, GRU did indeed outperform all other methods. The poor performance of LSTM and GRU at shorter distances may be explained by the limitations of the training dataset, which, due to being generated by manually removing the ringdown, did not have any examples of the multi-echo, and may have lacked other features that the networks have relied on.

RNN, on the other hand, showed optimal performance at both long and short distances, with results comparable to the traditional SPS and Bandpass methods. It is conventional to compare neural network performance according to loss function values achieved on training sets, and their training time. Training time-wise, the results were indeed as expected, with RNN taking significantly longer than the other two, and GRU being the fastest. For this problem, the significant difference between the training and testing sets makes the loss function evaluated on the training set irrelevant; instead, average errors across the testing set, presented in the previous section, are more illustrative about the methods' relative performance.

V. CONCLUSIONS

Six artifact removing methods were compared for the removal of ringdown artifacts from short distance A-mode ultrasound signals in water: the Bandpass filter, the Adaptive LMS filter, Spectrum Suppression, RNN, LSTM and GRU. Their performance was evaluated based on estimating the TOFs from filtered signals. Four methods for extracting TOFs were compared: Magnitude Threshold, Envelope Peak Detection, CC and STFT. Of the 24 analyzed method pairs, the three best were the RNN, Spectrum Suppression and Bandpass artifact removing methods, with each followed by the CC TOF estimation method. RNN with CC was the best performing method pair, and we select it as the best method for future work for preprocessing the surgical grasper's data in further classification algorithms. The RMSE of RNN with CC method pair was 3.60 % across distances of 0.8 cm and above. This proposed method pair makes it possible to use a single transducer configuration on surgical grasper jaws, which, beneficially, leaves room for additional sensors.

REFERENCES

- [1] T. D. Nagy and T. Haidegger, "Recent advances in robot-assisted surgery: Soft tissue contact identification," in *Proc. SACTI*, Timisoara, Romania, May 29–31, 2019, pp. 99–106.
- [2] P. Fish, *Physics and Instrumentation of Diagnostic Medical Ultrasound*. New York, NY, USA: John Wiley & Sons, 1990.
- [3] T. E. Doyle, A. P. Butler, M. J. Salisbury, M. J. Bennett, G. M. Wagner, H. A. Al-Ghaib, and C. B. Matsen, "High-frequency ultrasonic forceps for the in vivo detection of cancer during breast-conserving surgery," *J. Med. Devices*, vol. 14, no. 3, Sep. 2020, Art. no. 031004.
- [4] X. Qu, T. Azuma, H. Lin, H. Takeuchi, K. Itani, S. Tamano, S. Takagi, and I. Sakuma, "Limb muscle sound speed estimation by ultrasound computed tomography excluding receivers in bone shadow," *Proc. SPIE*, vol. 10139, pp. 313–320, Mar. 2017.
- [5] M. Wirtz and M. Hildebrandt, "IceShuttle Tereido: An ice-penetrating robotic system to transport an exploration AUV into the ocean of Jupiter's moon Europa," in *Proc. IAC*, Guadalajara, Mexico, Sep. 26–30, 2016.
- [6] S. Wagle and H. Kato, "Real-time measurement of ultrasonic waves at bolted joints under fatigue testing," *Exp. Mech.*, vol. 51, no. 9, pp. 1559–1564, Nov. 2011.
- [7] J. C. Norton, P. R. Slawinski, H. S. Lay, J. W. Martin, B. F. Cox, G. Cummins, M. P. Y. Desmulliez, R. E. Clutton, K. L. Obstein, S. Cochran, and P. Valdastrì, "Intelligent magnetic manipulation for gastrointestinal ultrasound," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019.
- [8] C. J. Barlow, R. J. Dickinson, and R. I. Kitney, "Medical ultrasound imaging," U.S. Patent 5 601 082, Feb. 11, 1997.
- [9] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Apr. 1979.
- [10] F. Chollet, *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications, 2018.
- [11] L. Svilainis, "Review of high resolution time of flight estimation techniques for ultrasonic signals," in *Proc. NDT*, Telford, England, UK, Sep. 10–12, 2013, pp. 1–12.
- [12] J. F. Figueroa, "An ultrasonic ranging system for robot end-effector position measurement," Ph.D. dissertation, Dept. Mech. Eng., Penn. St. Univ., State College, PA, USA, 1988.
- [13] L. Jia, B. Xue, S. Chen, H. Wu, X. Yang, J. Zhai, and Z. Zeng, "A high-resolution ultrasonic ranging system using laser sensing and a cross-correlation method," *Appl. Sci.*, vol. 9, no. 7, Apr. 2019, Art. no. 1483.
- [14] Z. Lu, F. Ma, C. Yang, and M. Chang, "A novel method for estimating time of flight of ultrasound echoes through short-time Fourier transforms," *Ultrasonics*, vol. 103, Apr. 2020, Art. no. 106104.
- [15] Y. Sosnovskaya, "Ultrasound waveforms with and without ringdown artifacts," IEEE Dataport, 2022. [Online]. Available: <https://doi.org/10.21227/z6v5-mf23>