

Visio-Acoustic Data Fusion for Structural Health Monitoring Applications

Chad R. Samuelson¹, Caitrin A. Duffy-Deno², Christopher B. Whitworth³, David D.L. Mascareñas⁴, Jeffery D. Tippmann⁴, and Alessandro Cattaneo⁴

¹Dept. of Mechanical Engineering
Brigham Young University, Provo, UT 84606

²Dept. of Mechanical and Aerospace Engineering
University of California San Diego, La Jolla, CA 92093

³Dept. of Mechanical Engineering
The University of Texas at Dallas, Richardson, TX 75080

⁴Los Alamos National Laboratory
Los Alamos, NM 87545

ABSTRACT

Structural health monitoring has been an expanding discipline due to its potential to decrease maintenance and downtime costs, detect failure early, extend life spans, and fulfill the increased need of safety and security. Acoustic source identification techniques have been used for remote structural health monitoring, but the applicability of each technique has been limited by factors ranging from achievable spatial resolution to hardware costs. This paper aims at mitigating current acoustic techniques' limitations by exploring the possibility of fusing acoustic and video data. This paper focuses on combining microphone acoustic measurements with vibrational information recovered from video-based measurements. Among acoustic methods, acoustic arrays have been used for remotely detecting, localizing and characterizing acoustic sources. Acoustic-array based techniques are limited in their ability to discriminate multiple closely-spaced acoustic sources from far-field acoustic pressure signals. On the contrary, video-based techniques have shown the ability to recover full-field, high resolution mode shapes, and the associated frequencies and damping ratios with virtually no dependence with the distance from the target. The challenge with video methods, applied to acoustic source identification, is that acoustic sources may occur in the kilohertz range requiring a higher frame per second sampling rate than most low cost cameras. Acoustic measurements provide additional information content that is used to recover the correct frequency content of an acoustically radiating structure from temporally-aliased (sub-Nyquist) video measurements. Experiments are conducted to show how combining acoustic and video data relaxes the hardware requirements for acoustic source detection and localization applications.

Keywords: Structural health monitoring; Vibro-acoustics; Video processing; Signal Aliasing; Sub-Nyquist sampling

I. INTRODUCTION

1.1. Background

In this work, we focus on the development of visio-acoustic data fusion techniques to bring together the complementary advantages of acoustics and video-based structural dynamics to enable new structural characterization techniques. The presence of damage on a structure may result in the change of its vibrational characteristics [1]. Many different vibro-acoustics methods have been developed to capture a change in the vibrational response of a structure. Microphone array-based techniques have the ability to detect, localize and characterize acoustically radiating structures [2, 3]. The acoustic techniques suffer both economical and

technical limitations. Off-the-shelf microphone-arrays and the commercial license of a given acoustic technique cost several thousand dollars. In addition, the spatial resolution achievable with acoustic techniques depends on the ability to acquire near field information (i.e. evanescent waves) [4]. To obtain high-resolution spatial acoustic characterization, the microphone-array needs to be placed well within 2 wavelengths from the acoustic source [3], with important limitations on the measurement setup. Laser Doppler vibrometers have been shown to capture accurate full-field structural vibration data [5], but they are only available at the cost of ten of thousands of dollars. Complex measurements setups implementing off-axis digital holography and advanced signal post-processing techniques [6] have been developed to image vibration fields. Another method, in order to obtain full-field information, involves attaching many accelerometers on a structure. Direct acceleration measurements cost an unnecessary amount of time and hardware, furthermore they might create unwanted load effects on lightweight and lightly damped structures.

In recent years, at the Los Alamos National Laboratory (LANL), research has been undergone to use both high-frame rate [7] and sub-Nyquist [8] video measurements to identify and characterize vibrational sources. The results achieved at LANL added original contributions to the other works available in the literature that proved the possibility to recover high-spatial resolution vibrations [9, 10] and even sound [11] from video measurements.

The cost and limitations of these vibro-acoustic methods are the driving motivators for our fusion idea.

1.2. Idea

We explored the possibility to combine video-based imaging techniques and microphone-data to identify which parts in a vibrating structure are contributing to sound radiation. Although video measurements enable to analyze/extract vibrations from a scene, not all vibrations in a scene are radiating sound waves. For this reason, it is not immediate to determine which vibrations are actually responsible to generate the sound field. In this work, we used the acoustic data collected with one microphone to identify which vibrations recorded in a high-frame-rate video of the scene correspond with the actual acoustic sound being emitted from the acoustic source. When combined, high-frame-rate videos and acoustic data enabled to create a high spatial resolution acoustic map of the scene. Further, we applied our visio-acoustic fusion technique to videos and sounds recorded with a commercial smartphone. We leveraged the information collected by the smartphone's microphone to produce acoustic maps from aliased videos. Capabilities and limitations of the proposed anti-aliasing technique are analyzed in light of the results obtained using high-frame-rate camera and smartphone's videos.

II. METHOD

2.1. Process Microphone Data

2.1.1. Find Acoustic Source Frequencies

The proposed method starts by transforming the microphone acoustic data from the temporal domain into the Fourier domain. The frequencies with the largest amplitudes are the most audible frequencies in an acoustic scene. Depending on the number of acoustic sources in the scene, a different number of peak frequencies in the microphone data can be found. In the experimental setups used in this work, no more than 5 different acoustic sources were present. We will move forward with the assumption that we are only detecting the 5 maximum frequencies in our data sets, but again this number can be adjusted depending on the acoustic scene.

2.1.2. Compute aliased frequencies

Most audible acoustic sources radiate in the lower end of the kHz frequency range. Since most video data is being sampled below the kHz frequency range (less than 1000 frames per second), aliasing will occur when detecting the motion of these vibro-acoustic objects in the scene. To overcome aliasing an expensive high-frame rate video camera can be used, as done in the Blind-identification paper [7], or some research

has been done to predict at what frequency an aliased frequency actually is radiating at [8]. A typical low-cost microphone has a minimum sampling rate of 44.1 kHz. We plan to overcome the aliasing affecting sub-Nyquist sampled video data by utilizing the high-sampling rate microphone data.

This is done by finding the main peaks in the spectrum of the microphone data that correlate to the vibrating objects in the video's camera scene. Based on the frame rate of the video, the peaks aliased frequency locations can be calculated as follows:

$$A = \frac{mic_{peaks}(f)}{\frac{f_s}{2}} \quad (1)$$

where f_s is the sampling rate of the camera. $Mic_{peaks}(f)$ is the peak frequencies found in section 2.1.1. The integer portion of A , A_{int} , is the number of times that the peak frequency is folded over to end up in the Nyquist frequency range. The decimal portion of A , A_{dec} , represents the fraction of f_s (i.e. where that frequency will show up in the aliased video data). Depending on if the integer portion of A is even or odd, multiply the decimal portion by f_s or 1 minus the decimal as shown below:

If A_{int} is even:

$$mic_{alias}(f) = A_{dec} * \left(\frac{f_s}{2}\right) \quad (2)$$

If A_{int} is odd:

$$mic_{alias}(f) = (1 - A_{dec}) * \left(\frac{f_s}{2}\right) \quad (3)$$

where $mic_{alias}(f)$ is the new calculated aliased frequency.¹

2.2. Process Video Data

2.2.1. Obtain motion from video

Vibrational motion is obtained based on a similar technique discussed in the Blind Identification paper [7]. The complex steerable pyramid [12] is applied to each frame of the video to filter the video building a pyramid and then collapsing it yields

$$I(x + \delta(x, t)) = \sum_{\omega=-\infty}^{\infty} R_{\omega}(x, t) = \sum_{\omega=-\infty}^{\infty} \rho_{\omega}(x, t) e^{j2\pi\omega_0(x+\delta(x,t))} \quad (4)$$

where x is the pixel location, t is the frame number (time index), δ contains the displacement of the structure, ω_0 is the spatial frequency, and $R_{\omega}(x, t)$ is the subband representation on the spatial scale ω defined as

$$R_{\omega}(x, t) = \rho_{\omega}(x, t) e^{j2\pi\omega_0(x+\delta(x,t))} \quad (5)$$

with the local amplitude $\rho_{\omega}(x, t)$ (corresponding to edge strength) and the local phase $\psi(x, t) = 2\pi\omega_0(x + \delta(x, t)) = 2\pi\omega_0x + 2\pi\omega_0\delta(x, t)$ that encodes the temporal vibration motion of the structure.

The specific subband, ω , is then chosen in the direction of motion most relevant for the specific recorded scene. The temporal mean $2\pi\omega_0x$ is subtracted from $\psi(x, t)$ to obtain only the displacement of the object from the reference phase, $2\pi\omega_0\delta(x, t)$. Lastly, combining $\delta(x, t)$ for every frame builds a 3-D Phase Matrix containing the vibrational motion of the structure.

2.2.2. Filter Video Data based on aliased frequencies from microphone data

In order to properly segment the vibro-acoustic scene, the 3-D Phase Matrix must be filtered at each aliased frequency found using the procedure described in *section 2.1.2*. The method proposed in our work applies a temporal bandpass filter [10] to the 3-D Phase Matrix at each found aliased frequency. Essentially a 4-D matrix is made where the first and second dimensions describe the pixel location, x , the third dimension contains all frames, and the fourth dimension index refers to the frequency filtered. In the case of finding 5

¹Note: The process is repeated for all the main frequency peaks identified in the spectrum.

frequency peaks in the spectrum of the microphone data, the size of the fourth dimension is 5.

2.3. Segment video data

2.3.1. Use Level Set Method

To segment anti-aliased video data, the level set method [13] can be employed to locate which pixels in the video have a certain power value. Power values can be visualized via a 3-D surface plot containing the locations of the pixels of the recorded scene on the 2-D plane and the vertical axis being the values of power for a given frequency band. To implement the level set method, the temporally filtered Phase Matrix is used to find the energy spectral density [14] for each pixel via

$$S(\omega) = |Y(\omega)|^2 \quad (6)$$

where ω represents frequency in radians per interval of sampling and Y is the amplitude of one pixel from the temporally filtered Phase Matrix in the Fourier domain. Parseval's Theorem [14], as shown below

$$\sum_{t=-\infty}^{\infty} |y(t)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) d\omega \quad (7)$$

where $y(t)$ is the signal in the time domain, can be applied to each pixel's displacement to find their average power across a frequency band. These power values would then be plotted in 3-D to show the magnitude of power with peaks indicating where the greatest vibration is occurring in the recorded scene for a specific frequency range.

2.3.2. Apply Segmentation Based on Power Values

The generated power spectrum plots can then be made into 2-D images/heat maps with the colors corresponding to the varying values of power in comparison to the lowest and maximum power values that are represented by blue and yellow colors, respectively. Segmentation of the peaks in the power data can then be applied based on these colors. This process includes using the k-means clustering method to find pixels that are similar in color to form a "mask". Edge detection is used on the mask to create an outline of the non-blue color regions. These outlines are overlapped with the power heat map to isolate the peak power areas which is then overlapped with an image of the recorded scene from the video to completely visually represent the results.

III. TESTING PERFORMED

There were two different experimental set-ups in testing the visio-acoustic technique proposed in this paper. Our first experiment involved multiple tuning forks (i.e. 128 Hz, 512 Hz, and 2048 Hz) vibrating at the same time recorded by the hi-speed camera and the smartphone, but in different tests. The second experiment had three tests which were recordings made by both the hi-speed camera (Edgertronic SC2+) and the smartphone (Samsung S9) at the same time of a vibrating 512 Hz tuning fork.

3.1 Experimental Set-Ups for the Recordings of Multiple Tuning Forks

3.1.1 Experimental Set-up for High-Frame Rate Camera

The first set-up was for using the hi-frame rate camera and a prepolarized, condenser microphone PCB 130A23 (see Figure 1). The camera was spaced about 1.8 meters from the tuning forks. The microphone was spaced within the near field, 2λ (λ represents acoustic wavelength), of the 2048 Hz fork, because this fork was the highest frequency fork and thus had the smallest wavelength (smaller near field). It follows that the microphone was within the near field of all tuning forks.

To avoid aliasing with the initial high-frame rate video, the video was sampled at 4940 frames per second (FPS). This avoided aliasing because double 2048 Hz (frequency of the highest frequency tuning fork) was

still less than 4940 Hz. This video data was processed by the above method just skipping sections 2.1.2 and changing section 2.2.2 to filtering based on original found frequencies from the microphone data. Results are shown in Figure 6.

The 4940 FPS video data was then down sampled by a down sampling factor of 6 reducing the frames per second from 4940 to 824 FPS. The downsampling caused the 512 and 2048 Hz acoustic information to be aliased mimicking a low-frame rate video. The above method was then implemented on the down sampled video data to produce the results shown in Figures 7 and 8.



Figure 1: Experimental Set-up for Hi-Frame Rate Camera. 4940 FPS used and Down Sampled to 823 FPS. Edgertronic SC2+ camera was used along with center microphone in microphone array. Record data based on vibration of 3 tuning forks with resonant frequencies (128 Hz - left, 512 Hz - center, 2048 Hz - right)

3.1.2 Experimental Set-Up for Smart Phone (Low-Frame Rate Camera and High Sampling Rate Microphone)

The second set-up was designed for using a low-frame rate smart phone camera and the smart phone's accompanying hi-sampling rate microphone (see Figures 2, 3). Only two (128 Hz and 512 Hz) of the three tuning forks were recorded in this experiment because we found from the first experiment that the 2048 Hz tuning fork wasn't vibrating enough for the camera to recognize its motion (more explained in the Results section). Again, when recording the tuning fork data with the microphone, the smart phone was placed in the near field of the 512 Hz tuning fork (the highest frequency tuning fork) (see Figure 3). The visio-acoustic technique proposed in this work was then implemented on the smart phone data. The smart phone experiment was repeated at 60, 120, 240 FPS. Results are shown in Figures 9, 10, and 11.



Figure 2: Video Experimental Set-up for Smart Phone (Low-Frame Rate Camera) with two tuning forks (512 Hz - left, 128 Hz - right)



Figure 3: Smart Phone Microphone Experimental Set-Up with Two Tuning Forks (512 Hz - left, 128 Hz - right)

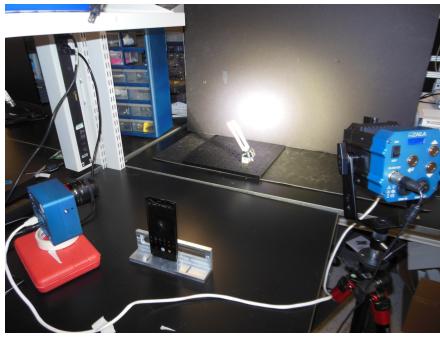


Figure 4: Experimental Set-Up for the Recording of a 512 Hz Tuning Fork with Both Camera Recording Simultaneously

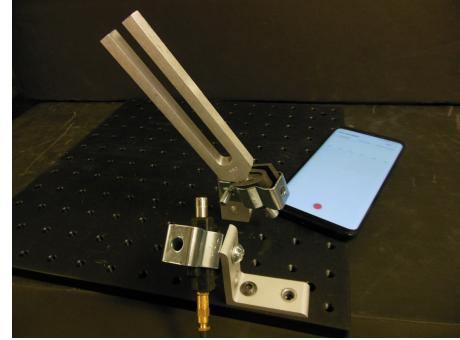


Figure 5: Set-Up of Simultaneously Recording Microphones

3.2. Experimental Set-Up for Direct Comparison of Simultaneous Recordings

A set of tests were performed with both the smartphone and the hi-speed camera recording a vibrating 512 Hz tuning fork with the settings of each camera being varied (see Figure 4). The distances that the cameras were placed from the tuning fork were kept constant for all tests and was found by matching the areas covered in the physical scene for each pixel for each camera. The number of pixels in the frame for the hi-speed camera were maximized, having a horizontal count of 1280 and a vertical count of 864 pixels. The smartphone was set to have a resolution of 4k where the number of horizontal pixels is 2160 and the number of vertical pixels is 3840.

From these calculations, it was found that with a fixed distance for the smartphone of 525 mm, the distance that the hi-speed camera should be and was placed was 624 mm. These distances were maintained the same for all the tests that were performed.

External lighting was used and was focused on the tuning fork. The tuning fork was hit with the rubber end of a hammer so as to decrease the magnitude of other modal frequencies besides 512 Hz that radiate. For recording the sound, both the smartphone microphone and the PCB microphone corresponding to the hi-speed camera were placed close to the tuning fork to assure that they were in the near field (see Figure 5).

3.2.1 Experimental Set-Up for First Test Performed

For the first test, both cameras had frame rates of 60 fps and shutter speeds of 1/24000 sec. The smartphone had an ISO value of 500 and the hi-speed camera was set to an ISO value of 2500.

3.2.2 Experimental Set-Up for Second Test Performed

The second test had all the same parameters as the first test except that the hi-speed camera had a frame rate of 2400 fps. The purpose of setting the frame rate on the hi-speed camera to 2400 fps was to compare segmentation results with the smartphone which was set at 60 fps. To process the video data of the hi-speed camera, the data was downsampled by a factor of 40 to equal the sampling rate of the smartphone.

3.2.2 Experimental Set-Up for Third Test Performed

The third test that was performed had both cameras' frame rates set to 60 fps and their shutter speeds set to 1/24000 sec while the ISO value on the smartphone was set to its maximum value of 800 and the hi-speed camera's ISO value was set to 10000. This was to see whether varying the amount of light that is captured while filming affects the motion detection by each pixel.

IV. RESULTS

4.1. Experimental Results for the Recordings of Multiple Vibrating Tuning Forks

4.1.1. Experimental Results for High-Frame Rate Camera

We were able to obtain segmentation results by leveraging the data obtained from the single microphone to identify the location and intensities of vibrations contained in the scene (see Figure 6). For the 4940 frames-per-second video, we located peaks in vibrations at 128 Hz and 512 Hz in each power spectrum that overlapped the tuning forks when segmented. There is a noticeable lack of peaks in the power spectrum for 2048 Hz and thus segmentation was not possible for that frequency. Similarly, for the downsampled and anti-aliased video, we were able to obtain similar results in that segmentation occurred as predicted for 128 Hz and 512 Hz but not for 2048 Hz (see Figure 8).

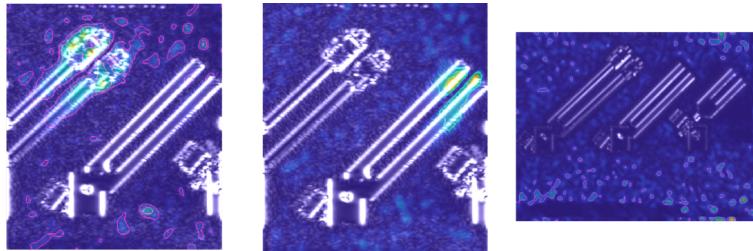


Figure 6: 4940 FPS Hi-Frame Rate Camera Segmentation Results (128 Hz - left, 512 Hz - center, 2048 Hz - right)

4.1.2. Experimental Results for Smart Phone (Low-Frame Rate Camera and High Sampling Rate Microphone)

When using the smart phone at 60 frames per second, we were able to segment at 128 Hz. However, segmenting did not work for the 512 Hz (see Figure 9). We believe this may be due to the minimal amount of movement captured by the video which could be caused by the smartphone camera possibly being set to a low shutter speed and/or resolution. The video taken at 120 frames per second (Figure 10) produced similar results as the 60 frames per second, however there was more noise and less noticeable peaks in the power spectrum. We believe the unsatisfactory outcome obtained at 120 frames per second could have been a result of not hitting the 512 Hz tuning fork with enough force. Noise appears at 60 and 120 frames per second for the 128 Hz segmentation image due to the corresponding aliased frequency being between 7 Hz and 8 Hz. This due to noise generally being more prevalent at lower frequencies. The 240 Hz filter was able to give a more detailed segmentation of the 128 Hz tuning fork and displayed less noise. The better results at 240 frames per second (Figure 11) are due to its corresponding aliased frequency being at 112.2 Hz where typically less noise occurs at higher frequencies.

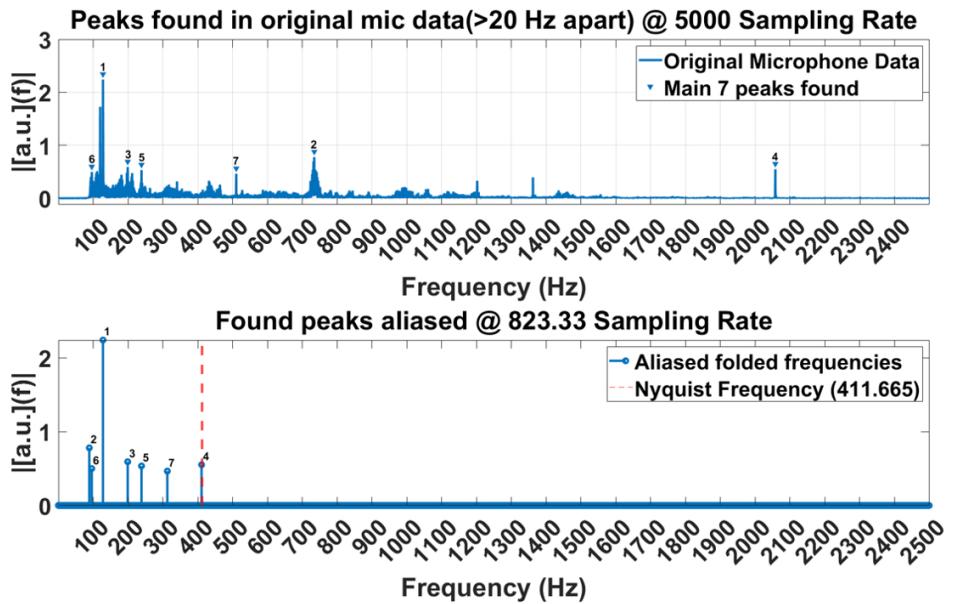


Figure 7: 7 Peaks found in Microphone Data and then Aliased Frequencies computed based on 824 FPS Down Sample Camera Data. Peaks 1, 7, and 4 correspond to 128, 512, and 2048 Hz forks.

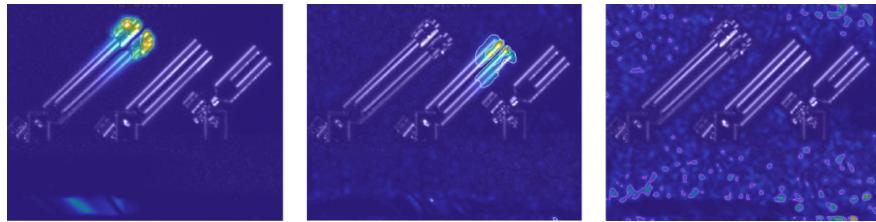


Figure 8: 824 FPS Down Sampled Hi-Frame Rate Camera Segmentation Results (128 Hz - left, 512 Hz - center, 2048 Hz - right)

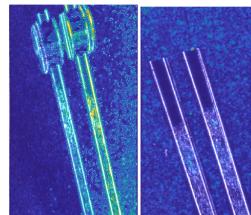


Figure 9: 60 FPS Smart Phone Segmentation Results (128 Hz - left, 512 Hz - right)

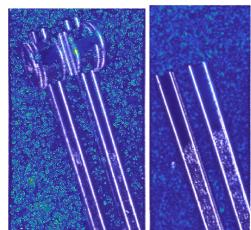


Figure 10: 120 FPS Smart Phone Segmentation Results (128 Hz - left, 512 Hz - right)

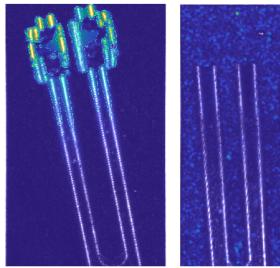


Figure 11: 240 FPS Smart Phone Segmentation Results (128 Hz - left, 512 Hz - right)

4.2. Experimental Results for Direct Comparison Experiments

4.2.1 Test 1

The results show that segmentation of the peaks in power at 512 Hz occurs for the videos captured by both the smartphone and the hi-speed camera. As seen in Figures 12 and 13, the areas that are segmented are not consecutive and are minute in size. Most likely, the scattered segmentation results are due to the low frame rate that the videos were recorded at. The outcome of Test 1 does indicate though that at nearly identical camera settings, the low performing smartphone can produce results that are similar to the hi-speed camera when it records at a low frame rate, in this case 60 fps. The difference in locations of the segmented areas on the tuning fork is most likely due to the fact that the cameras were directed towards the tuning fork at varying angles.

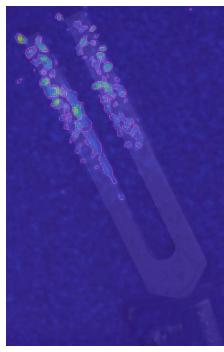


Figure 12: Test 1 Smartphone Camera Segmentation Results at 512 Hz (60 FPS, shutter speeds of 1/24000 sec, ISO 500)

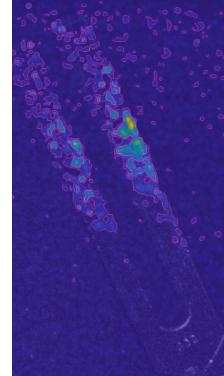


Figure 13: Test 1 Hi-Speed Camera Segmentation Results at 512 Hz (60 FPS, shutter speeds of 1/24000 sec, ISO 2500)

4.2.2 Test 2

For the results of the second test, segmentation of detected peaks in vibrations at 512 Hz was possible for both cameras, but there is a noticeable difference. The hi-speed camera video has segmentation of distinct, continuous regions of peaks in power values while the smartphone has sparse peaks. These results can be seen in Figures 14 and 15, respectively.

4.2.3 Test 3

The results show that for the smartphone, the segmentation at 512 Hz is worse when the video is recorded with the maximum ISO value possible than the segmentation of the video recorded in test two that had an ISO value of 500. The results are shown in Figures 16 and 17 for both cameras.

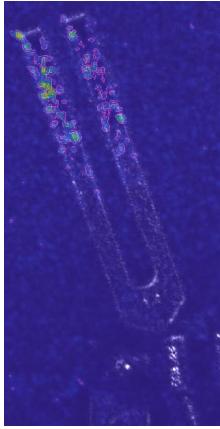


Figure 14: Test 2 Smartphone Camera Segmentation Results at 512 Hz (60 FPS, shutter speeds of 1/24000 sec, ISO 500)

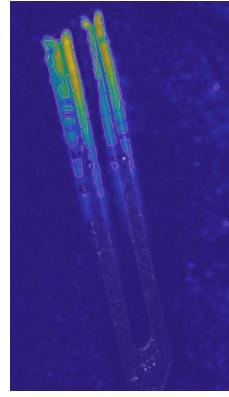


Figure 15: Test 2 Hi-Speed Camera Segmentation Results at 512 Hz (2400 FPS downsampled by a factor of 40, shutter speeds of 1/24000 sec, ISO 2500)

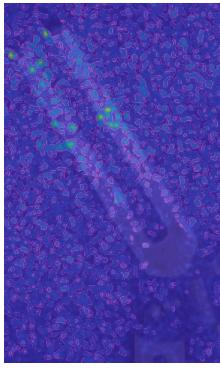


Figure 16: Test 3 Smartphone Camera Segmentation Results at 512 Hz (60 FPS, shutter speeds of 1/24000 sec, ISO 800)

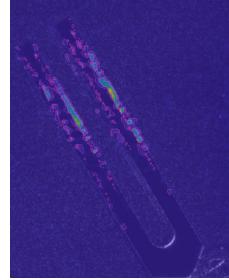


Figure 17: Test 3 Hi-Speed Camera Segmentation Results at 512 Hz (60 FPS, shutter speeds of 1/24000 sec, ISO 10000)

V. CONCLUSION

This paper explored a new vibro-acoustic structural health monitoring technique that is unsupervised, less costly and uses more accessible equipment than conventional vibro-acoustic techniques. From the results, we can conclude that our anti-aliasing technique was successful in segmenting the predicted areas of the scene at each frequency. This conclusion is based on comparing the segmented images from the original, high speed video using the hi-speed camera to the segmented images from the downsampled video as the 128 Hz and 512 Hz tuning forks were segmented for both video types. Being able to segment with the smart phone videos shows that a smart phone can indeed be used as a vibro-acoustic diagnostic tool and that a video with frame rates starting as low as 60 frames per second can be used. Multiple conclusions were made for both the high-speed camera and the smart phone which include them both being limited by the degree of motion that is captured and that using a higher frame rate improves spatial resolution. It was also concluded that the created algorithm which generated the segmented images is limited in its segmentation ability for the smartphone as the results for each camera of the second experiment's test two greatly vary.

VI. FUTURE WORK

Narrow band acoustic source signals are signals with only one frequency, while broadband signals are signals where multiple frequencies are embedded into one acoustic source. The anti-aliasing technique proposed in our work was developed for narrow band signals (just finding natural frequencies of tuning forks). Future work will be done in expanding the method to broad band signals. A microphone array can also be used,

instead of one microphone, to apply beamforming when recording the microphone data to reduce noise. Suppressing aliasing could also be improved using coded exposure. By randomizing the shutter speed on the camera, coded exposure effectively changes the windowing function applied to the recorded video. Coded exposure has found to reduce blurring in video data [15], and could help produce more detailed results with less noise in the vibro-acoustic segmented scene.

VII. ACKNOWLEDGMENTS

The authors are grateful for the support of Los Alamos National Laboratory and the Los Alamos Dynamics Summer School program. We thank Los Alamos National Laboratory for the fellowship funding provided for doing this research. We would also like to acknowledge Yongchao Yang and Wesley Scott for their previous work done in the area of motion identification and image processing here at Los Alamos National Laboratory along with David Mascareñas.

References

- [1] Charles R. Farrar, Scott W. Doebling, and David A. Nix. Vibrationbased structural damage identification. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 359(1778):131–149, 2001.
- [2] Alfredo Cigada, Francesco Ripamonti, and Marcello Vanali. The delay sum algorithm applied to microphone array measurements: Numerical analysis and experimental validation. *Mechanical Systems and Signal Processing*, 21(6):2645–2664, 2007.
- [3] W. A. Veronesi and J. D. Maynard. Nearfield acoustic holography (nah) ii. holographic reconstruction algorithms and computer implementation. *The Journal of the Acoustical Society of America*, 81(5):1307–1322, 1987.
- [4] Earl G. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Elsevier Science, 1999.
- [5] Hani H Nassif, Mayrai Gindy, and Joe Davis. Comparison of laser doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. *NDT and E International*, 38(3):213–218, 2005.
- [6] Sudheesh K. Rajput, Osamu Matoba, and Yasuhiro Awatsuji. Characteristics of vibration frequency measurement based on sound field imaging by digital holography. *OSA Continuum*, 1(1):200, 2018.
- [7] Yongchao Yang, Charles Dorn, Tyler Mancini, Zachary Talken, Garrett Kenyon, Charles Farrar, and David Mascareñas. Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. *Mechanical Systems and Signal Processing*, 85(C):567–590, 2017.
- [8] Yongchao Yang, Charles Dorn, Tyler Mancini, Zachary Talken, Satish Nagarajaiah, Garrett Kenyon, Charles Farrar, and David Mascareñas. Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-nyquist), video measurements. *Journal of Sound and Vibration*, 390(C):232–256, 2017.
- [9] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (TOG)*, 31(4):1–8, 2012.
- [10] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William Freeman. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013.
- [11] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham Mysore, Frédo Durand, and William Freeman. The visual microphone: passive recovery of sound from video. *ACM Transactions on Graphics (TOG)*, 33(4):1–10, 2014.
- [12] Eero P. Simoncelli and William T. Freeman. Steerable pyramid: a flexible architecture for multi-scale derivative computation. In *IEEE International Conference on Image Processing*, volume 3, pages 444–447. IEEE, 1995.

- [13] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.
- [14] Petre Stoica. *Spectral analysis of signals*. Pearson Prentice Hall, Upper Saddle River, N.J., 2005.
- [15] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Transactions on Graphics (TOG)*, 25(3):795–804, 2006.