A NOVEL APPROACH TO RECORD SOUND

Aaswad Anil Satpute[1], Sheetal Mangesh Pandrekar[2], Imdad A. Rizvi[3]

[1]*Junior Research Scientist, Computer Science, Hansa Embedded Systems Academy, Maharashtra, India*
[2]*Junior Research Scientist, Computer Science, Hansa Embedded Systems Academy, Maharashtra, India*
[3]*Associate Professor, Department of Electronics and Telecommunications Engineering, Terna Engineering College, University of Mumbai, Maharashtra, India*

Abstract

Analog sound recording is done with the help of a microphone when the diaphragm vibrates due to the changes in atmospheric pressure which are recorded as a graphic representation of the sound waves on a medium such as magnetic tape. Digital recording converts the analog sound signal picked up by the microphone to a digital form by a process of digitization. Both Analog and Digital Sound recording techniques have their pros and cons. However, the common disadvantage that both these techniques suffer from is the loss in the clarity of the signal due to distortion and attenuation of signal during propagation from the source of sound to the diaphragm of the microphone. This paper presents a novel approach that will get rid of the problems like signal attenuation and distortion while recording audio signals The approach is based on mapping the vibrations of the source with a High Speed camera by using Computer Vision techniques like Object Tracking. We prove that the approach is noise-resistant and can record a signal by maintaining a high quality, which is much better than the quality of the signal recorded by a microphone. Moreover, the approach can take sound recording and reproduction to a new level by making it possible to select a source when there are multiple sources producing audio signals simultaneously. A technology based on the proposed approach will particularly benefit researchers interested in recording the minute details of the audio signal of any frequency, without being restrained by the limitations of the recording system or the distance of the source of sound.

Keywords: *Acoustic Waves, Audio Recording, Computer Vision, Signal Reconstruction, Vibration Measurement*

----------***----------

1. INTRODUCTION

High Speed Imaging systems are used to measure vibrations, as described in [1, 2]. Displacements, as small as 0.18 mm, have been measured in [3] by using a high speed imaging system. In recent years, the technology of cameras has not only delivered an increase in the frame rate but also substantial improvements in the image quality, sensitivity, and lens performance. As established by Moore's Law, described in [4], the technology will continue to improve at an exponential rate. In 2011, researchers at MIT Labs brought about a major breakthrough in the field of high speed imaging. They invented a phenomenal camera which is capable of capturing a trillion frames per second. They named this phenomenon as Femtophotography, described in [5]. Motivated by this breakthrough and ability of high speed imaging to accurately map vibrations of an object, this paper attempts to explore the capabilities of high speed imaging in a novel technique for recording audio signals.

A vibrating object is a source of sound. It sets the molecules of the medium into vibrations and travels through the medium as a pressure wave. During its propagation, it is attenuated by the medium and distorted by other sound signals in the surroundings. Thus, with a microphone, it is difficult to record a clear and accurate sound wave. This paper introduces an approach that does not rely on the propagation of the signal from its source to the recording system. The approach aims at mapping how the source vibrates and synthesizing sound from its vibrations. The approach requires the use of a high speed camera to view the vibrations of the source.

In order to develop and test this approach for various cases, the simulations of tuning fork vibrating at different frequencies were generated. Simulations were generated for tuning fork vibrating at frequencies 98 Hz, 123.47 Hz, 174.61 Hz, 220 Hz, 349.23 Hz recorded with camera of 1000, 1750, 3350, 4250, 5000 fps respectively. These frame rates were chosen to ensure that the frame rate is not a direct multiple of the frequency of the vibrating tuning fork. In this paper, the readings and analysis of the 98 Hz tuning fork has been presented. The sound file generated after mapping the vibrations is compared with a 98 Hz sound sampled at 44.1 kHz (called reference file in this paper), the standard sampling rate of digital audio. The reference file depicts the sound that would be recorded by a microphone with sampling frequency 44.1 kHz. This reference file has been generated using Audacity software [6] instead of being recorded with a microphone. The reason for this is that to record an absolutely accurate audio signal with a microphone would require a completely noise free setup. In order to make useful consensus about the result, the sound file generated by mapping the vibrations is compared with the reference file. The reference file depicts the most accurate sound that can be recorded by a microphone with a sampling rate of 44.1 kHz in a completely noise free setup.

For periodic waveforms, including non-sinusoidal waveforms, the Fourier theorem, described in [7], states that "A periodic function which is reasonably continuous may be

expressed as the sum of a series of sine or cosine terms (called the Fourier series)." Thus, any periodic vibrations can be described in terms of sinusoidal waves and the sound may be synthesized. The approach may be extended to complex waveforms by using the Fourier Theorem if the results with simple waveforms are found to be beneficial.

Section 2 describes the experiment that was conducted by considering a case of a tuning fork vibrating at 98 Hz. A 1000 fps (frames per second) video clip of the tuning fork was processed, one frame at a time, to map the vibrations of the tuning fork. This mapped vibration was then used to write an audio file. Techniques that may be applied to improve the accuracy while recording the signal are described in section 2. The potential advantages of employing high speed imaging system for recording audio signal are discussed in section 5. Section 7 describes the ways in which the enormous potential of Femtophotography [5], can be employed to the proposed approach.

2. THE EXPERIMENT

A source of sound generally vibrates rapidly with small amplitude. In order to measure the amplitude of vibrations accurately, the camera has to be positioned such that the axis of the lens is perpendicular to the plane of vibration of the tuning fork.

For presenting the approach, a vibrating tuning fork of 98 Hz has been considered as the source of sound. The experiment is performed on a video clip of the vibrating tuning fork. The duration of the video is 1s and it comprises of 1000 frames. Thus, the video clip represents video recorded by a 1000 fps camera. The code for reading the frames of the video to map the vibrations and generating an audio file is written in .NET using OpenCV Libraries.

2.1 Preprocess on the Video

The video needs to be pre-processed in order to get rid of the irrelevant aspects. This is an important procedure as it will improve the efficiency of the remaining process. The video is preprocessed with the steps described in this section.

2.1.1 Conversion to Greyscale

As the process does not require color images, each frame in the video is converted to a gray scale image.

2.1.2 Conversion to Binary Image

The images are converted to binary image by using a suitable threshold. The suitable threshold is determined by using the Otsu Method which has been described in [8].

After obtaining the threshold t, the images are converted to binary by using,

$$F(x, y) = 1, \text{ when } F(x, y) \geq t \text{ , else, } F(x, y) = 0 .$$

2.1.3 Edge Detection

The edges of the object in the scene are determined in each frame of the video. The Canny Edge Detection Algorithm, described in [9] is used with the Scharr Edge Detection operator. This operator gives more enhanced edges than other gradient operators. The matrices for obtaining the horizontal and vertical gradient using Scharr Operator are given in [10].

This pre-processing step helps to obtain the points of interest for efficiently mapping the vibrations.

2.2 Reporting the Displacements into the File

The displacement of a point on the tuning fork is measured at in each frame of the camera. This displacement is measured in pixels. The instances at which this displacement is recorded are referred to as the timestamps. The time interval between consecutive timestamps must be as small as possible for greater accuracy in the measured amplitude.

2.3 Reporting the Displacements into the File

To ensure that the precision of the computing hardware does not affect the value of displacement, the readings have to be normalized. The normalization is performed in two steps:

- Subtracting the value of minimum displacement from all the readings
- Dividing each reading by maximum displacement

A part of the file containing the normalized readings of a source vibrating at 98 Hz is shown in Table 1. It shows the readings obtained for one vibration. As the source considered has a constant frequency, these reading are repeated for all the vibrations throughout the file. The file contains 1000 timestamps and the displacement at each timestamp.

2.4 Determining Frequencies of the Vibrations

The records in the file are read sequentially to determine the local maxima and the local minima of displacement. The value of displacement before which the displacements are in increasing order and after which the displacements begin to decrease, marks a change from increase to decrease in displacement. This value of displacement is the local maxima. Similarly, the value of displacement before which the displacements are in decreasing order and after which the displacements begin to increase, marks a change from decrease to increase in displacement. This value of displacement is the local minima.

From Table 1, we have local maximum at timestamp 9 (t_9) and local minima at t_{14}. As these readings are repeated, the next local maximum is obtained at t_{19}. The difference between local maximum and its following local minimum gives the amplitude of vibration. The time required by the source to cover a distance equal to its amplitude of vibration is equal to half times the time period of vibration. Thus, the difference between the timestamps of local maxima and local minima of displacement is equal to half times the time period

of vibration. If the vibrations of the source are not constant, then time period between every local maxima and local minima needs to be calculated.

Let t_{max} and t_{min} denote the timestamp at which local maximum occurs and the consecutive local minimum occurs, respectively. Let T denote the time period of the waveform.

Thus,

$$|t_{\max} - t_{\min}| = T/2 \quad (1)$$

Thus the frequency (f) of the vibration can be obtained as,

$$f = \frac{1}{2 \times |t_{\max} - t_{\min}|} \quad (2)$$

Finally, the frequency of the sound wave (f_s) is obtained as,

$$f_s = f \times fps \quad (3)$$

2.5 Preprocess for the Video

The normalized readings with frequencies at each timestamp are used to generate a suitable audio file. Due to support for high sampling rate, wave file format has been used. The final audio file is generated using wave file format at 1 MHz sampling rate.

3. TECHNIQUES TO IMPROVE ACCURACY

When the frequency of vibration is very high and a suitable camera is not available, there are certain procedures that may be applied to improve the waveform being recorded.

3.1 Interpolation

Interpolation techniques, from [11], can be applied to determine the intermediate displacements of the vibrating source.

3.2 Sub-Pixel Accuracy

If the vibrations are very small and the amplitude accounts for only a few pixels then it may be needed to determine the displacements at a sub-pixel level, as in [12]. With Sub-Pixel using Phase-Only Correlation (POC), it is possible to compute with up to $1/100^{th}$ pixel accuracy, as shown in [13].

4. EMPERICAL RESULTS

The wave file generated by experiment was compared with a reference file of 98 Hz audio signal. The comparison is done on GoldWave audio editing software, [14]. The comparison as obtained on GoldWave is shown in the Fig. 2 (a) and (b). Both the files are viewed at the same scale (2 units = 0.0002 on Time axis, and 2 units = 0.01 on Amplitude axis). It is observed that the reference file shows discontinuities whereas the one obtained by the experiment does not show any discontinuities at this scale. Fig. 2 (c) is the screenshot of the file obtained by the experiment, viewed at a scale of 1 unit = 0.000005 on Time axis, and 2 units = 0.0002 on Amplitude axis. The file begins to show noticeable discontinuities only at such a high magnification.

The average result of the comparison of the five sound files, one of each tuning fork, with Praat software, [15], is shown in the Table 2. The comparison of various characteristics of sound confirms that the experimentally generated waveforms of the various frequencies are better than the reference files, respectively.

The experiment was performed on a Windows 8 64-bit platform with 2.30 GHz processor and 8 GB RAM. The total time taken for preprocessing the 1000 frames, each 720 576 pixels, and reporting the displacements into a file, as described in section 2.1 and 2.2, is approximately, 1020159 ms. The time taken to determine the frequencies at each timestamp, as shown in section 2.4, is 0.0312137 ms. The time taken to generate the audio file, as described in section 2.5, is 0.2656429 ms.

Table -1: Part of the Normalized File (of Frequency 98Hz) with Frequencies determined

Timestamps	Displacement	Frequencies (Hz)
6	0.5	98
7	0.810526316	98
8	1	98
9	1	98
10	0.810526316	98
11	0.5	98
12	0.189473684	98
13	0	98
14	0	98
15	0.189473684	98

Table -2: Part of the Normalized File (of Frequency 98Hz) with Frequencies determined

File Char-acteristic	Empirically Generated File	Reference *Wave* File	Reference *MP3* File
Standard deviation of pitch	9×10^{-14} Hz	4×10^{-6} Hz	4×10^{-4} Hz
Jitter (local)	5×10^{-14} %	3×10^{-10} %	10^{-3} %
Jitter (local, absolute)	5×10^{-18} s	3×10^{-14} s	9.8×10^{-8} s
Jitter (rap)	3×10^{-14} %	2×10^{-10} %	6×10^{-4} %
Jitter (ppq5)	6×10^{-14} %	3×10^{-10} %	6×10^{-4} %
Jitter (ddp)	9×10^{-14} %	6×10^{-10} %	2×10^{-3} %

5. DISCUSSION

While recording an audio signal, background noise gets added to the desired signal when it propagates from the source to the recording system. This paper proposes an approach that does not rely on the propagated signal for recording the sound. Thus, it makes it possible to prevent the noise from being recorded. Apart from this, a number of potential strengths of the technique are highlighted in this section.

5.1 Recording Sound of Any Frequency

By Nyquist theorem, described in [16], a microphone with a sampling rate of 44.1 kHz can record a sound at most 22.05 kHz. The frequency response of the microphone is limited by the physical characteristics of the components that make up the microphone. Thus, there is a limit to the frequency of ultrasound that can be recorded by the microphone. The proposed approach may be able to record any frequency of ultrasound depending on the frame rate of the camera. With Femtophotography [5], it is possible to even record a photon of light moving. This means that it is possible to capture a vibrating object of 500 GHz. Thus, the approach can make it possible to record any ultrasound.

5.2 Elimination of Noise

Presently, active noise control, as described in [17, 18], is a method widely used for reducing unwanted sound in the environment. For Active noise control, the system needs to generate a secondary signal of nearly equal amplitude as the noise signal to interfere destructively with the noise, thereby cancelling the noise signal. In a complex field with many sources of sound, a number of secondary signals will need to be generated to cancel other signals in order to record only a particular sound signal.

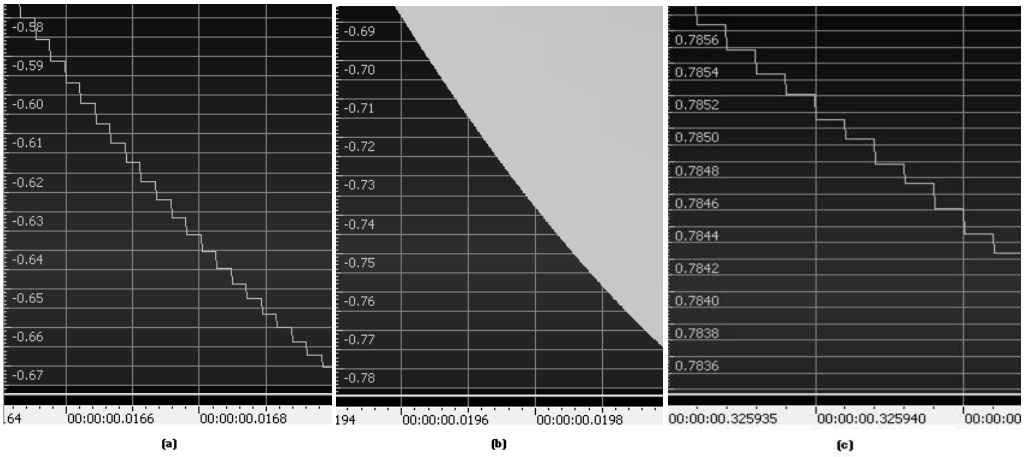

Fig -1: Comparison of the experimentally obtained wave file with a reference wave file of 98 Hz sound (a) Discontinuities of the reference wave file are clearly visible. (b) The experimentally obtained wave file does not show any discontinuities when viewed at the same scale as the reference file. (c) The experimentally obtained wave file begins to show noticeable discontinuities only when magnified to a much higher level.

The proposed approach overcomes some limitations of active noise control in the following ways:

- There is no introduction of a secondary source.
- Even in a complex field with many sources of sound, only one camera would be needed to eliminate unwanted signals. This can be done with the help of object tracking. The desired source can be selected and the remaining sources can be discarded. Thus, even though there would be interference of sound waves from the various sources, only the sound from the selected source will be recorded. Various Object Tracking algorithms have been described and compared in [19].

5.3 Selection of Desired Sound

By employing object tracking, we can track the desired source of sound. Thus, the proposed approach allows us to listen to a particular sound at a time when there are many sources producing sound simultaneously.

5.4 A Consistent Approach

Microphones come in a range of sizes and prices, having different directional properties. Each microphone is suited for a specific purpose. Like a microphone is used to record sound in air, a hydrophone, as described in [20], is used to record sound under water and a geophone, as shown in [21], is used to record vibrations under the ground. Every medium has a property called acoustic impedance, as defined in [22], which indicates how much sound pressure is generated by the vibration of the molecule of the medium at a particular frequency. Thus, different equipment has been designed for specific medium in which the sound is to be recorded.

The proposed approach is independent of the acoustic impedance of the medium. Thus, it will be possible to apply

the same approach in any medium without any major changes in the operation.

5.5 No Attenuation of Waves and Capable of Recording Very Small Amplitude

As the approach does not rely on the signal propagated from the source to record the sound, the sound will be recorded clearly irrespective of the attenuation of signal caused by the medium.

The proposed approach employed with a high resolution camera will be capable of recording extremely small vibrations. Thus, the sound signal of very low amplitude would be recorded in the presence of other high amplitude sound signals.

5.6 Vibrating Source in Vacuum

As the approach does not depend on the propagation of sound waves before recording the sound, it makes it possible to record a sound intended to be produced when the source is situated in vacuum.

5.7 Faster Recording of Sound

Velocity of sound differs from medium to medium. In certain applications, the sound may be needed instantly. Using the proposed approach; the delay in production of sound will be introduced by the camera, speed of light in the medium and computational cost in reproduction of sound. With the use of high speed processors and optimized computations, this delay can be made negligible as compared to the delay introduced by the speed of sound in most of the media. As the approach depends on the speed of light coming from the source rather than the speed of sound, it may be possible to record the sound instantaneously.

6. CONCLUSIONS

This paper proposes a novel approach for recording sound by mapping vibrations of the source. The experimental result confirms that the proposed algorithm applied on a 1000 fps, 1750 fps, 3350 fps, 4250 fps, and 5000 fps video clip of a tuning fork vibrating at 98 Hz, 123.47 Hz, 174.61 Hz, 220 Hz and 349.23 Hz, respectively, generates a smooth and accurate sound wave of the respective frequency. These experimentally generated sound file have significantly lower standard deviation and jitter as compared to a reference file of a the sound. The proposed technique employing high speed imaging can record the desired sound signal accurately by mapping the vibrations of the source and prevent the noise from being recorded. A number of potential advantages like consistency, no attenuation and capability to record any frequency of sound, etc. have been discussed.

FUTURE SCOPE

In this paper, tuning fork generating sinusoidal waves have been considered. As the result has been confirmed, the proposed technique needs to be tested for sound waves other than pure sinusoidal waves by considering more general vibrations. The technique needs to be developed to handle real time videos to enable real time sound recording. With the Ultra-fast imaging system, described in [23], developed at MIT, images of objects that are out of the line of sight of the camera can be captured. This system can be exploited for recording vibrations of the source which are not in the field-of-view of the camera.

REFERENCES

[1] D. Mas Candela, B. Ferrer Crespo, J. Espinosa Tomás, J. Pérez Rodríguez, A.B. Roig Hernández, and C. Illueca Contri, "High speed imaging and algorithms for non-invasive vibrations measurement," presented at The 4th International Conference on Experimental Vibration Analysis for Civil Engineering Structures, Varenna (Lecco), Italy, October 3-5, 2011.

[2] Q. Zhang, and X. Su, "High-speed optical measurement for the drumhead vibration," OSA's Optics InfoBase, Optics Express, Vol. 13, Issue 8, 2005, pp. 3110-3116.

[3] D.M. Freeman, and C.Q. Davis, "Using video microscopy to characterize micro-mechanics of biological and manmade micro-machines," Technical Digest of the Solid-State Sensor and Actuator Workshop, Hilton Head Island SC: Transducers Research Foundation, Inc, Jun. 1996, pp. 161-167.

[4] Gordon E. Moore, "Cramming more components onto integrated circuits", Electronics, Volume 38, Number 8, April 19, 1965

[5] R. Raskar, M.G. Bawendi, A. Velten, E. Lawson, A. Fritz, and D. Wu, et al., "Femto-Photography: visualizing Photons in motion at a trillion frames per second," Massachusetts Institute of Technology, 2011.

[6] Audacity. (2013, Jan. 21). Audacity (ver. 2.0.3) [Computer program] Available: http://audacity.sourceforge.net/

[7] Eli Maor, "Trigonometric Delights," Universities Press Ltd., 1998, pp. 200–202.

[8] T. Kurita, N. Otsu, N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models," Pattern Recognition, vol. 25, Issue 10, October 1992, pp. 1231–1240

[9] John Canny, "A Computational Approach to Edge Detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, Nov. 1986.

[10] G. Bradski, and A. Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library," O'Reilly Media, Inc., 2008, pp. 105.

[11] Steven Harrington, "Computer Graphics-A Programming Approach," 2nd ed., J. Peters, Ed. McGraw-Hill, 1987, pp. 400–417.

[12] Q. Tian, and M. Huhns, "Algorithms for sub-pixel registration," Computer Vision Graphics and Image Processing, vol. 35, 1986, pp. 220-233

[13] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-Accuracy Sub-pixel Image

Registration Based on Phase-Only Correlation," IEICE Trans. Fundamentals, vol. E86-A No. 8, 2003, pp. 1925-1933.

[14] GoldWave Inc. (2013, Jan. 24). GoldWave (ver. 5.68) [Computer program] Available: http://www.goldwave.com/

[15] Paul Boersma, and David Weenink. (2013, Feb. 09). Praat: doing phonetics by computer (Ver. 5.3.41) [Computer program]. Available: http://www.praat.org/

[16] Jerri, A.J., "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," Proceedings of the IEEE, 1979, vol. 67, Issue. 4.

[17] M. Tekavčič, "Active noise control," presented at the Seminar 2010-2011, Faculty for Mathematics in Physics, Section for Physics, Ljubljana, Slovenia, 2010, University of Ljubljana.

[18] S.M. Kuo, D.R. Morgan, "Active noise control: a tutorial review," Proceedings of the IEEE, vol.87, no.6, Jun 1999, pp. 943- 973.

[19] A. Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," ACM Computing Surveys, vol. 38, no. 4, 2006.

[20] O.A. Filatova, I.D. Fedutin, A.M. Burdin, and E. Hoyt, "Using a mobile hydrophone stereo system for real-time acoustic localization of killer whales (Orcinus orca)," Applied Acoustics, 2006, pp. 1243-1248.

[21] X. Zhang, X. Ke, and Z. Zhang, "Research on Micro-Electro-Mechanical-Systems Digital Geophone," Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on, vol. 3, 23-24 Oct. 2010, pp. 414-417.

[22] A.D. Pierce, "Acoustics-An Introduction to its Physical Principles and Applications," Melville: Acoustical Society of America, 1989, pp. 107-109.

[23] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M.G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nature Communications 3 (2012), pp. 745.

BIOGRAPHIES

Aaswad A. Satpute has pursued Computer Engineering from Sardar Patel Institute of Technology (SPIT), Mumbai, India in 2013. He has been keenly interested in research right from a very young age. He has successfully devised a method to Plot π on Real Number Line. His researches in the field of Mathematics include easy method to compute tables. His work has been appreciated by eminent people from India and around the world.

Sheetal M. Pandrekar has completed Computer Engineering from Sardar Patel Institute of Technology (SPIT), Mumbai, India in 2013. Since then, she has been working on many start-up endeavors and she also worked as a Junior Researcher at HESA (Hansa Embedded Systems Academy) in Mumbai, India. She takes profound interest in teaching and has conducted seminars and training sessions on various topics in Computer Science.

Dr. Imdad Ali Rizvi received his PhD from Indian Institute of Technology (IIT) Bombay, India in 2012. His areas of research include satellite image processing, object-based image analysis and multi-resolution algorithms for remote sensing and image processing. He is a life member of Indian Society for Remote Sensing and Resources Engineers Association. He has about 30 co-authored papers in international conferences and journals in the areas of image processing and analysis. His area of research is supervised image classification.