

Enhancing Fast Fourier Transform Algorithm for Keystroke Acoustic Emanation Denoising Strategy on Real-Time Scenario

Suleiman Ahmad*

Department of Cyber Security Science, Federal University of Technology, Minna, 920101, Nigeria

E-mail: ahmads@futminna.edu.ng

ORCID iD: <https://orcid.org/0000-0001-5506-8860>

*Corresponding Author

John Kolo Alhassan

Department of Computer Science, Federal University of Technology, Minna, 920101, Nigeria

E-mail: jkalhassan@futminna.edu.ng

ORCID iD: <https://orcid.org/0000-0002-9289-0404>

Shafi'I Muhammad Abdulhamid

Department of Cyber Security Science, Federal University of Technology, Minna, 920101, Nigeria

E-mail: shafii.abdulhamid@futminna.edu.ng

ORCID iD: <https://orcid.org/0000-0001-9196-9447>

Suleiman Zubair

Department of Telecommunication Engineering, Federal University of Technology, Minna, 920101, Nigeria

E-mail: zubairman@futminna.edu.ng

ORCID iD: <https://orcid.org/0000-0001-5242-3820>

Received: 14 May, 2023; Revised: 11 June, 2023; Accepted: 22 August, 2023; Published: 08 February, 2024

Abstract: The use of virtual keyboards in mobile devices such as smartphones and tablets has become an essential tool for inputting information. The sound of keystrokes has been observed in previous studies to be recorded along with ambient noises, such as those produced by uncontrolled student noise, fans, doors and windows, moving cars, and similar sources. The presence of such noises negatively affects the quality of the keystrokes signal, which in turn affects keystroke analysis. The traditional FFT-based denoising methods are vital but they are often limited by their inability to adapt to the varying characteristics of real-world audio and noises. This paper proposes an enhanced Fast Fourier Transform (FFT) with an adaptive threshold technique that reduces ambient noises. The adaptive threshold technique is developed to identify frequency bins that contain noise and set their sizes to zero or attenuate them to reduce the noise. The paper evaluates the performance of the enhanced FFT with adaptive threshold on keystrokes recorded audio and validates it through extensive experimentation. The results show that the enhanced FFT outperforms the traditional FFT in terms of speed and the amount of noise removed from the recorded audio signal, indicating a significant improvement.

Index Terms: Enhanced Fast Fourier Transform; Keystrokes; Smartphones; Denoising; Real-Time Environment.

1. Introduction and Related Work

Mobile devices are expanding worldwide, these smartphones are used for many daily and business activities, many of which contain sensitive information such as personally identifiable information (PII), bank and credit card numbers, passwords, health clearance files and address files. This means that malicious apps installed on smartphones can steal sensitive data and expose millions people to data theft. The operating systems of these smartphones prevent unauthorized access to applications/data using security techniques such as sandboxes and authorization. However, side channels such as sensors can bypass such security restrictions. Most modern smartphones have many sensors such as

microphones, cameras, gyroscopes and accelerometers that provide a better user experience, but these sensors leak information, allowing the attacker to provide confidential information and to steal sensitive information. Literature has established that even sound of keystrokes from virtual keyboard of smartphone can reveal personal information through side channel analysis [1].

The sound of typing on the keyboard was shown to leak information, indicating the possibility of acoustic side-channel attacks [2,3,4,5]. A keystroke sound is heard when the user presses a key on the keyboard while typing. These sounds from keystrokes will be recorded and analyzed to predict or infer what user type. However, the difference between the user ways of typing only leads to a change in the sound intensity of each keypress. Thus, a model can be developed to detect and classify the confidential information typed by the users [6]. It has been established in literature that each key on the keyboard emits acoustic with some level of unique characteristics that differentiate the keys [7]. Typing on both physical and virtual keyboards are inevitable as that is most popular way of inputting data into keyboard-related devices. The keyboard-related devices especially smartphones have become part of our daily lives. Smartphones are mostly used to carry out various day-to-day activities such as internet banking, normal typing of document, chatting, typing mails, filling various forms and many other typing activities. Therefore, the adversary captures acoustic emanated from some of these typing activities from their victim to launch keystrokes acoustic emanation attack.

Traditionally, the captured keystrokes acoustic emanations or the audio recording of the keystrokes in real time environments are usually mixed with ambient noise. Recorded audio signals are usually face with ever changing ambient noises. Keystrokes audio signals are either recorded in a controlled environment or in a real-world environment. The ambient noises have negative impact on both environments [8]. Acoustic noise is the most common noise present in our environment and this is better deal with using real-world condition than unrealistic controlled environment [9]. Examples are classroom noise, uncontrolled student noise, noise from fans, doors and windows, moving cars, people talking, wind, air-conditioners, rain, keyboard click, machines and other common noises around us. Consequently, robust technique(s) is/are required in order to produce clean audio signals as input for keystroke detection and classification. Different techniques have been used to either reduce the ambient noise or remove the noise since the noise perpetually affect the quality of audio and robust detection and classification of the keys [10], [11]. In the work of [12] changing environmental ambient noise is still an issue that needed more attention. One of these techniques which has been consistently used is the Fast Fourier Transform (FFT) however, not without deficit; especially in sparse data representation, and poor performance in denoising. Therefore, this research aims to enhance the traditional FFT as denoising strategy for keystroke acoustic emanation analysis. The subsequent sections of this paper include related works as section 2, while section 3 presents the overview of the experimental procedure. The proposed architecture for the enhanced FFT is presented by section 4. Lastly, sections 5 and 6 present experimental results and conclusion of the research work respectively.

2. Review of Previous Studies

Since the ambient noises are associated with keystroke acoustic emanation, which can evidently affect the expected results; there is need to advance research to either eliminate it completely or reduce it to the barest minimum. In view of this, the research of [12] acceded to the challenge of ambient noise, which has to be given the needed attention. The researchers achieved 85.4% to 75.6% accuracy in terms of detection and classification because of the influence of the ambient noise. Therefore, the search for more robust technique to further reduce the ever-changing environmental ambient noise is necessary.

In the work of [13,14], the authors considered denoising the PCG, computed Tomography (CT) and X-ray image as very important aspect of preprocessing because of the damaging effect of various ambient noises on the image signals. Specifically, [14] achieved 97.10% accuracy in their proposed classification methodology which they attributed to proper denoising technique, they however acknowledged that a large amount of data is needed for training set to obtain a higher accuracy. Consequently, the need for robust technique for denoising is highly desirable. The denoising of image signals is rated very vital aspect of the preprocessing in order to achieve a desired result. This assertion motivated the literature of [15] to consider the denoising technique in the field of remote sensing, medical and biometrics. The authors concluded that noise distorts the quality of expected information. Hence the need for continuous quest for an improved technique to either completely eradicates or reduces noise in audio signal. This is to achieve a clean signal for better information extraction and analysis of data. Naturally, the ecosystem is infiltrated with a lot of unwanted components capable to distort the expected results of any real life scenarios. For instance, the denoising technique has also been widely used on data obtained from the field of marine water (underwater). To this end, the work of [16] established that robust denoising technique is vital to marine acoustic signal processing. The authors noted that continuous improvement on denoising techniques especially on marine water is key because of the complex nature of marine environment.

In the field of signal processing, denoising is a common technique that involves using wavelets to enhance segmentation quality. This method entails breaking down a signal into high and low frequencies and applying a threshold to the wavelet's detail coefficients to eliminate noise. To optimize the signal-to-noise ratio (SNR) and remove ambient noise, various denoising methods, such as the FFT/IFFT filter, have been utilized. Robust denoising techniques

are significant in signal processing as noises accompanying signals acquired from systems like Holter monitoring of Electrocardiograph (ECG) can degrade signal quality.

Numerous studies have recognized the impact of ambient noise on signal quality and recommended denoising techniques to improve it. For example, the authors in [19] used the FFT/IFFT filter to denoise ECG signals, while those in [20] employed denoise techniques to eliminate noise from echo signals recorded from the speaker. Additionally, [21] established that denoising optical signals is necessary to enhance their quality. The authors recommended the use of various denoising techniques to improve signals. In conclusion, denoising signals is as crucial as detecting, recognizing, and classifying them, regardless of the signal type.

3. Overview of the Experimental Procedure

In this section, the procedures for entire experiment are presented. Firstly, generation and recording of keystrokes in the real environment are carefully built with the aim of recording the sound of each key on the virtual keyboard of different smartphones from different users, taking into account the changing environmental noise. The data generated and collected includes one hundred twenty (120) smartphones as presented by Table 2. In order to attain the desire goal of carrying out the experiment on real-life scenario, the participants were not subjected to any form of restrictions. While some participants were typing, other participants recorded their keystrokes under varying ambient noise conditions.

The audio recording test consists of 12 sessions. Each session consist of 20 participants among which 10 were typing while the other 10 were recording. Participants were selected across all levels of students from Federal University of Technology, Minna, Nigeria. Tables 1 and 2 show the details of the participants' biometric data and details of mobile phone used respectively.

Table 1. Demography of Participants

Gender	Age	Level of Education	Tribe	Social Status	Number
Male	18-24/24-50	Undergraduate/Postgraduate	Hausa, Yoruba, Igbo, Ebira & Other	Single/Married	90
Female	16-23/24-35	Undergraduate/Postgraduate	Hausa, Yoruba, Igbo, Ebira & Other	Single/Married	30

As shown in Table 1, these tests included men and women. All participants were undergraduate or graduate students in the age ranges shown in Table 1. There are 4,444 graduate students, most of whom are single, and most graduate students are married. Since this is a real world situation, there are different tribes across the country. 90 men and 30 women participated. In each session, each pair of changed their role as foe or victim.

Table 2. Details of Mobile Phone Used

Model of Phone	Number of phone model	Phone Platform	Sample of phone series	Version
Tecno	38	Android	Tecno CA7, Tecno K7, Tecno Camon x.pro, Tecno Pop 2 Power, etc.	>= 5.0.0
Infinix	34	Android	Infinix X626B, Infinix S, Infinix SMAT 3 Phy, etc.	>= 5.0.0
Samsung	19	Android	Samsung A20-SM520, Samsung Galaxy S8, Samsung Galaxy S7 Edge, etc.	>= 8.0.0
Huawei	4	Android/Honor	Huawei Y5Prime 2018, Huawei Y9S STK-L21, etc	>= 6.0.1
Gionee	6	Android	Gionee GN5001s, Gionee M7, Gionee M11, etc.	>= 7.0.0
Xiaomi Redmi	3	Android	Xiaomi Redmi Note 9, etc.	>= 9.0.0
Itel	8		Itel S32 Mini, Itel P33, Itel A56 Pro, etc.	>= 8.1.0
Oukile	1		Oukile C2	8.1
Nokia	1		Nokia 6.1	10
ASUS	1		ASUS-2012DB	8.0.0
Iphone	5		Iphone 6,	>= 12.5.1

As can be seen in Table 2, 120 different mobile phones were used in the experiment and files were produced in the real environment. Among them, Tecno users 31%, Infinix users 28%, Samsung users 17%, Itel users 7%, Gionee users 5% and Iphone users up to, Huawei users to count. For 3%, Xiaomi users are 2% and Oukile, ASUS and Nokia users are 1% each.

Similarly, audio of keystrokes of each key on virtual keyboard without considering the clicking style was collected. Each participant was allowed to either soft-click or hard-click depending on the participant's convenience. The participants were also allowed to use their natural speed so as to conform to the focus of this research work making the scenario completely real-world environment. Every participant filled out a bio-data and smartphone specification form

that was provided at the beginning of the experiment. The form is sectioned into two: first section contains gender and age of the participants, the second section contains name/model and version of the phones. The participants were paired up; one took the role of adversary while the other took the role of victim. In order to really take full advantage of the number of the participants, after each session, each pair interchange their roles. This really boosts both the size and varieties of our collected data. The noises were coming from walking in and out the halls by the students, ticking of tables and chairs, working fans.

4. Proposed Architecture for Enhanced FFT

Considering the FFT as one of the simplest and most commonly used in signal processing, it is enhanced to obtain a better performance compare to the traditional FFT. The architectural design for the enhancement of FFT is presented in this section.

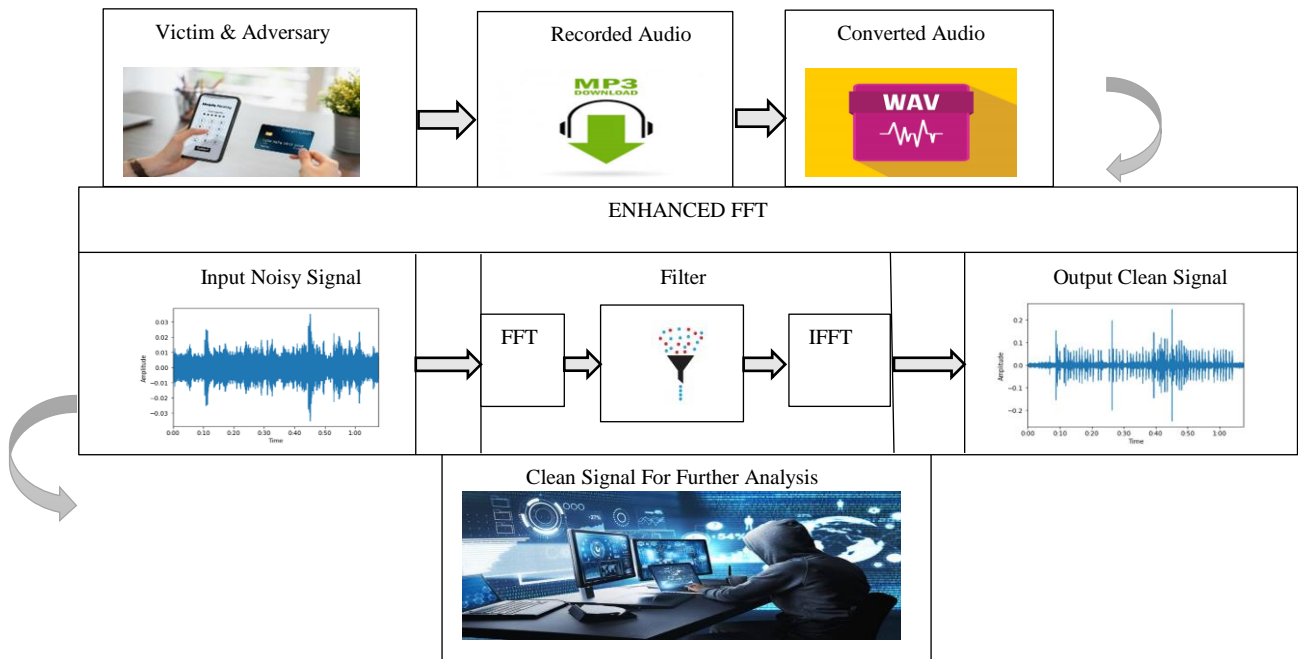


Fig. 1. Architecture for the Enhanced FFT

Figure 1 depicts the entire scenario of what happen from the victim (smartphone user) through the process of recorded noisy audio signal to the clean audio signal for further analysis by the adversary. The smartphone user was typing-in personal information freely in his/her phone. The user has no knowledge of the next person’s (adversary) intention of recording user’s keystrokes. The adversary on the other hand is quietly recording the keystrokes of the victim with the aid of already installed recording software. The adversary uses the built-in microphone of his/her smartphone for the recording. The audio recorded is inevitably corrupted by noise while acquiring, compressing and transferring into the system. This naturally could lead to loss of vital information from the original signal.

The recorded audio is in mp3 format and this is not appropriate format for signal analysis. The .WAV format, also known as Waveform Audio File Format, is an audio format that possesses the characteristic of being lossless. This implies that it is capable of maintaining the entirety of the initial audio data derived from the recording, without undergoing any form of compression or reduction in quality. In stark juxtaposition, the mp3 format, denoting MPEG-3, is deemed a lossy audio format, employing compression algorithms to reduce the file size by eliminating specific audio data that may not be readily discerned by the human auditory system.. In the realm of signal analysis, it is frequently imperative to operate with audio data that remains unadulterated, possessing the utmost level of quality, thus making .wav the preferred alternative. The audio is then converted to .wav format and then input into a system (laptop) where the enhanced FFT is loaded. The audio signal is converted from time domain to frequency domain by FFT as it is applied to the noisy signal, the frequencies associated with each coefficient is computed. The filtering process is activated where threshold value of 0.01 is set. If the coefficient is greater than or equal to the threshold, the algorithm would output it as the expected clean audio data. Else, the algorithm gives a leverage to loop once again to ascertain the validity of the initial judgment. Those confirm not to meet the requirement are dropped and when the process completed, the IFFT is computed. The signal is then converted from frequency domain to the time domain in order to obtain output in time-frequency signal.

5. Experimental Result

On the real-life raw data, (that is signal with noise), the result of FFT is figure 2 in the frequency domain. There is much noise in the output after the normal FFT denoising. However, the enhanced FFT removed noise as much as possible with the same raw audio data as can be seen in figure 3.

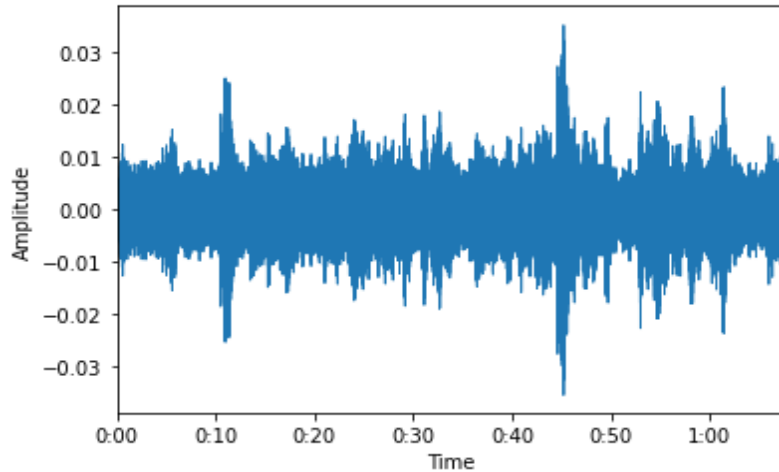


Fig. 2. Denoised raw data by Traditional FFT

It can be observe from figure 2 that the signal is blur with so much noise. The amplitudes could not be identified easily.

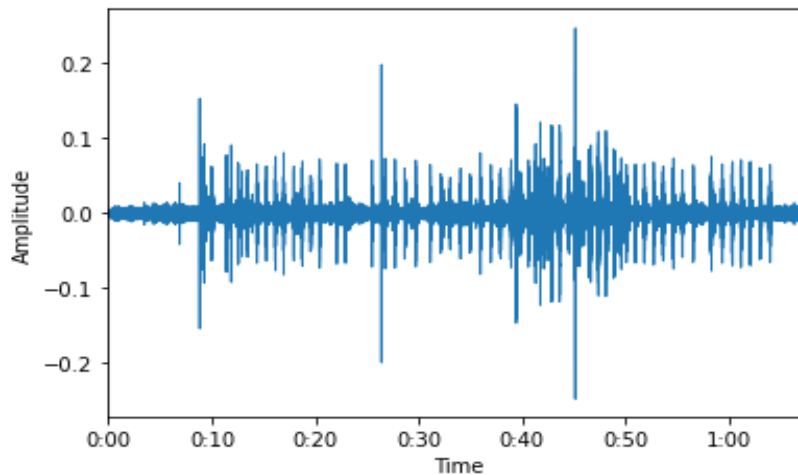


Fig. 3. Denoised raw data by Enhanced FFT

The fig. 3 cleans up the noise as much as possible. The amplitudes of the signals were clearly visible, cycle speeds and phases to correspond to any time signal were also clearly shown.

The table 1 display different amplitudes captured with their corresponding time taken to process the noise removal from the original audio signals.

Table 3. Comparing Sample Signals from Traditional and Enhanced FFT

Sample Signal		Conventional FFT	Enhanced FFT
Signal 1	Amplitude	0.015	0.015
	Time	0.06	0.03
Signal 2	Amplitude	0.02	0.02
	Time	0.10	0.08
Signal 3	Amplitude	0.025	0.025
	Time	0.14	0.10

Fig. 4 shows the appreciable difference in the time taken in denoising process. The conventional FFT took longer time compare to the enhanced FFT.

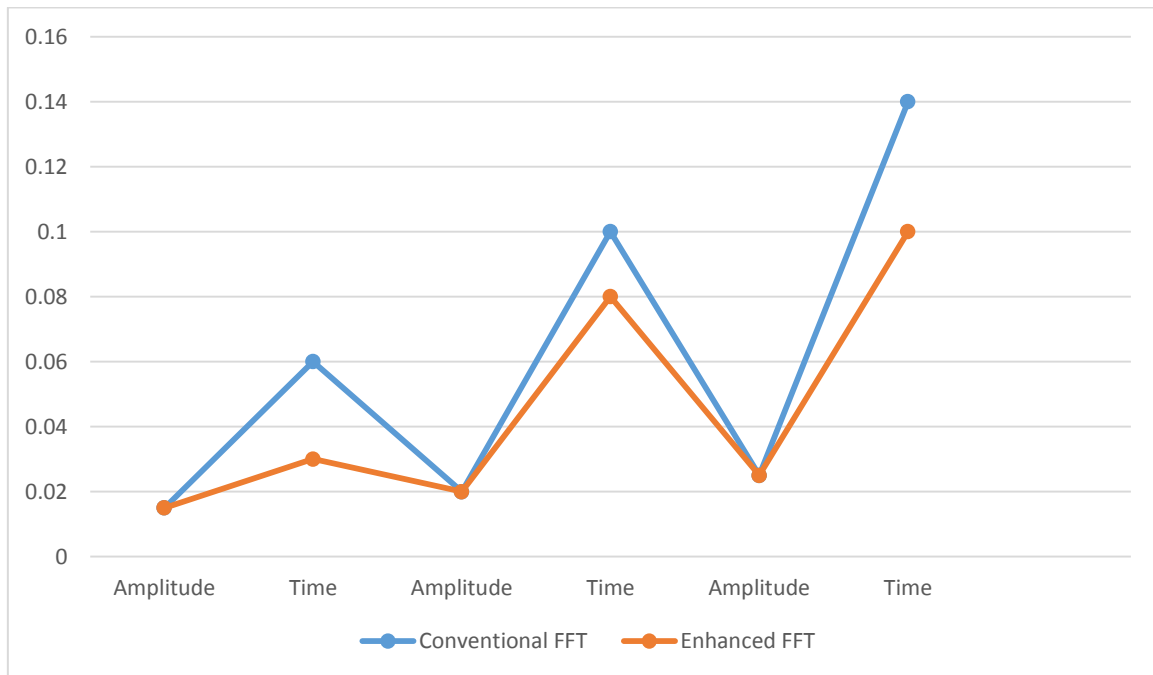


Fig. 4. Showing difference in time with same amplitudes

From fig. 3 and 4, it has been shown that the enhanced FFT outperform the conventional FFT in terms of processing speed and amount of noise been removed.

6. Conclusion

In this research work, an enhanced FFT is designed to reduce the influence of ever changing environmental ambient noise in real-world situation. The changing environmental ambient noise is still an issue that requires continuous research attention. FFT no doubt, over a period of time had been progressively exploited for noise removal due to simplicity and being very fast in terms of execution. However, it is often limited by its inability to adapt to the varying characteristics of real-world audio and noises, consequently leading to poor performance in denoising audio signals. This work generated data from one hundred and twenty (120) participants in the experiment with 120 smartphones that are commonly in use in the locality. The results showed that the enhancement to FFT improved the amount of noise removed from the raw signals and the processing speed has also been improved when compared to the conventional FFT.

References

- [1] I. Shumailov, L. Simon, J. Yan, and R. Anderson, "Hearing your touch: A new acoustic side channel on smartphones," pp. 1–23, 2019.
- [2] D. Asonov and R. Agrawal, "Keyboard acoustic emanations," *Proc. - IEEE Symp. Secur. Priv.*, vol. 2004, no. 1, pp. 3–11, 2004.
- [3] L. Zhuang, F. Zhou, and J. D. Tygar, "Keyboard Acoustic Emanations Revisited," pp. 373–382, 2005.
- [4] T. Halevi and N. Saxena, "Keyboard acoustic side channel attacks: exploring realistic and security-sensitive scenarios," *Int. J. Inf. Secur.*, vol. 14, no. 5, pp. 443–456, 2015.
- [5] A. Yeredor and R. Aviv, "Dictionary attacks using keyboard acoustic emanations," no. JANUARY 2006, 2016.
- [6] A. Zarandy, I. Shumailov, R. Anderson, and A. Alexa, "D ECODING SMARTPHONE SOUNDS WITH A VOICE ASSISTANT," 2020.
- [7] G. De Souza, F. Hae, and Y. Kim, "Differential audio analysis : a new side-channel attack on PIN pads," *Int. J. Inf. Secur.*, 2018.
- [8] D. Slater, S. Novotney, and J. Moore, "Robust Keystroke Transcription from the Acoustic Side-Channel," in *In 2019 Annual Computer Security Applications Conference (ACSAC '19)*, 2019, pp. 776–787.
- [9] A. M. A. Zaw Soe Yi, "Performance Comparison of Noise Detection and Elimination Methods For Audio Signals," vol. 03, no. 14, pp. 3069–3073, 2014.
- [10] A. Abuzneid, M. Uddin, S. A. Naz, and O. Abuzagheh, "An Algorithm to Remove Noise from Audio Signal by Noise Subtraction," pp. 5–10, 2008.
- [11] S. Lee and H. Kwon, "applied sciences A Preprocessing Strategy for Denoising of Speech Data Based on Speech Segment Detection," pp. 1–24, 2020.

- [12] H. Kim, B. Joe, and Y. Liu, "TapSnoop : Leveraging Tap Sounds to Infer Tapstrokes on Touchscreen Devices," 2020.
- [13] D. N. H. Thanh, "A Review on CT and X-Ray Images Denoising Methods Image formation in medical imaging systems and Poisson noise," vol. 43, pp. 151–159, 2019.
- [14] Y. Hu, "Time-Frequency Analysis , Denoising , Compression , Segmentation , and Classification of PCG Signals," vol. 8, 2020.
- [15] B. Goyal, A. Dogra, S. Agrawal, and B. S. Sohi, "Noise Issues Prevailing in Various Types of Medical Images," vol. 11, no. September, pp. 1227–1237, 2018.
- [16] Y. Li and L. Wang, "A novel noise reduction technique for underwater acoustic signals based on complete ensemble empirical mode decomposition with adaptive noise , minimum mean square variance criterion and least mean square adaptive filter," *Def. Technol.*, no. xxxx, 2019.
- [17] H. Abdelnasser, "MagStroke : A Magnetic Based Virtual Keyboard for Off-the-Shelf Smart Devices," 2020.
- [18] B. Nassi, Y. Pirutin, T. Galor, Y. Elovici, and B. Zadov, "Glowworm Attack : Optical TEMPEST Sound Recovery via a Device ' s Power Indicator LED," no. 3, 2021.
- [19] N. Zhang, Z. Nie, Y. Luo, L. Du, X. Wang, and L. Wang, "A Reconfigurable Overlapping FFT / IFFT Filter for ECG Signal De-noising," *2014 IEEE Int. Symp. Bioelectron. Bioinforma. (IEEE ISBB 2014)*, no. April 2014, pp. 1–4, 2020.
- [20] P. Cheng and I. Ethem, "SonarSnoop : active acoustic side-channel attacks," *Int. J. Inf. Secur.*, vol. 19, no. 2, pp. 213–228, 2020.
- [21] B. Nassi *et al.*, "Lamphone : Passive Sound Recovery from a Desk Lamp ' s Light Bulb Vibrations," 2022.

Authors' Profiles



Mr. Ahmad Suleiman, a PhD student and faculty member of the Department of Cyber Security Science, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.

Major interests: Cryptography, Cyber Security, Internet of Things Security, Digital Forensics & Investigations and Cloud Computing Security.



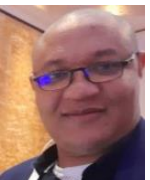
Prof. John Kolo Alhassan (BTech., MSc., PhD, MCPN, MNCS); had his Bachelor of Technology in Mathematics and Computer Science, from Federal University of Technology, Minna, Nigeria, in 2000. Masters of Science in Computer Science, from University of Ibadan, Nigeria in 2006. PhD research benchwork at United Institute of Informatics Problems, National Academy of Sciences of Belarus (UIIP NASB) Minsk, Republic of Belarus, 2012. Doctor of Philosophy (PhD) degree from Federal University of Technology, Minna, Nigeria in 2014.

Has taught many undergraduate and postgraduate courses. Has supervised many undergraduate projects, master and PhD theses. He is still actively involved in teaching and supervision of undergraduate and postgraduate students.

He has held various positions in the university such as School Examination Officer, Member, University Examination Misconduct Committee. Head, Department of Cyber Security Science, Member, Committee to Develop ICT Policy Document for Federal University of Technology, Minna, Nigeria. Deputy Director, Academic Planning Unit, Office of the Vice Chancellor.

He has published many articles in International Journals, Conference Proceedings, and book chapters. He has attended and presented papers at International and National conferences. He is a reviewer of many International and National Journals. He has served as a resource person for NECO, NOUN and NUC for accreditation exercises. He has served as external examiner for both undergraduate and postgraduate programmes in some Nigeria universities. Has won many national and international grants.

His areas of research are Artificial Intelligence, Information Security, Internet Technology, Software Engineering, Human Computer Interaction and Big Data Analytics.



Dr. Shafi'i Muhammad Abdulhamid was the former Head of Department (HOD) and a pioneer member of the first Cyber Security Science Department in Nigeria at the Federal University of Technology (FUT) Minna. He is presently a Visiting Assistant Professor of Cybersecurity in IT Department, Community College of Qatar (CCQ), Doha, Qatar. He received his PhD in Computer Science from the prestigious University of Technology Malaysia (UTM). His current research interests are in Cyber Security, Digital Forensics & Investigations, Cloud Computing Security, Internet of Things Security, Malware Detection and Cyber warfare. He has published more than 100 academic papers in reputable International journals, conference proceedings and book chapters. He has been cited more than 3360+ in *Google Scholar* with *h*-index of about 28 as at August 2023. Presently he is a Guest Editor in *Security and Communication Networks* journal (Hindawi/Wiley) to edit a Special Issue on *Advanced Fuzzy Logic in the Cloud and Network Security of Financial Applications* (Impact Factor of 1.964). Recently, he has edited a Scopus indexed book with Springer UK publishers titled *Machine Learning and Data Mining for Emerging Trends in Cyber Dynamic*. He has supervised both Masters and PhD students in Malaysia, Qatar and Nigeria.



Dr. Zubair Suleiman had his B.Eng degree in Electrical & Computer Engineering 2003, M.Eng. in Communication Engineering from the same university and PhD. degree in Wireless Sensor Networks from Faculty of Electrical Engineering MIMOS (Lab) Center of Excellence in Telecommunication Technology, Johor, UniversitiTeknologi Malaysia. He is a lecturer in the Department of Telecommunication Engineering, Federal University of Technology Minna, Nigeria. His research interest areas are, Wireless Sensor Networks, Embedded Systems, Data communication & Networking, Technological Development in third World Countries. Wireless communication, Optical Fiber Communications, Next Generation Networks and Biomedical Technology.

How to cite this paper: Suleiman Ahmad, John Kolo Alhassan, Shafi'I Muhammad Abdulhamid, Suleiman Zubair, "Enhancing Fast Fourier Transform Algorithm for Keystroke Acoustic Emanation Denoising Strategy on Real-Time Scenario", International Journal of Engineering and Manufacturing (IJEM), Vol.14, No.1, pp. 16-23, 2024. DOI:10.5815/ijem.2024.01.02