SN Computer Science

An Enhanced FFT Algorithm: Framework for Keystrokes Acoustic Emanations Denoising Approach on Real-Time Environment

--Manuscript Draft--

Manuscript Number:	SNCS-D-21-01617
Full Title:	An Enhanced FFT Algorithm: Framework for Keystrokes Acoustic Emanations Denoising Approach on Real-Time Environment
Article Type:	Original Research
Section/Category:	Security and Cryptology
Funding Information:	
Abstract:	In this digital age, the use of (physical or virtual) keyboard is indispensable as that is the most prevalent way of inputting data into keyboard-related-devices. Naturally, the captured keystrokes acoustic emanations or the audio recording of the keystrokes in real time environments are usually mixed with ambient noises; for examples, uncontrolled student noise, noise from fans, doors and windows, moving cars and the likes. However, to remove or reduce these ambient noises, some techniques which include FFT have been exploited in research because the noise perpetually affect the quality of audio and consequently, affects robust detection and classification of the keys. More so, FFT is considered in this research because is not expensive to obtain and it is nearly error-free measurement in audio suitable for waveform signal that is periodic. However, it is deficient in sparse representation of data, which leads to its poor performance in denoising. Therefore, this research work aims to design an enhanced algorithm for the FFT as an effective denoising framework for smartphones keystrokes acoustic emanations on a real-time scenario. This work is still in progress therefore, the researchers anticipate that upon successfully implementation of the algorithm, the enhanced FFT would perform better than the conventional FFT. The preliminary result shows significant improvement in speed of enhanced FFT over conventional FFT.
Corresponding Author:	Suleiman Ahmad, MSc Federal University of Technology Minna Minna, Niger State NIGERIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Federal University of Technology Minna
Corresponding Author's Secondary Institution:	
First Author:	Suleiman Ahmad, MSc
First Author Secondary Information:	
Order of Authors:	Suleiman Ahmad, MSc
	John Kolo Alhassan, PhD
	Shafi'i Muhammad Abdulhamid, PhD
	Suleiman Zubair, PhD
Order of Authors Secondary Information:	
Author Comments:	
Suggested Reviewers:	

An Enhanced FFT Algorithm: Framework for Keystrokes Acoustic Emanations Denoising Approach on Real-Time Environment

Suleiman Ahmad*
Dept. of Cyber Security Science
Federal University of Technology
Minna, Nigeria
ahmads@futminna.edu.ng

John Kolo Alhassan
Dept. of Computer Science
Federal University of Technology
Minna, Nigeria
jkalhassan@futminna.edu.ng

Shafi'I Muhammad Abdulhamid
Dept. of Cyber Security Science
Federal University of Technology
Minna, Nigeria
shafii.abdulhamid@futminna.edu

Suleiman Zubair
Dept. of Telecommunication Engineering
Federal University of Technology
Minna, Nigeria
zubairman@futminna.edu.

Abstract- In this digital age, the use of (physical or virtual) keyboard is indispensable as that is the most prevalent way of inputting data into keyboard-related-devices. Naturally, the captured keystrokes acoustic emanations or the audio recording of the keystrokes in real time environments are usually mixed with ambient noises; for examples, uncontrolled student noise, noise from fans, doors and windows, moving cars and the likes. However, to remove or reduce these ambient noises, some techniques which include FFT have been exploited in research because the noise perpetually affect the quality of audio and consequently, affects robust detection and classification of the keys. More so, FFT is considered in this research because is not expensive to obtain and it is nearly error-free measurement in audio suitable for waveform signal that is periodic. However, it is deficient in sparse representation of data, which leads to its poor performance in denoising. Therefore, this research work aims to design an enhanced algorithm for the FFT as an effective denoising framework for smartphones keystrokes acoustic emanations on a real-time scenario. This work is still in progress therefore, the researchers anticipate that upon successfully implementation of the algorithm, the enhanced FFT would perform better than the conventional FFT. The preliminary result shows significant improvement in speed of enhanced FFT over conventional FFT.

**Corresponding Author*

Keywords: Fast Fourier Transform; Keystrokes; Smartphones; Denoising; Real-Time Environment.

1. Introduction

It has been established in literature that each key on the keyboard emits acoustic with some level of unique characteristics that differentiate the keys [1]–[3]. 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 [4]. 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 [5]. 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 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 [6], [7]. In the work of [8] changing environmental ambient noise is still an issue that needed more attention. The work 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.

The Fast Fourier Transform (FFT) technique has been applied in various field of studies. It has been used as feature extraction and as a technique for denoising ambient noise [9]. The FFT is not expensive to obtain and it is nearly error-free measurement in audio. It is suitable for waveform signal that is periodic. Therefore, in this research work, the focus is to enhance FFT to reduce the influence of ever changing environmental ambient noise in real-world situation.

2. Related Works

Considering the importance of smartphones to human lives and its economic benefits, it is imminent to advance research on eavesdropping and side channel attacks based on audio recorded keystrokes. More importantly, how the inherent noises of the uncontrolled environment can be reduced or totally removed considering the focus of this research is accounted for by this section.

The research work of Motwani et al., 2003 and Swamy et al, 2020 present survey on denoising techniques. They compared the performance of Wavelet Transform and FFT techniques and concluded that wavelet transform performs better due to its multiscale, sparse and multiresolution characteristics. The authors pointed out that due to FFT's limitation in sparse representation of data, its performance in denoising is poor. However, in the research work of Alisha & Gnana, (2016) FFT has superiority over wavelet transform in terms of their performances.

The approaches to explore eavesdrop or side channel attack on what user type have been on the increase in recent time. The attacks have been based on information acquired from different sensors like video-based, camera-based, accelerometer-based, gyroscope-based, audio-based sensors and a host of others. In all these, there has been problem of ambient noise mixed with the desired signals especially audio-based signals [4].

Deep learning system was proposed using technique from speech transcription that performs end-to-end and audio-to-keystroke transcription. In this work, detection of keystrokes and keystroke classification were fused together. Dataset was generated using Seventy-six (76) typing sessions over Seventeen (17) users. Available datasets do not have typing audio with labeled keylogger transcripts, hence previous works rely on insufficiently small size dataset to train and evaluate deep learning models. This approach cannot handle overlapping keystrokes even after ambient noises removal. The performance of the approach was affected by minimal background noise [4]

3. Conceptual Framework for the Enhanced FFT

In this section, the concept of enhanced FFT framework is discussed. Considering the FFT as one of the most commonly used in signal processing, it is enhanced to perform better than the traditional FFT. Similarly, the algorithmic design for the enhancement of FFT is equally presented in this section.

Applying the traditional FFT to the raw data for denoising, the filter tends to remove all the clicks (include the desired signal) as a result of the condition of the threshold as shown by Figure 1.

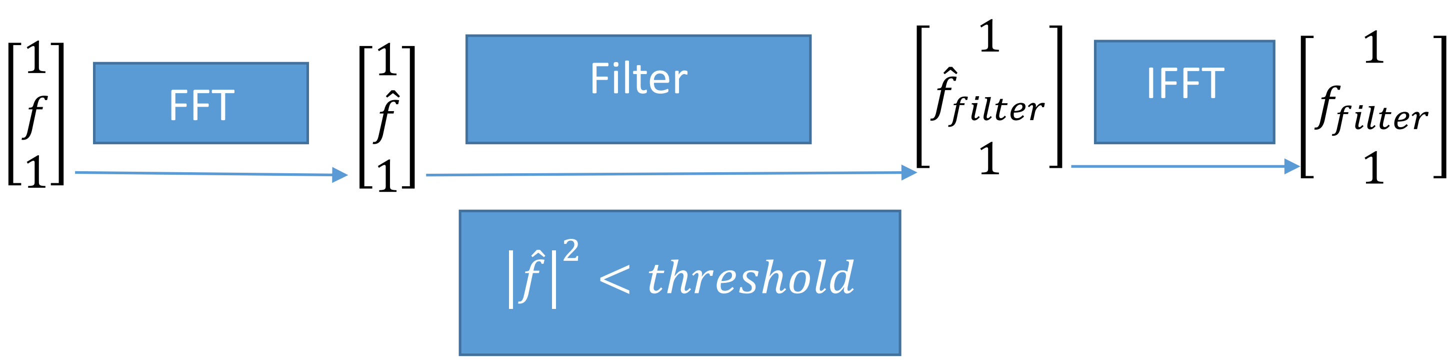

Figure 1. Traditional FFT

Considering Figure 1, the raw data (as recorded) is always mixed with ambient noise, there is increasing quest to get a better algorithm to separate the original signal from the noise. The raw data is inputted into the system as shown by the first matrix of the Figure 1. Consequently, FFT is applied on the signal (audio data) and the result is obtained as inverse matrix in preparation for filtering where the enhancement is done. The traditional or standard FFT algorithm determines the amplitude of the signal, there is a small threshold value, which is set as either 0.01 or 0.001. If an element from inverse matrix is greater or equal to this threshold value, then such element is inserted into the third matrix otherwise zero is inserted into that position in the third matrix. This filtering is done element-wise continuously until the whole data signal passing through is completed.

The new Matrix obtained from this process is the third matrix shown by the Figure 1. In order to undo the process that take place in converting FFT into inverse FFT, IFFT is applied to the matrix, and the result, which is the original signal data, is outputted. Conversely, Figure 2 shows the enhancement to the existing technique based on the condition of the threshold.

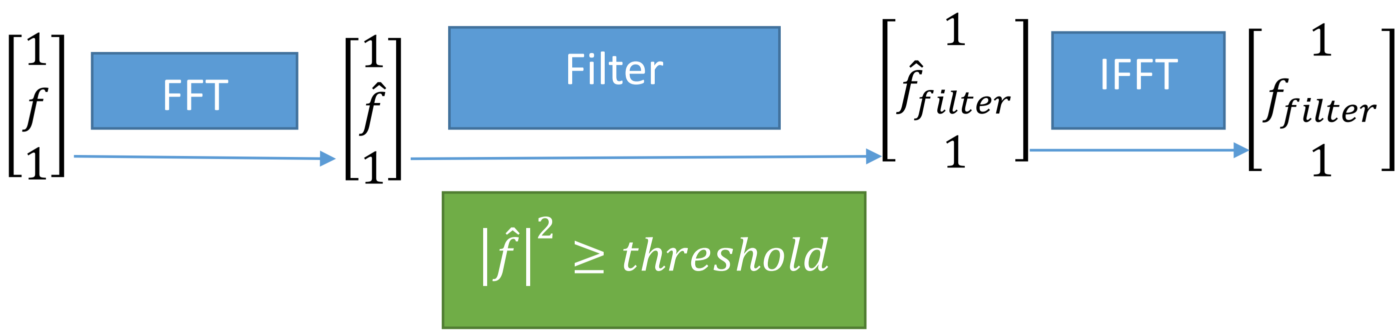

Figure 2: Enhanced FFT

With the enhanced FFT, it is believed that there will be improvement in the outcome and time taken in denoising when eventually carry out as the work is an ongoing. This enhanced framework is depicted by the Algorithm 1.

Algorithm 1: *FFT Enhancement*
Input: *R_{as}*
Output: *C_{as}*
Parameters: *Raw Audio Signal- R_{as}; Filtering_Process-FilP; Threshold-t; inverseFFT-iFFT CleanAudioSignal-C_{as}*

Procedure:
Step1: input [R_{as}]as array
Step2: generate $[R_{as}]^{-i}$
Step3: input $[R_{as}]^{i}$ to FFT
Step4: invoke FilP
Step5: set t =0.01,
Step6: forEach [R_{as}]
Step7: executes $[R_{as}]^{-i}$
Step8: if $[R_{as}]^{-i} \geq \boldsymbol{t}$
Step9: then iFFT and GOTO 2
Step10: else GOTO 12

Step11: endIf
Step12: output C_{as}
step13: endLoop

From Algorithm 1, the generated set of raw audio data from different smartphones are inputted as array data as shown by step1. Consequently, from step2 the inverse of the data is anticipated to be generated where it is being fed into the filter of the FFT technique as presented by step4. Threshold value of 0.01 is set and execution of the inverse audio files would be subsequently carried out. If the inverse audio file is less 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. It is important to mention that this work is still in progress in that the proposed algorithm has not been fully implemented. Therefore, as at the time of writing this research work; evaluation results are yet to be available.

4. Generation of Data from Real-World Environment

The generation and collection of data in a real-world environment considering the ever changing ambient noises was carefully designed with the intent of recording the audios of keystrokes of each key on virtual keyboard of different smartphones with different users. The data generation and collection scenario consist of one hundred and twenty (120) smartphones (see table 1b). Undergraduate lecture halls were used where the volunteers were not restricted in any form. Some participants recorded the keystrokes of other participants in those natural scenarios with ever changing ambient noises.

4.1 Set-up of the Experiment

The audio of keystrokes of each key on virtual keyboard were collected without considering the clicking style. 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 bio-data and smartphone specification forms, which were 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 as shown by Table 1.

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 role. This really boost both the size and varieties of the collected data. There were no specific instructions to participants on their typing speed and discussions were allowed to go on freely. Noises were not simulated but natural as the scenario was completely real-world environment. The noises were coming from walking in and out of the halls by the students, ticking of tables and chairs, working fans.

4.2 Demography of Participants/Phones Specification

The experiment for the collection of the audio data had 12 sessions. Each session consists of 10 pairs that is each session had 20 participants with 10 adversaries and 10 victims. The participants were taken from different levels of studies at the Federal University of Technology, Minna Nigeria during class sessions. The Table 1a and Table 1b respectively show the details of the participants' bio-data and phones specification.

Table 1a: 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

The experiment involved both male and female genders as shown by Table 1a. All the participants are either undergraduate or postgraduate students with the age brackets as appeared on the Table 1a. Undergraduates were majorly singles while the postgraduates were mainly married. Since it was real-world situation, there were diverse tribes from all part of the country. There were 90 male and 30 female participants. In every session, each pair interchanged their role as either adversary or victim.

Table 1b: Smartphones Specification

Model of Phone	Number of phone model	Sample of phone series	Version
Tecno	38	Tecno CA7, Tecno K7, Tecno Camon x.pro, Tecno Pop 2 Power, etc.	>= 5.0.0
Infinix	34	Infinix X626B, Infinix S, Infinix SMAT 3 Phy, etc.	>= 5.0.0
Samsung	19	Samsung A20-SM520, Samsung Galaxy S8, Samsung Galaxy S7 Edge, etc.	>= 8.0.0
Huawei	4	Huawei Y5Prime 2018, Huawei Y9S STK-L21, etc	>= 6.0.1
Gionee	6	Gionee GN5001s, Gionee M7, Gionee M11, etc.	>= 7.0.0
Xiaomi Redmi	3	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

From Table 1b, it is shown that 120 pieces of different off-the-shelf smartphones were used in the experiment to generate data in a real-world environment. The most common among them is Tecno model users account for 31%, Infinix users account for 28%, Samsung users account for 17%, Itel users are 7%, Gionee users are 5%, Iphone users are 4%, Huawei users are 3%, Xiaomi users account for 2% while Oukile, ASUS and Nokia users account for 1% each. For clearer representation of these statistics, Figure 3 presents a pie chart illustration.

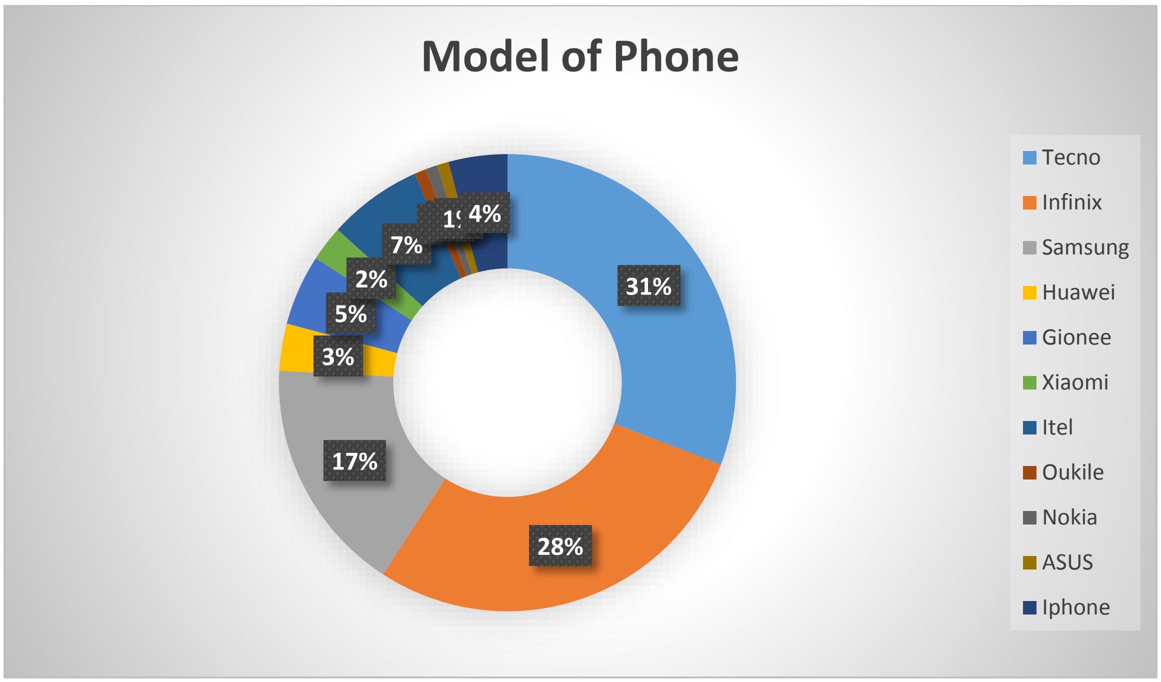

Figure 3. Pie Chart of Phone Model

From the phone model representation shown by Figure 3 during the course of data generation, it clearly shown the most popular smartphones among the students in this part of the country. The rationale behind it could not be ascertained in the course of this research work since it does not form part of the scope of this work. However, the statistic may open up further research interest.

5.0 Preliminary Results

In Table 2, some sample signals were selected where both amplitudes and corresponding time were considered. The parametric values of the signals were taken from both traditional and enhanced FFT. In the three sample signals, same amplitudes for both FFTs in each case but the corresponding time differ.

Table 2: Comparing Sample Signals from Traditional and Enhanced FFT

Sample Signal		Traditional FFT	Enhanced FFT
Signal 1	Amplitude	0.024	0.024
	Time	0.11	0.06
Signal 2	Amplitude	0.016	0.016
	Time	0.29	0.03
Signal 3	Amplitude	0.036	0.036
	Time	0.46	0.11

Figure 3 clearly shows the improvement in the enhanced FFT in term of speed. The time taken decreases significantly compare to the traditional FFT. This is also anticipated in retaining more features of the original signals after denoising.

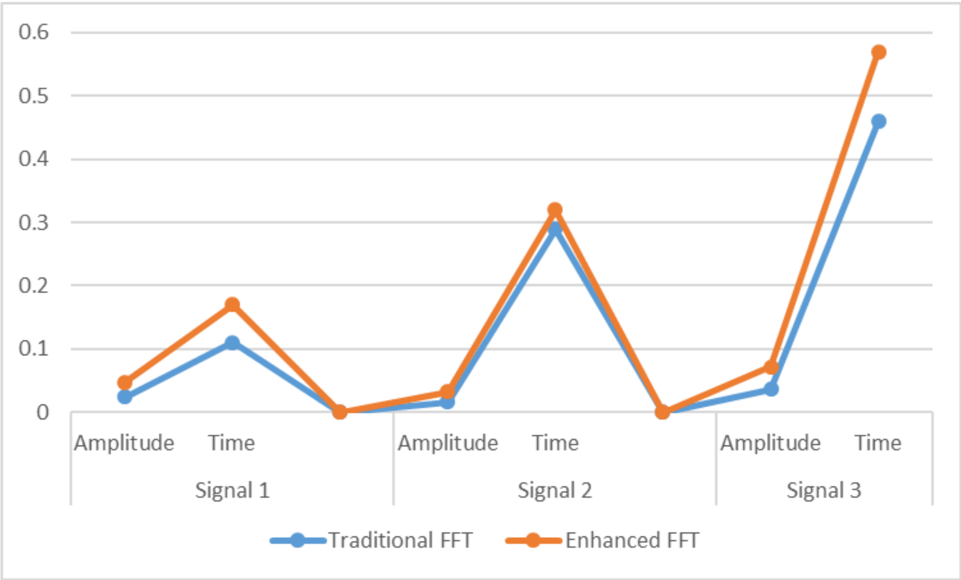

Figure 3. Showing difference in time with same amplitudes

6. Conclusion

In this research work, an enhanced algorithmic framework for 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 more research attention. FFT no doubt, over a period of time had been progressively exploited for noise removal; however, it is deficient in sparse representation of data consequently, leading to poor performance in denoising audio signals. This work considers one hundred and twenty (120) number of participants with 120 numbers of smartphones. The initial results showed that the enhancement will improve the accuracy of the classification of keystrokes with a cleaner data.

Declarations

Funding (Not applicable)

Conflicts of interest/Competing interests (On behalf of all authors, the corresponding author states that there is no conflict of interest)

Availability of data and material (data is available)

Code availability (software application or custom code)

Authors' contributions (optional: please review the submission guidelines from the journal whether statements are mandatory)

Ethics approval (Compliance with Ethical Standards)

Consent to participate (**Consent to participate**)

Consent for publication (**Consent to publish**)

References

[1] D. Asonov and R. Agrawal, "Keyboard acoustic emanations," *Proc. - IEEE Symp. Secur. Priv.*, vol. 2004, pp. 3–11, 2004, doi: 10.1109/SECPRI.2004.1301311.

[2] I. Shumailov, L. Simon, J. Yan, and R. Anderson, "Hearing your touch: A new acoustic side channel on smartphones," pp. 1–23, 2019.

[3] G. De Souza, F. Hae, and Y. Kim, "Differential audio analysis : a new side-channel attack on PIN pads," *Int. J. Inf. Secur.*, 2018, doi: 10.1007/s10207-018-0403-7.

[4] 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.

[5] Z. Y. ZAW and AYEAUNG MYINT, “Performance Comparison of Noise Detection and Elimination Methods For Audio Signals,” vol. 03, no. 14, pp. 3069–3073, 2014.

[6] A. Abuzneid, M. Uddin, S. A. Naz, and O. Abuzaghleh, “An algorithm to remove noise from audio signal by noise subtraction,” *Adv. Comput. Inf. Sci. Eng.*, pp. 5–10, 2008, doi: 10.1007/978-1-4020-8741-7_2.

[7] S. Lee and H. Kwon, “applied sciences A Preprocessing Strategy for Denoising of Speech Data Based on Speech Segment Detection,” pp. 1–24, 2020, doi: 10.3390/app10207385.

[8] H. Kim, B. Joe, and Y. Liu, “TapSnoop: Leveraging Tap Sounds to Infer Tapstrokes on Touchscreen Devices,” *IEEE Access*, vol. 8, pp. 14737–14748, 2020, doi: 10.1109/ACCESS.2020.2966263.

[9] N. A. Yusoff, M. Isa, H. Hamid, M. R. Adzman, M. Nur, and K. Hafizi, “Denoising Technique for Partial Discharge Signal : A Comparison Performance between Artificial Neural Network , Fast Fourier Transform and Discrete Wavelet Transform,” pp. 311–316, 2016.