Signal Processing Algorithm for Pre-processing the Surface Plasmon Resonance Signal Response

Isa Anshori
Lab-on-a-Chip Laboratory
Department of Biomedical Engineering
Institut Teknologi Bandung
Bandung, Indonesia
isaa@std.stei.itb.ac.id

Aminul Solihin
Department of Biomedical Engineering
Institut Teknologi Bandung
Bandung, Indonesia

Muhammad Harun Alrasyid
Department of Biomedical Engineering
Institut Teknologi Bandung
Bandung, Indonesia

Suksmandhira Harimurti
Lab-on-a-Chip Laboratory
Department of Biomedical Engineering
Institut Teknologi Bandung
Bandung, Indonesia

Gilang Gumilar
Advanced Functional Materials Lab.
Department of Engineering Physics
Institut Teknologi Bandung
Bandung, Indonesia

Muhammad Yusuf
Research Center for Molecular Biotechnology and Bioinformatics
Universitas Padjadjaran
Bandung, Indonesia

Silmina Prastriyati Sari
Research Center for Molecular Biotechnology and Bioinformatics
Universitas Padjadjaran
Bandung, Indonesia

Brian Yuliarto
Advanced Functional Materials Lab.
Department of Engineering Physics
Institut Teknologi Bandung
Bandung, Indonesia

Wyanda Arnafia
PT. Tekad Mandiri Citra
Bandung, Indonesia

***Abstract*— Surface plasmon resonance (SPR) is a versatile optical bio-sensing technique which has an ability to detect antibody-antigen molecular binding. In this work, we present a data processing algorithm that can process and analyze the data output from SPR equipment. The SPR data output is typically a non-periodic square wave, an indicator that a biological substance is captured, with continuous noises. To remove the outliers and smoothen the data, moving average and Savitzky-Golay Filter were employed. Then, a change point detection method and polynomial regression were used to isolate the buffer data as baseline and give baseline prediction which is later used to calculate and quantify the response. From this study, the algorithm is expected to give an accurate baseline prediction and response calculation. Based on the results, the algorithm was able to detect the SPR signal response (change point detection) with an error below 15%. Thus, this algorithm would enable the researcher to analyze and interpret the SPR data much faster and simpler.**

Keywords*— *Surface plasmon resonance; moving average; savitzky-golay change point detection; polynomial regression

I. Introduction

COVID-19 is a highly contagious disease caused by SARS-CoV2 which makes the world in pandemic condition. This airborne virus has the ability to survive in the droplet suspensions up to 16 hours [2]. This makes the spread of the virus is extremely hard to control.

The testing process is very essential to tackle the virus spreading. Based on the World Health Organization (WHO), the recommended test ratio for each country is 1000 tested over one million population or 0.001 tests/1M population. The PCR (polymerase chain reaction) technique is used as the gold-standard for the COVID-19 testing due to its ability to amplify the genetic molecules. But, it takes a long time to get the result. The testing process cannot be fully relied on the PCR, since we race with the fast spreading of the virus. Thus, rapid test is then used for the alternative test. But, it has lower sensitivity and specificity compared to PCR [3]. But, it is relatively cheap, so that it can boost up the testing number.

Currently, the test ratio in Indonesia is only 0.0001 [4], which is ten times below the WHO standard. This indicates that the testing capability of Indonesia is far from sufficient. The biosafety lab 2 is the minimum standard to perform PCR test and it is not evenly distributed in every region in Indonesia. It means that only certain regions that can perform the test directly, while the other regions have to send their sample instead. The real-time data is hard to be achieved since the sample transportation takes a long enough time.

Therefore, it is necessary to improve the testing process in order to get accurate results in a short period. As an alternative, the optical based biosensors can be utilized to overcome this issue. Commonly, they are widely used for the detection of molecular binding due to their advantages, such as high sensitivity and specificity, label-free, and real-time. In this case, the Surface Plasmon Resonance (SPR) is an optical based biosensor that can identify the interaction between antibodies and antigens [1-2]. Although it is a promising alternative test, the SPR raw data output is noisy and the vendor provided software cannot give a direct identification of the SPR signal response. Therefore, this paper proposes a signal processing algorithm to automatically remove the noise and simultaneously identify the SPR signal response.

II. Methods

A. Substrate and Sample Preparation

The substrate slide used in the experiment is a 20mm x 20mm x 1mm glass coated with an evaporated 45 nm thin film gold (Au), which has a 1.61 of refractive index. It was purchased from NanoSPR Devices Company, USA.

The IgY (Immunoglobulin-Y) antibody was used in substrate preparation. In the immobilization process, firstly a 10 mM MPA solution was applied to the surface of gold substrate and followed by 0.1 mM EDC and NHS. After that, IgY was dropped to the surface. 1% BSA was spread onto the surface drop by drop. Later, the gold substrate was incubated in the ambient temperature and pressure for 90 minutes. Then, it was flushed by immersing in the PBS buffer solution. The negative (IB and AI) and positive (RBD and Spike) sample were prepared in the PBS buffer solution. The concentration variation of each sample is shown in Table 1.

B. Data Acquisition

The NanoSPR-8 device (purchased from NanoSPR Company USA) was used in the data acquisition. Firstly, the device was turned on for approximately 30-60 minutes before the data acquisition. The prism glass was positioned in the holder and then the immersion liquid (1.61 refractive index) is applied on the top of the prism surface, followed by placing the bare substrate on the top surface. It is necessary to run the device for about 30 minutes with a buffer to remove the signal drift.

Generally, the data acquisition was done to know the response of the sample towards the buffer (baseline), whether ascending, descending, or no response. Each chamber of NanoSPR-8 was irrigated with a sample. The sequence of running processes was buffer, sample, and buffer. The first buffer was meant to be the reference line to calculate the response of the sample and the second buffer was for determining the dissociation of the antibody-antigen molecular binding.

C. Data Processing

The signal was a non-periodic square wave and disturbed with outlier spikes, continuous noise, and signal drift. Before calculating the sample ascending from baseline, the outlier spikes were removed by comparing the array with neighboring data. Savitzky-Golay Filter was used to smooth the signal due to its adaptability with polynomial function [3]. Buffer responses signal (baseline) and sample responses signal were separated using dynamic programming search method (DPSM) an offline change point detection function in ruptures python package [5]. The separated baseline data array was used for polynomial extrapolation baseline prediction under the sample. First derivative of each point was taken from the baseline prediction (BP) and used as the gradient for the perpendicular line towards the baseline (PL). The PLs were plotted along the x-axis (time). The ascending detection was done by calculating the mean of distance between two intersection points of PL and BP and PL and sample. The ascending data of an analyte was saved in Comma Separated Value (CSV) format. The whole process of this data processing is illustrated in Fig. 1.

TABLE I. ANALYTES USED IN THE EXPERIMENT

Analyte Type	*Analyte Name*	*Concentration (µg/mL)*
Positive control	RBD	0.4, 0.5, 1, 10
	Virus spike	0.1, 0.2, 0.4, 1
Negative control	IB	1/1000 TCID
	AI	1/1000 TCID

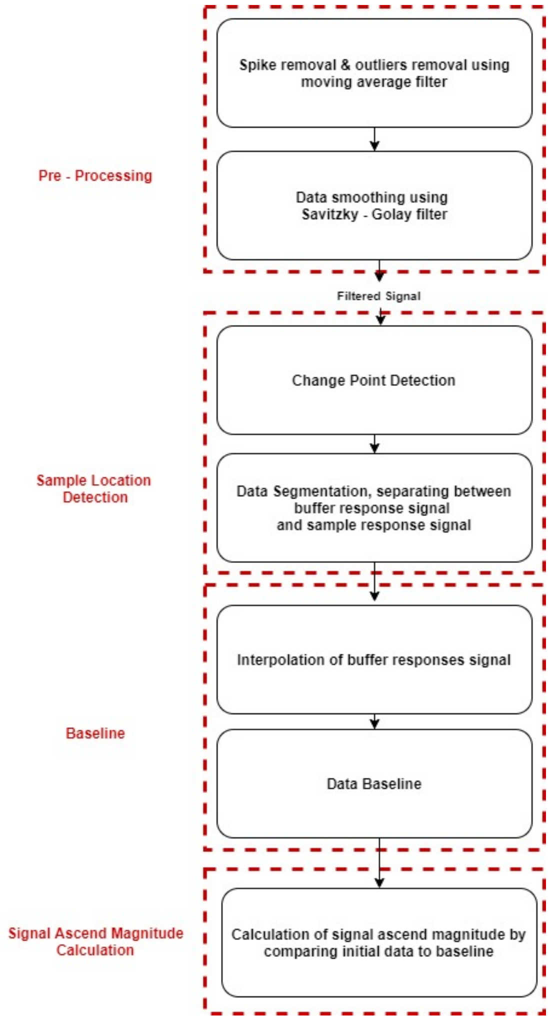

Fig. 1. Step-by-step process in data processing

III. RESULTS AND DISCUSSION

A. Pre-processing

The pre-processing was implemented in three steps, i.e. spike removal using despike algorithm, averaging with neighbouring data using moving average filter, and smoothing using Savitzky - Golay filter. In the first step, the despike algorithm which is adapted from the Raman spectra spike removal is employed to remove the unintended positive signal spike [10]. The despike algorithm calculates the median and median absolute deviation to determine the modified Z scores which contribute to the resolving spike signal. Then, the spike-removed signal is filtered using a moving average. This filter is commonly used in digital signal processing (DSP) due to its ability to reduce random noises while preserving the distinct step response [11]. The last, a Savitzky - Golay filter is utilized to smooth the signal. It is able to smooth the data in polynomial order using least-square smoothing signal. Results of the preprocessing stage can be seen in Fig.2(a) - Fig.2(c).

B. Sample Location Detection

This stage is implemented in Python using the Ruptures package and using the Dynamic Programming Search Method (DPSM) [9]. This method could solve nonlinear integer problems [12]. Then the change points obtained are used as references for data segmentation. The detected changes points are used as the point to separate the baseline and sample. The separated The algorithm at this stage was tested using 2 test data where the results obtained can be seen in Fig.3a and Fig.3b. The sample response area is marked in red, the buffer response area is marked in blue, and the orange line is the ground truth.

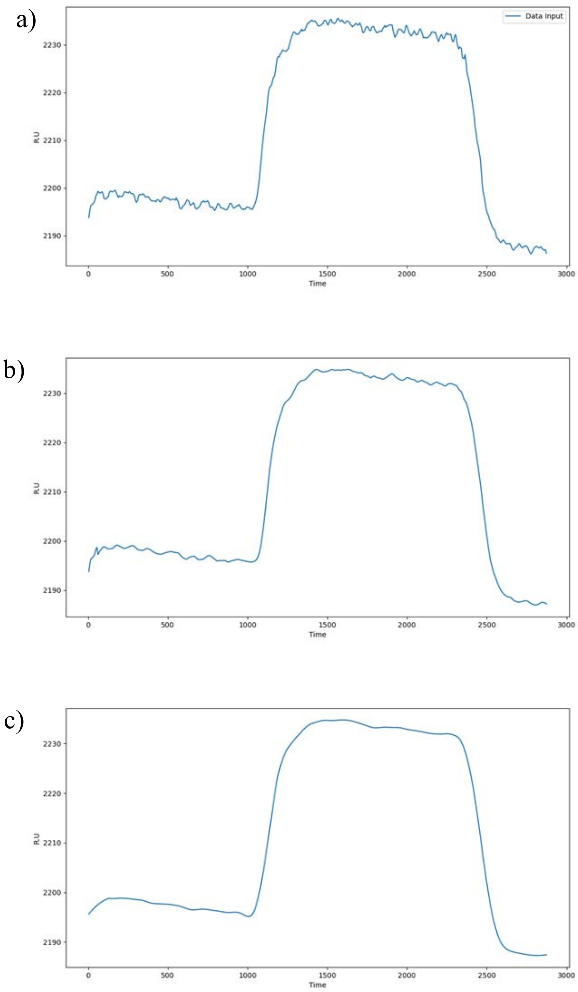

Fig. 2. Step-by-step data pre-processing after a) spike removal, b) moving average filtering, and c) signal smoothing using Savitzky – Golay filter

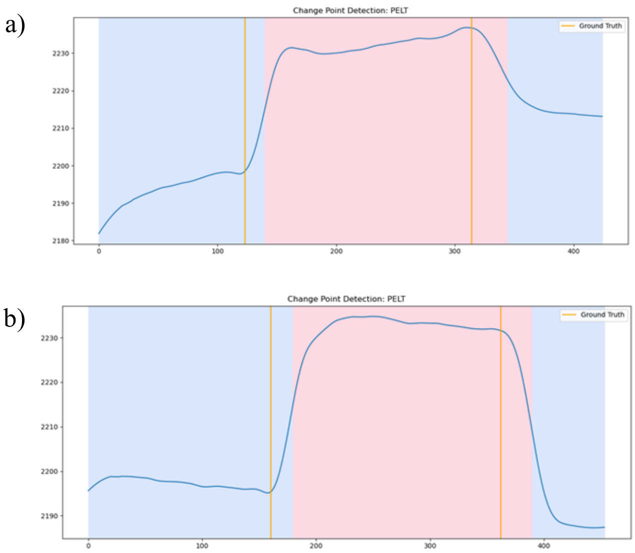

Fig. 3. The results of SPR signal response detection

TABLE II. CHANGE POINT AT 1 LOCATIONS

	Change Point 1 Location	
	Figure 3(a)	***Figure 3(a)***
GT	123	160
DPSM	140	180
Error (%)	13.8	12.5

TABLE III. CHANGE POINT AT 2 LOCATIONS

	Change Point 2 Location	
	Figure 3(a)	***Figure 3(b)***
GT	314	362
DPSM	345	390
Error (%)	9.9	7.7

C. Baseline Determination

Change point obtained before is used to exclude the sample response signal region, and the remaining data (buffer response signal region) are used for training data in the polynomial regression function (Fig.4(a)). Polynomial regression is implemented using Python Scikit-Learn Package. The result of the polynomial regression function is the baseline of the data, shown in Fig.4(b).

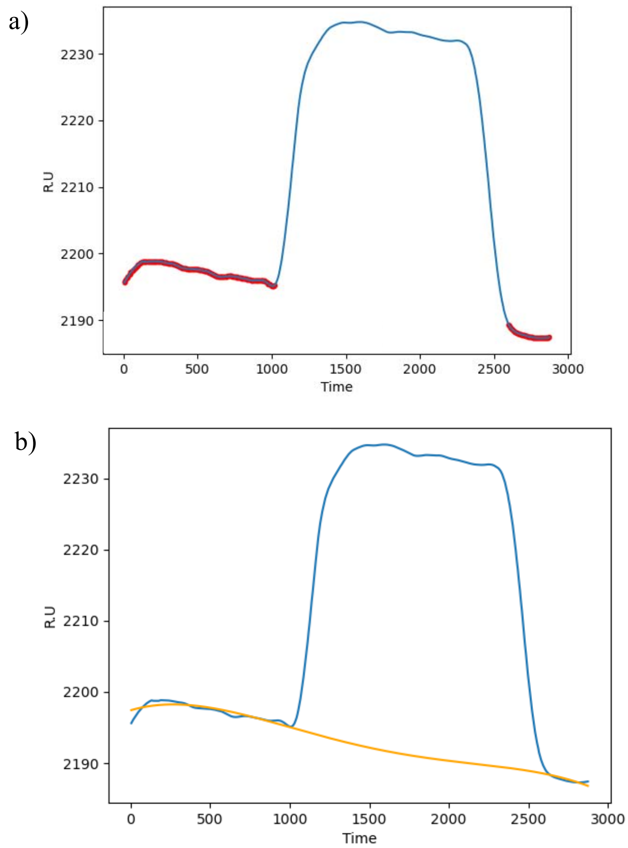

Fig. 4. The a) training data for and b) the results of baseline determination

In this step, there is minimal error because the program provides regression results based only on the training data provided. Meanwhile, the performance of this baseline prediction can be improved by increasing the accuracy of change point detection in the previous stage.

The calculation of the signal ascending is done by measuring the distance between baseline and the sample. The ripples in the calculated signal are caused by the inconsistency of the perpendicular line which intersects with the baseline and sample. The signal ascending is the average of the distance between sample and baseline, starting from the first change point to the second change point.

IV. CONCLUSIONS

In this study, an algorithm of processing SPR data output signal has been successfully developed. The proposed method is able to find the location of the sample response in the data, determines baseline of the given data, and calculates the magnitude of the signal ascend that will later be used for determining the positive or negative status of the sample.

ACKNOWLEDGMENT

The authors would like to thank to Badan Pengkajian dan Penerapan Teknologi (BPPT) for the funding support and the research project management of the National Task Force of the Republic of Indonesia for Covid-19 (TFRIC-19) Non-PCR test Subgroup as well as PT Tekad Mandiri Citra (TMC) for the generous support in providing the NanoSPR8 equipment for this research work. Additionally, the authors would also like to thank to Pusat Riset Bioteknologi Molekuler dan Bioinformatika Universitas Padjadjaran for providing the facilities required to conduct the whole research experiments.

REFERENCES

[1] L. Tate, "Tulane researchers find that coronavirus can survive in air for hours", Tulane News, 2020. [Online]. Available: https://news.tulane.edu/pr/tulane-researchers-find-coronavirus-can-survive-air-hours. [Accessed: 24- Sep- 2020].

[2] M. Lisboa Bastos et al., "Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis", BMJ, p. m2516, 2020. Available: 10.1136/bmj.m2516.

[3] "Coronavirus (COVID-19) Testing - Statistics and Research", Our World in Data, 2020. [Online]. Available: https://ourworldindata.org/coronavirus-testing. [Accessed: 24- Sep-2020].

[4] Killick, R., Fearnhead, P., & Eckley, I. A. (2012). *Optimal Detection of Changepoints With a Linear Computational Cost. Journal of the American Statistical Association, 107(500), 1590–1598.* doi:10.1080/01621459.2012.737745

[5] L. Fägerstam et al., “Detection of antigen—antibody interactions by surface plasmon resonance. Application to Epitope Mapping,”Journal of Molecular Recognition, vol. 3, no. 5-6, pp. 208-214, 1990.

[6] C. Baraniuk, “Covid-19 antibody tests: a briefing,”BMJ, 369, pp. m2284, 2020.

[7] J. Dombi and A. Dineva, “Adaptive Multi-round Smoothing Based on the Savitzky-Golay Filter,” In: Balas V., Jain L., Balas M. (eds) Soft Computing Applications. SOFA 2016. Advances in Intelligent Systems and Computing, vol 633. Springer, pp. 446-454, 2017.

[8] M. Henry, “Approximation by Polynomials: Interpolation and Optimal Nodes,”The American Mathematical Monthly, vol. 91, no.8, pp. 497-499, Oct. 1984.

[9] H. Ney and S. Ortmanns IEEE Signal Processing Magazine, vol. 16, no. 5, pp. 64-83, Sept. 1999.

[10] D. Whitaker and K. Hayes, "A simple algorithm for despiking Raman spectra", Chemometrics and Intelligent Laboratory Systems, vol. 179, pp. 82-84, 2018. Available: 10.1016/j.chemolab.2018.06.009 [Accessed 8 October 2020].

[11] S. Smith, "Moving Average Filters", Digital Signal Processing, pp. 277-284, 2003. Available: 10.1016/b978-0-7506-7444-7/50052-2 [Accessed 8 October 2020].

[12] K. Ng and N. Sancho, IIE Transactions, vol. 33, no. 12, pp. 1047-1058, 2001. Available: 10.1023/a:1010914518420 [Accessed 8 October 2020].