

# Memory and Processing Efficient Formula for Moving Variance Calculation in EEG and EMG Signal Processing

Mario Michael Krell<sup>1</sup>, Marc Tabie<sup>1</sup>, Hendrik Wöhrle<sup>2</sup> and Elsa Andrea Kirchner<sup>1,2</sup>

<sup>1</sup>University of Bremen, Robotics Lab, Bremen, Germany

<sup>2</sup>German Research Center for Artificial Intelligence (DFKI), Robotics Innovation Center, Bremen, Germany

Keywords: Variance, Mean, EMG, Electromyogram, EEG, Electroencephalogram, BCI.

Abstract: Adaptation of human-machine interaction devices by means of physiological data requires online analysis. We introduce new update formulas for otherwise time-demanding calculations of window based current mean and variance of the signal. Those were required for efficient realtime time series data processing. We discuss the formulas with the help of synthetic data. They differ from existing incremental calculations due to a *decremental* component, because of samples leaving the window of observation. An example application for EMG-based movement onset prediction is presented.

## 1 INTRODUCTION

For the improvement of human-machine interaction (HMI), many approaches make use of physiological data, e.g., gaze, electroencephalogram (EEG) or electromyogram (EMG) (Zander et al., 2010; Kirchner et al., 2013). Any data recorded during interaction has to be analyzed online to allow an online adaptation and improvement of the interface itself, e.g., by self-correction based on error potential detection (Zander et al., 2010), of the control of a machine or robotic device (Folgheraiter et al., 2012), or of the working condition of the user, e.g., adaption to high workload (Kohlmorgen et al., 2007). Recent approaches make use of hybrid brain-computer interfaces (BCIs) (Pfurtscheller et al., 2010) which support the human by detecting changes in different physiological data streams in parallel and by this do improve reliability and speed of the interface. Taking into account the growing demands in analysis of several data streams in parallel and that HMIs should even adapt before an interaction is carried out (Folgheraiter et al., 2012), the speed of data processing within the interface is critical. Further, for mobile support the analysis of physiological data must be enabled by means of *mobile* HMI devices (Wöhrle et al., 2013). Hence, new processing methods have to be applicable to small, low-cost processing devices with typically low memory resources. For these applications, specialized signal processing hardware like a Field Programmable Gate Array (FPGA) system is

a reasonable choice. For future HMI support, it is therefore required to review and improve existing data analysis algorithms regarding their applicability for the analysis of multiple data streams on mobile devices. In this paper, a new update formula for means and variance calculation was developed, which meets the before explained requirements of mobile processing.

Typically, one of the first processing steps in online analysis of physiological data is to calculate the mean and variance. Both can be used in two ways: as features or for an online standardization. A standardization might be needed, e.g., for EEG processing, to compensate for conductivity changes at the electrodes between human body and sensors or to reduce noise and artifact influence. The moving variance can be used as a feature for movement detection in HMI using EMG. Furthermore, in (Tabie and Kirchner, 2013) it is described how the variance in combination with an adaptive threshold can also be used to *predict* upcoming movements based on EMG signal analysis. This can be used to improve the control of an exoskeleton, as described in (Kirchner et al., 2013), where prediction times of 95 ms (28 Standard Error) on average could be achieved. An important requirement for online estimations of variance and mean is to update these values with each incoming sample and give the current up-to-date estimate which reflects their changing nature.

Earlier work handled this challenge only partially as described in the following. An incremental ap-

proach is given in (Welford, 1962) where the change in the “sum of squares of the deviations of the values about their mean” (short: sum of squares,  $S$ ) is calculated. In this approach, all recorded samples are used to estimate the mean and variance of the total signal. The expected changes over time, which we are interested in, are not monitored by this method. Thus, we added a *decremental* component to this approach, which enables the method to observe these changes. In (Chan et al., 1983) an algorithm for parallel computation of the sum of squares problem is given. This approach is very useful for fast offline calculations but requires extra memory and calculation resources on a mobile device. Formulas for an incremental approach are given as well and compared but no *decremental* approach is used. To allow forgetting of samples, it is suggested in (West, 1979) to use the weighted mean and variance. This approach would not be efficient enough because it requires two calculation steps.

To fulfill the requirements of mobile processing, we calculate the necessary formulas for mean and variance in Section 2. In Section 3 we give some empirical results showing the benefits of these new formulas. In Section 4 we present an experiment for EMG movement detection.

## 2 PROPOSED METHOD

In this section, the approach given in (Welford, 1962) is adapted to obtain the required relation for mean and variance. The mean  $m_0$  and the variance  $\frac{S_0}{n}$  are calculated with  $n$  samples  $x_0, \dots, x_{n-1}$ , the updated  $m_1$  and  $\frac{S_1}{n}$  with  $x_1, \dots, x_n$ .  $S_0$  and  $S_1$  are used for the sum of squares before and after the update. They have to be finally divided by  $n$  to get the variance. Thus,  $x_n$  is the incrementally added sample and  $x_0$  is the sample, which has to be forgotten in each calculation step.

There are two equivalent formulas for the mean:

$$\begin{aligned} m_1 &= m_0 + \frac{1}{n}(x_n - x_0) \text{ and} \\ m_1 n &= m_0 n + (x_n - x_0) \end{aligned} \quad (1)$$

To avoid rounding errors, induced by division, the sum of samples  $nm_1$  should be used for calculations. If the real mean is needed, this value has to be divided by the constant  $n$ . Next, the update formula for the mean 1 can be used to calculate the variance  $S_1$ :

$$S_0 = \sum_{i=0}^{n-1} (x_i - m_0)^2 \quad (2)$$

$$\begin{aligned} S_1 &= \sum_{i=1}^n (x_i - m_1)^2 \\ &= -(x_0 - m_1)^2 + (x_n - m_1)^2 + \sum_{i=0}^{n-1} (x_i - m_1)^2. \end{aligned} \quad (3)$$

The last part of equation (3) is separately transformed using the equations (2) and (1):

$$\begin{aligned} &\sum_{i=0}^{n-1} (x_i - m_1)^2 \\ &= \sum_{i=0}^{n-1} \left( (x_i - m_0) - \frac{1}{n}(x_n - x_0) \right)^2 \end{aligned} \quad (4)$$

$$= \sum_{i=0}^{n-1} \left( (x_i - m_0)^2 - \frac{2}{n}(x_n - x_0)(x_i - m_0) + \frac{(x_n - x_0)^2}{n^2} \right) \quad (5)$$

$$\begin{aligned} &= S_0 - 2(x_n - x_0) \left( \frac{1}{n} \sum_{i=0}^{n-1} x_i - \frac{1}{n} \sum_{i=0}^{n-1} m_0 \right) \\ &\quad + \frac{1}{n} \sum_{i=0}^{n-1} \frac{(x_n - x_0)^2}{n} \end{aligned} \quad (6)$$

$$= S_0 - 2(x_n - x_0)(m_0 - m_0) + \frac{(x_n - x_0)^2}{n} \quad (7)$$

$$= S_0 + \frac{(x_n - x_0)^2}{n}. \quad (8)$$

The result is now substituted into equation (3) and  $m_1$  is replaced using equation (1):

$$S_1 = S_0 + \frac{(x_n - x_0)^2}{n} - (x_0 - m_1)^2 + (x_n - m_1)^2 \quad (9)$$

$$\begin{aligned} &= S_0 + \frac{(x_n - x_0)^2}{n} \\ &\quad - (x_0^2 - 2m_1 x_0 + m_1^2) + (x_n^2 - 2m_1 x_n + m_1^2) \end{aligned} \quad (10)$$

$$= S_0 + \frac{(x_n - x_0)^2}{n} + (x_n^2 - x_0^2) - 2m_1(x_n - x_0) \quad (11)$$

$$= S_0 + (x_n - x_0) \left( \frac{x_n - x_0}{n} + (x_n + x_0) - 2m_1 \right) \quad (12)$$

$$= S_0 + (x_n - x_0) \left( \frac{n+1}{n}x_n + \frac{n-1}{n}x_0 - 2m_1 \right) \quad (13)$$

$$\begin{aligned} &= S_0 + (x_n - x_0) \left( \frac{n+1}{n}x_n + \frac{n-1}{n}x_0 - 2m_0 \right. \\ &\quad \left. - 2\frac{1}{n}(x_n - x_0) \right) \end{aligned} \quad (14)$$

$$= S_0 + (x_n - x_0) \left( \frac{n-1}{n}x_n + \frac{n+1}{n}x_0 - 2m_0 \right). \quad (15)$$

Finally the formula is multiplied with the constant  $n$ :  

$$nS_1 = nS_0 + (x_n - x_0)((n-1)x_n + (n+1)x_0 - 2nm_0). \quad (16)$$

For the calculation  $nS_0$  and  $nm_0$  as well as the last  $n$  samples have to be stored. Storing these samples might be also required for other signal processing steps like baseline correction, normalization and artifact correction. If the used samples are integers, the update can be calculated accurately without any rounding errors. This is usually the case for EEG and EMG signal processing. For the calculation with floating point or fixed point numbers, the multiplication with  $(x_n - x_0)$  might be problematic. To finally divide by  $n$  or  $n^2$  is not problematic, for two reasons: First of all, for EMG movement detection, we use the variance as a filter, hence, it does not matter if it is scaled by  $n^2$  and second we fixed  $n$  and therefore can calculate  $\frac{1}{n}$  and  $\frac{1}{n^2}$  beforehand with the required accuracy.

### 3 EMPIRICAL RESULTS

In this section, we compare the proposed method and the naive way of calculating the variance for a window of fixed size with  $n$  samples. In the proposed formula, the complexity is  $O(k)$  where  $k$  is the size of the time series. Compared to  $O(nk)$  for the naive way, here the complexity also depends on the window size since the variances have to be completely recalculated for each new sample.

In order to verify this and to point out the problem with the naive way, we performed tests on artificial data. The data, passed to the two algorithms, was a sinusoidal wave with a frequency of 5 Hz sampled with 5 kHz for 1 minute (300000 samples). With both approaches, the moving variance for each window using 8 different sizes (2, 10, 50, 100, 150, 200, 250, 300 samples) was calculated with 10 repetitions for each window size. The total execution time was measured. Results are shown in Table 1.

All calculations were performed on an Apple iMac with an Intel Core i5 CPU @2.7 GHz and 16 GB of RAM running Mac OS X 10.7.5. The implementation was performed in Python using version 2.6.8. The results show that for small window sizes up to 10 samples, the naive calculation of the variance outperforms the proposed method. With increasing window sizes, the computational time needed increases, too. For a size of 300 samples the time exceeds the length of the time series by 10.6s making an online processing with this approach impossible. For the proposed method, the results show that its calculation time is independent of the window sizes.

## 4 MOVEMENT PREDICTION WITH EMG

In a previous study, we could show that EMG signals can be used to reliably predict voluntary movements of the right human arm. We recorded EMG data from the right arm of 8 healthy, right-handed and normal or corrected-to-normal vision male subjects of age  $26.0 \pm 3.3$ . For each subject 6 runs containing 40 movements each were recorded. We investigated two movements speeds, slow and fast, and therefore 3 sets each, were recorded for every subject. For further details on the experiments please refer to (Tabie and Kirchner, 2013). The variance was used as non linear filter for preprocessing the EMG signals. The preprocessed signals were passed to an adaptive threshold for the detection of movement onsets. The adaptive threshold is given as

$$T_n(t) = \mu_n + p\sigma_n, \quad (17)$$

where  $\mu$  and  $\sigma$  are the mean and the standard deviation, for a window of  $n$  samples ending at time point  $t$ , respectively, and  $p$  is a sensitivity factor. The variance filter was calculated similar

$$V_m(t) = \sigma_m^2, \quad (18)$$

where  $\sigma_m^2$  is the variance for a window of  $m$  samples ending at time point  $t$ . The window sizes were optimized and resulted in values  $n = 20000$  (adaptive threshold) and  $m = 100$  (variance filter).

We analyzed the needed time for each run of this study for the here presented approach and the naive one. The results are shown in Table 2. The durations for each run varied in a range from 5.3 to 12.3 minutes. The needed computational times for our approach varied in range from 1.0 to 2.3 minutes and for the naive one from 404.8 to 944.3 minutes ( $\approx 6.7$  to 15.7 hours). Thus, the naive approach is not capable of online EMG processing for movement prediction in contrast to our approach. The long durations for the naive approach can be explained by the large window size (20000 samples) for the standard deviation in the adaptive threshold.

The presented results were calculated on the basis of results explained in Section 3. From those results we derived a formula to calculate the computational time for each method based on the window size and the length of a set. Even though the standard deviation used in the adaptive threshold is not exactly the variance its calculation is basically the same since it is the square root of the variance. Therefore, it is feasible to determine the needed time by looking at the time of the variance calculation.

Table 1: Execution time of the two methods for different window sizes. Time is given in seconds.

window size	2	10	50	100	150	200	250	300
time of our approach	5.5	5.5	5.6	5.5	5.5	5.5	5.5	5.6
standard deviation	0.03	0.02	0.1	0.08	0.05	0.05	0.04	0.07
time of naive approach	3.1	4.9	14.3	25.1	36.6	47.4	59.2	70.6
standard deviation	0.03	0.03	0.4	0.2	0.2	0.2	0.9	0.8

Table 2: Duration of sets (Run) recorded in (Tabie and Kirchner, 2013), and needed processing time with the presented approach (Our) and the naive approach (Naive) in minutes, respectively.

Set	Subject	1	2	3	4	5	6	7	8
1 slow	Run	8.2	8.2	9.7	9.3	8.2	10.7	7.9	8.2
	Our	1.5	1.5	1.8	1.7	1.5	2	1.5	1.5
	Naive	629.1	631	745	711.2	632.2	823.3	608.5	627.6
2 slow	Run	7.7	7.4	9.5	8.5	8.2	9.1	8.1	7.9
	Our	1.4	1.4	1.7	1.6	1.5	1.7	1.5	1.4
	Naive	589.6	571	726.9	649.2	632.4	699.3	618.3	602.6
3 slow	Run	7.1	7.3	9.4	9.2	7.7	7.9	7.4	7.7
	Our	1.3	1.3	1.7	1.7	1.4	1.5	1.4	1.4
	Naive	544	558.7	720	703.8	591.2	609	569.5	589.6
1 fast	Run	6.4	5.6	6.2	5.9	5.4	6.9	8.7	5.4
	Our	1.2	1	1.1	1.1	1	1.3	1.6	1
	Naive	488.7	426.1	475.9	454.9	414.3	527.4	664.4	414
2 fast	Run	6.3	5.9	6.7	11.3	4.9	6.6	12.3	6.6
	Our	1.2	1.1	1.2	2	0.9	1.2	2.3	1.2
	Naive	481	448.9	515.1	865.2	373.8	506.8	944.3	505
3 fast	Run	5.3	5.6	6.7	8.2	5.3	6.5	11.1	5.9
	Our	1	1	1.2	1.5	1	1.2	2.1	1.1
	Naive	410.1	428.8	511.9	627.7	404.8	500.9	852.8	455.9

## 5 CONCLUSIONS

In this paper, we motivated the necessity of moving mean and moving variance calculations for realtime muscle movement onset detection using EMG signals and for EEG data normalization. Formulas for an efficient calculation of mean and variance are derived in detail. The advantage of these new equations is shown in comparison with the naive approach. The resulting algorithm works in realtime and could be simply integrated into an FPGA to allow mobile adaptation of HMI systems by online analysis of physiological data. Furthermore, it was proven that the moving variance, when applied as a filter for EMG data, enables a very fast online movement detection. If there are other applications needing higher central moments (e.g., skewness), the same calculation scheme could be applied. This would result in update formulas with higher polynomial order.

## ACKNOWLEDGEMENTS

This work was funded by the *Federal Ministry of Economics and Technology* (BMW<sub>i</sub>, grant no. 50 RA 1012 and 50 RA 1011).

## REFERENCES

- Chan, T. F., Golub, G. H., and LeVeque, R. J. (1983). Algorithms for computing the sample variance: Analysis and recommendations. *The American Statistician*, 37(3):pp. 242–247.
- Folgheraiter, M., Jordan, M., Straube, S., Seeland, A., Kim, S. K., and Kirchner, E. A. (2012). Measuring the improvement of the interaction comfort of a wearable exoskeleton. *International Journal of Social Robotics*.
- Kirchner, E. A., Albiez, J., Seeland, A., Jordan, M., and Kirchner, F. (2013). Towards assistive robotics for home rehabilitation. In Chimeno, M. F., Solé-Casals, J., Fred, A., and Gamboa, H., editors, *In Proceedings of the 6th International Conference on Biomedical*

- Electronics and Devices (BIODEVICES-13)*, pages 168–177, Barcelona. SciTePress.
- Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Mueller, K.-R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., and Kincses, W. (2007). *Improving human performance in a real operating environment through real-time mental workload detection*, pages 409–422. MIT Press.
- Pfurtscheller, G., Allison, B., Brunner, C., Bauernfeind, G., Solis-Escalante, T., Scherer, R., Zander, T. O., Mueller-Putz, G., Neuper, C., and Birbaumer, N. (2010). The hybrid BCI. *Front. Neurosci.*, 4(30):1–11.
- Tabie, M. and Kirchner, E. A. (2013). EMG onset detection - comparison of different methods for a movement prediction task based on EMG. In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 242–247. SciTePress.
- Welford, B. P. (1962). Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4(3):pp. 419–420.
- West, D. H. D. (1979). Updating mean and variance estimates: an improved method. *Commun. ACM*, 22(9):532–535.
- Wöhrlé, H., Teiwes, J., Kirchner, E. A., and Kirchner, F. (2013). A framework for high performance embedded signal processing and classification of psychophysiological data. In *APCBEE Procedia. International Conference on Biomedical Engineering and Technology (ICBET-2013), 4th, May 19-20, Copenhagen, Denmark*. Elsevier.
- Zander, T. O., Kothe, C., Jatzev, S., and Gaertner, M. (2010). Enhancing human-computer interaction with input from active and passive brain-computer interfaces. *BrainComputer Interfaces*, pages 181–199.