Outlier Removal in Facial Surface Electromyography through Hampel Filtering Technique*

Susmit Bhowmik, Beth Jelfs, Sridhar P. Arjunan and Dinesh K. Kumar

Abstract— **The aim of the present investigation was to filter outliers in facial surface electromyography (fSEMG) originating from eye blinks, through a decision based filtering technique. Since, these outliers lie within the frequency range of electromyographic activity (30-300 Hz), conventional filtering methods fail to remove them. Hence, an application of an outlier filtering technique, Hampel filtering, has been introduced which is proficient at removing high frequency impulsive spikes (100-150 Hz) from facial sEMG. The Hampel filter removes the outliers without distorting the original data sequence and improves the quality of the signal as observed in time-frequency analysis.**

I. INTRODUCTION

There are 43 skeletal muscles in the face which are mostly controlled by the VII cranial nerve. The experimental technique through which the electrical activities of these muscles are measured is called facial surface electromyography (sEMG). It is the summation of the action potentials from the muscle fibres known as the motor unit action potential (MUAP) [1]. The facial musculature is unique and complicated as the muscles are closely knit together. Hence, crosstalk becomes an issue in facial sEMG recordings.

Compared to other skeletal muscles facial muscles are smaller in size, hence, their activation is of lower amplitude. Facial sEMG has certain limitations [2] due to the small size of the muscles it can be quite difficult to place electrodes accurately to isolate the targeted muscles. Thus, electrode placement leads to crosstalk and higher background activity [3], [4]. Conventional filtering methods have been successful in removing background noise but the issue with the occurrence of outliers due to crosstalk from adjacent muscles has not been properly addressed in facial sEMG before.

Different definitions of outlier definitions have been coined in statistical literature [5], [6]. Hawkins defined outlier as "an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism" [7]. Outliers may originate due to human error or other recording and environmental factors [8]. The influence of outliers on the recorded data can be immense, affecting the statistical parameters of the data by increasing the error variance and influencing the target sEMG recording. Hence, detection of outliers and their removal becomes an essential step when processing facial sEMG data.

The orbicularis oculi, the muscle responsible for eye blinks, often demonstrates crosstalk with surrounding muscle activity. For instance activity of the corrugator supercilii, which controls movement around the eyebrows such as frowning, has been demonstrated to be closely associated with that of the orbicularis oculi [9]. In the literature, motion artefacts in facial sEMG such as movements from eye blinks or eye ball rolling have been considered as low frequency components (0-20 Hz) and can be filtered by using high pass filters with cut-off frequencies in the region of 10 to 90 Hz [10]. These artefacts are a product of motion not activation produced from sEMG. However, it has been observed that due to the influence of crosstalk between muscles on the sEMG channels, outliers may also be produced from the muscle activity of endogenous eye blinks. These outliers are high frequency (130-200 Hz) in nature with frequency components which lie within the defined sEMG range (30-300 Hz). Hence, standard filtering does not remove the spikes and is inappropriate in this case.

In this paper, to address the problem with high frequency outliers, a statistical filtering method, the Hampel filter [11], has been introduced to remove the outliers from surface facial sEMG without overly affecting the raw data and thus improving the quality of the sEMG.

II. DECISION BASED FILTERING - HAMPEL IDENTIFIER

As highlighted in the Introduction, the presence of outliers distorts the statistical analysis of the data sequence significantly since the mean and the standard deviation are sensitive to the presence of outliers [12]. It has been found that the Hampel identifier is quite robust and effective in outlier removal from different biomedical applications. The Hampel identifier is calculated from the median value of the data sequence $\mathbf{x} = [x(1), x(2), \ldots, x(N)]$ and the median absolute deviation (MAD) from the median [12]. Hence, not only is it robust but also performs well in the presence of multiple outliers in the data. The Hampel identifier works on two parameters– a predefined threshold, T, and the median value of the chosen half window length of the given data sequence. The value of T controls the behaviour of the filter, as the threshold value increases fewer data points are identified as outliers whereas a decrease in the threshold leads to the identification of more outliers. The outliers detected in the window are replaced by the median value of the windowed data. If T approaches zero then, the filter will set all data points to the value of the median value of the window. The window of the data sequence, $\mathbf{x_w}(n)$, is given as $[x(n-k), \ldots, x(n+k)]$ as illustrated in Fig 1. The outliers are defined by the Hampel identifier as the

*The authors are with Biosignals lab, School of Engineering, RMIT University, Melbourne, AUSTRALIA (e-mail: s3579367@student.rmit.edu.au)

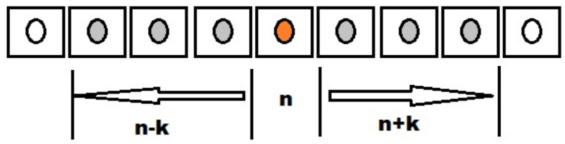


Fig. 1. Sliding window length for the Hampel identifier algorithm.

data points whose absolute difference from the median value $x'(n) = \text{median}(\mathbf{x_w}(n))$ is greater than the threshold and the MAD scale estimator, S, [13] such that

$$\text{Outlier}(n) = \begin{cases} 1 & \text{if } |x(n) - x'(n)| > TS, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where S is defined as

$$S(n) = 1.4286 \ \text{median}\{|(\mathbf{x_w}(n)) - x'(n)|\} \tag{2}$$

for $n = 1, \ldots, N$. The value 1.4286 is chosen so that the value of S equals the standard deviation of the normal distribution of the data and covers 50% of the standard normal cumulative distribution function. S becomes 0 when more than 50% of the data points have the same value as $x(n)$, which occurs in the case of crudely quantized data, and all other points in the sequence get rejected as outliers [14] regardless of how far they lie from $x'(n)$. The Hampel identifier has been introduced as a method to detect outliers occurring due to high frequency spikes in facial sEMG.

III. METHODS AND MATERIALS

A. Participants

Nine healthy participants without any facial musculature disorder volunteered to participate (age: 23-37 years, male = 6 female = 3) in the experiment, all provided written informed consent form. The experimental protocol was approved by Human Research Ethics Committee (HREC), RMIT University.

B. Equipment

The sEMG recording was performed using a wireless 4 channel Trigno Mini sensor (Delsys Inc). The dimension of the sensing head of the electrode is 25mm x 12mm x 7mm. These are ideally suited for recording sEMG from facial muscles due to their small size and inter-electrode distance of 10mm. The equipment uses a 16 bit resolution and a sampling rate of 2000 samples/sec.

C. Experimental protocol

An experiment was carried out to investigate the activation of the corrugator supercilii muscle and the influence of crosstalk from the orbicularis oculi when blinking. The electrode was placed on the corrugator supercilii as in Fig. 2 according to previous anatomical studies [15]. The skin was prepared with alcohol wipes to reduce skin impedance and remove dead cells. The participants were told to perform four short eye movements: (a) to stare at a fixed point without blinking; (b) to perform spontaneous controlled blinks; (c) perform a frowning action, in order to activate the corrugator supercilii without blinking; (d) to perform a combined forced blinking and frowning movement. These were performed to check for the presence of blinks as outliers in actual sEMG recording. Each recording was taken from the left eye for 5 seconds each.

Fig. 2. Electrode placement on the corrugator supercilii muscle.

D. Data analysis

The data analysis and then removal of the artefacts were performed via the following steps. Initially, a windowed moving standard deviation [16] of the data was calculated as

$$Y = \frac{1}{w}\left[\sum_{i=-m}^{m} x^2(n+i) - \frac{1}{2i+1}\left(\sum_{i=-m}^{m} x(n+i)\right)^2\right]^{\frac{1}{2}}, \tag{3}$$

where, $w = 2m + 1$ is the sliding window length and $n = m+1, m+2, .., N-m$. In the next step, the moving standard deviation was used to detect the start and end points of the outliers calculated by using the threshold value, T, which needs to be defined by the user manually. The algorithm detects the areas with values smaller than the threshold and sets them to zero so that the remaining areas with non-zero values correspond to the high spikes. The start and end points of the regions which were identified as outliers are then extracted. This is done throughout the data sequence to identify all the spikes. Having identified the outliers then the Hampel identifier algorithm is applied with the sliding window length, w, and the defined threshold, T, to remove the outliers.

IV. RESULTS

To illustrate the effect of the outliers on the data Fig. 3 provides a comparison of the average variance across all participants for experiment (c) where the participants were asked to frown and experiment (d) with the combined blink and frown. As can be seen there is a noticeable effect on the variance of the data due to eye blinks.

In order to remove the outliers using the Hampel identifier algorithm, it is very important to determine an appropriate Hampel window length. The window length must be long enough to capture the high frequency broad spikes. If the window length is too small there is high possibility that the outliers will go undetected and will be replaced by the median value of the adjacent samples. From the data

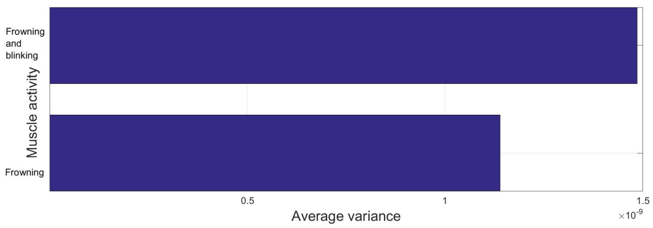


Fig. 3. Variance of data. Top bar shows the average variance across all participants when outliers (eye blinks) are included in corrugator activity and bottom bar shows the average variance in corrugator activity.

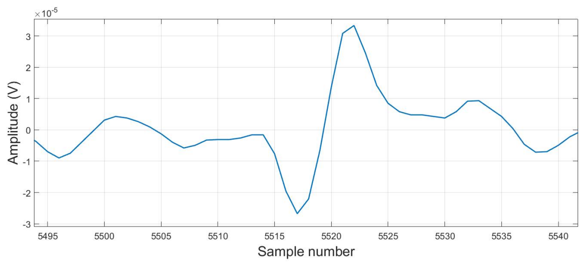
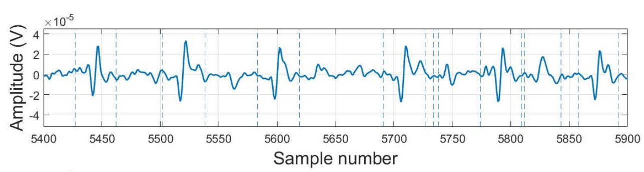


Fig. 4. The period of the activation of one eye blink consisting of around 20 samples with a frequency of 133.4 Hz.

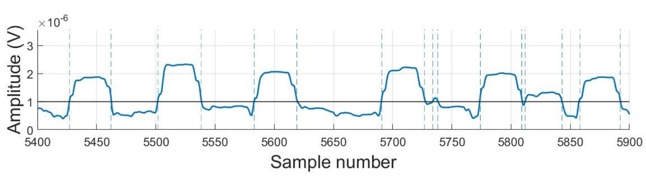
recorded from the experiment (2) with purely eye blinks and no activity of the corrugator supercilii, the eye blinks were determined to have a frequency range of $100 - 150$ Hz as illustrated in Fig. 4. In this case the half window length is chosen to be at least twice the width of the spectral peak of the eye blinks.

To determine the threshold, T, for the Hampel identifier the histogram of the data was used to identify the 95% confidence interval (CI). Having set the 95% CI the moving standard deviation was used to validate that the selected value of T successfully identified the outliers of the data as in Fig. 5.

Figure 6 demonstrates the performance of the Hampel identifier algorithm. Comparing the original and the filtered signals it can be seen that the large outliers have been removed. Having removed the outliers from the data a standard bandpass FIR filter with order 20 was used to filter the data from 30 to 300 Hz in order to remove any noise sources, for instance the low frequency eye movement. To compare the effect of the Hampel identifier vs. bandpass filtering alone a time-frequency analysis has been performed. It was observed that the power of the frequency components in the $100 - 150$ Hz range, which can be seen in Fig. 7(a) have been removed in Fig. 7(c) but not in Fig. 7(b).



(a) SEMG data with outliers identified by dashed vertical lines.



(b) Moving standard deviation with the selected threshold value shown by the horizontal line.

Fig. 5. Outlier identification using moving standard deviation.

V. Discussion

In this paper, a decision based filtering technique has been introduced to detect and remove outliers from facial sEMG. Very little research has been undertaken on the pre-processiong for removal of the outliers from the sEMG data of facial muscles, particularly from the upper half of the face. Van Boxtel [17] has shown that a bandpass of 0.4-512 Hz is sufficient to filter out the artefacts from sEMG data of peri-cranial muscles. However, this approach was inappropriate in the case discussed here where it was observed that the outliers occur due to artefacts arising from the crosstalk from adjacent muscles. As these artefacts are of a frequency within the frequency range of the sEMG they could not be removed by frequency domain filtering.

The Hampel filter has previously been demonstrated to improve the quality of the signal without adversely distorting the sEMG data from the target muscle [18]. The main advantages of this algorithm are the automatic detection and efficient removal of outliers. However, one drawback is the requirement for manual selection of the algorithm parameters. Previously Hampel filtering has predominantely been used in speech, audio and image analysis [19] but has not been investigated in the case of facial surface electromyography.

VI. Conclusion

Removal of outliers is very important when dealing with facial sEMG data. The filtering technique using the Hampel identifier is a promising and efficient method for removing the artefacts with frequency components within the range of the sEMG. We have successfully demonstrated the use of the Hampel identifier for removing the outliers originating from the adjacent muscles. This, in turn has increased the quality of the facial sEMG signal without affecting the raw data. To develop upon this work automatic selection of the threshold used within the Hampel filter would provide greater versatility of the proposed approach. This filtering method is not only applicable to sEMG but could also be of use for any biomedical signals which exhibit outliers in both time and spectral domain.

References

[1] D. Farina, M. Fosci, and R. Merletti, "Motor unit recruitment strategies investigated by surface emg variables," *Journal of Applied Physiology*, vol. 92, no. 1, pp. 235–247, 2002.

[2] A. Van Boxtel, "Facial emg as a tool for inferring affective states," in *Proceedings of measuring behavior*, 2010, Conference Proceedings, pp. 104–108.

[3] C.-N. H. C.-H. Chen and H.-Y. Chung, "The review of applications and measurements in facial electromyography," *Journal of Medical and Biological Engineering*, vol. 25, no. 1, pp. 15–20, 2004.

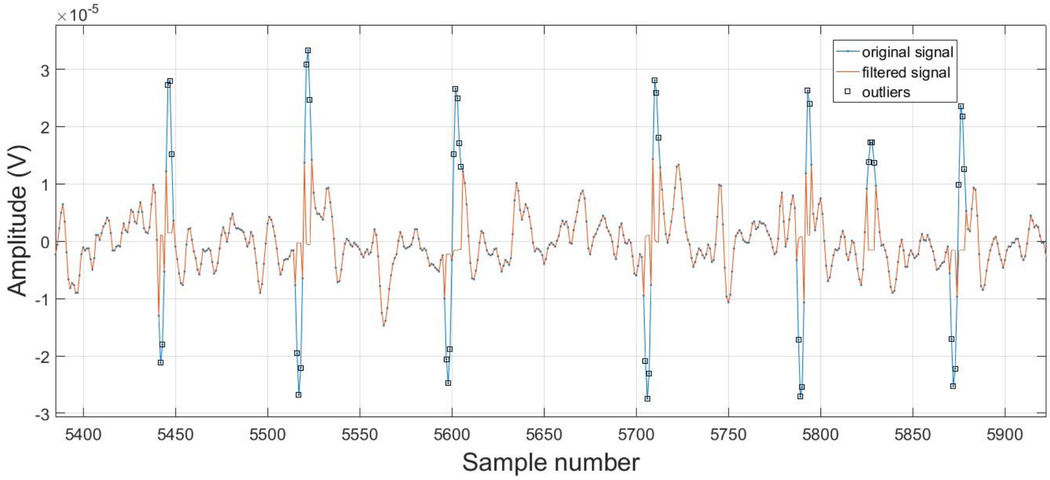
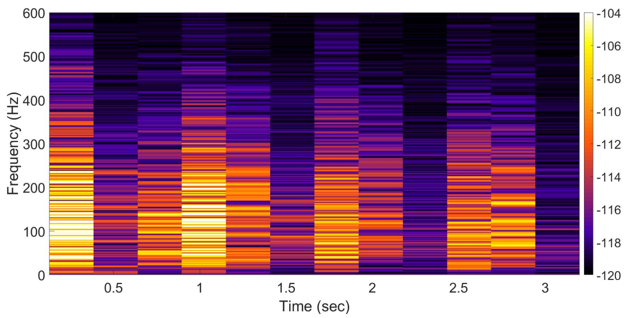
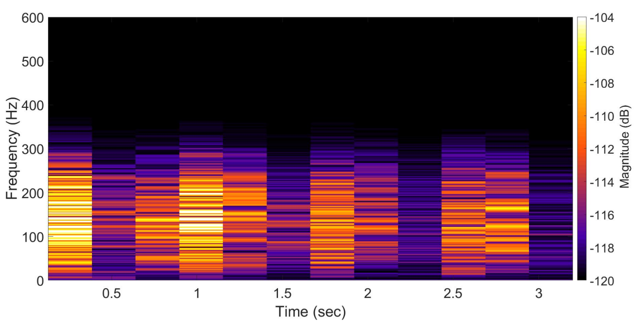


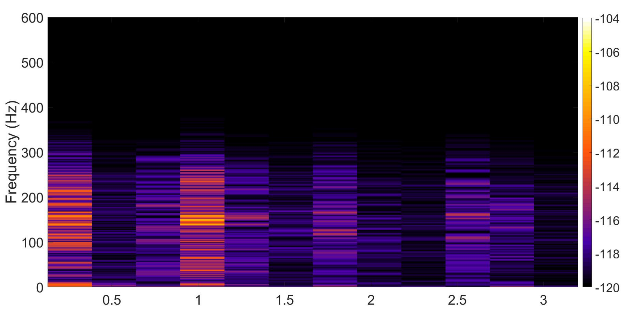
Fig. 6. Hampel filtering of EMG signal to remove outliers.



(a) Spectrogram of raw data.



(b) Spectrogram of bandpass filtered data.



(c) Spectrogram of Hampel and bandpass filtered data.

Fig. 7. Comparison of the quality of signal after filtering.

[4] B. G. Lapatki, D. F. Stegeman, and I. E. Jonas, "A surface emg electrode for the simultaneous observation of multiple facial muscles," *Journal of Neuroscience Methods*, vol. 123, no. 2, pp. 117–128, 2003.

[5] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[6] B. Iglewicz and D. C. Hoaglin, *How to detect and handle outliers*, 1993, vol. 16.

[7] D. M. Hawkins, *Identification of outliers*. Springer, 1980, vol. 11.

[8] J. W. Osborne and A. Overbay, "The power of outliers (and why researchers should always check for them)," *Practical assessment, research and evaluation*, vol. 9, no. 6, pp. 1–12, 2004.

[9] E. N. Marieb, "Human anatomy and physiology,(1995)," *Redwood City, Calif.: Benjamin/Cummings*, 1995.

[10] A. V. Boxtel, P. Goudswaard, and L. Schomaker, "Amplitude and bandwidth of the frontalis surface emg: effects of electrode parameters," *Psychophysiology*, vol. 21, no. 6, pp. 699–707, 1984.

[11] F. R. Hampel, "A general qualitative definition of robustness," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887–1896, 1971.

[12] R. K. Pearson, Y. Neuvo, J. Astola, and M. Gabbouj, "The class of generalized hampel filters," in *Signal Processing Conference (EUSIPCO), 2015*. IEEE, Conference Proceedings, pp. 2501–2505.

[13] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55–63, 2002.

[14] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, 2013.

[15] A. J. Fridlund and J. T. Cacioppo, "Guidelines for human electromyographic research," *Psychophysiology*, vol. 23, no. 5, pp. 567–589, 1986.

[16] F. Scholkmann, S. Spichtig, T. Muehlemann, and M. Wolf, "How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation," *Physiological Measurement*, vol. 31, no. 5, p. 649, 2010.

[17] A. Van Boxtel, "Optimal signal bandwidth for the recording of surface emg activity of facial, jaw, oral, and neck muscles," *Psychophysiology*, vol. 38, no. 01, pp. 22–34, 2001.

[18] D. P. Allen, E. L. Stegemller, C. Zadikoff, J. M. Rosenow, and C. D. MacKinnon, "Suppression of deep brain stimulation artifacts from the electroencephalogram by frequency-domain hampel filtering," *Clinical Neurophysiology*, vol. 121, no. 8, pp. 1227–1232, 2010.

[19] F. J. Gallegos-Funes and V. I. Ponomaryov, "Real-time image filtering scheme based on robust estimators in presence of impulsive noise," *Real-Time Imaging*, vol. 10, no. 2, pp. 69–80, 2004.