OM signal

Validation of Breathing Rate Algorithm During Running

Whitepaper by OMsignal
(Dated: December 13, 2017)

***Abstract*—The work described in this report is concerned with verifying the accuracy of the breathing rate algorithm (based on the inhale peak detections in the OMsignal box firmware). For this purpose, nine raw data files collected from men wearing OMshirts and thirty raw data files collected from women wearing OMbras had a portion of their useable recording manually scored by an engineer to identify true inhale peaks. The performance of the event detection algorithm is assessed based on sensitivity and positive predictive value (PPV). The accuracy of breathing rate was assessed using the 95th percentile of the absolute difference between the rate computed from the manual scores and the automatically detected IN-IN (inhale to inhale) intervals. The overall accuracy is 2.1 (+2.7) BPM which is within the desired tolerance of 5 BPM. Additionally, the per file average sensitivity and PPV is 94% and 94% respectively. The breathing rate algorithm is assessed to work well for both men and women in the context of running.**

I. Purpose and Scope

The work described in this report is concerned with verifying the accuracy of the breathing rate algorithm, while it is in use by runners wearing properly fitted OMsignal clothing. The breathing rate algorithm forms the basis of several features within the OMsignal infrastructure, such as AT/VT estimation. The breathing rate may also be displayed to the user in some form, and it should make sense to them based on what they were doing at the time. It is important to ensure that breathing rate is accurate for these uses. This report is not concerned with the verification of breathing depth, breathing regularity, or breathing rate during any use case except during a run (that may include brief periods of rest in between periods of running, as a real user may do). Additionally, this report is not concerned with the accuracy of breath event detection, though this is of indirect interest as errors here will generate errors in breathing rate.

II. Data Characteristics

There are four sub data sets, referred to as breakdowns, that are included in this analysis. TABLE I shows the number of files analyzed in each of the breakdowns and the number of corresponding users since several files may be analyzed from the same user.

The shirts dataset is comprised of recording of men running while wearing the OMshirt. The old bras breakdown is comprised of women running with early prototypes of the OMbra. The new bras round 1 and 2 breakdowns are comprised of women running with new OMbras with the finalized production design. Part of the algorithm was designed on the basis of new bras round 1, so round 2 was introduced to validate the modifications. The new bras round 1 breakdown includes 3 files that were recorded from women wearing bras that were intentionally incorrectly sized. These will be referred to as the ill-fitted bra files.

III. Exclusion Criteria

The criteria for excluding signal regions from analysis were as follows:

- 209 Artifact: This could be identified by a very sudden shift in the baseline of the signal that repeats every 209 seconds. These introduce errors in breathing detection, but they are predictable and occur infrequently. They are not important to consider in this context. Fig. 1 shows an example of the 209 artifact, and its resulting impact on inhale detection.
- No clear periodic signal: There is no periodic signal present in the pass band from 0.16 to 1.5 Hz. If this signal is not present, then either the breathing is very irregular, or more likely the region is very noisy. In either case, it is impossible to score the region with any degree of certainty. Fig. 2 shows an example of a signal obscured by noise. Fig. 3 shows an example of the same signal filtered to a pass band of 0.16Hz-1.5Hz. This reveals that there is no periodic pattern in the signal consistent with breathing.

IV. Manual Scoring Process

This section outlines the process used to decide what to manually label as an inhale. The end of an inhale is represented in the OM breathing signal as a local maximum of the signal. Other artifacts can also be local maxima though. To decide which local maxima, or peaks are real breaths and which are not, a bandpass filter is applied, with a band chosen to isolate frequencies between 0.1 and 1.5 Hz. The exact band is chosen on a case by case basis by the engineer based on what reveals a signal judged to most resemble breathing. The local maxima of this signal are then annotated as inhales, provided it passes the following tests:

Breakdown	Number of files analyzed	Number of users analyzed	Time analyzed (min)
Shirts	9	6	108.63
Old bras	9	6	49.29
New bras round 1	11	6	72.19
New bras round 2	10	9	204.92

TABLE I: List of the number of files and unique users analyzed in each breakdown of interest.

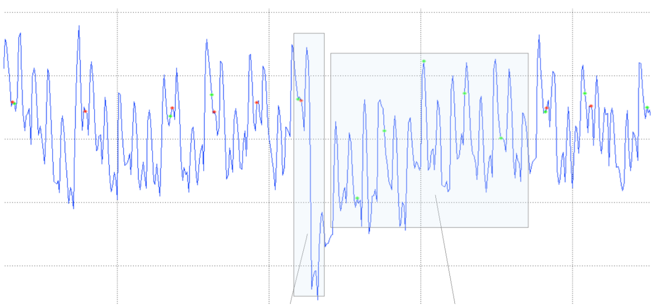

Fig. 1: A) 209 Artifact in first highlighted box. B) 209 Artifact induced detection failure in second highlighted box. Pairs of consecutive green dots indicate manual annotation of inhales. Red dots indicate automatically detected inhales. Note that the high frequency signal on top of the breathing is step noise; the signal as displayed is not filtered.

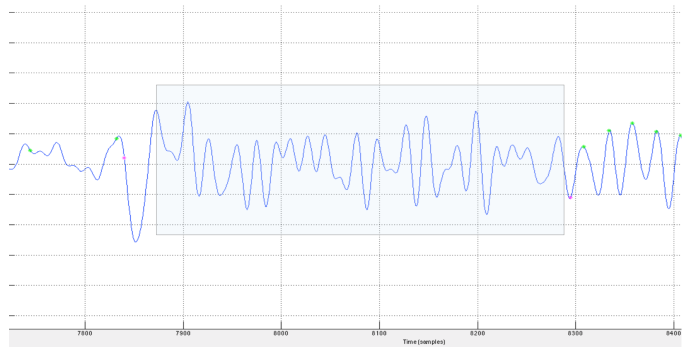

Fig. 3: Noisy region filtered with pass band 0.16Hz-1.5Hz. There is no residual periodic structure resembling breathing. Pairs of consecutive green dots indicate manual annotation of inhales. Red dots indicate automatically detected inhales.

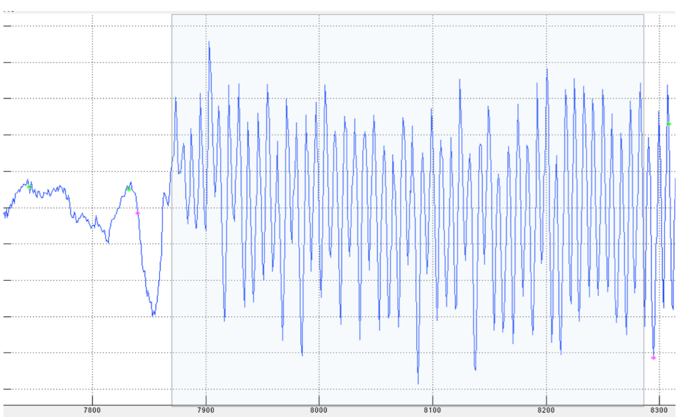

Fig. 2: Unfiltered signal with noisy region highlighted. Pairs of consecutive green dots indicate manual annotation of inhales. Red dots indicate automatically detected inhales.

- The IN-IN interval between the last peak and the candidate peak must not be less than 50% of the value of the previous IN-IN interval for the candidate peak to be scored as a peak. This value was chosen to exclude steps from being marked as breaths.
- If the IN-IN interval between the last peak and the candidate peak is more than 30% larger than the previous IN-IN interval, it should be accompanied by an increase in amplitude of the breath (indicating a deep breath) or should reflect the pattern consistent with talking to be scored.
- Sometimes a candidate peak appears to be a peak when a tight filter is used, but is identified as an artifact (usually due to talking coupled with steps, or the result of a baseline shift) when the filter is removed, or the pass band is loosened. These are not scored as peaks.

Note that: A longer than average breath that has a constant amplitude at the top after filtering out the noise, so that the sample to be marked as a peak is unclear, will be scored with the peak placed in the centre of the breath.

V. Metrics and Success Criteria

The overall desired performance is for the computed breathing rate to be within 5 bpm of the true breathing rate (determined by hand scoring), 95% of the time across all users. Put another way, there should be between 95% and 100% probability that the breathing rate is within 5 bpm given a random user at a random time during their run. More stringently, to ensure statistical significance of the conclusion, we want the upper 95% confidence limit of the 95th percentile of the absolute error to be less than or equal to 5. Note that it is unlikely that a bootstrap estimate (or any estimate) of the confidence bound will coincide with the true confidence bound given the limited number of files analyzed, and the fact that re-sampling should be done at least over files. However, this still provides a useful assessment of how sensitive the results may be to the specific sample used.

Secondary metrics of interest are the average sensitivity and PPV (positive predictive value) of the event detection algorithm. A TP (true positive) in this context is defined as any candidate inhale detected by the algorithm that is within 5 samples of a manually scored inhale. A FN (false negative) is any manually scored inhale that is not within 5 samples of a detected inhale. A FP (false positive) is any detected inhale that is not within 5 samples of a manually scored inhale.

Given these definitions sensitivity and PPV, as defined in (1) and (2), give measures of the detection algorithms propensity to over-detect and under-detect respectively.

$$sensitivity = TP/(TP + FN) \tag{1}$$

$$PPV = TP/(TP + FP) \tag{2}$$

The reported average sensitivities and PPVs are the average over files of the values from (1) and (2) calculated on the scores and detections from each file. Since different files can have different percentages of their use-able data scored, and the percent scored is correlated with the cleanliness of the data, averaging in this way serves to balance out this distortion.

VI. Results

The accuracy of the breathing rate is defined as the 95th percentile of the absolute difference between the breathing rate computed based on the IN-IN (inhale to inhale) intervals identified by the algorithm and the manually scored inhales. The prospectively desired success criterion is an accuracy of at most 5 bpm. Since breathing rates during running are expected to be between 30 and 60 BPM, this would correspond to an uncertainty interval encompassing one third of the range, which is likely acceptable for present purposes.

TABLE II has the summary performance statistics for each breakdown and several combinations of breakdowns. The confidence interval is computed by bootstrap estimation of the 97.5% confidence level with re-sampling over files. With the low number of files analyzed, the confidence bound is likely optimistic. All upper confidence bounds on accuracy are below 5 BPM. Moreover, all average sensitivities and PPVs are above 90% indicating good agreement between the detections of the algorithm and the manually scored breaths.

VII. Conclusion

There is no evidence that the current breathing rate algorithm is insufficient for men wearing OMshirts. The upper limit of the error confidence interval is below 5 BPM for all breakdowns that include men, which satisfies the success criterion.

There is no evidence that the current breathing rate algorithm is insufficient for women wearing OMbras. The upper limit of the error confidence interval is below 5 BPM for all breakdowns that include women, including the old bras data set and the new bras round 1 data set which includes the 3 ill-fitted bras. The success criterion is satisfied.

The only file which individually had an accuracy above 5 BPM was one of the ill-fitted bra files, with an accuracy of 6.5 BPM, and sensitivity and PPV both of 85%. Given the adverse conditions that the incorrect sizing creates, the fact that the other two files remain within the acceptance criterion, and the relatively small deviation from the single file that fails, demonstrates the robustness of the breathing rate algorithm.

Breakdown	Average Sensitivity	Average PPV	Accuracy (BPM)	CI+
Overall	94%	94%	2.1	2.7
Overall excl. NB 1	95%	95%	2.0	2.7
Men	96%	96%	1.5	1.8
Women	94%	94%	2.3	2.9
Women excl. NB 1	95%	95%	2.2	2.9
Women OB	93%	93%	2.1	2.8
Women NB	94%	94%	2.3	3.0
Women NB 1	93%	93%	2.6	4.8
Women NB 2	96%	96%	2.3	2.9

TABLE II: Overall performance statistics for all breakdowns and several combinations of breakdowns. The average for the sensitivity and PPV are taken over all files. The upper confidence bound is taken by bootstrap estimation with re-sampling over files, and represents the 97.5% confidence bound.