Electrohysterographic Detection of Uterine Contractions in Term Pregnancy

G.I. Bajlekov, C. Rabotti, S.G. Oei and M. Mischi

Abstract—Uterine-contraction detection is a fundamental component of pregnancy monitoring. Electrohysterography (EHG) provides a non-invasive and accurate alternative to intrauterine pressure (IUP) measurements, and several techniques provide an estimated IUP (eIUP) based on the EHG alone. Commonly, EHG contraction detection is based on amplitude thresholding of the eIUP. We aim at improving the reliability of contraction detection, such that automatic contraction detection can be realized. An algorithm for template-matching of the eIUP signal is proposed. This method is based on Bayesian evidence using a Gaussian likelihood function to classify uterine activity. Gaussian templates are matched to the input signal, with weights obtained empirically from manually-annotated contraction events in a training data-set. The results show an improvement in contraction detection accuracy compared to threshold-based methods. The template-matching method is adaptable to relevant features in the input training data, and is thus less sensitive to differences in eIUP derivation or measurement variability. The method allows for improved automatic uterine contraction detection in labor EHG data, while being extensible to e.g. preterm contraction detection.

I. INTRODUCTION

Uterine activity assessment is essential for evaluating labor progress, and can be used to predict preterm labor [1], [2]. Currently employed techniques for uterine activity monitoring include invasive intrauterine pressure (IUP) measurements and the use of an external tocodynamometer. While the first method is regarded as reliable, its invasive properties limit its application. External tocography is readily available, but is less reliable and often inconsistent due to the indirect mechanical nature of the strain gauge transducer used [2], [3], [4]. An alternative is electrohysterography (EHG), the measurement of electrophysiological signals resulting from the activity of uterine muscles. EHG provides uterine activity information which is more accurate and reliable than tocography [5], [6], [7] while maintaining the ease of use.

In order to assess the uterine activity from EHG measurements, the signal is processed to obtain a quantity correlating to the IUP [8], [9], [10]. In this estimated IUP (eIUP), uterine activity during contractions is selected according to various criteria [3], [8], [11], [7], which are then observed either manually or automatically. However, most of these criteria are based on amplitude thresholds analogous to those used for IUP assessment, not taking into account the differences in content between the invasive IUP and the derived eIUP signal [12]. As a result, these methods might be less suited for eIUP signals, and do not accurately capture information relevant to the detection process.

To overcome this problem, here we propose an approach for contraction detection that makes use of a template-matching algorithm. We aim for a more robust classification of uterine activity through adaptation to relevant features in the eIUP signal. A set of Gaussian templates with variable shape is compared to the eIUP signal based on a Gaussian likelihood function, resulting in new features indicating the similarity of the signal to each template shape. In a training process, each feature is evaluated in its ability to predict contractions in eIUP signals. Manually-annotated contraction events based on synchronized IUP signals are considered as ground truth. This training yields weights for each feature, used during the classification of novel measurements. The weights are applied to the same features obtained from the novel eIUP signal, resulting in a predictor for contractions in the novel measurement. By thresholding the resulting predictor, the contraction detection can be tuned for a trade-off between sensitivity and specificity.

II. METHODS

As proposed by Mubeen and Knuth in a different field [13], a template-matching method can be based on Bayesian evidence using a Gaussian likelihood function. Here, the likelihood of a template match given the input signal is expressed as the Gaussian of the mean squared difference between the two, compared to a dot product in convolution template-matching. This allows for a set of templates with varying amplitude, providing a non-linear relationship between the template amplitude and resulting likelihood.

Following this approach, we have developed a method for detection of uterine contractions in a derived eIUP signal derived from EHG analysis. Different from the original implementation by Mubeen and Knuth is the use of a training data-set from which the predictive ability of templates is obtained empirically. The training data-set consists of representative eIUP signals as well as manually annotated uterine contraction events, explained in chapter II-B.1. The weights from the training data were then used in classifying novel eIUP signals. Eventually, the resulting detection of uterine contractions was compared to an alternative method based on relative amplitude thresholding [8].

A. Template-matching

At the foundation of the template-matching algorithm are the templates. A family of Gaussian kernels, discretely
sampled at offset \( n \), have been used as templates. They were parameterized by their amplitude \( a \) and width \( \sigma \):

\[
T_{a,\sigma}[n] = a \cdot \exp \left( -\frac{(n - \frac{\text{max}}{2})^2}{\sigma^2} \right).
\]

The templates were limited to a size of \( \text{max} \) samples, within which the Gaussian peak was centered. In order for the template and signal indices to align properly, \( \text{max} \) was specified in which possible contraction events and their starts and ends were determined. In addition, margins were specified in which possible contraction events could be located.

The similarity between \( T_{a,\sigma} \) and a signal \( S \) at sample \( k \) was obtained by the similarity function \( f_{a,\sigma} \), with \( d_{k,n} \) being the squared difference between \( S \) and \( T_{a,\sigma} \) and a parameter \( \rho \) controlling the mismatch sensitivity:

\[
d_{k,n} = (S[n + k - \frac{\text{max}}{2}] - T_{a,\sigma}[n])^2,
\]

\[
f_{a,\sigma}[k] = \exp \left( -\frac{\rho}{a^2} \cdot \sum_{n=1}^{\text{max}} d_{k,n} \cdot W_{\sigma}[n] \right).
\]

\( W_{\sigma} \) is a normalized Gaussian kernel with width \( \sigma \):

\[
W_{\sigma}[n] = \frac{T_{1,\sigma}[n]}{\sum_{n=1}^{\text{max}} T_{1,\sigma}[n]},
\]

assigning a larger weight to the center of the template and reducing contributions of comparison to its tail.

This similarity \( f_{a,\sigma} \) represents the features of input signal \( S \) relevant for detecting shapes such as the peaks typical to the eIUP signal during contractions. Note that \( f_{a,\sigma} \) is a Gaussian-like function, such that a zero difference between \( S \) and \( T_{a,\sigma} \) \((d_{k,n} = 0)\) would result in \( f_{a,\sigma} = 1 \), and \( f_{a,\sigma} \to 0 \) for a mismatch between \( S \) and \( T_{a,\sigma} \).

1) Training: The proposed method consists of two phases: training and classification. During training, a data-set was used which in addition to an eIUP signal also contained annotated uterine contractions. The templates were evaluated in how well they match either contraction or baseline data, indicated by weights \( w_{a,\sigma} \) for \( \text{max} \) samples of training data:

\[
w_{a,\sigma} = \frac{\sum_{k=0}^{\text{max}} f_{a,\sigma}[k] \cdot c[k]}{\sum_{k=0}^{\text{max}} f_{a,\sigma}[k] \cdot (1-c[k]) + \sum_{k=0}^{\text{max}} f_{a,\sigma}[k] \cdot (1-c[k])},
\]

where \( f_{a,\sigma} \) is the similarity between a training data-set signal and \( T_{a,\sigma} \), and \( c[k] \) the probability of uterine activity obtained from the corresponding contraction annotations. Note that if the training data contains purely contraction \((c = 1)\) or baseline \((c = 0)\) signals, then \( w_{a,\sigma} \) is undefined.

2) Classification: Classification of a signal as either contraction or baseline data was performed through averaging all \( f_{a,\sigma} \) obtained from the classification data-set, weighted with \( w_{a,\sigma} \) from (5), providing a predictor \( p[k] \) of uterine activity for a novel eIUP signal:

\[
p[k] = \frac{1}{|P_a| \cdot |P_\sigma|} \sum_{a \in P_a} \sum_{\sigma \in P_\sigma} f_{a,\sigma}[k] \cdot w_{a,\sigma},
\]

where \( P_a \) and \( P_\sigma \) are the sets of parameters \( a \) and \( \sigma \), and \(|P_a|\) and \(|P_\sigma|\) the number of elements in each set.

**B. Validation**

The template-matching method was applied to a data-set containing eIUP signals obtained during labor. Manual annotations of uterine contractions were based on synchronized IUP recordings and were used as ground truth. Cross-validation was used to adjust the weights \( w_{a,\sigma} \) independently of the data being classified. These steps are described in detail below.

1) Data: The data used for the validation were obtained from the measurements described in [9], consisting of simultaneous EHG and IUP measurements during labor at term (9 patients). The EHG data were processed as described by [10], obtaining an eIUP signal by calculation of the Teager energy (TE) in both the horizontal and vertical direction sampled at 20 Hz. The eIUP data of only seven out of nine patients were used. The data from two patients were discarded due to artifacts in the EHG or IUP measurements.

From this data-set, ranges of equal time length were selected at random for each patient. This length, 1529 s, was equal to that of the shortest data-set, thus maximizing the amount of available data. All consequent analyses were performed on this selection. In addition, the eIUP data were normalized such that 10% of the data exceeded a value of 1. This was done to ensure the input is in a consistent range, while disregarding outliers. The time delay of the IUP signal compared to the EHG [9] was accounted for by observing the offset maximizing the cross-correlation between the IUP and eIUP signals. This offset was determined per patient, and the eIUP data was corrected for the median offset across all patients.

Uterine activity annotations for the data-set were obtained from visual analysis of the IUP. Three expert observers independently assessed and annotated the IUP signal, indicating start and end times of contraction events. In addition, margins were specified in which possible contraction events...
extended beyond the recording region, these margins were not considered in further analyses. The resulting annotations are illustrated in fig. 1. These annotations served as a ground truth for the presence of uterine activity and were aligned in time with the eIUP signals.

The annotations from the three observers were used in two ways. During training, the probability of uterine contraction events \( c[k] \) was calculated proportionally to the number of observers indicating a contraction at each sample \( k \). This probability was also used in evaluating the sample-wise classification performance. In addition, for the evaluation of complete contraction events, only the segments with full agreement among the three observers were used. No information regarding contraction events was used during classification of novel eIUP signals.

2) Assessment: The classification performance was cross-validated such that repeatedly \( n-1 \) patient data-sets (the training set) were used to obtain weights \( w_{a,\sigma} \), which related template similarity to contraction probability. The remaining patient data-set (the test set) was assessed using these weights, obtaining a scalar predictor \( p \). Predictors were obtained for each patient, with their data used in turn as the test set, and were compared to a global threshold to classify data between contraction and baseline signal. We obtained a receiver operating characteristic (ROC) curve that relates the rate of true positive classifications to the rate of false positive ones dependent on the threshold parameter.

All evaluated measures were compared to results of the method described by Jezewski [8]. This method calculates a threshold relative to an adaptively adjusted eIUP baseline. Contractions are detected when the eIUP amplitude exceeds this threshold for a minimum duration. This method was used as a reference for the performance of current contraction detection algorithms.

3) Parameters: The parameters used in the template-matching algorithm were selected to match the evaluated data. During contraction events in the recorded data an amplitude around \( a = 1 \) was expected due to the normalization of the eIUP signal, while \( \sigma = 1000 \text{ (50 s)} \) corresponded to a kernel width closely matching that of the average observed peaks during contractions. The parameters were centered around their expected values, while allowing for a wide range to capture features with a negative predictive ability as well. This led to a set of templates consisting of 25 amplitudes on a logarithmic scale with \( a \in \{0.1, \ldots, 10\} \), and 25 widths on a logarithmic scale with \( \sigma \in \{100, \ldots, 10000\} \). The remaining parameters were adjusted to enhance the area under the ROC curve: \( \rho = 10, \sigma' = 3 \cdot \sigma \). However, these parameters were not exhaustively optimized. For computational efficiency the kernel size was limited to \( n_{max} = 10000 \text{ (500 s)} \).

III. RESULTS

A. Contraction Annotations

In order to evaluate the quality of the obtained ground truth for uterine activity, we compared the annotations provided by each of the three observers. Out of 70 annotated contraction events, 6 events were not annotated by at least one observer.

5 of these due to the contraction partly exceeding the boundary of the recorded signal (Fig. 1). In two cases, a segment of the signal was annotated as a single event or two independent events by different observers. The agreement between the contraction event annotations of the independent observers was high (\( \alpha = 0.999 \)) according to Krippendorf’s \( \alpha \) coefficient [14].

B. Contraction Classification

The proposed method was assessed by the sample-wise classification performance of the obtained predictor \( p \). Varying the threshold of \( p \) and comparing the output to the manual annotations yields an ROC curve indicating the sensitivity–specificity trade-off, illustrated in fig. 2. Due to a limited choice of parameters of the reference method, the corresponding ROC curve could only be sparsely sampled. Additionally, the inter-subject variability was accounted for through bootstrapping [15], where subsets of the data randomized at the patient level were analyzed repeatedly in order to obtain a confidence region from the multiple observations. For the bootstrapping procedure and for all following results, the threshold was fixed to a value corresponding to a false positive rate of 5.09 %, equivalent to that observed by Jezewski’s reference method with default parameters. The template-matching method has a significantly higher true positive rate compared to the reference method at the same false positive rate (Template: Sensitivity 58.4 % (95 % CI: 54.3 %–62.8 %), Specificity 94.9 % (95 % CI: 91.7 %–98.0 %), Jezewski: Sensitivity 45.2 % (95 % CI: 38.4 %–51.2 %), Specificity 94.9 % (95 % CI: 92.0 %–97.5 %)).

IV. DISCUSSION

A. Result Relevance

The results from section III-B show the classification performance of the template-matching algorithm. Sample-wise classification is relevant for extracting parts of the
signal containing uterine activity for further analysis. In such scenarios, it is often important to have a high certainty of the selected samples being true positives, while there is no specific demand to select all uterine activity samples present in the recorded signal. Therefore, it is logical to prioritize the algorithm specificity, as was done for the presented results. Both the template and the reference method were analyzed at a false positive rate of 5.06% (set equal to that observed with Jezewski’s method). In this comparison, the novel algorithm attained a significantly better classification as seen in fig. 2. Furthermore, the sensitivity–specificity trade-off for the template-matching algorithm can be more finely tuned due to the continuous predictor $p$.

Based on these results, the proposed method seems beneficial for contraction detection in term labor. Its adaptable nature should make it suitable to uterine activity detection in different settings, such as preterm labor, where important features in the measured signals are less robust. Literature suggests that useful information for prediction of preterm delivery can be obtained from recorded EHG signals during uterine contractions [16], [17]. Due to changes in the maternal physiology during pregnancy [2], uterine activity in earlier stages is less pronounced and less regular. An improved method for automatic contraction detection during pregnancy will be valuable, allowing for automated continuous monitoring.

### V. Conclusion

The current work illustrates the advantage of the newly developed template-matching algorithm for uterine contraction detection in term labor. This method consistently outperforms previously employed methods. Additionally, its flexibility in the selection of relevant features makes it more suitable to a wide range of problems with less well-defined detection criteria.

### References


