**ABSTRACT**

The objectives of this study were (1) to develop an algorithm for the acceleration sensor of the Smartbow Eartag (Smartbow GmbH, Weibern, Austria) to distinguish between postures (lying and standing or locomotion) and to detect 6 kinds of activities (milk intake, water intake, solid feed intake, ruminating, licking or sucking without milk intake, and other activities) in dairy calves and (2) to evaluate this sensor for identifying these behaviors in dairy calves compared with observations from video. Accelerometers were applied to the left ears of 15 preweaned Holstein dairy calves. Calves were kept in a group pen and received milk replacer from an automatic calf feeder. Based on 38 h of acceleration data and video observation, an algorithm was established to detect the predefined behaviors. Using cross-validation, video recordings were used to analyze whether a behavior was detected correctly by the developed algorithm. For posture, sensitivity (94.4%), specificity (94.3%), precision (95.8%), and accuracy (94.3%) were high. Cohen’s kappa was calculated as 0.88. For the 6 defined activities, overall (i.e., aggregated for all activities) accuracy was 70.8% and kappa was calculated as 0.58. Some activities (e.g., ruminating, feed intake, other activities) were identified better than others. In conclusion, the developed algorithm based on the acceleration data of the Smartbow Eartag was successful in detecting lying behavior, ruminating, feed intake, and other activities in calves, but further development of the underlying algorithm will be necessary to produce reliable results for milk and water intake. **Key words:** accelerometer, evaluation, calf, behavior

**Technical Note**

During the last 40 yr, different types of accelerometers have been developed to detect certain behaviors in cows (Rutten et al., 2013). Previous work has evaluated the use of sensors in calves to monitor the following behaviors: lying (Bonk et al., 2013; Swartz et al., 2016), activity (Hill et al., 2017), locomotion (de Passillé et al., 2010; Swartz et al., 2016), eating (Hill et al., 2017), and rumination (Burfeind et al., 2011; Hill et al., 2017). Most devices, however, have been evaluated for a single parameter only (de Passillé et al., 2010; Burfeind et al., 2011; Bonk et al., 2013; Swartz et al., 2016). Systems for automated monitoring of feeding, drinking, and ruminating behaviors in calves have rarely been evaluated (Hill et al., 2017) and have not been implemented in practice so far, although these behavior data provide essential information regarding the animals’ health and growth (Appleby et al., 2001; Miller-Cushon and DeVries, 2015) and could complement information on daily feed and milk intake. Deviations detected by a monitoring system can help identify diseased calves and allow early intervention. This should contribute to a shorter duration of the disease, improved cure rates, and reduced negative effects on growth rates and weight gain.

The Smartbow acceleration sensor (Smartbow GmbH, Weibern, Austria) has been evaluated for real-time localization (Wolfgar et al., 2017), heat detection, rumination (Borchers et al., 2016; Reiter et al., 2018), and parturition (Krieger et al., 2017) in dairy cows. In calves, Breitenberger et al. (2015) described the general feasibility of detecting milk intake from a calf feeder by using Smartbow acceleration data but pointed out that further work is necessary to improve drinking recognition. Recently, Roland et al. (2018) conducted a pilot study to evaluate the Smartbow accelerometer for detecting drinking events in bucket-fed dairy calves.
Comprehensive systems designed to monitor more than one parameter—in particular, how calves budget their time among different activities—are still rare.

The aims of this study were (1) to generate an initial algorithm used by Smartbow GmbH for developing a comprehensive monitoring system (including proprietary commercial algorithms) that enables the differentiation between posture (lying and standing or locomotion) and detects 6 activities (milk intake, water intake, solid feed intake, ruminating, licking or sucking without milk intake, and other activities) based on the acceleration data of the Smartbow Eartag and (2) to perform an evaluation of the algorithm for detecting these postures and activities by comparison with video observations. The study was approved by the institutional ethics committee of the University of Veterinary Medicine, Vienna, Austria (ETK-03/09/2015), and by the State Office of Agriculture, Food Safety and Fisheries Mecklenburg-Vorpommern, Germany (7221.3-2-028/15).

The practical part of the study was performed on a commercial dairy farm in Mecklenburg-Vorpommern, Germany. Fifteen female Holstein-Friesian calves (39 ± 8 d of age; mean ± SD) were equipped with Smartbow sensors. Calves were housed in a stable of group pens with straw bedding (7.90 × 3.90 m). All 15 calves were kept in the same pen during the study. Milk replacer was fed from an automatic feeder (Calf Boy/Calf Master Kombi, Westfalia Landtechnik GmbH, Oelde, Germany, and Förster-Technik GmbH, Engen, Germany). Calves were allowed to retrieve 1/12 of the daily maximum amount of 9 L every 2 h.

The Smartbow Eartags consisting of an accelerometer and a transceiver were attached to the left ear 2 wk before the beginning of the study. The sensors continuously collected 10 triaxial acceleration values per second (10 Hz). The data were sent to wall-mounted receivers (Smartbow Indoor Receiver), which were connected to the local server (Smartbow Farm Server) on the farm where they were processed (Smartbow GmbH, 2018).

Three infrared video cameras recorded calf behavior during the data collection period. The group pen was covered by 2 cameras (Hikvision VR IR Bullet Network Camera, Hangzhou, China), and another camera provided a close-up view of the calf feeder. For the purpose of this study, videos recorded between 0700 and 1100 h were used because calves were fairly active during this period and therefore a greater variety of behaviors was more likely to be captured. Thus, the behavior of each calf was video classified for 4 h, resulting in 60 h of classification data. Videos were examined with Mangold Interact software (Mangold International, Arnstorf, Germany) by the same observer. Nonidentifiable behavior accounted for 36.2% of the observation time and was removed from the data set, resulting in 38 h 7 min of usable observation time in total (2 h 32 min ± 40 min/calf, mean ± SD; range: 1 h 6 min to 3 h 35 min/calf). Situations with nonidentifiable behavior (1 h 27 min ± 40 min/calf; range: 25 min to 2 h 54 min/calf) involved the observed calf not always being in clear view (e.g., the observed calf was blocked by other calves or was in an unsuitable position, such as rear toward camera). The behavior was excluded if it could not be classified unequivocally by the observer. It was considered better to have less but more reliable data than more data with potential observation errors or based on guessing. Two postures (lying and standing or locomotion) and 6 activities (milk intake, water intake, solid feed intake, ruminating, licking or sucking without milk intake, and other activities) were differentiated. A definition of each activity is provided in Table 1. Licking and sucking behavior without milk intake was included in the study to train the algorithm to differentiate between licking or sucking with and without milk or water intake.

To gain information about the reliability of video analyses, video observations were compared with direct visual observations. For this, 2 observers conducted 25 h of direct visual observation (1–2 h/calf) in the stable. Intra- and interobserver reliabilities were calculated with Mathematica (version 11.0; Wolfram Research Inc., Champaign, IL) using Cohen’s kappa (κ; Cohen, 1960) based on 0.1-s observation intervals. For direct versus video observation (n = 3 calves, 3 h), κ was 0.98 for posture and 0.90 for the 6 observed activities. Interobserver reliability (n = 7 calves, 8 h) for the direct observation was found to be 0.94 for both posture and activities. Intraobserver reliability for the video analysis was determined by letting one observer analyze the same video sequence (n = 1 calf, 4 h) twice and was calculated with a κ of 0.99 for posture and 0.94 for activities (Supplemental Table S1; https://doi.org/10.3168/jds.2018-14720).

For algorithm development, data were split 15 ways for a cross-validation approach (schema presented in Figure 1). In every split, data sets of 14 calves were considered as training sets for which the parameters or evaluation metrics or both were calculated and optimized. The remaining data set was used as the validation set on which the final procedure was evaluated. The results were summed to form a total confusion matrix. The algorithm consisted of 2 parts: the first part classified posture, and the second part classified the activities considering the estimated posture.

To classify posture, an a priori assumption that dispersion measures of the acceleration data were smaller during lying phases than during standing or locomotion.
phases was made. Every acceleration data set was split into 1-min segments, and each segment was considered to belong to the most frequent behavior during this time period. According to our video analysis, the average lying bout lasted about 33 min and lying bouts rarely lasted less than 1 min. Three parameters—dispersion measure, threshold discrimination lying from locomotion, and a lower time threshold to eliminate short lying intervals—were optimized to maximize accuracy for the whole training set. For the first parameter, different measures of dispersion (standard deviation, median absolute deviation, median deviation of the median, interquartile range) were defined. For every measure and for every day of the training set, the mean of the respective measure was calculated for all intervals. If the respective measure of an interval was above the threshold, which was calculated by multiplying the corresponding mean with the second parameter, it was considered as a standing or locomotion phase. If the respective measure was below this threshold, it was considered as a lying phase. A third parameter eliminated lying intervals below a certain time threshold and assigned them the state “standing or locomotion.”

To perform activity classification, a hidden Markov model (Rabiner, 1990) algorithm was used as a machine learning scheme. It was assumed that the frequency of an activity depends on the posture expressed during that time (e.g., in this study, water and milk intake was most unlikely to occur in a lying position). Each data set was split into 1-s intervals, and 2 discrete hidden Markov models (1 for lying and 1 for standing or locomotion) were built. The parameters of the hidden Markov model—namely, the emission probabilities, transition probabilities, and starting probabilities—were estimated with an expectation maximization algorithm using all sequences of true states in the training set during lying phases or standing phases, respectively. For emissions, the same set of dispersion measures as described above was chosen. The values were discretized by binning (i.e., segmentation of the values into smaller groups according to their value) with a certain fixed number of elements in each bin (except for the last bin). The combination of bin size and feature, which maximized the sum of the Hellinger distances between the empirical distribution of every pair of activities using this binning, was chosen to act as emissions for the current model.

To validate the agreement between the sensor data incorporating the algorithm and video analysis, the timing and duration of each behavior were compared using 0.1-s intervals. Based on the timing of an observation, the duration and percentage of true positive (TP; posture or activity identified correctly by the algorithm), true negative (TN; correctly predicted time period of no posture or activity), false positive (FP; falsely identified time period of a posture or activity), and false negative (FN; falsely identified time period of no posture or activity) events and the following test characteristics were calculated as follows.

- Sensitivity, calculated as sensitivity = TP/(TP + FN) (Lundorff Jensen and Kjelgaard-Hansen,

**Table 1.** Descriptive statistics of video analysis data based on 4 h of observation per calf

<table>
<thead>
<tr>
<th>Posture or activity</th>
<th>Mean observation time (h:min:s)</th>
<th>SD (h:min:s)</th>
<th>Mean duration of episode (h:min:s)</th>
<th>Observation time (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lying</td>
<td>02:19:48</td>
<td>00:24:17</td>
<td>00:32:46</td>
<td>57.9</td>
</tr>
<tr>
<td>Standing</td>
<td>01:34:25</td>
<td>00:21:44</td>
<td>00:00:48</td>
<td>39.1</td>
</tr>
<tr>
<td>Locomotion</td>
<td>00:07:03</td>
<td>00:02:55</td>
<td>00:00:04</td>
<td>2.9</td>
</tr>
<tr>
<td>Feed intake</td>
<td>00:35:54</td>
<td>00:12:15</td>
<td>00:01:22</td>
<td>14.8 (23.3)</td>
</tr>
<tr>
<td>Water intake</td>
<td>00:00:45</td>
<td>00:00:57</td>
<td>00:00:15</td>
<td>0.3 (0.5)</td>
</tr>
<tr>
<td>Milk intake</td>
<td>00:03:21</td>
<td>00:01:19</td>
<td>00:00:39</td>
<td>1.4 (2.2)</td>
</tr>
<tr>
<td>Ruminating</td>
<td>00:26:25</td>
<td>00:20:02</td>
<td>00:02:47</td>
<td>10.9 (17.1)</td>
</tr>
<tr>
<td>Licking/sucking</td>
<td>00:12:20</td>
<td>00:05:12</td>
<td>00:00:11</td>
<td>5.1 (8.0)</td>
</tr>
<tr>
<td>Other activities</td>
<td>01:15:30</td>
<td>00:26:36</td>
<td>00:00:54</td>
<td>31.2 (48.9)</td>
</tr>
<tr>
<td>Not identifiable</td>
<td>01:27:42</td>
<td>00:40:22</td>
<td>00:01:31</td>
<td>36.2</td>
</tr>
</tbody>
</table>

1Including nonidentifiable behavior of 1 h 27 min ± 40 min (mean ± SD) per calf.
2Percentage of observation time after excluding nonidentifiable activities is shown in parentheses.
3Uptake of hay, straw, or concentrate; feed intake or chewing movements visible on video.
4Calf is standing with muzzle in drinking trough; swallowing movements visible on video.
5Calf is standing in calf feeder sucking at the teat; swallowing movements visible on video.
6Calf is lying down; regular regurgitating and chewing movements (without feed intake) visible on video.
7Calf is licking or sucking at objects (e.g., teat, fence), other calves’ bodies, or its own body.
8Calf is engaged in activities other than those described in footnotes 1 and 3–7.
Figure 1. Schema illustrating the steps in algorithm development and quality assessment by use of a 15-fold cross-validation approach. The data set of 1 calf was excluded and used as a validation set, whereas the other 14 data sets were used as a training set. This procedure of splitting the data into training and validation sets followed by algorithm development was repeated 15 times, such that every calf’s data set was used as an excluded validation set once. For each of the 15 repetitions, the quality of the algorithm was evaluated by applying the determined model on the excluded validation set. Color version available online.
Classification of \( \kappa \)-values was made according to McHugh (2012), with \( \leq 0 \) as no agreement, 0.01 to 0.20 as slight agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial agreement, and 0.81 to 1.00 as almost perfect agreement.

To our knowledge, this is the first study that has evaluated the use of the Smartbow accelerometer for monitoring behaviors in dairy calves and one of the first studies to evaluate an accelerometer for more than one behavior in calves. Statistics describing the time budget of the calves observed in this study based on video analysis are presented in Table 1. The percentage of lying time (58%) was lower than reported by other authors (Morita et al., 1999; Chua et al., 2002; Hämmänen et al., 2005), but a time of day when calves were active and exhibited a large variety of behaviors was chosen intentionally. The percentage of milk intake time (1.4%) was comparable with the time calves occupied an automated feeding system in a study by Morita et al. (1999). The average duration of a milk drinking bout (39 s) was shorter than that reported in other studies (Veissier et al., 2002; Hill et al., 2017) because videos were classified on a 0.1-s basis in this study and thus captured short breaks from sucking at the teat, whereas other studies were based on observation intervals \( \geq 1 \) min. Water intake accounted for 0.3% of the time budget, which coincides with other studies during which researches rarely observed water intake (Hill et al., 2017). Rumination time (10.9%) was slightly lower than that reported by Burfeind et al. (2011); however, their study included calves that were older than the ones used in this study.

The results implicate that the Smartbow sensor reached a satisfying accuracy in detecting posture (i.e., lying vs. standing or locomotion). Sensitivity (94.4%), specificity (94.3%), precision (95.8%), and accuracy (94.3%) were high for detecting posture. Cohen’s kappa was calculated as 0.88, which indicates an almost perfect agreement. These results are similar to those reported by Bonk et al. (2013) and Swartz et al. (2016), who found high correlations between data logger and direct or video observation for lying behavior in calves. Discrimination between standing and locomotion was not yet feasible because the intervals during which calves were engaged in locomotion were too short (i.e., the mean duration of single locomotion episodes was 4 s, and episodes rarely lasted longer than 10 s) to be detected by the algorithm. de Passillé et al. (2010) were able to identify the gait patterns of calves with accelerometers under experimental conditions (sensors were taped to the hind legs and calves were encouraged to move in a large arena). The identification of locomotion and gait pattern under field conditions should be the subject of further research and could be supported by Smartbow’s localization function, which allows the real-time determination of an animal’s position in a barn.

The results for the 6 observed activities were more diverse and are presented in Table 2. Accuracy ranged between 81.2% (feed intake) and 99.4% (water intake). Total accuracy (i.e., correctly identified outcomes for all activities combined) was calculated as 70.8%; individual accuracies might be not meaningful for some activities (e.g., water intake) with an unbalanced data set. Sensitivity ranged between 2.7% (water intake) and 89.4% (ruminating), and specificity ranged between 83.7% (feed intake) and 99.9% (water intake). Because the data set was highly imbalanced in an absolute and relative sense (Parvin et al., 2013), the validity of these specificities and accuracies must be regarded as limited. Precision ranged from 10.0% (water intake) to 88.2% (other activities). Overall \( \kappa \) was calculated as 0.58, which indicates a moderate agreement between sensor data and video observation. The individual \( \kappa \)-values for each activity indicate slight to no agreement for water and milk intake, fair agreement for licking, moderate agreement for feed intake, and substantial agreement for rumination and other activities. These results indicate that some activities (i.e., ruminating, feed intake, other activities) were identified more precisely and exhibited a higher agreement than others (i.e., milk and water intake), presumably due to the relative rareness of liquid intake. Overall, calves spent only 0.3% of the
observation time drinking water, which considerably restricts the validity of this result. Although water intake made up only a small percentage of all activities, it was included in the algorithm development (as opposed to locomotion, which was combined with “standing” into an upright posture) because water intake was considered more important with regard to a calf’s overall health status than locomotion. Future studies should validate the algorithm for a more complex data set, generated by a higher number of animals and a longer observation time. Sensitivity and precision for detecting water and milk intake should be improved in particular.

Additional research should assess whether deviations in behavior detected by accelerometers can be used to identify disease or discomfort in dairy calves. Lying time, feeding, and rumination were identified reliably with the Smartbow sensor. This information can be used to set up a system for monitoring calves’ health status, which should be complemented with other parameters once they can be identified more accurately.

Previous research validating data loggers for calf behaviors did not involve accelerometers (Burfeind et al., 2011) or evaluate different parameters (Breitenberger et al., 2015; Hill et al., 2017). Hill et al. (2017) but came to similar conclusions in their evaluation of using an ear-attached accelerometer (SensOor, CowManager BV, Harmelen, the Netherlands) to record activities in calves. These studies reported the sensor to be a suitable tool for detecting eating, rumination, and inactivity but not for identifying drinking behavior. Bikker et al. (2014) evaluated the SensOor for dairy cows and calculated a κ-value of 0.77 for eating. The lower κ-value for feed intake reported in this study might be attributable to a higher degree of activity and a lower degree of standardization and repetition in the movement patterns of calves as well as a smaller jaw movement range compared with adult animals. In conclusion, the Smartbow accelerometer incorporating the algorithm developed in this study was successful in detecting lying behavior, rumination, feed intake, and other activities in group-housed dairy calves, but the results for milk and water intake were not reliable and require further adaptations.

ACKNOWLEDGMENTS

This work was supported by the COMET-K2 Center for Symbiotic Mechatronics of the Linz Center of Mechatronics (Linz, Austria) funded by the Austrian federal government and the federal state of Upper Austria. We thank the staff from Rinderzucht Peters KG (Körchow, Germany) for their technical assistance. The authors received no financial support from Smartbow GmbH (Weibern, Austria) for the research, authorship, or publication of this article.

REFERENCES


Table 2. Test characteristics of sensor data compared with video analysis for 6 activities as reference

<table>
<thead>
<tr>
<th>Test characteristic</th>
<th>Feed intake</th>
<th>Water intake</th>
<th>Milk intake</th>
<th>Ruminating</th>
<th>Licking/sucking</th>
<th>Other</th>
<th>Total (^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity (%)</td>
<td>73.2</td>
<td>2.7</td>
<td>4.2</td>
<td>89.4</td>
<td>42.3</td>
<td>71.6</td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>83.7</td>
<td>99.9</td>
<td>99.7</td>
<td>94.9</td>
<td>91.8</td>
<td>71.6</td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>81.2</td>
<td>99.4</td>
<td>97.6</td>
<td>93.9</td>
<td>87.7</td>
<td>81.6</td>
<td>70.8 (^2)</td>
</tr>
<tr>
<td>Precision (%)</td>
<td>57.9</td>
<td>10.0</td>
<td>27.1</td>
<td>78.5</td>
<td>30.2</td>
<td>88.2</td>
<td></td>
</tr>
<tr>
<td>F1 score (^3) (%)</td>
<td>64.7</td>
<td>4.2</td>
<td>7.3</td>
<td>83.6</td>
<td>34.7</td>
<td>79.0</td>
<td></td>
</tr>
<tr>
<td>Cohen’s kappa</td>
<td>0.52</td>
<td>0.04</td>
<td>0.07</td>
<td>0.80</td>
<td>0.28</td>
<td>0.63</td>
<td>0.58 (^4)</td>
</tr>
</tbody>
</table>

\(^1\) Additional performance measures, which are typically used to assess overall performance in a single value.
\(^2\) Percentage of correctly identified outcomes for all activities combined.
\(^3\) F1 score = TP/[TP + (FP + FN)/2]; TP = true positive events, FP = false positive events, FN = false negative events.
\(^4\) Kappa for all activities combined.


