Using image analysis and fast Fourier transform to predict respiration rate in unrestrained dairy cows

R. R. Mantovani[1] and J. R. R. Dórea[1,2,*]
[1]*Department of Animal and Dairy Sciences, University of Wisconsin-Madison, WI 53706, USA*
[2]*Department of Biological Systems Engineering, University of Wisconsin-Madison, WI 53706, USA*
[*]Corresponding author: João R. R. Dórea, joao.dorea@wisc.edu

Abstract

Respiration rate (RR) is a critical trait associated with animal physiology. It is commonly used to indicate heat stress, respiratory diseases, and welfare for dairy cattle and other animal species. However, this trait is difficult to collect in large-scale operations for both research and commercial settings. This study aimed to develop a computer vision system that accurately predicts the RR of lying Holstein cows using RGB (red, green, blue) and infrared images. Ninety-five videos of thirty lactating cows were collected and had a bounding box annotated over the lying animals' flank areas (region of interest; ROI). An image-processing pipeline was developed to capture the pixel intensity variation over the ROI and use it to predict the RR. This process utilized Fast Fourier Transform (FFT) to capture the original signal's frequency domain and select only the frequencies with the five highest power spectral densities. The inverse FFT was then performed on the data, and its peak count was used as the predicted RR. The Root Mean Squared Error of Prediction (RMSEP) and R^2 were 8.3 breaths/min, 15.8% (RMSEP/Mean), and 0.77, respectively. Applying FFT to the pixel intensity signals from RGB and infrared images was an accurate method to compute the RR of cows in unrestrained conditions.

Keywords: respiration rate, computer vision, spectral analysis, signal processing

Introduction

Respiration rate is a critical trait associated with animal physiology. It is commonly used as an indicator of heat stress, respiratory diseases, and welfare for dairy cattle (Gaughan et al., 2000). So far, the most traditional form of calculating the RR has been the visual observation of the cows' flank area movements. However, this process is labor intensive, requires specialized training, and has limited scalability, making it unsuitable for large-scale farm operations (Handa and Peschel, 2022). For this reason, many studies have developed automated technologies to assess RR in livestock, such as contact-based sensors (Eigenberg et al., 2002), thermal imaging from nostrils (Milan et al., 2016), and RGB (red, green, blue) imaging of abdominal movements (Wu et al., 2020). When comparing these methods, video-based approaches have shown a few advantages: lower cost, higher scalability, lower susceptibility to physical damage, and less stress caused to cows (Handa and Peschel, 2022). Furthermore, Wiede et al. (2017) demonstrated that using Fast Fourier Transform to analyze the average pixel intensity variation over the abdominal area of breathing humans effectively calculated their RR. However, the computer vision systems proposed for humans and cattle usually restrain the animals in order to reduce the noise and bad image quality captured in uncontrolled environments.

In view of the technologies presented above, this study aimed to apply FFT to the average pixel intensity of RGB and infrared videos of unrestrained lying dairy cows to assess their respiration rates. The goal was to develop a simple yet robust model to predict their RR through image analysis. For this reason, the lying position was chosen because it makes the respiration movements more visible and is less susceptible to random cattle movement than standing positions. The proposed method can be valuable for the automatic detection of lying cattle's RR in large-scale farming, contributing to the identification of cows under heat stress and with abnormal respiratory behaviors.

Materials and methods

Experimental data

The videos for this study were collected in July 2021 (from 6 pm to 6 am on days 8-9, 15-16, and 22-23) at the Dairy Cattle Center (DCC) of the University of Wisconsin-Madison. They were recorded using an Amcrest ASH42-W camera with a frame rate of 30 fps and a resolution of 2,560 pixels (horizontal) x 1,440 pixels (vertical). Each camera was positioned approximately 2 meters from the ground with an angle view and captured 1-4 resting Holstein cows. A total of 95; 30-second video segments were obtained, accounting for 193 observations. The ground truth (observed data) was collected by visually counting the RR and transforming its unit to breaths/minute.

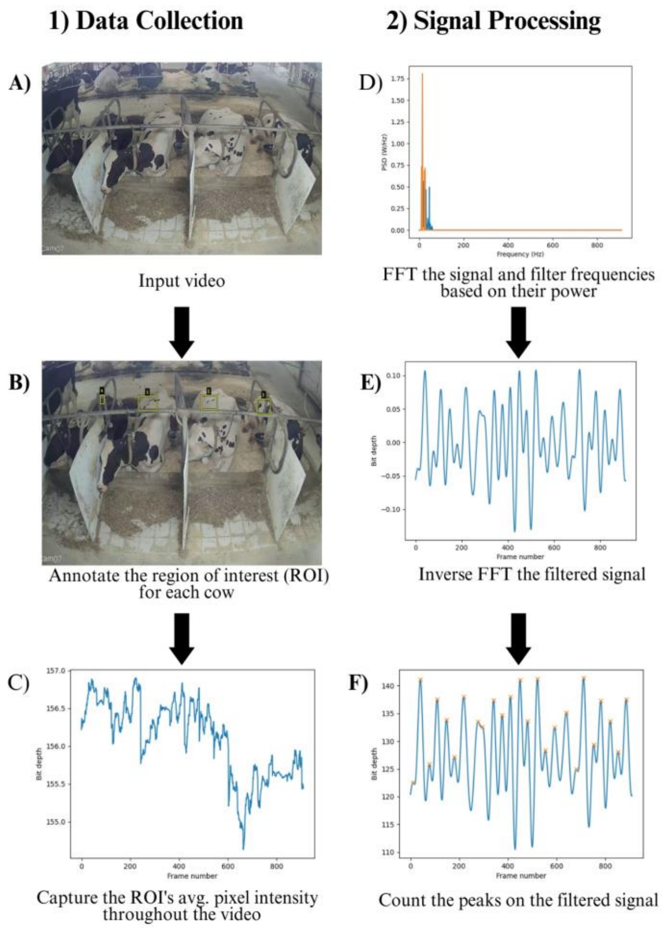

Figure 1: Summary of the proposed method. Example of an image captured for cows (A) and the annotations made (B). Average pixel intensity variation over the annotated region of interest (C). Power spectral density of each frequency in the data after performing Fast Fourier Transform (FFT), with the selected frequencies in orange (D). Denoised signal after filtering (E) and the peak count performed to calculate the respiration rate (F).

Image and signal processing

After collecting the videos, one frame from each recording was extracted. These were then exported to VGG Image Annotator, through which a rectangular bounding box was annotated over the part of the cows' flank areas (region of interest; ROI) where respiration could be observed. The videos were imported into Python using the cv2 library, and the annotated ROIs were captured throughout their *n* frames. Then, the average pixel intensity for each frame was calculated by extracting the mean pixel intensity of each image channel (R, G, B) per frame, adding these three values, and then dividing the result by three. Figure 1 shows a summary of the processing pipeline.

Fast Fourier Transform was applied to get the frequency domain of the pixel intensity's 2D signal (bit depth/frame). Two pre-processing steps were employed to clean the signal noise before computing the final RR. First, since healthy and severely heat-stressed cows' respiration rates usually lay between 26 and 120 breaths/minute (Becker et al., 2020), only the frequencies (Hz) between one-third and twice the video length were selected. Secondly, considering the power spectral density (PSD, W/Hz) outlines the most prominent frequencies in a signal, only the frequencies with the top 5 PSD values within the limited dataset were selected. Finally, Inverse Fast Fourier Transform was performed on the cleaned data to return it to its original unit (bit depth/frame), and the SciPy library was used to count the peaks on the resulting signal. The peak count was the predicted respiration rate (breaths/minute).

We split our dataset into three different training and testing groups in order to define the optimal set of the top *n* PSD values to select the respiration-related frequencies. In each split, the training set had all the observations from 20 cows, and the testing set had all the observations from the remaining 10 cows. To test the potential *n* values, the top *n* PSD values were considered hyperparameters, and we searched for $n = 50$, 25, 13, 5, 3, 1, where $n = 50$, for example, means the top 50 values of the PSD. The top *n* PSD values were tested for each training set, and the RMSEP and R^2 were calculated for each training set. The *n* value that yielded the best results (lower RMSEP and higher R^2) was then validated in the testing set.

Model assessment

After collecting the predicted respiration rates, z-scores were calculated to evaluate if there were any significantly outlying results within the data. By doing so, one video appeared to be an outlier, with a z-score greater than 5 ($p < .00001$). This video was removed from the model's analyses because, after rewatching it, it became evident that the atypical error was due to a bright light beam flashed in the camera over the ROI of the lying cows. For this reason, the dataset for assessing the model's performance comprised 94 videos and 191 observations. The metrics used for its evaluation were the coefficient of determination (R^2) and the Root Mean Squared Error of Prediction (RMSEP).

Results and discussion

After using training and testing sets to analyze how many top *n* PSD values must be selected to capture the RR most accurately, it was determined that the top 5 values yielded the best predictions across all different testing sets. These results can be seen in Table 1, where it is evident that the top 5 PSD values had the highest R^2 and lowest RMSEP for each dataset. Hence, the data was filtered, and only the five most prominent frequencies were selected to compose the cleaned signal, as shown in Figure 2.

Table 1: Model performance metrics for every top n PSD value tested in each different testing and training set

	Top 50		Top 25		Top 13		Top 5		Top 3		Top 1	
	R^2	RMSEP[a]	R^2	RMSEP[a]	R^2	RMSEP[a]	R^2	RMSEP[a]	R^2	RMSEP[a]	R^2	RMSEP[a]
Train[ab]1 (n[c] = 59)	.36	21.1	.39	18.8	.65	12.1	.78	8.4	.60	10.7	.32	15.7
Test[b]1 (n = 36)	.33	19.2	.31	18.3	.65	11.1	.76	8.0	.61	9.6	.39	12.6
Train 2 (n[c] = 65)	.37	20.3	.40	18.2	.68	11.5	.78	8.1	.56	11.0	.30	15.3
Test 2 (n[c] = 30)	.31	21.0	.32	19.3	.61	12.2	.75	8.5	.68	9.2	.41	13.8
Train 3 (n[c] = 66)	.31	20.2	.31	18.9	.62	11.7	.75	8.2	.64	9.4	.40	13.2
Test 3 (n[c] = 29)	.41	21.3	.48	18.1	.70	12.0	.80	8.3	.53	12.3	.23	17.8

[a]Root Mean Squared Error of Prediction (breaths/minute)
[ab]Training set with all observations of 20 cows
[b]Testing set with all observations of 10 cows
[c]Total number of observations (videos)

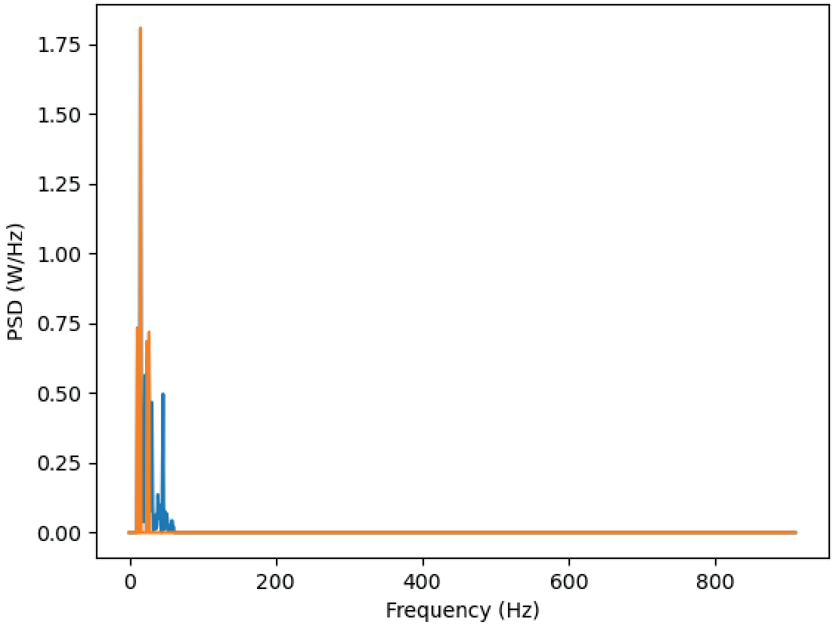

Figure 2: PSD vs. Frequencies - selected (orange) and disregarded (blue) frequencies for the inverse FFT

The original and cleaned signals for an arbitrary animal are shown in Figure 3. In most cases, the method could precisely identify the frequencies that compose the respiratory movements and return a clean signal through which it was possible to estimate the RR. Also displayed in Figure 4, the peak count performed on the cleaned signal with the Python function scipy.signal.find_peaks could precisely capture the animals' respiratory signals. This agrees with Anishchenko et al.'s (2019) study, which used the same function to estimate the RR of humans through bioradar signals and also achieved satisfactory results.

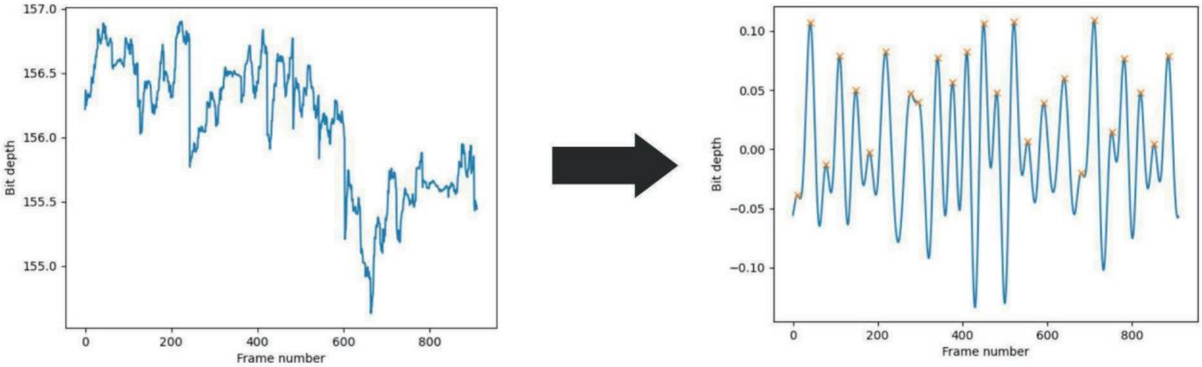

Figure 3: Bit depth vs. Frame number for the original (left) and the inverse Fast Fourier Transformed (right, with counted peaks) pixel intensity signals

The prediction results for the 30 dairy cows are shown in Figure 4 and Figure 5. The method's overall performance had an R^2 of 0.77 between the predicted and observed number of breaths of the cows over a 30-second video segment. The Root Mean Squared Error of Prediction (RMSEP) was 8.3 breaths/minute, 15.8% of the mean predicted respiration rate.

When testing the precision for RGB and infrared videos (night vision), the model performed slightly better for RGB ($R^2 = 0.81$) than for infrared videos ($R^2 = 0.74$). It is important to notice that there were 79 RGB and 112 infrared videos. These results suggest that night vision conditions imposed more challenging lighting to capture the breathing pattern. Nonetheless, the method still yielded great predictions and agreed with Wiede et al. (2017), which also demonstrated that pixel intensity monitoring in RGB images could precisely capture the RR.

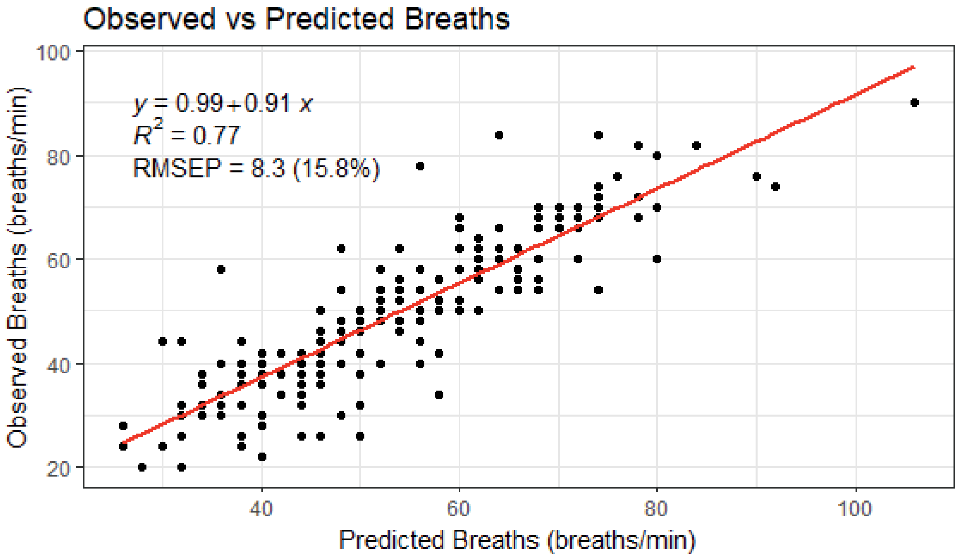

Figure 4: Observed vs. Predicted Breaths for all videos

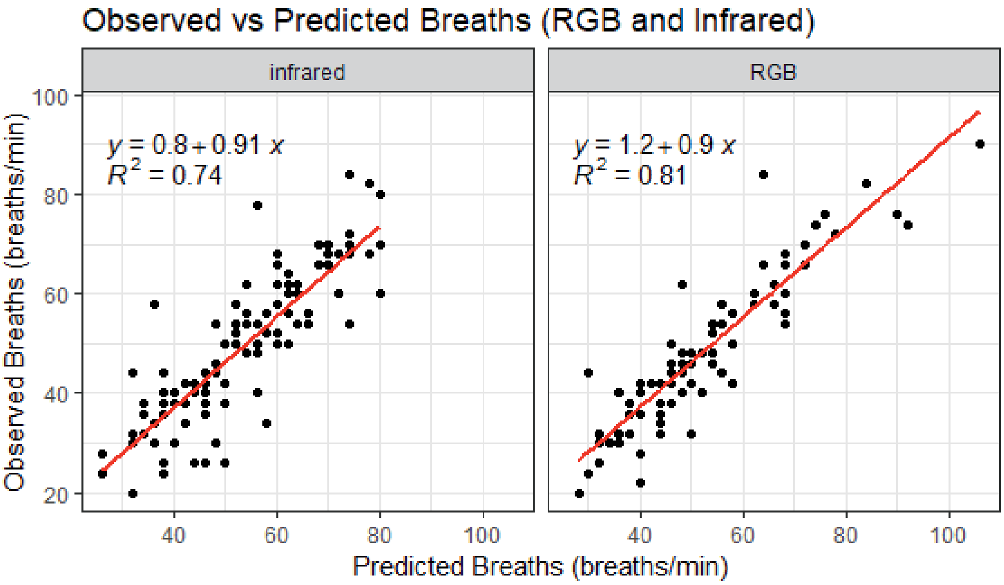

Figure 5: Observed vs. Predicted Breaths for RGB and infrared videos

Given that the proposed method is based on the pixel intensity variation throughout the videos, the influence of lighting changes and non-respiration-related movement could detrimentally affect the robustness of our method. As such, we analyzed a separate dataset with 170 observations, removing all annotations where cows moved or had any lighting disturbance over their ROI. The results, however, showed only a slight increase in R^2 (0.79) and a decrease in the RMSEP of only 0.2 breaths/minute (from 8.3 to 8.1) when compared with the previous analysis, as can be seen in Figure 6. These findings suggest that our method could be robust enough not to be affected by the cows' random movements in their stalls. Still, it is clear that automation to capture the ROI during lying time should be developed as one of the next steps to the full implementation of our system.

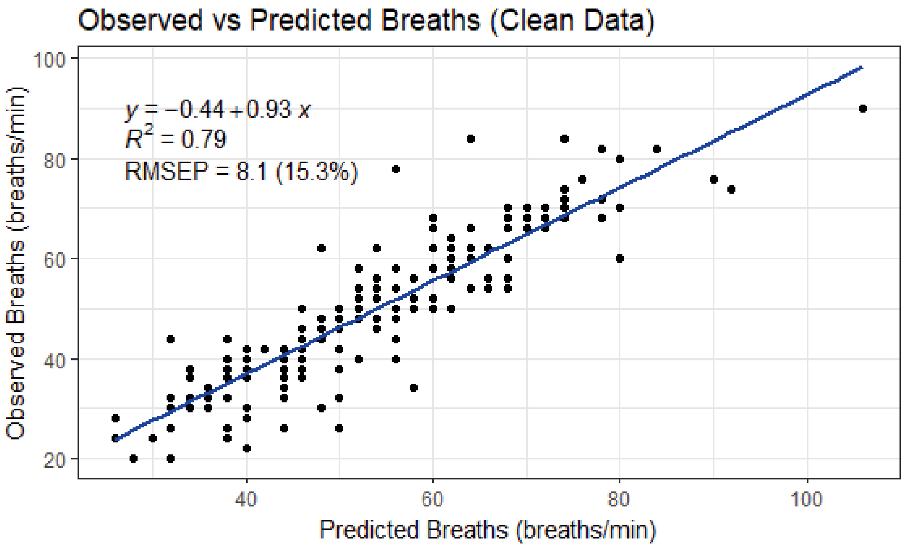

Figure 6: Observed vs. Predicted Breaths for the cleaned data

Compared to the thermal imaging approaches, such as the one developed by Milan et al. (2016), our study has the benefit of operating in a less controlled environment and not depending on placing cameras close to the cows' noses, which may be counter-productive in industrial settings. Regarding the RGB imaging method proposed by Wu et al. (2020), our model has two main advantages: (1) because it doesn't rely on deep learning neural networks, it is less data-hungry and may generalize better across different environments and imaging conditions, which was shown by its performance with calves and in both RGB and Infrared videos;
(2) it's demonstrated to have good predictions for cows in less controlled conditions, which can allow for a smoother transition into industrial settings and gives it great potential for high-throughput phenotyping and farm management.

Conclusions

The proposed method showed robustness by having similarly great predictive performance with different image types and videos with and without moving cows. This approach may have some advantages over other technologies that share the same purpose (i.e., measure RR), for example, the collection of breathing patterns. Future studies could be performed to train an object detector that can identify the ROI of cows without needing manual annotation so that this method can be applied in real-world scenarios. Furthermore, using an image magnifier in the videos may also allow for better detection of the RR by enhancing the significance of the cow's respiratory movements and may contribute to reducing the RMSEP.

Acknowledgments

This project was funded by the USDA National Institute of Food and Agriculture (Washington, DC; grant 2020- 67015-30831).

References

Gaughan, J.B., Holt, S.M., Hahn, G.L., Mader, T.L., and Eigenberg, R.A. (2000) Respiration rate - is it a good measure of heat stress in cattle. *Asian-Australasian Journal of Animal Science* 13,329–332.

Handa, D., and Peschel, J.M (2022) A Review of Monitoring Techniques for Livestock Respiration and Sounds. Front. *Frontiers in Animal Science* 3,904834.

Eigenberg, R.A., Brown-Brandl, T., and Nienaber, J.A. (2002) Development of a respiration rate monitor for swine. *Transactions of the ASAE* 45(5),1599-1603.

Milan, H.F.M., Maia, A.S.C., and Gebremedhin, K.G. (2016) Technical note: Device for measuring respiration rate of cattle under field conditions. *Journal of Animal Science* 94(12), 5434–5438.

Wu, D., Yin, X., Jiang, B., Jiang, M., Li, Z., and Song, H. (2020) Detection of the respiratory rate of standing cows by combining the deeplab v3+ semantic segmentation model with the phase-based video magnification algorithm. *Biosystems Engineering* 192, 72–89.

Wiede, C., Richter, J., Manuel, M., and Hirtz, G. (2017) Remote respiration rate determination in video data - vital parameter extraction based on optical flow and principal component analysis. *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* 326-333.

Anishchenko, L., Evteeva, K., Korostovtseva, L., Bochkarev, M., and Sviryaev, Y. (2019) Respiratory rate determination during sleep by CW Doppler radar. *2019 International Conference on Biomedical Innovations and Applications (BIA)* 1-4.

Becker, C.A., Woolums, A.R., and Stone, A.E. (2020) Recognizing heat stress in dairy cattle: How to visually record. Accessed March 31, 2023. *http://extension.msstate.edu/publications/recognizing-heat-stress-dairy- cattle-how-visually-record-respiration-rate.*