

# Automated non-contact detection of central apneas using video

Evelien E. Geertsema<sup>a,b</sup>, Gerhard H. Visser<sup>a</sup>, Josemir W. Sander<sup>a,c,d</sup>, Stiliyan N. Kalitzin<sup>a,b,\*</sup>

<sup>a</sup> Stichting Epilepsie Instellingen Nederland (SEIN), The Netherlands

<sup>b</sup> Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>c</sup> NIHR University College London Hospitals Biomedical Research Centre, UCL Institute of Neurology, Queen Square, London WC1N 3BG

<sup>d</sup> Chalfont Centre for Epilepsy, Chalfont St Peter, United Kingdom.

\* Corresponding author

e-mail address: skalitzin@sein.nl

postal address: Postbus 540, 2130 AM Hoofddorp, the Netherlands

**Key words:** epilepsy, event detection, pattern recognition, SUDEP, video analysis.

**Declarations of interest:** none

**Funding:** This work was supported by the Margaret Knip Fund and the Christelijke Vereniging voor de Verpleging van Lijders aan Epilepsie.

# Abstract

Central apneas occurring in the aftermath of epileptic seizures may lead to sudden death. The contact-sensors currently used to detect apneas are not always suitable or tolerated. We developed a robust automated non-contact algorithm for real-time detection of central apneas using video cameras. One video recording with simulated apneas and nine with real-life apneas associated with epileptic seizures, each recorded from 3-4 angles, were used to develop the algorithm. Videos were preprocessed using optical flow, from which translation, dilatation and shear rates were extracted. Presence of breathing motions was quantified in the dominant time-frequency spectrum by calculating the relative power in the respiratory range (0.1-1 Hz). Sigmoid modulation was calculated over different scales to quantify sigmoid-like drops in the respiratory range power. Each sigmoid modulation maximum constitutes a possible apnea event. Two event features were calculated to enable distinction between apnea events and movements: modulation maximum amplitude and total spectral power modulation at the time of the event. An ensemble support vector machine was trained to classify events using a bagging procedure and validated in a leave-one-subject-out cross validation procedure. All apnea episodes were detected in the signals from at least one camera angle. Integrating camera inputs capturing different angles increased overall detection sensitivity (>90%). Overall detection specificity of >99% was achieved with both individual cameras and integrated camera inputs. These results suggest that it is feasible to detect central apneas automatically in video, using this algorithm. When validated, the algorithm might be used as an online remote apnea detector for safety monitoring.

# 1. Introduction

Central apneas occurring during epileptic seizures are usually self-limiting [1], but when occurring in the aftermath of a seizure may lead to asystole and could trigger Sudden Unexpected Death in Epilepsy (SUDEP) [2]. Observed fatal SUDEP cases occurred within minutes after a convulsive seizure [2], but not all known cases were preceded by a seizure [3]. Interventions such as repositioning or clearing of airways may have a protective effect preventing SUDEP [4,5] and automated seizure detection may help alert carers. If a seizure was not detected or intervention could not prevent apnea, carers should be alerted to apneas to enable timely resuscitation.

Currently apneas are detected by a respiratory inductance plethysmogram (RIP), an impedance pneumogram, or a blood oxygen saturation measurement [6]. These contact-sensors are, however, not always suitable or tolerated. They would need to be worn continuously, may be uncomfortable and sensors could come loose during seizures [7]. Some populations such as children or people with intellectual disability may not tolerate wearable devices and may try to dislodge them. An alternative solution would be apnea detection using remote sensors.

There are several modes of remote respiratory monitoring using different sensors, but there are few to detect apneas. Algorithms aiming at remote apnea detection use radar [8], sonar [9], **capacitance sensors** [10], infrared sensors [11,12], depth sensors [13–15], and video [16–19]. Video cameras are suitable for safety monitoring, as they are relatively cheap and sensitive to movement, even at longer distance, provided they have suitable resolution.

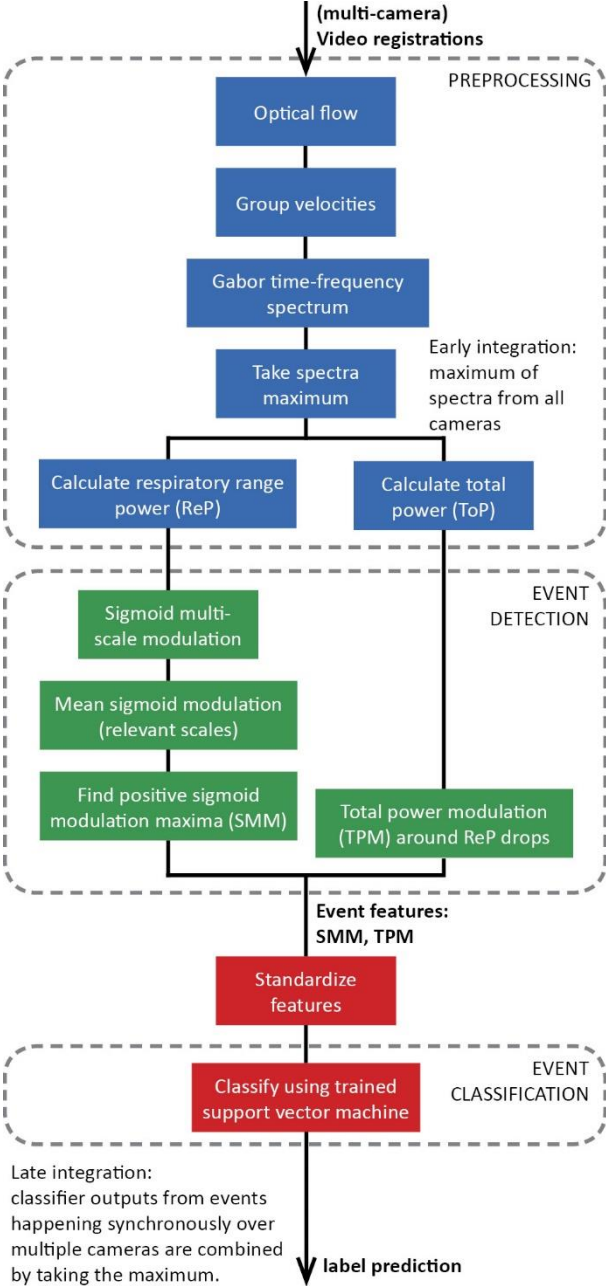
Respiration can be quantified in video by using the breathing motions of the trunk or the respiratory modulation of the photoplethysmogram (PPG). PPG techniques are unsuitable for our application, as they need skin regions that may not always be visible (covered or turned away from the camera), and color video, which cannot be used at night. Trunk breathing motions may be quantified, for example, by calculating translation rates with optical flow [20], tracking image points over time [21], or comparing breathing motion templates [16]. Most published algorithms aim to detect obstructive sleep apneas (OSA), often using the characteristic sounds and movements that conclude OSA for classification. These algorithms are unsuitable to detect central apneas; in which such sounds and movements are absent.

There is no algorithm available that can detect central apneas in video and is suitable for (postictal) monitoring. This algorithm needs to be 1) able to work near real-time, providing minimal time delay between apnea onset and its detection, 2) able to distinguish between apnea and gross body motion when breathing motions cease, 3) independent of camera imaging mode to enable daytime and nighttime use, and 4) independent of skin or chest visibility and tracking.

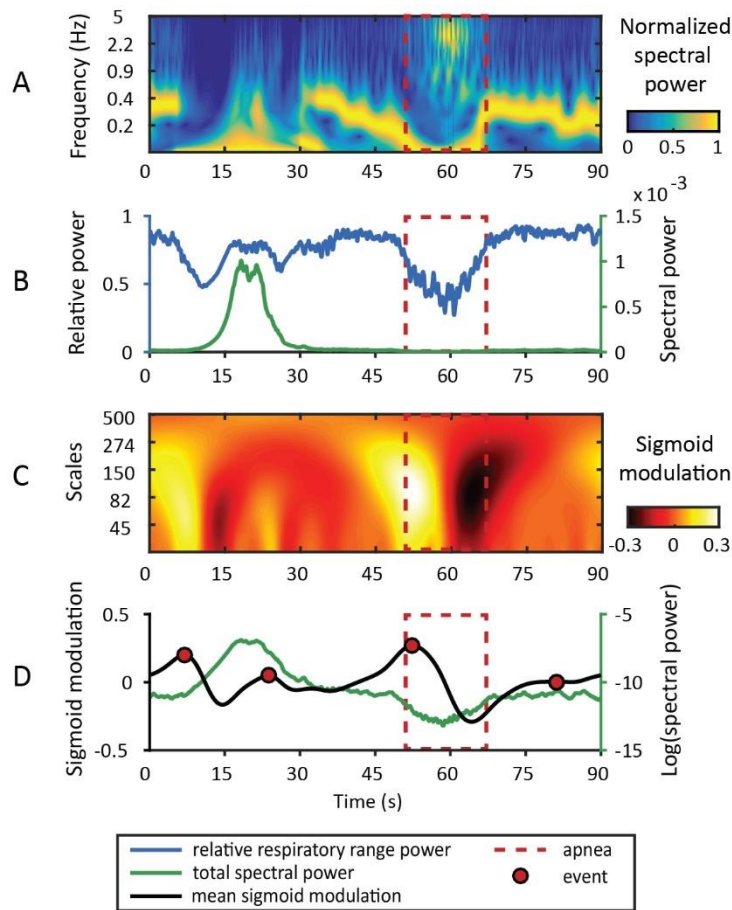
In this proof of concept study, we present a robust automated algorithm for real-time non-contact detection of central apneas in video registrations. Our algorithm detects cessation of the measured oscillatory movements of breathing, in the absence of gross body motion. **Recordings** of simulated apneas and real-life, seizure-related central apneas were used to develop and test the algorithm in a leave-one-subject-out cross-validation scheme. We investigated detection performance in a single-camera setup from different angles and propose two different strategies to combine signals from multiple cameras to improve performance.

# 2. Methods

The algorithm consists of three parts: preprocessing of the video signals, event detection and event classification. Video signals from multiple-camera setups were integrated in two ways: early, during preprocessing, or late, after classification. An overview of the algorithm is shown in Fig. 1. An example of algorithm output after different video processing steps is shown in Fig. 2.



**Fig. 1.** Apnea detection algorithm workflow. Total power modulation is only calculated at the time of a detected event, i.e., a positive mean sigmoid modulation maximum.



**Fig. 2.** Example of the output from different analysis steps in the apnea detection algorithm. A 90-second video fragment with a movement (at  $t \approx 15$  s) and an apnea (starting at  $t \approx 49$  s) is shown (subject SIM1, camera 1). (a) Normalized Gabor time-frequency spectrum, showing the dominant component from the five group velocity spectra. (b) Relative power in the respiratory range (ReP) plotted together with the total power. (c) Sigmoid modulation of the ReP, showing high power at significant signal amplitude drops. (d) Mean sigmoid power modulation, calculated over relevant sigmoid scales that fit apneas. Total power is scaled logarithmically here. Events were detected at positive mean sigmoid power modulation maxima.

## 2.1. Video recordings

One healthy subject, and nine subjects with epilepsy (but otherwise healthy), who had seizure-related apneas, were included in this study. The subjects with epilepsy were selected from a database which was prospectively collected (2016-2017) in an epilepsy monitoring unit for the development of seizure detection algorithms. Next to the standard long-term ( $\geq 24$  h) synchronized electro-encephalogram, video, sound, and electrocardiogram, respiration was quantified for this dataset using RIP belts. The study protocol was approved by the ethics committee of University Medical Center Utrecht and written informed consent was given by all participants or their guardians. All data were handled anonymously.

Subjects were registered with three or four cameras positioned around the room, so the face could be brought in view from at least one angle (**important for epilepsy diagnostics**). Subjects were not always in view of all cameras. Depending on the position of the subject moving freely in the room, subject-to-camera distance ranges 1-5 m. Camera rotation, zoom and pan could be controlled by nurses in a monitoring room. The registration system started a new video automatically after approximately 26 minutes, for storage purposes. Cameras were equipped with an infrared illuminator for night-time recordings. Videos were registered using a frame rate of 25 frames per second. Each frame, with a total resolution of 1080x1920 pixels, contains bordered sub-frames from

the four cameras, with a 535x959 pixel resolution. Sub-frames were separated after reading for further analyses.

### 2.1.1. Simulations

To be able to investigate the influence of different breathing patterns on apnea detection performance, one recording involved a healthy subject simulating different respiratory frequencies and depths leading up to apneas.

## 2.2. Data selection

As peri-ictal central apneas are common [1], the minutes around the seizures of each subject in the database were analyzed for the occurrence of apneas. During this screening, the signals from the respiratory belts were used to locate apneas visually. We presume peri-ictal apneas to be of central origin, as has been described in literature [1]. We did not confirm this with other sensors. For each subject that had peri-ictal apneas, one video was included in the dataset in which one or more apneas occurred. Recording information is shown in Table 1. For one subject (PAT9), more than one video recording was available for inclusion; we chose the video in which the subject was the only person in view of the cameras. One of the subjects was 16 years old at the time of registration, the remaining subjects were adults (mean age 38.3 years, SD: 15.8).

**Table 1.** Dataset information. Relevant apneas (duration  $\geq 8$  s, no gross body movement or other persons in view) were used for training and considered essential to detect. One (unique) apnea occurring in a subject may have been captured in multiple recordings, i.e. filmed from different angles with different cameras.

ID	Age	Sex	Recording length (min)	N cameras	N apneas (unique)	N relevant apneas (unique)
SIM1	29	F	16.0	4	36 (9)	30 (8)
PAT1	66	M	6.00	3	2 (1)	0 (0)
PAT2	45	F	17.7	3	3 (1)	0 (0)
PAT3	53	F	26.6	4	4 (1)	0 (0)
PAT4	45	F	26.6	4	3 (1)	2 (1)
PAT5	18	F	26.6	3	27 (9)	4 (2)
PAT6	38	M	26.6	4	3 (1)	2 (1)
PAT7	46	M	26.6	3	6 (2)	0 (0)
PAT8	27	F	19.9	4	4 (1)	3 (1)
PAT9	16	M	26.6	4	4 (1)	2 (1)
Total			219	36	92 (27)	43 (14)

## 2.3. Annotations

To provide ground truth information on the presence and timing of apnea, the beginning and end of each apnea was annotated. Annotations were based on the signals from one, or if available, both RIP belts. For each apnea we noted in which of the cameras the subject was visible, if the subject was moving at apnea onset, and whether there were other persons in view of one of the cameras. Apneas without gross subject movement or other persons in view, and a duration of  $\geq 8$  s were considered ‘relevant apneas’. Relevant apneas were used during algorithm training and were considered essential to detect during algorithm testing. An apnea may be considered relevant in a subset of the 3 or 4 cameras used in the registration, and non-relevant in other cameras; gross body movement or a second person may only be visible in a subset of camera views.

## 2.4. Preprocessing

Video recordings were preprocessed in order to extract: 1) respiratory range oscillatory power, which quantifies the presence of a breathing motion signal, and 2) total power, to quantify the presence of

gross body movement. The preprocessing procedure used here is similar to the procedure used in a convulsive seizure detection algorithm previously developed by our group [22]. A detailed description of the preprocessing procedure is provided in the referred paper. The main difference in the current preprocessing procedure is the focus on the presence of oscillatory motions in the frequency range of respiration, instead of the frequency range of oscillatory motions in convulsive seizures. We intend to combine our seizure and apnea detection algorithms into one safety monitoring system in the future. Similar preprocessing procedures for both algorithms allow computationally heavy analysis steps, such as optical flow calculation, to be executed once, presumably saving calculation time. All calculations were made with Matlab version 2018a on an Intel Core i7 2.50 GHZ processor with 8 GB RAM.

First, subject movement in the video was quantified with optical flow, using the Horn-Schunck method [23], implemented as standard in Matlab (version 2018a). From the resulting velocity fields five timeseries were derived, representing the spatial transformations, or group velocities; translation rates along the two image axes, dilatation rates and shear rates. Here, we chose to omit rotation, originally the sixth group velocity, for its lack of quantifying the respiratory signal.

Next, the time-dependent spectral content in each of the five group velocities was calculated using convolutions with 200 Gabor filters, or wavelet templates, with exponentially spaced central frequencies,  $\nu$ , in the range [0.08,5] Hz, and with  $\alpha = 0.5 * \nu$  as filter bandwidth. The Gabor filters are defined as:

$$g(t - t', \nu) = (e^{-\pi^2 \alpha^2 \nu^2 (t-t')^2 - i2\pi\nu(t-t')} - O_\nu) / N_\nu. \quad (1)$$

The normalization factor  $N$  and the offset factor  $O$  are chosen so that the functions have zero mean and unit 1-norm.

In order to obtain the dominant component of the group velocities ( $Q$ ), we take the maximum of the five Gabor time-frequency spectra ( $G_c$ ):

$$Q(t, \nu) = \max_c (G_c(t, \nu)), \quad c = 1, \dots, 5, \quad (2)$$

Where  $t$  is defined as time and  $c$  defines the corresponding group velocity. An example of the  $Q(t, \nu)$  spectrum at the time of breathing, movement, and apnea is illustrated in Fig. 2a. When early integration of video-signals was applied, the maximum in (2) was taken over the (15 or 20) group velocities from all cameras.

Relative power in the respiratory frequency band ( $ReP$ ) was calculated as the power in the 0.1-1 Hz band relative to the total Gabor power (0.08-5 Hz):

$$ReP(t) = \frac{\sum_{\nu=0.1}^1 Q(t, \nu)}{\sum_\nu Q(t, \nu)} \quad (3)$$

In video segments without gross body movement but with the presence of breathing motions, the  $ReP$  quantity will presumably be close to one. Note that we do not aim to quantify the respiration frequency with this quantity, but only the presence (and cessation) of respiration. Mean total power ( $ToP$ ) is calculated over all measured frequencies:

$$ToP(t) = \langle Q(t, \nu) \rangle_\nu. \quad (4)$$

An example of the  $ReP(t)$  and  $ToP(t)$  signals at the time of breathing, movement, and apnea is illustrated in Fig. 2b. Both the  $ReP$  and  $ToP$  signals were scaled logarithmically before further calculations.

## 2.5. Event detection

Apneas manifest as sigmoid-like amplitude drops in the  $ReP(t)$  signal. To quantify drop occurrence we calculated sigmoid modulation for different scales to detect drop events of different slopes and amplitudes in  $ReP(t)$ . This technique is analogous to computing a wavelet time-frequency spectrum. Here, we compute a sigmoid time-scale spectrum instead, using an aperture sigmoid as the generating wavelet template. In this way, the slope (scale) of the drop of  $ReP(t)$  and its amplitude (relative to the mean value) can be estimated. The technique is described in detail below.

First, we define a range of 200 scales  $s_k$  ( $k = 1, 2, \dots, 200$ ), with exponentially spaced values in the range [25,500]. For each scale, an aperture sigmoid template is defined for window  $\tau$ :

$$S(\tau, k) = N_k^{S-1} \frac{e^{\tau/s_k} - e^{-\tau/s_k}}{e^{\tau/s_k} + e^{-\tau/s_k}} e^{-|\tau|^2/s_k^2}, \tau = [-3s_k: 3s_k], \quad (5)$$

together with the Gaussian aperture template:

$$G(\tau, k) = N_k^{G-1} e^{-|\tau|^2/s_k^2}. \quad (6)$$

In Equations (5) and (6) L2 normalization was applied through the coefficients  $N_k^{S, G-1}$ , with  $N_k$  defined as the squared sum of the  $k^{\text{th}}$  aperture template. The time window in (5) and (6) is chosen to be of three scale lengths, as values outside this range are suppressed by the Gaussian aperture factor. Sigmoid time-scale modulation  $m$  can then be obtained using the convolutions between the filters and the  $ReP$  signal:

$$m(t, k) = \frac{\int_{-\infty}^{\infty} S(t-\tau) ReP(\tau) d\tau}{\int_{-\infty}^{\infty} G(t-\tau) ReP(\tau) d\tau} \quad (7)$$

An example of the sigmoid modulation at the time of breathing, movement, and apnea is illustrated in Fig. 2c.

To quantify the presence of significant respiratory range power drops, we calculated the mean sigmoid modulation  $M$  over the scales that correspond to observed drop times:

$$M(t) = \langle m(t, k) \rangle_{k \in s_{drop}} \quad (8)$$

Drop times were observed to be between 4.0 and 8.2 seconds (in the SIM1 recording), and correspond to filters  $s_{drop} = [s_{70}, s_{129}]$ . A positive mean sigmoid modulation maximum corresponds to a significant respiratory range power drop. Its amplitude constitutes our first event feature, to be used in classification. As the  $M(t)$  signal fluctuates, many positive maxima and thus events will be detected. High maximum values are however only found when there is a significant respiratory power drop. For the positive local maxima of  $M$  we define the set of time points  $t_M$  as:

$$t_M = \{t: M(t) > \max(M(t-1), M(t+1)); M(t) > 0\} \quad (9)$$

For each event, the sigmoid modulation maximum amplitude is defined as:

$$SMM(t_M) = M(t_M). \quad (10)$$

An example of the mean sigmoid modulation and its positive maxima forming the events, is shown in Fig. 2d.

When an apnea occurs, the total power stays low, or decreases from a low to an even lower value. At the time of events due to gross body movement on the other hand, the total power either increases (movement onset) or drops from a high to a lower value (movement end). An example of the total power change at the time of movement and apnea events is shown in Fig. 2d. A second classification feature quantifying the change of total power at the time of events may distinguish events due to apneas from events due to gross body movements. For each event we therefore calculated the total power modulation ( $TPM$ ), comparing the 2 s before, to the 2 s after the  $M$  maximum:



$$TPM(t_M) = \frac{\langle ToP \rangle_a - \langle ToP \rangle_b}{\langle ToP \rangle_a + \langle ToP \rangle_b},$$

$$b = [t_M - 2s, t_M], a = [t_M, t_M + 2s]. \quad (11)$$

Presumably, the  $TPM$  feature has a small and often negative value for apnea events, and a high value (positive or negative) for gross body movement events.

All events received a ground truth label according to the apnea annotations: 1 (apnea) or 0 (not an apnea). A five second annotation margin allowed detections slightly earlier than the annotated apnea onset. If multiple events were detected within an apnea period, the first event was labelled as the apnea event. Other events during the apnea period were disregarded; an apnea can only be detected once, but an extra detection during an apnea should not be considered a false detection. Relevant apneas were also labelled separately. If no events were found during an apnea, the apnea was missed, but as only events received a ground truth label, such a false negative would not be counted. We therefore corrected performance results by hand to make sure any missed apneas without an event are counted as a false negative.

## 2.6. Classification

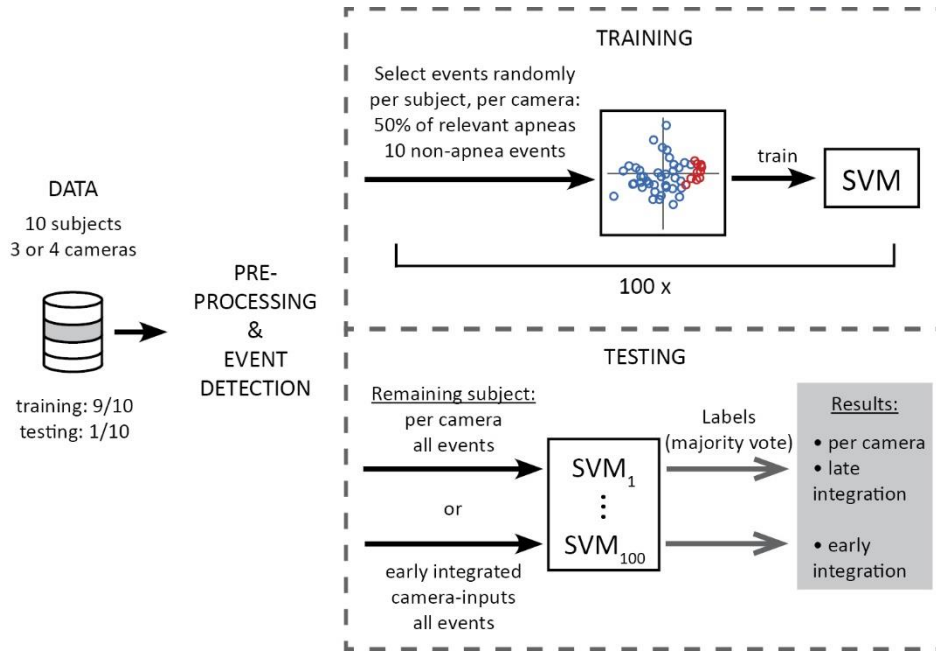
To automatically classify detected events as “apnea” or “not an apnea”, we used a support vector machine (SVM) with a radial basis function (RBF) kernel. Optimization of the SVM’s hyperparameters was followed by training, which provides SVM model parameters. During both hyperparameter and model parameter setting, precision and accuracy were optimized while ensuring 90% sensitivity for apnea events. Each time an SVM model was trained, the misclassification cost for a false negative was increased with 0.1 steps until 90% sensitivity on its training data was achieved.

### 2.6.1. Hyperparameter optimization

We found the SVM hyperparameters that provided high performance while generalizing well, using a grid-search. The soft margin constant and the RBF kernel scale were tuned to the combination (0.1 and 2, respectively) that performed best in cross validation.

### 2.6.2. Training and cross validation

A leave-one-subject-out (LOSO) cross validation procedure was used to train the event classifier and test its performance on new data not used for training. **Test results from LOSO cross validation reflect real-life algorithm performance, when the algorithm would be applied on a new subject.** In each of the ten folds the classifier is trained on data from nine/ten subjects and tested on the remaining subject. The training and testing procedure is depicted in Fig. 3.



**Fig. 3.** Leave-one-subject-out cross validation procedure. In each of the subject-folds training is performed 100 times with a bag of randomly selected training events from 9/10 subjects, resulting in an ensemble of 100 trained SVM models. Testing on the data from the remaining subject is done either 1) on all events per camera, providing predictions per camera, which may be integrated later, or 2) on the events from the early integrated camera inputs.

In each fold, an ensemble classifier was trained using a bagging (i.e. bootstrap aggregating) procedure. Applying a bagging procedure prevents overrepresentation of recurring events from a recording in the training set and helps to reduce variance, thus increasing classifier stability and accuracy. A sample, with replacement, from the training events (bag) was used to train a classifier. Each bag sample consisted of a random selection of 50% (with a minimum of one) of the relevant apnea events and 10 non-apnea events, selected from each camera recording in the training set separately. This process was repeated 100 times in each fold to obtain an ensemble classifier.

The classifier is subsequently tested on the recordings of the remaining subject of the fold, with two different sets of input events: the events found in the video signals from each individual camera, and the events found in the early integrated video signals. A prediction from an ensemble model was formed by the majority vote of the 100 predicted labels. Late integration of outputs was performed by combining synchronous events, i.e., the maximum output label was obtained for events that occurred within 5 s over two or more cameras. In practice, one camera providing a detection is enough to get a detection from a late integrated camera system.

## 2.7. Performance evaluation

Comparison of predicted labels with ground truth labels yielded the following counts: true positives ( $TP$ ); detected apneas, true negatives ( $TN$ ); events that were correctly classified as not an apnea, false positives ( $FP$ ); events incorrectly classified as an apnea, and false negatives ( $FN$ ); events that were incorrectly classified as not an apnea. With these counts we calculated apnea detection sensitivity, false positive rate (FPR), and precision, as follows:

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \quad (12)$$

$$\text{False positive rate} = \frac{FP}{D}, \quad (13)$$

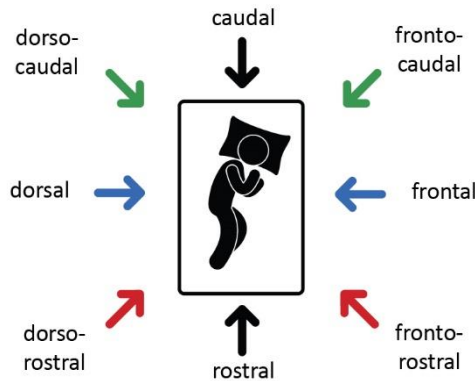
$$\text{precision} = \frac{TP}{TP+FP}. \quad (14)$$

with  $D$  defined as the **recording** duration in hours. Sensitivity for relevant apneas was analyzed separately. We also inspected algorithm output and video **recordings** at times of false positives to investigate their causes.

Grouped results were obtained by concatenating all events and comparing the predicted labels with the ground truth. Results from all individual cameras estimate the predicted performance when a random camera angle would be chosen. We also compared found performance with the performance using best and worst camera choices, determined post-hoc per subject according to 1) sensitivity and 2) false positive rate. Thus, we estimate performance when we would have known beforehand the best and worst camera positions to capture apneas. This is done for comparison, and does not pose a realistic situation.

We calculated detection latency for relevant apneas, defined as the time between the annotated start of the apnea and the detection. To estimate real-time latency, Gabor filter windows for respiratory frequencies need to be taken into account. These are calculated in the center of windows of 75-745 frames, effectively adding 1.5 to 15 seconds to real-time latency.

If only a single camera is available, it is important to know beforehand which camera placement gives the highest detection sensitivity. To answer this question, we analyzed the influence of camera perspectives with respect to the subject (as shown in Fig. 4) on apnea detection sensitivity.



**Fig. 4.** Probable camera angles when monitoring a subject. Opposite camera positions form a combination (indicated with an equal color), as they may represent the same view when the subject turns. Adapted from “Sleep by Gan Khoon Lay from the Noun Project”.

### 3. Results

Grouped algorithm **cross validation** test results using predicted labels from the ensemble SVMs are shown in Table 2. Detailed results per subject and camera can be found in Supplementary Table 1 in the Appendix. All relevant apneas from all subjects were detected in the signals from at least one of the used cameras. In three subjects relevant apneas were not detected in all cameras in which they were annotated. The apnea in the video of PAT9 was detected in camera signals without, but also with a second person in view (who was quietly sleeping in the background). With the best camera angles determined post-hoc, 100% sensitivity would have been achieved with a false positive rate of 1.09 per hour. Mean detection latency for detected relevant apneas was 1.7 s (SD: 1.4 s). Mean online latency is estimated to be in the range of 2.7-16 s.

**Table 2.** Grouped algorithm test results, using majority vote labels from the trained ensemble models.

Camera / integration method	Sensitivity (%) relevant apneas (N)	Sensitivity (%) all apneas (N)	False positive rate (/h)	Specificity (%) for events
-----------------------------------	---	-----------------------------------	--------------------------------	----------------------------------

All individual cameras	83.7	(43)	46.7	(92)	2.17	99.1
best choice	100	(14)	63.0	(27)	1.09	99.5
worst choice	80.0	(10)	37.0	(27)	3.28	98.5
late integrated	100	(14)	66.7	(27)	6.57	97.6
early integrated	92.9	(14)	51.9	(27)	1.64	99.3

With late camera input integration 100% sensitivity was achieved, with a FPR of 6.6 per hour. Late integration resulted in a higher FPR than found with randomly chosen cameras, and also higher than found with the post-hoc determined worst camera choices. With early camera input integration, 13/14 unique relevant apneas were detected. Closer inspection of the signals at the time of the missed apnea (PAT6) revealed a slight movement at apnea onset, mainly visible in one of the cameras. For this reason the apnea presumably was detected in the single-camera signals (and in the late integrated outputs), but not in the early integrated video signals, which included the movement as a dominant component.

Best single-camera sensitivity was achieved with angle combination frontocaudal-dorsocaudal (100%). Sensitivity was also high for the combination frontal-dorsal (92%). Three/six frontal view detections involved apneas of awake subjects sitting upright. Sensitivity was lower when the subject was registered from a more rostral view; 84% for the combination dorsorostral-frontorostral, and 50% from a rostral viewpoint. No **recordings** were made from a caudal view.

Inspection of algorithm output at times of false detections showed several causes for false detections. Coincidental detections, in which a very small movement precedes a larger movement that causes the respiratory power to drop, caused 68% of single-camera false detections. Of the 13 coincidental false detections, 3 were also in the early integrated results. When inspecting the respiratory belt signals, three “false” detections turned out to be short apneas (~5 s) and one was caused by a respiratory frequency drop to 0.1 Hz. In one camera **recording** providing a closeup of the subject’s face during sleep, eyelid myoclonias caused two false detections. These false detections were also seen in the results from the early integrated signal.

## 4. Discussion

**Our algorithm detected all apneas in all subjects in the signals of at least one camera, with acceptable false detection rates. The algorithm can quickly detect cessation of the breathing motion signal, while distinguishing cessations due to apneas from cessations due to gross body movements. Apneas were detected regardless of the breathing pattern (frequency, depth) preceding them. The algorithm functions with color and intensity video and does not need uncovered body regions in view to quantify breathing motions and detect apneas. The hardware that would be used in application of our algorithm needs to be able to acquire image sequences, and not necessarily record them, which favors privacy.**

When **multiple cameras are** available, integrating **their** signals early and late in the algorithm improves detection sensitivity, as respiration movements can be captured from different angles. Late integration, i.e., combining any synchronous detections over cameras, may, however, induce a high false detection rate: **an event causing** a false detection in any camera will **inevitably** cause a false detection in the late integrated results. Early integration, i.e., combining video time-frequency signals before event detection and classification, typically results in less false detections: **coincidental false detections are less likely, because small movements that cause the total power signal to fit the apnea model can be cancelled out during signal integration. Detection sensitivity with early signal integration** may however be affected by gross body movement at apnea onset in one of the camera

signals. If further study verifies that (near) SUDEP-related apneas do not involve subject movements at apnea onset, early integration may be preferred over late integration.

When only a single camera is used, detection sensitivity depends on the ability to capture respiratory movements, to detect their arrest. Best perspectives have camera axes perpendicular to **craniocaudal and dorsoventral** directions, **as trunk oscillatory movements of breathing are largest in these directions** [24]. Cameras should be placed high enough to get a good overview of the subject. Our results indicate that using a frontocaudal-dorsocaudal camera angle combination likely results in the best performance. Dorsal-frontal and dorsorostral-frontorostral viewpoints may also provide good detection sensitivity when the camera is placed high enough. Rostral views provided only mediocre apnea detection performance, and should be avoided. Closeup views of the subject's face and shoulder regions may provide a good respiratory signal, but false detections might be caused by small **muscle twitches**.

**An apnea was detected with a second person in view of the camera, which suggests that** apneas are not necessarily missed when there is a second person in view of the camera. The second person's breathing movements add to the respiratory range power, but if only one of the persons stops breathing, **respiratory** power will drop nonetheless. **To detect an apnea, the second person should not be showing gross body movements** and it will probably help if the subject to monitor is closest to the camera, as the subject's respiration will then influence respiratory range power the most. Four of ten subjects were covered by blankets when their apneas occurred, which did not prevent apnea detection. This suggests that apnea detection is not hampered by **the subject** being covered. The influence of covers on detection should be further investigated, as this is a likely occurrence in practice.

**To our knowledge**, there are no other central apnea detection algorithms **that** meet all requirements to be applicable for safety monitoring (**near real-time detection, distinguish between apnea and body motion, day and nighttime use, independent of skin visibility and tracking**). **One** reported algorithm **meets three/four requirements** [19]. Our approaches are similar; quantifying the presence of periodicity to detect apneas (periodicity cessation). In a dataset with 17 apneas, the algorithm detected 90% of apneas with a latency of 30 s and a specificity of 78%, which translates to an FPR of 79 /h with their windowing procedure. The lower specificity compared to our algorithm might be explained by the algorithm's inability to **distinguish between** gross **body** motion **and** apneas.

Our algorithm enables fast detection of apneas, with online detection latencies estimated to be between 3 and 16 seconds, while latencies of >20 s are common in other studies [18,19]. The difference in detection delay can be explained by the respiration-free signal window needed by most other algorithms to measure the absence of periodicity and detect an apnea. Our algorithm detects the moment of cessation of breathing motions, and does not need an large signal window free of breathing motions before it detects an apnea.

A possible limitation of our algorithm is that it can only detect apneas at their onset. If something/someone in view of the camera causes a gross motion during apnea onset, the apnea might be missed. **Such** movement influences might be prevented by confining camera view as much as possible to the subject. Subject segmentation before quantifying the presence of breathing motions may also improve algorithm performance. Segmentation may, however, also make the algorithm more sensitive to occlusions and heavier computationally.

Obstructive apneas occurring during sleep, **possibly indicating** an obstructive sleep apnea (OSA) disorder, may cause algorithm detections. These apnea detections could be considered true positives, and may help classify undiagnosed OSA cases. OSA is a common disorder, however, and people with the disorder can have frequent apneas throughout the night (severe cases have >30 /h)

[25], which may result in alarm fatigue. OSA detection with our algorithm was not studied, and should be further investigated.

A possible solution for high false detection rates due to OSA or other causes might be switching on apnea detection only after detection of a convulsive seizure (automatically, e.g. using [26], or by a carer). **Apnea monitoring during the half hour following each convulsive seizure might suffice to prevent SUDEP, as virtually all known cases of sudden death followed within 11 minutes after a convulsive seizure [2].**

Future work should focus on further improvement and validation of the algorithm. A larger dataset with central apneas captured from **appropriate** camera angles may improve algorithm performance and enable algorithm validation. Using flexible posterior probability outputs (prediction ranging from 0 to 1) **from the SVM** instead of label outputs might further optimize sensitivity, but this may also increase the false detection rate. **Higher false detection rates** might be acceptable if alerts can be easily verified by a caregiver who is close by, monitoring the video stream. Validation data should include central apneas occurring after convulsive seizures, to verify that these apneas indeed do not involve gross body movements. Long-term unselected **recordings** should be used to demonstrate robustness to the variety of (normal) breathing patterns and real-life false detection rates. Other application areas for the algorithm might also be explored, such as remote neonatal monitoring in the intensive care.

## 5. Conclusions

We present a novel automated algorithm to detect central apneas in video, based on the arrest of oscillatory breathing motions. **The algorithm meets important applicability requirements for safety monitoring; detecting apneas quickly, able to distinguish apneas from gross body motion, usable during day and nighttime, and independent of skin visibility and body tracking. Our results suggest that it is feasible to detect apneas automatically in video, using this algorithm.** When fully validated, it might be used as an online remote apnea detector, avoiding the use of contact sensors. Further studies are warranted.

## Acknowledgements

This work was supported by the Margaret Knip Fund and the Christelijke Vereniging voor de Verpleging van Lijders aan Epilepsie.

## References

- [1] N. Lacuey, B. Zonjy, J.P. Hampson, M.R.S. Rani, O. Devinsky, M. Nei, R.M. Harper, L. Bateman, A. Zaremba, R.K. Sainju, B.K. Gehlbach, S. Schuele, D. Friedman, L. Allen, B. Diehl, J.J. Millichap, M.A. Granner, D.N. Dragon, G.B. Richerson, S.D. Lhatoo, The incidence and significance of periictal apnea in epileptic seizures, *Epilepsia*. 59 (2018) 573–582. doi:10.1111/epi.14006.
- [2] P. Ryvlin, L. Nashef, S.D. Lhatoo, L.M. Bateman, J. Bird, A. Bleasel, P. Boon, A. Crespel, B.A. Dworetzky, H. Høgenhaven, H. Lerche, L. Maillard, M.P. Malter, C. Marchal, J.M.K. Murthy, M. Nitsche, E. Pataraiia, T. Rabben, S. Rheims, B. Sadzot, A. Schulze-Bonhage, M. Seyal, E.L. So, M. Spitz, A. Szucs, M. Tan, J.X. Tao, T. Tomson, Incidence and mechanisms of cardiorespiratory arrests in epilepsy monitoring units (MORTEMUS): A retrospective study, *Lancet Neurol*. 12 (2013) 966–977. doi:10.1016/S1474-4422(13)70214-X.
- [3] S.D. Lhatoo, M. Nei, M. Raghavan, M. Sperling, B. Zonjy, N. Lacuey, O. Devinsky, Non-seizure SUDEP: Sudden Unexpected Death in Epilepsy without preceding epileptic seizures, *Epilepsia*. 57 (2016) 1161–1168. doi:10.1111/epi.13419.

- [4] R. Surges, R.D. Thijs, H.L. Tan, J.W. Sander, Sudden unexpected death in epilepsy: risk factors and potential pathomechanisms, *Nat. Rev. Neurol.* 5 (2009) 492–504. doi:10.1038/nrneuro.2009.118.
- [5] F. Rugg-Gunn, J. Duncan, H. Hjalgrim, M. Seyal, L. Bateman, From unwitnessed fatality to witnessed rescue: Nonpharmacologic interventions in sudden unexpected death in epilepsy, *Epilepsia.* 57 (2016) 26–34. doi:10.1111/epi.13231.
- [6] S. Baillieux, B. Revol, I. Jullian-Desayes, M. Joyeux-Faure, R. Tamisier, J.-L. Pépin, Diagnosis and management of sleep apnea syndrome., *Expert Rev. Respir. Med.* 13 (2019) 445–557. doi:10.1080/17476348.2019.1604226.
- [7] A. Van de Vel, K. Cuppens, B. Bonroy, M. Milosevic, K. Jansen, S. Van Huffel, B. Vanrumste, L. Lagae, B. Ceulemans, Non-EEG seizure-detection systems and potential SUDEP prevention: state of the art., *Seizure.* 22 (2013) 345–55. doi:10.1016/j.seizure.2013.02.012.
- [8] A. Zaffaroni, B. Kent, E. O’Hare, C. Heneghan, P. Boyle, G. O’Connell, M. Pallin, P. De Chazal, W.T. McNicholas, Assessment of sleep-disordered breathing using a non-contact bio-motion sensor, *J. Sleep Res.* 22 (2013) 231–236. doi:10.1111/j.1365-2869.2012.01056.x.
- [9] R. Nandakumar, S. Gollakota, N. Watson, Contactless Sleep Apnea Detection on Smartphones, *Proc. 13th Annu. Int. Conf. Mob. Syst. Appl. Serv. - MobiSys ’15.* (2015) 45–57. doi:10.1145/2742647.2742674.
- [10] I.D. Castro, C. Varon, T. Torfs, S. van Huffel, R. Puers, C. van Hoof, Evaluation of a multichannel non-contact ECG system and signal quality algorithms for sleep apnea detection and monitoring, *Sensors (Switzerland).* 18 (2018) 1–20. doi:10.3390/s18020577.
- [11] M.M. Bani Amer, R. Az-Zaqah, A.K. Aldofash, A.Y. Mohammad, A.M. Dameer, Contactless method for detection of infant sleep apnoea, *J. Med. Eng. Technol.* 34 (2010) 324–328. doi:10.3109/03091902.2010.481034.
- [12] V. Hers, D. Corbugy, I. Joslet, P. Hermant, J. Demarteau, B. Delhougne, G. Vandermoten, J.P. Hermanne, New concept using Passive Infrared (PIR) technology for a contactless detection of breathing movement: A pilot study involving a cohort of 169 adult patients, *J. Clin. Monit. Comput.* 27 (2013) 521–529. doi:10.1007/s10877-013-9457-2.
- [13] A. Al-Naji, K. Gibson, S.-H. Lee, J. Chahl, Real Time Apnoea Monitoring of Children Using the Microsoft Kinect Sensor: A Pilot Study, *Sensors.* 17 (2017) 286. doi:10.3390/s17020286.
- [14] H. Garn, B. Kohn, C. Wiesmeyr, K. Dittrich, M. Wimmer, M. Mandl, G. Kloesch, M. Boeck, A. Stefanic, S. Seidel, 3D detection of the central sleep apnoea syndrome, *Curr. Dir. Biomed. Eng.* 3 (2017) 829–833. doi:10.1515/cdbme-2017-0174.
- [15] C. Yang, G. Cheung, V. Stankovic, K. Chan, N. Ono, Sleep Apnea Detection via Depth Video and Audio Feature Learning, *IEEE Trans. Multimed.* 19 (2017) 822–835. doi:10.1109/TMM.2016.2626969.
- [16] C.W. Wang, A. Hunter, N. Gravill, S. Matusiewicz, Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea, *IEEE Trans. Biomed. Eng.* 61 (2014) 396–404. doi:10.1109/TBME.2013.2280132.
- [17] S. Sharma, S. Bhattacharyya, J. Mukherjee, P.K. Purkait, A. Biswas, A.K. Deb, Automated detection of newborn sleep apnea using video monitoring system, 2015 Eighth Int. Conf. Adv. Pattern Recognit. (2015) 1–6. doi:10.1109/ICAPR.2015.7050675.
- [18] J. Jorge, M. Villarroel, S. Chaichulee, A. Guazzi, S. Davis, G. Green, K. McCormick, L. Tarassenko, Non-Contact Monitoring of Respiration in the Neonatal Intensive Care Unit, *Proc. - 12th IEEE Int. Conf. Autom. Face Gesture Recognit.* (2017) 286–293. doi:10.1109/FG.2017.44.
- [19] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, F. Pisani, Monitoring Infants by Automatic Video Processing: A Unified Approach to Motion Analysis, *Comput. Biol. Med.* 80 (2017) 158–165. doi:10.1016/j.compbimed.2016.11.010.
- [20] N. Koolen, O. Decroupet, A. Dereymaeker, K. Jansen, J. Vervisch, V. Matic, B. Vanrumste, G. Naulaers, S. Van Huffel, M. De Vos, Automated Respiration Detection from Neonatal Video Data, *Proc. 4th Int. Conf. Pattern Recognit. Appl. Methods.* (2015) 164–169. doi:10.5220/0005187901640169.
- [21] M.H. Li, A. Yadollahi, B. Taati, Noncontact Vision-Based Cardiopulmonary Monitoring in Different Sleeping Positions, *IEEE J. Biomed. Heal. Informatics.* 21 (2017) 1367–1375. doi:10.1109/JBHI.2016.2567298.
- [22] S. Kalitzin, G. Petkov, D. Velis, B. Vledder, F. Lopes da Silva, Automatic segmentation of episodes containing epileptic clonic seizures in video sequences., *IEEE Trans. Biomed. Eng.* 59 (2012) 3379–85. doi:10.1109/TBME.2012.2215609.
- [23] B.K.P. Horn, B.G. Schunck, Determining optical flow, *Artif. Intell.* 17 (1981) 185–203. doi:10.1016/0004-3702(93)90173-9.
- [24] A. de Groote, M. Wantier, G. Cheron, M. Estenne, M. Paiva, Chest wall motion during tidal breathing, *J. Appl. Physiol.* 83 (1997) 1531–1537. doi:10.1152/jappl.1997.83.5.1531.
- [25] C. V. Senaratna, J.L. Perret, C.J. Lodge, A.J. Lowe, B.E. Campbell, M.C. Matheson, G.S. Hamilton, S.C. Dharmage, Prevalence of obstructive sleep apnea in the general population: A systematic review, *Sleep Med. Rev.* 34 (2017) 70–81. doi:10.1016/j.smrv.2016.07.002.
- [26] E.E. Geertsema, R.D. Thijs, T. Gutter, B. Vledder, J.B. Arends, F.S. Leijten, G.H. Visser, S.N. Kalitzin, Automated video-based detection of nocturnal convulsive seizures in a residential care setting, *Epilepsia.* 59 (2018) 53–60.

# Appendix

**Supplementary Table 1.** Algorithm test results per camera for each subject. A dash in both sensitivity columns for relevant and all apneas means that the subject was not in view of that camera during the apnea(s). A dash in only the relevant apneas column implies that the apnea(s) registered from that camera angle were non-relevant. Cameras in which performance was best are underlined per subject.

ID	Camera	Sensitivity (%) relevant apneas (N apneas)	Sensitivity (%) all apneas (N apneas)	False positive rate (/hour)	Specificity for events (%)
SIM1	<u>1</u>	87.5 (8)	77.8 (9)	0	100
	2	87.5 (8)	77.8 (9)	0	100
	3	100 (8)	88.9 (9)	0	100
	4	33.3 (6)	22.2 (9)	0	100
	late integrated	100 (8)	88.9 (9)	0	100
	early integrated	100 (8)	88.9 (9)	0	100
PAT1	1	- (0)	- (0)	0	100
	<u>2</u>	- (0)	0 (1)	0	100
	3	- (0)	0 (1)	0	100
	late integrated	- (0)	0 (1)	0	100
	early integrated	- (0)	0 (1)	0	100
PAT2	1	- (0)	0 (1)	3.4	98.6
	<u>2</u>	- (0)	0 (1)	3.4	98.7
	3	- (0)	0 (1)	3.4	98.6
	late integrated	- (0)	0 (1)	3.4	98.9
	early integrated	- (0)	0 (1)	3.4	98.7
PAT3	<u>1</u>	- (0)	0 (1)	0	100
	2	- (0)	0 (1)	2.3	99.0
	<u>3</u>	- (0)	0 (1)	0	100
	4	- (0)	0 (1)	2.3	99.0
	late integrated	- (0)	0 (1)	4.5	98.5
	early integrated	- (0)	0 (1)	2.3	99.0
PAT4	<u>1</u>	100 (1)	100 (1)	6.8	97.2
	2	0 (1)	0 (1)	2.3	99.1
	3	- (0)	0 (1)	9.0	96.5
	4	- (0)	- (0)	6.8	97.1
	late integrated	100 (1)	100 (1)	22.6	92.5
	early integrated	100 (1)	100 (1)	0	100
PAT5	1	100 (2)	33.3 (9)	2.3	99.0
	<u>2</u>	100 (2)	33.3 (9)	0	100
	3	- (0)	0 (9)	2.3	99.0
	late integrated	100 (2)	44.4 (9)	4.5	98.3
	early integrated	100 (2)	33.3 (9)	2.3	98.9
PAT6	1	- (0)	- (0)	0	100
	<u>2</u>	100 (1)	100 (1)	0	100
	<u>3</u>	100 (1)	100 (1)	0	100
	4	- (0)	0 (1)	0	100
	late integrated	100 (1)	100 (1)	0	100
	early integrated	0 (1)	0 (1)	0	100
PAT7	1	- (0)	0 (2)	0	100
	<u>2</u>	- (0)	100 (2)	0	100
	3	- (0)	100 (2)	6.8	97.0
	late integrated	- (0)	100 (2)	6.8	97.4
	early integrated	- (0)	0 (2)	0	100
PAT8	1	100 (1)	100 (1)	9.0	96.2
	<u>2</u>	100 (1)	100 (1)	0	100
	<u>3</u>	100 (1)	100 (1)	0	100
	4	- (0)	0 (1)	0	100
	late integrated	100 (1)	100 (1)	9.0	96.7
	early integrated	100 (1)	100 (1)	9.0	96.2
PAT9	<u>1</u>	100 (1)	100 (1)	0	100
	2	- (0)	0 (1)	4.5	98.2
	3	- (0)	100 (1)	4.5	98.1
	4	100 (1)	100 (1)	2.3	99.1
	late integrated	100 (1)	100 (1)	6.8	97.7
	early integrated	100 (1)	100 (1)	0	100
Grouped	all individual cameras	83.7 (43)	46.7 (92)	2.2	99.1
	best choice	100 (14)	63.0 (27)	1.1	99.5



worst choice	80.0 (10)	37.0 (27)	3.3	98.5
late integrated	100 (14)	66.7 (27)	6.6	97.6
early integrated	92.9 (14)	51.9 (27)	1.6	99.3

---