Chronobiology International, Early Online: 1–20, (2012)

ISSN 0742-0528 print/1525-6073 online
DOI: 10.3109/07420528.2012.700669

informa
healthcare

An Improved Method for Estimating Human Circadian Phase Derived From Multichannel Ambulatory Monitoring and Artificial Neural Networks

Vitaliy Kolodyazhniy,[1] Jakub Späti,[2] Sylvia Frey,[3] Thomas Götz,[4] Anna Wirz-Justice,[3] Kurt Kräuchi,[3] Christian Cajochen,[3] and Frank H. Wilhelm[1]

[1]Division of Clinical Psychology, Psychotherapy and Health Psychology, Institute for Psychology, University of Salzburg, Salzburg, Austria, [2]Preclinical Laboratory for Translational Research into Affective Disorders (PLaTRAD), Clinic for Affective Disorders and General Psychiatry, Psychiatric University Hospital Zurich, Zurich, Switzerland, [3]Centre for Chronobiology, Psychiatric Hospital of the University of Basel, Basel, Switzerland, [4]Central Admission Unit, Psychiatric Hospital of the University of Basel, Basel, Switzerland

Recently, we developed a novel method for estimating human circadian phase with noninvasive ambulatory measurements combined with subject-independent multiple regression models and a curve-fitting approach. With this, we were able to estimate circadian phase under real-life conditions with low subject burden, i.e., without need of constant routine (CR) laboratory conditions, and without measuring standard circadian markers, such as core body temperature (CBT) or pineal hormone melatonin rhythms. The precision of ambulatory-derived estimated circadian phase was within an error of 12 ± 41 min (mean ± SD) in comparison to melatonin phase during a CR protocol. The physiological measures could be reduced to a triple combination: skin temperatures, irradiance in the blue spectral band of ambient light, and motion acceleration. Here, we present a nonlinear regression model approach based on artificial neural networks for a larger data set (25 healthy young males), including both the original data and additional data collected in the same protocol and using the same equipment. Throughout our validation study, subjects wore multichannel ambulatory monitoring devices and went about their daily routine for 1 wk. The devices collected a large number of physiological, behavioral, and environmental variables, including CBT, skin temperatures, cardiovascular and respiratory functions, movement/posture, ambient temperature, spectral composition and intensity of light perceived at eye level, and sleep logs. After the ambulatory phase, study volunteers underwent a 32-h CR protocol in the laboratory for measuring unmasked circadian phase (i.e., "midpoint" of the nighttime melatonin rhythm). To overcome the complex masking effects of many different confounding variables during ambulatory measurements, neural network–based nonlinear regression techniques were applied in combination with the cross-validation approach to subject-independent prediction of circadian phase. The most accurate estimate of circadian phase with a prediction error of −3 ± 23 min (mean ± SD) was achieved using only two types of the measured variables: skin temperatures and irradiance for ambient light in the blue spectral band. Compared to our previous linear multiple regression modeling approach, motion acceleration data can be excluded and prediction accuracy, nevertheless, improved. Neural network regression showed statistically significant improvement of variance of prediction error over traditional approaches in determining circadian phase based on single predictors (CBT, motion acceleration, or sleep logs), even though none of these variables was included as predictor. We, therefore, have identified two sets of noninvasive measures that, combined with the prediction model, can provide researchers and clinicians with a precise measure of internal time, in spite of the masking effects of daily behavior. This method, here validated in healthy young men, requires testing in a clinical or shiftwork population suffering from circadian sleep-wake disorders. (Author correspondence: vitaliy.kolodyazhniy@sbg.ac.at)

Keywords: Ambient light, Cross-validation, Curve fitting, Human circadian rhythms, Melatonin, Multichannel ambulatory recording, Multilayer perceptron, Skin temperatures

INTRODUCTION

The growing understanding of the function of circadian rhythms in human physiology and pathology has also led to emphasis on the important role of the external day-night cycle to daily synchronize internal with external time. Entrainment of endogenous circadian rhythms to the 24-h solar day (for a review see Roenneberg et al., 2003) is clearly evident in the daily rhythms of molecular, endocrinological, physiological, and

Submitted March 9, 2012, Returned for revision April 6, 2012, Accepted June 1, 2012

Address correspondence to Vitaliy Kolodyazhniy, Institute for Psychology, University of Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria. E-mail: vitaliy.kolodyazhniy@sbg.ac.at

psychological measures. These range from gene expression, core body temperature (CBT), heart rate, secretion of the pineal hormone melatonin and cortisol, to sleep timing, mood, as well as higher cognitive functions (for a review see Schmidt et al., 2007). Variables such as CBT, melatonin, and cortisol rhythms have been widely used to estimate parameters of the circadian system, such as phase, amplitude, and endogenous period, under controlled laboratory conditions (Klerman et al., 2002). However, in daily life situations such estimation is complicated by "masking" (Hiddinga et al., 1997; Minors & Waterhouse, 1989) induced by sleep, physical activity, meals, emotional activation, light, etc., all of which have different modifying effects on physiological variables (Wilhelm & Grossman, 2010; Wilhelm et al., 2006), and are also time-of-day dependent. The circadian profile of melatonin secretion is considered to be the most accurate circadian marker with least variability compared to CBT and cortisol (Klerman et al., 2002). Widespread use of this assay is jeopardized by the need for carefully controlled lighting conditions and the expense of multiple samples. CBT is more susceptible to masking effects than melatonin or cortisol and is measured via rectal probes, increasing subject burden.

Many researchers have discussed the complexity of estimating the relationship between measured data and endogenous circadian phase (Carrier & Monk, 1997; Eastman, 1992; Klerman et al., 1999; Minors & Waterhouse, 1989; Wever, 1979). They used only a limited number of physiological variables with demasking techniques that was univariate or based on relatively simple heuristics (e.g., subtracting or adding a value to CBT derived from the subject's activity) (Klerman et al., 1999).

One approach towards accurate estimation of endogenous circadian phase using ambulatory data would require multiple physiological and behavioral parameters as well as measurement of ambient light as the "zeitgeber" that contributes most to the entrainment of the human circadian clock. The relation between ambulatory measurements with their masking components and endogenous phase measured in the laboratory under constant routine (CR) conditions could be established using a "gold standard" circadian marker such as melatonin. A representative data set would identify which ambulatory variables provide most accurate estimation of endogenous circadian phase using statistical techniques. A prediction model would then allow estimation of circadian phase in real-life conditions using the minimal number of noninvasive variables, and no longer require laboratory validation.

In the European Union (EU) Framework Programme 6 integrated project EUCLOCK (2006–2011; www.euclock.eu), we developed an ambulatory multichannel circadian physiology monitoring system (www.phi-i.com) and a portable light sensor device (www.object-tracker.com) that was used to gather data from a group of 16 subjects with a broad range of chronotypes (Kolodyazhniy et al., 2011b). Subject-independent multiple regression models combined with curve-fitting procedures (Van Someren & Nagtegaal, 2007) and the center of gravity method (Wetterberg, 1998) allowed a rather precise prediction of circadian phase under ambulatory real-life conditions compared to the circadian phase of melatonin secretion in the gold standard CR protocol (Kolodyazhniy et al., 2011b).

However, the linear prediction models (Kolodyazhniy et al., 2011b), although simple and transparent, are too restrictive for estimating the complex relation between the endogenous circadian rhythm and ambulatory measurements of physiological, behavioral, and environmental variables, because the underlying relation between circadian rhythm and measured predictor variables can be substantially nonlinear. In the present work, we propose nonlinear models for predicting the circadian rhythm, as determined from melatonin secretion profile during CR, based on artificial neural networks (ANNs; Haykin, 1999) in combination with curve fitting and the center of gravity technique for phase detection (Kolodyazhniy et al., 2011b). ANNs are widely used for pattern recognition problems, such as regression and data classification (Bishop, 2006), and also for physiological data (Kolodyazhniy et al., 2007, 2011a), including applications related to sleep, fatigue, and alertness (Kiymik et al., 2004; Reifman, 2004; Vuckovic et al., 2002). To the best of our knowledge, the problem of detecting circadian phase from ambulatory physiological measurements based on multiple regression modeling as first addressed in our previous paper (Kolodyazhniy et al., 2011b) has not yet been solved using the ANN approach. In the present paper, we describe improved (nonlinear) subject-independent prediction models as well as significantly improved precision of estimation of endogenous circadian phase from ambulatory data. We attained our goal of reducing the number of measured variables to a minimal, noninvasive set that is practicable in a clinical or research setting.

MATERIALS AND METHODS

Protocol

The protocol was the same as described in (Kolodyazhniy et al., 2011b), but is here summarized briefly. Figure 1 illustrates the overall design of the ambulatory circadian monitoring validation study using a CBT data set. The protocol involved an ambulatory part of ~7 d under real-life conditions. In the subsequent laboratory part, an 8-h adaptation night of sleep at habitual bedtime was followed by 32 h of sustained wakefulness (CR).

Ambulatory measurements always began on a Friday between 11:00 and 17:00 h. During the 7-d ambulatory episode, subjects kept their usual bedtime regimen, which was registered by actimetry and sleep logs. In addition, they kept a daily log about their subjective well-being, sleepiness, etc., and electrocardiography (ECG), rectal temperature, and respiration were recorded

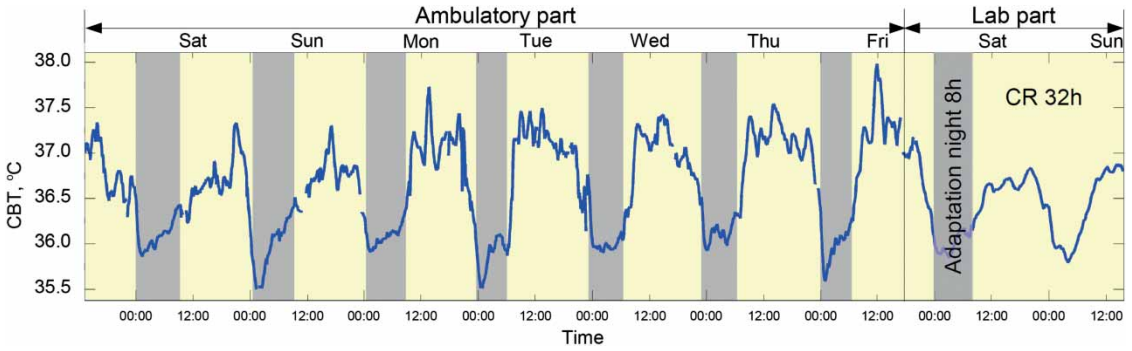

FIGURE 1. Overall protocol design and typical CBT measurements for one representative study volunteer: dark bars correspond to bedtimes; ticks on *x*-axis correspond to midnight and midday. In-laboratory part of the study protocol started on a Friday evening, scheduled according to the subject's usual bedtimes.

continuously with a wearable multichannel ambulatory monitoring device, and skin temperatures with miniature autonomous loggers. Ambient light was recorded by a miniature light sensor attached to the side of eyeglasses (either subjects' own or zero-dioptric with a normal frame provided by the investigators).

All ambulatory recordings continued in the laboratory part of the study that started on the following Friday between 19:00 and 21:00 h, and ended on Sunday between 15:00 and 18:00. Individual timing of the adaptation night (8 h) and of the following CR (32 h) was scheduled according to the sleep midpoint during the ambulatory week, as determined from sleep logs. Posture, ambient light (≤8 lux), and temperature levels were kept constant; food and water was distributed uniformly across the CR protocol in order to minimize masking effects and to enable us to quantify phase and amplitude of the circadian markers, as detailed in Cajochen et al. (2001).

The primary circadian marker was melatonin assayed in saliva samples collected every 60 min throughout the CR. CBT was continuously recorded throughout the entire 9-d protocol, and these data were used for an additional comparison of predictions of circadian phase provided by the proposed model, as were sleep logs acquired via electronic diaries.

The study protocol was approved by the local ethical committee (Ethikkommission beider, Basel), conformed to the Declaration of Helsinki, and met the ethical standards of the Journal (Portaluppi et al., 2010).

Subjects

Thirty male study volunteers of different chronotypes (assessed as midsleep on free days corrected for sleep deficit [MSFsc] using the Munich chronotype questionnaire [Roenneberg et al., 2007]) were included in the study. Only healthy nonsmoking men 19 to 35 yrs old, who lived on a normally entrained 24-h schedule and had no self-reported sleep problems, were selected. For practical reasons related to additional masking by the menstrual cycle, only men were studied. Participants signed a written informed consent form before being enrolled, and underwent a medical examination at the University Psychiatric Clinics (UPK) Basel prior to participation.

Data collection was started in July 2008. Our previous publication (Kolodyazhniy et al., 2011b) utilized data from 16 subjects collected until November 2009. Data from additional nine subjects were collected and the validation study finished in October 2010. The total number of participants included in the present analyses was N = 25, and the overall success rate in our demanding protocol was 83% (25 out of 30 participants, taking into account five subjects for whom data collection was not complete; see Kolodyazhniy et al., 2011b).

Ambulatory Monitoring Devices

Two ambulatory circadian physiology monitoring devices were used, "ClockWatcher" (Phi-I, Amsterdam) and Varioport (Becker Meditec, Karlsruhe) (see Figure 2 in Kolodyazhniy et al., 2011b). CBT, ECG, respiration, body movement, and posture were measured continuously throughout the protocol. For details of the ambulatory monitoring devices, our circadian software, and data organization, see Supplementary Material.

A specialized miniature ambulatory monitoring device "LightWatcher" (www.object-tracker.com) was also developed to record light in five spectral bands (infrared, red, green, blue, ultraviolet [UV]), three-dimensional (3D) acceleration, and ambient temperature (Figure 2 in Kolodyazhniy et al., 2011b). Eleven miniature wireless temperature sensors (Thermochron iButtons; Maxim, San Jose, CA, USA) were used to continuously record skin temperatures on hands and feet (distal) and upper and lower legs, shoulders, and thorax (proximal) throughout the protocol (Figure 2 in Kolodyazhniy et al., 2011b) as detailed in Supplementary Material.

An electronic diary was used for sleep logs and a number of other scales (see Supplementary Material). The procedures of checking and transferring the recorded data to a database for further analysis are explained in Supplementary Material.

Reference Circadian Phase

As previously reported (Kolodyazhniy et al., 2011b), the reference circadian phase for model development was based on the circadian rhythm of melatonin secretion

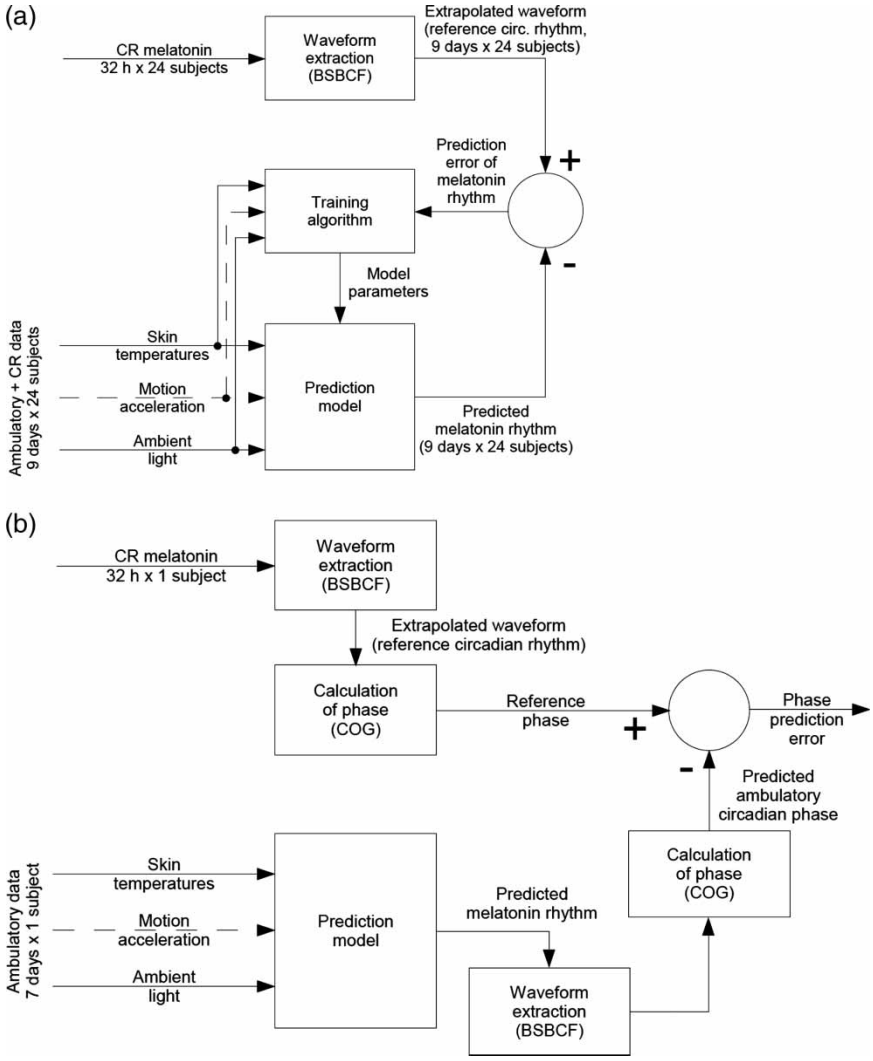

FIGURE 2. Modeling approach with cross-validation: (A) identification of a prediction model with ambulatory and CR data from 24 subjects (training); (B) validation of the prediction model with data from another subject who was not included in the training data set using ambulatory data only, and CR melatonin for comparison of predicted phase with reference phase. BSBCF = bimodal skewed baseline cosine function; COG = center of gravity method. The process is iterated 25 times such that the predictions are validated for each subject. Note: the training algorithm (least squares regression for linear models and resilient propagation for neural networks) determines the parameters of the prediction model via minimizing the sum of squared errors between the predicted melatonin rhythm and the extrapolated BSBCF waveform; the best neural network models (MLP115) did not include the motion variable.

during the 32-h CR protocol. Melatonin was used as a reference for the following reasons: it is known to be the most reliable circadian marker when measured under dim light conditions, also compared to CBT (Klerman et al., 2002). Reliable measurements of melatonin secretion were available for all 25 subjects throughout the CR, whereas the CBT measurements for 4 subjects contained artifacts. For one subject, the CBT measurements in the laboratory were not available for more than half of the CR due to a technical problem.

First, the reference waveform of melatonin secretion was determined based on melatonin levels from a total of 33 saliva samples taken every 60 min under CR conditions. This waveform was assumed to have a period of 24 h (i.e., entrained circadian period), and was identified by a bimodal skewed baseline cosine function (BSBCF; Van Someren & Nagtegaal, 2007). For details, see "Waveform Analysis" in Supplementary Material. For regression modeling, the BSBCF waveform of CR melatonin rhythm was extrapolated backwards on to the preceding days of the ambulatory part and normalized on to [0, 1] as in Kolodyazhniy et al. (2011b). The extrapolated melatonin waveform was used as the dependent variable and ambulatory measurements as independent variables for fitting prediction models of the circadian rhythm (linear regression and neural networks; see Figure 2A and description below).

The reference circadian phase was determined as the time corresponding to the center of gravity (COG) of the area (Wetterberg, 1998) under the periodic BSBCF curve for one period of 24 h (Kolodyazhniy et al., 2011b; see also "Calculation of Circadian Phase" in Supplementary Material). With the COG method, there is no need for thresholds to be defined for determining circadian

phase from the BSBCF curve, because the threshold is always equal to the baseline (mean level of melatonin after secretion offset and before onset; see Figure S1), in contrast to the original work (Van Someren & Nagtegaal, 2007).

Linear Multiple Regression Modeling

Estimation of ambulatory circadian phase in our previous work (Kolodyazhniy et al., 2011b) was based on linear multiple regression models for evaluation of circadian rhythm of CR melatonin from ambulatory measurements and a combination of the BSBCF, COG, and waveform extrapolation approaches mentioned above.

The best prediction of circadian rhythm was achieved with multiple regression models, including the skin temperatures (with lags from 0 to 5 h in 30-min steps), motion, and irradiance in the blue spectral band (both with lags from 0 to 24 h in 30-min steps). CBT was intentionally not included in the prediction variable set, because the goal of this validation study was to develop a device that imposes minimal subject burden by not using rectal probes. The linear models contained in total 164 predictor variables and 165 regression parameters counting all lags and the bias term.

To ensure subject-independent validation of the prediction models, we used the same cross-validation approach (Hastie et al., 2001) as before (Kolodyazhniy et al., 2011b). Briefly, we iteratively left one of the subjects (1 through 25) out of the data set and used the data from the other 24 subjects for identifying a prediction model via least squares regression, whereas the remaining subject was used for validation of the model. For determining the parameters of linear prediction models of the melatonin *rhythm*, data including both the ambulatory and laboratory parts from the 24 subjects were used (see Figure 2A). During validation (Figure 2B), only ambulatory data of the one remaining subject were used, and another BSBCF curve was fitted to the predicted melatonin rhythm for that subject. The predicted ambulatory *phase* was determined from this fitted BSBCF curve using the COG method (see "Calculation of Circadian Phase" in Supplementary Material) and compared to the reference phase of the same subject. Thus, information on the reference phase from CR melatonin was used only for fitting the models and estimation of performance of the subject-independent models during validation, but not for prediction of ambulatory circadian phase, itself. In the cross-validation loop, we identified 25 subject-independent linear multiple regression models, which can predict the melatonin rhythm from ambulatory data of subjects not included in the training data set without any information from CR melatonin. Using the BSBCF and COG methods, we derive the ambulatory circadian phase from the predicted melatonin rhythm. For more details, see the sections "Linear Prediction Models" and "Cross-validation" in Supplementary Material.

Neural Network–Based Nonlinear Modeling

An artificial neural network (ANN) consists of a number of elementary processing units called artificial neurons, which work in parallel in a way inspired by the parallel information processing in living brains. The desired input-output mapping of an ANN is realized via the so-called training algorithm based on input-output examples (input patterns and desired responses) from a representative training data set. In such a way, an ANN learns to approximate the distribution of data in the training set and to generalize for unknown data outside of the training set for an arbitrary unknown nonlinear function (Cybenko, 1989; Hornik, 1991).

We employ the most widely used type of artificial neural networks called "multilayer perceptron," or MLP (Haykin, 1999). Each artificial neuron in an MLP resembles a logistic regression model (Hastie et al., 2001), i.e., computes a weighted sum of its inputs plus a bias term and applies the so-called sigmoid activation function to the result to produce an output between 0 and 1 (for the popular logistic sigmoid activation function). Combining multiple neurons in layers and applying proper training algorithms, complex nonlinear input-output mappings can be learned based on the presented training data without explicitly knowing the underlying functional relation between the inputs and the output. For further detailed information see Haykin (1999) and Bishop (2006).

For our implementation of the MLP neural networks, we used the scientific computation package MATLAB R2011b (The MathWorks, Natick, MA, USA) with the Neural Network Toolbox. We applied neural network models with two different sets of input variables. The first set was exactly the same as for the previously described linear model, i.e., it included the same 164 variables (counting all lags). This model will be referred to as "MLP164" in the text below. The second set of input variables did not contain the motion variable, and the total number of input variables of MLP models was 115 (six skin temperature variables with lags from 0 to 5 h in .5-h steps plus the ambient blue light with lags from 0 to 24 h in .5-h steps). We subsequently call this model "MLP115" (see Figure 3 for a graphic representation of its structure).

The reason for exclusion of the motion variable was an attempt to develop a more parsimonious model with fewer parameters that might also lead to better prediction accuracy. We argued that the motion variable might be redundant in the nonlinear model and prevent improving accuracy, because the circadian information contained in the "zeitgeber" signal representing blue light and in the skin temperature rhythms might be sufficient if it can be captured by the nonlinear ANN approach as opposed to the previous linear technique.

Circadian phase was determined from the predicted melatonin rhythm via the same combination of the BSBCF and COG approaches as with the linear regression models from Kolodyazhniy et al. (2011b) described

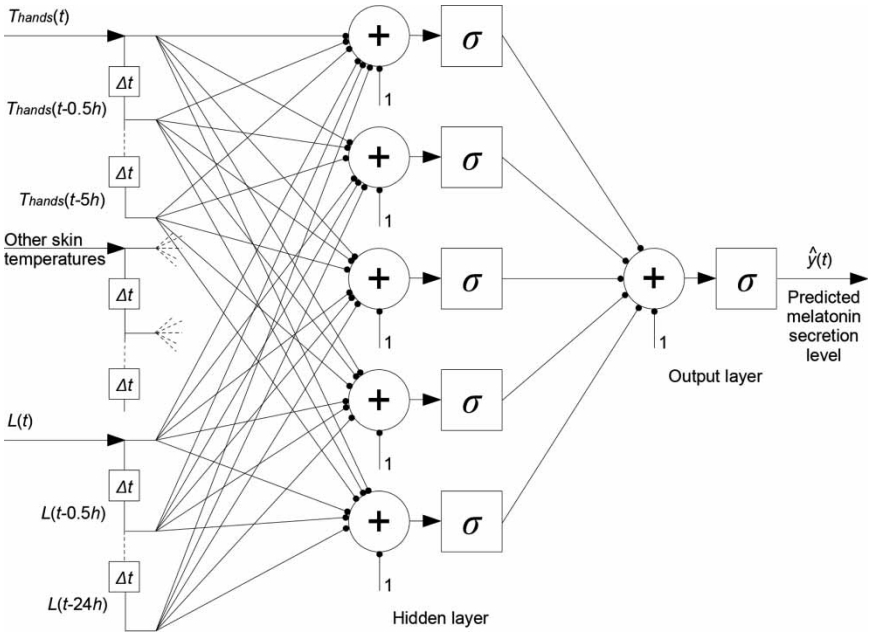

FIGURE 3. Best neural network prediction model for circadian rhythm of melatonin secretion: two-layer perceptron with five neurons in the hidden layer and sigmoid activation functions in both layers; dots (•) represent the adjustable weights of the neural network; circles with " + " are summation units; large square blocks are logistic sigmoid activation functions $\sigma(x) = 1/(1 + \exp(-x))$; small square blocks represent input lags with time step of $\Delta t = .5$ h; ones ("1") stand for constant bias inputs. The input variables (115 in total) include six skin temperatures (T_{hands}, T_{feet}, T_{thorax}, $T_{\text{shoulders}}$, $T_{\text{upper legs}}$, and $T_{\text{lower legs}}$) with lags up to 5 h and light variable L of irradiance in the blue spectral band with lags of up to 24 h.

above. The cross-validation setting with 25 subjects was also the same (see "Linear Multiple Regression Modeling" above and "Cross-validation" in Supplementary Material), i.e., both the ambulatory and laboratory data from 24 subjects were used for determining the parameters of subject-independent ANN prediction models of the melatonin *rhythm* (Figure 2A), and only the ambulatory data of the one remaining subject for the prediction of melatonin rhythm and its circadian phase during validation without using any data from the laboratory for the prediction itself (Figure 2B).

Both the MLP164 and MLP115 models contained five neurons in the hidden layer and one neuron in the output layer, all with logistic sigmoid activation function, and 831 and 586 adjustable weights, respectively. The weights were trained with the resilient propagation training algorithm (function "trainrp") of the MATLAB Neural Network Toolbox for 100 "epochs." The ratio of number of data points in the training sets to the number of adjustable weights was about 11:1 for MLP164 and about 15:1 for MLP115, which was quite reasonable in order to avoid undesired overfitting effects. For both MLP115 and MLP164, the 25-fold cross-validation procedures including the training of the neural networks were repeated 100 times, each time with a different random initialization of the MLP weights. This is the standard method of estimating performance of neural networks and selecting the best-performing models (Bishop, 2006; Haykin, 1999). More details are presented in the section "Neural Network–Based Prediction Models" in Supplementary Material.

Note: In a practical application after identifying the best random initial weights of the MLP model from the best cross-validation run, the MLP prediction model should be trained once again using these best initial weights and the *entire data set without iteratively splitting it into training and validation sets*. In this way, one obtains a *single prediction model* instead of multiple models from all iterations of cross-validation (one single MLP instead of 25 different MLPs in our case). This model can be further tested with another independent data set collected from subjects with a representative range of chronotypes (e.g., 10–20 subjects), including ambulatory data for prediction and CR melatonin for comparison of the predicted and actual circadian phases. At this stage, no training would be required. Afterwards, the obtained model can be used for prediction of ambulatory circadian phase without CR melatonin or retraining for any further subjects.

Comparison of Prediction of Ambulatory Circadian Phase With Different Methods

For comparison, we also computed predictions of ambulatory circadian phase using the three variables most often used as circadian markers: CBT, motion, and the sleep midpoint (as in Kolodyazhniy et al., 2011b). For CBT and motion, circadian phase was determined by harmonic regression analyses of ambulatory recordings as the minimum of a waveform with a period of 24 h. For both CBT and motion, one, two, and three harmonics were applied, and the best results (lowest standard deviation of prediction errors for CR melatonin phase) were

attained with the one harmonic regression for CBT and two harmonics for motion. Sleep midpoint was computed using the respective entries for "Lights On" and "Lights Off" in the electronic diary during the ambulatory part of the experiment.

In order to learn if the proposed prediction models provided a statistically significant improvement over the single predictors, we performed the Pitman-Morgan test of differences in variance (Mudholkar et al., 2003) of predictions by different methods. Compared to the original test, we computed one-tailed statistics in order to determine if the variance of prediction error of one method is smaller than the variance of the other method, rather than simply testing the inequality of the two. One-sided Pitman-Morgan tests are a reasonable way of testing the direction of difference in variance of two samples (Chen et al., 1996), and thus model fit improvement.

RESULTS

We studied a range of chronotypes, shown in Figure 4 as a distribution of circadian phases of study participants as determined by CR melatonin midpoint. Data for CR melatonin, blue light exposure, CBT, motion acceleration, and shoulder and foot skin temperatures for the subjects with the earliest and the latest phase of entrainment as determined by CR melatonin are shown in Figure 5. The earliest and latest phases were 01:45 and 08:29 h, respectively. Table 1 provides information on the quality of recordings from our monitoring devices as percentage of time with acceptable data as used by our prediction models.

In the analyses described in this paper, we used the data set from our previous publication (Kolodyazhniy et al., 2011b) as well as data from nine new study participants. Table 2 depicts the results of comparisons between the best neural network–based prediction models MLP115 and MLP164 (based on 100 cross-validation runs with different random initializations each), the previously developed linear multiple regression approach, and conventional approaches using CBT, sleep logs, and motion. The screening information on individual chronotype is presented as midsleep on free days corrected for sleep deficit (MSFsc; see Roenneberg et al., 2007). Additionally, mean bed- and wake-up times during the ambulatory part of the study with respective standard deviations are shown to give an idea of the variability in sleep timing of study participants with different chronotypes.

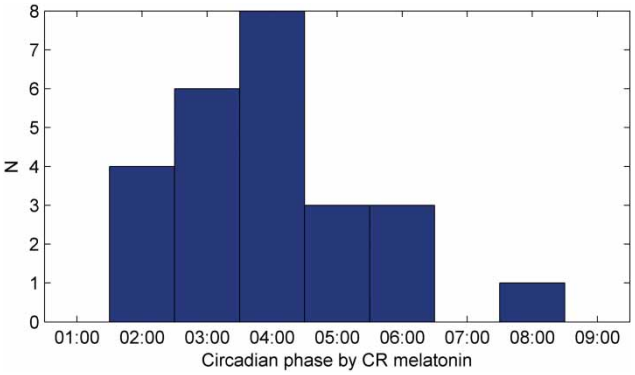

FIGURE 4. Distribution of individual circadian phase as assessed by the melatonin midpoint during the CR protocol of each study participant.

As can be seen from Table 2, the best prediction is achieved by the MLP115 models. This result was obtained in run 38 out of 100 cross-validation runs for this model. For both MLP115 and MLP164, we used the same sequence of the seed value of the random number generator of 0 to 99 corresponding to the run number minus one. The regression models MLP115 provided the most accurate prediction of circadian phase of CR melatonin, with the lowest mean prediction error, lowest standard deviation of error, highest correlation to CR melatonin phase, and smallest error range. The prediction error was determined as the reference phase of CR melatonin minus the predicted phase by the respective method. Interestingly, we found no association of prediction error with gaps in the data, as the subjects with the largest gaps (as given in the section "Interpolation of Missing Data" in Supplementary Material) were not those with the largest prediction error (see Table 2). The minimum amount of data for accurate predictions can be investigated separately, also for different chronotypes and different variability in sleep and activity timing.

For the linear prediction approach, we obtained a very similar estimate of prediction error (11 ± 42 min, mean ± SD) for the larger number of subjects (N = 25) as in our previous publication (Kolodyazhniy et al., 2011b), which was 12 ± 41 min for N = 16.

We compared the improvement of variance of the predicted circadian phase by the neural network approach over the linear models, CBT, sleep midpoint determined from the sleep logs, and motion acceleration using the one-sided Pitman-Morgan with correction for multiple comparisons according to the false discovery rate method (Curran-Everett, 2000). Thus, systematic offsets did not affect our comparison, although they are also reported in Table 2 as mean prediction errors for all methods. For the false discovery rate of $f_F = .05$ and the total of 803 comparisons (Table 3), 36% of all MLP115 models proved to have significantly lower variance of prediction error over the linear models, 25% over CBT, 70% over sleep midpoint, and 85% over motion acceleration. The error variance of MLP164 models was significantly smaller than that of the linear models in 17% cases, in 4% than of CBT, and in 93% cases than of sleep midpoint, and in 97% than of motion acceleration. The linear models, as previously (Kolodyazhniy et al., 2011b), proved to be significantly better than sleep midpoint and motion acceleration, but not CBT.

To estimate the improvement of MLP115 models over MLP164, we computed the one-sided Pitman-Morgan

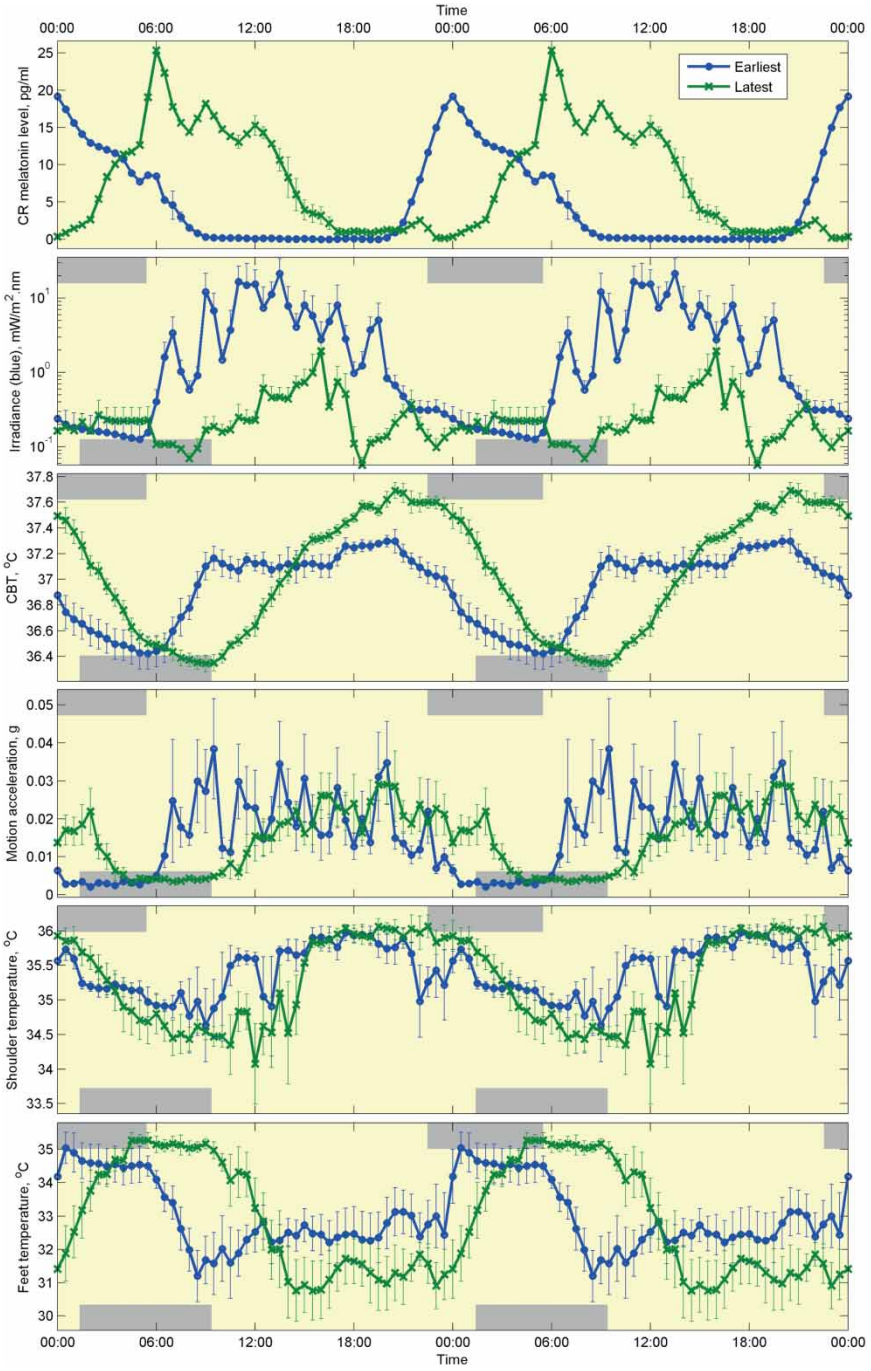

FIGURE 5. Salivary melatonin levels during a 32-h CR protocol (top graph) aligned with ambulatory data for 7 d (blue light irradiance, CBT, motion, shoulder and foot temperatures): double-plotted 30-min means ± SEM for two individual subjects with the earliest and the latest phase of entrainment. Gray bars represent average sleep times for the earliest and latest subject (top and bottom of ambulatory data plots, respectively).

test on the pooled prediction errors from 33 best cross-validation runs of MLP115 and MLP164 ranked by the standard deviation of prediction errors of the respective run. The test result was highly significant with $p < 10^{-8}$, i.e., the MLP models with 115 predictors were significantly better than all the other compared approaches.

In order to test if the accuracy of prediction can be further improved by reducing the number of skin temperatures, or if the number of skin temperature sensors can be reduced without sacrificing prediction accuracy, we tried iterative exclusion of each skin temperature variable for both MLP115 and MLP164 models. In each case, this led to increased standard deviation of prediction error. Therefore, MLP models reported here contain the same number of skin temperature variables as the linear approach (Kolodyazhniy et al., 2011b). Similarly, we tried to remove the light variable from the model, but this resulted in dramatic increase of prediction error. Thus, we have confirmed that our predictor variables were the optimal ones.

TABLE 1. Percentage of time with acceptable data for all monitoring devices with respect to the entire time of the protocol averaged for all analyzed subjects ($N = 25$)

Device	% time with acceptable data
ClockWatcher and Varioport	
Rectal temperature	89.7
Motion	97.6
iButtons	
Skin temperatures	99.28
LightWatcher	
Ambient light	95.1

Predictions of the normalized circadian rhythm of melatonin for the two subjects with the earliest and the latest phase of entrainment as determined by CR melatonin phase are shown in Figure 6 for the best prediction model MLP115. Prediction of the ambulatory circadian rhythm is available for 6 d starting on the second day of the ambulatory part of the study, because of the maximum lag of 24 h in the prediction model. Figure 6 also shows the reference circadian rhythm in form of a BSBCF curve for CR melatonin extrapolated onto the ambulatory part and scaled to [0, 1], and the BSBCF curve extracted from the prediction. Phases of these BSBCF curves determined using the COG method determine the reference phase from CR melatonin and the predicted ambulatory circadian phase. Note that the similarity of the predicted ambulatory circadian rhythm and the extrapolated CR melatonin waveform.

Scatter plots demonstrating the ambulatory circadian phase as predicted by the different methods versus the reference phase computed from melatonin under CR conditions are shown in Figure 7. It can be easily seen the best prediction is provided by the neural network models without the motion acceleration variable (MLP115; upper left plot in Figure 7): the points in this plot lie closest to the identity line representing the ideal prediction of zero error.

DISCUSSION

Our new results indicate that reliable detection of circadian phase in real-life conditions is feasible using only two noninvasive measures—even without including the "classical" circadian marker CBT or motion (the rest-activity cycle) as a predictor (in the best model MLP115). The validation study presented here is based on multiple regression models (linear and nonlinear) and data gathered from a multichannel ambulatory monitoring system. The predicted circadian phase compared very accurately to that of melatonin obtained in the laboratory under CR conditions. Variables that were included in the prediction models were obtained during everyday life behavior from small devices, which imposed little subject burden, and compared to the same variables measured under CR laboratory conditions. Possible reasons for the new results that exclude motion, as in the original modeling approach (Kolodyazhniy et al., 2011b), are discussed below.

Skin temperatures exhibit periodic patterns throughout the day that are related to an underlying circadian rhythm in thermoregulation but also to homeostatic thermoregulatory adjustments to changes in ambient and body temperatures. They are affected by many masking factors, so the distal-proximal temperature gradient (DPG) is a better indicator of the endogenous circadian rhythm than distal or proximal skin temperatures, per se (Kräuchi, 2007; Kräuchi et al., 1999). We did not simply use a linear combination of skin temperatures with fixed weights of 1 or –1, as is the case for calculating the DPG, but included six separate skin temperature variables in the prediction model with lags of up to 5 h. Thus, the model revealed individual weights of the separate skin temperature variables for best prediction of the endogenous circadian rhythm. Furthermore, in the neural network model, the six skin temperature variables are combined in a nonlinear fashion, which would have contributed to improving prediction accuracy.

Light is the major zeitgeber responsible for entraining the circadian clock to the external day (Roenneberg et al., 2003). Thus, it is crucial to measure daily light exposure in order to predict the day-to-day variations in circadian phase, as shown in a long-term measurement single case study (Watanabe & Kripke, 1988). Not just light, per se, but the spectral composition of light received provides even more specific information to the circadian clock: light in the blue wavelength appears to be most important for circadian entrainment (Cajochen et al., 2005; Lockley et al., 2003; Smith et al., 2009). In our regression models, we included irradiance measurements for ambient light corresponding to the blue spectral band with lags corresponding to one complete 24-h day-night cycle. It was also confirmed by cross-validation in Kolodyazhniy et al. (2011b) that blue light was a better predictor of circadian phase than the other four spectral bands (infrared, red, green, and ultraviolet), the combinations of all, or all the visible spectral bands. Blue light was also the only variable in our model that reflected season via the intensity of lighting and day length, as in our previous work (Kolodyazhniy et al., 2011b).

Motion was included in our initial prediction model based on linear regression (Kolodyazhniy et al., 2011b). Motion reflects the circadian rest-activity rhythm and actimetry has become a standard technique for noninvasive, 24-h, long-term measurements of rest-activity cycles in health and illness (Ancoli-Israel, 2003; Wirz-Justice, 2007). In our previous analysis (Kolodyazhniy et al., 2011b), we included motion as a predictor in the regression model with lags corresponding to one complete 24-h period of activity. Besides that, we assumed that inclusion of the motion variable helped in demasking the endogenous circadian rhythm from the skin temperatures by separating the influence of rest-activity cycles if the latter were not in phase with the endogenous rhythm, e.g., due to variability in

TABLE 2. Comparison of accuracy of different methods for prediction of ambulatory circadian phase versus CR melatonin phase

						Prediction error (min) error = CR melatonin phase – predicted phase					
						25-fold cross-validation					
Subject no.	End date, dd.mm.yy	CR melatonin phase, hh:mm	MSFsc,[a] hh:mm (screening prior to study)	Ambulatory mean bedtime, hh:mm ± SD (min)	Ambulatory mean wake-up time, hh:mm ± SD (min)	MLP115, best	MLP164, best	Linear models	CBT	Sleep midpoint	Motion
1	13.07.08	3:52	4:53	23:57 ± 43	08:17 ± 126	−6	2	78	−38	−16	88
2	27.07.08	3:29	3:32	23:17 ± 46	08:00 ± 32	−28	8	33	−10	−10	14
3	17.08.08	3:54	5:30	00:58 ± 51	08:15 ± 25	−2	−31	−25	−82	−44	−65
4	31.08.08	5:43	5:47	00:43 ± 43	08:33 ± 45	41	20	87	−92	65	73
5	16.11.08	3:55	4:27	00:03 ± 78	07:45 ± 21	20	17	39	−68	1	23
6	18.01.09	5:58	7:15	02:20 ± 33	09:41 ± 72	−9	11	32	−18	−3	48
7	01.03.09	8:29 (latest)	6:18	02:38 ± 85	09:54 ± 98	36	43	80	3	132	64
8	29.03.09	4:55	5:53	01:02 ± 45	08:18 ± 49	15	32	66	−45	15	62
9	26.04.09	2:57	4:02	23:15 ± 23	07:28 ± 30	−5	21	−4	−23	−25	−7
10	12.06.09	1:45 (earliest)	3:51	23:23 ± 40	06:31 ± 36	−18	−20	−9	−147	−73	−42
11	12.07.09	2:28	4:48	01:01 ± 34	08:30 ± 55	4	−29	−46	−180	−138	−167
12	23.08.09	2:47	6:32	01:36 ± 94	08:11 ± 70	−18	−29	−34	−97	−127	−117
13	13.09.09	2:01	4:11	23:48 ± 65	06:15 ± 36	−39	−38	−18	−14	−61	−45
14	27.09.09	4:06	5:00	00:48 ± 45	08:02 ± 101	20	23	28	−64	−19	35
15	11.10.09	2:36	4:13	22:59 ± 26	07:36 ± 69	−16	−34	−12	−75	−42	−42
16	08.11.09	3:44	7:32	00:56 ± 46	09:08 ± 20	−39	−29	−27	−94	−79	−41
17	20.12.09	3:36	4:58	01:30 ± 29	08:19 ± 36	−28	−33	−23	−95	−79	−64
18	17.01.10	3:56	5:04	00:08 ± 68	09:07 ± 74	13	−19	−29	−99	−42	−52
19	21.02.10	3:06	2:35	23:27 ± 35	08:16 ± 66	12	−18	25	−77	−44	−11
20	14.03.10	2:49	4:56	00:28 ± 31	07:33 ± 41	19	2	−16	−68	−72	−18
21	11.04.10	5:08	6:13	02:31 ± 59	08:44 ± 106	17	54	63	−127	−30	111
22	30.05.10	2:26	7:02	23:56 ± 42	07:48 ± 83	−7	−29	28	−33	−86	−51
23	04.07.10	3:49	5:22	00:41 ± 71	09:36 ± 78	−50	−38	−60	−92	−80	−29
24	10.10.10	5:32	7:13	01:50 ± 25	10:13 ± 113	−6	34	32	−50	−29	−80
25	24.10.10	5:13	6:37	02:42 ± 33	08:46 ± 38	−10	−3	−7	−110	−32	−15
Prediction error, mean ± SD (min)			−85 ± 83	192 ± 62	−265 ± 60	−3 ± 23	−3 ± 28	11 ± 42	−71 ± 44	−36 ± 56	−13 ± 65
Prediction error range (min)			407 (−277 to 130)	279 (71 to 350)	277 (−362 to −86)	91(−50 to 41)	92 (−38 to 54)	147 (−60 to 87)	183 (−180 to 3)	270 (−138 to 132)	278 (−167 to 111)
Correlation of predicted phase with CR melatonin phase			.51 $p < .01$	.71 $p < 10^{-4}$	.75 $p < 10^{-4}$	.97 $p < 10^{-14}$	.96 $p < 10^{-13}$	.89 $p < 10^{-8}$	.87 $p < 10^{-8}$	.80 $p < 10^{-5}$	.70 $p < 10^{-3}$

[a]Midsleep on free days corrected for sleep deficit, determined during screening using the MCTQ questionnaire.

TABLE 3. Comparison of variance of prediction errors with one-sided Pitman-Morgan test[a]

Compared prediction models						
Model A	Model B	Number of pairwise comparisons (total 803)	% p < .05	% p < .01	% p < .001	% rejected H_0 corrected for false discovery rate f_F = .05
Linear	CBT	1	0	0	0	0
Linear	Sleep midpoint	1	100	0	0	100
Linear	Motion	1	100	100	100	100
MLP_{115}	Linear	100	42	16	1	36
MLP_{115}	CBT	100	39	6	0	25
MLP_{115}	Sleep midpoint	100	76	62	45	70
MLP_{115}	Motion	100	93	75	58	85
MLP_{164}	Linear	100	35	4	1	17
MLP_{164}	CBT	100	9	0	0	4
MLP_{164}	Sleep midpoint	100	95	81	32	93
MLP_{164}	Motion	100	97	96	85	97

[a] $H_0 : \sigma_A^2 \geq \sigma_B^2$, $H_1 : \sigma_A^2 < \sigma_B^2$. The percentage of significant comparisons in the last column was computed for the total of 803 comparisons using the false discovery rate procedure (Curran-Everett, 2000).

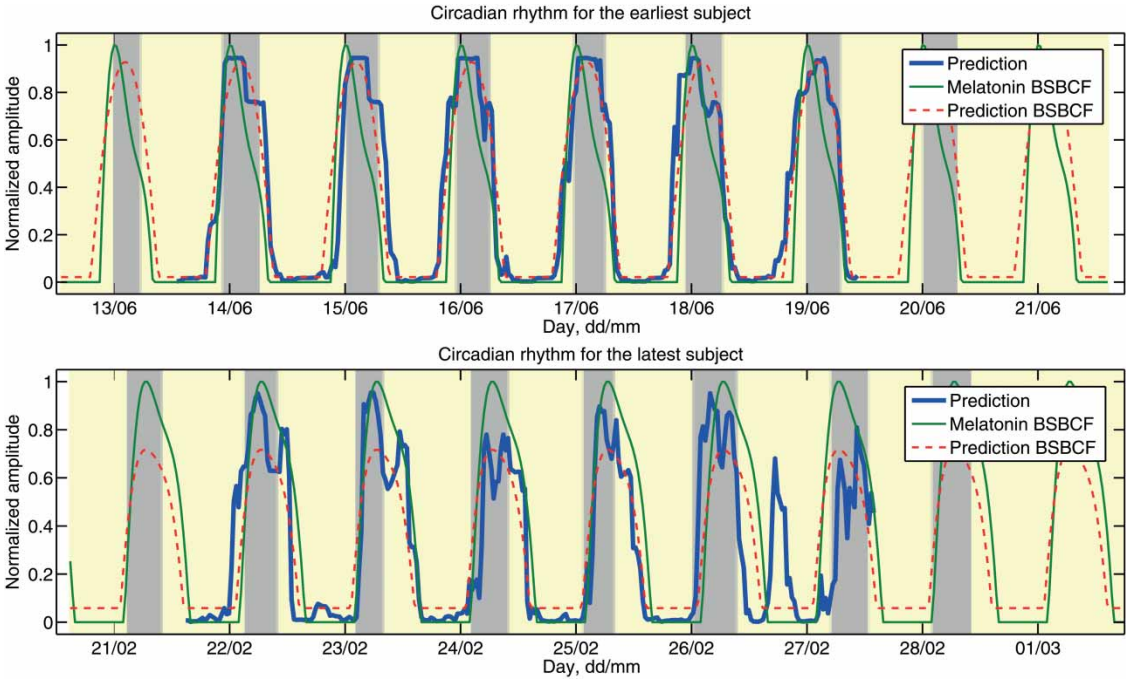

FIGURE 6. Prediction of ambulatory circadian rhythm of melatonin secretion for the subjects with the earliest and the latest phase of entrainment: thick solid line represents the prediction for the ambulatory part of the study, thin solid line represents the extrapolated BSBCF curve of CR melatonin scaled to [0, 1], dashed thin line represents the BSBCF curve extracted from the prediction, dark bars correspond to sleep, and ticks on the x-axis correspond to midnights.

bedtimes or late work hours. However, in the nonlinear neural network model, which can more accurately capture the complex dependencies of circadian rhythms on ambient light exposure and better extract the circadian component from skin temperatures, the motion variable might be redundant. Indeed, removing this variable with its lags of up to 24 h substantially simplified the nonlinear prediction model by reducing the number of adjustable weights by 42% (from 831 to 586), whereas at the same time accuracy was improved. Nevertheless, measuring the motion variable does not impose high subject burden, and it would make sense measuring it in further applied studies for different reasons. It can help estimate (without being a predictor in the model) the discrepancy of the actual rest-activity cycle (e.g., for night-shift workers or persons with

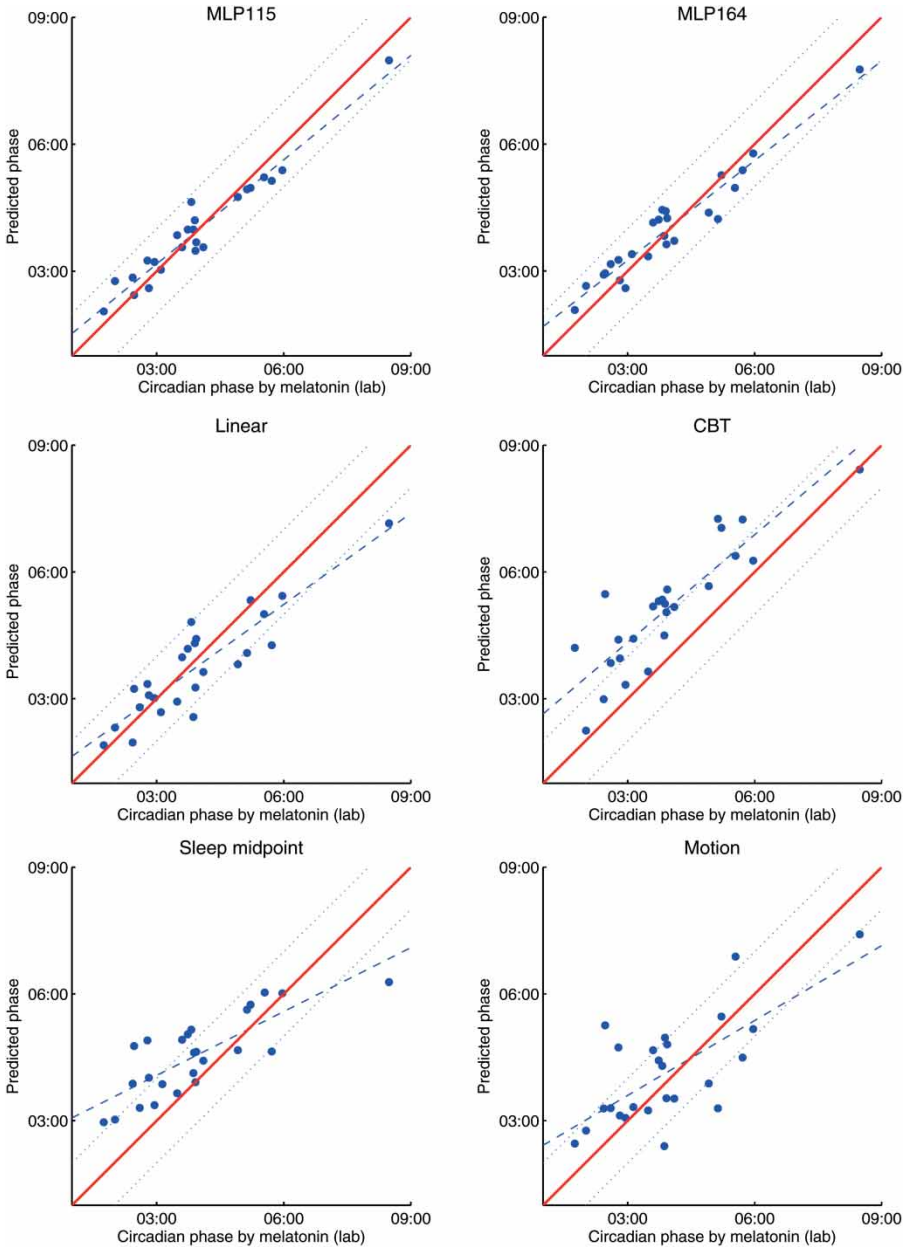

FIGURE 7. Prediction of ambulatory circadian phase for 25 subjects by six different methods versus the circadian phase of melatonin in the laboratory under CR. Thick solid line represents identity line of ideal prediction with zero error, the dashed line is the best fit, and the dotted lines encompass errors ±1 h of the identity line.

"social jet lag") and their endogenous circadian rhythm as estimated by the prediction model.

Our result showing skin temperatures and ambient light are the optimum predictors is also advantageous for the following reasons: some variables, such as heart rate and respiratory rate, are heavily influenced by artifacts that are inevitable in multiple-day recordings because of electrodes coming off (for ECG) or respiratory belts sliding down. In addition, wearing ECG electrodes, respiratory belts, or a leg movement sensor is less comfortable than wearing only small skin temperature sensors and spectacles with a light sensor. Moreover, respiratory belts require calibration, and the processing of respiratory channels as well as ECG requires more sophisticated algorithms with adjustable parameters for correct extraction of parameters, such as heart rate and respiratory rate (Grossman et al., 2010). Extraction of additional variables from ECG (e.g., heart rate variability) imposes yet higher requirements on data quality, which can be hardly feasible for ambulatory measurements over multiple days.

In general, the developed regression model can be considered as a nonlinear multivariate dynamic demasking technique (in contrast to the original linear approach of Kolodyazhniy et al., 2011b), because it is based not just on a single variable such as CBT or DPG taken at a single time point for determining the current value of the underlying circadian rhythm, but rather on a nonlinear

combination of multiple variables on a moving window via neural network regression. The model extracts the circadian variation from multiple skin temperature variables, "filtering out" the masking influences by taking into account an important factor determining entrainment—exposure to blue light. This, combined with the use of CR melatonin as the target variable and the advanced curve-fitting techniques (Van Someren & Nagtegaal, 2007) resulted in improved accuracy of prediction of entrained circadian phase compared to CBT, actimetry, DPG, or sleep logs.

As in Kolodyazhniy et al. (2011b), the approach presented in this paper is based on the assumption that there is no large systematic drift in the timing of sleep and daily activities throughout the ambulatory part, and that prior light history affects melatonin secretion in the CR (Hébert et al., 2002; Smith et al., 2004; Wehr, 1998). Influence of prior light history on CR melatonin was also confirmed by the importance of the blue light variable in our modeling.

If the timing of sleep and activity is drifting (e.g., as in rotating shiftwork), the phase of entrainment may not be determined reliably, irrespective of the method. At the same time, variations in sleep timing throughout ambulatory measurements without a large systematic drift would not have considerable adverse effects on prediction accuracy, if the circadian phase from the predicted melatonin rhythm is determined over multiple days as proposed using the BSBCF and COG approaches. Indeed, in our study, some subjects did have a considerable variation of their sleep timing, but the largest prediction errors were, nevertheless, not associated with the largest standard deviation of bed or wake-up times as can be seen from Table 2. For a better handling of the drift in the timing of sleep and activities, the proposed approach could be modified for weighting of the days over which circadian phase is determined. In this case, the first day would be weighted least and the last day most.

The proposed approach still needs to be validated with female participants and volunteers of different age groups. As we noted previously (Kolodyazhniy et al., 2011b), the best results can be expected for late and extremely late chronotypes, who are subject to "social jet lag" due to the discrepancy between their internal circadian phase and the phase of their rest-activity cycles as dictated by work or study hours, indicating the putative clinical applications in circadian sleep disorders. One further step would be to investigate modifications of the prediction model that would possibly allow reduction in the number of skin temperature sensors to minimize the subject burden even further.

However, the neural network model developed here is sufficient to estimate ambulatory circadian phase (without laboratory measurements such as melatonin) in a new data set, if the same devices for skin temperatures and blue light were used during an ambulatory period of ~1 wk. In contrast, replication of our results, including the construction of a new prediction model, would require data from a representative group of subjects (N = 15–30) with a range of chronotypes, using a similar protocol for ~1 wk of ambulatory measurements (skin temperatures and blue light on both free and workdays) followed by a >24-h constant routine (for a complete circadian profile of melatonin secretion as well as unmasked skin temperature data).

In conclusion, the practical equipment for our method to predict circadian phase would include only tiny skin temperature sensors (iButtons) and a light sensor for the blue spectral band. If required, a 3D accelerometer could be used to record the control variable motion. Although necessary for the validation study, most of the channels in the multichannel devices proved to be redundant. Thus, we attained our aim of finding the minimum number of measures required for an ambulatory device in the real world that, when combined with the prediction model, provides a remarkably accurate estimate of internal clock time. This noninvasive measuring equipment and the proposed prediction method holds promise in sleep medicine, psychiatry, or other clinical domains where knowing the exact endogenous circadian phase is important for accurate treatment timing, e.g., with light or melatonin, and where more invasive measurements and/or laboratory investigation are not possible. The improved nonlinear prediction models with higher accuracy potentially broaden the application area of our approach.

ACKNOWLEDGMENTS

We thank our collaborators, Prof. Domien Beersma from the University of Groningen for his valuable input on the design of the study protocol; Drs. Anand Kumar and Cees Lijzenga from the Personal Health Institute International (Phi-I) who designed the ClockWatcher; Luzian Wolf from Object Tracker who attended to LightWatcher development and improvement; Claudia Renz, Marie-France Dattler, and Giovanni Balestrieri from the Centre for Chronobiology for their help in data acquisition; Marielle Kappeler and all staff who took care of the constant routine in the chronobiology laboratory at the Psychiatric Hospital of the University of Basel, Basel, Switzerland.

Declaration of Interest: This study was supported by the EU FP6 integrated project 018741 EUCLOCK (www.euclock.eu). The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 26:342–392.

Bishop CM. (2006). *Pattern recognition and machine learning*. New York: Springer-Verlag. p. 738.

Cajochen C, Knoblauch V, Kräuchi K, Renz C, Wirz-Justice A. (2001). Dynamics of frontal EEG activity, sleepiness and body temperature under high and low sleep pressure. *Neuroreport* 12:2277–2281.

Cajochen C, Münch M, Kobialka S, Kräuchi K, Steiner R, Oelhafen P, Orgül S, Wirz-Justice A. (2005). High sensitivity of human melatonin, alertness, thermoregulation, and heart rate to short wavelength light. *J. Clin. Endocrinol. Metab.* 90:1311–1316.

Carrier J, Monk TH. (1997). Estimating the endogenous circadian temperature rhythm without keeping people awake. *J. Biol. Rhythms* 12:266–277.

Chen KW, Li G, Sun Y. (1996). A confidence region approach for assessing equivalence in variability of bioavailability. *Biometr. J.* 38:475–487.

Curran-Everett D. (2000). Multiple comparisons: philosophies and illustrations. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 279:R1–R8.

Cybenko G. (1989). Approximations by superpositions of sigmoidal functions. *Math. Control Signals Systems* 2:303–314.

Eastman CI. (1992). High-intensity light for circadian adaptation to a 12-h shift of the sleep schedule. *Am. J. Physiol.* 263:R428–R436.

FP6 project "EUCLOCK." 2006–2011. http://www.euclock.org. Created by David Lenssen, Accessed 9th March 2012, http://www.euclock.eu.

Grossman P, Wilhelm FH, Brutsche M. (2010). Accuracy of ventilatory measurement employing ambulatory inductive plethysmography during tasks of everyday life. *Biol. Psychol.* 84:121–128.

Hastie T, Tibshirani R, Friedman J. (2001). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer-Verlag.

Haykin S. (1998). *Neural networks: a comprehensive foundation*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall. p. 842.

Hébert M, Martin SK, Lee C, Eastman CI. (2002). The effects of prior light history on the suppression of melatonin by light in humans. *J. Pineal Res.* 33:198–203.

Hiddinga AE, Beersma DG, Van den Hoofdakker RH. (1997). Endogenous and exogenous components in the circadian variation of core body temperature in humans. *J. Sleep Res.* 6:156–163.

Hornik K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks* 4:251–257.

Kiymik MK, Akin M, Subasi A. (2004). Automatic recognition of alertness level by using wavelet transform and artificial neural network. *J. Neurosci. Methods* 139:231–240.

Klerman EB, Lee Y, Czeisler CA, Kronauer RE. (1999). Linear demasking techniques are unreliable for estimating the circadian phase of ambulatory temperature data. *J. Biol. Rhythms* 14:260–274.

Klerman EB, Gershengorn HB, Duffy JF, Kronauer RE. (2002). Comparisons of the variability of three markers of the human circadian pacemaker. *J. Biol. Rhythms* 7:181–193.

Kolodyazhniy V, Pfaltz MC, Wilhelm FH. (2007). Modeling of ambulatory heart rate using linear and neural network approaches. In Corrigan MS (ed.). *Pattern recognition in biology*. Hauppauge, NY: Nova Science Publishers, 191–206.

Kolodyazhniy V, Kreibig SD, Roth WT, Gross JJ, Wilhelm FH. (2011a). An affective computing approach to physiological emotion specificity: towards subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology* 48:908–922.

Kolodyazhniy V, Späti J, Frey S, Götz T, Wirz-Justice A, Kräuchi K, Cajochen C, Wilhelm FH. (2011b). Estimation of human circadian phase via a multi-channel ambulatory monitoring system and a multiple regression model. *J. Biol. Rhythms* 26:55–67.

Kräuchi K. (2007). The thermophysiological cascade leading to sleep initiation in relation to phase of entrainment. *Sleep Med. Rev.* 11:439–451.

Kräuchi K, Cajochen C, Werth E, Wirz-Justice A. (1999). Warm feet promote the rapid onset of sleep. *Nature* 401:36–37.

Lockley SW, Brainard GC, Czeisler CA. (2003). High sensitivity of the human circadian melatonin rhythm to resetting by short wavelength light. *J. Clin. Endocrinol. Metab.* 88:4502–4505.

Minors DS, Waterhouse JM. (1989). Masking in humans: the problem and some attempts to solve it. *Chronobiol. Int.* 6:29–53.

Mudholkar GS, Wilding GE, Mietlowski WL. (2003). Robustness properties of the Pitman-Morgan test. *Commun. Stat. Theory. Methods* 32:1801–1816.

Portaluppi F, Smolensky MH, Touitou Y. (2010). Ethics and methods for biological rhythm research on animals and human beings. *Chronobiol. Int.* 27:1911–1929.

Reifman J. (2004). Alternative methods for modeling fatigue and performance. *Aviat. Space Environ. Med.* 75(3, Suppl.): A173–A180.

Roenneberg T, Daan S, Merrow M. (2003). The art of entrainment. *J. Biol. Rhythms* 18:183–194.

Roenneberg T, Kuehnle T, Juda M, Allebrandt K, Gordijn M, Merrow M. (2007). Epidemiology of the human circadian clock. *Sleep Med. Rev.* 11:429–438.

Schmidt C, Collette F, Cajochen C, Peigneux P. (2007). A time to think: circadian rhythms in human cognition. *Cogn. Neuropsychol.* 24:755–789.

Smith AD, Crabtree DR, Bilzon, JL, Walsh NP. (2010). The validity of wireless iButtons and thermistors for human skin temperature measurement. *Physiol. Meas.* 31:95–114.

Smith KA, Schoen MW, Czeisler CA. (2004). Adaptation of human pineal melatonin suppression by recent photic history. *J. Clin. Endocrinol. Metab.* 89:3610–3614.

Smith MR, Revell VL, Eastman CI. (2009). Phase advancing the human circadian clock with blue-enriched polychromatic light. *Sleep Med.* 10:287–294.

Van Someren EJW, Nagtegaal E. (2007). Improving melatonin circadian phase estimates. *Sleep Med.* 8:590–601.

Vuckovic A, Radivojevic V, Chen ACN, Popovic D. (2002). Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. *Med. Eng. Phys.* 24:349–360.

Watanabe L, Kripke DF. (1988). Two months' illumination exposure of a young woman. *Sleep Res.* 17:401.

Wehr TA. (1998). Effect of seasonal changes in daylength on human neuroendocrine function. *Horm. Res.* 49:118–124.

Wetterberg L. (1998). Melatonin in adult depression. In Safii M, Safii SL (Eds.). *Melatonin in psychiatric and neoplastic disorders*. 1st ed. Washington, DC: American Psychiatric Press, 43–80.

Wever RA. (1979). *The circadian system of man: results of experiments under temporal isolation*. New York: Springer-Verlag. p. 276.

Wilhelm FH, Grossman P. (2010). Emotions beyond the laboratory: theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biol. Psychol.* 84:552–569.

Wilhelm FH, Pfaltz MC, Grossman P, Roth WT. (2006). Distinguishing emotional from physical activation in ambulatory psychophysiological monitoring. *Biomed. Sci. Instrum.* 42:458–463.

Wirz-Justice A. (2007). How to measure circadian rhythms in humans. *Medicographia* 29:84–90.

SUPPLEMENTARY MATERIAL

The contents of this supplementary material are based on the previous publication (Kolodyazhniy et al., 2011) with changes concerning the neural network prediction models and modifications of the curve fitting approach. The section "Variable Selection" is excluded, as it is not within the scope of the present publication. In other sections, only the variables relevant in the present analyses are considered. Additionally, the description of the ambulatory monitoring devices from the previous publication was moved to this supplement with minor updates. New parts include detailed descriptions of prediction models.

Sleep Logs

The subjects were instructed to keep their predefined bedtimes that they planned themselves and communicated to the study supervisor several days before the experiment. They were asked to fill out the sleep logs carefully for each day of the ambulatory phase, taking into account both their habitual wake times on weekends and workdays and their concrete plans for the ambulatory week. For the analysis presented in this paper, the times of switching the lights off before sleep and the wake-up times were used. Filling out the sleep log was scheduled 10 min after the planned wake-up time for each day. Subjects were notified to fill out the sleep log by a beep tone from the electronic diary, repeating itself every 5 min until the sleep log was filled out. Compliance in filling out the electronic sleep log was checked by the time stamps of the data entries, as well as by rest/activity and light data. Subjects were informed of these compliance checks and did not significantly deviate from their self-planned bedtimes. The sleep diary was checked before the constant routine (CR) study in the laboratory and used for scheduling the CR according to the average sleep midpoint during the ambulatory week.

Ambulatory Monitoring Devices

A prototype of the ambulatory circadian monitoring device "ClockWatcher" for studies within the EUCLOCK Project was developed by Personal Health Institute International (Phi-I) in Amsterdam (see Figure 2 in Kolodyazhniy et al., 2011). The ClockWatcher is designed to record the following variables over multiple days:

- core body temperature (CBT; measured with a disposable rectal probe);
- electrocardiogram (ECG; recorded with solid gel electrodes; each subject was given additional electrodes and instructed to replace them daily);
- respiration (recorded with two belts for thorax and abdomen);
- body movement and posture (measured with a three-dimensional [3D] accelerometer);
- leg movement (measured with a 1D accelerometer); and
- event markers (going to sleep/taking shower).

Additionally, an "off-the-shelf" monitoring device (Varioport, Becker Meditec, Karlsruhe, Germany) was acquired in order to expedite the development of the ambulatory circadian models (see Figure 2 in Kolodyazhniy et al., 2011). The VarioPort device had custom-made signal preamplifiers that allowed them to be used for recording the same set of variables as with the ClockWatcher. Out of the 30 participants, 8 wore the ClockWatcher (5 complete data sets) and 22 the VarioPort (20 complete sets) Based on feedback from this validation study, a new miniaturized version of ClockWatcher with improved reliability was developed by Phi-I (www.phi-i.com).

Using calibration and our MATLAB software (The MathWorks, Natick, MA, USA), all data from both devices were brought to the same ranges to facilitate further model development irrespective of the device type. For details of our circadian software and data organization, see the section "Circadian Software Toolbox and Database" below.

For recording ambient light, a specialized miniature ambulatory monitoring device "LightWatcher" was also developed within the EUCLOCK Project by Sowoon (Lausanne, Switzerland) (Figure 2). It has the following measurement channels (Kolodyazhniy et al., 2011):

- light in five spectral bands (infrared, red, green, blue, ultraviolet);
- 3D acceleration (motion along axes *X*, *Y*, and *Z*, used for compliance check); and
- ambient temperature.

Eleven miniature wireless temperature sensors (DS 1922L Thermochron iButtons; diameter × height: 17 × 6 mm, accuracy .0625°C; Maxim, San Jose, CA, USA) were used to record skin temperatures continuously in 2.5-min intervals throughout the protocol (Smith et al., 2010). The iButtons were fixed to the skin with thin, air-permeable adhesive surgical tape (Fixomull® Beiersdorf, Hamburg, Germany) on the left and right side of the body (hands, feet, except for thorax) (see Figure 2 in Kolodyazhniy et al., 2011). The temperature sensors can be worn under normal life conditions, including taking a shower and doing sports. The iButtons were applied by the experimenters at the beginning of the protocol and were worn throughout. Each subject was given ca. 50 pieces of adhesive tape 50 × 50 mm in size and instructed to replace them when necessary, so the temperature sensors stayed in good contact with skin and at the same locations. All sensors were numbered and the subjects received a diagram showing the correct locations of the sensors on skin according to the numbering, which were checked again in the laboratory before the CR.

Handheld computers of type Palm Tungsten E (Palm Inc., Sunnyvale, CA, USA) were used as an electronic diary. Questionnaires for sleep logs and a number of other scales and fill-in forms were programmed with Pendragon Forms software v. 4.0 (Pendragon Software, Buffalo Grove, IL, USA).

Checking and Transferring Data

The clocks of all devices were synchronized at the beginning of the measurements with the master personal computer (PC) that was used for collecting and processing data. Deviation of the clocks from the master clock after completing the entire protocol for each subject did not exceed ±3 min.

For the ClockWatcher and VarioPort devices, batteries needed to be replaced in the middle of the ambulatory part (3 or 4 d after the start) and at its end. Replacement of batteries was done by the study supervisor. At the same time, data were downloaded for inspection and backup, although the memory cards had enough capacity to store data for the entire protocol (about 1 GB). At the same time, data from the light sensor were checked. The entire procedure including battery replacement took ~20 min. Data were transferred and inspected using a portable computer in the presence of the test subject. Import of data from all devices, including the skin temperature sensors, into the database for analysis was done after the end of the entire protocol for each subject.

Circadian Software Toolbox and Database

For multiple regression modeling based on data from multiple multichannel ambulatory monitoring devices, it is essential that the data from heterogeneous sources are synchronized and properly preprocessed. For storing the preprocessed data, a database running on MySQL Server v.5.5.9 (Oracle Corp, Redwood City, CA, USA) was created. Import of study data into the database, preparation of the data for analysis, and analysis, including the linear and neural network–based regression modeling approaches, were done using a specially developed circadian software toolbox running under the general-purpose numerical computing environment MATLAB R2011b (The MathWorks). Fitting of bimodal skewed baseline cosine function (BSBCF) curves for melatonin and predictions of the circadian rhythm required the use of the MATLAB Optimization Toolbox (see also the section "Waveform Analysis"). For the new analyses with neural network models, functions of the MATLAB Neural Network Toolbox were used.

Recordings from "slow" data sources with sampling periods of ≥1 min were imported into the database without any preprocessing (iButtons, LightWatcher, E-diary, melatonin levels). Data from the ClockWatcher and VarioPort devices with sampling rates up to 512 Hz were preprocessed prior to importing them into the database. In the preprocessing step, sampling periods for all channels were reduced to 30 s after some transformation of the recorded signals, primarily the following: heart rate and respiratory rate were derived from the respective raw signals (VarioPort recordings already contained a heart rate channel) and accelerations along axes *X*, *Y*, and *Z* were divided into the respective motion and posture components, and leg movement was computed from the respective 1D accelerometer. In the present publication, only melatonin, CBT, motion, skin variables, blue light, and sleep log data are analyzed. Therefore, only these variables are described in this and the following section.

A direct double-antibody radioimmunoassay (RIA) was used for the melatonin assay, validated by gas chromatography–mass spectroscopy (Bühlmann Laboratories, Schönenbuch, Switzerland). The minimum detectable dose of melatonin (analytical sensitivity) was determined to be .2 pg/mL. The functional least-detectable dose using the less than 20% coefficient of interassay variation criterion was <.65 pg/mL (Weber et al., 1997).

For obtaining the motion variable, the low-frequency components X_{post}, Y_{post}, and Z_{post} corresponding to posture were computed from the respective acceleration channels X_{acc}, Y_{acc}, and Z_{acc} using a second-order IIR filter with a 3 dB cutoff at .15 Hz. The high frequency components corresponding to motion were then extracted via rectifying the difference of the acceleration and posture:

$$X_{mov} = abs(X_{acc} - X_{post})/2,\ Y_{mov} = abs(Y_{acc} - Y_{post})/2,\ Z_{mov} = abs(Z_{acc} - Z_{post})/2.$$

An integrated variable for motion was also computed as square root of sum of squared values of motion along the respective coordinates: $M = \sqrt{X_{mov}^2 + Y_{mov}^2 + Z_{mov}^2}$.

Via downsampling of the original high sampling rate to the sampling period of 30 s, the overall amount of data was dramatically reduced from more than 1 GB per subject to less than 10 MB. At the same time, the sampling period of 30 s allowed us to visually detect CBT probe slips as a fast decline in temperature and edit them using the graphical user interface of our software. The software uses real time and date for all data records in the database, such that all time stamps from multiple heterogeneous data sources (including multichannel physiological recordings, electronic diaries, and melatonin assays) are synchronized. Data imported into the database are retrieved by the software and resampled "on the fly" to an arbitrarily common sampling period that can be chosen from 30 s to 6 h, independent of the original sampling rate of the raw data or the preprocessed data in the database. This provides a convenient means for multivariate modeling and visualization.

Data Preparation for Regression Modeling

The following steps of data preparation were performed with the data previously imported into the database to form the data set for multiple regression modeling with the linear and neural network-based approaches:

1. *Skin temperatures.* For skin temperatures measured with iButtons, six variables were produced from the 11 sensors: averaging of the left- and right-hand side sensors (except for thorax) resulted in skin temperature variables for *shoulders*, *hands*, *feet*, *thorax*,

upper legs, and *lower legs*. Temperature data were detrended by subtracting from each temperature channel its moving average with a window of 24 h. The moving average was computed by averaging the data in the interval of ±12 h from the current point (i.e., from the center of the moving window), so that no phase distortion was introduced. Detrended data were *z*-transformed for each participant. In the *z*-transformed data, all values outside the range of [–2, 2], i.e., ±2 standard deviations, were discarded as outliers, and the gaps were interpolated using cubic polynomials. For details on interpolation, also for other measured variables mentioned in items 3 and 4 below, see the section "Interpolation of Missing Data."

2. *Ambient light.* Irradiance values <.01 were replaced with .01, and the variable was $\log_{10}$-transformed to accommodate the very broad range of irradiance between darkness and bright sunlight.
3. *Motion.* Only the integrated variable for motion along axes *X*, *Y*, and *Z* was used. Missing values were interpolated using cubic polynomials.
4. *CBT.* This variable was edited for probe slips. Missing values were interpolated using cubic polynomials. Probe slips were determined as a decline in the temperature within several minutes below the CBT minimum of the entire recording.
5. *Melatonin.* A periodic BSBCF waveform was extracted from the CR-derived melatonin rhythm for each subject and extrapolated backwards onto the 7 d of the ambulatory part preceding the in-laboratory experiment. Then, after scaling to the range of [0, 1] and adding normally distributed random noise with a standard deviation of .01, this variable was used as the target for fitting a multiple regression model (linear or neural network based), with input variables from the multichannel recordings of both the ambulatory and laboratory parts of the experiment that were 9 d long for each subject. The scaling of the BSBCF waveform was performed in order to eliminate individual differences in levels of melatonin secretion that complicate the development of a subject-independent circadian rhythm model, and the random noise improved model identification.

To model a realistic situation of real-world application, no recordings that were used in the analysis were manually edited, i.e., data that were included in the analysis were automatically processed raw data. The only editing was of the CBT recordings, which were used for comparison of the prediction of circadian phase with the linear and nonlinear regression models and were used neither as a predictor nor as the dependent variable in the models themselves.

All channels for each subject were aligned based on the respective time stamps, and the data were further resampled to 30-min bins. To avoid phase distortion in resampling, the time stamp of all bins on the time axis corresponded to the center of the bin. From each subject, data from both the ambulatory and laboratory parts were used for fitting a prediction model using the least squares method for the linear models or the resilient propagation training algorithm ("trainrp") of the MATLAB Neural Network Toolbox for the neural network models.

Interpolation of Missing Data

For interpolation of missing points in data from the monitoring devices, piecewise cubic Hermite interpolating polynomials were used (MATLAB function "interp1" with method "cubic"). The polynomials use three data points on each side of the gap to fit the interpolating curve. In our software, we set the threshold for the maximum length of interpolated gaps to 12 h.

In the entire data set there were three large gaps in ambulatory data caused by technical problems in the multichannel measurement devices: ca. 48 h for subject 1 (CBT, motion), ca. 30 h for subject 8 (ambient light), and ca. 80 h for subject 17 (ambient light). We did not exclude these participants, because our threshold for the usability of collected data was the presence of at least 50% of ambulatory recordings in all channels. For the other four subjects who were excluded at the initial stage of our study due to technical problems as described in (Kolodyazhniy et al., 2011), this criterion did not hold. The technical problems were subsequently solved by the respective manufacturers in newer versions of the devices. Most of the other gaps were only several minutes long and occurred when the subjects took shower or during replacement of the batteries and memory cards in the ClockWatcher and VarioPort devices.

Waveform Analysis

The bimodal skewed baseline cosine function (BSBCF; Van Someren & Nagtegaal, 2007) is defined as follows:

$$f(t) = b + \frac{H}{2(1-c)}(\cos(t-\varphi+\nu\cos(t-\varphi)) + m\cos(2t-2\varphi-\pi)-c+|\cos(t-\varphi + \nu\cos(t-\varphi)) + m\cos(2t-2\varphi-\pi)-c|),$$

where t is time (in radians, $2\pi = 24$ h), b is the baseline, H is the peak level (>0), c is the peak width (≥–1 and <1), φ is phase (in radians, 0 to 2π), ν is skewness (between ≥–.5 and ≤.5), and m is bimodality (≥0 and <1). Fitting of the BSBCF curve was done using the MATLAB optimization toolbox using data with a sampling period of 30 min. The fitted data were either melatonin data or the predicted circadian rhythm from the regression models. Hourly melatonin data were resampled to match the timestamps of other data channels first by upsampling to 1-min sampling rate using linear interpolation and then by downsampling to 30-min bins. An example of the curve fitting is shown in Figure S1. The minimized

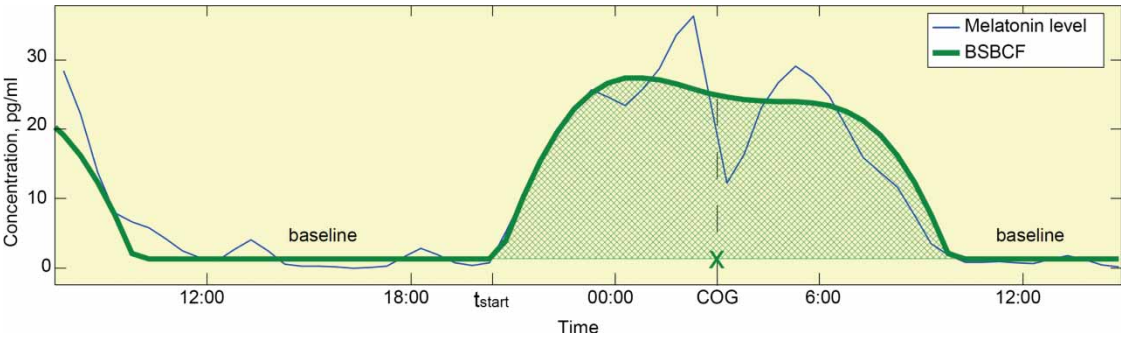

FIGURE S1. Determining circadian phase from CR melatonin using bimodal skewed baseline cosine function (BSBCF) and center of gravity (COG) of area under BSBCF curve for one period (shaded area in the plot): t_{start} is the first time point where the BSBCF curve differs from the baseline, and marks the beginning of the interval for computing the COG; the baseline is found from fitting the BSBCF curve.

objective function in the optimization procedure in MATLAB was

$$J(Y, \hat{Y}) = 1 - R^2(Y, \hat{Y}),$$

where Y is the data to be fitted (melatonin data or the predicted melatonin rhythm), $\hat{Y}$ is a vector of values of the fitted BSBCF function $f(t)$ for time t from the start to the end of the fitted data with a step of 1 min, and R^2 is the coefficient of determination. For more stable results, we slightly modified our original approach (Kolodyazhniy et al., 2011) by "folding" the data to be fitted via computing 1-min averages for time from 00:00 to 23:59 h for the entire length of the data and fitting the BSBCF function to the folded data.

Calculation of Circadian Phase

Circadian phase was determined as the time of day corresponding to the center of gravity (COG) of the area (Wetterberg, 1998) under the periodic BSBCF curve for one period of 24 h (Kolodyazhniy et al., 2011):

$$\mathrm{COG} = \frac{\sum_{t=t_{start}}^{t_{start}+24} t \cdot (f(t) - b)}{\sum_{t=t_{start}}^{t_{start}+24} (f(t) - b)},$$

where t is time in hours, b is the baseline found from fitting the BSBCF function, and t_{start} is the time point when the 24-h period starts (Figure S1).The start of the 24-h period (t_{start}) was found as the first time point where the BSBCF curve differed from the baseline. For precise computing of circadian phase, the fitted BSBCF curve was resampled with a sampling period of 1 min. The combination of the BSBCF and COG approaches can be interpreted as average circadian phase for each subject, i.e., for 32 h of the reference melatonin data (1.33 d with two offsets and one onset) or for the 6-d-long prediction of the circadian rhythm from the multiple regression model (linear or neural network). With the COG method, there is no explicit threshold parameter for determining circadian phase from the BSBCF curve, because the "threshold" is always equal to the baseline b. Van Someren and Nagtegaal (2007) defined the phase of the BSBCF curve as midpoint at threshold level of 25%. This threshold may need to be adjusted if there is a narrow peak or two peaks very different in amplitude, with one of them below 25%.

Linear Prediction Models

Linear multiple regression models for estimation of circadian rhythm of CR melatonin from ambulatory measurements were described by the following equation (Kolodyazhniy et al., 2011):

$$\begin{aligned}\hat{y}(t) &= \alpha_{\mathrm{hands},0} T_{\mathrm{hands}}(t) + \alpha_{\mathrm{hands},1} T_{\mathrm{hands}}(t-0.5) \\ &+ \ldots + \alpha_{\mathrm{hands},10} T_{\mathrm{hands}}(t-5) + \alpha_{\mathrm{feet},0} T_{\mathrm{feet}}(t) \\ &+ \alpha_{\mathrm{feet},1} T_{\mathrm{feet}}(t-0.5) + \ldots + \alpha_{\mathrm{feet},10} T_{\mathrm{feet}}(t-5) \\ &+ \alpha_{\mathrm{thorax},0} T_{\mathrm{thorax}}(t) + \alpha_{\mathrm{thorax},1} T_{\mathrm{thorax}}(t-0.5) + \ldots \\ &+ \alpha_{\mathrm{thorax},10} T_{\mathrm{thorax}}(t-5) + \alpha_{\mathrm{shoulders},0} T_{\mathrm{shoulders}}(t) \\ &+ \alpha_{\mathrm{shoulders},1} T_{\mathrm{shoulders}}(t-0.5) + \ldots \\ &+ \alpha_{\mathrm{shoulders},10} T_{\mathrm{shoulders}}(t-5) + \alpha_{\mathrm{upper\,legs},0} T_{\mathrm{upper\,legs}}(t) \\ &+ \alpha_{\mathrm{upper\,legs},1} T_{\mathrm{upper\,legs}}(t-0.5) + \ldots \\ &+ \alpha_{\mathrm{upper\,legs},10} T_{\mathrm{upper\,legs}}(t-5) + \alpha_{\mathrm{lower\,legs},0} T_{\mathrm{lower\,legs}}(t) \\ &+ \alpha_{\mathrm{lower\,legs},1} T_{\mathrm{lower\,legs}}(t-0.5) + \ldots \\ &+ \alpha_{\mathrm{lower\,legs},10} T_{\mathrm{lower\,legs}}(t-5) + \alpha_{\mathrm{motion},0} M(t) \\ &+ \alpha_{\mathrm{motion},1} M(t-0.5) + \ldots + \alpha_{\mathrm{motion},48} M(t-24) \\ &+ \alpha_{\mathrm{light},0} L(t) + \alpha_{\mathrm{light},1} L(t-0.5) + \ldots + \alpha_{\mathrm{light},48} L(t-24) \\ &+ \beta,\end{aligned}$$

where $\hat{y}(t)$ is the predicted value of normalized melatonin level at time t, T_{hands}, T_{feet}, T_{thorax}, $T_{\mathrm{shoulders}}$, $T_{\mathrm{upper\,legs}}$, and $T_{\mathrm{lower\,legs}}$ were the respective skin temperatures, M was the integrated variable for motion, L was blue light. All data were downsampled to 30-min bins. The models contained 165 parameters each: 164 coefficients $\alpha_{\mathrm{hands},0}, \ldots, \alpha_{\mathrm{light},48}$ for the predictor variables with their respective lags plus the bias term β. The first index of the coefficients alpha means sensor type or location, whereas the second one designates the lag of the respective input variable in 30-min steps from 0 (no lag) through 48 (a lag of 24 h). The parameters $\alpha_{\mathrm{hands},0}, \ldots, \alpha_{\mathrm{light},48}$ and β were determined using linear least squares regression.

Neural Network–Based Prediction Models

The artificial neural network (ANN) model with the same predictor variables as the linear model above (MLP164) is described as

$$\hat{y}(t) = f_{MLP164}\{T_{\text{hands}}(t), T_{\text{hands}}(t-0.5), \ldots, T_{\text{hands}}(t-5),$$
$$T_{\text{feet}}(t), T_{\text{feet}}(t-0.5), \ldots, T_{\text{feet}}(t-5),$$
$$T_{\text{thorax}}(t), T_{\text{thorax}}(t-0.5), \ldots, T_{\text{thorax}}(t-5),$$
$$T_{\text{shoulders}}(t), T_{\text{shoulders}}(t-0.5), \ldots, T_{\text{shoulders}}(t-5),$$
$$T_{\text{upper legs}}(t), T_{\text{upper legs}}(t-0.5), \ldots, T_{\text{upper legs}}(t-5),$$
$$T_{\text{lower legs}}(t), T_{\text{lower legs}}(t-0.5), \ldots, T_{\text{lower legs}}(t-5),$$
$$M(t), M(t-0.5), \ldots, M(t-24),$$
$$L(t), L(t-0.5), \ldots, L(t-24)\}.$$

Here, f_{MLP164} is the nonlinear function "learned" by the MLP neural network with 164 inputs based on the same training data as the linear model in (Kolodyazhniy et al., 2011). The meaning of the variables is the same as for the linear model described above.

The ANN model without the motion variable (MLP115) is represented by the following equation:

$$\hat{y}(t) = f_{MLP115}\{T_{\text{hands}}(t), T_{\text{hands}}(t-0.5), \ldots, T_{\text{hands}}(t-5),$$
$$T_{\text{feet}}(t), T_{\text{feet}}(t-0.5), \ldots, T_{\text{feet}}(t-5),$$
$$T_{\text{thorax}}(t), T_{\text{thorax}}(t-0.5), \ldots, T_{\text{thorax}}(t-5),$$
$$T_{\text{shoulders}}(t), T_{\text{shoulders}}(t-0.5), \ldots, T_{\text{shoulders}}(t-5),$$
$$T_{\text{upper legs}}(t), T_{\text{upper legs}}(t-0.5), \ldots, T_{\text{upper legs}}(t-5),$$
$$T_{\text{lower legs}}(t), T_{\text{lower legs}}(t-0.5), \ldots, T_{\text{lower legs}}(t-5),$$
$$L(t), L(t-0.5), \ldots, L(t-24)\}.$$

Here, f_{MLP115} is the nonlinear function of the MLP network with 115 inputs.

The weights of the hidden layer represent the influence of each variable (e.g., T_{hands}) for each specific lag on each neuron of the hidden layer. In turn, the weights of the output layer represent the influence of the hidden layer neurons on the output of the ANN, i.e., on the resulting prediction. Both layers are fully connected, i.e., all inputs are connected via weights to all neurons of the hidden layer, and all outputs of the hidden layer are connected to the output layer neuron. The number of weights in the hidden layer equals the number of inputs plus one for bias weights times the number of hidden layer neurons (five for both MLP164 and MLP115), and in the output layer the number of hidden neurons (one for both MLP164 and MLP115) plus one. This gives in total $(164+1)\times 5+(5+1)=831$ weights for MLP164 and $(115+1)\times 5+(5+1)=586$ weights for MLP115. The weights were trained with the resilient propagation training algorithm (function "trainrp") for 100 epochs.

Cross-validation

Prediction of circadian phase of melatonin secretion with multiple regression models (linear and neural networks) was based on the subject-independent cross-validation approach, where data from 24 subjects out of 25 were used for identifying a model, and data from another subject was used for validation. In the first iteration, subjects N2–N25 were used for model identification, and subject N1 for validation; in the second iteration subjects N1, N3–N25 for model identification, and subject N2 for validation, and so on (see Figure S2). Thus, in each of the iterations, one model was identified using data from all but one subject and validated with the data from that subject "unseen" during model identification, i.e., *there were 25 fits with 24 subjects each validated against an unknown subject*. The approach to subject-independent cross-validation was similar to that widely accepted in the literature (e.g., Ho et al., 2009; Howard et al., 2009; Zhao & Lu, 2005).

For model identification, data with both ambulatory and laboratory parts was used, whereas validation was done only with ambulatory data of the respective subject in order to provide a realistic estimate of circadian phase in real-life conditions. Special care was taken to ensure no data from the CR and the adaptation night were included into the data set for validation, also taking into account the detrending method for skin temperatures described above. Thus, the predictions of the ambulatory circadian phase started 24 h after the protocol start due to the maximum lag of 24 h in the prediction model, and ended 12 h before the adaptation night in the laboratory due to the detrending method for skin temperatures.

In each iteration of cross-validation, data were restandardized to simulate a real situation where unknown observations with unknown means and standard deviations arrive and are used for prediction of circadian phase based on statistical characteristics of the known data that were used for fitting the regression model. Data standardization in each of cross-validation iterations was done as follows: first, means and standard deviations were computed for the respective variables in the entire data set used for fitting the model comprising data from 24 subjects. Then, these means and

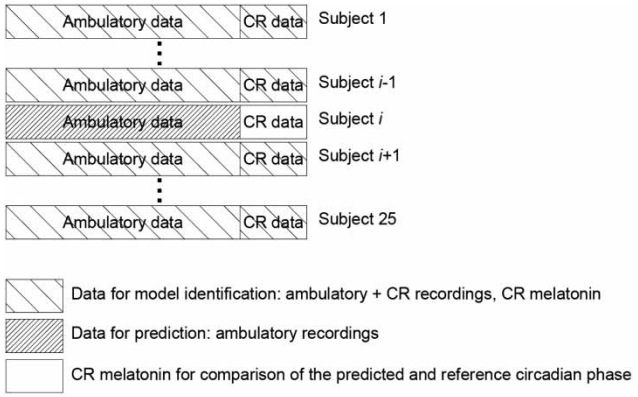

FIGURE S2. Data split for 25-fold cross-validation with 25 subjects: for subject number i ($i = 1, \ldots, 25$), complete ambulatory and CR data from the other 24 subjects are used for identification of the ith prediction model, and the ith subject's own ambulatory data and CR melatonin are used for validation of the ith model.

standard deviations were used to perform *z*-transformation of both the data used for fitting the model (24 subjects) and of the validation data from another subject. For artificial neural networks, the 25-fold cross-validation procedure, including training of a neural network, was repeated 100 times, each time using a different initialization of the random number generator.

References

Ho TC, Brown S, Serences JT. (2009). Domain general mechanisms of perceptual decision making in human cortex. *J. Neurosci.* 29:8675–8687.

Howard J, Plailly J, Grueschow M, Haynes JD, Gottfried JA. (2009). Odor quality coding and categorization in human posterior piriform cortex. *Nat. Neurosci.* 12:932–938.

Kolodyazhniy V, Späti J, Frey S, Götz T, Wirz-Justice A, Kräuchi K, Cajochen C, Wilhelm FH. (2011). Estimation of human circadian phase via a multi-channel ambulatory monitoring system and a multiple regression model. *J. Biol. Rhythms* 26:55–67.

Van Someren EJW, Nagtegaal E. (2007). Improving melatonin circadian phase estimates. *Sleep Med.* 8:590–601.

Weber JM, Schwander JC, Unger I, Meier D. (1997). A direct ultrasensitive RIA for the determination of melatonin in human saliva: comparison with serum levels. *J. Sleep Res.* 26:757.

Wetterberg L. (1998). Melatonin in adult depression. In Safii M, Safii SL (eds.). *Melatonin in psychiatric and neoplastic disorders*. 1st ed. Washington, DC: American Psychiatric Press, pp. 43–80.

Zhao Q, Lu H. (2005). GA-driven LDA in KPCA space for facial expression recognition. In Wang L, Chen K, Ong YS (eds.). *Lecture notes in computer science*, Vol. 3611. Berlin: Springer-Verlag, pp. 28–36.