

Pattern Recognition of Sleep in Rodents Using Piezoelectric Signals Generated by Gross Body Movements

Aaron E. Flores, Judith E. Flores, Hrishikesh Deshpande, Jorge A. Picazo, Xinmin (Simon) Xie, Paul Franken, H. Craig Heller, Dennis A. Grahn, and Bruce F. O'Hara*

Abstract—Current research on sleep using experimental animals is limited by the expense and time-consuming nature of traditional EEG/EMG recordings. We present here an alternative, noninvasive approach utilizing piezoelectric films configured as highly sensitive motion detectors. These film strips attached to the floor of the rodent cage produce an electrical output in direct proportion to the distortion of the material. During sleep, movement associated with breathing is the predominant gross body movement and, thus, output from the piezoelectric transducer provided an accurate respiratory trace during sleep. During wake, respiratory movements are masked by other motor activities. An automatic pattern recognition system was developed to identify periods of sleep and wake using the piezoelectric generated signal. Due to the complex and highly variable waveforms that result from subtle postural adjustments in the animals, traditional signal analysis techniques were not sufficient for accurate classification of sleep versus wake. Therefore, a novel pattern recognition algorithm was developed that successfully distinguished sleep from wake in approximately 95% of all epochs. This algorithm may have general utility for a variety of signals in biomedical and engineering applications. This automated system for monitoring sleep is noninvasive, inexpensive, and may be useful for large-scale sleep studies including genetic approaches towards understanding

sleep and sleep disorders, and the rapid screening of the efficacy of sleep or wake promoting drugs.

Index Terms—Activity, automated, behavior, classification, instrumentation, mice, noninvasive, respiration, wake.

I. INTRODUCTION

WE spend approximately one third of our lives asleep, yet the basic functions of sleep are unknown. Further, sleep disturbances afflict hundreds of millions of individuals throughout the world [1]. A serious limitation to progress in understanding fundamental questions of sleep and sleep disorders is technological. Current sleep analysis techniques use electroencephalograms (EEG) and electromyograms (EMG) as input. EEG/EMG recording in rodents (the dominant experimental organisms) requires extensive surgery, recovery, and the attachment of cables to the subjects during the experiments. We developed an alternative technique for monitoring sleep in rodents that is noninvasive, inexpensive, and could be used for large-scale sleep studies. Large-scale studies in rodents can take advantage of the huge breakthroughs in mouse and human genetics to find genes that influence sleep and sleep disorders and that may in turn shed light on the basic functions of sleep, or provide new drug targets for better sleep and wake-promoting drugs.

Our system utilizes piezoelectric materials configured as highly-sensitive motion detectors incorporated into the bottom of an animal cage. Gross body movements change in a predictable manner with changes in arousal state. Wakefulness is characterized by high frequency and erratic movements associated with volitional movement. This activity pattern produces a chaotic output from the piezoelectric transducer that we designate here as “irregular.” During sleep, the primary gross body movements are related to respiration and are rhythmic in nature, producing a quasi-periodic wave pattern from the piezoelectric transducer that we refer to as “regular.”

Previous work by others has demonstrated the use of movement and respiratory patterns to discriminate sleep versus wake in humans and experimental animals using static charge sensitive beds or electronic activity monitors [2], [3]. Computer scoring of these signals was also accomplished [4]. However, none of these methods was developed for high-throughput sleep scoring, especially in mice, where many large-scale genetic screening programs are now underway [5]. Piezoelectric materials are more sensitive and more conducive to these attempts. Other behavioral monitoring systems that determine gross

Manuscript received January 21, 2006; revised July 1, 2006. This work was supported in part by the National Institutes of Health (NIH) under Grant DA13349, Grant RR017182, and Grant MH067752, in part by a grant from Deltagen, Inc., in part by the Department of Defense under Grant from DoD FA9550-05-1-0464, and in part by the National Science Foundation (NSF) under Functional Genomics Infrastructure Award EPS-0132295. Asterisk indicates corresponding author.

A. E. Flores was with the Department of Electrical Engineering of Stanford University, Stanford, CA 94305 USA. He is now with the R&D Department, Advertising.com, Mountain View, CA 94041 USA.

J. E. Flores was with the Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020 USA. She is now with Genentech, Inc., South San Francisco, CA 94080 USA.

H. Deshpande was with the Department of Computer Sciences, Stanford University, Stanford, CA 94305 USA. He is now with Sony Computer Entertainment, Foster City, CA 94404 USA.

J. Picazo was with the Department of Statistics of Stanford University, Stanford, CA 94305 USA. He is now with the Quantitative Research Group, Credit Suisse, New York, NY 10010 USA.

X. Xie was with the Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020 USA. He is now with AfaSci, Inc., Burlingame, CA 94010 USA.

P. Franken was with the Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020 USA. He is now with the Center for Integrative Genomics, the University of Lausanne, Lausanne 1015, Switzerland.

H. C. Heller and D. A. Grahn are with the Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020 USA.

*B. F. O'Hara is with the Department of Biology, University of Kentucky, Lexington, KY 40506-0225 USA (e-mail: bohara@uky.edu).

Color versions of Figs. 3, 4, 6, and 8 are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBME.2006.886938

Piezoelectric Output (top) vs. Impedance Pneumograph (bottom)

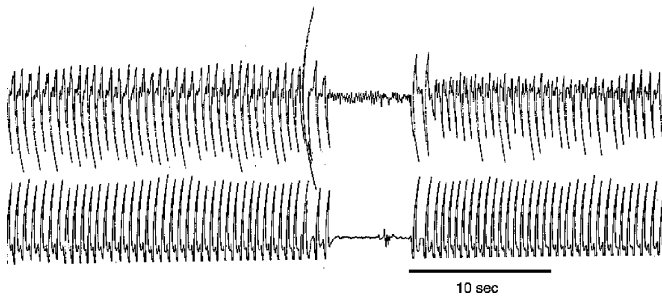


Fig. 1. Simultaneous output signals of the piezoelectric transducer (top) and impedance pneumograph (bottom) demonstrated that the piezoelectric signal provided an accurate respiratory trace during sleep in rodents. Note the 7-s period of apnea in the middle of this period that is accurately reflected in both recordings.

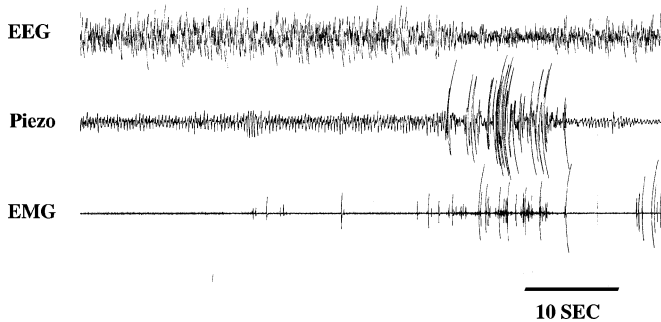


Fig. 2. Simultaneous EEG, EMG and piezoelectric recordings from an adult mouse. This segment is composed primarily of SWS interrupted by a 15-s wake bout about two thirds into the recording period. Note the corresponding changes in the EEG, EMG and piezo recordings. The EEG shows the typical shift from higher amplitude, lower frequency SWS to wake during the 15-s wake bout, while the piezo signal changes from a more regular pattern to erratic. In fact, the piezo signal is the easiest of the three to score for sleep and wake.

motor activity by infra-red methods, photo-beam breaking, or wheel running, are clearly inadequate for sleep scoring since a considerable amount of waking behavior (such as grooming) is not detected as movement or activity.

Sleep can be further classified as slow wave sleep (SWS) or rapid-eye-movement-sleep (REMS) by standard EEG/EMG criteria [1]. Our system does not differentiate between SWS and REMS, as both produce regular patterns in the piezoelectric signal, but it does distinguish Wake from Sleep, which could be either SWS or REMS.

To confirm that our piezoelectric system provides an accurate respiratory trace during sleep, we compared the piezo signal to an impedance pneumograph output (Fig. 1). Validation of the piezoelectric signal as a sleep scoring marker was achieved by comparing the visual scoring of the piezoelectric signal with simultaneously recorded EEG/EMG traces as shown in Fig. 2. Although we had initial concern that quiet wake might produce respiratory traces similar to those during sleep, it turns out that mice spend extremely little time in such a quiet state. A complete lack of nonrespiratory movements is almost always indicative of sleep, since during even quiet wake the mouse is making postural adjustments, head movements, or grooming that are picked up by the very sensitive piezoelectric film. This was clear from visual observations of mice, but is also supported by the results we present here.

Visual scoring of the piezoelectrical signal is tedious and time consuming; hence, our system automates the scoring using pattern recognition and statistical learning techniques. Previous work has been done for the extraction of periodic components from signals using Fourier and wavelet analysis [6], periodicity transforms [7], and singular value decomposition [8], [9]; however, none of that work was concerned with the classification of the signals as “regular” or “irregular” in the sense we defined before. In [10], they developed an optimal Bayes detector for certain periodic waveforms that requires *a priori* knowledge of the periodic waveform, which we do not have in our application. In [11] and [12], they used measures of similarity of waveforms in a neighborhood of the signal to identify the period and amplitude of periodic components; our techniques described in Section III-B are in the same spirit of that work, where we used the affine invariant pseudometric in [13] to measure similarity between waveforms. It is important to note that the shapes of the regular waveforms corresponding to respiration vary considerably throughout the record, even within 1 s. So, although our problem might look simple, it turned out that signal sections regarded as “regular” by the human were not straightforward to characterize mathematically, especially near transitions from regular to irregular segments.

The paper is structured as follows: In Section II, we present the structure of the overall system which consists of two parts, the *Feature Extractor* explained in Section III, and the *Classifier*, described in Section IV. The different experiments performed and their results are explained in Section V, followed by the conclusions in Section VI.

II. SYSTEM OVERVIEW

The automated sleep scoring system is designed to classify the piezoelectric signals as “Sleep” or “Wake” with reliability equal to, or greater than, scoring by visual inspection. The transducer consisted of a piece of piezoelectric film (Measurement Specialties, Inc., Wayne, Pennsylvania). The piezo film was composed of PVDF polymer 52 μm thick and coated with a thin layer of NiCu (part number – 0-1000420-0). The surface area is approximately 7.5 cm \times 15 cm. Two such films were placed side by side and wired in series (total cage flooring is, therefore, 15 cm \times 15 cm). Piezoelectric output was amplified and filtered with steep cut-offs below 0.5 Hz and above 15 Hz. The signal was converted using an A-D board from National Instruments, Inc. (Austin, Texas – part number 778075-01), using a sampling rate of 100 Hz. The signal input was processed and graphically displayed using commercially available software (LabView, National Instruments). Signal analysis was performed using the algorithms described in Sections III and IV.

EEG/EMG signal scoring is typically based on 4-s or 10-s epochs. We chose a 1-s epoch resolution for the analysis of the piezo signals. Each epoch was deemed to correspond to *sleep* if it contained a regular periodic-like signal characteristic of respiratory movement; a 1-s epoch that lacked a regular periodic component was deemed to be *wake*. The raw signal was initially visually scored by a trained sleep researcher.

The rules that the human expert follows in the classification of 1-s epochs as sleep or wake are very hard to specify and are rather arbitrary. This makes the automated classification process suitable for the use of a classifier that learns from examples, as

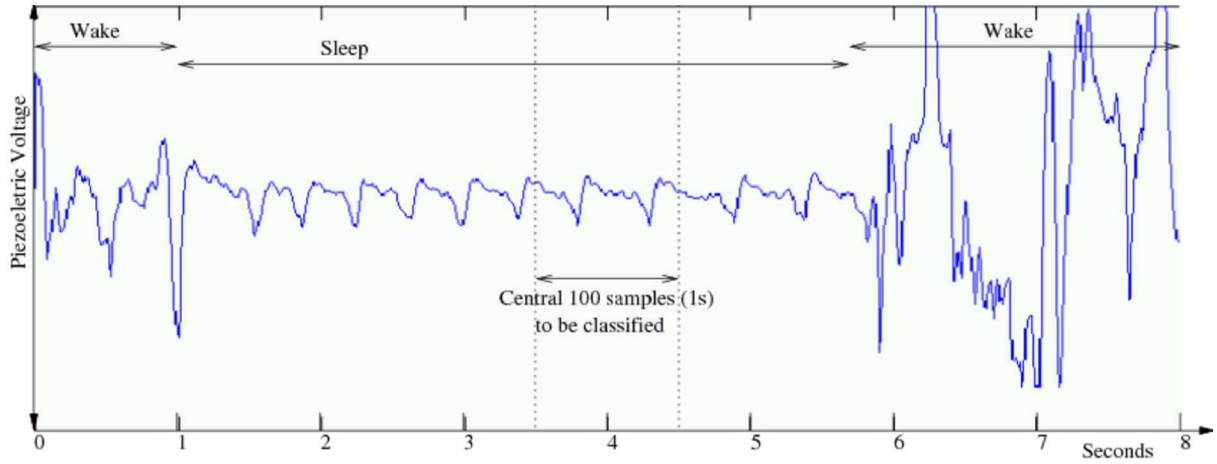


Fig. 3. Example of 8-s piezoelectric signal window used by the system to score the 1-s epoch in the center. Regions where the signal is regular are deemed sleep, whereas irregular sections are deemed wake.

opposed to one that would use a set of predefined (knowledge based) rules. Our classification system consists of two parts: the *feature extractor* and the *classifier*. The feature extractor quantifies certain characteristics of the signal by generating a feature vector for each 1-s epoch. Each component of the feature vector is a function of the signal contained in an 8-s window centered around that designated 1-s epoch. The feature vector is then fed to the classifier, which maps it into a binary label corresponding to sleep or wake. The classifier is implemented with a neural network, which is trained with a representative large sample of data scored by a human expert.

III. FEATURE EXTRACTION

The objective of the feature extractor is to produce for each 1-s epoch a collection of features that quantify certain characteristics of the signal to enable discrimination between sleep and wake. The features are stacked into a feature vector \mathbf{f} that is used as input to the classifier. The feature vector allows for encoding of some domain knowledge that is difficult for the classifier to recognize using raw data alone. The usual breathing rate of mice is between 2 and 4 Hz, hence, the breathing cycles are up to one-half second long. In order to identify these waveforms, the feature extractor uses a *context* data window of eight ss, centered around the 1-s epoch to be classified, as shown in Fig. 3.

All of the features are designed based on heuristic arguments. We started with a basic set of features based on Fourier and wavelet analysis complemented with several time domain features. We then sought the subset of those features to achieve the best classification accuracy. Each feature subset we tried was evaluated using the best classifier we could obtain for that subset; parameters involved in the feature extraction were adjusted using a south-well relaxation technique. Although this was highly computational intensive, it provided an effective way to select a good feature set, which we describe next.

We show intuitively how each feature gives some information not captured by the others. We denote by $x_n, n = 0, 1, 2, \dots, 799$ the 8-s context signal window centered around the 1-s epoch under consideration, which corresponds to $x_i, 350 \leq i \leq 449$. Since the magnitude of the signals from the recordings varies a lot, and has minimal effect on the scoring

decision used by the experts, all the features were designed to be scale invariant.

A. Features Based on Fourier Techniques

The obvious immediate candidates for feature extraction are Fourier and wavelet analysis. We expect that in the case of sleep signals, most of the spectral power of the signal is concentrated in the lower frequencies corresponding to breathing rates, whereas in wake, the animal moves and produces high-frequency signals. So we computed the 200-point DFT X_k of the signal in the 2-s window centered around the 1-s epoch in consideration (The signal was premultiplied by a Hanning window before calculating its DFT). The signal was sampled at 100 Hz, hence, we had 0.5 Hz frequency bins, and since the piezoelectric output was filtered above 15 Hz, we restricted our attention to X_k for $0 \leq k \leq 30$. The features are intended to be scale invariant, hence, the DC term X_0 was dropped and the power spectrum was normalized, obtaining $Y_k \triangleq |X_k|^2 / \sum_{n=1}^{30} |X_n|^2$, $k = 1, 2, \dots, 30$. Using the values Y_k directly as components of our feature vector proved insufficient to achieve enough accuracy; mostly because the spectral features were not suitable to classify epochs near transitions due to the limited time resolution using a 2-s window; using a shorter window did not improve the performance since frequency resolution was decreased and less context regularity information was included. Speculating that high dimensionality of the power spectrum vector could be a problem for the classifier, we tried using principal components analysis to reduce its dimensionality; this however, did not provide a significant improvement.

The percentage of spectral power corresponding to breathing frequencies (2–4 Hz) was considered too, but this was not a useful feature since the complexity of many waveforms made the power spectra of the signal to spread outside the breathing frequency range. This suggested the design of other features to characterize the spectrum distribution. Observing that $\sum_{k=1}^{30} Y_k = 1$, we can regard Y_k as a probability mass function and use its centroid $\sum_{n=1}^{30} nY_n$ as a feature, the frequency centroid tended to have smaller values for sleep compared to wake episodes. The spectrum of sleep signals is likely to be concentrated into a few sharp spikes, whereas for wake, is more spread out; so we used

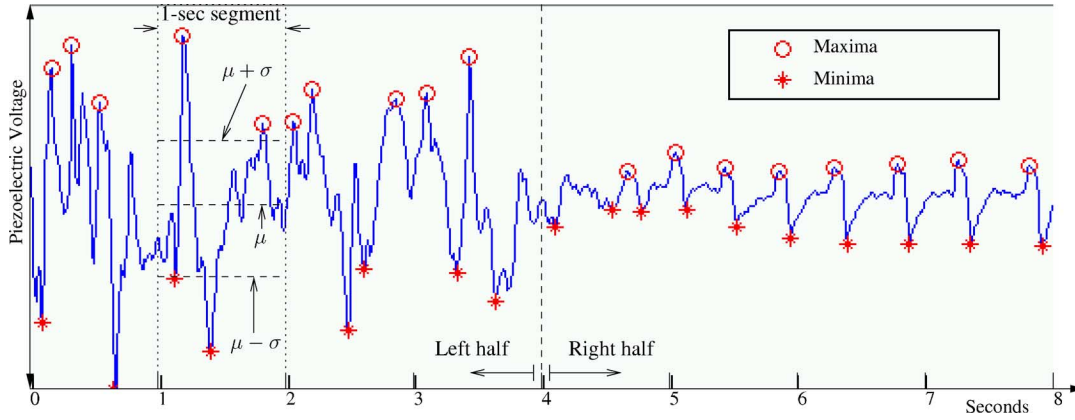


Fig. 4. Example of peaks in the piezo signal found by the peak-detection algorithm and description of the regions used for the peaks statistics calculation. Notice the regularity in amplitude and interspacing for the peaks in the right half corresponding to sleep, versus the irregularity in the left half corresponding to a wake episode.

$-\sum_{n=1}^{399} Y_n \log Y_n$, the entropy of Y_k , as a measure of spectrum spread. Since the entropy of a probability distribution is higher when it is closer to the uniform distribution, and smaller when it consists of a few spikes, the entropy of Y_k tended to have smaller values for sleep than for wake cases. Other more sophisticated functions of the Fourier spectrum were considered, such as scaled transforms in frequency [14] and phases of integrals of the bispectrum [15]. Also, in an attempt to counteract the limited frequency-time resolution of Fourier analysis, feature extraction was done with the wavelet coefficient using Db10, Daubechies compact support orthogonal wavelet with 20 taps [6]. Using only these features based on Fourier and wavelet analysis surprisingly gave limited results (with error rates often exceeding 20%); thus, other features were designed based on time domain analysis. When combined with the time domain features, the frequency centroid and frequency entropy were the most effective transform domain features, constituting the first two components f_1 and f_2 , respectively, of the feature vector implemented in our system. The other features based on Fourier and wavelet analysis did not survive the feature selection process as they seemed to impair the ability of the classifier to generalize. This was due perhaps to their relatively high dimensionality. Next, we describe the features based on time domain analysis that were included in the final feature vector.

B. Waveform Similarity Features

These features are based on the similarity between waveforms in the 1-s epoch and waveforms in the vicinity, and were designed along the same spirit in [11] and [12]. Let $s_{l,k}$ be the vector corresponding to the signal segment $2l$ samples long, centered at the $400 + k$ sample of the context window. Using a distance criterion $d(x, y)$, we compare the central waveform $s_{l,0}$ to the waveforms $s_{l,k}$ in a neighborhood around it, and define as a feature the minimum distance found, i.e.,

$$f_3 \triangleq \min_{l,k} d(s_{l,0}, s_{l,k}), \quad 10 \leq l \leq 50, \quad 2l \leq |k| \leq 3l.$$

This feature tended to have low values for sleep epochs and higher values for wake cases. Since we were interested in shape similarity only, the distance criterion $d(x, y)$ used was the affine invariant pseudometric defined in [13], which is the minimum

euclidean distance between x and all affine transformations on y . The search interval chosen for l is based on the fact that the period of breathing waveforms ranges from 0.25 s to 1 s. The interval search for k is set to consider segments that are 1–1.5 segment lengths apart from the central one. The value of l for which the minimum distance was found is an estimate of the period of the signal when regularity is present. This period estimate is our second waveform similarity feature f_4 , and it is used further as a parameter for the extraction of the adaptive filtering feature explained later.

C. Features Based on Extrema

The features previously explained still proved to be insufficient for our classification task. This was because in order for the signal to be scored as sleep by the human experts, it is often sufficient for the signal to exhibit a regular pattern of “peaks” uniformly spaced and having approximately the same “amplitude,” even if the corresponding waveforms are dissimilar. For this reason, we developed an algorithm to identify the local extrema regarded as peaks by a human expert, and then compute statistical measures of the uniformity in their time separation and height, constituting our set of features based on extrema.

1) *Finding the Extrema*: Trying to mimic the human experts, we decided upon the following methodology for identifying extrema they regard as peaks. The methodology for maxima and minima are symmetric, so we will only describe it for maxima. We divide the signal into 1-s segments overlapping by 50%. For each signal segment we calculate its mean μ and standard deviation σ , and mark off regions with a signal value larger than the threshold $\mu + \sigma$. The maximum signal point in each of these regions was considered a peak. We then prune the set of peaks found with the criteria that if multiple peaks lie within a 0.1-s interval, we keep only the largest one. Peaks corresponding to minima are identified in an analogous way. An example of the above process is shown in Fig. 4.

2) *Peaks Statistics*: Once the peaks in the signal are detected, the next step is to measure the uniformity in amplitude and spacing between the peaks in the following 4-s signal segments: the left half of the context window, the central 4-s segment, and the right half. In many cases of sleep epochs, the peak regularity is only observed for either the maxima or the minima, hence,

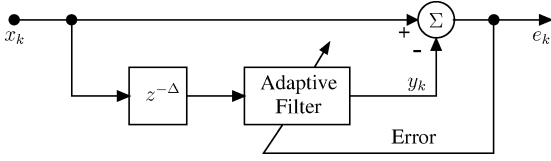


Fig. 5. Diagram of the adaptive linear predictor used to obtain a measure of predictability of the piezo signal. Predictability is regarded as a feature indicative of sleep episodes.

they were treated separately for peak regularity measurement. To measure uniformity in amplitude of a set of peaks, we use the ratio of their standard deviation to their mean, also known as the standard coefficient. This is a unitless measure of variability, so the smaller this number, the more uniform the amplitude is. For each of the three segments described above, the standard coefficient for their maxima is computed. The minimum of those coefficients defines f_5 , our first feature based on extrema. Repeating this procedure for their minima, we obtain the next feature f_6 . Substituting amplitudes with time separation between consecutive peaks in the above calculations, we obtain two more features f_7 and f_8 as measures of uniform spacing for maxima and minima, respectively. The rationale for these features based on peaks is that sleep epochs will have low values for one or several of these features, while wake epochs will tend to have higher values.

D. Feature Based on Adaptive Prediction

The features described so far proved to be very good, but not always sufficient to classify an epoch, especially near transitions. For example, for the three epochs in the center of Fig. 4 the features based on Fourier and on extrema will be rather similar, even though two of those epochs correspond to wake and the third one to sleep. The waveform similarity feature f_3 may not be too useful either since the waveform in the sleep epoch is not similar to the next waveforms. Hence, one more feature was designed to alleviate this problem. When the signal exhibits a regular pattern or periodicity, it is reasonable to expect a good linear prediction of the signal from previous samples. Based on this heuristic argument, we decided to use a measure of the signal's predictability as our last feature. Since the signal is nonstationary, an adaptive linear predictor [16], [17] is applied to the signal as shown in Fig. 5. The output of the adaptive filter is given by $y_k = W_k^T X_k$, where $X_k \triangleq [x_{k-\Delta} x_{k-\Delta-1} \cdots x_{k-\Delta-L+1}]^T$, Δ is the prediction time, and $W_k \in \mathbb{R}^L$. The weight vector W_k is adapted according to the Normalized LMS algorithm

$$\begin{aligned} \varepsilon_k &= x_k - y_k \\ W_{k+1} &= W_k + 2 \frac{\alpha}{\|X_k\|^2} \varepsilon_k X_k \end{aligned}$$

where α is an adaptation constant.¹ For each epoch, the measure of predictability was defined as the ratio γ of the power of the error ε_k to the power of the signal x_k in that epoch. Small values of γ indicate high predictability, high values indicate low predictability. It is clear that for the calculation of γ

¹Since all the signal sections with values very close to zero are removed before the feature extraction, the magnitude square $\|X_k\|^2$ is never close to zero, hence, instability of the algorithm was avoided.

for a given epoch, the algorithm is only needed to run with the data in the first 4.5 s of the context window. The initial weight vector W_0 is set to the least squares solution of the prediction problem corresponding to that data. The predictability of the signal from past samples can be low for epochs lying in transitions from wake to sleep, as is the case in Fig. 4. However, the signal at the center could be predicted well from “future” samples, hence, we decided to drive the adaptive linear predictor with the time-reversed signal $\tilde{x}_k = x_{800-k}$ and obtain its measure of predictability $\tilde{\gamma}$. Our last feature was then defined as $f_9 \triangleq \min(\gamma, \tilde{\gamma})$. Since a good choice for Δ is the period of the signal when it is periodic, we decided to make $\Delta = f_4$. The values of α and L were tweaked according to a south-well relaxation procedure, resulting in $\alpha = 0.1$ and $L = 2f_2$ for the best classification accuracy. This feature was useful in cases where the other features were insufficient for the discrimination between wake and sleep near transitions.

IV. CLASSIFIER

The feature vector for a given 1-s epoch is fed to the classifier which outputs a binary label corresponding to sleep or wake. The classifier is implemented with a feedforward artificial neural network (ANN),² trained and tested with data scored visually by human experts. The data set consisted of the feature vectors obtained from every 1-s epoch and their associated label assigned by the expert. Since we want the output of the ANN to be mapped to a binary label, its output layer consisted of a single neuron with a hyperbolic tangent activation function $\tanh(x)$, and the desired response for each input feature vector was set to +1 if it was labeled as sleep by the expert, and to -1 if regarded as wake. We partitioned our data set into three disjoint sets: a training set, a validation set and a test set. The training set was used to train the classifier (neural network) using the Levenberg-Marquadt algorithm [19], [20]. After each iteration of the training process, the misclassification rate of the classifier on the Validation set was calculated. Training was stopped when the misclassification rate on the validation set did not improve for 20 consecutive iterations. The misclassification rate for the Test set was used as an estimate for the performance of the trained classifier with data it has never been exposed to. The architecture of the neural network was designed using the network growing technique [18, p. 218]; we started with a single neuron, and as it was insufficient to achieve good results, we added a hidden layer; we increased the number of neurons in the hidden layer until adding more neurons did not improve the classification task. This procedure yielded a two-layer ANN, with 20 neurons in the first (hidden) layer and one neuron in the second (output) layer.³ All neurons used hyperbolic tangent activation functions and a bias weight. The ANNs were implemented and trained with the Neural Networks Toolbox from Matlab.

Once the neural network was trained, it was used as a binary classifier by taking the sign of its output. If its output for a given feature vector was positive, a sleep label was assigned to the corresponding 1-s epoch; if its output was negative a wake label was given.

²also known as feedforward multilayer perceptron [18].

³Adding one more hidden layer did not offer an advantage over the two-layer structure.

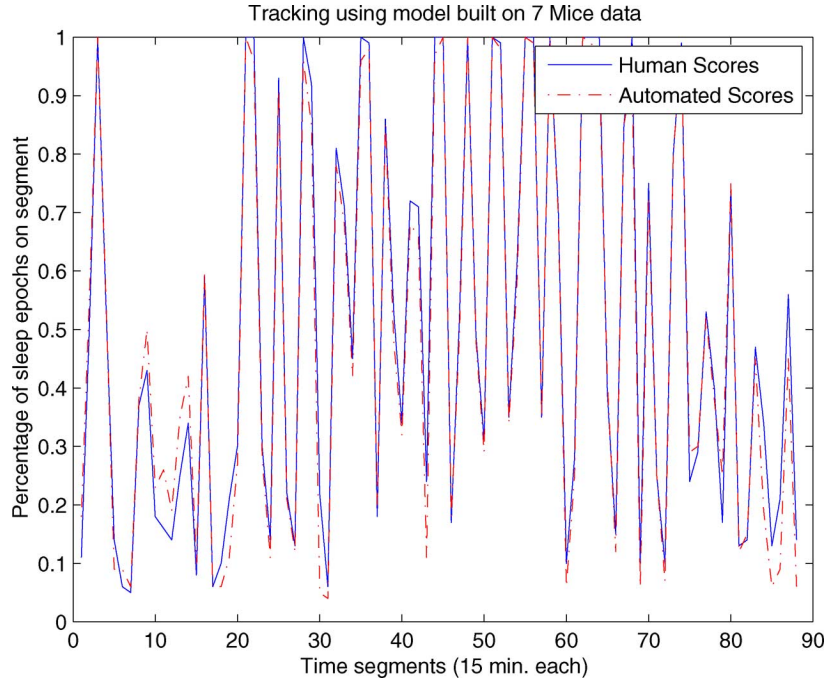


Fig. 6. Percentage of sleep time scored by human visual observation of signals versus automated scoring.

TABLE I
AGREEMENT MATRIX FOR TEST SET IN EXPERIMENT 1

	Sleep	Wake	% Sensitivity
Sleep	38932	1270	96.1 %
Wake	1891	39804	95.2 %
% Specificity	95.1%	96.2 %	

We also used Multiple Additive Regression Trees (MART) [21], [22] using the software described in [23] to benchmark our ANN. The MART classifier was particularly useful during the feature selection, since it provided measures of their “importance” in the classification tree; suggesting which features to keep, to remove, or to improve. Once our final feature set was defined, the results provided by MART were about the same as with ANNs. For reasons of implementation, we chose the ANN as the classifier for the final system.

V. EXPERIMENTS AND RESULTS

Piezo-electric signals from eight different mice were collected with similar recording conditions and prefiltering as described in Section II. The mice were commonly used inbred strains: C57BL/6J, AKR/J, and DBA/2J. These data included both sleep and wake episodes, i.e., regular and nonregular signal waveforms. Each mouse was recorded for 24 hrs, and for seven of the mice, the first 10 min of each hour was scored by a human expert. This provided 14 400 1-s scored epochs per mouse. For the eighth mouse, the entire 24 hrs (86 400 1-s epochs) was scored by another human expert. To test the validity of our automated piezo system for distinguishing sleep versus wake, we first tested automated scoring versus human scoring of the same piezo signals (experiment 1 and 2, below). We next compared the automated piezo scoring versus EEG/EMG scoring in mice recorded simultaneously with both systems (experiment 3).

A. Experiment 1

For each of the first seven mice, 80% of the wake and sleep epochs were chosen at random to build a training set, the remaining 20% was used for validation. Wake vs sleep was determined visually by the regularity of the piezo signal based on our experience with simultaneous EEG/EMG recordings and visual observations of the mouse. The data from the 8th mouse (24 hours of data scored by a human expert) was used as the Test set. The accuracy achieved by the classifier trained in this experiment was 95.7% on the test set; the specificity and sensitivity ranged between 95% and 96% for both sleep and wake epochs, as shown in the agreement matrix obtained in Table I.

The percentage of time the rodent is sleeping as opposed to awake is an important statistic for sleep studies, and is under genetic control [24]. Hence, we divided the 24 hrs of data from the eight mice into 15-min nonoverlapping intervals, and for each of them we calculated the percentage of 1-s epochs the mouse was sleeping according to both the human score and the automated score given by the classifier. Fig. 6 shows the ability of the automated system to track the percentage of sleep behavior of the mouse. This indicates that the performance of our system is consistent through time, with errors distributed in a more or less uniform manner throughout the whole recording.

B. Experiment 2

In order to test the robustness of the classifier training methodology against subject variability, we designed the following experiment. The data from one of the first seven mice was used as the Test set. The data from the remaining six mice was partitioned at random in a 80%-20% proportion as described in experiment 1, to build the training and validation set. In this manner, for each mouse a classifier was trained with data from the remaining mice. The classification accuracy obtained

TABLE II
OVERALL ACCURACY, SENSITIVITY AND SPECIFICITY IN EXPERIMENT 2

Mouse tested	1	2	3	4	5	6	7
Overall % Accuracy	95.5	96.1	95.2	95.1	94.5	95.3	95.7
% Sens. Sleep	94.6	96.2	96.7	95.2	95.5	94.0	95.9
% Sens. Wake	96.4	95.8	93.2	95.0	92.8	96.6	92.9
% Spec. Sleep	96.5	96.6	95.0	96.9	95.2	96.5	94.4
% Spec. Wake	94.4	95.4	95.5	92.2	93.4	94.1	94.8
# epochs Sleep	7599	7610	8172	7696	7885	7692	7711
# epochs Wake	7198	9198	6092	4576	5356	7686	6190

in the test mouse for each classifier is shown in Table II. The overall accuracy for all mice was about the same as in the previous experiment. Sensitivity and specificity for sleep epochs was mostly between 95% and 96%, whereas for wake, both sensitivity and specificity were in the range of 94% to 96%, with a few cases around 93%. The test sets used in experiment 2 contained between 7000 and 8000 epochs each, and although they are approximately six times smaller than the test set used in experiment 1, the results obtained showed good consistency of our system across subjects. Our results illustrate the ability of our system to perform well on recordings it has never been exposed to, given that it has been trained with data from a variety of different subjects under similar recording conditions. This indicates that further training of our system should only be necessary after significant changes in the recording conditions have occurred, such as hardware changes or noisier outputs of the piezoelectric signal. We should also mention that a thorough examination of the epochs where mistakes of the automated classifier were observed indicated that most of the errors were made in the state transitions; this was no surprise since even the human scoring of such epochs can be rather ambiguous.

The results in experiments 1 and 2 show that the accuracy of our automated classification system is between 95% and 96% in the test sets, comparable to the scoring agreement between human experts which is around 95%.

C. Experiment 3

As another test of our system, we next validated our piezo system against the gold standard of sleep/wake scoring by comparing EEG/EMG and piezo signals simultaneously from the same mice. EEG/EMG methodologies were as described in Franken *et al.* [24], while piezo scoring was done by the automated classifier as described above. Fig. 7 shows combined data from all eight mice. Both EEG/EMG and piezo systems were used for a simple two state classification of sleep versus wake. Overall, EEG/EMG and piezo scoring matched approximately 90% of the time. This similarity is quite good since even EEG/EMG scoring of the same record by two different individuals will differ to some extent. This result also supports our earlier contention that quiet wake is not a substantial problem for our system presumably because of the subtle movements mice make during quiet wake that are picked up by the piezo

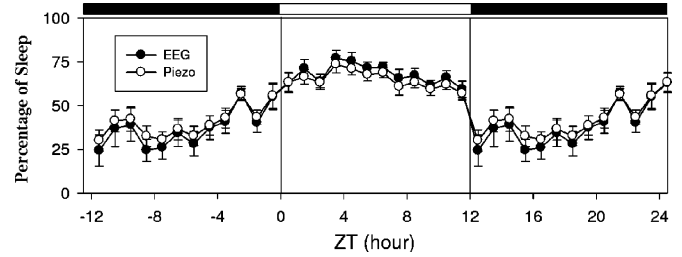


Fig. 7. Simultaneous piezo versus EEG/EMG output was classified and compared as described in the text. The graph represents a compilation of data from eight different mice. The y-axis shows the percentage of time (in one hour bins) classified as sleep by the piezo system (open circles) and by EEG/EMG scoring (closed circles). EEG/EMG was done according to Franken *et al.* 1999. Note the high correspondence between the two methods, supporting the fact that the piezo system can accurately distinguish sleep from wake.

film, and also because periods of extremely quiet wake are relatively rare in mice. Occasional movements occur during sleep as well, but to a much lesser degree.

D. Experiment 4

Lastly, as a first test of our system for the screening of sleep and wake promoting drugs, we chose the commonly used stimulant caffeine. As expected, caffeine injection (20 mg/kg, ip) dramatically increased wakefulness following the injection for three to four hours, and this was accurately reflected by the automated scoring with the piezo system as shown in Fig. 8. The automated scoring was consistent with hand scoring of the piezo signals and visual observations of the mice, but no simultaneous EEG/EMG recordings were conducted in this study.

VI. CONCLUSION AND FUTURE WORK

We have successfully developed an automated classifier of the piezoelectric signal into sleep and wake episodes. Its accuracy is comparable to the percentage agreement between human scorers. Implementation of this technology for high-throughput screening of sleep and wake in mice has already begun at the Tennessee Mouse Genome Consortium [5] to investigate genetic aspects of sleep and sleep disorders using both mutant mice and large sets of recombinant inbred strains. However, the piezo system is still an initial screening tool that may require EEG/EMG follow-up studies on the most interesting mice, since it is possible that certain genetic alterations may decrease waking movements or increase sleep movements in a way that increases classification errors (i.e., sleep/wake scoring). EEG/EMG follow-up studies would also provide details on REM sleep versus SWS that has not yet been achieved using our piezo signal (but, see next paragraph). On the plus side, our piezo system may have certain advantages over EEG/EMG. In addition to the identification of genes that influence total sleep time, sleep bout durations, circadian, and time of day variables, etc., we should also be able to easily spot any periods of interrupted breathing during sleep since the piezo signal provides an accurate respiratory trace during sleep. Our system should also be useful for rapid initial screens of sleep and wake promoting drugs, but again with the caveat that certain drugs may alter movements during wake or sleep in a way that increases classification errors. In general, this system should allow the analysis of a large number of rodents in a relatively

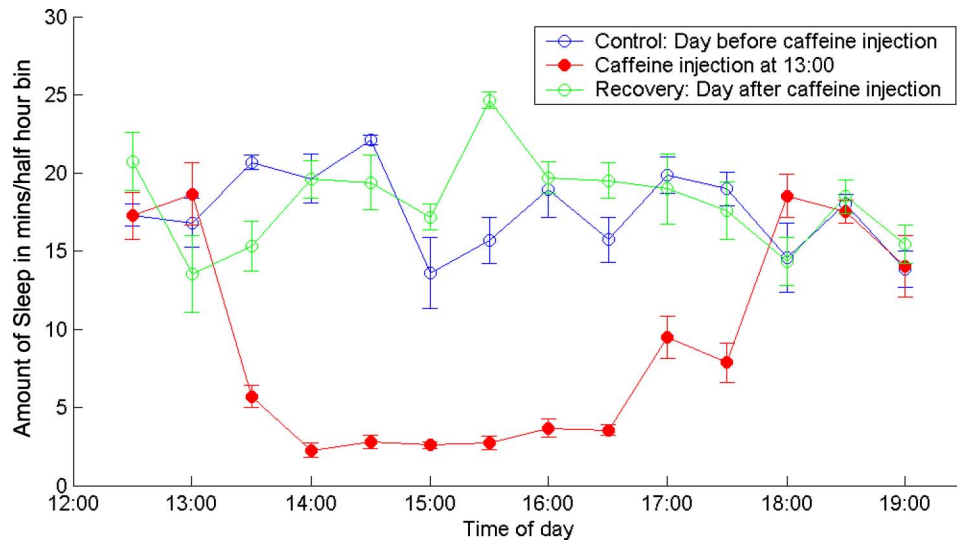


Fig. 8. Piezo scoring of increased wakefulness following caffeine in mice (20 mg/kg, ip). The time-course plot shows average amount of sleep in minutes per 30-min bin in each mouse on the day before caffeine injection, the day of injection, and the day following injection (mean of four mice, error bar = SEM).

short time, which was previously not feasible due to the time and cost involved with traditional EEG/EMG methodology.

It has been known for decades that respiration becomes less regular during REMS than SWS raising the possibility that we can also distinguish these states with our piezo system. This change in respiratory pattern is often visually apparent in our piezo signals, however, the change is quite subtle. Therefore, in this initial system we have simply differentiated sleep from wake. A high-throughput system that can make even this simple distinction is still of great value as an initial screen in studies requiring large numbers of mice. Follow-up studies using EEG and EMG can also be done. We will also continue development of this system with the goal of distinguishing SWS and REMS in the next generation system.

It is important to note that some of the techniques we used for feature extraction can be used in other pattern recognition problems where the objective is to detect *regular* patterns or periodicity in signals, regardless of the shape of the patterns. Further refinement of these approaches might allow us to distinguish different types of waking behavior such as grooming and rearing that appear to have unique signal characteristics. This could make our piezo system even more useful in a wide variety of behavioral studies.

ACKNOWLEDGMENT

The authors would like to thank G. Hagiwara for her invaluable work recording and scoring the piezo signals, and for her insightful comments that helped to develop our system. They would also like to thank B. Widrow for generous help and assistance on the algorithms and manuscript.

REFERENCES

- [1] M. Kryger, T. Roth, and W. Dement, *Principles and Practice of Sleep Medicine*. Philadelphia, PA, USA: Saunders, 2000.
- [2] E. B. Thoman and W. D. Tynan, "Sleep states and wakefulness in human infants: Profiles from motility monitoring," *Physiol. Behavior*, vol. 23, no. 3, pp. 519–526, 1979.
- [3] M. Erkinjuntti, P. Kero, J. P. Halonen, H. Mikola, and K. Sainio, "SCSB method compared to EEG-based polygraph in sleep state scoring of newborn infants," *Acta Paediatrica Scandinavica*, vol. 79, no. 3, pp. 274–279, Mar. 1990.
- [4] D. Carroll, V. Denenberg, and E. B. Thoman, "Reliability and validity of computer scoring of behavioral sleep-wake states in rats and rabbits," *Physiol. Behavior*, vol. 54, no. 2, pp. 269–273, Aug. 1993.
- [5] D. Goldowitz, W. N. Frankel, J. S. Takahashi, M. Holtz-Vitaterna, C. Bult, W. A. Kibbe, J. Snoddy, Y. Li, J. P. S. Yates, and D. J. Swanson, "Large-scale mutagenesis of the mouse to understand the genetic bases of nervous system structure and function," *Brain Res. Mol. Brain Res.*, vol. 132, no. 2, pp. 105–115, Dec. 2004.
- [6] V. Venkatachalam and J. L. Aravena, "Detecting periodic behaviour in nonstationary signals," in *International Symposium on Time-Frequency and Time-Scale Analysis*, Oct. 6–9, 1998, pp. 497–500.
- [7] W. A. Sethares and T. W. Staley, "Periodicity transforms," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2953–2964, Nov. 1999.
- [8] P. P. Kanjilal and S. Palit, "Extraction of multiple periodic waveforms from noisy data," in *Proc. ICASSP '94. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Adelaide, SA, Australia, Apr. 1994, vol. 2, pp. II/361–II/364.
- [9] —, "On multiple pattern extraction using singular value decomposition," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1536–1540, Jun. 1995.
- [10] R. M. Gagliardi, "Noncoherent detection of periodic signals," *IEEE Trans. Inf. Theory*, vol. IT/21, pp. 282–285, May 1975.
- [11] G. Spence and I. Clarke, "A time-domain analysis of multiple periodic signals," in *Proc. Inst. Electr. Eng. Seminar Time-Scale and Time-Frequency Analysis and Applications*, London, U.K., 2000, pp. 2/1–2/10.
- [12] P. Vosschulte, G. Kurlmann, and E. Harms, "Identification of periodic patterns in biosignals by computation of recurrence times of unstable periodic orbits," in *Proc. Annu. Int. Conf. IEEE Engineering in Medicine and Biology*, Sep. 1995, vol. 17, pp. 1473–1474.
- [13] D. Nelson, "Special purpose correlation functions for improved signal detection and parameter estimation," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP '93)*, Minneapolis, MN, USA, Apr. 27–30, 1993, vol. 4, pp. 73–76.
- [14] N. Marinovich, L. Cohen, and S. Umesh, "Joint representation in time and frequency scale for harmonic type signals," in *Proc. IEEE-SP Int. Symp. Time-Frequency and Time-Scale Analysis*, Philadelphia, PA, USA, Oct. 25–28, 1994, pp. 84–87.
- [15] V. Chandran and S. L. Elgar, "Pattern recognition using invariants defined from higher order spectra-one-dimensional inputs," *IEEE Trans. Signal Process.*, vol. 41, no. 1, pp. 205–12, Jan. 1993.
- [16] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1985.
- [17] B. Widrow, J. Glover, Jr., J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, Jr., and R. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

- [18] S. Haykin, *Neural Networks, a Comprehensive Foundation*. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [19] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 11, pp. 989–993, Nov. 1994.
- [20] M. Hagan, H. Demuth, and M. Beale, *Neural Network Design*. Boston, MA: PWS Publishing, 1996.
- [21] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
- [22] —, "Stochastic gradient boosting," *Computational Statist. Data Anal.*, vol. 38, no. 4, pp. 367–78, Feb. 2002.
- [23] —, Tutorial: Getting Started with MART in R 2002 [Online]. Available: <http://www-stat.stanford.edu/~jhf/r-mart/tutorial/tutorial.pdf>
- [24] Franken, A. Malafosse, and M. Tafti, "Genetic determinants of sleep regulation in inbred mice," *Sleep*, vol. 22, pp. 155–169, 1999.



Aaron E. Flores received the B.S. degree in electronic systems from ITESM campus Guadalajara, Mexico in 1995. He received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2005, respectively. He is currently doing research developing algorithms for optimal allocation of inventory in the internet advertising business. His research interests include adaptive signal processing, statistical learning, and data mining.

Dr. Flores received the Centennial Teaching Assistant Award from Stanford University in 2005.



Judith E. Flores received the B.S. degree in chemistry, pharmacy, and biology from Guadalajara University, Guadalajara, Mexico, in 2001.

From 2001–2003, she was a Life Science Research Assistant with the Sleep Research Lab in the Biological Sciences Department, Stanford University, Stanford, CA. In 2003, she worked as a Research Assistant in the Sleep Research and Neurobiology Lab at Stanford Research Institute International until April 2006. She is currently a Research Associate in the Biomedical Imaging group at the Tumor Biology and

Angiogenesis Department, Genentech, Inc., South San Francisco, CA.



Hrishikesh Deshpande received the B.Tech. degree from the Indian Institute of Technology, Bombay, India, in 2000 and the M.S. degree in computer science from Stanford University, Stanford, CA, in 2002.

His publications are in the areas of machine and statistical learning, computer vision, and computer graphics. He is currently a Research Engineer with Sony Computer Entertainment (PlayStation), Foster City, CA, working on computer vision problems related to the EyeToy video camera for console game

applications. He holds patents for various user interface applications of image processing and computer vision techniques.



Jorge A. Picazo received the B.Sc. degree (with Honors) in actuarial sciences from the National Autonomous University of Mexico (UNAM), Mexico City, Mexico, in 1994; the M.S. degree in Engineering-Economic Systems and Operations Research in 1999 and the Ph.D. degree in statistics in 2002 from Stanford University, Stanford, CA.

He is currently doing research and developing statistical end econometric models in the financial industry. His research interests include data mining, prediction algorithms, and Monte Carlo Simulations.



Xinmin (Simon) Xie received the M.D. degree from the University of Bern, Bern, Switzerland, in 1989 and the Ph.D. degree from the University of London, London, U.K., in 1992.

He is the founder and Chief Scientific Officer of AfaSci, Inc., Burlingame, CA, which specializes in and is committed to developing HTS tools for behavioral phenotyping and drug screens *in vivo*. He also serves as Manager of the HTS Lab, SRI International, Menlo Park, CA. Prior to the establishment of AfaSci in 2003, he worked at GlaxoWellcome, Stevenage, U.K., Theravance, South San Francisco, CA, and Deltagen, San Carlo, CA, where he led electrophysiologic studies and phenotyping of gene-knockout mice, and participated in HTS assay development and drug discovery.



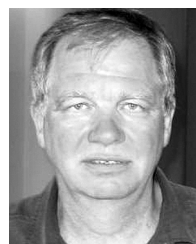
Paul Franken received the Ph.D. degree from the University of Groningen, Groningen, Netherlands, in 1993.

Between 1993 and 1996, he was a Postdoctoral Fellow with Stanford University, Stanford, CA. He then joined the University of Geneva, Geneva, Switzerland, where he began to genetically map sleep and EEG traits in mice. In 2000, he moved to Stanford University and established an independent research group studying the genetics of sleep homeostasis and the molecular interactions between circadian rhythms, sleep homeostasis, and brain metabolism. In 2005, he joined the Center for Integrative Genomics at the University of Lausanne.



H. Craig Heller received the Ph.D. degree from Yale University, New Haven, CT, in 1970.

He joined Stanford University, Stanford, CA, in 1972, where he is the Lorry I. Lokey/Business Wire Professor in Biological Sciences. His laboratory works on the neurobiology of sleep and circadian rhythms. His sleep work has focused on functions of sleep states, development of sleep, and the genetics of sleep. His circadian work has focused on mechanisms of entrainment. His laboratory also studies regulation of body temperature, mammalian hibernation, and human exercise physiology focusing on the effects of body temperature on physical conditioning and performance.



Dennis A. Grahn received the B.S. degree from the University of Oregon, Eugene, in 1975, the M.S. degree from Portland State University, Portland, OR, in 1984, and the Ph.D. degree in biological sciences from Stanford University, Stanford, CA, in 1989.

He is a Senior Research Scientist in the Department of Biological Sciences, Stanford University. His research interests are the control of mammalian homeostatic processes and the interactions between these processes, specifically temperature regulation, sleep, and circadian mechanisms. A variety of animals models have been used in the research including elephant seals, hibernating black bears, ground squirrels, and rat pups. His current research focuses on temperature regulation and physical activity in humans.



Bruce F. O'Hara received the Ph.D. degree in human genetics from the Johns Hopkins University, Baltimore, MD, in 1988.

He did post-doctoral work at the National Institute on Drug Abuse and then joined the Stanford Center for Sleep and Circadian Neurobiology, Stanford, CA, in 1991 focusing on molecular and genetic approaches towards an understanding of sleep, circadian rhythms and hibernation. He joined the faculty of the University of Kentucky, Lexington, in 2003, as an Associate Professor in biology, where he has continued these research efforts, including the development of technology for noninvasive, high throughput monitoring of sleep and wake in mice.