Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 108 (2007) 150–170
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Interacting with human physiology

I. Pavlidis *, J. Dowdall, N. Sun, C. Puri, J. Fei, M. Garbey

Department of Computer Science, University of Houston, Houston, TX 77204, USA

Received 17 October 2005; accepted 21 November 2006
Available online 31 January 2007
Communicated by Mathias Kolsch

Abstract

We propose a novel system that incorporates physiological monitoring as part of the human–computer interface. The sensing element is a thermal camera that is employed as a computer peripheral. Through bioheat modeling of facial imagery almost the full range of vital signs can be extracted, including localize blood flow, cardiac pulse, and breath rate. This physiological information can then be used to draw inferences about a variety of health symptoms and psychological states. Our research aims to realize the notion of desktop health monitoring and create truly collaborative interactions in which humans and machines are both observing and responding.

Keywords: Human–computer interaction; Thermal imaging; Facial tracking; Blood flow; Cardiac pulse; Breath rate; Stress; Sleep apnea

1. Introduction

Computers are primarily used for information exchange. They help users to perform a task, solve a problem, be entertained or educate themselves. Perhaps there is a more dramatic role for computers. Significant portions of the population in the developed world spend most of their daily lives interacting with computers. This is a tremendous time investment. Thinking in this context, computers may not give back to their users as much as they should.

We propose to add a new dimension in Human–Computer Interaction (HCI), namely, to monitor the physiology of computer users on a sustained basis and take appropriate actions when warranted. Our research aspires to use the abundant computing resources at home and the office in combination with novel sensing, algorithmic, and interface methods to enhance the user's experience and at the same time create a new preventive medicine paradigm.

Specifically, we collect sensory data through a thermal camera. This is a high quality Mid-Wave Infra-Red (MWIR) sensor (Phoenix model) from FLIR Inc. [1], sensitive in the spectral range 3–5 μm and with noise equivalent detection temperature (NEDT) 0.01 °C. Thermal imaging is a passive modality, which means that it depends solely on radiation emitted naturally by the subject. The thermal camera is connected to the computer as a typical peripheral and pointed towards the face of the user (see Fig. 1(a)). Imaging of the face is not only convenient but also advantageous. The face of the computer user is typically exposed. It also features a thin layer of tissue, which facilitates observation in a surface modality such as thermal infrared. Since sensing is done through imaging, it is contact free.

In a series of conference papers, we have demonstrated that one can extract a variety of physiological variables from the facial thermal video through bioheat modeling. At the moment, we are able to compute superficial blood flow [2,3], cardiac pulse [4], and breath rate [5]. In typical scenarios a user's face moves freely, therefore successful application of our bioheat models depends on reliable facial tracking [6]. This initial work of ours has established the feasibility and laid the foundation for contact-free physiological computation, an idea we first postulated in [7].

* Corresponding author. Fax: +1 713 743 3335.
E-mail address: ipavlidi@central.uh.edu (I. Pavlidis).

1077-3142/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2006.11.018

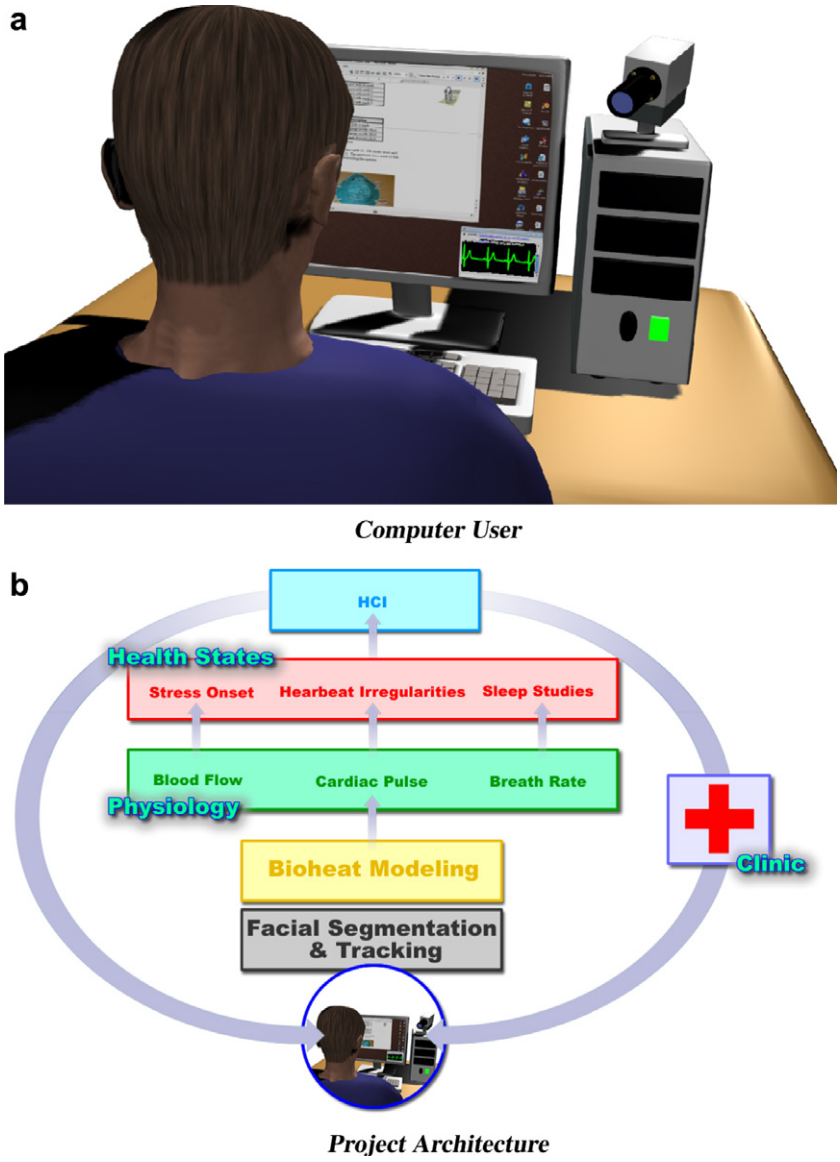

Fig. 1. (a) Typical setting where a computer user is being imaged through a thermal camera. (b) The relationship between the various components of our project.

We use localized blood flow, cardiac pulse, and breath rate signals to monitor incidents of stress onset, heartbeat irregularities, and sleep apnea respectively. We have discovered in previous work that instantaneous stress brings about an increase in the periorbital blood flow [8–10] and sustained stress is associated with elevated blood flow in the forehead [11]. Therefore, both stress conditions are detectable within our framework. In the cases of heartbeat irregularities and sleep apnea, we did not really discover new physiological variables for monitoring these pathological situations. We simply introduce novel ways of measuring the standard vital signs associated with them (i.e., cardiac pulse and breath rate).

The computational and biomedical aspects of our work are meant to facilitate the HCI component of the project (see Fig. 1(b)). Specifically, we place great emphasis on using stress as an emotional indicator of frustrated computer users. Based on the non-intrusive nature of our methodology and the success forecasted by our pilot studies, the value of pinpointing sources of computer user frustration cannot be underestimated. We also discuss appropriate machine responses to alleviate such stress. Fundamental HCI questions that stem from presenting and acting upon real-time health information (e.g., incidents of sleep apnea) are also addressed.

The implications of this research may be far-reaching. It will help to pinpoint and better understand the sources of frustration for the computer user. At a more general level, our research leverages novel computer vision technology to redefine the way people think and practice health care. Under the new paradigm, part of health care will not be administered periodically, off-line, and at special locations. But, it will be administered on a continuous basis, on-line, in a highly automated fashion, at home and the office.

Certain aspects of health care may become an integral part of HCI.

In the remaining paper, we first contrast our work with the existing body of research (Section 2). Then, we describe our facial tissue tracking algorithm and the physiological measurement methods (Sections 3–6). In Section 7 we discuss the application of facial blood flow measurements to stress quantification. In Section 8 we describe recent results from the application of the technology in sleep studies and the detection and management of sleep apnea incidents. We conclude the paper in Section 9 where we discuss our ongoing work on harvesting the interactivity potential of the technology.

2. Relation to the present state of knowledge in the field

To the best of our knowledge, this is a unique research effort and there is no previous work with regard to the totality of the proposed undertaking. Because this work represents a fundamental shift in the HCI paradigm, the related literature is massive. For the sake of brevity we will focus on the most relevant areas: facial segmentation and tracking; bioheat modeling; health state determination; and the role of emotion in HCI.

We apply our bioheat computations on the subject's face. The face is usually exposed and within the field of view of the thermal camera attached to the computer. The face of a typical computer user moves freely all the time making facial tracking necessary for the meaningful application of bioheat modeling. Although, facial tracking is a relatively well researched issue [12,13], none of the methods reported in the literature is appropriate for the application at hand. Most of them apply to the visible, not the thermal infrared band. All of them are meant to facilitate face recognition or surveillance not physiological modeling. Typically, in the aforementioned methods, tracking accuracy can fluctuate, and as long as it is regained momentarily, the effect on end applications (e.g., surveillance) is minimal. However, in physiological monitoring applications such as ours, even temporary loss of tracking creates serious problems. It results in spikes in the thermal signal of the tracked tissue, which are falsely indicative of strong physiological responses. Even when tracking is maintained, its accuracy typically degrades over time. We have developed a new facial tracker that addresses to a significant degree the aforementioned problems and works closely with our bioheat models [6].

Many physiological variables, such as superficial blood flow and cardiac pulse, are related in some way to the heat transfer mechanism of the human body. Due to this interrelationship, it is possible to compute such physiological variables from the raw thermal imagery. Relatively little work exists in the literature towards this direction. The most relevant reference is [14] where a simple heat transfer model is proposed for the computation of superficial blood flow out of thermal video. There are several problems with the approach in [14]. First, it does not provide for automated subject registration and tracking. As a result, the subject has to stand still during the examination, something out of the question in our case. Second, the bioheat model is simplistic and introduces a significant error in the presence of major vessels close to the skin. In contrast, we propose a novel pair of bioheat models, one for tissue with no major vessels [2] and one for tissue featuring at least one major vessel [3]. The models are based on partial differential equations (PDE). In addition to computing more accurately blood flow, our models allow the measurement of cardiac pulse [4].

The time evolution of 2D blood flow, cardiac pulse, and breath rate can reveal important clues about many health problems. As initial examples, we focus our attention to stress, heartbeat irregularities, and sleep apnea. In the past, thermal imaging was used for breast tumor detection often on an empirical basis [15–21]. In contrast, our approach is based on rigorous scientific methodology. We screen for a variety of health symptoms and not the manifestations of a specific ailment (e.g., cancer). Moreover, our screening is automated, continuous, and based on rigorous mathematical modeling not heuristic interpretations [7].

Several studies exist in the literature regarding the physiological profile of sleep apnea [22,23] and heartbeat irregularities [24–26]. None of these efforts, however, approaches the problem from the angle of continuous contact-free physiological monitoring. Since we can retrieve the breath rate and the cardiac pulse waveform through bioheat modeling, we can also incorporate monitoring of sleep apnea and heartbeat irregularities in the proposed HCI scheme. Our work may offer not only an elegant way of screening and monitoring sleep apnea and heartbeat irregularities but provide a deeper insight into the ailments, due to the unprecedented volume of information that is capable of producing.

We have recently developed methods for contact-free detection of elevated stress levels [8–11]. Such methods can be integrated in the proposed HCI scheme because they are based on physiological variables computable from bioheat modeling of the face. Stress as an emotional indicator is of particular importance and relevance to HCI. It is a widespread belief in the human–computer interaction community that new computing ideas should focus on human needs [27] and on the importance of affect and emotion [28]. The role of emotion in HCI has gained so much attention that one author recently used the term "emotion commotion" to describe the current state of affairs [29]. This activity is largely driven by the observation that, in spite of decades of effort refining user interfaces, computers still routinely anger, annoy, insult, and generally frustrate users. Early attempts to build more "friendly" and "helpful" interfaces, such as the ill-fated *Bob* and *Mr. Clippy* efforts were found by many users to be even more annoying than the interfaces they replaced. One explanation for this phenomenon is that these interfaces conveyed apparent emotion but lacked the ability to detect the emotions of the user [30]. Inappropriate responses to emotional states

are worse than no response at all [31,32]. The interest in emotions is supported by observations that emotions can be potent facilitators of cognitive processes [33] and that systems that create a positive affect can improve performance [34]. Furthermore, emotional state is an important contributor to many illnesses. Computers that could assist in the management of negative emotions could prove widely beneficial [35].

Before computers can respond adequately, they must be able to detect the emotions. A number of approaches have been researched. Nakatsu, Nicholson, and Tosa have applied neural network approaches to the analysis of speech input, achieving a 50% speaker-independent recognition rate [36]. Facial affect recognition has been studied by a number of groups, including Lisetti and Nasoz, who reported that facial segmentation can improve recognition rates [37]. A major limitation with both voice and facial-based approaches is that humans can be quite skilled at masking emotion conveyed through these modes. A number of physiological parameters have been explored as emotion indicators including heart rate, galvanic skin response [37], blood volume pulse (BVP), and electromyelography (EMG) [38]. An advantage of the physiological emotion indicators is that they are primarily under control of the autonomic nervous system and are less susceptible to conscious control. A major limitation to current physiological approaches is the need for sensors in direct contact with the user, or even implanted [39]. As a result, such sensors are impractical for most routine user environments. Sensor-based approaches have achieved user-specific accuracy rates of 60–80% in experimental conditions [38]. To be widely deployed, such systems would need to advance from user-specific to user-independent modes.

Recognizing the emotional state of the user is only the first step. The system must also respond appropriately [35]. As expected, there is considerably less information regarding this second step in the process. It has been shown that simply acknowledging the emotion is not adequate [31]. Picard proposes the "person-supporting-you test" [35]. Put simply, "How would I want a person who found me in this emotional state to respond?" For situations where mitigation of a negative emotion is the goal, Klein and Picard have proposed an eight-step approach [31]. For other situations the answer is likely to be highly application and context-dependent.

3. Facial tracking

The core of this research depends on temperature recordings at specific parts of the human face. The outcome is 2D time-varying temperature signals. These signals are used within a signal processing framework or as boundary conditions in partial differential equations (PDE) for the computation of vital signs. Typical cases include temperature recordings on the temporal area for pulse computation or on the periorbital area for determining the rate of blood perfusion in the orbital muscle. Obviously, these recordings need to be done on the same tissue area for a period of time in order to be meaningful. However, the face is in constant motion, which makes such a recording task challenging to say the least. The solution can be provided only through tracking technology.

In standard computer vision tracking, accuracy can fluctuate. However, in physiological monitoring applications such as ours, persistent high quality tracking is very important. To address this issue, we have introduced the concept of *tandem tracking*. Simultaneously with the small facial region of interest (e.g., temporal, periorbital, or nasal) we track an extended central facial region that is rich in contrasting features. The latter is more invariant to out of plane rotations. Therefore, the quality of tracking for this central region is better on average and can be used to correct the estimate of the regional tracker. This leads to the extraction of a superior quality temperature signal (see Fig. 2).

We refer to the central facial region tracker as the *tandem tracker* (TT) and the specific tissue tracker, where the physiological measurement is performed, as the *measurement tracker* (MT). The tandem tracking scheme works as follows:

(1) The system selects two rectangular regions on the initial frame of the thermal clip—the central facial and a specific tissue region. Depending on the physiological measurement, the specific tissue region may be centered on the periorbital (blood flow), the forehead (blood flow), the carotid (blood flow/cardiac pulse), the temporal (cardiac pulse) or the nostrils (breath rate).
(2) The central facial and specific tissue regions are tracked by two independent CONDENSATION trackers, the TT and MT, respectively. The feedback measurement for each tracker is based on a template. In essence, this is a sub-sampling of the tracker's initial rectangular region. We have determined experimentally that a surprisingly small sub-sample of the original area produces the same tracking result as the full area.
(3) The relative spatial position of the central and specific regions is established in the initial frame.
(4) In subsequent frames the CONDENSATION trackers come up with independent position estimates. The position estimate of the TT tracker is computed first and is used to influence the MT tracker. This combined estimate provides for a robust solution and helps the MT tracker to overcome local extrema (see Fig. 3).

3.1. Information fusion

The essence of the tandem tracking mechanism is based on the notion of information fusion. It uses the information from a robust high-level tracker (the TT tracker) to

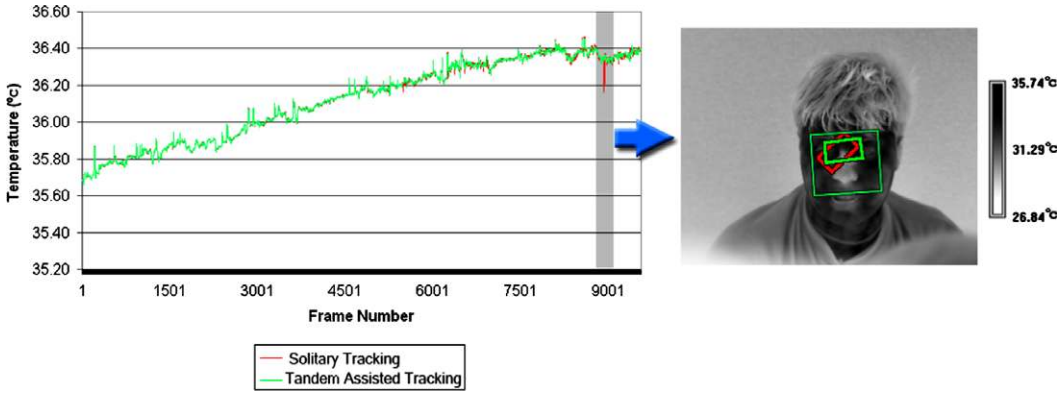

Fig. 2. Example of tandem-assisted tracking resulting in superior temperature measurement. On the graph the green line is the temperature extracted for each frame using the tandem assisted tracking, and the red line is the temperature extracted for each frame using standard single region tracking. Note the highlighted portion of the graph where there is a spike at frame 8934 in the temperature extracted from the single region tracker. This is a result of a tracking failure from the *solitary* tracker because of an abrupt movement by the subject. The tandem assisted tracking was able to correctly track the subject through this motion and therefore extract the correct temperature at this frame. The grey scale thermograph on the right shows both the tandem assisted (green), and the solitary (red) regional trackers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

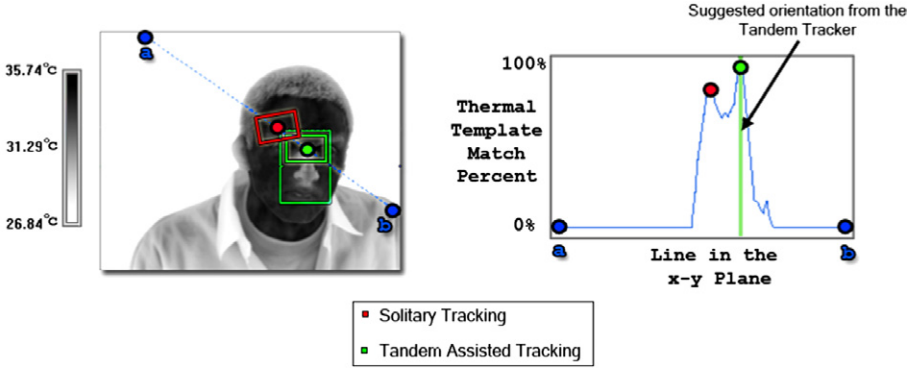

Fig. 3. Tandem facial tracking. The subject in the thermograph (shown here in gray-scale) on the left has had his periorbital region tracked by both the solitary (red) and the tandem assisted (green) trackers. The solitary periorbital region tracker has become confused because of a local maximum in the template match response space. However, the MT tracker in the periorbital region that is assisted by the TT tracker in the tandem configuration has maintained the proper track. This is because of the suggestion of the TT tracker, which is shown in the graph on the right as the vertical green line. The response of the template matching along the blue line with endpoints a and b in the thermograph (see the graph on the right) has been generated to highlight this point. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

boost the stability of another high-level tracker (the MT tracker). The benefit of the tandem relationship is enhanced by tuning the dynamics model for the TT tracker to be more tolerant to large changes in motion (by increasing the stochastic component of its dynamics model), and tuning the dynamics model for the MT tracker to allow for more localization (by decreasing the stochastic component of its dynamics model).

The TT tracker influences the MT tracker's posterior density. To describe the influence mechanism in some detail we need first to proceed with some definitions. Let us assume that $B_n = (b_1, b_2, \ldots, b_k)^{\mathrm{T}}$ is the general parameterized representation of the MT tracker at frame n, for parameters b_1, b_2, …, b_k, where k can be as large as necessary to describe the tracker for a given representation. Similarly, let us assume that $C_n = (c_1, c_2, \ldots, c_k)^{\mathrm{T}}$ represents the TT tracker at frame n. In our representation we model the tracked regions by their center of mass positions and orientations. The state space $\mathscr{S}$ for these is:

$$\mathscr{S} = \{(x, y, \theta) | x, y \in \mathbb{R}, \theta \in [0, 2\pi)\}, \tag{1}$$

where x and y are the coordinates of the center of mass of the tracked region, and θ is the angle of the tracked region with respect to the x-axis of the image coordinate system. The TT tracker uses its current state, C_n, coupled with information derived during initialization, C_0 and B_0, to derive the information about the possible state of the MT tracker B_n:

$$\hat{B}_n = C_n + (B_0 - C_0). \tag{2}$$

In Eq. (2), $\hat{B}_n$ represents the projected current state of the MT tracker based on the current state of the TT tracker. $\hat{B}_n$ is generated from the maximum-likelihood sample of the TT tracker. The method uses the maximum-likelihood sample instead of the k-highest weighted samples in order

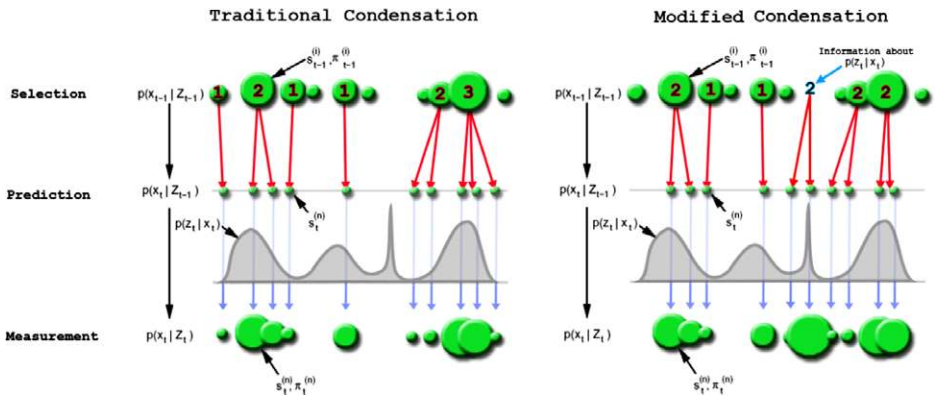

Fig. 4. Tandem modifications to the CONDENSATION framework. On the left is the traditional implementation of the CONDENSATION framework, on the right is the modified version of the CONDENSATION framework. Notice that during the selection step for the modified version, the additional information is used as the basis for producing new samples, which would have not been spawned if we used the traditional method.

to exert all the influence of the tandem tracker in one location. Thus, it explores the parameterization space around $\hat{B}_n$ to the maximum extent possible by virtue of the stochastic component of the dynamics model.

Once $\hat{B}_n$ has been calculated, it must be incorporated into the CONDENSTATION framework of the MT tracker. For this purpose, the CONDENSATION framework features a modified selection step to automatically pre-select a portion of its samples based on the knowledge of $\hat{B}_n$. The information fusion methodology is depicted in Fig. 4, where Z is an image and X is the object parameterization; $p(X_{t-1}|Z_{t-1})$ is the posterior probability from the previous time step derived from the sample set $\{(s_{t-1}^{(i)}, \pi_{t-1}^{(i)}), i = 1, \ldots, N\}$, where $s^{(i)}$ is one of N samples with corresponding weight $\pi^{(i)}$; $p(X_t|Z_{t-1})$ is the prior for time step t; $p(Z_t|X_t)$ is the observation density at time step t; $p(X_t|Z_t)$ is the posterior probability for the current time step derived from the sample set $\{(s_t^{(n)}, \pi_t^{(n)}), n = 1, \ldots, N\}$.

The tandem tracking method assigns a percent weight α to the additional information from the TT tracker. During the selection step of the CONDENSATION framework the method chooses $(1 - \alpha)\%$ of the samples normally and $\alpha\%$ based on the additional information. All of the samples selected using the additional information are set equal to its suggested state, then during the prediction step only the stochastic component of the dynamics model is applied to these samples.

3.2. Observation model

We use thermal templating to construct the observation model of tandem tracking. The fundamental idea behind template matching is to sub-sample the object of interest, the template, from the initial frame of video and then to find the region in subsequent frames that most closely resembles the template. The underlying assumption when using templates to represent the object of interest is that the appearance of the object will remain relatively constant throughout the course of the video. For facial physiology this assumption is likely to hold true only if the monitoring period is not excessively long (minutes rather than hours).

The thermal template $T(\mathbf{c})$ is extracted from a subregion R_0 of the initial video frame $V(\mathbf{c}, 0)$, which contains the object of interest. To speed up computation a subset of the pixels within the initial region of interest is used for the template. We chose to compose the set of points in our template $T(\mathbf{c})$ from the union of two subsets, $U(\mathbf{c})$ and $M(\mathbf{c})$ (see Fig. 5).

The set $U(\mathbf{c})$ is composed of points uniformly distributed over the region R_0. The set $M(\mathbf{c})$ is composed of the coldest and hottest pixels of region R_0. The motivation for composing $T(\mathbf{c})$ as the combination of $U(\mathbf{c})$ and $M(\mathbf{c})$ is to have both regular and variance spatial information encoded within the template.

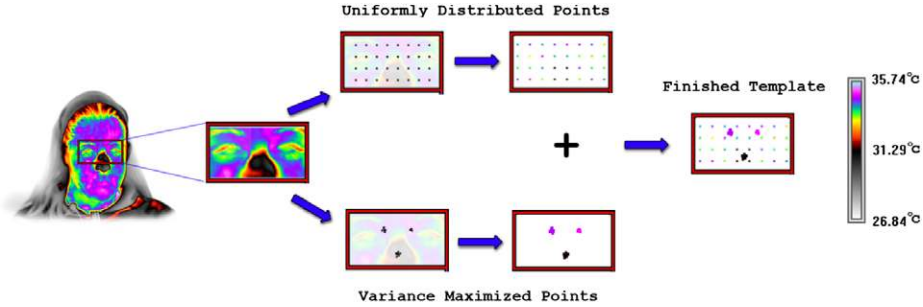

Fig. 5. Creation of the thermal template from the combination of a uniform grid of points and a variance maximized set of points.

4. Blood flow computation

We compute superficial blood flow on the face by using a pair of bioheat models. One model computes blood perfusion in homogeneous tissue regions that are void of large superficial vessels (*perfusion model*). The other model computes directional blood flow in a major superficial vessel, like the external carotid (*directional model*).

We reported the perfusion model in a number of papers [2,7,40] and tested it extensively in human experiments with excellent results. We routinely apply this model in monitoring instantaneous stress, a quintessential operation in the physiological HCI framework.

We reported the directional model briefly in the 2004 CVPR Proceedings [3]. Since the model measures directional blood flow in a major vessel, it is related to the heart output and can yield the cardiac pulse waveform. The role of this waveform in monitoring heartbeat irregularities cannot be underestimated. The scant information about the directional model in combination with its importance in physiological HCI warrant its more detailed examination in this article.

In general, bioheat modeling on the surface of the body starts with the construction of the energy balance equation on a control volume that extends several millimeters in depth. The dynamic form of this equation is a PDE with boundary conditions that are determined by environmental and other factors, including the sensor output (skin temperature). A major term in the energy balance equation is convective heat produced by blood flow (perfused or directional). The solution of the PDE form of the equation yields the evolution of the blood flow variable.

Specifically, the directional blood flow model assumes that the vessel acts as a heat source for the surrounding four-layer control volume tissue. These layers are successively, in positive z direction, the skin, the fat, the muscle, and the core (see Fig. 6(a)). We assume that each layer is isotropic with respect to thermal conductivity $K(z)$, metabolic heat rate $q^{\mathrm{M}}(z)$, density ρ, and specific heat c of the tissue. The heat effect of the vessel on the skin temperature depends on the vessel's depth and shape as well as the blood temperature and flow rate.

We consider a single large vessel running along a direction x parallel to the skin layer (see Fig. 6(a)). The heat conduction in the tissue surrounding the vessel is dominant in directions parallel (x) and perpendicular (z) to the skin. In the remaining y direction we can neglect heat transfer because of the presence of other vessels, periodically arranged and similar to that considered. Therefore, we introduce the following 2D PDE model:

$$\rho c\frac{\partial \Theta}{\partial t}-\frac{\partial}{\partial x}\left(K(z)\frac{\partial \Theta}{\partial x}\right)-\frac{\partial}{\partial z}\left(K(z)\frac{\partial \Theta}{\partial z}\right) = q^{\mathrm{BL}}(x,t)+q^{\mathrm{M}}(x,z),(x,z)\in(0,L)\times(0,D), \tag{3}$$

where q^{M} is the volumetric metabolic heat while q^{BL} is the heat due to blood flow speed u_{bl} in a vessel assimilated to a line source $z = S(x)$. $\Theta(x,z,t)$ is the temperature distribution function in the control volume tissue over time, $K(z)$ is the thermal conductivity of a particular layer within the volume, while ρ and c are the tissue density and specific heat, respectively.

The heat produced by the terms on the right hand side of Eq. (3) conducts through the control volume tissue and fluctuates over time as expressed by the spatial and time derivatives respectively of the left hand side. The structure of the PDE model clearly reveals its energy balance origin.

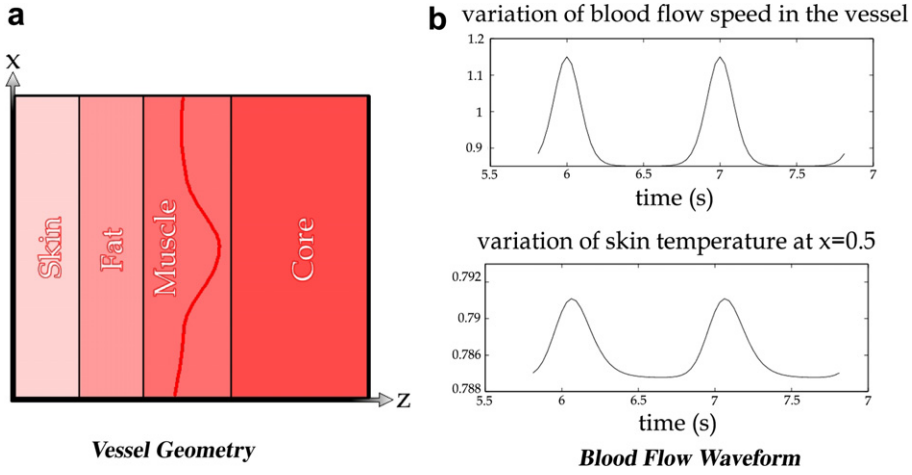

Fig. 6. (a) Four-layer tissue structure hypothesized by our model along with the coordinate system convention. The red curve represents the assumed position and shape of the vessel. (b) The graph at the top shows the time variation of the blood flow speed in a major superficial vessel, as it is computed by the PDE model. The graph at the bottom shows the corresponding input to the model, that is, the time variation of the skin temperature. Such measurements were taken from the carotid or temporal areas of the subjects at a distance of 10 ft and were validated with a traditional contact measurement device (LaserFlo BPM^2, Vasamedics [42]). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

We impose the following boundary conditions:

$$\Theta(x, D, t) = \Theta_{\text{core}}, x \in (0, L), \tag{4}$$

$$\Theta(x, 0, t) = \Theta_{\text{skin}}(x, t), x \in (0, L), \tag{5}$$

$$\frac{\partial \Theta}{\partial z}(x, 0, t) = \lambda(\Theta(x, 0, t) - \Theta_{\text{air}}) + q_{\text{ir}}, x \in (0, L), \tag{6}$$

$$\frac{\partial \Theta}{\partial x}(0/L, z, t) = 0, z \in (0, D). \tag{7}$$

λ is the convection heat transfer coefficient, which depends on air flow. According to [41]: $\lambda = 2.7 + 7.4(v_{\text{air}})^{0.67}(W/m^2K)$, where v_{air} is the air speed in (m/s). q_{ir} is the radiation heat flux: $q_{\text{ir}} = \sigma\epsilon(\Theta^4_{\text{skin}} - \Theta^4_{\text{wall}})$, where σ is the Stefan-Boltzmann constant and ϵ is the skin emissivity. Θ_{wall} is approximated by the temperature of the air.

The heat source term associated with blood flow is assumed to have the decomposition $q^{\text{BL}} = u_{\text{bl}}(t)r(x, z)$, where u_{bl} is the unknown blood flow speed in the vessel. We assume that the vessel is centered on the curve $z = S(x)$. Then, we take for $r(x, z)$ the modified bell function:

$$r(x, z) = \mu \exp\left(-\frac{(z - S(x))^2}{\pi v^2_{\text{app}}}\right).$$

v_{app} is the apparent radius of the vessel seen as a heat source. μ is defined as follows:

$$\mu = \rho_{\text{bl}} c_{\text{bl}} \frac{A}{V}(\Theta_{vessel}(x, z, t) - \Theta(x, z, t)) \ (\text{J/m}^4), \tag{8}$$

where ρ_{bl} and c_{bl} are the density and the specific heat of blood, respectively, A is the vessel cross section, and V is the control volume of tissue. We assume that the temperature of the blood in the vessel is the same as the core temperature $\Theta_{\text{vessel}} = \Theta_{\text{core}}$.

The mathematical problem is to retrieve the blood flow speed $u_{\text{bl}}(t)$ from the skin temperature Θ_{skin} obtained with a thermal infrared camera. We can make educated assumptions about all the other parameters involved in Eqs. (3)–(7) based on physiological facts [3]. In extensive direct simulation and limited application of the inverse solution on real data (two subjects) the model behaved impeccably [3] (see Fig. 6(b)). The inverse solution correlated very well with ground-truth data collected through a contact blood flow meter from Vasamedics [42].

By observing the normalized waveforms in Fig. 6(b) one may understand why it could be beneficial from the image analysis point of view to operate on the blood flow domain versus the thermal domain. The blood flow operational range is at least an order of magnitude larger than the thermal operational range. In addition, the patho-physiological meaning of normalized unit measurements of superficial blood flow have been extensively studied and understood in the context of oximetry. Therefore, by transforming thermal to blood flow signals, one may capitalize upon this rich body of knowledge. Still one may choose to use the direct thermal measurements and his conclusions would be of some relevance, as heat dissipation at the surface of the body is strongly correlated to blood flow and perfusion.

5. Cardiac pulse computation

In addition to the amplitude of the blood flow waveform as computed by Eq. (3) we can compute the frequency of the blood flow pulsation (cardiac pulse). Both amplitude and frequency are important in potential applications of this technology, like desktop monitoring of heartbeat irregularities. Our method is based on the assumption that temperature modulation due to pulsating blood flow produces the strongest thermal signal on a superficial vessel. This signal is affected by physiological and environmental thermal phenomena. Therefore, the resulting thermal signal that is being sensed by the infrared camera is a composite signal, with the pulse being only one of its components. Our effort is directed into recovering the frequency of the component signal with the highest energy content. This is consistent with our hypothesis of pulse dominance in the thermal field of a superficial vessel.

We reported briefly a contact-free pulse measurement methodology in the Proceedings of CVPR 2005 [4]. In the present article, we elaborate further, as cardiac pulse is a vital sign and quintessential to the proposed desktop health monitoring system. A prerequisite to accurate pulse measurement is proper localization and motion tracking. We select the pulse taking location in the first frame of the thermal video on a major facial superficial vessel (usually carotid or temporal). The selection is performed by a mouse point and drag operation along the length of the vessel's thermal imprint. This results in a measurement line, which is the virtual equivalent of taking the pulse by placing the index finger on the vessel.

To maintain consistently the pulse measurement line over the course of time, one needs to track the facial tissue as it moves due to natural head motion. We use our tandem tracking methodology described in Section 3 to this effect. The MT tracker is focused on the vessel's region of interest grown symmetrically around the selected measurement line. Based on the outcome of Fourier analysis an estimation function computes the cardiac pulse. Fig. 7 illustrates the general steps of our methodology.

Considering that the blood vessel is a long, narrow structure, the pulse propagation phenomenon causes slight phase shift on the temperature profiles along the blood vessel. This may weaken the signal if we use conventional signal recovery methods in the time domain. Each pixel along the blood vessel has a unique periodical temperature profile, which is shifted with respect to the others. As Fig. 8 shows, averaging these temperature profiles may weaken the signal. Although, the temperature profiles of the pixels along the blood vessel are shifted in the time domain, their frequency should remain the same (unshifted). Therefore, by operating on the frequency domain and combining appropriately the power spectra of these temperature profiles we can reinforce the signal instead of weakening it. We

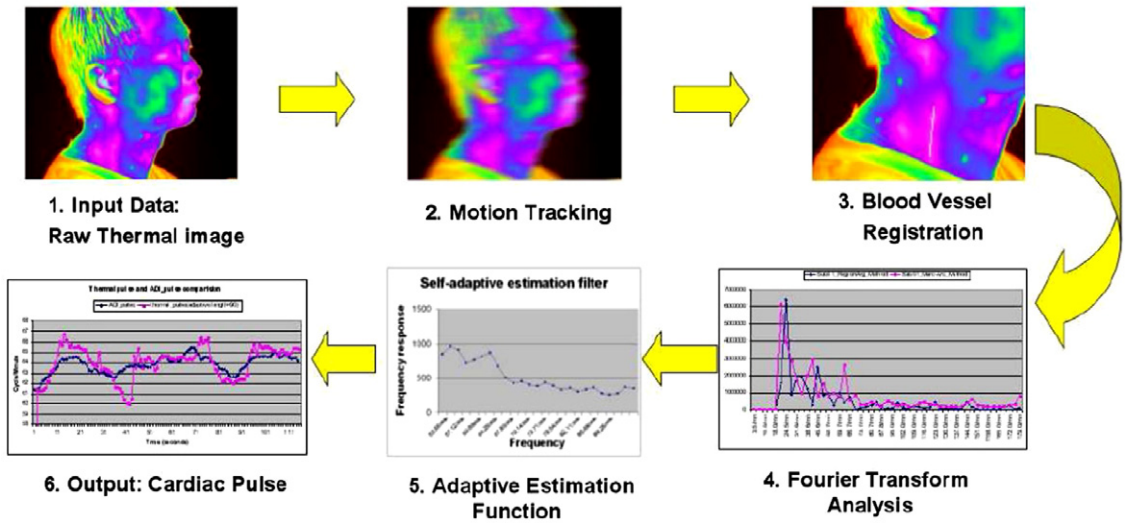

Fig. 7. Pulse measurement methodology. The white line along the carotid vessel is the measurement line selected by the operator.

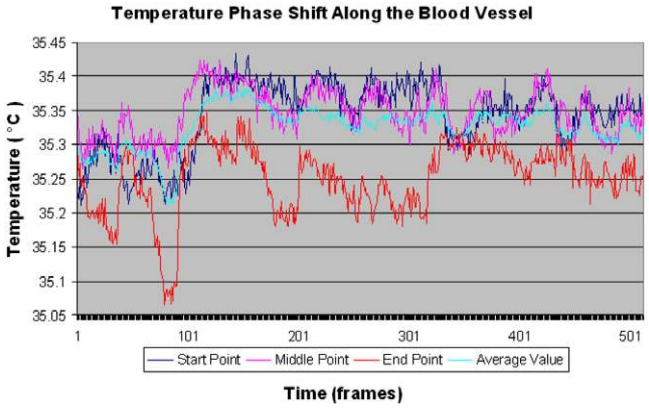

Fig. 8. Temperature profiles of three different pixels along the exposed blood vessel compared to the average temperature profile.

apply Fourier analysis in a novel manner to capitalize upon the pulse propagation effect and extract the dominant pulse frequency:

First step. We name L the straight segment selected by the operator along the center line of a large superficial blood vessel (e.g., carotid). The algorithm expands symmetrically L into an elongated rectangle R. The width of this rectangle depends on the width of the vessel on the thermal imagery. For a subject imaged at 6 ft with a 50-mm lens the rectangle's width is 3–7 pixels. By convention, we place the x axis of our coordinate system along the width and the y axis along the length of the vessel (see Fig. 9).

Second step. We record the time evolution of the pixel matrix delineated by rectangle R for N frames ($N = 256$ or 512). Thus, we produce a 3D matrix $A(x,y,t)$, where $0 \leqslant x \leqslant R_x, 0 \leqslant y \leqslant R_y$ is the spatial extent of rectangle R and $0 \leqslant t \leqslant N-1$ is the timeline.

Third step. We average the pixel temperatures along the x dimension. Thus, we derive a 2D matrix:

$$A'(y,t) = \frac{1}{R_x}\sum_{x=0}^{R_x} A(x,y,t) \tag{9}$$

where $0 \leqslant y \leqslant R_y$, $0 \leqslant t \leqslant N-1$. This reduces the noise and "shrinks" the rectangular vessel region R into an *effective line*, upon which the signal measurement will be performed.

Fourth step. For each *effective* pixel on the measurement line we obtain the time evolution signal of its temperature:

$$\forall y : S_y(t) = A'(y,t), 0 \leqslant t \leqslant N-1. \tag{10}$$

We apply the fast fourier transform (FFT) on each of these signals to obtain the respective power spectra:

$$\forall y : \mathscr{P}_y = \mathscr{F}(S_y(t)). \tag{11}$$

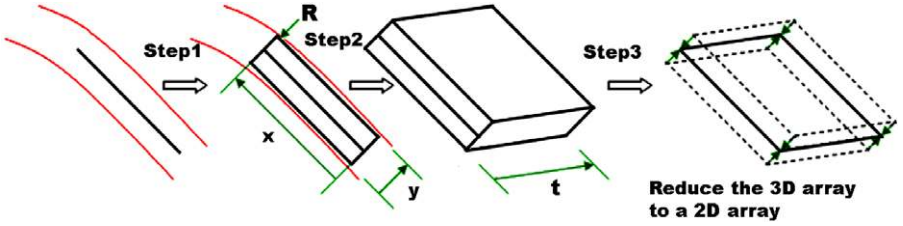

Fig. 9. Schematic diagram of the first three steps in the Fourier analysis of the vessel temperature signal.

Specifically, we apply a classical decimation-in-time (Cooley and Tukey) 1D base-2 FFT method given in [43].

Fifth step. We average all the power spectra computed in the previous step into a composite power spectrum:

$$\bar{\mathscr{P}} = \frac{1}{R_y} \sum_{y=0}^{R_y} \mathscr{P}_y. \tag{12}$$

We still refrain, however, from finding the dominant frequency in power spectrum $\bar{\mathscr{P}}$ and declare it as the pulse frequency.

5.1. Adaptive estimation function

A fundamental question is what we report as the effective pulse along the timeline. The instantaneous computation described in Section 5 is not to be trusted literally since it may be affected occasionally by thermoregulatory vasodilation [44] and creeping noise. To address this problem we use an estimation function that takes into account the current measurement as well as a series of past measurements.

The current power spectrum $\bar{\mathscr{P}}_0$ of the temperature signal is being computed over the previous N frames ($N = 256$ or 512) by applying the process outlined in Section 5. We convolve the current power spectrum with a weighted average of the power spectra computed during the previous M time steps. We chose $M = 60$, since at the average speed of 30 fps sustained by our system, there is at least one full pulse cycle contained within 60 frames even in extreme physiological scenarios. Therefore, the historical contribution to our estimation function remains meaningful at all times.

Specifically, the historical frequency response at a particular frequency is given as the summation of all the corresponding frequency responses for the M spectra, normalized over the total sum of all the frequency responses for all the historical M spectra:

$$H(f) = \sum_{i=1}^{M} \bar{\mathscr{P}}_i(f) \Big/ \sum_{i=1}^{M} \sum_{j=1}^{F} \bar{\mathscr{P}}_i(j). \tag{13}$$

Finally, we convolve the historical power spectrum $\bar{H}$ with the current power spectrum to filter out transient features. We then designate as pulse the frequency f_{pulse} that corresponds to the highest energy value of the filtered spectrum within the operational frequency band.

5.2. Experimental results of pulse measurements

We have used a high quality thermal imaging system for data collection [1], to rigorously quantify the performance of the pulse measurement method. We recorded 25 thermal clips from five subjects while resting in an armchair. Concomitantly we recorded ground-truth pulse signals with PowerLab/4SP from AD Instruments featuring an MLT 1010 piezoelectric pulse transducer [45]. The data set features subjects of both genders, different ages, and with varying physical characteristics.

Although, the emphasis of our effort is on facial measurements, we attempted a few measurements on the wrist (radial artery), because it was easy to alter the pulsative

Table 1
Ground truth (GT) and thermal imaging (TI) pulse measurements

Subject No.	Video file	Time length (s)	Vessel	Status of pressure on forearm	GT (bpm)	TI (bpm)	% accuracy
Subject 01	D005-001	132.3	Carotid	N/A	63.7	63.1	98.28
Subject 01	D005-002	121.4	Carotid	N/A	60.3	60.8	99.22
Subject 01	D005-003	123.3	Radial	No pressure	62.1	62.9	98.68
Subject 01	D005-011	120.7	Radial	No pressure	67.9	68.0	99.88
Subject 01	D005-012	210.4	Radial	Pressure	61.3	61.7	99.43
Subject 01	D005-013	240.5	Radial	After pressure	62.7	62.9	99.66
Subject 02	D005-016	120.7	Carotid	N/A	82.7	80.5	98.01
Subject 02	D005-017	120.7	Carotid	N/A	73.3	73.8	99.37
Subject 02	D005-018	120.6	Radial	No pressure	75.7	75.7	99.92
Subject 02	D005-019	180.7	Radial	Pressure	74.8	74.6	99.76
Subject 02	D005-020	180.6	Radial	After pressure	78.5	78.2	99.72
Subject 02	D005-040	120.2	Carotid	N/A	68.0	68.6	99.11
Subject 03	D005-041	122.5	Carotid	N/A	65.6	66.5	98.68
Subject 03	D005-042	120.7	Radial	No pressure	63.8	64.3	99.32
Subject 03	D005-044	123.7	Radial	After pressure	68.2	68.3	99.34
Subject 03	D005-046	122.1	Temporal	N/A	67.4	67.8	99.34
Subject 03	D005-060	125.1	Carotid	N/A	67.3	67.2	99.76
Subject 04	D005-062	121.0	Radial	No pressure	65.2	64.4	98.72
Subject 04	D005-063	180.8	Radial	Pressure	63.7	64.7	98.54
Subject 04	D005-064	181.3	Radial	After pressure	72.1	71.1	98.66
Subject 04	D005-066	183.4	Temporal	N/A	66.8	67.3	99.27
Subject 04	D005-079	120.9	Carotid	N/A	76.2	76.1	99.83
Subject 05	D005-080	117.1	Carotid	N/A	72.6	71.9	99.03
Subject 05	D005-081	120.8	Radial	No pressure	70.7	71.7	98.67
Subject 05	D005-085	188.6	Radial	After pressure	73.2	73.5	99.59

blood flow and test the robustness of the method under various physiological conditions. Specifically, we have taken wrist measurements before, during, and after the application of a forearm cuff. We have found no significant differences in the accuracy of the thermal imaging measurements among these scenarios.

Because our thermal imaging system and the PowerLab/4SP data acquisition system have different frequency of sampling and perform measurements on a vastly different theoretical basis, we normalized the experimental data from the two modalities in order to compare them.

The PowerLab/4SP data acquisition system (ground truth) collects 100 samples per second, while our thermal imaging system acquires 30 frames per second. We averaged the ground truth output data every ten samples while the thermal imaging data every three samples (frames). Based on this normalization, we compared the average cardiovascular pulse rate computed by our imaging method to that reported by the ground-truth instrument for all the subjects in our data set.

Table 1 shows the detailed profile of our comparative experiment and the average pulse measurements reported by the two modalities. The overall agreement between the two measurement methods is 98%. To quantify the linear correlation between the two measurement modalities, we have used the high Pearson product moment measure P_c [46]. By applying the Pearson formula on the data of Table 1, we find that $P_c = 0.994$, which indicates a strong degree of correlation.

6. Breath rate computation

Human breathing consists of expiration and inspiration phases. In expiration, air that was heated through its contact with the lungs flows via the nostrils to the environment. Conversely, in inspiration environmental air flows via the nostrils to the lungs. This creates a periodic thermal signal in the vicinity of the nostrils that oscillates between high (expiration) and low (inspiration) values [47]. In traditional pulmonary studies a thermistor is attached near the nostrils to capture this phenomenon and produce a representative breath signal. The thermal imager can be viewed as a virtual thermistor, since it captures the same phenomenon, but at a distance [48]. Fig. 10 shows a thermal snapshot of a subject during the expiration phase. One can observe the contrast between the hot expired air, next to the nasal area and the lower intensity surrounding background. During the inspiration phase, the hot expired air is absent.

We reported briefly a contact-free breath rate measurement methodology in the Proceedings of EMBC 2005 [5]. In the present article, we elaborate further, as breath rate is a vital sign and quintessential to the proposed desktop health monitoring system. As in the case of pulse, a prerequisite to accurate breath measurement is proper localization and motion tracking. We select the breath Measurement Region Of Interest (MROI) in the first frame

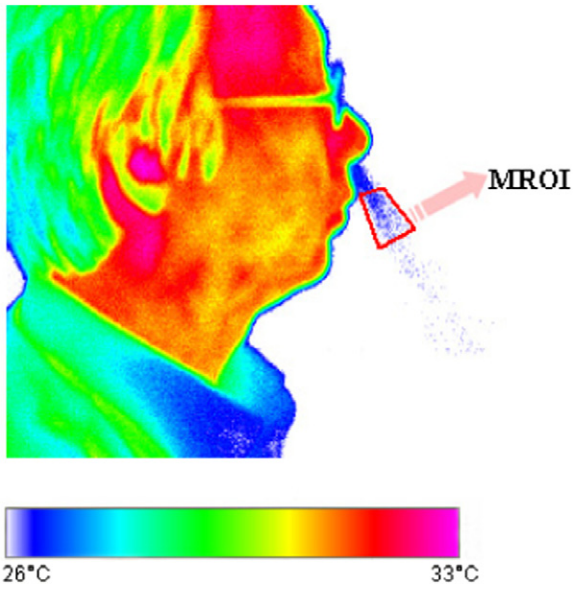

Fig. 10. Thermal snapshot of expiration. The Measurement Region Of Interest (MROI) is depicted as a polygon next to the nasal-mandibular region.

of the thermal video near the nasal-mandible area. We compute the mean temperature within the MROI in each frame. Along the timeline, this produces a periodic thermal signal, which is indicative of the breathing function. As we explained previously, this imaging type of sensing is the equivalent of a virtual thermistor. To maintain the virtual thermistor in place over time, one needs to track the facial tissue as it moves due to natural head motion. We use our tandem tracking methodology described in Section 3 to this effect.

As a periodic signal, the breath signal can be analyzed through Fourier transformation. Since we operate on the discrete domain we use FFT. We perform FFT analysis on sliding segments (windows) of the normalized breath thermal signal. We adjust the size of the window as the timeline evolves. Initially, the window size is small, but then it expands as time permits. The goal is to start reporting breath rate as soon as possible, while incrementally improving the computational accuracy over time.

One would think that the breath signal analysis may be very similar to the pulse signal analysis as both signals are periodic thermal undulations. However, breath signal analysis can be simpler and faster, as breath is a sharp on (expiration) and off (inspiration) phenomenon in contrast to the gradual thermal fluctuation produced by pulsation.

Indeed, in breath signal analysis the only pre-processing step we apply is down-sampling of the raw stream to 10 fps, as a way to reduce high frequency noise and fix a constant rate for FFT application. Then, we produce the FFT power spectrum for each sliding time window. From the resulting power spectra we remove responses corresponding to frequencies outside the range 5–40 cpm (cycles per minute). We consider frequencies outside this range unlikely to occur in our experimental scenarios, where individ-

uals are either at rest or undergoing mild aerobic exercise. We select the dominant frequency in the power spectrum of each sliding window as the likely breath rate at the time. Fig. 11 illustrates the major steps of our methodology.

We define as $\mathbf{V}_i[t]$, $t \in [0, \cdots, N]$, the down-sampled breathing thermal signal framed in sliding window i. We normalize the signal $\mathbf{V}_i[t]$ as follows:

$$\mathbf{V}'_i[t] = \frac{\mathbf{V}_i[t] - \mu_i}{\sigma_i}, \tag{14}$$

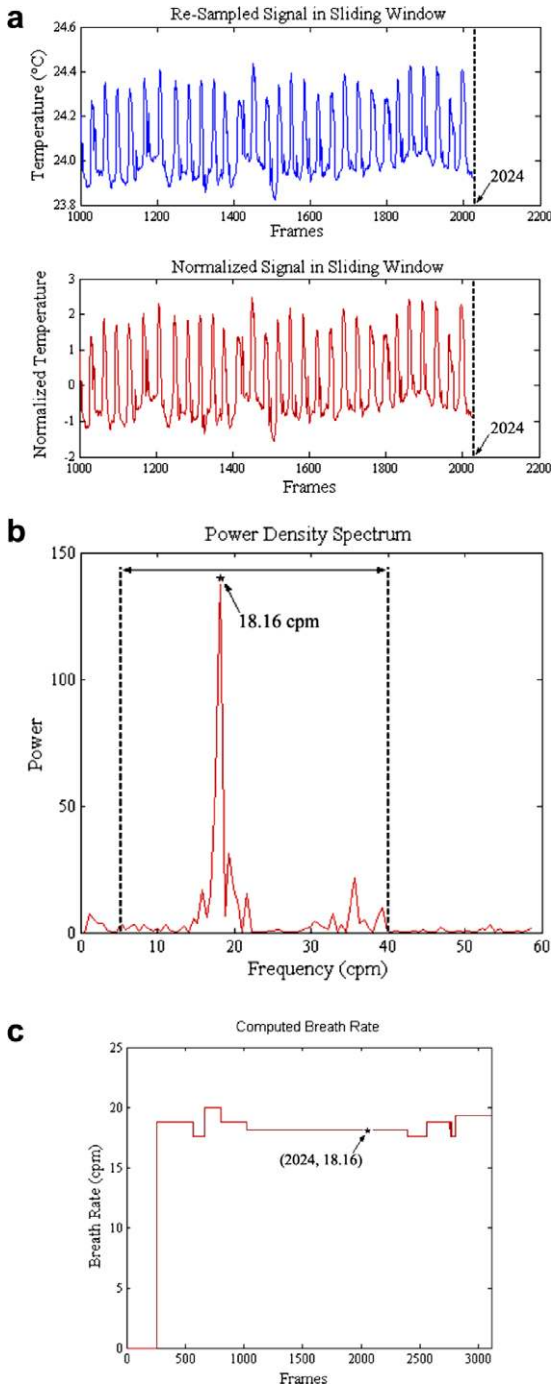

Fig. 11. (a) Normalization of windowed signal. (b) Power density spectrum. (c) Breath rate signal.

where μ_i and σ_i are the mean and standard deviation of $\mathbf{V}_i[t]$, respectively. The normalization transforms signal $\mathbf{V}_i[t]$ to $\mathbf{V}'_i[t]$ that features mean $\mu'_i = 0$ and standard deviation $\sigma'_i = 1$ (see Fig. 11(a)).

We apply FFT on the autocorrelation sequence $\phi_{\mathbf{V}'_i\mathbf{V}'_i}[t]$ of signal $\mathbf{V}'_i[t]$, to compute the power density spectrum:

$$\Phi_{\mathbf{V}'_i\mathbf{V}'_i}(e^{j\omega}) = \sum_{t=-\infty}^{\infty} \phi_{\mathbf{V}'_i\mathbf{V}'_i}[t] \mathrm{e}^{-j\omega t}. \tag{15}$$

In Fig. 11(b), we present the power spectrum density of signal $\mathbf{V}'_i[t]$ that corresponds to the window ending in frame $t = 2024$. The cut-off lines exclude the portion of the power spectrum that corresponds to uninteresting frequencies (outside the 5–40 cpm range). The dominant frequency of 18.16 *cpm* is indicated in the diagram.

Sliding window $i = 1$ applies to the normalized signal $\mathbf{V}'_1[t]$, $t \in [0, \cdots, N_1]$. We set $N_1 = 2^8 = 256$ samples, which means that breathing rate is reported for the first time after 25.6 s, since the down-sampled frame rate is 10 fps. After that, breathing rate is reported every 0.1 s, as the window slides one sample to the right with every incoming frame. The window maintains the size N_1, until the total number of processed frames becomes $N_2 = 2 * N_1 = 2^9 = 512$, which is the next power of two. Then, it automatically adjusts to the larger power size. The size of the window adjusts one final time when the total number of processed frames reaches $N_3 = 2 * N_2 = 2^{10} = 1024$ and retains this value for the remaining monitoring period. Therefore, the method achieves top accuracy after 102.4 s of operation.

6.1. Experimental results of breath rate measurements

As in the case of pulse measurements, we used a high quality thermal imaging system and a ground truth apparatus to rigorously quantify the performance of the breathing measurement method. The ground-truth apparatus consisted of a PowerLab/4SP from AD Instruments [45] featuring a respiratory belt transducer. The transducer was fitted around the subject's chest wall. By measuring the up and down movement of the thoracic cavity, the sensor formed the ground-truth breath signal and sent it to the computer through the PowerLab/4SP unit. Equivalently, we could have used a thermistor as a ground-truth sensor, but this may had interfered with the imaging measurement, as it had to be attached in the subject's nostrils.

We recorded 19 thermal clips of 9 subjects. Seventeen of the clips were recorded while the subjects were at rest. The remaining 2 clips were recorded after the subjects undergone 2 min of moderate aerobic exercise. For many subjects we recorded more than one clip at different times. All the thermal clips are ~5 min in length.

Table 2 shows the detailed experimental results for all the thermal clips and all three windowing stages of the FFT-based computation. The imaging results are juxtaposed with the corresponding ground-truth measurements obtained through the respiratory belt transducer.

Table 2
Comparison of ground truth (GT) and thermal imaging (TI) breath rate measurements

Video file	Time length (s)	Stage 1			Stage 2			Stage 3		
		GT (cpm)	TI (cpm)	% accuracy	GT (cpm)	TI (cpm)	% accuracy	GT (cpm)	TI (cpm)	% accuracy
D005-007	302.60	19.67	18.75	95.32	15.73	17.71	87.41	16.50	17.28	95.27
D005-008	302.60	7.96	11.30	85.95	6.74	7.64	95.91	4.96	5.39	96.91
D005-010	300.80	13.89	16.53	86.98	14.73	15.23	97.92	11.69	13.27	90.23
D005-012	304.00	13.57	15.08	84.19	15.96	16.34	92.34	14.12	14.81	97.78
D005-016	300.80	37.16	37.50	98.58	32.03	34.65	88.80	27.10	28.97	99.57
D005-017	302.20	14.17	11.93	58.04	10.83	10.00	86.65	9.02	8.82	91.33
D005-019	315.60	26.68	27.06	36.13	21.79	24.23	91.40	21.11	21.20	97.07
D005-021	306.70	20.43	22.03	22.68	20.77	20.75	71.12	24.28	24.81	93.59
D005-022	310.40	16.44	18.75	80.99	18.11	18.85	96.61	17.78	18.33	86.48
D005-023	305.40	4.76	7.80	88.28	9.42	8.61	94.96	7.52	7.74	87.43
D005-024	305.90	10.86	12.45	82.81	10.93	11.81	86.45	11.92	12.03	90.98
D005-026	307.00	14.93	16.40	99.09	15.54	16.40	91.82	15.99	16.37	93.10
D005-027	305.20	19.83	21.09	88.87	20.38	21.00	97.62	19.91	21.43	95.11
D005-028	306.32	16.59	18.75	71.13	17.33	17.69	83.88	14.63	16.06	98.24
D005-029	305.40	16.90	18.88	92.17	16.87	17.72	99.90	13.60	15.31	97.82
D005-033	360.80	19.91	23.43	85.36	21.34	22.65	91.95	21.04	20.05	99.08
D005-037	401.00	4.41	7.82	90.15	4.64	5.98	94.47	6.40	5.99	97.62
D005-038	306.20	16.00	18.75	93.65	13.95	15.84	96.96	12.20	13.30	92.37
D005-039	325.60	10.91	14.06	82.32	10.30	11.96	93.86	10.79	10.98	95.29

We can observe that the imaged breath rate in stage 3 is closer to ground-truth than that in stages 2 and 1. In turn, the imaged breath rate in stage 2 is more accurate than that in stage 1. The breath rates corresponding to clips D005-016 and D005-019 are higher, since these clips were recorded after the subjects undergone moderate aerobic exercise.

We use the Pearson product measurement R [46] to evaluate the linear correlation between the imaged breath rate (Y) and the ground-truth breath rate (X). The higher the value of R, the stronger the correlation of X and Y is.

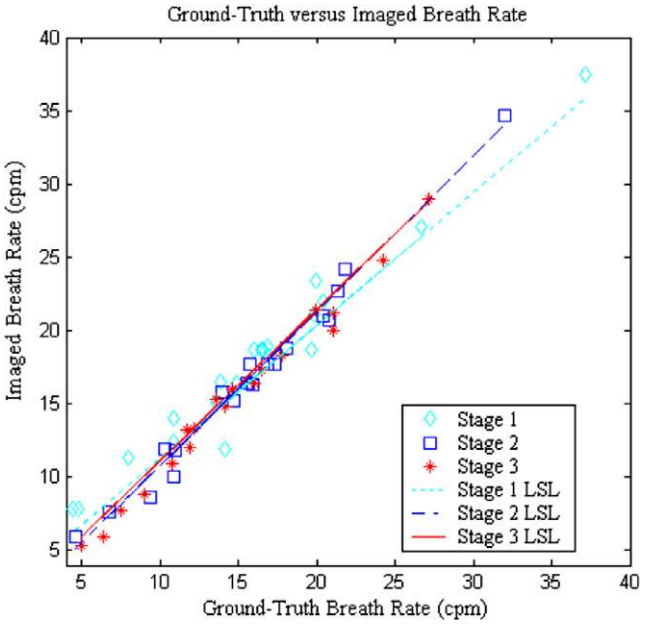

Fig. 12. Linear correlation of ground-truth breath rate and imaged breath rate. Stage 1: $R = 0.9810$, $p = 1.52\text{E}{-}013$; stage 2: $R = 0.9895$, $p = 1.52\text{E}{-}016$; stage 3: $R = 0.9906$, $p = 3.68\text{E}{-}017$.

We also compute the p-value to test the hypothesis of no correlation. If the p-value is small, normally less than 0.05, then the correlation R is significant. Fig. 12 illustrates the linear correlation of ground-truth and imaged breath rate, based on the results listed in Table 2.

The breath rates from the two modalities are highly correlated in stage 3, with $R = 0.9906$ and $p = 3.68E - 017$. Even in stages 1 and 2 the correlation is strong, but not as strong as in stage 3. Thus, by applying multi-stage FFT analysis on the breath temperature signal, we obtain greater accuracy as the window size increases. Furthermore, the methodology appears to perform well either the subject is at rest or in an elevated metabolic state.

7. Applications—desktop stress monitoring

The technology of computing vital signs at a distance in a highly automated manner can be applied to monitoring a variety of chronic or transient health conditions including stress, heartbeat irregularities, respiratory problems, and others. This monitoring can typically take place at the desktop or at home. So far, we have used the contact-free blood flow measurement (see Section 4) as the basis of a stress monitoring method. The importance of detecting elevated stress levels in HCI cannot be underestimated, since this emotion is often provoked by computer usage itself.

We have found that during mental stress, there is an increase in blood flow to the forehead region of a subject. The increase in blood flow is centered on the frontal vasculature of the forehead at and just above the corrugator or "frowning muscle." For each subject we select a Region of Interest (ROI) in the forehead that includes the frontal vessels (see Fig. 13(a)). Our tracking algorithm (see Section 3) registers this ROI throughout the course of experiments.

Tracking allows meaningful application of physiological computations despite subject motion. The computation is performed on the 10% of the hottest pixels in the selected ROI. We have found experimentally that this typically coincides with the frontal vessels in the forehead ROI (see Fig. 13(b)). We compute the mean temperature of the 10% hottest pixels on the ROI for every frame in the thermal clip. We thus produce a forehead temperature signal (see Fig. 14). We then use the bioheat model we described in Section 4 (Eq. 3) to compute the blood flow on the frontal vessels based on the dynamic thermal input from the ROI.

7.1. Experimental design for stress monitoring

To evaluate our stress monitoring methodology, we induced stress in 12 subjects and we measured accordingly. For this study, we utilized the computerized version of a psychological tool called the Stroop Color Word Conflict Test [49,50]. The Stroop test is a well-established method for inducing stress. During the Stroop, the user is presented with the names of colors that are displayed in a discordant coloration. For example, if a word is spelled as "BLUE," but is colored green, then the subject is expected to report green as an answer (see Fig. 15). As the Stroop test progresses, the time allotted to report a color for an answer is reduced. The diminished time acts to artificially induce stress by decreasing the required response time and increasing the error rate of the subject in reporting the colors.

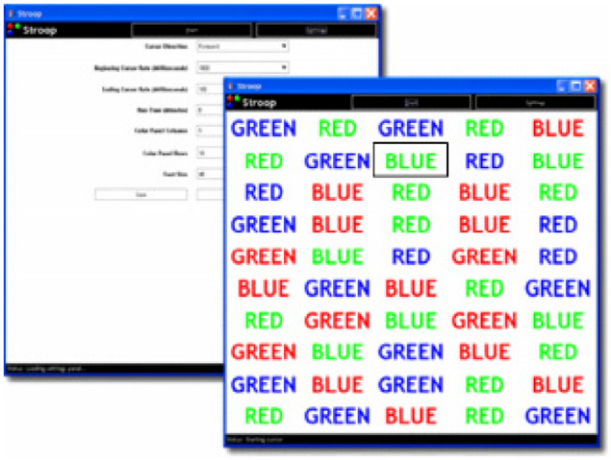

Fig. 15. Screenshot of the Stroop test application. The boxed word indicates to the subject which color to report, irrespectively of how the word is spelled [50]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Every subject underwent two sequential test sessions. In the first session we outfitted the subject with a metabolic rate measurement device to gauge Energy Expenditure (EE). In the second session we recorded the subject's face with our thermal imaging system. Every session included a baseline part (10 min) where the subject was at rest and a supervised Stroop testing part (10 min). We have included the EE measurement session as a validation standard for our novel thermal imaging approach to stress

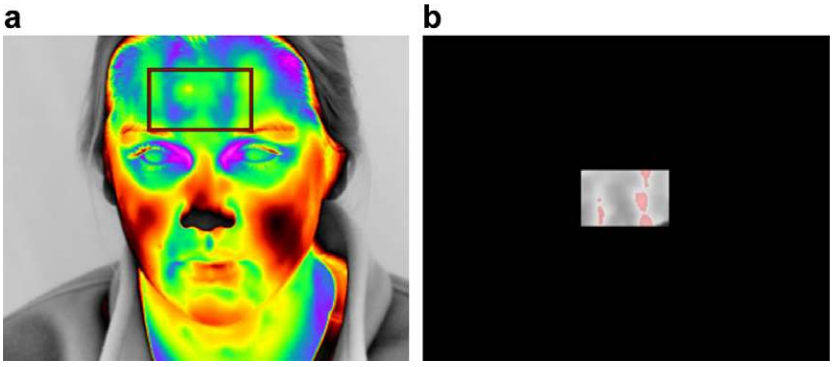

Fig. 13. (a) Region of interest (ROI) on the subject's forehead. (b) The frontal vessels (10% hottest pixels in ROI) are marked in pink. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

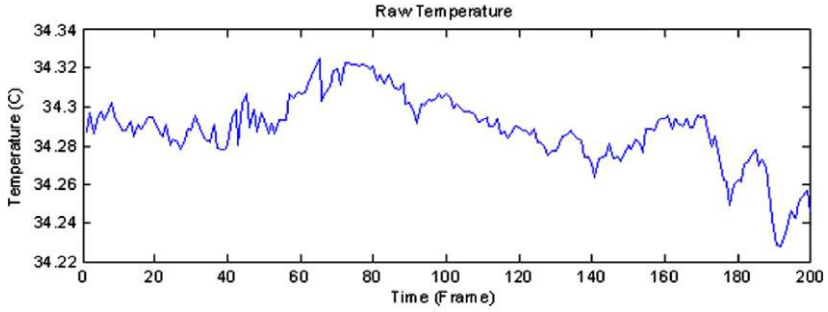

Fig. 14. Forehead temperature signal.

quantification. It has been documented in the literature that EE is a reliable stress indicator [51]. Ideally, we should have carried out thermal image recording of the face and EE measurement simultaneously on every subject. However, this was impractical due to the gas masks worn by the subjects in the EE measurements (see Fig. 16) and can cause spurious noise to be interjected into the measurement collection process.

We quantified EE by analyzing the subject's respiratory activity using a cardiopulmonary stress test device, the V_{max} Spectra from SensorMedics. The measurement is based on the premise that the volume of oxygen consumed by a subject is proportional to the current EE of the individual. We express the EE, or metabolic rate measurement, in Calories. For the facial thermal image recording, we used the same thermal imaging system as in all other experiments.

7.2. Experimental results of stress monitoring

The experimental subject set included 12 individuals of varying ethnicity and distributed between 7 males and 5

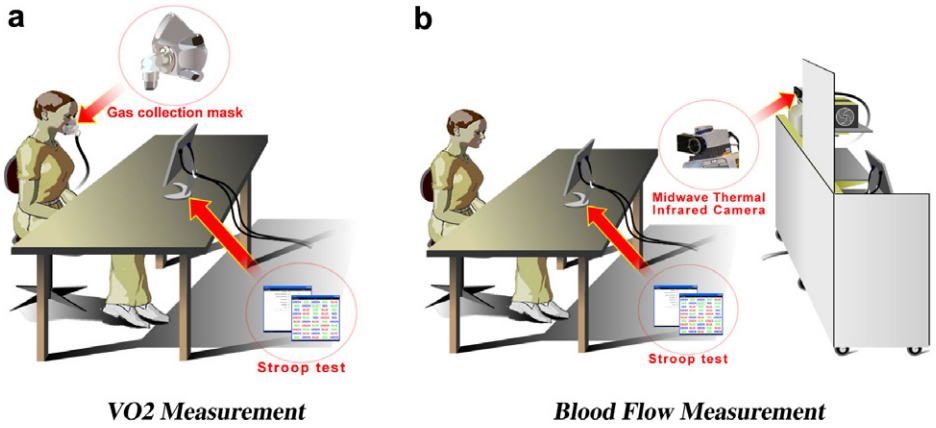

Fig. 16. (a) Monitoring the energy expenditure of a subject during Stroop testing. (b) Recording thermal facial information of a subject during Stroop testing.

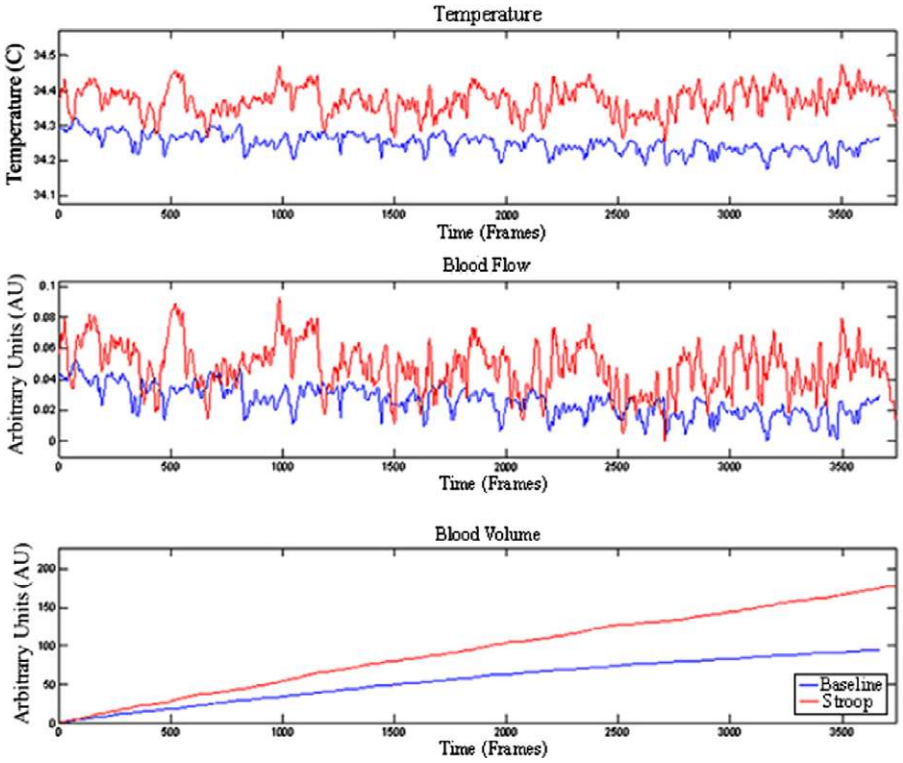

Fig. 17. (Top) Original forehead temperature signals. (Middle) Blood flow signals. (Bottom) Blood volume signals. Red denotes signals for the Stroop session while blue for the baseline session. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

females. The room temperature at the time of the experiments was measured between 19.4–22.2 °C.

For each subject, we produced six curves, three curves for the baseline and three curves for the Stroop measurements: temperature, blood flow, and blood volume. We used the resultant blood volume curves to compute the thermal stress indicators (see Fig. 17). Due to the differences in units for the EE and the thermal imaging measurements, the results have been normalized for comparison (see Table 3 and Fig. 18). In all but one subject, the comparison metric showed a difference no greater than 16 points from the respective ground truth metric; that is the corresponding EE. In most cases the difference was within 8 points (subjects 1, 2, 3, 4, 5, 6, 8, 9, and 11). The correlation of the thermal infrared information to the ground truth data can be found and quantified by computing the Pearson correlation factor, r, $(-1 \leqslant r \leqslant 1)$ [46]. For all subjects, the Pearson correlation value is $r = 0.52$. If the one outlier subject (subject 12) is excluded from our computation, the correlation value becomes $r = 0.91$. The results indicate that the thermal imaging methodology correlates very well with the ground truth EE data for typical subjects. The blatant outlier drives the correlation sharply downwards due to the small size of the set.

Table 3
Ground truth (EE) data compared to thermal imaging (TI) data

Subject	Ground truth (EE)	Thermal imaging (TI)
1	48	47.750770
2	4	10.427090
3	30	22.554500
4	60	56.136000
5	21	24.208160
6	9	6.267478
7	28	17.334160
8	19	16.778860
9	33	40.234130
10	19	3.3304120
11	34	27.127570
12	80	14.692870

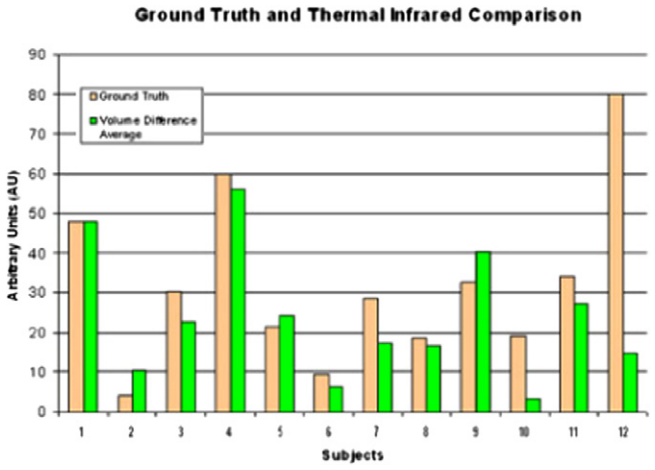

Fig. 18. Comparison bar graph of the difference measures indicating relative stress levels measured using the ground truth EE and the novel thermal methodology, respectively.

We do not have a good explanation for the outlier. The most probable scenario is human error by the clinical technician who was operating the instrument at the time. There is some indication of improperly calibrated thermal values in the field of view for that subject. No such incidence has occurred since then in numerous stress experiments which were performed as part of the interactivity work (see Section 9). In fact, what we measure is metabolic activation of the corrugator muscle, which is concomitant to mental processes. This appears to be in accordance with the observation theories of Paul Ekman, who coded this facial action unit in his FACS scheme of primitive non-verbal communications, as thought driven [52]. In other words, our stress quantification work can be viewed as the physiological support of certain visual FACS observations.

8. Applications—sleep studies

Recently and in cooperation with the Sleep Lab at the University of Texas Medical School, we started to investigate the application of the technology in the detection and management of incidents of sleep apnea. This time, the thermal imaging sensor is fixed above the bed of the subject and communicates with the desktop computer via the Ethernet. We continue to benchmark the performance of the image-based breath rate quantification algorithm against a ground-truth sensor (respiratory belt).

Several subjects have been processed so far and the performance of the system appears to exceed expectations. First, the thermal signal produced in the vicinity of the nasal area is detectable when subjects are at sleep and with help form the tracker is recovered in pristine form. In Fig. 19(a) one can observe in the encircled region of interest, how the temperature fluctuates between expiratory state (green color in the nostril) and inspiratory state (red color in the nostril). This produces the periodic breath signal using the method described in Section 6. This signal is depicted in Fig. 19(b) and is in perfect sync with the ground-truth breath signal. In contrast to what is reported in Section 6, the measurement is performed on the nasal-mandible tissue itself and not in the environmental vicinity. For this reason the polarity of the signal is reversed. The measurement region of interest is dictated by the application scenario, as the subject's face is more accessible in frontal view.

Initial skepticism about the accuracy of the imaging-based method has started to relax. The method appears to stand on an equal footing with the ground-truth sensing method. The weak point of this new technology is that the signal can be lost if the nasal measurement area disappears from the field of view. Fig. 20 exemplifies this scenario, where the subject's face is covered almost entirely by the blanket. However, we have found that the imaging method outperforms the ground-truth sensing method when the subject moves in his sleep. Then, the noise introduced to

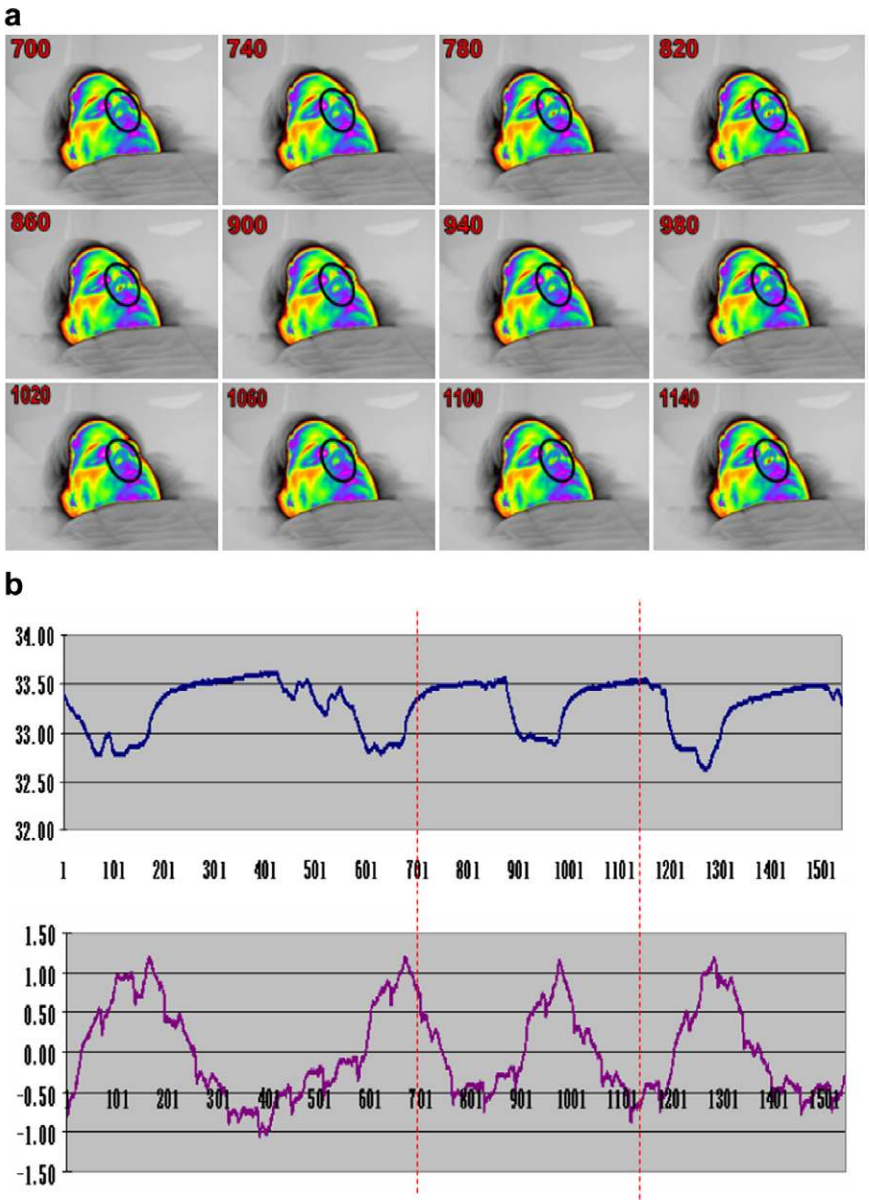

Fig. 19. (a) Thermal snapshots during consecutive expiration and inspiration phases of a normal sleep subject. (b) Normal breath waveforms produced by the thermal imaging system (top) and the respiratory belt (bottom). The waveform between the red lines corresponds to the frame sequence depicted in (a).

the ground-truth signal is typically much more severe than the corresponding noise in the imaging signal. In general, it is impossible to maintain continuous and accurate measurements throughout the sleep session with either method. The goal is to get accurate measurements most of the time or some of the time and these are typically enough to paint the medical picture. From this point of view, the imaging method appears to perform as well as the ground-truth sensing method. In fact, the imaging method accurately recorded incidents of sleep apnea concomitantly with the ground-truth sensing instrument (see Fig. 21).

Where the imaging method really shines, however, is the highly automated and totally unobtrusive nature of its operation. By contrast, Fig. 22 depicts a fully outfitted sleep studies subject using legacy sensing. The discomfort caused by the monitoring method may interfere with the sleep routine of the subject, thus, contaminating the experiment.

9. The interactivity potential

We have presented three novel methods for quantifying at a distance three vital signs, that is, blood flow on a superficial blood vessel, pulse, and breath rate respectively. The common thread of these three methods is that they depend on the same sensing modality, which is thermal imaging. The methods are based on bioheat modeling of physiological signals embedded in the thermal radiation emitted from the human face. A robust facial tissue tracking algorithm allows the measurements to be taken in the

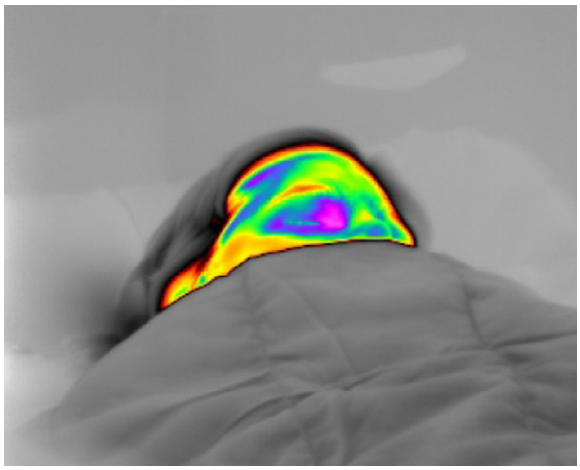

Fig. 20. Nasal area is blocked by the blanket from the camera's field of view.

presence of natural head motion. The methodology is ideal for sustained physiological monitoring, since the sensing element is passive and does not require contact and the measurements are carried out in a highly automated manner. In fact, the system can operate at the desktop level and at home and be incorporated into the human–computer interface.

Analysis of the so measured vital signs can provide clues about the presence and nature of various health problems and psychological states. The methodology is ideal for monitoring chronic conditions that are difficult to diagnose and monitor otherwise, because they manifest themselves stochastically. Examples of such ailments are heartbeat irregularities and sleep apnea that correlate with cardiac pulse and breath rate correspondingly. Medical studies are ongoing with our partner lab at Mayo Clinic regarding heartbeat irregularities. But, we have a clear indication of the method's usefulness in sleep studies and the monitoring of sleep apnea incidents.

Ultimately, this technology may be used at the desktop to monitor the computer user's health and psychological state and interact accordingly. However, such a widespread application of the technology is not feasible in the present time due to the high cost of the sensing element ($60 K for an entry level camera). Initially, the technology is expected to find home in specialized HCI applications where it has distinct advantages and the high cost can be tolerated.

We have established that our technology can quantify stress. Stress monitoring is clearly an area where the system is very appealing. It is not only unobtrusive but also capable of quantifying stress through unique localized channels (i.e., blood supply to the corrugator muscle), which are presumably more predictive. Traditionally, stress is quantified through global channels, like pulse and breath rate, which our technology is also capable of monitoring. However, such systemic vital signs are sometimes non-predictive as they are affected by patho-physiological conditions that are non-stress related (e.g., respiratory ailments).

Currently, we are evaluating the interactivity potential of the technology in driver usability studies. The subject operates a car driver simulator while is being monitored by the thermal imaging system. During the course of his simulated driving is being distracted by conversation initiated by another subject, who emulates a car passenger, cell

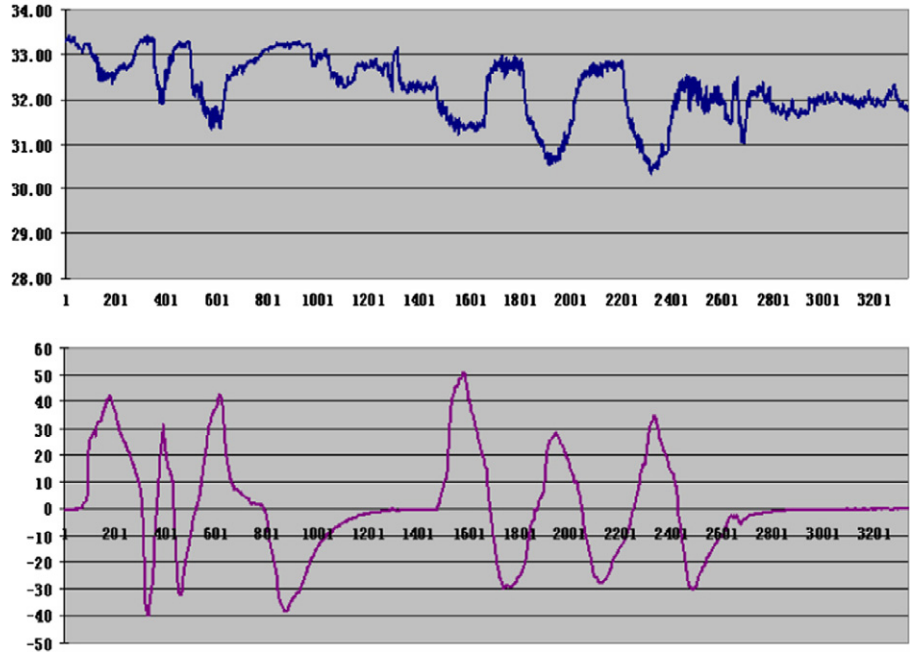

Fig. 21. Abnormal breath waveforms computed though via the imaging system (top) and recorded via the ground-truth respiratory belt (bottom). The sleep apnea incident is manifested in the lengthy flat of the breath waveform and the instantaneous drop of the breathing rate.

Fig. 22. Patient outfit in classical sleep studies. Courtesy of the Sleep Lab of the University of Texas Medical School.

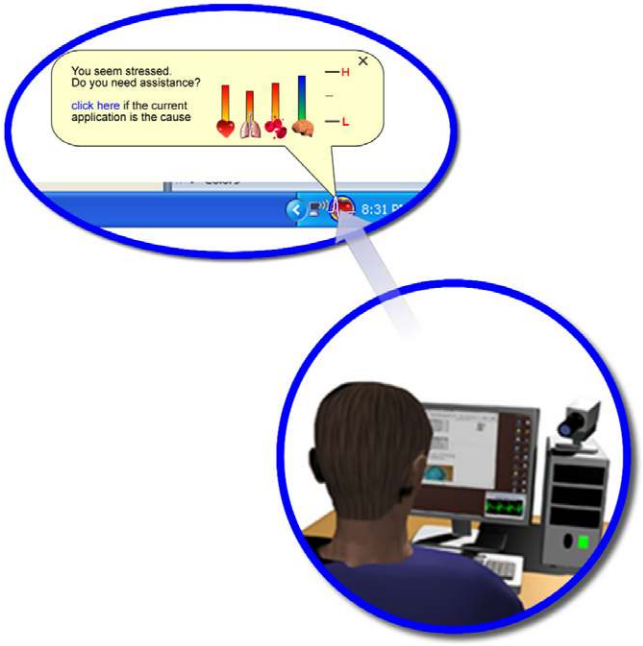

Fig. 23. Stress alert based on measurements of the physiological signs on the face. The alert is delivered through the task bar and the user may momentarily interrupt his current activity to provide feedback.

phone calls, music, random noise, and engine and environmental indicators. The purpose of the study is:

(1) Determine the breaking point for the average driver beyond which he cannot drive effectively and safely.
(2) Find the best ergonomic scheme possible (e.g., position and appearance of indicators) to optimize the average driver's performance within the limits of tolerable mental load.

The test software incrementally adds mental load, consults with the physiological output of our system and if stress indicators change significantly in response to a new stimulus (e.g., high RPM indicator), it interacts with the user and solicits his input (see Fig. 23) Then, it compares this with his driving score, produced by the car simulator and reports the results to the study team. The same scenario is repeated for the same subject at a later time but the stimulus is delivered in a different manner. For example the RPM indicator is placed in a different position in the graphical user interface (GUI) and assumes a different form. The physiological indicators include pulse, breath rate, periorbital blood perfusion, and blood flow in the frontal vessels on the forehead that supply the corrugator muscle. The pulse and breath rate are traditional systemic stress indicators, which are measured in a novel manner in our desktop HCI scheme. Their relative weight is small in our stress scoring system. The periorbital perfusion is indicative of the 'fight or flight syndrome' (surprise) [10] and the frontal blood flow is indicative of mental stress (brain icon in Fig. 23). The latter has the most weight in our stress scoring scheme, as it is highly predictive of the mental load caused by competing stimuli. Such competing stimuli were superbly emulated in our experiments via the Stroop test (see Section 7) and are characteristic of interactive software systems, such as simulators.

The driver usability studies is one application example for this technology. There are many other potential HCI uses for stress monitoring. Measurements during conventional usability testing could help identify user interface features that increase user stress, even if the user is not consciously aware of them. Use of our system during computerized testing could identify questions that are unusually stress-inducing. Monitoring of software users on a routine basis, could identify software issues not uncovered by traditional usability testing. In the future, software or operating systems could adapt dynamically to the user's emotional state, for example, suspending non-essential processes when the users are under considerable stress.

In the sleep studies application, we are currently working towards building a unique interactivity scheme not possible before the advent of this technology. The computer processing the thermal data on the subject's bed may adjust the sleep number of the mattress automatically, depending on the apnea scoring through the night. At the same time, depending on the statistical frequency and severity of the apnea incidents, the system may determine the recom-

mended dose of sip-up medication for the next day, based on the doctor's guidelines encoded in the software. Medicine dosage is presented as an alert through the task-bar in our GUI scheme, which we found to be a convenient way to interact with the user without cluttering his desktop.

In conclusion, the potential benefit to HCI and preventive medicine of this novel and unique technology can be enormous. We understand that this benefit will be realized incrementally, starting from specialized usability applications and ultimately precipitating in widespread desktop applications only as the cost of the technology drops. At the same time, we also realize that desktop physiological monitoring raises privacy and ethical issues, beyond the scope of this paper [53].

Acknowledgments

We thank the National Science Foundation (Grant No. IIS-0414754) and Dr. Ephraim Glinert, for their support and encouragement during the lifetime of this burgeoning research project. Equally, we also thank the Computer Science Department of the University of Houston for providing additional support. The views expressed by the authors in this paper do not necessarily reflect the views of the funding agencies.

A number of individuals and labs have contributed in various ways in this project including the Endocrine Research Lab of Dr. James Levine at Mayo Clinic, the Medical Usability Lab of Dr. Justin Starren at Columbia University, and the Sleep Lab of Dr. Jay Murthy at the University of Texas Medical School were most of the human experiments are carried out. Special thanks to Dr. Marc Garbey, Dr. Arcangelo Merla, and Dr. Panagiotis Tsiamyrtzis for providing mathematical, biomedical, and statistical support, respectively.

References

[1] FLIR Systems, 70 Castilian Dr., Goleta, California 93117. Available from: <http://www.flir.com>.

[2] I. Pavlidis, J. Levine, Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph testing, in: Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 3, Istanbul, Turkey, 2001, pp. 2826–2829.

[3] M. Garbey, A. Merla, I. Pavlidis, Estimation of blood flow speed and vessel location from thermal video, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, Washington D.C., 2004, pp. 356–363.

[4] N. Sun, M. Garbey, A. Merla, I. Pavlidis, Imaging the cardiovascular pulse, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, 2005.

[5] J. Fei, Z. Zhu, I. Pavlidis, Imaging breathing rate in the CO_2 absorption band, in: Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Shanghai, China, 2005.

[6] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. Frank, P. Ekman, Lie detection—recovery of the periorbital signal through tandem tracking and noise suppression in thermal facial video, in: E. Carapezza (ed.), Proceedings of SPIE Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV, vol. 5778, Orlando, FL, 2005.

[7] I. Pavlidis, Continuous physiological monitoring, in: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2, Cancun, Mexico, 2003, pp. 1084–1087.

[8] I. Pavlidis, J. Levine, P. Baukol, Thermal image analysis for anxiety detection, in: Proceedings of the 2001 IEEE International Conference on Image Processing, vol. 2, Thessaloniki, Greece, 2001, pp. 315–318.

[9] J. Levine, I. Pavlidis, M. Cooper, The face of fear, The Lancet 357 (9270) (2001) 1757.

[10] I. Pavlidis, N. Eberhardt, J. Levine, Human behavior: seeing through the face of deception, Nature 415 (6867) (2002) 35.

[11] C. Puri, L. Olson, I. Pavlidis, J. Levine, J. Starren, Stresscam: non-contact measurement of users' emotional states through thermal imaging, in: Proceedings of the 2005 ACM Conference on Human Factors in Computing Systems (CHI), Portland, OR, 2005, pp. 1725–1728.

[12] M. Yang, D. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (1) (2002) 34–58.

[13] C. Eveland, D. Socolinsky, L. Wolff, Tracking human faces in infrared video, in: Proceedings 2001 IEEEWorkshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications, Kauai, HI, 2001.

[14] I. Fujimasa, T. Chinzei, I. Saito, Converting far infrared image information to other physiological data, IEEE Engineering in Medicine and Biology Magazine 19 (2000) 71–76.

[15] J. Keyserlingk, P. Ahlgren, E. Yu, N. Belliveau, M. Yassa, Functional infrared imaging of the breast, IEEE Engineering in Medicine and Biology Magazine 19 (2000) 30–41.

[16] J. Head, F. Wang, C. Lipari, R. Elliot, The important role of infrared imaging in breast cancer, IEEE Engineering in Medicine and Biology Magazine 19 (2000) 52–57.

[17] J. Keyserlingk, P. Ahlgren, M. Yassa, N. Belliveau, Functional infrared imaging as part of a multi-imaging strategy for breast cancer detection and therapeutic monitoring, in: The Second Joint EMBS-BMES Conference, Houston, TX, 2002.

[18] A. Boyd, S. Maloney, Digital infrared thermal imaging as biofeedback tool: Monitoring chemotherapy response in a young female with breast cancer mediastinal secondaries, in: The Second Joint EMBS-BMES Conference, Houston, TX, 2002.

[19] M. Frize, C. Herry, R. Roberge, Processing of thermal images to detect breast cancer: Comparison with previous work, in: The Second Joint EMBS-BMES Conference, Houston, TX, 2002.

[20] M. Anbar, C. Brown, L. Milescu, J. Babalola, L. Gentner, The potential of dynamic area telethermometry in assisting breast cancer, IEEE Engineering in Medicine and Biology Magazine 19 (2000) 58–62.

[21] H. Qi, P. Kuraganti, Z. Liu, Early detection of breast cancer using thermal texture maps, in: 2002 IEEE International Symposium on Biomedical Imaging: Macro to Nano, Washington D.C., 2002.

[22] M. Partinen, J. Alihanka, J. Hasan, Detection of sleep apneas by the static chargesensitive bed, in: Proceedings of the 6th European Congress on Sleep Research, Zurich, Switzerland, 1982.

[23] K. Nepal, E. Biegeleisen, T. Ning, Apnea detection and respiration rate estimation through parametric modeling, in: Proceedings of the 28th IEEE Annual Northeast Bioengineering Conference, Philadelphia, PA, 2002.

[24] H. Shukla, E. James, J. Schutz, Window to the heart: the value of a native and paced qrs duration. current perspective and review, Journal of Interventional Cardiac Electrophysiology: An International Journal of Arrhythmias and Pacing 9 (3) (2003) 333–342.

[25] E. Schulze-Bahr, G. Mnnig, L. Eckardt, The long qt syndrome: considerations in the athletic population, Current Sports Medicine Reports 2 (2) (2003) 72–78.

[26] K. Sybrandy, M. Cramer, C. Burgersdijk, Diagnosing cardiac contusion: old wisdom and new insights, Heart 89 (5) (2003) 485–489.

[27] B. Shneiderman, Leonardo's Laptop, The MIT Press, Cambridge, MA, 2002.
[28] B. Hayes-Roth, G. Ball, C. Lisetti, R. Picard, A. Stern, Panel on affect and emotion in the user interface, in: Proceedings of the 3rd International Conference on Intelligent User Interfaces, ACM Press, NY, UsSA, 1998, pp. 91–94.
[29] A. Marcus, The emotion commotion, Interactions 10 (6) (2003) 28–34.
[30] R. Picard, Perceptual user interfaces: affective perception, Communications of ACM 43 (3) (2000) 50–51.
[31] J. Klein, Y. Moon, R. Picard, This computer responds to user frustration, Interacting with Computers 14 (2002) 119–140.
[32] R. Picard, J. Klein, Computers that recognise and respond to user emotion: Theoretical and practical implications, Interacting with Computers 14 (2002) 141–169.
[33] J.M. Artz, The role of emotion in reason: and its implications for computer ethics, SIGCAS Computer Society 30 (1) (2000) 14–16.
[34] D. Norman, Emotion & design: attractive things work better, Interactions 9 (4) (2002) 36–42.
[35] R.W. Picard, Affective medicine: Technology with emotional intelligence, in: R.G. Bushko (Ed.), Future of Health Technology, vol. 80 of Studies in Health Technology and Informatics, IOS Press, Amsterdam, The Netherlands, 2002.
[36] R. Nakatsu, J. Nicholson, N. Tosa, Emotion recognition and its application to computer agents with spontaneous interactive capabilities, in: Proceedings of the Seventh ACM International Conference on Multimedia (Part 1), ACM Press, NY, USA, 1999, pp. 343–351.
[37] C. Lisetti, F. Nasoz, Maui: a multimodal affective user interface, in: Proceedings of the 10th ACM International Conference on Multimedia, ACM Press, NY, USA, 2002, pp. 161–170.
[38] J. Scheirer, R. Fernandez, J. Klein, R. Picard, Frustrating the user on purpose: a step toward building an affective computer, Interacting with Computers 14 (2002) 93–118.
[39] Project Cyborg. Available from: <http://www.rdg.ac.uk/KevinWarwick/index.html>.
[40] I. Pavlidis, J. Levine, Thermal image analysis for polygraph testing, IEEE Engineering in Medicine and Biology Magazine 21 (6) (2002) 56–64.
[41] J. Werner, M. Buse, Temperature profiles with respect to inhomogeneity and geometry of the human body, Journal of Applied Physiology 65 (3) (1988) 1110–1118.
[42] Vasamed, 7615 Golden Triangle Drive, Eden Prairie, Minnesota 55344. Available from: <http://www.vasamed.com>.
[43] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C, second ed., Cambridge University Press, New York, 1992, pp. 504–521.
[44] N. Charkoudian, Skin blood flow in adult human thermoregulation, Mayo Clinic Proceedings 78 (2003) 603–612.
[45] ADInstruments, 2205 Executive Circle, Colorado Springs, Colorado 80906. Available from: <http://www.adinstruments.com>.
[46] K. Pearson, Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia, Philosophical Transactions of the Royal Society of London. Series A 187 (1896) 253–318.
[47] J. Lindemann, R. Leiacker, G. Rettinger, T. Keck, Nasal mucosal temperature during respiration, Clinical Otolaryngology 27 (2002) 135–139.
[48] Y. Zhang, Quantitative measurement of radiation properties for opaque and semitransparent greybodies, Infrared Physics 30 (2) (1990) 149–153.
[49] D. Olson, Computerized stroop ver. 1.0, Rochester, Minnesota (2004).
[50] J. Stroop, Studies of interference in serial verbal reactions, Journal of Experimental Psychology 18 (1935) 643–662.
[51] G. Seematter, M. Dirlewanger, V. Rey, P. Schneiter, L. Tappy, Metabolic effects of mental stress during over and underfeeding in healthy women, Obesity Research 10 (2002) 49–55.
[52] P. Ekman, The face of man. Expressions of universal emotions in a New Guinea village, Garland STMP Press, New York and London, 1980.
[53] C. Reynolds, R. Picard, Affective sensors, privacy, and ethical concerns, in: CHI2004, Vienna, Austria, 2004.