

# A Survey of Approaches to Unobtrusive Sensing of Humans

JOSÉ MARCELO FERNANDES, University of Coimbra, CISUC, Portugal

JORGE SÁ SILVA, University of Coimbra, INESC, Portugal

ANDRÉ RODRIGUES, Polytechnic of Coimbra, Coimbra Business School Research Centre|ISCAC,  
Portugal and University of Coimbra, CISUC, Portugal

FERNANDO BOAVIDA, University of Coimbra, CISUC, Portugal

The increasing amount of human-related and/or human-originated data in current systems is both an opportunity and a challenge. Nevertheless, despite relying on the processing of large amounts of data, most of the so-called smart systems that we have nowadays merely consider humans as sources of data, not as system beneficiaries or even active “components.” For truly smart systems, we need to create systems that are able to understand human actions and emotions, and take them into account when deciding on the system behavior. Naturally, in order to achieve this, we first have to empower systems with human sensing capabilities, possibly in ways that are as inconspicuous as possible. In this context, in this article we survey existing approaches to unobtrusive monitorization of human beings, namely, of their activity, vital signs, and emotional states. After setting a taxonomy for human sensing, we proceed to present and analyze existing solutions for unobtrusive sensing. Subsequently, we identify and discuss open issues and challenges in this area. Although there are surveys that address some of the concerned fields of research, such as healthcare, human monitorization, or even the use-specific techniques like channel state information or image recognition, as far as we know this is the first comprehensive survey on unobtrusive sensing of human beings.

CCS Concepts: • **General and reference** → **Estimation; Experimentation**; • **Hardware** → **Sensor applications and deployments; Wireless devices; Wireless integrated network sensors**;

Additional Key Words and Phrases: Unobtrusive sensing, signal processing, IoT, data processing, HiTL

## ACM Reference format:

José Marcelo Fernandes, Jorge Sá Silva, André Rodrigues, and Fernando Boavida. 2022. A Survey of Approaches to Unobtrusive Sensing of Humans. *ACM Comput. Surv.* 55, 2, Article 41 (January 2022), 28 pages. <https://doi.org/10.1145/3491208>

This work is funded by national funds through the FCT - Foundation for Science and Technology, I.P., within the scope of the project CISUC - UID/CEC/00326/2020 and by European Social Fund, through the Regional Operational Program Centro 2020. Also, the research presented in this article was carried out in the scope of the EEA Grants Portugal 2014–2021–Bilateral Initiative number 63 “PrivacyCoLab” and the project iFriend - Supervisión inteligente del estado de salud en personas mayores con insuficiencia renal mediante dispositivos inalámbricos” supported by Fundación CSIC, Interreg Portugal-Espanha. José Marcelo Fernandes wishes to acknowledge the Portuguese funding institution FCT - Foundation for Science and Technology for supporting his research under the Ph.D. grant SFRH/BD/147371/2019.

Authors’ addresses: J. M. Fernandes, University of Coimbra, CISUC, Pólo II - Pinhal de Marrocos, Coimbra, Coimbra, 3030-290, Portugal; email: jmfernandes@dei.uc.pt; J. S. Silva, University of Coimbra, INESC, Coimbra, Coimbra, Portugal; email: sasilva@deec.uc.pt; A. Rodrigues, Polytechnic of Coimbra, Coimbra Business School Research Centre|ISCAC, Coimbra, Coimbra, Portugal and University of Coimbra, CISUC, Coimbra, Coimbra, Portugal; email: andre@iscac.pt; F. Boavida, University of Coimbra, CISUC, Coimbra, Coimbra 3030-290, Portugal; email: boavida@dei.uc.pt.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Association for Computing Machinery.

0360-0300/2022/01-ART41 \$15.00

<https://doi.org/10.1145/3491208>

## 1 INTRODUCTION

Now more than ever, we have extremely large amounts of data at our disposal. For instance, a supermarket chain can deal with hundreds of thousands of products equipped with **Radio Frequency Identifiers (RFID)**, and use **RFID** readers to scan these items every second, generating about 12.6 GB per second and about 544 TB per day [13]. Building on the available data, the number of smart systems is also increasing. In [63], the **European Technology Platform on Smart Systems Integration (EPoSS)** defines smart systems as systems that “*are able to sense, diagnose, describe, qualify and manage a given situation, ... They are able to interface, interact and communicate with users, their environment and with other Smart Systems.*”

Smart systems should be able to perform and incorporate functions of sensing, actuation, and control, in order to describe and analyze a specific situation and make decisions based on the available data. In most cases, the smartness of the system can be attributed to autonomous operation based on closed-loop control, energy efficiency, and networking capabilities. However, despite the fact that current systems are “*smart*” in many ways, the majority of these systems ignore the human factor or treats humans as an external factor [62]. Nevertheless, in order to achieve a truly smart system, we need to work in the integration of humans as part of the closed-loop process. That is, humans need to be taken into account in every phase of the loop, be it the acquisition of data, processing/inference, or the actuation phase.

At the same time, we are witnessing the emergence of new sensor techniques, such as virtual and social sensors [9], that show the importance of humans within the sensing system. In our opinion, only a system that is able to extend its capabilities to perceive and adapt to human actions, intentions, and emotional states, can be considered a truly smart system. Only then we can enter the realm of **Human-in-the-Loop-Cyber-Physical-Systems (HiTLCPS)**, where humans and machines interact, cooperate, coexist, and enhance our current systems.

In order to achieve this goal, we first need to endow our systems with sensory capabilities that are tailored to perceive humans. Human actions are, most of the times, unpredictable when analyzed by a common observer. However, trained observers can perceive certain indicators that give them information to classify those actions. These indications are often accompanied by involuntary physiological reactions, e.g., fluctuations in the **heart rate (HR)**, or an irregular **breathing rate (BR)**. As such, by creating and incorporating sensors that are able to gather information about those physiological responses in our systems, we are, in turn, moving toward empowering them with the ability to perceive and understand humans.

Most of the work and effort done toward the development and integration of human-specific sensors was carried out in the field of wearable sensors, like, for example, the works in [10, 49, 66, 81]. These devices allow for long-term monitoring that, alternatively, would require long-term hospitalizations and/or ambulatory environments that are much more expensive to set up and maintain [14]. This represents a great advance not only in clinical terms but also when it comes to gathering more information about human lives. The success of this type of technology paved the way for the emergence of several commercial solutions (e.g., [25, 38, 91]). Other approaches like Body Sensor Networks or Body Area Networks, have also largely advanced in the last few years, improving medical solutions, the training of professionals (e.g., military, sport athletes), or even other aspects of our daily lives like virtual reality gaming [6, 19]. However, even when considering the advances in the last decade when it comes to miniaturization of devices and wireless technologies, these devices can be bothersome to use, and most of them require some cooperation from the user. Even with solutions like [10], which rely on wearable sensors that are embedded in clothing, we still have to take into account that the user has to wear that specific piece of cloth and have some special care in order to maintain the system in good condition.

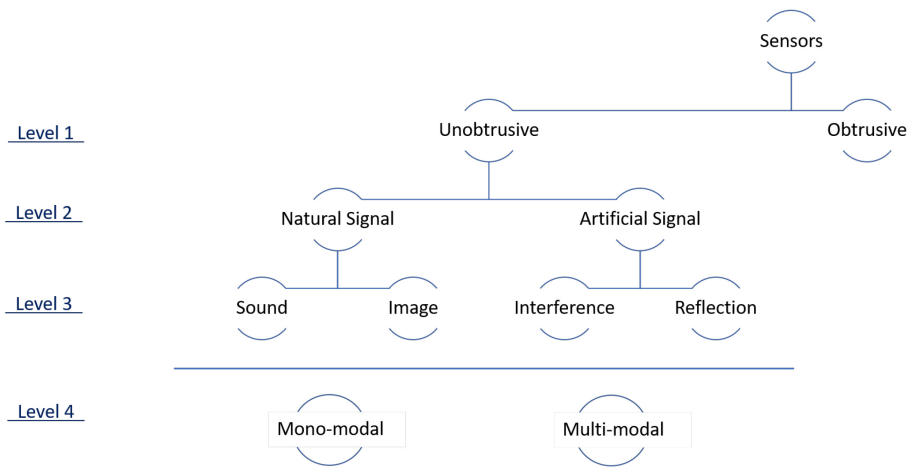


Fig. 1. Proposed human sensing taxonomy.

Nevertheless, despite being around for more than one decade, wearable devices and associated sensing techniques have yet to be explored in real clinical scenarios and have not yet been approved for medical use [70], with even the most recent versions of commercial solutions having considerable errors when compared to a straightforward, traditional **electrocardiogram (ECG)** [11]. Recent surveys showed that 32% of wearable device users stopped using the device after 6 months, and 50% of users stopped using it after the first year [48]. Additionally, it has also been shown that people that tend to possess and purchase wearable devices, are the ones that are already leading a healthy lifestyle [12]. This shows that we still need different approaches in order to perform long-term monitorization of humans.

Motivated by the problems and limitations associated with wearable, obtrusive sensing, there is currently some work in the field of unobtrusive sensing of humans' physical and emotional parameters. The aim and contributions of this article are a comprehensive review of the state-of-the-art in this area, and the discussion and analysis of remaining challenges and open issues.

The rest of the article is organized as follows: In the Section 2, we present our taxonomy for unobtrusive sensing. This taxonomy will be used as a guide and framework for the presentation of the state-of-the-art for unobtrusive sensing, in Section 3. In Section 4, we identify open issues and discuss possible research directions. Lastly, in Section 5, we present the conclusions.

## 2 A TAXONOMY FOR HUMAN SENSING

Before we proceed to identify and present existing approaches to unobtrusive sensing of human beings, it is essential to define a general taxonomy for human sensing that can set the terminology and serve as a guide for the discussion that follows. The proposed taxonomy can be seen in Figure 1. In this taxonomy, we propose a three-level division, with a fourth common level. These are briefly presented below.

In the first level, we divide sensing into obtrusive and unobtrusive. On one hand, unobtrusive sensing solutions are able to monitor the users in a continuous way or during large temporal windows, in a contactless way and without requiring any specific actions from the users. On the other hand, unobtrusiveness cannot be defined as just being contactless, or user-interaction free, as the negative emotional effects of a user's perception of the sensing technique must also be considered obtrusive. For instance, some studies have addressed the negative effects of surveillance,

and how it correlates with anxiety [64]. However, in this review we specifically address physical obtrusiveness. As such, we consider any solution that requires physical contact or explicit human actions as being obtrusive (e.g., wearables, ECG). In this article, we will focus mainly on solutions that set as their goal to monitor humans' physiological signals (e.g., heart rate, respiratory rate, activity detection) and emotional states (e.g., happiness, sadness).

The second-level branch considers the nature/origin of the signal being used for sensing. Two clear patterns were noticed during the review of existing techniques, corresponding to two possible origins of the sensed signal. The signal can be natural or it can be artificial. Natural signals are originated by human beings themselves (e.g., heartbeat sound, respiration sound, human body motion images) and are then captured by the sensing technique. Artificial signals are generated for the purpose of monitoring humans. They are created by the sensing technique itself, interact with their target, and are then captured by a sensing system (e.g., radar waves).

At the third level, for the natural signals branch, the division is on how the signal is being captured. Much like human beings perceive information from the environment through their senses, systems have been empowered with capabilities that emulate those senses, such as cameras to emulate the vision, microphones to emulate our hearing, or pressure sensors that emulate the sense of touch. For the purpose of unobtrusive sensing of physical and emotional states, only image and sound are used.

Still at the third taxonomy level, when considering signals of an artificial nature, the signal is always external to the human body. When that signal interacts with the human body it is transformed and then perceived by the used technology. Upon analyzing the state-of-the-art, two physical phenomena seem to be the most commonly used: signal interference, and signal reflection. Therefore, those were the considered classes for the Artificial Signals branch.

Last but not least, every third-level node will be evaluated in terms of how many modalities it can work on, that is, if the technique can capture more than one physiological signal or if it can be used to perceive one or more states (e.g., physiological, emotional). For example, a technique that can only be used to monitor breathing rate, will be monomodal while one that can monitor both the breathing rate and the heart rate will be considered multimodal. The same happens for one technique that can capture only the heart rate, against one that captures both the heart rate and stress level.

In the next section, each of the nodes of the unobtrusive branch of the taxonomy will be explored.

### 3 EXISTING APPROACHES TO UNOBTRUSIVE SENSING

This section provides an overview of the state-of-the-art of unobtrusive sensing of humans. For this, the taxonomy proposed in Section 2 is used. In Subsection 3.1, we present and discuss approaches that rely on natural signals, whereas in Subsection 3.2, approaches that use artificial signals are addressed. Subdivisions of these subsections are also done according to the taxonomy.

#### 3.1 Natural Signals

We consider a signal to be natural if it is directly produced by the human body (e.g., heat, breathing sound, body's image, speech). In this regard, sensing techniques may explore two types of signals, namely, sound and image.

*3.1.1 Sound-Based Techniques.* The work by Zhenhua et al. [39] uses a geophone [79] to capture the changes in a person's heart rate during sleep. Geophone is more commonly used to monitor earthquakes, but these devices can also generate a noticeable response to sounds such as those created by the sound of a beating heart. Other authors have proposed solutions based on custom

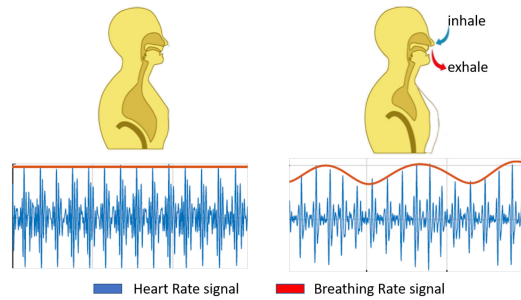


Fig. 2. Signal modulated in amplitude by the respiratory rate. The signal at the left was obtained while the subject held his/her breath. The right-hand side signal was obtained while normally breathing. Image adapted from [40].

bed or altered bed parts to monitor humans' heart rate, such as the works of [17, 87]. However, these solutions require specific alterations to someone's bed and are not suited for wide adoption.

Geophone is insensitive to low-frequency movements. Due to this feature, it automatically filters out any signal caused by low-frequency movements, such as respiration. However, a geophone still captures other movements, such as rolling on the bed or even other people walking next to the bed. As such, these movements need to be filtered out in order to obtain a clear signal that can be compared to the heart rate signal. In [39], the authors were able to obtain the number of sound peaks in a time window and, by extrapolation of that value to a 1-minute window, the heart rate was obtained.

Two experiments were also conducted: one in a controlled environment and a second one in real apartments. The first experiment comprised 34 healthy subjects, including 26 males and 8 females, with ages between 22 and 65 years. An average heart rate estimation error of 1.30% was obtained when the subjects were lying still, as opposed to an average estimation error of 3.87% when the subjects were asked to perform some movements during the experiment. In the second scenario, the system was installed in the homes of nine different subjects during a period of 25 nights, and a mean estimation error of 8.25% was obtained. All of these results were obtained by comparing the results with those from a finger oximeter, and correlating the data with camera footage in the second experiment.

Some of the authors of the previous work decided to extend their work in [40], by demonstrating that this technique was able to estimate the breathing rate and, furthermore, that it was able to monitor two persons sleeping in the same bed at the same time.

Concerning the first problem, due to the geophone being insensitive to low-frequency signals, it cannot normally detect the breathing rate. Nevertheless, after some experimentation and observations, the authors concluded that respiration modulates the geophone signal in amplitude. This can be seen in Figure 2, where the graphic on the left corresponds to the signal obtained while the subject held his breath, and the graphic on the right corresponds to the signal obtained while the subject breathed normally. This can be explained because breathing changes the amount of air in the chest and, in turn, its stiffness, and, consequently, the amount of energy that the heart-beat sound loses while crossing it. This phenomenon can be leveraged, in turn, to estimate the breathing rate.

The second problem addressed in [40] was the use of the technique developed in [39] to perform the detection of the heart rate and breathing rate for two persons while sharing a bed. The used system representation can be seen in Figure 3. By using two synchronized geophones, and by knowing each geophone's relative location to each subject (i.e., the geophone G1 is closer to H1

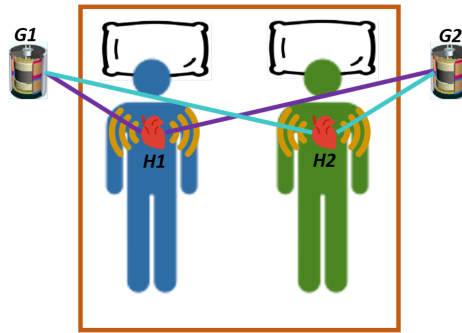


Fig. 3. Representation of the system used in [40] for two people in the same bed. With two geophones (G1 and G2) and two heartbeats sources H1 and H2.

than to H2, and the inverse situation happens to geophone G2), it was possible to separate the signals into two distinct signals, making it possible to estimate both the heart rate and the breathing rate of each of the two individuals. In this work, tests were also performed with 86 participants, and a breathing rate estimation error of 0.38 breaths per minute was obtained, for a single person. For tests with two persons, sharing the same bed, an average estimation error of 1.90 beats per minute was obtained for the heart rate estimation, and 2.62 breaths per minute for the breathing rate estimation.

Smartphones are now ubiquitous devices, carried by most people and kept near them even while sleeping. Commercial sensors that aim to monitor sleep quality, such as [25, 38], also leverage smartphones to store and process data. The work by Ren [73, 74] discards the use of wearable sensors and proposes a new framework for breathing rate and sleep quality monitoring. In this framework, the authors used a common off-the-shelf smartphone and an earphone to monitor the sleep activity and estimate the breathing rate. Although smartphones have a microphone, earphones have several advantages: firstly, earphone microphones have better audio quality; secondly, some users are reluctant to keep a smartphone close to them due to radiation; thirdly, earbuds can be used as microphones as well, thus increasing the recording capabilities and allowing them to record in stereo; lastly the use of earphones can increase the distance at which the device works. The authors also argue that despite the fact that earphones are additional devices to the system, they are a very common device that normally comes with the smartphone itself, and, as such, they are only reusing existing devices.

The framework presented in [73] comprises three modules, namely, noise reduction, breathing rate detection, and sleep event detection. Although most people sleep in a relatively quiet environment, there are always some sources of background noise, such as other electronic devices, pets, or even outside noises like cars. The first step to noise filtering is to apply a band-pass filter to remove high and low frequencies that are not in the range of the breathing frequencies. The biggest difference between ambient noise and respiration noise is stability, as the amplitude of ambient noise does not vary during small periods of time. As such, by computing the variability of the signal, it is possible to detect the frames that only contain ambient noise. The final step of this procedure is to subtract the noise from the signal that contains the breathing information and obtain a clean signal. After cleaning the signal, the authors first extracted the envelope from the acoustic sound. They then used the strong correlation relationship between the breathing cycles to capture the time length between the breathing cycles and, from there, they derived the breathing rate. The system was tested with nine people throughout a period of 6 months, and the average estimation error for the breathing rate was 0.5 breaths per minute.

In [73], the system was also applied in a case study to detect and classify the level of sleep apnea. The results of the system correlate with the truth, since the system was able to correctly classify the three subjects involved in the study. The focus on the clinical area also shows another possible applicability of this type of technique.

In addition, to using sound-based techniques to detect physiological states, there is some work that explores the use of sound for detecting emotions. Apart from visual clues, one of the most reliable ways for humans to detect emotion is through speech. The goal of automatically detecting emotion through voice has also been around for quite some time, with some works, such as [57], showing results in this field. Other authors have applied sound-based techniques to detect fear-type emotions, in dangerous or unpleasant situations for human beings [20], showing that these systems can also be used for increasing and promoting human safety.

This area has also seen some developments in recent years with the emergence of deep neural networks [33]. In the latter work, the authors focused on emotion recognition not only from speech but also from non-verbal sounds. Additionally, there is also the approach of converting speech to text and processing the text in order to infer sentiment [15]. However, with this approach we lose information, such as non-verbal sounds and sentiment, and we consider that the sensing is not direct. As such, we chose not to include this area in this survey.

*3.1.2 Image-Based Techniques.* Several studies exploit image and video techniques in order to monitor human activities, showing that this technique is quite effective [65]. Nevertheless, there is still an ongoing effort to use images and videos in order to monitor more than just activities.

The widespread use of camera-equipped devices (e.g., laptops, cameras in the dashboard of cars, smartphones) opens up opportunities to leverage the physiological phenomenon of skin color changes in order to unobtrusively capture the human heart rate. Some studies have already explored the possibility of exploiting cameras and flash as sensors [77], by analyzing variations of light reflection with the change of blood volume in fingers. However, this requires the user to hold his/her finger against the camera's flash and thus falls under the category of obtrusive sensing.

Because cardiac pulsation leads to subtle skin color changes, a **photoplethysmography (PPG)** signal can be measured through video analysis. One of the body parts that is more susceptible to color changes is the face, and some studies try to exploit this in order to capture the heart rate signal. Works such as the one from Kwon et al. [45], already proved that smartphones can indeed be used to monitor the heart rate through video recording. In that work, the author developed a mobile application that by recording 20 seconds of human face could estimate the heart rate of a person with an average error rate of 1.04% for the 10 trial participants. However, this technique has limitations as an unobtrusive technique, since the person has to stay still for the duration of the monitorization. Furthermore, high-frequency parameters such as the heart rate variability could only be extracted from smartphones that supported high frame rates. However, current day smartphones have drastically evolved, especially in terms of hardware. Nowadays, smartphones can have up to four rear cameras with more than 40 Megapixels, record videos in 4k, and even capture slow-motion at more than 480 frames per second, which could lead to even better results.

The studies in [46] and [51] both use mounted cameras in order to explore the detection of the heart rate based on skin color changes in realistic scenarios, i.e., scenarios with movement, bad illumination, and noise. Both works try to detect the **Region of Interest (ROI)** for facial detection. The results from [46] show that the forehead and both cheeks are good candidates for computationally efficient ROI, while the chin and nose are less suitable. These findings can address some of the limitations of [45], as the fact that the cheeks are a suitable ROI allows one to monitor the heart rate not only in a frontal facial position but also from both facial profiles. In [51], the authors also explored the possibility to monitor the heart rate of a person while playing a video game. This

could be used to monitor user experience throughout the game, and help game developers in their design work.

In [56], the authors focused on detecting the breathing rate and patterns by using a camera to monitor the movements of a pit in the neck. The authors present results for breath-by-breath respiratory rate, which is estimated from the processed breathing pattern. In addition, the effect of image resolution on monitoring breathing patterns and respiratory rate is also addressed, by comparing different camera resolutions. This system was tested on a group of 12 participants that were all healthy. The system showed a mean average error of 1.53 breaths per minute with the worst resolution, and of 0.55 breaths per minute with the best resolution.

Other than image techniques that use the visual spectrum, there have also been efforts toward using infrared imaging to monitor peoples' physiological signals. In [68], the authors employed a long wave infrared camera to capture the breathing rate of people. In this work, the authors explored the previously discussed technique of **ROIs** detection to segment the patient nose. Secondly, the authors segmented a second ROI in the region of the nostrils. The temperature around the nostrils fluctuates during the respiratory cycle (inspiration and expiration), making it possible to monitor the breathing rate from the subtle temperature changes ( $0.3^{\circ}\text{C}$ ). Experimentation of this technique was also performed with 11 healthy subjects, obtaining a mean average error of 0.71 breaths per minute.

Previous work had also proved that thermography could be used to detect the heart rate [26]. In this work, the authors based their work on the fact that the variance of skin temperature is strongest along the superficial blood vessels. The authors were able to detect the heart rate changes from major superficial vessels, such as those on the face, the carotid artery, and the radial vessel complex. After segmenting the area of interest of the vessels and filtering the image, the authors computed the Fast Fourier Transform and found out the most relevant frequency between 0.67 Hz and 1.67 Hz (40–100 bpm). The system was tested with 34 subjects obtaining a performance of 90.33%. Other authors have also employed the use of thermal cameras to detect the heart rate and breathing rate of people during their sleep [32]. This system used an array of infrared lighting and a thermal imaging camera, and a custom-made deep learning model to detect **ROI** and infer peoples' heart rate and breathing rate. The created system was tested with 26 and 25 sleeping subjects for the breathing rate and heart rate estimation, respectively, obtaining a mean estimation error of 1.865 breaths per minute and a mean estimation error of 4.293 beats per minute. This shows that this kind of system can also be used in sleep monitorization, where normal imaging techniques cannot be used due to lack of proper illumination.

Other authors focused on the practicality of these techniques in the medical area. Currently, to monitor newborns' vital signs, such as heart rate, breathing rate, or oxygen saturation, it is necessary to have sensors and electrodes sticking to the skin. This can bruise their vulnerable skin, cause infections, cause stress, or even pain. The use of unobtrusive techniques could therefore bring several advantages to the current state-of-the-art. In [44] and [2], in addition to the use of video recording to measure the heart rate, the previously discussed infrared technique of estimating the respiration rate by capturing the temperature around the nostrils was also used to monitor infants' breathing cycles. However, in [82] the authors were able to monitor newborns' heart rate, breathing rate, and oxygen's blood saturation with just a normal video camera mounted over their respective incubators. They were able to achieve this by filtering the image signal in the frequency domain. In this work, the authors went even further into the clinical realm, and were able to detect bradycardia accompanied by a major desaturation.

The use of unobtrusive techniques relying on natural signals can also be applied to emotional state inference. Some features extracted from physiological signals, such as heart rate variability, can be used to infer emotional states, as several studies proved [8, 47]. However, some works



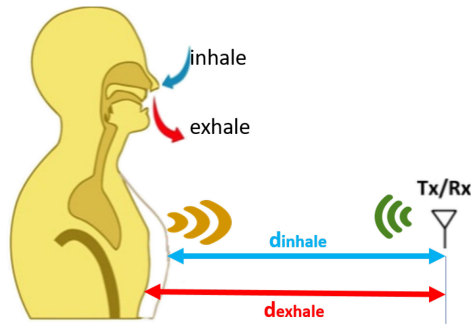


Fig. 4. Scheme adapted from [5]: Representation of the inhale and exhale events in the respiration cycle, and the respective distance from the chest to the radio antenna in both events.

focus on directly detecting the emotional state using the facial expression of people. This happens in [24], where the authors use a multimodal framework for smart homes, constituted by three different components to perceive emotions from the users. The used modules enable facial emotion detection through video, behavior detection through video, and valence/arousal detection with physiological wearable sensors. The fusion of the three modules happens at decision level, that is, each module works separately and a decision rule system then processes each output and gives a final classification. Although the latter module does not fit our classification of unobtrusive sensors, since the fusion happens at decision level, we can envision a framework constituted by the other two modules only, which would be unobtrusive. Additionally, the aim of this work is not only to detect emotions, but also to influence them, through music and color/light, thus providing an example on how unobtrusive sensing techniques can open ways to improve our daily lives and mindsets.

### 3.2 Artificial Signals

Contrary to the techniques presented in the previous subsection, there are techniques that resort to signals that are external to the human body, in order to monitor humans. Those techniques fall into the “Artificial Signals” branch of our taxonomy. There are two types of signal-body interactions used by these unobtrusive techniques, namely, signal reflection and signal interference. These are addressed below.

**3.2.1 Reflection-Based Techniques.** One of the studies in the area of unobtrusive sensing that uses the phenomenon of signal-body reflection is presented in [4]. In this study, a **Frequency Modulated Continuous Wave (FMCW)** radio wave was used to monitor the location of people in a 3D setup, by using the reflection effect of radio waves and T-shaped antenna arrays. By modulating the signal in frequency and capturing its reflection, the authors were able to capture the time of flight of the wave, which directly correlates to the distance to that object. Furthermore, by using one radio and three antennas, they were able to triangulate the signal and pinpoint the location of an object or person in three dimensions. Due to the fact that radio waves can cross walls, in this work they were also able to detect people location and movements even when the device had no line of sight to the subject.

Building on the previous work and on the fact that the respiratory cycle creates chest movements that affect the distance from the subject to the radio source, as can be seen in Figure 4, the authors developed the work in [5]. Using the system in [4], they transmitted a low-power wireless signal and measured the time it took for the signal to travel to the human body and reflect back to

its originating antenna. With this, they were able to easily compute the distance to the subject as well as the distance fluctuations caused by breathing cycles and, thus, could easily determine the breathing rate. The output signal of their system corresponds to the phase of the signal that returned to the radar after ricocheting in the human body. In this signal, there is information about the breathing rate and the heart rate, and by resorting to signal theory to filter it in the frequency domain, they were able to extract this information. In this work, the authors used the system to monitor humans, while they stood still or performed actions that did not require considerable movement (e.g., work on their computer, read, browsing the web on their phone). The system has quasi-static requirements in order to function properly, due to the fact that the movements of other parts of the body, other than the chest, caused by moving around or doing a certain activity (e.g., walking, exercising), cause too much distortion in the signal. The measures were correlated with results from wearable sensors that were attached to the users. The results show that it is possible to obtain an accuracy of 99.4% and 99% for the breathing rate and heart rate, respectively.

In [93], Zhao et al. build on the work of [4] and [5] by using the information gathered by the system to also monitor human emotions. Using the previously explained technique, they were able to construct a signal for the heart rate and breathing rate in the time domain, from which they then derived the features necessary to perform emotion detection. The features extracted from both signals were based on the work of [43]. They then classified the subjects in a two-dimensional model, whose axes were valence and arousal,<sup>1</sup> based on four emotions: sadness (negative valence and negative arousal), anger (negative valence and positive arousal), pleasure (positive valence and negative arousal), and joy (positive valence and positive arousal). The obtained results were compared with two different techniques: ECG-based and vision-based emotion recognition. They obtained results slightly worse than those of the ECG-based technique, and better results than the ones from the vision-based technique, for three of the four emotions under consideration.

Other studies, like the work of Lee et al. [50] and Lu et al. [54], focused on the detection of the breathing rate and heart rate, respectively, by employing microwave sensors to perform contactless monitoring. The characteristics of the microwaves allow them to work at some distance away from the human body, and even go through clothes, being then reflected by the human body. The work in [54] also had as its primary goal the measuring of heart rate variability based on this technique. Several recordings with 5-minute duration were performed, for both the microwave sensor and for an ECG. The tests were made with 16 different male subjects aged between 19 and 27, and no significant difference between the two techniques was found in both the frequency and time domains, as well as in non-linear dynamic analysis of heart rate variability measurements. Although the results were promising, all the tests were performed under controlled research conditions, and all subjects were healthy. The authors claim that this work can prove to be a practical alternative to ECG for heart rate variability analysis, as it eludes the negative aspects of wires and body sensors.

Another work that addressed the use of signal reflection for determining a person's breathing rate is the work of Nandakumar et al. [60]. In this work, the authors present a solution for detecting sleep apnea events with smartphones. To achieve this, they developed a solution where an off-the-shelf smartphone is turned into an active sonar system that emits frequency-modulated sound waves and captures reflections. In this work, they use the same approach employed in [5], that is, the detection of the minute chest movements caused by breathing. The speaker of the smartphone is used to emit an FMCW signal with frequencies between 18 kHz and 20 kHz, and the reflection of these signals is then caught by the smartphone's microphone. This range of frequencies is very close to the threshold of the human hearing and, as such, this work also addressed the issue

---

<sup>1</sup>Both of the definitions for arousal and valence can be found in [71] along with more information in the 2D model.

of whether the sound was audible. They found out that the majority of people were not able to hear it, while a small minority were able to hear a small noise in quiet environments. This aspect is also important when considering an unobtrusive solution, since an acute sound can generate discomfort and/or become bothersome and intrusive.

Additionally, in [60], the **FMCW** technique was used for monitoring more than one person at the same time, as long as they were separated by at least 20 cm. The authors claim that this technique works even when the subject uses a blanket, and that the system can maintain its accuracy even when the blanket is some centimeters thick. The system obtained a mean estimation error of 0.11% for one person with the device up to a meter away from the subject. The error increased with the distance to the device and when monitoring more than one person. This work also presented the results of a clinical trial with 37 subjects, where the solution was used for monitoring the breathing rate, as well as for detecting and classifying apnea events. The system correctly classified 32 out of the 37 subjects, for four types of sleep apnea. An average detection error of 1.9 apnea events per hour was achieved. This shows that this type of technique can also be leveraged in clinical context.

Other authors also proposed similar solutions based on commonly available devices. In [85], the authors proposed a system based on a normal microphone and a computer speaker. Additionally, in [84], the authors also proposed a similar system based on Smart Speakers. These devices are becoming more common as we witness an increasing trend in the use of smart devices and in the deployment of smart home solutions. This also addresses the fact that we may not need new devices to power humans' unobtrusive sensing, as many of our existing devices can be used or modified to work as sensing solutions.

Additionally, there are also some commercial solutions that use this sensing technique, such as [69]. While this solution primarily focuses on detecting objects through walls, the development version of this solution already offers an off-the-shelf breathing **Application Programming Interface (API)** that is able to use this board to detect people breathing, using **FMCW**. The company responsible for the device in [36], which is mainly used for radar applications such as presence detection and security, also proposed that this device could be used for monitoring the heart rate and breathing rate. This is quite remarkable, since the device is roughly the size of a coin. Furthermore, even smartphone companies are starting to see these radar solutions as an opportunity. The latest version of Google's smartphone flagship, the Google Pixel 4, has an incorporated radar sensor [37]. Although details on the availability of **APIs** for developers have not yet been disclosed, this could pave the way for new mobile applications for healthcare.

**3.2.2 Interference-Based Techniques.** Several techniques use artificial signals and the signal-body interference phenomenon in order to capture humans' physical or emotional states. The human body allows high-frequency signals to pass through it. However, the signal that enters the body is different from the one that leaves it. In Figure 5, we can see a representation of this phenomenon, where the signal that enters the body is modulated by its minute movements. The signals can be attenuated and suffer interference, which affects one or more frequency components.

In [41], the authors leverage this phenomenon, by using a single commercial off-the-shelf transmitter-receiver pair to monitor peoples' respiratory rate, through the interference caused in the **received signal strength (RSS)**. In this work, the authors used a single transmitter node and a single receiver node, where the receiver antenna is also connected to a real-time spectrum analyzer to obtain the baseline. The tests were conducted in a sleeping scenario where one person was lying on a king-size bed, and the antennas were placed at each side of the bed, two meters apart; the antennas were also placed at 0.2 meters above the chest. This careful positioning of the equipment creates the best signal-to-noise ratio scenario. Additionally, pre-filtering was also exploited to increase

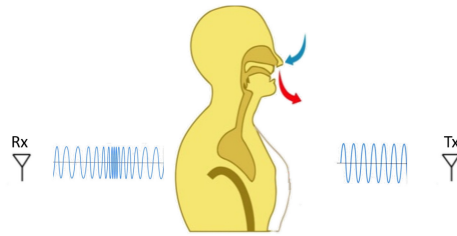


Fig. 5. Illustrative scheme of the interference phenomena caused by minute movement of the human body in radio frequency signals.

the signal-to-noise ratio in the **RSS** measurements. The obtained results were compared with those of a real-time spectrum analyzer, to prove that the system could obtain comparable results to those of a system that costs three to four times more. The system was able to achieve a mean absolute error as low as 0.12 breaths per minute. In this work, however, the authors obtained the breathing frequency in the frequency domain using **Power Spectral Density (PSD)** and, therefore, they were unable to obtain information in the time domain. Furthermore, in instances of time where the subject is moving, breathing estimation cannot be performed. The authors believe that the work can still be enhanced by exploiting channel diversity to improve the breathing detection ability.

In [41], the authors stated that they were building on the work done in [67] by Patwari et al., since they believed that by only using two sensors this would reduce the system complexity and increase its feasibility. However, due to the complexity of the system in [67], other sensing opportunities emerge. In this work, in one of the experiments the authors used a 33-node wireless sensor network to monitor an apartment of  $7 \times 8$  meters. Although the obtained results were less precise than the results in their follow-up work, the use of several nodes allowed them to perform breathing estimation and location of people in two dimensions. They were able to detect the location of a breathing person with a mean error of 2 meters. Although this value seems high for applications that require a precise location, these techniques can be useful in conditions such as the ones in search and rescue missions after an earthquake or a tsunami, where information about a relative location (even with a 2-meter error) can be useful.

Other authors have also based their approach on the fact that the human body minute movements interfere with radio signals, in a way that is related to vital signs. In one of those studies [52], the authors proposed the use of off-the-shelf Wi-Fi devices to track human vital signs. In this case, the system is based on the use of the **channel state information (CSI)** of Wi-Fi signals, which is more suitable for this task than **RSS**-based approaches. The reason behind this is the fact that **RSS** is already an aggregation of all of the sub-carriers' signal strength in order to mitigate interference in the signal. The sub-carriers are affected differently depending on their frequency and, as such, there are some sub-carriers that suffer more visible interference caused by human movements. Furthermore, recent comparative studies prove that Wi-Fi **CSI** measurements provide more robust estimations of breathing rates when compared to other radio frequency measurements [31].

In this work, only one laptop and one **access point (AP)** were used. In a first instance, the **CSI** data is collected, by using a **CSI** tool in the laptop [1]. That data is then filtered, and the system runs an algorithm that takes into account moments during which the person moves and moments during which the person is almost static. Similarly, to what happens in the studies that use the interference phenomena to detect human movements, this technique requires the **signal-to-noise ratio (SNR)** to be high. As such, moments during which the subject moves around or moves a part of the body prevent the detection of the breathing rate and heart rate. Although the authors claim that this system can be used in any scenario where quasi-static moments occur, the

system was only tested with people in their sleep. In this particular scenario, events like turning over in the bed, or getting in and out of the bed, can contain precious information about the quality of one's sleep and, as such, the authors use this information as well. After filtering the movements, the system uses the filtered signal to estimate the breathing rate and heart rate.

In the mentioned study, the sub-carriers with greater variance in signal strength were selected, as those are the ones that are more affected by the human body in the frequency domain. After the signal is filtered, the peaks of the signal are detected for each of the selected sub-carriers. The mean value for the location of each peak in all sub-carriers is computed and the respiration rate estimation is then given by calculating the number of peaks in 1 minute. In this work, it was also demonstrated that it is possible to detect the respiration rate for two people while sleeping in the same bed, without increasing the number of devices. In order to achieve this, the authors used the **PSD** technique. A strong sinusoidal signal, such as the respiration cycle, generates a frequency peak corresponding to the period of the sinusoidal **PSD** signal. When two people are being monitored, two strong frequency peaks will appear in the **PSD**, corresponding to the breathing rate of each person. By applying this technique to each of the selected sub-carriers, the authors used the K-means technique in order to find the two peak clusters that correspond to the breathing rate of each person.

Using the same approach, it is also possible to detect the heart rate of a person. The movements caused by the heartbeat are smaller than those caused by breathing and, as such, are more difficult to detect. Nevertheless, the heart rate is higher than the breathing rate. This means that in the frequency domain the heart rate will be represented in higher frequencies than those of the breathing rate and, thus, it is possible to separate both signals in the frequency domain. The heart rate also generates a strong sinusoidal signal, which means that the **PSD** technique can then be applied to find the stronger component in all of the sub-carriers' signals. The mean value for all of those components can be calculated, which corresponds to the heart rate. Although in this work the authors do not demonstrate it, they also propose that the same approach used for the breathing rate can be used to obtain the heart rate for two people at the same time. However, contrary to what happens with the respiratory rhythm module, it is not possible to obtain a signal that corresponds to the cardiac rhythm in the temporal domain.

Most of the techniques that use the interference phenomenon are only applied to monitor one or two persons at the same time, as it is difficult to interpret interference without knowing the signal propagation path. The work in [86], however, leverages **CSI** phase difference data to intelligently estimate the breathing rate of several people at once. The proposed system applies tensor decomposition, namely, canonical polyadic decomposition, to obtain multi-persons breathing rates. In this study, they demonstrate that, while normal **CSI** cannot be used to detect more than two persons' breathing rate at the same time with accuracy, the proposed technique can be used to estimate the breathing rate of five people with accuracy. The system was tested for different temporal size windows and sampling rates, obtaining better results when both attributes increase. The system was also tested in three different situations with several line-of-sight and non-line-of-sight situations, since the technique works through walls and objects as well. The system obtained an absolute estimation error of 0.9 breaths per minute even through walls for one person, with that value increasing for two breaths per minute for five people within the same confined space.

The **CSI** technique can also be leveraged for detecting emotion, as happens in [27], where EmoSense was presented. Contrary to what was done in [93], where physiological signs unobtrusively extracted were used to perform emotion detection, in this work the physical expression of the subject was captured in order to determine emotions, through **CSI** measurements. There were three major findings for this study: firstly, **CSI** was indeed able to capture emotional expression; secondly, the performance of the system depends on the experimental setup; and thirdly, the

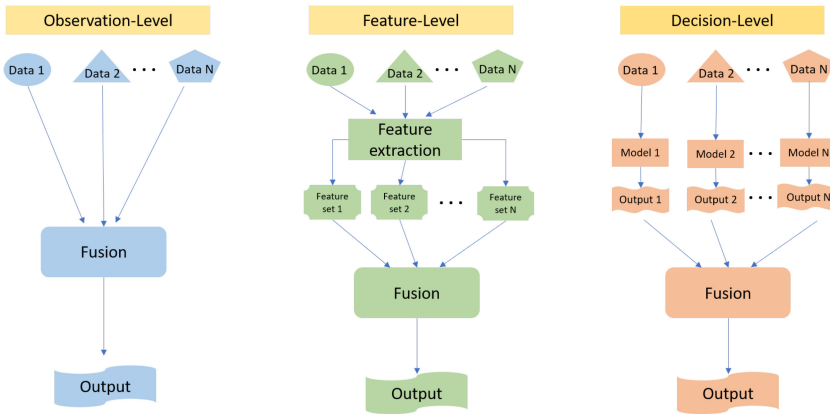


Fig. 6. Types of data fusion.

performance is person-dependent. The system was based on a data-driven architecture, where the CSI measures were sent to a server. Classification models were then used to infer one out of four basic emotions, namely, happiness, sadness, anger, and fear. The created system was compared to the sensor-based approach and was not able to achieve the same performance. Specifically, the sensor-based approach elicited an accuracy of 95.83%, while EmoSense only reached an accuracy of 80.48%.

### 3.3 Data Multimodality

All of the work presented in the previous sections explores monomodal sensing, that is, the use of a single source of data to infer the humans' physical or emotional states. However, in order to achieve a more robust system, work that fuses/mixes different sources of data should also be considered. As can be seen in Figure 6, when considering data fusion techniques there are three main levels of fusion, namely, observation level, feature level, and decision level [28]. Observation-level fusion is made directly on the input of the system, that is, the raw data are directly combined. On the other hand, feature-level fusion explores the preliminary extraction of several representative features from each input. Lastly, decision-level fusion occurs when an output or pre-output from independent models is first obtained and then fused to derive a final output. In this area of unobtrusive sensing, most work is still experimental and, as such, focuses only on developing the sensing technology or exploring new algorithms. In this field of research, work that uses multimodal approaches is scarce. However, if we consider pieces of work that do not focus primarily on detecting physiological and emotional states, we can find several approaches to data fusion that could also be reused for this purpose. As such, in this section we explore available multimodal approaches to unobtrusive sensing, as well as relevant works to data fusion with the explored types of data sources.

One study used a multimodal approach with sound and image-based techniques to predict humans' states. In [7], the authors propose a multimodal system to predict emotion in human-robot interaction. The proposed system is based on decision-level fusion. In this case, emotion detection through voice and through facial expression detection work as independent systems, and each of these gives a prediction for the human emotion. These predictions and their respective confidence intervals are then processed by a decision rule mechanism that elicits a final emotion classification. This work also demonstrates another dimension to the applicability of this kind of system, showing that they can also be used in emerging and fast-evolving fields such as robotic systems.

The authors tested their work on a real case scenario where a robot interacted with 16 people, one at a time, and asked them to express several emotions. The tests demonstrated that the results obtained with the multimodal approach outperformed both classifiers when considered as standalone solutions. Furthermore, other studies proved that some emotions such as anger, happiness, surprise, and dislike are detected more easily by visual appearance, while other emotions such as sadness and fear were more evident through speech [21]. These results indicate that multimodal approaches can lead to better emotion detection approaches.

Although [7] only applied the multimodal approach to emotional states inference, we can envision the use of decision-level fusion in order to combine the approaches taken in two or more monomodal systems, such as the ones presented in the previous subsections. In decision-level fusion, the sensing mechanisms work independently until they reach a classification or inference. At that point, the output of each sensing mechanism is fed into a decision system that weights each output and reaches a final classification or inference.

Other authors have created a mixed approach of speech and thermal cameras to infer humans' emotions during speech. In [90], the authors proposed a system that infers human emotion through vowel judgment and facial recognition, using thermal images. The main objective of the proposed system was to empower applied robotics with the capability of detecting humans' emotions as they speak even under varying lighting conditions. In this work, data fusion is made at feature-selection level, as the authors mainly use speech recognition to track the best frames from the thermal feed to perform emotion inference. The authors tested their system with three subjects with different genders and ages and also tested the system in the presence of glasses. The test results demonstrated a 79.8% accuracy while detecting facial expressions of the three subjects.

As already mentioned, work exploring multimodal approaches is still scarce. Nevertheless, it is possible to find approaches that fuse the same types of data with different purposes. This is the case of [76], where **RSS** and **CSI** are fused together to build a more robust localization model. In this work, the authors used a feature-level fusion, to extract features from both data sources. The authors also proved that the multimodal system was able to surpass both a system based only on **RSS** and a system based only on **CSI**. As was presented above, there are works in the interference taxon that use **RSS** or **CSI** to infer the breathing rate of one person, for instance. As such, it is also possible to envision a system with a similar multimodal approach as the one in [76] to unobtrusively sense humans' physiological signals, that could also show performance improvements when compared to a monomodal approach.

In addition to combining different data streams using one of the mentioned data fusion techniques, it is also possible to use different techniques in a complementary way. For instance, techniques from the image taxon that use facial video feeds for heart rate monitoring are less affected by body movements and, conversely, more affected by rapid head movements. On the other hand, techniques from the reflection and interference taxa are more accurate when the user is stationary or quasi-stationary and are quite affected by rapid body movements, whereas they are unaffected by head movements. Thus, it is possible to envision a system that uses both modalities to address these limitations. One such approach is used in [42], where the authors proposed a system that uses both **CSI** and a video camera to generate a video stream. The overall aim of this system is to be able to generate video frames from **CSI** at points in time at which the video camera fails or is unavailable, for instance during an attack or while the camera line of sight is blocked by an object. In this work, the authors also explore a different approach, which only uses the video feed to train a deep-learning model that works with the **CSI** as input. Complementary techniques such as the ones that are mentioned above can, thus, be valid approaches for unobtrusive sensing.

In [18], the authors proposed a feature-level fusion method that uses several physiological signals, such as the heart rate and respiration rate, to monitor drivers' stress. This shows the

Table 1. Distribution of the Reviewed Unobtrusive Sensing Solutions/Approaches per Taxonomy Branch and Application Field

	Natural		Artificial	
	Sound	Image	Reflection	Interference
Physical States	[39, 40, 73]	[45, 46, 51, 56], [26, 32, 44, 68] [2, 82]	[5, 50, 54, 60], [84]	[41, 52, 67, 86]
Emotional States	[7, 33, 57, 90]	[7, 24, 90]	[93]	[27]
Multimodality	[40]	[2, 44, 82]	[5, 93]	[52]
Medical Field	[73]	[82]	[60]	--
Sleep Monitoring	[39, 40, 73]	[32]	[60]	[41, 52]

possibility to fuse several physiological signals, and we believe that this approach could also be leveraged in the field of unobtrusive sensing. For instance, there is the possibility of fusing the estimated heart rate and breathing rate from different sensing techniques. Additionally, given the fact that while we are driving our body movements are limited, many of the techniques presented above are quite suited and, as such, this could also be one of the areas that could greatly benefit from the use of unobtrusive techniques. Furthermore, there are also studies that proved that using multimodal physiological data can also be leveraged in deep-learning models to detect other emotions, leading to results that are far better than those of single modalities [89].

## 4 DISCUSSION

### 4.1 Overview

The distribution of the various solutions/approaches presented in the previous section can be seen in Table 1. In this table, we identify not only the taxonomical branch that applies to the work, according to the taxonomy proposed in Section 2, but also the application field.

Most techniques aim at sensing the physical state of the subjects, such as heart rate, breathing rate, position, and so forth, either by using natural or artificial signals. A smaller number of studies/proposals address the assessment of emotional states.

Furthermore, when considering multimodality, only seven of the reviewed studies explored it, and most of these use multimodality for the monitorization of heart rate and breathing rate. Moreover, only [93] explores multimodality for monitoring both physiological signals and emotional states.

As we can see in Table 1, only three proposals focus on the goal of using unobtrusive solutions for clinical purposes. Furthermore, one of them focuses on bradycardia detection in neonatal monitorization, while the other two focus on sleep apnea monitorization. Sleep monitoring is, in fact, one of the major focuses in this field of research. This is partly due to the lower intensity and frequency of movements of people while they sleep. Most of the reviewed techniques are highly



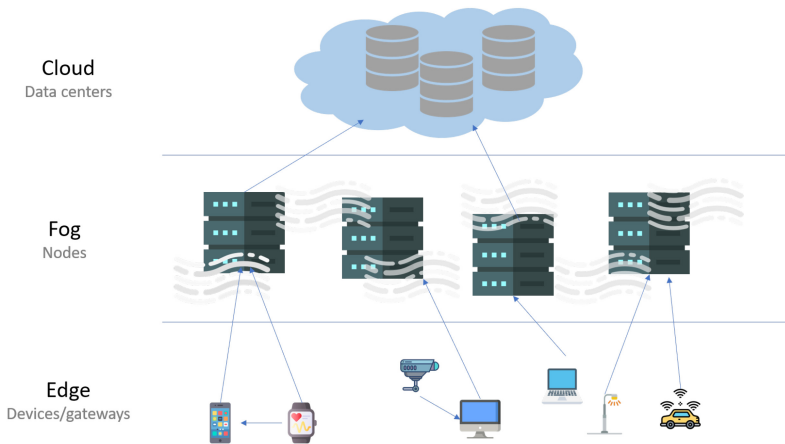


Fig. 7. Cloud, fog, and edge representation.

sensitive to noise generated by large movements. By concentrating on sleep monitoring, we are, in fact, avoiding the problems, difficulties, and limitations caused by such movements.

## 4.2 Computing Architectures

Distributed systems can use one or more architectural approaches, namely, edge, fog, or cloud, as can be seen in Figure 7. Cloud computing is a model for the provision of remote resources over the Internet, providing high computing power, elastic storage capabilities, and high scalability. Fog computing, in turn, offers computing capabilities closer to the network edge, on a smaller scale and with less scalability, but with considerable gains in terms of latency. Lastly, edge computing refers to performing data processing on devices inside the edge network, i.e., next to end user devices, such as sensor nodes that acquire the data, or gateways to which sensor devices are connected.

As previously stated, several proposals focus on the development of new sensing techniques, and little considerations are made about optimal architectural designs. That is, most pieces of work are made in experimental setups and the data processing is made offline. However, when considering similar work outside the field of unobtrusive sensing of physiological and emotional states, it is possible to find approaches that take into consideration the underlying distributed computing architecture. As such, in this subsection we will overview the existing designs and address what we believe would be the optimal case for each of the taxonomy nodes.

In [24], the authors proposed an architecture for the regulation and detection of emotions, through the fusion of behavior data, emotional data, and valence/arousal data. This work was based on the software architecture from [23]. In order to achieve real-time feedback, and since the sensor processing for emotion detection tends to be the more demanding part, the authors propose that the information from each sensor should be processed in a dedicated node. The authors also propose a decision-level fusion of the heterogeneous outputs from the sensor nodes. As such, they proposed the existence of a central node in the system, dedicated to this functionality. This central node is also in charge of communicating with the actuation nodes, which, in turn, interact with the physical devices in the environment. The authors describe their system as a hybrid-distributed system and do not specify the location of their central node, assuming a mixed approach of edge and cloud/fog. We also believe that this is the best approach for this type of system, as it relies on performing pre-processing and dedicated inference in each sensor node (i.e., at the edge), while fusing and dealing with the more complex models in a more robust infrastructure, at fog/cloud

level. This approach has also been widely used in other systems such as smart cities [22], and has proved to be one of the best approaches in terms of system performance and scalability.

Additionally, most of the works in the image taxon that use video feeds or thermal video feeds could also benefit from performing processing sensor data in a dedicated node. Transferring video feeds can lead to considerable traffic load and decrease the chance of performing real-time inference. Although it is not trivial, it has been proved that processing video feeds in dedicated nodes [92] is feasible. However, other studies also indicate that the hierarchical architecture <camera/device, private cluster, cloud> is the most common and the only feasible approach for large-scale analyses of video feeds [35]. We believe that in the case of unobtrusive sensing, since some systems may be dealing with preventive detection of pathologies, which are related to the subjects' history, the best approach should also be a mixed approach of edge-fog-cloud. In this case, edge nodes should pre-process and extract the more relevant features from the acquired data, while upstream nodes should deal with computing the history, training more complex models, and making inferences. However, in some systems such as those of a clinical nature (e.g., [2, 44, 82]), privacy is a must, and for those systems we believe that the correct architecture should be edge-fog, to maintain the personal data in architectural levels that are under direct control of data owners (i.e., users) or primary data processors (i.e., medical staff).

As mentioned in Section 3, one of the fields in which techniques for unobtrusive sensing have been used is the field of robotics and human-computer interaction. The systems developed with this purpose try to create the illusion of interacting with other human beings and thus the response needs to be provided in real time. In [7], the authors proposed a robotic system to detect emotions from facial expressions and speech in human-robot interactions. In this system, the model was created offline, from previously collected data, and then inserted in the robot. Thus, all inferences from the system were directly made on the edge. However, we believe that a more complex robotic system for human interaction should also benefit from a mixed edge-fog/cloud approach. A complex robotic system should be able to perform reactions to user inputs on the fly, but it should also be able to keep track of all the users and the history of each user interaction, and learn from them. As such, the most appropriate approach should be to perform inferences based on existing models and additionally retrain those models in a fog or cloud infrastructure.

Smartphones have also been widely used as a tool for creating unobtrusive sensing solutions, as seen in Section 3. Most of the works that are based on smartphones use an offline approach, that is, the data is collected by a smartphone and then processed at a later date on a computer. However, in [45], the authors also developed a mobile application capable of detecting the user's heart rate using a 20-second video feed directly on the smartphone. Other authors also demonstrated that it is possible to run complex models on current smartphones [16]. We believe that techniques based on smartphones should also use a mixed edge-cloud architecture, where pre-trained models and simpler processing tasks could run directly on the smartphones, while training of more complex models could be offloaded to a cloud server.

In [67], the authors propose the use of a wireless sensor network to monitor the breathing rate and find the location of people. Although in this work the authors processed the data offline, in their architecture they used a sink node that collected and stored all of the data. In this system, we can envision a fog computing architecture, where a more resource-capable node on the fog would perform the same tasks as the current sink node and, concurrently, process all data and product inferences. Additionally, in [40] the use of two synchronized geophones could also resort to a fog architecture, as the feed from both devices is needed in order to perform the inference.

Last but not least, it should be noted that all of the reviewed works that used CSI were also made offline. However, in [27] the authors proposed an architecture where the data would be sent to and processed in a remote server. Additionally, there are works that show that it is possible to

Table 2. Possible Computing Architectures Applicable to the Reviewed Works

	Natural		Artificial	
	Sound	Image	Reflection	Interference
Cloud-Fog-Edge	[7, 90]	[7, 51, 56, 90], [26, 32, 68]	----	----
Cloud-Edge	[33, 40]	----	[60, 84]	[27, 52, 86]
Fog-Edge	[40]	[2, 44, 82]	[5, 50, 54, 93]	[27, 41, 52, 67]
Edge	[39, 73]	[45, 46]	----	----

run complex models directly in the access points [80]. As such, when considering works based on CSI, we believe that a mixed approach edge-cloud/fog could also be taken.

Based on the considerations made above, in Table 2 we provide an insight on the possible computational architectures, applicable to the various works in unobtrusive sensing, reviewed in this article.

### 4.3 Open Issues and Challenges

There are several open research lines and opportunities to further explore existing unobtrusive sensing techniques and improve them. Table 3 will guide us through the discussion of the most relevant open issues and challenges, in what concerns unobtrusive sensing of humans. For this, in the first column we identify the main open issues and challenges in this area, namely, noise reduction, multi-person monitorization, emotional states detection, use of machine learning techniques, privacy, access to open datasets, use in the medical field, use of data fusion, and adoption of HiTL-CPS approaches. For each of these, in the second column of Table 3, we identify the predominant fields of application according to the proposed taxonomy.

All of the presented techniques are subject to noise, which may lead to significant errors. Sound-based techniques may be affected by static noise and by environment noise, while image-based techniques can suffer from interference caused by lighting and moving artifacts. In the artificial branch of the taxonomy, we observe noise created by movement or by obstructions for both the reflection-based and the interference-based techniques. As we have seen from the review of the state-of-the-art, some approaches already explore these limitations and compare results with and without noise sources (e.g., [51, 52]). This noise-prone nature of the used sensing mechanisms has an impact on the achievable results, and this is one of the reasons why some studies focus on sleep monitoring or on near-static environments (as is the case of neonatal monitorization). We believe that this is one of the fields that should be further explored, by exploring new techniques to increase system performance in noisy environments, by filtering noise sources, finding new ways to selectively sense the phenomena, or even by adopting a multimodality approach, as previously discussed. This, in turn, would open new opportunities for extending the applicability of existing work to new fields.

Another challenging area for which there is considerable lack of work is the determination or assessment of emotional state, especially in what concerns the use of artificial signals. Most approaches focus on physiological aspects only, and do not address emotional states. Furthermore,

Table 3. Open Issues, Challenges, and Most Relevant Areas

<b>Open Issues and Challenges</b>	<b>Predominant Fields</b>
Noise Reduction	Natural Branch: Sound Taxon Artificial Branch
Multi-Person Monitorization	Natural Branch: Sound Taxon Artificial Branch
Emotional States Detection	Artificial Branch
Machine Learning Approach	Natural Branch: Image Taxon Artificial Branch: Inference Taxon
Privacy	Natural Branch: Image Taxon Artificial Branch
Open Datasets	Natural Branch: Physiological Data Artificial Branch
Medical Field Exploration	All
Multimodality and Data Fusion	All
Standardization	All
HiTLCPS Approach	All

several approaches that do address the emotional state of people, extract the required data from physiological data (e.g., [93]). This, in turn, shows that several approaches developed for the purpose of detecting physiological state can also be repurposed to address emotional states. Several studies point to a positive correlation between positive emotional states and physical well-being [75], while others present results that indicate that mixed emotions with a balance between negative and positive emotional states can be beneficial to the physical well-being [30]. Nevertheless, information about human psychic state and emotions can be as important as physiological state. Furthermore, we believe that in the future this kind of solution could lead to the possibility of analyzing people's mental state at a society scale, and as a group indicator. This could be particularly important, for instance, in a pandemic situation, where the emotions and mental states could alert government to the negative effects of certain decisions and restrictions. Thus, we believe that the exploration of unobtrusive techniques to monitor the emotional states of people is one of the most important, promising, and challenging research fields.

Most of the reviewed approaches can only deal with one or two persons at the same time, with [86] being the exception, as it is able to simultaneously monitor up to five people. Nevertheless, some image-based techniques are applicable to more than one person at a time, such as the techniques described in [46] and [51]. Despite this, approaches that address the monitorization of multiple persons typically lead to results that deteriorate as the number of monitored subjects increases. As such, we believe that this is still one of the areas that should be further explored. We believe that techniques that explore tagging/identification and tracking mechanisms inside the sensing environment could contribute to improve the results of the sensing solutions, and lead to more reliable and robust solutions. Furthermore, the identification of people could grant the sensing systems access to historic information of a certain subject's physiological and emotional states. This, in turn, could be used to take more accurate long-term decisions, or even, for instance, project a future outcome for a subject's relative.

As stated before, many techniques are prone to noise and, as such, in most cases the solution is to apply signal filtering. Although in most situations this is indeed the best approach, the possibility of using machine learning techniques can also be explored. We believe that approaches such as the use of deep-learning in the case of image-based sensing would greatly enhance the system performance. This approach could also be used in **CSI**-based techniques, where large amounts of data are produced. Additionally, deep-learning is now accepted as a valid approach for systems that use **CSI** for a variety of purposes, such as activity recognition [34, 94] or occupancy detection [53]. In [42], it was also proven that it is possible to implement a domain translator to generate video frames from **CSI**, proving the similarity of these types of data. We believe that future research could benefit from the work done in deep-learning in other applications of **CSI**, and from the extensive work done in deep-learning with images and videos in order to improve many of the reviewed solutions.

The unobtrusive sensing techniques presented in this article open up exciting prospects in what concerns better and more affordable health care and e-health systems. However, they also raise several concerns in terms of privacy and security. Questions such as "How do we control who reads our vital signs?," "How can we keep our privacy, when people can see us and track our movements even through walls?," or "How can we make sure that these technologies are not misused?," and many more, must be answered. Unobtrusive sensing techniques create privacy vulnerabilities like no other before. For instance, when dealing with the possibility of tracking movements and vital signs through walls, how can we make sure that this is not used by burglars to rob houses or by governments to keep track of our movements and actions? Furthermore, this could also be used by health insurance companies to keep track of our clinical conditions and deny insurance in high-risk cases. Furthermore, when considering computing architectures many of the presented

solutions could leverage the use of distributed edge computing. This, in turn, could also lead to several security and privacy breaches [61]. Thus, assessing data privacy in terms of the different solution architecture options is also important. As we previously discussed, the medical field is one of the fields that would leverage more from the employment of unobtrusive solutions. However, it is also one of the fields where data privacy is of the utmost importance. Other authors have explored the concerns of privacy in the medical field and proposed new approaches to lead with attacks [58], but we believe that new solutions should be proposed with the aim of specifically including these new sensing techniques. As such, we firmly believe that privacy assurance is one of the most important open issues in this field and should be further explored in future research.

One important drawback in this field of research is the lack of open databases/datasets. Although there are several image and video repositories and datasets, those are not labeled for the recognition of physiological states. Apart from image-based and sound-based emotion detection techniques, to the best of our knowledge there are not any open datasets in this field of research. This creates a significant drawback in this field of research, as this kind of system implementations have high complexity, and the nonexistence of datasets forces researchers in this area to implement their own systems from scratch, and perform trials for data acquisition, which requires considerable effort. Clearly, a collective effort should be made to develop better data acquisition protocols, with well-known and easily replicable factors and constraints. This would allow for independent, yet comparable data acquisition actions to occur, and would lead to the creation of common datasets, largely available to the research community. Additionally, many of the studies presented in Section 3 were performed with a small group of subjects. The existence of larger datasets would allow the validation of these techniques in a more reliable and meaningful way.

Although some of the reviewed approaches address the detection and monitorization of medical conditions, most of them focus on physiological data acquisition. Moreover, although several approaches allow the monitoring of the heart rate signal, they do not compare their results with the results obtained through electrocardiogram, which is the golden standard for heart rate monitorization. In order to move from the labs to clinical scenarios, we need more evidence that the signal obtained with those techniques is reliable, maintains the information in the time domain, and can be correlated with the clinical metrics. As such, we believe that, as future work, researchers should make an effort to evaluate their solutions against medically approved and certified methods and equipment, as opposed to ad-hoc, and non-approved methods and experimental equipment, such as wearables. This is one of the most pressing research opportunities in this area, as only then we can bring these solutions to real system level, and apply them to improve the medical field. For instance, in [59] the authors propose the use of robots to monitor the breathing rate and aid in the monitorization of patients with infectious diseases such as COVID-19. The use of techniques such as those depicted in the artificial branch of our taxonomy, that can retrieve these values even through walls, could greatly help in these tasks. Thus, the development of these techniques and the effort to get them certified and approved for their use in the medical field could be an important step in fighting future pandemics.

Additionally, as mentioned in Section 3.3, the use of multimodality approaches could bring forth more robust systems. This is still a largely unaddressed field, and, as such, this should also be subject to future work in the area of unobtrusive sensing. In the work that we reviewed in Section 3.3, feature-level fusion seems to be the more common approach when creating multimodal approaches. Additionally, some systems use decision-level fusion, where several sub-systems work separately to generate their own outputs, and a decision mechanism weights each output to generate a final classification. This allows the involved systems to mitigate each other's errors (e.g., speech is better suited to detect certain emotions while facial expressions are better suited for others). Although these approaches could raise other issues, such as dealing with data

heterogeneity and data synchronization, we believe that exploring several levels of data fusion can be an interesting opportunity in the field of unobtrusive sensing of humans' emotions and physiological states. The multimodality of sensing techniques could also be used to deal with some of the other open issues, such as creating systems to detect multiple persons at the same time. Cameras and video feeds have proven to be one of the best solutions to detect, identify, and track people [83]. As such, we believe that in the future the combination of video feeds and unobtrusive technique could lead to systems that are more accurate and can be used to track multiple persons at the same time.

In addition to data fusion, there is also the option of using several systems to complement each other. For instance, the works in the artificial branch of the taxonomy are less precise when there is significant body movement, but are unaffected by head movements or temperature fluctuations, while works from the image taxon exhibit the opposite behavior. When considered separately, these techniques may lead to windows of time during which it is not possible to perform inferences. However, an approach that combines both types of techniques may overcome the mentioned problems. As such, we believe that this should also be an issue to address in future work.

Another issue that should be addressed in the field of unobtrusive sensing, should be the issue of standardization. Standardization helps to bring consensus between industry, the research community, and organizations in general. It also defines benchmarks that set performance minima for these technologies. Additionally, standardization helps with building the trust of customers in products and systems/services, which, in turn, will help the dissemination of these technologies. There are already some approaches to this issue. For instance, the new 6G design proposals already include the standardization of the cellular signal as a sensing method [88]. In this respect, it is proposed that the Radio Frequency sensing capabilities should be natively integrated into the system design of 6G, and that base stations should use the same spectrum for both communication and sensing purposes.

The standardization of this type of technique is also an important step toward answering the issues we raised about privacy, since privacy enforcing measures could be built into the standard. Furthermore, standardization could also improve the acceptance of these techniques in the medical field. For instance, the use of medically approved protocols for communications such as HL7 and DICOM [29], could be built into the standards for these new sensing systems, making their integration with existing medical information systems easier, and improving their acceptance by managers of these systems. As such, we believe that standardization is one of the open issues that should be addressed in order to improve the field of unobtrusive sensing.

Lastly, as mentioned in the Introduction, systems that take into consideration human intents, actions, emotions, and physical state, i.e., **HiTLCPS** systems, maximize the benefits for the users. In order to develop such systems, we need to close the loop and actuate on humans, Although some approaches already show efforts in this area, as is the case of [24], where the authors' goal is not only to detect emotions but also to regulate them by automatically changing the environment that surrounds the user. Most of the work in the state-of-the-art still ignores this component.

In Figure 8, we present a possible high-level framework for unobtrusive **HiTLCPS** sensing. As we can see in the figure, the framework is composed of (i) a physical interface to the real world, which could include one or more modalities according to the proposed taxonomy; (ii) a state inference phase, where past, present, and future states are inferred; and (iii) an actuation phase to close the loop, where the actuation should involve both humans and the environment. This is a framework that could serve as a next step for many of the solutions addressed in this review, as most of them had the goal of developing sensing techniques and were not intended as fully operational and/or readily deployable systems. However, we believe that this should be the next step for many of the addressed solutions, as the interactions of the users with these systems and their perception

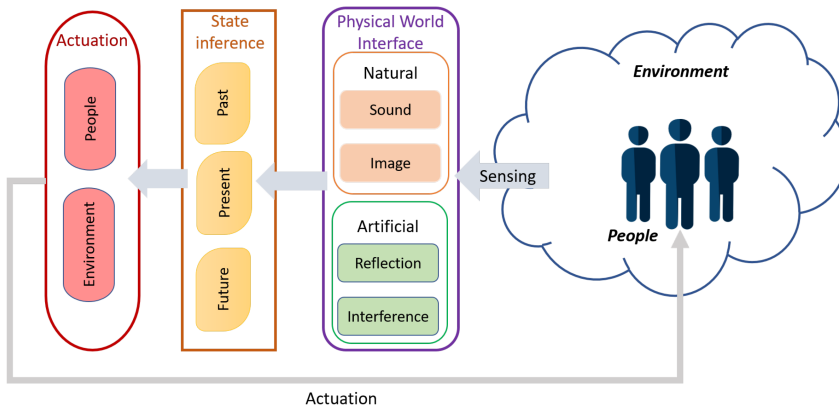


Fig. 8. Proposed unobtrusive HiTLCPS sensing high-level framework.

of the system itself is an essential step to better understand and achieve true unobtrusiveness. In the future, this kind of system, using unobtrusive solutions, will be part of several of our society's systems. We envision that they will be in healthcare systems, security systems, or even in our industry, enhancing human beings well-being, and could in turn enhance any system of which they are part of. As such in theory, unobtrusive sensing of human beings' physiological and emotional states could improve every existing system.

## 5 CONCLUSIONS

In this article, we provided a comprehensive review of the state-of-art in unobtrusive sensing of humans' physical and emotional states. In order to support our analysis, we proposed a taxonomy for human sensing that considered the type of sensing (obtrusive or unobtrusive), the type of signals used for sensing (natural or artificial), the specific signals of methods for sensing (sound, image, interference, reflection), and the number of signals to be considered (monomodal or multimodal).

After identifying and discussing the most relevant techniques and approaches and classifying them according to the taxonomy and the application area, we identified a set of open issues, challenges, and concerns. Although there are some surveys in specific fields such as healthcare systems [72], human monitorization solutions [3], Internet of Things [78], or even surveys that review the work around a specific sensing technique (for instance, the work presented in [55], where the authors address the use of Wi-Fi networks' CSI for a variety of sensing goals), as far as we know, this is the first survey to address the state-of-the-art of physically unobtrusive sensing of humans' physical and emotional states.

The reviewed unobtrusive techniques and approaches show promising results when compared to their obtrusive counterparts (e.g., wearables). They also show that there are many approaches that can be used to create systems that are able to monitor humans in an unobtrusive way. Nevertheless, several concerns were raised that should be addressed in future research. We believe that in the next few years the identified research lines will lead to substantial developments in unobtrusive human monitorization and e-health systems.

## REFERENCES

- [1] Linux 802.11n CSI Tool. 2018. Retrieved October 10, 2018 from <https://dhalperi.github.io/linux-80211n-csitool/>.
- [2] Abbas K. Abbas, Konrad Heimann, Katrin Jergus, Thorsten Orlikowsky, and Steffen Leonhardt. 2011. Neonatal non-contact respiratory monitoring based on real-time infrared thermography. *Biomedical Engineering Online* 10, 1 (2011), 93.



- [3] Giovanni Acampora, Diane J. Cook, Parisa Rashidi, and Athanasios V. Vasilakos. 2013. A survey on ambient intelligence in healthcare. *Proceedings of the IEEE* 101, 12 (2013), 2470–2494.
- [4] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C. Miller. 2014. 3D tracking via body radio reflections. In *11th {USENIX} Symposium on Networked Systems Design and Implementation (NSDI'14)*, 317–329.
- [5] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 837–846.
- [6] Ghada Almashaqbeh, Thayer Hayajneh, Athanasios V. Vasilakos, and Bassam J. Mohd. 2014. QoS-aware health monitoring system using cloud-based WBANs. *Journal of Medical Systems* 38, 10 (2014), 1–20.
- [7] Fernando Alonso-Martin, Maria Malfaz, Joao Sequeira, Javier F. Gorostiza, and Miguel A. Salichs. 2013. A multimodal emotion detection system during human–robot interaction. *Sensors* 13, 11 (2013), 15549–15581.
- [8] Bradley M. Appelhans and Linda J. Luecken. 2006. Heart rate variability as an index of regulated emotional responding. *Review of General Psychology* 10, 3 (2006), 229–240.
- [9] Ngombo Armando, André Rodrigues, Vasco Pereira, Jorge Sá Silva, and Fernando Boavida. 2018. An outlook on physical and virtual sensors for a socially interactive internet. *Sensors* 18, 8 (2018), 2578.
- [10] Asli Atalay, Ozgur Atalay, Muhammad D. Husain, Anura Fernando, and Prasad Potluri. 2017. Piezofilm yarn sensor-integrated knitted fabric for healthcare applications. *Journal of Industrial Textiles* 47, 4 (2017), 505–521.
- [11] Simone Benedetto, Christian Caldato, Elia Bazzan, Darren C. Greenwood, Virginia Pensabene, and Paolo Actis. 2018. Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS One* 13, 2 (2018), e0192691.
- [12] N. Bhas. 2013. *Smart Wearable Devices: Fitness Healthcare Entertainment and Enterprise 2013–2018*. Technical Report Juniper Research, Basingstoke, UK.
- [13] Shen Bin, Liu Yuan, and Wang Xiaoyi. 2010. Research on data mining models for the internet of things. In *2010 International Conference on Image Analysis and Signal Processing*. IEEE, 127–132.
- [14] Paolo Bonato. 2003. Wearable sensors/systems and their impact on biomedical engineering. *IEEE Engineering in Medicine and Biology Magazine* 22, 3 (2003), 18–20.
- [15] Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems* 31, 2 (2016), 102–107.
- [16] Liang Cao, Yufeng Wang, Bo Zhang, Qun Jin, and Athanasios V. Vasilakos. 2018. GCHAR: An efficient Group-based Context-Aware human activity recognition on smartphone. *Journal of Parallel and Distributed Computing* 118 (2018), 67–80.
- [17] Yongjoon Chee, Jooman Han, Jaewoong Youn, and Kwangsuk Park. 2005. Air mattress sensor system with balancing tube for unconstrained measurement of respiration and heart beat movements. *Physiological Measurement* 26, 4 (2005), 413.
- [18] Lan-lan Chen, Yu Zhao, Peng-fei Ye, Jian Zhang, and Jun-zhong Zou. 2017. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Systems with Applications* 85 (2017), 279–291.
- [19] Min Chen, Sergio Gonzalez, Athanasios Vasilakos, Huasong Cao, and Victor C. M. Leung. 2011. Body area networks: A survey. *Mobile Networks and Applications* 16, 2 (2011), 171–193.
- [20] Chloé Clavel, Ioana Vasilescu, Laurence Devillers, Gaël Richard, and Thibaut Ehrette. 2008. Fear-type emotion recognition for future audio-based surveillance systems. *Speech Communication* 50, 6 (2008), 487–503.
- [21] Liyanage C. De Silva, Tsutomu Miyasato, and Ryohei Nakatsu. 1997. Facial emotion recognition using multi-modal information. In *Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications*, Vol. 1. IEEE, 397–401.
- [22] Rong Du, Paolo Santi, Ming Xiao, Athanasios V. Vasilakos, and Carlo Fischione. 2018. The sensible city: A survey on the deployment and management for smart city monitoring. *IEEE Communications Surveys & Tutorials* 21, 2 (2018), 1533–1560.
- [23] Antonio Fernández-Caballero, José Carlos Castillo, María T. López, Juan Serrano-Cuerda, and Marina V. Sokolova. 2013. INT3-Horus framework for multispectrum activity interpretation in intelligent environments. *Expert Systems with Applications* 40, 17 (2013), 6715–6727.
- [24] Antonio Fernández-Caballero, Arturo Martínez-Rodrigo, José Manuel Pastor, José Carlos Castillo, Elena Lozano-Monator, María T. López, Roberto Zangróniz, José Miguel Latorre, and Alicia Fernández-Sotos. 2016. Smart environment architecture for emotion detection and regulation. *Journal of Biomedical Informatics* 64 (2016), 55–73.
- [25] Fitbit. 2018. Retrieved October 10, 2018 from <https://www.fitbit.com/eu/home>.
- [26] Marc Garbey, Nanfei Sun, Arcangelo Merla, and Ioannis Pavlidis. 2007. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering* 54, 8 (2007), 1418–1426.
- [27] Yu Gu, Tao Liu, Jie Li, Fuji Ren, Zhi Liu, Xiaoyan Wang, and Peng Li. 2018. Emosense: Data-driven emotion sensing via off-the-shelf WiFi devices. In *2018 IEEE International Conference on Communications (ICC'18)*. IEEE, 1–6.

- [28] David L. Hall and James Llinas. 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE* 85, 1 (1997), 6–23.
- [29] Kun-Hee Han and Woo-Sik Bae. 2016. Proposing and verifying a security-enhanced protocol for IoT-based communication for medical devices. *Cluster Computing* 19, 4 (2016), 2335–2341.
- [30] Hal E. Hershfield, Susanne Scheibe, Tamara L. Sims, and Laura L. Carstensen. 2013. When feeling bad can be good: Mixed emotions benefit physical health across adulthood. *Social Psychological and Personality Science* 4, 1 (2013), 54–61.
- [31] Peter Hillyard, Anh Luong, Alemayehu Solomon Abrar, Neal Patwari, Krishna Sundar, Robert Farney, Jason Burch, Christina Porucznik, and Sarah Hatch Pollard. 2018. Experience: Cross-technology radio respiratory monitoring performance study. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 487–496.
- [32] Menghan Hu, Guangtao Zhai, Duo Li, Yezhao Fan, Huiyu Duan, Wenhan Zhu, and Xiaokang Yang. 2018. Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation. *PLoS One* 13, 1 (2018), e0190466.
- [33] Kun-Yi Huang, Chung-Hsien Wu, Qian-Bei Hong, Ming-Hsiang Su, and Yi-Hsuan Chen. 2019. Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'19)*. IEEE, 5866–5870.
- [34] Si Huang, Dong Wang, Run Zhao, and Qian Zhang. 2019. Wiga: A WiFi-based contactless activity sequence recognition system based on deep learning. In *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN'19)*. IEEE, 69–74.
- [35] Chien-Chun Hung, Ganesh Ananthanarayanan, Peter Bodik, Leana Golubchik, Minlan Yu, Paramvir Bahl, and Matthai Philipose. 2018. Videododge: Processing camera streams using hierarchical clusters. In *2018 IEEE/ACM Symposium on Edge Computing (SEC'18)*. IEEE, 115–131.
- [36] Socionext America. 2018. Socionext CMOS 24-GHz Radar Sensor Retrieved October 10, 2018 from <http://socionextus.com/wp-content/uploads/2016/07/CN00010AE-SOCIONEXT.pdf>.
- [37] Infineon's innovative XENSIV™ 60 GHz radar chip enables things to see and revolutionizes the Human Machine Interface. 2018. Retrieved October 10, 2018 from <https://www.infineon.com/cms/en/product/promopages/60GHz/?redirId=59739>.
- [38] Jawbone. 2018. Retrieved October 10, 2018 from <https://jawbone.com/up>.
- [39] Zhenhua Jia, Musaab Alaziz, Xiang Chi, Richard E. Howard, Yanyong Zhang, Pei Zhang, Wade Trappe, Anand Sivasubramaniam, and Ning An. 2016. HB-phone: A bed-mounted geophone-based heartbeat monitoring system. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN'16)*. IEEE, 1–12.
- [40] Zhenhua Jia, Amelie Bonde, Sugang Li, Chenren Xu, Jingxian Wang, Yanyong Zhang, Richard E. Howard, and Pei Zhang. 2017. Monitoring a person's heart rate and respiratory rate on a shared bed using geophones. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. 1–14.
- [41] Ossi Kaltiokallio, Hüseyin Yiğitler, Riku Jäntti, and Neal Patwari. 2014. Non-invasive respiration rate monitoring using a single COTS TX-RX pair. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (IPSN'14)*. IEEE, 59–69.
- [42] Mohammad Hadi Kefayati, Vahid Pourahmadi, and Hassan Aghaeinia. 2020. Wi2Vi: Generating video frames from WiFi CSI samples. *IEEE Sensors Journal* 20, 19 (2020), 11463–11473.
- [43] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 12 (2008), 2067–2083.
- [44] John H. G. M. Klaessens, Marlies Van Den Born, Albert Van Der Veen, Janine Sikkens-Van De Kraats, Frank A. M. van den Dungen, and Rudolf M. Verdaasdonk. 2014. Development of a baby friendly non-contact method for measuring vital signs: First results of clinical measurements in an open incubator at a neonatal intensive care unit. In *Advanced Biomedical and Clinical Diagnostic Systems XII*, Vol. 8935. International Society for Optics and Photonics, 89351P.
- [45] Sungjun Kwon, Hyunseok Kim, and Kwang Suk Park. 2012. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2174–2177.
- [46] Sungjun Kwon, Jeehoon Kim, Dongseok Lee, and Kwangsuk Park. 2015. ROI analysis for remote photoplethysmography on facial video. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'15)*. IEEE, 4938–4941.
- [47] Richard D. Lane, Kateri McRae, Eric M. Reiman, Kewei Chen, Geoffrey L. Ahern, and Julian F. Thayer. 2009. Neural correlates of heart rate variability during emotion. *Neuroimage* 44, 1 (2009), 213–222.
- [48] Dan Ledger and Daniel McCaffrey. 2014. Inside wearables: How the science of human behavior change offers the secret to long-term engagement. *Endeavour Partners* 200, 93 (2014), 1.

- [49] Young-Dong Lee and Wan-Young Chung. 2009. Wireless sensor network based wearable smart shirt for ubiquitous health and activity monitoring. *Sensors and Actuators B: Chemical* 140, 2 (2009), 390–395.
- [50] Yee Siong Lee, Pubudu N. Pathirana, Robin J. Evans, and Christopher L. Steinfert. 2015. Noncontact detection and analysis of respiratory function using microwave doppler radar. *Journal of Sensors* 2015 (2015).
- [51] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen. 2014. Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4264–4271.
- [52] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015. Tracking vital signs during sleep leveraging off-the-shelf WiFi. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 267–276.
- [53] Yang Liu, Tiexing Wang, Yuexin Jiang, and Biao Chen. 2020. Harvesting ambient RF for presence detection through deep learning. *IEEE Transactions on Neural Networks and Learning Systems* (2020).
- [54] Guohua Lu, Fang Yang, Yue Tian, Xijing Jing, and Jianqi Wang. 2009. Contact-free measurement of heart rate variability via a microwave sensor. *Sensors* 9, 12 (2009), 9572–9581.
- [55] Yongsan Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi sensing with channel state information: A survey. *ACM Computing Surveys (CSUR)* 52, 3 (2019), 1–36.
- [56] Carlo Massaroni, Daniel Simões Lopes, Daniela Lo Presti, Emiliano Schena, and Sergio Silvestri. 2018. Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: A single camera approach. *Journal of Sensors* 2018 (2018).
- [57] Sinéad McGilloway, Roddy Cowie, Ellen Douglas-Cowie, Stan Gielen, Machiel Westerdijk, and Sybert Stroeve. 2000. Approaching automatic recognition of emotion from voice: A rough benchmark. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.
- [58] Weizhi Meng, Kim-Kwang Raymond Choo, Steven Furnell, Athanasios V. Vasilakos, and Christian W. Probst. 2018. Towards Bayesian-based trust management for insider attacks in healthcare software-defined networks. *IEEE Transactions on Network and Service Management* 15, 2 (2018), 761–773.
- [59] H. Mo, S. Ding, S. Yang, X. Zheng, and A. V. Vasilakos. 2020. The role of edge robotics as-a-service in monitoring COVID-19 infection.
- [60] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. 45–57.
- [61] Jianbing Ni, Kuan Zhang, and Athanasios V. Vasilakos. 2020. Security and privacy for mobile edge caching: Challenges and solutions. *IEEE Wireless Communications* 28, 3 (2021), 77–83.
- [62] David Sousa Nunes, Pei Zhang, and Jorge Sá Silva. 2015. A survey on human-in-the-loop applications towards an internet of all. *IEEE Communications Surveys & Tutorials* 17, 2 (2015), 944–965.
- [63] EPoSS—The European Technology Platform on Smart Systems Integration. 2013. EPoSS Response to the Self Assessment Exercise Launched by the European Commission for Renewed Recognition as European Technology Platform.
- [64] Alexander J. O'Connor and Farhana Jahan. 2014. Under surveillance and overwrought: American Muslims' emotional and behavioral responses to government surveillance. *Journal of Muslim Mental Health* 8, 1 (2014).
- [65] Preksha Pareek and Ankit Thakkar. 2021. A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications. *Artificial Intelligence Review* 54, 3 (2021), 2259–2322.
- [66] Juha Parkka, Miiikka Ermes, Panu Korpipaa, Jani Mantyjarvi, Johannes Peltola, and Ilkka Korhonen. 2006. Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine* 10, 1 (2006), 119–128.
- [67] Neal Patwari, Lara Brewer, Quinn Tate, Ossi Kaltiokallio, and Maurizio Bocca. 2013. Breathfinding: A wireless network that monitors and locates breathing in a home. *IEEE Journal of Selected Topics in Signal Processing* 8, 1 (2013), 30–42.
- [68] Carina Barbosa Pereira, Xinchu Yu, Michael Czaplík, Rolf Rossaint, Vladimir Blazek, and Steffen Leonhardt. 2015. Remote monitoring of breathing dynamics using infrared thermography. *Biomedical Optics Express* 6, 11 (2015), 4378–4394.
- [69] Alan Pierce. 2017. Walabot DIY can see into walls. *Tech Directions* 76, 5 (2017), 8.
- [70] Lukasz Piwek, David A. Ellis, Sally Andrews, and Adam Joinson. 2016. The rise of consumer health wearables: Promises and barriers. *PLoS Medicine* 13, 2 (2016), e1001953.
- [71] Jonathan Posner, James A. Russell, and Bradley S. Peterson. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* 17, 3 (2005), 715.
- [72] Yazdan Ahmad Qadri, Ali Nauman, Yousaf Bin Zikria, Athanasios V. Vasilakos, and Sung Won Kim. 2020. The future of healthcare internet of things: A survey of emerging technologies. *IEEE Communications Surveys & Tutorials* 22, 2 (2020), 1121–1167.

- [73] Yanzhi Ren, Chen Wang, Yingying Chen, Jie Yang, and Hongwei Li. 2019. Noninvasive fine-grained sleep monitoring leveraging smartphones. *IEEE Internet of Things Journal* 6, 5 (2019), 8248–8261.
- [74] Yanzhi Ren, Chen Wang, Jie Yang, and Yingying Chen. 2015. Fine-grained sleep monitoring: Hearing your breathing with smartphones. In *2015 IEEE Conference on Computer Communications (INFOCOM'15)*. IEEE, 1194–1202.
- [75] Peter Salovey, Alexander J. Rothman, Jerusha B. Detweiler, and Wayne T. Steward. 2000. Emotional states and physical health. *American Psychologist* 55, 1 (2000), 110.
- [76] David Sánchez-Rodríguez, Miguel A. Quintana-Suárez, Itziar Alonso-González, Carlos Ley-Bosch, and Javier J. Sánchez-Medina. 2020. Fusion of channel state information and received signal strength for indoor localization using a single access point. *Remote Sensing* 12, 12 (2020), 1995.
- [77] Christopher G. Scully, Jinseok Lee, Joseph Meyer, Alexander M. Gorbach, Domhnall Granquist-Fraser, Yitzhak Mendelson, and Ki H. Chon. 2011. Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering* 59, 2 (2011), 303–306.
- [78] Zhengguo Sheng, Shusen Yang, Yifan Yu, Athanasios V. Vasilakos, Julie A. McCann, and Kin K. Leung. 2013. A survey on the IETF protocol suite for the internet of things: Standards, challenges, and opportunities. *IEEE Wireless Communications* 20, 6 (2013), 91–98.
- [79] Geophone sm 24. 2018. Retrieved October 10, 2018 from <https://www.sparkfun.com/products/11744>.
- [80] Elahe Soltanaghaei, Rahul Anand Sharma, Zehao Wang, Adarsh Chittilappilly, Anh Luong, Eric Giler, Katie Hall, Steve Elias, and Anthony Rowe. 2020. Robust and practical WiFi human sensing using on-device learning with a domain adaptive model. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 150–159.
- [81] Emmanuel Munguia Tapia, Stephen S. Intille, William Haskell, Kent Larson, Julie Wright, Abby King, and Robert Friedman. 2007. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *2007 11th IEEE International Symposium on Wearable Computers*. IEEE, 37–40.
- [82] Mauricio Villarroel, Alessandro Guazzi, João Jorge, Sara Davis, Peter Watkinson, Gabrielle Green, Asha Shenvi, Kenny McCormick, and Lionel Tarassenko. 2014. Continuous non-contact vital sign monitoring in neonatal intensive care unit. *Healthcare Technology Letters* 1, 3 (2014), 87–91.
- [83] K. Visakha and Sidharth S. Prakash. 2018. Detection and tracking of human beings in a video using Haar classifier. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA'18)*. IEEE, 1–4.
- [84] Anran Wang, Dan Nguyen, Arun R. Sridhar, and Shyamnath Gollakota. 2021. Using smart speakers to contactlessly monitor heart rhythms. *Communications Biology* 4, 1 (2021), 1–12.
- [85] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.
- [86] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 1 (2017), 1–27.
- [87] Kajiro Watanabe, Takashi Watanabe, Harumi Watanabe, Hisanori Ando, Takayuki Ishikawa, and Keita Kobayashi. 2005. Noninvasive measurement of heartbeat, respiration, snoring and body movements of a subject in bed via a pneumatic method. *IEEE Transactions on Biomedical Engineering* 52, 12 (2005), 2100–2107.
- [88] Thorsten Wild, Volker Braun, and Harish Viswanathan. 2021. Joint design of communication and sensing for beyond 5G and 6G systems. *IEEE Access* 9 (2021), 30845–30857. <https://doi.org/10.1109/ACCESS.2021.3059488>
- [89] Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, and Jianhua Zhang. 2017. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer Methods and Programs in Biomedicine* 140 (2017), 93–110.
- [90] Yasunari Yoshitomi, Taro Asada, Kyouhei Shimada, and Masayoshi Tabuse. 2011. Facial expression recognition of a speaker using vowel judgment and thermal image processing. *Artificial Life and Robotics* 16, 3 (2011), 318–323.
- [91] Zephyr. 2018. Retrieved October 10, 2018 from <https://www.zephyranywhere.com/>.
- [92] Qingyang Zhang, Hui Sun, Xiaopei Wu, and Hong Zhong. 2019. Edge video analytics for public safety: A review. *Proceedings of the IEEE* 107, 8 (2019), 1675–1696.
- [93] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 95–108.
- [94] Han Zou, Yuxun Zhou, Jianfei Yang, and Costas J. Spanos. 2018. Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning. *Energy and Buildings* 177 (2018), 12–22.

Received November 2020; revised October 2021; accepted October 2021