Multimodal Classification of Stressful Environments in Visually Impaired Mobility Using EEG and Peripheral Biosignals

Charalampos Saitis and Kyriaki Kalimeri, *Member, IEEE*

Abstract—In this study, we aim to better understand the cognitive-emotional experience of visually impaired people when navigating in unfamiliar urban environments, both outdoor and indoor. We propose a multimodal framework based on random forest classifiers, which predict the actual environment among predefined generic classes of urban settings, inferring on real-time, non-invasive, ambulatory monitoring of brain and peripheral biosignals. Model performance reached 93% for the outdoor and 87% for the indoor environments (expressed in weighted AUROC), demonstrating the potential of the approach. Estimating the density distributions of the most predictive biomarkers, we present a series of geographic and temporal visualizations depicting the environmental contexts in which the most intense affective and cognitive reactions take place. A linear mixed model analysis revealed significant differences between categories of vision impairment, but not between normal and impaired vision. Despite the limited size of our cohort, these findings pave the way to emotionally intelligent mobility-enhancing systems, capable of implicit adaptation not only to changing environments but also to shifts in the affective state of the user in relation to different environmental and situational factors.

Index Terms—Visual impairment, mobility, affective state, cognitive load, multimodal recognition, data fusion

✦

1 INTRODUCTION

MOBILITY in indoor and outdoor environments can be a challenging and emotionally stressful task for visually impaired people (VIP), especially when navigating in unfamiliar sites. Despite an increasing number of assistive technologies that help individuals with sight loss to augment their spatial awareness and wayfinding abilities when in move, very few systems provide a high degree of independence beyond known environments that would allow VIP to significantly achieve mobility and integrate into everyday active life [1], [2]. Placing the visually impaired in the center of attention and exploiting recent developments in pervasive physiological sensing for affective computing, two mobility "in the wild" studies were designed to better understand how people with sight loss perceive and interact with the urban and indoor space as manifested in their management of cognitive load and stress.

Orientation and mobility (O&M) in humans heavily relies on sight, which provides instantaneous, effortless access to anticipatory (e.g., stairs, turns, signs) and proactive (e.g., moving people, poles) information at various distances simultaneously [3]. Visually impaired pedestrians learn to obtain critical environmental information primarily through touch (sensing the ground surface with a white cane) and hearing (identifying and localizing events and landmarks

through sound). Mobility challenges can be summarized in four main problems: avoiding obstacles (e.g., people moving or standing in the way, pillars, tree branches, doors opening outwards, improperly parked cars); detecting ground level changes (e.g., stairs, ramps, pavement edge); negotiating street crossings (e.g., lack of curbs, traffic lights, or sound signaling); finding entrance/exit points (e.g., automated doors, elevators); and adapting to light variation (e.g., abrupt changes between different environments) [4], [5].

Although these problems generally diminish with increased experience of an environment (e.g., own living or working space, route from home to work), they still make traveling in unfamiliar settings particularly challenging, often preventing VIP from going outdoors or visiting new sites altogether. The limitations and dependence on others in daily living and mobility often have profound consequences for the psychological health of VIP, generating increased anxiety and motivating social isolation and depression [6], [7]. An increasing number of navigation and access technologies has removed many barriers to independent wayfinding for VIP, greatly advancing their personal and social well-being. However, the degree of independence offered by these technologies can be limited when the user encounters unfamiliar situations that stress and inhibit them [1]. A less than optimal presentation of information may cause unnecessary mental burden, increasing emotional stress through imposing cognitive load on working memory [8].

Despite a significant amount of research on understanding the perceptual and neurocognitive mechanisms by which people with sight loss access and process wayfinding information [9], there is still little practical knowledge of how the management of cognitive load and psychological stress relates to the wayfinding process itself. This is a critical aspect of human-computer interaction that has only recently

- C. Saitis is with the Audio Communication Group, Technische Universität Berlin, Germany. E-mail: *charalampos.saitis@campus.tu-berlin.de*.
- K. Kalimeri is with the Institute for Scientific Interchange (ISI Foundation), Turin, Italy. E-mail: *kalimeri@ieee.org*.

Stressors

Environment
Obstacles
Perception
Decisions
Self-awareness

Stress monitoring

Mobile EEG
Wearable biosensor
Camera
GPS unit
WIFI points
Questionnaire

Data streams

Brain signals
Skin conductance
Blood volume pulse
Heart rate
Skin temperature
Motion-based activity
Video recordings
Geolocation
Mobile network data
Verbal responses

Neuropsychophysiological models of emotional stress

Multimodal analysis

Stress-relevant features
Supervised classification
Kernel density estimation
Unsupervised learning

Applications

Usability/UX
Quantified Self
Healthcare
Navigation/O&M
Urban planning
Marketing
Entertainment

Stress visualization

On a map
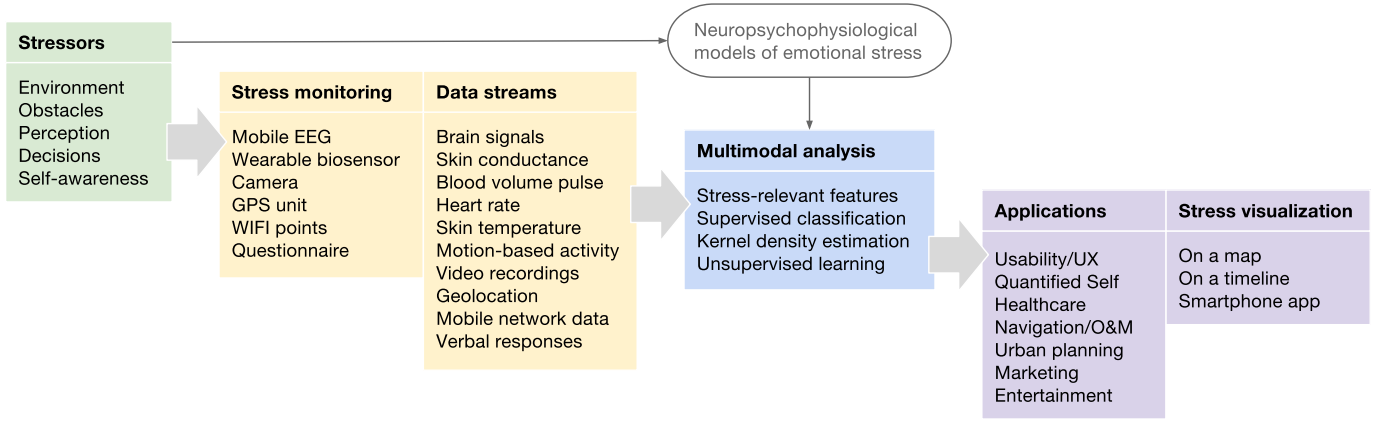On a timeline
Smartphone app

Fig. 1. Multimodal data analysis framework for detecting stress during mobility tasks with visually impaired people. Adapted from Fig. 1 in [11].

been considered essential in designing *emotionally intelligent* mobility-enhancing systems that are capable of implicitly adapting not only to changing environments but also to shifts in the user's affective experience in relation to environmental factors [10]. For example, to activate additional information if a visually impaired traveler feels unsafe and stressed due to an insufficient representation of the surroundings of the system without increasing the mental workload of the user, or to reduce the amount of channeled information if a VIP feels relaxed and confident in a certain environment, or to evaluate changing priorities.

One way of detecting emotion and psychological stress is through identifying patterns in the central and peripheral physiological modalities, the most common being electrodermal activity (EDA), cardiovascular activity, and electroencephalography (EEG). Electrodermal activity is a well-known indicator of physiological arousal and stress activation in affective computing [12], [13]. It is more sensitive to emotion-related variations in arousal as opposed to physical stressors, which can be better reflected in measurements of heart rate (HR). Blood volume pulse (BVP) patterns can also reflect transient processes in arousal and cognitions [14]. Two outdoor mobility studies in the early 1970s suggested that some form of psychological rather than physical stress is responsible for increased HR in visually impaired versus sighted pedestrians [15], [16]. However, certain mobility tasks (e.g., stair climbing) may result in an interactive effect of psychological stress and momentary physical workload, thus cardiovascular measures may be less suitable than EDA.

Electroencephalography, on the other hand, can provide neurophysiological markers of cognitive-emotional processes induced by stress due to imposed cognitive load, indicated by changes in brain activity. The latter is characterized by rhythmic patterns across distinct frequency bands, the definition of which can vary somewhat among studies. Hereafter we consider six EEG bands, namely delta (0.5–4 Hz), theta (4–7 Hz), alpha-1 (7–10 Hz), alpha-2 (10–13 Hz), beta (13–30 Hz), and gamma (30–60 Hz). Gamma waves are thought to be involved in higher cognitive functions such as multimodal processing or object representation [17]. Beta waves are associated with psychological and physical stress [18], whereas theta and alpha-1 frequencies reflect response inhibition and attentional demands such as phasic alertness [19]. Alpha-2 is related to task performance in terms of speed, relevance, and difficulty [20], [21]. Reduced activity at alpha frequencies has been repeatedly associated with increasing cognitive load in a variety of task demands (see [22] for a review). Further evidence suggests that asymmetry in frontal alpha band power varies dependent on affective disposition and engagement, with more activity in left alpha indicating positive approach/motivation and emotions, whereas increased right frontal alpha showed withdrawal/avoidance and negative emotions [23]. EEG delta activity has been reported to indicate attention to internal processing during performance of mental tasks [24].

In recent years, the advent of ubiquitous mobile and sensing technologies, consumer brain-computer interfaces (BCI), and the quantified self movement has driven the development of wireless wearable multi-sensor systems (from devices to smartphone apps) for easy and reliable automatic collection of brain and peripheral biosignal data streams, making it possible to monitor human affective states in virtually any real-world situation [11], [25]. Massot and colleagues [26] used a custom mobile biosensor to collect EDA from 27 blind pedestrians as they walked through urban environments of varying complexity. Examination of arousal-relevant EDA features showed that VIP experience increased psychological stress when walking on busy shopping streets, passing through large open areas, and crossing junctions. In another study [27], analyses of EEG signals recorded from 12 VIP during outdoor travel using a commercial BCI headset [27] further indicated that busy streets, open spaces, and street crossings induce larger cognitive engagement than quieter and less complex urban settings.

Expanding on previous work by the authors and colleagues [28], [29], [30], this paper presents a multimodal framework to automatic inference of stressful environmental conditions affecting visually impaired mobility based on ambulatory monitoring and multimodal fusion of EEG, EDA, and BVP signals, taking advantage of their inherent and complementary properties (Fig. 1). The goal of the research was twofold: to discover biomarkers that can be used to detect shifts in emotional stress and cognitive load between different settings and situations, and to develop and understanding of which environmental factors increase cognitive load and stress during visually impaired mobility.

TABLE 1
Descriptions of Mobility Environments

ID	Description	Challenges
Outdoor Route		
A	Shopping street	People, ads, chairs, tables, poles, ramps
B	Small street	People, poles, ads, ramps, blocked passageway
C	Narrow alley	People, chairs, tables, street ads, trash bins, flower planters, stairs going down, parked cars
D	Urban park	People, tree branches, poles, flower planters, blocked passageway
E	Open space	People, flower planters, stairs going up, blocked passageway
F	Crossing main road with traffic lights	People
G	Crossing small street without traffic lights	People, uneven pavement, detecting edges
H	Construction alley	People, ramps, construction
Indoor Route		
A	Entering through automated doors (two hinged and one rotating)	Finding the pushbutton (hinged doors only), finding where and when to enter the rotating door, other people going through the door at the same time
B	Using an elevator to move between floors	Finding the pushbuttons (calling the elevator, selecting floor), other people exiting/entering
C	Walking along a narrow corridor	Moving people, noise, classroom doors opening suddenly
D	Moving across an open space	Moving people, standing people, tables, chairs, trash bins, pillars, people talking, loud noises
E	Using stairs to move between floors	People using the stairs in the opposite direction, walk down a large spiral flight of stairs

Such knowledge can help design *emotionally intelligent* O&M systems, which are capable of implicitly adapting not only to changing environments but also to shifts in the internal experience of the user in relation to environmental factors. The proposed framework thus differs fundamentally from context-based approaches to environment recognition, for example, GPS-based geolocation. While the latter allow a certain degree of independence for VIP, identifying dynamic stressors in different or the same environments can lead to even more independent mobility systems that recognize, interpret, and adapt to the affective states of the user.

Using state-of-the-art portable sensor devices, EEG, EDA, and BVP signals were collected from a group of VIP and from two normally sighted individuals and as they walked through outdoor and indoor environments of varying complexity and difficulties. A number of multimodal features ranging from low-level signal descriptors to indexes of higher cognitive and emotional functions were extracted and used in unimodal and multimodal classification experiments. While the relationship between unimodal biosignals and psychological arousal has been studied extensively, the detection of stress from fusing multimodal biosignal streams has not been comparatively investigated. To better understand the relationship of stress biomarkers with the environmental and situational factors that evoke them, the most predictive features were examined in relation to variables such as type of environment/situation, amount of vision loss, and impaired versus normal sight with a linear mixed model analysis. A technique for visualizing geographical and temporal density distributions of biomarkers using weighted kernel density estimation [31] and dynamic time warping [32] was developed.

2 MATERIALS AND METHODS

2.1 Participants

A total of ten healthy visually impaired adults with different degrees of sight loss participated in the two mobility studies (6 female; average age = 41 yrs, range = 22–53 yrs). One participant was fully blind, three had visual acuity less than

2%, four had visual acuity less than 5%, and two had visual acuity between 5% and 10%. Eight of them were congenitally or early blind (first 2–3 yrs of life) and two had become blind later in life (generally after the age of 3). To help make the VIP feel comfortable and safe, they were encouraged to walk as usual using their white canes if they wished so, and were accompanied by their familiar O&M instructor. Participants were instructed to avoid smoking normal or e-cigarettes and consuming caffeine or sugar (e.g., coffee, cola, chocolate) approximately one hour prior to the walk. Recruitment was based on volunteering and all VIP were capable of giving free and informed consent. The study was approved by the National Bioethics Committee of Iceland. All data was anonymized before analysis.

All visually impaired participants actively experience indoor environments other than where they reside on a daily basis: four work full-time, three are part-time employees, and three attend educational or vocational establishments. All VIP reported traveling alone outdoors on an "almost daily" basis, but six of them would not feel confident enough to do so in unfamiliar spaces and routes. Two participants, who reported regular use of a white cane when mobile, felt safe enough to walk without any aid. When asked to describe their feelings regarding the ease of mobility over the previous year, four VIP believed that it has not changed, an equal number thought that it has become easier, while two considered it to have become less easy. Eight of the participants walked both the outdoor and indoor routes, one took part only in the outdoor study, and two completed only the indoor task.

Two healthy normally sighted individuals (1 female; age = 31 and 40 yrs) were further recruited. They walked only the outdoor route and their data were used only in the linear mixed model analysis (Sec. 2.8).

2.2 Mobility Environments

The outdoor and indoor routes were planned with the assistance of caretakers and O&M instructors to take the VIP through circumstances of varying complexity and difficulty, where different levels of stress were likely to occur.

The outdoor route was charted in Reykjavik's city center between the City Hall and the port, which consists of smaller and larger streets, narrower and wider sidewalks, street crossings with and without traffic lights, as well a number of open spaces. Accordingly, the route was divided into seventeen parts grouped in eight distinct urban environments (see Fig. 1 in [28]). These were defined so as to cluster environmental and situational factors expected to elicit similar affective reactions (Table 1). For example, participants had to walk on a busy shopping street (environment A1), pass through an urban park-square (D1), and cross a major junction (F1). The route was approximately 1 km long and took on average 13 min 44 sec to walk (range = 9–19 min).

The Háskólatorg building of the University of Iceland campus in Reykjavik houses various service units for students, a bookstore, two restaurants, classrooms, and reading rooms. As such, it provided both uncomplicated and sufficiently complex indoor scenes for the purposes of our study. The charted route linked the entrance at the back of the building (START) to the main entrance at its front (END) and comprised five distinct environments representable of a variety of indoor mobility challenges (see Fig. 2 in [29] and Table 1). Indicatively, participants had to enter through automated doors (scenes A2 and A2*), use an elevator (B2), move across a busy open space (D2—main entrance hall), and walk down a large spiral staircase (E2*). The route was approximately 200 meters in length and took on average 5 minutes to walk (range = 4–8 minutes).

2.3 Multimodal Biosignals

Brain electrical signals were recorded using the Emotiv EPOC+ mobile headset (http://emotiv.com/epoc/), which provides 14+2 dry electrodes registering over the 10-20 system locations AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, FC4, P3, and P4 (sampling rate $f_s = 128$ Hz). The last two locations are used for CMS (Common Mode Sense) and DRL (Driven Right Leg) reference electrodes, respectively. Given the practical constraints involved in monitoring brain electrical activity in the wild, EPOC+ was chosen because it provides a good compromise between performance (i.e., number of channels and scientific validity of the acquired EEG signals) and usability (i.e., portability, preparation time and user comfort) with respect to other commercial wireless EEG systems [25], [33], [34], [35].

Along with the Emotiv headset, participants were asked to wear the Empatica E4 wristband (https://www.empatica.com/e4-wristband) [36]. E4 measures EDA as skin conductance through 2 ventral (inner) wrist electrodes ($f_s = 4$ Hz) and BVP through a dorsal (outer) wrist photoplethysmography (PPG) sensor ($f_s = 64$ Hz). E4 further reports HR, extracted on board from BVP interbeat intervals. The wristband also includes an infrared thermopile sensor and a 3-axis accelerometer. E4 is currently the only commercial multi-sensor device developed based on extended scientific research in the areas of psychophysiology and affective computing. Additionally, it has a cable-free, watch-like design, which makes it easier and more aesthetically pleasing to wear, and thus better fitted to use in the wild compared to other wearable biosignal devices. Participants were asked to wear the wristband on the non-dominant hand to minimize motion artifacts related to handling the white cane [37].
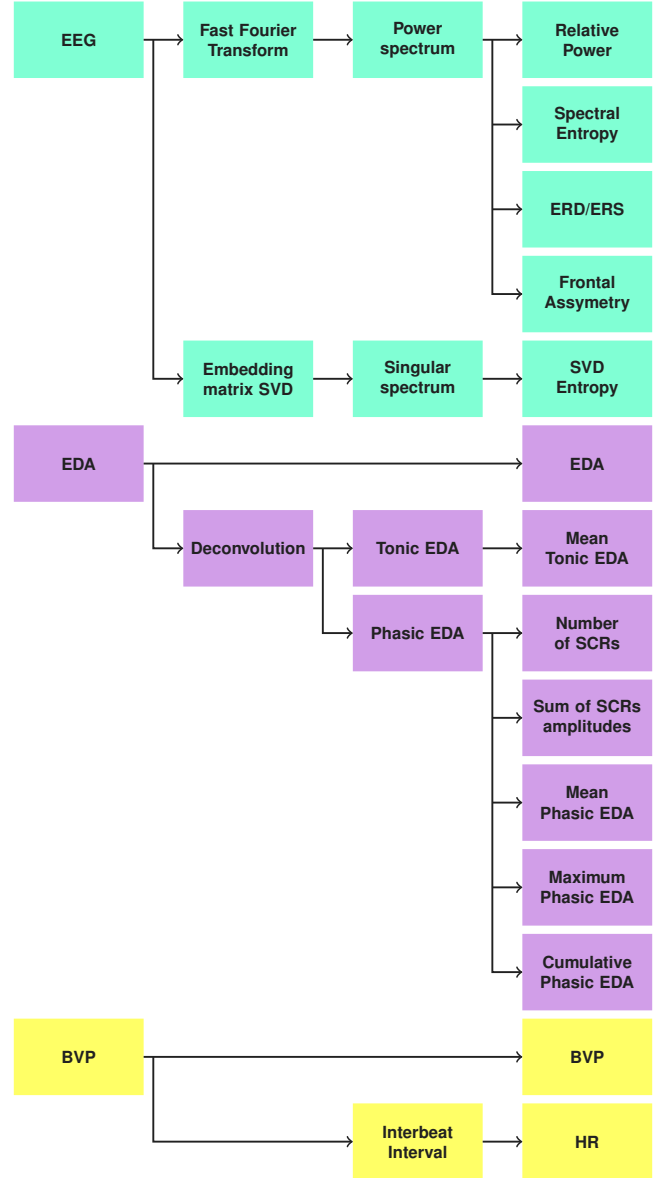

Fig. 2. Features extracted from the EEG, EDA, and BVP modalities.

2.4 General Procedure

Participants walked the outdoor route twice and the indoor route three times for training purposes. In both studies directions were only provided during the first walk to help the VIP familiarize with the route. They were instructed to avoid unnecessary head movements and hand gestures as well as talking to their O&M instructor unless there was an emergency. Video and audio were registered by means of a smartphone camera to facilitate data annotation (observing behaviors across the different environments and situations) and synchronization (start/end of walk, environments, and obstacles). In the outdoor study, GPS coordinates were additionally logged using a Garmin GPSMAP-64s unit at a rate of 1 registration per second. Upon completing the last walk, participants were asked to describe stressful moments they experienced along the route.

2.5 EEG Feature Extraction

The Emotiv EPOC+ system involves a number of internal signal conditioning steps. Analogue signals are first high-pass filtered with a 0.16 Hz cut-off, pre-amplified, low-pass filtered with an 83 Hz cut-off, and sampled at 2048 Hz. Digital signals are then notch-filtered at 50/60 Hz and down-sampled to 128 Hz prior to transmission. For approximately half of the participants, EEG data obtained from the headset was first time-domain interpolated using the Fast Fourier Transform (FFT) to account for missing samples due to connectivity issues. Subsequently, all signals were baseline-normalized. For visually impaired participants, this involved subtracting for each individual and for each channel the mean of resting state registrations obtained during a series of laboratory studies with the same participants [38], [39]. For normally sighted participants, for whom no resting state EEG data were available due to technical issues, the mean signal value was instead subtracted for each individual and for each channel. Finally, min-max scaling was applied to reduce inter-individual variance.

A number of features related to signal power and complexity were extracted using the PyEEG open source Python module [40]. As before [29], we computed for each of the 14 EEG channels relative spectral power in the delta (0.5–4 Hz), theta (4–7 Hz), alpha-1 (7–10 Hz), alpha-2 (10–13 Hz), beta (13–30 Hz), and gamma (30–60 Hz) bands, and the entropy of the power spectrum across bands:

$$\text{PSI}_k = \sum_{i=|N(f_k/f_s)|}^{|N(f_{k+1}/f_s)|} |X_i|$$

$$\text{RIR}_k = \frac{\text{PSI}_k}{\sum_{j=1}^{K-1} \text{PSI}_j}, \quad k = 1, 2, \ldots, K-1$$

$$\text{H} = -\frac{1}{log\,(K)} \sum_{i=1}^{K} \text{RIR}_i \, log\,(\text{RIR}_i)$$

where f_s is the sampling rate, N is the time series length, $|X_1, X_2, \ldots, X_N|$ is the FFT of the series, and K is the total number of bands.

For this paper we also computed SVD Entropy, which measures entropy over the spectrum of eigenvalues in a singular value decomposition (SVD) of an embedding matrix formed by consecutive delay vectors extracted from the signal [41]:

$$\text{H}_{\text{SVD}} = -\sum_{i=1}^{M} \overline{\sigma_i} log_2 \overline{\sigma_i}$$

where M is the number of singular values and $\overline{\sigma_1}, \overline{\sigma_2}, \ldots, \overline{\sigma_M}$ are normalized singular values such that $\overline{\sigma_i} = \sigma_i / \sum_{j=1}^{M} \sigma_j$. PyEEG uses 20 embedding dimensions and 2-point delay.

We further computed, for each electrode and for each frequency band, the event-related (de-) synchronization (ERD/ERS), which reflects the decrease (event-related desynchronization; ERD) or increase (event-related synchronization; ERS) in band power while performing a task relative to a reference baseline without any task demands, in our case the resting state [42]. Using the RIR function of PyEEG (relative band power) the ERD/ERS index is defined as

$$\text{ERD/ERS}_k = \frac{\text{resting RIR}_k - \text{RIR}_k}{\text{resting RIR}_k} * 100$$

TABLE 2
Classification Schemes

Exp.	Description
I	Single-class classification using as predictors the unimodal features extracted from the EEG signals ($N = 198$).
II	Single-class classification using as predictors the unimodal features extracted from the EDA signals ($N = 6$).
III	Single-class classification using feature-level multimodal fusion of raw EDA, BVP, and HR signals ($N = 3$).
IV	Single-class classification using feature-level multimodal fusion of EEG and EDA features ($N = 204$).
V	Single-class classification using feature-level multimodal fusion of EEG, EDA, BVP, and HR features ($N = 206$).
VI	Multi-class classification using as predictors the unimodal features extracted from the EEG signals ($N = 198$)
VII	Multi-class classification using as predictors the unimodal features extracted from the EDA signals ($N = 6$)
VIII	Multi-class classification using feature-level multimodal fusion of raw EDA, BVP, and HR signals ($N = 3$)
IX	Multi-class classification using feature-level multi-modal fusion of EEG and EDA features ($N = 204$).
X	Multi-class classification using feature-level multimodal fusion of EEG, EDA, BVP, and HR features ($N = 206$)

where RIR and $k = 1, 2, \ldots, K$ are as previously. Positive ERD/ERS values indicate a decrease in band power (ERD), while negative values indicate an increase (ERS).

Lastly, a Frontal Asymmetry Index (FAI) was computed by subtracting the log-transformed relative alpha power in the F3 channel (left frontal) from the log-transformed relative alpha power in the F4 channel (right frontal) [43].

$$\text{FAI} = log\left(\frac{\text{RIR(F4)}_{alpha}}{\text{RIR(F3)}_{alpha}}\right)$$

Because increased brain activity suppresses alpha waves, higher values on this index reflect relatively higher left activity (i.e., lower left alpha power) and thus positive feelings and higher engagement. The FAI index was calculated for both the alpha-1 (lower alpha) and alpha-2 (upper alpha) bands. In total 198 features were extracted from each individual EEG recording, using time windows equal to one second of the continuous signal.

2.6 EDA and BVP Feature Extraction

A measurement of skin conductance is traditionally characterized by two types of behavior: short-lasting phasic responses (can be thought of as rapidly changing peaks in EDA) and long-term tonic level in the absence of phasic responses (can be thought of as the underlying slow-changing level of EDA). In terms of physiology, a skin conductance response (SCR) is a sudden rise in the electrical conductance of the skin due to secretion from the skin's sweat glands (sweat contains electrolytes) in response to sympathetic nervous activation. Another characteristic of the signal is the superposition of subsequent SCRs (i.e., one SCR emerges on top of the preceding one), typically observed in states of high arousal.

Skin conductance data obtained from the E4 was first low-pass filtered (1st order Butterworth, $f_c = 0.6$ Hz) to remove steep peaks stemming from artifacts and subsequently min-max normalized to reduce inter-individual variance [44]. Using Ledalab (http://www.ledalab.de/),

TABLE 3
Outdoor Scene Classification Average Weighted AUROC and Standard Deviation Over 5-Fold Cross-Validation

Scene	EEG	EDA	{EDA,BVP}	{EEG,EDA}	{EEG,EDA,BVP}
Single-Class Classification	**Exp. I**	**Exp. II**	**Exp. III**	**Exp. IV**	**Exp. V**
Construction alley (H)	92 (1.0)	91 (1.4)	96 (0.8)	95 (0.6)	97 (0.5)
Crossing main road (F)	93 (1.0)	90 (0.8)	94 (1.0)	95 (1.4)	97 (0.5)
Crossing small street (G)	72 (2.2)	76 (2.1)	88 (3.1)	74 (2.0)	78 (2.3)
Narrow alley (C)	81 (1.3)	88 (0.5)	93 (0.9)	88 (1.2)	93 (0.6)
Open space (E)	85 (1.0)	86 (0.4)	93 (0.9)	89 (0.9)	94 (0.5)
Shopping street (A)	89 (0.6)	88 (0.9)	95 (0.3)	94 (0.4)	98 (0.4)
Small street (B)	82 (1.2)	84 (0.7)	91 (0.7)	86 (1.2)	91 (0.8)
Urban park (D)	85 (1.8)	89 (0.8)	95 (0.5)	91 (1.3)	95 (1.0)
Multi-Class Classification	**Exp. VI**	**Exp. VII**	**Exp. VIII**	**Exp. IX**	**Exp. X**
All outdoor	83 (0.5)	85 (0.7)	93 (0.2)	86 (0.7)	92 (0.1)

conditioned SC signals were then decomposed into two continuous components of phasic and tonic EDA by means of deconvolution using the biexponential function as impulse response and estimating tonic activity, and implicitly phasic activity, through inter-impulse fitting of the deconvolved SC data [45]. As in our previous work [28], [29], six features were extracted: number of SCRs and sum of their amplitudes; average, maximum, and cumulative phasic EDA; and mean tonic EDA. Here we further used the EDA signal directly as reported from the E4 wristband, applying only min-max scaling to lessen inter-individual variation [44].

To index cardiovascular activity, we used the BVP and HR data as streamed by the Empatica E4 wristband. HR is derived from BVP interbeat intervals. The raw BVP signal is preprocessed on board using a proprietary motion artifact removal technique [36]. No further conditioning was implemented to either of the signals. However, given that cardiovascular markers can be highly dependent on physical activity (e.g., when climbing stairs), BVP and HR were min-max normalized prior to analysis.

2.7 Classification Analysis

In order to identify automatically the affective meaning of an urban space based on biosignals recorded from VIP walking through it, we postulated the study as a supervised classification process. A widely-used ensemble learning method for classification was employed, namely Random Forest (RF) classifier [46], selected due to its ability to deal with imbalanced classes as well as because it provides a straightforward assessment of the variable importances. For each of the distinct environments described in Table 1, each time point of the corresponding biosignal data was annotated based on a binary schema per second, where "1" signaled the presence of the participant in the given environment at the given time point and "0" otherwise.

Generally, multimodal information can be fused at three levels: (raw) data, extracted features, and decision based on separate unimodal classifiers. Decision-level fusion is more appropriate when there are time scale differences between modalities. For different biosignals measured in synchronized time scale as in this paper, feature-level fusion offers a more convincing way to build a single multimodal classifier [47], [48]. A series of experiments were thus designed to assess and compare the predictive power of each modality (EEG, EDA or BVP) as well as of their feature-level fusion in both single-class and multi-class scenarios (Table 2). By means of grid search parameter estimation with 5-fold cross-validation, the optimum number of estimators was 300 and the maximum number of features was set equal to the total number of features for each experiment. The Gini impurity function was used to assess the relative importance of each feature to the predictability of the target variable [46].

All data from all times each participant walked a route were used. While overall familiarity might have gradually increased, individual environments still retained a dynamic complexity due to "new" stressors such as people coming from the opposite direction (outdoor) or out of the elevator (indoor), classroom doors opening (indoor), bicycles or cars being parked in different spots (outdoor), or chairs and tables being displaced (indoor). With regard to the outdoor collected dataset, there were 10,340 data points in total; the eight classes were significantly imbalanced, ranging from 3,278 (most frequent) to 460 (least frequent) data points. The indoor dataset comprised 6,412 data points in total; again the five classes were imbalanced ranging from 1,964 (most frequent) to 570 (less frequent) data points.

Sequential data points were split randomly in training and testing subsets (which, as a result, no longer contain sequential points). We trained one model for each of the single-class cases and one for the multi-class experiment following a 5-fold cross-validation scheme, where the 80% of the data points were used for training and the 20% for testing, with data shuffling in order to avoid dependencies in consecutive data points. The best model is chosen as the one that maximized the weighted area under receiver operating characteristic (AUROC) statistic, taking into account the lack of balance between the class labels.

2.8 Linear Mixed Model Analysis

We examined the role of vision impairment in the perception of environmental and situational stressors with a statistical analysis of features that emerged as the most predictive in the multimodal classification experiments. A linear mixed model method was used, which performs a regression-like analysis while controlling for random variance caused by differences in factors such as participant and electrode [49].

TABLE 4
Indoor Scene Classification Average Weighted AUROC and Standard Deviation Over 5-Fold Cross-Validation

Environment	EEG	EDA	{EDA,BVP}	{EEG,EDA}	{EEG,EDA,BVP}
Single-Class Classification	**Exp. I**	**Exp. II**	**Exp. III**	**Exp. IV**	**Exp. V**
Door (A)	76 (0.6)	76 (1.2)	86 (1.1)	79 (0.9)	86 (1.0)
Elevator (B)	87 (0.8)	79 (0.3)	88 (0.8)	89 (0.8)	93 (0.8)
Corridor (C)	72 (1.6)	66 (1.9)	79 (1.3)	75 (0.9)	76 (0.8)
Open space (D)	76 (0.8)	67 (1.0)	85 (0.9)	77 (1.2)	85 (0.7)
Stairs (E)	78 (1.2)	69 (1.1)	84 (1.0)	82 (1.2)	92 (0.4)
Multi-Class Classification	**Exp. VI**	**Exp. VII**	**Exp. VIII**	**Exp. IX**	**Exp. X**
All indoor	78 (0.4)	71 (0.6)	87 (1.2)	81 (0.5)	84 (0.5)

Fixed factors examined in the analyses included environment and vision. For the latter, three categories were considered: normal (visual acuity greater than 30%), severe impairment (visual acuity less than 10% but greater than 2%), and blind (visual acuity less than 2%). When fitting EDA and BVP data, a random intercept for each participant was added. When fitting EEG data, a further random intercept for electrode was included. Type III Wald F-tests were used to test the significance of the fixed factors and their interaction. Pairwise comparisons of group means were carried out with t-tests, using Bonferroni-adjusted p-values where appropriate.

We previously described that EEG signals collected from visually impaired participants were baseline-normalized by subtracting the mean of resting state registrations obtained in a series of laboratory experiments, whereas for those of normally sighted participants the mean signal value was subtracted instead (Sec. 2.5). To facilitate comparison between the two groups, we re-baseline-normalized the EEG data of the VIP that walked the outdoor route to match those of the normally sighted.

3 RESULTS AND DISCUSSION

Due to temporary dysfunctions of the equipment, incomplete data recorded from two VIP during the first walk of the outdoor route and from two other participants during the first walk of the indoor route were discarded.

3.1 Stressful Environment Prediction

The average weighted AUROC and standard deviation over five folds for all outdoor scene classification experiments are reported in Table 3. In the one-versus-all scenario (Exp. I–V), the EEG and EDA modalities (Exp. I and II, respectively) were both predictive of the distinct scenes (classes) and with highly similar performance. Fusing the two modalities (Exp. IV) gave marginally improved results. The fusion of the EDA and BVP modalities (Exp. III) boosted the performance of the classifier compared to using only EDA or EEG features or, to a lesser extent, both modalities. Combining features from all three modalities achieved almost perfect accuracy across all scenes. Similar trends were observed for the multi-class classification experiments (Exp. VI–X). Including BVP (Exp. VIII and X) considerably improved performance over considering only EEG, EDA, or their fusion.

Table 4 summarizes the mean weighted AUROC and standard deviation over five folds for all indoor scene classification experiments. Overall performance for the indoor

scenes was quite satisfactory, but not as high as for the outdoor scenes. In the one-versus-all scenario (Exp. I–V), the EEG and EDA modalities (Exp. I and II, respectively) were both predictive of the distinct scenes, with EEG performing considerably better than EDA. Fusing the two modalities (Exp. IV) resulted in considerably better results, particularly of the elevator class, the detection of which improved substantially compared to when using only EDA features. Combining EDA with BVP (Exp. III) achieved substantially better accuracy than using only EDA. Adding EEG (Exp. V) only improved results for two classes (elevator and stairs). In the multi-class scenario, EEG (Exp. VI) performed better than EDA (Exp. VIII). Their fusion (Exp. IX) marginally increased accuracy. Combining EDA with BVP gave the best outcome, softly outperforming the fusion of all three modalities (Exp. X).

As a means of assessing the qualitative performance of the multi-class multimodal fusion model (Exp. X), Fig. 3 shows the weighted ROC curves for each outdoor and indoor scene in a one-against-all binary scenario. In both cases, the trained model was able to learn all different scenes equally well, providing proof of the stability of the multimodal approach.

Feature importances were estimated for all multi-class experiments. The most predictive ones appeared always with the highest ranks. In the outdoor scene classification, mean tonic EDA emerged as the most predictive feature in Exp. VII (EDA features only), Exp. IX (fusion of EEG and EDA features), and Exp. X (fusion of EEG, EDA, and BVP features). The same results were obtained for the indoor scene experiments, except for Exp. X where HR
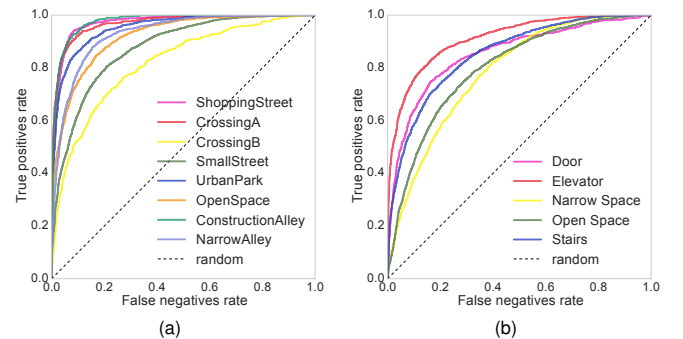

Fig. 3. One-against-all weighted ROC curves for the multi-class multimodal classification of (a) outdoor and (b) indoor environments.
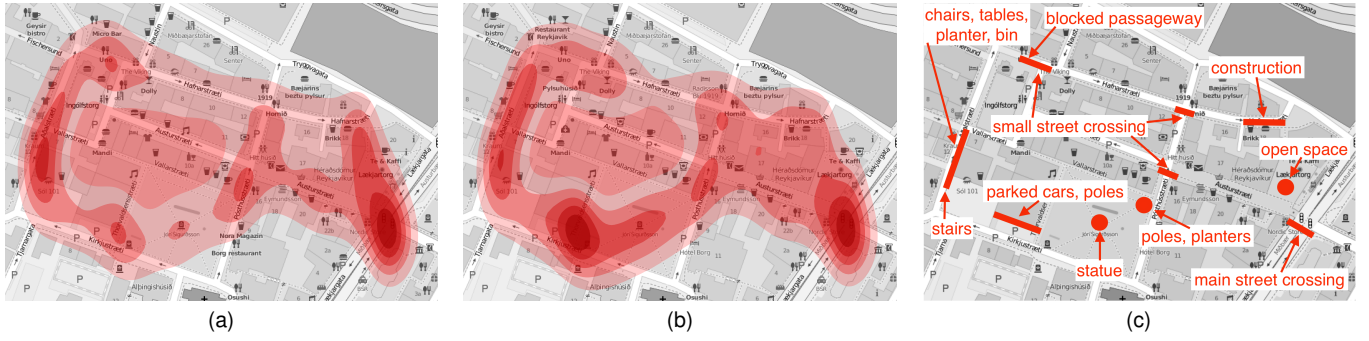
Fig. 4. Spatial density distributions of (a) mean tonic EDA and (b) number of SCRs (the darker the color, the higher the density of the distribution). (c) Annotated obstacles and situations along the route. Reprinted with permission from Springer. Map image: https://www.openstreetmap.org/.

performed marginally better than mean tonic EDA. The emergence of the latter as the most predictive feature in the present experiments confirms previous findings showing skin conductance tonic level to be a highly relevant index of stress-induced physiological arousal.

Among the eight most predictive features resulting from the {EEG,EDA,BVP} fusion models were the EDA and HR signals registered by the Empatica E4 wristband, which were used "as is" (but min-max scaled to reduce inter-individual variation, see Sec. 2.5). This result illustrates the big potential of using existing state-of-the-art sensors such as the E4 for real-time prediction of human affective states from peripheral physiological signals during less controlled experimental conditions, which could be employed to design emotionally intelligent mobility aids for visually impaired travelers.

Predictions involving EEG features, alone or fused with other modalities, were dominated by changes in relative spectral power (i.e., ERD/ERS, see Sec. 2.5) in the delta band of the F3/4, T7/8, P7/8, and, to a lesser extent, FC5/6 and O1/2 channels. Although real-time EEG acquisition may be subject to very noisy signals, this finding is in line with previous evidence (a) reporting increased delta activity during mental tasks as a result of attention to internal processing [24] and (b) suggesting that the 10-20 system locations F3/4, F7/8 and T7/8 may be suitable enough to monitor brain activity under cognitive-emotional stress [50].

EDA and BVP features were generally better in predicting stressful urban scenes, whereas the EEG modality performed better in indoor scene classification. This might have resulted from differences in the how the outdoor and indoor routes were designed. Whereas the former focused mainly on passive walking, the latter involved active wayfinding, for example, turning and going towards the door, finding where stairs begin, negotiating orientation while climbing down the spiral stairs and after exiting the elevator. Therefore, whereas EDA and BVP features reflected the more general stressful situations across urban settings, EEG features traced changes in cognitive load during specific indoor wayfinding tasks. This difference between the two studies proves further evidence of the complementary nature of the three modalities in assessing human affective states and thus supports the need for multimodal approaches to stress detection in visually impaired mobility.

To assess person-specific effects on the performance of the unimodal and multimodal classifiers, we further performed leave-one-participant-out cross-validation. The complete results are reported in Appendix A (Tables 7 and 8). Overall, average weighted AUROC values were greatly affected, dropping in some cases as much as 50%. Indoor scene classifiers performed generally better than their outdoor counterparts, with the {EEG,EDA} and {EEG,EDA,BVP} fusion models performing as good as the unimodal EEG classifier. These results are not surprising and individual reactions should be considered when using biosignals for automatic affective state recognition in navigation aid systems. Nevertheless, the proposed models are a viable fit for personalized systems, where after a short period of user-specific training almost perfect accuracies can be achieved.

3.2 Visualizing Biomarker Density Distributions

To better understand how mean tonic EDA relates to environmental and situational factors, as well as the intensity of the cognitive and emotional response it expresses, its geographical (outdoor route) and temporal (indoor route) density distributions were assessed by means of weighted kernel density estimation [28], [29] and contrasted with that of the number of SCRs, the least predictive EDA feature.

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be an independent random sample drawn from some distribution with density function $f(\mathbf{x})$
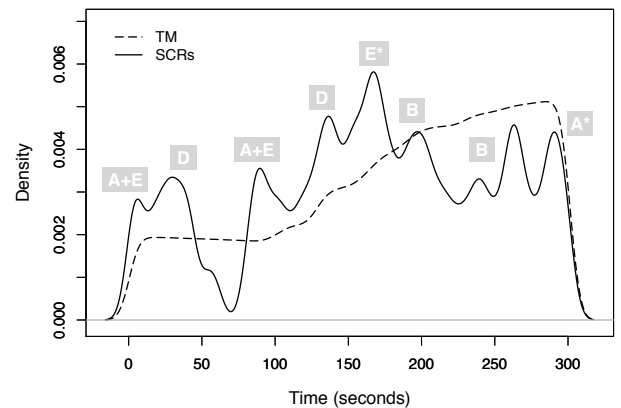

Fig. 5. Temporal density distributions of mean tonic EDA and number of SCRs (indoor route). Letters depict the different indoor scenes reported in Table 1 (A*: rotating door; E*: large spiral stairs).

TABLE 5
Linear Mixed Model Type III Wald Tests Comparing Categories of Visual Impairment

	df	F	p	df	F	p
		Outdoor			**Indoor**	
		Mean tonic EDA				
Intercept (I)	1, 6.00	24.47	0.003	1, 7.03	29.27	< 0.001
Vision (V)	1, 6.00	0.72	0.430	1, 7.03	0.96	0.360
Scene (S)	7, 89.01	0.76	0.621	4, 108.00	1.25	0.294
V × S	7, 89.01	0.33	0.937	4, 108.00	0.02	0.999
		HR				
I	1, 6.00	33.22	0.001	1, 7.06	76.41	< 0.001
V	1, 6.00	1.16	0.323	1, 7.06	0.04	0.848
S	7, 89.01	4.83	< 0.001	4, 108.01	0.79	0.533
V × S	7, 89.01	2.58	0.018	4, 108.01	0.24	0.914
		ERD/ERS, delta				
I	1, 6.06	23.55	0.003	1, 9.55	100.24	< 0.001
V	1, 6.08	0.66	0.447	1, 7.51	2.48	0.157
S	7, 639.20	3.92	< 0.001	4, 832.31	16.87	< 0.001
V × S	7, 639.77	5.63	< 0.001	4, 831.92	0.52	0.719
		ERD/ERS, alpha-2				
I	1, 7.97	1027.09	< 0.001	1, 9.74	903.41	< 0.001
V	1, 6.05	7.38	0.035	1, 7.37	1.23	0.302
S	7, 833.86	5.39	< 0.001	4, 888.74	20.92	< 0.001
V × S	7, 833.90	7.30	< 0.001	4, 887.90	1.16	0.325
		FAI				
I	1, 6.00	0.17	0.699	1, 7.04	0.12	0.741
V	1, 6.00	1.45	0.274	1, 7.04	3.97	0.087
S	7, 89.01	0.07	0.999	4, 108.00	1.15	0.335
V × S	7, 89.01	0.10	0.998	4, 108.00	0.15	0.964

defined on $\mathbb{R}^d$. The weighted kernel density estimate of f is defined as

$$\hat{f}_H(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} w(\mathbf{x}_i, \mathbf{w}) \, K_H(\mathbf{x} - \mathbf{x}_i)$$

where K is a kernel function, $H > 0$ is a symmetric $d \times d$ matrix which controls the bandwidth (or smoothing) of the estimate, $K_H(\mathbf{x}) = |H|^{-1/2} K(H^{-1/2}\mathbf{x})$, and w is a function weighting each data point in the sample with a value from $\mathbf{w} \in \mathbb{R}^m$, $m \le d$. A popular choice for K is the Gaussian (or normal) kernel, which was also applied here.

For the outdoor scenes, feature values were assigned to pairs of latitude and longitude coordinates based on recorded timestamps. Using values as weights (w with $m = 1$) for GPS points ($\mathbf{x}$ with $d = 2$) and a bandwidth of $H(\mathbf{x}) = 0.0008$ helped estimate the feature-weighted density of GPS points on a 500×500 grid and thus obtain a density distribution contour plot for each participant (VIP only). Figure 4 shows the resulting contours aggregated for all participants and walks and plotted on top of a map.

Increased stress-elicited arousal along the different urban scenes of the route is immediately observed when the VIP had to cross a main road (scene F), pass along parked cars in a narrow alley after the urban park (C), walk up and down stairs (E), or pass through a narrow area between construction works (H). These arousal "hotspots" are in full agreement with the scenes reported as stressful by the participants themselves at the end of the study. Furthermore, they coincide with the presence of certain obstacles and situations that can be less or more stressful in visually impaired mobility, and which were taken into account when designing the outdoor route.

To visualize density distribution along the indoor route, feature values were assigned to 1-second steps from the start point based on recorded timestamps. Due to different walking speeds and behaviors, individual walk times varied between participants and trials, ranging from 4 to 8 minutes with an average length of 5 minutes. To temporary align all features so that same times corresponded to same environments we performed dynamic time warping [32], postulating that a certain environment induced similar biomarker patterns. Each feature vector was warped to a reference vector that was 300 seconds (5 minutes) long. Using warped feature values as weights (w with $m = 1$) for 1-second time steps ($\mathbf{x}$ with $d = 1$) and a bandwidth of $H(\mathbf{x}) = 5.59$ helped estimate the feature-weighted density of time points (temporal distances) on a 400-point grid across all participants and walks.

Figure 5 shows the resulting density distributions for the mean tonic EDA and number of SCRs features plotted together and annotated with the different indoor scenes. Tonic EDA appears to gradually increase towards the second half of the walk, which is where the most stressful parts of the route were according to reports from all participants at

the end of the study. As expected, the estimated density of SCRs generally followed the same trend while allowing to observe localized rises in arousal. These suggest the presence of a higher number of instantaneous stressors, for example, when safely entering/exiting an elevator or passing through a rotating door or walking up/down stairs while others try to do the same, or maintaining direction amidst loud noises and people moving in an open space indoors.

3.3 Linear Mixed Model Analysis

Mean tonic EDA and HR, the two most predictive biomarkers, were first analyzed. We then examined ERD/ERS in delta frequencies, an EEG feature that further dominated predictions. In addition, because higher alpha desynchronization has been consistently associated with increased cognitive load [22] and asymmetry in frontal alpha activity has been found to relate to the positive/negative disposition and engagement [23], ERD/ERS and FAI values in the upper alpha band were also analyzed. Before averaging across conditions, a logarithmic transformation of single-condition feature values was applied to improve their distributional characteristics, except for FAI as its definition involves such a transformation already. Average biomarker values for each scene and vision group are provided in Appendix B (Fig. 6).

Type III Wald F-tests comparing two categories of visual impairment (severe versus blind, see Sec. 2.8) are displayed in Table 5. Vision alone was only a significant predictor of upper alpha ERD/ERS in the outdoor route and a marginally significant predictor of FAI in the indoor route, although the interaction of vision and scene was significantly influential for both delta and upper alpha ERD/ERS as well as for HR in the outdoor route. The scene alone had a significant effect on delta and alpha ERD/ERS in both outdoor and indoor models, as well as on HR in the outdoor model.

Post-hoc paired samples t-tests showed that HR was significantly lower for severely impaired than for blind individuals when crossing a major intersection [$t(13.21) = -2.45, p = 0.029$] and when walking in a shopping street [$t(13.21) = -2.39, p = 0.033$]. ERD/ERS in the delta band varied significantly between the two VIP groups during crossing traffic lights [Severe > Blind, $t(31.29) = 2.43, p = 0.021$], a small street [Severe < Blind, $t(13.54) = -2.99, p = 0.010$], and an open space [Severe < Blind, $t(26.80) = -2.20, p = 0.037$]; and marginally significantly when crossing a construction alley [Severe < Blind, $t(14.14) = -1.80, p = 0.094$]. For the indoor environments, delta EDR/ERS was only marginally significantly higher for blind than for severely impaired individuals when navigating through automated moving doors [$t(28.00) = 1.95, p = 0.062$]. ERD/ERS in the upper alpha band was significantly lower for severely impaired than for blind individuals for the outdoor environments A [$t(13.52) = -2.35, p = 0.035$], B [$t(20.51) = -4.13, p = 0.001$], E [$t(15.85) = -2.14, p = 0.048$], G [$t(25.92) = -4.20, p < 0.001$], and H [$t(24.07) = -2.47, p = 0.021$]. No significant within-scene differences emerged between the two VIP groups in indoor environments. FAI was only marginally significantly higher for severely impaired than for blind individuals when walking along a narrow corridor indoors [$t(15.63) = 1.95, p = 0.069$].

When averaging across the two VIP groups, HR was significantly lower when crossing a main traffic junction

TABLE 6
Linear Mixed Model Type III Wald Tests Comparing Impaired to Normal Vision (Outdoor Study Only)

	df	F	p
Mean tonic EDA			
Intercept (I)	1, 6.95	27.58	0.001
Vision (V)	2, 6.99	0.56	0.594
Scene (S)	7, 112.00	4.92	< 0.001
V × S	14, 112.00	1.63	0.082
HR			
I	1, 6.91	31.51	< 0.001
V	2, 6.93	1.49	0.291
S	7, 112.01	5.64	< 0.001
V × S	14, 112.01	1.42	0.154
ERD/ERS, delta			
I	1, 7.08	13.32	0.008
V	2, 7.01	0.53	0.610
S	7, 831.65	6.35	< 0.001
V × S	14, 831.47	4.17	< 0.001
ERD/ERS, alpha-2			
I	1, 7.46	514.59	< 0.001
V	2, 7.00	1.53	0.281
S	7, 1069.18	5.88	< 0.001
V × S	14, 1069.05	5.03	< 0.001
FAI			
I	1, 6.92	0.42	0.540
V	12, 6.94	0.87	0.460
S	7, 112.01	0.14	0.995
V × S	7, 112.01	0.10	1.000

than in any other outdoor environment [$-4.21 < t(89.00) < -3.27, p < 0.045$], except for when walking in a shopping street. The latter resulted in significantly lower HR than when walking across an open space [$t(89.00) = 3.49, p = 0.021$] and in marginally significantly lower HR when passing a construction alley [$t(89.00) = 3.08, p = 0.078$]. Delta ERD/ERS in the shopping street, narrow alley, and small street crossing environments was significantly lower than when passing a construction alley [$-4.36 < t(646.41–649.18) < -3.54, p < 0.018$]. Upper alpha ERD/ERS was significantly higher when crossing a main traffic junction [$3.68 < t(833.64–834.56) < 4.44, p < 0.007$] and marginally significantly higher when walking along a narrow alley [$2.99 < t(830.96–834.45) < 3.64, p < 0.079$] than when passing through the shopping street, small street, and small street crossing scenes. Significantly higher upper alpha ERD/ERS was also observed for the urban park scene compared to the small street environment [$t(833.74) = 3.31, p = 0.027$]. For the indoor route, delta ERD/ERS was significantly lower when entering through moving doors, using an elevator, and using stairs than when walking along a narrow corridor [$-6.89 < t(830.64–838.12) < -4.12, p < 0.001$] and across an open space [$-5.40 < t(831.39–837.41) < -2.91, p < 0.038$]. Upper alpha ERD/ERS was significantly higher when using automated moving doors and when taking the elevator than in the other indoor environments [$3.19 < t(887.33–893.52) < 7.13, p < 0.015$ and $3.22 < t(886.08–893.01) < 7.27, p < 0.014$, respectively].

Table 6 reports type III Wald F-tests comparing impaired

to normal vision (outdoor route only), specifically normal vision to severe visual impairment on the one hand, and normal vision to blindness on the other. No significant differences emerged between the three categories of vision for the examined biomarkers. Scene was a significant predictor in all models but FAI. The two factors appeared to influence each other significantly for delta and alpha ERD/ERS, and nearly significantly for mean tonic EDA.

Post-hoc paired samples t-tests showed no significant within-scene differences between pairs of the three vision groups for neither of the tested biomarkers. The significant interaction between vision and scene in the TM and ERD/ERS models was a result of differences between severely impaired and blind individuals rather than between these and normally sighted persons. Accordingly, we merged the Severe and Blind participants into a single VIP group and run new t-tests against the group of normally sighted. Still, no significant within-scene differences emerged between normal and impaired vision. When averaging across scenes, no significant differences emerged between normal and impaired vision either. This result could have two different origins considering the circumstances related to the particular study. First, the VIP who took part are super-achievers: they have a job or attend college, and travel alone outdoors on an almost daily basis. Second, they were accompanied by their familiar O&M instructor to help make them feel comfortable and safe. These experimental factors may have prevented the elicitation of increased psychological stress.

4 CONCLUSIONS

Mobility aids for visually impaired people should be capable of implicitly adapting not only to changing environments but also to shifts in the affective state of the user in relation to different environmental and situational factors. To this end, this paper presents a framework for real-time automatic assessment of the cognitive-emotional experience of VIP while navigating in unfamiliar outdoor and indoor environments, based on ambulatory monitoring and fusion of brain and peripheral biosignal data. Different multimodal fusion scenarios were compared, aiming to address the robustness of the model as well as emerging differences in the perception and interaction of VIP with their surroundings.

The consistently high prediction rates in the multimodal classification experiments (81–93% weighted AUROC) are very encouraging of the proposed approach. Even if the chosen city and building sites did not represent all possible different outdoor and indoor environments and situations in terms of complexity and difficulty, the charted routes were designed so as to combine most of the mobility challenges faced by VIP. Indeed, the most predictive biomarkers indicated spaces and situations as stressful and cognitively demanding "hotspots" in perfect agreement with the self-reported experience of the visually impaired participants.

Reported findings, despite being promising, should be considered with caution due to the limited number of participants, which did not allow for an in-depth analysis of specific stressors in each category of vision impairment. Furthermore, the well-established Emotiv EPOC+ EEG headset has certain limitations with respect to the quality of the recorded signal during experiments involving physical activity "in the wild"

such as those presented in this paper. The number of the provided electrodes is limited and hence the EEG markers discussed in this paper are meant to provide only insights on the most predictive features and their connection to specific tasks and conditions.

A rich multimodal dataset has been collected, which will be made openly available in order to maximize the impact of the work and encourage further investigations. Future steps of the present study include refining the predictive model through exploring novel multimodal biosignal features and comparing different classifiers. Such findings hopefully pave the way to emotionally intelligent mobile technologies that take the concept of navigation one step further, accounting not only for the shortest path but also for the most effortless, least stressful and safest one.

REFERENCES

[1] N. A. Giudice and G. E. Legge, "Blind navigation and the role of technology," in *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*, A. Helal, M. Mokhtari, and B. Abdulrazak, Eds. John Willey & Sons, 2008, pp. 479–500.

[2] J. R. Marston and R. G. Golledge, "The hidden demand for participation in activities and travel by persons who are visually impaired," *J. Vis. Imp. Blind.*, vol. 97, pp. 475–488, 2003.

[3] S. Millar, *Understanding and Representing Space: Theory and Evidence from Studies with Blind and Sighted Children*. New York: Oxford: Clarendon, 1994.

[4] D. R. Geruschat and A. J. Smith, "Low vision for orientation and mobility," in *Foundations of Orientation and Mobility*, 3rd ed., W. R. Wiener, R. L. Welsh, and B. B. Blasch, Eds. New York: AFB Press, 2010, vol. I: History and Theory.

[5] P.-A. Quiñones, T. C. Greece, R. Yang, and M. W. Newman, "Supporting visually impaired navigation: A needs-finding study," in *Proc. ACM CHI*, Vancouver, BC, 2011, pp. 1645–1650.

[6] M. I. Wallhagen, W. J. Strawbridge, S. J. Shema, J. Kurata, and G. A. Kaplan, "Comparative impact of hearing and vision impairment on subsequent functioning," *J. Am. Geriatr. Soc.*, vol. 49, pp. 1086–1092, 2001.

[7] G. Rees, H. W. Tee, M. Marella, E. Fenwick, M. Dirani, and E. L. Lamoureux, "Vision-specific distress and depressive symptoms in people with vision impairment," *Invest. Ophthalmol. Vis. Sci.*, vol. 51, pp. 2891–2896, 2010.

[8] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learn Instr.*, vol. 4, pp. 295–312, 1994.

[9] Z. Cattaneo, T. Vecchi, C. Cornoldi, I. Mammarella, D. Bonino, E. Ricciardi, and P. Pietrini, "Imagery and spatial processes in blindness and visual impairement," *Neurosci. Biobehav. Rev.*, vol. 32, pp. 1346–1360, 2008.

[10] R. L. Welsh, "Improving psychosocial functioning for orientation and mobility," in *Foundations of Orientation and Mobility*, 3rd ed., W. R. Wiener, R. L. Welsh, and B. B. Blasch, Eds. New York: AFB Press, 2010, vol. 2.

[11] E. Kanjo, L. Al-Husain, and A. Chamberlain, "Emotions in context: examining pervasive affective sensing systems, applications, and analyses," *Pers. Ubiquit. Comput.*, vol. 19, pp. 1197–1212, 2015.

[12] J. Healey, R. W. Picard *et al.*, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, pp. 156–166, 2005.

[13] C. Setz, B. Arnrich, J. Schumm, R. L. Marca, and G. Tröster, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, pp. 410–417, 2010.

[14] E. Peper, R. Harvey, I.-M. Lin, H. Tylova, and D. Moss, "Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony?" *Biofeedback*, vol. 35, pp. 54–61, 2007.

[15] P. Peake and J. A. Leonard, "The use of heart rate as an index of stress in blind pedestrians," *Ergonomics*, vol. 14, pp. 189–204, 1971.

[16] R. J. Wycherley and B. H. Nicklin, "The heart rate of blind and sighted pedestrians on a town route," *Ergonomics*, vol. 13, pp. 181–192, 1970.

[17] A. Keil, M. M. Müller, W. J. Ray, T. Gruber, and T. Elbert, "Human gamma band activity and perception of a gestalt," *J. Neurosci.*, vol. 19, pp. 7152–7161, 1999.

[18] S. K. Jena, "Examination stress and its effect on EEG," *Int. J. Med. Sci. Public Health*, vol. 11, pp. 1493–1497, 2015.

[19] W. J. Ray and H. W. Cole, "EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes," *Science*, vol. 228, pp. 750–752, 1985.

[20] A. Gevins, M. E. Smith, L. McEvoy, and D. Yu, "High-resolution eeg mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice." *Cerebral Cortex*, vol. 7, pp. 374–385, 1997.

[21] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res. Rev.*, vol. 29, pp. 169–195, 1999.

[22] P. Antonenko, F. Paas, R. Grabner, and T. Van Gog, "Using electroencephalography to measure cognitive load," *Educ. Psychol. Rev.*, vol. 22, pp. 425–438, 2010.

[23] E. Harmon-Jones, P. A. Gable, and C. K. Peterson, "The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update," *Biol. Psychol.*, vol. 84, pp. 451–462, 2010.

[24] T. Harmony, T. Fernández, J. Silva, J. Bernal, L. Díaz-Comas, A. Reyes, E. Marosi, M. Rodríguez, and M. Rodríguez, "EEG delta activity: an indicator of attention to internal processing during performance of mental tasks," *Int. J. Psychophysiol.*, vol. 24, pp. 161–171, 1996.

[25] S. Debener, F. Minow *et al.*, "How about taking a low-cost, small, and wireless EEG for a walk?" *Psychophysiology*, vol. 49, pp. 1449–1453, 2012.

[26] B. Massot, N. Baltenneck, C. Gehin, A. Dittmar, and E. McAdams, "EmoSense: An ambulatory device for the assessment of ANS activity—application in the objective evaluation of stress with the blind," *IEEE Sens. J.*, vol. 12, pp. 543–551, 2012.

[27] P. Mavros, K. Skroumpelou, and A. H. Smith, "Understanding the urban experience of people with visual impairments," in *Proc. GIS Research UK 2015*, Leeds, UK, 2015, pp. 401–406.

[28] C. Saitis and K. Kalimeri, "Identifying urban mobility challenges for the visually impaired with mobile monitoring of multimodal biosignals," in *Universal Access in Human-Computer Interaction: Users and Context Diversity*, M. Antona and C. Stephanidis, Eds. Cham, Switzerland: Springer, 2016, pp. 616–627.

[29] K. Kalimeri and C. Saitis, "Exploring multimodal biosignal features for stress detection during indoor mobility," in *Proc. ACM ICMI*, Tokyo, Japan, 2016, pp. 53–60.

[30] C. Saitis, M. Z. Parvez, and K. Kalimeri, "Cognitive load assessment from EEG and peripheral biosignals for the design of visually impaired mobility aids," *Wirel. Commun. Mob. Comput.*, vol. 2018, Article ID 8971206, 2018.

[31] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London: Chapman & Hall, 1995.

[32] T. Giorgino, "Computing and visualizing dynamic time warping alignments in R: The dtw package," *J. Stat. Softw.*, vol. 31, pp. 1–24, 2009.

[33] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, "Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs," *PeerJ*, vol. 19, 2013.

[34] W. D. Hairston, K. W. Whitaker, A. J. Ries, J. M. Vettel, J. Cortney Bradford, S. E. Kerick, and K. McDowell, "Usability of four commercially-oriented EEG systems," *J. Neural Eng.*, vol. 11, 2014.

[35] J. I. Ekandem, T. A. Davis, I. Alvarez, M. T. James, and J. E. Gilbert, "Evaluating the ergonomics of BCI devices for research and experimentation," *Ergonomics*, vol. 55, pp. 592–598, 2012.

[36] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *Proc. EAI Mobihealth*, 2014, pp. 39–42.

[37] W. Boucsein, *Electrodermal Activity*. New York: Springer, 2012.

[38] S. Spagnol, C. Saitis, M. Bujacz, O. I. Johannesson, K. Kalimeri, A. Moldoveanu, A. Kristjansson, and R. Unnthorsson, "Model-based obstacle sonification for the navigation of visually impaired persons," in *Proc. DAFx*, Brno, Czech Republic, 2016, pp. 309–316.

[39] S. Spagnol, C. Saitis, K. Kalimeri, O. I. Johannesson, and R. Unnthorsson, "Sonificazione di ostacoli come ausilio alla deambulazione di non vedenti (Obstacle sonification as navigation aid for the visually impaired)," in *Proc. XXI Colloquio di Informatica Musicale (Colloquium on Music Informatics)*, Cagliari, Italy, 2016.

[40] F. S. Bao, X. Liu, and C. Zhang, "PyEEG: An open source Python module for EEG/MEG feature extraction," *Comput. Intell. Neurosci.*, vol. 2011, Article ID 406391, 2011.

[41] S. J. Roberts, W. Penny, and I. Rezek, "Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing," *Med. Biol. Eng. Comput.*, vol. 37, pp. 93–98, 1999.

[42] G. Pfurtscheller and F. H. Lopes Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, 1999.

[43] J. J. B. Allen, J. A. Coan, and M. Nazarian, "Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion," *Biol. Psychol.*, vol. 67, pp. 183–218, 2004.

[44] J. T. Cacioppo and L. G. Tassinary, "Inferring psychological significance from physiological signals," *Am. Psychol.*, vol. 45, pp. 16–28, 1990.

[45] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, pp. 80–91, 2010.

[46] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[47] J. Kim and E. André, "Emotion-specific dichotomous classification and feature-level fusion of multichannel biosignals for automatic emotion recognition," in *Proc. IEEE MFI*, Seoul, Korea, 2008, pp. 114–119.

[48] M. Naji, M. Firoozabadi, and P. Azadfallah, "Classification of music-induced emotions based on information fusion of forehead biosignals and electrocardiogram," *Cogn. Comput.*, vol. 6, pp. 241–252, 2014.

[49] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, vol. 38, pp. 963–974, 1982.

[50] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, pp. 162–175, 2015.

Charalampos Saitis holds an MA in Sonic Arts from Queen's University Belfast and a PhD in Music Technology from McGill University. He is currently Humboldt Research Fellow at the Technical University of Berlin. His work focuses on psychoacoustics and auditory semantics, particularly on approaches relating to embodied cognition and crossmodal processing. Other research interests include haptics in music and affective biosignal computing for cognitive load assessment.

Kyriaki Kalimeri holds a PhD in Brain and Cognitive Science from the University of Trento and a Diploma in Electrical and Computer Engineering from the Technical University of Crete. She is currently Researcher at Fondazione ISI. Her research lies at the intersection of social science and engineering, employing machine learning techniques for predicting psychometric and affective profiles as well as behavioral nudging from multimodal data such as smartphones, social media, and biosignals.