Breeze: Smartphone-based Acoustic Real-time Detection of Breathing Phases for a Gamified Biofeedback Breathing Training

CHEN-HSUAN (IRIS) SHIH, ETH Zürich, Switzerland
NAOFUMI TOMITA, Geisel School of Medicine at Dartmouth, USA
YANICK X. LUKIC, ETH Zürich, Switzerland
ÁLVARO HERNÁNDEZ REGUERA, Universidad de Sevilla, Spain
ELGAR FLEISCH, ETH Zürich and University of St. Gallen, Switzerland
TOBIAS KOWATSCH, ETH Zürich and University of St. Gallen, Switzerland

Slow-paced biofeedback-guided breathing training has been shown to improve cardiac functioning and psychological well-being. Current training options, however, attract only a fraction of individuals and are limited in their scalability as they require dedicated biofeedback hardware. In this work, we present Breeze, a mobile application that uses a smartphone's microphone to continuously detect breathing phases, which then trigger a gamified biofeedback-guided breathing training. Circa 2.76 million breathing sounds from 43 subjects and control sounds were collected and labeled to train and test our breathing detection algorithm. We model breathing as inhalation-pause-exhalation-pause sequences and implement a phase-detection system with an attention-based LSTM model in conjunction with a CNN-based breath extraction module. A biofeedback-guided breathing training with Breeze takes place in real-time and achieves 75.5% accuracy in breathing phases detection. Breeze was also evaluated in a pilot study with 16 new subjects, which demonstrated that the majority of subjects prefer Breeze over a validated active control condition in its usefulness, enjoyment, control, and usage intentions. Breeze is also effective for strengthening users' cardiac functioning by increasing high-frequency heart rate variability. The results of our study suggest that Breeze could potentially be utilized in clinical and self-care activities.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**; • **Computing methodologies → Machine learning approaches**; • **Applied computing → Health informatics**; • **Hardware → Signal processing systems**;

Additional Key Words and Phrases: breathing training, breathing detection, real-time, smartphone microphone, acoustic signal processing, deep learning, gamified biofeedback

ACM Reference Format:
Chen-Hsuan (Iris) Shih, Naofumi Tomita, Yanick X. Lukic, Álvaro Hernández Reguera, Elgar Fleisch, and Tobias Kowatsch. 2019. Breeze: Smartphone-based Acoustic Real-time Detection of Breathing Phases for a Gamified Biofeedback Breathing Training. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 152 (December 2019), 30 pages. https://doi.org/10.1145/3369835

Authors' addresses: Chen-Hsuan (Iris) Shih, ETH Zürich, Zürich, Switzerland, ishih@ethz.ch; Naofumi Tomita, Geisel School of Medicine at Dartmouth, Lebanon, USA; Yanick X. Lukic, ETH Zürich, Zürich, Switzerland, ylukic@ethz.ch; Álvaro Hernández Reguera, Universidad de Sevilla, Sevilla, Spain, alvherreg@alum.us.es; Elgar Fleisch, ETH Zürich and University of St. Gallen, Zürich, St. Gallen, Switzerland; Tobias Kowatsch, ETH Zürich and University of St. Gallen, Zürich, St. Gallen, Switzerland.

2474-9567/2019/12-ART152
https://doi.org/10.1145/3369835

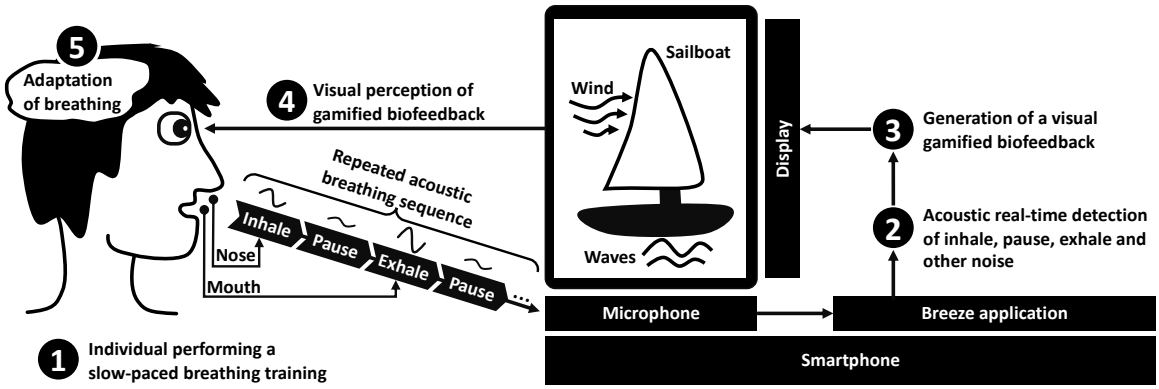


Fig. 1. Overview of Breeze, a mobile gamified biofeedback breathing training.

1 INTRODUCTION

Slow-paced breathing training is an effective method to strengthen cardiac functioning [26, 45, 62] and psychological well-being [27, 65, 70]. It thus addresses two of the most pressing global health challenges: chronic disease [42] and mental illness [55]. Slow-paced breathing training can be self-paced (e.g., during mindfulness meditation) [65], externally-guided by following acoustic, seismic, or visual instructions [62], or biofeedback-guided where individuals consciously control their own heart rate variability or breathing rate by monitoring the bio-signals [26]. In addition to various physiological benefits [16, 18, 26, 33, 45, 49, 51], biofeedback-guided training, if mastered, can also improve self-efficacy [68, 74] (i.e., an individual's believe in their ability to perform an action), which is a relevant predictor of health behavior [66, 67].

There are, however, limitations that hinder both the effectiveness and the reach of slow-paced breathing training. First, biofeedback-guided breathing training is not scalable because it requires trained therapists and additional equipment that senses bio-signals [16, 18, 49], which is particularly critical in developing countries [76]. Second, the use of interventions that incorporate slow-paced breathing training is less prevalent in males, less-educated individuals, and physically inactive individuals [11].

To address these limitations, the current work presents Breeze, a mobile application that uses a smartphone's microphone to continuously detect breathing phases in quasi-real-time (i.e., inhale, exhale, and pause between inhale and exhale), which are then used to trigger a gamified biofeedback-guided breathing training. An overview of Breeze is depicted in Figure 1. Breeze is a scalable mobile application as it relies solely on a smartphone, which is curently available to circa 76% of individuals in high-income countries and to circa 45% of individuals in middle and low-income countries [71]. Breeze therefore offers the benefits of biofeedback without requiring additional equipment due to recent advancements in smartphone-based acoustic signal processing [6, 12, 44, 54]. Moreover, gamified biofeedback in Breeze (e.g., a successful slow-paced breathing increases the speed of a sailboat) targets experiential outcomes (e.g., enjoyment) in addition to instrumental outcomes of strengthening cardiac functioning or psychological well-being [12, 18, 50, 59, 81]. This experiential value of Breeze can be used to reach those who are less motivated to perform slow-paced breathing training [11].

The following research questions (RQs) guide this work and are based on limited evidence of the technical performance, user acceptance, and the physiological effect of a smartphone-based gamified biofeedback breathing training [12]:

(1) *How accurate are breathing phases detected in quasi-real-time with a smartphone's microphone?*
(2) *How efficiently does this real-time detection algorithm run on a smartphone?*
(3) *How is a smartphone-based gamified biofeedback breathing training perceived?*

(4) *Does a smartphone-based gamified biofeedback breathing training positively impact physiology?*

This work must address several challenges and to answer these RQs. First and foremost, an acoustic breathing phases detection algorithm has to be developed that is able to process and classify continuously incoming breathing data. Second, the detection algorithm requires the collection of appropriate acoustic training data to generalize well to different mobile microphone technologies and various inter-individual and intra-individual characteristics of individuals performing a slow-paced breathing training. Third, the algorithm has to be designed and implemented on a smartphone that process the quasi-real-time detection with low detection latency. Fourth, the false positive rate of the detection has to be minimized, because inaccurate biofeedback may induce stress in individuals and, thus, could have a negative effect on the physiological and psychological outcomes, which, in turn, could lead to non-usage of the application. Fifth, related work on the design of a smartphone-based detection algorithm, particularly in the context of slow-paced breathing training, is limited (see Section 2.2). Finally, to the best of our knowledge, there are no evidence-based guidelines for the design of an effective gamified biofeedback visualization that is delivered through a smartphone screen (see Section 2.3).

Against this background, the current work has the following objectives:

(1) **To review related work:** The primary purpose of this review is to lay the groundwork for Breeze, the smartphone-based gamified biofeedback breathing training. To this end, we review literature related to slow-paced breathing (Section 2.1), detection of breathing sound with a smartphone's microphone (Section 2.2), and technology-supported breathing training (Section 2.3).
(2) **To present the breathing detection function of Breeze:** Here, we describe the design of the acoustic detection pipeline that classifies continuously incoming acoustic data using a smartphone's microphone. The pipeline consists of a pre-processing step followed by a breath extraction module that uses a convolutional neural network to filter out non-breath sound. In a third step, a breathing phases detection module uses an attention-based bidirectional long short-term memory to continuously detect acoustic inhale-pause-exhale-pause sequences, followed by a post-processing step that filters out detection noise. In addition, we describe the implementation of the breathing detection on a smartphone (Section 3.1).
(3) **To present the gamified biofeedback of Breeze:** Justificatory knowledge from gamified biofeedback literature and research on environmental well-being is used to describe the gamified elements and mechanics of the biofeedback visualization component of Breeze (Section 3.2).
(4) **To report and discuss the evaluation of Breeze:** A data collection study with 43 subjects (Section 4.1) and a pilot study with 16 new subjects (Section 4.2) were carried out to assess the detection performance (RQ1, Section 4.3) and the efficiency of the smartphone implementation (RQ2, Section 4.4) of Breeze. Moreover, user acceptance (RQ3, Section 4.5) and the physiological impact (RQ4, Section 4.6) of Breeze are assessed in comparison to a validated active control condition. The results are discussed, and limitations and suggestions for future work are outlined (Section 5).

In successfully pursuing these objectives, this work presents, for the very first time, implementation details and empirical evidence of the technical performance, user acceptance and efficacy of a scalable gamified biofeedback breathing training delivered solely by a smartphone. It thus contributes to the interdisciplinary field of digital health at the intersection of computer science, biological psychology and behavioral medicine with the potential to positively impact chronic disease and mental illness by strengthening cardiac functioning and psychological well-being.

2 RELATED WORK

2.1 Relevance, Working Mechanisms and Techniques of Slow-paced Breathing Training

Mental illness and chronic disease impose a significant burden on individuals' well-being and health economic costs worldwide [10, 42, 55, 80]. Even worse, mental illness worsens the impact of comorbid chronic disease

[21, 32, 53]. Socioeconomic, cultural, behavioral (lifestyle) and genetic factors are the driving forces behind these global health challenges and require appropriate preventive and therapeutic interventions [37, 42].

To this end, slow-paced breathing training has been proposed, because it strengthens cardiac functioning [26, 45, 62] and with it, psychological well-being [27, 65, 70]. Cardiac functioning can be quantified as high frequency heart rate variability (HF-HRV), i.e., the portion of the beat-to-beat interval between 0.15 and 0.4 Hz [43]. In fact, decreased HF-HRV is linked not only to stress [40], depression [38] or anxiety disorders [13] but also to type-2 diabetes mellitus [7], chronic obstructive pulmonary diseases [60] or hypertension [48]. Moreover, lower values of HF-HRV have been associated with death and disability [75]. Evidence suggests that slow-paced breathing increases HF-HRV [45, 62] and, with it, improves mental health outcomes such as negative affectivity [65], stress, anxiety symptoms [27] and depression [70]. Moreover, the positive impact of slow-paced breathing training has also been shown in chronic pain, asthma, chronic obstructive pulmonary diseases, cardiac rehabilitation, coronary artery disease, and hypertension [26]. It has even been demonstrated that slow-paced breathing training has the potential to treat substance use disorders [19] and it is used in various mindfulness-based stress reduction interventions [34, 39, 65].

Several working mechanisms of slow-paced breathing have been suggested, such as strengthening the function of the baroreceptors (i.e., sensors of blood pressure), stimulation of the parasympathetic nervous system to reduce stress reactions, increased gas exchange efficiency and mechanical stretching of the airways (e.g., relevant to individuals with respiratory diseases), meditation effects through focused breathing, and anti-inflammatory effects [45, 46]. So far, most empirical studies show evidence for the baroreceptor mechanism [45]. Here, slow-paced breathing leads to heart rate oscillations that are significantly higher compared to a condition at rest. This is due to the coherence of a particular breathing frequency and oscillations in blood pressure and heart rate. Individual characteristics of the cardiovascular system and activity in the baroreceptors also contribute to this effect [45]. Studies have shown that maximum heart rate oscillations are triggered by breathing frequencies that lie around six breaths per minute in adults and children [45, 62, 77].

Additionally, it has been shown that a particular breathing technique, i.e., four seconds of inhalation, followed by two seconds of exhalation and four seconds of pause, is effective in increasing HF-HRV [62]. We thus adopt this 4-2-4 breathing sequence in the current work and, as a resulting requirement, Breeze has to detect the corresponding acoustic breathing phases, i.e., inhale, exhale and pause between the two. The detection of these phases allows Breeze not only to provide and leverage the benefits of biofeedback but also to calculate derived parameters, such as breathing frequency, average inhale and exhale duration, and inhalation-exhalation ratios [56, 62]. These additional parameters can be monitored over time and used as indicators of training performance by clients and health professionals alike.

2.2 Breathing Detection with a Smartphone Microphone

Advancements in smartphone-based acoustic signal processing have led to various applications of respiratory sound detection. An overview of related work is shown in Table 1. Almost all applications focus on the detection of abnormal breathing sound linked to respiratory diseases or sleep disorders. In these works, variability in audio sources or smartphones is barely studied, except for the work by Fisher et al. [22], which examines the feasibility of detecting snore and breath phases with participants' own smartphones. Results from seven participants showed high accuracy on snore detection, yet low sensitivity on normal breath inhalation (sensitivity = 0.34). Fisher et al. further address the challenges of detecting nearly inaudible inhalation sound and the variation in sound quality across different smartphones. The problem of the usage of inhalation sound has also been discussed by Larson et al. [44], whose work focuses on the estimation of the flow rate from forced exhalation sound alone to infer lung function (mean error = 0.05). To ensure the audibility of breathing sound, some works place the smartphone headset microphone directly underneath the nose to record nasal sound and, thus, to detect breathing [6, 54].

Most of the related work shares conventional approaches towards acoustic breathing detection; steps include signal pre-processing, feature engineering, and machine learning. Filtering and de-noising are often used to ensure sound quality [2, 6, 57]. Breath sound features are mostly derived from frequency domain and Mel-Frequency Cepstral Coefficients (MFCCs), which model the signals based on the human sense of hearing, is widely used in different applications [4, 31, 57, 69]. Support-vector machines (SVMs) are the most applied method in the work listed in Table 1. While [2] involves breath phases, the phases are pre-segmented by physicians as categories to evaluate the best phase for asthmatic breath detection. Most of the approaches from the related work are designed for offline analysis. Specifically, the signals are processed as event-based (e.g., each label represents a 4-second sound event) and case specific (e.g., tuned the features with defined thresholds [9] or additional recorded sensor features such as actigraphy [6]), instead of as a continuous detection for general breath monitoring. Real-time applications in the field face challenges of capturing short time frames information, and extra setup is often required for sound calibration. For example, Larson et al. use the participant's height to estimate the arm length to perform inverse radiation modeling to compensate for the sound pressure losses due to distance. RunBuddy [31] uses breathing sound as features to measure running rhythm and achieve an accuracy of 0.927. A pre-recording step was required to capture participants' breathing profiles for calibration to perform breath detection. BreathPrint [14], on the other hand, uses the fact that breathing sounds differ across individuals to perform user authentication via breathing gestures.

Table 1. Overview of applications that use a smartphone's microphone to detect breathing sound.

Authors	Application	#D	Distance	Algorithm	BPD	RTD	ACC
Larson et al. '12 [44]	Lung volume estimation	1	In front of face	BDT+k-means	no	Exhale flow	*
Behar et al. '14 [6]	OSA screening	1	Close to nose	SVM	no	OSA/not-OSA	92%
Hao et al. '15 [31]	Running rhythm detection	1	Close to mouth	Threshold+LRC	no	BR.& Strides	93%*
Nam et al. '15 [54]	Breath rate estimation	1	Close to nose	Autoregression	no	no	94%
Ren et al. '15 [57]	Sleep apnea monitoring	1	Close to head	Envelop+SVM	no	*unclear*	96%
Bokov et al. '16 [9]	Wheezing recognition	1	$5-10cm$ to mouth	SVM	no	no	80%
Fischer et al. '16 [22]	OSA monitoring	6	Unknown	RBs+ANN	yes	no	*
Chauhan et al. '17 [14]	User authentication	2	Close to nose	GMM	no	no	94%
Azam et al. '18 [2]	Abnormal breath detection	1	$25 \pm 5cm$ to mouth	SVM	no	no	75%
Carlier et al. '19 [12]	Breathing training	1	Close to mouth	(no details)	no	"ooom" sound	-
Romero et al. '19 [58]	SDB detection	1	Close to head	DNN	no	no	95%

Note: ACC = Accuracy of detection, ANN = Artificial Neural Network, BDT = Bagged Decision Tree, BPD = Breathing Phases Detection, #D = Number of devices used to record breath noise, DNN = Deep Neural Network, GMM = Gaussian Mixture Model, LRC = Locomotor Respiratory Coupling, OSA = Obstructive Sleep Apnoea, RBs = Robust Boost, RTD = Real=time Detection, SDB = Sleep-disordered Breath, SVM = Support Vector Machine.

Altogether, it is challenging to capture information from normal breathing phases sound collected with a distant smartphone microphone, and no satisfactory result to detect inhalation sound via smartphones has yet been shown. Furthermore, there are a limited number of works where the proposed systems are suitable for performing automatic real-time breathing phases detection with robust interactive biofeedback as a cross-platform smartphone application.

2.3 Technology-supported Breathing Training

Breathing instructions can be communicated via personal consultations, group-based programs or digital technology such as audio CDs or portable music players [15, 16, 41, 82]. Various other technology-supported applications have been proposed and/or implemented and assessed in the field to support slow-paced breathing training. An overview of these technology-supported approaches is listed in Table 2.

Most of these training instructions leverage the benefits of biofeedback and a significant amount employs gamified elements. In addition to smartphones, various computer screens, augmented/virtual-reality visualizations,

and auditory feedback, training instructions can also be delivered by electronically-enhanced everyday objects (e.g., stuffed toys, pendants or fidget spinners). The vast majority of these instructions, however, require dedicated sensors for the presentation of biofeedback and are therefore not scalable. Only the approach by Carlier et al. [12] is close to our work, but their study lacks technical details of the implementation and empirical validation. Moreover, the breathing detection by the "om" sound during exhalation may lead to hoarseness over time and may have a negative impact on individuals nearby.

Against this background, a validated and scalable biofeedback-guided and gamified breathing training with all the potential benefits as outlined in the introduction is still not available today. In an attempt to overcome this shortcoming, we describe the design and evaluation of Breeze in the following sections.

Table 2. Overview of technology-supported slow-paced breathing training.

Author(s)	Mode of instruction	Biofeedback sensor(s)	Gamified	Positive impact on
Grossman et al. '01 [28]	AUD by wearable	RES via chest belt	no	systolic blood pressure
Elliott et al. '04 [20]	AUD by wearable	RES via chest belt	no	systolic blood pressure
Chittaro & Sioni '14a [16]	VIS by SP	(no biofeedback)	no	perceived relaxation
Uratani et al. '14 [77]	Belly movement of toy	(no biofeedback)	yes	(no efficacy study)
Dillon et al. '16 [18]	VIS by SP	SC via mobile device	yes	perceived stress & HR
Sonne & Jensen '16 [72]	VIS by PC	THE on RFduino	yes	HRV
Hao et al. '17 [30]	(future work)	MOT on SW	no	(no efficacy study)
Roo et al. '17 [59]	VIS by AR & VR devices	RES on torso belt & HR by SW	no	perceived mindful state
Lin '18 [49]	VIS by SP	HRV via chest belt	yes	HRV parameters
Frey et al. '18 [24]	VIS, AUD, HAP by pendant	MOT on pendant	no	HRV parameters
Liang et al. '18 [47]	VIS by fidget spinner & SP	HRV via fidget spinner	yes	perceived stress
Carlier et al. '19 [12]	VIS by SP	'om' exhalation noise	yes	(no efficacy study)

Note: AR = augmented reality, AUD = audio modality, HAP = haptic modality, HR = heart rate, HRV = heart rate variability, MA = muscle activity, MOT = motion, PC = personal computer, RES = respiratory, SC = skin conductance, SP = smartphone, SW = smartwatch, THE = thermistor, VIS = visual modality, VR = virtual reality.

3 DESIGN OF THE SMARTPHONE-BASED BREATHING TRAINING

3.1 Design and Implementation of the Real-time Breathing Detection Pipeline

Figure 2 illustrates a flow-chart of Breeze's real-time breathing detection pipeline of Breeze. For every one-second, incoming signals are processed through the following blocks: *Pre-processing and Feature Extraction*, *Breath Extraction Module (BEM)*, and *Breathing Phases Detection Module (BPDM)*. Pre-processing applies band-pass filters to eliminate noise signals and extract relevant features. The BEM separates breath sound from not breath noise and the BPDM detects inhale-pause-exhale-pause events, which represents the backbone for generating biofeedback. The detailed design of the algorithms is explained in the following subsection.

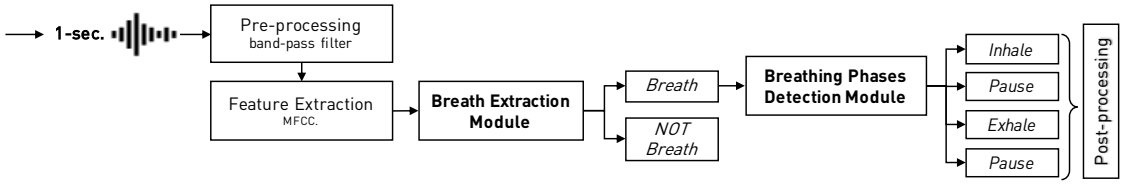


Fig. 2. Processing pipeline of the real-time breathing detection. The incoming signal is buffered and processed every second. The Breath Extraction Module passes only breathing-related noise to the Breathing Phases Detection Module.

3.1.1 Pre-processing & Feature Extraction. To simulate a real-time detection process, the labeled data is segmented into one-second frames, to which a band-pass filter is applied and from which features are extracted. Within each one-second frame, *logMel* and MFCC features are calculated based on sliding windows of 44ms with 75% overlap using the Hamming function. Specifically, the design of the band-pass filter and the *logMel* are inspired by the data annotation process (details in Section 4.1.6). While *logMel* represents the logarithm of the power spectrum mapped with mel-scale frequency as $10 * log10(Mel^2/1.0)$, MFCC adds cepstral analysis, on top of which the Discrete Cosine Transform (DCT) is used to inverse the calculated spectrum, and the resulting coefficients in low resolutions are discarded. The differentials and acceleration coefficients are included, the combination of which has been shown to improve the automatic speech detection [29].

3.1.2 Breath Modeling. We have designed a two-step breath detection model to achieve the objective of detecting accurate breathing phases. It consists of two modules. First, the *Breath Extraction Module (BEM)* is an automatic calibration step for filtering out the non-breath noise sound. Second, the *Breathing Phases Detection Module (BPDM)* which further processes the breath sound to detect four breathing phases (inhale, pause, exhale, and after exhale pause). In the following, we describe the details of these two modules.

Breath Extraction Module: This module serves as a control gate that distinguishes breathing sound from all other sounds in the environment, including speaking, reading, coughing, and laughing (categorizes as *noise*). As a result, only the identified breath sound buffers are transferred to the BPDM to process biofeedback. The module also evaluates if the current environment is suitable for a breathing exercise. For example, when only noises are detected, a recommendation to "find a quieter place" could be triggered. To maximize the information that can be derived from a one-second sound, we propose a Convolutional Neural Network (CNN). CNNs have already been successfully applied to different audio event detection applications [1, 5, 63, 83]. For BEM to identify breath sound, we employ a single layer CNN to extract information from the entire one-second sound features. A convolutional layer operates on a spectral-content image which is converted from sound features, and extracts local patterns by multiplying filters over the whole image. Figure 3 illustrates the CNN architecture. Specifically, ten filters with size $12 \times 12 \times 1$ are applied, and the Rectified Linear Unit (ReLU) activation function is used to accelerate the convergence of gradient descent. A fully connected layer with the softmax function outputs the probability distribution over a set of classes.

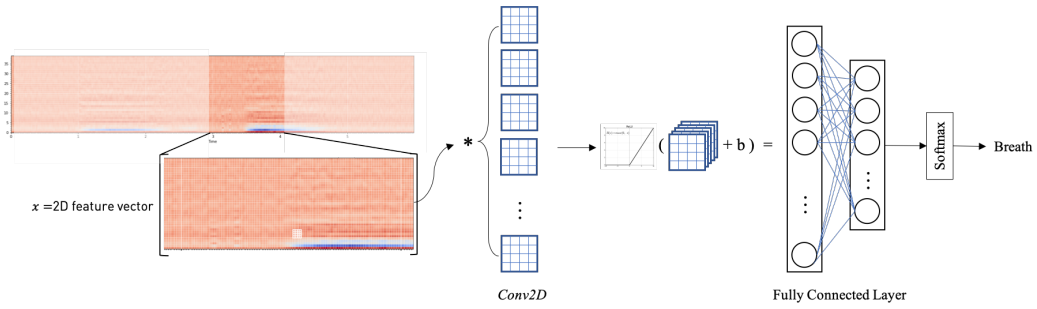


Fig. 3. CNN model.

Breathing Phases Detection Module: In this module, we train a real-time detection model that can continuously monitor the breathing phases. After the one-second breathing signal passes BEM, this module detects the breathing phases (i.e., inhale, exhale, pause, or mix). It has been shown that it is challenging to differentiate inhalation and exhalation due to the similarity of their audio characteristics [17], regardless of the variance between individuals. Therefore, instead of treating the signal segments as independent events, we take into account their relationships

and analyze them as a sequence. Here, we explore three sequential modeling methods including Hidden Markov Model (HMM), Long Short-term Memory Network (LSTM), and attention-based LSTM to evaluate the feasibility of detecting breathing phases. The details of these models are as following:

- *HMM* extends the concept of the Markov-chain to not only model the probability based on an observable sequence (i.e., features vector from breathing signals), but also to incorporate the hidden states (i.e., inhale, exhale, pauses in between or mix). HMM is widely applied in recognizing temporal patterns and speech [25]. We use a supervised HMM based on the nature of breathing cycles, which can be seen as sequences with repetitive patterns. We hypothesize that a single dependency from two consecutive time steps represents the breath patterns. In detail, a supervised HMM calculates the transition and emission probability based on the labels of all the training data. Since the posteriors are from the labels, we only need the maximization step during the iteration of the maximum likelihood estimation.
- *LSTM* networks, a variant of recurrent neural networks, are powerful sequential modeling methods which have the capability of learning long-term dependencies. A single LSTM cell consists of input, output, and forget gates. Over any time interval, the cell maintains the dependencies, and the gates control the flow of input and output information. For each time step, we use 44ms of breathing signal as input. Inhalation, exhalation or pauses are greater than or equal to one second, which an equivalent of a sequence with 87 samples is sufficient to represent one breathing phase. Thus, LSTMs have the advantage of learning the breath phase from the whole one second. We evaluate a Bi-directional LSTM (BiLSTM) architecture, where not only the forward cell contributes to the current hidden state, but the backward cell in the LSTM layer also contributes, as illustrated in Figure 4. Thus, the output is

$$\hat{y}_t = softmax(weight_y * h_t + bias_y), h_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]$$

  where *softmax* is a generalized sigmoid function that outputs the probability of the classes, and h_t is the forward and backward states.
- *Attention-based LSTM* is a model where an attention mechanism is introduced to enhancing memory of the network by focusing on certain part of the long sequences [3]. We add an attention layer on top of BiLSTM network to learn which segments of the breath sequence are of importance and closely related to the current segment. As shown in Figure 4, the attention function calculates attention scores and form a context vector as follow:

$$\text{Attention score: } \alpha_{t,i} = align(y_t, x_i)$$

$$\text{Context vector: } c_t = \sum_{i=t-1}^{N} \alpha_{t,i} h_i$$

  where t indicates the current time step and the *align* function calculates how well the input is matched with the output through a single layer feedforward network. Then, a fully connected layer aggregates the output from the attention layer and the softmax function outputs the probability of the classes.

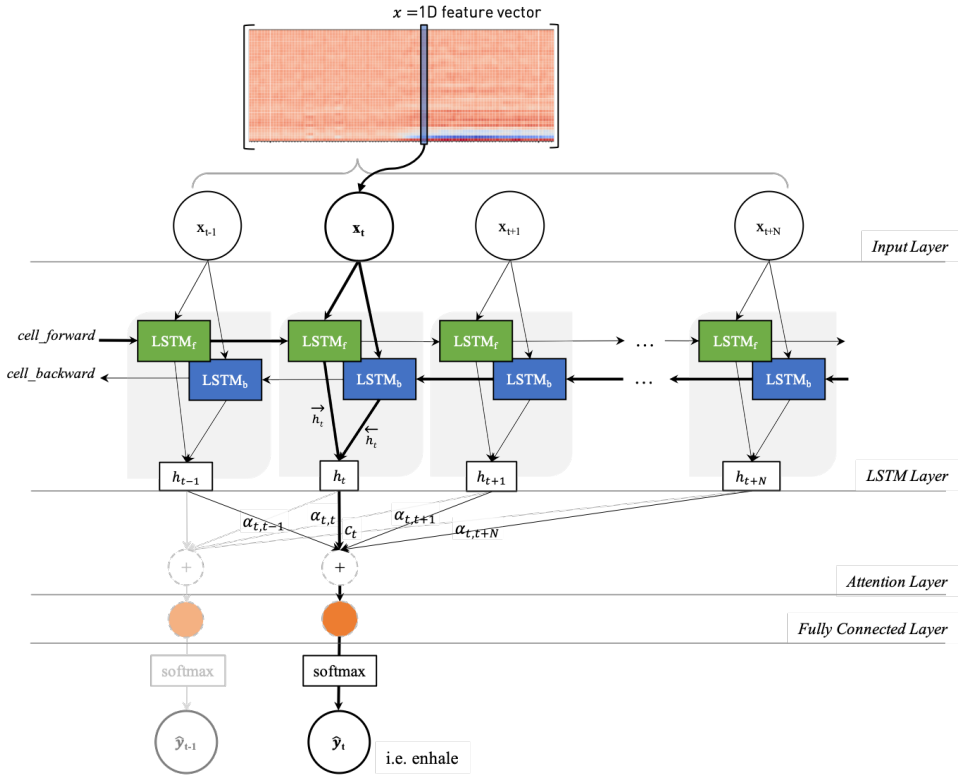


Fig. 4. Attention based bidirectional LSTM model.

Postprocessing: To optimize the breathing phases detection performance for real-time biofeedback, we employ a post-processing step to smooth the model's output. For each detected one-second breath sound, BPDM outputs a sequence of 87 predictions. Based on the 4-2-4 breathing instruction, the majority of the one-second sound should present one breathing phase. Thus, we take the mean of the BPDM's output probabilities, *logits*, and the $argmax(logits)$ is the resulting breath phase for each one-second input sound.

3.1.3 Smartphone implementation. We use the TensorFlow library to train the models with the data collected from the data collection study (Section 4.1). The trained models are frozen and ported into a serialized file that can be read in an Android application using the TensorFlow Lite API. We implement the detection pipeline in Java and the Android API to support native execution on the Android operating system, which reduces overhead at run-time. Breathing phases detection is performed with the ported models in the pipeline to eventually trigger visual biofeedback events in Breeze, which we describe in the following section.

3.2 Design of the Gamified Biofeedback-guided Visualization

Breeze uses the results from the breathing phases detection to facilitate a gamified, biofeedback-guided visualization for slow-paced breathing training. In line with prior work [12, 18, 59, 81], gamification elements are used to increase the experiential value (e.g., enjoyment) of this training [50]. It is assumed that this experiential value increases the reach of slow-paced breathing training in population groups with lower prevalence rates of usage,

such as males, less-educated, or physically inactive individuals [11]. Moreover, a biofeedback-driven visualization was implemented to improve the self-control perception of individuals and, thus, self-efficacy [68, 74].

3.2.1 Design Choices. The users' objective in Breeze is to repeatedly speed-up a sailing boat by following the 4-2-4 breathing pattern and, thus, get the sailboat to travel a greater distance in a six-minute breathing training. During the breathing training, the boat sails further in distance, and a constantly changing landscape provides additional motivation. Considering previous work, which indicates a calming and meditative effect of gardens, trees, and general green spaces [35, 59], the landscape is set in a natural environment. To further promote calmness and stimulate an approach motivation, mostly cool colors (e.g., green, blue) are used [52, 73].

Aside from the static environment, the game uses two dynamic objects: the sailing boat and the wind. The movements of these objects mediate the guidance for the breathing cycles as well as to provide biofeedback.

3.2.2 Reference Visualization. Visual changes to the dynamic objects indicate the active reference breathing phase. During the inhalation phase, the wind, represented as tailed particles, moves against the movement direction of the boat (towards the view) and during the exhalation phase the wind moves in the direction of the boat's movement (away from the view). During the pause phase, the wind stops blowing and a few tailed particles are shown without moving in a specific direction. Furthermore, the speed of the boat and its sail behave in accordance with the wind direction and strength. Specifically, the sail inflates and the boat accelerates during the exhalation phase, while the boat decelerates until it reaches a minimal movement speed in the pause and inhalation phases. To infer the current reference breathing phase, the sail takes a neutral position in the pause phase and deflates in the inhalation phase. A semi-transparent UI element, showing an arrow or a horizontal line, is additionally used to help the users identify the current reference breathing phase. The arrow in upwards and downward directions represents the inhalation and exhalation phases, respectively, and a horizontal line indicates the pause phase. (Figure 5).

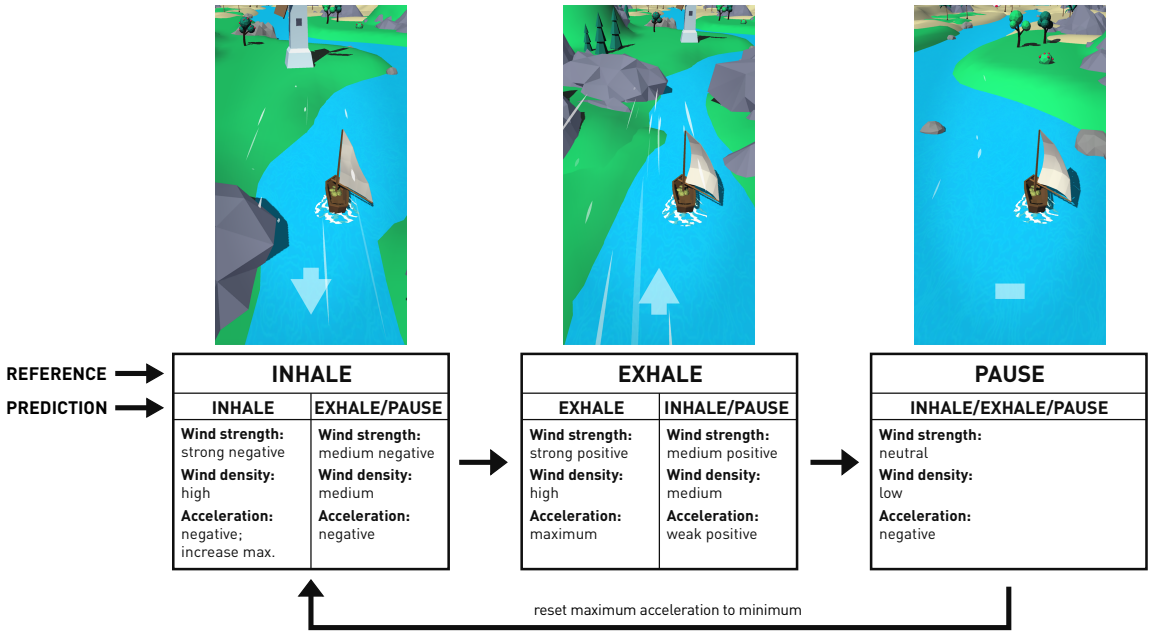


Fig. 5. Exact responses of the biofeedback-driven visualization to detected breathing phase sounds.

3.2.3 Biofeedback. The biofeedback is incorporated into the dynamic reference visualizations that serve as guidance for the breathing cycles. The logic is shown in Figure 5. It consists of two aspects: the wind strength and density, and the acceleration and speed of the boat. As long as a correct inhalation is detected during the reference inhalation phase, the wind becomes stronger, indicated by more tailed particles with longer tails. The same applies if an exhalation is detected during the reference exhalation phase. The power of the acceleration is determined during the preceding inhalation phase. If an inhalation is detected during this phase, the possible acceleration is linearly increased until it reaches a predefined maximum value. If no inhalation was detected during the inhalation phase, an exhalation during the subsequent exhalation phase just applies the minimum acceleration. The minimum acceleration is still higher than the basic acceleration during the exhalation phase, thus, a correct exhalation has a noticeable effect on the boat's acceleration even when no correct inhalation was detected in the preceding inhalation phase. During the pause phase or if no correct inhalation is detected, a negative acceleration is applied to the boat to slow it down until it reaches the minimum speed or is again accelerated due to a correct exhalation.

3.2.4 Implementation. The gamified biofeedback-guided visualization of Breeze is implemented using the Unity game engine (Version 2019.1.8). For communication between the models running in the Java part of the application and the game, Unity's AndroidJNI module is used. All 3D models have been custom-made using the Blender 3D modeling software.

4 EVALUATION OF BREEZE

Two studies were carried out to answer the four research questions (RQs) of the current work. First, a data collection study (N=43) was conducted to train Breeze's two breathing detection models and to assess their performance in an offline setting. Second, a pilot study (N=16) including a pretest (N=3) was carried out to assess the smartphone-based quasi-real-time detection performance of Breeze. In addition, the pilot study was used to assess the technical efficiency of the smartphone implementation, technology acceptance and impact of Breeze on physiological parameters. In the following subsections, we describe the two studies in detail (Section 4.1 and 4.2) and report the findings with respect to the detection performance (RQ1, Section 4.3), technical efficiency (RQ2, Section 4.4), technology acceptance (RQ3, Section 4.5), and physiological effects (RQ4, Section 4.6).

4.1 Data Collection Study

4.1.1 Design. A within-subject crossover design was used for the data collection study. The flow chart and the setup of the mobile recording devices of the study are shown in Figure 6 and Figure 7. To develop a robust and device-agnostic breathing detection algorithm for Breeze, variance of breathing sounds was increased experimentally by instructing the subjects of the study to perform two breathing techniques (i.e., normal chest breathing and deep abdominal breathing). Moreover, several intentional control sounds were recorded from each subject (e.g., coughing, reading a text, and laughing). Further control sounds were recorded (e.g., traffic or multi-person chatting sounds). In addition, four different mobile devices were used to account for different form-factors and manufacturers of microphone sensors.

4.1.2 Procedure. Each session of the data collection study, it was carried out in the following steps: First, subjects were asked to sit quietly on a chair in front of the mobile devices that were positioned at arm's-length distance on a table with the microphone facing upwards as depicted in Figure 7. Second, subjects were equipped with a belt that measured their breathing pattern by the expansion and contraction of the belly. Third, subjects were randomly assigned to one of two groups that differed in the order of the two breathing tasks. That is, the first group was instructed to start with a slow-paced abdominal breathing followed by a normal chest breathing. In contrast, the order of the other group was reversed. This cross-over design was employed to counterbalance any

order effects and to increase the variance of the acoustic breathing sound. During each of the two breathing tasks, subjects were instructed to inhale through their nose and to exhale through their mouth for three minutes. After the two breathing tasks, subjects were asked to produce control sounds. Examples include inhalation through the mouth and exhalation through the nose, throat clearing, coughing, laughing, and reading a text aloud. These control sounds were used to increase the performance of Breeze's breathing detection.

Overall, each session lasted approximately 50 minutes and each subject received a small monetary compensation. Details of the audio and breathing recordings, the study population, and the data labeling are outlined in the next subsections.

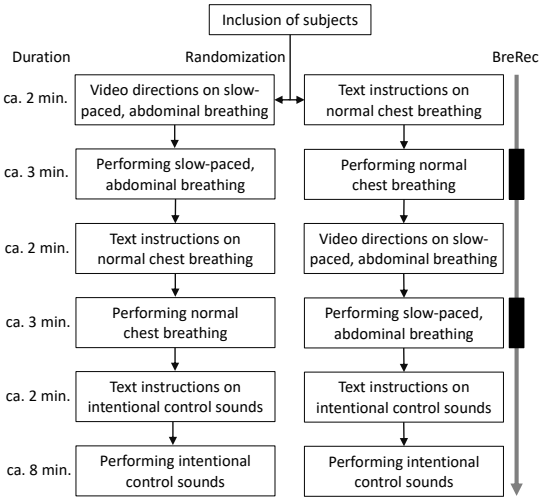


Fig. 6. Flow chart of the data collection study. Note: Intentional control sounds were, for example, coughing, laughing or reading a text; BreRec = Breathing recording with a respiration sensor.

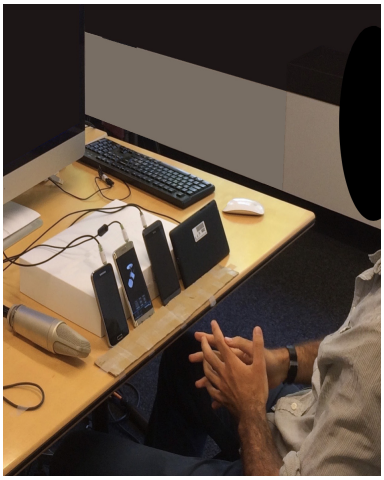


Fig. 7. Recording setup. Note: A computer screen was used to provide the instructions; the mobile devices were positioned with their microphones facing upwards; From left to right: NT1000, iPhone 5, Galaxy S5, One M8, and Nexus 7

4.1.3 Audio recordings. During each experimental session, audio was recorded simultaneously for the following four mobile devices: Apple iPhone 5, HTC One M8, Samsung Galaxy S5, and Google Nexus 7. A Rode studio microphone (NT1000) was added as a reference in terms of audio quality. An audio interface (Focusrite Scarlett 18i20 USB) and audio software (Audacity, V2.1.3) were used for this purpose. The configuration for the recordings was set to the best resolutions capable for a smartphone application, i.e., a sampling rate of 44.1 kHz with a resolution of 16 bits PCM.

4.1.4 Physiological recording of breathing. To support the labeling process of the acoustic breathing sounds, Mindmedia's NeXus respiration sensor was attached around the belly of each subject. The sensor was connected to the medical device NeXus 10 and, in combination with the Biotrace software (V2016), raw breathing data was recorded with a sampling frequency of 32 Hz.

4.1.5 Study population. Subjects were sampled from within the first author's institution. Consistent with prior work and to reduce error variance in the breathing data, subjects were excluded if they have cardiovascular diseases, severe respiratory diseases, or mental health diseases [62]. Overall, data from 43 subjects, 31 (70.5%) female, with an average age of 25.9 (SD = 4.80) was used.

4.1.6 Data labeling. Overall, the audio recordings consist of circa 19 hours of breathing sound, 8 hours of intentional control sound and 20 hours of surrounding noise including traffic sound, sound of individuals walking around, door slamming sound and indistinct multi-person chatting sound. While the non-breath sounds are

labeled as noise, the rest requires further processes. In order to provide biofeedback that is specific to inhalation sound, exhalation sound and pause, each individual's breathing phases have to be identified and labeled. Thus, we propose a semi-automatic labeling method, in two steps, to label the breathing phases. In the first step, breathing cycles are identified, and in the second step, individual breathing phases are labeled. Details of the two steps are outlined in the following paragraphs.

Step 1 - Identification of the breathing cycles: To identify each completed breath cycle, we compute the envelopes of mel-scale spectrograms of the audio signals over time and validated the extraction with the data from the respiration sensor. Since all recordings were synchronized, we take the data from the Rode studio microphone as a reference to identify the breathing cycles. The mel-scale models the none-linearity of human-hearing by emphasizing the lower frequency changes through a mel-scale filter bank. The spectrogram is a visual representation of sound which shows the power across frequencies over time. Given the mel-scale spectrogram matrix as M, we computed the $logMel = 10\log_{10}(M/ref)$ which converts the power amplitude into decibels (dB) with reference $ref = 1.0$ representing 0 dB. From the resulting spectrogram, we hypothesize that the parts with higher dB values represent either inhalation or exhalation. To verify this assumption, we align the data from the respiration sensor with the $logMel$. As shown on the left of Figure 8, the blue curve is the respiratory belt signal, where the ascent and descent are the movement of inhalation and exhalation. With the confirmation from the ground truth data, we further segment each breathing cycle by first extracting the envelope of frequencies over time as follows:

$$Env(t) = \text{Savitzky-Golay}\left(\sum_{j=0}^{t} logMel_{ij}, i \in \{0, ..., n\}\right)$$

Savitzky-Golay filter was applied to smooth the summed signal; t indicates the frames in time from the input signal, and $n = 60$ represents the number of defined mel bands. To extract the segments, a $step$-wise local minimum detection is performed on the Env. Note that the variable $step$ in the detection process is dynamic because breath rate varies among individuals, thus adaptation is required based on the ground truth data, i.e., either to add or to remove the missing or additional indexes respectively. Based on the ground truth, as the example shown in Figure 8, every second local minimum is a cut for a complete breath cycle. Overall, 16,227 complete breathing cycles were extracted from the 43 subjects.

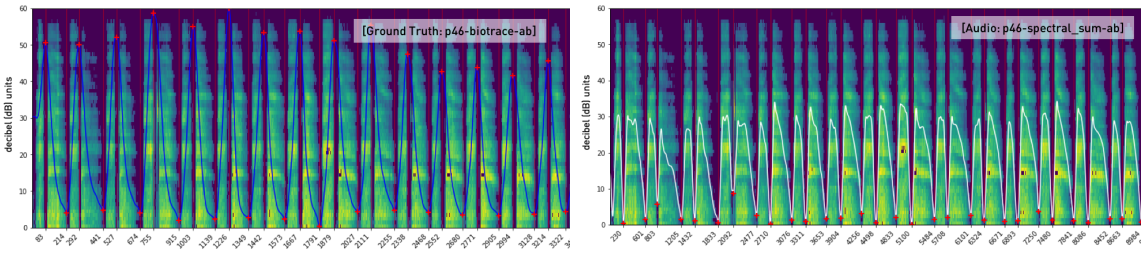


Fig. 8. Inhalation (higher values) and exhalation (lower values) of the respiration sensor (blue line on the left) and the sum of frequencies over time (white line on the right) are drawn on top of the mel-scale spectrogram.

Step 2 - Labeling of the breathing phases: For each breath cycle, we further annotate the breathing phases (i.e., the inhale, inhale pause, exhale, and exhale pause) using the Viterbi decode algorithm. In the previous step, the unclear segments (i.e., breath cycles) can be compensated or eliminated by the ground truth data, yet a similar approach cannot be applied in this step due to the low sampling rate of the respiratory belt which does not capture the short time pause moments. Thus, a different approach was adopted to label the phases, and Figure 9

shows the steps towards the result. In more detail, we first use a band-pass filter with the cutoff values at 1kHz and 8kHz to eliminate the non-essential frequencies. The cutoff values are selected based on observations through the collected breath audios. Then, for the Viterbi decoding process, we extract the signal envelope by calculating the frame-based root-mean-square (RMS) and use logistic mapping to convert the raw RMS into likelihood as follow:

$$p[S = 1|RMS] = \exp(RMS - \tau)/1 + \exp(RMS - \tau)$$

The decision threshold τ is 0.001 in the case shown in Figure 9, and the $S = 1$ indicates either inhale or exhale phases while 0 indicates pauses. Furthermore, the transition matrix is defined as $transition[i, j] = (1 - [0.5, 0.6])$ for all $j \neq i$, which assumes that there is a slightly higher chance to observe inhale/exhale after the inhale/exhale frame, but for pauses frame is equally likely to be followed by either pause or inhale/exhale. Given the observation p and the encoding $transition$ matrix, the Viterbi algorithm [23] calculates the most likely sequence of states (i.e., pause or inhale/exhale) by the observations. An example output is shown as the last step of Figure 9. Note that breath sound varies across individuals, thus the threshold τ is never a static value; instead, it requires manual adjustment to retrieve the optimal cutting indexes. Based on the identified sequence of phases, we generated reference standard labels, where the first cutting index indicates the start of inhalation. Moreover, the annotated inhale-pause-exhale-pause phases have $\tau_{mean} = 0.0035$ and $\tau_{std} = 0.0118$, indicating a wide distribution and a strong skewness of the collected data.

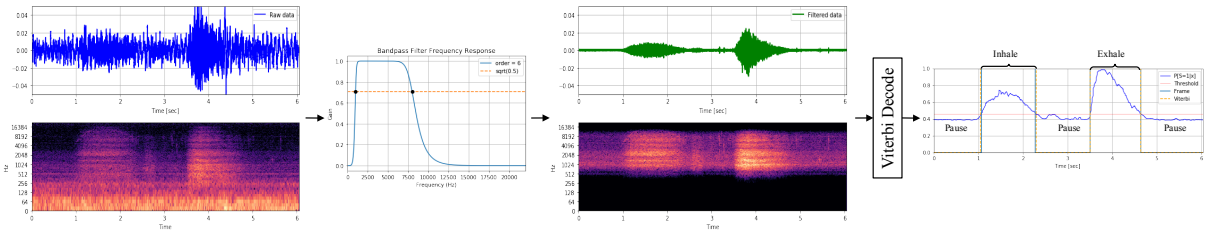


Fig. 9. One complete breathing cycle labeling process. From raw data in both time and frequency domain (left) to designed filter and filtered data, then to the Viterbi decode algorithm to extract the final result (right).

Table 3. Amount of labeled breathing phases.

Device	Inhale	In-Pause	Exhale	Ex-Pause
NT1000	183,285 (23%)	157,867 (22%)	274,759 (21%)	146,710 (23%)
iPhone 5	114,752 (15%)	101,572 (14%)	185,444 (14%)	87,433 (14%)
Galaxy S5	181,513 (23%)	167,398 (23%)	318,299 (24%)	145,457 (23%)
One M8	124,343 (16%)	120,986 (17%)	217,707 (17%)	101,594 (16%)
Nexus 7	182,784 (23%)	168,720 (24%)	319,348 (24%)	146,513 (23%)
Total (100%)	627,707	786,677	716,543	627,707

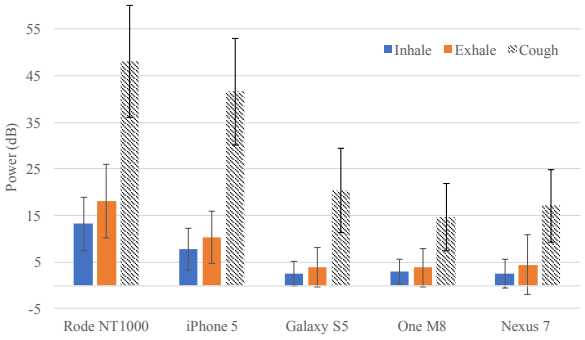


Fig. 10. Mean values and standard deviation (error bars) for the SNR of inhalation, exhalation and cough sounds.

Collected Audios & Additional Noise: Table 3 summarizes the labels that are derived from the data annotation process. The data across devices are unbalanced due to the unsuccessful export of the audio files. Figure 10 shows the mean and the standard deviation of the signal-to-noise ratio (SNR) across devices from all participants. Specifically, the power in decibel of breathing phases signals in relation to background noise is defined as

$$SNR_{dB} = 20log_{10}(Amp_{signal}/Amp_{noise})$$

where *Amp* is the root-mean-square of signal amplitude in time. Since the data are collected in a relatively quiet setup, a silence signal segment is selected as background noise to calculate the SNR of the breathing sound quality. In addition to breath sound, Figure 10 also includes the SNR of the collected cough sound with the same silence segment as background noise as comparison. Even though the signals were recorded in a quiet environment, the SNRs of breathing phases are low, especially in those recordings from the Galaxy S5, One M8, and the Nexus7. There are differences in sound quality across devices, as well as high variance among the participants. Nevertheless, in order to evaluate the robustness of the models, we augment the collected data with white noise. While it is often unknown how the external sound sources are acquired (i.e., under which frequency range), we use white noise, which is a random signal with constant power spectrum density. The white noise is added based on a set of different variance $\sigma = \{0.022, 0.012, 0.007, 0.004, 0.002\}$ which corresponds to the SNR_{dB} range from −15 to +5, calculated as follow:

$$\sigma = \sqrt{10^{log_{10}(Signal_{avg.} - SNR_{dB}^{target})}}, SNR_{dB}^{target} = \{-15, -10, -5, 0, 5\}$$

Note that the $Signal_{avg}$ is the average power of breath signal of all devices and participants, and the final white noises are added equally to all the collected signals. We chose a 20 dB range of corresponding SNR based on the approximate noise level of a quiet room to a noisy office.

4.2 Pilot Study

4.2.1 Design. Consistent with related work [62], a within-subject crossover design was used for the pilot study. The flow chart of the study is shown in Figure 11. A validated breathing training, which has been shown to be effective in increasing HF-HRV [62] by guiding individuals to follow breath cycles of 4 seconds of inhalation, followed by 2 seconds exhalation and a resting period of 4 seconds, was used as a reference for the assessment of Breeze. This active control condition was an externally-guided breathing training without biofeedback adapted and tailored to a smartphone screen as depicted in Figure 12. In particular, subjects had to follow the animation of a white circle. A growing / shrinking circle indicated inhalation / exhalation while no animation indicated that subjects had to hold their breath.

4.2.2 Procedure. Each session of the pilot study was carried out in several steps. First, included subjects were randomly assigned to two groups that differed in the order of the two breathing tasks to counterbalance any order effects. Second, subjects were asked to sit quietly on a chair for the duration of six minutes and a baseline measurement of the physiological data was taken. Third, participants were taught how to perform a slow-paced breathing training via a standardized video clip. Breathing instructions for Breeze or the circle-based breathing training for the active control condition were then provided. Then, subjects had to perform the corresponding breathing training for six minutes. Afterwards, subjects were asked to assess the breathing training via a first self-report assessment. After a washout period of five minutes, instructions for the other breathing training were provided and subjects had to perform and assess that training in a second self-report assessment step. Finally, subjects had to compare both version of the breathing training and were asked to provide demographic data and to indicate prior experience with breathing exercises and biofeedback. Overall, each session lasted approx. 50 minutes and each subject received a monetary compensation worth US $20 for their participation. Details of

the self-reports, physiological measures, log data of Breeze and the study population are outlined in the next subsections.

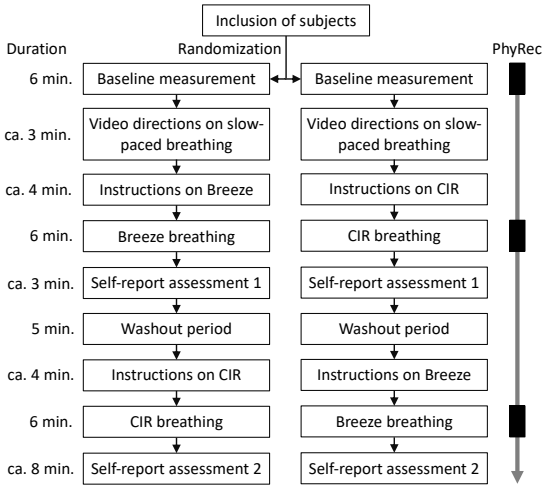


Fig. 11. Flow chart of the pilot study. Note: CIR = circle-based breathing training, a validated active control condition adapted from prior work [62]; PhyRec = Physiological recordings.

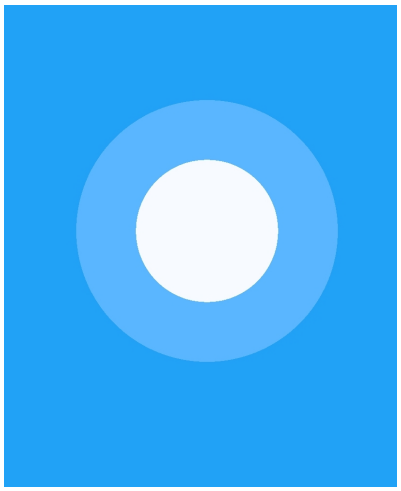


Fig. 12. Screenshot of the circle-based breathing training. Note: A growing / shrinking of the white circle indicates inhalation / exhalation; no animation indicates a resting period [62].

4.2.3 Self-reports. Technology acceptance was measured with five questionnaire items adapted from technology acceptance research [36, 78, 79]. These were perceived ease of use (*It was easy for me to follow this breathing task.*); perceived usefulness (*I was able to relax by following this breathing task.*); perceived enjoyment (*Following the breathing task was enjoyable.*); perceived control (*I felt in control while following this breathing task.*); and the behavioral intention to use (*I would perform this breathing task in my everyday life to better manage stressful situations.*). Each subject had to assess these items after each breathing task. Consistent with prior work [8], subjects were asked to rate single-item statements per construct on 7-point Likert scales ranging from strongly disagree (1) to strongly agree (7). Additionally and during the second self-report assessment step, subjects were asked to indicate their preference for either Breeze or the active control condition via binary forced-choice questions regarding the very same constructs (e.g., for perceived ease of use, the item was formulated as follows: *Which of the versions of the breathing training was the easiest to follow?*). Moreover, subjects had to rate on the same 7-point Likert scales their experience with breathing training (*I am experienced in performing breathing training.*) and biofeedback *I am experienced with biofeedback training*. The subjects were finally asked to indicate details about their age, gender and level of education.

4.2.4 Physiological recordings. High frequency heart rate variability (HF-HRV) and breathing frequency were recorded with BIOPAC's AcqKnowledge software (V4.2), as well as the wireless sensors attached around the belly and via one-lead electrocardiogram electrodes. The sampling rate was set to 2.000 Hz. AcqKnowledge was used to calculate the breathing rate and Kubios HRV Premium (V3.3) for the calculation of the HF-HRV scores during the baseline step and the two breathing tasks for each participant.

4.2.5 Breathing Log of Breeze. The reference of the 4-2-4 breathing pattern, the detected breathing phases and the distance traveled with the sailboat were logged in the Breeze application in combination with a timestamp for each session of the six-minute breathing training.

4.2.6 Study population. Subjects were sampled from the first and last authors' institutions. To reduce error variance in the physiological recordings, subjects were excluded if they were pregnant or reported a history of respiratory disease, cardiovascular disease, gastrointestinal disorder, depression, anxiety disorder or panic disorder. Three male subjects at the average age of 31 (SD = 3.00) with Master degree and almost no experience with breathing training (Mean = 2.33, SD = 0.58) and biofeedback (Mean = 1.33, SD = 0.57) participated in a pretest prior to the pilot study to test the feasibility of the study protocol and the first prototype of Breeze. Sixteen subjects, 7 (43.8%) females, at the average age of 24.6 (SD = 3.44), participated in the pilot study. The level of education ranged from upper secondary education (N=1) to individuals with a bachelor's degree (N=9) or a master's degree (N=6). Their experience with breathing training (Mean = 3.81, SD = 1.94) and biofeedback (Mean = 2.81, SD = 1.87) was limited.

4.3 Performance of the Breathing Detection

Data of the data collection study is split into a training set with 26 (60%) subjects and a test set with 18 (40%) subjects. Different experiments were conducted to evaluate the two-step breath model's performance, including baseline comparison, the detection capabilities on unseen device's recordings, and the robustness to white noises. Moreover, evaluation on the data collected from the pilot study is also presented. Details about the performance metrics, experiment process, and evaluation results are outlined in the following subsections.

4.3.1 Evaluation Metrics. The two-step breath model is evaluated by the following metrics:

- True Positive Rate: $TPR = TP/(TP + FN)$
- False Positive Rate: $FPR = FP/(FP + TN)$
- Positive Predictive Value: $PPV = TP/(TP + FP)$
- Accuracy: $ACC = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i)$
- Mean Absolute Error: $MAE = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples-1}} (\hat{z}_i - z_i)$

It is crucial for BEM to correctly detect breath sound with the fewest false alarm on noise. Thus, we focus the evaluation for BEM on TPR and FPR in the form of a Receiver Operating Characteristic Curve (ROC) and Area Under Curve (AUC). We use PPV, TPR and ACC to assess the detection performance of BPDM regarding the predicted breathing sequence. Here, ACC indicates how well the predicted sequence ($\hat{y}_i$) matches with the corresponding true sequence (y_i). The FPR and PPV are critical metrics for the detection of inhalation and exhalation sounds as they assess the degree to which real-time biofeedback can be provided that is not misleading. A confusion matrix is also used as an overview of the detection performance. We use MAE to evaluate the models' performance on the data collected from the pilot study by looking at the ground truth breathing cycles (z_i) and the detected breathing cycles that are derived from the continuous breath phases detection ($\hat{z}_i$).

4.3.2 Comparison to baseline methods. To evaluate BEM for breath sound extraction, we compare the proposed methods with three methods inspired from related work. The Support Vector Machine (SVM) method is the most applied method to detect abnormal breathing sound. In prior work, MFCC and SVM were used to detect respiratory events versus noise sound and achieved the highest accuracy [57]. The Random Forest (RF) method is used to detect breathing phases and noise [61]. However, the referenced work utilizes features that include the duration of the breathing phases, which cannot be applied in our real-time application. Thus, we exclude features that are related to the dynamic length of the input data to evaluate the performance. This resulted in 107 energy and spectral features. Then, for each decision tree in the RF classifier, a subset of features is selected, and the *Gini index* is used as a cost function to evaluate the split of decisions. The initial values of 13 features as subset per tree and 400 total trees are selected based on [61]. To explore the best possible outcome, we implemented a Gaussian Mixture Model (GMM) with Gammatone Frequency Cepstral Coefficients (GFCC), which is inspired by

BreathPrint [14]. In GFCC, the cepstral coefficients are derived by using equivalent rectangular bandwidth scale instead of mel-scale to emphasize the low frequency signals of breathing. In combination with GMM, we assume that this approach is a reasonable baseline to model the breathing sound so that it is distinguishable from noise.

As a first step, we analyze BEM's performance by training and testing on the studio microphone's data. The method which performs the worst is excluded for the evaluation of the mobile-devices' data. The resulting ROCs and AUCs are depicted in Figure 13, in which CNN with MFCC features as input can extract breathing segments with an AUC of 0.92, followed by CNN with the logMel feature. Based on the results of BEM, we perform the analysis for BPDM with the logMel and MFCC feature set. Table 4 lists the best performance of the baseline method, HMM, and proposed methods in combination with the feature input for the detection of the breathing phases.

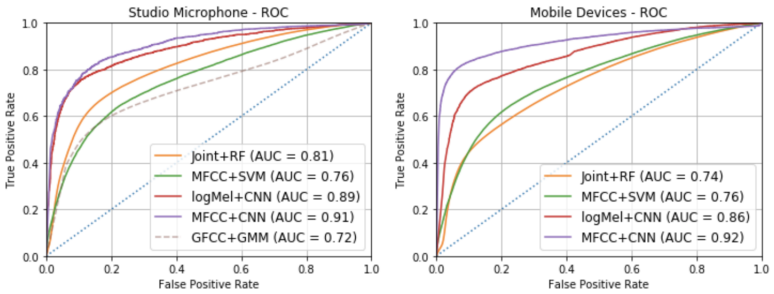
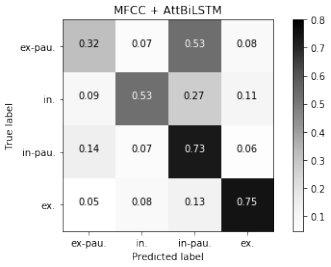


Fig. 13. ROC curve for RF, SVM, CNN, and GMM models for breathing sounds and no-breathing noise detection on data recorded by the studio microphone and the four mobile-devices.

Table 4. Detection performance of BPDM. The confusion matrix of the best performing MFCC with AttBiLSTM on the mobile-devices' data is shown on the right.

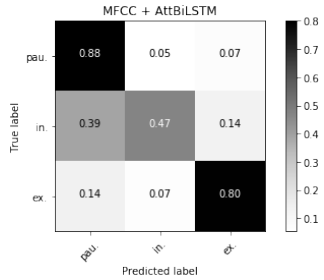
	Studio microphone			Mobile devices		
In./In-Pau./Ex./Ex-Pau.	PPV	TPR	ACC	PPV	TPR	ACC.
MFCC+HMM	0.623	0.631	0.662	0.391	0.387	0.423
logMel+BiLSTM	0.734	0.746	0.746	0.533	0.521	0.522
MFCC+AttBiLSTM	**0.749**	**0.756**	**0.762**	**0.639**	**0.618**	**0.617**



From the confusion matrix, we find that most of the errors occur in the pause phase after exhalation, and fewer errors in detecting exhalation. Thus, we look further at the performance by classifying only three classes (i.e., pause, inhale, exhale). Results are shown in Table 5 with the confusion matrix. MFCCs with Attention-based BiLSTM (AttBiLSTM) with 0.755 ACC achieves the best result. We also find that classifying three classes performs better than directly combining the predictions of pauses after inhale and exhale from the 4-classes detection scenario, which results in an ACC of 0.742. For inhalation, there is a performance drop on TPR, but a slight increase in PPV and reduce on FPR, from 0.681 to 0.702 and 0.073 to 0.059 respectively. The errors occur, in this case, in between inhalation and pause. This suggests that taking the maximum index of the output probabilities is not sufficient to derive biofeedback. Post-processing is, thus, required to optimize the output based on the distribution of the probabilities.

Table 5. Detection performance of BPDM for three breathing phases. The confusion matrix of the best performing MFCC with AttBiLSTM is shown on the right.

	Studio microphone			Mobile devices		
Pau./In./Ex.	PPV	TPR	ACC.	PPV	TPR	ACC.
MFCC+HMM	0.809	0.809	0.813	0.504	0.507	0.498
logMel+BiLSTM	**0.862**	0.838	0.841	0.624	0.610	0.611
MFCC+AttBiLSTM	**0.862**	**0.852**	**0.851**	**0.750**	**0.714**	**0.755**



4.3.3 Evaluation on Unseen Device. In this experiment, we evaluate the performance of BEM and BPDM on unseen devices. Specifically, we retrain the best-performed models on all devices' data but one device's data in the training set, and test the models on the excluded device's test set. In this way, we evaluate the models' robustness on the unseen device and unseen participants. The first row in Figure 14 shows the resulting ROC for BEM and PPV, TPR, and FPR for BPDM. On average, BEM seems to be robust across device with $AUC_{avg} = 0.907$, however, the results vary in BPDM with $ACC_{avg} = 0.617$. From Table 3 and Figure 10, we summarize that the use of tablet as the test device performs the worst due to the unbalanced sound quality data in the training set and the high variance within the data itself.

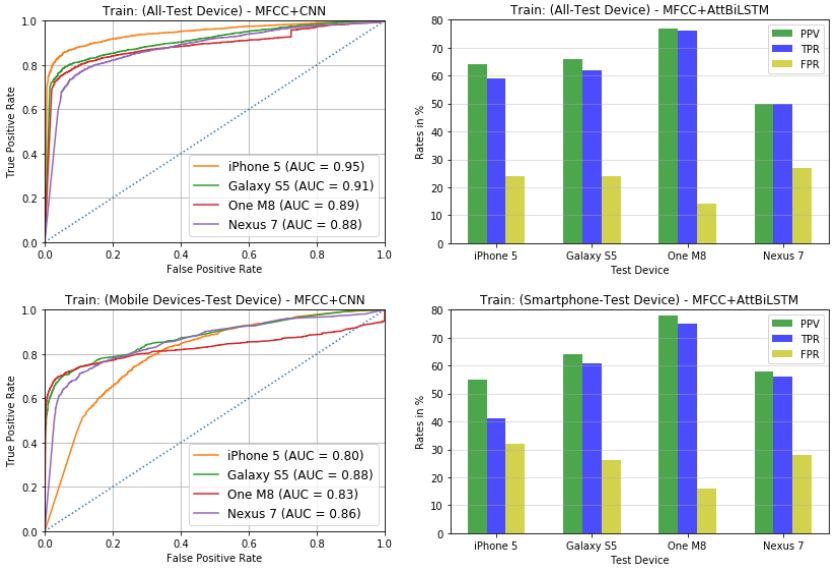


Fig. 14. Evaluation results for unseen device on both BEM and BPDM models.

To evaluate whether the results are influence by the inclusion of the studio microphone's data, we analyze the performance by training on the mobile-devices' data only. In other words, we use the same procedure from the previous evaluation, but exclude the studio microphone's recordings. In the second row of Figure 14, we observe a definite decrease in performance with iPhone as the test device. While low quality signals hurt testing performance, the best quality signals obtained from iPhone also negatively affect the performance of the models due to the discrepancy of data quality in training and test set.

4.3.4 Robustness to Noise. As described in section 4.1.6, we augmented the test set with different variances of white noises to evaluate the robustness of BEM to various level of noises. From the results in Figure 15, we observe a drop of AUC when the noise variance increases from 0.004 to 0.007 (i.e., SNR from 0dB to -5dB). Since the white noise adds random signals to all frequencies, it is expected that the FPR for breath detection increases. We also evaluate the performance by training the model with noisy data (i.e., added $\sigma = 0.002$ white noise). However, the model becomes less robust than if trained on the clean data for breath detection.

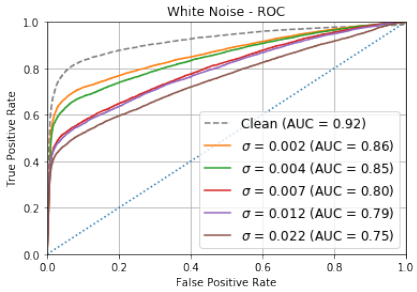


Fig. 15. Performance of BEM on different levels of white noise.

Based on the result from BEM, we further evaluate the performance of BPDM on added white noise variance of 0.002 and 0.004. It is challenging for BPDM to perform well when the SNR is down to 5dB, especially for exhalation. We further evaluate the performance by retraining the BPDM on augmented training data with added $\sigma = 0.002$ white noise. Results of the model tested on the noise variance of 0.004 show a high FPR on pause detection.

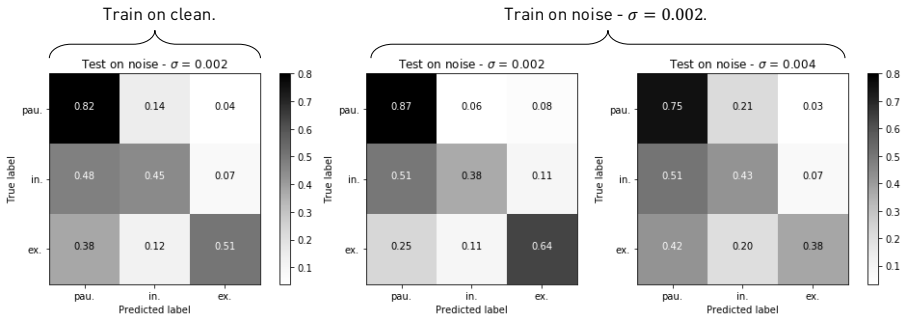


Fig. 16. Performance of BPDE on different levels of white noise.

4.3.5 Preliminary Evaluation of Breath Detection in Practice. To compare against the ground truth breathing cycles, we evaluate different algorithms by counting breathing cycles from the predictions of the detection models. The process consists of two phases: a pre-processing of the predictions and a counting rule-set. All the evaluated algorithms share the same pre-processing phase, which assigns each second a prediction based on majority vote of raw predictions. We use two heuristic rules on top of the voting. First, if the average probability for inhalations during a second exceeds an empirical threshold of 0.3, the prediction is set to inhalation. This threshold is chosen based on the results of the BPDM on the test data from the data collection study. Second, if the most often occurring prediction is noise, it is set to the prediction that occurs the second-most. If there is no second-most, the prediction of the preceding second is used.

After pre-processing, a counting ruleset is employed. We have evaluated four such rulesets: Counting the number of inhalations, exhalations, pauses, and counting pauses following exhalation. For the three approaches that count breath cycles based on single breathing phases, we increase the count for each one-second occurrence of this prediction that is not immediately following a one-second prediction of the same breathing phase. The final ruleset counts the number of pauses that follow an exhalation. It is not required that a pause immediately follows the exhalation to be counted, but per exhalation, only the next occurrence of a pause is considered. This ruleset is motivated by the evaluation of the BPDM, where the highest accuracies are achieved for exhalation and pause. With a MAE of 2.27, overall participants, this ruleset achieves the best results when compared to 4.00, 3.01, and 3.02 for the inhalation, exhalation, and pause rulesets, respectively. The complete results of this evaluation are presented in Table 6.

The evaluation of derived breathing cycle detection shows that breathing phases are identified. Consistent with this result, the subjects have been able to travel between 5489 and 9891 in-game meters with their sailboat. Of note, the minimally and maximally reachable distances are 2765 and 10500 in-game meters, respectively. This implies that the detection has triggered active biofeedback.

Table 6. Detection of breathing phases and frequency per minute during the 6-minute training with Breeze.

		Detected breathing phases per minute				Mean absolute error (MAPE)			
#	BFM	EX→Pause	IN	EX	Pause	EX→Pause	IN	EX	Pause
1	6.00	7.17	0.00	7.33	7.33	1.17 (19.4%)	6.00 (100.0%)	1.33 (22.2%)	1.33 (22.2%)
2	6.00	6.67	2.83	8.67	6.67	0.67 (11.1%)	3.17 (52.78%)	2.67 (44.4%)	0.67 (11.1%)
3	5.83	2.67	0.00	2.83	2.83	3.17 (54.3%)	5.83 (100%)	3.00 (51.4%)	3.00 (51.4%)
4	6.00	8.50	4.00	9.17	10.50	2.50 (41.7%)	2.00 (33.3%)	3.17 (52.8%)	4.50 (75.0%)
5	5.83	4.33	0.00	4.33	4.33	1.50 (25.7%)	5.83 (100.0%)	1.50 (25.7%)	1.50 (25.7%)
6	6.00	3.33	0.00	3.50	3.50	2.67 (44.4%)	6.00 (100.0%)	2.50 (41.7%)	2.50 (41.7%)
7	5.83	6.33	1.83	6.67	7.00	0.50 (8.6%)	4.00 (68.6%)	0.83 (14.3%)	1.17 (20.0%)
8	5.83	3.67	0.00	3.83	3.67	2.17 (37.1%)	5.83 (100.0%)	2.00 (34.3%)	2.17 (37.1%)
9	5.83	13.83	1.00	13.83	14.33	8.00 (137.1%)	4.83 (82.9%)	8.00 (137.1%)	8.50 (145.7%)
10	6.17	6.00	8.83	9.50	7.83	0.17 (2.7%)	2.67 (43.2%)	3.33 (54.1%)	1.67 (27.0%)
11	6.00	8.83	0.17	8.83	9.00	2.83 (47.2%)	5.83 (97.2%)	2.83 (47.2%)	3.00 (50.0%)
12	6.00	7.17	6.67	9.83	8.67	1.17 (19.4%)	0.67 (11.1%)	3.83 (63.9%)	2.67 (44.4%)
13	6.00	8.50	1.33	9.00	9.17	2.50 (41.7%)	4.67 (77.8%)	3.00 (50.0%)	3.17 (52.8%)
14	6.00	5.33	7.50	5.67	10.33	0.67 (11.1%)	1.50 (25.0%)	0.33 (5.6%)	4.33 (72.2%)
15	6.00	6.00	5.17	8.33	7.17	0.00 (0.0%)	0.83 (13.9%)	2.33 (38.9%)	1.17 (19.4%)
16	6.00	12.67	1.67	13.50	13.00	6.67 (111.1%)	4.33 (72.2%)	7.50 (125.0%)	7.00 (116.7%)
Mean	5.96	6.94	2.56	7.80	7.83	2.27 (38.3%)	4.00 (67.4%)	3.01 (50.5%)	3.02 (50.8%)
SD	0.10	3.08	2.98	3.25	3.29	2.23 (36.6)	1.94 (32.0)	2.08 (34.0)	2.17 (35.6)

Note: # = number of subjects, BFM = breathing frequency per minute derived from the ground truth respiration sensor, MAPE = Mean absolute percentage error, IN = Inhalation, EX = Exhalation, SD = Standard deviation.

4.4 Efficiency of the Smartphone-based Implementation

In this section, we present the execution cost, runtime, and energy profile of BEM and BPDM when running on the OnePlus 6 Snapdragon 845 processor. Table 7 summarizes the computation complexity to execute the two models to perform the detection. The models are designed to be lightweight, thus resulting in a requirement of only 2.65 and 10.07 million FLOPS per inference, and the execution times are on average 1.6ms and 6.9ms over a continuous six-minute detection.

Furthermore, we use Android Profiler to record the application performance during the six-minute breathing sessions with Breeze. An output focused on the breath detection thread is presented in Table 8, including the total CPU time, memory, energy characteristics, and averaged wall-time cost per inference. The prepossessing

step (i.e., filtering and feature calculation) requires the most runtime and memory in this thread compare to the rest of the processes. The category of *Others* includes post-processing and loggings of the final results. The *light* energy is an indicator from the profiler that measures whether the running component requires more energy than needed. In the real-time process, every second is buffered with a gap of 30ms. Except the first prediction, which has a latency of around 1.08 seconds (i.e., 1+ total wall-time), the rest of the predictions come with around 30ms. The total wall-time results in 119ms detection latency. These are estimations because the audio recording and buffering themselves have latency. While the detection thread is running, the Unity thread for the biofeedback visualization can run smoothly in parallel with 30 frames per second (fps).

Table 7. Requirements for model execution.

Module	# Parameters	Size	FLOPS
BEM	662.77k	1.7MB	2.65×10^6
BPDM	624.79k	0.8MB	10.07×10^6

Note: FLOPS = Floating point operations per second

Table 8. System performance of the breathing detection thread.

	CPU(s)	Memory	Energy	Wall-Time(ms)
Pre-processing	274.77	0.440 MB	*light*	78.02
Detection	15.78	1552 bytes	*light*	8.56
Others	62.95	7140 bytes	*light*	2.32
Total	353.50	0.448 MB	*light*	88.90

4.5 Technology Acceptance

The descriptive statistics for the technology acceptance ratings of the pretest and the pilot study are listed in Table 9. The results of the pretest indicate positive ratings for both Breeze and the active control condition positive ratings with average scores above the neutral scale value of four. However, the active control condition was clearly preferred with the exception of perceived enjoyment by the three subjects. Qualitative feedback revealed clearly that the visual design of when to inhale, to exhale and to hold the breath was not clear enough. Therefore, we added explicit arrows to the visualization as outlined in Figure 5 before conducting the pilot study.

Results of the pilot study indicate that both Breeze and the circle-based breathing training are assessed positively, with average item scores above the neutral scale value of four. In contrast to the pretest, Breeze is now preferred with respect to all variables with the exception of perceived ease of use and is rated higher in all dimensions. Two-sided paired t-tests are performed for all five variables to identify any significant differences between Breeze and the control condition. Only two significant differences are found: perceived ease of use scores are significantly higher for the circle-based active control condition ($p = .004$), and perceived enjoyment scores are significantly higher for Breeze ($p = .005$).

Table 9. Technology acceptance ratings of the pretest and pilot study.

	Pretest with 3 subjects			**Pilot study with 16 subjects**		
Construct	**BRE Mean (SD)**	**CIR Mean (SD)**	**BRE Pref. %**	**BRE Mean (SD)**	**CIR Mean (SD)**	**BRE Pref. %**
Perceived ease of use	5.00 (1.00)	6.33 (1.15)	0.0%	5.81 (1.05)	6.62 (0.50)	31.3%
Perceived usefulness	5.33 (1.53)	6.00 (1.00)	0.0%	5.88 (0.96)	6.00 (0.82)	56.3%
Perceived enjoyment	5.00 (0.00)	5.00 (1.00)	66.6%	5.75 (1.06)	4.69 (1.49)	87.5%
Perceived control	5.33 (1.15)	5.33 (2.08)	33.3%	5.62 (1.26)	5.88 (0.89)	68.8%
Intention to use	4.33 (0.58)	4.67 (1.53)	33.3%	4.88 (1.50)	4.62 (1.93)	75.0%

Note: BRE = Breeze; BRE Pref. % = percentage of individuals who preferred Breeze over CIR. CIR = circle-based breathing training, validated active control condition, see [62]; SD = standard deviation; 7-point Likert scales ranging from strongly disagree (0) to strongly agree (7) were used.

4.6 Impact on Breathing Frequency and High Frequency Heart Rate Variability (HF-HRV)

The descriptive statistics of the HF-HRV scores and the breathing frequencies are calculated from the electrocardiogram and respiration sensors. They are listed in Table 10 for the six-minute baseline measurement and the

six-minute slow-paced breathing training sessions with both Breeze and the active control condition. Our results support the findings of prior work [62] regarding the baseline measurement and the active control condition. Breathing frequency per minute was reduced during both training sessions from circa 12 breathing cycles per minute to six cycles per minute, which is the intended effective target breathing frequency [62]. Consistently, a linear effects model with post-hoc Tukey contrasts resulted in significantly lower breathing rates during the Breeze training (b = -6.27, p < .001) and the active control condition (b = -3.24, p < .001). In addition, the use of Breeze leads to an even stronger increase in HF-HRV from the baseline measurement compared to the active control condition. A linear effects model with post-hoc Tukey contrasts confirmed this result when HF-HRV scores of baseline measurement were compared to Breeze (b = 0.85, p < .001) and the active control condition (b = 0.56, p < .002).

Table 10. Mean values and standard deviations (SD) of the physiological data for the 16 subjects of the pilot study.

Physiological measure	**Baseline Mean (SD)**	**BRE Mean (SD)**	**CIR Mean (SD)**
High frequency heart rate variability (HF-HRV)	6.60 (0.94)	7.45 (1.02)	7.16 (1.13)
Breathing frequency per minute	12.23 (4.10)	5.96 (0.10)	5.99 (0.04)

Note: The unit of HF-HRV is the log of ms^2 [62]; BRE = Breeze; CIR = Circle-based breathing training, validated active control condition, see [62].

5 DISCUSSION, LIMITATIONS AND FUTURE WORK

In the following subsections, we discuss the results of the current work, outline its limitations and suggest future work for each of the four research questions.

5.1 Research Question 1: Detection Performance of Breeze

We have demonstrated the approaches and evaluated the feasibility of detecting breathing phases utilizing acoustics breathing sounds recorded through various devices and from various individuals. Results from the models that are trained on the mobile-devices' data collected from the data collection study present acceptable results on unseen individuals. There are, however, limitations for BPDM to detect accurate breathing phases on unknown devices that have lower recording qualities and with devices' noises. The performance of the detection in the pilot study indicates the generalizability of pre-trained models' across new individuals, devices, and environmental settings. Result shows that it is challenging for inhalation detection.

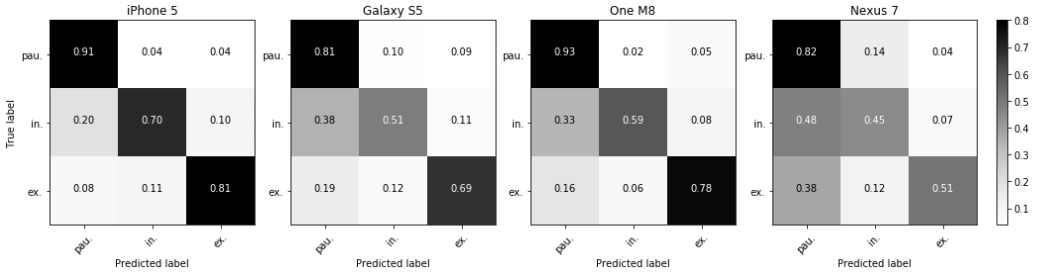


Fig. 17. Confusion matrices of the BPDM for each of the four mobile-devices.

Collected Data. Sound travels in all directions, so detection performance is influenced not only by the recording quality of mobile devices or their distance to the source, but also the environmental setting. Surrounding objects, for example, matter as they either reflect and/or attenuate specific frequencies of the signal. In the data collection study, we had a universal set up for all five devices. That is, differences in sound quality across devices (Figure

10) is due to the devices themselves. Thus, there are differences in detection performance if the devices were evaluated independently, especially for the detection of inhalation sound (Figure 17). In this scenario, detection of inhalation can be very challenging. We fully realized this challenge during the pilot study when minor changes in the layout of the room influenced the detection of inhalation sound measurably.

Mobile Recordings. It is fundamentally challenging to capture information from sound that is recorded with very low SNR (i.e., inhalation, in our case). However, devices such as studio microphones and iPhones which are able to record better quality of sound, can reach higher accuracy on the inhalation detection. To further improve the performance, we conjecture that converting low quality signals to higher quality and clearer signals by using a generative model could remedy the inconsistency of recording qualities. For example, a Cycle-Consistent Adversarial Networks [84] learns a mapping function $f : X \rightarrow Y$, which maps the good-quality signals $f(X)$ to the low ones Y and vice versa, such that the distribution of the two sets of data are indistinguishable. Thus, the new low-quality signals can be *translated* into signals with better quality to retrieve higher accuracy.

Environmental Noise. The evaluation of the robustness of the models is demonstrated using white noises with different variances. However, such a specific experiment is still limited to generally presenting the models' capabilities in real-world settings. Adding white noise introduces noise to all frequencies. Besides the expected drop on models' performance, further insights into the possible confusion noises (e.g., public transport, air condition, or fans in office noises), which potentially share similar acoustic features with breathing sounds, remain unknown. Restrictions came from the two lab studies conducted under artificial settings, and the use of the open audio data sources with unknown recording configurations. For Breeze to provide adequate biofeedback in the wild, it is crucial, as a next step, to conduct experiments outdoor to gain knowledge on the feasibility of detecting breathing phases against specific environmental noises and the natural detection boundaries. Such that we can enhance the usage of Breeze, in none-quiet environments, by training noise agnostic models or by applying corresponding breathing sound normalization to increase the detectability of breathing phases.

In summary and to answer the first research question (*How accurate are breathing phases detected in quasi-real-time with a smartphone's microphone?*), we conclude that the accuracy of 0.755 is achieved with the data from the data collection study on unseen participants. Furthermore, the performance of the detection *in practice* shows the ability of breathing phases detection, and, therefore, the capability to trigger biofeedback. Breathing cycles derived from the detected breathing phases are used as a comparison against the ground truth with a mean absolute error of 2.27.

5.2 Research Question 2: Technical Efficiency of Breeze

We have implemented and demonstrated the feasibility of a real-time breath detection pipeline running on a OnePlus 6 smartphone. The two models were designed to be lightweight, thus enabling the detection with low latency while running the visualizations of the biofeedback in parallel. Since this work focuses primarily on the prototype development and feasibility evaluation, further investigation into enhancing the detection performance and the scalability for a real-world product is still required.

We have encountered a challenge in the trade-off in between efficiency and detection performance. The detection performance of BPDM increases slightly (i.e., ACC from 0.755 to 0.760) when MFCC is calculated from the power of the spectrum instead of the amplitude. However, the behavior of BEM is the opposite with more significant differences. Since the feature calculation uses up the majority of the CPU time in the detection thread, the decision was made to only use one set of the feature. Thus, it is essential for the future work to focus on improving of the detection accuracy without introducing more complexity. It is foreseeable that an algorithm such

as Fast Dynamic Time Wrapping [64] could be explored to model the BPDM's output to improve the performance without adding significant latency.

In summary, and to answer the second research question (*How efficiently does this real-time detection algorithm run on a smartphone?*), we conclude that the developed breathing detection model is able to perform real-time detection smoothly with low latency on a smartphone, and at the same time, process an interactive biofeedback visualization.

5.3 Research Question 3: Technology Acceptance of Breeze

The results of the pilot study are promising with respect to the technology acceptance of Breeze. That is, all self-reported scores lie above the neutral scale value of four. Moreover, when compared to a validated active control condition from prior work [62], Breeze was preferred in four out of five dimensions by the majority of subjects, i.e., in terms of perceived usefulness, perceived enjoyment, perceived control, and the behavioral intention to use the application. In fact, the rated perceived enjoyment of Breeze is significantly higher than that of the active control training. This result shows that the gamified design of Breeze was positively acknowledged by the subjects. Also, perceived control was rated above-average on the answer scale. This indicates that the visualization of the biofeedback of Breeze is reasonably understandable.

With respect to the perceived ease of use, Breeze was rated significantly lower by the subjects. This can be explained by the fact that the major effort devoted to Breeze is related to the real-time detection of the breathing phases. In our future work, we plan to analyze feedback from individuals qualitatively to derive implications for revisions of the biofeedback visualization. An integrated tutorial in combination with video directions on how to use Breeze are further additions for future work to improve Breeze's perceived ease of use. Finally, there is still a lot to research about how Breeze is adopted and used in the everyday life of individuals with chronic conditions and mental illness. Thus, further studies are required to assess the reach, acceptance, and daily or weekly adherence of Breeze in specific populations that are also less likely to perform sessions of slow-paced breathing [11].

In summary and to answer the third research question (*How is a smartphone-based gamified biofeedback breathing training perceived?*), we conclude that our prototype Breeze was perceived positively by higher-educated test subjects from two universities, who had limited experience in breathing training and biofeedback.

5.4 Research Question 4: Physiological Impact of Breeze

The impact of the Breeze training on the breathing frequency and HF-HRV is comparable to an already validated active control condition in which subjects had to follow an animated circle [62]. In particular, HF-HRV as measure of cardiac functioning was significantly increased with an even higher score during the slow-paced breathing training with Breeze compared to the active control condition. Moreover, the intended and effective target frequency of six breathing cycles per minute was also achieved when subjects followed the breathing pattern triggered by the gamified biofeedback visualization of Breeze.

However, the pilot study was limited to 16 subjects from two universities and, thus, external validity has to be added to the promising physiological impact Breeze was able to induce. Therefore, we plan a fully-powered efficacy study in a laboratory setting consistent with prior work [62] and with a more diverse population. We further plan to assess physiological parameters that are relevant to the efficacy such as the heart rate, (very) low frequency HRV, and the low frequency to high frequency ratio [43]. Finally, several longitudinal clinical trials are planned to assess the long-term impact of slow-paced breathing on various patient populations.

In summary and to answer the final research question (*Does a smartphone-based gamified biofeedback breathing training positively impact physiology?*), we conclude that Breeze is able to help subjects follow a slow-paced breathing technique and, as a result, to help them strengthen their cardiac functioning.

6 CONCLUSION

In the current work, we propose Breeze, a scalable smartphone-based gamified biofeedback breathing training. Breeze has the overall goal of extending the reach of therapeutic and preventive health interventions for chronic disease and mental illness by strengthening cardiac functioning. Its technical implementation, acceptance and physiological impact are promising topics for future longitudinal studies and implementations in health care systems. This work contributes to the interdisciplinary field of digital health at the intersection of computer science, biological psychology, and behavioral medicine, and responds to the pressing need for scalable digital health interventions that improve the self-management capabilities of individuals with and without chronic disease or mental illness in their everyday lives.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers of the paper, Filipe Barata and Marcia Nissen from ETH Zurich for supporting us in the data collection, and Helen Galliker from CSS Insurance for supporting the design of the biofeedback visualization. We would also like to thank Maksim Bolonkin and Behnaz Abdollahi from Dartmouth College for providing feedback on early draft versions. The work is part-funded by the Swiss National Science Foundation (http://p3.snf.ch/project-159289) and CSS Insurance.

REFERENCES

[1] Shahin Amiriparian, Maurice Gerczuk, Sandra Ottl, Nicholas Cummins, Michael Freitag, Sergey Pugachevskiy, Alice Baird, and Björn W Schuller. 2017. Snore Sound Classification Using Image-Based Deep Spectrum Features. In *Proc. Interspeech 2017*. 3512–3516. https://doi.org/10.21437/Interspeech.2017-434

[2] Muhammad Awais Azam, Aeman Shahzadi, Asra Khalid, Syed M Anwar, and Usman Naeem. 2018. Smartphone Based Human Breath Analysis from Respiratory Sounds. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 445–448. https://doi.org/10.1109/EMBC.2018.8512452

[3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. (2014). arXiv:arXiv preprint arXiv:1409.0473

[4] Mohammed Bahoura. 2009. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in biology and medicine* 39, 9 (2009), 824–843. https://doi.org/10.1016/j.compbiomed.2009.06.011

[5] Filipe Barata, Kevin Kipfer, Maurice Weber, Peter Tinschert, Elgar Fleisch, and Tobias Kowatsch. 2019. Towards Device-Agnostic Mobile Cough Detection with Convolutional Neural Networks. In *7th IEEE International Conference on Healthcare Informatics (ICHI 2019)*. https://doi.org/10.3929/ethz-b-000348112

[6] Joachim Behar, Aoife Roebuck, Mohammed Shahid, Jonathan Daly, Andre Hallack, Niclas Palmius, John Stradling, and Gari D Clifford. 2015. SleepAp: an automated obstructive sleep apnoea screening application for smartphones. *IEEE Journal of Biomedical and Health Informatics* 19, 1 (2015), 325–331. https://doi.org/10.1109/JBHI.2014.2307913

[7] Thomas Benichou, Bruno Pereira, Martial Mermillod, Igor Tauveron, Daniela Pfabigan, Salwan Maqdasy, and Frédéric Dutheil. 2018. Heart rate variability in type 2 diabetes mellitus: A systematic review and metaâĂŞanalysis. *PLOS ONE* 13, 4 (04 2018), 1–19. https://doi.org/10.1371/journal.pone.0195166

[8] Timothy W. Bickmore, Suzanne E. Mitchell, Brian W. Jack, Michael K. Paasche-Orlow, Laura M. Pfeifer, and Julie OâĂŹDonnell. 2010. Response to a relational agent by hospital patients with depressive symptoms. *Interacting with Computers* 22, 4 (2010), 289–298. https://doi.org/10.1016/j.intcom.2009.12.001

[9] Plamen Bokov, Bruno Mahut, Patrice Flaud, and Christophe Delclaux. 2016. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Computers in Biology and Medicine* 70 (2016), 40–50. https://doi.org/10.1016/j.compbiomed.2016.01.002

[10] Paul Brennan, Markus Perola, Gert-Jan van Ommen, and Elio Riboli. 2017. Chronic disease research in Europe and the need for integrated population cohorts. *European Journal of Epidemiology* 32, 9 (2017), 741–749. https://doi.org/10.1007/s10654-017-0315-2

[11] Adam Burke, Chun Nok Lam, Barbara Stussman, and Hui Yang. 2017. Prevalence and patterns of use of mantra, mindfulness and spiritual meditation among adults in the United States. *BMC Complementary and Alternative Medicine* 17, 1 (2017), 316. https://doi.org/10.1186/s12906-017-1827-8

[12] Stéphanie Carlier, Sara Van der Paelt, Femke Ongenae, Femke De Backere, and Filip De Turck. 2019. Using a Serious Game to Reduce Stress and Anxiety in Children with Autism Spectrum Disorder. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'19)*. ACM, New York, NY, USA, 452–461. https://doi.org/10.1145/3329189.3329237

[13] John A. Chalmers, Daniel S. Quintana, Maree J. Anne Abbott, and Andrew H. Kemp. 2014. Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis. *Frontiers in Psychiatry* 5 (2014), 80. https://doi.org/10.3389/fpsyt.2014.00080

[14] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2017. BreathPrint: Breathing Acoustics-based User Authentication. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '17)*. ACM, New York, NY, USA, 278–291. https://doi.org/10.1145/3081333.3081355

[15] S. Chen, P. Sun, S. Wang, G. Lin, and T. Wang. 2016. Effects of heart rate variability biofeedback on cardiovascular responses and autonomic sympathovagal modulation following stressor tasks in prehypertensives. *Journal of Human Hypertension* 30, 2 (2016), 105–11. https://doi.org/10.1038/jhh.2015.27

[16] Luca Chittaro and Riccardo Sioni. 2014. Evaluating mobile apps for breathing training: The effectiveness of visualization. *Computers in Human Behavior* 40 (2014), 56–63. https://doi.org/10.1016/j.chb.2014.07.049

[17] Eliran Dafna, Ariel Tarasiuk, and Yaniv Zigel. 2014. Automatic Detection of Whole Night Snoring Events Using Non-Contact Microphone. *PLOS ONE* 8, 12 (12 2014), 1–14. https://doi.org/10.1371/journal.pone.0084139

[18] Alison Dillon, Mark Kelly, Ian H. Robertson, and Deirdre A. Robertson. 2016. Smartphone Applications Utilizing Biofeedback Can Aid Stress Reduction. *Frontiers in Psychology* 7, 832 (2016). https://doi.org/10.3389/fpsyg.2016.00832

[19] David Eddie, Evgeny Vaschillo, Bronya Vaschillo, and Paul Lehrer. 2015. Heart rate variability biofeedback: Theoretical basis, delivery, and its potential for the treatment of substance use disorders. *Addiction research theory* 23, 4 (2015), 266–272. https://doi.org/10.3109/16066359.2015.1011625

[20] William J Elliott, Joseph L Izzo, William B White, Douglas R Rosing, Christopher S Snyder, Ariela Alter, Benjamin Gavish, and Henry R Black. 2004. Graded blood pressure reduction in hypertensive outpatients associated with use of a device to assist with slow breathing. *The Journal of Clinical Hypertension* 6, 10 (2004), 553–559. https://doi.org/10.1111/j.1524-6175.2004.03553.x

[21] Alize J. Ferrari, Fiona J. Charlson, Rosana E. Norman, Scott B. Patten, Greg Freedman, Christopher J.L. Murray, Theo Vos, and Harvey A. Whiteford. 2013. Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. *PLOS Medicine* 10, 11 (11 2013), 1–12. https://doi.org/10.1371/journal.pmed.1001547

[22] Tim Fischer, Johannes Schneider, and Wilhelm Stork. 2016. Classification of breath and snore sounds using audio data recorded with smartphones in the home environment. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 226–230. https://doi.org/10.1109/ICASSP.2016.7471670

[23] G. David Forney. 1973. The viterbi algorithm. *Proc. IEEE* 61, 3 (1973), 268–278. https://doi.org/10.1109/PROC.1973.9030

[24] Jérémy Frey, May Grabli, Ronit Slyper, and Jessica R. Cauchard. 2018. Breeze: Sharing Biofeedback Through Wearable Technologies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 645, 12 pages. https://doi.org/10.1145/3173574.3174219

[25] Mark JF. Gales. 1998. Maximum likelihood linear transformations for HMM-based speech recognition. *Computer speech & language* 12, 2 (1998), 75–98. https://doi.org/10.1006/csla.1998.0043

[26] Richard Gevirtz. 2013. The Promise of Heart Rate Variability Biofeedback: Evidence-Based Applications. *Biofeedback* 41, 3 (2013), 110–120. https://doi.org/10.5298/1081-5937-41.3.01

[27] Vera C Goessl, Joshua E Curtiss, and Stefan G Hofmann. 2017. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. *Psychol Med* 47, 15 (2017), 2578–2586. https://doi.org/10.1017/s0033291717001003

[28] E. Grossman, A. Grossman, MH. Schein, R. Zimlichman, and B. Gavish. 2001. Breathing-control lowers blood pressure. *Journal of Human Hypertension* 15, 4 (2001), 263–9. https://doi.org/10.1038/sj.jhh.1001147

[29] Brian A. Hanson and Ted H. Applebaum. 1990. Robust speaker-independent word recognition using static, dynamic and acceleration features: Experiments with Lombard and noisy speech. In *International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 857–860 vol.2. https://doi.org/10.1109/ICASSP.1990.115973

[30] Tian Hao, Chongguang Bi, Guoliang Xing, Roxane Chan, and Linlin Tu. 2017. MindfulWatch: A Smartwatch-Based System For Real-Time Respiration Monitoring During Meditation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 57 (Sept. 2017), 19 pages. https://doi.org/10.1145/3130922

[31] Tian Hao, Guoliang Xing, and Gang Zhou. 2015. RunBuddy: A Smartphone System for Running Rhythm Monitoring. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 133–144. https://doi.org/10.1145/2750858.2804293

[32] Richard I. G. Holt and Wayne J. Katon. 2012. Dialogue on Diabetes and Depression: Dealing with the double burden of co-morbidity. *Journal of Affective Disorders* 142 (2012), S1–S3. https://doi.org/10.1016/S0165-0327(12)00632-5

[33] Claudio Imperatori, Miranda Mancini, Giacomo Della Marca, Enrico Maria Valenti, and Benedetto Farina. 2018. Feedback-Based Treatments for Eating Disorders and Related Symptoms: A Systematic Review of the Literature. *Nutrients* 10, 11 (2018). https://doi.org/10.3390/nu10111806

[34] Math Janssen, Yvonne Heerkens, Wietske Kuijer, Beatrice van der Heijden, and Josephine Engels. 2018. Effects of Mindfulness-Based Stress Reduction on employees' mental health: A systematic review. *PLOS ONE* 13, 1 (2018), 1–37. https://doi.org/10.1371/journal.pone.0191332

[35] Bin Jiang, Chun-Yen Chang, and William C Sullivan. 2014. A dose of nature: Tree cover, stress reduction, and gender differences. *Landscape and Urban Planning* 132 (2014), 26–36. https://doi.org/10.1016/j.landurbplan.2014.08.005

[36] Arnold Kamis, Marios Koufaris, and Tziporah Stern. 2008. Using an Attribute-Based Decision Support System for User-Customized Products Online: An Experimental Investigation. *MIS Quarterly* 32, 1 (2008), 159–177. https://doi.org/10.2307/25148832

[37] David L. Katz, Elizabeth P. Frates, Jonathan P. Bonnet, Sanjay K. Gupta, Erkki Vartiainen, and Richard H. Carmona. 2018. Lifestyle as Medicine: The Case for a True Health Initiative. *American Journal of Health Promotion* 32, 6 (2018), 1452–1458. https://doi.org/10.1177/0890117117705949

[38] Andrew H. Kemp, Daniel S. Quintana, Marcus A. Gray, Kim L. Felmingham, Kerri Brown, and Justine M. Gatt. 2010. Impact of Depression and Antidepressant Treatment on Heart Rate Variability: A Review and Meta-Analysis. *Biological Psychiatry* 67, 11 (2010), 1067–1074. https://doi.org/10.1016/j.biopsych.2009.12.012

[39] Bassam Khoury, Manoj Sharma, Sarah E Rush, and Claude Fournier. 2015. Mindfulness-based stress reduction for healthy individuals: A meta-analysis. *Journal of psychosomatic research* 78, 6 (2015), 519–28. https://doi.org/10.1016/j.jpsychores.2015.03.009

[40] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investig* 15, 3 (2018), 235–245. https://doi.org/10.30773/pi.2017.08.17

[41] Sang-Dol Kim and Hee-Seung Kim. 2005. Effects of a relaxation breathing exercise on anxiety, depression, and leukocyte in hemopoietic stem cell transplantation patients. *Cancer nursing* 28, 1 (2005), 79–83.

[42] Joseph C Kvedar, Alexander L Fogel, Eric Elenko, and Daphne Zohar. 2016. Digital medicine's march on chronic disease. *Nature Biotechnology* 34, 3 (2016), 239–46. https://doi.org/10.1038/nbt.3495

[43] Sylvain Laborde, Emma Mosley, and Julian F. Thayer. 2017. Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research - Recommendations for Experiment Planning, Data Analysis, and Data Reporting. *Frontiers in Psychology* 8, 213 (2017). https://doi.org/10.3389/fpsyg.2017.00213

[44] Eric C. Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N. Patel. 2012. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 280–289. https://doi.org/10.1145/2370216.2370261

[45] Paul M Lehrer and Richard Gevirtz. 2014. Heart rate variability biofeedback: How and why does it work? *Frontiers in Psychology* 5, 756 (2014). https://doi.org/10.3389/fpsyg.2014.00756

[46] Paul M. Lehrer, Evgeny Vaschillo, and Bronya Vaschillo. 2000. Resonant Frequency Biofeedback Training to Increase Cardiac Variability: Rationale and Manual for Training. *Applied Psychophysiology and Biofeedback* 25, 3 (01 Sep 2000), 177–191. https://doi.org/10.1023/A:1009554825745

[47] Rong-Hao Liang, Bin Yu, Mengru Xue, Jun Hu, and Loe M. G. Feijs. 2018. BioFidget: Biofeedback for Respiration Training Using an Augmented Fidget Spinner. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 613, 12 pages. https://doi.org/10.1145/3173574.3174187

[48] Duanping Liao, Jianwen Cai, Ralph W. Barnes, Herman A. Tyroler, Pentti Rautaharju, Ingar Holme, and Gerardo Heiss. 1996. Association of cardiac automatic function and the development of hypertension: The ARIC study. *American Journal of Hypertension* 9, 12 (1996), 1147–1156. https://doi.org/10.1016/S0895-7061(96)00249-X

[49] I. Mei Lin. 2018. Effects of a cardiorespiratory synchronization training mobile application on heart rate variability and electroencephalography in healthy adults. *International Journal of Psychophysiology* 134 (2018), 168–177. https://doi.org/10.1016/j.ijpsycho.2018.09.005

[50] De Liu, Radhika Santhanam, and Jane Webster. 2017. Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Q.* 41, 4 (Dec. 2017), 1011–1034. https://doi.org/10.25300/MISQ/2017/41.4.01

[51] Ross W. May, Gregory S. Seibert, Marcos A. Sanchez-Gonzalez, and Frank D. Fincham. 2019. Self-regulatory biofeedback training: an intervention to reduce school burnout and improve cardiac functioning in college students. *Stress* 22, 1 (2019), 1–8. https://doi.org/10.1080/10253890.2018.1501021

[52] Ravi Mehta and Rui Juliet Zhu. 2009. Blue or red? Exploring the effect of color on cognitive task performances. *Science* 323, 5918 (2009), 1226–1229. https://doi.org/10.1126/science.1169144

[53] Saba Moussavi, Somnath Chatterji, Emese Verdes, Ajay Tandon, Vikram Patel, and Bedirhan Ustun. 2007. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *The Lancet* 370, 9590 (2007), 851–858. https://doi.org/10.1016/S0140-6736(07)61415-9

[54] Yunyoung Nam, Bersain A Reyes, and Ki H Chon. 2015. Estimation of respiratory rates using the built-in microphone of a smartphone or headset. *IEEE journal of biomedical and health informatics* 20, 6 (2015), 1493–1501. https://doi.org/10.1109/JBHI.2015.2480838

[55] Vikram Patel, Shekhar Saxena, Crick Lund, Graham Thornicroft, Florence Baingana, Paul Bolton, Dan Chisholm, Pamela Y. Collins, Janice L. Cooper, Julian Eaton, Helen Herrman, Mohammad M. Herzallah, Yueqin Huang, Mark J. D. Jordans, Arthur Kleinman, Maria Elena Medina-Mora, Ellen Morgan, Unaiza Niaz, Olayinka Omigbodun, Martin Prince, Atif Rahman, Benedetto Saraceno, Bidyut K. Sarkar, Mary De Silva, Ilina Singh, Dan J. Stein, Charlene Sunkel, and JÜrgen UnÜtzer. 2018. The Lancet Commission on global mental health and sustainable development. *The Lancet* 392, 10157 (2018), 1553 – 1598. https://doi.org/10.1016/S0140-6736(18)31612-X

[56] Stephen W. Porges. 2007. The polyvagal perspective. *Biological psychology* 74, 2 (2007), 116–143. https://doi.org/10.1016/j.biopsycho.2006.06.009

[57] 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)*. 1194–1202. https://doi.org/10.1109/INFOCOM.2015.7218494

[58] Hector E Romero, Ning Ma, Guy J Brown, Amy V Beeston, and Madina Hasan. 2019. Deep Learning Features for Robust Detection of Acoustic Events in Sleep-disordered Breathing. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 810–814. https://doi.org/10.1109/ICASSP.2019.8683099

[59] Joan Sol Roo, Renaud Gervais, Jeremy Frey, and Martin Hachet. 2017. Inner Garden: Connecting Inner States to a Mixed Reality Sandbox for Mindfulness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1459–1470. https://doi.org/10.1145/3025453.3025743

[60] Adriano L. Roque, Vitor E. Valenti, Thais Massetti, Talita Dias da Silva, Carlos Bandeira de Mello Monteiro, Fernando R. Oliveira, ÃĄlvaro Dantas de Almeida Junior, Sheylla Nadjane Batista Lacerda, Gustavo Carreiro Pinasco, Viviane Gabriela Nascimento, Luiz Gonzaga Granja Filho, Luiz Carlos de Abreu, David M. Garner, and Celso Ferreira. 2014. Chronic obstructive pulmonary disease and heart rate variability: a literature update. *International archives of medicine* 7, 1 (2014), 43. https://doi.org/10.1186/1755-7682-7-43

[61] T Rosenwein, Eliran Dafna, Ariel Tarasiuk, and Yaniv Zigel. 2014. Detection of breathing sounds during sleep using non-contact audio recordings. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 1489–1492. https://doi.org/10.1109/EMBC.2014.6943883

[62] Matthew E. B. Russell, April B. Scott, Ian A. Boggero, and Charles R. Carlson. 2017. Inclusion of a rest period in diaphragmatic breathing increases high frequency heart rate variability: Implications for behavioral therapy. *Psychophysiology* 54, 3 (2017), 358–365. https://doi.org/10.1111/psyp.12791

[63] Tara N. Sainath and Carolina Parada. 2015. Convolutional neural networks for small-footprint keyword spotting. In *Proc. Interspeech 2015*. 1478–1482. http://www.isca-speech.org/archive/interspeech_2015/i15_1478.html

[64] Stan Salvador and Philip Chan. 2007. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis* 11, 5 (2007), 561–580. https://doi.org/10.3233/IDA-2007-11508

[65] Maya C Schumer, Emily K Lindsay, and J David Creswell. 2018. Brief mindfulness training for negative affectivity: A systematic review and meta-analysis. *J Consult Clin Psychol* 86, 7 (2018), 569–583. https://doi.org/10.1037/ccp0000324

[66] Ralf Schwarzer. 1992. *Self-Efficacy Thought Control Of Action*. Taylor Francis, New York. https://doi.org/10.4324/9781315800820

[67] Ralf Schwarzer. 2008. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied psychology* 57, 1 (2008), 1–29. https://doi.org/10.1111/j.1464-0597.2007.00325.x

[68] Elizabeth K. Seng and Kenneth A. Holroyd. 2010. Dynamics of changes in self-efficacy and locus of control expectancies in the behavioral and drug treatment of severe migraine. *Ann Behav Med* 40, 3 (2010), 235–47. https://doi.org/10.1007/s12160-010-9223-3

[69] Nandini Sengupta, Md Sahidullah, and Goutam Saha. 2016. Lung sound classification using cepstral-based statistical features. *Computers in biology and medicine* 75 (2016), 118–129. https://doi.org/10.1016/j.compbiomed.2016.05.013

[70] Martin Siepmann, Volkan Aykac, Jana Unterdörfer, Katja Petrowski, and Michael Mueck-Weymann. 2008. A pilot study on the effects of heart rate variability biofeedback in patients with depression and in healthy subjects. *Appl Psychophysiol Biofeedback* 33, 4 (2008), 195–201. https://doi.org/10.1007/s10484-008-9064-z

[71] Laura Silver. 2019. *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally*. Report. Washington, DC.

[72] Tobias Sonne and Mads Møller Jensen. 2016. ChillFish: A Respiration Game for Children with ADHD. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '16)*. ACM, New York, NY, USA, 271–278. https://doi.org/10.1145/2839462.2839480

[73] Nancy J. Stone. 2003. Environmental view and color for a simulated telemarketing task. *Journal of Environmental Psychology* 23, 1 (2003), 63–78. https://doi.org/10.1016/S0272-4944(02)00107-X

[74] Martin Teufel, Kerstin Stephan, Axel Kowalski, Saskia Käsberger, Paul Enck, Stephan Zipfel, and Katrin E Giel. 2013. Impact of biofeedback on self-efficacy and stress reduction in obesity: a randomized controlled pilot study. *Appl Psychophysiol Biofeedback* 38, 3 (2013), 177–84. https://doi.org/10.1007/s10484-013-9223-8

[75] Julian F. Thayer, Shelby S. Yamamoto, and Jos F. Brosschot. 2010. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *International Journal of Cardiology* 141, 2 (2010), 122–131. https://doi.org/10.1016/j.ijcard.2009.09.543

[76] Graham Thornicroft, Somnath Chatterji, Sara Evans-Lacko, Michael Gruber, Nancy Sampson, Sergio Aguilar-Gaxiola, Ali Al-Hamzawi, Jordi Alonso, Laura Andrade, Guilherme Borges, Ronny Bruffaerts, Brendan Bunting, Jose Miguel Caldas de Almeida, Silvia Florescu, Giovanni de Girolamo, Oye Gureje, Josep Maria Haro, Yanling He, Hristo Hinkov, Elie Karam, Norito Kawakami, Sing Lee, Fernando Navarro-Mateu, Marina Piazza, Jose Posada-Villa, Yolanda Torres de Galvis, and Ronald C. Kessler. 2017. Undertreatment of people with major depressive disorder in 21 countries. *The British journal of psychiatry : the journal of mental science* 210, 2 (2017), 119–124. https://doi.org/10.1192/bjp.bp.116.188078

[77] Hiroki Uratani, Kohzoh Yoshino, and Mieko Ohsuga. 2014. Basic Study on the Most Relaxing Respiration Period in Children to Aid the Development of a Respiration-Leading Stuffed Toy. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and*

Biology Society. 3414–3417. https://doi.org/10.1109/EMBC.2014.6944356

[78] H. van der Heijden. 2004. User Acceptance of Hedonic Information Systems. *MIS Quarterly* 28, 4 (2004), 695–704. https://doi.org/10.2307/25148660

[79] Viswanath Venkatesh, James Y. L. Thong, and Xin Xu. 2012. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly* 36, 1 (2012), 157–178. https://ssrn.com/abstract=2002388

[80] Theo Vos, Amanuel Alemu Abajobir, Kalkidan Hassen Abate, Cristiana Abbafati, Kaja M Abbas, Foad Abd-Allah, Rizwan Suliankatchi Abdulkader, Abdishakur M Abdulle, Teshome Abuka Abebo, Semaw Ferede Abera, et al. 2017. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* 390, 10100 (2017), 1211–1259. https://doi.org/10.1016/S0140-6736(17)32154-2

[81] Joanneke Weerdmeester, Marieke van Rooij, Owen Harris, Niki Smit, Rutger C.M.E Engels, and Isabela Granic. 2017. Exploring the Role of Self-efficacy in Biofeedback Video Games. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17 Extended Abstracts)*. ACM, New York, NY, USA, 453–461. https://doi.org/10.1145/3130859.3131299

[82] Wellington P Yamaguti, Renata C Claudino, Alberto P Neto, Maria C Chammas, Andrea C Gomes, João M Salge, Henrique T Moriya, Alberto Cukier, and Celso R Carvalho. 2012. Diaphragmatic breathing training program improves abdominal motion during natural breathing in patients with chronic obstructive pulmonary disease: a randomized controlled trial. *Archives of physical medicine and rehabilitation* 93, 4 (2012), 571–577. https://doi.org/10.1016/j.apmr.2011.11.026

[83] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. 2017. DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 351–360. https://doi.org/10.1145/3038912.3052577

[84] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. In *The IEEE International Conference on Computer Vision (ICCV)*. 2223–2232.