

Article

# Predicting Emotion and Engagement of Workers in Order Picking Based on Behavior and Pulse Waves Acquired by Wearable Devices

Yusuke Kajiwara <sup>1,\*</sup> , Toshihiko Shimauchi <sup>2</sup> and Haruhiko Kimura <sup>1</sup>

<sup>1</sup> Department of Production Systems Engineering and Sciences, Komatsu University, Shichomachi Nu1-3, Komatsu, Ishikawa 923-8511, Japan; haruhiko.kimura@komatsu-u.ac.jp

<sup>2</sup> Department of Creative Community, Komatsu College, Shichomachi Nu1-3, Komatsu, Ishikawa 923-8511, Japan; shimauchi@komatsu-c.ac.jp

\* Correspondence: yusuke.kajiwara@komatsu-u.ac.jp; Tel.: +81-761-48-3152

Received: 24 November 2018; Accepted: 30 December 2018; Published: 4 January 2019



**Abstract:** Many logistics companies adopt a manual order picking system. In related research, the effect of emotion and engagement on work efficiency and human errors was verified. However, related research has not established a method to predict emotion and engagement during work with high exercise intensity. Therefore, important variables for predicting the emotion and engagement during work with high exercise intensity are not clear. In this study, to clarify the mechanism of occurrence of emotion and engagement during order picking. Then, we clarify the explanatory variables which are important in predicting the emotion and engagement during work with high exercise intensity. We conducted verification experiments. We compared the accuracy of estimating human emotion and engagement by inputting pulse wave, eye movements, and movements to deep neural networks. We showed that emotion and engagement during order picking can be predicted from the behavior of the worker with an accuracy of error rate of 0.12 or less. Moreover, we have constructed a psychological model based on the questionnaire results and show that the work efficiency of workers is improved by giving them clear targets.

**Keywords:** order picking; emotion; engagement; the wearable sensor; deep neural network; flow experience

---

## 1. Introduction

Many logistics companies adopt a manual order picking system. Order picking is a task of picking items on the shelves according to the order form and delivering them [1]. To reduce costs [2], related research has been conducted to reduce human errors and improve the work efficiency of workers. This related research has revealed the effects of stress, motivation, and work satisfaction on order picking. The worker often experiences a positive flow when the difficulty of work and the ability of the worker are balanced and he/she receives a clear target and feedback. The worker in a flow experience is engaged in the task and evokes a pleasant emotion. There are two kinds of affect: emotion and mood. Emotions are defined as affective states that change during a short time. Mood is defined as a relatively long-lasting affective state, which is different from emotion, as mood is less specific and less tense [3]. Engagement is defined as when “all the attention of workers is directed to completing a task, and workers feel the transformation of time sense” [4]. Pleasant emotions increase the concentration of a worker and encourage positive behavior [5]. Furthermore, pleasant emotions have an effect to increase the immune strength of the whole body, and they contribute to the maintenance of the physical health of the worker [6]. Engagement improves the performance of the worker [7] and gives a sense of

fulfillment [8,9]. Pleasant emotions and engagement contribute to a higher job retention rate [10,11]. By managing a worker's psychological state, the organization can maintain the motivation of workers and improve the productivity of order picking. However, it is difficult to measure the psychological state after each order picking using a questionnaire because it interrupts the work and reduces the productivity of the work. In addition, often when we conduct a questionnaire frequently, the worker will embrace unpleasant emotions and distracts attention from the work. Therefore, it is desirable to develop a method to predict the psychological state of the workers for each task with less burden on the worker.

In this research, we propose a system that utilizes wearable sensors and deep neural networks to predict the degree of worker's emotion and engagement. First, behaviors and pulse waves are acquired with wearable sensors. Second, time series features are calculated from the behaviors and pulse waves. Third, we predict the emotion and engagement by inputting the time series feature to deep neural networks. In this research, emotions are represented by Russell's circular model of affect [12]. Finally, we maximize the work efficiency and minimize the human error of workers by improving work environment based on the prediction results and psychology model. The contributions of this research are as follows:

- (1) Emotion and engagement of workers were predicted from behavior and pulse waves acquired with wearable devices and the accuracy of the prediction was clarified. Furthermore, we identified important variables to predict emotion and engagement.
- (2) We construct a psychology model of worker and clarified effects of emotion and engagement.

Section 2 describes the definition of flow experience, the method of measuring the psychological effects of flow with references to related research. Section 3 describes the method to predict the emotion and engagement using wearable sensors and deep neural networks. In Section 4, we show the prediction accuracy and clarify the explanatory variables which are important in predicting the emotion and engagement during work with high exercise intensity. In addition, we show the validity of psychology model of the worker from questionnaire results. Section 5 draws conclusions and outlines the future directions of our work.

## 2. Related Research

### 2.1. Definition of Flow Experience

A worker measures flow experience by a flow state scale [13] (Table 1). The scale consists of a balance of task difficulty and worker skill, a fusion of behavior and consciousness, control behavior, instruction of clear target and feedback, concentration, a transformation of time sense, evocation of emotion.

**Table 1.** Item of flow state scale.

Symbol	Item
$q_{FE}^{(1)}$	I felt that the task difficulty and worker skill of the work are balanced.
$q_{FE}^{(2)}$	I was fused with behavior and consciousness.
$q_{FE}^{(3)}$	I had a clear target.
$q_{FE}^{(4)}$	I had a clear feedback.
$q_{FE}^{(5)}$	I paid all the attention to the work.
$q_{FE}^{(6)}$	I had a sense of controlling the behavior.
$q_{FE}^{(7)}$	I could move my body naturally.
$q_{FE}^{(8)}$	I feel the transformation of time sense.
$q_{FE}^{(9)}$	I enjoyed a task.

Each item of the flow state scale is evaluated on the Likert scale [14] with 5-point scales or 7-point scales. The degree of flow experience is evaluated on the average value of items of the flow state scale. The evocation of emotion is measured by the circumplex model of affect [12] and the general affect scale [15].

Flow experience has been shown to be effective in various fields. Flow experience occurs mainly when the task difficulty and the worker's skill are balanced [16]. When this balance is achieved, the worker is fused with behavior and consciousness, which reduces the cognitive load and improves job performance [17]. The worker who behaves autonomously with a clear goal has high job satisfaction [18,19]. The worker with flow experience evokes pleasant emotions and displays increased subjective well-being [20,21].

## 2.2. Measurement of Emotion

Emotion is measured from behavior and vital signs using wearable devices. For the indicators of behaviors, the pitch of the voice [22] and facial expressions [23] have been used. For the indicators of the vital signs, pulse rate variability, respiratory variability, and the electrodermal activity have been employed [24–26]. These indices directly or indirectly express the degree of activity of autonomic nerves. Organisms have the property of controlling the autonomic nervous system when an external/internal stimulus is applied and keeping physiological parameters (heart rate, blood glucose level, etc.) within an allowable range. This function is called homeostasis [27]. Internal stimuli include affects such as anxiety and anger. Therefore, affects are predicted by measuring the activity of the autonomic nervous system controlled by homeostasis. However, the physiological parameters kept by homeostasis change depending on the type and condition of stimulation [28]. Therefore, when the same stressor is given under different condition, the response of the vital sign often differs. In a related research, a video was shown to the subject to evoke a specific emotion [29] and predicted the emotion from the change of the vital signs and behavior on real-time basis using wearable devices [30]. However, these researches were directed to emotions of resting subjects. Since the vital signs and behavior is an indirect indicator of emotion, it is not always possible to estimate the emotion during order picking by the indicator of the vital signs and behavior at rest. A related study predicts pleasant and unpleasant moods in everyday life with an accuracy of 60% by acquiring gait and pulse rate variability using a wristwatch pulse wave meter and smartphone [31]. However, this related study predicts the mood of the day, Hence, it is difficult to predict the emotions that change in a short time like during order picking.

Since the worker performs a task alone and is appraised by the outcome, there are few workers that talk to others worker and change facial expressions during order picking. In addition, the worker works in order picking while moving the whole body. Therefore, it is meaningful to explore new variables for effectively predicting emotions in order picking.

## 2.3. Measurement of Engagement

The effects of engagement on the worker are less cognitive load and full attention to the task. The cognitive load was defined as the mental resources a person possesses for solving problems or completing tasks at a given time [32]. The cognitive load and attention are measured from behavior and vital signs using wearable devices such as inertial sensors [33,34], eye tracker [35,36], electroencephalographs [37], pulse wave meters [38]. A related research loaded working memory [39] of subjects on cognitive tasks and attention task such as semantic fluency task, elementary cognitive tests, and visual search task, and predicted the cognitive load from the change of the vital signs and behavior on a real-time basis using wearable devices [40]. The related research estimated the user's cognitive load and attention during rest such as personal computer work [40] and driving a car [41]. In addition, the related research predicted the cognitive load of pedestrians using inertial sensors attached to the waist. The related research predicted engagement from behavior and vital signs when a worker is at rest. There are studies that predict cognitive load during action, but its

behavior is simple. The worker performs complicated behavior. Therefore, it is meaningful to explore new variables for effectively predicting engagement during order picking.

### 3. Predicting Emotion and Engagement Using Deep Neural Networks

Figure 1 shows the prediction of emotion and engagement using deep neural networks. In this research, we calculate features from behaviors and pulse waves acquired from wearable devices and input them into deep neural networks, thereby predicting the emotion and engagement. The emotion is represented by a circumplex model of affect as illustrated in Figure 2. In the circumplex model of affect, emotions are represented by mutually orthogonal pleasant-unpleasant axis and arousal-sleepy axis. Workers can easily describe their current emotions by simply marking the corresponding square in the model, which includes  $7 \times 7$  squares.

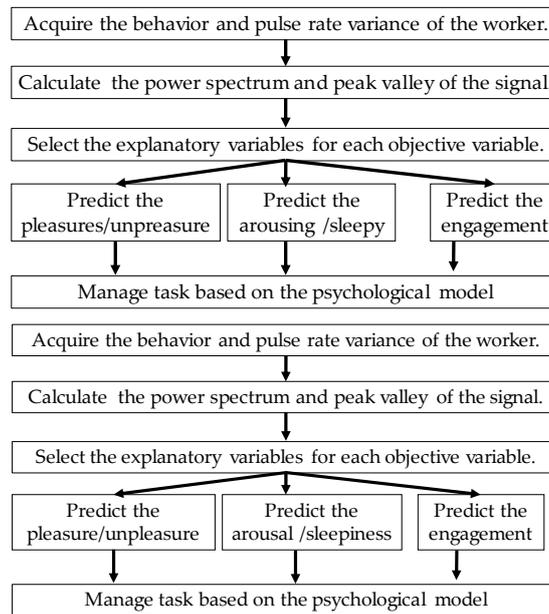


Figure 1. Task management using the prediction of pleasure, arousal, and engagement.

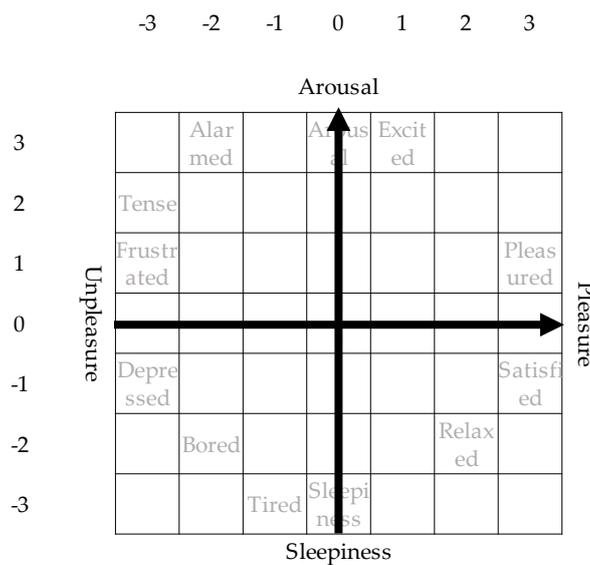


Figure 2. A questionnaire based on the circumplex model of affect.

The model drastically shortens the time required for a subject to answer compared with PANAS [15]. PANAS is a questionnaire asking about five positive emotions (relaxed, happy, concentrated, interested, and active) and five negative emotions (tired, stressed, sleepy, angry, and depressed). Therefore, the model is appropriate for situations where we often ask workers about their current emotions. The prediction values of emotion and engagement are visualized and used as an indicator of business improvement. Business improvement is conducted based on a psychology model.

### 3.1. Behavior and Pulse Wave Feature Calculation

To predict the emotion and engagement, we acquired the behavior and pulse wave of the worker from the beginning to the end of assigned task by using wearable devices. In this research, we use a pulse wave sensor, the motion detectors, and an eye tracker to acquire the time series data listed in Table 2. When the workers evoked an emotion, they activate the autonomic nerve and change the pulse wave [24]. When the workers' pay full attention to a task, they change the movement of the waist while walking [33], pulse rate variability [38], pupil diameter [35], head movement [35], and eye-gaze [36]. The eye tracker used was the Tobii Pro Glasses 2 made by Tobii (Tokyo, Japan). The tracker is equipped with an acceleration sensor, an angular velocity sensor, two cameras for recording eye movements and viewpoint. The sampling rate of the cameras was 40 Hz. The acceleration and angular velocity of XYZ coordinates were acquired. The sampling rate of the acceleration sensor and angular velocity sensor was 100 Hz. The pulse wave sensor is a vital meter made by TAOS (Kanagawa, Japan). The sensor is equipped with a photoelectric pulse wave meter, an acceleration sensor, and an angular velocity sensor. The pulse wave meter acquires a pulse wave. The sampling rate of the pulse wave were 1000 Hz. The RR interval is calculated from the pulse wave. The frequency component of the RR interval is defined as very low frequency (VLF) at 0.04 Hz or less, low frequency (LF) at 0.04 Hz to 0.15 Hz, and high frequency (HF) at 0.15 Hz to 0.4 Hz. The combined power spectrum density of VLF, LF, and HF is defined as total power (TP). The acceleration and angular velocity sensors were put in the breast pocket. The sampling rate of the acceleration and angular velocity were 1000 Hz. Entropy and Lyapunov exponent are calculated from the acceleration and angular velocity. To acquire the time series data of VLF, LF, HF, TF, Lyapunov exponent, and entropy, time window was prepared and slid the time window at 1 s intervals. The time window size was 30 s. Therefore, the sampling rate of the VLF, LF, HF, TF, Lyapunov exponent, and entropy was 1 Hz. The motion detector is Neuron made by NOITOM (Beijing, China). The sampling rate of the motion detector was 50 Hz. The pulse wave sensor, eye tracker, and motion detector have sufficient sampling rate to calculate time series features.

**Table 2.** Time series features are calculated from time series data acquired with wearable devices.

Device	Types and Part of the Body		
	Power Spectrum Value	Peak Valley Value	Time
Pulse wave sensor	HF, LF, VLF, TP, Heart rate (HR), Lyapunov exponent, Entropy	VLF, LF, HF, TP, Heart rate (HR), R-R interval (RR), Lyapunov exponent, Entropy	-
Eye tracker	Eye movement in XYZ coordinate system(3D), Point of view(2D), Pupil diameter (PD), Triaxial accelerometer, Triaxial angular velocity	Eye movement in XYZ coordinate system(3D), Point of view(2D), Pupil diameter (PD), Triaxial accelerometer, Triaxial angular velocity	Eye movement, Eye blink (BK), Saccade
Motion detector	Left and right hands(LH, RH), Left and right elbow(LA, RA), Left and right shoulder(SL, SR), Buttocks(BT), Lumbar spine(S1), Thoracic spine(S2), Cervical spine(S3), head(HD) of the rotation angle	Left and right hands(LH, RH), Left and right elbow(LA, RA), Left and right shoulder(SL, SR), Buttocks(BT), Lumbar spine(S1), Thoracic spine(S2), Cervical spine(S3), head(HD) of the rotation angle	-

According to Table 2, the task complexity and time series features are calculated from items and time series data of the behaviors and pulse waves. The mathematical symbols corresponding to the time series features calculated in this research are shown in Table 3. Type described in  $u^{(type)}$  means type of time series data acquired by eye tracker such as eye movement in XYZ coordinate system(3D), point of view(2D), and pupil diameter (PD). Type described in  $p^{(type)}$  means type of time series data acquired by pulse wave sensor such as VLF, LF, and HF. Part in described in  $a^{(part)}_{axis}$ ,  $g^{(part)}_{axis}$ , and  $r^{(part)}_{axis}$  means the part of the body to which the sensor is attached. The time series features are average, standard deviation, minimum value, maximum value, the power spectrum density, peak-valley value, and entropy. The power spectrum density, which represents the periodic feature of the signal, is calculated by Fourier transform. In this research, the frequency resolution was set to 1/10th of the Nyquist frequency. The peak-valley value, which represents the kurtosis of the signal, is calculated by the following procedure: First, the extremum of the signal is acquired. Second, the acquired extremum is normalized to the maximum value. Third, we apply a threshold to the extreme and gain a significant change in the signal. When  $extremum \leq 0.5 + threshold$  or  $extremum \geq 0.5 - threshold$ , the extremum is adopted. Finally, the average value, standard deviation, detection time, minimum value, and maximum value are calculated from the adopted extremum. In this paper, time series features were described with reference to Tables 2 and 3. For example,  $\sigma(v(e(a^{S3}), 0.25))$  means the standard deviation of the peak valley value of threshold = 0.25 of the entropy of the cervical spine acceleration.

**Table 3.** Correspondence table of each feature quantity and mathematical symbol.

Mathematical Symbols	Meaning of Symbols
$m(data)$	Average of d time series data
$\sigma(data)$	Standard deviation of time series data
$\eta(data)$	Minimum value of time series data
$\zeta(data)$	Maximum value of time series data
$c(data)$	Number of extremes of time series data
$v(data, threshold)$	Peak valley value of time series data when setting threshold
$f(data, frequency)$	Power spectrum value at frequency of time series data
$e(data)$	Entropy of time series data
$a^{(part)}_{axis}$	Acceleration of each axis of acceleration sensors attached to a part of the body
$g^{(part)}_{axis}$	Angular velocity of each axis of angular velocity sensors attached to a part of body
$r^{(part)}_{axis}$	Rotation angle of each axis of motion detectors attached to a part of body
$p^{(type)}$	Types of time series data acquired by the pulse wave sensor.
$u^{(type)}$	Types of time series data acquired by eye tracker

The power spectrum density of VLF represents breathing variability. The power spectrum density of LF represents the activity of the sympathetic and parasympathetic nervous systems. The power spectrum density of HF represents the activity of the parasympathetic nervous system. An excited worker activates the sympathetic nervous system and increases the power spectrum density of LF. A relaxed worker activates the parasympathetic nervous system and increases the power spectrum density of HF. The power spectral density of LF and HF represent the relaxed interval of the worker. The Lyapunov exponent and entropy, which represent the stability of order picking, are calculated from the acceleration sensor attached to the pulse wave sensor. Pulse waves change with behaviors and emotions evoked. Lyapunov exponent and entropy change with behaviors. Therefore, the power spectrum density of the pulse wave sensor shows that the emotions and behaviors change at a regular interval. In addition, the peak-valley value of the pulse wave sensor represents a significant change in behaviors and emotions.

The motion detectors acquire the rotation angle of the body joints. The rotation angle of joints changes with behaviors. The power spectrum value of the rotation angle of joints represents the rhythm of the work. In addition, the peak-valley value of the rotation angle of joints represents a significant change in behaviors.

The eye tracker acquires the eye-gaze, the pupil diameter, the number of saccades, and the number of blinks. The movements of the eye-gaze and the head change with order picking. The power spectral densities of eye-gaze and head represent the rhythm of work. The peak-valley values of eye-gaze and the head represent a significant change in behaviors. The pupil diameter represents a change of emotion. The power spectrum density of the pupil diameter shows that the emotion changes at a regular interval. The peak-valley value of pupil diameter represents a significant change in emotion. The number of blinks represents the degree of arousal.

The task complexity is represented by items and degree of overlap of items listed in the order sheets. When an area  $j$  is listed in order sheets, a dummy variable  $l_j (j = 1, 2, \dots, 6)$  is set to 1, otherwise, it is set to 0. Second, to calculate the degree of overlap, when an area  $j$  is listed in an order sheet  $l$ , a dummy variable  $l_j^{(l)}$  is set to 1, otherwise, it is set to 0. The degree of overlap is calculated as the average  $m(l)$  of  $l_j^{(l)}$  and the standard deviation  $\sigma^{(l)}$  of  $l_j^{(l)}$ . In addition, we acquire working time  $t$ . In this research, there are six areas in the experiment environment. Therefore, the task complexity is represented by nine dimensions which combine  $l_j$ ,  $m(l)$ ,  $\sigma^{(l)}$ , and  $t$ . The task complexity is added to the time series features of the pulse wave sensor, the motion detectors, and the eye tracker.

### 3.2. Variable Selection by a Hypothesis Test

From Tables 2 and 3, we select variables effective in predicting the emotion and engagement by the Brunner Munzel test. First, an emotion and engagement are divided into two groups. The horizontal axis is divided into pleasant (PL) and unpleasant (UPL). We defined as UPL when the value in the pleasant-unpleasant axis was less than 0, otherwise, it was defined as PL. The vertical axis is divided into arousal (AR) and sleepiness (SE). We defined as SE when the value in arousal-sleepy axis was less than 0, otherwise, it was defined as AR. Engagement is divided into boredom (BR) and concentration (CR). The engagement was measured with a value of 1 to 5. We defined as BR when the questionnaire of engagement was less than 3, otherwise, it was defined as CR. We make a null hypothesis that there is no difference in a feature element of both groups of an emotion and engagement and conduct the Brunner Munzel test. The rejection region is 0.01. When the null hypothesis is rejected, a feature element has a significant difference between the two groups. Therefore, the feature element is selected as a variable effective in predicting emotion and engagement. Explanatory variables are consisting of the feature elements selected by Brunner Munzel test.

### 3.3. Prediction of Emotion and Engagement by Deep Neural Networks

In this research, we construct a regression model using deep neural networks. Deep neural networks using auto-encoder [42] learn the explanatory variables when emotion and engagement occurs. Deep learning has a high feature extraction ability, and excellent prediction accuracy are shown in papers of various fields. Therefore, we use deep neural networks to predict pleasant emotion, arousal, and engagement. Deep neural networks consist of an input layer, hidden layers, and an output layer. The explanatory variables are inputted to the input layer. First, deep neural networks adjust the weight of each unit in layers using auto-encoder. The auto-encoder adjusts the weight to output the same signal as the input signal. Therefore, in the deep neural network of the auto encoder type, the features of the input layer are aggregated and input to the output layer. Deep neural networks adjust the weight of each unit in the output layer by fine-tuning.

Deep neural networks calculate the effect size of each explanatory variables from the weight of each unit in layers. A variable with a high effect size is an important variable to predict emotion and engagements. The followings are the parameters for this study: Hidden layer size was 3. Unit size was 100. Activation function was Rectifier. Dropout ratio was 0.5.

## 4. Experimental Results

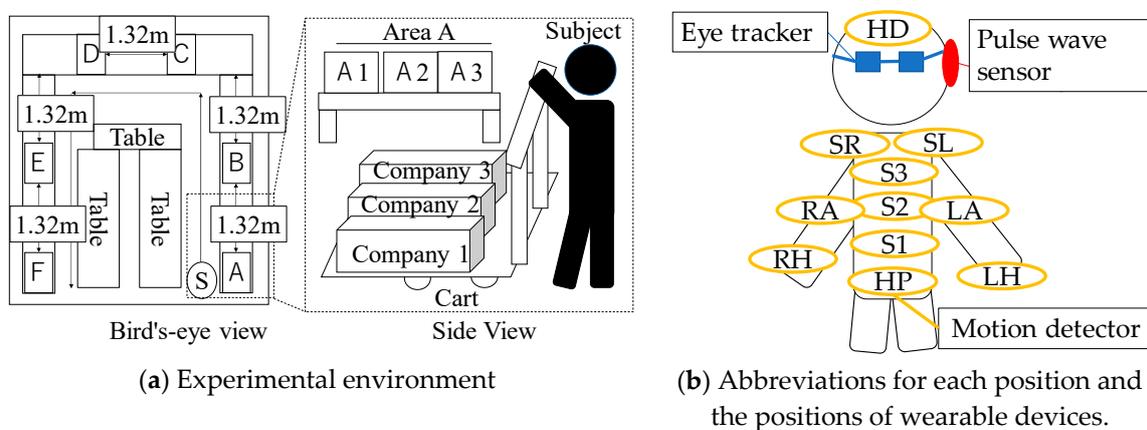
First, to show that emotion and engagement contribute to the productivity of the order picking, we construct the psychological model from the questionnaire results and clarify the effect of emotion

and engagement on work efficiency and human error. Second, we show the accuracy of prediction of pleasant, arousal, and engagement by deep neural networks. Then, it shows that pleasant, arousal, and engagement can be predicted from behaviors and pulse waves acquired by wearable devices. Furthermore, we identify important variables to predict pleasant, arousal, and engagement.

#### 4.1. Experimental Procedure

We conducted the experiment in an environment illustrated in Figure 3a. The subjects, 15 males and two females, were in good health condition. Subjects did not have order picking experience before the experiment. They were paid by the hour. The subjects picked items according to lists. The items were installed in the item box. The items to be picked were put in a delivery box on a hand truck. The subject visited the work area only once and picked the items. We instructed the subjects to finish order picking within one minute. After each picking session, the subjects answered pleasant, arousal, engagement, and subjective task difficulty experienced during the session. The subjects plotted the emotions into the  $7 \times 7$  squares according to the circumplex model of affect in Figure 2. The subjects answered the level of engagement and subjective task difficulty in 5-point scales.

The delivery boxes and the lists are prepared for company 1–3. We prepared each product list for each delivery BOX. The number of areas was set to six, from Area A to F. Each item box contained four items to be picked. The size of the item box and the delivery box was  $250 \text{ mm} \times 160 \text{ mm} \times 150 \text{ mm}$ . To adjust the complexity of order picking, we created a simple list and a complex list according to the following rules. The type of items and the number of items is determined by a uniform random number. The maximum number of items is 3. The subjects should pick all the items on the list by visiting three areas. A simple item list is set so that the items and areas do not overlap between lists. A complex list is set so that the items and area overlap between the lists.



**Figure 3.** The experimental environment and the position of wearable devices.

The subjects ran 32 order picking sessions. The subjects rested for 10 min after the 16th session to remove the effects of physical fatigue. Sessions from 1st to 16th are defined as the first period, and from 17th to 32nd as the second period. To verify the effect of the target on flow experience, we gave a clear target before the second period began. The subjects are told that “the wages are paid in full for the first period regardless of the outcome of work. However, for the second period, the wages are adjusted according to the outcome of work”. The merit-based pay system is widely used in various fields and clarifies the intention of the management to achieve a target. To clarify the effects of the target, we objectively evaluated the outcome by evaluating human error and work efficiency. The human error refers to the number of items that are picked incorrectly. Work efficiency is calculated by the Equation (1):

$$w = (1 - t/60), \quad (1)$$

where  $t$  is the working time. Since the worker performed order picking within 60 s, we divided  $t$  by 60.  $t$  and human error were recorded after each order picking. However, in fact, the wages are paid in full for the first and the second periods regardless of the outcome of work. We explained the fact to the subjects and got the consent of subjects.

To verify the effect of feedback on flow experience, the subjects were randomly given feedback in the first or the second period. All the feedback was positive. Nine subjects received feedback in the first period and the remaining eight subjects received feedback in the second period.

To verify the effect of flow experience, the subjects responded to the items of the flow state scale [13] in Table 1 at the ends of the first period and the second periods based on their task performance. The subjects wore an eye tracker, a pulse wave sensor, and motion detectors as illustrated in Figure 3b. Table 3 shows that the time series features are calculated from the acquired signals by the wearable devices.

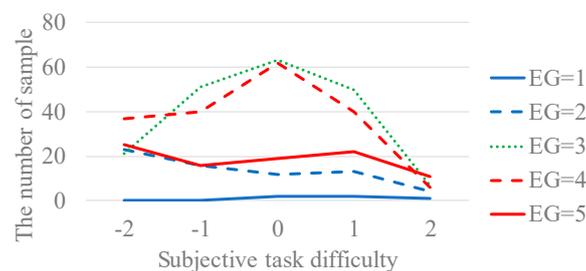
#### 4.2. Validation of the Psychological Model

Table 4 shows the sample sizes of each square in the circumplex model of affect. The largest square is the first quadrant, followed by the second, the fourth and the third quadrants. From Table 4, there are many in the order of the first quadrant, the second quadrant, the fourth quadrant, and the third quadrant of circumplex model of affect.

**Table 4.** The sample size of each square of circumplex of model of affect.

Arousal/Pleasure	−3	−2	−1	0	+1	+2	+3
+3	2	5	1	10	11	43	21
+2	4	4	28	31	36	79	7
+1	2	5	3	39	32	25	5
0	1	4	4	47	9	6	5
−1	2	3	15	15	9	3	4
−2	0	1	2	0	11	4	3
−3	0	0	0	2	0	0	0

We analyzed the effect of subjective task difficulty on engagement. Table 4 shows the sample sizes of the questionnaire result of subjective task difficulty and engagement. From Figure 4, the engagement level of 3 and 4 occurred most frequently when subjective task difficulty was 0, while their occurrence decreased when subjective task difficulty is less than 0 or greater than 0. The engagement level of 1, 2, and 5 were uniformly distributed regardless of subjective task difficulty.

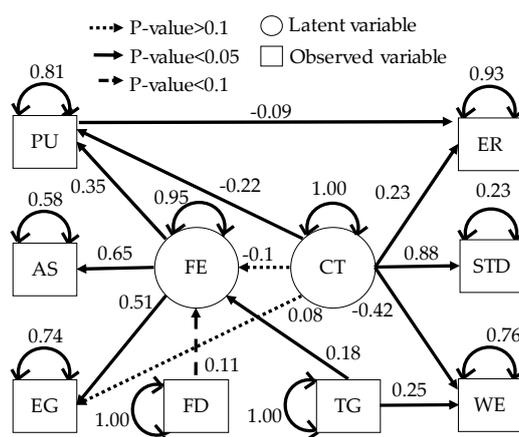


**Figure 4.** The sample size of engagement (EG) and subjective task difficulty.

Subjective task difficulty and engagement were uncorrelated. From the above, adjusting the subjective task difficulty is effective to make the worker reach engagement level of 3 and 4. From these results, it is suggested that engagement level of 3 and 4 of the workers can be controlled by adjusting subjective task difficulty. However, the engagement was 1, 2, and 5 of the workers cannot be controlled by adjusting subjective task difficulty.

The psychological model was constructed from the results of the hypothesis test and factor analysis and it has verified the goodness of fit of the model by covariance structure analysis. The results of the hypothesis test and factor analysis are shown in Appendix A.

Predicted emotion and engagement contribute to the maximization of work efficiency and minimization of human error. The sample size was 544 (the number of subjects  $\times$  the number of session =  $17 \times 32$ ). The goodness of fit of the model was evaluated by the chi-square test. As a result, the  $p$ -value was 0.61, and the null hypothesis was adopted. The comparative fit index (CFI) was 1.00, and the Tucker-Lewis index (TLI) was 1.00. RMSEA was 0.00. We make a null hypothesis that RMSEA is less than 0.05 and tested. As a result, since the  $p$ -value was 1.00, the null hypothesis was adopted. The results proved the validity of the model in Figure 5.



**Figure 5.** Covariance structure analysis result of the psychological model (PU: Pleasant emotion, AS: Arousal, EG: engagement, FE: flow experience, CT: complexity of the task, TG: target, FD: feedback, ER: human error, STD: subjective task difficulty, WE: work efficiency.).

Since the  $p$ -value of the effect of the complexity of the task on engagement and flow experience was 0.05 or more, the effects were not statistically significant. Since the  $p$ -value of the effect of feedback on flow experience was 0.05 or more, the effect was not statistically significant. The results were equal to engagement decreased as subjective task difficulty is less than 0 or greater than 0. Therefore, the results show that flow experience is adjusted by the subjective task difficulty. Pleasant emotions gave vitality to workers and reduced human error. Target improved work efficiency. Flow experience mainly affected arousal and engagement. On the other hand, pleasant emotion was affected by flow experience and complexity of the task. The effect of complexity of the task on pleasant emotion suggests that negative emotions were evoked when complicated work was given while wage is reduced by the outcome. It suggests that pleasant emotion is changed by the cognitive process of the worker.

The related research [43] has shown that flow experience mainly occurs when the task difficulty and worker's skill are balanced. On the other hand, the effect of flow experience on work efficiency and human error in order picking is not shown. The effect of feedback depends on the way of feedback. The effect of feedback depends on how the feedback is given to a worker. Human error is affected by the complexity of the task and pleasant emotions. Flow experience does not directly affect human error. However, Figure 5 showed that pleasant emotion evoked by flow experience affects human error. Therefore, organizations and supervisors can reduce human error by managing the flow experience of the worker. Since the complexity of the task affects human error, flow experience, and pleasant emotion, they can be controlled by adjusting the complexity of the task. Since target affects flow experience and work efficiency, target contributes to work efficiency. Target and complexity of the task are externally controllable factors. Therefore, organizations and supervisors can generate preferable flow experience by managing target and complexity of the task to improve work efficiency in order picking and reduce the human error of the worker.

#### 4.3. Prediction of Pleasant Emotion, Arousal, and Engagement Using Deep Neural Networks

We clarify the prediction accuracy with deep neural networks which learned the time series features of the pulse wave sensor, the eye tracker, and the motion detectors when the worker evoked pleasant emotion, arousal, and engagement. In addition, we identify the important variables for predicting pleasant emotion, arousal, and engagement. We compared the accuracy of the generalized model with the accuracy of the personalized model. The comparison results show that we acquired knowledge for improving the accuracy of the generalized model.

There are two types of predictive models of supervised learning. The first type is the generalized model. The sample size was 544 (the number of subjects  $\times$  the number of session =  $17 \times 32$ ). The generalized model is constructed by training deep neural networks with the questionnaire results and time series features of workers other than the target person. The target person can receive prediction results simply by inputting the time series features to the generalized model. The second type is a personalized model. The personalized model is constructed by training deep neural networks with the questionnaire results and time series features of the target person by deep neural networks. Since the personalized model requires the target person to answer the questionnaire, the usability of the personalized model is less than that of the generalized model. However, the prediction accuracy of the personalized model is greater than that of the generalized model. Therefore, by analyzing the difference between the generalized and personalized models, we can acquire knowledge for improving the accuracy of the generalized model.

##### 4.3.1. Accuracy of Generalized Model

We constructed a generalized model by training deep neural networks with the questionnaire results and time series features of workers other than the target person. The target person can receive prediction results simply by inputting the time series features to the generalized model. The accuracy of the generalized model was verified. We conducted 17-fold-cross-validation with a worker as the test set and the other as the training set. The test set is the time series features of a subject. The training set is time series features of subjects other than the subject selected as the test set. The prediction accuracy is evaluated by the mean squared error normalized in Equation (2):

$$Error = (y' - y)^{1/2} / (\zeta(y') - \eta(y')), \quad (2)$$

where *Error* is a value which divided a difference between the questionnaire result  $y'$  and the predicted value  $y$  by a difference between the maximum value  $\zeta(y')$  and the minimum value of the questionnaire result  $\eta(y')$ . Mean squared error is traditionally used as an effective index for evaluating regression models. Therefore, in this research, the prediction model was evaluated using mean squared error.

In order to verify the effectiveness of variable selection in Section 3.2, all explanatory variables in Tables 2 and 3 were input to deep neural networks, and the prediction results were shown in Figure 6. The accuracy of predicting pleasant emotion, arousal, and engagement by the deep neural networks is shown in Figures 7–9. The prediction accuracy in Figures 7–9 was improved by 0.05 maximum in pleasant, 0.05 in maximum in arousal, and up to 0.03 in engagement compared to the accuracy in Figure 6. Therefore, the variable selection in this research was effective for improving prediction accuracy. The prediction accuracy of pleasant emotion and arousal was highest when predicted from the time series features of the pulse wave sensor. The prediction accuracy of engagement was highest when predicted from the time series features of the motion detectors. The results were shown that the prediction accuracy among individuals was small when the pleasant emotion, arousal, and engagement were predicted by inputting time series features of motion detectors into generalized model.

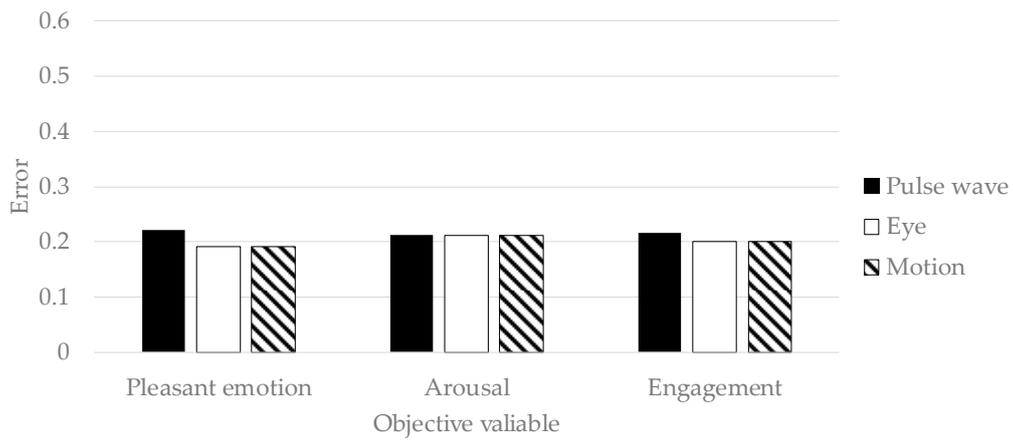


Figure 6. The accuracy of deep neural networks when inputted all time series features.

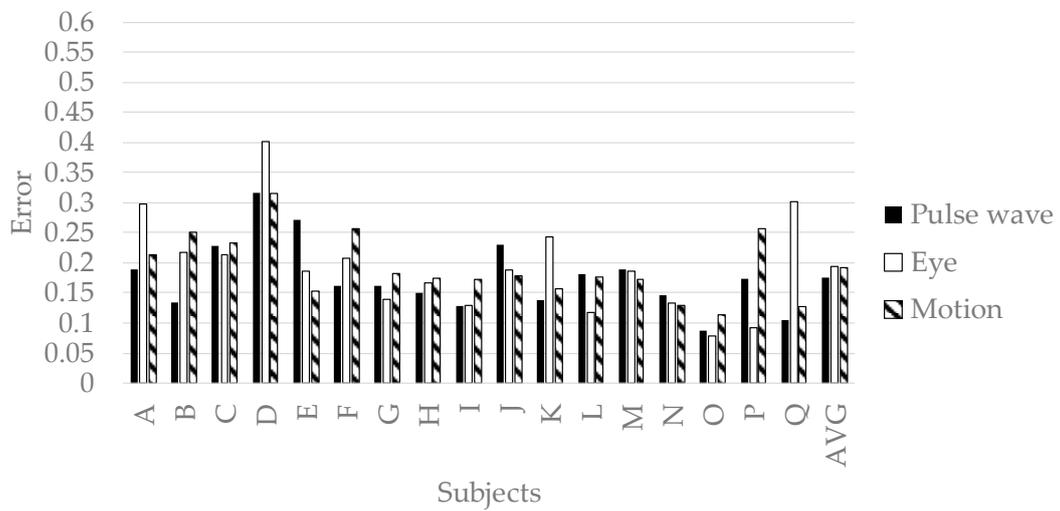


Figure 7. The accuracy of generalized model of pleasant emotion.

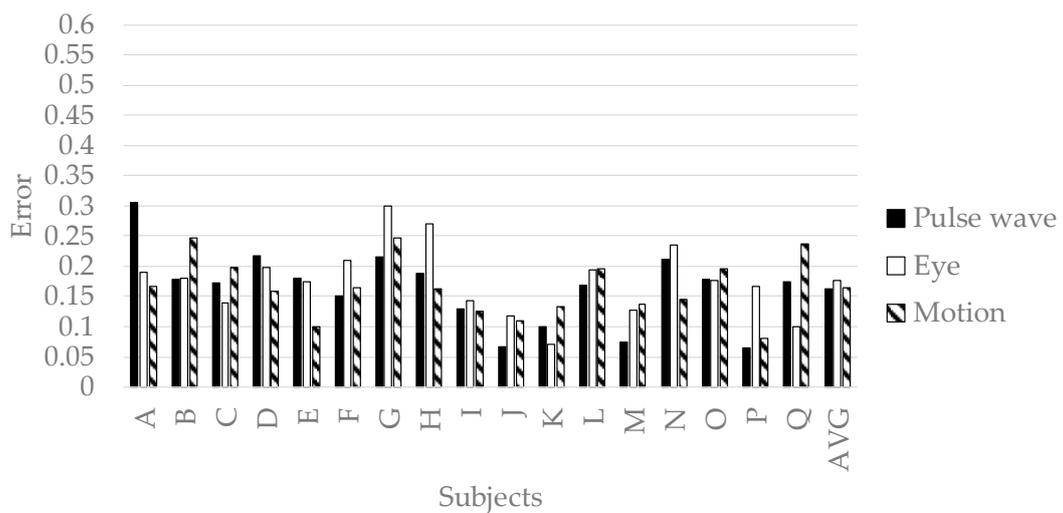


Figure 8. The accuracy of generalized prediction model of arousal.

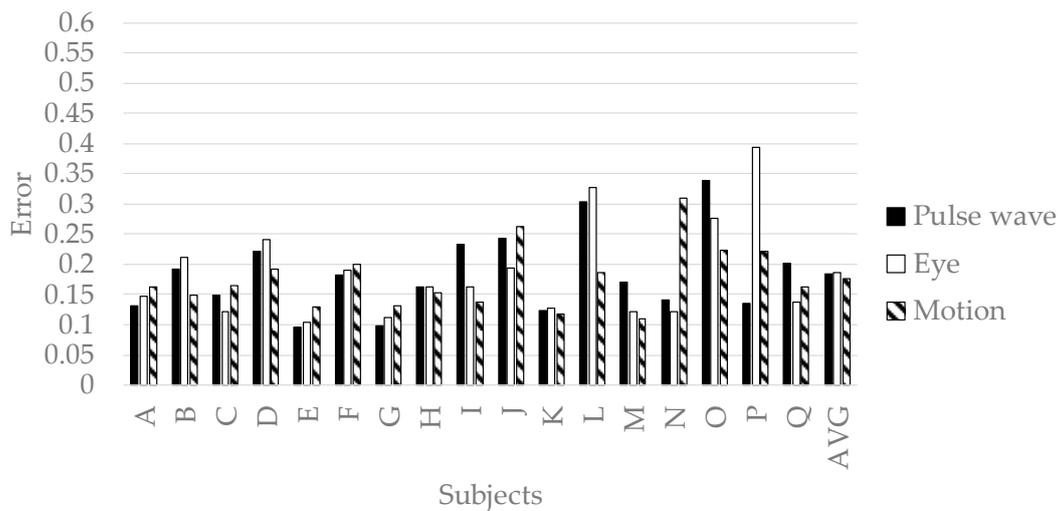


Figure 9. The accuracy of generalized model of engagement.

We show important variables for predicting the pleasant emotion, arousal, and engagement in Tables 5–7. Table 5 shows the movements of the chest and waist are important variables for predicting the pleasant emotion.  $f(e(a^{S3}), x)$  ( $x = 0.1, 0.15, 0.25, 0.3$ ) in the pulse wave sensor at PL is smaller than that at UPL. Moreover,  $\sigma(v(e(a^{S3}), 0.25))$  in pulse wave sensor at PL is greater than that at UPL.  $f(r_z^{BT}, x)$  ( $x = 7.5, 10, 12.5, 15, 17.5$ ) in the motion detectors at PL is smaller than that at UPL. The rotation angle of the buttocks and the acceleration of the chest changed in conjunction with the movement of limb and arms during walking and order picking.  $\eta(u^{PD})$  in the eye tracker at PL is greater than that at UPL. Therefore, there is little unnecessary movement during order picking when a pleasant emotion is evoked. These time series features represent behavior control. The control of behavior is a factor that produces flow experience. It is suggested that deep neural networks predict the occurrence probability of the flow experience from the behavior control and indirectly predict pleasant emotions, arousal, and engagement as secondary reactions from the occurrence probability of the flow experience.

Table 5. Top five of the important variables for predicting the pleasant emotion in the generalized model.

Pulse Wave Sensor			Eye Tracker			Motion Detector		
Variable	UPL	PL	Variable	UPL	PL	Variable	UPL	PL
$f(e(a^{S3}), 0.1)$	0.29	0.24	$\eta(u^{PD})$	2.91	3.12	$f(r_z^{BT}, 17.5)$	2.46	0.77
$f(e(a^{S3}), 0.15)$	0.16	0.12	$m(v(u_z^{3D}, 0.9))$	457	415	$f(r_z^{BT}, 15)$	2.85	0.87
$f(e(a^{S3}), 0.25)$	0.09	0.06	$m(u^{PD})$	3.79	3.98	$f(r_z^{BT}, 12.5)$	3.31	1.07
$f(e(a^{S3}), 0.3)$	0.07	0.05	$\sigma(v(u_z^{3D}, 0.75))$	417	435	$f(r_z^{BT}, 10)$	4.23	1.42
$\sigma(v(e(a^{S3}), 0.25))$	4.42	3.17	$m(v(u_z^{3D}, 0.75))$	523	461	$f(r_z^{BT}, 7.5)$	5.94	2.06

Table 6. Top five of the important variables for predicting the arousal in the generalized model.

Pulse Wave Sensor			Eye Tracker			Motion Detector		
Variable	SE	AR	Variable	SE	AR	Variable	SE	AR
$t$	47.3	43	$f(u_x^{2D}, 6)$	1039	948	$c(v(r_z^{LA}, 0.75))$	82.1	55.9
$c(v(p^{RR}, 0.5))$	5	3.64	$f(u_x^{2D}, 8)$	6238	5162	$c(v(r_y^{LA}, 0.75))$	69.5	51.7
$m(v(p^{HR}, 0.9))$	902	966	$m(v(a_z^{HD}, 0.75))$	2.84	3.9	$m(r^{LA})$	−39.1	−20.4
$m(v(p^{HR}, 0.5))$	892	945	$m(u^{BK})$	326	182	$c(v(r_z^{LA}, 0.25))$	220	194
$c(v(p^{HR}, 0.25))$	13.2	11.1	$f(a_z^{HD}, 45)$	455	235	$\eta(r^{LA})$	−81.1	−82.1

**Table 7.** Top five of the important variables for predicting the engagement in the generalized model.

Pulse Wave Sensor			Eye Tracker			Motion Detector		
Variable	BR	CR	Variable	BR	CR	Variable	BR	CR
$m(v(p^{LF}, 0.9))$	2325	1484	$f(u_x^{2D}, 2)$	942	1664	$f(r_z^{BT}, 10)$	1.4	1.81
$\sigma(v(p^{LF}, 0.9))$	3695	2711	$f(u_x^{2D}, 6)$	1068	944	$f(r_y^{BT}, 5)$	2.4	2.35
$m(v(p^{LF}, 0.25))$	2620	1708	$f(a_z^{HD}, 35)$	333	440	$m(v(r_y^{SL}, 0.9))$	0.49	0.2
$\sigma(v(p^{HF}, 0.9))$	4782	3317	$f(a_z^{HD}, 30)$	492	664	$m(v(r_y^{SL}, 0.25))$	0.36	0.14
$\sigma(v(p^{LF}, 0.75))$	3570	2633	$f(a_y^{HD}, 35)$	238	223	$m(v(r_x^{S2}, 0.25))$	1.57	3.39

Tables 2 and 3 list the abbreviations used in this study. The pulse wave sensor in Table 6 shows that  $t$ ,  $c(v(p^{RR}, 0.5))$ , and  $c(v(p^{HR}, 0.25))$  at arousal was smaller than that at sleepiness.  $m(v(p^{HR}, 0.5))$  and  $m(v(p^{HR}, 0.9))$  at arousal was greater than that at sleepiness. The eye tracker in Table 6 shows that  $f(u_x^{2D}, 6)$ ,  $f(u_x^{2D}, 8)$ , and  $m(v(a_z^{HD}, 0.75))$  at arousal was smaller than those at sleepiness.  $m(v(a_z^{HD}, 0.75))$  and  $m(u^{BK})$  at arousal were greater than those at sleepiness. The motion detector in Table 6 shows that  $m(r^{LA})$  at arousal was greater than that at sleepiness.  $c(v(r_z^{LA}, 0.25))$ ,  $c(v(r_y^{LA}, 0.75))$ ,  $c(v(r_z^{LA}, 0.75))$ , and  $\eta(r^{LA})$  at the arousal were smaller than those of the sleepiness. The results show that the movement of the body at arousal was agiler than that at sleepiness. The agile movements have high exercise intensity and increase the heart rate of a worker. Therefore, the heart rate at the arousal temporarily increases. The peak-valley value is normalized to the maximum value. The temporary increase in heart rate caused the threshold of peak-valley to be high. As a result, many small peak-valley values were excluded. Therefore, the number of peak-valley detected was small.

The pulse wave sensor in Table 7 shows that psychological state of a worker was stable since  $m(v(p^{LF}, 0.25))$ ,  $m(v(p^{LF}, 0.9))$ ,  $\sigma(v(p^{LF}, 0.75))$ ,  $\sigma(v(p^{LF}, 0.9))$ , and  $\sigma(v(p^{HF}, 0.9))$  during concentration were smaller than those during boredom. The eye tracker in Table 7 shows that  $f(u_x^{2D}, 2)$ ,  $f(a_z^{HD}, 35)$ , and  $f(a_z^{HD}, 30)$  during concentration were greater than those during boredom.  $f(u_x^{2D}, 6)$  and  $f(a_y^{HD}, 35)$  during concentration were smaller than those during boredom. The motion detector in Table 7 shows that  $f(r_z^{BT}, 10)$  and  $m(v(r_x^{S2}, 0.25))$  during concentration were greater than those during boredom.  $m(v(r_y^{SL}, 0.9))$ ,  $f(r_y^{S1}, 5)$ ,  $m(v(r_y^{SL}, 0.25))$ , and  $m(v(r_x^{S2}, 0.25))$  during concentration were smaller than those during boredom. These data show that there is little unnecessary movement in the order picking when the worker was engaged.

#### 4.3.2. Accuracy of the Personalized Model

We verified the prediction accuracy of the personalized model. We conducted 32-fold-cross-validation with an OP of the picker as the test set and the other as the training set. The test set is a time series features of a subject. The training set is time series features of the subjects other than the subject selected as the test set. The prediction accuracy is evaluated by the mean squared error normalized in Equation (2). Figures 10–13 show the accuracy of the personalized model. The prediction accuracy of pleasant emotion and engagement was highest when predicted from the time series features of the motion detectors. The prediction accuracy of arousal was highest when predicted from the time series features of the eye tracker. The prediction accuracy of the personalized model was improved by 0.08 maximum in pleasant, 0.07 in maximum in arousal, and up to 0.05 in engagement compared to the generalized model. The results showed that the prediction accuracy among individuals was small when the pleasant emotion, arousal, and engagement were predicted by inputting time series features of motion detectors into personalized model. On the other hand, there were subjects with remarkably low accuracy of pleasant emotion when inputting the time series features of pulse wave sensor into the personalized model.

We show important variables for predicting pleasant emotion, arousal, and engagement using the personalized model in Tables 8–10. The tables show that both the generalized and personalized models selected the same variables as important. However, the accuracy of the personalized model

was higher than that of the generalized model. The results showed that there are individual differences in the speed of walking and order picking.

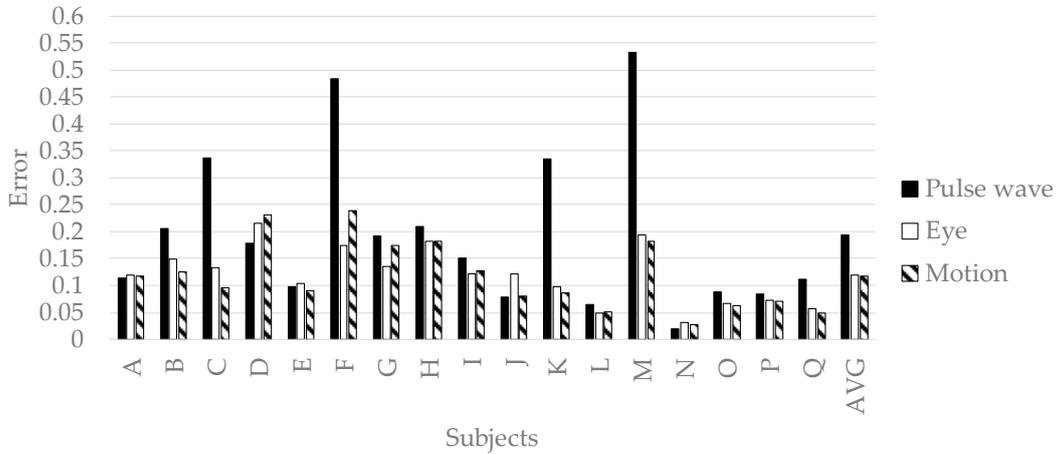


Figure 10. The accuracy of personalized model of pleasant emotion.

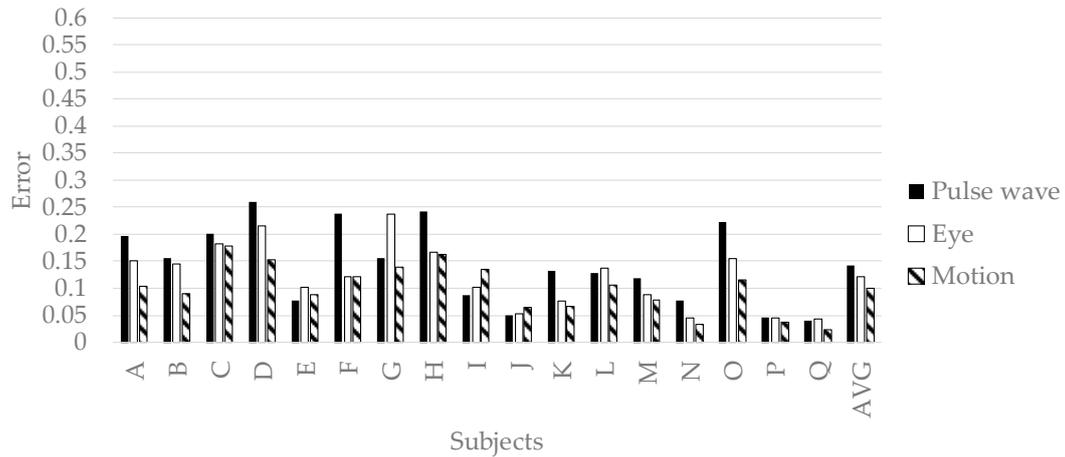


Figure 11. The accuracy of personalized prediction model of arousal.

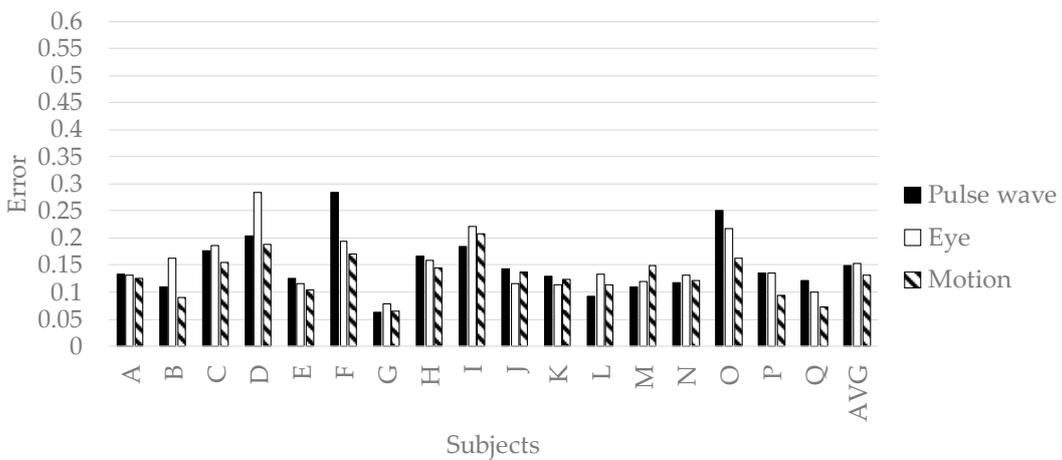


Figure 12. The accuracy of personalized model of engagement.

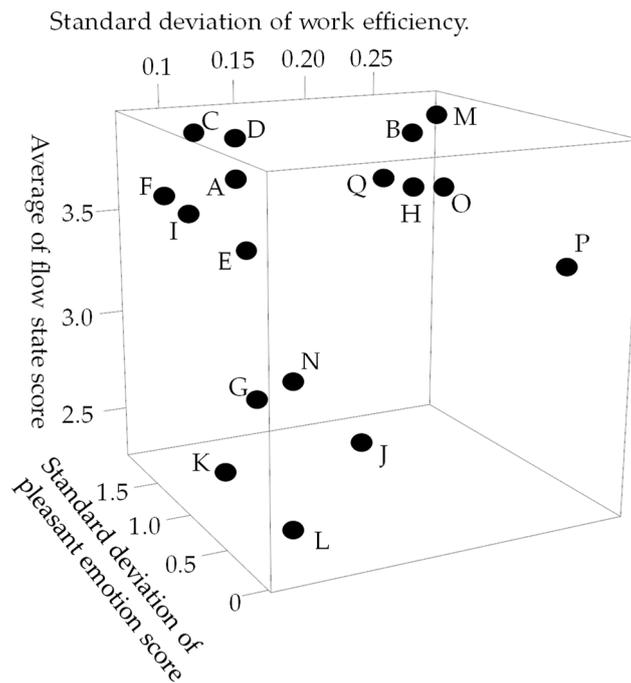


Figure 13. The accuracy of personalized model of pleasant emotion.

Table 8. Top five of the important variables for predicting the pleasant emotion in the personalized model.

Pulse Wave Sensor			Eye Tracker			Motion Detector		
Variable	UPL	PL	Variable	UPL	PL	Variable	UPL	PL
$\sigma(v(e(a^{S3}),0.9))$	0.39	0.30	$\eta(u^{PD})$	2.91	3.12	$f(r_z^{BT}, 15)$	2.85	0.87
$\sigma(v(e(a^{S3}),0.75))$	0.36	0.28	$m(u^{PD})$	3.79	3.98	$f(r_z^{BT}, 17.5)$	2.46	0.77
$f(e(a^{S3}),0.25)$	0.07	0.05	$m(v(u_z^{3D}, 0.75))$	523	461	$f(r_z^{BT}, 10)$	4.23	1.42
$f(e(a^{S3}),0.15)$	0.16	0.12	$m(v(u_z^{3D}, 0.9))$	457	415	$f(r_z^{BT}, 7.5)$	5.94	2.06
$m(v(e(a^{S3}),0.9))$	4.85	3.55	$\zeta(u_z^{3D})$	851	769	$f(r_z^{BT}, 5.0)$	10.5	3.82

Table 9. Top five of the important variables for predicting the arousal in the personalized model.

Pulse Wave Sensor			Eye Tracker			Motion Detector		
Variable	SE	AR	Variable	SE	AR	Variable	SE	AR
$c(v(p^{RR},0.5))$	5	3.64	$f(g_x^{HD}, 15)$	1040	949	$c(v(r_y^{LA}, 0.75))$	69.5	51.7
$t$	47.3	43	$T$	47.3	43	$c(v(r_y^{LA}, 0.5))$	140	113
$m(v(p^{HR},0.5))$	892	945	$u^{BK}$	327	182	$c(v(r_x^{LA}, 0.5))$	123	101
$m(v(p^{HR},0.9))$	902	966	$f(a_z^{HD}, 50)$	406	208	$c(v(r_x^{LA}, 0.25))$	204	180
$\sigma(p^{HR})$	410	512	$f(a_z^{HD}, 35)$	683	384	$m(v(r_x^{LA}, 0.9))$	-38.4	-28.2

Table 10. Top five of the important variables for predicting the engagement in the personalized model.

Pulse Wave Sensor			Eye Tracker			Motion Detector		
Variable	BR	CR	Variable	BR	CR	Variable	BR	CR
$\sigma(v(p^{LF},0.9))$	3695	2711	$\sigma(v(a_y^{HD}, 0.25))$	2.39	2.08	$m(v(r_y^{SL}, 0.9))$	0.49	0.2
$m(v(p^{LF},0.9))$	2325	1484	$m(a_z^{HD})$	5.80	5.27	$\sigma(v(r_x^{SL}, 0.25))$	4.34	3.87
$\sigma(v(p^{LF},0.75))$	3570	2633	$\sigma(v(a_y^{HD}, 0.5))$	1.94	1.73	$m(v(r_y^{SL}, 0.25))$	0.36	0.14
$m(v(p^{LF},0.25))$	2620	1708	$m(v(a_y^{HD}, 0.5))$	-6.58	-7.08	$\sigma(r^{S3})$	0.29	0.09
$m(v(p^{LF},0.75))$	2315	1454	$m(v(a_y^{HD}, 0.25))$	-7.26	-7.66	$m(v(r_x^{S2}, 0.25))$	1.57	3.39

## 5. Discussion

Deep neural networks have the power and ability of automatic feature extraction. However, from Figures 6–9, automatic feature extraction does not work well, and the prediction accuracy with variable selection was better than the prediction accuracy without variable selection. The reason for this is that the amount of information included in the biological signal and behavior is less than the image showing the effectiveness in previous research [44]. Therefore, when inputting the time series features of biological signals and behavior to deep neural networks, it is necessary to select a variable.

In this research, we measured the level of pleasant emotion and arousal in 7-point scales and the level of engagement in 5-point scales. The spatial resolution of the questionnaire is 1. When the level of pleasant emotion and arousal differs by one-point, the mean squared error normalized is 0.14. When the level of engagement differs by one-point, the mean squared error normalized is 0.2. On the other hand, we showed that pleasant emotion, arousal, and engagement during order picking can be predicted from the behavior of the worker with an accuracy of error rate 0.12 or less. Considering these facts, this research can predict pleasant emotion, arousal, and engagement with high accuracy. In addition, the personalized model has sufficient accuracy to classify pleasant and unpleasant, arousal and sleepy, concentration and boredom.

From Figure 5, we predict pleasant emotion, arousal, and engagement as a secondary reaction of the flow experience. From Tables 5–10 the worker in flow no longer experiences wasteful movements not required for work such as left arm and head. This fact shows that the flow state leads the worker to minimize the energy consumption of the body. As a result, worker enhances pleasure emotion, arousal, and engagement. Therefore, we predicted pleasant emotion, arousal, and engagement indirectly by time series features (power spectrum value and peak valley value) representing waste of behavior. On the other hand, personalized models are generally more accurate than generalized models. However, when time series features of the pulse wave sensor were input to deep neural networks, the accuracy of the personalized model was significantly lowered in the four subjects (C, F, K, and M) than in the generalized model. The reason for this is related to the characteristics of the pulse wave and the specificity of the subjects C, F, K and M. It is known that patterns of different emotions and biomedical signals appear in related studies even under the same stimulation depending on the situation. The related research pointed out that patterns of different vital signs appear depending on the situation even with the same stimulus. In addition, the time series features of the pulse wave appear to be complicated patterns due to the influence of exercise and the activity of the autonomic nervous system. Machine learning yields high accuracy with respect to objective variables and explanatory variable patterns present in the training data set, but accuracy for other patterns becomes unstable. In general, machine learning predicts highly accurately in patterns of target variables and explanatory variables present in the training set, but prediction accuracy in other patterns is unstable. Therefore, when the variation of emotion is large and the sample size is small, the number of patterns not existing in training data set increases. As a result, prediction accuracy decreases remarkably. Subjects C, F, K, and M have specificity in standard deviation of pleasant emotion, standard deviation of work efficiency, and average of flow state score (Figure 13). A small standard deviation of work efficiency shows that the worker has a high skill. C, F, and K have larger standard deviation of pleasant emotion compared to other subjects, and the skill of workers is high. M has larger standard deviation of pleasant emotion compared to other subjects, and the skill is low. In addition, since C, F, and M had a high average of flow state score, the standard deviation of pleasant emotion was caused by a flow experience. On the other hand, since the average of flow state score of K was remarkably low, the standard deviation of pleasant emotion was caused by factors other than the flow experience. The results in Figure 10 indicate that it is not possible to create a good predictive model for pleasant emotion by machine learning due to the small sample size when time series features of pulse waves of subjects with specificity were used. On the other hand, when pleasant emotion was predicted using the time series feature of behavior, the accuracy is stable and high. Therefore, it can be said that the sample size is sufficient when behavior is used as an index. From this, it is shown that if we take short

emotions of work, we can predict emotion, arousal, and immersion stably by obtaining behavior from vital signs. In the case of using the time series feature of the pulse wave, as generalized model, the generalized model is more accurate, so let us learn data of similar users not only between individuals by multi task learning. There is a high possibility that the accuracy can be improved. There was a difference between a person who is likely to increase the pulse by exercise and a person who is difficult to rise, and the accuracy became unstable. On the other hand, as Table 5 shows, there is a possibility that a remarkable decrease in precision due to characteristics of an individual hard to appear in a pulse wave can be solved by considering other person's characteristics using multi task learning.

## 6. Conclusions and Future Work

The contributions of this research are as follows: First, we show that the organizations and supervisors can generate preferable flow experiences by giving a clear target and managing task complexity. In addition, pleasant emotions gave vitality to workers and decrease human error. Second, we showed that deep neural networks accurately predicted pleasant emotion, arousal, and engagement during order picking. The accuracy of the prediction model was the highest when predicting flow experience from the time series features of behavior.

In this research, the workers wore multiple devices. However, to put the monitoring system into practical use, it is necessary to decrease the number of devices because the multiple devices interrupt the order picking task and increase the cost of the monitoring system. One possible way to solve this dilemma is to install the sensors on the belt so that the worker can utilize the monitoring system without additional load.

The prediction accuracy of the generalized model was smaller than that of the personalized model. The prediction accuracy of the personalized model was the highest when predicting flow experience from the time series features of behavior. Since the important variables were the same for the generalized model and the personalized model, it was suggested that the picking speed, the walking speed, and the step are different for each individual. Therefore, by taking differences among individuals into consideration and conducting multi task learning, improvement of accuracy of the generalized model can be expected. In the future, we plan to devise measures to address these individual differences using multi-task learning. Since the psychological model constructed in this research is a short-term order picking, we will investigate the effect of long-term order picking.

**Author Contributions:** Conceptualization, Y.K., H.K.; methodology, Y.K.; software, Y.K.; validation, Y.K.; formal analysis, Y.K.; investigation, Y.K. and T.S.; resources, Y.K.; data curation, Y.K.; writing—original draft preparation, Y.K.; writing—review and editing, T.S.; visualization, T.S.; supervision, H.K.; project administration, Y.K.; funding acquisition, Y.K.

**Funding:** The APC was funded by KONICA MINOLTA, INC.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

The following appendix indicate the validity of the hypothesis of the psychological model. To clarify the effect of the target on pleasant emotion, arousal, engagement, subjective task difficulty, and flow experience after work, we compared those of the first period with those of the second period.

We instructed a clear target to the worker only in the second period. The Brunner-Munzel test was conducted to show a significant difference between the values of the two groups of the first and second periods. The rejection region was set to 0.05. The sample size of pleasant emotion, arousal, and engagement were 544, the sample size of flow state scale was 34 (the number of periods  $\times$  the number of subjects =  $2 \times 17$ ). Table A1 shows the average and test results of the first period and the second period. Bold letters in the table show the items with a *p*-value of 0.05 or less.

**Table A1.** Effect of the target on pleasant emotion, Arousal, subjective task difficulty, engagement, human error, work efficiency, and flow experience.

Symbol	Average			Brunner Munzel Test	
	First Period	Second Period	Total	Statistics	<i>p</i> -Value
Pleasant	0.69	0.91	0.8	1.67	0.1
Arousal	<b>1.11</b>	<b>1.36</b>	<b>1.24</b>	<b>2.03</b>	<b>0.04</b>
Subjective task difficulty	−0.33	−0.24	−0.28	0.91	0.36
Engagement	<b>3.43</b>	<b>3.64</b>	<b>3.54</b>	<b>2.73</b>	<b>0.01</b>
Human error	0.55	0.43	0.49	−1.46	0.15
Work efficiency	<b>0.23</b>	<b>0.32</b>	<b>0.27</b>	<b>−6.26</b>	<b>0</b>
$q^{(1)}_{FE}$	<b>3.2</b>	<b>3.8</b>	<b>3.5</b>	<b>2.1</b>	<b>0.05</b>
$q^{(2)}_{FE}$	3.4	4.1	3.8	1.9	0.07
$q^{(3)}_{FE}$	<b>3.2</b>	<b>4.1</b>	<b>3.7</b>	<b>3</b>	<b>0.01</b>
$q^{(4)}_{FE}$	<b>2.8</b>	<b>3.6</b>	<b>3.2</b>	<b>2.5</b>	<b>0.02</b>
$q^{(5)}_{FE}$	3.8	3.9	3.9	0.2	0.85
$q^{(6)}_{FE}$	3.4	3.6	3.5	0.86	0.4
$q^{(7)}_{FE}$	2.5	2.6	2.6	0.51	0.61
$q^{(8)}_{FE}$	2.9	3.4	3.2	1.2	0.26
$q^{(9)}_{FE}$	3.3	3.8	3.6	1.1	0.26

Work efficiency in Table A1 is calculated using Equation (1). The test results in Table A1 suggested that, when the target was given: (1) arousal, engagement and work efficiency were improved and (2) fulfillment was improved. From Table A1, when a clear target was given, the worker increased  $q^{(1)}_{FE}$ ,  $q^{(3)}_{FE}$ , and  $q^{(4)}_{FE}$ . The average score of the items of flow state scale is 3.5. It was suggested that most workers were in flow experience.

To clarify the effect of feedback on pleasant emotion, arousal, engagement, subjective task difficulty, and flow experience, Brunner-Munzel test was conducted to show a significant difference between the values of the two groups, with one group receiving feedback and the other not.

The rejection region was set to 0.05. The sample size of pleasant emotion, arousal, and engagement were 544, the sample size of flow state scale was 34 (the number of periods  $\times$  the number of subjects =  $2 \times 17$ ). Table A2 shows the effect of feedback on average and test results. The test results in Table A2 show that (1) the worker improved engagement when feedback was given, and (2) the difference between the two groups was not statistically significant.

**Table A2.** Effect of feedback on pleasant emotion, Arousal, subjective task difficulty, engagement, human error, work efficiency, and flow experience.

Symbol	Average		Brunner Munzel Test	
	Feedback	Natural	Statistics	<i>p</i> -Value
Pleasant	0.91	1	0.84	0.4
Arousal	1.4	1.2	−1.4	0.17
Subjective task difficulty	2.8	2.7	−0.79	0.43
Engagement	<b>3.6</b>	<b>3.4</b>	<b>−2.5</b>	<b>0.01</b>
Human error	0.46	0.52	0.66	0.51
Work efficiency	0.28	0.27	0.92	0.36
$q^{(1)}_{FE}$	3.5	3.5	0.28	0.78
$q^{(2)}_{FE}$	4	3.5	−1.6	0.13
$q^{(3)}_{FE}$	4	3.4	−1.8	0.09
$q^{(4)}_{FE}$	3.5	2.9	−1.7	0.09
$q^{(5)}_{FE}$	4.1	3.7	−0.52	0.61
$q^{(6)}_{FE}$	3.5	3.5	0.14	0.89
$q^{(7)}_{FE}$	2.6	2.5	−0.54	0.59
$q^{(8)}_{FE}$	3.2	3.1	−0.61	0.55
$q^{(9)}_{FE}$	3.7	3.4	−1.02	0.31

To clarify the effect of pleasant emotion, arousal and engagement on human error, the Brunner-Munzel test was conducted to show a significant difference between the values of the two groups, with one group human error occurred and the other not. The rejection region was set to 0.05. The sample size of pleasant emotion, arousal, and engagement were 544. Table A3 shows the effect of the average and test results of pleasant emotion, arousal, and engagement on human error. The test results in Table A3 show that the workers with negative emotions committed more human errors.

**Table A3.** Effect of pleasant emotion, arousal, and engagement on human error.

Item	Average		Brunner Munzel Test	
	Error	Success	Statistics	p-Value
Pleasant	0.53	0.91	−2.5	0.01
Arousal	1.1	1.3	−1.2	0.25
Engagement	3.1	2.6	0.26	0.8

Factor analysis shows the effect of latent variables on observed variables. The observed variables were pleasant emotion, arousal, engagement, subjective task difficulty, human error, and work efficiency. First, the eigenvalues of the correlation matrix were acquired. The number of latent variables was determined from the change in slope of eigenvalues. The change in slope of eigenvalues illustrated was at a maximum value when the number of latent variables was three. Therefore, the number of latent variables was set to 2. Since latent variables were weakly correlated with 0.21 correlation, we conduct the factor analysis of Promax rotation. Table A4 show the factor analysis. Since one of the latent variables affects positively to pleasant emotion, arousal, and engagement, the latent variable represents flow experience. Another latent variable represents task complexity because the latent variable affects positively to subjective task difficulty, work efficiency, and human error.

**Table A4.** The result of factor analysis.

	Pleasant	Arousal	Subjective Task Difficulty	Engagement	Work Efficiency	Human Error
Flow experience	0.47	0.7	-	0.54	-	-
Complexity of the task	−0.16	-	0.78	0.12	0.46	0.29
Uniquenesses	0.71	0.53	0.42	0.73	0.9	0.78

## References

- Battini, D.; Faccio, M.; Persona, A.; Sgarbossa, F. Design of the optimal feeding policy in an assembly system. *Int. J. Prod. Econ.* **2009**, *121*, 233–254. [\[CrossRef\]](#)
- Tompkins, J.A.; White, Y.A.; Bozer, E.H.; Tanchoco, J.M.A. *Facilities Planning*, 4th ed.; Wiley: Hoboken, NJ, USA, 2010.
- Thayer, R.E. *The Biopsychology of Mood and Arousal*; Oxford University Press: Oxford, UK, 1990.
- Bakker, A.B.; Schaufeli, W.B.; Leiter, M.P.; Taris, T.W. Work engagement: An emerging concept in occupational health psychology. *Work Stress* **2008**, *22*, 187–200. [\[CrossRef\]](#)
- Sharot, T. The optimism bias. *Curr. Biol.* **2011**, *21*, R941–R945. [\[CrossRef\]](#) [\[PubMed\]](#)
- Steptoe, A.; Wardle, J.; Marmot, M. Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6508–6512. [\[CrossRef\]](#) [\[PubMed\]](#)
- Woodruffe, C. The crucial importance of employee engagement. *Hum. Resour. Manag. Int. Dig.* **2006**, *14*, 3–5. [\[CrossRef\]](#)
- Schaufeli, W.B.; Bakker, A.B.; van der Heijden, F.M.; Prins, J.T. Workaholism among medical residents: It is the combination of working excessively and compulsively that counts. *Int. J. Stress Manag.* **2009**, *16*, 249. [\[CrossRef\]](#)
- White, B. *Employee Engagement Report*; Blessing White: Princeton, NJ, USA, 2011; Retrieved.
- Alshallah, S. Job satisfaction and motivation: How do we inspire employees? *Radiol. Manag.* **2004**, *26*, 47–51.

11. Caplan, S.; Whittemore, R. Barriers to treatment engagement for depression among Latinas. *Issues Ment. Health Nurs.* **2013**, *34*, 412–424. [[CrossRef](#)]
12. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [[CrossRef](#)]
13. Jackson, S.A.; Marsh, H.W. Development and validation of a scale to measure optimal experience: The Flow State Scale. *J. Sport Exerc. Psychol.* **1996**, *18*, 17–35. [[CrossRef](#)]
14. Likert, R. A Technique for the Measurement of Attitudes. *Arch. Psychol.* **1932**, *140*, 55.
15. Watson, D.; Clark, L.A.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* **1988**, *54*, 1063–1070. [[CrossRef](#)]
16. Mahnke, R.; Benlian, A.; Hess, T. Flow Experience in Information Systems Research: Revisiting its Conceptualization, Conditions, and Effects. *ICIS* **2014**. [[CrossRef](#)]
17. Punj, G. Consumer decision making on the web: A theoretical analysis and research guidelines. *Psychol. Mark.* **2012**, *29*, 791–803. [[CrossRef](#)]
18. Petrou, P.; Demerouti, E.; Peeters, M.C.W.; Schaufeli, W.B.; Hetland, J. Crafting a job on a daily basis: Contextual correlates and the link to work engagement. *J. Organ. Behav.* **2012**, *33*, 1120–1141. [[CrossRef](#)]
19. Spector, P.E.; Cooper, C.L.; Sanchez, J.I.; O’Driscoll, M.; Sparks, K.; Bernin, P.; Büssing, A.; Dewe, P.; Hart, P.; Lu, L.; et al. Locus of control and well-being at work: How generalizable are western findings? *Acad. Manag. J.* **2002**, *45*, 453–466. [[CrossRef](#)]
20. Reis, H.T.; Sheldon, K.M.; Gable, S.L.; Roscoe, J.; Ryan, R.M. Daily well-being: The role of autonomy, competence, and relatedness. *Personal. Soc. Psychol. Bull.* **2000**, *26*, 419–435. [[CrossRef](#)]
21. Fullagar, C.J.; Kelloway, E.K. Flow at work: An experience sampling approach. *J. Occup. Organ. Psychol.* **2009**, *82*, 595–615. [[CrossRef](#)]
22. Vayrynen, E.; Kortelainen, J.; Seppanen, T. Classifier-based learning of nonlinear feature manifold for visualization of emotional speech prosody. *IEEE Trans. Affect. Comput.* **2013**, *4*, 47–56. [[CrossRef](#)]
23. Zeng, Z.; Fu, Y.; Roisman, G.I.; Wen, Z.; Hu, Y.; Huang, T.S. Spontaneous emotional facial expression detection. *J. Multimedia* **2006**, *1*, 1–8. [[CrossRef](#)]
24. Poh, M.Z.; Swenson, N.C.; Picard, R.W. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1243–1252. [[PubMed](#)]
25. Kreibig, S.D. Autonomic nervous system activity in emotion: A review. *Biol. Psychol.* **2010**, *84*, 394–421. [[CrossRef](#)] [[PubMed](#)]
26. Choi, J.; Ahmed, B.; Gutierrez-Osuna, R. Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 279–286. [[CrossRef](#)] [[PubMed](#)]
27. Cannon, W.B. *Bodily Changes in Pain, Hunger, Fear and Rage*, 2nd ed.; D. Appleton & Company: New York, NY, USA, 1929.
28. Van Den Broek, E.L.; Lisý, V.; Janssen, J.H.; Westerink, J.H.; Schut, M.H.; Tuinenbreijer, K. Affective man-machine interface: Unveiling human emotions through biosignals. In *International Joint Conference on Biomedical Engineering Systems and Technologies*; Springer: Berlin/Heidelberg, Germany, January 2009; pp. 21–47.
29. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* **2012**, *3*, 42–55. [[CrossRef](#)]
30. Schmidt, P.; Reiss, A.; Duerichen, R.; Van Laerhoven, K. Wearable affect and stress recognition: A review. *arXiv* **2018**, arXiv:1811.08854.
31. Kajiwara, Y.; Yonekura, S.; Kimura, H. Prediction of Future Mood Using Majority Vote Based on Certainty Factor. *Sens. Mater.* **2018**, *30*, 1473–1486. [[CrossRef](#)]
32. Oviatt, S. Human-centered design meets cognitive load theory: Designing interfaces that help people think. In *Proceedings of the 14th ACM international conference on Multimedia*; ACM: New York, NY, USA, 2006; pp. 871–880.
33. Uemura, Y.; Kajiwara, Y.; Shimakawa, H. Figuring out Distraction Degree from Working Memory Consumption for Pedestrian Safety. *Int. J. Internet Things* **2017**, *6*, 1–8.
34. Won, B.Y.; Lee, H.J.; Jiang, Y.V. Statistical learning modulates the direction of the first head movement in a large-scale search task. *Atten. Percept. Psychophys.* **2015**, *77*, 2229–2239. [[CrossRef](#)]
35. Kang, O.; Wheatley, T. Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *J. Exp. Psychol. Gen.* **2017**, *146*, 569. [[CrossRef](#)]

36. Kalogeropoulou, Z.; Rolfs, M. Saccadic eye movements do not disrupt the deployment of feature-based attention. *J. Vis.* **2017**, *17*, 4. [[CrossRef](#)]
37. Nikolin, S.; Boonstra, T.W.; Loo, C.K.; Martin, D. Combined effect of prefrontal transcranial direct current stimulation and a working memory task on heart rate variability. *PLoS ONE* **2017**, *12*, e0181833. [[CrossRef](#)] [[PubMed](#)]
38. McDuff, D.; Gontarek, S.; Picard, R. Remote measurement of cognitive stress via heart rate variability. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2014**, 2957–2960. [[CrossRef](#)]
39. Baddeley, A. Working memory. *Science* **1992**, *255*, 556–559. [[CrossRef](#)] [[PubMed](#)]
40. Ferreira, E.; Ferreira, D.; Kim, S.; Siirtola, P.; Röning, J.; Forlizzi, J.F.; Dey, A.K. Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), Orlando, FL, USA, 9–12 December 2014; pp. 39–48.
41. Solovey, E.T.; Zec, M.; Garcia Perez, E.A.; Reimer, B.; Mehler, B. Classifying driver workload using physiological and driving performance data: Two field studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2014; pp. 4057–4066.
42. Bengio, Y. Learning deep architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–127. [[CrossRef](#)]
43. Csikszentmihalyi, M. *Flow and the Psychology of Discovery and Invention*; HarperPerennial: New York, NY, USA, 1997; p. 39.
44. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).