S & M 1599

Prediction of Oxygen Saturation by Pulse Oximetory from Image and Sound Data with Long Short-term Memory Recurrent Neural Network

Takehiro Kasahara,[1,2*] Yuji Yonezawa,[2] Yoshihiro Ueda,[2] Masatoshi Saito,[3]
Koji Kojima,[3] Yuki Fujimoto,[3] Hirohisa Toga,[3] and Hidetaka Nambo[1]

[1]Kanazawa University, Kakuma, Kanazawa, Ishikawa 920-1192, Japan
[2]Industrial Research Institute of Ishikawa, 2-1 Kuratsuki, Kanazawa, Ishikawa 920-8203, Japan
[3]Kanazawa Medical University, 1-1 Daigaku, Uchinada, Kahoku, Ishikawa 920-0293, Japan

(Received December 30, 2017; accepted March 14, 2018)

Keywords: camera, mic, IoT sensors, sleep apnea, SpO_2, prediction, LSTM, recurrent neural network

An Internet of Things (IoT) communication function was attached to inexpensive sensors such as cameras and microphones and was used for data acquisition and analysis. In this study, the value of saturation of blood oxygen measured by pulse oximetry (SpO_2), which is used for sleep apnea syndrome (SAS) detection, was estimated from the data obtained from the camera and microphone. SpO_2 was recorded by a pulse oximeter worn by subjects suspected of having SAS when sleeping overnight. The camera and microphone were located on the side of the bed to record the data. The SpO_2 value was learned using long short-term memory (LSTM), which is one of the deep neural network methods that have shown excellent results as a method of analyzing time series data. When evaluated by leave-one-out cross validation using the data of four persons, it was found that the amplitude of the estimated SpO_2 was about half. The cause seems to be individual differences in time from apnea occurrence to SpO_2 declination. By doubling the estimation result, it was confirmed that the SpO_2 value was well estimated.

1. Introduction

Applications of inexpensive sensors such as cameras and microphones with Internet of Things (IoT) are expanding. Such sensors are suitable for simple and inexpensive data collection and analysis. In this study, the data obtained from these sensors were estimated using long short-term memory (LSTM) for deep learning, which is suitable for handling time series data.

Sleep apnea syndrome (SAS) induces drowsiness during the day, and it causes not only traffic accidents but also adverse effects on hypertension and heart disease.(1) Some detection methods for sleep apnea using sound(2,3) and others using images(4,5) have been reported in the literature. We have also proposed a method that analyzes data obtained from a camera and a microphone and integrates the results.(6) In addition, it has been reported that a strong

[*] Corresponding author: e-mail: kasaharat@irii.jp

http://dx.doi.org/10.18494/SAM.2018.1883

ISSN 0914-4935 © MYU K.K.
http://myukk.org/

correlation has been found between the frequency of blood oxygen saturation decline and SAS,[7] and therefore, a method for diagnosing SAS from the value of SpO_2 has also been established.

In the field of recognition and analysis, deep learning has produced excellent results, and attempts to use it to analyze medical data have also been reported.[8] Among them, LSTM is suitable for analyzing time series data.[9] Therefore, we proposed in this study an analytical method using LSTM.

In this work, we estimated SpO_2 using LSTM and contactless sensors such as cameras and microphones instead of contact sensors such as a pulse oximeter and an electroencephalography (EEG). Sensor installation is expected to be more flexible if SpO_2 can be estimated with a camera or a microphone often used as an IoT sensor. Such an installation will also contribute to making a diagnostic system more convenient for the user.

2. Materials and Methods

In data collection using IoT, it is important to keep the amount of data associated with communication as small as possible. In addition, because processing units attached to sensors are often selected to have limited functions in order to reduce costs, it is also important to slightly suppress the need for processing on the device side. For these reasons, we adopted a low data traffic volume and light processing on the acquired data as a policy.

2.1 Image data processing

Video data were recorded in MJPEG format at a resolution of 752 × 480 pixels and an average frame rate of 5 frames per second (fps). Considering that the fastest physiologically important body movements associated with SAS during sleep are respiratory movements and that the respiratory rate for adults at rest is on average 12–18 breaths per minute, the frame rate was reduced to 1 fps. The overall motion signal was derived by calculating the sum of the absolute pixel difference between successive frames and converted to a log scale (Fig. 1). Prior to input into LSTM, the data were normalized to average 0, variance 1. An example of video data is shown in Fig. 2(a).

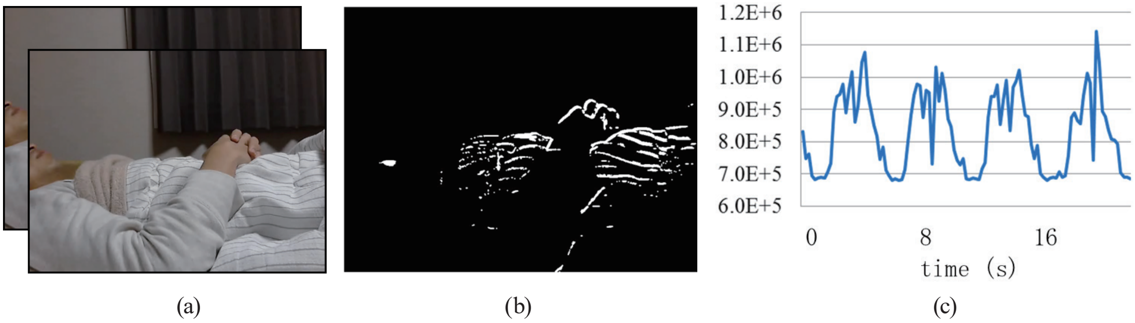
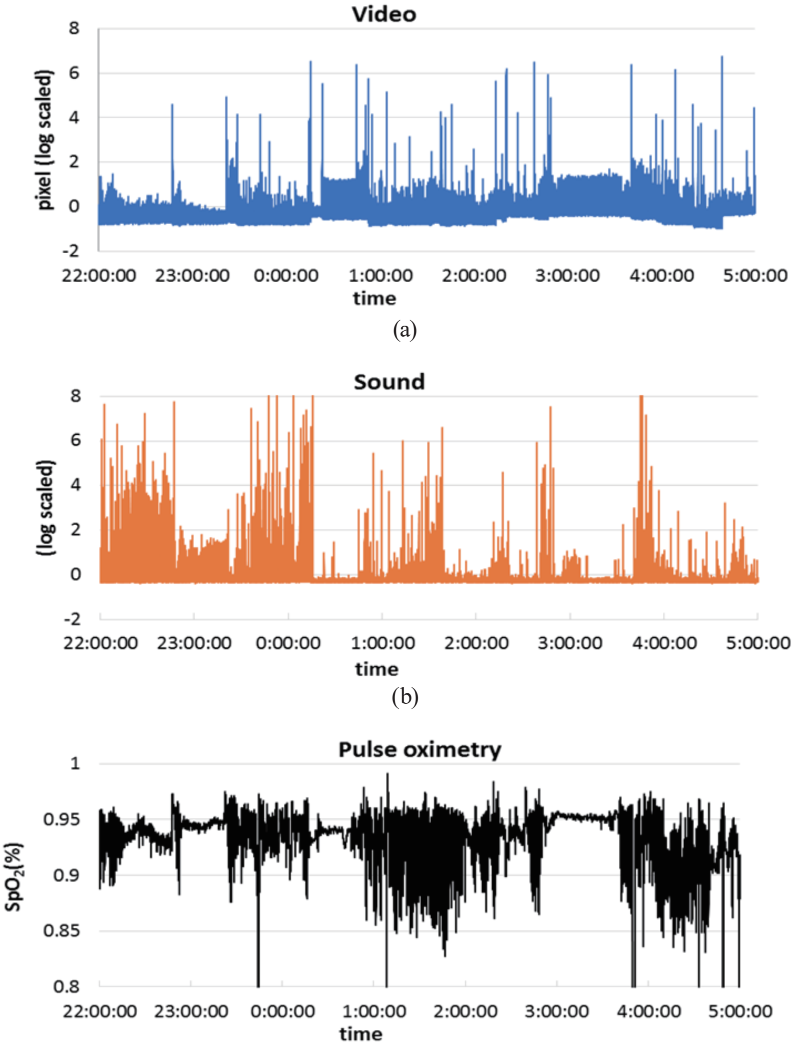


Fig. 1. (Color online) (a) Two successive frames, (b) absolute pixel difference, and (c) conversion to log scale. Plot to time series.



(a)

(b)

(c)

Fig. 2. (Color online) (a) Video data from one night, (b) sound data from one night, and (c) pulse oximetry data from one night.

2.2 Sound data processing

Sound data were recorded in WAV format. The sound data were applied to a band pass filter at 150–5000 Hz and the maximum absolute value per second was used as sound input data. Prior to input into LSTM, the data were normalized to average 0, variance 1. An example of a sound signal is shown in Fig. 2(b).

2.3 Pulse oximetry sensor data processing

The SpO_2 data obtained by a pulse oximeter was used as supervisor data. A pulse oximeter outputs one SpO_2 value every second. It is reported that the value of SpO_2 has a deep correlation with the symptoms of SAS,(7) and simple SAS detection is carried out using this value. If it is possible to predict the value of SpO_2 from sound or video data, it could be substituted as a simple SAS detector without using a pulse oximeter. Before using the SpO_2 data as the LSTM supervisor, the data were normalized to mode 0. An example of pulse oximetry data is shown in Fig. 2(c).

2.4 Estimation using LSTM

LSTM is proposed as a deep neural network that is effective for the analysis of time series data. The number of hidden layer units and learning rate can be changed as parameters. There are also reports that an LSTM stacked vertically is effective and that a bidirectional LSTM is also effective. Two values calculated from two sensors are used as input values and the value of SpO_2 obtained from a pulse oximeter is used as the supervisor value for learning. The result of the estimation is an output value from the output layer of the LSTM. Learning by the LSTM is performed by optimizing the weights of the network to minimize the error that is the difference between the value of SpO_2 and the result of estimation at each time.

2.5 Data acquisition

Data from four subjects who were suspected of having SAS and then diagnosed as positive for it as the result of overnight polysomnography at Kanazawa Medical University were used. The subjects gave informed consent for their data to be acquired. The sound was collected using a microphone (Sony ECM-360), and video was acquired using a camera (IDS UI-1220LE-M-GL) and a lens (TAMROM 12 VM 412 ASIR), and SpO_2 data was acquired by pulse oximetry (Konica Minolta PULSOX-300i).

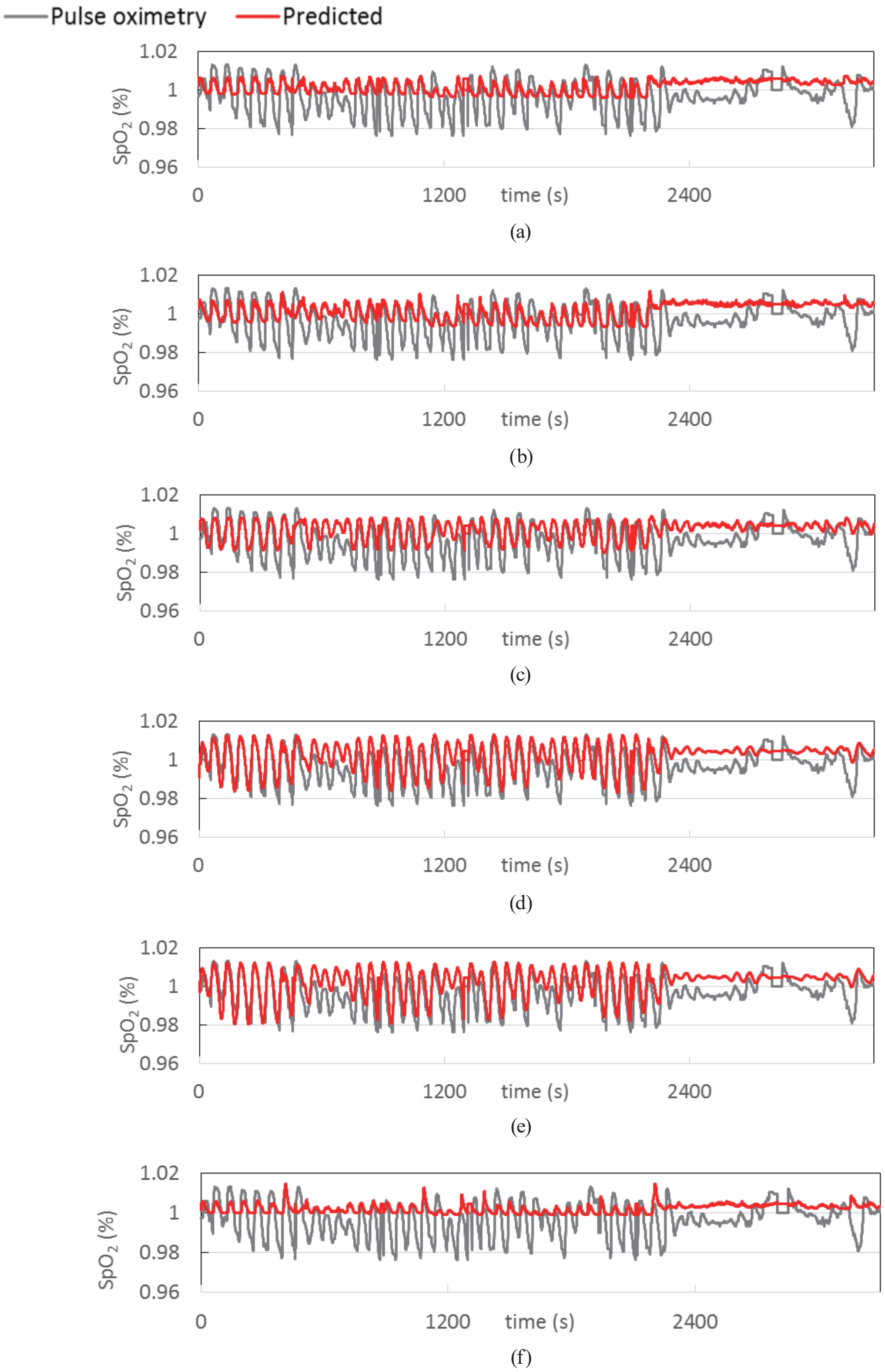
2.6 Evaluation

To evaluate the estimation method, a two-step assessment was performed. First, the first 5 h of overnight data for one person was used for training data and the last 1 h was used as testing data. In this evaluation, to find the optimal number of hidden layer units, the number of hidden layers was changed from 10 to 120.

Next, four overnight data sets were evaluated using the leave-one-out (LOO) cross validation method. In short, three of four data sets were used as training data and one remaining data set was used as testing data.

3. Results

As a result of the first evaluation, the best result was obtained when the number of hidden layer units was 80 [Fig. 3(d)]. In the selection of the learning rate, the best result was obtained at a learning rate of 0.1 [Fig. 3(j)]. In the results of the experiment using four persons' overnight data, SpO_2 could be estimated (Fig. 4). However, the amplitude of the estimated value became smaller than the amplitude of the estimation result using one person's data.



(a)

(b)

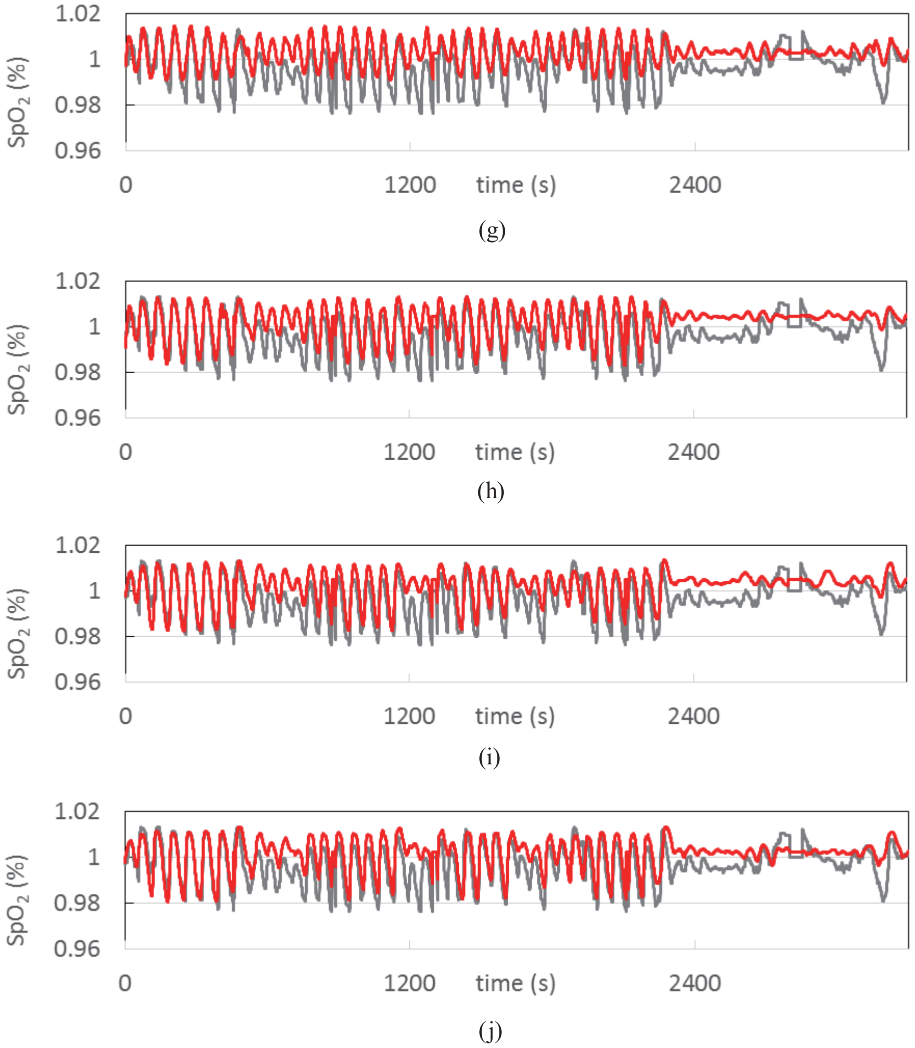
(c)

(d)

(e)

(f)

Fig. 3. (Color online) SpO_2 value from pulse oximetry and predicted from LSTM. (a) Learning rate = 0.01, H unit = 3, E = 0.01332, (b) learning rate = 0.01, H unit = 5, E = 0.01265, (c) learning rate = 0.01, H unit = 10, E = 0.01277, (d) learning rate = 0.01, H unit = 80, E = 0.01245, (e) learning rate = 0.01, H unit = 120, E = 0.01249, and (f) learning rate = 0.001, H unit = 80, E = 0.01353.



(g)

(h)

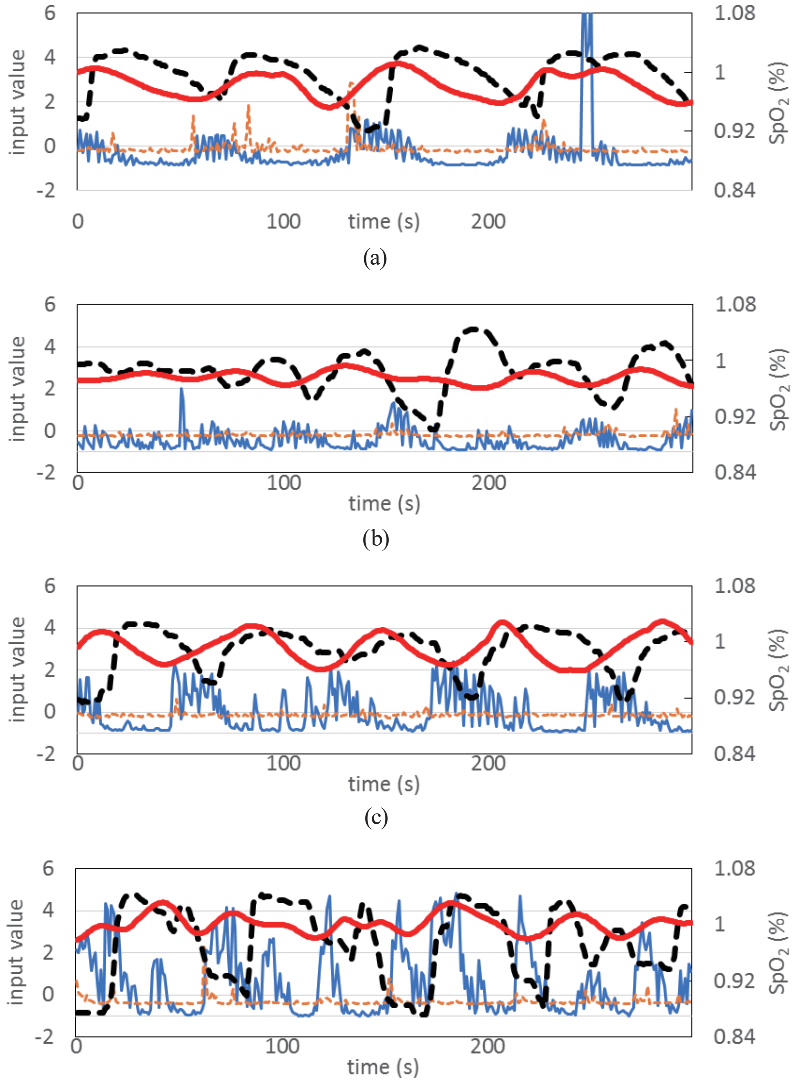
(i)

(j)

Fig. 3. (continued) (g) learning rate = 0.05, H unit = 80, E = 0.01292, (h) learning rate = 0.01, H unit = 80, E = 0.01245, (i) learning rate = 0.05, H unit = 80, E = 0.01238, and (j) learning rate = 0.1, H unit = 80, E = 0.01186. Figure 3(j) shows the best result, which has learning rate = 0.1 and H unit = 80.

4. Discussion

As a result of LOO evaluation using four persons' data, the amplitude of the estimated result became smaller than the result using one person's data. This difference causes an error in the estimation of SAS, which is done by counting the number of drops of 4.0 points in SpO_2. Table 1 shows the number of drops more than the threshold in the SpO_2 data and the estimated data. According to Table 1, if the threshold is set as 1.6 to 2.3 points, a similar number of times of a drop of more than 4.0 points in the SpO_2 data is obtained. Furthermore, by multiplying the estimated result by 2.0, it is shown that the count of drops can be estimated with a ratio of 96.7% (Data1: 190/197, Data2: 224/257, Data3: 201/188, Data4: 275/295).



(a)

(b)

(c)

(d)

Fig. 4. (Color online) Plot of video, sound, pulse oximetry, and predicted data. (a) Data set 1, (b) data set 2, (c) data set 3, and (d) data set 4.

Table 1
Number of drops more than the threshold in SpO_2 data and estimated data.

	SpO_2	Estimated												
Threshold	4.0	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	3.0	4.0
Data 1	197	230	227	218	207	196	190	179	171	161	152	148	113	68
Data 2	257	266	259	254	248	236	224	215	203	192	181	165	74	4
Data 3	188	219	215	209	204	203	201	197	194	190	185	178	152	65
Data 4	295	319	309	300	290	283	275	263	254	241	231	219	166	97

Consider the cause that the amplitude of the estimated result becomes small by learning the data of multiple people. When an apnea occurs, the SpO_2 value begins to fall, and when the respiration is restored, the SpO_2 value returns to its original value. For a person with long

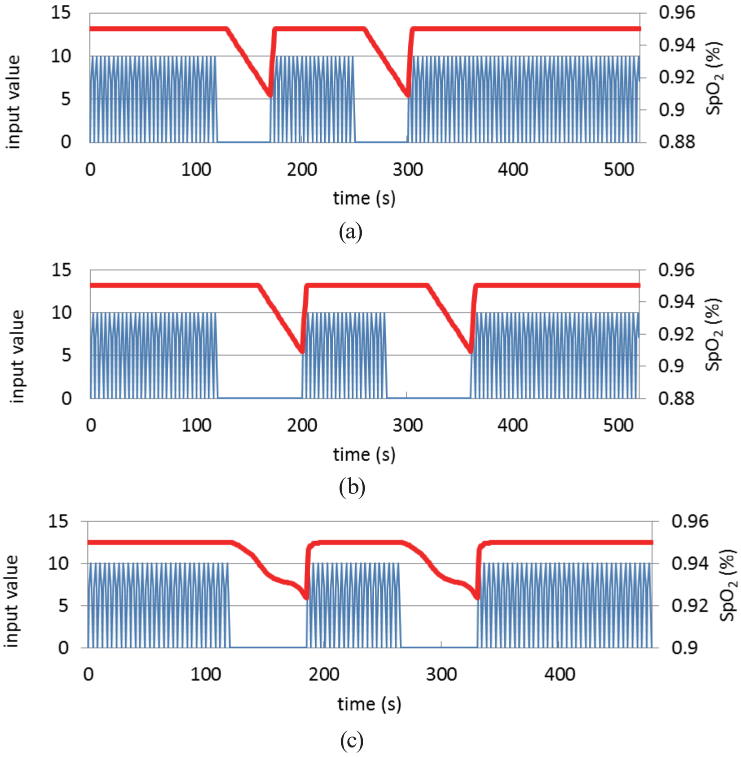


Fig. 5. (Color online) Simulated body movement signal and SpO_2. (a) SpO_2 drops with short delay. (b) SpO_2 drops with long delay. (c) Predicted SpO_2 learned from (a) and (b), which have small amplitude.

arms or a low heart rate, it takes a long time for the blood of the lungs to reach the fingertips. Therefore, the time required to return the SpO_2 value to the original value after respiration is restored is greater than that of other people and is considered to be long. Therefore, we created and evaluated two data assuming that the arm lengths are different. As a result, even if the falling amplitude of SpO_2 is the same, if we learn two data [Figs. 5(a) and 5(b)] with different timings when declines occur, the amplitude of the estimation result becomes small [Fig. 5(c)]. From this result, it is considered that if LSTM learning is performed using data of multiple examinees having different arm lengths and heart rate trends as it is, the amplitude of the estimation result will be small.

5. Conclusions

We proposed a method to estimate the blood oxygen saturation value using inexpensive sensors such as microphones and cameras often used as IoT sensors. We calculated one value per second from the video data obtained by recording body movements. We also calculated one value per second from the sound data recording the breathing sound. Also, one SpO_2 value was recorded per second. It was shown that the SpO_2 value can be obtained by regression by learning by LSTM using the two values of body motion and respiratory sound obtained in this way as the input and the SpO_2 value as the output. In this study, we evaluated four subjects, but we want to increase the reliability of the results by increasing the number of subjects. In addition, in this study, we aimed to realize the method by simple processing, but in order to improve the accuracy, it is effective to analyze the frequency of the respiration sound and to use a method to estimate respiratory motion from body motion data.

Acknowledgments

The authors would like to thank Cosmo Summit Co., Ltd. for providing the data, analysis software, and equipment used in this study.

References

1 K. Kario: Hypertens. Res. **32** (2009) 428.
2 A. Azarbarzina and Z. Moussavi: Med. Eng. Phys. **35** (2013) 479.
3 T. Emoto, U. R. Abeyratne, T. Kusumoto, M. Akutagawa, E. Kondo, I. Kawada, T. Azuma, S. Konaka, and Y. Kinouchi: Trans. J. Soc. Med. Biol. Eng. **48** (2010) 115 (in Japanese).
4 E. Gederi and G. D. Clifford: Biomed. Health Inf., 2012 IEEE-EMBS Int. Conf. (2012) 890.
5 Y. Nishida, Y. Mori, H. Mizoguchi, and T. Sato: J. Rob. Soc. Jpn. **16** (1998) 274 (in Japanese).
6 T. Kasahara, K. Nomura, Y. Ueda, Y. Yonezawa, M. Saito, H. Toga, Y. Fujimoto, K. Kojima, H. Kimura, and H. Nambo: Proc. 11th Int. Conf. Management Science and Engineering Management (2017) 804.
7 J. R. Stradling and J. H. Crosby: Thorax **46** (1991) 85.
8 S. Biswal, J. Kulas, H. Sun, B. Goparaju, M. B. Westover, M. T. Bianchi, and J. Sun: arXiv preprint arXiv:1707.08262 (2017).
9 F. A. Gers, J. Schminhuber, and F. Cummins: Learning to Forget: 9th Int. Conf. Artificial Neural Networks **2** (1999) 850.