

# System for Collecting Biosignal Data from Multiple Patient Monitoring Systems

Dukyong Yoon, MD, PhD<sup>1,2</sup>, Sukhoon Lee, PhD<sup>3</sup>, Tae Young Kim, BE<sup>1</sup>, JeongGil Ko, PhD<sup>4</sup>,  
Wou Young Chung, MD<sup>5</sup>, Rae Woong Park, MD, PhD<sup>1,2</sup>

<sup>1</sup>Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea; <sup>2</sup>Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea; <sup>3</sup>Department of Software Convergence Engineering, Kunsan National University, Gunsan, Korea; <sup>4</sup>Department of Software and Computer Engineering, College of Information Technology, Ajou University, Suwon, Korea; <sup>5</sup>Department of Pulmonology and Critical Care Medicine, Ajou University Hospital, Suwon, Korea

**Objectives:** Biosignal data include important physiological information. For that reason, many devices and systems have been developed, but there has not been enough consideration of how to collect and integrate raw data from multiple systems. To overcome this limitation, we have developed a system for collecting and integrating biosignal data from two patient monitoring systems. **Methods:** We developed an interface to extract biosignal data from Nihon Kohden and Philips monitoring systems. The Nihon Kohden system has a central server for the temporary storage of raw waveform data, which can be requested using the HL7 protocol. However, the Philips system used in our hospital cannot save raw waveform data. Therefore, our system was connected to monitoring devices using the RS232 protocol. After collection, the data were transformed and stored in a unified format. **Results:** From September 2016 to August 2017, we collected approximately 117 patient-years of waveform data from 1,268 patients in 79 beds of five intensive care units. Because the two systems use the same data storage format, the application software could be run without compatibility issues. **Conclusions:** Our system collects biosignal data from different systems in a unified format. The data collected by the system can be used to develop algorithms or applications without the need to consider the source of the data.

**Keywords:** Biosignal, Database, Intensive Care Units, Electrocardiography, Photoplethysmography

**Submitted:** September 28, 2017

**Revised:** October 20, 2017

**Accepted:** October 20, 2017

## Corresponding Author

Dukyong Yoon, MD, PhD

Department of Biomedical Informatics, Ajou University School of Medicine, 206 Worldcup-ro, Yeongtong-gu, Suwon 16449, Korea.  
Tel: +82-31-219-4476, E-mail: d.yoon.ajou@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2017 The Korean Society of Medical Informatics

## 1. Introduction

A biosignal is any continually measured and monitored signal from the body, such as electrocardiograph (ECG), continually measured arterial blood pressure (ABP) or central venous pressure (CVP), and photoplethysmograph to measure peripheral oxygen saturation (SpO<sub>2</sub>). These data can provide important physiological data in real time with high resolution and are becoming increasingly important [1].

The data types described above can be used to predict or determine patient status. Because these data are mainly captured in intensive care units (ICUs), much research has focused on predicting patient outcomes, such as mortality within 24 hours, bleeding, urgent intubation, or sepsis [2,3]. Another type of study aimed to determine the serum potas-

sium level without blood sampling by analyzing single-lead ECG signals [4].

A team working at the Laboratory for Computational Physiology of the Massachusetts Institute of Technology has collected and constructed a biosignal database named MIMIC and have made these data available to the public [5,6]. This database has contributed not only to biomedical research but also to electronic, engineering, and computer science research. However, the data source was limited to one institute, which hinders cross validation or implementation in other institutes. Therefore, we developed a biosignal database, the contents of which were gathered mainly from Korean populations.

Several issues must be overcome to develop an integrated biosignal database incorporating data from different ICUs and systems. First, different monitoring systems may be in use within the same institute. Moreover, one ICU might have a central monitoring system that can temporarily store raw waveform data. However, systems in other ICUs may lack the ability to store raw data. Second, monitoring systems are typically independent of Electronic Medical Records (EMRs). Therefore, if biosignal and EMR data cannot be mapped, their utility for biomedical research, such as the development of disease prediction models or the discovery of novel biomarkers, is limited.

We aimed to develop a biosignal database comprising data collected from different systems. Use of a unified data for-

mat enables biosignal data to be accessed without the need to consider its source. We demonstrated the utility of our system using two software applications.

## II. Case Description

The study protocol was approved by the Ajou University Hospital Institutional Review Board (No. AJIRB-MED-MDB-16-155).

The study was performed in a tertiary teaching hospital with 1,185 beds in Suwon, Korea. The hospital has 12 ICUs with 205 beds, in which General Electronics, Nihon Kohden, and Philips patient monitoring systems are used. The Nihon Kohden (30 beds) and Philips (49 beds) systems were used in this study.

To integrate the data from the two monitoring systems, we defined a unified data format (Table 1) comprising (1) meta-data, such as measurement location, time, type of biosignal etc., which are stored in a relational database, and (2) raw signal waveform data that cannot be stored in a relational database because of the high frequency of the measurements. For example, ECG signals are measured at 250 and 500 Hz by the Nihon Kohden and Philips systems, respectively. Therefore, raw waveform data are stored in CSV files. To reduce redundancy, we recorded only the measurement start time in the CSV file name, not the timestamp of each measurement. Because every 10-second data was recorded

Table 1. Metadata collection specification

Table	Column	Type, precision	Description
NUMERIC_INFO	ID	Integer	Unique identifier for numeric records
	PATIENT_ID	Varchar(40)	Unique identifier for patients
	TIMESTAMP	Varchar(20)	Date and time when the data were recorded
	OBSERVATION	Varchar(20)	Type of observation (e.g., heart rate, respiratory rate, SpO <sub>2</sub> , etc.)
	OBSERVATION_VALUE	Float	Value of observation
	OBSERVATION_VALUE_UNIT	Varchar(20)	Unit for value of observation
	ROOM	Varchar(20)	Identifier for ICU where the patient was admitted
	BED_NO	Integer	Identifier for bed in each ICU
	WAVEFORM_INFO	ID	Integer
PATIENT_ID		Varchar(40)	Unique identifier for patients
STARTTIME		Varchar(20)	Date and time when data recording started
ENDTIME		Varchar(20)	Date and time when data recording ended
WAVETYPE		Varchar(100)	Type of observation (e.g., ECG, ABP, respiration, etc.)
FILEPATH		Varchar(255)	Location of file where the waveform raw data were written

SpO<sub>2</sub>: peripheral oxygen saturation, ICU: intensive care unit, ECG: electrocardiograph, ABP: arterial blood pressure.

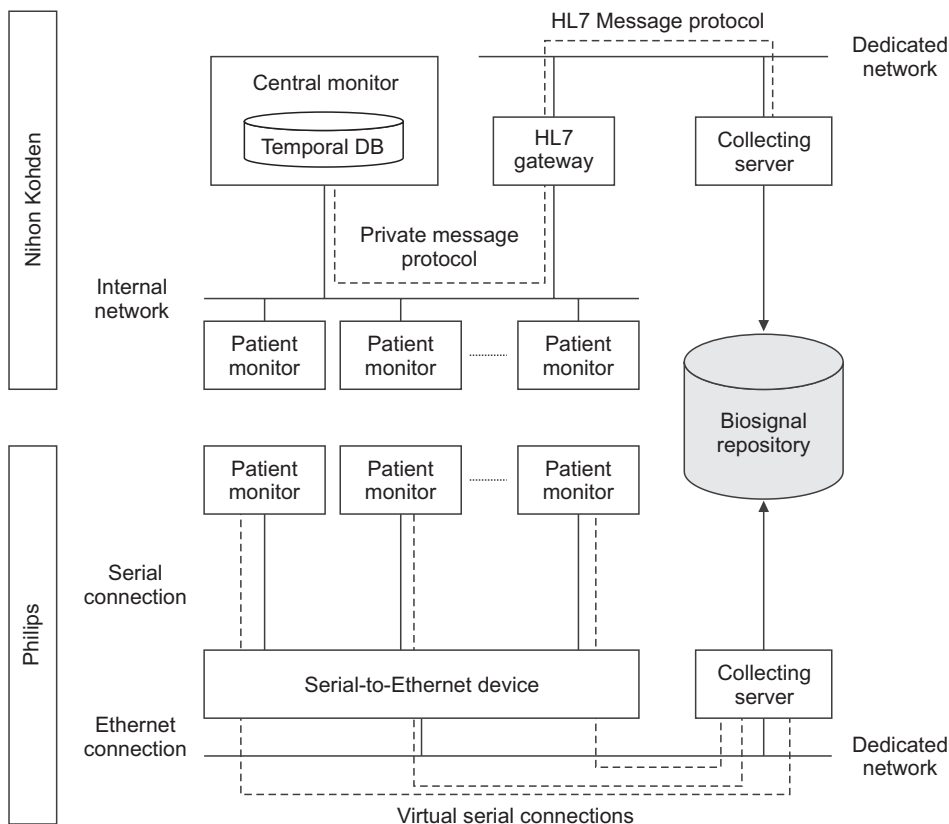


Figure 1. Schematic diagram of data collection from the Nihon Kohden (top) and Philips (bottom) patient monitoring systems.

in a single row, the user can easily recognize the frequency at which the data were recorded. For example, if there are 2,500 values in a single row, it indicates that the data were recorded at 250 Hz. Therefore, data recorded at a specific time can be accessed using the file name and the location of the value in the file. Each file was separated in every 10-minute recording period.

The Nihon Kohden system has a central monitoring system that collects and saves all signal waveform data for up to 5 days, and it transmits data using the HL7 message protocol (Figure 1). Temporarily stored waveform data were transferred to permanent storage using software based on the HL7 protocol.

The Philips system is also a central monitoring system; however, it does not have storage or transfer functions. To overcome this, we installed an RS232 interface board on each patient monitor to communicate with the data collection server in real time (Figure 1). Biosignal data are stored on the server temporarily, transformed into a unified format, and transferred to permanent storage every 10 minutes, as with the Nihon Kohden system.

From September 1, 2016 to August 31, 2017, we collected around 117 patient-years of biosignal data from 1,268 patients. Details on the data collected are provided in Table 2.

We developed two application software packages for our

biosignal database: a dashboard that provides information on the current status of the database, and a data viewer (Figure 2). Both applications functioned as expected, and we experienced no issues with compatibility.

### III. Discussion

We report here a system for collecting and storing biosignal data from different patient monitoring environments, unification of those data formats, and the utility of these data for further study or applications.

To be clinically useful, biosignal data should be linked to an EMR database. To achieve this, we used two strategies. First, ICU nurses enter the patient identification (ID) into the monitoring system, which is encrypted using the same algorithm applied to EMR data. Because IDs may be incorrect, our system can recognize the bed location where each device was installed along with the timestamp and link this with a specific patient ID.

The database can be used for prediction or determination of disease state (e.g., cardiac arrest or sepsis) or a specific value (e.g., serum potassium level) [2,4]. It will likely also contribute to drug post-market surveillance. Prolongation of the QT interval on ECG is a critical adverse drug event, but determining the relationship between a certain drug and QT

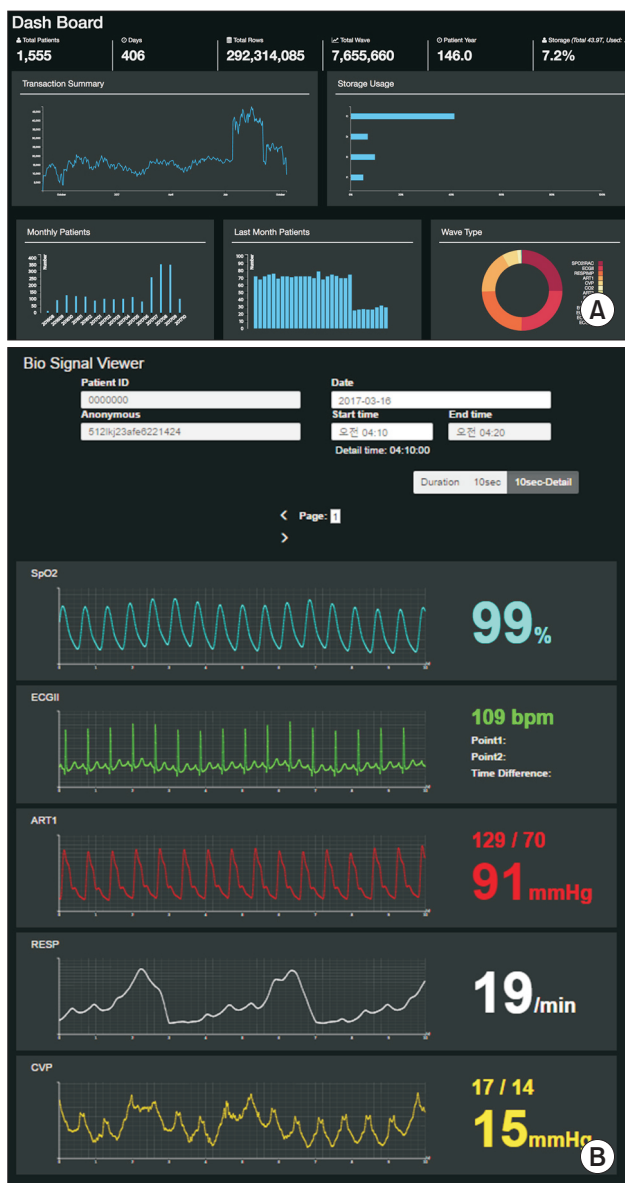


Figure 2. Two applications used in the integrated biosignal database: (A) dashboard and (B) data viewer.

prolongation is generally hampered by the lack of ECG data at the time of drug administration. However, our database can provide ECG data for a patient throughout their ICU stay.

We plan to develop diverse prediction and determination models using the database. We also plan to make the database (including both waveform and EMR data) available to the public to foster collaboration among researchers in various disciplines. We expect that our system will overcome the problems associated with MIMIC databases, such as the lack of cross validation data.

In this paper, we described the current status biosignal data collection in a tertiary teaching hospital in Korea, and

Table 2. Data collected

	Nihon Kohden	Philips
Number of ICUs	2	3
Number of beds	30	49
Data collection start date	September 2016	July 2017
Number of patients	954	314
Amount of numeric observation		
Heart rate	12,120,625	2,675,540
Respiratory rate	12,093,247	2,638,390
SpO <sub>2</sub>	12,073,501	2,720,840
Blood pressure	10,643,480	1,500,110
Central venous pressure	3,821,204	-
End-tidal CO <sub>2</sub>	853,640	-
Intracranial pressure	510,219	-
Length of collected waveforms (patient-year)		
PPG	24.3	5.1
ECG (lead II)	24.3	5.0
Respiration	23.1	5.1
Arterial blood pressure	17.1	3.2
Central venous pressure	7.4	-

ICU: intensive care unit, SpO<sub>2</sub>: peripheral oxygen saturation, CO<sub>2</sub>: carbon dioxide, PPG: photoplethysmograph, ECG: electrocardiograph.

introduced how we overcame it with our system. While biosignal data is an important source of artificial intelligence for health care [7-9], discussions are still ongoing to collect and store large volumes of biosignal data from different monitoring systems and to prepare them as data sets for future analysis [10]. We expect that this study could provide an opportunity advance these discussions.

Our system had several limitations. First, there is a 10-minute interval between data collection and storage in the repository, which is necessary for the unification of data formats from different monitoring systems. This limits the ability of the system to provide real-time analysis and feedback. However, this issue does not hamper the use of our system for research, such as the development of prediction algorithms. For practical application of the developed algorithm, the interval required for data collection and unification must be reduced. Second, data measurement and transmission involves the generation of noise or a faulty signal. We plan to develop a machine learning-based filtering model by tagging randomly sampled signals; this model will be capable of

determining whether a signal is suitable for use. We aim to use this model to estimate the probability of a signal being unsuitable when the waveform data are made available to the public.

In conclusion, we developed a system for collecting waveform data from different patient monitoring systems, on which software applications can be run without compatibility problems due to the use of unified data format. The database will facilitate the development of diverse algorithms and contribute to improved patient health outcomes.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. HI16C0982 and HI17C0970).

## References

1. Yoon D. What we need to prepare for the fourth industrial revolution. *Healthc Inform Res* 2017;23(2):75-6.
2. Moss TJ, Lake DE, Calland JF, Enfield KB, Delos JB, Fairchild KD, et al. Signatures of subacute potentially catastrophic illness in the ICU: model development and validation. *Crit Care Med* 2016;44(9):1639-48.
3. Lehman LW, Adams RP, Mayaud L, Moody GB, Malhotra A, Mark RG, et al. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. *IEEE J Biomed Health Inform* 2015;19(3):1068-76.
4. Attia ZI, DeSimone CV, Dillon JJ, Sapir Y, Somers VK, Dugan JL, et al. Novel bloodless potassium determination using a signal-processed single-lead ECG. *J Am Heart Assoc* 2016;5(1):e002746.
5. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035.
6. PhysioNet [Internet]. [place unknown]: PhysioNet; c2017 [cited at 2017 Oct 25]. Available from: <https://physionet.org/>.
7. mRhythm Study [Internet]. [place unknown]: mRhythmStudy.org; c2017 [cited at 2017 Oct 25]. Available from: <http://www.mrhythmstudy.org/>.
8. Lee H, Shin SY, Seo M, Nam GB, Joo S. Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks. *Sci Rep* 2016;6:32390.
9. Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks [Internet]. Ithaca (NY): arXiv.org; c2017 [cited at 2017 Oct 25]. Available from: <https://arxiv.org/abs/1707.01836>.
10. Kokkinaki A, Chouvarda I, Maglaveras N. Searching biosignal databases by content and context: Research Oriented Integration System for ECG Signals (ROISES). *Comput Methods Programs Biomed* 2012;108(2):453-66.