

## Estimation of Heart Rate from Vocal Frequency Based on Support Vector Machine

Motoki Sakai\*

Chiba University of Commerce, Faculty of Policy Informatics, Japan

### \*Correspondence Info:

Motoki Sakai,  
1-3-1 Konodai Ichikawa-shi Chiba,  
272-8512, Japan  
E-mail: [m.sakai510@gmail.com](mailto:m.sakai510@gmail.com)

### Abstract

Heart rate (HR) is one of the vital signs used to assess our physical condition; it would be beneficial if HR could easily be obtained without special medical instruments. In this study, a feature of vocal frequency was used to estimate HR, because it can easily be recorded with a common device such as a smartphone. Previous studies proposed that a support vector machine (SVM) that adopted the inner product as the kernel function was efficient for estimating HR to a certain extent. However, these studies did not present the effectiveness of other kernel functions, such as the hyperbolic tangent function. Therefore, this study identified a combination of kernel functions of the kernel ridge regression (KRR). In addition, features of vocal frequency to effectively estimate HR were investigated. To evaluate the effectiveness, experiments were conducted with two subjects. In the experiment, 60 sets of HRs and voice data were measured per subject. To identify the most effective kernel function, four kernel functions (the inner function, Gaussian function, polynomial function, and hyperbolic tangent function) were compared. Moreover, effective features of vocal frequency were selected with the sequential feature selection (SFS) method. As a consequence, the hyperbolic tangent function worked best, and high-frequency components of voice were efficient. However, results of this research indicated that effective vocal spectrum components to estimate HR differ depending on prediction models.

**Keywords:** Heart rate, vocal frequency, support vector machine, kernel function.

### 1. Introduction

Heart rate (HR) is one of the vital signs used to assess a patient's physical condition. In recent years, HR is often measured with a simplified measuring instrument when outside a hospital. For instance, mobile cardio-tachometers are used for prophylaxis or for treatment of lifestyle diseases, such as obesity. Obesity is responsible for many disorders, including diabetes, hypertension, and hyperlipidemia, and can lead to death [1]. For treatment of obesity, adequate exercise is effective. However, hard-intensity exercise is not advised for obese patients, and it is important to exercise under good conditions and appropriate exercise intensities. Thus, it is important to assess the physical condition and exercise intensity. HR is recommended as a simple and effective way to evaluate a patient's physical condition [2,3].

Currently, there are many applications to measure HR without special medical equipment. For example, HR can be measured with a smartphone camera or some wearable appliances. To maintain our health, these tools are useful because our physical condition should routinely and easily be assessed.

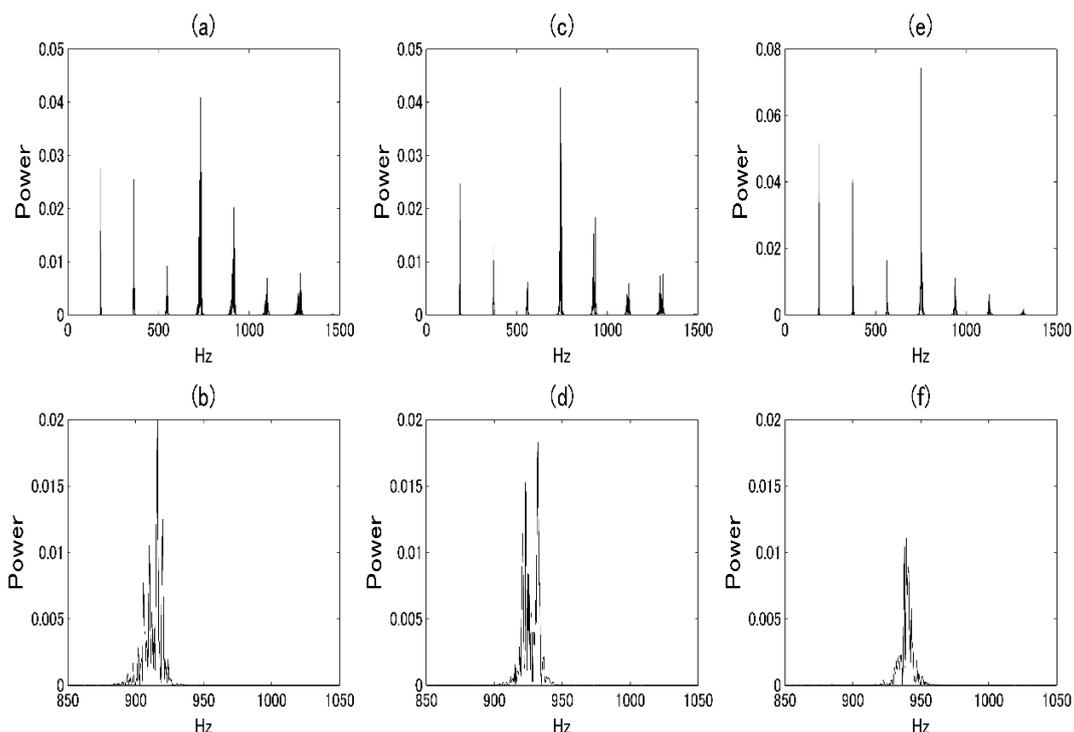
However, existing cardio-tachometers that use smartphone cameras have drawbacks from a practical perspective. For example, cardio-tachometers using smartphone cameras are easily influenced by the user's body motion or external light. Therefore, this research proposes a method to more easily estimate HR by measuring the human voice.

The relationship between HR and voice has already been observed, because HR and voice have been simultaneously measured in stress assessments and other similar assessments [4]. As shown in Fig. 1, a change in vocal frequency relates to HR variation. Fig. 1 illustrates that formants are shifted to the right (higher frequency) with a decrease in HR, and the amplitude of each formant also changes with HR variation. In sum, the higher the HR is, the larger the amplitude of the fifth formant. Although factors that change vocal frequency are not limited to HR, this example indicates that HR is one factor that influences vocal frequency.

Given this relationship between HR and vocal frequency, some methods to estimate HR from voice has been proposed [5-7], and [8]. Studies [5] and [6] take advantage of fluctuations synchronized with R-

waves of the ECG signal in the vocal spectrogram. In [5], HR is roughly estimated by image processing for 2D images of a vocal spectrogram. In [6], the authors chose the component of vocal frequency that is most affected by R-waves in one-dimensional scanning of the speech spectrogram; this algorithm realized accurate HR estimations. On the other hand, studies [7] and [8] proposed HR estimation methods based on the support vector machine (SVM), and HR values were estimated to some level. In [7] and [8], the inner product was used as the kernel function for the SVM, but evaluation results obtained by other kernel

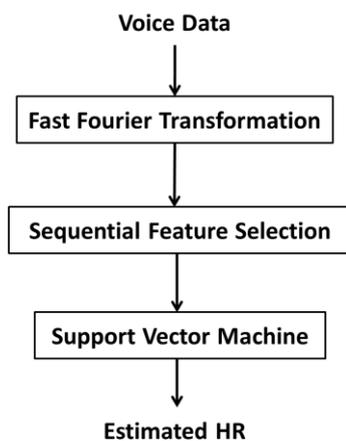
functions, such as the Gaussian or hyperbolic tangent function, were not presented. Generally, it is known that estimation accuracy depends on an adopted kernel function. Therefore, this research first compares several kernel functions and identifies the most suitable one to estimate HR. In addition, frequency bands that effectively estimate HR value are also examined. If a more effective combination of kernel function and voice feature can be identified, it facilitates the practical use of HR estimation from voice.



**Figure 1: Example of formant movement associated with a change in HR for a male subject: (a) Vocal frequency when HR is 105 bpm, (b) vocal frequency near 950 Hz when HR is 105 bpm, (c) vocal frequency when HR is 95 bpm, (d) vocal frequency near 950 Hz when HR is 95 bpm, (e) vocal frequency when HR is 78 bpm, (f) vocal frequency near 950 Hz when HR is 78 bpm**

**2. Support vector machine-based method to estimate HR from voice data**

Fig. 2 shows a processing diagram of HR estimation from voice data. The proposed HR estimation method consists of following three steps. First, the recorded voice signal is transformed into frequency representations by using the fast Fourier transform (FFT). Next, vocal frequency components to effectively estimate HR are selected by the sequential feature selection (SFS) method [9], and selected vocal frequency components are used as explanatory variables in the subsequent estimation step. Finally, an SVM-based [10,11] HR estimation is performed using the explanatory variables. (Generally, the term “explanatory variable” is also called “feature,” but the term “feature” means a variable mapped into an unknown feature space in the context of the SVM.)



**Figure 2: Processing diagram of HR estimation from voice data**

**2.1 Selecting effective explanatory variables**

In this research, explanatory variables to effectively estimate HR are selected with the SFS method. (Explanatory variables are vocal frequency components ranging from 120 Hz to 10000 Hz. In this research, time-domain features are not used.) In other words, the SFS method is performed to select the best subset to increase the accuracy of HR estimation. The SFS method is performed according to the following steps:

- I. SFS begins with an evaluation of the accuracy of the kernel ridge regression (KRR)-based estimation for each explanatory variable, identifying one explanatory variable to most accurately estimate HR.
- II. SFS is performed using a pair of explanatory variables. The explanatory variable selected in Step I is coupled with each of the other variables from the remaining explanatory variable subset, identifying the pair yielding the highest estimation accuracy.
- III. SFS next evaluates three variables: the pair from Step II and another selected from the remaining subset.
- IV. Steps I-III are repeated with four or more explanatory variables.

Usually, the estimation accuracy peaks when the number of selected explanatory variables reaches a certain number. In sum, the best explanatory variable subset can be singled out for maximum accuracy in the SFS procedure. (Note that the “accuracy” is defined as a correlation coefficient between the estimated and measured HRs in this research.)

**2.1 Support vector machine-based regression method**

The SVM, which was proposed by Vapnik [12], can realize an effective regression in the feature space mapped by nonlinear transformation. SVM performs implicit mapping of the input data into a high-dimensional feature space using a kernel function. This is defined as an inner product in the feature space and does not require an understanding of the types of features that are being used. In this research, KRR, which is one of the applications of SVM algorithms, was used to estimate HR from the vocal frequency. KRR is a ridge-regression method based on some kernel matrices, and it can also be derived from a modified least-squares method.

Suppose that  $x_t \in \mathbb{R}^n$  and  $y_t \in \mathbb{R}$ , ( $t = 1, \dots, L$ ) are training sets, and  $x_s \in \mathbb{R}^n$  is an evaluation set. In this research,  $x_t$  and  $x_s$  are explanatory variables of

vocal frequencies selected by the SFS method, and  $y_t$  is a vector of the measured HR value. KRR predicts the HR value  $f(x_s)$  from a new unlabeled example  $x_s$  by

$$f(x_s) = y'(K + \lambda I)^{-1}k, \tag{1}$$

Where  $k$  is a vector whose elements are  $k_t = K(x_t \cdot x_s)$ , ( $t = 1, \dots, L$ ), and  $K$  is a matrix whose elements are  $K_{t1,t2} = K(x_{t1}, x_{t2})$ , ( $t1, t2 = 1, \dots, L$ ). This vector and matrix are used to construct the linear regression in the feature space. In this research,  $f(x_s)$  is the estimated HR, and  $y$  is the measured HR.

A kernel function  $K(x, z)$  computes the inner product of the two vectors  $x$  and  $z$  in a given feature mapping, which are realized in several forms such as

$$K(x, z) = ax^T \cdot z, \tag{2}$$

$$K(x, z) = \exp\left(\frac{-a\|x - z\|^2}{\sigma^2}\right), \tag{3}$$

$$K(x, z) = (x^T \cdot z + c)^m, \tag{4}$$

$$K(x, z) = \tanh(ax^T \cdot z + b). \tag{5}$$

The kernel functions described in equations (2) through (5) are called the inner function, Gaussian function, polynomial function, and hyperbolic tangent function, respectively. In these functions, parameters  $a$ ,  $b$ ,  $c$ ,  $m$ , and  $\sigma$  must be optimized depending on obtained datasets, and there are some algorithms to decide them. In this research, these parameters have been empirically set.

In the following evaluation step, these four kernel functions were tested to estimate HR, and accuracies of HR estimation obtained by these kernel functions are compared.

**3. Experiment**

To evaluate the proposed method to estimate HR from vocal frequency, experiments measuring the HR and voice were conducted involving two subjects: a 33-year-old male and a 25-year-old female.

In the measurement experiments, factors that might affect recorded vocal frequency were controlled. The measurement experiments were conducted in a sufficiently quiet room; moreover, subjects always sat with their head upright in a relaxed manner, because

physical posture can influence the production of sound.

First, the subject's voice was recorded using a voice recorder (Sony ICD-TX50). The voice recorder was kept 5 cm from the subject's mouth, and the subject was asked to sustain the sound [ã] (Japanese vowels) for 7 s. The sampling frequency of the voice recording was 44.1 kHz with 16-bit resolution. Next, immediately following the voice recording, HR was measured with a cardio-tachometer (Omron HEM-1010), assuming that the HR does not change suddenly if there are no external stimuli. In fact, subjects were not given any stimuli after the voice measurement. These two steps were repeated 60 times for each subject.

#### 4. Evaluation and results

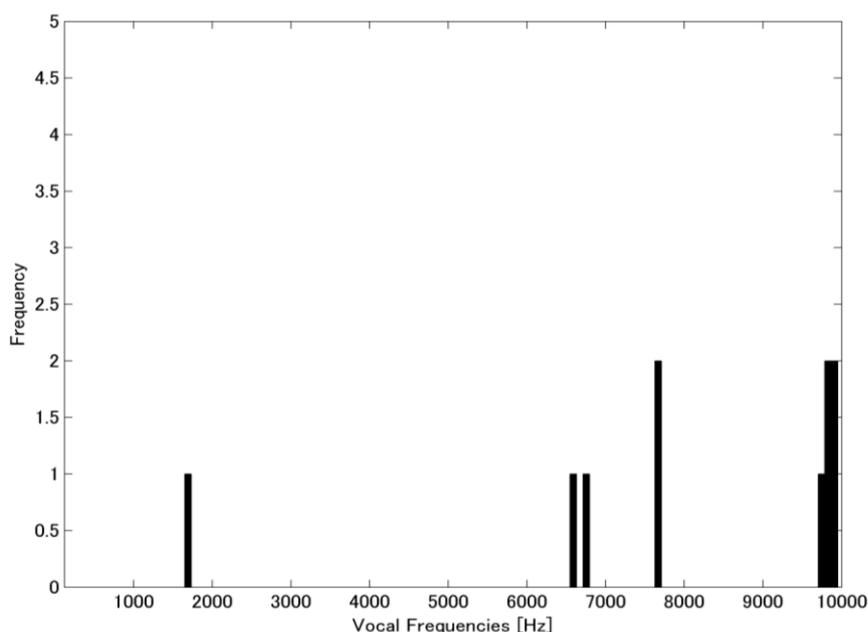
As described above, KRR-based HR estimations were performed using the four kernel functions described in equations (2) through (5). To evaluate the estimation accuracies, the correlation coefficient between measured and estimated HRs is computed for each kernel function. For each subject, one data set is initially selected as evaluation data from the 60 trials. The remaining 59 are used as the training sets, and the HR value is estimated. Next, another data set that has not been selected as evaluation data is selected as the new evaluation data, and HR is estimated with the remaining 59 training sets. This same selection and estimation procedure is repeated 60 times, and the 60 estimated HR values are computed. Table 1 shows these correlation coefficients for the male subject, female subject, and both subjects.

**Table 1: Correlation coefficients between measured and estimated HRs**

	Inner product	Gaussian	Polynomial function	Hyperbolic tangent
Male subject	1.00	0.99	0.82	1.00
Female subject	1.00	1.00	0.99	1.00
Both subjects	0.51	0.92	0.34	1.00

Figs. 3 through 8 show the vocal frequencies selected by the SFS. In particular, Figs. 3 to 5 show histograms of the top nine most effective frequency components for the male, female and both subjects' HRs, respectively. These results were obtained by

KRR-based estimation with the hyperbolic tangent kernel function. On the other hand, Figs. 6 through 8 show histograms obtained by the Gaussian kernel function.



**Figure 3: Histogram of nine frequency components selected by the SFS for the male subject's HR**

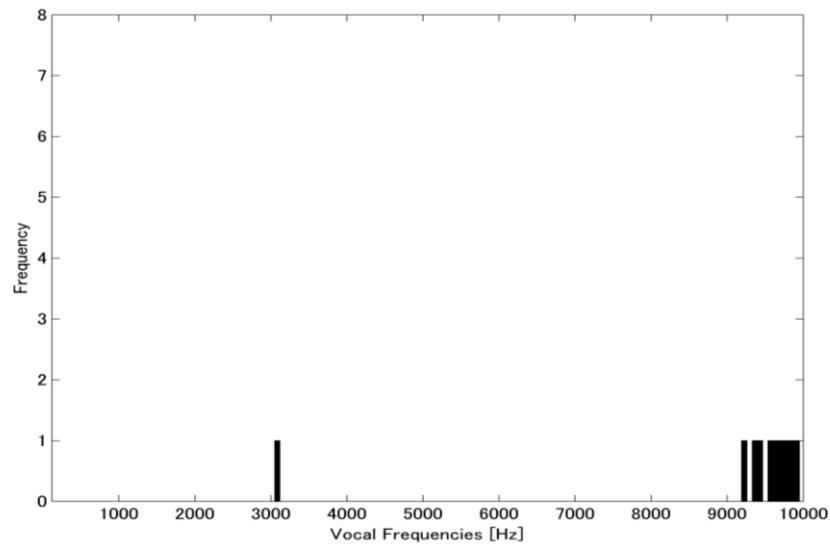


Figure 4: Histogram of nine frequency components selected by the SFS for the female subject's HR

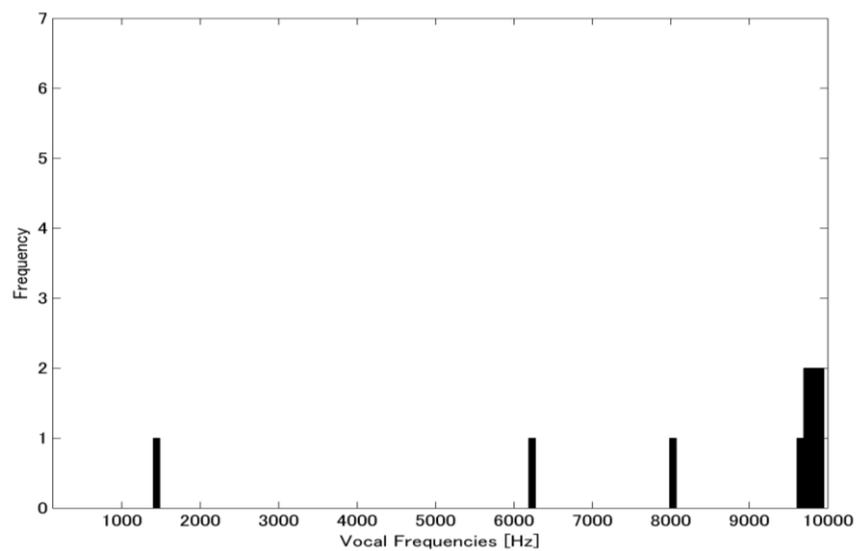


Figure 5: Histogram of nine frequency components selected by the SFS for both subjects' HRs

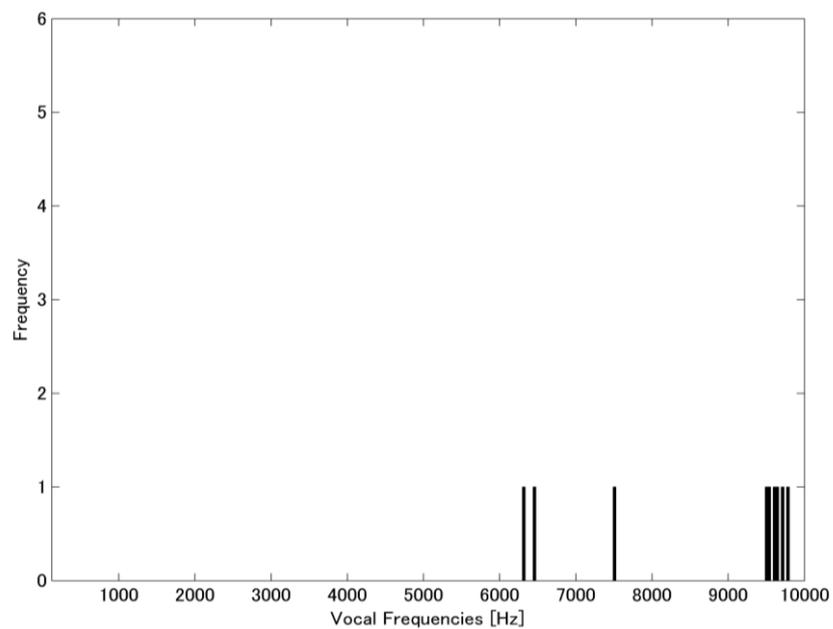


Figure 6: Histogram of nine frequency components selected by the SFS for the male subject's HR

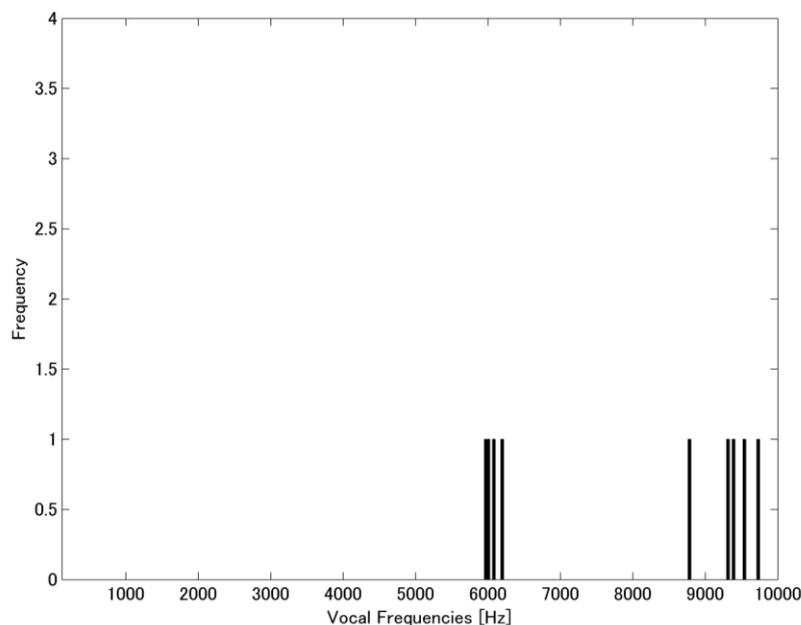


Figure 7: Histogram of nine frequency components selected by the SFS for the female subject's HR

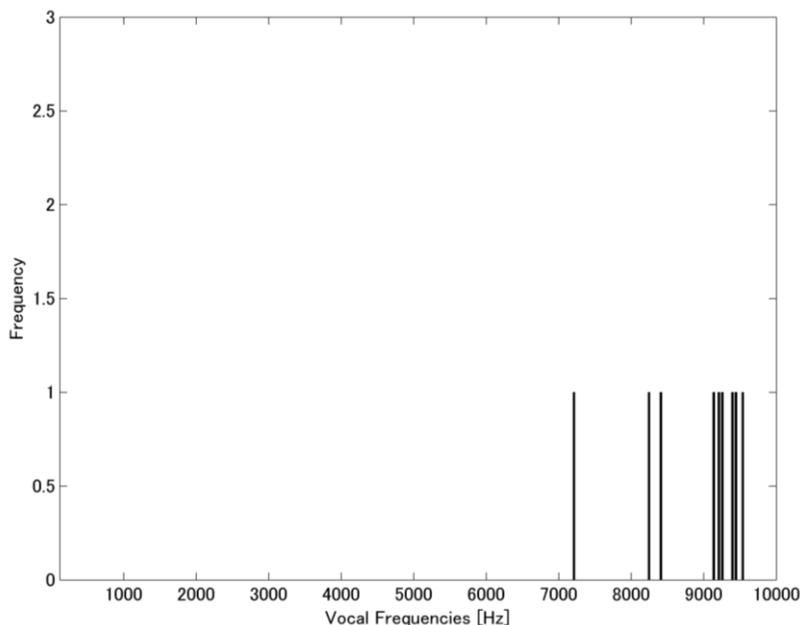


Figure 8: Histogram of nine frequency components selected by the SFS for both subjects' HRs

### 5. Discussion

To estimate the HR value from the vocal frequency, the KRR was used in this research, and four kernel functions (inner function, Gaussian function, polynomial function, and hyperbolic tangent function) were compared to examine which works best for the HR estimation.

As shown in Table 1, the hyperbolic tangent function yielded the most accurate estimation of HR for the male, female, and both subjects (correlation coefficient = 1.0), and the Gaussian function was the second-most effective (correlation coefficient > 0.9). By contrast, the inner function and polynomial function were effective for each subject's (male and

female) HR estimation, but these did not produce excellent results when the learning machines for the KRR were generated with both male and female subjects' vocal frequency components.

Figs. 3 through 8 show histograms of the vocal frequency components that were selected as effective features to estimate HR by the SFS method when the hyperbolic tangent function or Gaussian function was used as the kernel function. These figures illustrate that the vocal frequency components higher than approximately 9000 Hz were most selected. (For example, Fig. 4 shows that eight out of the nine effective frequency components were selected from frequencies higher than 9000 Hz.) In addition, these

figures illustrate that frequency components ranging from 6000 Hz 7000 Hz were the second-most effective, behind frequencies higher than 9000 Hz. These results indicate that high-frequency components of voice are effective for HR estimation.

Meanwhile, the previous study [13], which proposed an explicit model of the relationship between HR and vocal frequency components, proposed that low- and medium-frequency components were effective for HR estimation. However, an account for these differences is unknown. Therefore, in future work, physical or physiological interpretation of these differences of effective vocal frequency components between prediction models will be investigated with more subjects.

## 6. Conclusion

The goal of this research is to identify an effective combination of kernel function of the KRR and features of vocal frequency aimed at HR estimation. To evaluate the effectiveness of kernel functions and features of vocal frequency, HR and voice data were recorded from male and female subjects. To identify the most suitable kernel function, the inner function, Gaussian function, polynomial function, and hyperbolic tangent function were compared. In addition, effective vocal spectrum features were selected with the SFS method. As a result, the hyperbolic tangent function realized the most accurate HR estimation, and it is implied that high-frequency voice components were most efficient. However, results of this research indicated that effective vocal spectrum components to estimate HR differ substantially depending on prediction models.

## References

- [1] Halpern, M.C. Mancini, M.E. Magalhães, M. Fisberg, R. Radominski, M.C. Bertolami, A. Bertolami, M.E. de Melo, M.T. Zanella, M.S. Queiroz, and M. Nery, "Metabolic syndrome, dyslipidemia, hypertension and type 2 diabetes in youth: from diagnosis to treatment," *Diabetology & Metabolic Syndrome*, 2010.
- [2] M.H. Davenport, S. Charlesworth, D. Vanderspank, M.M. Sopper, and M.F. Mottola, "Development and validation of exercise target heart rate zones for overweight and obese pregnant women," *Applied Physiology, Nutrition, and Metabolism.*, vol. 33, no. 5, pp. 984–989, 2008.
- [3] S.C. Franckowiak, D.A. Dobrosielski, S.M. Reilley, J.D. Walston, and R.E. Andersen, "Maximal heart rate prediction in adults that are overweight or obese," *Journal of Strength & Conditioning Research*, vol. 25, no. 5, pp. 1407–1412, 2011.
- [4] J. Hada and Y. Takeuchi, "Evaluation of mental stress by voice analysis during simulated landing," *Japanese Journal of Ergonomics*, vol. 44, no. 3, pp. 171–174, 2008.
- [5] D. Skopin and S. Baglikov, "Heartbeat feature extraction from vowel speech signal using 2D spectrum representation," *Proceedings of the 4th International Conference Information Technology*, 2009.
- [6] Mesleh, D. Skopin, S. Baglikov and Anas Quteishat, "Heart rate extraction from vowel speech signals," *Journal of Computer Science and Technology*, vol. 27, no. 6, pp. 1243–1251, 2012.
- [7] Schuller, F. Friedmann, and F. Eyben, "The Munich biovoice corpus: effects of physical exercising, heart rate, and skin conductance on human speech production," *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pp. 1506–1510, 2014.
- [8] Schuller, F. Friedmann, and F. Eyben, "Automatic recognition of physiological parameters in the human voice: heart rate and skin conductance," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7219–7223 2013.
- [9] M.O. Mendez, J. Corthout, S. Van Huffel, M. Matteucci, T. Penzel, S. Certti, and A. M. Bianchi, "Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis," *Physiol Meas*, vol. 31 no. 3, pp. 273–289, 2010.
- [10] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Information Processing*, vol. 10, no. 10, pp. 203–224, 2007.
- [11] N. Vapnik, "Statistical learning theory," In A. Gammerman, editor, *Computational Learning and Probabilistic Reasoning*. Wiley, 1996.
- [12] Saunders, A. Gammerman, and V. Vovk, "Ridge regression learning algorithm in dual variables," *Proceedings of the 15th International Conference on Machine Learning, ICML'98*, pp. 515–521, 1998.
- [13] M. Sakai, "Modeling the relationship between heart rate and features of vocal frequency," *International Journal of Computer Applications*, vol. 120, no. 6, pp. 32–37, 2015.