ICIC Express Letters
Volume **13**, Number **1**, January **2019**
ICIC International ©2019 ISSN 1881-803X
pp. 11–18

DETERMINING THE HEART RATE BY USING SPEECH SIGNAL

Fergyanto Efendy Gunawan and Mellisa Pratiwi
Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
fgunawan@binus.edu

Received June 2018; accepted September 2018

Abstract. *Heart disease is known to be one of the primary causes of death worldwide. In the United States, the disease is the leading cause of death for people of most ethnicities. In Indonesia, the disease contributes 17% of the total mortality. It is also well-known that a regular or continuous monitoring the heart rate is necessary to minimize the fatality. At the current technological state, the monitoring method is expensive and often requires a wearable sensor, which is inconvenient. This study intends to establish a relationship between the heart rate and speech signals empirically by using regression and correlation analysis. The speech signals are analyzed through a frequency analysis using fast Fourier transform. The results suggest that vowel 'e' speech signal has a strong correlation to the heart rate.*
Keywords: Heart rate, Vowel speech signals, Fast Fourier transform, Relative error

1. **Introduction.** Around seventy million people worldwide suffer from heat disease [1]. Most of them died because they did not receive appropriate medical treatment in a timely manner. According to the WHO report, more than 200000 Indonesian people died because of a heart attack, which contributes 17% of the total deaths [2]. Early detection is believed to be a solution to reduce the mortality rate of heart-disease patients. Early detection is highly needed [3].

A standard diagnostic tool to assess human heart rate condition is electrocardiogram (EKG or ECG) [4]. ECG is a non-invasive contact tool to check the electrical activity of the heart based on the bio-potential signals acquired from human skin surface [5]. It requires patients to wear adhesive gel patches or chest straps that may cause skin irritation and discomfort [6].

Lately, along with the increasing number of heart disease patients who require long-term and continuous heart rate monitoring, the demand for contact-less and easy-to-use heart-rate measurement devices increased significantly. Some previous researchers have shown that heart rate can be identified from a distance. For example, [7] found that English vowel speech signals could be related to the human heart rate. The previous research about heart rate extraction using English vowel speech signal showed that heart activity could be accurately estimated by applying Fourier analysis, including short-time Fourier transform (STFT) and discrete Fourier transform (DFT) to detect the formant maximum peaks from the English vowel-speech signals [7]. [8] has proposed a novel non-contact-less technique for heart-rate extraction. The human speech signals contain heart-beat modulation in which the relevant ECG information can be extracted.

Although the number of works in this area is limited, the conjecture that speech signal is somewhat related to the heart rate has rather sound ground. The model of the source-filter theory of speech production suggests that the speech production is a result of air molecules moving across the human vocal tract. The produced speech is controlled by the vocal tract

DOI: 10.24507/icicel.13.01.11

volume, shape, and length. The larynx and vocal cord, essential components of the vocal tract, contain muscles covered by many blood vessels connected to the circulatory system. Thus, the heart rate is dynamically related to the variations of vocal cord parameters and is directly related to the acoustic properties of human speech [7, 9].

The objectives of this work are of the following. The first is to explore pattern that may relate the heart rate to the speech signals, particularly, vowels in Indonesian language. The second is to establish a mathematical model of the relationship and to evaluate the level of conformity of the model.

The manuscript contains the following sections. Section 2, Research Methods, presents the research procedure, data collection methods, and data analysis methods. Section 3, Results and Discussion, presents the obtained patterns relating heart rate to speech signal. In addition, the patterns are also mathematically quantified, and the model validity is statistically examined. Section 4, Conclusions, briefly presents the current research focus, related findings, and a recommendation for the future work.

2. **Research Methods.** The procedure of this research is of the following. Firstly, we perform the data acquisition process. Secondly, we analyze the data to extract essential features. Thirdly, we perform a statistical analysis to establish a connection between the data.

Two essential types of data to support this study are the heart-rate data and the associated speech signal. The heart-rate data and vowel speech signals were collected from 20 adults, ten males and ten females. They were asked to speak the vowels: ‘a’, ‘i’, ‘u’, ‘e’, and ‘o’ for a duration about 6 s, and their voices were recorded using a smartphone.

The collected speech signal was transformed from the time domain t to the frequency domain f by applying the Fourier transform:

$$X(f) = \int_{-\infty}^{+\infty} x(t) \exp\left(-2\jmath\pi f t\right) \,\mathrm{d}t, \tag{1}$$

where $\jmath = \sqrt{-1}$, $x(t)$ is the speech signal data in the time domain, and $X(f)$ is the data in the frequency domain.

Prior to the transformation, the signal was smoothed out by applying the fifth order of the Butterworth filter with the cutoff frequency of 1 kHz and 6 dB roll-off. Mathematically, the gain of an n-order Butterworth low-pass filter is

$$G^2(\omega) = \frac{G_0^2}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}, \tag{2}$$

where ω is the angular frequency, ω_c is the cutoff frequency, and G_0 is the DC gain.

In addition to the filtering, the signal was also passed through a Hamming window. The window formula is

$$w(n) = \alpha - \beta \cos\left(\frac{2\pi n}{N-1}\right), \tag{3}$$

where $\alpha = 0.54$, $\beta = 0.46$, and the index $n \in [0, N-1]$.

From the signal spectra $X(f)$, we identified the frequencies related to the five most dominant spectra. Subsequently, we determined the differences in those frequencies, and the average of these frequency differences was considered to be the most relevant feature for heart rate prediction.

The heart-rate data were collected by means of Pulse Oximeter. To obtain the data for the condition of high heart rate, the participants were asked to perform some light exercises, jogging or jumping, prior to the data acquisition. In total, 100 speech signals were collected.

We assumed that the heart rate is linearly related to the average frequency difference or

$$\text{bpm} = \beta_0 + \beta_1 \cdot \Delta f, \tag{4}$$

where β_0 and β_1 are the model coefficients, bpm is the heart rate in bpm and Δf is the average frequency difference in Hz.

The validity of the proposed linear model (4) would be checked by using the two statistics: t-statistic to measure the significance of β_1 and the coefficient of determination R^2. The t-statistic has the following hypotheses:

$$\mathrm{H}_0: \quad \beta_1 = 0, \text{ and} \tag{5}$$

$$\mathrm{H}_\mathrm{a}: \quad \beta_1 \neq 0. \tag{6}$$

Thus, we expect that the empirical data to support the case where $\beta_1 \neq 0$. The coefficient of determination would measure to which extent the assumed linear model fits to the empirical data. The relevant formulas are

$$R^2 = 1 - \frac{\mathrm{SS}_\mathrm{res}}{\mathrm{SS}_\mathrm{tot}}, \tag{7}$$

$$\mathrm{SS}_\mathrm{tot} = \sum_i \left(\text{bpm}_i - \overline{\text{bpm}}\right)^2, \text{ and} \tag{8}$$

$$\mathrm{SS}_\mathrm{res} = \sum_i \left(\text{bpm}_i - \widehat{\text{bpm}}_i\right)^2, \tag{9}$$

where bpm_i is the ith data of bpm, $\overline{\text{bpm}}$ is the average bpm, and $\widehat{\text{bpm}}_i$ is the predicted bpm.

Eighty-five percents of the data were used in the training phase where the coefficients β_0 and β_1 were estimated by using the least-squares method. The remaining data were used to evaluate the accuracy of the developed linear model.

3. Results and Discussion.

3.1. **Speed data in time and frequency domains.** As mentioned in Section 2, we begin this research by recording voices of participants pronouncing the letters 'a', 'i', 'u', 'e', and 'o' separately in Indonesian language for a short duration for each signal. The voice signals are stored in '*.wav' files. An example of the recorded voice signals, i.e., for vowel 'e', is shown in Figure 1. Then, the signals are converted into frequency domain by using Equation (1). For the example speech signal of Figure 1, the frequency-domain

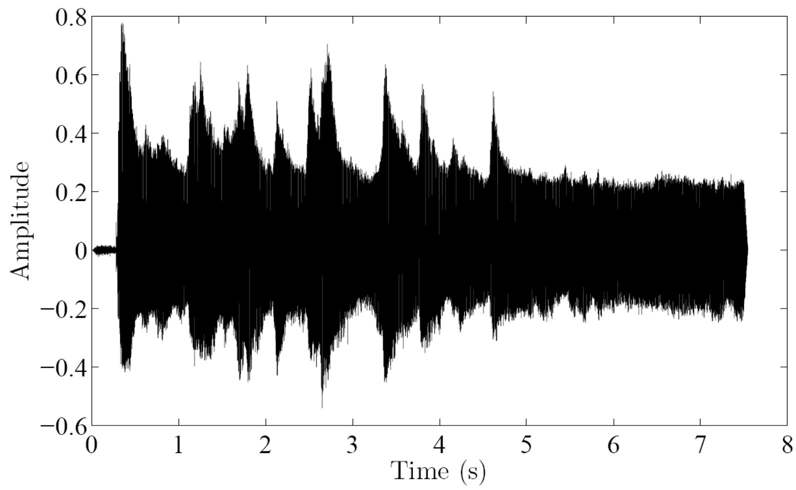

Figure 1. The example of the speech signal for vowel 'e' in Indonesian language

results are shown in Figure 2. We note that the signals in the frequency domain are presented in complex numbers. Only the magnitude of the signal is presented in Figure 2.

In the frequency domain, we observe and estimate the distance between two subsequent dominant-frequencies. Graphically, this distance, labeled as Δf, is shown in Figure 2. From each signal, we take the average of Δfs to represent the frequency difference of the speech signal.

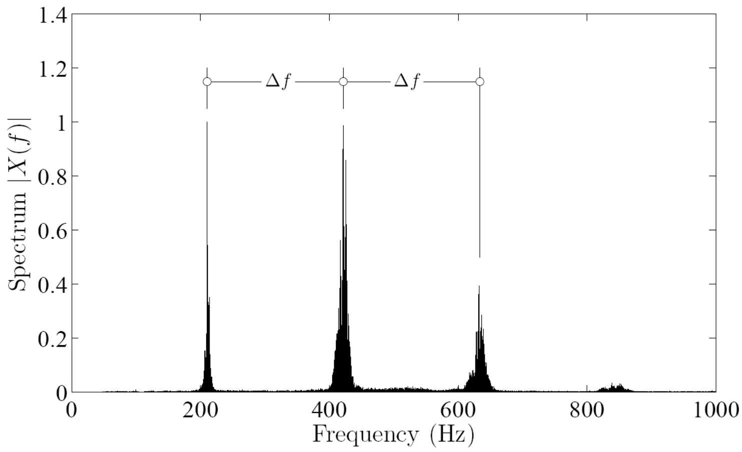

FIGURE 2. The spectra of the speech signal shown in Figure 1 and the definition of frequency difference Δf

3.2. **Linear model development.** The main result of this work is the relation between the heart rate and the frequency difference. These relations are graphically shown in Figures 3-7 and are numerically evaluated for its fitness in Table 1.

From these results, we conclude the following. The hypothesis that the heart rate is linearly related to the frequency difference is strongly supported by the empirical data. This strong supports can be seen from the values of the computed coefficient of determinations, which are in the range of 0.699-0.869, except for the case of 'o' speech signal. The t-statistic also supports the hypothesis that can be seen from the p-statistics, whose values are extremely small, except for the case of 'o' speech signal.

3.3. **Testing of the regression model.** The best model established in the previous section is the case of 'e' speech signal. The model has the fitness level of 86.9% as measured by the coefficient of determination R^2. This model is better than the second-

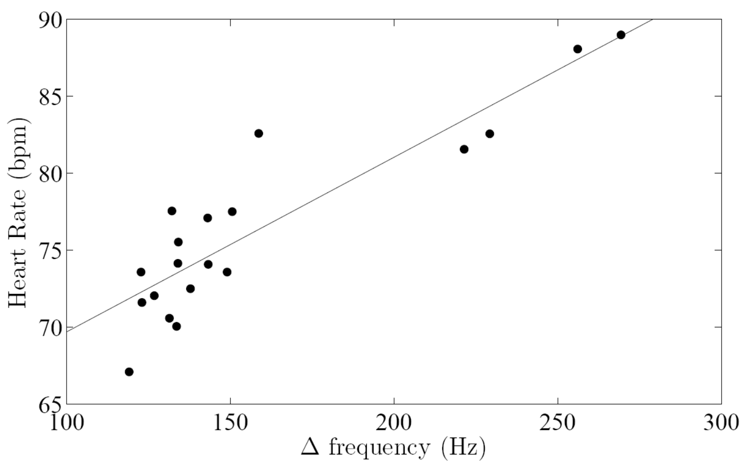

FIGURE 3. The average frequency differences in Hz versus the heart rate in bpm for the Indonesian speech signal for the character 'a'

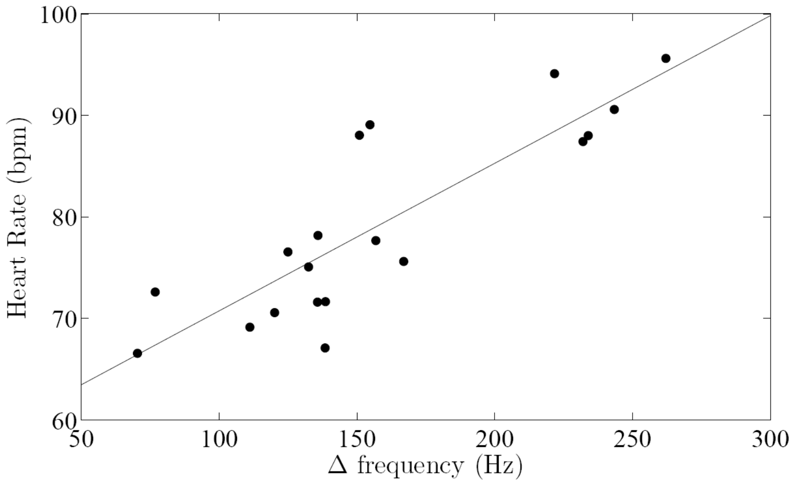

Figure 4. The average frequency differences in Hz versus the heart rate in bpm for the Indonesian speech signal for the character 'i'

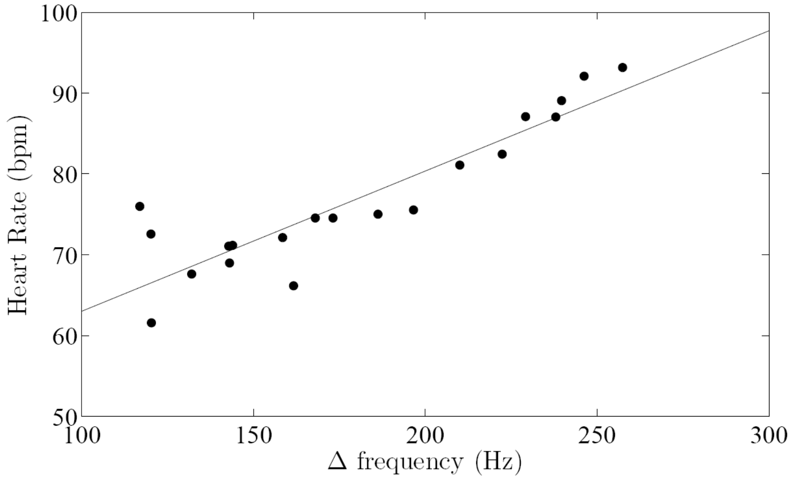

Figure 5. The average frequency differences in Hz versus the heart rate in bpm for the Indonesian speech signal for the character 'u'

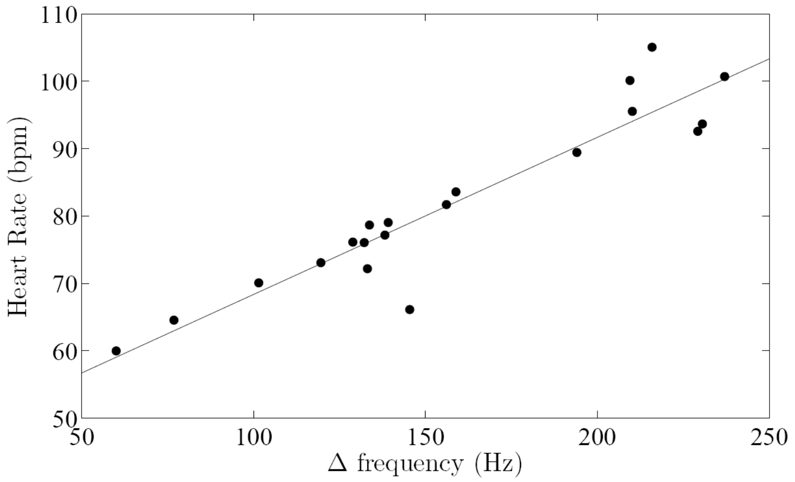

Figure 6. The average frequency differences in Hz versus the heart rate in bpm for the Indonesian speech signal for the character 'e'

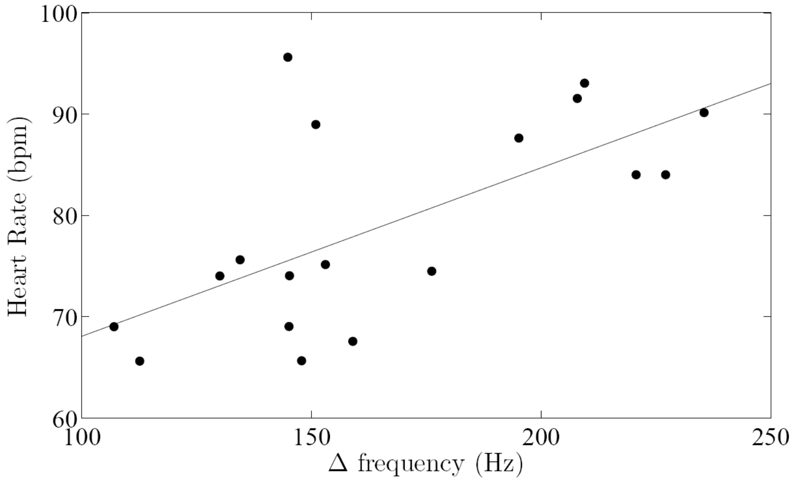

FIGURE 7. The average frequency differences in Hz versus the heart rate in bpm for the Indonesian speech signal for the character 'o'

TABLE 1. Some statistics describing the relations between the frequency difference Δf in Hz and the heart rate in bpm. The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates of β_0 and β_1 in Equation (4). The statistic p-value measures the significance of $\hat{\beta}_1$; p-value smaller than 5% is considered significant.

Speech Signal	$\hat{\beta}_0$	$\hat{\beta}_1$	R^2	adj-R^2	p-value
'a'	58.358	0.113	0.801	0.789	0.000
'i'	56.206	0.145	0.716	0.699	0.000
'u'	45.653	0.174	0.813	0.803	0.000
'e'	45.057	0.233	0.876	0.869	0.000
'o'	51.449	0.166	0.422	0.385	0.004

TABLE 2. The results of the testing of the model bpm $= 45.06 + 0.233 \cdot \Delta f$, which is obtained for the vowel 'e'

No.	Age	Gender	Heart Rate (bpm)		
			Measurement	Estimation	Error (%)
1	23	M	69	77	8
2	22	M	66	80	14
3	61	M	70	77	7
4	21	M	80	80	0
5	54	F	81	78	3
6	23	F	86	79	7
7	23	F	105	96	9
8	23	F	109	98	11
9	22	F	121	99	22
10	24	M	66	68	2
11	34	M	72	77	5
12	22	M	73	72	1
13	23	F	85	79	6
14	70	F	81	77	4
15	23	M	78	78	0

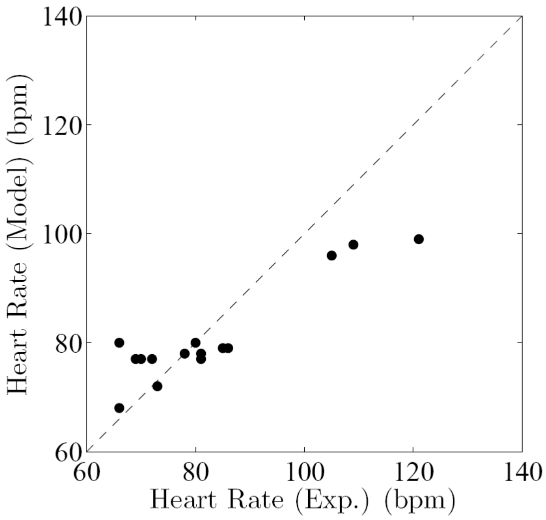

FIGURE 8. A comparison of the measured heart rate and the predicted heart rate using a linear model. The coefficient of determination R^2 of these data is 0.85.

and third-best models by about 6% and 8% (see Table 1 for the detail). This case provides the model of bpm $= 45.06 + 0.233 \cdot \Delta f$. In this section, we test the accuracy of this model. We use the testing data set, which is collected separately from the data set utilized in the model training. The results are tabulated in Table 2 and graphically are shown in Figure 8. With the coefficient determination of 0.85, these results are rather satisfactory.

4. **Conclusions.**

From the medical perspective, the capability to continuously monitor the heart rate is of interest. The continuous monitoring of the heart rate is arguably crucial to prevent the sudden death due to a heart attack. The existing techniques are rather impractical. This work intends to establish the relationship between the speech signal frequency and the heart rate. This particular relation would facilitate the continuous heart rate monitoring. The results suggest that the heart rate may be somehow related to the speech signal frequency by a linear model. This model is rather accurate for vowel speech signals of ‘a’, ‘i’, ‘u’, and ‘e’, but is less accurate for the vowel speech signal of ‘o’.

In this work, the relationship between the feature of the Indonesian speech signal and the heart rate is assumed to be linear. The applicability of this assumption may also be evaluated for other languages. We leave this issue for the future work.

REFERENCES

[1] F. Zennifa, H. Kamil, K. Iramina et al., Prototype early warning system for heart disease detection using android application, *The 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp.3468-3471, 2014.

[2] LeDuc, *Indonesia: Coronary Heart Disease*, http://www.worldlifeexpectancy.com/indonesia-coronary-heart-disease, Retrieved on February 2, 2015.

[3] D. S. Celermajer, C. K. Chow, E. Marijon, N. M. Anstey and K. S. Woo, Cardiovascular disease in the developing world: Prevalences, patterns, and the potential of early disease detection, *Journal of the American College of Cardiology*, vol.60, no.14, pp.1207-1216, 2012.

[4] H. C. Hung and D. Wei, Muscle noise cancellation in four-lead ecg system for ubiquitous healthcare, *Joint Conferences on Pervasive Computing (JCPC)*, pp.115-118, 2009.

[5] A. A. A. Bakar, Y. L. Lim, S. J. Wilson, M. Fuentes, K. Bertling, T. Bosch and A. D. Rakic, Electrocardiographic signal detection using self-mixing interferometer technique with customized electro-optic phase modulator, *The 3rd International Conference on Photonics*, 2012.

[6] D.-Y. Chen, J.-J. Wang, K.-Y. Lin, H.-H. Chang, H.-K. Wu, Y.-S. Chen and S.-Y. Lee, Image sensor-based heart rate evaluation from face reflectance using Hilbert-Huang transform, *IEEE Sensors Journal*, vol.15, no.1, pp.618-627, 2015.

[7] A. Mesleh, D. Skopin, S. Baglikov and A. Quteishat, Heart rate extraction from vowel speech signals, *Journal of Computer Science and Technology*, vol.27, no.6, pp.1243-1251, 2012.
[8] D. Skopin and S. Baglikov, Heartbeat feature extraction from vowel speech signal using 2D spectrum representation, *Proc. of the 4th Int. Conf. Information Technology*, 2009.
[9] I. R. Titze, *Principles of Voice Production*, National Center for Voice and Speech, 2000.