Bayesian Modelling of The Distribution of Heart Rate (BPM) Among Nigerian Students

Adedoyin O. Taiwo¹, Abiodun A. Taiwo²

Abstract:- Heart Rate (bpm) is widely regarded as a very significant element in every human being. Numerous studies have been so far undertaken in this field, but a very few have really tried to touch the distribution of students' Heart Rate (bpm). The primary aim of this study is to investigate the heart rate distribution of faculty of science students at the University of Ibadan.

A total number of 347 students' information was used in the analysis. The data was obtained in the faculty of science, a descriptive analysis of the data collected was used to investigate the distribution of the dataset, mean, standard deviation, minimum value and maximum value. The Kolmogorov - Smirnov's test was used to validate the distribution of the likelihood. The Chi Square was used to test for the association between Gender and the distribution of Heart Rate.

The descriptive analysis of the collected data revealed that the average heart rate and corresponding standard deviation of the observed data were 71.46 and 4.977, with the minimum value being 58 bpm and the maximum value being 87 bpm. The Kolmogorov-Smirnov test revealed that the data fits well with the normal distribution. The Chi Square test of independence revealed that there is no relationship between Gender and Heart Rate Distribution (bpm).

For the Conjugate prior (which is informative), where the prior, the likelihood and the posterior was found to follow the normal distribution, the posterior mean (Bayes estimate) of the distribution of heart rate (bpm) is 71.55 while the posterior standard deviation is 4.65. The posterior credible interval for average heart rate (bpm) is [62.486, 80.71]. For the positive uniform prior (which is noninformative), the posterior follows a normal distribution, where the posterior mean (Bayes estimate) of the distribution of heart rate (bpm) was 71.46 while the posterior standard deviation was 4.977.

The posterior means for both the Uniform and Normal priors are almost equal along with their credible intervals being close to the point estimate. But, considering the standard errors, the Normal prior performs slightly better than the Uniform prior. Hence, the normal likelihood is recommended for analyzing Heart Rate (bpm) in the Faculty of Science. *Keywords:*- *Heart Rate (BPM), Uniform Prior, Normal Prior, Credible Interval, Normal Distribution.*

I. INTRODUCTION

One of the most important organs in the human body is the heart. It functions as an impeller by circulating oxygen through the body. It also helps in the circulation of blood throughout the body to keep it running (Forte *et al.*, 2022). This circulated blood as well transports byproducts made by the body to the kidneys (Legrand *et al.*, 2022). Whenever the body exerts itself, the frequency with which the heart beats varies proportionally to the amount of the exerted effort. This rate with which the heart beats can be easily noticed used for a variety of health purposes by detecting the voltage created by its beating (Danilin *et al.*, 2022).

The total number of times that your heart beats per minute is referred to as your heart rate (Catai *et al.*, 2020). Reportedly the heart, just like other muscles in the body, requires physical activity to continue to stay in shape. A normal resting heart rate has been estimated to be between 60 and 100 beats per minute (bpm) (American Heart Association, 2018). This will, however, differ considerably based on when it is measured and what you the person was doing just before the heart rate reading. It will be higher when you walk, for example, than when you sit and rest. This is majorly because the body requires more energy when you are active, when that happens it means that the heart must work harder than normal (Shaffer and Ginsberg, 2017).

Regular exercise has been verified to help improve the overall heart health (de Geus *et al.*, 2019). This regular exercise also helps to improve many several of the identified "risk factors" for cardiovascular disease in many instances (Stone *et al.*, 2021). It is quite critical to know if an individual exercises at the appropriate level required, consistent measurement of the heart rate can help track fitness level (Shaffer and Ginsberg, 2017). The heart rate is then critical in determining an individual's blood pressure because blood pressure is the force of blood against the artery walls as it circulates through the body. Blood pressure is represented by two numbers. The very first (systolic) figure means vascular pressure when the heart is beating. The 2nd (diastolic) count refers to the pressure in the vessels between heartbeats (Forte *et al.*, 2022).

Heart rate (HR) data is by far the most significant and consistent physiological measurement by many COTS wearables as it has a variety of applications and is comprehensible for the vast bulk of end-users (Malik et al., 1996; Chalmers et al., 2022; Oyeleye et al., 2022). HR values indicate general functionality in the human body numerically, with smaller values indicating a body at rest (negligible stressors). Elevated HR values typically reflect higher metabolic requirements or a decline in efficiency improvement during physical effort or stress (e.g., suppressed running economy due to fatigue) (Jachymek et al., 2022). Variations in HR over period seem to be reflective of stress responses, which can be favorable (e.g., enhanced fitness producing lowered HR during exercise) or critical (e.g., poor nutrition and sleep vielding enhanced HR at rest), and thus are frequently seen as a metric for defining general health in healthcare, sporting events, and wellness settings. Furthermore, because cardio - vascular health problems are the major cause of death in the U.s., HR is a strong predictor of mortality (Chalmers et al., 2022; Danilin et al., 2022; Legrand et al., 2022; Nuuttila et al., 2022).

➢ Aim and Objectives

The primary aim of this study to explore the distribution of heart rate by using the Bayesian inferential approach.

The objective of this study then involves:

- To evaluate the distribution of the likelihood function of the observed data.
- To determine a better prior distribution to be used for study into heart rate.
- To estimate the posterior-mean of the heart rate (bpm) and its corresponding variance using the Uniform Prior approach.
- To estimate the posterior-mean of the heart rate (bpm) and its corresponding variance using the Conjugate Prior approach.

II. METHODOLOGY

A total number of 347 students' information was used in the analysis. The data was obtained in the faculty of science, using a face-to-face quantitative methodology. The heart rate (bpm) of students was measured using a standard stethoscope. And the bayesian inference was used to estimate the posterior estimates.

➢ Bayesian Inference

Bayesian inference is an analytical method that utilizes the popular Bayes' theorem to inform our belief in the incidence of certain events; in other words, it utilizes a specific tool known as the Bayes' theorem to revise our conviction considering the evidence (Katsuki, Torii and Inoue, 2012). Bayesian approach allows for direct possibility statements about just the parameter, which are much more helpful than Frequentists' self-belief statements. Bayesian analytical method also has a single tool, Bayes' theorem, that is used in all applications. In contrast to this, the Frequentists procedure necessitates the use of numerous tools for various applications (De Luca, Magnus and Peracchi, 2021). Bayes improvement is a key method in statistics, particularly computational statistics. Almost any inferential question can be easily answered from a bayesian point by performing an appropriate analysis of the posterior distribution. After obtaining the posterior distribution, one could compute parametric point and interval estimates, forecasts, and stochastic hypothesis evaluation (Darbon and Langlois, 2021).

The basic steps involved in carrying out Bayesian analysis are outlined below:

> The Formulation Of Probability Model For The Data Set

The very first phase in Bayesian method of analysis is to select a probabilistic model for the set of data. This step seems to be similar to the traditional method of selecting models for just a data set. If the assumptions are met, a decision has to be made on a probability density function for the data. Assume random selection of size n are observed, i.e. $y_1, y_2, ..., y_n$, as well as the vector of unknown parameters is denoted by θ , and we then choose a probability function for the data set, presuming the observations are independents (Spiegelhalter *et al.*, 2002).

$P(Y_i/\theta)$

In some cases where the dataset is seen to consists of covariate information, x_i for the *ith* observation, as in the traditional regression models, we would then go ahead to choose a probability function of the model form, which is denoted by

 $P(y_i/x_i,\theta)$

Peradventure, the observed data are not independent given the known parameters and the corresponding covariates, we must then go ahead to ensure that we specify the joint probability distribution function of the model

 $P(y_1, y_2, \dots, y_n/x_1, x_2, \dots, x_n, \theta)$

> Decision On The Choice Of The Prior Distribution

As soon as the probability of the observed data has been duly specified, the analytical approach involving the Bayesian methodology usually demands the specification of a certain prior distribution for the obviously unknown population parameters (Ghosh, Li and Mitra, 2018). The parameters of this prior distributions are often referred to as hyper parameters, this is done to easily distinguish them from the parameters of the model. This is denoted by

 $P(\theta)$

> The Types of Prior Distributions

There are two major known types of prior distribution in Bayesian analysis, the non-informative prior and the informative prior (Suh and Kim, 2015).

• Non-informative

A non-informative prior is a prior distribution that allocates equal likelihood to all potential parameter values. A prior distribution is non-informative if it is flat with respect to the likelihood function. This means that a prior distribution is referred to as non-informative if its influence on the prior distribution is quite negligible. This is also known as the vague, diffuse, or flat prior.

• Informative prior

A prior that is influenced by the posterior distribution but is not influenced by the likelihood. It compiles evidence from various sources about the parameters in question.

Constructing The Likelihood Function

Once we have the observed data, we can then go ahead to construct the likelihood function. The likelihood function of the data usually measures the support provided for each possible value of the parameter by the observed data. This means that all the information coming directly from the data about the parameter is adequately contained in the likelihood of the data. Suppose $y_1, y_2, ..., y_n$ are data obtained independently after observation, then, the likelihood function is denoted by;

$$P(\theta/y_i) = L(\theta/y_1, ..., y_n)$$

= $P(y_1, ..., y_n/\theta)$
= $\prod P(y_i/\theta)$

Say we have a series of observation $y_1, y_2, ..., y_n$ which are samples that were obtained from a population with probability density function $P(y_i/\theta)$, where each observation is assumed to be independently and identically distributed (iid). The joint density function for $y_1, y_2, ..., y_n$ is then given as the product of each probability density function.

$$P(y_1, \dots, y_n/\theta) = \prod_{i=1}^n P(y_i/\theta)$$

Obtain The Posterior Distribution

The likelihood function combines information from the data under consideration with information from the prior distribution to produce the posterior distribution. It represents the state of knowledge about the parameter after acquiring new information from measurements. The posterior distribution is represented by the Bayes theorem as

$Posterior \propto Likelihood \times Prior$

Mathematically given as:

$$P(\theta/y_i) \propto P(y_i/\theta) \times P(\theta)$$

$$P(\theta/y_i) = \frac{P(y_i) \times P(\theta)}{P(y_i)}$$

Where,

 $P(y_i)$

Is the evidence (normalization constant useful for selection of Bayesian Model)

For a discrete function

$$P(y_i) = \sum P(y_i/\theta) \times P(\theta)$$

For a continuous function

$$P(y_i) = \int P(y_i/\theta) \times P(\theta)\delta\theta$$

> Posterior Summary

Inferential deductions can thus be summarized with an appropriate analysis once the posterior distribution has been fully determined. The mean or mode of the posterior distribution is commonly used to compute point estimates of parameters. Interval estimates can also be computed.

• Posterior Mean and Variance Using Conjugate Prior

Say the observation y is a random variable obtained from a normal distribution with mean μ and variance σ^2 which is assumed known. Then, we have a prior distribution that is normal with mean \bar{x} and variance s^2 . The shape of this prior density is then illustrated by

$$g(\mu) \propto e^{-\frac{1}{2s^2}(\mu-m)^2}$$

What is happening here is that we are ignoring the part that doesn't involve μ this is majorly because if we multiply the prior by any constant of proportionality will cancel out in the posterior. The shape of the likelihood is then given by

$$f(y|\mu) \propto e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

where we have ignored the part that doesn't depend on the parameter μ because multiplying the likelihood by any constant whatsoever will cancel out in the posterior. The prior times likelihood is

$$g(\mu) \times f(y|\mu) \propto e^{-\frac{1}{2}\left[\frac{(\mu-m)^2}{s^2} + \frac{(y-\mu)^2}{\sigma^2}\right]}$$

Putting all the terms in exponent over the common denominator and expanding them out then gives us

$$\propto e^{-\frac{1}{2}\left[\frac{\sigma^2(\mu^2-2\mu m+m^2)+s^2(y^2-2y\mu+\mu^2)}{\sigma^2s^2}\right]}$$

We go ahead to combine the like terms

$$\propto e^{-\frac{1}{2} \left[\frac{(\sigma^2 + s^2)\mu^2 - 2(\sigma^2 m + s^2 y)\mu + m^2 \sigma^2 + y^2 s^2}{\sigma^2 s^2} \right]}$$

and factor out $\sigma^2 + s^2/\sigma^2 s^2$. Completing the square and absorbing the part that doesn't depend on μ into the proportionality constant, we have

$$\propto e^{-\frac{1}{2\sigma^2 s^2/(\sigma^2+s^2)} \left[\mu^2 - 2\frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2} \mu + (\frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2})^2\right]} \\ \propto e^{-\frac{1}{2\sigma^2 s^2/(\sigma^2+s^2)} \left[\mu - \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2}\right]^2}$$

We recognize from this shape that the posterior is a normal distribution having mean and variance given by

$$m' = \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2}$$
 and $(s')^2 = \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2}$

respectively. We started with a *normal* (m, s^2) prior, and ended up with a *normal* $[m', (s')^2]$ posterior. This shows that the *normal* (m, s^2) distribution is the conjugate family for the normal observation distribution with known variance.

• Credible Interval

A credible interval is an interval that contains the probability of the parameter. It is an interval in the domain of a posterior probability distribution used for interval estimation.

$$= E(\theta/y) \pm Z_{\frac{\alpha}{2}}SD(\theta/y)$$

Where,

 $E(\theta/y) = Posterior Mean,$ $SD(\theta/y) = Posterior Standard Deviation$ And,

$$Z_{\frac{\alpha}{2}} = z - value$$

III. RESULTS AND DISCUSSION

Table 1: Summary statistics of data collected on Heart Rate

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
HEART RATE	347	58	87	71.46	4.977	24.769
N	347					

Therefore, $\mu = 71.46$ And,

 $\sigma = 4.977$

The mean and standard deviation of the data are 71.46 and 4.977 respectively.

> Testing For The Distribution Heart Rate (Bpm) Follows H_0 = The distribution of the heart rate (bpm) follows normal distribution

 H_1 = The distribution of the heart rate (bpm) does not follow normal distribution.

The average heart rate (bpm) in the area under study is = 71.46

The probability mass function for a normal distribution is given as:

$$P(Y = y | \mu) = e^{-\frac{(y-\mu)^2}{2\sigma^2}}, y \ge 0$$

Table 2 Showing the outcome of the Kolmogorov – Smirnov's
test

Kolmogorov-Smirnov								
Sample Size	347							
Statistic	0.05687							
P-Value	0.20377							
Rank	1							
A	0.2	0.1	0.05	0.02	0.01			
Critical Value	0.0576	0.06565	0.0729	0.08149	0.08745			
Reject?	No	No	No	No	No			

From figure 2 above, it was observed that normal distribution was rank the best among other distribution that was in line with the data and was accepted at 80%, 90%, 95%, 98% and 99% Confidence Interval.

> Conclusion

We do not reject the null hypothesis and thereby conclude that the data of the distribution of heart rate (bpm) of faculty of science students follows Normal Distribution.

> Bayesian Analysis

A non-informative and informative prior will be used to ascertain the average heart rate of students in faculty of science, university of Ibadan.

Construction Of The Likelihood

Since there are 347 observations and each of the it is independently and identically distributed, the joint density function is the product of the individual p.d.f's. Therefore, the likelihood follows a normal distribution (as shown in section 4.2). Let y_i , i = 1, 2, ..., 347 be the number of observation.

$$P(y_1, y_2, \dots, y_{347} | \theta) = \prod_{\substack{i=1 \ i=1}}^{n} P(y_i | \theta).$$

\$\approx (2\pi \sigma^2)^{-n/2} e^{-\frac{\sigma(x-\mu)^2}{2\sigma^2}}\$

Posterior Mean And Standard Deviation For Conjugate Prior

Using the conjugate prior of a Normal Distribution which is also Normal Distribution with a mean and variance parameters. It is given as:

$$f(y|\mu) = e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

The Posterior Mean (Bayes Estimate) of y is

$$E(\theta|y) = m' = \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2} = 71.55$$

Where Posterior Variance is

$$V(\theta|y) = (s')^2 = \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2} = 41.62$$

Therefore, the **Posterior Standard Deviation** is = 4.65

The credible interval = Posterior mean of point estimate $\pm Z_{\alpha/2} \sqrt{(Posterior Precision)}$

$$CI(\theta/y) = E(\theta/y) \pm Z_{\frac{\alpha}{2}}SD(\theta/y)$$

Where,

 $E(\theta/y) = 71.6,$ And,

$$SD(\theta/y) = 4.65$$

 $Z_{\frac{\alpha}{2}} = 1.96$

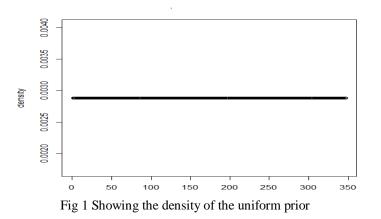
The 95% credible interval for y (heartbeat) = $CI(\theta/y) = 70.9 \pm 1.96$ (4.65) $CI(\theta/y) = 71.6 \pm 9.114$ $CI(\theta/y) = [62.486, 80.71]$

Prior Distribution For Non – Informative Prior

Choosing a positive uniform prior (which is a noninformative prior) in order for a data to have a say, in the sense that we have no idea of what the value of θ might be before making the observation. Therefore, the probability of selecting an individual is

$$P(\theta) = \frac{1}{347} = 0.002, \ \theta \ge 0$$

Since that, we have 347 observations. Therefore, the prior is as follows;



 Posterior Mean And Standard Deviation For Uniform Prior The Bayes Estimate (Posterior mean) of heart rate is;
 E(θ|y_i) = Σ_{i=1}³⁴⁷ θ × P(θ|y_i) = 71.46 And the Posterior Variance is

$$V(\theta|y_i) = E(\theta^2|y_i) - [E(\theta|y_i)]^2$$

Therefore, $V(\theta|y_i) = 24.7009$ Also,

Posterior Standard Deviation =
$$\sqrt{V(\theta|y_i)}$$

= $\sqrt{24.7009}$
= 4.977

IV. SUMMARY

Having used the data containing the heart rate (bpm) of faculty of science students in this research, Bayesian Inference was used to know how much my opinion changes about the distribution of heart rate (bpm), and the effect of age on it.

A total number of 347 students' information was used in the analysis. The data was obtained in the faculty of science, a descriptive analysis of the data collected showed that the mean and standard deviation of the data was 71.46 and 4.977 respectively, it was observed that the minimum value was 58 bpm while the maximum value is 87 bpm, it was observed that the data followed a normal distribution after an histogram chart was plotted. The Kolmogorov smirnov's test also showed that the data followed a normal distribution.

For the Conjugate prior (which is informative), where the prior, the likelihood and the posterior distribution follow the normal distribution, the posterior mean (Bayes estimate) of the distribution of heart rate (bpm) is 71.55 while the posterior standard deviation is 4.65. The posterior credible interval for average heart rate (bpm) is [62.486, 80.71]. For the positive uniform prior (which is non-informative), the posterior follows a normal distribution, where the posterior mean (Bayes estimate) of the distribution of heart rate (bpm) was 71.46 while the posterior standard deviation was 4.977.

V. CONCLUSION

This study vividly outlines the basic target of the Bayesian inference approach, this is defined as a composite index of previous knowledge about the parameter (represented by the prior distribution) and knowledge about the parameter heart rate (bpm) contained in the observational data (which is represented by the likelihood function). It was found out that the distribution of heart rate (bpm) in the faculty of science follows a normal distribution, also the average heart rate (bpm) in the faculty of science was obtained to be 71.55, while the standard deviation of the heart rate (bpm) in the faculty of science was obtained to be 4.65 and the credible Interval for posterior mean was obtained to be [62.486, 80.71], also there is no association between gender and distribution of heartbeat (bpm) of faculty of science students.

RECOMMENDATION

Just in tandem with the various analysis the researcher has conducted thus far in this study, the researcher hereby posits the following as recommendations that are made for health practitioners and educators

- For subsequent studies into heart rate, the normal distribution should be maintained as the informative prior for studies involving heart rate (bpm).
- The normal distribution should also be considered as likelihood for studies involving heart rate (bpm).
- Health practitioners should still maintain the standard information which is that the heartbeat follows a normal distribution.

REFERENCES

- [1]. American Heart Association (2018) 'All About Heart Rate (Pulse)', *Heart.Org* [Preprint].
- [2]. Catai, A.M. *et al.* (2020) 'Heart rate variability: are you using it properly? Standardisation checklist of procedures', *Brazilian Journal of Physical Therapy*. doi:10.1016/j.bjpt.2019.02.006.
- [3]. Chalmers, T. *et al.* (2022) 'Stress watch: The use of heart rate and heart rate variability to detect stress: A pilot study using smart watch wearables', *Sensors*, 22(1). doi:10.3390/s22010151.
- [4]. Danilin, L.K. *et al.* (2022) 'Heart rate and heart rate variability in patients with chronic inflammatory joint disease: the role of pain duration and the insular cortex', *BMC Musculoskeletal Disorders*, 23(1). doi:10.1186/s12891-022-05009-1.
- [5]. Darbon, J. and Langlois, G.P. (2021) 'On Bayesian Posterior Mean Estimators in Imaging Sciences and Hamilton–Jacobi Partial Differential Equations', *Journal* of Mathematical Imaging and Vision, 63(7). doi:10.1007/s10851-021-01036-0.

- [6]. Forte, G. *et al.* (2022) 'Heart Rate Variability and Pain: A Systematic Review', *Brain Sciences*. doi:10.3390/brainsci12020153.
- [7]. de Geus, E.J.C. *et al.* (2019) 'Should heart rate variability be "corrected" for heart rate? Biological, quantitative, and interpretive considerations', *Psychophysiology*. doi:10.1111/psyp.13287.
- [8]. Ghosh, J., Li, Y. and Mitra, R. (2018) 'On the use of Cauchy prior distributions for Bayesian logistic regression', *Bayesian Analysis*, 13(2). doi:10.1214/17-BA1051.
- [9]. Jachymek, M. *et al.* (2022) 'Wristbands in home-based rehabilitation—Validation of heart rate measurement', *Sensors*, 22(1). doi:10.3390/s22010060.
- [10]. Katsuki, T., Torii, A. and Inoue, M. (2012) 'Posteriormean super-resolution with a causal gaussian markov random field prior', *IEEE Transactions on Image Processing*, 21(7). doi:10.1109/TIP.2012.2189578.
- [11]. Legrand, N. *et al.* (2022) 'The heart rate discrimination task: A psychophysical method to estimate the accuracy and precision of interoceptive beliefs', *Biological Psychology*, 168. doi:10.1016/j.biopsycho.2021.108239.
- [12]. De Luca, G., Magnus, J.R. and Peracchi, F. (2021)
 'Sampling properties of the Bayesian posterior mean with an application to WALS estimation', *Journal of Econometrics* [Preprint]. doi:10.1016/j.jeconom.2021.04.008.
- [13]. Malik, M. *et al.* (1996) 'Heart rate variability: Standards of measurement, physiological interpretation, and clinical use', *Circulation*, 93(5). doi:10.1161/01.cir.93.5.1043.
- [14]. Nuuttila, O.P. *et al.* (2022) 'Validity of the wrist-worn polar vantage v2 to measure heart rate and heart rate variability at rest', *Sensors*, 22(1). doi:10.3390/s22010137.
- [15]. Oyeleye, M. et al. (2022) 'A Predictive Analysis of Heart Rates Using Machine Learning Techniques', International Journal of Environmental Research and Public Health, 19(4). doi:10.3390/ijerph19042417.
- [16]. Shaffer, F. and Ginsberg, J.P. (2017) 'An Overview of Heart Rate Variability Metrics and Norms', *Frontiers in Public Health*. doi:10.3389/fpubh.2017.00258.
- [17]. Spiegelhalter, D.J. et al. (2002) 'Bayesian measures of model complexity and fit', Journal of the Royal Statistical Society. Series B: Statistical Methodology, 64(4). doi:10.1111/1467-9868.00353.
- [18]. Stone, J.D. *et al.* (2021) 'Assessing the Accuracy of Popular Commercial Technologies That Measure Resting Heart Rate and Heart Rate Variability', *Frontiers in Sports and Active Living*, 3. doi:10.3389/fspor.2021.585870.
- [19]. Suh, Y. and Kim, H. (2015) 'Probabilistic Class Histogram Equalization Based on Posterior Mean Estimation for Robust Speech Recognition', *IEEE Signal Processing Letters*, 22(12). doi:10.1109/LSP.2015.2490202.