

## МЕТОД АНАЛИЗА ЗАВИСИМОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН ПРИ МАЛОМ ЧИСЛЕ НАБЛЮДЕНИЙ

© 2022 А. И. Михальский<sup>1</sup>, Я. А. Туровский<sup>1,2</sup>, С. В. Борзунов<sup>✉2</sup>

<sup>1</sup>Институт проблем управления им. В. А. Трапезникова РАН  
ул. Профсоюзная, 65, 117997 Москва, Российская Федерация

<sup>2</sup>Воронежский государственный университет  
Университетская пл., 1, 394018 Воронеж, Российская Федерация

**Аннотация.** При нелинейном характере зависимости случайных величин использование коэффициента корреляции может приводить к неверным выводам. Более универсальной характеристикой зависимости является взаимная информация, не требующая предположения о линейности связи случайных величин. Ее можно интерпретировать как среднее количество информации относительно случайной величины, содержащейся в распределении некоторой другой случайной величины. В статье описан метод непараметрической оценки величины взаимной информации по эмпирическим данным, заключающийся в минимизации регуляризованного квадратичного функционала в гильбертовом пространстве с воспроизводящим ядром. Используется замена скалярного произведения в обычном гильбертовом пространстве на значение неотрицательно определенной функции двух переменных. Аналитическим решением является линейная комбинация значений воспроизводящего ядра с коэффициентами, вычисляемыми как решение регуляризованной системы линейных уравнений. Приведены результаты оценки связи между вариабельностью сердечного ритма и успешностью работы оператора системы управления через взаимную информацию и через коэффициент корреляции по данным реального эксперимента. На примере выяснения зависимости успешности выполнения задания оператором системы управления от вариабельности сердечного ритма на малых интервалах времени показана эффективность использования предложенной оценки взаимной информации.

**Ключевые слова:** зависимые случайные величины, оценка взаимной информации, гильбертово пространство с воспроизводящим ядром, регуляризация, вариабельность сердечного ритма и успешность выполнения задания оператором.

### ВВЕДЕНИЕ

Вопрос о поиске зависимостей между случайными величинами важен и сложен как с теоретической, так и с практической стороны. Оценка взаимосвязи случайных величин является одной из центральных задач анализа данных. Неправильный учёт наличия статистической зависимости между факторами приводит к плохой обусловленности матрицы плана при построении регрессионных зависимостей [1] и к получению статистически

ненадежных оценок влияния рассматриваемых факторов. При этом часто делается неверный вывод о характере существующей зависимости. Широко используемая мера связи случайных величин — коэффициент корреляции эффективен при наличии линейной зависимости между изучаемыми величинами. В случае нелинейной зависимости коэффициент корреляции дает неверное значение и даже может ошибочно свидетельствовать об отсутствии зависимости.

Другой характеристикой зависимости случайных величин является величина взаимной информации, не требующая предположения о линейности связи случайных вели-

✉ Борзунов Сергей Викторович  
e-mail: [sborzunov@gmail.com](mailto:sborzunov@gmail.com)



Контент доступен под лицензией Creative Commons Attribution 4.0 License.

The content is available under Creative Commons Attribution 4.0 License.

чин, которую можно интерпретировать как среднее количество информации относительно случайной величины  $X$ , содержащейся в распределении случайной величины  $Y$  [2].

В настоящей работе рассматривается метод построения непараметрической оценки величины взаимной информации по малому числу наблюдений и приводятся результаты оценки связи между вариабельностью сердечного ритма (ВСР) и успешностью работы оператора системы управления

## 1. МЕТОДЫ ИССЛЕДОВАНИЯ

Оценка взаимной информации ищется как точка минимума регуляризованного эмпирического квадратичного функционала в гильбертовом пространстве с воспроизводящим ядром (RKHS). Аналитическим решением этой задачи является линейная комбинация значений воспроизводящего ядра [3, 4] с коэффициентами, вычисляемыми как решение регуляризованной системы линейных уравнений. Приведены результаты оценки связи между вариабельностью сердечного ритма и успешностью работы оператора системы управления через взаимную информацию и через коэффициент корреляции по данным реального эксперимента.

### 1.1. Метод оценки взаимной информации

Формально величина взаимной информации между случайными величинами  $Z$  и  $T$ , имеющими совместное распределение  $P(z, t)$  с плотностью распределения  $p(z, t)$  определяется соотношением

$$I(Z, T) = \int p(z, t) \ln \frac{p(z, t)}{p(z)p(t)} dz dt.$$

Из определения следует, что для независимых случайных величин взаимная информация равна нулю, поскольку справедливо соотношение  $p(z, t) = p(z)p(t)$  и логарифм под интегралом равен нулю.

Простой метод оценки взаимной информации состоит в замене интегрирования усреднением по  $k$  независимым выборочным значениям случайной величины  $z \in \{z_1, \dots, z_k\}$

и  $m$  независимым выборочным значениям случайной величины  $t \in \{t_1, \dots, t_m\}$

$$\hat{I}(Z, T) = \frac{1}{km} \sum_{i=1}^k \sum_{j=1}^m p(z_i, t_j) \ln \frac{p(z_i, t_j)}{p(z_i)p(t_j)}.$$

Недостатком такого подхода является необходимость прямого использования плотностей распределения  $p(z, t)$ ,  $p(z)$  и  $p(t)$ , для достаточно точной оценки которых необходимо большое число экспериментальных данных.

В [5] задачу оценки взаимной информации  $I(Z, T)$  авторы рассматривают как задачу минимизации квадратичного функционала

$$J(\hat{w}) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (\hat{w}(z_i, t_j) - 1)^2, \quad (1)$$

где  $\hat{w}(z, t) = p(z, t) / p(z) / p(t)$ . Преимущества такого подхода в том, что задача оценивания взаимной информации сводится к минимизации квадратичного функционала, а оценка взаимной информации вычисляется по формуле

$$\hat{I}(Z, T) = \frac{1}{n} \sum_{i=1}^n \ln \hat{w}(z_i, t_i).$$

### 1.2. Минимизация эмпирического функционала в RKHS

Минимизация функционала (1) требует параметрического задания вида функции  $\hat{w}(z, t)$ , однако, этого можно избежать, если искать оценку взаимной информации в бесконечномерном гильбертовом пространстве с воспроизводящим ядром [6]. Смысл рассмотрения таких пространств заключается в замене скалярного произведения в обычном гильбертовом пространстве на значение неотрицательно определённой функции  $K(s, r)$  двух переменных — элементах RKHS. По определению неотрицательно определённой функцией называют симметричную непрерывную функцию двух переменных  $K(s, r)$  такую, что для любой последовательности  $t_1, \dots, t_n$  и любых чисел  $a_1, \dots, a_n$  справедливо соотношение

$$\sum_{j=1}^n \sum_{i=1}^n a_i a_j K(t_i, t_j) \geq 0.$$

Доказана теорема, что используя неотрицательно определённую функцию в качестве

ядра можно построить гильбертово пространство с воспроизводящим ядром (теорема Мерсера). Другая теорема, обобщённая теорема о представителе [3], позволяет найти оценку взаимной информации в непараметрическом виде без необходимости параметрического задания вида функции  $\hat{w}(z, t)$ . В силу этой теоремы минимум функционала

$$J(\hat{w}, \lambda) = J_e(\hat{w}) + \frac{\lambda}{2} \|\hat{w}\|_K^2 \quad (2)$$

в гильбертовом пространстве с воспроизводящим ядром  $K$  достигается на функции вида

$$\hat{w}(z, t) = \sum_{i=1}^n a_i K((z, t), (z_i, t_i)). \quad (3)$$

В (2)  $J_e(\hat{w})$  обозначает эмпирический функционал, величина  $\|\hat{w}\|_K^2$  — квадрат нормы решения в RKHS с ядром  $K$ ,  $\lambda$  — параметр регуляризации. Коэффициенты  $a_1, \dots, a_n$  в (3) вычисляются в процессе минимизации функционала (2). Если эмпирический функционал  $J_e(\hat{w})$  квадратичный, то коэффициенты в (3) вычисляются аналитически.

### 1.3. Непараметрическая оценка взаимной информации

Чтобы построить непараметрическую оценку взаимной информации, введём обозначение  $\hat{w}(z, t) = p(z, t) / p(z) / p(t)$ , рассмотрим квадратичный функционал с регуляризатором

$$J_\lambda(\hat{w}) = \frac{1}{2} \iint (w(z, t) - \hat{w}(z, t))^2 \times \\ \times p(z) p(t) dz dt + \frac{\lambda}{2} \|\hat{w}\|_K^2$$

и его представление при замене интегрирования суммированием по элементам экспериментальных данных — эмпирический риск

$$J_\lambda^e(\hat{w}) = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \hat{w}^2(z_i, t_j) - \\ - \frac{1}{n} \sum_{i=1}^n \hat{w}(z_i, t_i) + \frac{\lambda}{2} \|\hat{w}\|_K^2 + C. \quad (4)$$

Слагаемое  $C$  обозначает независящий от искомого решения член. Минимум функционала (4) ищется путём решения регуляризованной системы линейных уравнений [4]

$$(\mathbf{H} + \lambda \mathbf{K}) \mathbf{a} = \mathbf{h},$$

где элементы матрицы  $\mathbf{K}$  вычисляются по формуле

$$K_{ij} = K(z_i, t_i, z_j, t_j),$$

элементы матрицы  $\mathbf{H}$  вычисляются по формуле

$$H_{lm} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(z_i, t_j, z_l, t_l) K(z_i, t_j, z_m, t_m),$$

$\mathbf{h}$  — вектор с элементами

$$h_i = \frac{1}{n} \sum_{j=1}^n K(x_i, y_i, x_j, y_j).$$

Величина  $\lambda$  является параметром метода и выбирается методом cross-validation [4].

## 2. Анализ связи коротких записей ВСР с эффективностью работы оператора робототехнической системы

Среди методов неинвазивной диагностики функционального состояния человека и, в частности, сердечно-сосудистой системы широкое распространение получил метод анализа variability сердечного ритма [7]. Основанный на анализе изменения R-R интервалов на электрокардиограмме пациента этот метод позволяет количественно оценить влияние каждого из звеньев регуляции сердечного ритма на синусовый узел сердца, который определяет частоту сердечных сокращений (ЧСС). Многочисленные исследования доказали высокую прогностическую значимость показателей ВСР в определении риска сердечно-сосудистых осложнений и внезапной кардиальной смерти [8–11]. Ведутся исследования связи ВСР с состоянием как больных [12, 13], так и здоровых людей, находящихся под влиянием стресса: пилоты, машинисты локомотивных бригад, студенты и даже полицейские [14–17].

ВСР используются при профессиональном отборе операторов АЭС [18], оценке функционального состояния операторов систем управления [19, 20]. С появлением человеко-машинных интерфейсов возникли новые задачи по использованию ВСР в работе оператора. Поскольку ВСР отражает эмоциональное состояние оператора, уровень

концентрации внимания, то по локальным изменениям ВСР можно судить о корректности действий и команд, даваемых оператором управляющему устройству. Существенной особенностью такой постановки задачи является то, что для получения надёжных оценок статистических свойств ВСЗ необходимо использовать записи ЭКГ не менее 5–10 минут [7], а для оперативного управления роботизированным устройством требуется принятие решения в течение более коротких интервалов времени.

Эффективность оценки зависимости случайных величин при малом числе наблюдений с помощью взаимной информации продемонстрирована путём исследования связи характеристик ВСЗ, записанных на интервалах длительности 0,5 минут, с эффективностью работы оператора при управлении самоходным шасси с помощью окулографического интерфейса [21].

### **2.1. Организация эксперимента по управлению робототехнической системы с помощью окулографического интерфейса**

Окулографический интерфейс основан на бесконтактной регистрации движения глаз. Отслеживая движение глаз, устройство определяет на каких объектах сфокусировано внимание пользователя, положение зрачка и, анализируя паттерны этих движений, формирует команды по выполнению необходимых действий при реабилитации моторных и когнитивных нарушений различной этиологии; управлении экзоскелетными решениями; управлении подвижными объектами; обеспечении взаимодействия пользователей с электронными и электронно-механическими устройствами [21].

Основной частью интерфейса является аппаратная часть с камерой, которая, детектируя положение зрачка, передает изображение в программную часть, где оно сначала упрощается, а затем по упрощенному изображению происходит распознавание координаты направления взгляда.

До начала управления производится определение базисных точек направления глаза оператора — калибровка. И в процессе работы, после получения координат расположения глаза — они сравниваются с базисом для определения команды, которую программная часть должна передать робототехнической системе. Именно на этапе определения команды возможно программное улучшение качества управления с помощью привлечения ВСР, зарегистрированных на коротких интервалах времени.

В экспериментах, проведённых в лаборатории медицинской кибернетики ФКН (CS) ВГУ, приняли участие 40 добровольцев: 22 юноши и 18 девушек в возрасте от 19 до 23 лет. Испытуемые не имели неврологической и психиатрической патологии, не принимали лекарств, влияющих на координацию движений и скорость принятия решения, имели нормальное или скорректированное до нормального зрение. Испытуемые находились перед экраном монитора (22") на комфортном для себя расстоянии. Видеоокулографический интерфейс позволял на основе стандартной видеокамеры с извлеченным инфракрасным (ИК) фильтром получить более четкое по сравнению с камерой с оставленным ИК фильтром изображение области глаза. После проведения калибровки на экране монитора запускалось экспериментальное поле, представляющее собой область старта, прямоугольные границы ряда объектов и область цели. В инструкции, которая в устной и письменной форме излагалась испытуемому, запрещалось совмещать маркер с объектами, обозначенными черными прямоугольниками на экране. Совмещение маркера с таким объектом расценивалось как ошибка управления. Программа фиксировала и распознавала движение зрачка глаза в четырех направлениях: «вверх», «вниз», «влево», «вправо». Каждому испытуемому было предоставлено четыре попытки для прохождения эксперимента, длительностью 2 мин. Каждая новая команда могла генерироваться пользователем не чаще 1 раза в секунду, при регистрации положения зрачка 25 раз в секунду. Перед каждым экс-



периментом проводилась необходимая калибровка устройства [22].

Регистрация значений ВСП производилась по ЭКГ прибором «Поли-Спектр 12» (производство ООО «Нейрософт») при частоте дискретизации 1 кГц и включенном режекторном фильтре и фильтре дрейфа изолинии. Электроды фиксировались на конечностях испытуемого согласно стандартной схеме [22]. Для дальнейшего анализа запись ВСП, совпадающая по времени с временем проведения эксперимента, разделялась на четыре непересекающихся участка по 0,5 минут, для которых рассчитывались оценки базовых статистических характеристик: среднего значения и среднеквадратического отклонения. Эффективность работы оператора характеризовалась бинарным признаком 0 — ошибка при выполнении задания и 1 — задание выполнено без ошибок. Попытки выполнения задания, при которых были зафиксированы записи ВСП короче 2 минут, не рассматривались.

## 2.2. Результаты эксперимента

Анализируемые данные о статистической связи характеристик ВСП, записанных на интервалах длиной 0,5 минут и показателя успешности выполнения оператором задания с помощью окулографического интерфейса записаны в таблицу, содержащую 9 колонок и 32 строки. Число строк соответствует числу экспериментов, в которых содержатся записи ВСП длиной не менее 2 минут. В первых четырёх колонках содержатся средние значения ВСП в четырёх последовательных интервалах по 0,5 минут, записанные во время выполнения задания. В следующих четырёх колонках содержатся среднеквадратичные значения ВСП, записанные на этих же интервалах. Последний девятый столбец содержит признак успешности выполнения задания оператором: 0 — при выполнении задания допущена ошибка, 1 — задание выполнено без ошибок.

На рис. 1 представлены характеристики разброса среднего значения ВСП в исходных

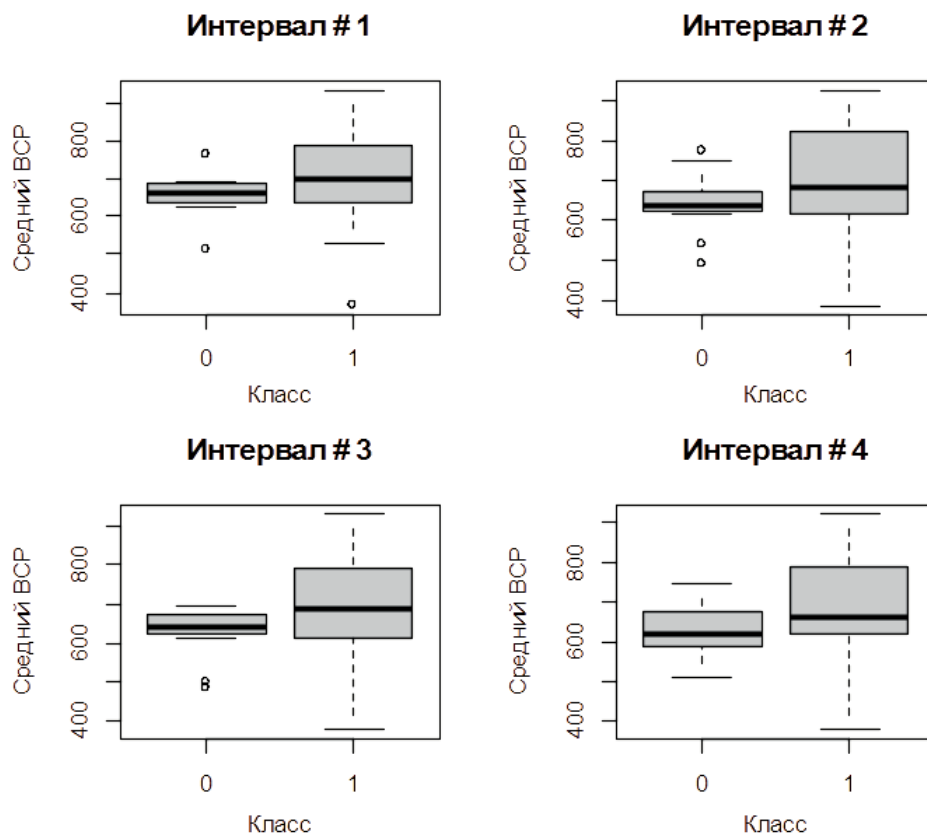


Рис. 1. Сравнительные характеристики распределения среднего значения ВСП в зависимости от порядкового номера интервала измерения и успешности выполнения задания оператором [Fig. 1. Comparative characteristics of the distribution of the average HRV value depending on the ordinal number of the measurement interval and the success of the task by the operator]

данных для каждого из четырёх интервалов. Аналогичные характеристики для среднеквадратичных значений ВСР в исходных данных для каждого из четырёх интервалов представлены на рис. 2.

Из рисунков видно, что для всех интервалов измерения области между нижним и верхним квартилями, на рисунках изображены заштрихованными прямоугольниками, в двух классах перекрываются. Это означает, что прямое использование среднего значения ВСР, как и его среднеквадратического отклонения, непригодно для определения успешности выполнения задания оператором.

В табл. 1 приведены результаты оценки коэффициента корреляции между характеристикой успешности выполнения задания оператором и средним значением и среднеквадратическим отклонением ВСР на различных интервалах измерения. Там же приведены соответствующие значения  $p$ -value. Из таблицы

следует, что для всех интервалов наблюдения ни среднее значение ВСР, ни его среднеквадратическое отклонение не коррелирует с успешностью выполнения задания оператором, поскольку значения  $p$ -value, характеризующие достоверность гипотезы об отсутствии корреляции, не принимают значения ниже 0.1.

Иная картина складывается, если сравнить распределения признаков в двух классах. На рис. 3 представлены эмпирические оценки функции распределения среднего значения ВСР для каждого из четырёх интервалов и каждого класса.

Из рисунка видно, что распределение среднего значения ВСР различно в двух классах. Если среднее значение ВСР ниже некоторого порога, то распределения в классах близки, а если выше, то различаются существенно. Видимо, в этом и заключается природа нелинейности в зависимости успешности управления от среднего значения ВСР.

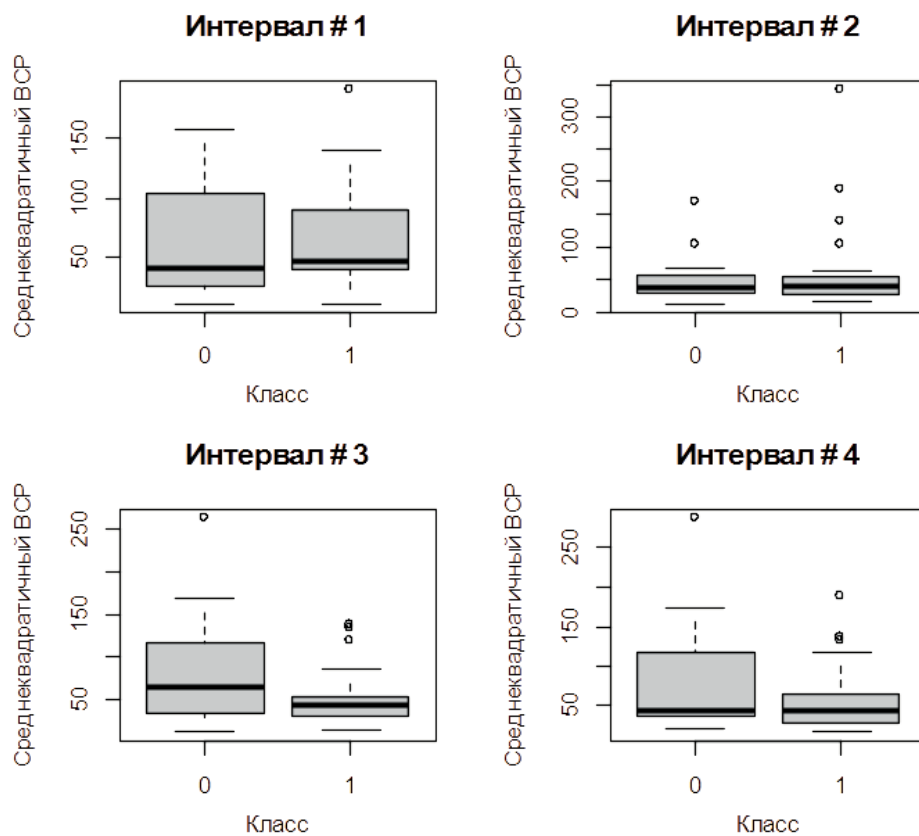


Рис. 2. Сравнительные характеристики распределения среднеквадратического отклонения ВСР в зависимости от порядкового номера интервала измерения и успешности выполнения задания оператором

[Fig. 2. Comparative characteristics of the distribution of the standard deviation of HRV depending on the ordinal number of the measurement interval and the success of the task by the operator]

Таблица 1. Оценки коэффициентов корреляции и  $p$ -value между успешности выполнения задания оператором и средним значением или среднеквадратическим отклонением ВСР на различных интервалах измерения

[Table 1. Estimates of correlation coefficients and  $p$ -value between the success of the task by the operator and the mean value or standard deviation of HRV at various measurement intervals]

Признак	Среднее значение ВСР		Среднеквадратическое отклонение ВСР	
	Коэффициент корреляции	$p$ -value	Коэффициент корреляции	$p$ -value
1	0.18	0.31	0.01	0.95
2	0.24	0.18	0.08	0.68
3	0.26	0.16	-0.29	0.10
4	0.22	0.23	-0.22	0.22

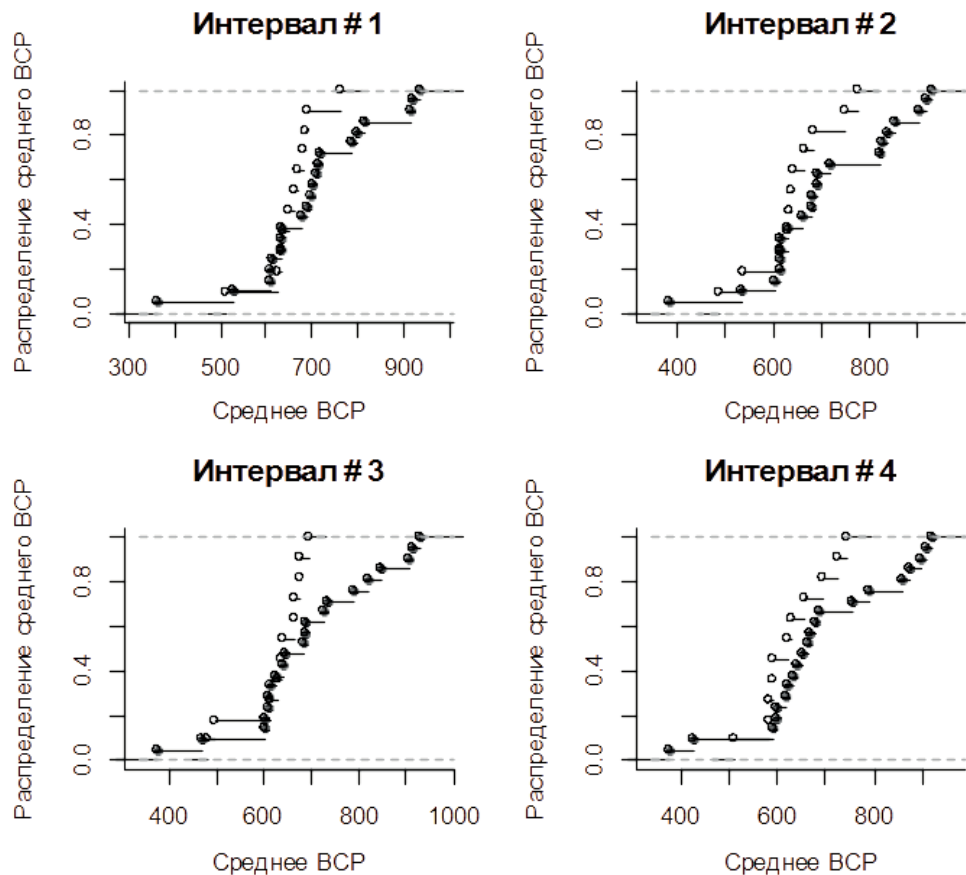


Рис. 3. Эмпирические оценки функции распределения среднего значения ВСР в классе 0 (открытые символы) и в классе 1 (чёрные символы) для записей на четырёх интервалах [Fig. 3. Empirical estimates of the distribution function of the mean HRV value in class 0 (open symbols) and in class 1 (black symbols) for entries at four intervals]

Эмпирические оценки функции распределения среднеквадратического значения ВСР для каждого из четырёх интервалов и каждого класса показаны на рис. 4.

В отличие от распределения среднего значения ВСР, распределения среднеквадра-

точного значения в классах 0 и 1 совпадают только для интервалов записи ВСР номер 1 и 2. Для интервалов номер 3 и 4 распределения различаются если среднеквадратическое значение ВСР превышает некоторый порог.

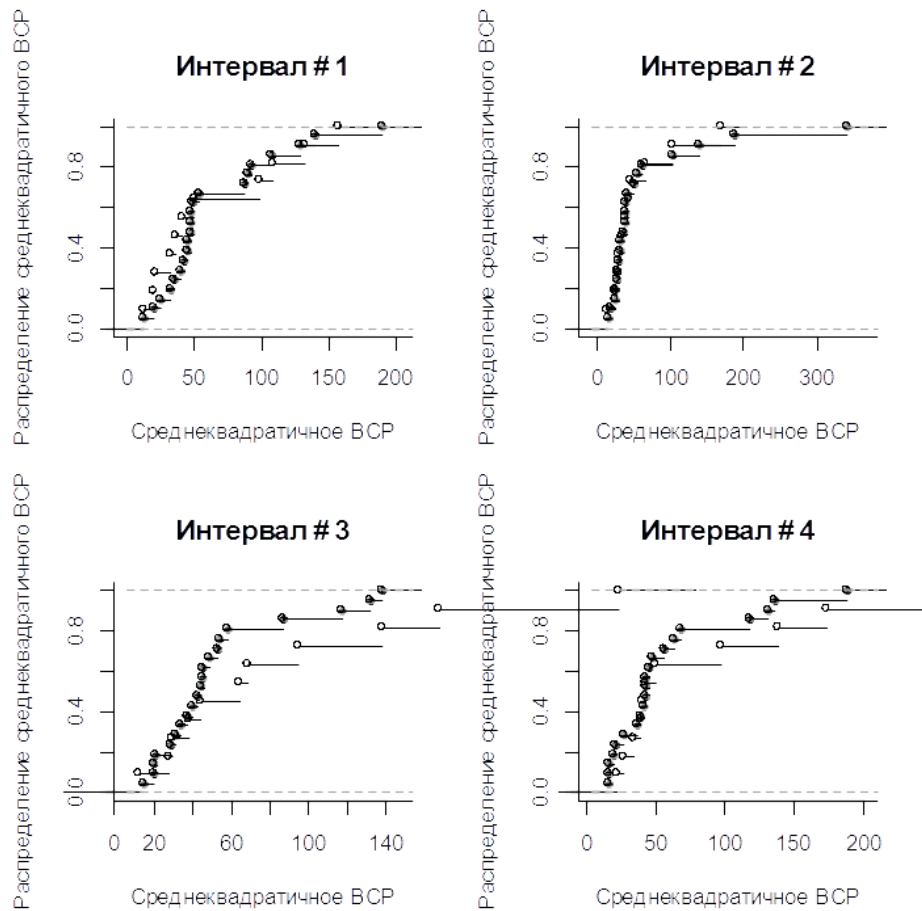


Рис. 4. Эмпирические оценки функции распределения среднеквадратичного значения ВСР в классе 0 (открытые символы) и в классе 1 (чёрные символы) для записей на четырёх интервалах [Fig. 4. Empirical estimates of the distribution function of the root-mean-square value of HRV in class 0 (open symbols) and in class 1 (black symbols) for entries at on four intervals]

### 2.3 Результаты оценки взаимной информации

Для количественной оценки статистической связи ВСР, записанной на коротких интервалах времени, с успешностью выполнения задания оператором вычислялись непараметрические оценки величины взаимной информации между средним значением ВСР на интервале и меткой успешности выполнения задания оператором. Вычисления проводились по описанной выше методике в RKHS. При вычислениях использовалось экспоненциальное ядро вида

$$K(x_i, y_i, x_j, y_j) = \exp\left[-\gamma\left(\frac{(x_i - x_j)^2}{100} + (y_i - y_j)^2\right)\right],$$

где  $x_i$  и  $x_j$  — средние значения ВСР на интервале для операторов номер  $i$  и  $j$ , а  $y_i$  и  $y_j$  —

признаки успешности выполнения задания этими операторами, принимающие значение 0 или 1. Параметр  $\gamma$  является параметром метода, определяется методом cross-validation. Данные подвергались предобработке, в результате которой значений ВСР ниже 500 не рассматривались

Для повышения точности изучаемых оценок использовалась бутстреп-процедура, заключающаяся в генерировании случайным образом дополнительных данных [23]. При каждой генерации вычислялась оценка взаимной информации. По полученному набору оценок строились гистограммы. На рис. 5 приведены гистограммы, построенные по результатам оценки взаимной информации при числе генераций в бутстреп-процедуре, равном 50. В целях нормировки оценки взаимной информации каждый раз нормировались на величину энтропии распределения при-



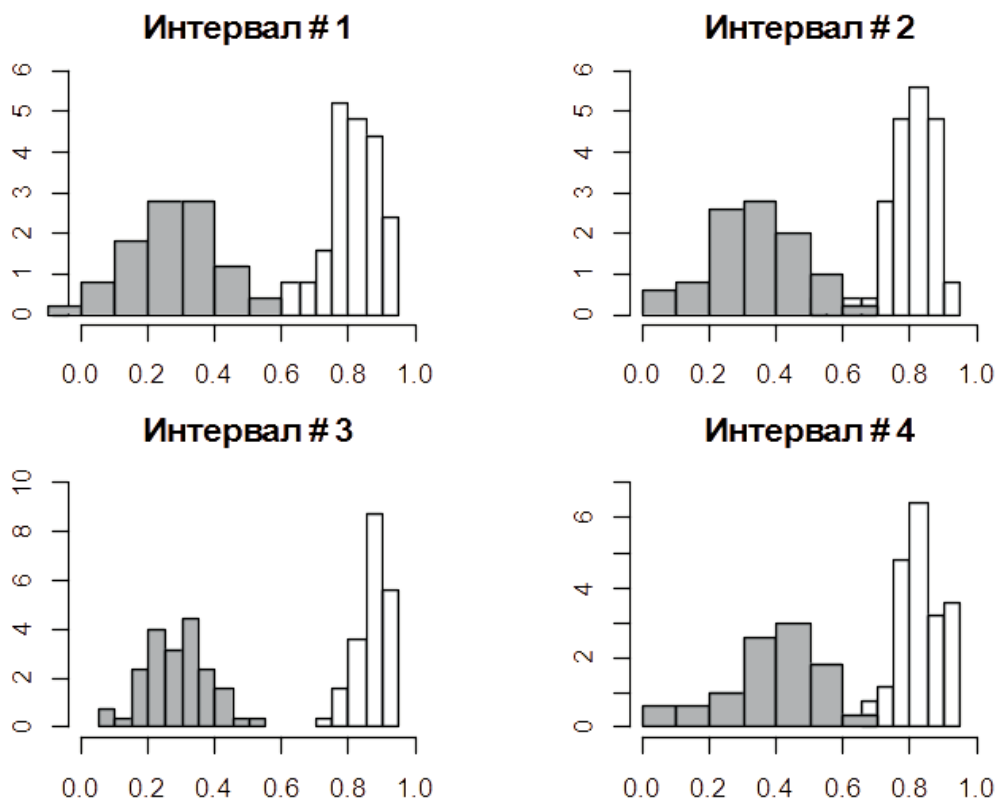


Рис. 5. Гистограммы распределения нормированной оценки взаимной информации (белые) и коэффициента корреляции (серые), для четырёх интервалов записи ЭКГ [Fig. 5. Distribution histograms of normalized assessment of mutual information (white) and correlation coefficient (gray), for four EEG recording intervals]

знака успешности выполнения задания у. Одновременно для сравнения рассчитывался и коэффициент корреляции. Гистограммы распределений полученных значений приведены также на рис. 5.

Из рисунка видно, что для всех интервалов нормированная оценка взаимной информации удалена от нуля дальше, чем коэффициент корреляции, что говорит о наличии нелинейной зависимости, которую коэффициент корреляции в данном эксперименте определяет с низкой надёжностью.

### ЗАКЛЮЧЕНИЕ

Представленные в статье результаты демонстрируют эффективность использования непараметрической оценки величины взаимной информации при выяснении наличия статистической связи между двумя случайными величинами при малом числе наблюдений. Графики, представленные на рис. 3, ясно демонстрируют различие распределений

среднего значения ВСП в двух классах. В то же время, вычисление величины коэффициента корреляции, представленное в табл. 1 никакой зависимости не обнаруживает.

На рис. 5 приведены гистограммы распределений оценок взаимной информации и коэффициента корреляции, построенные по 50 реализациям, сгенерированным процедурой бутстреп при значениях ВСП не ниже 500. Коэффициент корреляции увеличился, но его доверительный интервал оказался существенно шире доверительного интервала для нормированной величины взаимной информации. Это явилось следствием наличия регуляризирующего компонента в алгоритме оценки, а нормированная величина взаимной информации оказалась близка к единице.

### БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке РФФИ, проекты № 19-07-01037 А, 19-29-01156 мк.

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

## СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Исследование зависимостей. – М. : Финансы и статистика, 1985. – 487 с.
2. Габидулин Э. М., Пилипчук Н. И. Лекции по теории информации. – М. : МФТИ, 2007. – 214 с.
3. Scholkopf B., Herbrich R., Smola A. J. A generalized representer theorem // Proceedings of the 14th Annual Conference on Computational Learning Theory and 5th European Conference on Computational Learning Theory, COLT '01/ EuroCOLT '01. – 2001. – P. 416–426.
4. Mikhalskii A. I., Petrov I. V., Tsurko V. V., Anashkina A. A., Nekrasov A. N. Application of mutual information estimation for predicting the structural stability of pentapeptides // RUSS J NUMER ANAL M. – 2020. – V. 35, No. 5. – P. 263–271.
5. Suzuki T., Sugiyama M., Kanamori T., Sese J. Mutual information estimation reveals global associations between stimuli and biological processes // BMC Bioinformatics. – 2009. – V.10 (Suppl 1):S52.
6. Manton J. H., Amblard P. O. A Primer on Reproducing Kernel Hilbert Spaces // Foundations and Trends in Signal Processing. – 2014. – V.8, Is.1. – P. 1–26.
7. Бабуни И. В., Мириджанян Э. М., Машаех Ю. А. Азбука анализа вариабельности сердечного ритма. – Ставрополь : Принт-мастер, 2002. – 112 с.
8. Баевский Р. М., Иванов Г. Г. Вариабельность сердечного ритма: теоретические аспекты и возможности клинического применения // Ультразвуковая функциональная диагностика. – 2001. – № 3. – С. 108–127.
9. Рябыкина Г. В., Соболев А. В. Вариабельность ритма сердца. – Москва : Оверлей, 2001. – 196 с.
10. Зиен Б., Таратухин Е. О. Возможности методики вариабельности сердечного ритма // Российский кардиологический журнал. – 2011. – № 3. – С. 69–75.
11. Седов А. В., Царегородцев Д. А., Сулимов В. А. Неинвазивные электрофизиологические маркеры высокого риска фатальных событий у больных с дилатационной кардиомиопатией // Российский кардиологический журнал. – 2017. – № 2. – С. 50–61.
12. Александрова М. Р., Лукашенко А. А., Юровская И. И., Артемьева М. С. Роль психоэмоционального стресса в генезе жизнеугрожающих нарушений ритма сердца // Вестник РУДН, серия Медицина. – 2014. – № 2. – С. 35–41.
13. Зуйкова А. А., Красноуцкая О. Н., Ханина Е. А. Изменение биологических ритмов организма человека на фоне стрессового воздействия // Вестник новых медицинских технологий. – 2010. – Т. 17, № 2. – С. 195–197.
14. Меерсон Ф. З., Букина Т. Н., Васильев Л. А. Роль профессионального и социального стресса в возникновении неишемических аритмий у пилотов // Кардиология. – 1991. – № 1. – С. 60–62.
15. Дементьев М. В., Чибисов С. М., Халаби Г., Агарвал Р. К., Ходорович Н. А., Харлицкая Е. В. Десинхронизация сердечно-сосудистой системы у машинистов локомотивных бригад // The Journal of scientific articles «Health and Education Millennium». – 2015. – Vol. 17, № 4. – P. 20–25.
16. Павленко С. И., Ведясова О. А., Кротова И. Г. Связь показателей внешнего дыхания и вариабельности сердечного ритма при умственной нагрузке у студентов с разными хронотипами // Физиология человека. – 2021. – Т. 47, № 2. – С. 45–55.
17. Zefferino R., L'Abbate N., Facciorusso A., Potenza A., Lasalvia M., Nuzzaco A., Di Biase M., Ambrosi L. Assessment of heart rate variability (HRV) as a stress index in an emergency team of urban police // G Ital Med Lav Ergon. – 2003. – Vol. 25, Suppl(3). – P. 167–169.
18. Машин В. А., Машина М. Н., Шмелева И. А. Психофизиологические исследования эмоциональной лабильности операторов АЭС // Вопросы психологии. – 1997. – № 4. – С. 95–103.

19. *Машин В. А.* Психическая нагрузка, психическое напряжение и функциональное состояние операторов систем управления // Вопросы психологии. – 2007. – № 6. – С. 86–96.

20. *Машин В. А.* К вопросу классификации функциональных состояний человека // Экспериментальная психология. – 2011. – Т. 4, № 1. – С. 40–56.

21. *Туровский Я. А., Кургалин С. Д., Алексеев А. В.* Анализ движения глаз человека при управлении самоходным шасси с использованием системы видео-окулографического интерфейса // Сенсорные системы. – 2017. – Т. 31, № 1. – С. 51–58.

22. *Туровский Я. А., Алексеев А. В., Лесных И. Е., Мартыненко Е. В.* Частотно-временные особенности движения глаз при использовании видеоокулографического интерфейса в задачах управления эргатическими системами // Сенсорные системы. – 2021. – Т. 35, № 1. – С. 30–37.

23. *Эфрон Б.* Нетрадиционные методы многомерного статистического анализа. – М.: Финансы и статистика, 1988. – 263 с.

**Михальский Анатолий Иванович** — д-р биол. наук, канд. техн. наук, старший научный сотрудник, главный научный сотрудник ИПУ РАН  
E-mail: ipuran@yandex.ru  
ORCID ID: <https://orcid.org/0000-0003-1998-1315>

**Туровский Ярослав Александрович** — д-р техн. наук, канд. мед. наук, ведущий научный сотрудник ИПУ РАН, зав. лаб. Воронежского государственного университета.  
E-mail: yaroslav\_turovsk@mail.ru  
ORCID ID: <https://orcid.org/0000-0002-5290-885X>

**Борзунов Сергей Викторович** – канд. физ.-мат. наук, доцент кафедры цифровых технологий Воронежского государственного университета.  
E-mail: sborzunov@gmail.com  
ORCID ID: <https://orcid.org/0000-0002-5099-9655>

DOI: <https://doi.org/10.17308/sait/1995-5499/2022/2/25-37>  
Received 10.03.2022  
Accepted 29.06.2022

ISSN 1995-5499

## A METHOD FOR ANALYZING THE DEPENDENCE OF RANDOM VARIABLES WITH A SMALL NUMBER OF OBSERVATIONS

© 2022 A. I. Mikhalski<sup>1</sup>, Ya. A. Turovsky<sup>1,2</sup>, S. V. Borzunov<sup>✉2</sup>

<sup>1</sup>*V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences  
65, Profsoyuznaya Street, 117997 Moscow, Russian Federation*

<sup>2</sup>*Voronezh State University  
1, University Square, 394018 Voronezh, Russian Federation*

**Annotation.** With the nonlinear nature of the dependence of random variables, the use of the correlation coefficient can lead to incorrect conclusions. A more universal characteristic of the dependence is mutual information, which does not require the assumption of the linearity of the relationship of random variables. It can be interpreted as the average amount of information about a random variable contained in the distri-

✉ Borzunov S. V.  
e-mail: sborzunov@gmail.com

bution of some other random variable. The article describes a method for nonparametric estimation of the value of mutual information based on empirical data, which consists in minimizing a regularized quadratic functional in a Hilbert space with a reproducing kernel. The replacement of the scalar product in the standard Hilbert space by the value of a nonnegatively defined function of two variables is used. The analytical solution is a linear combination of the values of the reproducing kernel with coefficients calculated as the solution of a regularized system of linear equations. The results of the evaluation of the relationship between heart rate variability and the success of the operator of the control system using mutual information and using the correlation coefficient according to the data of the real experiment are presented. By the example of finding out the dependence of the success of the task by the operator of the control system on the variability of the heart rhythm on small time intervals, the effectiveness of using the proposed mutual information assessment is shown.

**Keywords:** dependent random variables, mutual information estimation, Hilbert space with reproducing kernel, regularization, heart rate variability and the success of the task by the operator.

## ACKNOWLEDGEMENTS

This work was supported by the RFBR, Grants No 19-07-01037 A, 19-29-01156 МК.

## CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

## REFERENCES

1. Ayvazyan S. A., Enyukov I. S., and Meshalkin L. D. (1985) Issledovanie zavisimostej [The study of dependencies]. *Moscow : Finance and Statistics*. 487 p. (in Russian)
2. Gabidulin E. M. and Pilipchuk N. I. (2007) Lekcii po teorii informacii [Lectures on information theory]. *Moscow : MIPT*. 214 p. (in Russian)
3. Scholkopf B., Herbrich R., and Smola A. J. (2001) A generalized representer theorem. *Proceedings of the 14th Annual Conference on Computational Learning Theory and 5th European Conference on Computational Learning Theory, COLT '01/EuroCOLT '01*. P. 416–426.
4. Mikhalskii A. I., Petrov I. V., Tsurko V. V., Anashkina A. A. and Nekrasov A. N. (2020) Application of mutual information estimation for predicting the structural stability of pentapeptides. *RUSS J NUMER ANAL M*. V. 35, No. 5. P. 263–271.
5. Suzuki T., Sugiyama M., Kanamori T. and Sese J. (2009) Mutual information estimation reveals global associations between stimuli and biological processes. *BMC Bioinformatics*. V. 10 (Suppl 1):S52.
6. Manton J. H. and Amblard P. O. (2014) *A Primer on Reproducing Kernel Hilbert Spaces*. Foundations and Trends in Signal Processing. V. 8, Is. 1. P. 1–26.
7. Babunts I. V., Mirijanyan E. M. and Mashayeh Yu. A. (2012) Azbuka analiza variabel'nosti serdechnogo ritma [Abc of analysis of heart rate variability]. *Stavropol : Print-master*. 112 p. (in Russian)
8. Baevsky R. M. and Ivanov G. G. (2001) Variabel'nost' serdechnogo ritma: teoreticheskie aspekty i vozmozhnosti klinicheskogo primeneniya [Heart rate variability: theoretical aspects and possibilities of clinical application]. *Ultrasound functional diagnostics*. No 3. P. 108–127. (in Russian)
9. Ryabykina G. V., and Sobolev A. V. (2001) Variabel'nost' ritma serdca [Heart rate variability]. *Moscow : Overlay*. 196 p. (in Russian)
10. Ziep B. and Taratukhin E. O. (2011) Vozmozhnosti metodiki variabel'nosti serdechnogo ritma [Possibilities of heart rate variability methodology]. *Russian Journal of Cardiology*. No. 3. P. 69–75. (in Russian)
11. Sedov A. V., Tsaregorodtsev D. A. and Sulimov V. A. (2017) Neinvazivnye elektrofiziologicheskie markery vysokogo riska fatal'nyh sobytij u bol'nyh s dilatacionnoj kardiomiopatiej [Non-invasive electrophysiological markers of high risk of fatal events in patients with dilated cardiomyopathy]. *Russian Journal of Cardiology*. No. 2. P. 50–61. (in Russian)
12. Alexandrova M. R., Lukashenko A. A., Yurovskaya I. I. and Artemyeva M. S. (2014) Rol' psioemocional'nogo stressa v geneze zhizneugrozhayushchih narushenij ritma serdca [The



role of psychoemotional stress in the genesis of life-threatening cardiac arrhythmias]. *Bulletin of the RUDN, Medicine series*. No. 2. P. 35–41. (in Russian)

13. *Zuikova A. A., Krasnorutskaya O. N. and Khanina E. A.* (2010) *Izmenenie biologicheskikh ritmov organizma cheloveka na fone stressovogo vozdejstviya* [Changes in the biological rhythms of the human body against the background of stress]. *Bulletin of New Medical Technologies*. V.17(2). P. 195–197. (in Russian)

14. *Meerson F. Z., Bukina T. N. and Vasiliev L. A.* (1991) *Rol' professional'nogo i social'nogo stressa v vzniknovenii neishemicheskikh aritmij u pilotov* [The role of professional and social stress in the occurrence of non-ischemic arrhythmias in pilots]. *Cardiology*. No. 1. P. 60–62. (in Russian)

15. *Dementiev M. V., Chibisov S. M., Halabi G., Agarval R. K., Khodorovich N. A. and Kharlitskaya E. V.* (2015) *Desinhronoz serdechno-sudustoj sistemy u mashinistov lokomotornyh brigad* [Desynchronization of the cardiovascular system in locomotive crew drivers]. *The Journal of scientific articles «Health and Education Millennium»*. Vol. 17(4). P. 20–25. (in Russian)

16. *Pavlenko S. I., Vedyasova O. A. and Kretova I. G.* (2021) *Svyaz' pokazatelej vneshnego dyhaniya i variabel'nosti serdechnogo ritma pri umstvennoj nagruzke u studentov s raznymi hronotipami* [The relationship of external respiration and heart rate variability under mental stress in students with different chronotypes]. *Human Physiology*. 47(2). P. 45–55. (in Russian)

17. *Zefferino R., L'Abbate N., Facciorusso A., Potenza A., Lasalvia M., Nuzzaco A., Di Biase M. and Ambrosi L.* (2003) *Assessment of heart rate variability (HRV) as a stress index in an emergency team of urban police*. *G Ital Med Lav Ergon*. Vol. 25, Suppl(3). P. 167–169.

18. *Mashin V. A., Mashin M. N. and Shmeleva I. A.* (1997) *Psihofiziologicheskie issledovaniya emocional'noj labil'nosti operatorov AES* [Psychophysiological studies of emotional lability of NPP operators]. *Questions of psychology*. No. 4. P. 95–103. (in Russian)

19. *Mashin V. A.* (2007) *Psihicheskaya nagruzka, psichicheskoe napryazhenie i funkcional'noe sostoyanie operatorov sistem upravleniya* [Mental load, mental stress and functional state of control system operators]. *Questions of psychology*. No. 6. P. 86–96. (in Russian)

20. *Mashin V. A.* (2011) *K voprosu klassifikacii funkcional'nyh sostoyanij cheloveka* [On the classification of human functional states]. *Experimental Psychology*. V. 4(1). P. 40–56. (in Russian)

21. *Turovsky Ya. A., Kurgalin S. D. and Alekseev A. V.* (2017) *Analiz dvizheniya glaz cheloveka pri upravlenii samohodnym shassi s ispol'zovaniem sistemy video-okulograficheskogo interfejsa* [Analysis of human eye movement when driving a self-propelled chassis using a video-oculographic interface system]. *Sensor Systems*. V. 31(1). P. 51–58. (in Russian)

22. *Turovsky Ya. A., Alekseev A. V., Lesnykh I. E. and Martynenko E. V.* (2021) *Chastotno-vremennye osobennosti dvizheniya glaz pri ispol'zovanii videookulograficheskogo interfejsa v zadachah upravleniya ergaticheskimi sistemami* [Frequency-time features of eye movement when using a video oculographic interface in control tasks of ergatic systems]. *Sensor systems*. V. 35(1). P. 30–37. (in Russian)

23. *Efron B.* (1988) *Netradicionnye metody mnogomernogo statisticheskogo analiza* [Unconventional methods of multidimensional statistical analysis]. *Moscow : Finance and Statistics*. 263 p. (in Russian)

**Mikhalski A. I.** — DSc in Biological Sciences, PhD in Technical Sciences, Senior Researcher, Chief Researcher of V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences. E-mail: [ipuran@yandex.ru](mailto:ipuran@yandex.ru) ORCID ID: <https://orcid.org/0000-0003-1998-1315>

**Turovsky Ya. A.** — DSc in Technical Sciences, PhD in Medical Sciences, Leading Researcher of V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences; Head of Lab., Voronezh State University. E-mail: [yaroslav\\_turovsk@mail.ru](mailto:yaroslav_turovsk@mail.ru) ORCID ID: <https://orcid.org/0000-0002-5290-885X>

**Borzunov S. V.** – PhD in Physics and Mathematics, Associate Professor, Digital Technologies Department, Voronezh State University. E-mail: [sborzunov@gmail.com](mailto:sborzunov@gmail.com) ORCID ID: <https://orcid.org/0000-0002-5099-9655>