

Оригинальная статья / Original article

<https://doi.org/10.21869/2223-1560-2019-23-6-225-240>

Модели и методика определения речевой активности пользователя социо-киберфизической системы

Е.Е. Усина¹ ✉, А.Р. Шабанова¹, И.В. Лебедев¹

¹ Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 14 линия В.О., д. 39, Санкт-Петербург, 199178, Российская Федерация

✉ e-mail: Lizzzi96@mail.ru

Резюме

Цель исследования. Статья посвящена разработке модельно-алгоритмического обеспечения процесса определения речевой активности пользователя социо-киберфизической системы. Предложена топологическая модель распределенной подсистемы аудиозаписи, реализуемой в ограниченных физических пространствах (помещениях), позволяющая оценить качество воспринимаемых аудиосигналов для случая распределения микрофонов в таком помещении. На основе данной модели разработана методика определения речевой активности пользователя социо-киберфизической системы, максимизирующая качество воспринимаемых аудиосигналов при перемещении пользователя в помещении за счет определения координат установки микрофонов.

Методы. Для наиболее полного анализа и формального описания распределенной подсистемы аудиозаписи был использован математический аппарат теории графов и теории множеств. С целью определения координат размещения микрофонов в одном помещении была разработана соответствующая методика, которая предполагает проведение таких операций, как излучение в помещении речевого сигнала с помощью акустического оборудования и замер уровней сигнала в предполагаемых для установки микрофонов местах с использованием шумомера.

Результаты. Были рассчитаны зависимости коэффициента корреляции объединенного сигнала и исходного тестового сигнала от расстояния до источника сигнала для различного количества микрофонов. Полученные зависимости позволяют определить минимально необходимое количество разнесенных микрофонов для обеспечения качественной записи речи пользователя. Результаты апробации разработанной методики речевой активности в конкретном помещении свидетельствуют о возможности и высокой эффективности определения речевой активности пользователя социо-киберфизической системы.

Заключение. Использование предложенной методики определения речевой активности пользователя социо-киберфизической системы позволит повысить качество записи аудиосигнала и, как следствие, его последующей обработки с учетом возможного перемещения пользователя.

Ключевые слова: социо-киберфизическая система; речь; микрофоны; распределенная аудиозапись

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

Для цитирования: Усина Е.Е., Шабанова А. Р., Лебедев И.В. Модели и методика определения речевой активности пользователя социо-киберфизической системы // Известия Юго-Западного государственного университета. 2019; 23(6): 225-240. <https://doi.org/10.21869/2223-1560-2019-23-6-225-240>.

Поступила в редакцию 11.11.2019

Подписана в печать 04.12.2019

Опубликована 23.12.2019

Models and a Tecnique for Determining the Speech Activity of a User of a Socio-Cyberphysical System

Elizaveta E. Usina ¹ ✉, Alexandra R. Shabanova ¹, Igor V. Lebedev ¹

¹ St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences
39, 14-th Linia VI, St. Petersburg 199178, Russian Federation

✉ e-mail: Lizzzi96@mail.ru

Abstract

Purpose of reseach. The article presents the development of the model-algorithmic support for the process of determining the speech activity of a user of a socio-cyberphysical system. A topological model of a distributed subsystem of audio recordings implemented in limited physical spaces (rooms) is proposed; the model makes it possible to assess the quality of perceived audio signals for the case of distribution of microphones in such a room. Based on this model, a technique for determining the speech activity of a user of a socio-cyberphysical system, which maximizes the quality of perceived audio signals when a user moves in a room by means of determining the installation coordinates of microphones has been developed.

Methods. The mathematical tools of graph theory and set theory was used for the most complete analysis and formal description of the distributed subsystem of the audiorecording. In order to determine the coordinates of the placement of microphones in one room, a relevant technique was developed; it involves performing such operations as emitting a speech signal in a room using acoustic equipment and measuring signal levels using a noise meter in the places intended for installing microphones.

Results. The dependences of the correlation coefficient of the combined signal and the initial test signal on the distance to the signal source were calculated for a different number of microphones. The obtained dependences allow us to determine the minimum required number of spaced microphones to ensure high-quality recording of the user's speech. The results of testing the developed technique for determining speech activity in a particular room indicate the possibility and high efficiency of determining the speech activity of a user of a socio-cyberphysical system.

Conclusion. Application of the proposed technique for determining the speech activity of a user of a socio-cyberphysical system will improve the recording quality of the audio signal and, as a consequence, its subsequent processing, taking into account the possible movement of a user.

Keywords: socio-cyberphysical system; speech; microphones; distributed audio recording.

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

For citation: Usina E. E., Shabanova A. R., Lebedev I. V. Models and a Tecnique for Determining the Speech Activity of a User of a Socio-Cyberphysical System. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2019, 23(6): 225-240 (In Russ.). <https://doi.org/10.21869/2223-1560-2019-23-6-225-240>.

Received 11.11.2019

Accepted 04.12.2019

Published 23.12.2019

Введение

В условиях непрерывного технического и технологического прогресса концепция Интернета вещей [1] – связанных в вычислительную сеть физических устройств – является основой киберфизических систем [2, 3]. Последние интегрируют доступные вычислительные ресурсы и обеспечивают совместную работу элементарных частей физического и кибернетического пространств. При введении в указанные системы человека,

зических систем [2, 3]. Последние интегрируют доступные вычислительные ресурсы и обеспечивают совместную работу элементарных частей физического и кибернетического пространств. При введении в указанные системы человека,

как части киберфизических систем, образуется социо-киберфизическая система [4-6].

Рассматривая элементы физического пространства — датчики социо-киберфизических систем, необходимо отметить большое разнообразие их видов, которое определяется измеряемыми физическими величинами (давление, температура, плотность, масса и т.д.), их совокупностями и видами воспринимаемой информации (акустическая, текстовая, видео и т.д.).

Так как в реалиях современного мира общение между людьми и управление различными системами все чаще происходит посредством голоса — широкий спектр возможностей по комфортному общению и голосовому управлению в любой точке ограниченных физических пространств (помещений) может быть предоставлен с использованием множества распределенных микрофонов¹ [7-9]. Например, одним из вариантов их применения может стать разговор пользователя в помещении без использования им смартфона и других дополнительных устройств детектирования и записи голоса.

В связи с вышеизложенным возникает необходимость размещения микрофонов в некотором помещении с учетом того, чтобы *распределенная подсистема*

аудиозаписи (РПА) могла детектировать перемещающегося пользователя, распознавать его голосовые команды и выполнять их. Под РПА в данном случае необходимо понимать распределенную физическую подсистему социо-киберфизической системы записи, обработки, распознавания команд пользователя и взаимодействия с ним, которую составляют: множество микрофонов, устройства аудиозаписи с микрофонов, устройство управления включением микрофонов, устройство обработки и объединения аудиозаписей.

Передовыми российскими исследователями и разработчиками систем высококачественной записи, обработки и анализа аудиоинформации являются компании «ЦРТ» и «ЦРТ-инновации» [10], занимающиеся, в том числе, и разработкой систем распределенной аудиозаписи («Незабудка», «Нестор» и др.) и АО «ОКБ «Октава» (разработчик микрофонов, микрофонных капсулей и гарнитур) [11]. При разработке указанных систем большое внимание уделено алгоритмам шумоочистки и совместной обработки записанных аудиосигналов, однако вопросы размещения микрофонов в помещениях остаются недостаточно проработанными.

Следовательно, необходимо разработать модельно-алгоритмическое обеспечение для создания распределенной подсистемы аудиозаписи, позволяющей определять речевую активность пользователя социо-киберфизической системы и необходимой для определения наи-

¹ Ронжин А.Л. Методы и программные средства многоканальной дистанционной обработки речи и их применение в интерактивных многомодальных приложениях: дис. ... д-ра техн. наук. СПб., 2010. 330 с.

лучшего расположения микрофонов на инженерных конструкциях здания для качественной записи и распознавания речи с учетом возможного перемещения пользователя.

Проектирование РПА в настоящее время основано на нескольких способах размещения микрофонов:

- гарнитурный (непосредственное расположение возле рта);
- петличный (расположение микрофона вблизи рта на одежде диктора);
- настольный/напольный (перед диктором в направлении распространения звуковой волны);
- распределенное множество микрофонов¹ [7-101].

Первые три способа связаны с определенной подготовкой пользователя для записи голоса (закрепление гарнитуры на голове, установка микрофона на предметы одежды, расположение рта пользователя в непосредственной близости от микрофона), при этом пользователь постоянно испытывает некоторый дискомфорт, вызываемый необходимостью подстраиваться под требования систем записи речи (переносить на себе элементы датчиков, располагаться статично, говорить четко по направлению к микрофону).

Четвертый способ является наиболее подходящим для разработки РПА,

так как не требует от пользователя дополнительной подготовки. Множество направленных микрофонов представляет возможность использовать превосходства бинаурального эффекта [12-14].

Рассмотренные схемы расположения распределенного множества микрофонов, существующие в настоящее время, являются пригодными лишь для статичного пользователя. Но при разработке РПА необходимо учитывать, что пользователь может перемещаться по помещениям и зданию. Поэтому возникает необходимость создания научно-методического инструментария для определения мест установки микрофонов РПА.

Материалы и методы

1. Топологическая модель распределенной подсистемы аудиозаписи

Моделирование структуры РПА произведем на примере одноэтажного здания, схема которого представлена на рис.1. В указанном здании границами перемещений пользователя являются ограждающие конструкции: внутренние и внешние стены, двери, окна, части перекрытий, пол, потолок.

Наиболее полно анализ РПА можно провести с помощью математического аппарата теории графов, разработав топологическую структуру РПА. Соответствующая примерной схеме здания топологическая структура в виде графа показана на рис. 2.

Для графа топологической структуры РПА $G_{\text{ТОП}}(V, E, I_G)$ с множеством вершин $V=\{v_i\}$, $i=1,..., N_y$), соответ-

¹ Ронжин А.Л. Методы и программные средства многоканальной дистанционной обработки речи и их применение в интерактивных многомодальных приложениях: дис. ... д-ра техн. наук. СПб., 2010. 330 с.

ствующих элементам (подсистемам) РПА (N_y – число микрофонов), множеством ребер $E = \{e_{ij}\}$, $i, j = 1, \dots, N_y$, соответствующих направлениям прохождения сигналов, и связностью I_G , задаются взаимное расположение микро-

нов a_i в виде координат $\{x_i, y_i\}$ в помещении и конструктивное исполнение отдельных узлов и блоков средства, что позволяет анализировать (рассчитывать) энергетические характеристики акустических сигналов и т. п.

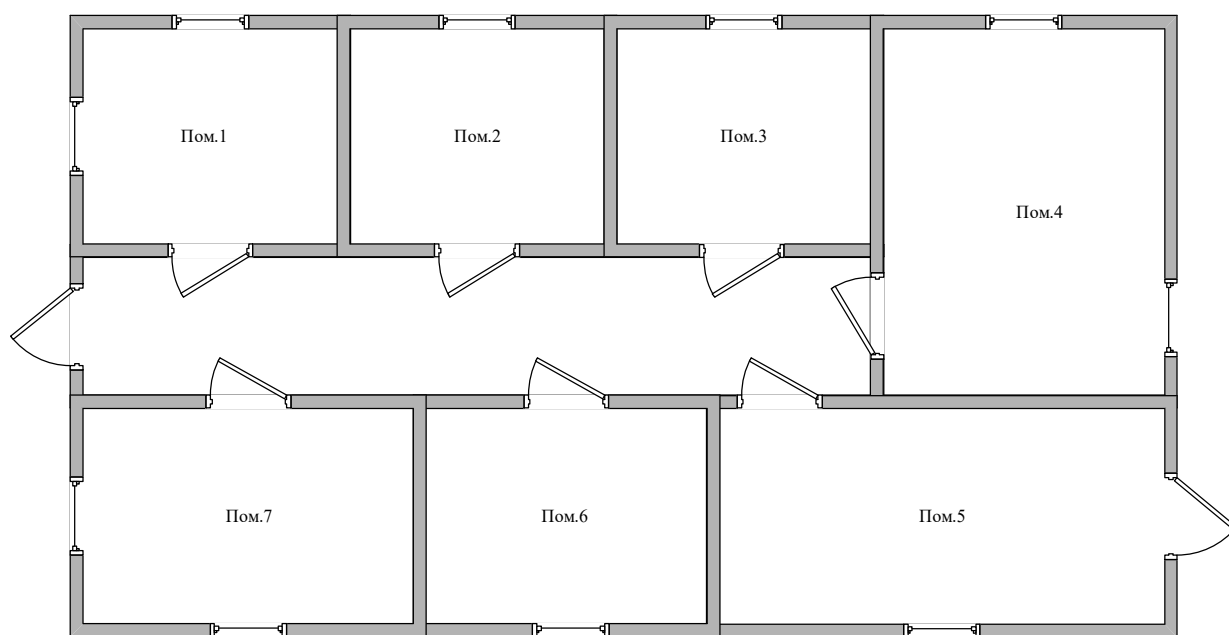


Рис. 1. Примерная схема одноэтажного здания

Fig. 1. Approximate layout of a one-floor building

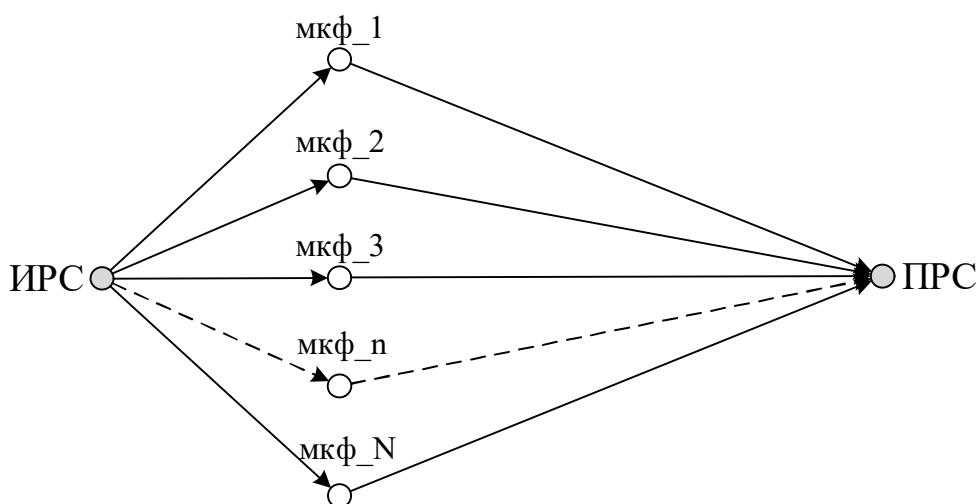


Рис. 2. Топологическая структура РПА: ИРС – источник речевого сигнала; ПРС – приемник речевого сигнала

Fig. 2. Topological structure of DARS: ИРС – speech signal source; ПРС – speech signal receiver

С учетом представленной структуры для формального описания РПА введем следующие множества:

– множество $D = \{d_h\}$, $h = \{1, \dots, N\}$ устройств преобразования речевых сигналов;

– множество задач $i = \{1, \dots, I\}$ (этапов $z = \{1, \dots, z_i\}$) РПА и их вариантов $k = \{1, \dots, K\}$;

– множество преобразований $W = \{w_o\}$, $o = \{1, \dots, O\}$, выполняемых при обработке речевой информации;

– множества потоков сигналов $S = \{s_e\}$, $e = \{1, \dots, E\}$, использующихся для анализа входных $S^0 = \{S_1^0, S_2^0, \dots, S_{N_s}^0\}$ и синтеза выходных $S' = \{S_1', S_2', \dots, S_{N_s}'\}$ сигналов;

– множество допустимых системо-технических решений $\Delta_\alpha = \{\alpha = \langle mm, mo, al, ap \rangle \mid mm \in MM, mo \in MO, al \in AL, ap \in AP\}$ включающее подмножества математических моделей MM , математических методов MO , алгоритмов AL и программно-аппаратных реализаций процесса объединения перехваченных сигналов;

– множество узлов функциональных структур $j = \{1, \dots, J\}$ и вариантов их построения $p = \{1, \dots, P\}$.

Топологическая структура РПА включает в себя объединенные структурные элементы с фиксированным расположением в здании и может быть

преобразована для дальнейшего анализа в функциональную структуру РПА.

Для всей группы каналов N микрофонов сигнал можно представить в виде следующей системы:

$$\begin{cases} s_1(t) = \gamma_1 \cdot s(t - \tau_1) + N_1(t) \\ s_2(t) = \gamma_2 \cdot s(t - \tau_2) + N_2(t) \\ \vdots \\ s_n(t) = \gamma_n \cdot s(t - \tau_n) + N_n(t) \\ \vdots \\ s_N(t) = \gamma_N \cdot s(t - \tau_N) + N_N(t), \end{cases} \quad (1)$$

где γ_n – постоянный коэффициент передачи канала; τ_n – задержка в канале n -ого микрофона; $N_n(t)$ – аддитивный шум в канале n -ого микрофона.

Для объединенного сигнала $s'(t)$ в приемнике можно получить следующее выражение:

$$s'(t) = s_1(t) + s_2(t) + \dots + s_n(t) + \dots + s_N(t) = \sum_N s_N(t). \quad (2)$$

Общее выражение для сигнала в приемнике будет выглядеть следующим образом:

$$s'(t) = \sum_n \left(\gamma_n \cdot s(t - \tau_n) + N_n(t) \right). \quad (3)$$

Тогда соотношение «сигнал-шум» на входе анализатора (решающего устройства) примет следующую форму:

$$\Delta = \frac{\left(\sum_n \left(\gamma_n \cdot s(t - \tau_n) \right) \right)^2}{D \left[\sum_n N_n(t) \right]}. \quad (4)$$

Учитывая при расчетах возможность синхронизации сигналов в каждом из n каналов можно прийти к выводу о равенстве задержек при распространении ре-

чевого сигнала, то есть $\sum_n \tau_n = 0$. Тогда

при $M\left[\sum_n N_n(t)\right] = 0$ и коэффициентах

корреляции $r_{N_i N_j} \ll r_{s_x(t) s_y(t)}$ ($i \neq j$, $x \neq y$) выражение (4) преобразуется к следующему виду:

$$\Delta = \frac{\left(n \cdot s(t) \cdot \sum_n (\gamma_n)\right)^2}{\sum_n D\left[\sum_n N_n(t)\right]}. \quad (5)$$

Суммарный уровень шума в каждом из каналов $\sum_n N_n(t)$ можно обозначить как $N_n(t)$:

$$\Delta = \frac{n^2 \cdot s^2(t) \cdot \left(\sum_n \gamma_n\right)^2}{\sum_n D[N_n(t)]}. \quad (6)$$

При равенстве параметров распределений шумовых сигналов $N_n(t)$ [15] выражение для соотношения «сигнал-шум»:

$$\begin{aligned} \Delta &= \frac{m^2 \cdot s^2(t) \cdot \left(\sum_n \gamma_n\right)^2}{m \cdot D[N(t)]} = \\ &= \frac{m \cdot s^2(t) \cdot \left(\sum_n \gamma_n\right)^2}{\sigma^2}. \end{aligned} \quad (7)$$

Отношение «сигнал-шум» с учетом одного канала (микрофона) определяется следующей формулой:

$$\Delta_0 = \frac{P_c}{P_{ш}} = \frac{s^2(t) \cdot \gamma_n^2}{\sigma^2}. \quad (8)$$

При коэффициентах передачи каналов $\gamma_n \rightarrow 1$ за счет распределенной аудиозаписи можно получить выигрыш по соотношению сигнал шум в n раз, где n –

количество одновременно записывающих микрофонов, что определяет теоретический максимум повышения значения отношения «сигнал-шум» для РПА.

Установлено, что с ростом числа каналов (микрофонов) повышается коэффициент корреляции между речевым сигналом источника и совокупным сигналом приемника (рис. 3).

Для расчета коэффициента корреляции между дискретизированными речевым сигналом источника $s(t)$ и совокупным сигналом приемника $s'(t)$ необходимо пользоваться следующим выражением [16]:

$$\begin{aligned} r_{ss'} &= \frac{\text{cov}(s, s')}{\sigma_s \cdot \sigma_{s'}} = \\ &= \frac{M\{[s - M(s)] \cdot [s' - M(s')]\}}{\sqrt{D(s)} \cdot \sqrt{D(s')}}, \end{aligned} \quad (9)$$

где $M(...)$ – математическое ожидание; $D(...)$ – дисперсия.

Результаты моделирования свидетельствуют о возможности повышения качества воспринимаемого РПА сигнала для случая распределения микрофонов в помещении.

2. Методика размещения микрофонов распределенной подсистемы аудиозаписи

Для описания методики размещения микрофонов РПА необходимо ввести некоторые ограничения и допущения:

– звук от источника распространяется с одинаковой скоростью и одинаковым затуханием во все стороны сферы R (R – максимальное расстояние от источника, на котором возможно детек-

тирование и аудиозапись с помощью микрофонов);

- источник один, перемещается в горизонтальной плоскости (на одной высоте $h=1,7$ м без изменения вертикального положения);

- уровень речевого сигнала постоянен (75 дБА) или сигнал отсутствует;

- уровень шума ограничивается отметкой 50 дБА (тихо или слегка шумно);

- коэффициенты поглощения материалов мебели и отделки помещений менее 0,2 (коэффициентами поглощения для обеспечения точности расчетов

$\pm 1\%$ можно пренебречь);

- время реверберации в помещении не превышает 50 мс;

- объем помещений ограничивается плоским потолком, плоскими стенами и плоским полом;

- микрофоны РПА размещаются на стенах или потолке;

- траектория движения источника сигнала известна и задается матрицей $Tr(x, y)$;

- при невозможности размещения датчиков РПА на стенах, они размещаются на потолке.

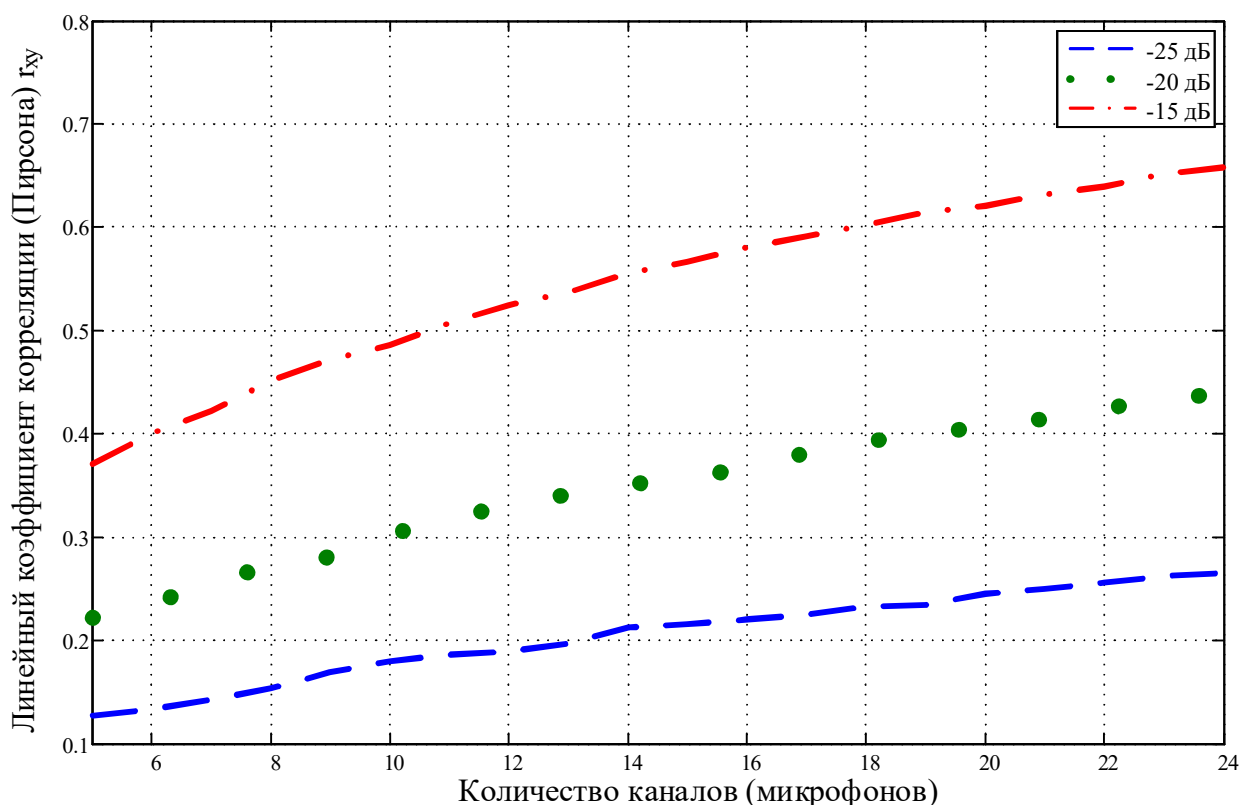


Рис. 3. Зависимость коэффициента корреляции между исходным сигналом и совокупным сигналом от количества каналов (микрофонов) для различного отношения сигнал-шум

Fig. 3. Correlation of source signal and aggregate signal on the number of channels (microphones) for different signal-noise ratios

Методика размещения микрофонов РПА предполагает проведение следующих операций:

- излучение в помещении с помощью акустического оборудования речевого сигнала (тестового);

– замер с использованием шумомера уровней сигнала в предполагаемых для установки микрофонов местах (стены и потолок);

– определение расстояния R .

Приборы и оборудование:

– акустическая широкополосная звуковоспроизводящая система с возможностью воспроизведения сигналов с интегральным уровнем не менее 90 дБ;

– шумомер (например, «ЭкоФизика») для настройки акустической звуковоспроизводящей системы на уровень, соответствующий разговорной речи;

– набор аудиофайлов с записями тестового речевого сигнала для воспроизведения через акустическую систему.

Требования к тестовому речевому сигналу, применяемому для методики размещения микрофонов РПА:

– озвученная диктором запись тестовых фраз из ГОСТ 50840¹;

– параметры аудиофайла: формат аудиофайла WAV, частота дискретизации не ниже 22 кГц, квантование 16 бит, продолжительность от 10 с до 20 с, моно;

– средства воспроизведения: акустическая система, позволяющая воспроизводить звук в диапазоне 100 Гц – 20 кГц, отношение сигнал-шум не ниже 75 дБ.

Для определения сферы R для РПА необходимо выбрать микрофоны. Характеристики современных микрофонов приведены в табл.

Исходные данные:

– план помещений, в которых предполагается размещение микрофонов РПА;

– характеристики выбранных микрофонов;

– матрица координат $Tr(x,y)$ траектории движения источника сигнала (рис. 4):

$$Tr = \begin{bmatrix} 2 & 3 & 3,5 & 3,5 & 2 & 0,8 & 0,5 & 1 & 0,7 & \Lambda \\ 5 & 5,3 & 6,5 & 7,5 & 7,8 & 7,8 & 7 & 6,5 & 5,3 & \end{bmatrix}.$$

Таблица. Характеристики современных микрофонов

Table. Modern microphones characteristics

Название	STELBERRY					
	У М-10	М-20	М-30	М-40	М-50	М-60
Тип микрофона	Высокочувствительный активный микрофон	Высокочувствительный активный микрофон	Высокочувствительный активный микрофон	Высокочувствительный активный микрофон	Высокочувствительный активный микрофон	Сверхчувствительный MEMS
Ветровая защита	Есть	Есть	Есть	Есть	Есть	Есть

¹ ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. М.: Издательство стандартов, 1995. 234 с.

Название	STELBERRY					
	У М-10	М-20	М-30	М-40	М-50	М-60
Металлический корпус от электромагнитных помех	Есть	Есть	Есть	Есть	Есть	Есть
Регулировка усиления	Нет	Ручная	Автоматическая	Ручная/ автоматическая	Ручная/ автоматическая	
Скорость срабатывания АРУ, с	—	—	0,25	0,25	0,0007	
Стабилизатор напряжения	Нет	Нет	Есть	Есть	Есть	
Цифровой					Да	
Отношение сигнал/шум						63 дБ (1400 раз)
Реалистичный звук (эффект присутствия)						Да
Полоса пропускания, Гц	100...6100	100...7200	100...8300	100...9200	270...4000	80...16000

Методика размещения микрофонов РПА для одного помещения включает в себя следующие шаги.

Шаг 1. Определение точки $(x_i; y_i)$ траектории движения источника, ближайшей ко входу в помещение.

Шаг 2. Излучение акустического сигнала из точки $(x_i; y_i)$ в пространство помещения.

Шаг 3. Нахождение с помощью шумомера точки с максимальным уровнем речевого сигнала на фоне присутствующих акустических (вибракустических) шумов в районе поверхности стены и потолка (нахождение точки R_i).

Шаг 4. Сравнение характеристики микрофона (динамический диапазон, чувствительность) с уровнем речевого сигнала в точке R_i .

Шаг 5. Если уровень речевого сигнала U_i достаточен (соответствует) для корректной записи микрофоном ($U_i > U_{min}$, где U_{min} – минимально возможный для записи микрофоном уровень речевого сигнала), то точка R_i фиксируется в качестве места расположения микрофона.

Шаг 6. Если уровень речевого сигнала в точке R_i не достаточен (не соответствует) для корректной записи микрофоном, то необходимо выбрать микрофон с другими характеристиками (или расположение микрофона не на стене или потолке).

Шаг 7. Передвижение по траектории $Tr(x, y)$ к следующей точке $(x_{i+1}; y_{i+1})$ с постоянными измерениями уровня речевого сигнала на фоне присутствующих акустических (вибракустических) шумов в фиксированной для установки микрофона точке R_i пока $U_i > U_{min}$.

Шаг 8. Если точка $(x_{i+1}; y_{i+1})$ в матрице Tr существует (не вся траектория пройдена), то переход на **Шаг 2**.

$$R = \begin{bmatrix} 2,5 & 3,2 & 3,1 & 3,7 & 2,9 & 2 & 1 & 0 & 1,5 & \Lambda \\ 5 & 5 & 6,3 & 7,6 & 8 & 7,5 & 8 & 7 & 5,5 & \end{bmatrix}.$$

На плане пом. 1 одноэтажного здания (рис. 1) нанесены места установки

В результате работы алгоритма создается матрица координат установки микрофонов R для плана помещений:

микрофонов в соответствии со значениями координат матрицы R (рис. 5).

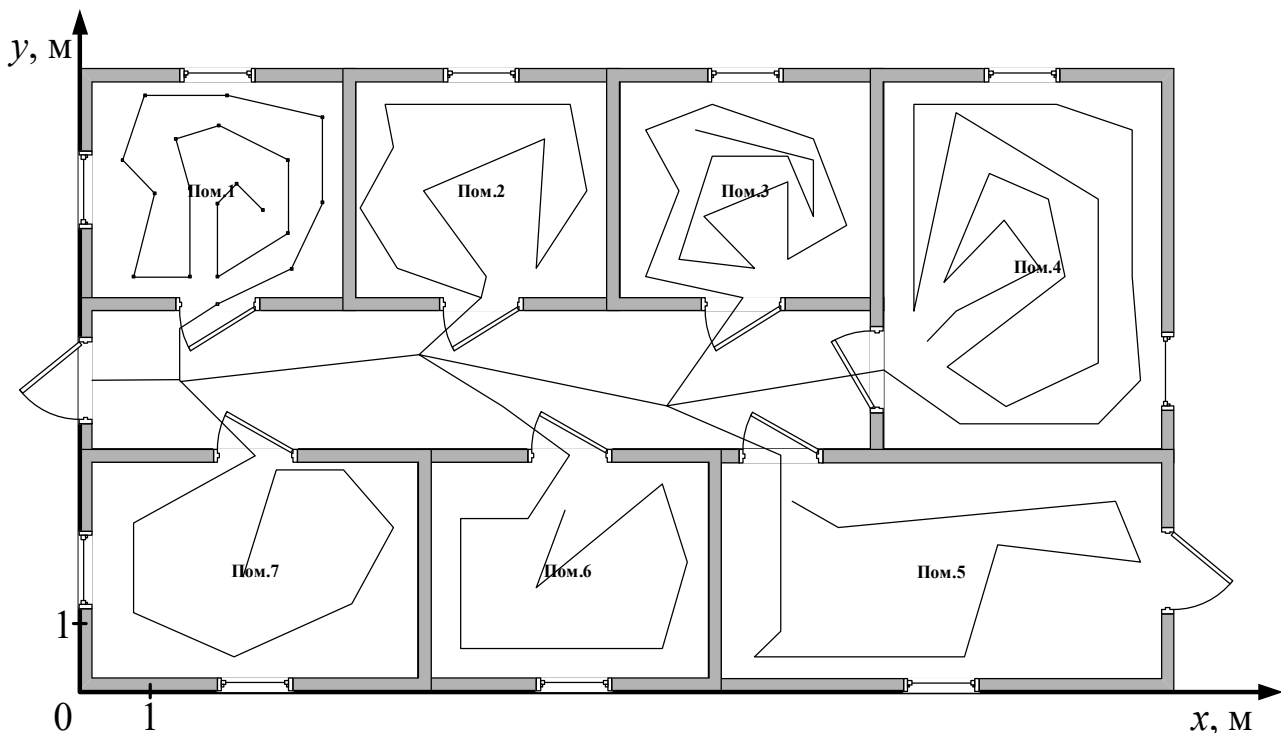


Рис. 4. Примерная схема одноэтажного здания с нанесенной траекторией движения источника

Fig. 4. Approximate layout of a single-storey building with plotted movement pattern of a signal source

Результаты и их обсуждение

По результатам работы алгоритма были получены координаты установки микрофонов. С использованием различного количества микрофонов, установленных в соответствии с координатами соседних столбцов матрицы R , были записаны тестовые фразы от источника сигнала, находящегося на расстоянии от 1 до 5 м. Для каждого записанного объединенного сигнала рассчитано значение коэффициента корреляции Пирсона

на, являющегося показателем качества записанного сигнала¹ [16], получены зависимости коэффициента корреляции объединенного сигнала и исходного тестового сигнала от расстояния до источника сигнала для различного количества микрофонов, по которым были построены соответствующие зависимости (рис. 6).

¹ ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. М.: Издательство стандартов, 1995. 234 с.

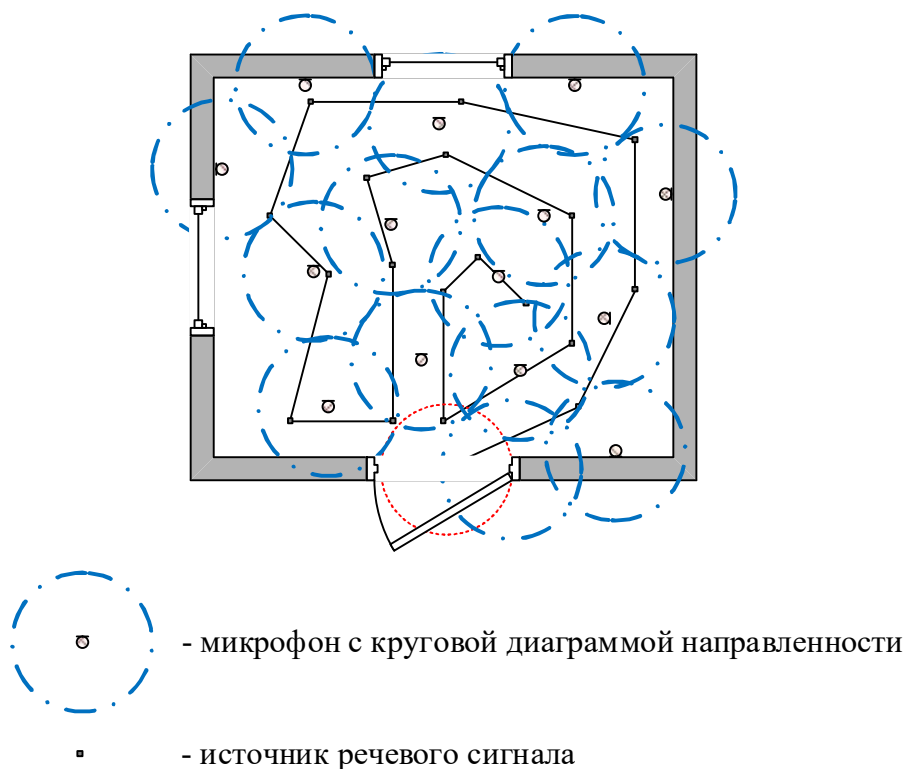


Рис. 5. Иллюстрация методики на примере пом. 1 одноэтажного строения

Fig. 5. Methodology illustrated on example of Room 1 of one-floor building

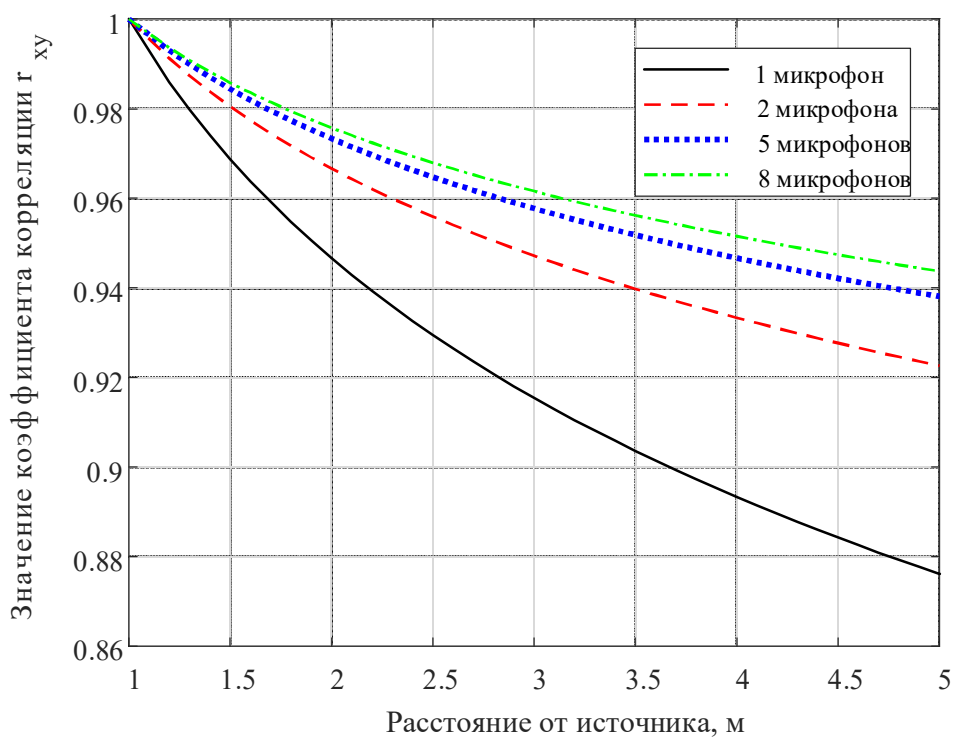


Рис. 6. Графики зависимости разборчивости от количества микрофонов для аудиозаписи

Fig. 6. Dependency diagram of speech intelligibility on the number of audio-recording microphones

Полученные зависимости позволяют определить минимально необходимое количество разнесенных микрофонов для обеспечения качественной записи речи пользователя в РПА. В данном случае с учетом ограничений и допущений, описанных в начале раздела, необходимо использовать не менее пяти микрофонов. Увеличение количества задействованных микрофонов приводит к незначительному выигрышу в качестве сигнала.

Выводы

В современном мире при решении вопросов организации эффективного взаимодействия физических, кибернетических подсистем и человека, а также оптимального распределения имеющихся ресурсов между ними прочные позиции заняли социо-киберфизические системы.

На сегодняшний день одним из основных видов обрабатываемой информации в таких системах является акустическая, а именно речь человека [17, 18]. Разработчики систем высококачественной записи, обработки и анализа аудиоинформации большое внимание уделяют разработке алгоритмов шумочистки и совместной обработки записанных аудиосигналов, тогда как открытым остается вопрос размещения микрофонов в ограниченных физических пространствах для записи голоса пользователя.

Применение разработанной методики определения речевой активности пользователя социо-киберфизической системы позволит повысить качество записи аудиосигнала и, как следствие, его последующей обработки с учетом возможного перемещения пользователя.

Список литературы

1. Internet of Things, IoT European Research Cluster. [процитировано 6 ноября 2019]. URL: http://www.internet-of-things-research.eu/about_iot.htm
2. Teaching Smart Production: An insight into the Learning Factory for Cyber-Physical Production Systems (LVP) / L. Merkela, J. Atuga, L. Merhara, C. Schultza, S. Braunreuthera, G. Reinharta // *Procedia Manufacturing*. 2017. №9. P. 269-274. <https://doi.org/10.1016/j.promfg.2017.04.034>
3. Knowledge-Based Decision Making in a CyberPhysical Production Scenario / J. Klöber-Koch, S. Pielmeier, J. Grimm, M. Brandt, M., Schneider G. Reinhart // *7th Conference on Learning Factories*. 2017. №7. P. 167-174. <https://doi.org/10.1016/j.promfg.2017.04.014>
4. Jiang P., Ding K., Leng J. Towards a cyber-physical-socialconnected and service-oriented manufacturing paradigm: Social Manufacturing // *Manufacturing Letters*. 2016. №7. P. 15-21. <https://doi.org/10.1016/j.mfglet.2015.12.002>
5. Cassandras C.G. Smart Cities as Cyber-Physical Social Systems // *Engineering*. 2016. №2. P. 156-158. <https://doi.org/10.1016/J.ENG.2016.02.012>

6. Смирнов А.В., Левашова Т.В. Приобретение знаний в социокиберфизических системах в процессе информационного взаимодействия ресурсов // Информационно-управляющие системы. 2017. №6. Р. 113–122.
7. Мазуренко И.Л. Многоканальная система распознавания речи // Сборник трудов VI всероссийской конференции «Нейрокомпьютеры и их применение». М., 2000.
8. Beyond Close-talk – Issues in Distant Speech Acquisition, Conditioning Classification, and Recognition / V. Stanford, C. Rochet, M. Michel, J.N Garofolo / Proc. ICASSP 2004 Meeting Recognition Workshop. 2004. Р. 123-127.
9. Pfau T., Ellis D. P. W., Stolcke A. Multispeaker speech activity detection for the ICSI meeting recorder // IEEE Workshop on Automatic Speech Recognition and Understanding. 2001. Р. 107-110. <https://doi.org/10.1109/ASRU.2001.1034599>
10. Центр речевых технологий. 2019 [протитировано 6 ноября 2019]. URL: <http://www.speechpro.ru>
11. АО «ОКБ «Октава». 2019 [протитировано 6 ноября 2019]. URL: <https://www.окбоктава.рф>
12. Акустика / Ш.Я. Вахитов, Ю.А. Ковалгин, А.А. Фадеев, Ю.П. Щевьев. М.: Горячая линия, 2009.
13. Разработка многомодального информационного киоска / А.Л. Ронжин, А.А. Карпов, А.Б., Леонтьева Б.Е. Костюченко // Труды СПИИРАН. 2007. №5(1). С. 227-245.
14. Ронжин А.Л., Карпов А.А., Кагиров И.А. Особенности дистанционной записи и обработки речи в автоматах самообслуживания // Информационно-управляющие системы. 2009. №42(5). С. 32-38.
15. Харкевич А.А. Борьба с помехами. Изд. 4-е. М.: Книжный дом «ЛИБРОКОМ», 2013.
16. Склад Б. Цифровая связь. Теоретические основы и практическое применение. Изд. 2-е, испр. М.: Издательский дом «Вильямс», 2003.
17. Ogunfunmi T., Togneri R., Narasimha M. Speech and audio processing for coding, enhancement and recognition. New York : Springer, 2015.
18. Марковников Н.М., Кипяткова И.С. Аналитический обзор интегральных систем распознавания речи // Труды СПИИРАН. 2018. №3. С. 77-110. <https://doi.org/10.15622/sp.58.4>

References

1. Internet of Things, IoT European Research Cluster. [Quoted November 6, 2019]. Available at: http://www.internet-of-things-research.eu/about_iot.htm
2. Merkela L., Atuga J., Merhara L., Schultza C., Braunreuthera S., Reinharta G. Teaching Smart Production: An insight into the Learning Factory for Cyber-Physical Production

Systems (LVP). *Procedia Manufacturing*, 2017, no. 9, pp. 269-274. <https://doi.org/10.1016/j.promfg.2017.04.034>

3. Klöber-Koch J., Pielmeier S. Grimm J., Brandt M., Schneider M., Reinhart G. Knowledge-Based Decision Making in a CyberPhysical Production Scenario. *7th Conference on Learning Factories*, 2017, no. 7, pp. 167-174. <https://doi.org/10.1016/j.promfg.2017.04.014>

4. Jiang P., Ding K., Leng J. Towards a cyber-physical-socialconnected and service-oriented manufacturing paradigm: Social Manufacturing. *Manufacturing Letters*. 2016, no. 7, pp. 15-21. <https://doi.org/10.1016/j.mfglet.2015.12.002>

5. Cassandras C.G. Smart Cities as Cyber-Physical Social Systems. *Engineering*, 2016, no. 2, pp. 156-158. <https://doi.org/10.1016/J.ENG.2016.02.012>

6. Smirnov A.V., Levashova T.V. Priobretenie znaniy v sotsiokiberfizicheskikh sistemakh v protsesse informatsionnogo vzaimodeystviya resursov [The acquisition of knowledge in sociocyberphysical systems in the process of information interaction of resources]. *Informatsionno-upravlyayushchie sistemy = Information and Control Systems*, 2017, no.6, pp. 113–122 (In Russ.)

7. Mazurenko I.L. Mnogokanal'naya sistema raspoznavaniya rechi [Multi-channel speech recognition system]. *Sbornik trudov VI vserossiiskoi konferentsii "Neirokomp'yutery i ikh primeneniye"* [Proceedings of the VI All-Russian conference "Neurocomputers and their application"]. Moscow, 2000. (In Russ.).

8. Stanford V., Rochet C., Michel M., Garofolo J. Beyond Close-talk – Issues in Distant Speech Acquisition, Conditioning Classification, and Recognition. *Proc. ICASSP 2004 Meeting Recognition Workshop*, 2004, pp. 123-127.

9. Pfau T., Ellis D. P. W., Stolcke A. Multispeaker speech activity detection for the ICSI meeting recorder. *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2001, pp. 107-110. <https://doi.org/10.1109/ASRU.2001.1034599>

10. Centr rechevykh tekhnologiy [Speech technology center]. 2019 [Quoted November 6, 2019]. Available at: <http://www.speechpro.ru> (In Russ.).

11. AO "OKB "Oktava". 2019 [Quoted November 6, 2019]. Available at: <https://www.okboktava.pf> (In Russ.).

12. Vakhitov Sh.Ya., Kovalgin Yu.A., Fadeev A.A., Shcheviev Yu.P. *Akustika* [Acoustics]. Moscow, *Goryachaya liniya Publ.*, 2009. (In Russ.).

13. Ronzhin A.L., Karpov A.A., Leontyeva A.B., Kostyuchenko B.E. Razrabotka mnogomodal'nogo informatsionnogo kioska [The development of the multimodal information kiosk]. *Trudy SPIIRAN = SPIIRAS Proceedings*, 2007, no. 5(1), pp. 227-245 (In Russ.)

14. Ronzhin A.L., Karpov A.A., Kagiroy I.A. Osobennosti distantsionnoi zapisi i obrabotki rechi v avtomatakh samoobsluzhivaniya [Features of remote recording and speech processing in self-service machines]. *Informatsionno-upravlyayushchie sistemy = Information and control systems*, 2009, no. 42(5), pp. 32–38 (In Russ.)

15. Kharkevich A.A. *Bor'ba s pomekhami* [Struggle against interference]. Moscow, Knizhnyi dom "LIBROKOM" Publ., 2013 (In Russ.)
16. Sklar B. *Tsifrovaya svyaz'. Teoreticheskie osnovy i prakticheskoe primeneniye* [Digital communication. Theoretical foundations and practical application]. Moscow, Izdatel'skii dom "Vil'yams" Publ., 2003 (In Russ.)
17. Ogunfunmi T., Togneri R., Narasimha M. *Speech and audio processing for coding, enhancement and recognition*. New York, Springer Publ., 2015.
18. Markovnikov N.M., Kipyatkova I.S. Analiticheskii obzor integral'nykh sistem raspoznavaniya rechi [An Analytic Survey of End-to-End Speech Recognition Systems]. *Trudy SPIIRAN = SPIIRAS Proceedings*, 2018, no. 3, pp. 77-110 (In Russ.). <https://doi.org/10.15622/sp.58.4>.

Информация об авторах / Information about the Authors

Усина Елизавета Евгеньевна, младший научный сотрудник лаборатории технологий больших данных социкиберфизических систем, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: Lizzzi96@mail.ru

Elizaveta E. Usina, Junior Researcher, Laboratory of Big Data Technologies of Sociocyberphysical Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation, e-mail: Lizzzi96@mail.ru

Шабанова Александра Романовна, младший научный сотрудник лаборатории технологий больших данных социкиберфизических систем, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, Санкт-Петербург, Российская Федерация, e-mail: iialex.shabanovaii@gmail.com

Alexandra R. Shabanova, Junior Researcher, Laboratory of Big Data Technologies of Sociocyberphysical Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation, e-mail: iialex.shabanovaii@gmail.com

Лебедев Игорь Владимирович, младший научный сотрудник лаборатории технологий больших данных социкиберфизических систем, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: igorlevedev@gmail.com

Igor V. Lebedev, Junior Researcher, Laboratory of Big Data Technologies of Sociocyberphysical System, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation, e-mail: igorlevedev@gmail.com