DOI 10.15826/urej.2017.1.1.004 УДК 004.522

Голосовой замок

А. Е. Жданов[1, 2], *Л. Г. Доросинский*[2*]
[1] Университет имени Фридриха — Александра в Эрлангене и Нюрнберге, Эрланген, Германия

[2] Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, Екатеринбург, Россия
[*] l.dorosinsky@mail.ru

Аннотация. В данной статье представлено обобщенное решение задачи идентификации личности говорящего по речевому сигналу. В среде программирования MATLAB авторами была разработана программа, реализующая различение (дискриминацию) речевых сигналов по коэффициенту корреляции Пирсона между корреляционной функцией входного речевого сигнала и корреляционными функциями эталонных речевых сигналов.

Разработанная программа определяет истинный речевой сигнал с вероятностью 60—100 % при коэффициенте шума 0—0,3. Достоинством выведенного алгоритма является быстрая скорость вычислений, а недостатком — отсутствие дополнительного программного алгоритма шумопонижения входного речевого сигнала.

Ключевые слова. Идентификация речевого сигнала, дискриминантный анализ распознавания речи, корреляционная функция, коэффициент корреляции Пирсона.

Voice Lock

Aleksei E. Zhdanov[1, 2], *Leonid G. Dorosinsky*[2*]
[1] Friedrich — Alexander University Erlangen-Nürnberg, Erlangen, Germany

[2] Ural Federal University named after the first President of Russia B. N. Yeltsin, Ekaterinburg, Russia
[*] l.dorosinsky@mail.ru

Abstract. The aim of this article is to identify the speaker's personality by a speech signal. The speech recognition system identifies people's individuality by comparison the speech signal of a speaker and the standard speech signal.

The article shows an overview of speech signal recognition methods, such as the dynamic programming, the discriminant analysis (which is based on Bayesian discrimination) and the hidden Markov's model method, as well as an artificial neural network.

To solve the speaker's personality determination aim, we choose the discriminant analysis method which is based on the comparison of the speaker's speech signal and the standard speech signal.

Using the MATLAB programming software, we have developed a program which discriminates speech signals based on the Pearson correlation coefficient between the correlation function of the input speech signal and the correlation functions of the standard speech signals.

If a noise factor is ranges from 0 to 0.3, the program determines the correct speech signal with a probability range from 60 to 100 %. The advantage of the developed algorithm is the fast computational speed. However, the absence of an additional software algorithm for noise reduction of the input speech signal is the disadvantage. Thus, the presented article considers a generalized solution of this problem.

Keywords. Speech signal identification, discriminant analysis of speech recognition, correlation function, Pearson correlation coefficient.

Актуальность

В данной статье рассматривается задача определения личности говорящего. Решение этой задачи может найти применение в криминалистике, радиоразведке, контрразведке, обеспечении безопасности доступа к физическим объектам, информационным и финансовым ресурсам.

К основным задачам обработки речи относятся: распознавание и синтез речи, выделение ключевых слов, определение личности говорящего, компрессия речи. Остановимся подробнее на задаче определения личности говорящего.

Классы задач, в которых применяется задача определения личности говорящего:

1. Проверка прав доступа к различным системам (вычислительные системы, банковские счета, различные устройства и механизмы (транспортные средства, оружие и т. п.) и т. д.);

2. Криминалистическая экспертиза (например, анализ записей переговоров).

Установление индивидуальности по голосу при решении задач подобного рода обеспечивает безопасность, в том числе информационную. Установить индивидуальность личности по голосу можно с помощью системы распознавания речевых сигналов на основе сравнения с эталоном.

Обзор научной литературы

Распознавание речевых сигналов (РРС) — это автоматическое отнесение предъявленного речевого сигнала к одному из заранее выбранных классов. Решение задачи РРС означает нахождение способа

классификации речевых сигналов, наиболее точно соответствующего классификации, осуществляемой человеком.

РРС в широком смысле — это фонемное перекодирование речевого акустического сигнала. Классами речевых сигналов в этом случае являются фонемы. Понятие «фонема» определяется как обозначение всех тех элементарных звуков речи, которым соответствует при написании в фонетической транскрипции одна и та же буква или символ.

РРС в узком смысле — это решение частных задач распознавания речи, когда с целью облегчения решения задачи распознавания искусственно ограничиваются условия, при которых производится классификация. Такой задачей является, например, распознавание изолированно произнесенных слов из заранее выбранного словаря. В зависимости от поставленной цели ответом при РРС может быть не только фонема или слово, но также индивидуальность диктора (идентификация личности по ее голосу), его эмоциональное состояние и др. [1]

На сегодняшний день существует четыре метода распознавания речи на основе сравнения с эталоном:

1. Динамическое программирование (ДП) — это один из наиболее мощных и широко известных математических методов современной теории управления, был предложен в конце 50-х годов американским математиком Р. Беллманом и быстро получил широкое распространение.

В основе метода лежит идея рассмотрения исходной задачи как представителя семейства сходных с ней задач. ДП связано с многошаговым (многоэтапным) процессом принятия решений. При этом под многошаговым процессом принятия решений понимается деятельность, при которой принимаются последовательные решения, направленные на достижение одной цели. Методу ДП посвящено множество публикаций, в которых достаточно подробно рассмотрена техника решения задач методом динамического программирования [2–4].

2. Дискриминантный анализ, основанный на Байесовской дискриминации.

Дискриминантный анализ — это раздел математической статистики, посвященный разработке методов решения задач различения (дискриминации) объектов наблюдения по определенным признакам. Например, разбиение совокупности речевого сигнала на несколько однородных групп по значениям каких-либо показателей тональности сигнала.

Методы дискриминантного анализа находят применение в различных областях: медицине, социологии, психологии, экономике и т. д. При наблюдении больших статистических совокупностей часто появляется необходимость разделить неоднородную совокупность на однородные группы (классы). Такое разделение в дальнейшем, при проведении статистического анализа, дает лучшие результаты моделирования зависимостей между отдельными признаками.

Параметрические методы распознавания для поиска оптимальных дискриминационных функций используют аппроксимацию функции вероятностного распределения исходных данных и сводятся к определению отношения правдоподобия в различных областях многомерного пространства признаков. Классификатор может быть непосредственно построен из формулы условных вероятностей Байеса и апеллирующей к априорным вероятностям принадлежности объектов к тому или иному распознаваемому классу и условным плотностям распределения значений вектора признаков [5].

Если априорные вероятности появления каждого класса равны, то вероятность того, что вектор x принадлежит классу y_i, равна

$$P_i = \frac{P(x / y_i)}{\sum_{k=1}^{p} P(x / y_k)}.$$

Очевидно, что наибольшая из величин $P(x / y_i)$ и будет обеспечивать наименьшую вероятность неправильной классификации или наименьший средний риск. Решающее правило можно сформулировать следующим образом: вектор измерений x принадлежит классу y_i, если:

$$P(x / y_i) > P(x / y_j) \forall i \neq j.$$

3. Скрытая Марковская модель (СММ) — это статистическая модель, имитирующая работу процесса, похожего наМарковский процесс с неизвестными параметрами, и задачей ставится разгадывание неизвестных параметров на основе наблюдаемых. Полученные параметры могут быть использованы в дальнейшем анализе, например для распознавания образов. СММ может быть рассмотрена как простейшая байесовская сеть доверия.

Первые заметки о скрытых Марковских моделях опубликовал Баум в 1960-х, и уже в 70-х их впервые применили при распознавании речи. С середины 1980-х СММ применяются при анализе биологических последовательностей, в частности ДНК.

Основное применение СММ получили в области распознавания речи, письма, движений и в биоинформатике [6].

Рис. 1, представленный ниже, показывает общую структуру СММ. Овалы представляют собой переменные со случайным значением. Случайная переменная $x(t)$ представляет собой значение скрытой переменной в момент времени *t*. Случайная переменная $y(t)$ — это значение наблюдаемой переменной в момент времени *t*. Стрелки на диаграмме символизируют условные зависимости.

Из рис. 1 становится ясно, что значение скрытой переменной $x(t)$ (в момент времени *t*) зависит только от значения скрытой переменной

$x(t-1)$ (в момент $t-1$). Это называется свойством Маркова. Хотя в то же время значение наблюдаемой переменной $y(t)$ зависит только от значения скрытой переменной $x(t)$ (обе в момент времени t).

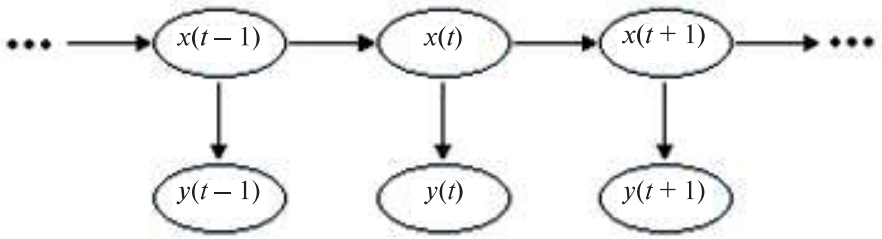

Рис. 1. Структура СММ

Вероятность увидеть последовательность $Y = y(0), y(1), \ldots, y(L-1)$ длины L равна:

$$P(Y) = \sum_{X} P(Y|X) P(X).$$

Предполагается, что сумма пробегает по всем возможным последовательностям скрытых узлов $X = x(0), x(1), \ldots, x(L-1)$.

4. Искусственная нейронная сеть (ИНС, нейронная сеть) — это набор нейронов, соединенных между собой. Как правило, передаточные функции всех нейронов в нейронной сети фиксированы, а веса являются параметрами нейронной сети и могут изменяться. Некоторые входы нейронов помечены как внешние входы нейронной сети, некоторые выходы — как внешние выходы нейронной сети. Подавая любые числа на входы нейронной сети, мы получаем какой-то набор чисел на выходах. Таким образом, работа нейронной сети состоит в преобразовании входного вектора в выходной вектор, причем это преобразование задается весами нейронной сети.

Практически любую задачу можно свести к задаче, решаемой с помощью нейронной сети. В качестве примера сформулируем в терминах нейронной сети задачу распознавания рукописных букв.

Пусть дано растровое черно-белое изображение буквы размером 30×30 пикселов (входной вектор из 900 двоичных символов 30×30 = = 900). Требуется определить, какая это буква (в алфавите 33 буквы). Задача заключается в том, чтобы построить нейронную сеть с 900 входами и 33 выходами, которые помечены буквами. Если на входе нейронной сети присутствует изображение буквы «А», то максимальное значение выходного сигнала достигается на выходе «А». Так же нейронная сеть работает для всех 33 букв. В определенном смысле уровень сигнала на выходе «А» — это достоверность того, что на вход нейронной сети была подана рукописная буква «А» [7].

Описание методологии исследования

Для решения поставленной задачи определения личности говорящего нами был выбран метод дискриминантного анализа распознавания речи на основе сравнения с эталоном.

В среде программирования MATLAB нами была разработана программа, реализующая различение (дискриминацию) речевых сигналов по коэффициенту корреляции Пирсона между корреляционной функцией входного речевого сигнала и корреляционными функциями эталонных речевых сигналов.

Алгоритм разработанной нами программы можно условно разделить на пять этапов:

1. Загрузка данных эталонных аудиофайлов в систему MATLAB.

В программу были загружены три эталонных речевых сигнала, записанные в студии звукозаписи тремя разными голосами, а именно одним женским и двумя мужскими. В качестве примера была записана фраза «Привет, это я», длительность каждого аудиофайла составляет три-четыре секунды. По умолчанию программа считывает файлы в формате wav, данный формат является стандартным несжатым звуковым форматом в Windows. Также по умолчанию установлена разрядность звука (тип сэмпла) только для 32-битных файлов, так как высокая разрядность обеспечивает более широкий динамический диапазон и уменьшает вероятность искажений, но увеличивает размер файла и время обработки. На рис. 2 показана зависимость амплитуды речевого сигнала от количества отсчетов, взятых при считывании данных из аудиофайла (в формате wav) системой MATLAB.

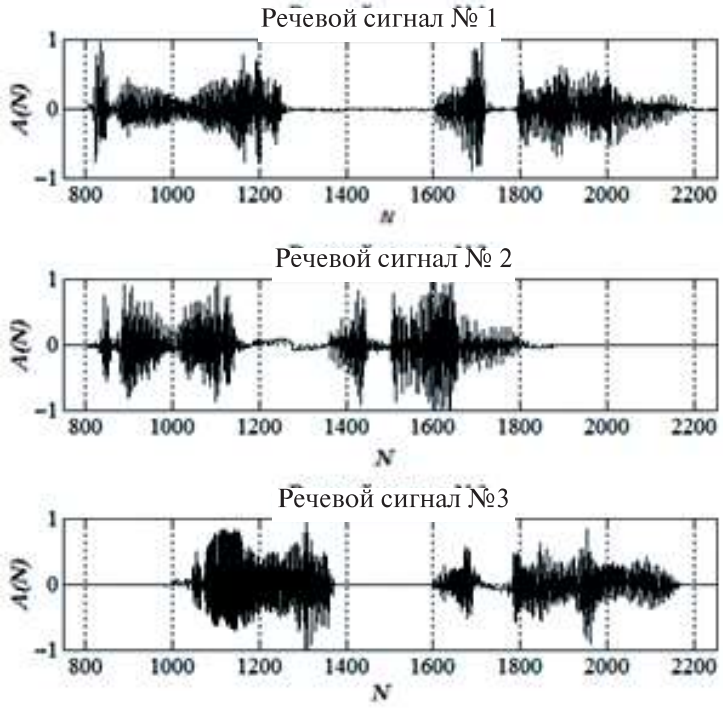

Рис. 2. Зависимость амплитуды речевого сигнала *A* от количества отсчетов *N*

2. Расчет корреляционных функций эталонных речевых сигналов.

Процесс расчета корреляционных функций эталонных речевых сигналов включает в себя расчет автокорреляционных функций, нормирование автокорреляционных функций по абсолютному максимуму, а также вычисление абсолютных значений элементов автокорреляционных функций. На рис. 3 показана зависимость пространственных координат $R1$, $R2$ и $R3$ (соответствующих первому, второму и третьему эталонному речевому сигналу) от времени τ, эквивалентного количеству отсчетов взятых при считывании данных из аудиофайла.

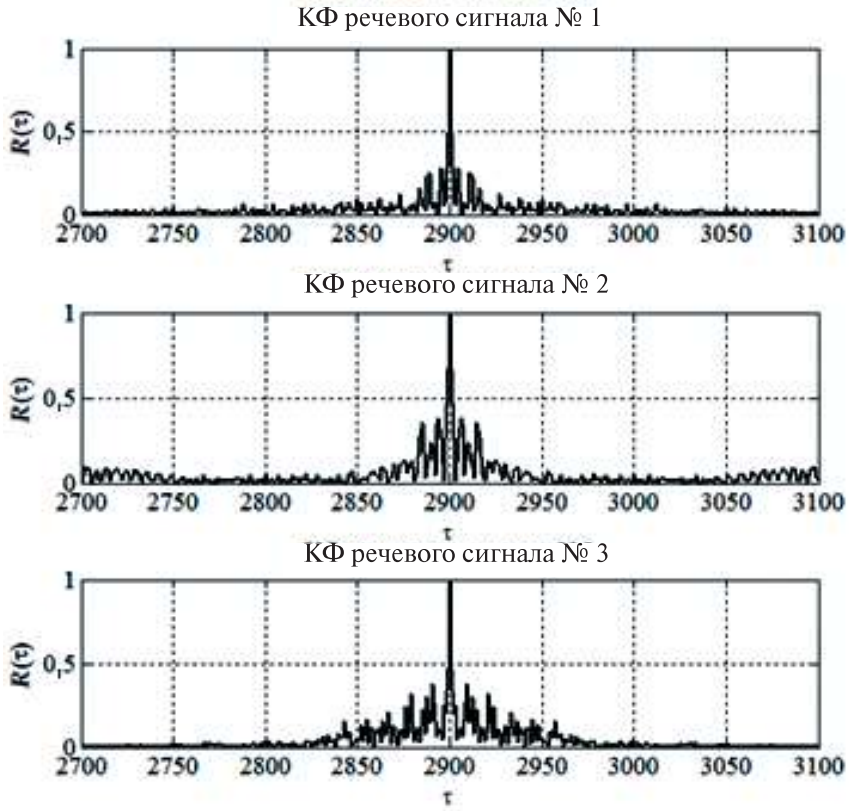

Рис. 3. Корреляционные функции эталонных речевых сигналов

3. Моделирование и расчет корреляционных функций входных речевых сигналов.

В качестве входного сигнала используются эталонные речевые сигналы с наложением шума, распределенного по нормальному закону с математическим ожиданием 0 и среднеквадратическим отклонением 1. В данной программе уровень шума регулируется коэффициентом шума K в пределах от 0 до 1.

Процесс расчета корреляционных функций входных речевых сигналов идентичен процессу расчета корреляционных функций эталонных речевых сигналов, описанному в пункте 2. В качестве примера на рис. 4 представлены корреляционные функции входных сигналов с $K = 0{,}1$.

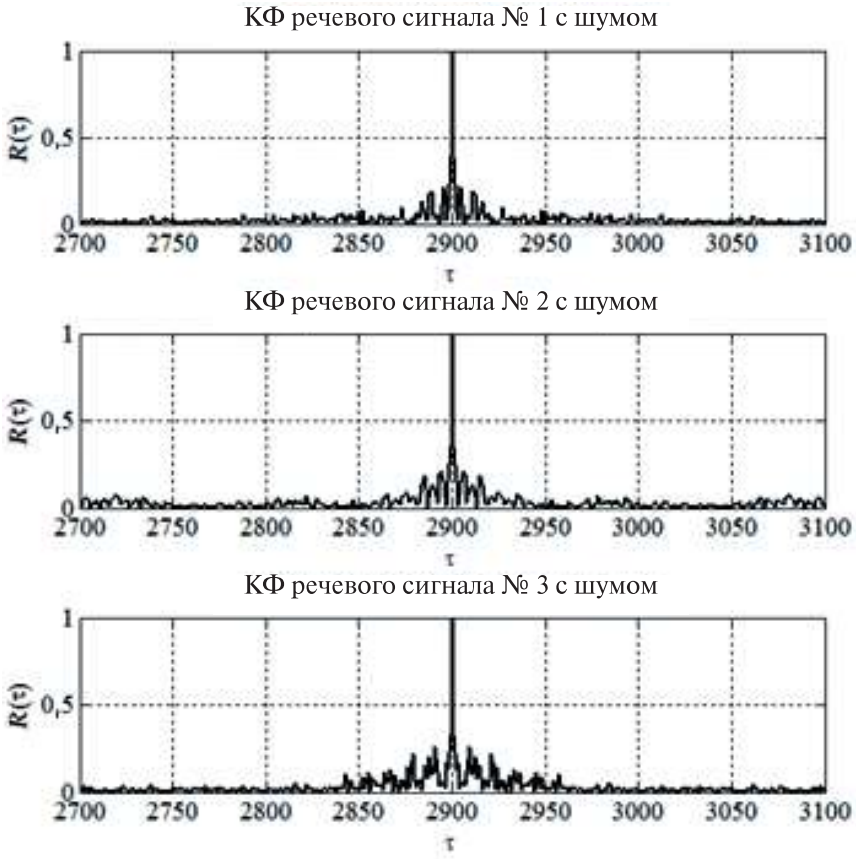

Рис. 4. Корреляционные функции входных сигналов с коэффициентом шума $K = 0{,}3$

4. Расчет взаимной корреляции между корреляционной функцией входного сигнала и корреляционными функциями эталонных речевых сигналов.

В разработанной программе признаком различения (дискриминации) речевых сигналов является коэффициент корреляции Пирсона между корреляционной функцией входного речевого сигнала и корреляционными функциями эталонных речевых сигналов.

В программе был реализован алгоритм расчета коэффициента корреляции Пирсона [8]:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{x}\right)\left(Y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2 \sum_{j=1}^{n}\left(Y_j - \bar{y}\right)^2}}, \qquad (4)$$

где $\left(x_1, y_1\right), \left(x_2, y_2\right), \ldots, \left(x_n, y_n\right)$ – выборка из n наблюдений пар переменных $\left(X, Y\right)$, а $\bar{x}$ и $\bar{y}$ — выборочные средние, определяющиеся по следующим формулам (5):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i ,$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i . \qquad (5)$$

5. Обработка результатов расчетов.

Моделирование и расчет корреляционных функций входных речевых сигналов с коэффициентом шума K и последующий расчет взаимной корреляции между корреляционной функцией входного сигнала и корреляционными функциями эталонных речевых сигналов выполняется 1000 раз. Заданный коэффициент шума $K \in 0{,}05–0{,}8$ с шагом, равным 0,05. Программа идентифицирует речевой сигнал по максимальному коэффициенту корреляции Пирсона и принимает решение об идентификации личности. Полученный результат сохраняется в переменной, соответствующей номеру речевого сигнала. После выполнения 1000 итераций программа выводит на экран таблицу принятия решений того или иного речевого сигналав процентах. На основании данных, полученных в результате расчетов, нами был построен график зависимости коэффициента шума K от вероятности определения истинного речевого сигнала, представленный на рис. 5.

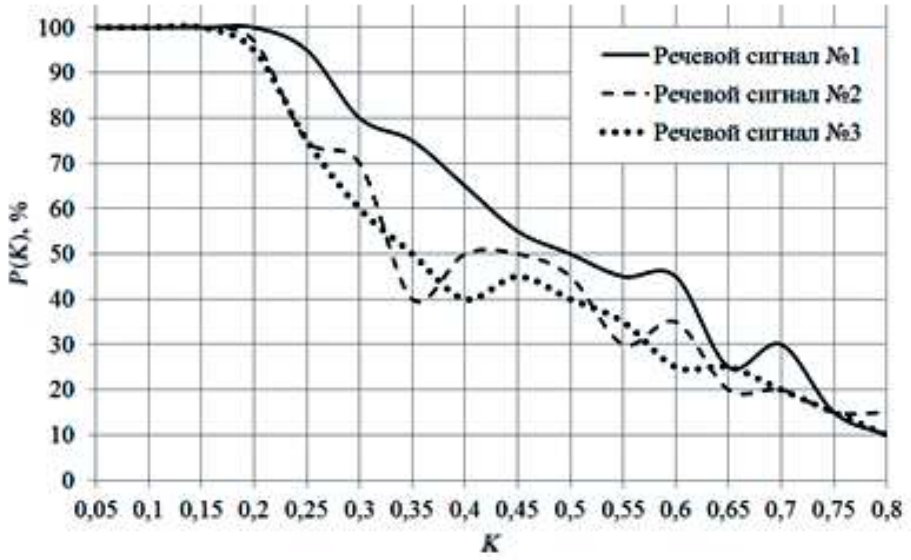

Рис. 5. График зависимости коэффициента шума K от вероятности определения истинного речевого сигнала $P(K)$

Выводы и дальнейшие перспективы исследования

Из полученных нами результатов расчетов, представленных на рис. 5, видно, что с увеличением коэффициента шума уменьшается вероятность определения истинного речевого сигнала. Разработанный нами алгоритм определяет истинный речевой сигнал с вероятностью 60–100 % при коэффициенте шума 0–0,3, что видно из рис. 5.

Достоинством разработанного алгоритма является быстрая скорость вычислений, производимых над 32-битнымиаудио файлами. Причиной выбора такого типа файлов является высокая разрядность, которая уменьшает вероятность искажений. Недостатком разработанного нами алгоритма является отсутствие дополнительного программного алгоритма шумопонижения входного речевого сигнала, предназначенного для увеличения отношения сигнал/шум за счет избыточности либо понижения разрядности или разрешения сигнала. Таким образом, представленная статья рассматривает обобщенное решение поставленной задачи.

Мы планируем продолжить исследования в этом направлении. Это могла бы быть реализация идентификации речевого сигнала не только методом дискриминантного анализа на основе сравнения с эталоном, но и методом динамического программирования, скрытой Марковской модели или методом искусственной нейронной сети.

Литература

1. Глушков В. М., Амосов Н. М., Артеменко И. А. Энциклопедия кибернетики. Киев, 1974. Т. 2. С. 267.

2. Беллман Р. Динамическое программирование. М. : Издательство иностранной литературы, 1960. С. 400.

3. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. М. : Наука, 1965. С. 458.

4. Беллман Р., Калаба Р. Динамическое программирование и современная теория управления. М. : Наука, 1969. С. 118.

5. Шитиков В. К., Розенберг Г. С., Зинченко Т. Д. Количественная гидроэкология: методы системной идентификации. Тольятти : ИЭВБ РАН, 2003. С. 344.

6. Rabiner L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition [Electronic resource]. URL: http://www.cs.cornell.edu/courses/cs4758/2012sp/materials/hmm_paper_rabiner.pdf (date of access: 7.09.2017).

7. Доросинский Л. Г. Основы теории принятия решений. Саарбрюккен : LAP LAMBERT Academic Publishing, 2014. С. 47.

8. Гмурман В. Е. Теория вероятностей и математическая статистика : учебное пособие для вузов // М. : Высшая школа, 2004. С. 251.

Referents

1. Glushkov V. M., Amosov N. M., Artemenko I. A. *Entsiklopediya kibernetiki* (tom 2). Kiev, 1974. 267 p.

2. Bellman R. *Dinamicheskoe programmirovanie*. Moscow, Izdatelstvo inostrannoy literaturyi Publ., 1960. 400 p.

3. Bellman R., Dreyfus S. *Prikladnyie zadachi dinamicheskogo programmirovaniya*. Moscow, Nauka Publ., 1965. 458 p.

4. Bellman R., Kalaba R. *Dinamicheskoe programmirovanie i sovremennaya teoriyau pravleniya*. Moscow, Nauka Publ., 1969. 118 p.

5. Shitikov V. K., Rozenberg G. S., Zinchenko T. D. *Kolichestvennaya gidroekologiya: metody i sistemnoy identifikatsii*. Tolyatti, IEVBRAN Publ., 2003. 344 p.

6. Rabiner L. A. Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. IEEE XploreDigital Library, 1989.

7. Dorosinskiy L. G. *Osnovy i teorii prinyatiya resheniy*. Saarbrucken, LAP LAMBERT Academic Publishing, 2014. 47 p.

8. Gmurman V. E. *Teoriya veroyatnostey i matematicheskaya statistika*. Moscow, Uchebnoe posobie dlya vuzov Publ., 2004. 251 p.

Информация об авторах

Жданов Алексей — магистрант университета имени Фридриха — Александра в Эрлангене и Нюрнберге (Шосплатс 4, Эрланген, Германия).

Доросинский Леонид Григорьевич — профессор Уральского федерального университета имени первого Президента России Б. Н. Ельцина (ул. Мира, 32, Екатеринбург, Россия).

Information about the authors

Aleksei E. Zhdanov is master student at Friedrich-Alexander University Erlangen-Nürnberg (Schlossplatz 4, Erlangen, Germany).

Leonid G. Dorosinskiy is a Professor of the Department of Radioelectronics and Communications of Engineering School of Information Technologies, Telecommunications and Control Systems of Ural Federal University the first President of Russia B. N. Yeltsin (Mira street 32, Ekaterinburg, Russian Federation).