

*Г. К. Бердибаева, О. Н. Бодин, Н. В. Громков,
В. В. Козлов, К. А. Ожикенов, Я. А. Пижонков*

ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД

*G. K. Berdibaeva, O. N. Bodin, N. V. Gromkov,
V. V. Kozlov, K. A. Ozhikenov, Ya. A. Pizhonkov*

APPLICATION OF ARTIFICIAL NEURAL NETWORKS FOR VOICE RECOGNITION

А н н о т а ц и я. *Актуальность и цели.* Целью работы является оценка применения искусственных нейронных сетей для распознавания речевых команд в системах голосового управления. *Материалы и методы.* Применяемый метод основан на анализе голосовых команд нейронными сетями, каждая из которых обучена на соответствующие фонемы естественного языка. Выделенные сегменты речевых команд параллельно подаются на вход каждой сети. В случае распознавания фонемы соответствующая сеть выдаст сигнал, затем группа распознанных фонем анализируется и сравнивается с базовым словарем речевых команд. *Результаты.* Рассмотрены современные методы и средства голосового управления, анализ которых показал необходимость использования метода фонемного декодирования при распознавании речевых команд. Приведено описание примера работы программы разделения голосового сообщения на составляющие фонемы, в результате преобразования речевая команда разделяется на фонемы, являющиеся входными данными для нейросетевого анализа. *Выводы.* Предлагаемый авторами подход при построении систем голосового управления повышает достоверность распознавания речевых команд за счет выявления частотных признаков фонем и последующего их нейросетевого анализа. При этом повышается восприимчивость нейронной сети к входным данным по сравнению с непосредственной подачей на вход нейронной сети речевых команд.

A b s t r a c t. *Background.* It is considered a possibility application of artificial neural networks for voice recognition in voice control applications. *Materials and methods.* The applied method is based on the analysis of the voice commands, neural networks, each trained on respective phonemes of a natural language. Selected segments of speech commands in parallel applied to the input of each network. In the case of the recognition of phonemes corresponding to the network emit a signal, then a group of recognized phonemes are analyzed and compared with the basic vocabulary of voice commands. *Results.* The article considers modern methods and means of voice control, the analysis of which showed the need to use phonemic decoding method in the speech recognition. It is provided, an example, of the work of separation of a voice message into its phonemes. By converting a voice command is divided into phonemes, which are the input to the neural network analysis. *Conclusions.* The proposed approach in the construction of systems increases the accuracy of voice recognition by identifying the frequency characteristics of phonemes and their subsequent neural network analysis. This increases the susceptibility of the neural network to the input data in comparison with the direct input of the neural network speech commands.

К л ю ч е в ы е с л о в а : голосовое управление, распознавания речевых команд, искусственная нейронная сеть, классификация, фонема.

К e y w o r d s : voice control, voice recognition, artificial neural network, classification, phoneme.

Обзор современных методов и средств голосового управления

Системы голосового управления занимают важное место в разработках таких гигантов ИТ-индустрии, как Google, Microsoft, IBM, Intel. Область применения систем голосового управления постоянно расширяется: только в европейском сообществе объем продаж систем голосового управления составляет несколько миллиардов долларов [1]. Решение задачи распознавания речевых команд является центральным элементом системы голосового управления и одним из самых приоритетных в направлениях исследования искусственного интеллекта. Типичная структурная схема системы голосового управления приведена на рисунке 1.

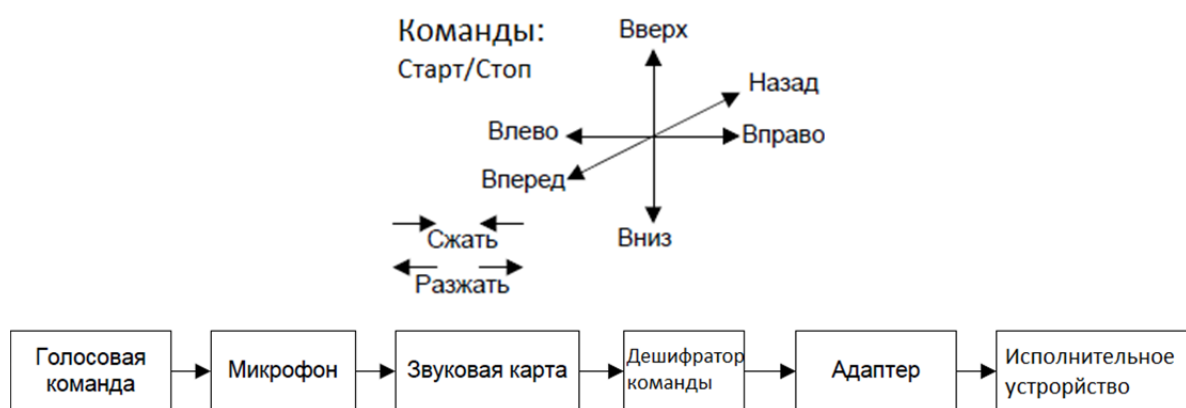


Рис. 1

Как следует из анализа рис. 1, процесс голосового управления включает четыре основных действия:

- *регистрация* речевой команды, реализуемая с помощью микрофона и звуковой карты персонального компьютера;
- *обработка и распознавание* речевой команды, реализуемые с помощью дешифратора команды;
- *исполнение* речевой команды, реализуемое с помощью исполнительного механизма.

Кратко опишем эти действия. На этапе регистрации речевая команда в виде звукового потока данных через микрофон поступает на вход звуковой карты, в которой стандартными средствами осуществляется фильтрация и аналого-цифровое преобразование (АЦП). Критерием фильтрации является изменение уровня громкости. Известно, что во время произношения речевых команд громкость постоянно изменяется, например, гласные звуки громче согласных, при произношении человек делает паузы между слогами. Таким образом, по колебаниям уровня громкости можно отделить речь от шума. При фильтрации речи от шума весь спектр воспринимаемых звуков делится на частотные диапазоны – каналы. В каждом из каналов находится анализатор громкости, отличающий речь от шума, и если в каком-то из каналов появляется звук, по характеристикам похожий на шум, то усиление в этом канале уменьшается. Если же характеристики звука говорят о том, что это речь, то усиление достигает необходимого уровня.

Полученный после фильтрации достаточно «чистый» поток полезной информации поступает на вход АЦП. Количество разрядов АЦП n определяется по формуле

$$D = 6n + 1,8,$$

где $D = 50$ дБ – динамический диапазон (в дБ), изменяющийся от шепота (20 дБ) до громкого разговора (70 дБ);

Очевидно, что обработка и распознавание речевой команды – это основные действия, определяющие функциональные возможности системы голосового управления. Обработка и распознавание речевой команды осуществляются цифровыми средствами обработки сигналов. Разработка эффективной системы голосового управления в настоящее время является важной задачей, требующей создания методов, позволяющих получать высокую достоверность распознавания речевых команд [2, 3].

Предлагаемый подход к построению систем голосового управления

Целью настоящей статьи является оценка применения искусственных нейронных сетей (НС) для распознавания речевых команд.

Для распознавания речевых команд используют различные методы детектирования команд (ДК) [4–6]. При этом алгоритм ДК обеспечивает классификацию сегментов речевого сигнала по типу «речь» или «не речь». В большинстве случаев используют простые и быстрые алгоритмы, построенные на основе пороговых сравнений кратковременных энергий, количества переходов через ноль, корреляционных параметров, энергий спектральных подполос и т.п. На практике чаще имеют дело с нестационарными фоновыми шумами (паразитные хлопки, шелчки и др.), иногда – с шумами значительной интенсивности, например, шум в кабине самолета, автомобиля. В этих случаях задача правильной сегментации речевого сигнала на команды значительно усложняется [7]. Установлено, что простой детектор речевой активности на основе пороговой классификации не способен достоверно детектировать команду.

Задача распознавания речевых команд схожа с задачей классификации, поэтому для решения поставленной задачи наиболее подходят НС, решающие задачи классификации данных [8, 9]. Одной из таких структур является сеть Кохонена, которая может распознавать кластеры в данных, а также устанавливать близость классов [10]. Таким образом, можно улучшить понимание структуры данных, чтобы затем уточнить нейросетевую модель. Если в данных распознаны классы, то их можно обозначить, после чего сеть сможет решать задачи классификации. Сети Кохонена можно использовать и в тех задачах классификации, где классы уже заданы, тогда преимущество будет в том, что сеть сможет выявить сходство между различными классами и должна будет отнести, к какому классу относятся входные данные.

Авторами предлагается алгоритм распознавания речевых команд на основе нейросетевого анализа фонем как минимальных смысловозначительных единиц языка. Суть предлагаемого подхода заключается в следующем: в системе голосового управления реализуется N НС, каждая из которых обучена на соответствующие фонемы естественного языка. Выделенные сегменты речевых команд параллельно подаются на вход каждой из обученных на «свою» фонему НС. В случае распознавания фонемы соответствующая НС выдаст сигнал, и распознанная фонема будет помещена в стек фонем. Стек работает по принципу *FIFO* (*first in – first out*). Группа распознанных фонем анализируется и сравнивается с базовым словарем речевых команд. При совпадении группы распознанных фонем с командой из базового словаря речевых команд, принимается решение о выполнении речевой команды, и она подается на исполнение исполнительному механизму. На основе вышесказанного авторами разработан алгоритм распознавания речевых команд, схема которого приведена на рис. 2.

Для того чтобы можно было подавать большой набор данных напрямую на вход НС, необходимо провести предварительную обработку этих данных с целью уменьшения их объема, а также для выявления участков, по которым будет производиться классификация.

Для этого голосовое сообщение разделяется на составляющие – фонемы, по которым можно выделить пики возрастания и спада уровня сигнала. Предварительная обработка производится в системе MATLAB [9] с помощью функции *envelope()*, которая возвращает верхние и нижние пределы входной последовательности, от величины его аналитического сигнала. Аналитический сигнал находится путем дискретного преобразования Фурье и преобразования Гильберта. Используемая функция *envelope()* сначала удаляет среднее значение, а затем добавляет его обратно после вычисления максимальных и минимальных значений.

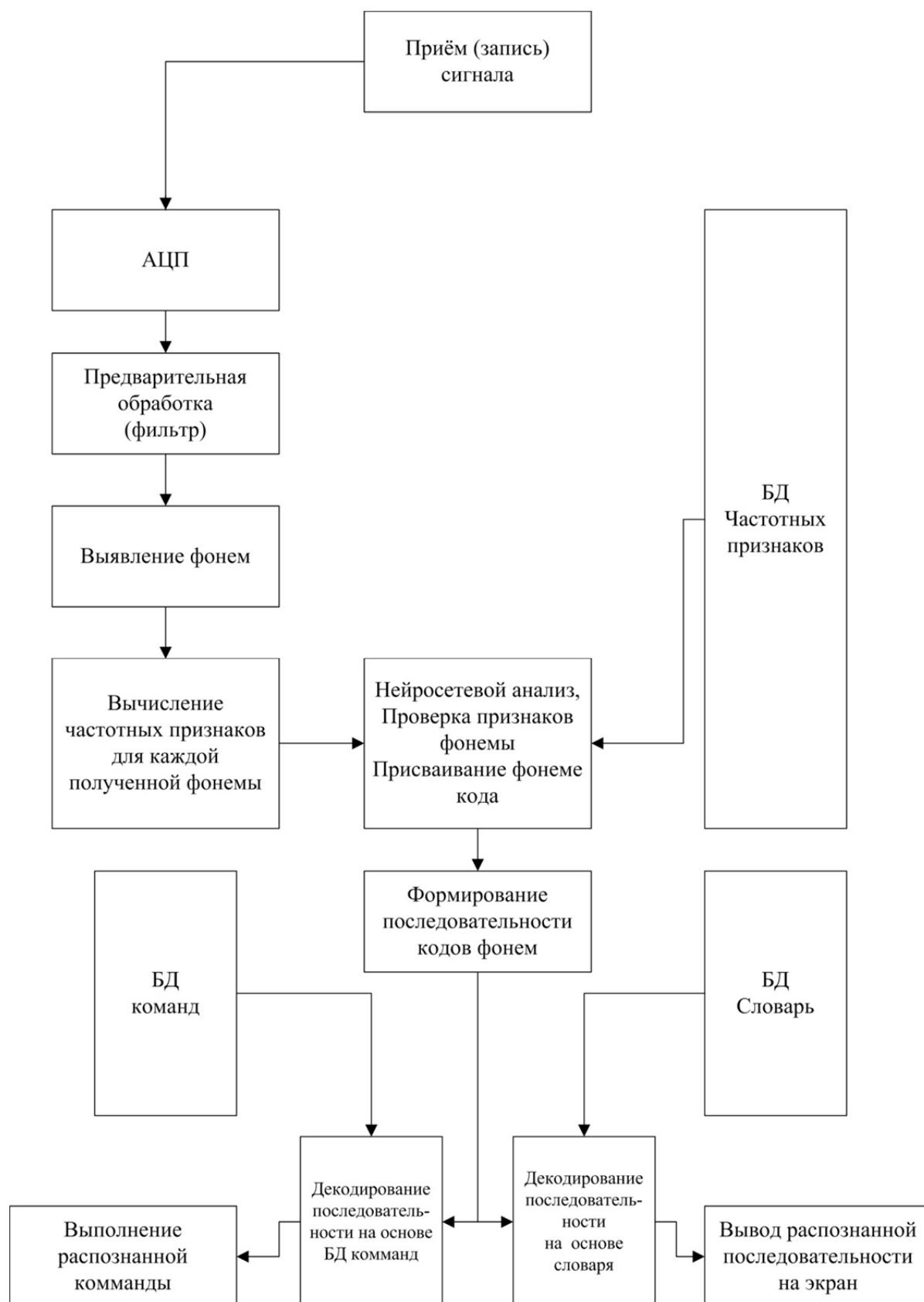


Рис. 2

Исходный сигнал речевой команды «вправо» (см. рис. 1), «оцифрованный» частотой дискретизации 8 кГц, показан на рис. 3.

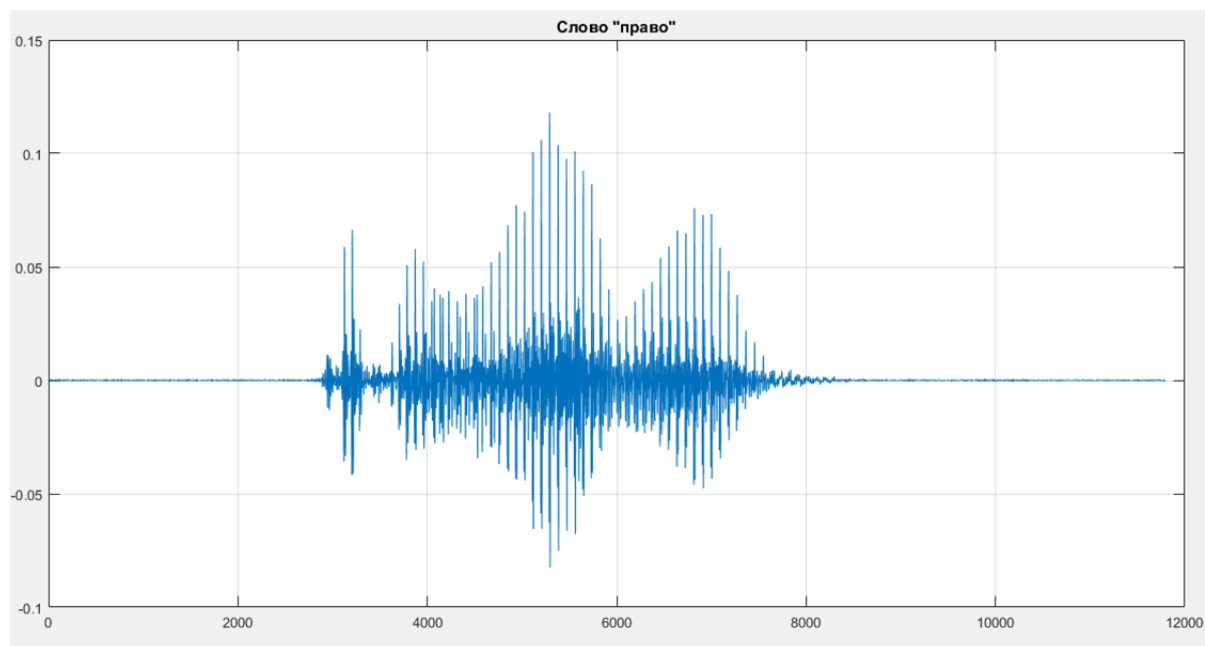


Рис. 3

Затем выделяется огибающая сигнала речевой команды. Первая огибающая строится на среднеквадратичных значениях сигнала. При использовании функции

envelope(nsignal,ceil(length(nsignal)/27.2),'rms');

получается огибающая сигнала речевой команды, представленная на рис. 4.

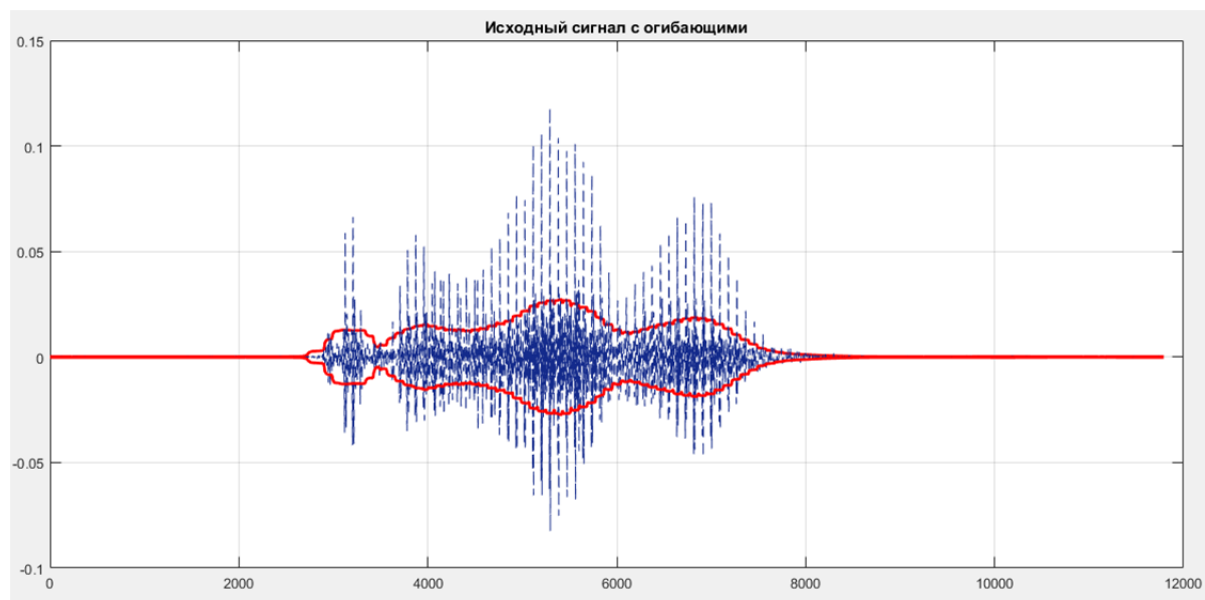


Рис. 4

Затем находится вторая огибающая сигнала речевой команды, которая строится по пикам полученной ранее огибающей. При использовании функции

envelope(up,ceil(length(nsignal)/12.2),'peak');

получается огибающая, представленная на рис. 5.

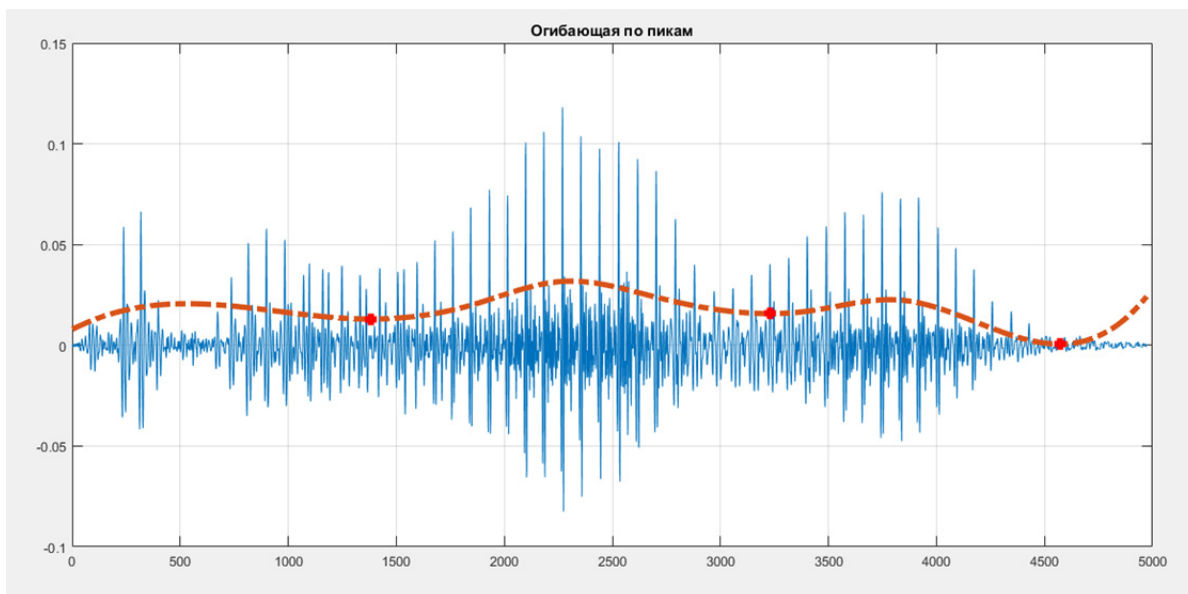


Рис. 5

На рис. 6 точками отмечены координаты, по которым будет происходить разделение на составляющие, т.е. выделение фонем, по которым будет происходить классификация голосового сообщения. Пример работы программы разделения голосового сообщения на составляющие фонемы представлен на рис. 6.

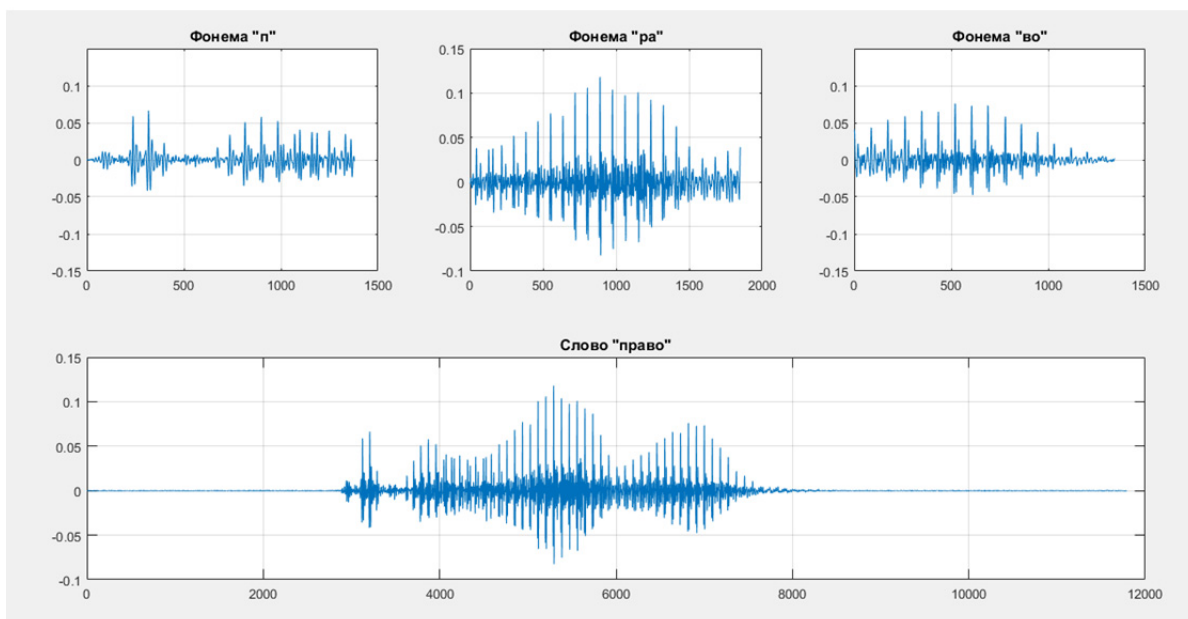


Рис. 6

В результате указанного преобразования речевая команда разделяется на фонемы, являющиеся входными данными для нейросетевого анализа.

Заключение

Проведен критический анализ современных методов и средств голосового управления, который показал необходимость использования метода фонемного декодирования при распознавании речевых команд.

Предлагаемый авторами подход при построении систем голосового управления повышает достоверность распознавания речевых команд за счет выявления частотных признаков

фонем и последующего их нейросетевого анализа. При этом повышается восприимчивость нейронной сети к входным данным по сравнению с непосредственной подачей на вход нейронной сети речевых команд.

Библиографический список

1. Schuster, M. Speech Recognition for Mobile Devices at Google / M. Schuster // LNCS. – 2010. – Vol. 6230. – P. 8–10.
2. Козлов, В. В. Определение параметров гармонических сигналов в условиях действия шумов и помех на основе метода разложения сигнала на собственные числа / В. В. Козлов // Современные проблемы науки и образования. – 2013. – № 6. – URL: <http://www.science-education.ru/113-10860>.
3. Козлов, В. В. Исследование погрешности определения параметров гармонического сигнала на основе метода разложения на собственные числа / В. В. Козлов, Б. Н. Маньжов, Е. А. Ломтев // Измерения. Мониторинг. Управление. Контроль. – 2012. – № 1. – С. 50–55.
4. Пат. 2466468 Российская Федерация. Система и способ распознавания речи / Кочаров Д. А., Хомяков А. Б. – заявл. 30.06.2011 ; опубл. 10.11.2012, Бюл. № 31.
5. Pat. US 8175883 B2, Int. C1 G10L21/00 (2006.01). Speech recognition system and method / Inventor: Grant R., Gregor P. – Assignee: Nuance Communications Inc. – Pub. Date 08.05.2012.
6. Вишнякова, О. А. Алгоритм фонемной сегментации на основе анализа скорости изменения энергии дискретного вейвлет-преобразования / О. А. Вишнякова, Д. Н. Лавров // Информационные технологии. – 2011. – № 4. – С. 146–152.
7. Савченко, В. В. Распознавание речевых команд методом фонетического декодирования слов с подавлением фонового шума / В. В. Савченко // Информационные технологии. – 2016. – № 1. – С. 76–80.
8. Уоссермен, Ф. Нейрокомпьютерная техника: Теория и практика : пер. с англ. / Ф. Уоссермен. – М. : Мир, 1992. – 118 с.
9. Медведев, В. С. Нейронные сети. MATLAB 6 / В. С. Медведев, В. Г. Потемкин ; под общ. ред. В. Г. Потемкина. – М. : Диалог-МИФИ, 2002. – 496 с.
10. Сергиенко, А. Б. Цифровая обработка сигналов / А. Б. Сергиенко. – СПб. : Питер, 2003. – 604 с.

Бердибаева Гульмира Куанышбаевна

докторант,
Казахский национальный исследовательский
университет имени К. И. Сатпаева
(Казахстан, г. Алматы, ул. Сатпаева, 22)
E-mail: horli@mail.ru

Berdibaeva Gul'mira Kuanyshbaevna

doctoral student,
Kazakh National Technical University
named after K. I. Satpayev
(22 Satpayev street, Almaty, Kazakhstan)

Бодин Олег Николаевич

доктор технических наук, профессор,
кафедра информационно-измерительной техники
и метрологии,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: bodin_o@inbox.ru

Bodin Oleg Nikolaevich

doctor of technical sciences, professor,
sub-department of information and measuring
equipment and metrology,
Penza State University
(40 Krasnaya street, Penza, Russia)

Громков Николай Валентинович

доктор технических наук, профессор,
кафедра информационно-измерительной техники
и метрологии,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: ngrom@bk.ru

Gromkov Nikolay Valentinovich

doctor of technical sciences, professor,
sub-department of information and measuring
equipment and metrology,
Penza State University
(40 Krasnaya street, Penza, Russia)

Козлов Валерий Валерьевич

кандидат технических наук, доцент,
кафедра информационно-измерительной техники
и метрологии,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: iit@pnzgu.ru

Ожикенов Касымбек Адилбекович

кандидат технических наук, доцент,
заведующий кафедрой робототехники
и автоматизации оборудования,
Казахский национальный исследовательский
университет имени К. И. Сатпаева
(Казахстан, г. Алматы, ул. Сатпаева, 22)
E-mail: horli@mail.ru

Пижонков Ярослав Андреевич

студент,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: iit@pnzgu.ru

Kozlov Valeriy Valer'evich

candidate of technical sciences, associate professor,
sub-department of information and measuring
equipment and metrology,
Penza State University
(40 Krasnaya street, Penza, Russia)

Ozhikenov Kasymbek Adilbekovich

candidate of technical sciences, associate professor,
head of sub-department of robotics and automation
equipment,
Kazakh National Technical University
named after K. I. Satpayev
(22 Satpayev street, Almaty, Kazakhstan)

Pizhonkov Yaroslav Andreevich

student,
Penza State University
(40 Krasnaya street, Penza, Russia)

УДК 004.934

Применение искусственных нейронных сетей для распознавания речевых команд /
Г. К. Бердибаева, О. Н. Бодин, Н. В. Громков, В. В. Козлов, К. А. Ожикенов, Я. А. Пижонков // Изме-
рение. Мониторинг. Управление. Контроль. – 2017. – № 2 (20). – С. 77–84.