

УДК 519.2

DOI 10.18413/2411-3808-2019-46-2-367-375

ПОСТРОЕНИЕ СИСТЕМЫ РАСПОЗНАВАНИЯ РЕЧЕВЫХ СИГНАЛОВ

CONSTRUCTION OF THE SYSTEM OF RECOGNITION OF VOICE SIGNALS

Хеин Мин Зо¹, В.М. Довгаль¹, В.А. Кудинов²
Hein Min Zaw¹, V.M. Dovgal¹, V.A. Kudinov²

¹Курский государственный университет,
Россия, 305000, г. Курск, ул. Радищева, д. 33

²Курская государственная сельскохозяйственная академия имени проф. И.И. Иванова,
Россия, 305021, г. Курск, ул. Карла Маркса, д. 70

¹Kursk state University, 33 Radishchev St., Kursk, 305000, Russia

²Kursk State Agricultural Academy n.a. prof. I.I. Ivanov,
70 Karl Marx St., Kursk, 305021, Russia

E-mail: heinminzaw13@gmail.com, dovgalvm.prof@yandex.ru, kudinovva@yandex.ru

Аннотация

В работе описаны подходы к разработке системы распознавания речевых сигналов, которая представлена как сложная совокупность различных блоков обработки, для которой устанавливается набор параметров, обеспечивающих наиболее точное описание речевого сигнала в условиях неблагоприятной помеховой обстановки. Рассмотрены два основных подхода при анализе в частотной области – анализ спектра с помощью блока полосовых фильтров и анализ на основе кодирования с линейным предсказанием (LPC-анализ). Для практического примера рассмотрен логарифмический спектр сигнала, полученный обычным преобразованием Фурье и с помощью LPC-коэффициентов). Отмечены основные преимущества LPC-метода в сравнении с анализом с помощью блока фильтров.

Abstract

The paper describes approaches to the development of a speech recognition system, which is presented as a complex set of different processing units, for which a set of parameters is established that provide the most accurate description of a speech signal in an unfavorable jamming environment. Two main approaches are considered when analyzing in the frequency domain - analyzing the spectrum using a bandpass filter unit and analyzing coding with linear prediction (LPC analysis). It was determined that in order to ensure that the parameters are informative with acceptable computational costs for their processing, the spectrum obtained using the discrete Fourier transform is divided into frequency bands and the average signal energy value is calculated in each band. For the parametric description of a speech signal, linear prediction coefficients and reflection coefficients are used. Applying the inverse Fourier transform to the logarithm of the linear prediction spectrum allows us to obtain cepstral coefficients, the advantage of which is their uncorrelatedness and the possibility of reducing the influence of the transmission channel on the parameters of the speech signal. For a practical example, the logarithmic spectrum of the signal obtained by the usual Fourier transform and using LPC coefficients is considered. The main advantages of the LPC-method in comparison with the analysis with the help of a filter block are noted.

Ключевые слова: речевой сигнал, система распознавания, преобразование Фурье, вспомогательные коэффициенты, LPC-коэффициенты, z-характеристика.

Keywords: speech signal, recognition system, Fourier transform, auxiliary coefficients, LPC coefficients, z-characteristic.

В последнее десятилетие успех в развитии вычислительной техники и новейших информационных технологий определяет устойчивую тенденцию к нарастанию сложности систем управления технологическими объектами, а также различными видами человеко-машинных систем. Возможность взаимодействия человека-оператора с технологической машиной на языке, максимально приближенном к естественному, является важной, поскольку общение на естественном языке позволяет организовать эффективное и удобное взаимодействие оператора с технологическим объектом. В настоящее время сфера внедрения систем речераспознавания и голосового управления значительно расширяется, проникая в различные отрасли административной, производственной и бытовой деятельности.

Управление голосом основывается на технологиях распознавания речи – система получает входную информацию о звуковых колебаниях воздуха при помощи микрофона, затем сравнивает полученные команды с данными, записанными в системе и, в случае совпадения, выполняет предписанные указания [Червяков, Кучукова, 2016].

Существующие технологии распознавания речи не имеют пока достаточных возможностей для их широкого использования, но сегодня проводится интенсивный поиск возможностей употребления коротких многозначных слов-процедур для облегчения понимания.

Голосовое управление техническими объектами и вычислительными машинами имеет ряд неоспоримых преимуществ:

- позволяет освободить руки;
- разгружает зрение;
- обеспечить независимость от механических воздействий (вибраций) и условий освещения.

Несмотря на многолетнюю историю и значительные достижения в области создания систем речераспознавания, в полном объеме данная задача остается нерешенной. Особенность современного развития речевых технологий заключается в переосмыслении многолетнего опыта и наработок в данной области на современной элементной и технологической базе.

Для управления сложными техническими объектами человеку-оператору приходится осваивать, а иногда и заново изучать так называемый «язык интерфейса», на что часто тратится достаточно много времени. Поэтому требование эргономичности к разработке человеко-машинных интерфейсов заключается в обеспечении наиболее естественных форм взаимодействия, одной из которых и является речевой диалог [Ломакин, Лифиренко, 2014; Хейн Мин Зо, 2017]. Кроме того, компьютерные системы становятся более доступными людям с нарушением зрения [Серебровский, Филист, 2014; Смоленцев, 2010].

В настоящее время технологии речевого управления уже получили широкое распространение и развиваются в направлениях, указанных на рис. 1.



Рис. 1. Основные направления развития технологий речевого управления

Fig. 1. The main directions of development of speech technology

Сравнение речевого интерфейса с наиболее распространенными в настоящее время средствами взаимодействия пользователя с компьютером (дисплей и клавиатура) позволяет выделить его принципиальные отличия:

1. Недостаток дисплея и клавиатуры заключается в том, что для взаимодействия с компьютером человеку-оператору необходимо пройти специальную подготовку. В то время как речь – это естественный процесс для любого, даже неподготовленного человека. В значительной степени речь снижает психологическое расстояние между человеком и машиной. При появлении речевого интерфейса у машины круг ее пользователей становится практически неограниченным.

2. Речевое управление сокращает физическое расстояние между человеком и машиной, а речь может быть связана с ней через системы коммуникаций, например, микрофон. Это дополнительно расширяет круг потенциальных пользователей такой машиной и делает речевое управление идеальным средством для создания комплексов массового информационного обслуживания.

3. Голосовое управление позволяет общаться с компьютером или технологической машиной в условиях занятости рук рычагами управления, в полной темноте, с закрытыми глазами, с завязанными руками и в другой экстремальной обстановке. При получении информации данное свойство обеспечивает не только оперативность и мобильность общения, но и освобождение рук, разгрузку зрительного канала восприятия, а также облегчает общее психоэмоциональное состояние оператора. Это может быть важно, например, для водителя пассажирского автобуса, диспетчера аэропорта, оператора большой энергетической системы, пилота самолета и т. п.

Несмотря на большое количество проведенных к настоящему времени исследований, современные системы речераспознавания остаются недостаточно изученными, многие проблемы, связанные с процессом автоматического распознавания речи и созданием систем голосового управления, остаются нерешенными [Жилияков, Белов, 2014]. Все это свидетельствует об актуальности проблемы изучения процесса речераспознавания и разработки алгоритмов и методов обеспечения данного процесса.

Современная система речераспознавания – это сложная совокупность различных блоков обработки информации. В условиях действия тех или иных помех достоверность распознавания зависит от работы блока подавления помех, от используемой системы параметров, от выбранного алгоритма распознавания речи и методов адаптации модели. Структура системы речераспознавания представлена на рис. 2.

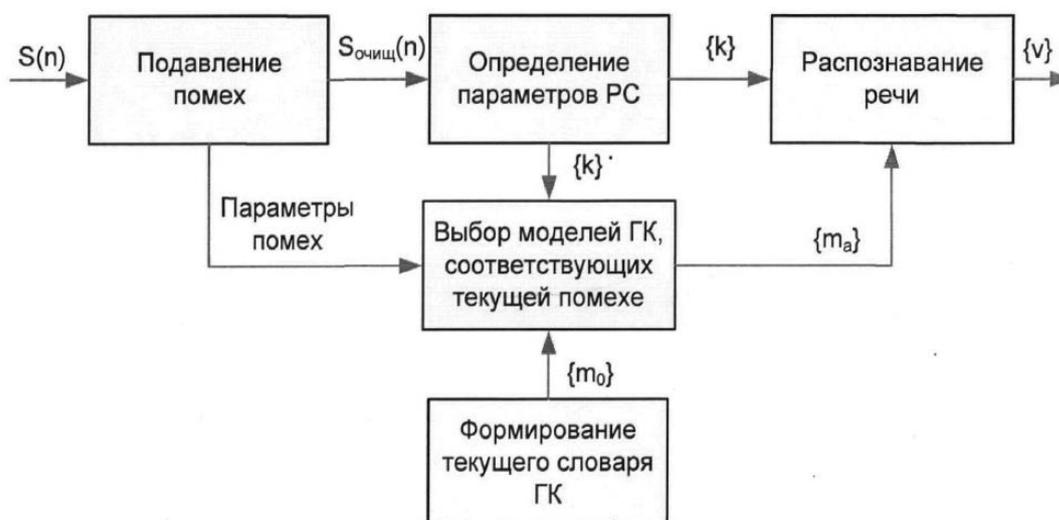


Рис. 2. Структура системы речераспознавания
 Fig. 2. Structure of the recognition system

На вход блока подавления помех поступает дискретизированный речевой сигнал $S(n)$. На вход блока определения параметров сигнала, который может быть введен в состав блока подавления помех, поступает очищенный речевой сигнал $S_{\text{очищ}}(n)$. Тогда оценка помеховой обстановки может выполняться не до оценки параметров речевого сигнала, а на основе анализа его параметров. Выбор набора моделей голосовых команд, соответствующих текущему типу голоса (мужской, женский, детский и пр.) и помеховому фону происходит на основе анализа параметров $\{k\}$ речевого сигнала и сведений о помехах. Учитывая, что системы речераспознавания позволяют быстро менять словарь распознаваемых команд, то в общем случае для каждого распознаваемого способа произнесения может быть подготовлен свой словарь $\{m_o\}$ [Сорокин, 2004].

Заключение $\{v\}$ о соответствии произнесенной команды некоторой модели происходит в блоке распознавания, на вход которого поступают адаптированные к помехе модели $\{m_a\}$. В данном блоке происходит сопоставление изменения параметров речевого сигнала во времени с моделями голосовых команд $\{m_a\}$.

Для обеспечения быстродействия и повышения эффективности речераспознавания необходимо выбрать набор параметров, который обеспечивает наиболее точное описание речевого сигнала в условиях неблагоприятной помеховой обстановки. При этом основными параметрами, которые используются при речераспознавании, являются параметры огибающей спектра речевого сигнала.

Для анализа в частотной области используют [Хейн Мин Зо, 2017; Рылов, 2003] два доминирующих подхода: 1) анализ спектра с помощью блока полосовых фильтров; 2) анализ на основе кодирования с линейным предсказанием (linear predictive coding – LPC).

В основе анализа спектра с помощью блока фильтров лежит кратковременное преобразование Фурье [Lawrence Rabiner, 1993; Сидоренко, Кускова, 2015]. Обычно спектр, полученный с помощью дискретного или быстрого преобразования Фурье, разбивается на полосы частот, а затем вычисляется среднее значение энергии сигнала в каждой полосе. С целью обеспечения максимальной информативности параметров при удовлетворительных вычислительных затратах по их обработке, применяются равномерное, логарифмическое и некоторые другие виды разбиений по полосам. В последнее время наиболее часто используется мелкочастотное преобразование спектра сигнала, которое учитывает особенности человеческого восприятия различных частот [Ефремов, Ефремова, 2014]:

$$Mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

Если сигнал прошел через канал связи, то удобнее работать не со спектром сигнала, а с его кепстром:

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{j\omega})| e^{j\omega n} d\omega, \quad (2)$$

где $X(e^{j\omega})$ – спектр сигнала.

При прохождении речевого сигнала через канал связи он умножается на его частотную характеристику. После перехода в кепстральную область влияние канала передачи можно выразить в прибавлении некоторого постоянного значения. Снизить влияние канала передачи на параметры речевого сигнала можно, если вычесть постоянную составляющую кепстра [Steve Young, 2001; Жилияков, Лихолоб, 2016]. Кепстральный анализ стал активно применяться при обработке речевого сигнала лишь в последние годы, поскольку он требователен к возможностям вычислительной техники.

Описание речевого сигнала на основе кодирования с линейным предсказанием, в основе которого лежит модель речеобразования, также используется при распознавании речевого сигнала. При использовании данного метода речь представляется как отклик

фильтра с переменными параметрами, на вход которого подается сигнал возбуждения. Передаточная функция такой системы имеет следующий вид:

$$H(z) = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}}, \tag{3}$$

где G – коэффициент усиления сигнала возбуждения, a_k – коэффициенты линейного предсказания (КЛП).

Для минимизации математического ожидания квадрата ошибки предсказания, получаемой на выходе анализа, необходимо осуществить подбор коэффициентов a_k соответствующим образом. КЛП производится согласно следующему алгоритму [Рабинер, 1981; Жилияков, Жилиякова, 2015]:

1) Определяются первые $p+1$ элементов автокорреляционной последовательности для заданного окна отсчетов входного сигнала $\{S_n, n=l, N\}$:

$$r_i = \sum_{j=1}^{N-j} S_j S_{i+j}, \tag{4}$$

где $i=0, p$.

2) Рекурсивно находятся КЛП с использованием вспомогательных коэффициентов – коэффициенты отражения (КО) или коэффициенты частной корреляции (КЧО).

В начале ошибка предсказания E устанавливается равной r_0 . Если КО и КЛП обозначить как $\{k_j^{(i-1)}\}$ и $\{a_j^{(i-1)}\}$ (где $i-1$ – порядок фильтра), то КЛП первого порядка определяются в три этапа:

– вычисляется новый набор коэффициентов отражения:

$$k_j^{(i)} = k_j^{(i-1)}, \text{ где } j=1, i-1 \text{ и } k_i^{(i)} = \left\{ r_i + \sum_{j=1}^{i-1} a_j^{i-1} r_{i-j} \right\} / E^{(i-1)}; \tag{5}$$

– определяется значение энергии остатка предсказания

$$E^{(i)} = (1 - k_i^i k_i^j) E^{(i-1)}; \tag{6}$$

– вычисляются коэффициенты линейного предсказания:

$$a_j^{(i)} = a_j^{(i-1)} - k_i^i a_{i-j}^{(i-1)}, \text{ для } j=1, i-1 \text{ и } a_i^{(i)} = -k_i^{(i)}. \tag{7}$$

Этот процесс требуется повторить от $i=j$ до требуемого порядка фильтра $i=p$. Сигнал остатка предсказания $E^{(p)}$, полученный на последнем шаге, соответствует сигналу возбуждения в модели речеобразования.

Для того чтобы осуществить параметрическое описание речевого сигнала, можно использовать как КЛП, так и КО. Для определения дополнительных параметров речи, например, частоты основного тона, используется сигнал остатка предсказания.

Коэффициенты кепстра линейного предсказания (ККЛП) могут являться альтернативой для использования КЛП. ККЛП могут быть получены путем применения обратного преобразования Фурье к логарифму спектра линейного предсказания. При этом спектр линейного предсказания вычисляется как преобразование Фурье от КЛП фильтра. Применение дискретного косинусного преобразования вместо дискретного преобразования Фурье позволяет значительно сократить вычислительные затраты [Баден, Макаров, 2004; Rabiner, Schafer, 2007].

Если сравнивать КЛП с ККЛП, то достоинство последних заключается в их некоррелированности, что дает возможность использовать диагональные матрицы ковариации в

моделях скрытых марковских процессов. Также использование КЛП приводит к уменьшению влияния канала передачи на параметры речевого сигнала.

Расчет КЛП для некоторого сигнала можно осуществить, используя специальный алгоритм расчета [Жиляков, Белов, 2015], позволяющий получить набор числовых коэффициентов, описывающих полюсный фильтр. Полученные коэффициенты позволяют получить выражение для выхода полюсного фильтра как во временной области, так и общий вид его z -характеристики. Z -характеристика дает возможность выполнить полноценный анализ полученного фильтра во временной и частотной областях [Жиляков, Лихолоб, 2016]. Численные значения частот, соответствующих полюсам данного фильтра, можно получить разложением на множители знаменателя полученной z -характеристики. Кроме того, полученные значения позволяют аппроксимировать формантные частоты речевого тракта анализируемого сегмента речевого сигнала. Сигнал-ошибка, полученный с помощью КЛП-метода, которым аппроксимируется процесс звукопреобразования, можно в дальнейшем использовать при компрессии звукового сигнала [Савченко, 2014].

Логарифмические спектры речевого сигнала, полученные обычным преобразованием Фурье (FFT-спектр) и с помощью КЛП-коэффициентов (LPC-спектр) представлены на рис. 3.

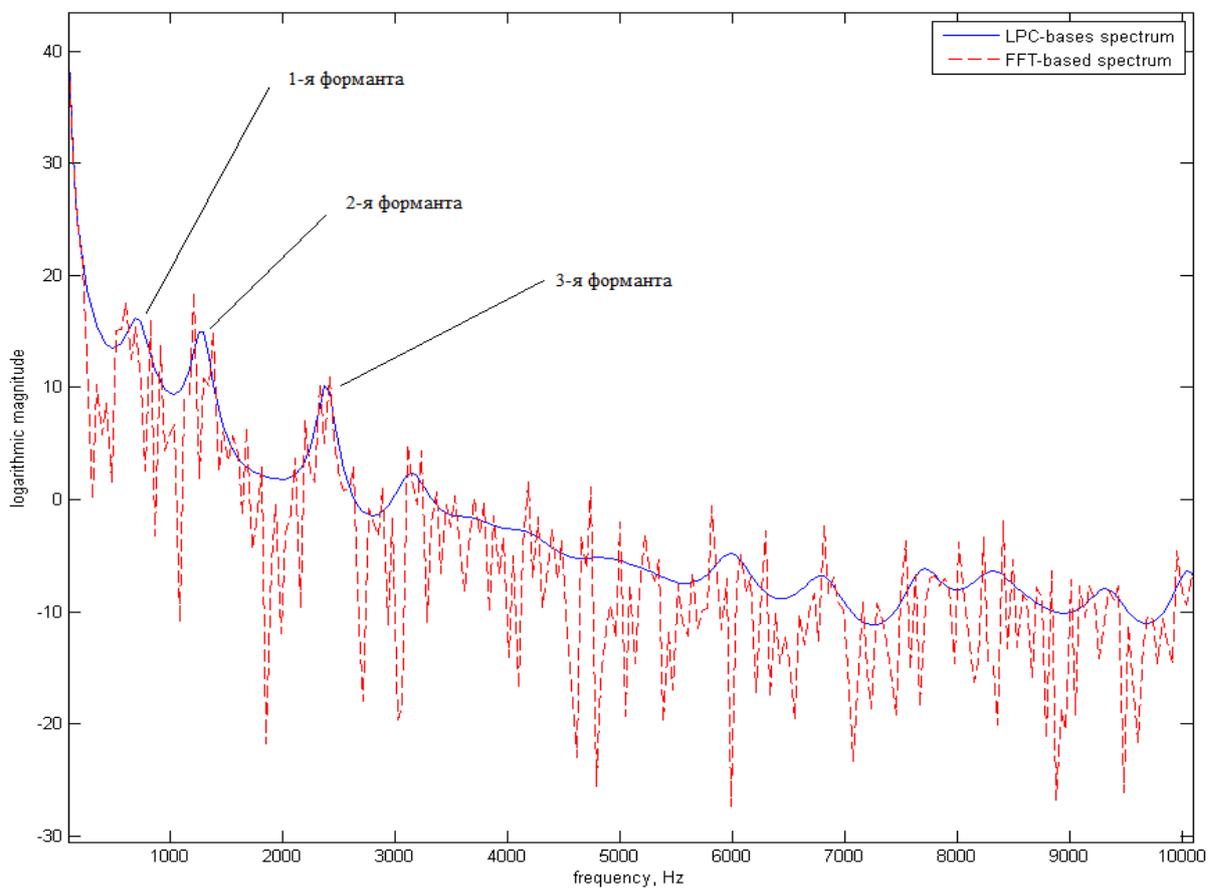


Рис.3 Логарифмический спектр речевого сигнала (анализировалась «а», произнесенная мужским голосом)

Fig. 3 Logarithmic spectrum of the speech signal (the vowel «a» was pronounced, pronounced by a male voice)

В сравнении с анализом звукового сигнала с помощью блока фильтров КЛП-анализ имеет следующие преимущества при обработке речи:

– модель речевого сигнала на основе линейного предсказания может быть представлена аналитически;

- КЛП-анализ дает хорошую аппроксимацию огибающей спектра голосового тракта, что повышает эффективность выполняемого анализа речевых сигналов;
- анализ на основе линейного предсказания позволяет отделить источник возбуждения голосового тракта и голосовой тракт, что позволяет компактно представлять характеристики речевого сигнала и в конечном итоге снижает поток обрабатываемых данных.

Список литературы

References

1. Баден П., Макаров И.С., Сорокин В.Н. 2004. Алгоритм вычисления площадей поперечных сечений речевого тракта. Акустический журнал. 50(6): 739–745.

Baden P., Makarov I.S., Sorokin V.N. 2004. Algorithm vychisleniya ploshchadej poperechnyh sechenij rechevogo trakta [Algorithm for calculating the areas of cross sections of the vocal tract]. Akusticheskij zhurnal [Acoustic magazine]. 50(6): 739–745.

2. Ефремов В.В., Ефремова И.Н., Серебровский В.В., Черепанов А.А. 2014. Информационные системы обработки и сжатия текста. Научные ведомости БелГУ. История. Политология. Экономика. Информатика. 1(172): 182–183.

Efremov V.V., Efremova I.N., Serebrovskij V.V., Cherepanov A.A. 2014. Information systems for text processing and compression. Belgorod State University Scientific Bulletin. History. Political science. Economics. Information technologies. 1(172): 182–183.

3. Жилияков Е.Г., Жилиякова Е.Т., Белов С.П., Белова О.В. 2015. Парные сравнения при анализе фрагментов речи. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 13(210): 145–149.

Zhilyakov E.G., Zhilyakova E.T., Belov S.P., Belova O.V. 2015. Pair comparisons when analyzing speech fragments. Belgorod State University Scientific Bulletin. Economics. Information technologies. 13(210): 145–149.

4. Жилияков Е.Г., Белов С.П. 2014. Об оценивании параметров линейных моделей многомерных сигналов. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 8(179): 83–88.

Zhilyakov E.G., Belov S.P. 2014. On the estimation of the parameters of linear models of multidimensional signals. Belgorod State University Scientific Bulletin. Economics. Information technologies. 8(179): 83–88.

5. Жилияков Е.Г., Белов С.П., Белов А.С., Фирсова А.А. О сегментации речевых сигналов на однородные отрезки. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 7(204): 194–199.

Zhilyakov E.G., Belov S.P., Belov A.S., Firsova A.A. 2015. About segmentation of speech signals into homogeneous segments. Belgorod State University Scientific Bulletin. Economics. Information technologies. 7(204): 194–199.

6. Жилияков Е.Г., Лихолоб П.Г., Курлов А.В., Медведева А.А. 2016. Об однозначности определения идентификационно-значимой частотной полосы в звуках русской речи, подверженных влиянию шума. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 2(223): 167–173.

Zhilyakov E.G., Liholob P.G., Kurlov A.V. Medvedeva A.A. 2016. On the uniqueness of certain identifying significant frequency bands in the sound russian speech exposed to noise. Belgorod State University Scientific Bulletin. Economics. Information technologies. 2(223): 167–173.

7. Жилияков Е.Г., Лихолоб П.Г., Медведева А.А., Прохоренко Е.И. 2016. Исследование чувствительности некоторых мер оценки качества скрытия информации в речевых сигналах. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 9(230): 174–179.

Zhilyakov E.G., Liholob P.G., Medvedeva A.A., Prokhorenko E.I. 2016. Research of sensitivity of some measures quality assessment hidden information in the speech signal. Belgorod State University Scientific Bulletin. Economics. Information technologies. 9(230): 174–179.

8. Ломакин В.В., Лифиренко М.Ф. 2014. Система поддержки принятия решений с автоматизированными средствами корректировки суждений экспертов. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 1(172): 114–120.



Lomakin V.V., Lifirenko M.F. 2014. Decision support system with automated tools for adjusting expert judgment. *Belgorod State University Scientific Bulletin. Economics. Information technologies.* 1(172): 114–120.

9. Рабинер Л.Р., Шафер Р.В. 1981. Цифровая обработка речевых сигналов. Радио и связь, М., 496.

Rabiner L.R., Shafer R.V. 1981. *Cifrovaya obrabotka rechevyh signalov* [Digital processing of speech signals]. Radio i svyaz', M., 496.

10. Рылов А.С. 2003. Анализ речи в распознающих системах. Минск: Бестпринт: 264.

Rylov A.S. 2003. *Analiz rechi v raspoznaushchih sistemah* [Speech analysis in recognition systems]. Minsk: Bestprint: 264.

11. Савченко В.В. 2014. Анализ эмоционального состояния диктора по голосу на основе фонетического детектора лжи. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 21(192): 86–92.

Savchenko V.V. 2014. Analysis of the speaker's emotional state by voice based on phonetic lie detector. *Belgorod State University Scientific Bulletin. Economics. Information technologies.* 21(192): 86–92.

12. Серебровский В.В., Филист С.А. Шаталова О.В., Черепанов А.А. 2014. Генерация структуры и параметров экспертных информационных систем. Научные ведомости Белгородского государственного университета. Экономика. Информатика. 1(172): 150–153.

Serebrovskij V.V., Filist S.A. Shatalova O.V., Cherepanov A.A. 2014. Generation of the structure and parameters of expert information systems. *Belgorod State University Scientific Bulletin. Economics. Information technologies.* 1(172): 150–153.

13. Сидоренко И.А., Кускова П.А. 2015. О повышении точности спектрального анализа фонем при использовании звуковых редакторов. Научные ведомости Белгородского государственного университета. Экономика. Информатика. 7(204): 188–193.

Sidorenko I.A., Kuskova P.A. 2015. On improving the accuracy of the spectral analysis of phonemes using sound editors. *Belgorod State University Scientific Bulletin. Economics. Information technologies.* 7(204): 188–193.

14. Смоленцев Н.К. 2010. Введение в теорию вейвлетов. Ижевск: РХД: 292.

Smolencev N.K. 2010. *Vvedenie v teoriyu vejvletov* [Introduction to Wavelet Theory]. Izhevsk: RHD: 292.

15. Сорокин В.Н. 2004. Структура проблемы автоматического распознавания речи. Информационные технологии и вычислительные системы. М., 2: 25–40.

Sorokin V.N. 2004. *Struktura problemy avtomaticheskogo raspoznavaniya rechi* [The structure of the problem of automatic speech recognition]. *Informacionnye tekhnologii i vychislitel'nye sistemy* [Information technology and computing systems]. M., 2: 25–40.

16. Хеин Мин Зо. 2017. Основные преимущества использования вейвлет-анализа в процессе обработки речевых сигналов и изображений. Физико-математические и технические науки как постиндустриальный фундамент эволюции информационного общества: Сб. статей Межд. науч.-практ. конф. Уфа, АЭТЕРНА: 214–218.

Hein Min Zo. 2017. *Osnovnye preimushchestva ispol'zovaniya vejvlet-analiza v processe obrabotki rechevyh signalov i izobrazhenij. Fiziko-matematicheskie i tekhnicheskie nauki kak postindustrial'nyj fundament ehvolyucii informacionnogo obshchestva* [The main advantages of using wavelet analysis in the processing of speech signals and images. Physical, mathematical and technical sciences as a post-industrial foundation for the evolution of the information society]: *Sb. statej Mezhd. nauch.-prakt. konf. Ufa, AETHERNA*: 214–218.

17. Хеин Мин Зо. 2017. Современное состояние проблемы анализа речевых сигналов. Воздействие научно-технической революции на характер связи науки с производством: Сб. статей Междунар. науч.-практ. конф. Уфа: АЭТЕРНА: 99–102.

Hein Min Zo. 2017. *Sovremennoe sostoyanie problemy analiza rechevyh signalov. Vozdejstvie nauchno-tekhnicheskoy revolyucii na harakter svyazi nauki s proizvodstvom* [The current state of the problem of analyzing speech signals. The impact of scientific and technological revolution on the nature of the relationship of science with production]: *Sb. statej Mezhdunar. nauch.-prakt. konf. Ufa: AETHERNA*: 99–102.

18. Червяков Н.И., Кучукова Н.Н. 2016. Проблемы автоматического распознавания слитной речи. Методы обработки исходного речевого сигнала. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 23(244): 148–154.



Chervyakov N.I., Kuchukova N.N. 2016. Problems of automatic recognition of continuous speech. Methods for processing the original speech signal. Belgorod State University Scientific Bulletin. Economics. Information technologies. 23(244): 148–154.

19. Lawrence Rabiner, Bing-Hwang Juang. 1993. Fundamentals of speech recognition. Prentice Hall PTR, Englewood Cliffs, NJ 07632: P5070.

20. Rabiner L.R., Schafer R.W. 2007. Introduction to Digital Speech Processing. Foundations and Trends R in Signal Processing, 1(1–2): 1–194.

21. Steve Young, Gunnar Evermann, Dan Kershaw and other. 2001. The HTK book (for HTK Version 3.1). Speech group, Cambridge University Engineering Department, December.