УДК 534.734

О ПОВЫШЕНИИ ТОЧНОСТИ СПЕКТРАЛЬНОГО АНАЛИЗА ФОНЕМ ПРИ ИСПОЛЬЗОВАНИИ ЗВУКОВЫХ РЕДАКТОРОВ

ON IMPROVING THE ACCURACY OF SPECTRAL ANALYSIS OF PHONEMES USING AUDIO EDITORS

И.А. Сидоренко, П.А. Кускова
I.A. Sidorenko, P.A. Kuskova

Белгородский государственный национальный исследовательский университет, Россия, 308015, Белгород, ул. Победы, 85
Belgorod State National Research University, 85 Pobeda St, Belgorod, 308015, Russia

e-mail: Sidorenko@bsu.edu.ru

Аннотация. В статье предложен способ повышения точности анализа формантной структуры звуков речи на основе формирования «идеализированных» фонем, полученных из квазистационарных фрагментов исходных фонем. Показано, что на основе предложенного подхода можно детализировать спектр любого фрагмента фонемы, как для вокализованных, так и не вокализованных звуков речи, а также любых стохастических сигналов при использовании стандартной процедуры быстрого преобразования Фурье в звуковых редакторах.

Resume. The article proposes a method to improve the accuracy of the analysis of the formant structure of speech sounds on the basis of the formation of the "idealized" phonemes obtained from quasi-stationary fragments of source phonemes. It is shown that the proposed approach can be further defined range of any fragment of phonemes for both vocalized and not vocalized speech sounds, as well as any stochastic signals using the standard procedure of the fast Fourier transform in audio editors.

Ключевые слова: имитационное моделирование, транспортный поток, технологический процесс, исчисление объектов, язык моделирования производственных процессов, системно-объектный подход «Узел-Функция-Объект».

Keywords: фонема, форманта, звуковой редактор, быстрое преобразование Фурье.

Введение

Самым естественным и востребованным средством общения между людьми была и остается речь, поэтому интерес к разработке технологий обработки речевых сигналов остается в центре внимания специалистов в области инфокоммуникационных систем. Об этом свидетельствует наличие научных публикаций на эту тему, например [1-5]. Сама речь по своей природе является уникальным сигналом, а сущность процесса речевого общения людей так до конца и не раскрыта. Именно поэтому в основе технологий применяемых для разработки инфокоммуникационных систем используются различные подходы, в том числе основанные на анализе фонемной и формантной структур речи [2, 3]. При подготовке специалистов в области речевых инфокоммуникационных технологий полезно продемонстрировать фонемную и формантную структуры речи. Для этих целей логично было бы применить звуковые редакторы, позволяющие в реальном масштабе времени производить анализ и обработку звуковых сигналов. Однако, попытка их применения для спектрального анализа звуков речи не увенчалась успехом [1].

В статье [1] были рассмотрены причины, не позволившие продемонстрировать формантную структуру звуков русской речи при использовании звуковых редакторов Adobe Audition®, Sound Forge® , Audacity® и им подобным. Для спектрального анализа звуков в указанных редакторах применяется стандартная процедура быстрого преобразования Фурье (БПФ), которая предусматривает разбиение исследуемого звукового сигнала на сегменты заданной размерности отсчётов N, причём кратной целой степени числа 2, т.е. $N=2^n$. Как правило, в звуковых редакторах могут задаваться значения $n \geq 6$, что позволяет формировать сегменты звуков речи из 64, 128, ... , 1024 и т.д. отсчетов. Это ограничение на размер сегмента приводит к невозможности точного согласования размера выборки с длительностью отдельных звуков речи, которая в общем случае является случайной величиной. В результате спектральному анализу подвергаются некие сегменты звуков, размер которых меньше, либо (чаще всего) больше длительности реальных фонем.

Кроме этого, в [1] было показано, что сегменты звуков речи подвергаются дополнительным искажениям, возникающим в результате применения оконных функций, необходимых для устранения эффекта Гиббса. Эти искажения можно было бы существенным образом ослабить, увеличив размер выборки N таким образом, чтобы она заключала в себе большое число периодов исследуе-

мой фонемы. Однако это редко оказывается возможным по двум причинам. Во-первых, невокализованные звуки речи (краткие согласные «н», «к» и т.п.) имеют малую длительность, поэтому выделить сегмент звука, содержащий несколько одинаковых фонем практически невозможно. Во-вторых, даже у вокализованных звуков речи на различных участках их звучания форма фонемы претерпевает существенные изменения, которые всегда сопровождаются перераспределением энергии в спектральной области. Кроме этого, дополнительный интерес представляет анализ изменения формантной структуры фонемы на различных участках её существования: атаке, стационарной части, затухания. Попытки улучшить ситуацию варьированием размером выборки N, повышением частоты дискретизации звуковых сигналов, применением различных оконных функций не привели к положительным результатам. Поэтому в статье [1] был сделан вывод о том, что использование в учебном процессе звуковых редакторов для демонстрации формантной структуры фонем речи невозможно, поскольку вычисляемые ими спектры не отражают известных результатов. Для решения этой задачи следует использовать программное обеспечение, предоставляющее пользователю полную свободу в выборе параметров размера выборки анализируемого фрагмента и вычисляющего спектр по алгоритмам, не предусматривающим обязательную кратность размера окна БПФ степени числа 2. Такие условия могут быть реализованы, например, в программе MATLAB®, дающей возможность произвольного задания параметров дискретного преобразования Фурье.

Целью статьи является изложение способа повышения точности отображения формантной структуры звуков речи, при использовании стандартной процедуры вычисления спектра на основе БПФ, применяемой в звуковых редакторах.

Постановка задачи

Для проведения исследований была использована прикладная программа MATLAB®, в которой реализованы все доступные алгоритмы вычисления спектра сигналов.

Прежде всего, нужно было решить вопрос с выбором критерия для оценки точности вычисления спектра. Известно [6], что при спектральном анализе детерминированных сигналов можно аналитически рассчитать точные значения коэффициентов ряда Фурье, которые следует использовать в качестве эталонных величин при анализе результатов вычисления спектра таких сигналов различными способами. Фонемы же не являются детерминированными сигналами, поскольку не обладают устойчивой формой, которая всегда изменяется в зависимости от места звука в слове, особенностей голоса говорящего и ряда других факторов. В этом смысле фонему часто сравнивают с буквами, написанными людьми с разным почерком. Следовательно, спектр каждой фонемы является уникальным и не предсказуемым. Усреднённые оценки спектра возможны и широко используются при разработке инфокоммуникационных технологий и устройств обработки речи. Однако в ряде случаев практический интерес представляет именно точная оценка формантной структуры звуков речи. Например, при решении задачи идентификации голоса диктора важно знать именно уникальные, присущие конкретному человеку особенности произношения звуков или даже отдельных их частей. Именно такую цель, - сохранение уникальных спектральных параметров исследуемых звуков речи и ставили перед собой авторы данной статьи. Учитывая изложенное, было принято решение осуществлять оценку точности вычисления спектра следующим образом: выбрать один из алгоритмов вычисления спектра в качестве эталонного, определив предварительно условия его применения, а затем сравнивать полученные с его помощью контрольные результаты со значениями, полученными при использовании других алгоритмов.

В статье [1] было сделано предположение о том, что искажения спектра исследуемых звуков речи обусловлено применением алгоритма БПФ, а именно: не возможностью согласовать размер анализируемой выборки с длительностью исследуемой фонемы и последующим применением оконных функций для устранения эффекта Гиббса. Следовательно, алгоритм БПФ не может быть выбран в качестве эталонного, а, наоборот, полученные с его помощью результаты должны сравниваться с контрольными значениями, полученными с помощью алгоритма, выбранного эталонным. Дальнейший ход рассуждений состоял в следующем: для устранения искажений, вызванных эффектом Гиббса, следует избавиться у исследуемых сегментов звука от точек разрыва первого рода, т.е. скачков напряжения сигнала на границах сегмента. Тогда не потребуется применение оконных функций, искажающих форму анализируемого сегмента. Достичь желаемого результата можно в том случае, если сегментирование звука выполнять в ручную, выделяя начало и конец исследуемой фонемы в моменты времени перехода графика уровнеграммы через ноль. У такого способа сегментирования есть и ещё одно преимущество: можно проводить исследование спектральной структуры звука на любом отрезке его существования - атаке, стационарной части или затухания. Очевидно, что размер полученной таким образом выборки сигнала не будет характеризоваться числом, кратным степени двойки, поэтому применение стандартного для звуковых ре-

дакторов алгоритма БПФ не возможно. В этом случае спектр может быть вычислен по формуле дискретного преобразования Фурье:

$$F(k)=\sqrt{\left(\sum_{i=0}^{N-1}x[i]\cos(2\pi ki/N)\right)^2+\left(-\sum_{i=0}^{N-1}x[i]\sin(2\pi ki/N)\right)^2} \quad (1)$$

Результаты расчета по формуле (1) и следует использовать в качестве контрольных, а сам метод вычисления ДПФ считать эталонным.

Результаты вычислительных экспериментов

Для проведения вычислительных экспериментов были использованы звуковой редактор Adobe Audition® и программа MATLAB®, в которой спектр вычислялся двумя способами: с помощью встроенной функции fft(.), реализующей алгоритм БПФ с прямоугольным окном; с помощью написанной программы, реализующей вычисление ДПФ по формуле(1).

На рис.1 представлены выделенная указанным выше способом фонема звука «*и*» с частотой дискретизации Fд=8 кГц (*а*) и результаты её спектрального анализа методом БПФ с использованием окна Хэннинга, N=64 в редакторе Adobe Audition® (*б*), вычисленные в программе MATLAB® с помощью встроенной функции БПФ, N=64 (*в*) и с помощью полной формулы ДПФ (*г*).

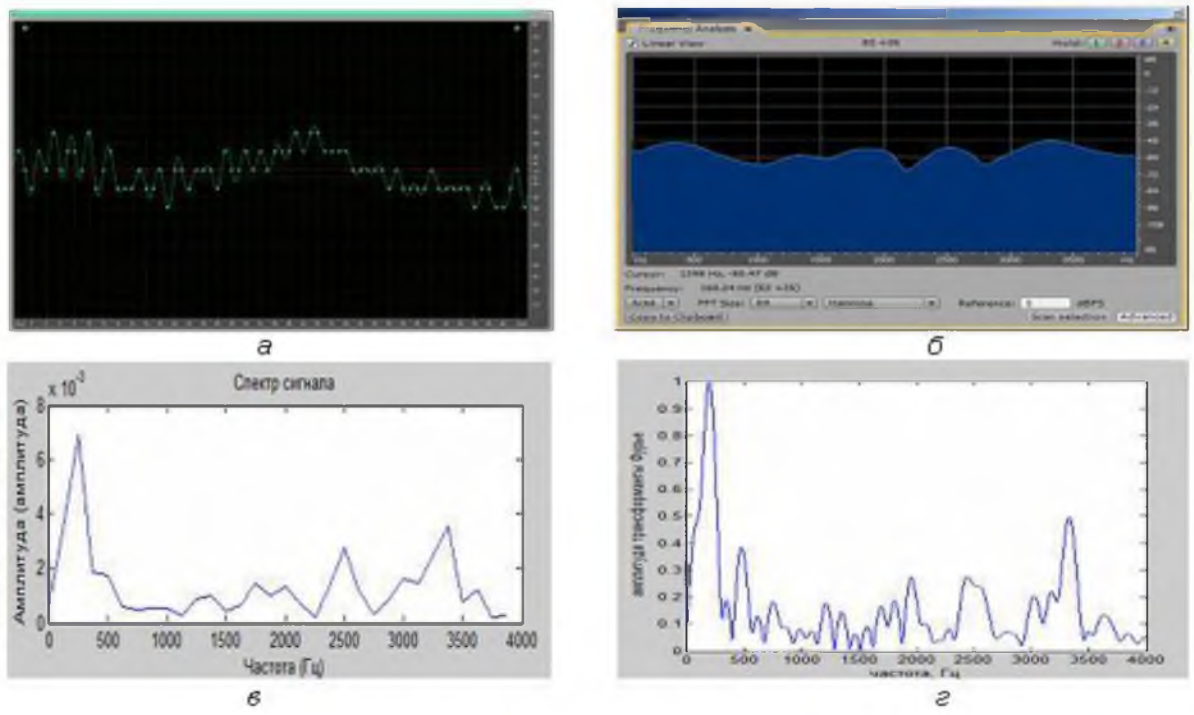

Рис.1. Уровнеграмма и вычисленные спектры фонемы звука «и»
Fig.1. Signalosome and calculated spectra phonemes sound "i"

Сравнение спектров на рисунке 1,*а,б,в*, показывает их существенное различие. На рисунке 1,*б* спектр выглядит как флуктуирующая функция с примерно одинаковым, равномерным распределением энергии во всей полосе частот сигнала. Спектры, вычисленные в программе MATLAB®, дают совершенно другую картину распределения энергии – сильно выделяется область в районе 250 Гц, просматриваются также локальные формантные области на частотах 2000 Гц, 2500 Гц и 3400 Гц, имеющие явно более низкую концентрацию энергии. Для правильной интерпретации этих спектров следует провести огибающую линию, плавно соединяющую пиковые значения спектральных компонент. Очевидно, что применение звукового редактора Adobe Audition® не даёт удовлетворительного результата и может вызвать неправильную оценку распределения энергии фонемы в ходе образовательного процесса. Похожая картина наблюдалась и при вычислении спектра фонем других звуков речи. На рис. 2 приведены аналогичные результаты для фонемы звука «*з*».

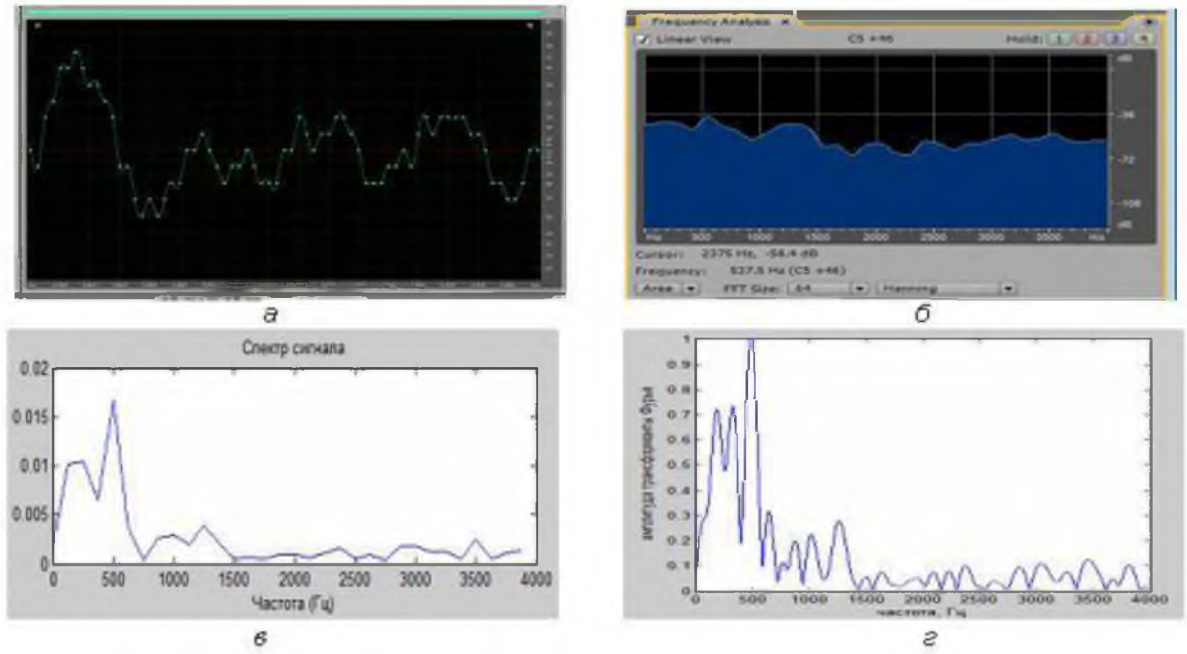

Рис. 2. Уровнеграмма и вычисленные спектры фонемы звука «з»
Fig.2. Signalosome and calculated spectra phonemes sound "z"

На следующем этапе исследований было сделано предположение о том, что для повышения точности спектрального анализа с помощью звуковых редакторов можно синтезировать «идеальный» для анализа звук, образованный многократным копированием одного образца фонемы, не имеющей точек разрыва (скачков напряжения) в начале и в конце. Такая операция легко реализуема с помощью опции «вставка», имеющейся во всех звуковых редакторах.

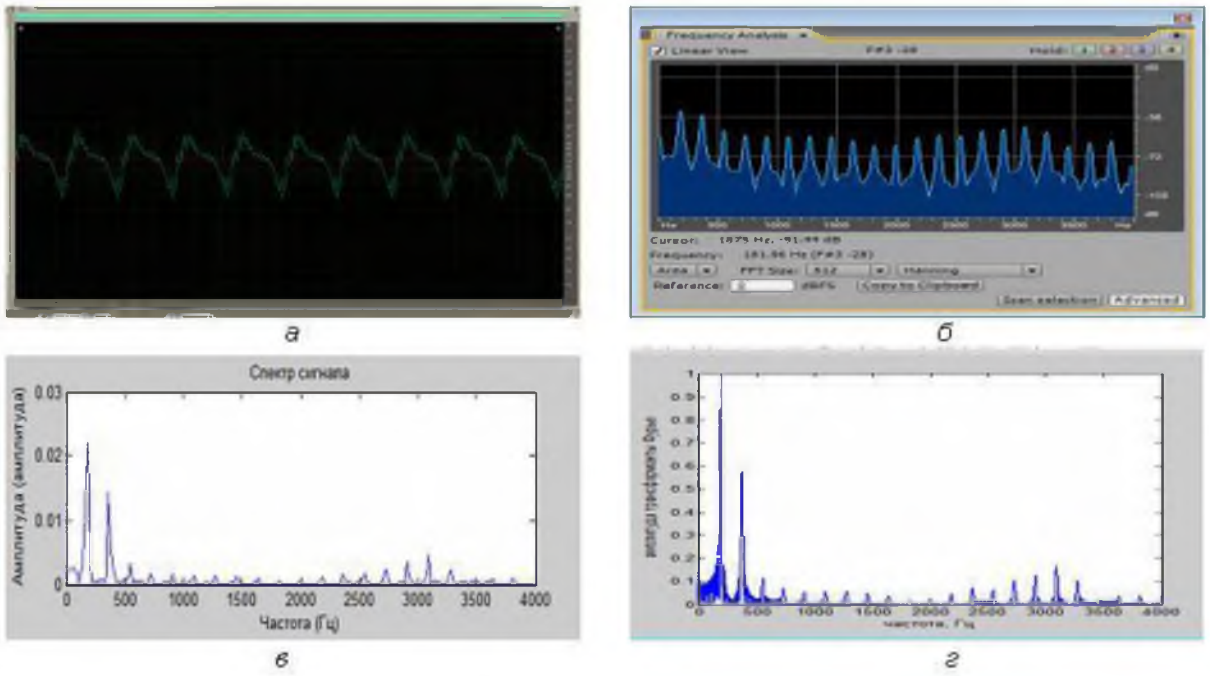

Рис.3. Уровнеграмма и вычисленные спектры 10 фонем звука «и»
Fig.3. Signalosome and calculated spectra of 10 phonemes sound "i"

На рис. 3 представлены результаты вычисления спектра звукового сигнала, синтезированного из 10 периодов фонемы звука «и ». Общая длительность полученного сигнала равнялась 440 отсчетам. Для анализа спектра методом БПФ в обоих случаях выбирался ближайший из больших размер окна N=512>440. Соответственно, ДПФ вычислялось точно для 440 отсчетов.

Сравнение соответствующих результатов вычисления спектров на рис. 1 и рис.3 показывает, что синтезированный звук даёт более четкую картину спектра, который приобретает ярко выраженную периодическую структуру и облегчает локализацию формантных областей. При этом результаты расчетов для БПФ и ДПФ, выполненные в программе MATLAB® для синтезированной «идеальной» фонемы очень близки, что свидетельствует о повышении точности вычисления спектра при БПФ, несмотря на различие в объеме выборки сигнала и размера окна анализа. Существенно улучшился и вид спектра в программе Adobe Audition® - огибающая спектра стала более выразительной и в целом повторяет поведение огибающей спектров, полученных в программе MATLAB®. Для большей выразительности распределения энергии следует использовать не логарифмическую шкалу амплитуд, а линейную, как в программе MATLAB®.

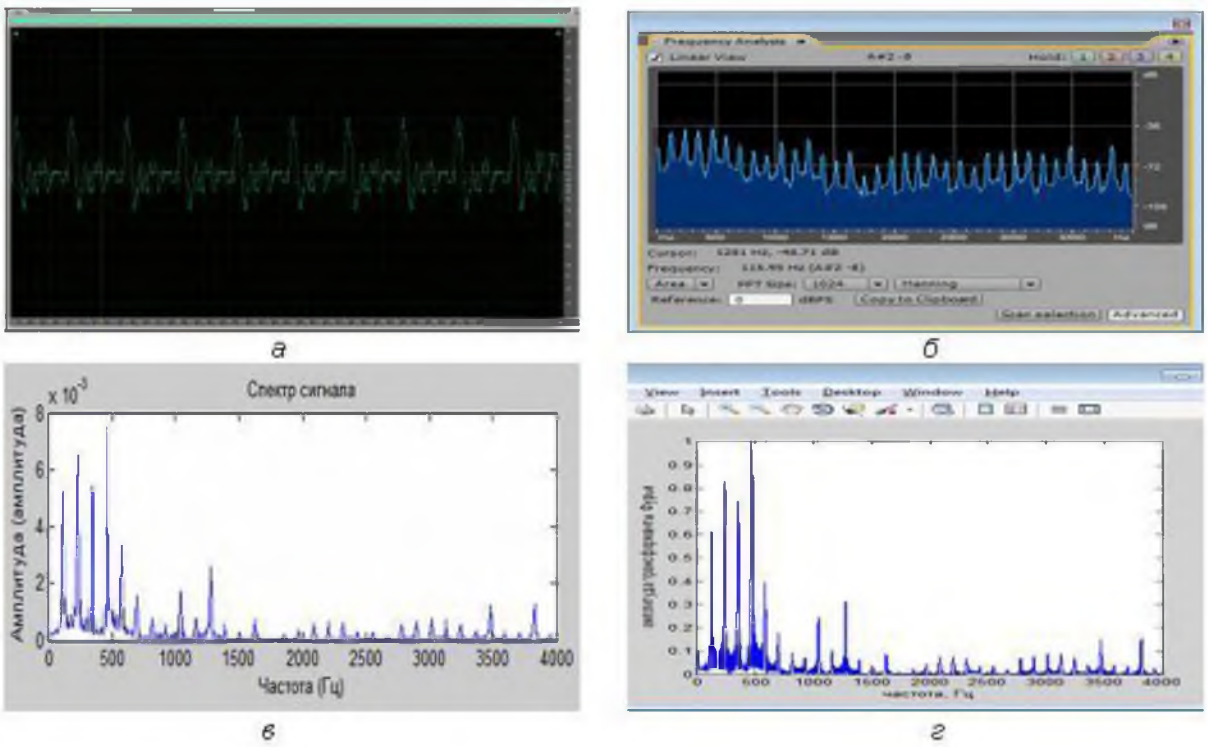

Рис. 4. Уровнеграмма и вычисленные спектры 10 фонем звука «з»
Fig. 4. Signalosome and calculated spectra of 10 phonemes sound "z"

На рис.4. приведены результаты вычисления спектра для «идеальной» фонемы звука «з», состоящей из 10 периодов фонемы, общим объемом 690 отсчетов. Размер окна для БПФ выбирался равным 1024 отсчетам. Также как и в предыдущем случае заметно существенное улучшение локализации формантных областей.

Заключение

Предложенный способ вычисления формантных областей, предусматривающий формирование «идеальной» фонемы путем многократного копирования одного периода исследуемой фонемы позволяет повысить точность спектрального анализа при применении стандартной процедуры БПФ в общем случае и при использовании звуковых редакторов в частности. Необходимо также отметить универсальность предложенного подхода для спектрального анализа любых нестационарных или коротких по длительности сигналов, если предметом исследования является их мгновенный спектр. Фактически предложенный подход частично устраняет самый главный недостаток преобразования Фурье – присущую ему частотно-временную неопределенность.

Список литературы
References

1. Сидоренко И.А., Кускова П.А. О спектральном анализе фонем с использованием звуковых редакторов // Научные ведомости Белгородского государственного университета №1(144) 2013 выпуск 25/1, серия Информатика, Белгород, 2013г. – стр. 246-250.

Sidorenko I.A., Kuskova P.A. O spektral'nom analize fonem s ispol'zovaniem zvukovyh re-daktorov // Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta №1(144) 2013 vypusk 25/1, serija Informatika, Belgorod, 2013g. – str. 246-250.

2. Жиляков Е.Г., Прохоренко Е.И. Частотный анализ речевых сигналов // Научные ведомости Белгородского государственного университета. №2(31) 2006, выпуск 3, серия Информатика и прикладная математика. – Белгород, 2006 г.- стр. 201-208.

Zhilyakov E.G., Prohorenko E.I. Chastotnyj analiz rechevyh signalov // Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta. №2(31) 2006, vypusk 3, serija Informatika i prikladnaja matematika.- Belgorod, 2006g.- str. 201-208.

3. Жиляков Е.Г., Фирсова А.А. Оценивание периода основного тона звуков русской речи // Научные ведомости Белгородского государственного университета №1(144) 2013 выпуск 25/1, серия Информатика, Белгород, 2013г. – стр. 173-181

Zhilyakov E.G., Firsova A.A. Ocenivanie perioda osnovnogo tona zvukov russkoj rechi. // Nauch-nye vedomosti Belgorodskogo gosudarstvennogo universiteta №1(144) 2013 vypusk 25/1, serija Informa-tika, Belgorod, 2013g. – str. 173-181.

4. Бабаринов С.Л., Будникова М.А. О распознавании речи // Научные ведомости Белгородского государственного университета №21(192) 2014 выпуск 32/1, серия Информатика, Белгород, 2014. – стр. 182-185

Babarinov S.L., Budnikova M.A. O raspoznavanii rechi // Nauchnye vedomosti Belgorodskogo gosudar-stvennogo universiteta №21(192) 2014 vypusk 32/1, serija Informatika, Belgorod, 2014g. – str. 182-185

5. Савченко В.В., Васильев Р.В.Анализ эмоционального состояния диктора по голосу на основе фонетического детектора лжи // Научные ведомости Белгородского государственного университета №21(192) 2014 выпуск 32/1, серия Информатика, Белгород, 2014г. – стр. 186-195

Savchenko V.V., Vasil'ev R.V.Analiz jemocional'nogo sostojanija diktora po golosu na osnove fo-neticheskogo detektora lzhi // Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta №21(192) 2014 vypusk 32/1, serija Informatika, Belgorod, 2014g. – str. 186-195

6. Приходько А.И. Детерминированные сигналы. Учеб. пособ. для вузов. – М.:Горячая линия-Телеком, 2013.-326 с.: ил.

Prihod'ko A.I. Determinirovannye signaly. Ucheb. posob. Dlja vuzov.-M.:Gorjachaja linija-Telekom, 2013. – 326 s.: il.