

РОССИЙСКАЯ АКАДЕМИЯ НАУК  
Отделение нанотехнологий и информационных технологий

САНКТ-ПЕТЕРБУРГСКИЙ  
ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ РАН

# ТРУДЫ СПИИРАН

[proceedings.spiiras.nw.ru](http://proceedings.spiiras.nw.ru)



ВЫПУСК 1 (32)



Санкт-Петербург  
2014

18+

# Труды СПИИРАН

Выпуск № 1(32), 2014

Научный, научно-образовательный, междисциплинарный журнал с базовой специализацией в области информатики, автоматизации и прикладной математики

Журнал основан в 2002 году

## Учредитель и издатель

Федеральное государственное бюджетное учреждение науки  
Санкт-Петербургский институт информатики и автоматизации Российской академии наук  
(СПИИРАН)

## Главный редактор

Р.М. Юсупов, чл.-корр. РАН, д-р техн. наук, проф., С-Петербург, РФ

## Редакционная коллегия

**А.А. Ашимов**, академик национальной академии наук Республики Казахстан д-р техн. наук, проф., Алматы, Казахстан

**С.Н. Баранов**, д-р физ.-мат. наук, проф., С.-Петербург, РФ

**Н.П. Веселкин**, академик РАН, д-р мед. наук, проф., С.-Петербург, РФ

**В.И. Городецкий**, д-р техн. наук, проф., С.-Петербург, РФ

**О.Ю. Гусихин**, Ph.D., Диаборн, США

**В. Делич**, д-р техн. наук, проф., Нови-Сад, Сербия

**А.Б. Долгий**, Dr. Habil., проф., Сент-Этьен, Франция

**Д.А. Иванов**, д-р экон. наук, проф., Берлин, Германия

**С.О. Ипатов**, д-р техн. наук, проф., С.-Петербург, РФ

**В.П. Леонов**, д-р пед. наук, проф., С.-Петербург, РФ

**Г.А. Леонов**, член-корр. РАН, д-р физ.-мат. наук, проф., С.-Петербург, РФ

**К.П. Марков**, Ph.D., доцент, Аизу, Япония

**Ю.А. Меркурьев**, член-корр. Латвийской академии наук, Dr. Habil., проф., Рига, Латвия

**Н.А. Молдовян**, д-р техн. наук, проф., С.-Петербург, РФ

**А.А. Петровский**, д-р техн. наук, проф., Минск, Беларусь

**В.В. Попович**, д-р техн. наук, проф., С.-Петербург, РФ

**В.А. Путилов**, д-р техн. наук, проф., Апатиты, Россия

**А.Л. Ронжин** (зам. главного редактора), д-р техн. наук, проф., С.-Петербург, РФ

**А.И. Рудской**, член-корр. РАН, д-р техн. наук, проф., С.-Петербург, РФ

**В.А. Сарычев**, д-р техн. наук, проф., С.-Петербург, РФ

**В. Гурев**, академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария

**В.А. Скормин**, Ph.D., проф., Бингемптон, США

**А.В. Смирнов**, д-р техн. наук, проф., С.-Петербург, РФ

**Б.Я. Советов**, академик РАО, д-р техн. наук, проф., С.-Петербург, РФ

**В.А. Соифер**, член-корр. РАН, д-р техн. наук, проф., Самара, РФ

**Б.В. Соколов**, д-р техн. наук, проф., С.-Петербург, РФ

**А.В. Тимофеев**, д-р техн. наук, проф., С.-Петербург, РФ

**Л.В. Уткин**, д-р техн. наук, проф., С.-Петербург, РФ

**А.Л. Фрадков**, д-р техн. наук, проф., С.-Петербург, РФ

**Н.В. Хованов**, д-р физ.-мат. наук, проф., С.-Петербург, РФ

**Д.С. Черешкин**, д-р техн. наук, проф., Москва, РФ

**Л.Б. Шереметов**, д-р техн. наук, Мехико, Мексика

**А.В. Язенин**, д-р техн. наук, профессор, Тверь, РФ

## Адрес редакции

199178, Санкт-Петербург, 14-я линия, д. 39,

email: [publ@iias.spb.su](mailto:publ@iias.spb.su), сайт: <http://www.proceedings.spiiras.nw.ru/>

Подписано к печати 31.03.2014. Формат 60x90 1/16. Усл. печ. л. 12,9. Заказ № 166. Тираж 200 экз., цена свободная  
Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

Журнал зарегистрирован Федеральной службой по надзору в сфере связи и массовых коммуникаций,  
свидетельство ПИ № ФС77-41695 от 19 августа 2010 г.

Подписной индекс 29393 по каталогу «Почта России»

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

© Федеральное государственное бюджетное учреждение науки

Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 2014

Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания-журнала «Труды СПИИРАН» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания-журнала «Труды СПИИРАН»

# SPIIRAS Proceedings

Issue № 1(32), 2014

Scientific, educational, and interdisciplinary journal primarily specialized  
in computer science, automation, and applied mathematics

Trudy SPIIRAN ♦ Founded in 2002 ♦ Труды СПИИРАН

---

## Founder and Publisher

Federal State Budget Institution of Science  
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences  
(SPIIRAS)

---

## Editor-in-Chief

R.M. Yusupov, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia

---

## Editorial Board Members

**A.A. Ashimov**, Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan  
**S.N. Baranov**, Prof., Dr. Sci., St. Petersburg, Russia  
**N.P. Veselkin**, Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia  
**V.I. Gorodetski**, Prof., Dr. Sci., St. Petersburg, Russia  
**O.Yu. Gusikhin**, Ph. D., Dearborn, USA  
**V. Delic**, Prof., Dr. Sci., Novi Sad, Serbia  
**A. Dolgui**, Prof., Dr. Habil., St. Etienne, France  
**D.A. Ivanov**, Prof., Dr. Habil., Berlin, Germany  
**O.S. Ipatov**, Prof., Dr. Sci., St. Petersburg, Russia  
**V.P. Leonov**, Prof., Dr. Sci., St. Petersburg, Russia  
**G.A. Leonov**, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia  
**K.P. Markov**, Assoc. Prof., Ph.D., Aizu, Japan  
**Yu.A. Merkur'yev**, Prof., Dr. Habil., Corr. Member of the Latvian Academy of Sciences, Riga, Latvia  
**N.A. Moldovian**, Prof., Dr. Sci., St. Petersburg, Russia  
**A.A. Petrovsky**, Prof., Dr. Sci., Minsk, Belarus  
**V.V. Popovich**, Prof., Dr. Sci., St. Petersburg, Russia  
**V.A. Putilov**, Prof., Dr. Sci., Apatity, Russia

**A.L. Ronzhin** (Deputy Editor-in-Chief), Prof., Dr. Sci., St. Petersburg, Russia  
**A.I. Rudskoi**, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia  
**V.A. Saruchev**, Prof., Dr. Sci., St. Petersburg, Russia  
**V. Sgurev**, Prof., Dr. Sci., Academician of the Bulgarian academy of sciences, Sofia, Bulgaria  
**V. Skormin**, Prof., Ph.D., Binghamton, USA  
**A.V. Smirnov**, Prof., Dr. Sci., St. Petersburg, Russia  
**B.Ya. Sovetov**, Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia  
**V.A. Soyfer**, Prof., Dr. Sci., Corr. Member of RAS, Samara, Russia  
**B.V. Sokolov**, Prof., Dr. Sci., St. Petersburg, Russia  
**A.V. Timofeev**, Prof., Dr. Sci., St. Petersburg, Russia  
**L.V. Utkin**, Prof., Dr. Sci., St. Petersburg, Russia  
**A.L. Fradkov**, Prof., Dr. Sci., St. Petersburg, Russia  
**N.V. Hovanov**, Prof., Dr. Sci., St. Petersburg, Russia  
**D.S. Chereshekin**, Prof., Dr. Sci., Moscow, Russia  
**L.B. Sheremetov**, Assoc. Prof., Dr. Sci., Mexico, Mexico  
**A.V. Yazenin**, Prof., Dr. Sci. Tver, Russia

---

## Editorial Board's address

14-th line VO, 39, SPIIRAS, St. Petersburg, 199178, Russia,  
email: [publ@ias.spb.su](mailto:publ@ias.spb.su), web: <http://www.proceedings.spiiras.nw.ru/>

Signed to print 31.03.2014

Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

---

The journal is registered in Russian Federal Agency for Communications and Mass-Media Supervision,  
certificate ПИ № ФС77-41695 dated August 19, 2010 r.

Subscription Index 29393, Russian Post Catalog

© Federal State Budget Institution of Science  
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 2014

## СОДЕРЖАНИЕ

Азаров И.С., Вашкевич М.И., Лихачев Д.С., Петровский А.А. ИЗМЕНЕНИЕ ЧАСТОТЫ ОСНОВНОГО ТОНА РЕЧЕВОГО СИГНАЛА НА ОСНОВЕ ГАРМОНИЧЕСКОЙ МОДЕЛИ С НЕСТАЦИОНАРНЫМИ ПАРАМЕТРАМИ	5
Басов О.О., Носов М.В., Шалагинов В.А. ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК ДЖИТТЕРА ПЕРИОДА ОСНОВНОГО ТОНА РЕЧЕВОГО СИГНАЛА	27
Тиунов С.Д., Мещеряков Р.В., Черных Д.В. ОПТИМИЗАЦИЯ ВЫЧИСЛЕНИЯ ОДНОВРЕМЕННОЙ МАСКИРОВКИ РЕЧЕВОГО СИГНАЛА	45
Евдокимова В.В. АНАЛИЗ СПЕКТРА ГЛАСНЫХ НА ОСНОВЕ НЕРАВНОМЕРНОЙ ПСИХОАКУСТИЧЕСКОЙ ШКАЛЫ ЭРБОВ ДЛЯ ОПРЕДЕЛЕНИЯ СЛОВЕСНОГО УДАРЕНИЯ	58
Качковская Т.В. ИСПОЛЬЗОВАНИЕ ТЕМПОРАЛЬНЫХ ХАРАКТЕРИСТИК ДЛЯ СЕГМЕНТАЦИИ РЕЧЕВОГО ПОТОКА НА КРУПНЫЕ СМЫСЛОВЫЕ ЕДИНИЦЫ (НА МАТЕРИАЛЕ РУССКОГО ЯЗЫКА)	68
Захарьев В.А., Петровский А.А., Лобанов Б.М. СИСТЕМА СИНТЕЗА РЕЧИ ПО ТЕКСТУ С ВОЗМОЖНОСТЬЮ НАСТРОЙКИ НА ГОЛОС ЦЕЛЕВОГО ДИКТОРА	82
Ляско Е.Е., Григорьев А.С., Ситдииков В.М., Огородникова Е.А. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ МОДЕЛИ «INFANT.MAVS» ДЛЯ ИЗУЧЕНИЯ ПСИХОФИЗИОЛОГИЧЕСКОГО СОСТОЯНИЯ ДЕТЕЙ ПЕРВОГО ГОДА ЖИЗНИ	99
Мещеряков Р.В., Костюченко Е.Ю., Балацкая Л.Н., Чойнзонов Е.Л. СТРУКТУРА И БАЗА ДАННЫХ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ОЦЕНКИ КАЧЕСТВА И РАЗБОРЧИВОСТИ РЕЧИ В ПРОЦЕССЕ РЕАБИЛИТАЦИИ ПОСЛЕ ОПЕРАЦИИ ПРИ ЛЕЧЕНИИ РАКА ПОЛОСТИ РТА И РОТОГЛОТКИ, ЧЕЛЮСТНОЛИЦЕВОЙ ОБЛАСТИ	116
Крючков Б.И., Карпов А.А., Усов В.М. ПЕРСПЕКТИВНЫЕ ПОДХОДЫ К ПРИМЕНЕНИЮ СЕРВИСНЫХ РОБОТОВ В ОБЛАСТИ ПИЛОТИРУЕМОЙ КОСМОНАВТИКИ	125
Басов О.О., Сайтов И.А. КАЧЕСТВО ФУНКЦИОНИРОВАНИЯ И ЭФФЕКТИВНОСТЬ ПОЛИМОДАЛЬНЫХ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ	152
Батенков К.А. МОДЕЛИРОВАНИЕ НЕПРЕРЫВНЫХ КАНАЛОВ СВЯЗИ В ФОРМЕ ОПЕРАТОРОВ ПРЕОБРАЗОВАНИЯ НЕКОТОРЫХ ПРОСТРАНСТВ	171
Левоневский Д.К., Фаткиева Р.Р. ИССЛЕДОВАНИЕ КОМБИНИРОВАННЫХ АТАК КЛАССА «ОТКАЗ В ОБСЛУЖИВАНИИ»	199
Трегубов Р.Б., Алексиков Ю.Г. МОДЕЛЬ ПРОЦЕССА МУЛЬТИПЛЕКСИРОВАНИЯ ПРОТОКОЛЬНЫХ БЛОКОВ ДАННЫХ В КАНАЛЕ СВЯЗИ УПРАВЛЕНИЯ	210

## CONTENTS

Azarov E.S., Vashkevich M.I., Likhachov D.S., Petrovsky A.A. PITCH MODIFICATION OF SPEECH SIGNAL USING HARMONIC MODEL WITH TIME-VARYING PARAMETERS	5
Basov O.O., Nosov M.V., Shalaginov V.A. PITCH-JITTER ANALYSIS OF THE SPEECH SIGNAL	27
Tiunov S.D., Meshcheryakov R.V., Chernyh D.V. OPTIMIZATION OF CALCULATION OF FREQUENCY MASKING OF SPEECH SIGNAL	45
Evdokimova V.V. VOWEL SPECTRAL ANALYSIS USING THE PSYCHOACOUSTIC ERB SCALE FOR THE WORD STRESS DETECTION	58
Kachkovskaia T.V. TEMPORAL ASPECTS OF INTONATIONAL PHRASING (EVIDENCE FROM RUSSIAN)	68
Zahariev V.A., Petrovsky A.A., Lobanov B.M. TEXT TO SPEECH SYNTHESIS SYSTEM WITH THE TARGET SPEAKER VOICE CUSTOMIZATION CAPABILITY	82
Lyakso E.E., Grigorev A.S., Sitdikov V.M., Ogorodnikova E.A. SOFTWARE MODEL «INFANT.MAVS» FOR STUDYING PSYCHOPHYSIOLOGICAL STATE OF INFANTS	99
Meshcheryakov R.M, Kostyuchenko E.Yu., Balatskaya L.N., Choinzonov E.L. STRUCTURE AND DATABASE OF SOFTWARE FOR SPEECH QUALITY AND INTELLIGIBILITY ASSESSMENT IN THE PROCESS OF REHABILITATION AFTER SURGERY IN THE TREATMENT OF CANCERS OF THE ORAL CAVITY AND OROPHARYNX, MAXILLOFACIAL AREA	116
Kryuchkov B.I., Karpov A.A., Usov V.M. PROMISING APPROACHES FOR THE USE OF SERVICE ROBOTS IN THE DOMAIN OF MANNED SPACE EXPLORATION	125
Basov O.O., Saitov I.A. FUNCTIONING QUALITY AND EFFECTIVENESS OF POLYMODAL INFOCOMMUNICATIONAL SYSTEMS	152
Batenkov K.A. CONTINUOUS CHANNEL MODELING IN SHAPE OF SOME SPACE TRANSFORMATION OPERATORS	171
Levonevskiy D.K., Fatkieva R.R. RESEARCH OF COMBINED “DENIAL OF SERVICE” ATTACKS	199
Tregubov R.B., Aleksikov J.G. MODEL OF PROCESS OF MULTIPLEXING THE PROTOCOL DATA UNITS IN MANAGEMENT CHANNEL	210

И.С. АЗАРОВ, М.И. ВАШКЕВИЧ, Д.С. ЛИХАЧЕВ, А.А. ПЕТРОВСКИЙ  
**ИЗМЕНЕНИЕ ЧАСТОТЫ ОСНОВНОГО ТОНА РЕЧЕВОГО  
СИГНАЛА НА ОСНОВЕ ГАРМОНИЧЕСКОЙ МОДЕЛИ С  
НЕСТАЦИОНАРНЫМИ ПАРАМЕТРАМИ**

---

*Азаров И.С., Вашкевич М.С., Лихачев Д.С., Петровский А.А.* **Изменение частоты основного тона речевого сигнала на основе гармонической модели с нестационарными параметрами.**

**Аннотация.** В статье предлагается решение задачи изменения частоты основного тона речевого сигнала. Необходимость решения данной задачи возникает во многих речевых приложениях таких как конверсия голоса, коррекция акцента, обеспечение конфиденциальности диктора и др. Разработанная схема обработки локализованной части речевого сигнала основывается на гармонической модели с нестационарными (изменяющимися в каждый момент времени) параметрами. Для повышения частотного разрешения модели оценка параметров выполняется при помощи узкополосной фильтрации в искривленном масштабе времени, согласованном с контуром мгновенной частоты основного тона. На основании субъективной оценки результатов показано, что разработанный способ обеспечивает высокую натуральность и разборчивость синтезированной речи и может применяться как в широкополосных так и в узкополосных каналах связи с различными стандартами кодирования (в том числе с кодеками G.711 и GSM).

**Ключевые слова:** гибридная модель речевого сигнала, оценка основного тона, изменение просодических характеристик речи.

*Azarov E., Vashkevich M., Likhachov D., Petrovsky A.* **Pitch modification of speech signal using harmonic model with time-varying parameters.**

**Abstract.** The paper presents a solution to the problem of pitch modification of speech. The problem occurs in different speech processing applications such as voice conversion, accent correction, hiding speaker's personality and other. Developed processing scheme for voiced part of speech is based on the harmonic model with nonstationary (time-varying) parameters. In order to improve frequency resolution of the model parameters are extracted using narrowband filtering in warped time domain aligned to instantaneous pitch frequency. Using subjective listening tests it is shown that developed system provides high naturalness and intelligibility of reconstructed speech and can be applied to wideband and narrowband communication channels with various coding standards (including G.711 and GSM).

**Keywords:** hybrid speech modeling, pitch estimation, prosody modification.

---

**1. Введение.** Эффект изменения частоты основного тона звукового сигнала может достигаться различными методами. Самым простым из них является изменение скорости воспроизведения, что приводит к смещению частоты всех составляющих сигнала. Однако, во-первых, это приводит к изменению длительности сигнала и потому не может быть использовано в приложениях, работающих в реальном масштабе времени, а во-вторых, это сильно искажает тембр голоса. Одним из наиболее популярных альтернатив является фазовый вокодер [1] и разнообразные методы на его основе, выполняющие смеще-

ние компонент сигнала в частотной области при помощи прямого и обратного преобразования Фурье. Применение фазового вокодера позволяет изменять высоту звучания сигнала без изменения длительности и сохранять тембр путем коррекции спектральной огибающей. Данный подход может быть использован для любых звуковых сигналов. Вокодер может быть реализован в режиме реального времени, поскольку выполняет обработку входного сигнала последовательно фрейм за фреймом. При применении его к речи и певческому голосу метод имеет существенные ограничения: 1) при обработке не используется какая-либо модель голосообразования, что при значительном изменении частоты основного тона приводит к неестественному звучанию; 2) выделяемые частотные составляющие сигнала не соответствуют гармоникам основного тона, что приводит к потере натуральности и звонкости голоса; 3) сигнал не разделяется на вокализованный и невокализованный, что приводит к неестественному звучанию некоторых звуков (в частности невокализованных шипящих ‘с’, ‘ш’, а так же смешанных, частично вокализованных, ‘з’ и ‘ж’).

Для повышения натуральности звучания и расширения доступного диапазона изменения основного тона необходимо использовать более сложные решения, основанные на гибридной (детерминистской/стохастической) модели речевого сигнала. Гибридная обработка звуковых сигналов предложена в работах [2,3], где используется три отдельные составляющие: периодическая, шумовая и транзиентная. Поскольку при изменении основного тона голоса обрабатывается только вокализованная (детерминистская) часть сигнала, в данном случае нет необходимости различать между собой шумовые и транзиентные составляющие – вместе их можно отнести к невокализованной (стохастической) части сигнала [4,5,6].

В настоящей работе предлагается решение задачи изменения основного тона, основанное на гибридном представлении речевого сигнала. Частотно-временной анализ выполняется с учетом контура частоты основного тона, что позволяет выделять параметры отдельных гармоник и разделять сигнал на вокализованную и невокализованную составляющие в частотной области. Моделирование вокализованного сигнала выполняется при помощи нестационарных (т.е. изменяющихся во времени) параметров. В экспериментальной части работы приводятся результаты субъективной оценки качества обработанной речи. С целью оценить применимость метода в существующих каналах связи для экспериментов использовались как широкополосные, так и узкополосные речевые сигналы с применением кодеков G.711 и GSM.

**2. Гибридная модель и общая схема обработки речевого сигнала.** Речевой сигнал представляется в виде суммы двух составляющих: вокализованной и невокализованной. Для описания вокализованной части сигнала используется синусоидальная модель [7,8]:

$$s(n) = \sum_{k=1}^K A_k(n) \cos \varphi_k(n) + r(n), \quad (1)$$

где  $A_k(n)$  – мгновенная амплитуда  $k$ -ой гармоники,  $K$  – число гармоник,  $\varphi_k(n)$  – мгновенное значение фазы  $k$ -ой гармоники,  $r(n)$  – шумовая составляющая сигнала. Мгновенная частота  $f_k(n)$  связана с мгновенной фазой следующим соотношением:

$$\varphi_k(n) = \sum_{i=0}^n \frac{2\pi f_k(i)}{F_s} + \varphi_k(0),$$

где  $F_s$  – частота дискретизации и  $\varphi_k(0)$  – начальная фаза  $k$ -ой гармоники. Приблизительно можно считать, что частота каждой гармоники является кратной частоте основного тона (как показано на рисунке 1), т.е.

$$f_k(n) \approx F_0(n)k,$$

где  $F_0(n)$  – основной тон.

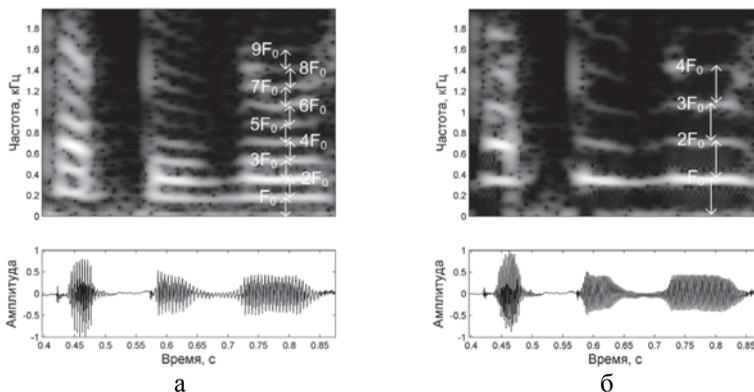


Рис. 1. Изменение основного тона речевого сигнала;  
а – исходный речевой сигнал; б – обработанный сигнал

Частота основного тона определяет высоту звучания голоса. Контур частоты основного тона (его изменение в зависимости от времени на большом протяжении) определяет интонацию речи.

Для того чтобы сохранить исходный тембр диктора при изменении основного тона должна сохраняться спектральная огибающая, которая формируется из мгновенных амплитуд гармоник основного тона. Спектральная огибающую можно рассматривать как функцию от номера отсчета и частоты  $E(n, f)$ , которая принимает значения мгновенных амплитуд гармоник основного тона в соответствующих точках  $E(n, f_k(n)) = A_k(n)$ . Для произвольных  $n$  и  $f$  функция вычисляется путем линейной интерполяции ближайших к ним амплитудных значений.

Синтез голоса с модифицированным контуром частоты основного тона может быть выполнен по следующей формуле:

$$s(n) = \sum_{k=1}^K E(n, \bar{F}_0(n)k) \cos \bar{\varphi}_k(n) + r(n),$$

где фазы гармонических компонентов  $\bar{\varphi}_k(n)$  рассчитываются в соответствии с новым контуром частоты основного тона  $\bar{F}_0(n)$  следующим образом:

$$\bar{\varphi}_k(n) = \sum_{i=0}^n \frac{2\pi \bar{F}_0(i)}{F_s} + \bar{\varphi}_k^\Delta(n).$$

Дополнительный фазовый параметр  $\bar{\varphi}_k^\Delta(n)$  используется для сохранения относительных фаз гармоник по отношению к фазе частоты основного тона. Данный параметр вычисляется как

$$\bar{\varphi}_k^\Delta(n) = \varphi_k(n) - k\varphi_0(n).$$

Основной тон присутствует только в вокализованных сегментах речи, т.е. только тогда когда задействованы голосовые связки диктора. Такие звуки как 'а', 'о', 'ж' являются вокализованными, в то время как звуки 'с', 'ш', 'щ' являются невокализованными. Как было показано выше в спектре речевого сигнала вокализованность проявляется в виде спектральных компонент кратной частоты. Для того чтобы сохранить исходное качество звучания невокализованных звуков алгоритм изменения тона должен автоматически выделять области вокализованности и выполнять обработку только в этих областях. Шумовая часть  $r(n)$  выделяется из исходного сигнала вычитанием выделенных вокализованных звуков как показано на рисунке 2.

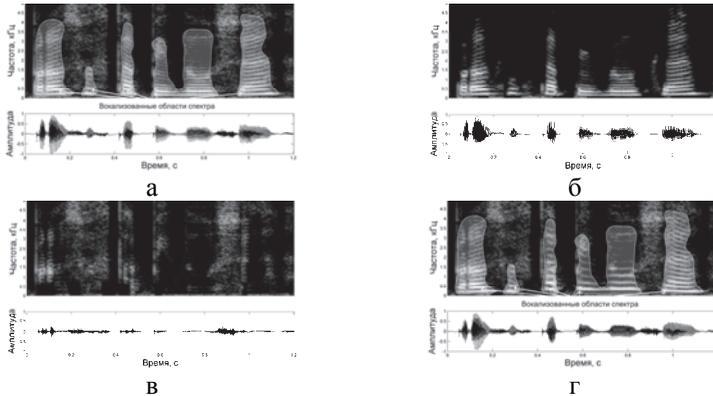


Рис. 2. Разделение сигнала на периодическую и стохастическую составляющие; а – исходный речевой сигнал; б – выделенная вокализованная компонента; в – сигнал-остаток  $r(n)$ ; г – обработанный выходной сигнал

Алгоритм обработки сигнала можно кратко описать в виде последовательности действий, схематически изображенной на рисунке 3:

- 1) Определение частоты основного тона  $F_0(n)$ ;
- 2) Преобразование речевого сигнала в параметрический вид т.е. оценка мгновенных гармонических параметров  $A_k(n)$ ,  $f_k(n)$  и  $\varphi_k(n)$ ,  $k = 1, 2, \dots, K$ ;
- 3) Оценка вокализации каждой тройки гармонических параметров и отбор только тех, которые относятся к вокализованным областям спектра.
- 4) Синтез исходной вокализованной компоненты сигнала и ее вычитание из исходного речевого сигнала для получения шумовой составляющей  $r(n)$ .
- 5) Синтез вокализованной компоненты с измененным основным тоном и ее сложение с шумовой составляющей  $r(n)$ .

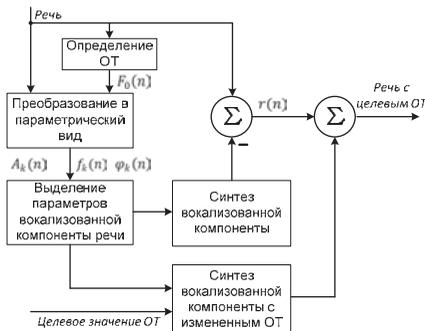


Рис. 3. Общая схема обработки сигнала

Оценка параметров модели выполняется с учетом контура мгновенной частоты основного тона. Входной речевой сигнал масштабируется во времени для того, чтобы обеспечить его стационарность, а затем выполняется узкополосная фильтрация, разделяющая гармоники основного тона. В результате фильтрации формируются аналитические (комплексные) сигналы, которые описываются при помощи параметров синусоидальной модели. На основе анализа смежных значений мгновенной частоты определяется степень вокализации каждой гармоники. Затем выполняется синтез квазипериодического сигнала по полученным параметрам, который вычитается из исходного речевого сигнала. В результате выполняется разделение речевого сигнала на вокализованную и невокализованную части. Одновременно синтезируется квазипериодический сигнал с целевой частотой основного тона, который складывается с полученной невокализованной частью.

Целевой контур основного тона формируется согласно требованиям конкретного приложения. Для экспериментов использовались заданные профили изменения исходного основного тона, приведенные в соответствующем разделе.

**3. Оценка параметров речевого сигнала.** Оценка основного тона выполняется при помощи алгоритма, изложенного в работах [9,10]. Особенностью алгоритма является возможность определения мгновенной частоты. Эта возможность достигается за счет использования специальной функции оценки периодичности аналогичной автокорреляционной функции [11], вычисляемой из мгновенных гармонических параметров субполосных составляющих сигнала:

$$\phi_{inst}(n, p) = \frac{\sum_{k=1}^K A_k^2(n) \cos(f_k(n)p)}{\sum_{k=1}^K A_k^2(n)}, \quad (2)$$

где  $p$  – длина периода кандидата основного тона. В отличие от автокорреляционной функции,  $\phi_{inst}(n, p)$  нечувствительна к любым изменениям частоты основного тона в окрестности отсчета  $n$  при условии, что используемые гармонические параметры получены достаточно точно.

Оценка параметров отдельных гармоник выполняется при помощи узкополосной фильтрации. Для того, чтобы повысить частотное разрешение анализа масштаб времени сигнала изменяется согласованно с полученным контуром мгновенной частоты основного тона. Сигнал дискретизируется таким образом, чтобы на каждый период основного тона приходилось равное количество отсчетов  $N_{f_0}$ .

Каждому входному отсчету речевого сигнала  $s(n)$  ставится в соответствие фаза периода основного тона  $\phi(n)$

$$\phi(n) = \sum_{i=0}^n f_0(i),$$

где  $f_0(i)$  – нормализованная круговая частота основного тона в момент времени  $i$ . Новые моменты времени  $m$  в которые необходимо перерасчитать входной сигнал определяются как

$$m = \phi^{-1}(p/N_{f_0}),$$

где  $p$  – это индекс отсчета в масштабированной временной области. Обратная функция  $\phi^{-1}(\cdot)$  вычисляется при помощи линейной интерполяции ее известных значений для целых  $n$  как показано на рисунке 4.

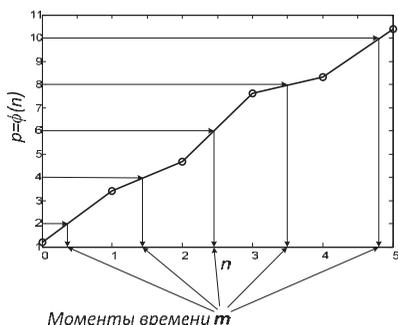


Рис. 4. Расчет новых моментов времени для масштабирования сигнала

Вычисление значений сигнала в заданный момент времени выполняется по теореме Котельникова (рисунок 5), согласно которой

$$s(t) = \sum_{n=-\infty}^{\infty} s(nT) \operatorname{sinc}\left(\frac{t - nT}{T}\right), \quad (3)$$

где  $\operatorname{sinc}(x) = \frac{\sin \pi x}{\pi x}$ ,  $T$  – интервал дискретизации, а  $x(nT)$  – значения сигнала в дискретные моменты времени  $nT$ . Для простоты в дальнейшем примем  $T = 1$  и вместо  $s(nT)$  будем писать  $s(n)$ .

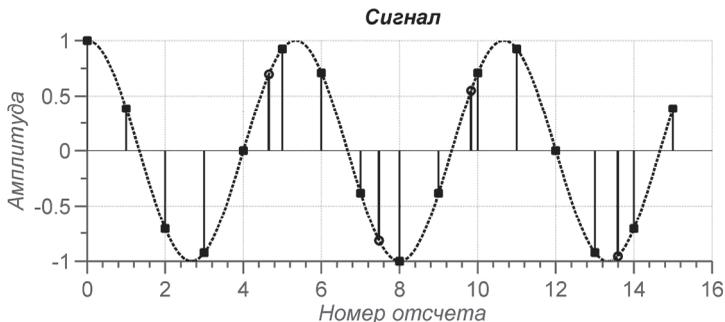


Рис. 5. Вычисление значений сигнала в произвольные моменты времени (пунктирная линия – непрерывный сигнал; квадратный маркер – сигнал дискретизированный с равным временным интервалом; круглый маркер – значения сигнала в моменты времени не кратные интервалу дискретизации)

Выражение (3) нельзя использовать на практике из-за суммирования в бесконечных пределах. Поэтому выбирается конечное число точек сигнала  $N_{pt}$  предшествующих моменту  $t$  и  $N_{pt}$  последующих точек:

$$s(t) = \sum_{n=-N_{pt}}^{N_{pt}} s(n) \text{sinc}(t - n) w(t - n), \quad (4)$$

где  $w(\cdot)$  – оконная функция с центром симметрии в точке 0. Для каждого выходного отсчета временные отметки пересчитываются таким образом, чтобы текущий момент  $t$  всегда попадал в диапазон от 0 до 1 как показано на рисунке 6. Это позволяет использовать таблицу с заранее рассчитанными значениями функции  $\text{sinc}$ . Использование табличных значений дает возможность существенно сократить вычислительные затраты.

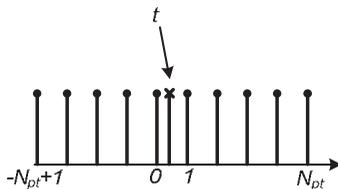


Рис. 6. Интерполяция сигнала функциями  $\text{sinc}$  со смещением от 0 до 1

После временного масштабирования основной тон становится постоянным и каждая гармоника частоты основного тона имеет фикс-

сированное положение в частотном диапазоне. Для разделения гармоник и оценки их параметров используется ДПФ-модулированный банк фильтров анализа, который представляет собой совокупность фильтров  $H_k(z)$ ,  $0 \leq k < M$ , раскладывающих входной сигнал на  $M$  субполосных сигналов [12]. Амплитудно-частотные характеристики (АЧХ) фильтров формируются путем сдвига в частотной области фильтра-прототипа. В результате полоса частот от нуля до частоты дискретизации перекрывается гребенкой из  $M$  фильтров, как показано на рисунке 7.

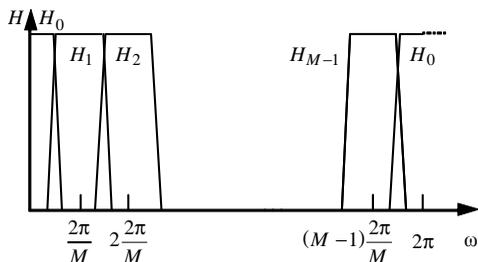


Рис. 7. АЧХ каналов ДПФ-модулированного банка фильтров

Число каналов в банке выбирается равным  $N_{f0}$ , поскольку в этом случае центральная частота каждого канала совпадает с положением соответствующей гармоники масштабированного сигнала  $s(m)$ . Импульсные характеристики фильтров анализа определяются выражением:

$$h_k(n) = h(n)W_M^{-kn}, \quad k = 0, 1, \dots, M - 1, \quad (5)$$

где  $h(n)$  – фильтр-прототип нижних частот с нормированной частотой среза  $\pi/M$ ,  $W_M = e^{-j2\pi/M}$ . Передаточная функция каждого фильтра анализа выражается через передаточную функцию фильтра-прототипа следующим образом:

$$H_k(z) = \sum_{n=0}^{N-1} h_k(n)z^{-n} = \sum_{n=0}^{N-1} h(n)(zW_M^{-k})^{-n} = H(zW_M^{-k}). \quad (6)$$

Фильтрация сигнала с использованием импульсных характеристик (5) является весьма неэффективной и приводит к большой вычислительной нагрузке. Более эффективный способ фильтрации можно получить, если воспользоваться полифазным представлением фильтра-

прототипа, в котором импульсная характеристика  $h(n)$  перегруппирована в  $M$  подпоследовательностей  $e_l(n)$

$$H(z) = \sum_{l=0}^{M-1} z^{-l} E_l(z^M), 0 \leq l < M, \quad (7)$$

где

$$E_l(z) = \sum_{n=0}^{N/M-1} e_l(n) z^{-n} = \sum_{n=0}^{N/M-1} h(l + Mn) z^{-n}.$$

Далее объединим выражения (6) и (7). С учетом ва  $W_M^{kM} = 1$  получим

$$H_k(z) = \sum_{n=0}^{M-1} z^{-l} W_M^{-lk} E(z^M W_M^{kM}) = \sum_{l=0}^{M-1} z^{-l} W_M^{-kl} E_l(z^M). \quad (8)$$

Уравнение (8) можно записать в виде произведения вектора строки на вектор-столбец

$$H_k(z) = \begin{bmatrix} 1 & W_M^{-k} & W_M^{-2k} & \dots & W_M^{-(M-1)k} \end{bmatrix} \begin{bmatrix} E_0(z^M) \\ z^{-1} E_1(z^M) \\ z^{-2} E_2(z^M) \\ \vdots \\ z^{-(M-1)} E_{M-1}(z^M) \end{bmatrix}.$$

Все  $M$  уравнений (для каждого  $H_k$ ) в матричной форме можно записать следующим образом:

$$\begin{bmatrix} H_0(z) \\ H_1(z) \\ H_2(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_M^{-1} & W_M^{-2} & \dots & W_M^{-(M-1)} \\ 1 & W_M^{-2} & W_M^{-4} & \dots & W_M^{-2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_M^{-(M-1)} & W_M^{-2(M-1)} & \dots & W_M^{-(M-1)^2} \end{bmatrix} \begin{bmatrix} E_0(z^M) \\ z^{-1} E_1(z^M) \\ z^{-2} E_2(z^M) \\ \vdots \\ z^{-(M-1)} E_{M-1}(z^M) \end{bmatrix}. \quad (9)$$

Выражение (9) является основой для построения эффективной полифазной структуры ДПФ-модулированного банка фильтров, поскольку матрица  $M \times M$  в правой части выражения является матрицей дискретного преобразования Фурье (ДПФ). Структура банка, использованная в настоящей работе, приведена на рисунке 8. При реализации ДПФ-модулированного банка фильтров порядок фильтра-прототипа выбирался равным:  $N = kM + 1$ , где  $k$  – некоторое целое число.

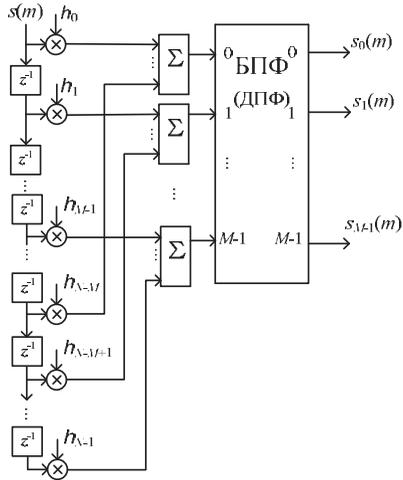


Рис. 8. Схема эффективной реализации ДПФ-модулированного банка фильтров анализа

Выходы банка фильтров  $s_k(m)$  являются узкополосными аналитическими сигналами, которые интерпретируется как отдельные синусоидальные компоненты модели (1). Параметры синусоидальных компонент вычисляются с интервалом 5 мс при помощи следующих выражений:

$$A_k(m) = \sqrt{\operatorname{Re}^2(s_k(m + \Delta)) + \operatorname{Im}^2(s_k(m + \Delta))}, \quad (10)$$

$$\varphi_k(m) = \arg s_k(m + \Delta), \quad (11)$$

$$f_k(m) = \frac{\arg s_k(m + \Delta + 1) - \arg s_k(m + \Delta)}{2\pi}, \quad (12)$$

где  $\Delta = \frac{N-1}{2}$  – групповая задержка банка фильтров.

**4. Модификация гармонических параметров и синтез выходного сигнала.** Полученные параметры гармонической ли  $A_k(m)$ ,  $\varphi_k(m)$  и  $F_0(m)$  изменяются в соответствии с новым значением мгновенной частоты основного тона  $\bar{F}_0(m)$ . В каждый момент времени амплитуды гармоник входного сигнала  $A_k$  определены на сетке частот  $kF_0$ . Поскольку основной тон изменяется, но сохраняется

исходная спектральная огибающая, то необходимо перерасчитать амплитуды на новой частотной сетке  $k\bar{F}_0$  как показано на рисунке 9. Перерасчет осуществляется при помощи линейной интерполяции.

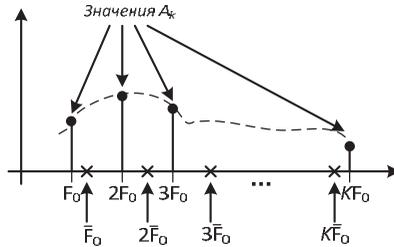


Рис. 9. Вычисление амплитудных значений гармоник новой частоты основного тона

Мгновенные фазовые значения изменяются таким образом, чтобы сохранить их непрерывность и синхронизацию фазы между гармониками:

$$\bar{\varphi}_k(m) = \bar{\varphi}_k(m-1) + \frac{\bar{F}_0(m)}{F_s} 2\pi\Delta t + \varphi_k(m) - kF_0(m),$$

где  $F_s$  – частота дискретизации,  $\Delta t$  – шаг (в отсчетах) между двумя последовательными оценками параметров модели.

Восстановление сигнала из параметрического представления выполняется путем синтезирования децимированных субполосных сигналов  $\hat{s}_k(m)$  из измененных параметров для каждой гармоники. Затем эти сигналы подаются на вход банка фильтров синтеза, схема которого показана на рисунке 10. Через  $\uparrow S$  обозначен экспандер, добавляющий  $S-1$  нулевых значений между последовательными отсчетами сигнала  $\hat{s}_k(m)$ . Предполагается, что число  $S$  является лем  $M$ . Через  $F_k(z)$  на схеме обозначены фильтры синтеза, импульсные характеристики которых определяются следующим образом:

$$f_k(n) = h(n)W_M^{-kn}, \quad k = 0, 1 \dots M-1, \quad n = 0, 1 \dots N-1. \quad (13)$$

Прямая реализация данной схемы требует значительных вычислительных затрат. Её сложность можно уменьшить если учесть способ получения импульсных характеристик фильтров (13) и то, что большинство поступающих в фильтры отсчетов являются нулевыми. Полифазное представление фильтра-прототипа имеет вид:

$$H(z) = \sum_{l=0}^{M-1} z^{-(M-1-l)} R_l(z^M), \quad 0 \leq l \leq M, \quad (14)$$

где  $R_l(z) = E_{M-1-l}(z)$ . Объединяя (13) и (14) получим:

$$F_k(z) = \sum_{l=0}^{M-1} z^{-(M-1-l)} W_M^{-(M-1-l)} R_l(z^M) = \sum_{l=0}^{M-1} z^{-(M-1-l)} W_M^{-kl} R_l(z^M).$$

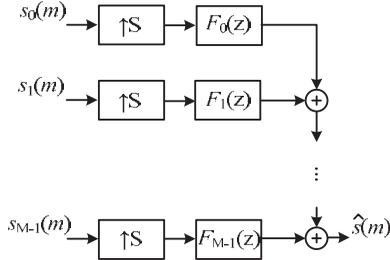


Рис. 10. Банк фильтров синтеза

Выражение для банка фильтров синтеза можно представить в матричной форме:

$$\hat{S}(z) = [F_0(z) \quad F_1(z) \quad F_2(z) \quad \dots \quad F_{M-1}(z)] \begin{bmatrix} S_0(z) \\ S_1(z) \\ S_2(z) \\ \vdots \\ S_{M-1}(z) \end{bmatrix}.$$

Его же можно записать в следующем виде:

$$\hat{S}(z) = [z^{-(M-1)} R_{M-1}(z^M) \quad z^{-(M-2)} R_{M-2}(z^M) \quad z^{-(M-3)} R_{M-3}(z^M) \quad \dots \quad R_0(z^M)] \times$$

$$\times \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_M^{-1} & W_M^{-2} & \dots & W_M^{-(M-1)} \\ 1 & W_M^{-2} & W_M^{-4} & \dots & W_M^{-2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_M^{-(M-1)} & W_M^{-2(M-1)} & \dots & W_M^{-(M-1)^2} \end{bmatrix} \times \begin{bmatrix} S_0(z) \\ S_1(z) \\ S_2(z) \\ \vdots \\ S_{M-1}(z) \end{bmatrix}.$$

В этом выражении присутствует матрица размера  $M \times M$  обратного дискретного преобразования Фурье (ОДПФ). В соответствии с этим строится эффективная полифазная структура ДПФ-модулированного банка фильтров синтеза, показанная на рисунке 11.

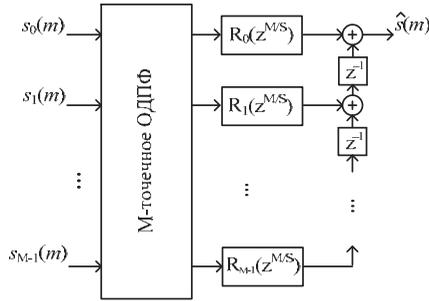


Рис. 11. Эффективная реализация банка фильтров синтеза

После реконструкции сигнала  $\hat{s}(m)$  выполняется его обратное временное масштабирование с учетом целевого контура основного тона, в результате чего формируется выходной сигнал  $\hat{s}(n)$  с постоянным шагом дискретизации.

## 5. Результаты экспериментов

**5.1. Профили изменения контура основного тона.** Для тестирования качества работы системы было разработано четыре эффекта (профиля) изменения частоты основного тона. Первый эффект заключался в повышении основного тона. Целевая частота основного тона  $\bar{F}_0(n)$  и исходная  $F_0(n)$  связываются соотношением:

$$\bar{F}_0(n) = F_0(n)k, \quad k > 1,$$

где  $k$  – коэффициент на который происходит повышение тона. Высокая натуральность синтезированного голоса достигается, если коэффициент  $k < 2$ . На рисунке 12 показан пример повышения основного тона.

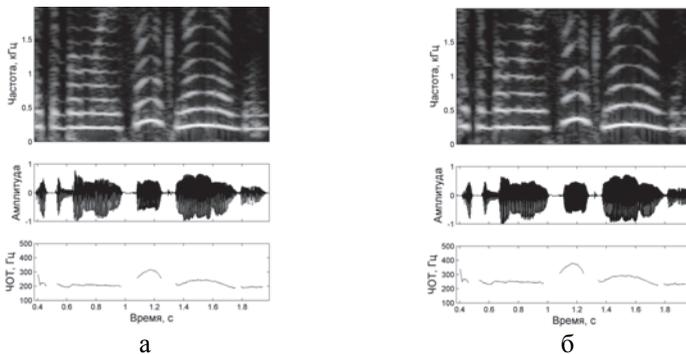


Рис. 12. Эффект повышения частоты основного тона;  
а – входной сигнал, б – обработанный сигнал

Второй эффект заключался в понижении основного тона. Целевая частота основного тона определялась из исходной как:

$$\bar{F}_0(n) = \frac{F_0(n)}{k}, \quad k > 1,$$

где  $k$  – коэффициент, на который происходит понижение тона. Высокая натуральность синтезированного голоса достигается, если коэффициент  $k$  не превышает 2. На рисунке 13 показан пример повышения основного тона.

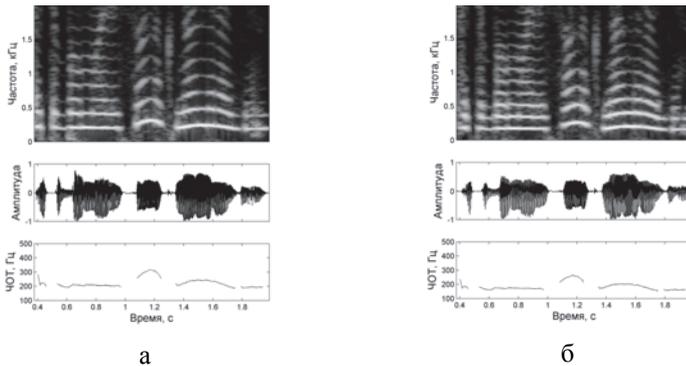


Рис. 13. Эффект понижения частоты основного тона;  
а – входной сигнал, б – обработанный сигнал

Третий эффект ‘*cycle*’ выполняет периодическое повышение тона. Профиль согласно которому происходит повышение задается функцией  $f_{profile}(n)$ , которая определена на конечном интервале от 0 до  $N - 1$ . Пример такой функции показан на рисунке 14.

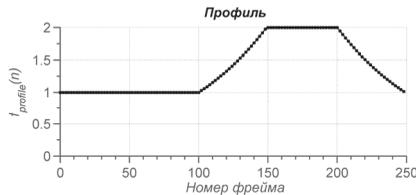


Рис. 14. Функция-профиль для изменения частоты основного тона

При реализации эффекта ‘*cycle*’ целевая частота основного тона определяется следующим образом:

$$\bar{F}_0(n) = F_0(n) \cdot f_{profile}(n \bmod N).$$

На рисунке 15 показан результат применения эффекта ‘cycle’ к речевому сигналу.

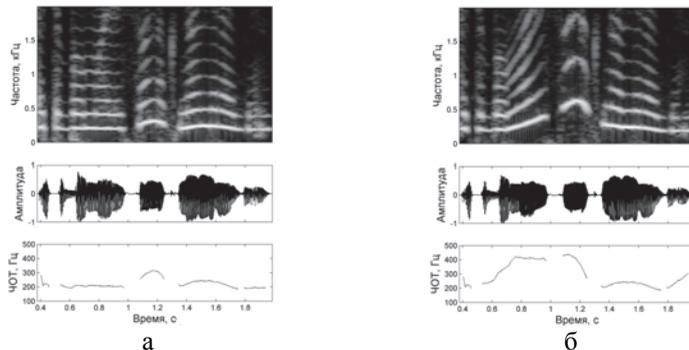


Рис. 15. Эффект ‘cycle’;  
а – входной сигнал, б – обработанный сигнал

Четвертый эффект ‘sinus’ позволяет придать голосу дрожание, которое иногда появляется у человека в момент волнения. Для достижения этого эффекта частота основного тона сигнала преобразуется следующим образом:

$$\bar{F}_0(n) = F_0(n) + 20 \cos\left(2\pi n \frac{12}{F_s}\right),$$

где  $F_s$  – частота дискретизации сигнала. На рисунке 16 показан результат применения эффекта ‘sinus’ к речевому сигналу.

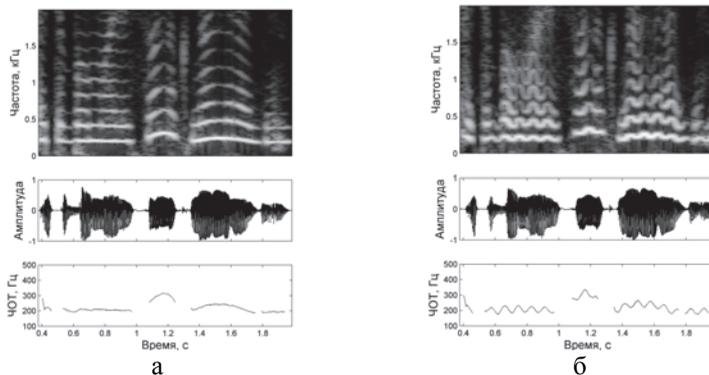


Рис. 16. Эффект ‘sinus’;  
а – входной сигнал, б – обработанный сигнал

**5.2. Речевая база и оценка качества обработки.** Тестирование системы изменения основного тона речевого сигнала выполнялось на речевой базе, начитанной 9 дикторами (4 мужских голоса и 5 женских). Длительность записи каждого диктора приблизительно составляла 1 мин 10 сек. Чтобы оценить качество работы системы в телекоммуникационной системе на ряду с исходными сигналами обрабатывались сигналы декодированные кодеками GSM и G.711.

Образцы сигналов, обработанные четырьмя различными эффектами, прослушивались группой экспертов. Задачей экспертов было оценить натуральность речи по 5 бальной шкале ("5" – натуральная речь; "4" – речь звучит натурально, заметны отдельные артефакты; "3" – синтезированная речь, имеются артефакты; "2" – ненатуральное звучание, отдельные слова не воспринимаются; "1" – полная потеря разборчивости). В таблице 1 приводятся средние оценки, полученные в результате эксперимента.

Таблица 1. Оценка натуральности образцов речи, обработанных в системе

Образец \ Эффект	Без обработки	Повышение тона $k = 1.2$	Понижение тона $k = 1.2$	Эффект <i>cycle</i>	Эффект <i>sinus</i>
Широкополосная речь	5	4.80	4.75	4.30	3.95
Речь после кодека GSM	4.25	4.15	4.20	3.85	3.70
Речь после кодека G.711	4.15	4.05	4.15	3.85	3.75
Речь после кодеров G.711 и GSM	4.10	4.00	4.10	3.80	3.60

На вход системы подавались сигналы трех различных категорий: широкополосная речь, узкополосная речь и узкополосная речь, декодированная после сжатия кодеками GSM и G.711. Таким образом, для каждой категории сигнала и каждого эффекта изменения тона было синтезировано 9 различных образцов по числу дикторов. Каждый из полученных образцов оценивался группой экспертов, а в таблицу заносились только средние значения оценок.

Очевидно, что натуральность речи будет тем выше, чем меньше изменений вносится в контур частоты основного тона. Это подтверждается и экспериментальными данными. Эффекты понижения и повышения вносят наименьшие изменения и по этой причине они получили более высокую оценку. Эффект '*cycle*' имеет временные интервалы, на которых динамика изменения контура частоты основного тона остается неизменной, поэтому натуральность этого эффекта выше чем

у эффекта ‘*sinus*’, который вносит систематическое изменение в натуральную форму контура.

**6. Заключение.** В работе предложено решение задачи изменения основного тона речи на основе гибридной модели речевого сигнала. Обработка вокализированной составляющей сигнала выполняется в измененном масштабе времени, что позволяет выполнять точное разделение гармоник и оценивать их мгновенные параметры при помощи узкополосной фильтрации. Для снижения вычислительной сложности используется полифазная реализация ДПФ-модулированного банка фильтров. Полученные результаты моделирования свидетельствуют о применимости предложенного решения в широкополосных и узкополосных каналах связи.

### Литература

1. *Flanagan J.L., Golden R. M.* Phase vocoder // Bell System Technical Journal, 1966. vol. 45, pp. 1493-1509.
2. *Levine S., Smith J.* A sines+transients+noise audio representation for data compression and time/pitch scale modifications // Signal processing: proceedings of 105th AES convention, San Francisco, USA, San Francisco, Preprint 1998. № 4781. 21 p.
3. *Serra X.* A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition: PhD thesis // Stanford, 1989. 178 p.
4. *Kawahara H., Takahashi T., Morise M., Banno H.* Development of exploratory research tools based on TANDEM-STRAIGHT // Proc. APSIPA, Sapporo, Japan, Oct. 2009.
5. *Kawahara H., Morise M.* Analysis and synthesis of strong vocal expressions: extension and application of audio texture features to singing voice // Proc. ICASSP'2012, Kyoto, Japan, March 2012. pp. 5389–5392.
6. *Erro D., Sainz I., Navas E., Hernaez I.* Improved HNM-based vocoder for statistical synthesizers // Proc. INTERSPEECH, Florence, Italy, Aug. 2011.
7. *Painter T., Spanias A.* Sinusoidal analysis-synthesis of audio using perceptual criteria // EURASIP Journal on Applied Signal Processing. 2003. № 1. pp. 15-20.
8. *Degottlex G., Stylianou Y.* A full-band adaptive harmonic representation of speech // Proc. INTERSPEECH, Portland, Oregon, USA, Sep. 2012.
9. *Azarov E., Vashkevich M., Petrovsky A.* Instantaneous pitch estimation based on RAPT framework // Proc. EUSIPCO, Bucharest, Romania, Aug. 2012, pp. 2787-2791.
10. *Азаров И.С., Вашкевич М.И., Петропский А.А.* Алгоритм оценки мгновенной частоты основного тона речевого сигнала // Цифровая обработка сигналов. Москва: 2012. №4. С. 49-57.
11. *Talkin D.* A Robust Algorithm for Pitch Tracking (RAPT) // Speech Coding & Synthesis, W B Kleijn, K K Paliwal eds, Elsevier ISBN 0444821694, 1995.
12. *Vaidynathan P.P.* Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial // Processing of the IEEE. January, 1990. vol. 78, no 1. pp. 56–93.

### References

1. Flanagan J.L., Golden R. M. Phase vocoder. Bell System Technical Journal, 1966. vol. 45, pp. 1493-1509.

2. Levine S., Smith J. A sines+transients+noise audio representation for data compression and time/pitch scale modifications. Signal processing: proceedings of 105th AES convention, San Francisco, USA, San Francisco, Preprint 1998. no. 4781. 21 p.
3. Serra X. A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition: PhD thesis. Stanford, 1989. 178 p.
4. Kawahara H., Takahashi T., Morise M., Banno H. Development of exploratory research tools based on TANDEM-STRAIGHT. Proc. APSIPA, Sapporo, Japan, Oct. 2009.
5. Kawahara H., Morise M. Analysis and synthesis of strong vocal expressions: extension and application of audio texture features to singing voice. Proc. ICASSP2012, Kyoto, Japan, March 2012. pp. 5389–5392.
6. Erro D., Sainz I., Navas E., Hernaez I. Improved HNM-based vocoder for statistical synthesizers. Proc. INTERSPEECH, Florence, Italy, Aug. 2011.
7. Painter T., Spanias A. Sinusoidal analysis-synthesis of audio using perceptual criteria. EURASIP Journal on Applied Signal Processing. 2003. no. 1. pp. 15–20.
8. Degottlex G., Stylianou Y. A full-band adaptive harmonic representation of speech. Proc. INTERSPEECH, Portland, Oregon, USA, Sep. 2012.
9. Azarov E., Vashkevich M., Petrovsky A. Instantaneous pitch estimation based on RAPT framework. Proc. EUSIPCO, Bucharest, Romania, Aug. 2012, pp. 2787–2791.
10. Azarov E., Vashkevich M., Petrovsky A. [An algorithm for instantaneous pitch estimation of speech]. *Cifrovaja obrabotka signalov – Digital Speech Processing*. Moscow: 2012. no. 4, pp. 49–57. (In Russ.).
11. Talkin D. A Robust Algorithm for Pitch Tracking (RAPT). Speech Coding & Synthesis, W B Kleijn, K K Paliwal eds, Elsevier ISBN 0444821694, 1995.
12. Vaidynathan P.P. Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial. Processing of the IEEE. January, 1990. vol. 78, no 1. pp. 56–93.

**Азаров Илья Сергеевич** — к-т тех. наук., доцент кафедры электронных вычислительных средств БГУИР. Область научных интересов: цифровая обработка речевых сигналов. Число научных публикаций — 42. azarov@bsuir.by, www.bsuir.by; БГУИР, ул. П. Бровки 6, г. Минск, 220013, РБ; р.т. +375 (17) 293-8805.

**Azarov Elias** — Ph.D., associate professor of computer engineering department, BSUIR. Research interests: digital speech processing. The number of publications — 42. azarov@bsuir.by, www.bsuir.by; BSUIR, 6, P.Brovky str., 220013, Minsk, RB; office phone +375 (17) 293-8805.

**Вашкевич Максим Иосифович** — к-т тех. наук., доцент кафедры электронных вычислительных средств БГУИР. Область научных интересов: цифровая обработка сигналов. Число научных публикаций — 36. vashkevich@bsuir.by, www.bsuir.by; БГУИР, ул. П. Бровки 6, г. Минск, 220013, РБ; р.т. +375 (17) 293-8805.

**Vashkevich Maxim** — Ph.D., associate professor of computer engineering department, BSUIR. Research interests: digital signal processing. The number of publications — 36. vashkevich@bsuir.by, www.bsuir.by; BSUIR, 6, P.Brovky str., 220013, Minsk, RB; office phone +375 (17) 293-8805.

**Лихачев Денис Сергеевич** — к-т тех. наук., доцент, доцент кафедры электронных вычислительных средств БГУИР. Область научных интересов: цифровая обработка сигнала-

лов. Число научных публикаций — 40. [likhachov@bsuir.by](mailto:likhachov@bsuir.by), [www.bsuir.by](http://www.bsuir.by); БГУИР, ул. П. Бровки 6, г. Минск, 220013, РБ; р.т. +375 (17) 293-8805.

**Likhachov Denis** — Ph.D., associate professor; associate professor of computer engineering department, BSUIR. Research interests: digital speech processing. The number of publications — 40. [vashkevich@bsuir.by](mailto:vashkevich@bsuir.by), [www.bsuir.by](http://www.bsuir.by); BSUIR, 6, P.Brovky str., 220013, Minsk, RB; office phone +375 (17) 293-8805.

**Петровский Александр Александрович** — д-р тех. наук, профессор, заведующий кафедрой электронных вычислительных средств БГУИР. Область научных интересов: цифровая обработка сигналов. Число научных публикаций — более 600. [palex@bsuir.by](mailto:palex@bsuir.by), [www.bsuir.by](http://www.bsuir.by); БГУИР, ул. П. Бровки 6, г. Минск, 220013, РБ; р.т. +375 (17) 293-2340.

**Petrovsky Alexander** — Ph.D., Dr. Sci., professor, head of computer engineering department, BSUIR. Research interests: digital speech processing. The number of publications — more than 600. [palex@bsuir.by](mailto:palex@bsuir.by), [www.bsuir.by](http://www.bsuir.by); BSUIR, 6, P.Brovky str., 220013, Minsk, RB; office phone +375 (17) 293-2340.

## РЕФЕРАТ

*Азаров И.С., Вашкевич М.С., Лихачев Д.С., Петровский А.А.* **Изменение частоты основного тона речевого сигнала на основе гармонической модели с нестационарными параметрами.**

В статье предлагается решение задачи изменения частоты основного тона речевого сигнала. Необходимость решения данной задачи возникает во многих речевых приложениях таких как конверсия голоса, коррекция акцента, обеспечение конфиденциальности диктора и др.

Разработанная схема обработки вокализованной части речевого сигнала основывается на гармонической модели с нестационарными (изменяющимися в каждый момент времени) параметрами. Оценка параметров модели выполняется с учетом контура мгновенной частоты основного тона. Входной речевой сигнал масштабируется во времени для того, чтобы обеспечить его стационарность, а затем выполняется узкополосная фильтрация, разделяющая гармоники основного тона. В результате фильтрации формируются аналитические (комплексные) сигналы, которые описываются при помощи параметров синусоидальной модели. На основе анализа смежных значений мгновенной частоты определяется степень вокализации каждой гармоники. Затем выполняется синтез квазипериодического сигнала по полученным параметрам, который вычитается из исходного речевого сигнала. В результате выполняется разделение речевого сигнала на вокализованную и невокализованную части. Одновременно синтезируется квазипериодический сигнал с целевой частотой основного тона, который складывается с выделенной невокализованной частью.

На основании субъективной оценки результатов показано, что разработанный способ обеспечивает высокую натуральность и разборчивость синтезированной речи и может применяться как в широкополосных так и в узкополосных каналах связи с различными стандартами кодирования (в том числе с кодеками G.711 и GSM).

## SUMMARY

*Azarov E., Vashkevich M., Likhachov D., Petrovsky A.* **Pitch modification of speech signal using harmonic model with time-varying parameters.**

The paper presents a solution to the problem of pitch modification of speech. The problem occurs in different speech processing applications such as voice conversion, accent correction, hiding speaker's personality and other.

Developed processing scheme for voiced part of speech is based on the harmonic model with nonstationary (time-varying) parameters. Parameters of the model are estimated synchronously with instantaneous contour of pitch. Input speech is warped in time domain in order to ensure stationarity of pitch and then narrow-band filtering is applied which separates individual pitch harmonics. The filtering results in analytical (complex) subband signals which can be represented by means of sinusoidal modeling. Vocalization degree of each harmonic is extracted by analysis of instantaneous frequency values of adjacent frames. Vocalized part of the signal is synthesized using extracted sinusoidal parameters and subtracted from the source signal. That results in deterministic/stochastic decomposition of the signal. The new vocalized part of the signal with target pitch values is synthesized and added to the extracted stochastic part.

Using subjective listening tests it is shown that developed system provides high naturalness and intelligibility of reconstructed speech and can be applied to wideband and narrowband communication channels with various coding standards (including G.711 and GSM).

О.О. БАСОВ, М.В. НОСОВ, В.А. ШАЛАГИНОВ  
**ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК ДЖИТТЕРА ПЕРИОДА  
 ОСНОВНОГО ТОНА РЕЧЕВОГО СИГНАЛА**

*Басов О.О., Носов М.В., Шалагин В.А. Исследование характеристик джиттера периода основного тона речевого сигнала.*

**Аннотация.** Стремление к повышению качества кодирования обусловило необходимость более глубокого анализа характеристик речевого сигнала и синтеза его прецизионных моделей. В работе предложен научно-методический инструментарий оценки периода основного тона речевого сигнала и разделения его джиттера на периодическую и случайную компоненты. Результаты его использования позволили улучшить качество субъективного восприятия речевого сигнала за счет уточнения модели формирования значений периода основного тона и установить наличие корреляции периодического джиттера с пульсом говорящего человека.

**Ключевые слова:** джиттер, основной тон, речевой сигнал, дискретное преобразование Фурье.

*Basov O.O., Nosov M.V., Shalaginov V.A. Pitch-jitter analysis of the speech signal.*

**Abstract.** The desire to improve encoding quality determined a deeper analysis of the characteristics of speech signal and the synthesis of its precision models. This paper presents a scientific-methodological tool for assessing the tone of the speech signal and its division of jitter by the periodic and random components. Results of its use have improved the quality of the subjective perception of the speech signal by specifying the values of the pitch and established an existence of correlation of periodic jitter with the speaking person's pulse.

**Keywords:** jitter, pitch, speech signal, discrete Fourier transform.

**1. Введение.** В задачах анализа речевого сигнала (при его кодировании) достаточно часто возникает ошибка определения вокализации кадра речи, связанная с наличием коротких изолированных тонов [10]. С целью уменьшения указанной ошибки в кодеках речи, основанных на линейном предсказании со смешанным возбуждением [25], кадр речевого сигнала классифицируется на три типа (рис. 1): невокализованный, вокализованный и слабо вокализованный (вокализованный с джиттером, jittery voiced). Последний класс соответствует случаю, когда возбуждение аperiodическое, но не случайное, и характерен для переходных кадров. Для отнесения кадра речи к слабо вокализованному необходимо определение наличия джиттера периода основного тона (ОТ).

Под джиттером периода ОТ речевого сигнала понимается [24] изменение данного периода, оцениваемое путем вычитания из каждого значения последовательности значений периода ОТ его ближайшего (предшествующего) значения или их комбинации:

$$Jitter_i^{(1)} = T_i - T_{i-1}, \quad (1)$$

$$Jitter_i^{(2)} = T_i - \frac{1}{N_0} \sum_{j=n_0}^{n_0+N_0-1} T_j, \quad i = 1 \dots n_0 \dots N, \quad (2)$$

где  $T_i$  – значение периода ОТ, вычисленное на  $i$ -м кадре речевого сигнала;  $n_0$  – номер кадра, соответствующий началу последовательности значений периода ОТ (вокализованному фрагменту речи);  $N_0$  – длина такой последовательности;  $N$  – длительность анализируемого речевого сигнала (в отсчетах).

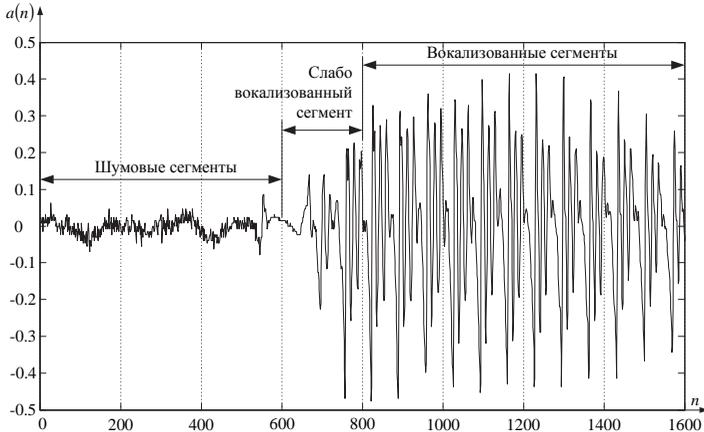


Рис. 1. Классификация речевых кадров во временной области

Использование характеристик джиттера периода ОТ при решении задач синтеза речевого сигнала позволяет повысить качество его восстановления. Значение периода ОТ для слабо вокализованных кадров речи, как правило, представляется в виде:

$$T_{jitter} = T(1 + Jitter \cdot x), \quad (3)$$

где  $T$  – декодированное и интерполированное значение периода основного тона;  $Jitter$  – значение джиттера, принятое в [16] равным 0,25 для «большинства практических ситуаций»;  $x$  – случайная величина, равномерно распределенная в интервале  $[-1, 1]$ .

В [21–23] приводятся экспериментальные данные, указывающие на то, что джиттер периода ОТ (2)–(3) имеет непараметрическое распределение и может быть разделен на постоянную и случайную компоненты. Указанный факт может быть объяснен как акустическими особенностями речеобразования: характером колебаний голосовых

связок (включая нерегулярность), состоянием артикуляционного аппарата при речеобразовании а также определенным воздействием пульсации кровотока [7], так и фонетическими – влиянием интонационного рисунка озвученной фразы.

Выявленное противоречие определяет необходимость разработки методического инструментария точного выделения периода ОТ речевого сигнала, разделения его джиттера на случайную и периодическую компоненты и исследования их характеристик.

**2. Определение периода ОТ речевого сигнала.** Несмотря на огромное число предложенных алгоритмов [1, 13, 15, 20], задача практического построения помехоустойчивых анализаторов ОТ, надежно функционирующих в присутствии акустического шума или при ограничении частотного диапазона речи, еще далека до окончательного решения. Для использования в различных условиях человеко-машинного взаимодействия разработан и реализован в [14] представленный ниже алгоритм определения периода ОТ (рис. 2).

Исходными данными для его функционирования являются:

$K$  – длительность временного окна анализа;

$N$  – размерность преобразования Фурье;

$\mu$  – допустимое отклонение траектории ОТ;

$s(n)$  – анализируемый речевой сигнал.

Особенностями разработанного алгоритма определения ОТ по сравнению с существующими решениями [1, 20] являются:

– ограничение поиска траектории ОТ конечным числом точек с целью реализации анализатора в реальном масштабе времени;

– использование комбинированного метода независимой оценки траектории ОТ [3] по интервалам прошлых и будущих кадров РС с последующим выбором лучшего результата, эффективно снижающего ошибки определения ОТ в начале и конце вокализованных звуков;

– определение дробного периода ОТ для его оптимального целочисленного значения.

Оценка качества разработанного алгоритма определения ОТ, согласно методики [1], основана на расчете процента грубых ошибок от общего числа выдаваемых оценок для заданных тестовых сигналов:

$$GPE = \frac{100}{F} \sum_{f=1}^F \left\{ \begin{array}{l} 1, \text{ если } |NPE_f| \geq \varepsilon \\ 0, \text{ если } |NPE_f| < \varepsilon \end{array} \right\} \% ; NPE_f = NP_f - 1 ; NP_f = \frac{T_{OT}^{(f)}}{T_{crit}^{(f)}} ,$$

где  $F$  – количество измерений ОТ;  $NPE_f$  – нормированная ошибка оценки ОТ;  $\varepsilon$  – порог разделения грубых ошибок и малых отклонений

в оценке ОТ;  $NP_f$  – нормированная оценка периода ОТ;  $T_{OT}^{(f)}$  – оценка задержки ОТ на выходе анализатора ОТ;  $T_{ctr}^{(f)}$  – контрольное значение задержки ОТ для  $f$ -й точки измерения, известно заранее.

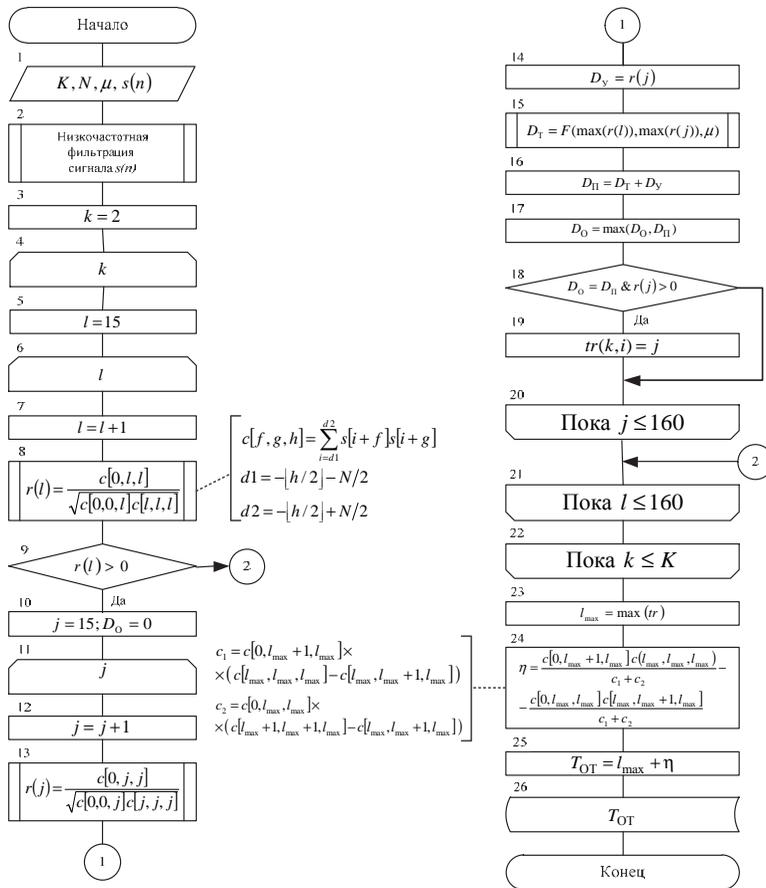


Рис. 2. Алгоритм определения периода ОТ речевого сигнала

Результаты сравнения разработанного алгоритма с помехоустойчивостью существующих выделителей ОТ на записях (частота дискретизации 8 кГц) речи трех дикторов, читающих текст газетных статей (табл. 1), свидетельствуют о его превосходстве по точности в различных акустических условиях.

Таблица 1. Сравнение точности различных алгоритмов определения периода основного тона речевого сигнала

Характеристика акустических условий	<i>GPE</i> , %			
	G.729	G.723	FS 1017	Разработанный алгоритм
Отсутствие акустического шума	10.2	10.8	4.5	4.3
Смесь речевого сигнала и белого шума (ОСШ 0 дБ)	29.5	31.4	47.3	6.9

Результаты применения разработанного алгоритма (рис. 2) и расчета джиттера периода ОТ, согласно выражений (1) и (2), представлены ниже (рис. 3).

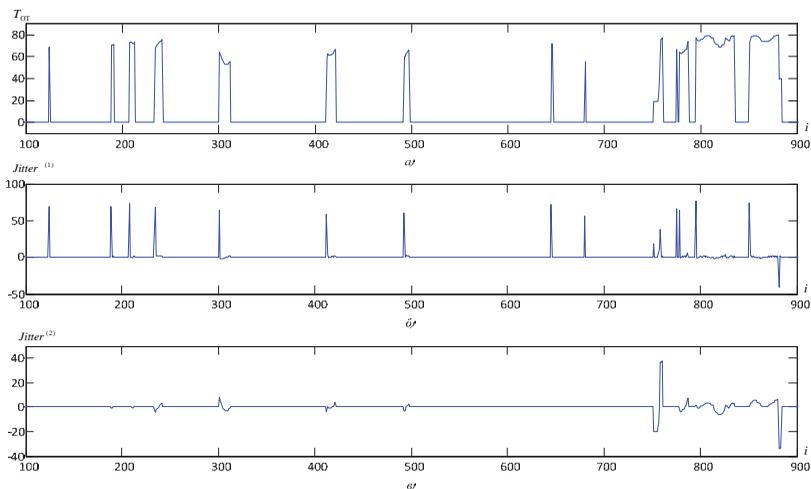


Рис. 3. Результаты определения периода ОТ речевого сигнала и его джиттера: *a* – траектория основного тона; *б, в* – оценки джиттера периода ОТ согласно выражений (1) и (2) соответственно

**3. Разделение общего джиттера периода ОТ на периодическую и случайную составляющие.** Для отделения периодического джиттера (*PJ* – Periodic Jitter) от случайного (*RJ* – Random Jitter) широко используются спектральные методы [3, 5, 17, 19, 26, 27], основанные на дискретном преобразовании Фурье (ДПФ).

Однако анализируемый сигнал (рис. 3, *б, в*) может содержать неизвестные значения общего джиттера периода ОТ, например, в случае отсутствия периода ОТ на некокализованных кадрах речи. Их наличие не позволяет напрямую использовать спектральный метод отде-

ления периодического джиттера  $PJ$  от случайного  $RJ$ , так как для последовательности с неизвестными значениями нельзя выполнить ДПФ.

В [3, 5, 17, 19, 26, 27] для заполнения неизвестных значений предлагается использовать интерполяцию по ближайшим известным значениям. При этом может использоваться линейная, полиномиальная или сплайновая интерполяции (рис. 4).

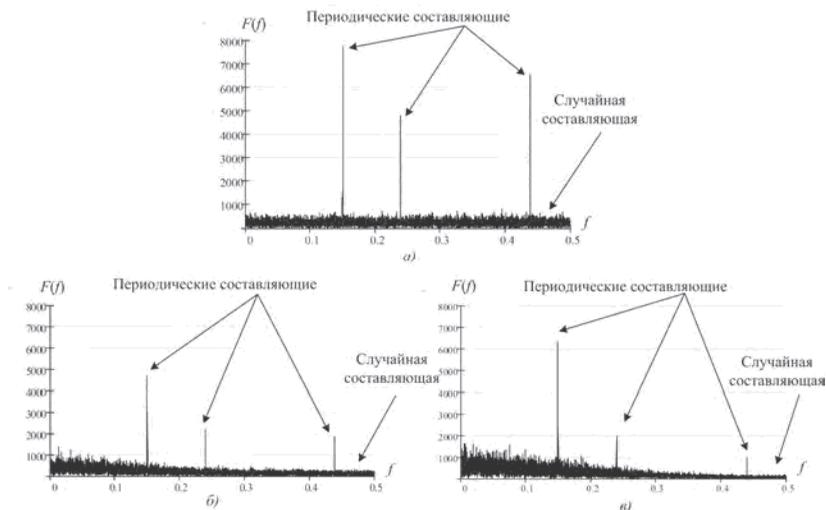


Рис. 4. Амплитудный спектр:

$a$  – известной последовательности значений джиттера;

$б$  – линейно интерполированной последовательности;

$в$  – последовательности, интерполированной кубическими сплайнами

Использование интерполяции неизвестных значений по ближайшим известным искажает спектр сигнала общего джиттера в сторону нижних частот с подавлением амплитуд высокочастотных составляющих. Такое искажение спектра ведет к ухудшению точности оценки периодического  $PJ$  и случайного  $RJ$  джиттера при использовании спектрального метода разделения.

Из известной последовательности значений джиттера (случайная составляющая – случайный гауссовский процесс с нулевым математическим ожиданием и среднеквадратическим отклонением  $\sigma = 2$ ; периодическая составляющая – три гармонических сигнала с относительными частотами  $f_1 = 0,15$ ,  $f_2 = 0,24$ ,  $f_3 = 0,44$ , амплитудами  $A_1 = 1,38$ ,  $A_2 = 0,563$ ,  $A_3 = 0,876$  и случайными фазами  $\varphi_1$ ,  $\varphi_2$  и  $\varphi_3$ ;

объем выборки  $N = 2^{14}$ ) удалены (с вероятностью 0,5) некоторые значения, которые затем линейно интерполированы по ближайшим известным значениям (рис. 4, б).

Из сравнения спектров (рис. 4, а и 4, б) видно, что применение интерполяции неизвестных значений по ближайшим известным искажает спектр последовательности значений джиттера в сторону нижних частот с подавлением амплитуд высокочастотных составляющих. Такой же эффект наблюдается при использовании полиномиальной интерполяции и интерполяции сплайнами (рис. 4, в). Такое искажение спектра ведет к ухудшению точности оценивания периодического  $PJ$  и случайного  $RJ$  джиттера при использовании спектрального метода разделения.

Для устранения недостатков существующих способов предлагается алгоритм разделения периодического и случайного джиттера (рис. 5), состоящий из следующих этапов [2, 4, 9].

I. Выполняется переход из временной в частотную область анализа. При этом последовательность значений джиттера, в которой неизвестные значения заменены на нули (шаг 3), взвешивается на окно данных  $PJ(i)$ , например, окно Блэкмана (шаг 4) для устранения эффекта растекания спектра при выполнении операции БПФ (шаг 5).

II. Определяются пиковые значения амплитудного спектра последовательности значений джиттера, соответствующие периодическому джиттеру  $PJ$ .

Периодический джиттер  $PJ$  представляет собой узкополосную помеху на одной или нескольких частотах. Для его отделения от случайного джиттера используются методы, основанные на фильтрации в частотной области. Для определения частотной характеристики (ЧХ) соответствующего фильтра необходимо определить частоты, на которых сосредоточен периодический джиттер.

Для поиска пиковых значений спектра, соответствующих частотам  $PJ$ , осуществляются следующие шаги.

Шаг 6. Вычисляется спектральная плотность мощности (СПМ) входной последовательности, сглаженной оконной функцией  $w(n)$ :

$$S_0(f_k) = \frac{1}{U} \left| \sum_{n=0}^{N-1} x(n)w(n)e^{-j\frac{2\pi}{N}kn} \right|^2,$$

где  $U = \sum_{n=0}^{N-1} w^2(n)$  – энергия используемого окна.

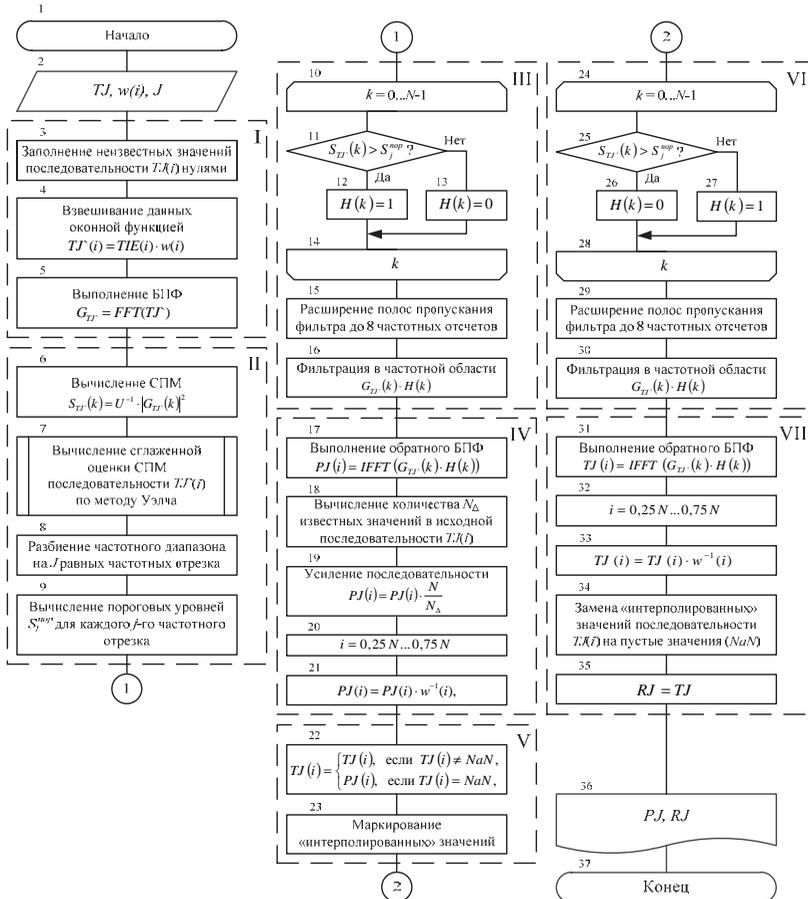


Рис. 5. Алгоритм разделения периодического и случайного джиттера ОТ

Шаг 7. Вычисляется smoothed оценка СПМ входной последовательности  $\hat{S}(f_k)$  по методу Уэлча:

$$\hat{S}(f_k) = \frac{1}{P} \sum_{p=1}^P I_p(f_k) = \frac{1}{PU} \sum_{p=1}^P |\dot{X}_p(k)|^2.$$

Как показано в [11], данная оценка СПМ является смещенной и состоятельной. Чем больше сегментов  $P$ , тем меньше дисперсия, но больше смещение оценок. Количество сегментов выбирается с учетом требуемой гладкости оценки СПМ и требуемой разрешающей способ-

ности по частоте. При этом имеется компромисс между смещением или спектральной разрешающей способностью и дисперсией оценки при постоянной длине выборки входной последовательности.

Шаг 8. Весь частотный диапазон разбивается на  $J$  одинаковых отрезка. Количество частотных отрезков  $J$  выбирается таким образом, чтобы более точно описать спектральную плотность мощности случайной составляющей общего джиттера.

Шаг 9. Вычисляются пороговые уровни  $S_j^{nop}$  с заданной доверительной вероятностью  $\alpha$  для каждого  $j$ -го отрезка как верхние границы доверительного интервала СПМ, полученные по одной реализации входной последовательности длиной  $N$ :

$$S_j^{nop} = \hat{S}_j \frac{2}{\chi_{\alpha/2}^2(2)},$$

где  $\chi_{1-\alpha/2}^2(2)$  –  $\chi^2$ -распределение при заданной доверительной вероятности  $\alpha$  и числе степеней свободы  $\mu = 2$ .

Если СПМ  $S_0(f_k)$  входной последовательности на данной частоте, превышает пороговый уровень  $S_j^{nop}$  в соответствующем частотном диапазоне, то считается, что на данной частоте сосредоточена узкополосная помеха, соответствующая периодической составляющей  $PJ$  общего джиттера. Так значения СПМ  $S_0(f)$  на относительных частотах  $f_1 = 0,016$  и  $f_2 = 0,418$  (рис. 6), превышающие пороговые значения  $S_j^{nop}$  при  $J = 16$ , соответствуют периодическому джиттеру.

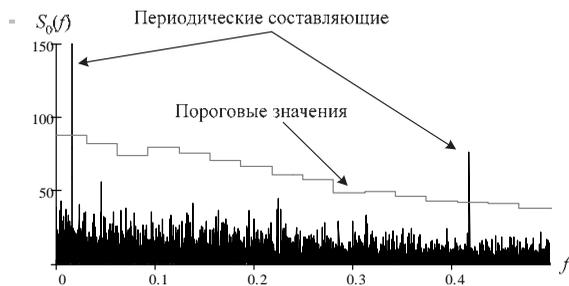


Рис. 6. Определение частот периодического джиттера

III. Формируется ЧХ фильтра и используется фильтрация с помощью БПФ для выделения постоянной составляющей общего джиттера. После определения частот, на которых сосредоточены узкополосные помехи, соответствующие  $PJ$ , необходимо отфильтровать их от случайного джиттера. Так как заранее количество и частоты узкополосных помех неизвестны, то необходимо использовать фильтрацию на основе ДПФ [6]. Особенностью данной фильтрации (шаги 10-16) является наличие «неверных» отсчетов на концах полученной после фильтрации последовательности.

IV. Выполняется обратное ДПФ от спектра, в котором остались только пиковые значения (шаг 17). Полученная последовательность, содержащая только периодические составляющие, усиливается (шаги 18-19) и взвешивается на обратную оконную функцию (шаг 21), чем устраняется эффект взвешивания окном.

V. В исходной последовательности значений общего джиттера неизвестные значения заменяются на соответствующие значения из полученной последовательности, описывающей  $PJ$ , и помечаются как «интерполированные» (шаги 22-23).

После заполнения неизвестных значений в последовательности значений джиттера выполняется удаление периодической составляющей  $PJ$  из смеси  $PJ+RJ$  спектральным методом за счет фильтрации в частотной области.

VI. Формируется ЧХ фильтра, который задерживает периодические компоненты джиттера и пропускает случайный джиттер (шаги 24-29). На шаге 30 выполняется фильтрация в частотной области путем перемножения спектра взвешенной входной последовательности и сформированной частотной характеристики фильтра.

VII. Выполняется обратное ДПФ (шаг 31) от спектра, содержащего только случайный джиттер. На шаге 32 исключаются крайние  $N/4$  отсчета, согласно метода фильтрации с помощью ДПФ, а получившаяся последовательность умножается на соответствующую обратную оконную функцию (шаг 33). Значения последовательности ошибок значений джиттера, помеченные как «интерполируемые», заменяются обратно (шаг 34) на неизвестные значения ( $NaN$ ).

Полученные последовательности (шаг 36) значений  $RJ$  и  $PJ$  используются для дальнейшего анализа.

**4. Исследование характеристик случайного и периодического джиттера.** В результате экспериментальных исследований установлено, что случайных джиттер  $RJ$  периода  $OT$ , соответствующий  $Jitter^{(1)}$  и  $Jitter^{(2)}$  (выражения (1) и (2) соответст-

венно), при критическом уровне значимости  $\alpha = 0,01$  (по критерию А. Н. Колмогорова [8]) имеет нормальное распределение (рис. 7).

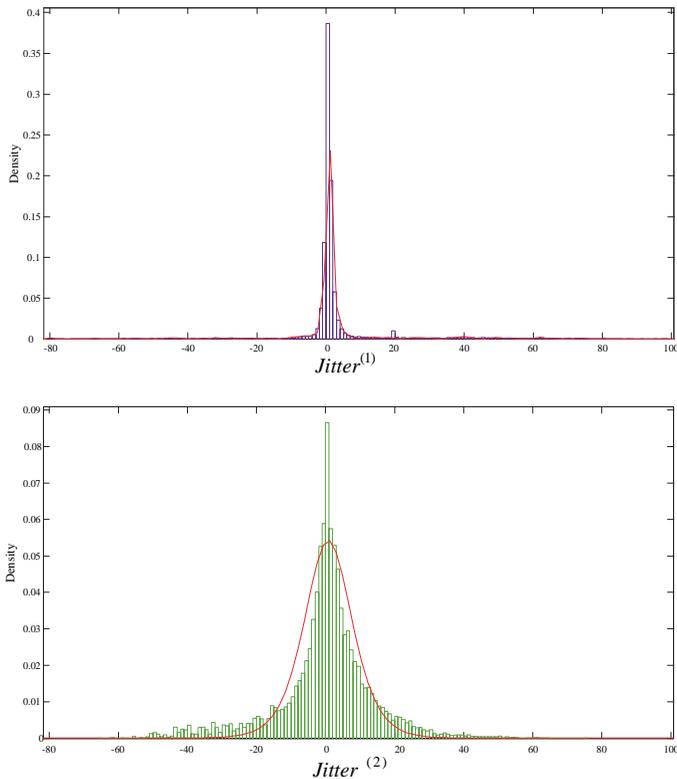


Рис. 7. Плотность распределения вероятности значений случайного джиттера

Представленные данные, а также наличие периодических составляющих  $PJ$  джиттера (рис. 8) позволяют модифицировать модель (3) периода ОТ, соответствующую формальному представлению джиттера  $Jitter^{(2)}$  (2), следующим образом:

$$T_{jitter} = T \left( 1 + Jitter \cdot x + \sum_j A_j \sin(2\pi f_j) \right), \quad (4)$$

где  $x$  – нормально распределенная случайная величина;  $A_j$  и  $f_j$  – амплитуды и частоты  $PJ$ .

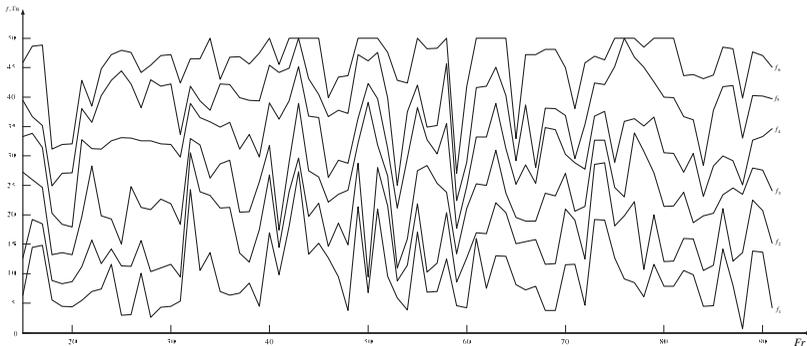


Рис. 8. Изменения частоты первых 6 максимумов спектра джиттера периода ОТ на длительности  $N$  речевого сигнала

Модификация процедур кодирования (разработанным алгоритмом определения периода ОТ речевого сигнала (рис. 2)) и декодирования (моделью восстановления периода ОТ (4)) по стандарту G.729 [12] позволила повысить субъективное качество синтезированного речевого сигнала. Однако их применение требует использования дополнительного информационного объема кадра передачи (для квантованных значений  $A_j$  и  $f_j$ ), либо поиска закономерностей формирования гармонических составляющих общего джиттера периода ОТ.

Использование полученных данных для других задач анализа речевого сигнала (например, оценки психофизиологического состояния человека по голосу) ограничивается существенной зависимостью оценок (1) и (2) от периода ОТ, значения которого на длительности  $N$  изменяются в широких пределах. Свободными от указанного недостатка являются джиттер, определенный в [18] как

$$Jitter_i^{(3)} = \frac{T_i - T_{i-1}}{T_i}, \quad (5)$$

а также введенный авторами

$$Jitter_i^{(4)} = \frac{T_i - T_{i-1}}{\frac{1}{N_0} \sum_{i=n_0}^{n_0+N} T_i}, \quad (6)$$

основанные на усреднении (1) относительно текущего значения и их комбинации (последовательности значений периода ОТ) соответственно.

Предварительные результаты статистического анализа свидетельствуют о том, что распределения вероятности значений случайных

составляющих джиттера (5) и (6) являются непараметрическими, а корреляционного – о наличии значительных связей между периодическими составляющими и значением пульса говорящего человека. Их подтверждение и строгое обоснование требуют проведения дальнейших исследований.

**7. Заключение.** В работе представлен научно-методический инструментарий точной оценки периода ОТ речевого сигнала, определения джиттера указанного периода и его разделения на периодическую и случайную компоненты. Результаты его применения подтверждают возможность отделения периодического джиттера от случайной составляющей общего джиттера периода ОТ и свидетельствуют о нормальном распределении последней.

Использование полученных результатов в задачах анализа и синтеза речевого сигнала позволяют улучшить качество его субъективного восприятия, однако требуется проведение дальнейшего исследования периодической составляющей джиттера (5), имеющего значительные корреляционные связи с пульсом говорящего человека.

## Литература

1. *Бабкин В.В.* Помехоустойчивый выделитель основного тона речи // Цифровая обработка сигналов и ее применение: материалы 7-й международной конф. М.: ИПУ РАН, 2005. Доклады, X-1. С. 175–178.
2. *Басов О.О., Шалагинов В.А., Офицеров А.И., Богданов С.П., Зацепин А.В.* Способ разделения джиттера периода основного тона речевого сигнала // патент № 2419166; Российская Федерация: МПК G 10 L 11/00, G 01 R 13/00 – № 2009144611/09; заявл. 01.12.09; опубл. 20.05.2011, Бюл. № 14. 14 с.: ил.
3. *Беллман Р., Дрейфус С.* Прикладные задачи динамического программирования / пер. с англ.; под ред. А. А. Первозванского. М.: Наука, гл. ред. физ.-мат. лит.-ры, 1965. 459 с.
4. *Большаков А.А., Каримов Р.Н.* Методы обработки многомерных данных и временных рядов : учебное пособие для вузов // М. : Горячая линия – Телеком, 2007. 522 с.
5. *Дамм В.А., Шалагинов В.А., Елифферевский В.В., Кутузов А.В.* Восстановление пропущенных значений ошибок временных интервалов при разделении компонентов общего фазового дрожания цифрового сигнала // Вестник РГРТУ. 2008. № 4 (26).
6. *Дамм В.А., Шалагинов В.А., Кутузов А.В., Королев М.В.* Частотный фильтр // патент на полезную модель № 76185; Российская Федерация: МПК H04B 1/69. № 2008115603; заявл. 21.04.08; опубл. 10.09.08, Бюл. № 35. 3 с.: ил.
7. *Калинцев Ю.К.* Разборчивость речи в цифровых вокодерах // М.: Радио и связь, 1991. 220 с.
8. *Кобзарь А.И.* Прикладная математическая статистика. Для инженеров и научных работников // М.: ФИЗМАТЛИТ, 2006. 816 с.
9. *Корольков А.В., Дамм В.А., Шалагинов В.А.* Разделение компонентов общего фазового дрожания цифрового сигнала данных // Вестник РГРТУ, 2009. № 3 (29).
10. *Маркел Дж.Д., Грей А.Х.* Линейное предсказание речи: пер. с англ. // под ред. Ю.Н. Прохорова и В.С. Звездина. М.: Связь, 1980. 308 с.: ил.
11. *Оппенгейм А.В., Шафер Р.В.* Цифровая обработка сигналов // М.: «Связь», 1979. 621 с.

12. Официальный сайт Международного союза электросвязи URL: [www.itu.int](http://www.itu.int) (дата обращения: 26.10.2013).
13. *Пирогов А.А.* Вокодерная телефония // М.: Связь, 1974. 536 с.
14. *Рыболовлев А.А., Басов О.О., Афанасьев А.А., Илюшин М.В., Катков О.Н.* Анализатор основного тона и сигнала тон-шум // патент на полезную модель № 78977; Российская Федерация: МПК G 10 L 11/00. – № 2008126595; заявл. 30.06.08; опубл. 10.12.08, Бюл. № 34. 2 с.: ил.
15. *Соболев В.Н.* Информационные технологии в синтетической телефонии // моногр. М.: ИРИАС, 2007. 360 с.
16. *Wai C. C.* Speech coding algorithms: Foundation and evolution of standardized coders // John Wiley & Sons, Inc. Hoboken, New Jersey, USA, 2003. 558 p.
17. *Draving S.D.* Method and apparatus for decomposing signal jitter using multiple acquisitions // US Patent № US 6898535 B2. May 24, 2005.
18. *National Committee for Information Technology Standardization (NCITS).* Fiber Channel – Methodologies for Jitter and Signal Quality Specification // Working draft for Rev. 10. Washington, DC, 2003. 228 p.
19. *Guenther M.L.* Method for decomposing timing jitter on arbitrary serial data sequences // US Patent № US 7254168 B2. Aug. 7, 2007.
20. *Huang X., Acero A., Hon H.-W.* Spoken language processing: a guide to theory, algorithm and system development // New Jersey, Prentice Hall, Inc, 2001. 980 p.
21. *Dong L.* Time series analysis of jitter in sustained vowels // ICPHS XVII, Hong Kong, 17-21 August, 2011. pp. 603–606.
22. *Schoentgen J., Guchteneere R.D.* Predictable and random components of jitter // Speech Communication. 1997. vol. 21. pp. 255–272.
23. *Schoentgen J., Guchteneere R. D.* Time series analysis of jitter // Journal of Phonetics. 1994. vol. 23(1-2), pp. 189–201.
24. *Silva D.G., Oliveira L.C., Andrea M.* Jitter Estimation Algorithms for Detection of Pathological Voices // EURASIP Journal on Advances in Signal Processing. Hindawi Publishing Corporation, 2009. pp. 1–9.
25. *Supplee L.M., Cohn R.P., Collura J. S., McCree A. V.* MELP: The new Federal Standard at 2400 bps // IEEE ICASSP-97 Conference. Munich, Germany. pp. 1591–1594.
26. *Tabatabaei S.* Jitter spectrum analysis using random sampling // US Patent Application Publication № US 2007/0110146 A1. May 17, 2007.
27. *Ward B.A., Tan K., Guenther M.L.* Apparatus and method for spectrum analysis-based serial data jitter measurement // US Patent № US 6832172 B2. Dec. 14, 2004.
28. *Wilstrup J.B., Petrich D.M.* Method and apparatus for jitter analysis // US Patent № US 6356850 B1. Mar. 12, 2002.

## References

1. Babkin V.V. [Noise-immune extractor of the speech pitch]. *Cifrovaja obrabotka signalov i ee primeneniye: materialy 7-j mezhdunarodnoj konf* [Digital signal processing and its application: proceedings of the 7th international conference]. М.: IPU RAN, 2005. Doklady, X-1. pp. 175–178.
2. Basov O.O., Shalaginov V.A., Oficerov A.I., Bogdanov S.P., Zacepin A.V. *Sposob razdeleniya dzhittera perioda osnovnogo tona rechevogo signala* [The way to detachment of a pitch jitter period of speech signal]. Patent RF no. 2419166, 2011.
3. Bellman R., Dreyfus S. *Applied Dynamic Programming*. Princeton University Press. 1962. 388 p. (Russ. ed.: Pervozvanskogo A. A. *Prikladnye zadachi dinamicheskogo programmirovaniya*. М.: Nauka, gl. red. fiz.-mat. lit-ry, 1965. 459 p.)
4. Bol'shakov A.A., Karimov R.N. *Metody obrabotki mnogomernykh dannykh i vremennykh ryadov: uchebnoe posobie dlja vuzov* [Methods of processing of multidimensional data

- and time series: a textbook for high schools] M. : Gorjachaja linija – Telekom. 2007. 522 p.
5. Damm V.A., Shalaginov V.A., Eliferevskij V.V., Kutuzov A.V. [Recovering missing values errors slots in the separation of components in common jitter digital signal]. *Vestnik RGRTU – Herald RGRTU*, 2008. no. 4 (26).
  6. Damm V.A., Shalaginov V.A., Kutuzov A.V., Korolev M.V. *Chastotnyj fil'tr* [Frequency filter]. Patent RF no. 76185, 2008.
  7. Kalincev Ju.K. Razborchivost' rechi v cifrovyh vokoderah [Speech intelligibility in digital vocoders]. M.: Radio i svjaz', 1991. 220 p.
  8. Kobzar' A.I. Prikladnaja matematicheskaja statistika. Dlja inzhenerov i nauchnyh rabotnikov [Applied Mathematical Statistics. For engineers and scientists]. M.: FIZMATLIT, 2006. 816 p.
  9. Korol'kov A.V., Damm V.A., Shalaginov V.A. [Separation of components total jitter digital data signal]. *Vestnik RGRTU – Herald RGRTU*, 2009. no. 3(29).
  10. Markel J.D., Grey A.H. *Linear Prediction of Speech*. 1976. 288 p. (Russ. ed.: Prohorova Ju.N., Zvezdina V.S. *Linejnoe predskazanie rechi*. M.: Svjaz'. 1980. 308 p.)
  11. Oppengejm A.V., Shafer R.V. *Cifrovaja obrabotka signalov* [Digital Image Processing]. M.: «Svjaz'». 1979. 621 p.
  12. *Oficial'nyj sajt Mezhdunarodnogo sojuza jelektrsvjazji* [Official web site of International Telecommunication Union]. Available at: [www.itu.int](http://www.itu.int). (accessed 26.10.2013). (In Russ.).
  13. Pirogov A.A. *Vokodernaja telefonija* [Vocoder telephony]. M.: Svjaz'. 1974. 536 p.
  14. Rybolovlev A.A., Basov O.O., Afanas'ev A.A., Iljushin M.V., Katkov O.N. *Analizator osnovnogo tona i signala ton-shum* [Analyzer pitch and signal of the tone-to-noise]. Patent RF no. 78977. 2008.
  15. Sobolev V.N. *Informacionnye tehnologii v sinteticheskoy telefonii* [Information technology in synthetic telephony]. M.: IRIAS, 2007. 360 p.
  16. Wai C. C. *Speech coding algorithms: Foundation and evolution of standardized coders*. John Wiley & Sons, Inc. Hoboken, New Jersey, USA, 2003. 558 p.
  17. Draving S.D. Method and apparatus for decomposing signal jitter using multiple acquisitions. US Patent no. US 6898535 B2. May 24, 2005.
  18. National Committee for Information Technology Standardization (NCITS). *Fiber Channel – Methodologies for Jitter and Signal Quality Specification*. Working draft for Rev. 10. Washington, DC, 2003. 228 p.
  19. Guenther M.L. Method for decomposing timing jitter on arbitrary serial data sequences. US Patent no. US 7254168 B2. Aug. 7, 2007.
  20. Huang X., Acero A., Hon H.-W. *Spoken language processing: a guide to theory, algorithm and system development*. New Jersey, Prentice Hall, Inc, 2001. 980 p.
  21. Dong L. Time series analysis of jitter in sustained vowels. ICPhS XVII, Hong Kong, 17-21 August, 2011. pp. 603–606.
  22. Schoentgen J., Guchteneere R.D. Predictable and random components of jitter. *Speech Communication*. vol. 21, 1997. pp. 255–272.
  23. Schoentgen J., Guchteneere R.D. Time series analysis of jitter. *Journal of Phonetics*. vol. 23(1-2), 1994. pp. 189–201.
  24. Silva D.G., Oliveira L.C., Andrea M. Jitter Estimation Algorithms for Detection of Pathological Voices. *EURASIP Journal on Advances in Signal Processing*. Hindawi Publishing Corporation, 2009. pp. 1–9.
  25. Supplee L.M., Cohn R.P., Collura J. S., McCree A. V. MELP: The new Federal Standard at 2400 bps. *IEEE ICASSP-97 Conference*. Munich, Germany. pp. 1591–1594.
  26. Tabatabaei S. Jitter spectrum analysis using random sampling. Patent US no. US 2007/0110146 A1. May 17, 2007.
  27. Ward B.A., Tan K., Guenther M.L. Apparatus and method for spectrum analysis-based serial data jitter measurement. Patent US no. US 6832172 B2. Dec. 14, 2004.

28. Wilstrup J.B., Petrich D.M. Method and apparatus for jitter analysis. Patent US no. US 6356850 B1. Mar. 12, 2002.

**Басов Олег Олегович** — к-т техн. наук, докторант Академии ФСО России. Область научных интересов: обработка и кодирование речевых и иконических сигналов, проектирование полимодальных инфокоммуникационных систем. Число научных публикаций — 135. oobasov@mail.ru; Академия ФСО России, Приборостроительная, 35, г. Орел, 302034, РФ; р.т. +7(4862)549533, факс +7(4862)541325. Научный консультант — И.А. Сайтов.

**Basov Oleg Olegovich** — Ph.D., Doctoral of the Academy of Federal Agency of Protection of Russian Federation. Research interests: processing and coding of speech and iconic signals, polymodal infocommunicational systems design. Number of scientific publications — 135. oobasov@mail.ru; Academy of Federal Agency of protection of Russian Federation, Priborostroitel'naya Street, 35, Orel, 302034, Russia, office phone +7(4862)549533, fax +7(4862)541325. Scientific adviser – Saitov I.A.

**Носов Максим Васильевич** — сотрудник Академии ФСО России. Область научных интересов: человеко-машинное взаимодействие, проектирование полимодальных интерфейсов. Число научных публикаций — 6. Nosovm@mail.ru; Академия ФСО России, Приборостроительная, 35, г. Орел, 302034, РФ; р.т. +7(4862)549801, факс +7(4862)541325. Научный руководитель — О.О. Басов.

**Nosov Maksim Vasilevich** — Member of the Academy of Federal Agency of Protection of Russian Federation. Research interests: human-computer interaction, design of polymodal interfaces. The number of publications — 6. Nosovm@mail.ru; Academy of Federal Agency of protection of Russian Federation, Priborostroitel'naya Street, 35, Orel, 302034, Russia, office phone +7(4862)549533, fax +7(4862)541325. Scientific director — O.O.Basov.

**Шалагинов Владимир Александрович** — к-т техн. наук, старший преподаватель Академии ФСО России. Область научных интересов: техническая защита информации. Число научных публикаций — 33. Vovila2003@rambler.ru; Академия ФСО России, Приборостроительная, 35, г. Орел, 302034, РФ; р.т. +7(4862)549801, факс +7(4862)541325.

**Shalaginov Vladimir Aleksandrovich** — Ph.D., Senior Teacher of the Academy of Federal Agency of Protection of Russian Federation. Research interests: technical protection of information. The number of publications — 33. Vovila2003@rambler.ru; Academy of Federal Agency of protection of Russian Federation, Priborostroitel'naya Street, 35, Orel, 302034, Russia, office phone +7(4862)549801, fax +7(4862) 541325.

## РЕФЕРАТ

### *Басов О.О., Носов М.В., Шалагинов В.А.* **Исследование характеристик джиттера периода основного тона речевого сигнала.**

Для повышения качества кодирования речи, основанного на линейном предсказании со смешанным возбуждением, возникает задача уменьшения ошибки отнесения кадра речевого сигнала к слабо вокализованному. Данная задача может быть решена на основе точного определения наличия джиттера периода основного тона.

Во втором разделе статьи предложен алгоритм, позволяющий повысить помехоустойчивость определения периода основного тона речевого сигнала по сравнению с алгоритмами работы существующих выделителей основного тона, что подтверждено результатами имитационного моделирования.

В третьем разделе произведен анализ существующих подходов к оценке джиттера периода основного тона речевого сигнала, указаны их недостатки. Авторами предложен свой способ определения джиттера, основанный на усреднении изменения периода основного тона относительно текущего значения.

В четвертом разделе предложен алгоритм разделения джиттера периода основного тона на случайную и периодическую составляющие. Особенность данного алгоритма заключается в возможности использования дискретного преобразования Фурье над последовательностью периодов основного тона при наличии в ней неизвестных значений, возникающих на невокализованных кадрах речи. Путем имитационного моделирования показано преимущество предлагаемого подхода заполнения неизвестных значений периода основного тона по сравнению с существующими методами, основанными на интерполяции по ближайшим известным значениям.

Проведенные экспериментальные исследования, результаты которого представлены в пятом разделе, позволили модифицировать модель периода основного тона, соответствующую формальному представлению джиттера. Использование полученных в статье результатов в задачах анализа и синтеза речевого сигнала позволит улучшить качество его субъективного восприятия.

Авторами выявлено направление дальнейшего исследования периодического джиттера периода основного тона речевого сигнала, имеющего значительные корреляционные связи с пульсом говорящего человека.

## SUMMARY

*Basov O.O., Nosov M.V., Shalaginov V.A.* **Pitch-jitter analysis of the speech signal.**

To increase the quality of speech coding based on a linear prediction with a mixed excitement, a problem arises to reduce the error classifying the speech signal frame to weakly voiced. This problem can be solved on the basis of a precise definition of the presence of jitter of the pitch period.

In the second section of the paper is proposed an algorithm that allows to increase the immunity determine the pitch period of the speech signal in comparison with the existing algorithms of highlighters pitch that is confirmed by the results of the simulation.

In the third section the analysis of existing approaches to measuring jitter period the main tone of the voice signal is executed, their shortcomings are identified. The authors propose their way to determine the jitter based on averaging the change of the pitch period relative to the current value.

In the fourth section, an algorithm for separation of jitter pitch period at random and periodic components is proposed. The peculiarity of this algorithm is the ability to use the discrete Fourier transform on the sequence of pitch periods in the presence therein of unknown values, arising at the unvoiced speech frames . By simulation study the advantage of the proposed approach of filling the unknown values of the pitch period in comparison with existing methods based on interpolation of the nearest known values is showed.

The experimental studies, the results of which are presented in the fifth section, allowed to modify the model of the pitch period corresponding to the formal presentation of jitter. The use of the obtained results in this paper for the analysis and synthesis of the speech signal will improve the quality of its subjective perception.

The authors identified the direction of future research of the periodic pitch-jitter of the speech signal which has significant correlations with the pulse of the speaker.

С.Д. ТИУНОВ, Р.В. МЕЩЕРЯКОВ, Д.В. ЧЕРНЫХ  
**ОПТИМИЗАЦИЯ ВЫЧИСЛЕНИЯ ОДНОВРЕМЕННОЙ  
МАСКИРОВКИ РЕЧЕВОГО СИГНАЛА**

---

*Тиунов С.Д., Мещеряков Р.В., Черных Д.В.* **Оптимизация вычисления одновременной маскировки речевого сигнала.**

**Аннотация.** В настоящей статье предлагается оптимизация вычисления одновременной маскировки речевого сигнала для реализации в задачах реального времени. Показана трудоемкость циклической свертки для итерационного алгоритма Тоома-Кука длины 4 и алгоритма на основе БПФ. Делается вывод об эффективности предлагаемых решений по вычислительной сложности и по объемам занимаемой памяти.

**Ключевые слова:** речевой сигнал, одновременная маскировка, циклическая свертка, оптимизация.

*Tiunov S.D., Meshcheryakov R.V., Chernykh D.V.* **Optimization of calculation of frequency masking of speech signal.**

**Abstract.** Paper proposes optimization of calculation of frequency masking of speech signal for use in real time applications. The complexity of circular convolution is shown for iterative Toom-Cook algorithm with length 4 and an algorithm based on the FFT. The conclusion about the efficiency of the proposed solutions is drawn on computation complexity and memory.

**Keywords:** speech signal, frequency masking, circular convolution, optimization.

---

**1. Введение.** Обработка речи включает в себя различные затратные операции над звуковым сигналом [1, 2]. Вместе с тем, использование новых подходов и методов обработки сигнала требует значительных вычислительных ресурсов и не может быть проведено в реальном масштабе времени [2-5]. Это связано с одной стороны с тем, что при разработке новых алгоритмов для обработки речи на первом месте стоит вопрос точности, а не вопрос вычислительной эффективности. Вместе с тем некоторые приложения речевых технологий подразумевают обработку сигнала в реальном времени. Данная статья предлагает эффективный численный метод для вычисления одновременной маскировки из модели слуховой системы [2, 6], что позволяет применять ее в программном комплексе для анализа речевого сигнала в норме и при патологии [7-10].

Одновременная маскировка является особенностью восприятия звука слуховой системой человека. Эта особенность заключается в том, что из двух близких частотных составляющих ухом будет воспринята только та, которая имеет большую интенсивность [10]. Моделирование эффекта одновременной маскировки позволяет выделить в спектрограмме сигнала наиболее важные (информативные) составляющие с точки зрения слуховой системы человека. Моделирование одновременной маскировки и связанные алгоритмы подробно описаны в [11].

Алгоритм одновременной маскировки может быть представлен в виде функции  $m$ , которая принимает на вход матрицу спектрограммы  $S_{KN}$  и дает на выходе двоичную матрицу одновременной маскировки  $M_{KN}$  (единица означает, что данная составляющая воспринимается слуховой системой). При этом для своей работы алгоритм использует матрицу  $H_{KL}$  – маскирующую матрицу (см. рисунок 1).



Рис. 1. Вычисление одновременной маскировки речевого сигнала как черный ящик

Формула для вычисления функционала  $m$  может быть записана следующим образом:

$$\forall k = 0 \dots K - 1, \forall n = 0 \dots N - 1:$$

$$M_{kn} = \left( \sum_{l=0}^{L-1} |S_{k'n}|^2 \cdot H_{kl} \right) \cdot b > |S_{kn}|^2, \quad (1)$$

$$\text{где } k' = \begin{cases} 0, & \text{если } k + l - L/2 < 0 \\ k + l - L/2, & \text{если } 0 \leq k + l - L/2 \leq K - 1 \\ K - 1, & \text{если } K - 1 < k + l - L/2. \end{cases}$$

$$b = \text{const}$$

Предполагая, что квадрат модуля спектра  $|S_{kn}|^2$  вычислен заранее, запишем вычислительную сложность (количество умножений  $M$  и количество сложений  $A$ ) «наивного» алгоритма одновременной маскировки (т. е. алгоритма, использующего формулы (1) непосредственно):

$$M(K, N, L) = N \cdot K \cdot (L - 1), \quad (2)$$

$$A(K, N, L) = N \cdot K \cdot (L + 1), \quad (3)$$

$$M + A = 2 \cdot N \cdot K \cdot L. \quad (4)$$

Рассчитаем требуемую вычислительную мощность для наивного алгоритма при различных значениях параметров (таблица 1). Для получения значения в FLOPS (количество вещественных операций в

секунду) необходимо принять  $N = f_q$ , где  $f_q$  – частота дискретизации сигнала.

Таблица 1. Расчеты мощности вычислительной машины для расчета одновременной маскировки по наивному алгоритму

$\xi$	Необходимая вычислительная мощность в GFLOPS								
0,1	0,014	0,022	0,029	0,061	0,092	0,123	0,238	0,356	0,475
0,01	0,023	0,034	0,045	0,086	0,129	0,172	0,319	0,479	0,639
0,001	0,027	0,040	0,053	0,102	0,154	0,205	0,385	0,578	0,770
0,0001	0,031	0,046	0,061	0,119	0,178	0,238	0,451	0,676	0,901
$K$	128	128	128	256	256	256	512	512	512
$N=f_q$	8	12	16	8	12	16	8	12	16

Измерение  $K$  – это количество частотных каналов в спектрограмме, оно зависит от приложения. Измерение  $L$  маскирующей матрицы  $H$  зависит от требуемой точности вычислений и используемой шкалы частот  $f_k$ . Представленные значения вычислительной сложности приведены в таблице 1 позволяют сделать вывод о необходимости создания новых алгоритмов, оптимизирующих вычислительную сложность «наивного» алгоритма одновременной маскировки. Необходимо отметить, что для повышения точности за счет увеличения числа каналов вычислительная сложность растет нелинейно.

**2. Вычисление одновременной маскировки путем цифровой фильтрации.** В общем случае формула одновременной маскировки не сводится к цифровой фильтрации, и оптимальный алгоритм вычисления ненамного превосходит наивный. Однако если ввести ограничения на маскирующую матрицу:

$$\forall k_1, k_2, l: H_{k_1 l} = H_{k_2 l}, \quad (5)$$

то матрица  $H_{KL}$  может быть заменена любой своей строкой  $H_L$ . С учетом особенностей вычисления  $k'$  можно заменить  $S_{k'}$  на расширенный вектор  $S'_{k+L}$ :

$$\forall k = 0 \dots K - 1, l = L - 1: S'_{k+l} = S_k. \quad (6)$$

Тогда формула (1) для вычисления одновременной маскировки принимает вид:

$$M_k = \sum_{l=0}^{L-1} |S'_{k+l}|^2 \cdot H_l > |S_k|^2 = (|S'_{k+l}|^2 \otimes H_L)_{k+L/2} > |S_k|^2, \quad (7)$$

где  $\otimes$  – оператор линейной свертки. Поскольку существуют эффективные методы вычисления линейной свертки, то введенное ограничение (7) приводит к снижению требуемой вычислительной мощности алгоритма. Для выяснения условий, при которых ограничение (7) будет выполняться, необходимо описать вычисление матрицы  $H_{KL}$ . Матрица  $H_{KL}$  вычисляется из маскирующей функции  $H(x, y)$ , связанной со шкалой частот  $f(x)$  следующими соотношениями:

$$\forall x, y \in \mathbb{R}: \\ H(x, y) = \exp\left(-\frac{(f(x+y) - f(x))^2}{2\sigma(x)^2}\right), \quad (8)$$

$$\sigma(x) = \frac{f(x)}{2.4 \cdot Q(x) \cdot \delta}, \quad (9)$$

$$Q(x) = B \cdot \frac{f(x)}{\alpha \cdot f(x) + \beta}, \quad (10)$$

$$\alpha, \beta, \delta, B = \text{const.} \quad (11)$$

Элементы матрицы  $H_{KL}$  вычисляются следующим образом:

$$\forall k = 0 \dots K - 1, l = 0 \dots L_{max} - L_{min}:$$

$$H_{kl} = H(k, k + L_{min}) / \sum_{i=L_{min}}^{L_{max}} H(k, i).$$

Значения  $L_{min}$  и  $L_{max}$  вычисляются исходя из требуемой точности  $\xi$ . Исходя из ограничения (5) произведем подстановку:

$$\forall k_1, k_2, l \in \mathbb{Z}: H_{k_1 l} = H_{k_2 l}, \\ \forall x_1, x_2, y \in \mathbb{R}: H(x_1, y) = H(x_2, y), \\ \exp\left(-\frac{(f(x_1+y) - f(x_1))^2}{2\sigma(x_1)^2}\right) = \exp\left(-\frac{(f(x_2+y) - f(x_2))^2}{2\sigma(x_2)^2}\right), \\ \frac{f(x_1+y) - f(x_1)}{\alpha \cdot f(x_1) + \beta} = \frac{f(x_2+y) - f(x_2)}{\alpha \cdot f(x_2) + \beta}.$$

При  $\alpha \neq 0$  справедливо:

$$\frac{\alpha \cdot f(x_1+y) + \beta}{\alpha \cdot f(x_1) + \beta} = \frac{\alpha \cdot f(x_2+y) + \beta}{\alpha \cdot f(x_2) + \beta}.$$

Введем следующие обозначения:

$$\gamma = \frac{\beta}{\alpha}, \quad g(x) = f(x) + \gamma. \quad (12)$$

Получаем:

$$\begin{aligned} \frac{g(x_1 + y)}{g(x_1)} &= \frac{g(x_2 + y)}{g(x_2)}, \\ \frac{g(x + y)}{g(x)} &= h(y). \end{aligned} \quad (13)$$

Условие (13) накладывает ограничение на используемую шкалу частот. Очевидно, что степенная функция  $g(x) = \exp(ax + b)$  удовлетворяет этому условию. Используя  $g_1(x) = \ln g(x)$ ,  $h_1(x) = \ln h(x)$ ,  $g_2(x) = g_1(x) - g_1(0)$  получаем:

$$\begin{aligned} \frac{g(x + y)}{g(x)} &= h(y), \\ g_1(x + y) - g_1(x) &= h_1(y), \\ g_1(0 + y) - g_1(0) &= h_1(y), \\ g_1(x + y) - g_1(x) &= g_1(y) - g_1(0), \\ g_2(x + y) - g_2(x) &= g_2(y), \\ g_2(x + y) &= g_2(x) + g_2(y). \end{aligned} \quad (14)$$

Условие (14) – это функциональное уравнение Коши, которое имеет только одно непрерывное вещественное решение  $g_2(x) = ax$ , где  $a \in \mathbb{R}$  [12]. Следовательно,  $f(x)$  имеет вид:

$$\begin{aligned} f(x) = g(x) - \gamma &= \exp(g_1(x)) - \gamma = \exp(g_2(x) + g_1(0)) - \gamma, \\ f(x) &= \exp(ax + b) - \gamma, \quad a, b, \gamma = \text{const}. \end{aligned} \quad (15)$$

При выполнении условия (15) имеется возможность вычисления одновременной маскировки через цифровую фильтрацию.

Оптимальным вариантом для цифровой фильтрации для больших значений длины окна  $L$  является использование теоремы о свертке и быстрого преобразования Фурье. В этом случае может быть использован метод перекрытия с накоплением [13,14]. Выведем формулу для количества операций, необходимых алгоритму, основанному на цифровой фильтрации, для обработки входного массива, соответствующего одной секунде. Пусть входной сигнал имеет частоту дискретизации  $f_q$ , спектрограмма имеет  $K$  каналов, а окно фильтрации

для расчета одновременной маскировки имеет размер  $L$ . Поскольку на каналах с номерами  $k < L$  и  $k \geq K - L$  необходимо «расширение» области частот (см. также формулу (6)), то к каждому входному вектору  $M_K$  необходимо добавить по  $L/2$  отсчетов с каждой стороны (см. рисунок 2). Тогда входная матрица размера  $f_q \times K$  для фильтрации преобразуется в вектор длиной  $f_q \times (K + L)$ .

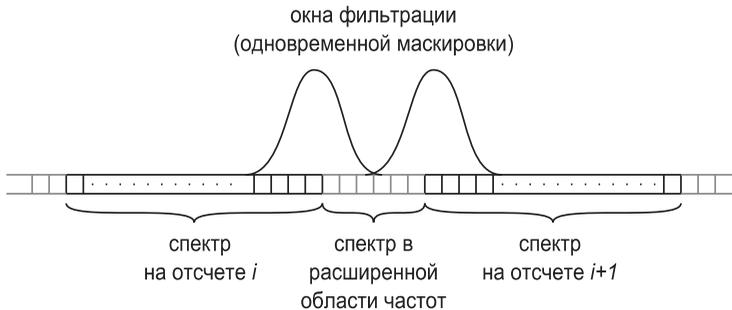


Рис. 2. Иллюстрация расширения области частот спектра для расчета одновременной маскировки

Пусть далее используется алгоритм расчета циклической свертки размера  $N$ , который требует  $M(N)$  умножений,  $A(N)$  сложений,  $S(N) = M(N) + A(N)$ . Тогда для обработки одной секунды входных данных, необходимо  $P$  операций, где:

$$P(K, L, N, f_q) = S(N) \cdot f_q \cdot (K + L) / (N - L). \quad (16)$$

Член  $S(N)$  в формуле (16) зависит от алгоритма вычисления циклической свертки размера  $N$ . Сравним алгоритмы с точки зрения минимизации количества операций.

При использовании алгоритма циклической свертки на основе БПФ вычисления ведутся фактически с комплексными числами. Поэтому данный алгоритм можно использовать для вычисления двух вещественных циклических свертки одновременно. Это обстоятельство следует учесть при расчете вычислительной сложности алгоритма на основе БПФ. Формулы (17) содержат описание данного алгоритма:

$$\begin{aligned} s_1 &= g \otimes d_1 \quad s_2 = g \otimes d_2 \quad G = \Phi(g), \\ d &= d_1 + i \cdot d_2 \quad D = \Phi(d), \\ S &= G \times D \quad s = \widehat{\Phi}(S), \\ s_1 &= \Re_s \quad s_2 = \Im_s. \end{aligned} \quad (17)$$

В формулах (17) оператор  $\otimes$  означает циклическую свертку, оператор  $\times$  – поточечное умножение векторов, операторы  $\Phi$  и  $\hat{\Phi}$  – прямое и обратное преобразование Фурье.

В таблице 2 приведен расчет вычислительной мощности в GFLOPS, необходимой для вычисления одновременной маскировки в реальном времени, для двух различных алгоритмов циклической свертки различной длины: итерационного алгоритма Тоома-Кука для  $n=4$  и алгоритма на основе БПФ, а также для наивного алгоритма.

Таблица 2. Расчеты требуемой мощности вычислительной машины для расчета одновременной маскировки в реальном времени

Размер ( $N$ )	Необходимая вычислительная мощность в GFLOPS								
	Наивный алгоритм								
—	0,014	0,034	0,053	0,061	0,129	0,205	0,238	0,479	0,770
Алгоритм циклической свертки на основе БПФ Кули-Тьюки									
16	0,027	0,075	0,168	0,486					
32	<b>0,026</b>	0,047	0,070	0,075	0,179	0,380	0,854		
64	0,028	<b>0,047</b>	<b>0,066</b>	<b>0,066</b>	0,115	0,172	0,184	0,394	0,783
128	0,032	0,051	0,071	0,069	<b>0,112</b>	<b>0,157</b>	<b>0,157</b>	0,267	0,398
256	0,037	0,058	0,079	0,076	0,120	0,165	0,161	<b>0,258</b>	<b>0,362</b>
512	0,042	0,065	0,088	0,085	0,132	0,180	0,174	0,272	0,375
Итерационный алгоритм циклической свертки Тоома-Кука по основанию 4									
16	<b>0,039</b>	0,108	0,244	0,702					
32	0,057	<b>0,104</b>	0,155	0,167	0,395	0,840	1,887		
64	0,064	0,107	<b>0,150</b>	<b>0,150</b>	<b>0,262</b>	<b>0,391</b>	<b>0,420</b>	0,898	1,786
128	0,100	0,160	0,220	0,215	0,349	0,490	0,491	<b>0,834</b>	1,239
256	0,124	0,194	0,265	0,257	0,404	0,556	0,545	0,870	<b>1,222</b>
512	0,186	0,290	0,393	0,379	0,589	0,803	0,779	1,216	1,673
$K$	128	128	128	256	256	256	512	512	512
$f_q$ (кГц)	8	12	16	8	12	16	8	12	16
$\xi$	0,1	0,01	0,001	0,1	0,01	0,001	0,1	0,01	0,001
$L$	7	11	13	15	21	25	29	39	47

Из таблицы 2 видно, что алгоритм на основе БПФ требует меньше вычислений, чем итерационный алгоритм. В некоторых случаях – для малой длины  $L$  – наивный алгоритм оказывается лучше алгоритма БПФ. Однако, при необходимости вычислений с высокой точностью ( $\xi$ ) или разрешающей способностью ( $K$ ) значение  $L$  будет достаточно большим, и использование фильтрации позволит сократить вычислительную сложность до 2 раз.

В разработанном алгоритме (см. рисунок 3) используются специальные объекты входного потока  $istr$  и выходного потока  $ostr$ , которые инкапсулируют в себе функционал вычисления расширенных входных

векторов по формуле 6 перед фильтрацией и последующего «сужения» после фильтрации.

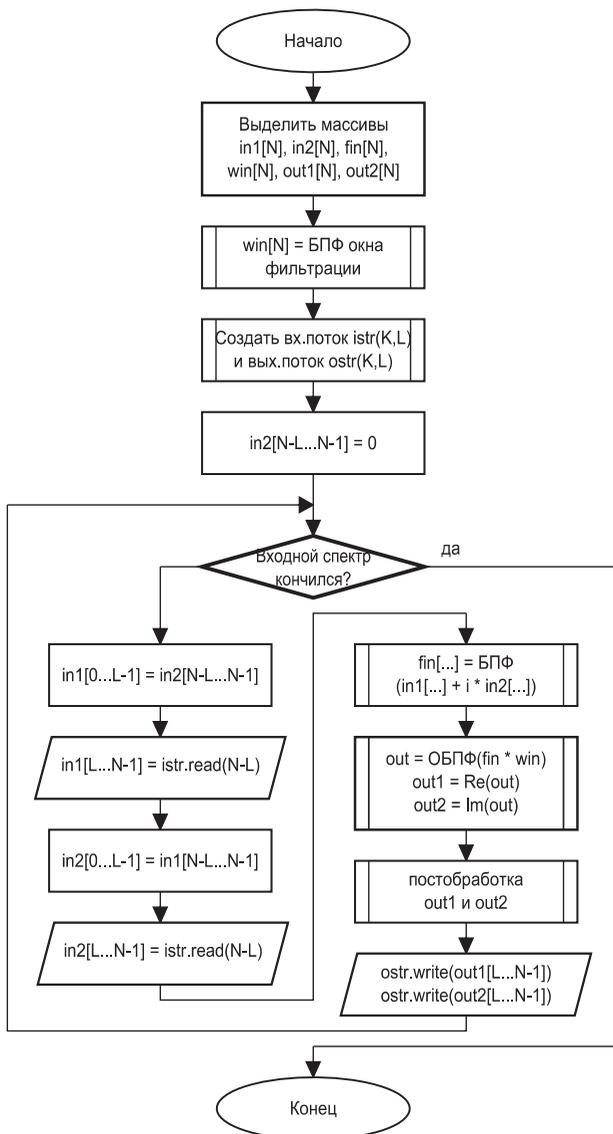


Рис. 3. Алгоритм быстрого вычисления одновременной маскировки на основе БПФ

Использование таких потоков позволяет сделать независимыми границы входных блоков, соответствующих временным отсчетам, (размера  $K$ ) и блоков для фильтрации (размера  $N$ ). В алгоритме за одну итерацию вычисляется два выходных блока, т.е.  $2(N - L)$  чисел. При этом, поскольку используется метод пересечения с накоплением, входные блоки пересекаются – поэтому используется копирование из одной части входного блока в другую. Постобработка выходных векторов (предпоследний шаг в итерации) заключается в выполнении операции сравнения (оператор  $>$ ) из формулы (7).

**3. Заключение.** Как следует из таблицы 2, предложенный метод позволяет вычислять одновременную маскировку быстрее, чем оригинальный (наивный) алгоритм при определенных условиях. Эти условия заключаются в выполнении ограничения (15) на используемую шкалу частот и достаточно больших значениях  $K$  и  $\xi$  (см. таблицу 2). Требуемая вычислительная мощность при этом снижается до 2 раз.

Значения вычислительной сложности как для оригинального (до 0,77 GFLOPS), так и для предлагаемого (до 0,36 GFLOPS) алгоритмов позволяют вычислять одновременную маскировку в реальном времени с учетом вычислительной мощности современных цифровых процессоров (от 3 GFLOPS до более чем 100 GFLOPS). Однако вычисление одновременной маскировки, как правило, не является конечной целью в прикладных задачах, а является лишь промежуточной ресурсоемкой задачей. Поэтому снижение сложности вычисления одновременной маскировки позволяет повысить шансы на работу всей системы в реальном времени, а также снизить требования к ее аппаратному обеспечению.

## Литература

1. *Сапожков М.А.* Речевой сигнала в кибернетике и связи // М.: Государственное издательство литературы по вопросам связи и радио. 1963. 452 с.
2. *Бондаренко В.П., Коуцубинский В.П., Мецержаков Р.В.* Нестационарные модели в обработке речевых сигналов // Акустика речи. Медицинская и биологическая акустика. Архитектурная и строительная акустика и вибрации. Сб. трудов XVIII сессии Российского акустического общества. Т.3. М.: ГЕОС, 2006. С. 8-11.
3. *Ronzhin A.L., Budkov V.Yu.* Multimodal Interaction with Intelligent Meeting Room Facilities from Inside and Outside // Springer-Verlag Berlin Heidelberg, S. Balandin et al. (Eds.): NEW2AN/ruSMART 2009, LNCS 5764, 2009, pp. 77–88.
4. *Yusupov R.M., Ronzhin An.L., Prischepa M., Ronzhin Al.L.* Models and Hardware-Software Solutions for Automatic Control of Intelligent Hall // Automation and Remote Control. 2011. vol. 72. no. 7. pp. 1389–1397.
5. *Ronzhin A., Karpov A., Kiryatkova I., Zelezny M.* Client and Speech Detection System for Intelligent Infokiosk // Springer-Verlag Berlin Heidelberg, Petr Sojka et al. (Eds.): TSD 2010, LNAI 6231, 2010. pp. 560–567.
6. *Meshcheryakov R.V., Bondarenko V.P.* Dialogue as a basis for construction of speech systems // Cybernetics and Systems Analysis. 2008. no. 44(2). pp. 175-184.
7. *Meshcheryakov R. V., Balatskaya L. N., Choinzonov E. L., Chizevskaya S. Yu., Kostyuchenko E.U.* Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw // Proceedings of the 15th International Conference, SPECOM 2013, September 1-5, 2013, Pilsen, Czech Republic, 2013. pp 294-301.

8. Bondarenko V.P., Kvasov A.N., Konev A.A., Meshcheryakov R.V., Choinzonov E.L., Chizhevskaya S.Y. Software for speech signal research in patients with malignant diseases of the throat // *Biomedical Engineering*, 2009. no. 43(4). pp. 183-186.
9. Ponizov A.G., Meshcheryakov R.V. Implementation of Tuning Fork Hearing Tests Using Electronic Devices // *Biomedical Engineering*. 2012. no. 46(1). pp. 32-35.
10. Меццержаков Р.В. Структура систем синтеза и распознавания речи // *Известия Томского политехнического университета*. 2009. Т. 315. № 5. С. 121-126.
11. Bondarenko V.P., Ponomarev A.A., Rogozinskaya E.A. Модель одновременной маскировки // *Интеллектуальные системы в управлении, конструировании и образовании*. Томск: STT, 2004. С. 167-174.
12. Huang X., Acero A., Hon H.-W. Spoken Language Processing: a guide to theory, algorithm, and system development // Prentice Hall, 2001. 980 p.
13. Jung S. Hyers-Ulam-Rassias Stability of Functional Equations in Nonlinear Analysis // Springer optimization and its applications. Springer, 2011. URL: <http://books.google.ru/books?id=NcClS-Hm19kC>.
14. Traunmüller H. Analytical expressions for the tonotopic sensory scale // *The Journal of the Acoustical Society of America*. 1990. vol. 88, no. 1. pp. 97-100. URL: <http://scitation.aip.org/content/asa/journal/jasa/88/1/10.1121/1.399849>.

## References

1. Sapozkov M.A. *Rechevoj signala v kibernetike i svyazi* [Speech signal in cybernetics and communication]. M., Gosudarstvennoe izdatelstvo literatury po voprosam svyazi I radio, 1963. 452 p. (In Russ.).
2. Bondarenko V.P., Kotsubinsky V.P., Meshcheryakov R.V. [Nestacionarnie modeli v obrabotke rechevix signalov]. *Acustica rechi. Medicinskaja i biologicheskaja akustika. Arhitekturnaja i stroitel'naja akustika i vibracii. Sbornik trudov XVIII sessii Russain acoustic society* [Acoustics of speech. Medical and biological acoustics. Architectural and building acoustics and vibration. Proceedings of the 17th Session of the Russian Acoustical Society]. V.3, M: GEOS, 2006. pp. 8-11. (In Russ.).
3. Ronzhin A.L., Budkov V.Yu. Multimodal Interaction with Intelligent Meeting Room Facilities from Inside and Outside // Springer-Verlag Berlin Heidelberg, S. Balandin et al. (Eds.): NEW2AN/ruSMART 2009, LNCS 5764, 2009, pp. 77–88.
4. Yusupov R.M., Ronzhin An.L., Prischepa M., Ronzhin A.L.L. Models and Hardware-Software Solutions for Automatic Control of Intelligent Hall. *Automation and Remote Control*. 2011. vol. 72, no. 7. pp. 1389–1397.
5. Ronzhin A., Karpov A., Kipyatkova I., Zelezny M. Client and Speech Detection System for Intelligent Infokiosk. Springer-Verlag Berlin Heidelberg, Petr Sojka et al. (Eds.): TSD 2010, LNAI 6231. 2010. pp. 560–567.
6. Meshcheryakov R.V., Bondarenko V.P. Dialogue as a basis for construction of speech systems. *Cybernetics and Systems Analysis*, 2008. no. 44 (2). pp. 175-184.
7. Meshcheryakov R. V., Balatskaya L. N., Choinzonov E.L., Chizevskaya S.Yu., Kostyuchenko E.U. Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw. Proceedings of the 15th International Conference, SPECOM 2013, September 1-5, 2013, Pilsen, Czech Republic, 2013. pp. 294-301.
8. Bondarenko V.P., Kvasov A.N., Konev A.A., Meshcheryakov R.V., Choinzonov E.L., Chizhevskaya S.Y. Software for speech signal research in patients with malignant diseases of the throat. *Biomedical Engineering*, 2009. no. 43 (4). pp. 183-186.
9. Ponizov A.G., Meshcheryakov R.V. Implementation of Tuning Fork Hearing Tests Using Electronic Devices. *Biomedical Engineering*, 2012. no. 46 (1) , pp. 32-35.
10. Meshcheryakov R.V. [Structure of speech synthesis and recognition systems]. *Isvestiya Tomskogo politehnicheskogo universiteta – Tidings Tomsk Polytechnic University*. 2009. vol. 315. no. 5. pp 121-126. (In Russ.).
11. Bondarenko V.P., Ponomarev A.A., Rogozinskaya S.I. [Model odnovenmennoy maskirovki]. *Intellektualniye sistemi v upravlenii, konstruirovanii i obrazovanii – Intelligent systems in the management, engineering and education*. Tomsk, STT. 2004. pp. 167-174. (In Russ.).

12. Huang X., Acero A., Hon H.-W. Spoken Language Processing: a guide to theory, algorithm, and system development. Prentice Hall, 2001. 980 p.
13. Jung S. Hyers-Ulam-Rassias Stability of Functional Equations in Nonlinear Analysis. Springer optimization and its applications. Springer, 2011. Available at: <http://books.google.ru/books?id=NcClS-Hm19kC>.
14. Traummuller H. Analytical expressions for the tonotopic sensory scale. The Journal of the Acoustical Society of America. 1990. vol. 88, no. 1. pp. 97-100. Available at: <http://scitation.aip.org/content/asa/journal/jasa/88/1/10.1121/1.399849>.

**Тиунов Сергей Дмитриевич** — аспирант кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: обработка речевого сигнала, математическое моделирование. Число научных публикаций — 17. [t5d@mail.ru](mailto:t5d@mail.ru); г. Томск, пр. Ленина, 40, ауд. 210, тел. +7 (913) 823 32 56.

**Tiunov Sergey Dmitrievich** — post-graduate student, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech signal processing, mathematical modeling. The number of publications — 17. [t5d@mail.ru](mailto:t5d@mail.ru); KIBEVS Dept. TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; phone +7 (913) 832 32 56.

**Мещеряков Роман Валерьевич** — д-р техн. наук, профессор кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: системный анализ, анализ и синтез речи информационная безопасность, вопросы обработки информации в интеллектуальных системах, особое внимание уделяется вопросам создания информационно-безопасных систем. Число научных публикаций — 247. [mrv@security.tomsk.ru](mailto:mrv@security.tomsk.ru); 634050, г. Томск, пр. Ленина, 40, ауд. 210; п.т. +7 (3822) 900-111, факс +7 (3822) 900-111.

**Meshcheryakov Roman Valerievich** — Ph.D., Dr. Sci., professor, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech analysis, speech recognition, medical technology, information security. The number of publications — 247, IEEE Senior Member. [mrv@security.tomsk.ru](mailto:mrv@security.tomsk.ru); KIBEVS Dept. TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; office phone +7(3822)413-426, fax +7(3822)900-111.

**Черных Дмитрий Владимирович** — аспирант кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: обработка речевого сигнала, математическое моделирование. Число научных публикаций — 9. [dmit-cher@mail.ru](mailto:dmit-cher@mail.ru); г. Томск, пр. Ленина, 40, ауд. 210, тел. +7 (952) 805 97 59.

**Chernyh Dmitriy Vladimirovich** — post-graduate student, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech signal processing, mathematical modeling. The number of publications — 9. [dmit-cher@mail.ru](mailto:dmit-cher@mail.ru); KIBEVS Dept. TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; phone +7 (952) 805 97 59.

**Поддержка исследований.** Работа выполнена при финансовой поддержке РФФИ (проект № 14-07-00449 -а), государственного задания Министерства образования и науки РФ, проект № 1220 «Фундаментальные основы проектирования информационно-безопасных систем».

**Acknowledgements.** This research is supported by RFBR (grant 14-07-00449 -а) and the state contract of the Ministry of Education and Science №1220 “Fundamentals of information-security systems designing”.

## РЕФЕРАТ

### *Тиунов С.Д., Мещеряков Р.В., Черных Д.В.* **Оптимизация вычисления одновременной маскировки речевого сигнала.**

При обработке речи проводятся различные операции над звуковым сигналом и, как правило, это требует значительных вычислительных ресурсов и не может быть проведено в реальном масштабе времени. Это связано с одной стороны с тем, что при разработке новых алгоритмов для обработки речи на первом месте стоит вопрос точности, а не вопрос вычислительной эффективности. Вместе с тем некоторые приложения речевых технологий подразумевают обработку сигнала в реальном времени. Предлагается эффективный численный метод для вычисления одновременной маскировки из модели слуховой системы, что позволяет применять ее в программном комплексе для анализа речевого сигнала в норме и при патологии.

Одновременная маскировка – это особенность восприятия звука слуховой системой человека. Эта особенность заключается в том, что из двух близких частотных составляющих ухом будет воспринята только та, которая имеет большую интенсивность. Моделирование эффекта одновременной маскировки позволяет выделить в спектрограмме сигнала наиболее важные (информативные) составляющие с точки зрения слуховой системы человека.

В работе вводится ограничение на используемую шкалу частот, что позволяет свести вычисление одновременной маскировки к вычислению линейной свертки. Поскольку существуют эффективные методы вычисления линейной свертки, то введенное ограничение приводит к снижению вычислительной сложности алгоритма.

Предложенный метод позволяет вычислять одновременную маскировку быстрее, чем оригинальный (наивный) алгоритм при определенных условиях. Эти условия заключаются в выполнении ограничения на используемую шкалу частот и определенные (высокие) значения разрешающей способности. Требуемая вычислительная мощность при этом снижается до 2 раз.

Значения вычислительной сложности как для оригинального (до 0,77 GFLOPS), так и для предлагаемого (до 0,36 GFLOPS) алгоритмов позволяют вычислять одновременную маскировку в реальном времени с учетом вычислительной мощности современных цифровых процессоров (от 3 GFLOPS до более чем 100 GFLOPS). Однако вычисление одновременной маскировки, как правило, не является конечной целью в прикладных задачах, а является лишь промежуточной ресурсоемкой задачей. Поэтому снижение сложности вычисления одновременной маскировки позволяет повысить шансы на работу всей системы в реальном времени, а также снизить требования к ее аппаратному обеспечению.

## SUMMARY

### *Tiunov S.D., Meshcheryakov R.V., Chernykh D.V.* **Optimization of calculation of frequency masking of speech signal.**

Speech processing is associated with various operations on sound signal and typically requires significant computational resources and may not be conducted in real time. On the one hand this is due to the fact that the development of new algorithms for speech processing in the first place is a question of accuracy, not a question of computational efficiency. However, some applications involve speech signal processing techniques in a real time. An efficient numerical method for calculating the frequency masking of the auditory system model is proposed, which allows using it in a software package for the analysis of speech signal in normal and pathological conditions.

Frequency masking is a feature of the perception of sound in human auditory system. It states that of the two neighboring frequency components only one that has a greater intensity will be heard. Simulation of frequency masking allows selecting the most important (informative ones) components from the spectrogram of the signal from the perspective of human auditory system.

The constraint on the used frequency scale is proposed and considered, that allows calculating frequency masking by calculating linear convolution. Since the efficient methods of convolution calculation are known, the proposed constraint leads to reduce the computational complexity of the algorithm.

The proposed method allows calculating the frequency masking faster than the original (naive) algorithm under certain conditions. These conditions include the constraint on the used frequency scale and certain (high) resolution parameters. The computational power required in this case is reduced up to 2 times.

Computational complexity for the original (up to 0,77 GFLOPS), and for the proposed (up to 0,36 GFLOPS) algorithms allows us to calculate the frequency masking in real time based on computational power of modern CPUs (3 GFLOPS to more than 100 GFLOPS). However, the calculation of frequency masking is usually not an end goal in applications, but only an intermediate resource-consuming task. Therefore, reducing the complexity of calculating the frequency masking can increase the chances of the entire system real time work, as well as reduce its hardware requirements.

В.В. ЕВДОКИМОВА  
**АНАЛИЗ СПЕКТРА ГЛАСНЫХ НА ОСНОВЕ  
НЕРАВНОМЕРНОЙ ПСИХОАКУСТИЧЕСКОЙ ШКАЛЫ ЭРБОВ  
ДЛЯ ОПРЕДЕЛЕНИЯ СЛОВЕСНОГО УДАРЕНИЯ**

---

*Евдокимова В.В.* **Анализ спектра гласных на основе неравномерной психоакустической шкалы эрбов для определения словесного ударения.**

Аннотация. Рассматривается возможность определения в слове ударного слога путем изучения особенностей формантной картины. На основе анализа массива логатомов выявляются особенности формантной картины, причем сравниваются энергетические характеристики отдельных формант. Опираясь на допущение об оптимальности способа формирования диктором речевой посылки, спектр гласных разбивается по психоакустической шкале эрбов. Для удобства обработки материала полученный массив частичных дисперсий кодифицируется. Для выявления признака ударности полученный массив кодов разбивается по признаку гласной и по признаку ударности. Сравнение частичных массивов дает возможность подтвердить существование исследуемого признака ударности и выявить фонетические особенности этого явления, существующие в формантной картине. Сделан вывод о возможности применения признака ударности по особенностям формантной картины гласной как эффективно дополняющего определение ударности слога в слове по иным общепринятым признакам.

**Ключевые слова:** фонетика, акустический анализ речевого сигнала, формантная картина гласного, шкала эрбов, ударность гласного.

*Evdokimova V.V.* **Vowel spectral analysis using the psychoacoustic ERB scale for the word stress detection.**

**Abstract.** The paper presents a possibility to determine the stressed syllable in the word by examining features of the vowel formant structure. Based on the analysis of the array of logatoms the vowel formant structure is identified. The energy characteristics of individual formants of different vowels in one word are compared. Spectrum is split by the vowels psychoacoustic ERB scale. For ease of handling of the material resulting array of partial dispersions is codified. To determine the stressed vowel the resulting array codes are broken on the basis of the stressed/unstressed vowel characteristic and phoneme type. Partial arrays comparison enables to confirm the existence of the stress feature characteristic and identify phonetic features of this phenomenon, existing in the vowel formant structure. The vowel stress characteristic is defined by the vowel formant pattern and can effectively complement the determination of stressed syllables being added to conventional stressed vowel characteristics.

**Keywords:** phonetics, speech acoustic analysis, vowel formant structure, ERB scale, vowel stress.

---

**1. Введение.** В современных системах автоматической обработки и анализа речевого материала существует проблема определения границ слов в потоке речи. Одним из приемов решения этой проблемы является определение места словесного ударения, которое «объединяет звуки, образующие облик слова, а если слово состоит более чем из одного слога, то ударный слог связывает его в единое целое» [5].

Как отмечает Ю.С. Маслов, «словесное ударение заключается в том, что в слове с помощью тех или иных звуковых средств подчёркивается один вполне определённый слог» [7].

Известно, что информативность словесного ударения обеспечивается всей совокупностью речевых явлений [1, 3]. Ударный слог выделяется интенсивностью речевого сигнала, его длительностью, вариациями частоты основного тона и формантной картины, причем способ такого комплексного выделения ударения диктором всегда оптимизирован по критерию его восприятия слушателем.

Основным коррелятом словесного ударения в русском языке является длительность, что было доказано в диссертации Л.В. Златоустовой [6]. Ударный слог обычно является самым длинным в слове [1]. Этот признак достаточно устойчив и выделяется многими исследователями [4, 7, 9], но в ряде случаев длительность первого предупредительного слога может оказаться сравнимой с ним и даже превышать его длительность. Дополнительную неточность вносит и удлинение последнего слога в синтагме [2].

Другим общепринятым признаком ударения является повышенная интенсивность ударного слога, которая проявляется как в увеличении его акустической мощности, так и в повышении частоты основного тона из-за большей напряженности речевого тракта [3]. Однако признак интенсивности также не может считаться универсальным. Замеряемая акустическая мощность в значительной степени зависит от типа гласной в ударном слоге. Так, акустическая мощность ударного слога с закрытой гласной может оказаться меньше, чем у соседнего предупредительного либо первого заударного слога с открытой гласной. В свою очередь, частота основного тона существенно зависит от интонационного рисунка речи, который тоже вносит искажение в определение ударного слога.

Когда появилась возможность создавать системы автоматического синтеза речи, было выявлено, что акустически предупредительные, ударные и заударные аллофоны имеют разные спектральные картины [1, 3, 8], т. е. отмечено, что ударение формируется не только действием голосового источника, реализующего отмеченные выше признаки, но и деятельностью артикуляторного аппарата, трансформирующего формантную картину гласной [1]. Ряд исследователей обратил внимание на особенности спектра речевого сигнала: в ударных слогах повышена интенсивность низкочастотных первой и второй формант [11, 12], что соответствует участку высокого уровня слухового восприятия и способствует лучшему информационному контакту со слушателем.

Данная статья посвящена исследованию процесса формирования формантной картины гласных с целью оценки возможности использования особенностей трансформации этой картины в качестве признака ударности слога.

**2. Оценка влияния вариативности формантной картины гласного на словесное ударение.** Предметом исследования стал записанный одним диктором в условиях лаборатории экспериментальной фонетики Санкт-Петербургского государственного университета массив из пяти тысяч трехсложных логатомов [14].

Чтобы исключить влияние интонационных особенностей построения фразы, в данной работе было решено исследовать характеристики ударных и безударных гласных в изолированно произнесённых словах - логатомах. Каждый логатом представлял собой повторение трех слогов типа CVCVCV, где при одном и том же наборе в каждом слоге гласных и согласных ударение ставилось попеременно на одну из трех гласных в слове.

Логатомы составлялись таким образом, чтобы можно было анализировать гласные во всех акустически значимых фонетических контекстах [8]. Для этого массивы логатомов были записаны для гласных /a/, /i/, /u/, /e/, /o/, /ɨ/. Для каждой согласной записаны логатомы с различным местом ударного слога в логатоме. Общее количество аллофонов гласных фонем в материале составляет 5263 единицы. Общее количество слов в материале составило 1748 единиц. В дальнейшем они были сгруппированы как по гласным, так и по месту ударного слога в логатоме.

Далее производилась сегментация полученных файлов и их обработка с помощью стандартного программного обеспечения Wave Assistant. Спектральные характеристики гласных вычислялись при помощи программы «Фонетическая лаборатория v.3». Результатом обработки явились спектрограммы гласных, отражающие основные особенности их формантных картин [1]. Площади пиков на спектрограмме соответствовали энергии формант. Спектрограммы были построены в шкале эрбов [13], учитывающей психофизиологические особенности речи диктора. Выбор именно этой шкалы обусловлен тем, что она дает оптимальный с точки зрения дальнейшей обработки уровень градации спектрограммы на участке трансформации первой форманты (280...650 Гц), который соответствует седьмому-двенадцатому эрбам (284...643 Гц). Предварительно этот диапазон трансформации первой форманты был уточнен для данного диктора по полученным спектрограммам. Альтернативные шкалы барков и мелов также могли быть использованы, но они не имеют на этом участке естественной градации удобной для дальнейшей обработки.

При обработке спектральных плотностей гласных нормированные частичные дисперсии определялись для каждого эрба по формуле:

$$D_i = \frac{1}{D} \int_{f_{1i}}^{f_{2i}} S(f)df, \quad (1)$$

где  $i$  – номер эрба ( $i = 1...30$ ),  $D$  – дисперсия сигнала, рассчитанная по спектрограмме,  $f_{1i}$  и  $f_{2i}$  – частотные границы  $i$ -го эрба,  $S(f)$  – спектральная плотность (спектрограмма) гласной, полученная при обработке.

Таким образом, для каждой гласной была получена последовательность из тридцати частичных дисперсий, характеризующая распределение энергии по шкале эрбов. После этого проводилось сопоставление последовательностей, полученных для одного ударного и двух безударных слогов одного и того же логотома. Пример такого сопоставления приведен в таблице 1.

Таблица 1. Интенсивность спектра по эрбам

Вид логатома	Номер эрба								
	1	...	7	8	9	10	11	...	30
бЕбебе	0,191	...	0,026	0,082	<b>0,063</b>	0,011	0,155	...	<b>0,012</b>
бЕбебе	0,082	...	0,048	0,098	0,049	0,006	<b>0,362</b>	...	0,0075
бЕбебе	<b>0,312</b>	...	<b>0,160</b>	<b>0,195</b>	0,055	<b>0,041</b>	0,189	...	0,0012
бебЕбе	0,112	...	0,040	<b>0,155</b>	<b>0,091</b>	0,0119	0,184	...	<b>0,017</b>
бебЕбе	0,038	...	0,039	0,077	0,042	0,009	<b>0,302</b>	...	0,001
бебЕбе	<b>0,225</b>	...	<b>0,135</b>	0,121	0,038	<b>0,033</b>	0,127	...	0,001
беббЕ	0,011	...	0,128	0,031	0,031	0,076	0,258	...	0,002
беббЕ	0,056	...	0,134	<b>0,154</b>	<b>0,041</b>	0,065	<b>0,272</b>	...	0,001
беббЕ	<b>0,105</b>	...	<b>0,184</b>	0,067	0,017	<b>0,163</b>	0,056	...	<b>0,015</b>

Каждый опыт (каждый логатом) представлен тремя строками таблицы, в которых приведены последовательности распределения энергии, рассчитанные для первого, второго и третьего его слогов соответственно. По каждому эрбу проведено сравнение относительных дисперсий между тремя слогами данного логатома (в пределах столбца из трех строк логатома). Наиболее интенсивный слог выделен на каждом эрбе жирным шрифтом).

Таблица дает возможность сопоставить распределение интенсивности сигнала для различных слогов одного и того же логатома по каждому участку спектра, соответствующему одному эрбу. Появляется возможность перехода с ее помощью к вторичным показателям оценки относительной интенсивности в различных слогах

одного логатома. Для этого после сравнения по каждому эрбу относительных частичных дисперсий трех слогов трехстрочного столбца, выбирается слог с максимальной дисперсией и полученному таким путем событию присваивается код 1, 2 или 3 по номеру строки в этом трехзначном столбце. Таким образом, каждый опыт рассматривается как сложное событие, описываемое тридцатиразрядным троичным кодом [10]. Пример преобразования таблицы дискретизации приведен в таблице 2.

Таблица 2. Интенсивность спектра по эрбам

Вид логатома	Код дискретизации по эрбам
бЕбебе	3...33132...1
бебЕбе	3...31132...1
бебебЕ	3...32232...3

Собранный информационный массив велик и структурно избыточен для решения задачи оценки связи словесного ударения со структурой формантной картины. Исходя из поставленной задачи исследования и из известных особенностей физиологии работы речевого тракта, выберем структуру аппарата анализа.

Низкочастотный участок спектра речевого сигнала описывает работу голосового источника. Считаем, что структура первых шести эрбов (частный диапазон 30...284 Гц) в основном определяется вариациями частоты основного тона [1, 2]. Их исследование выходит за пределы данной работы, направленной на изучение трансформации формантной картины, поэтому рассмотрение начинаем с седьмого эрба, поставив задачу поиска особенностей формирования кодов – вторичных признаков трансформации формантной картины.

В соответствии с поставленной задачей сосредоточим внимание на участке спектра с седьмого до двенадцатого эрба (частотный диапазон 284...643 Гц). Проведенные пробные расчеты по большему числу эрбов показали, что увеличение их числа не только не помогает выявлению признака ударности, но наоборот нивелирует результаты. Причина этого в том, что начиная с диапазона второй форманты существенно сказываются различия формантных картин разных гласных.

Каждый из опытов рассматривается, таким образом, как событие, описываемое шестиразрядным кодом (см. таблицу 2) по числу эрбов на отобранном участке спектра (7-й...12-й эрбы).

Полученный код представляет собой формализованное отражение фонетических процессов, соответствующих определенному в конкретном опыте логатому и дает возможность оценки по вторичным признакам картины трансформации первой форманты в ударных слогах.

Разделим полученный по результатам обработки эксперимента общий числовой массив на три по признаку первого, второго и третьего ударных слогов, затем каждый из них – на шесть массивов по числу исследуемых гласных фонем русского языка /a/, /i/, /u/, /e/, /o/, /i/.

Сгруппируем массивы разных фонем по признаку ударного слога и сопоставим их. Оценим степень их перекрытия, т.е. наличие одинаковых кодов, что в фонетическом плане соответствует одинаковому функционированию артикуляторного аппарата. В таблицах 3-5 приведены оценки степени взаимоперекрывания этих массивов, а именно присутствия в массиве **у** кодов гласной /y/, кодов гласной /x/. Показатели приведены в процентной форме в соответствии с формулой:

$$y_{xy} = \frac{N_{xy}}{N_x} \times 100, \quad (2)$$

где  $y_{xy}$  - степень присутствия кодов массива **У** в массиве **X**,  $N_x$  - общее число опытов в массиве **X**,  $N_{xy}$  - число опытов массива **У**, имеющих коды, присутствующие в массиве **X**.

В таблицах 3-5 введено обозначение  $m_j$ , где  $m$  – одна из шести гласных,  $j$  – номер слога, на который падает ударение в логатоме. Как уже указано, эти таблицы отражают определенную по формуле (2) степень перекрытия пространства кодов по каждой гласной, т.е. случаев, когда формантная картина одной гласной неотличима от основного массива формантных картин другой гласной.

Таблица 3. Ударение выпадает на первый слог логатома

		Массив Y					
Массив X	%	e1	i1	a1	u1	o1	i1
	e1	X	19	16	21	15	10
	i1	47	X	8	14	8	12
	a1	36	14	X	13	37	5
	u1	30	27	14	X	16	25
	o1	23	11	26	11	X	10
	i1	40	11	15	54	30	X

Таблица 4. Ударение выпадает на второй слог логатома

		Массив Y					
Массив X	%	e2	i2	a2	u2	o2	i2
	e2	X	13	58	46	52	21
	i2	28	X	14	11	10	6
	a2	45	17	X	20	63	11
	u2	52	3	43	X	53	22
	o2	33	2	34	29	X	20
	i2	34	8	24	41	40	X

Оценим попарно по таблицам 3-5 степень взаимопроникновения исследуемых массивов кодов гласных. Качественный результат сравнения сведем в таблицу 6.

Количественно уровень взаимопроникновения максимален при втором ударном слоге, минимален при третьем.

Таблица 5. Ударение выпадает на третий слог логатома

		Массив Y						
		%	e3	i3	a3	u3	o3	i3
Массив X	e3	X	16	52	42	49	18	
	i3	24	X	13	12	9	7	
	a3	32	17	X	20	63	11	
	u3	48	5	42	X	51	23	
	o3	32	3	37	26	X	18	
	i3	36	9	22	41	42	X	

Таблица 6. Степень взаимопроникновения массивов кодов гласных

Ударный слог логатома	Массивы кодов гласных	
	Максимальное взаимопроникновение	Минимальное взаимопроникновение
Первый слог	/u/ и /i/, /o/ и /a/, /i/ и /e/	/o/ и /i/, /i/ и /a/, /i/ и /i/
Второй слог	/a/ и /e/, /a/ и /o/, /e/ и /u/	/o/ и /i/, /u/ и /i/, /i/ и /i/
Третий слог	/u/ и /i/, /o/ и /a/, /i/ и /u/	/o/ и /i/, /a/ и /i/

Результаты сравнения хорошо согласуются с известными положениями фонетики о близости артикуляции определенных пар гласных (а – е, е – і, о – u) [1], что подтверждает допустимость выбранного инструмента исследования.

Проведем сравнение массивов различных ударных слогов, так как именно перекрестная связь между этими массивами представляет собой погрешность при использовании алгоритма определения в логатоме ударного слога по кодам. Для этого объединим массивы гласных по признаку общего ударного слога (первый гласный в слове ударный, второй или третий) и сравним эти массивы. Допустимость такого обобщения обусловлена заложенным в постановку задачи ограничением частотного диапазона, которое исключает разброс кодов разных гласных от более высокочастотных формант.

По результатам сравнения можно отметить, что в каждом из трех массивов выделяются три основные группы кодов. Процентное отношение этих групп к общему числу кодов в данном массиве приведено в таблице 7. В каждом из них существует группа кодов, соответствующая исключительно этому массиву. Другой важной группой являются коды, многократно повторяющиеся во всех трех массивах. Фонетически они соответствуют наиболее привычным диктору режимам работы речевого тракта. Перечислим их: 311211, 311311, 311312, 311313, 311321, 313312, 313313, 321311, 321321, 331211, 331311, 331321, 332211, 332312, 332312, 332322.

Таблица 7. Перекрестное присутствие кодов

Вид кодов	Номер ударного слога		
	Первый слог (%)	Второй слог (%)	Третий слог (%)
Коды, неповторяющиеся в других массивах	43	33	53
Коды, однократно повторяющиеся в других массивах	7	8	6
Коды, многократно повторяющиеся во всех массивах	52	59	41

Кроме этих двух групп существует промежуточная группа, которая соответствует однократному появлению каком-то массиве кода, присутствующего также и в другом массиве.

**3. Заключение.** Проведенное исследование позволяет сделать следующие выводы. Использованный в работе метод кодирования формантных картин гласных в исследуемых логатомах дает возможность перевода набранного речевого материала в форму, удобную для проведения его анализа.

Поставленный эксперимент подтверждает, что ударному слогу в слове присуща общая для всех гласных характерная трансформация формантной картины. Примерно в 30-50% опытов в ударном слоге формируется комбинация кодов, которая позволяет выделить ударный слог в слове. В остальных случаях формируются коды, соответствующие привычным для диктора режимам работы речевого тракта, которые присутствуют как в ударных, так и в безударных слогах.

Анализ результатов подтверждает, что характерная трансформация формантной картины присутствует в ударном слоге, но этот признак недостаточно устойчив и может быть использован при определении ударения только как дополнительный к описанным выше.

## Литература

1. *Бондарко Л.В.* Фонетика современного русского языка. СПб. 1998. 275 с.
2. *Вольская Н.Б., Степанова С.Б.* Предпаузальное удлинение в русском языке // Экспериментально-фонетический анализ речи: проблемы и методы. Фонетическое многообразие языков мира. СПб. 2004. № 5. С. 48-55
3. *Кодзасов С.В., Кривнова О.Ф.* Общая фонетика. М.: РГГУ, 2001, 592 с.
4. *Кривнова О.Ф.* Реализация словесного ударения в связанном тексте // Фонетические чтения в честь 100-летия со дня рождения Л.Р. Зиндера, Санкт-Петербург, СПбГУ, 2004. С. 150-154.
5. *Зиндер Л.Р.* Общая фонетика. СПб. 2007. 576 с.
6. *Златоустова Л.В.* Фонетическая природа русского словесного ударения: дисс. канд. филол. Наук. Л., 1953.
7. *Маслов Ю.С.* Введение в языкознание. М., 1998. 272 с.
8. *Скрелин П.А.* Фонетические аспекты речевых технологий. СПб, 1999. 76 с.
9. *Скрелин П.А.* Формальные методы анализа речи: проблемы интерпретации результатов // Материалы XXXVI международной филологической конференции, Санкт-Петербург, 2007. С. 3-14.
10. Цифровое кодирование URL: [http://fmi.asf.ru/library/book/network2\\_2.html](http://fmi.asf.ru/library/book/network2_2.html) (дата обращения: 26.10.2013).
11. *Sluijter A.M.C., van Heuven V.J.* Spectral balance as an acoustic correlate of linguistic stress // *Proceedings of the Acoustical Society of America*, 1996, vol. 100, pp. 2471-85. Труды SPIIRAS. 2014. Вып. 1 (62). ISSN 2078-9181 (печ.), ISSN 2078-9500 (онлайн).

12. *Sluijter A.M.C., van Heuven V.J., Pacilly J.J.A.* Spectral balance as a cue in the perception of linguistic stress // *Journal of the Acoustical Society of America*, 1997. vol. 101, pp. 503-513.
13. *Smith J.O., Abel J.S.* Bark and ERB Bilinear Transforms // *IEEE Trans. on Speech and Audio Proc.*, Nov 1999. vol. 7, no. 6, pp. 697-708.
14. *Welge-Lüssen A., Hauser R., Erdmann J., Schwob Ch., Probst R.* Sprachaudiometrie mit Logatomen // *Laryngo-Rhino-Otologie*, 2008. vol. 76. №2. p. 57–64.

## References

1. Bondarko L.V. *Fonetika sovremennogo russkogo jazyka* [The Phonetics of Contemporary Standard Russian]. SPb. 1998, 275 p. (In Russ.).
2. Volskaya N.B., Stepanova S.B. [Prepouse extension in Russian]. *Eksperimentalno-foneticheski analiz: problem i metody – Experimental-phonetics apeech analysis: problems and methods*. Spb. 2004. no. 5. pp. 48-55 (in Russ.).
3. Kodzasov S.V., Krivnova O.F. *Obshhaja fonetika* [General Phonetics]. M., 2001. 592 p. (In Russ.).
4. Krivnova O.F. [Realization of word stress in continuous speech]. *Foneticheskije chtenija v chest' 100-letije so dnja rozhdenija L.R.Zindera*. [Phonetic readings in honour of 100-th anniversary of Zinder L.R.] Spb. SPbSU. 2004. pp. 150-154. (in Russ.).
5. Zinder L.R. *Obshhaja fonetika* [General phonetics]. SPb. 2007. 576 p. (in Russ.).
6. Zlatoustova L.V. *Foneticheskaja priroda russkogo slovesnogo udarenija* [Phonetic nature of Russian word stress] PhD thesis in Philological sciences. Leningrad. 1953. (in Russ.).
7. Maslov Y.S. *Vvedenie v jazykoznanie* [Introduction to linguistics]. M. 1998. 272 p. (in Russ.).
8. Skrelin P.A. *Foneticheskie aspekty rechevyh tehnologij* [Phonetic aspects of speech technologies]. SPb. 1999. 76 p. (in Russ.).
9. Skrelin P.A. [Formal methods of speech analysis: problems of the result interpretation]. *Materialy XXXVI mezhdunarodnoj filologicheskoi konferencii* [XXXVI international philological conference]. SPb. 2007. pp. 3–14. (In Russ.).
10. Cifrovoe kodirovanie [Digital encoding]. Available at: [http://fmi.asf.ru/library/book/network/2\\_2.html](http://fmi.asf.ru/library/book/network/2_2.html) (accessed 26.10.2013). (in Russ.).
11. *Sluijter A.M.C., van Heuven V.J.* Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 1996. vol. 100, pp. 2471–85.
12. *Sluijter A.M.C., van Heuven V.J., Pacilly J.J.A.* Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America*, 1997. vol. 101, pp.503-513.
13. *Smith J.O., Abel J.S.* Bark and ERB Bilinear Transforms. *IEEE Trans. on Speech and Audio Proc.*, Nov 1999. vol. 7, no. 6, pp. 697-708.
14. *Welge-Lüssen A., Hauser R., Erdmann J., Schwob Ch., Probst R.* Sprachaudiometrie mit Logatomen. *Laryngo-Rhino-Otologie*, 2008. vol. 76. no. 2, pp. 57–64.

**Евдокимова Вера Вячеславовна** — к-т филол. наук., ст.преподаватель кафедры фонетики и методики преподавания иностранных языков филологического факультета СПбГУ. Область научных интересов: Речевые технологии, фонетика, языкознание, фонетический анализ речевого сигнала. Число научных публикаций — 30. [postmaster@phonetics.spbu.ru](mailto:postmaster@phonetics.spbu.ru), [www.phonetics.spbu.ru](http://www.phonetics.spbu.ru); СПбГУ, Университетская наб., д. 11, г. Санкт-Петербург, 199034, РФ; p.t. +7(812)328-9565.

**Evdokimova Vera Viatcheslavovna** — Ph.D., senior teacher, Department of Phonetics and Foreign Languages Teaching Methodology, Philological Faculty, SPbSU. Research interests: speech technology, phonetics, linguistics, phonetic analysis of speech. The number of publications — 30. [postmaster@phonetics.spbu.ru](mailto:postmaster@phonetics.spbu.ru), [www.phonetics.spbu.ru](http://www.phonetics.spbu.ru); SPbSU, Universitetskaya emb., 11, St. Petersburg, 199034, Russia; office phone +7(812)328-9565.

## РЕФЕРАТ

### ***Евдокимова В.В. Анализ спектра гласных на основе неравномерной психоакустической шкалы эрбов для определения словесного ударения.***

Рассматривается возможность определения в слове ударного слога путем анализа особенностей формантной картины. Приводится описание известных признаков ударности и работ авторов, обративших внимание на особенности формирования формант при ударении.

Для проведения исследования используется специально сформированный массив трехсложных логатомов, в которых каждый из трех слогов состоит из определенных гласной и согласной, а ударение поочередно делается на первый, второй и третий слог. Для каждого слога, как ударного, так и безударного строится спектр гласной, причем делается это в шкале эрбов, чтобы учесть присущее речи стремление диктора оптимизировать речевую посылку для лучшего ее восприятия слушателем. Полученный спектр разбивается по эрбам и определяются соответствующие каждому участку частичные дисперсии. Для удобства работы с материалом проводится сравнение частичных дисперсий по каждому эрбу между слогами логатома и последующее кодифицирование этих дисперсий по величине. Полученный массив кодов позволяет при соответствующей группировке по гласным и по ударности слога провести сравнение и выявить особенности, присущие ударному слогу.

Делается вывод о том, что признак ударности по особенностям формантной картины существует и может быть использован для дополнения действия общепринятых признаков.

## SUMMARY

### ***Evdokimova V.V. Vowel spectral analysis using the psychoacoustic ERB scale for the word stress detection.***

The paper presents a possibility to determine the stressed syllable in the word by examining features of the vowel formant structure. Based on the analysis of the array of logatoms the vowel formant structure is identified. The energy characteristics of individual formants of different vowels in one word are compared. Spectrum is split by the vowels psychoacoustic ERB scale. For ease of handling of the material resulting array of partial dispersions is codified. To determine the stressed vowel the resulting array codes are broken on the basis of the stressed/unstressed vowel characteristic and phoneme type. Partial arrays comparison enables to confirm the existence of the stress feature characteristic and identify phonetic features of this phenomenon, existing in the vowel formant structure. The vowel stress characteristic is defined by the vowel formant pattern and can effectively complement the determination of stressed syllables being added to conventional stressed vowel characteristics.

Т.В. КАЧКОВСКАЯ  
**ИСПОЛЬЗОВАНИЕ ТЕМПОРАЛЬНЫХ ХАРАКТЕРИСТИК ДЛЯ  
СЕГМЕНТАЦИИ РЕЧЕВОГО ПОТОКА НА КРУПНЫЕ  
СМЫСЛОВЫЕ ЕДИНИЦЫ  
(НА МАТЕРИАЛЕ РУССКОГО ЯЗЫКА)**

---

*Качковская Т.В.* **Использование темпоральных характеристик для сегментации речевого потока на крупные смысловые единицы (на материале русского языка).**

**Аннотация.** Настоящая работа посвящена исследованию темпоральных характеристик гласных в словах, расположенных на концах крупных смысловых единиц — синтагм и фраз. Исследование выполняется на материале корпуса CORPRES, содержащего записи чтения текстов профессиональными дикторами; общее время звучания проанализированного материала составляет около 12 часов. Результаты анализа значений нормализованной длительности гласных показали, что в словах, расположенных на конце синтагмы или фразы, удлиняется ударный гласный, а также заударный, если он находится в абсолютном конце слова; это верно даже в тех случаях, когда фразовое ударение реализуется не на последнем слове синтагмы. На степень удлинения влияет наличие последующей паузы, «глубина» границы, наличие фразового ударения на последнем слове синтагмы и тип интонационного контура. Кроме того, описаны особенности темпорального оформления незавершенности по типу ИК-3 и логического ударения по типу ИК-2 в позиции перед границей синтагмы.

**Ключевые слова:** темпоральный контур, длительность гласных, сегментация речевого потока, предпаузальное удлинение.

*Kachkovskaia T.V.* **Temporal Aspects of Intonational Phrasing (Evidence from Russian).**

**Abstract.** The present paper investigates temporal parameters of vowels occurring before the boundaries of major intonation units—intonational phrases and utterances. The research is based on the analysis of approximately 12 hours of speech from the Corpus of Professionally Read Speech (CORPRES). Our data show that before major prosodic boundaries the stressed vowel of the last word is lengthened, and so does the post-stressed vowel if immediately preceding the boundary; this is true even when the last word does not bear nuclear stress. The degree of lengthening is influenced by the presence of a pause after the boundary, the boundary “depth”, the location of nuclear stress in the unit, and pitch movement type. The temporal organization of two different pitch contours realized before a prosodic boundary is described as well.

**Keywords:** temporal organization of speech, vowel duration, intonation phrasing, phrase-final lengthening.

---

**1. Введение.** Сегментация речевого потока на крупные смысловые единицы — синтагмы и фразы — осуществляется за счет комплексного взаимодействия мелодических, темпоральных (временных), динамических и тембральных характеристик. По данным исследований, проведенных на материале различных языков (см. ниже), временные характеристики занимают высокое место в иерархии маркеров просодических границ.

Данная работа посвящена исследованию длительности гласных, расположенных перед границей синтагмы. Увеличение длительности звуков в этой позиции отмечается на материале многих языков [13] и получило название предпаузального (финального, конечного) удлинения (продления). Как правило, удлинение затрагивает последний слог перед границей; иногда оно также распространяется на последний ударный слог, если он не является конечным. Для русского языка это явление изучено не достаточно: выводы о месте и степени предпаузального удлинения основываются на результатах анализа сравнительно небольшого материала, записанного в лабораторных условиях [15,17]. Таким образом, вопрос о локализации предпаузального удлинения пока остается открытым.

Данное исследование представляет собой попытку установить, как удлинение гласного зависит от его положения относительно ударения и от расстояния до конца синтагмы, а также сделать предварительные выводы о влиянии типа мелодического движения на степень предпаузального удлинения гласных.

**2. Маркеры границ крупных смысловых единиц.** Существует достаточно много работ, посвященных исследованию мелодических, временных (темпоральных), динамических и тембральных характеристик на границах крупных смысловых единиц; при этом рассматриваются следующие акустические ключи (граничные маркеры):

- наличие паузы и ее длина;
- резкое изменение частоты основного тона;
- восстановление линии деклинации;
- предпаузальное удлинение;
- понижение амплитуды;
- наличие ларингализации;
- наличие придыхательной фонации.

Данные о том, какой ключ является наиболее существенным, расходятся. По результатам исследования [3] на материале американского варианта английского языка темпоральные характеристики (паузация и предпаузальное удлинение) имеют большее значение, чем мелодические и динамические характеристики. В работе [8] предпаузальному удлинению приписывается ведущая роль по сравнению с паузацией и мелодикой. Об этом же свидетельствуют данные для французского языка [4], где наиболее важным параметром оказалась длительность последнего гласного (в сравнении с параметрами энергии гласного и частоты основного тона).

Просодические границы часто маркируются сочетанием нескольких акустических ключей. По данным работы [10], выполненной

на материале голландского языка и посвященной взаимодействию паузации, предпаузального удлинения (только для звука /ə/), резкого изменения частоты основного тона и возобновления линии деклинации, единственный граничный маркер, который может встречаться сам по себе, — это резкое изменение частоты основного тона. Остальные маркеры, как правило, встречаются в сочетаниях: если граница маркируется возобновлением линии деклинации, то на месте границы реализуется еще и пауза; если граница маркируется паузой, то в подавляющем большинстве случаев наблюдается также изменение частоты основного тона.

Спектральные изменения звуков, предшествующих просодической границе, исследуются в работе [2]. Помимо длительности, частоты основного тона и амплитуды, авторы рассматривают два акустических ключа, соотносящихся с ларингализацией и придыханием; применимость этих параметров (наряду с длительностью, частотой основного тона и амплитудой) для распознавания границ просодических единиц была доказана на практике. О функции ларингализации как маркера просодических границ говорится также и в [16].

Таким образом, очевидно, что временные характеристики, и предпаузальное удлинение в частности, занимают если не первое, то достаточно высокое место в иерархии граничных маркеров, и, следовательно, должны учитываться как при автоматическом распознавании границ, так и при синтезе речи.

**3. Факторы, влияющие на длительность звуков.** Увеличение длительности сегмента или сегментов может указывать не только на наличие/отсутствие просодической границы. Оно может также:

- маркировать фонемные различия;
- указывать на определенные признаки соседних сегментов (например, более высокая длительность гласных перед звонкими согласными в английском языке);
- участвовать в оформлении противопоставления ударных и безударных гласных;
- быть связанным с длиной слога, слова и синтагмы;
- участвовать в оформлении интонационного контура;
- говорить о наличии/отсутствии эмфатического ударения;
- в случае наличия просодической границы указывать на ее «глубину» (конец фразы/конец синтагмы внутри фразы).

Это значит, что если измеренная длительность какого-либо сегмента будет существенно отличаться от референсного значения (например, среднего по корпусу), мы не сможем легко определить, о чем

говорит такое удлинение. Поэтому при анализе данных о длительности сегмента необходимо учитывать следующие факторы:

- характеристики данного типа сегмента (собственная длительность звука, его растяжимость);
- фонетический контекст;
- положение по отношению к словесному ударению;
- длину слога, слова, синтагмы;
- наличие/отсутствие фразового ударения и в случае наличия — тип интонационного контура;
- наличие/отсутствие эмфатического ударения;
- положение по отношению к границам синтагмы, и в случае близости к концу синтагмы — «глубину» границы и наличие/отсутствие паузы после синтагмы.

**4. Различные подходы к измерению темпоральных характеристик.** В работах, посвященных описанию темпоральной организации речи, чаще всего вычисляются абсолютные значения длительности сегментов (звуков, слогов, слов), а удлинение или сокращение указывается в процентах. Основной недостаток этого метода заключается в том, что вычисленные коэффициенты удлинения (или сокращения) будут отличаться в зависимости от типа звука — это связано с тем, что разные звуки имеют разную собственную длительность (например, звук /a/ имеет большую длительность, чем звук /i/) и разную растяжимость.

Один из возможных вариантов учета этих параметров является вычисление нормализованной длительности по следующей формуле:

$$\bar{d}_i = \frac{d_i - \mu_p}{\sigma_p}, \quad (1)$$

где  $\bar{d}_i$  — нормализованная длительность звука,  $d_i$  — измеренная длительность звука,  $\mu_p$  и  $\sigma_p$  — среднее значение и стандартное отклонение длительности для данного типа звука (вычисленные по всему корпусу для данного диктора). В таком случае собственная длительность звука будет учитываться за счет среднего значения, а растяжимость — за счет стандартного отклонения.

В исследованиях по автоматическому определению синтаксических границ и моделированию темпоральной структуры фразы [5, 9, 11] этот метод используется достаточно широко. Недостатком такого подхода является то, что формула (1) не учитывает изменения

темпа речи. Чтобы получить нормализованную длительность с учетом темпа, формулу (1) можно преобразовать в следующий вид [1]:

$$\bar{d}_i = \frac{d_i - \alpha \mu_p}{\alpha \sigma_p}, \quad (2)$$

добавив в нее коэффициент темпа  $\alpha$ , который вычисляется по формуле:

$$\alpha = \frac{1}{N} \sum_{i=1}^N \frac{d_i}{\mu_{p_i}}, \quad (3)$$

где  $N$  — число звуков в анализируемом отрезке,  $d_i$  — длительность  $i$ -го звука в отрезке,  $\mu_{p_i}$  — средняя длительность соответствующего типа звука.

Такой метод вычисления длительности делает возможным сравнение звуков разных типов (обладающих разной собственной длительностью и растяжимостью), что в свою очередь позволяет увеличить объем выборки.

**5. Материал и метод исследования.** Данное исследование было выполнено на материале русскоязычного корпуса CORPRES [12], содержащего записи чтения текстов различных стилей 8 профессиональными дикторами (4 муж., 4 жен.); при этом некоторые тексты записывались всеми дикторами, а некоторые только двумя (1 муж., 1 жен.) или четырьмя (2 муж., 2 жен.). Общее время звучания составляет около 60 часов. Приблизительно 40 % корпуса содержат многоуровневую разметку, включая уровни идеальной и акустической транскрипции, уровень меток основного тона, уровень орфографических слов и уровень интонационной транскрипции. На уровне интонационной транскрипции указывается тип мелодического движения и тип паузы; на уровне орфографических слов отмечаются слова, несущие фразовое ударение и слова, несущие акцентное выделение. Для проведения данного исследования были выбраны 4 диктора (2 муж., 2 жен.); у трех из них набор текстов совпадает полностью, у четвертого диктора один из текстов отличается.

В ходе обработки файлов аннотации корпуса (с использованием языка Python) для каждого гласного звука в акустической транскрипции были получены значения следующих параметров:

- позиция слова, из которого взят этот гласный, в синтагме (начало/середина/конец);
- позиция синтагмы, из которой взят этот гласный, во фразе (конец/не конец);

- наличие/отсутствие паузы после синтагмы;
- наличие/отсутствие переноса фразового ударения с последнего слова синтагмы;
- наличие/отсутствие акцентно-выделенных слов в синтагме;
- тип мелодического движения в синтагме (по системе, используемой в корпусе [14]);
- длина синтагмы (в слогах и в фонетических словах);
- упрощенная запись фонетического слова (с помощью символов «с» для согласного, «v» для безударного гласного, «V» для ударного гласного; например, для слова *собака* — «сvсVсv»);
- темп речи, вычисленный для всей синтагмы (см. формулу (3));
- нормализованная длительность гласного (см. формулу (2)).

Средние значения длительности и значения стандартного отклонения, необходимые для вычисления коэффициента темпа речи и нормализованной длительности, вычислялись отдельно для ударных, заударных и преударных аллофонов гласных.

Из полученных данных были выбраны значения для гласных, удовлетворяющих следующим условиям:

- синтагма, из которой взят гласный, имеет длину от 3 до 6 фонетических слов и не менее 6 слогов;
- слово, из которого взят гласный, находится в позиции конца синтагмы;
- слово в упрощенной записи оканчивается на «сV», «сVс», «сVсv» или «сVсvс», а гласный является либо ударным, либо заударным (и, следовательно, находится вблизи границы синтагмы).

Так, например, из последнего слова синтагмы *я всегда был уверен*, оканчивающейся на «сVсvс» и имеющей длину 4 фонетических слова (в произнесении диктора М) и 7 слогов, можно выбрать ударный /e/ и заударный /i/.

Все анализируемые синтагмы были разделены на три группы:

- синтагмы, находящиеся в начале/середине фразы, за которыми следует пауза;
- синтагмы, находящиеся в начале/середине фразы, за которыми не следует пауза;
- синтагмы, завершающие фразу (в таких случаях далее практически всегда следует пауза — в силу особенностей корпуса; поэтому случаи, когда пауза отсутствует, не рассматривались).

**6. Результаты исследования.** Сначала были проанализированы значения нормализованной длительности гласных в синтагмах, где фразовое ударение находится *не* на последнем слове — то есть не на слове, из которого взят гласный. Данные о нормализованной длитель-

ности ударных и заударных гласных приведены в таблице 1. В первом столбце указан контекст гласного (рассматриваемый гласный подчеркнут) на конце синтагмы; во втором столбце указано положение синтагмы во фразе (Фр. — конец фразы, С. — начало или середина фразы) и наличие паузы.

Таблица 1. Нормализованная длительность гласных ( $\bar{d}$ ) вблизи конца синтагмы и размер выборки ( $N$ ) в зависимости от контекста для четырех дикторов; гласные взяты из синтагм с фразовым ударением *не* на последнем слове

		Диктор С		Диктор К		Диктор М		Диктор А	
Конец синтагмы	Тип границы	$\bar{d}$	$N$	$\bar{d}$	$N$	$\bar{d}$	$N$	$\bar{d}$	$N$
c <u>V</u>	Фр.+пауза	0,40	16	0,41	27	0,09	26	-0,40	25
c <u>V</u>	С.	0,10	14	-0,09	28	0,55	27	0,89	25
c <u>V</u>	С.+пауза	1,02	16	0,50	24	1,60	11	0,55	33
c <u>V</u> c	Фр.+пауза	0,14	39	0,31	38	-0,11	51	-0,20	24
c <u>V</u> c	С.	-0,44	14	-0,11	23	-0,01	21	0,04	16
c <u>V</u> c	С.+пауза	0,61	19	0,29	22	0,18	26	0,25	23
c <u>V</u> cv	Фр.+пауза	0,28	34	0,72	35	0,28	48	-0,07	27
c <u>V</u> cv	С.	0,38	13	0,34	21	0,15	25	-0,19	7
c <u>V</u> cv	С.+пауза	0,40	21	0,63	22	0,40	23	0,25	29
c <u>V</u> c <u>v</u>	Фр.+пауза	0,42	34	0,04	35	-0,06	48	-0,34	27
c <u>V</u> c <u>v</u>	С.	-0,10	13	0,09	21	0,20	24	0,73	7
c <u>V</u> c <u>v</u>	С.+пауза	1,09	21	0,22	22	0,27	23	0,38	29
c <u>V</u> cvc	Фр.+пауза	0,20	20	0,56	24	-0,15	21	0,23	8
c <u>V</u> cvc	С.	-0,21	5	-0,12	11	-0,11	12	-0,62	2
c <u>V</u> cvc	С.+пауза	0,10	12	0,54	11	0,48	12	0,38	14
c <u>V</u> cyc	Фр.+пауза	-0,35	20	-0,13	24	-0,39	21	-0,53	8
c <u>V</u> cyc	С.	-0,26	5	0,09	11	-0,09	12	-1,05	2
c <u>V</u> cyc	С.+пауза	-0,07	12	-0,33	11	-0,44	12	-0,18	14

Полученные данные позволяют сделать вывод о том, что в словах, расположенных на конце синтагмы или фразы, но не несущих фразовое ударение, удлиняются ударные гласные как в последнем, так и в предпоследнем слоге; исключение составляет контекст «cVcvc» в конце синтагмы без паузы. (Это согласуется с тем, что отсутствие паузы соотносится с меньшей степенью удлинения — см. ниже.) Заударные гласные удлиняются сильнее в позиции абсолютного конца слова (и в данном случае синтагмы); если за заударным гласным следует согласный, то значения нормализованной длительности оказываются отрицательными.

Для ударных гласных, а также безударных гласных в позиции абсолютного конца синтагмы, наличие паузы соотносится с более высокими значениями длительности.

Наиболее высокие значения длительности достигаются в некоторых синтагмах перед паузой (обозначенных как «С.+пауза»). Получается, что *менее* «глубокая» граница — по сравнению с границей, соответствующей концу фразы, — маркируется *сильнее*.

Далее значения нормализованной длительности были вычислены для гласных в синтагмах с фразовым ударением на последнем слове — то есть на слове, из которого взят гласный. Для анализа были выбраны две самые частотные модели: 11 — незавершенность по типу ИК-3 в системе Е. А. Брызгуновой, — и 02 — логическое ударение по типу ИК-2. Модель 11 представляет собой восходящий контур; часто она реализуется в более высоком темпе. Модель 02 описывается как нисходящая; она реализуется с повышением интенсивности и часто — с увеличением длительности (в зависимости от индивидуальных предпочтений говорящего). Таким образом, эти модели противопоставлены как по направлению движения, так и по темпоральным характеристикам.

Данные о нормализованной длительности ударных и заударных гласных приведены в таблице 2, организованной аналогично таблице 1, с указанием типа мелодического движения.

Сравнивая данные для моделей 11 и 02, можно заметить, что для всех дикторов, кроме диктора М, длительность ударных гласных оказывается выше в модели 02. Безударные гласные, наоборот, оказываются длиннее для модели 11 — и это же верно для всех дикторов. Возможно, это говорит об определенной компенсации значений длительности в пределах фонетического слова.

Гласные, находящиеся в абсолютном конце синтагмы, имеют большую длительность, чем гласные, за которыми следует согласный. (То же самое было верно и для синтагм, в которых фразовое ударение находится не на последнем слове — см. выше). Можно предположить, что более сильному удлинению подвергается последний звук, и во втором случае это будет согласный; основания для такого предположения дают данные об удлинении согласных непосредственно перед синтагматической границей [6, 7].

Поскольку модель 11 оформляет незавершенность, она почти никогда не встречается в позиции конца фразы. Этим объясняется отсутствие данных в соответствующих ячейках таблицы (обозначение: nan).

Таблица 2. Нормализованная длительность гласных ( $\bar{a}$ ) вблизи конца синтагмы и размер выборки ( $N$ ) в зависимости от контекста для четырех дикторов; гласные взяты из синтагм с фразовым ударением на последнем слове.

			Диктор С		Диктор К		Диктор М		Диктор А	
Конец синтагмы	Модель	Тип границы	$\bar{a}$	$N$	$\bar{a}$	$N$	$\bar{a}$	$N$	$\bar{a}$	$N$
cV	02	Фр.+пауза	0,94	26	1,04	13	0,90	17	0,98	6
cV	02	С.	0,58	27	0,96	13	0,57	10	1,68	6
cV	02	С.+пауза	2,01	14	1,27	26	1,33	19	1,45	18
cV	11	Фр.+пауза	nan	0	0,77	1	1,94	2	nan	0
cV	11	С.	0,32	34	0,46	50	0,67	69	1,39	24
cV	11	С.+пауза	1,15	46	0,52	44	1,93	47	1,24	53
cVc	02	Фр.+пауза	0,69	19	1,08	27	0,18	26	0,33	7
cVc	02	С.	0,42	12	0,54	11	1,00	7	0,34	15
cVc	02	С.+пауза	0,99	21	1,08	25	0,45	14	0,92	21
cVc	11	Фр.+пауза	nan	0	nan	0	0,29	4	-0,23	2
cVc	11	С.	-0,10	50	0,22	58	0,53	60	0,34	30
cVc	11	С.+пауза	0,35	66	0,56	61	0,68	72	0,34	80
cVcv	02	Фр.+пауза	0,77	31	1,32	23	0,63	25	0,71	9
cVcv	02	С.	0,99	19	1,30	15	0,47	10	1,21	8
cVcv	02	С.+пауза	0,87	22	1,40	37	0,80	29	0,60	26
cVcv	11	Фр.+пауза	0,68	3	1,68	1	0,58	1	nan	0
cVcv	11	С.	0,47	56	0,68	86	0,71	91	0,41	30
cVcv	11	С.+пауза	0,55	60	0,77	86	0,88	33	0,38	72
cVcy	02	Фр.+пауза	0,16	31	0,28	23	0,43	25	-0,11	9
cVcy	02	С.	0,26	19	-0,10	15	-0,06	10	0,55	8
cVcy	02	С.+пауза	0,66	22	0,34	37	0,61	29	-0,02	26
cVcy	11	Фр.+пауза	2,41	3	0,68	1	0,18	1	nan	0
cVcy	11	С.	0,55	56	0,08	86	0,41	90	1,41	30
cVcy	11	С.+пауза	1,72	60	0,80	86	1,20	33	0,90	72
cVcvc	02	Фр.+пауза	0,73	21	1,18	15	0,45	9	0,88	4
cVcvc	02	С.	0,29	9	1,21	7	0,47	3	0,43	5
cVcvc	02	С.+пауза	1,05	14	1,61	11	0,50	4	0,66	15
cVcvc	11	Фр.+пауза	nan	0	nan	0	nan	0	-0,95	1
cVcvc	11	С.	0,07	29	0,57	40	0,51	40	0,51	13
cVcvc	11	С.+пауза	0,50	38	0,62	38	0,52	21	0,20	38
cVcyc	02	Фр.+пауза	-0,13	21	0,06	15	-0,23	9	-0,10	4
cVcyc	02	С.	0,22	9	-0,35	7	0,00	3	0,07	5
cVcyc	02	С.+пауза	-0,14	14	-0,19	11	-0,20	4	-0,40	15
cVcyc	11	Фр.+пауза	nan	0	nan	0	nan	0	0,32	1
cVcyc	11	С.	0,28	29	-0,13	40	-0,01	40	-0,11	13
cVcyc	11	С.+пауза	0,38	38	0,06	38	0,01	21	0,13	38

Зависимость длительности гласных от «глубины» границы и наличия паузы в данном случае представляется более сложной. Тенденция, обнаруженная для синтагм с фазовым ударением не на последнем слове — более высокая длительность в неконечных синтагмах перед паузой (обозначенных как «С.+пауза») (см. выше), — наблюдается не для всех дикторов (исключение составляет диктор А, для которого более высокой длительностью обладают гласные в неконечных синтагмах без последующей паузы) и не во всех контекстах (иначе распределяются длительности безударных гласных в синтагмах, оканчивающихся на «*Vcyc*», для модели 02; однако для данного контекста размер выборки слишком маленький). Тем не менее, для большинства дикторов и контекстов эта тенденция сохраняется.

**7. Заключение.** Анализ длительности гласных в позиции перед синтагматической границей, выполненный на материале большого речевого корпуса, позволил выявить следующие особенности темпорального оформления конца синтагмы.

1. Предпаузальное удлинение распространяется на ударный гласный последнего слова синтагмы, а также на заударный гласный, если он расположен в абсолютном конце слова.

2. В подавляющем большинстве случаев наличие паузы после синтагмы соотносится с большей степенью предпаузального удлинения гласных.

3. Как правило, в неконечных синтагмах степень предпаузального удлинения оказывается выше, чем в конечных.

4. Логическое ударение по типу ИК-2 реализуется с бóльшим удлинением ударного гласного и меньшим удлинением заударного (если он есть).

5. Для незавершенности по типу ИК-3 характерно меньшее удлинение ударного гласного и бóльшее удлинение заударного (если он есть).

В дальнейшем описанный метод планируется использовать для анализа длительности согласных в словах, расположенных на конце синтагм и фраз. Это позволит нам получить более полное описание особенностей временной организации крупных смысловых единиц.

### Литература

1. *Wightman C.W., et. al. Segmental durations in the vicinity of prosodic phrase boundaries // J. Acoust. Soc. Am. 1992. vol. 91. № 3. pp. 1707–1717.*
2. *Choi J.-Y., Hasegawa-Johnson M., Cole J. Finding intonational boundaries using acoustic cues related to the voice source // J. Acoust. Soc. Am. 2005. vol. 118. № 4. pp. 2579–2587.*
3. *Streeter L.A. Acoustic determinants of phrase boundary perception // J. Acoust. Soc. Am. 1978. vol. 64. № 6. pp. 1582–1592.*

4. *Bartkova K.* Rhythmic patterns and their automatic retrieval in spontaneous French // Proceedings of the Speech Prosody 2008 Conference, 2008. pp. 351–354.
5. *Campbell W.N.* Evidence for a Syllable-Based Model of Speech Timing // Proceedings of the International Conference on Spoken Language Processing, 1990. pp. 9–12.
6. *Cooper W.E., Danly M.* Segmental and Temporal Aspects of Utterance-Final Lengthening // *Phonetica*. 1981. vol. 38. № 1-3. pp. 106–115.
7. *Horne M., Strangert E., Heldner M.* Prosodic Boundary Strength In Swedish: Final Lengthening And Silent Interval Duration // Proceedings of the XIIIth International Congress of Phonetic Sciences. Stockholm: Congress organisers at KTH and Stockholm University, 1995. pp. 170–173.
8. *Lehiste I., Olive J.P., Streeter L.A.* Role of duration in disambiguating syntactically ambiguous sentences // *J. Acoust. Soc. Am.* 1976. vol. 60. № 5. pp. 1199–1202.
9. *Ostendorf M., et. al.* The Use of Relative Duration in Syntactic Disambiguation // Proceedings of the Workshop on Speech and Natural Language HLT '90. Stroudsburg, PA, USA: Association for Computational Linguistics, 1990. pp. 26–31.
10. *Pijper J.R. de, Sanderman A.A.* On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues // *J. Acoust. Soc. Am.* 1994. vol. 96. № 4. pp. 2037 – 2047.
11. *Price P.J., Ostendorf M., Wightman C.W.* Prosody and Parsing // Proceedings of the Workshop on Speech and Natural Language HLT '89. Stroudsburg, PA, USA: Association for Computational Linguistics, 1989. pp. 5–11.
12. *Skrelin P.A., et. al.* A Fully Annotated Corpus of Russian Speech // Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10). : European Language Resources Association (ELRA), 2010. pp. 109 – 112.
13. *Vaisière J.* Language-Independent Prosodic Features // *Prosody: Models and Measurements* Springer Series in Language and Communication. / edited by. D.A. Cutler, D.D.R. Ladd.: Springer Berlin Heidelberg, 1983. pp. 53–66.
14. *Volskaya N.B., Skrelin P.A.* Prosodic model for Russian // Proceedings of Nordic Prosody X. Helsinki: Frankfurt am Main: Peter Lang, 2009. pp. 249–260.
15. *Вольская Н.Б., Степанова С.Б.* Предпаузальное удлинение в русском языке // Экспериментально-фонетический анализ речи: проблемы и методы. Фонетическое многообразие языков мира. СПб., 2004. Вып. 5. С. 48–55.
16. *Кривнова О.Ф.* Глоттализация на границах фонетических составляющих в женской речи (на русском материале) // Проблемы и методы экспериментально-фонетических исследований. СПб., 2002. С. 165–189.
17. *Кривнова О.Ф.* Ритмизация и интонационное членение текста в «процессе речи-мысли»: опыт теоретико-экспериментального исследования. Диссертация на соискание ученой степени доктора филологических наук. // М., 2007. 347 с.

## References

1. *Wightman C.W., et. al.* Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 1992. vol. 91. no. 3. pp. 1707–1717.
2. *Choi J.-Y., Hasegawa-Johnson M., Cole J.* Finding intonational boundaries using acoustic cues related to the voice source. *J. Acoust. Soc. Am.* 2005. vol. 118. no. 4. pp. 2579–2587.
3. *Streeter L.A.* Acoustic determinants of phrase boundary perception. *J. Acoust. Soc. Am.* 1978. vol. 64. no. 6. pp. 1582–1592.
4. *Bartkova K.* Rhythmic patterns and their automatic retrieval in spontaneous French. Proceedings of the Speech Prosody 2008 Conference. 2008. pp. 351–354.
5. *Campbell W.N.* Evidence for a Syllable-Based Model of Speech Timing. Proceedings of the International Conference on Spoken Language Processing, 1990. pp. 9–12.
6. *Cooper W.E., Danly M.* Segmental and Temporal Aspects of Utterance-Final Lengthening. *Phonetica*. 1981. vol. 38. no. 1-3. pp. 106–115.
7. *Horne M., Strangert E., Heldner M.* Prosodic Boundary Strength In Swedish: Final Lengthening And Silent Interval Duration. Proceedings of the XIIIth International Con-

- gress of Phonetic Sciences. Stockholm: Congress organisers at KTH and Stockholm University, 1995. pp. 170–173.
8. Lehiste I., Olive J.P., Streeter L.A. Role of duration in disambiguating syntactically ambiguous sentences. *J. Acoust. Soc. Am.* 1976. vol. 60. no. 5. pp. 1199–1202.
  9. Ostendorf M., et. al. The Use of Relative Duration in Syntactic Disambiguation. Proceedings of the Workshop on Speech and Natural Language HLT '90. Stroudsburg, PA, USA: Association for Computational Linguistics, 1990. pp. 26–31.
  10. Pijper J.R. de, Sanderma A.A. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *J. Acoust. Soc. Am.* 1994. vol. 96. no. 4. pp. 2037–2047.
  11. Price P.J., Ostendorf M., Wightman C.W. Prosody and Parsing. Proceedings of the Workshop on Speech and Natural Language HLT '89. Stroudsburg, PA, USA: Association for Computational Linguistics, 1989. pp. 5–11.
  12. Skrelin P.A., et. al. A Fully Annotated Corpus of Russian Speech. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), 2010. pp. 109–112.
  13. Vaissière J. Language-Independent Prosodic Features. *Prosody: Models and Measurements* Springer Series in Language and Communication. Eds. D.A. Cutler, D.D.R. Ladd. Springer Berlin Heidelberg, 1983. pp. 53–66.
  14. Volskaya N.B., Skrelin P.A. Prosodic model for Russian. Proceedings of Nordic Prosody X. Helsinki: Frankfurt am Main: Peter Lang, 2009. pp. 249–260.
  15. Volskaya N.B., Stepanova S.B. [Prepausal lengthening in Russian]. *Eksperimental'no-foneticheskij analiz: rechi: problemy i metody. Foneticheskoe mnogoobrazie yazykov mira – Experimental phonetic analysis of speech: issues and methods. Phonetic systems of world languages.* St. Petersburg, 2004. no. 5. pp. 48–55. (In Russ.).
  16. Krivnova O.F. [Glottalization at prosodic boundaries in women's speech (evidence from Russian)]. *Problemy i metody jeksperimental'no-foneticheskikh issledovanij – Issues and methods of experimental phonetic research.* St. Petersburg, 2002. pp. 165 – 189. (In Russ.).
  17. Krivnova O.F. *Ritmizacija i intonacionnoe chlenenie teksta v «processe rechi-mysli»: opyt teoretiko-jeksperimental'nogo issledovanija. Dissertacija na so-iskanie uchenoj stepeni doktora filologicheskikh nauk [Rhythmization and intonational phrasing in the “process of speech and thought”: a theoretical and experimental study. Dissertation for the degree of Doctor of Philological Sciences].* Moscow, 2007. 347 p. (In Russ.).

**Качковская Татьяна Васильевна** — младший научный сотрудник лаборатории экспериментальной фонетики Института филологических исследований СПбГУ, аспирант кафедры фонетики и методики преподавания иностранных языков СПбГУ. Область научных интересов: фонетика, речевые технологии. Число научных публикаций — 4. [tania200288@rambler.ru](mailto:tania200288@rambler.ru); СПбГУ, Филологический факультет, кафедра фонетики и методики преподавания иностранных языков, Университетская наб. 11, 199034, Санкт-Петербург, РФ; р.т. +7(812)328-9565.

**Kachkovskaia Tatiana Vasilievna** — junior researcher, Laboratory of Experimental Phonetics, St. Petersburg State University; PhD student, Department of Phonetics and Foreign Language Teaching Methodology, St. Petersburg State University. Research interests: phonetics, speech technology. Number of publications — 4. [tania200288@rambler.ru](mailto:tania200288@rambler.ru); St. Petersburg State University, Faculty of Philology, Department of Phonetics and Foreign Language Teaching Methodology, 11 Universitetskaya emb., St. Petersburg, 199034, Russia; office phone +7(812) 328-9565.

**Поддержка исследований.** Работа выполнена при финансовой поддержке СПбГУ (проект № 31.0.145.2010).

**Acknowledgements.** The research is supported by Saint Petersburg State University (grant 31.0.145.2010).

## РЕФЕРАТ

### *Качковская Т.В.* **Использование темпоральных характеристик для сегментации речевого потока на крупные смысловые единицы (на материале русского языка).**

Данная статья посвящена проблеме сегментации речевого потока на крупные смысловые единицы — синтагмы и фразы.

В статье рассматривается вопрос о зависимости длительности ударного и заударного гласного в слове от позиции слова по отношению к границе синтагмы или фразы, а также от типа мелодического движения, «глубины» границы и наличия последующей паузы. Исследование представляет собой анализ данных, полученных на материале корпуса профессиональной русской речи.

Статья состоит из 7 частей. Во введении дается определение понятия предпаузального удлинения; автор указывает на то, что это явление представляет собой один из ключевых маркеров границ крупных смысловых единиц. Вторая часть посвящена обоснованию этого тезиса и представляет собой обзор работ, в которых рассматривается взаимодействие граничных маркеров. Далее приводится список факторов, которые могут влиять на длительность звуков и которые необходимо учитывать при анализе данных.

В четвертой части описываются основные подходы к измерению темпоральных характеристик, и обосновывается выбор нормализованной длительности в качестве основного параметра для анализа данных.

В пятой части описывается аннотированный корпус профессиональной русской речи CORPRES, используемый в качестве материала для анализа, и метод исследования — получение из корпуса всех необходимых данных.

Основной частью работы является раздел 6, в котором приводятся данные о нормализованной длительности ударных и заударных гласных в словах, заканчивающихся на «*v*», «*vс*», «*vсv*» и «*vсvс*». Данные приводятся отдельно для трех типов позиций слова: на конце синтагмы, за которой следует пауза; на конце синтагмы, за которой не следует пауза; на конце фразы, за которой следует пауза.

Сначала приводятся данные о длительности гласных в тех случаях, когда последнее (рассматриваемое) слово синтагмы или фразы не несет фразового ударения. Анализируя эти данные, автор показывает, что предпаузальное удлинение связано именно с конечным положением слова, а не наличием интонационного центра на последнем слове. Далее рассматриваются синтагмы и фразы с интонационным центром на последнем слове, и выполняется сравнение значений нормализованной длительности гласных для двух типов мелодического движения: логического ударения по типу ИК-2 и незавершенности по типу ИК-3.

Данные, приведенные в статье, позволяют сформулировать список факторов, влияющих на степень удлинения гласных в словах, расположенных на концах крупных смысловых единиц: наличие последующей паузы, «глубина» границы, наличие фразового ударения на последнем слове синтагмы и тип интонационного контура.

Кроме того, результаты исследования позволяют сделать вывод о локализации предпаузального удлинения в русском языке: оно распространяется на ударный гласный последнего слова синтагмы, а также на заударный гласный, если он расположен в абсолютном конце слова.

## SUMMARY

### *Kachkovskaia T.V. Temporal Aspects of Intonational Phrasing (Evidence from Russian).*

The present paper addresses the problem of speech segmentation into major prosodic units: intonational phrases and utterances.

The paper focuses on factors influencing the duration of stressed and post-stressed vowels in words immediately preceding prosodic boundaries. The factors in question are pitch movement type, boundary "depth" and the presence of a following pause. This is performed through analysis of data obtained from the Corpus of Professionally Read Speech (CORPERS) for Russian.

The paper is divided into 7 sections. In the introduction the notion of final lengthening is defined, which is claimed to be one of the key markers of major prosodic boundaries. Section 2 provides evidence for this statement based on the analysis of papers dealing with the issue of prosodic boundary markers. In the next section the factors are listed which might influence segment duration and therefore must be taken into account when analyzing durational data.

Section 4 provides an overview of different approaches for duration and tempo measurement, justifying the use of normalized duration as the main parameter.

Section 5 describes the annotated corpus of Professionally Read Speech, which the research is based on, and the methods of handling the corpus data.

The essential part of the paper is section 6, which provides normalized duration values for vowels in words ending in «cV», «cVC», «cVcV», and «cVcvc». The values are grouped according to the type of prosodic unit (intonational phrase vs. utterance) and the presence/absence of a pause.

The cases where the word in question (the last word in the unit) does not bear nuclear stress are analyzed first, showing that final lengthening is indeed caused by the word's position. The following analysis of units with nuclear stress on the last word aims at comparing duration values for two types of pitch movement: the contour used for logical stress and rising-falling intonation used in non-final phrases.

The data described in the paper enables to list the factors influencing the degree of vowel lengthening in words immediately preceding prosodic boundaries. These are the presence of a following pause, boundary "depth", the presence of nuclear stress on the last word, and pitch movement type.

The data also lead to a conclusion that the locus of final lengthening in Russian is the last stressed vowel and the last post-stressed vowel when not followed by a consonant.

В.А. ЗАХАРЬЕВ, А.А. ПЕТРОВСКИЙ, Б.М. ЛОБАНОВ  
**СИСТЕМА СИНТЕЗА РЕЧИ ПО ТЕКСТУ С ВОЗМОЖНОСТЬЮ  
НАСТРОЙКИ НА ГОЛОС ЦЕЛЕВОГО ДИКТОРА**

---

*Zahariev V.A., Petrovskiy A.A., Lobanov B.M. Система синтеза речи по тексту с возможностью настройки на голос целевого диктора.*

**Аннотация.** В статье представлена схема построения мультиголосового синтезатора речи, основанная на использовании синергетического эффекта от интеграции системы синтеза речи по тексту и конверсии голоса. Такая организация даёт возможность одновременно выполнять действия синтеза и модификации речевого сигнала на основе комплексного подхода, позволяя снизить количество ошибок и артефактов, которые влияют на качество речевого сигнала. Применение данного подхода обеспечивает реализацию функции настройки синтезатора речи на голос целевого диктора без существенных затрат трудоёмкости на обучение речевой базы данных, для добавления новых голосов.

**Ключевые слова:** синтез речи по тексту, конверсия голоса, функция конверсии голоса

*Zahariev V.A., Petrovskiy A.A., Lobanov B.M. Text to speech synthesis system with the target speaker voice customization capability.*

**Abstract.** A scheme of constructing multi-voice speech synthesizer based on the use of the synergies of integration of the text to speech and voice conversion systems are presented in this article. Such organization of the system allows simultaneous synthesis and modification actions in speech signal, based on an integrated approach to its treatment, significantly reducing the number of errors and artifacts that affect the resulting quality. Applying this approach let to implement a function for multivoice speech synthesizer without significant labor costs for training speech database to add new voices.

**Keywords:** text to speech synthesis, voice conversion, voice conversion function.

---

**1. Введение.** На данном этапе развития систем синтеза речи по тексту (ССРТ) ставится вопрос уже не столько об обеспечении хороших уровней основных показателей систем этого класса, например, разборчивости синтезируемой речи, сколько о более сложных характеристиках, таких как, натуральность синтезируемой речи, поддержка множества языков и различных голосов дикторов. Последний аспект — создание мультиголосовых систем синтеза речи по тексту — требует особого подхода и внимания, поскольку в существующих ССРТ перенастройка системы на нового диктора требует больших материальных и временных затрат от разработчиков системы [1].

В данной статье предлагается рассмотреть возможность решения задачи построения мультиголосового синтезатора речи с использованием технологии конверсии голоса, на основе системы синтеза речи по тексту, разработанной авторами.

**2. Архитектура мультиголосовой ССРТ.** Система конверсии голоса (КГ) базируется на технологии обработки речевого сигнала,

позволяющей реализовать процесс трансформации параметров голоса, характеризующих речь исходного диктора (ИД), в параметры целевого диктора (ЦД). Объектами конверсии голоса, как технологии обработки сигналов, являются стабильные во времени свойства говорящего, проявляющиеся в речевом сигнале через изменение его акустических параметров [2–4].

Вполне очевидной является идея применения данной технологии в синтезаторах речи по тексту для решения задачи добавления функций мультиголосового синтеза. В простейшем случае системы СРТ и КГ являются полностью независимыми: выходной сигнал, поступающий от ССРТ используется в качестве входного сигнала для системы КГ [5,6]. К положительным сторонам данного подхода можно отнести устойчивость и универсальность данной архитектуры, в её рамках легко может быть заменена любая составляющая, без нарушения общей работоспособности всей системы. Например, при смене модели синтеза или конверсии на более совершенную.

Однако, реализация данного подхода может привести к существенной потере качества т.к. изменение просодических характеристик речи (частоты основного тона и длительности звуков) осуществляется дважды: первый раз просодическим процессором синтезатора речи по тексту и второй раз системой конверсии голоса.

Второй подход [7], развиваемый авторами в данной работе, подразумевает тесное объединение двух типов систем, путём включения элементов системы конверсии голоса в состав синтезатора речи по тексту, и использования ожидаемого синергетического эффекта от такого варианта построения интегрированной архитектуры системы. Реализация данного подхода достигается путём внедрения модуля конверсии голоса в блок акустического процессора и рационального разделения задач конверсии параметров голоса между двумя видами систем с учётом особенностей этапов обработки информации в каждой из них.

Предлагаемая архитектура мультиголосовой ССРТ представлена на рисунке 1. В ней были учтены следующие важные нюансы. Во-первых, аспекты конверсии голоса учитываются при выборе единиц компиляции. Во-вторых, все алгоритмы преобразования и конверсии (спектральные и просодические) выполняются единым блоком, это означает, что характеристики сигнала модифицируются только один раз. В-третьих, конкатенация и реконструкция синтезированного речевого сигнала выполняются после конверсии голоса исходного диктора в голос целевого.

**3. Модель представления речевого сигнала.** В качестве основного метода для выполнения операций анализа-синтеза и параметризации речевого сигнала, мы предлагаем использовать модель сигнала на базе адаптивной интерполяции взвешенного спектра — Speech Transformation and Representation by Adaptive Interpolation of weiGHTed spectrum (STRAIGHT), первоначально разработанную и построенную профессором Хайдеки Кавахара для исследования механизмов восприятия речи человека, основанную на её вокодерном представлении. STRAIGHT использует процедуры, которые могут быть сгруппированы в три подсистемы: экстрактор информации о сигнале возбуждения, экстрактор сглаженного частотно-временного представления сигнала, и движок синтеза, состоящий из источника возбуждения и системы фильтров с изменяющимися во времени характеристиками [8]. Результаты работы двух первых блоков вычисления параметров речевого сигнала представлены на рисунке 2.

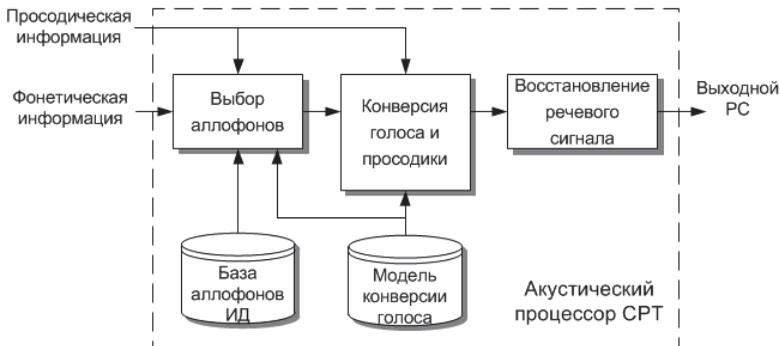


Рис. 1. Архитектура мультиголосового CPT

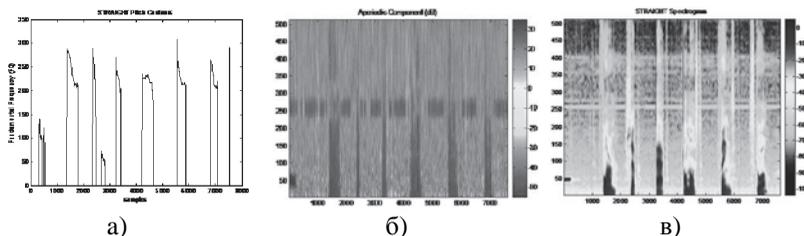


Рис. 2. Параметры сигнала, полученные на основе модели STRAIGHT: а — контур ЧОТ; б — спектрограмма аperiodической компоненты; в — спектрограмма периодической компоненты

Рассмотрим основные особенности этапов работы подсистем данной модели. Для того чтобы иметь возможность адекватно описывать, трансформировать, и восстанавливать речь, очень важной является возможность точного определения контура частоты основного тона (ЧОТ), который бы не имел никаких следов интерференций, вызванных неправильным размером и формой окна анализа сигнала.

На первом этапе подсистема извлечения сигнала возбуждения определяет частоту основного тона как мгновенную частоту фундаментальной компоненты сигнала. Она определяется как отображение фиксированных точек в область мгновенной частоты из коэффициентов оконного преобразования Фурье.

Процедура, основанная на использовании нормализованной автокорреляционной функции, была интегрирована с предыдущей процедурой поиска мгновенной частоты для уменьшения числа ошибок определения ЧОТ. Представленный способ экстракции контура ЧОТ предполагает, что сигнал имеет близкую к гармонической структуру. Тем не менее, практически всегда существуют отклонения от строгой периодичности, которые вводят в сигнал дополнительные негармонические компоненты. Таким образом, существует возможность найти меру аperiodичности, принимая за её величину отношение энергии составляющих на частотах, нарушающих гармонический строй, к полной энергии сигнала. Аperiodическая компонента  $A(w, t)$  оценивается как остаток между исходным сигналом и его периодической составляющей [9].

Второй этап, экстракция сглаженного частотно-временного представления (спектрограммы) сигнала является “ядром” данной модели. В случае с простыми вокодерными системами периодическое возбуждение фильтров является эффективной стратегией для передачи резонансной информации и увеличения отношения сигнал-шум. Однако, это периодическое возбуждение вносит дополнительные интерференции как во временной так и в частотной областях, как это показано в левой части рисунка 3. Поэтому возникает необходимость в реконструкции гладкой поверхности частотно-временного плана из искаженного представления.

Следующие два шага последовательно выполняются для решения данной проблемы. Первым шагом является сокращение оставшейся временной периодичности из-за фазовой интерференции между соседними гармоническими составляющими.

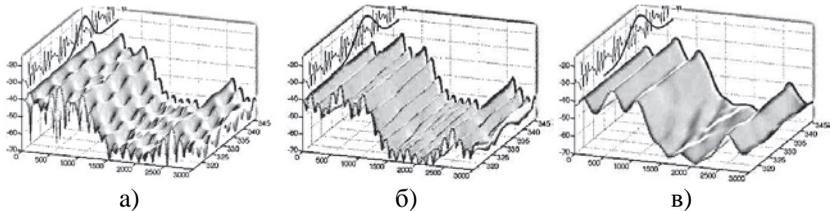


Рис. 3. Процесс экстракции сглаженного частотно-временного представления: а — спектр оригинального сигнала; б — спектр сигнала с устранённой темпоральной волатильностью; в — спектр сигнала в модели STRAIGHT

Это достигается за счёт применения, помимо основного окна анализа, дополнительного комплементарного временного окна. Комплементарное окно  $w_c(t)$  к окну анализа  $w(t)$  определяется согласно выражению:

$$w_c(t) = w(t) \sin \frac{\pi t}{T_0},$$

где  $T_0$  — период основного тона сигнала. Спектр комплементарного сигнала  $P_c(w, t)$ , вычисленный с помощью комплементарного окна, имеет пики там, где спектр оригинального сигнала  $P(w, t)$ , вычисленный с помощью обычного окна (например, на рисунке 3а с помощью окна Гаусса), имеет провалы. Далее рассчитывается спектр сигнала с устранённой темпоральной волатильностью  $P_R(w, t)$ , который определяется путем вычисления взвешенной суммы двух типов спектров  $P(w, t)$  и  $P_c(w, t)$  с весовым коэффициентом  $\xi$ , отыскиваемым в процессе решения задачи оптимизации [10]. Рисунок 3б показывает спектр сигнала  $P_R(w, t)$ , полученного при оптимальном значении весового коэффициента. Легко видеть, что на данном рисунке отсутствуют провалы в частотно-временном плане, наличие которых очевидно на левом графике, где представлен спектр оригинального сигнала  $P(w, t)$ .

Вторым шагом является обратная фильтрация на основе сплайнов. Она необходима для того, чтобы удалить оставшиеся паразитные периодичности в частотной области, одновременно с сохранением распределения первоначальных спектральных уровней по частотам

гармоник. STRAIGHT реализует операцию сглаживания с использованием базисной функции на основе Б-сплайна второго порядка. Поскольку спектр сигнала, вычисленный с использованием комплементарного набора окон, не представляет собой линейный спектр, а также дает некоторое размытие энергии сигнала по частоте, используется соответствующее сглаживающее ядро  $h_{\Omega}$  для восстановления “смазанных” значений на частотах гармоник. Использование такой весовой функции  $h_{\Omega}$  приводит к тому, что мы получаем спектр, менее чувствительный к ошибкам определения ЧОТ, а также с хорошей локализацией энергии по частоте, практически соответствующий линейному спектру. Следующее выражение является расчётным для определения восстановленной spectroграммы  $P_{ST}(w, t)$  (spectrogramмы STRAIGHT):

$$P_{ST}(w, t) = \left( \int h_{\Omega}(\lambda / w_0) P_R^{\gamma}(w - \lambda, t) d\lambda \right)^{1/\gamma},$$

где  $w_0$  — циклическая частота от  $F_0$ . Параметр  $\gamma$  определяет нелинейность характеристики и был выбран равным 0,3 на основе экспериментов. Форма весовой функции  $h_{\Omega}$  вычисляется путём решения системы линейных уравнений, полученных из  $w(t)$ ,  $w_c(t)$ ,  $\xi$  и  $\gamma$  [10]. В крайней правой части рисунка 3в показан спектр сигнала в модели STRAIGHT для гласного звука /a/, произнесённого диктором мужского рода. Необходимо обратить внимание, что в процессе выполнения вышеописанных шагов (при переходе от графика к графику слева направо) постепенно снижается степень проявления интерференции и различных искажений частотно-временного плана, с одновременным сохранением детализации спектральной картины на частотах гармонических составляющих сигнала.

После выполнения всех необходимых преобразований для реконструкции сигнала, аппарату синтеза необходим соответствующий набор параметров (спектр STRAIGHT  $P_{ST}(w, t)$ , карты аперидичности  $A(w, t)$ , контура ЧОТ  $F_0$  с отметками о вокализованности  $f_0(t)$ ). Все эти параметры имеют строго определённый физический смысл, связанный с процессом речеобразования, и в тоже время позволяют проводить с ними разнообразные и независимые манипуляции, без опасения внесения различного рода несоответствий между параметрами.

**4. Модель конверсии голоса.** Для настройки мультиголосового СРТ на целевого диктора необходимо предварительное обучения модели конверсии голоса. Данный этап состоит из ряда шагов:

— Подготовки априорной речевой информации об ИД и ЦД в виде набора фонограмм. Вся необходимая информация об ИД, позволяющая сгенерировать произвольный набор обучающих аудиозаписей фраз, уже имеется в акустической базе аллофонов ССРТ. Для ЦД возможны варианты в зависимости от выбранного пользователем способа обучения системы. В случае текстозависимого способа, обучающая выборка аудиозаписей формируется путём озвучивания фонетически сбалансированного текста, предлагаемого пользователю модулем обучения системы. В случае текстонезависимого способа, пользователь может сам вводить в систему любой текст, параллельно произнося его в микрофон, или загрузить произвольный набор фонограмм и их стенограмм. Например, фрагменты аудиокниги и соответствующие им фрагменты текста. В случае невыполнения условия фонетической сбалансированности, предоставляемого пользователем речевого и текстового материала, система автоматически сформирует пользователю запрос на предоставление дополнительной информации [11,12].

— Проведения анализа и параметризации данного набора фонограмм с помощью модели STRAIGHT и получения последовательностей векторов параметров характеризующих голос ИД и ЦД.

— Определения соответствия между этими последовательностями векторов параметров с помощью метода динамического программирования, при текстозависимом обучении, или аппарата скрытых Марковских моделях, для текстонезависимого способа обучения [13,14]. Объединения данных двух отмасштабированных и выровненных последовательностей в общую матрицу, фактически, представляющую собой совместное пространство векторов параметров дикторов.

— Кластеризации получившегося совместного пространства векторов параметров с помощью статистических методов, в частности методов мягкой классификации на основе модели Гауссовых смесей (ГС). Характеристики данных классов или кластеров, выступают в дальнейшем в качестве параметров функции конверсии, которая является ядром модели и выполняет сам процесс трансформации признаков голоса.

Подробно шаги этапа обучения описаны в работах авторов и литературе [15], а далее рассмотрим особенности реализации функции

конверсии как наиболее важного элемента используемого на этапе выполнения системы.

За основу функции конверсии, являющейся ядром системы, было принято выражение на основе линейной регрессионной модели первого порядка, которая использует в качестве своих коэффициентов регрессии параметры ГС, полученные на этапе обучения [16]. Эта функция доказала свою эффективность, особенно по сравнению с подходами, основанными на жесткой кластеризации пространства акустических параметров дикторов [17]. Однако, детальный её анализ выявил ряд недостатков. Сложность выбора порядка модели ГС, с ростом которой увеличивается вычислительная сложность, а также быстро деградирует выходное качество речевого сигнала в результате переобучения и чрезмерного усреднения параметров моделью. Кроме того, статистические отношения рассматриваются исключительно между одной парой векторов параметров ИД и ЦД в каждый  $i$ -ый момент времени (рис. 4а). В данной работе предлагается расширенный вариант функции конверсии, основанный на модели ГС и свойствах эргодичности сигнала. Это позволяет учитывать не только пространственные, но и временные корреляционные связи между соседними векторами параметров ИД и ЦД, тем самым придавая процедуре конверсии свойства Марковского процесса (рис 4б):

$$y_i = \sum_{q=1}^Q p_q(x_i, y_{i-1}, x_{i+1}) [\mu_q + \Phi_q \bar{x}_i^q + \Psi_q \bar{y}_{i-1}^q + \Omega_q \bar{x}_{i+1}^q],$$

где  $i$  — порядковый номер фрейма сигнала  $i=1, \dots, T$ ,  $q$  — индекс смеси в составе модели ГС  $q=1, \dots, Q$ ,  $x=[x_1, x_2, \dots, x_T]^T$ ,  $y=[y_1, y_2, \dots, y_T]^T$  — последовательность векторов параметров ИД и ЦД  $x_i, y_i \in \mathbb{R}^{l \times p}$ ,  $p$  — размерность вектора параметров сигнала,  $p_q(x_i, y_{i-1}, x_{i+1})$  — апостериорная вероятность того, что входной вектор  $x_i$  и вектора  $y_{i-1}$  и  $x_{i+1}$  принадлежат к  $q$ -ой компоненте ГС,  $\mu_q = [\mu_1, \mu_2, \dots, \mu_Q]^T$  — вектор математических ожиданий для каждой компоненты смеси ЦД  $\mu_q \in \mathbb{R}^{l \times p}$ ,  $\{\Phi, \Psi, \Omega\}$  — матрицы регрессионных коэффициентов для всех компонент смеси независимой перемен-

ной  $x_i$ , а также предикторов  $x_{i+1}$  и  $y_{i-1}$  где  $\Phi_q, \Psi_q, \Omega_q \in \mathbb{R}^{p \times p}$ . Тогда задача нахождения неизвестных параметров  $\{\mu, \Phi, \Psi, \Omega\}$  формулируется как задача оптимизации, решение которой можно найти по методу наименьших квадратов [18].

Вторым улучшением модели конверсии стало применение в её рамках метода спектрального взвешивания, имеющего глубинную связь с физической природой речевого сигнала. Это позволяет на базе имеющейся информации о входном векторе параметров  $x$ , производить конверсию спектральной огибающей с последующим её взвешиванием специальной масштабирующей функцией в виде линейной комбинации  $Q$  базисных функций  $\{W_q(f)\}$ , используя апостериорные вероятности  $p_q$  как коэффициенты в данной линейной комбинации.

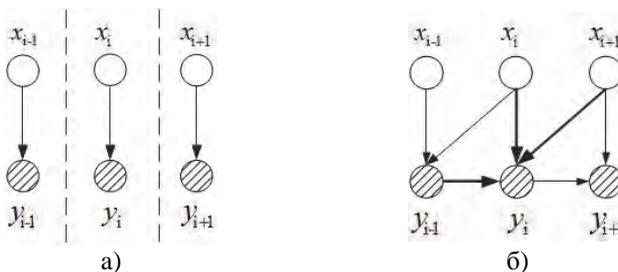


Рис. 4. Отношения между парами векторов в обучающей последовательности: а — независимая модель; б — Марковский процесс

Таким образом, осуществляется обработка параметров сигнала после процедуры конверсии с целью решения проблемы чрезмерного усреднения огибающей в результате конверсии или возможного переобучения системы.

Для преобразования просодических характеристик голоса диктора предлагается применить новый метод, использующий параметрическое представление контура частоты основного тона. Он основан на поиске особых точек контура частоты, в соответствии с методикой предложенной Паттерсоном [19]. Далее, используя кусочно-линейную функцию конверсии ЧОТ специального вида, которая имеет различные свойства в различных областях значений ЧОТ, выполняется схема преобразования просодических единиц [20]. Эта манипуляция обеспечивает лучшее качество конверсии просодических элементов без су-

щественного увеличения вычислительных затрат, по сравнению со стандартной процедурой нормализации статистических оценок ЧОТ в классических алгоритмах конверсии.

Блок реконструкции формы сигнала производит восстановление речевого сигнала из набора параметров на основе модели STRAIGHT. Таким образом, по входному тексту на выходе системы, мы получаем реконструированную фразу озвученную голосом целевого диктора, что позволяет с помощью представленных методов решить поставленную задачу мультиголосового синтеза речи по тексту.

**5. Экспериментальная часть.** В данном разделе статьи представлена экспериментальная оценка качества работы мультиголосового синтезатора речи по тексту. Основная идея эксперимента заключалась в определении того, насколько речь синтезируемая голосом целевого диктора, с использованием конверсии, соответствует голосу реального целевого диктора, или же в речевом сигнале всё еще присутствуют черты исходного диктора. Такой вывод можно сделать, основываясь на степени близости между речевым сигналом синтезируемым системой мультиголосового синтеза и оригинальных аудиозаписей тестовых фраз для целевых дикторов с одной стороны, и озвученных исходным диктором (т.е. синтезированных системой, с уже имеющимся в ней диктором) с другой. Степень близости, в свою очередь мы предлагаем определять метрикой на базе мел-кепстрального расстояния, как одной из наиболее адекватных объективных оценок определения акустического расстояния между векторами [21]. Более подробно методика эксперимента изложена далее.

Из специального фонетически сбалансированного текста для создания макси-набора аллофонов, диаллофонов аллослогов русской речи [1] сформированы один обучающий (30 предложений), а также пять тестовых (по 4 предложения) наборов текстов различных фраз. На основе данных наборов обучающих и тестовых текстов фраз были записаны их фонограммы для четырёх различных целевых дикторов: двух женщин (ДЖ1, ДЖ2) и двух мужчин (ДМ1, ДМ2). Полученные аудиофайлы были закодированы в формате wav, с частотой дискретизации 16000кГц и разрядностью сетки квантования в 16 бит. Далее, на основании имеющихся фонограмм и стенограмм тренировочной выборки фраз было проведено обучение и настройка системы мультиголосовой ССРТ в текстозависимом режиме, на каждого из четырёх целевых дикторов. Затем, для каждого тестового набора фраз, уже непосредственно в режиме функционирования мультиголосовой системы, был осуществлён их синтез, вначале голосом исходного диктора,

имеющегося в синтезаторе, а затем голосами целевых дикторов с использованием конверсии голоса. Далее, над результирующими аудиоданными, полученными на тестовом наборе фраз, т.е. оригинальными и синтезированными фонограммами для целевых дикторов, а также фонограммами синтезированным голосом исходным диктора системы, выполнялся мел-кепстральный анализ сигнала, с размерностью вектора кепстральных коэффициентов, для одного фрейма, равным двадцати. В параметрическом представлении наборы фраз были объединены в последовательности векторов параметров мел-кепстральных коэффициентов. А затем для каждого диктора по каждому набору фраз рассчитывались две оценки близости, синтезированного системой сигнала, к исходному диктору  $\varepsilon_{ИД}$ , и вторая — к целевому  $\varepsilon_{ЦД}$ . Они рассчитывались как средняя квадратичная ошибка между векторами кепстральных коэффициентов для всей последовательности векторов одного набора фраз, согласно выражению [21]:

$$\varepsilon_{ИД} = \frac{1}{N} \sum_{n=1}^N \|c_n^{synt} - c_n^{src}\|^2,$$

$$\varepsilon_{ЦД} = \frac{1}{N} \sum_{n=1}^N \|c_n^{synt} - c_n^{trg}\|^2,$$

где  $N$  — количество векторов кепстральных коэффициентов в одном наборе фонограмм параметризованных аудиозаписей тестовых фраз,  $c_n^{src}$  —  $n$ -ый вектор кепстральных коэффициентов параметризованной последовательности фонограмм, одного набора, синтезированных голосом исходного диктора (имеющегося в синтезаторе),  $c_n^{synt}$  —  $n$ -ый вектор кепстральных коэффициентов параметризованной последовательности фонограмм, одного набора, синтезированных голосом текущего диктора (не которого настроен синтезатор в текущий момент),  $c_n^{trg}$  —  $n$ -ый вектор кепстральных коэффициентов параметризованной последовательности фонограмм, одного набора, озвученных голосом целевого диктора (реального диктора, на которого была настроена система). Результирующие оценки представлены в таблице.

Таблица. Результаты экспериментальной оценки качества работы системы

Диктор	Набор тестовых фраз									
	I		II		III		IV		V	
	$\varepsilon_{ид}$	$\varepsilon_{цд}$	$\varepsilon_{ид}$	$\varepsilon_{цд}$	$\varepsilon_{ид}$	$\varepsilon_{цд}$	$\varepsilon_{ид}$	$\varepsilon_{цд}$	$\varepsilon_{ид}$	$\varepsilon_{цд}$
ДЖ1	0,89	0,62	0,82	0,76	0,90	0,67	0,87	0,63	0,80	0,59
ДЖ2	0,99	0,78	0,96	0,91	0,99	0,78	0,98	0,72	0,91	0,68
ДМ1	0,75	0,56	0,71	0,57	0,78	0,57	0,78	0,52	0,70	0,50
ДМ2	0,73	0,54	0,74	0,63	0,80	0,62	0,73	0,49	0,76	0,44
<b>Среднее:</b>	0,84	0,63	0,81	0,72	0,87	0,66	0,84	0,59	0,79	0,55

Исходя из таблицы, общие среднеквадратические ошибки по всем наборам фраз составляют  $\varepsilon_{ид} = 0,84$  для голосов исходного и синтезированного системой диктора, а также  $\varepsilon_{цд} = 0,63$  для синтезированного и целевого дикторов. По результатам экспериментов можно наблюдать следующую картину: речевой сигнал, генерируемый системой мультиголосового синтеза речи, настроенной на целевого диктора, оказывается в параметрическом представлении на тридцать процентов ближе к речевому сигналу реального целевого диктора, чем речь, произведённая этой же системой, голосом имеющегося в ней исходного диктора. Это следствие, в свою очередь, позволяет сделать вывод о том, что речь генерируемая системой мультиголосового синтеза будет, с большей степенью вероятности, воспринята слушателем как произнесённая голосом целевого диктора. Данный результат подтверждает возможность создания системы синтеза речи по тексту с возможностью настройки на голос целевого на базе предложенной архитектуры системы, а также методов и моделей её реализации.

**6. Заключение.** В работе изложены новые подходы и принципы создания многоголосого синтезатора речи по тексту с использованием технологии конверсии голоса. Представлено обоснование выбора рациональной архитектуры системы на базе интерактивного подхода между двумя типами систем. Предлагаемая схема, на основе синергетического эффекта от интеграции двух типов систем, позволяет в полной мере использовать полезные свойства обоих, и решает проблему создания многоголосого СРТ с улучшенными показателями качества конверсии. Расширенная регрессионная функция конверсии применяется для повышения точности преобразования векторов параметров, а метод спектрального взвешивания — для решения проблемы чрезмерного усреднения спектральной огибающей. Всё это в итоге повышает качественные характеристики результирующего восстановленного сигнала, сохраняя вместе с тем высокие уровни узнаваемости голоса целевого диктора для преобразованной речи.

## Литература

1. *Лобанов Б. М., Цирульник Л. И.* Компьютерный синтез и клонирование речи // Минск: Белорусская наука, 2008. 344 с.
2. *Abe M., Nakamura S., Shikano K.* Voice conversion through vector quantization // Proc. of International Conference on Acoustics, Speech and Signal Processing. New York, 1988. pp. 655–658.
3. *Valbret H., Moulines E., Tubach J.P.* Voice transformation using PSOLA technique // Proc. of International Conference on Acoustics, Speech and Signal Processing. 1992. vol. 1. pp. 145–148.
4. *Moulines E., Sagisaka Y.* Voice conversion: State of the art and perspectives // Speech Communication. 1995. pp. 125–224.
5. *Kain A., Macon M. W.* Text-to-speech voice adaptation from sparse training data // Proc. of International Conference on Spoken Language Processing. 1998. pp. 2847 – 2850.
6. *Sundermann D., Hoge H., Bonafonte A.* Text-independent voice conversion based on unit selection // Proc. of International Conference on Acoustics, Speech and Signal Processing. 2006. vol. 1.
7. *Azarov E, Petrovsky A.A, Lobanov B, Tsurulnik L.* Text-to-speech system with acoustic processor based on the instantaneous harmonic analysis // SPECOM. 2009. pp. 414–418.
8. *Kawahara H., Morise M.* Technical foundations of tandem-straight, a speech analysis, modification and synthesis framework // SADHANA. Academy Proceedings in Engineering Sciences, 2011. pp. 713–722.
9. *Kawahara H., Katayose H., Cheveigne A.* Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity // Proc. Eurospeech'99. 1999. pp. 2781–2784.
10. *Kawahara H., Masuda I., Cheveigne A.* Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous–frequency–based F0 extraction // Speech Communication. 2004. pp. 187–207.
11. *Arslan L.* Speaker transformation algorithm using segmental codebooks // Speech Communication. 1999. vol. 28, no. 3. P. 211–226.
12. *Narendranath M., Murthy H., Rajendran S., Yegnanarayana N.* Transformation of formants for voice conversion using artificial neural networks // Speech Communication. 1995. vol. 16, no. 2. pp. 207–216.
13. *Рабинер Л. П.* Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: Обзор // Труды института инженеров по электронике и радиоэлектронике. 1989. Т. 77, № 2. С. 86–120.
14. *Азаров И. С., Петровский А. А.* Система конверсии голоса в реальном масштабе времени с текстонезависимым обучением на основе гибридного параметрического описания речевых сигналов // Цифровая обработка сигналов. 2012. № 2. С. 15–23.
15. Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика (с MATLAB примерами) / под редакцией А. А. Петровского. // Минск: Бестпринт, 2009. 456 с.
16. *Stylianou, Y., Cappé O., Moulines E.* Statistical methods for voice quality transformation // Proc. of European Conference on Speech Communication and Technology. Madrid, 1995. pp. 447–450.
17. *Павловец А.С., Лившиц М.З., Личачев Д. С., Петровский А. А.* Конверсия голоса с использованием модели сепарации речевого сигнала на компоненты “гармоники + шум” и переходные фреймы // Речевые технологии. 2008. №4. С. 37–50.
18. *Stylianou Y., Cappé O., Moulines E.* Continuous probabilistic transform for voice conversion // Proc. of International Conference on Acoustics, Speech and Signal Processing. 1998. pp. 2451–2455.
19. *Patterson D.* A linguistic approach to pitch range modeling // PhD dissertation. Scotland: University of Edinburgh, 2000. 201 p.

20. *Захарьев В.А., Петровский А.А.* Конверсия просодических характеристик диктора на основе методов параметризации контура частоты основного тона. Доклады БГУИР // Минск, 2013. С. 39–46.
21. *Hu Y., Loizou P.C.* Evaluation of objective quality measures for speech enhancement // IEEE Transactions on Audio, Speech & Language Processing. 2008. vol. 16, no. 1. pp. 229–238.

## References

1. Lobanov B. M., Cirul'nik L. I. *Komp'yuternyj sintez i klonirovanie rechi* [Computer speech synthesis and cloning] Minsk: Belorusskaja nauka, 2008. 344 p. (In Russ.).
2. Abe M., Nakamura S., Shikano K. Voice conversion through vector quantization. Proc. of International Conference on Acoustics, Speech and Signal Processing. New York, 1988. pp. 655–658.
3. Valbret H., Moulines E., Tubach J.P. Voice transformation using PSOLA technique. Proc. of International Conference on Acoustics, Speech and Signal Processing. 1992. vol. 1. pp. 145–148.
4. Moulines E., Sagisaka Y. Voice conversion: State of the art and perspectives. Speech Communication. 1995. pp. 125–224.
5. Kain A., Macon M. W. Text-to-speech voice adaptation from sparse training data. Proc. of International Conference on Spoken Language Processing. 1998. pp. 2847–2850.
6. Sundermann D., Hoge H., Bonafonte A. Text-independent voice conversion based on unit selection. Proc. of International Conference on Acoustics, Speech and Signal Processing. 2006. vol. 1.
7. Azarov E., Petrovsky A.A., Lobanov B., Tsurulnik L. Text-to-speech system with acoustic processor based on the instantaneous harmonic analysis. SPECOM. 2009. pp. 414 – 418.
8. Kawahara H., Morise M. Technical foundations of tandem-straight, a speech analysis, modification and synthesis framework. SADHANA. Academy Proceedings in Engineering Sciences, 2011. pp. 713–722.
9. Kawahara H., Katayose H., Cheveigne A. Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity. Proc. Eurospeech'99. 1999. pp. 2781–2784.
10. Kawahara H., Masuda I., Cheveigne A. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous–frequency–based F0 extraction. Speech Communication. 2004. pp. 187–207.
11. Arslan L. Speaker transformation algorithm using segmental codebooks. Speech Communication. 1999. vol. 28, no. 3. pp. 211–226.
12. Narendranath M., Murthy H., Rajendran S., Yegnanarayana N. Transformation of formants for voice conversion using artificial neural networks. Speech Communication. 1995. vol. 16, no. 2. pp. 207–216.
13. Rabiner L.R. [Hidden Markov Models and their applications in speech recognition]. *Trudy instituta inzhenerov po jelektronike i radiojelektronike – Proceedings of the Institute of Engineers on Electronics*. 1989. vol. 77, no. 2. pp. 86–120. (In Russ.)
14. Azarov I.S., Petrovskij A.A. [Real-time voice conversion system with text independent learning based on the hybrid parametric descriptions of speech signals]. *Cifrovaja obrabotka signalov – Digital Signal Processing*. 2012. no. 2. pp. 15–23. (In Russ.).
15. *Analizatory rechevyyh i zvukovyh signalov: metody, algoritmy i praktika (s MATLAB primerami)* [Analyzers of speech and audio signals: methods, algorithms and practice (with MATLAB examples)]. edited by A.A. Petrovskii. Minsk: Bestprint, 2009. 456 p. (In Russ.).
16. Stylianou, Y., Cappe O., Moulines E. Statistical methods for voice quality transformation. Proc. of European Conference on Speech Communication and Technology. Madrid, 1995. pp. 447–450.
17. Pavlovec A.S., Livshic M.Z., Lichachev D. S., [Voice conversion using a model of the speech with signal separation into "harmonic + noise" components and transitional frames]. *Rechevye tehnologii – Speech Technologies*. 2008. no. 4. pp. 37–50.

18. Stylianou Y., Cappe O., Moulines E. Continuous probabilistic transform for voice conversion. Proc. of International Conference on Acoustics, Speech and Signal Processing. 1998. pp. 2451–2455.
19. Patterson D. A linguistic approach to pitch range modeling. PhD dissertation. Scotland: University of Edinburgh. 2000. 201 p.
20. Zahar'ev V.A., Petrovskij A.A. Konversija prosodicheskikh karakteristik diktora na osnove metodov parametrizacii kontura chastoty osnovnogo tona [Conversion of prosodic speaker features based on parameterization of the pitch contour]. *Doklady BGUIR – Proceedings – Reports of the BSUIR*. Minsk, 2013. pp. 39–46. (In Russ.).
21. Hu Y., Loizou P.C. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech & Language Processing*. 2008. vol. 16, no. 1. pp. 229–238.

**Захарьев Вадим Анатольевич** — аспирант кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники (БГУИР). Область научных интересов: цифровая обработка сигналов, методы машинного обучения, конверсия голоса, синтез и распознавание речи. Число научных публикаций — 3. zahariev@bsuir.by; БГУИР, П. Бровки, д. 6, г. Минск, 220013, РБ; р.т. +37(517) 293-8946.

**Zahariev Vadim Anatol'evich** — Ph.D. student of Computer Engineering Department of the Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: digital signal processing, machine learning, voice conversion, speech synthesis and recognition. The number of publications — 3. zahariev@bsuir.by; BSUIR, P. Brovkiy 6, Minsk, RB; office phone +37(517) 293-8946.

**Петровский Александр Александрович** — д-р техн. наук., профессор, заведующий кафедрой электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники (БГУИР). Область научных интересов: цифровая обработка сигналов, системы кодирования аудио и видео данных, мультимедийные системы реального времени для мультимедиа приложений. Число научных публикаций — более 600. palex@bsuir.by; БГУИР, П. Бровки, д. 6, г. Минск, 220013, РБ; р.т. +37(517) 293-8946.

**Petrovsky Aleksandr Aleksandrovich** — Ph.D., Dr. Sci., professor, head of Computer Engineering Department of the Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: digital signal processing, speech and video coding systems, multi-processor real-time systems for multimedia applications. The number of publications — more than 600. palex@bsuir.by; BSUIR, P. Brovkiy 6, Minsk, RB; office phone +37(517) 293-8946.

**Лобанов Борис Мефодьевич** — д-р техн. наук., профессор, главный научный сотрудник лаборатории распознавания и синтеза речи Объединённого института проблем информатики Национальной академии наук Республики Беларусь (ОИПИ НАН РБ). Область научных интересов: методы автоматического анализа, синтеза и распознавания речевых сигналов, человеко-машинные системы речевого общения. Число научных публикаций — 350. lobanov@newman.bas-net.by; ОИПИ НАН РБ, Сурганова 6, г. Минск, 220012, РБ; р.т. +37(517) 284-2773, факс +37(517) 331-8403.

**Lobanov Boris Mefod'evich** — Ph.D., Dr. Sci., professor, chief researcher of the United Institute of Informatics Problems National Academy of Sciences of Belarus (UIIP NAS RB). Research interests: methods of automatic analysis, synthesis and speech recognition, human-machine speech communication systems. The number of publications — 350. lobanov@newman.bas-net.by; UIIP NAS RB, Surganova 6, Minsk, 220012, RB, office phone . +37(517) 284-2773, fax +37(517) 331-8403.

## РЕФЕРАТ

### *Захарьев В.А., Петровский А.А., Лобанов Б.М.* Система синтеза речи по тексту с возможностью настройки на голос целевого диктора.

В данной статье предлагается один из подходов построения систем синтеза речи по тексту, обладающих возможностью настройки на голос целевого диктора, и получивших название мультиголосовых, с применением методов и моделей, использующихся в речевых системах другого класса — конверсии голоса. Система конверсии голоса позволяют автоматически выполнять процесс преобразования параметров голоса, характеризующих речь исходного диктора, в параметры целевого. Объектами конверсии голоса, как технологии обработки сигналов, являются стабильные во времени свойства говорящего, проявляющиеся в речевом сигнале через изменение его акустических параметров. Объединение данных двух типов систем позволяет в контексте их взаимодействия эффективно решать задачу быстрой настройки системы синтеза на произвольного диктора, без внесения существенных изменений в существующую систему синтеза, или длительного и трудозатратного процесса подготовки речевой базы для нового диктора.

Представлена архитектура системы на базе интерактивного взаимодействия между двумя типами систем, которая позволяет получить определённый синергетический эффект от их интеграции. Основными особенностями предлагаемого типа архитектуры являются: выполнение операций конкатенации и реконструкции синтезированного речевого сигнала после конверсии голоса исходного диктора в голос целевого, аспекты конверсии голоса учитываются при выборе единиц компиляции, все алгоритмы преобразования и конверсии (спектральные и просодические) выполняются единым блоком, это означает, что характеристики сигнала модифицируются только один раз.

В статье рассмотрены нюансы выбранной модели представления и параметризации речевого сигнала, понимание принципов работы которой являются ключевым момент для успешной реализации этапов анализа, конверсии и восстановления результирующего речевого сигнала со свойствами целевого диктора. В нашем случае за основу была взята модель вокодерного типа на основе взвешенной интерполяции спектра — STRAIGHT. Она позволяет разложить сигнал на три относительно независимые параметризуемые компоненты: контур ЧОТ, а также периодическую и аperiodическую составляющую сигнала, и предоставляет большие возможности по гибкому изменению параметров речевого сигнала в широком диапазоне.

Также в статье были представлены расширенная регрессионная функция конверсии, применяемая для повышения точности преобразования векторов параметров, и методика постобработки сигнала на основе спектрального взвешивания — для решения проблемы чрезмерного усреднения спектральной огибающей. Всё это в итоге повышает качественные характеристики результирующего восстановленного сигнала, сохраняя вместе с тем высокие уровни узнаваемости голоса целевого диктора для преобразованной речи.

## SUMMARY

### *Zahariev V.A., Petrovsky A.A., Lobanov B.M. Text to speech synthesis system with the target speaker voice customization capability.*

An approach of constructing of the text to speech synthesis systems is proposed in this article. Such synthesizers have the abilities to customize the original system voice on the target speaker, using the methods and models of voice conversion technology. Voice conversion system can automatically perform the process of converting voice parameters characterizing the source speaker speech to the target speaker parameters. Voice conversion sites as signal processing technology, works with a stable over time speaker properties, manifested in the speech signal through a changes in its acoustic parameters. Combining these two types of systems we can effectively solve the problem of tuning the synthesizer on the arbitrary speaker without significant changes in the existing synthesis system, or a long and labor-intensive process of preparation of the speech database for a new speaker .

The architecture of the multivoice synthesizer based on the interactivity between the two types of systems, which allows us to get some synergy effect from such integration was presented. The main features of the proposed type of architecture are: implementation of concatenation operations and reconstruction of the synthesized speech signal after the conversion of the original speaker's voice into the voice of the title, aspects of the voice conversion are considered when choosing compilation units, all conversion and transformation algorithms ( spectral and prosodic ) are performed in a single block is means that the signal characteristic is modified only once.

In the article the nuances of chosen the representation and parameterization model of the speech signal were discussed. Understanding of these principles is a key to successful implementation of analysis, conversion stages and recovery of the resulting speech signal with a voice features of the target speaker. In our case speech model was based on the vocoder-type method known as a speech transformation and representation by adaptive interpolation of weighted spectrum — STRAIGHT. It allows us to decompose signal into three relatively independent parameterized components: pitch contour, maps of periodic and aperiodic component of the signal, and provides a great opportunity for flexible changing of the speech signal characteristics in a wide range.

Also in the article the extended regression conversion function that is used to improve the accuracy of the transformation parameter vectors, and a signal post-processing method based on spectral weighing to solve the problem of excessive averaging spectral envelope, were presented. All of this ultimately increases the quality characteristics of the resulting reconstructed signal, maintaining at the same time high levels of similarity of the target speaker voice for the transformed speech.

Е.Е. ЛЯКСО, В.М. СИТДИКОВ, А.С. ГРИГОРЬЕВ, Е.А. ОГОРОДНИКОВА  
**ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ МОДЕЛИ «INFANT.MAVS»  
ДЛЯ ИЗУЧЕНИЯ ПСИХОФИЗИОЛОГИЧЕСКОГО  
СОСТОЯНИЯ ДЕТЕЙ ПЕРВОГО ГОДА ЖИЗНИ**

---

*Ляксо Е.Е., Ситдигов В.М., Григорьев А.С., Огородникова Е.А.* Программное обеспечение модели «INFANT.MAVS» для изучения психофизиологического состояния детей первого года жизни.

**Аннотация.** Создана модель мультимодальной сенсорной среды «INFANT.MAVS», включающая две базы стимулов разной перцептивной сложности – простых (визуальных, звуковых, тактильных и графических) и комплексных, синтезированных на основе сочетания простых. Программное обеспечение модели включает компонент управления созданными базами данных и саму базу данных. Компонент управления разработан на языке Microsoft Visual Basic v.6.0 и предназначен для работы под управлением операционных систем семейства MS Windows.

**Ключевые слова:** модель мультимодальной сенсорной среды, программное обеспечение, психофизиологическое состояние ребенка, раннее когнитивное развитие.

*Lyakso E., Sitdikov V., Grigorev A., Ogorodnikova E.* Software model «INFANT.MAVS» for studying psychophysiological state of infants.

**Abstract.** A model multimodal sensory environment «INFANT.MAVS», comprising two base stimuli of different perceptual complexity - simple (visual, audible, tactile and graphic) and base of complex stimuli synthesized on the basis of a combination of simple, is elaborated. The software includes a management component model created database and the database itself. Management component is created with Microsoft Visual Basic v.6.0 and is designed to run on operating systems of MS Windows.

**Keywords:** A model multimodal sensory environment, software, child psychophysiological state, early cognitive development.

---

**1. Введение.** В последние десятилетия изучению вопроса мультимодального взаимодействия человека и компьютера [3, 11] отводится много внимания. На материале русского языка создана модель синтеза аудиовизуальной речи («говорящая голова»), произведена разработка модели синхронизации речевой и визуальной модальностей речи для данной модели [4]. Разработана информационная система для исследования и развития слухоречевой функции у детей с нарушениями слуха и речи, внедренная в клиническую (реабилитация пациентов после операции кохлеарной имплантации) и коррекционную практику [7, 1].

Однако, применительно к обучению детей раннего возраста, созданию у них определенных состояний и условий сенсорно-когнитивного взаимодействия с окружающим миром, такие системы являются единичными. Большее распространение получили детские социальные сети, рассчитанные на детей младшего школьного возраста.

та. Их главной особенностью является преимущественная направленность на развлечение и обучение маленького пользователя через игру; родительский контроль над деятельностью ребенка в сети.

В помощь родителям широко применяются коммерческие программы «Видеоняня-Аудионяня». Это специальные устройства, с помощью которых можно наблюдать за своим ребенком («Видеоняня») и/или слышать его голос в любом месте помещения («Аудионяня»). В практику внедрена система «Why Cry Baby Analyzer HC-WHYCRY» [12]. Она оценивает состояние ребенка на основании анализа его плача длительностью не менее 20 с и позволяет выделить состояние голода, боли, сонливости, стресса у младенца. Система разработана в помощь родителям и не может быть использована в медицинских целях для тонкой дифференциации патологического состояния. Эти разработки носят скорее вспомогательный характер, чем обучающий. Полученные нами данные об особенностях построения вокально-речевого взаимодействия в диадах «мать-ребенок» и о специфике долингвального этапа развития детей в доме ребенка [5, 6], легли в основу построения модели «Виртуальной матери» [9], предназначенной для детей, воспитывающихся в условиях домов ребенка, и детей с ограниченными возможностями. Обучающая компьютерная программа позволяет стимулировать ребенка к большей звуковой активности в первом полугодии жизни; что приводит во втором полугодии к качественному усложнению вокализаций: расширению звукового репертуара, появлению слоговых структур и лепетных конструкций, обеспечивая переход на последующую стадию речевого развития – появлению первых слов. Однако отсутствует специализированное программное обеспечение для создания адекватных условий для развития сенсомоторных и когнитивных способностей младенцев.

Целью настоящей работы явилось создание модели мультимедийной сенсорной среды с элементами интерактивности для детей раннего возраста и разработка программного обеспечения для работы с моделью.

Модель предназначена для предотвращения и снятия негативных последствий сенсорно-коммуникативной депривации и нормализации психофизиологического состояния младенцев в условиях длительного отсутствия контакта с матерью, ограничения социального общения.

Задачи исследования заключались: в подборе стимулов разных модальностей и создании базы данных стимульного материала, дифференцированного по степени сложности; в разработке программного обеспечения базы; в тестировании программы с целью определения

эффекта, вызываемого предъявлением стимулов разной модальности из созданной базы данных.

**2. Подбор и организация стимульного материала.** Создана модель мультимодальной сенсорной среды «INFANT.MAVS», включающая базу стимулов разной перцептивной сложности (рис. 1) и программное обеспечение для работы с ней.



Рис. 1. Программа «INFANT.MAVS»:  
окно загрузки программы (сверху); заставка программы (снизу)

База состоит из двух частей – простых стимулов (БПС) и сложных стимулов (БСС). БПС содержит 1380 файлов (объем – 1,47 Гб), заключенных в разделы визуальных, звуковых, тактильных и графических стимулов (рис. 2).

БПС содержит каталоги визуальных (915 файлов; 732,3 Мб), звуковых (401 файл; 533,4 Мб) и тактильных (64 файлов; 229,6 Мб) стимулов.

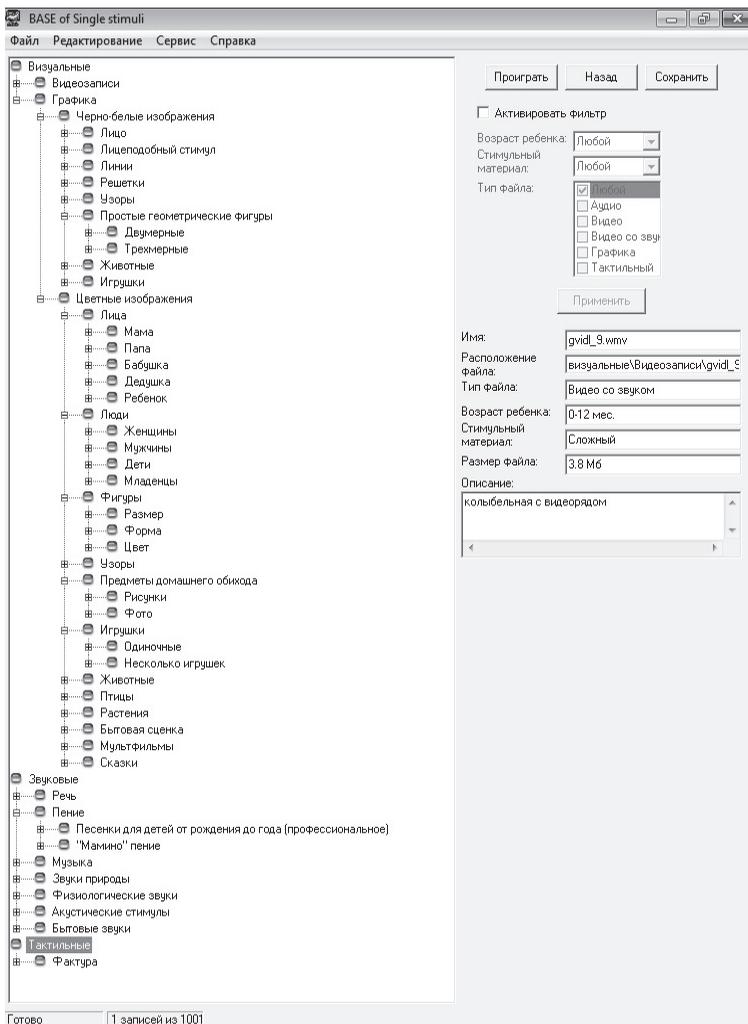


Рис. 2. Окно со структурой базы стимулов

Раздел «Визуальные стимулы» содержит подразделы видеозаписей и графики. Подраздел «Графика» включает черно-белые и цветные изображения. Каталог «Черно-белые изображения» - это лица и лицеподобные стимулы, содержащие все элементы (глаза, нос, рот, волосы), лицеподобные стимулы с тремя и/или двумя элементами, представленными в разных комбинациях. В этот подраздел входят ли-

нии разной толщины и ориентации (вертикальные, горизонтальные, наклонные); решетки; узоры - простые и сложные (состоящие из набора простых узоров); геометрические фигуры - простые, двумерные и трехмерные; изображения животных и игрушек. В каталог «Цветные изображения» входят фотографии и рисунки людей, животных, птиц, игрушек, растений, геометрических фигур, предметов домашнего обихода и бытовых сценок. В нем представлены мультфильмы и иллюстрации сказок.

Раздел «Звуковые стимулы» включает подразделы музыки и речи. В музыкальный подраздел входят песенки и музыкальные мелодии, «мамино» пение и колыбельные, звуки природы; физиологические и бытовые звуки; акустические стимулы. Подраздел речь содержит вокальные конструкции комфортного состояния младенцев; успокаивающие, привлекающие внимание и стимулирующие к имитации образцы «материнской и отцовской речи»; потешки и стихи. Раздел «Тактильные стимулы» - представляют фотографии поверхностей и текстур.

Используется буквенное обозначение, соответствующее типу стимула, с последующим подразделением на подтипы, в соответствии с разбиением, используемом в каталоге стимулов. Модальность стимула: а - аудио, v - видео, g – графика. Аудио стимулы: sp - речь: ms - материнская речь (с – успокаивающие, at - привлекающие внимание, i – звуки, стимулирующие к имитации, r – стихи). Цифрой указывается номер записи: например, aspmsi\_1 - аудиостимул 1, материнская речь, стимулирующий к имитации.

БСС сложных стимулов содержит комплексы разномодальных стимулов (177 сложных стимулов, 15 – комплексных), синтезированных путем использования простых стимулов и служащих основой для создания аудиодорожки и/или видеодорожки.

Видеоряд может быть представлен последовательностью видеороликов, статичных картинок и анимаций заданной продолжительности. Комплексные стимулы синтезируются на основе сложных стимулов. Звуковые и речевые стимулы представлены в формате - \*.WAV, музыка - \*.MP3; видео – \*.MPG.

Организация базы сложных стимулов материала подразумевает хранение уже готовых комбинаций стимулов и возможность создания новых стимульных комбинаций пользователем (рис. 3).

Стимульный материал подобран в соответствии с перцептивными возможностями младенцев. Так, раздел «Стимулы для детей от 0 до 6 месяцев» включает подразделы – лица, черно-белые изображения (аудиографические и видео), колыбельные (смена картинок с соот-

ветствующим музыкальным сопровождением). Раздел «Стимулы для детей от 6 до 12 месяцев» содержит все простые стимулы разных модальностей в разнообразных комбинациях. Специальный раздел «Для пользователя» включает подраздел «Для взрослых пользователей», предназначенный для релаксации, снятия усталости и формирования положительного настроения у взрослого персонала, работающего с детьми, и у родителей.

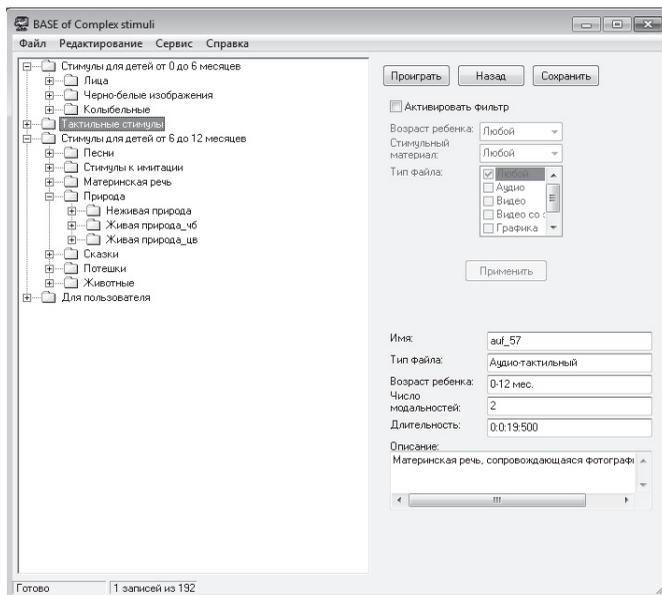


Рис. 3. Окно программы базы сложных стимулов с основными каталогами

**3. Разработанное программное обеспечение.** Для работы с базой стимульного материала разработана программа, позволяющая осуществлять: конструирование стимульного материала; систематизацию хранения и выбор стимульного и исходного мультимедийного материала; воспроизведение и редактирование стимульного и исходного мультимедийного материала.

Данный программный комплекс включает в себя компонент управления базами данных и саму базу данных. Базу данных стимульного материала можно разделить на две большие относительно самостоятельные (с возможностью взаимодействия друг с другом) части. Одна часть - база данных простых стимулов. Она служит для упорядоченного хранения и выборки медиа-данных, служащих основой для

конструирования сложных (составных) стимулов. Сложные стимулы находятся во второй части базы данных. Они представляют собой комбинации медиа-материалов, предъявляемых пользователю в определенном порядке и заданным способом.

Компонент управления программы разработан на языке Microsoft Visual Basic v.6.0 и предназначен для работы под управлением операционных систем семейства Windows (9x, NT, ME, 2000, XP, Vista, 7). Используются внешние оболочки: Microsoft Visual Basic Runtime (интегрирован в пакет установки); видео и аудио кодеки, установленные в операционную систему пользователя, позволяющие работать со следующими стандартами сжатия: для видео - MPEG (2,4), AVC, H.265; для звука - MP3 (MPEG-1 Layer I, II, III), AAC, WMA (рис. 4). Интерфейс программы графический.

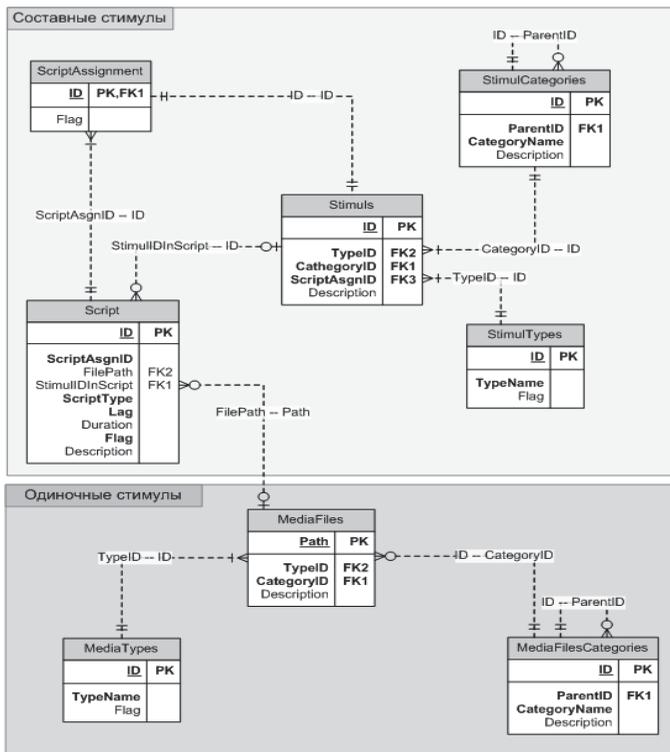


Рис. 4. Схема базы данных

Описание полей базы данных представлено в таблицах 1-8.

Таблица 1. Информация о сложных стимулах (Stimulus)

Имя поля	Описание
ID	идентификатор, ключевое поле
Type ID	тип стимула, хранит ссылку на таблицу Stimul Types (типы стимулов)
Category ID	категория, хранит ссылку на таблицу Stimul Categories (категории сложных стимулов)
Script Asgn ID	ссылка на таблицу Script Assignment
Description	Описание

Таблица 2. Типы сложных стимулов (Stimul Types)

Имя поля	Описание
ID	идентификатор, ключевое поле
Type Name	имя типа
Flag	константа, определяющая тип в программе

Таблица 3. Структура категорий стимулов (Stimul Categories)

Имя поля	Описание
ID	идентификатор, ключевое поле
Parent ID	ссылка на значение поля ID, указывающая на родительскую категорию в иерархической структуре
Category Name	имя категории
Description	Описание

Таблица 4. Информация о структуре стимула, в которой отражен порядок воспроизведения медиа-файлов (Script)

Имя поля	Описание
ID	идентификатор, ключевое поле
File Path	путь к файлу, ссылка на простой стимул из таблицы Media Files
StimulIDInScript	ссылка на сложный стимул из таблицы Stimulus
Script Asgn ID	ссылка на таблицу Script Assignment
Script Type	указание на тип стимула: простой или сложный
Lag	задержка относительно предыдущего элемента
Duration	длительность воспроизведения
Flag	флаг, содержащий дополнительные параметры воспроизведения

Таблица 5. Информация о параметрах воспроизведения стимула (Script Assignment)

Имя поля	Описание
ID	идентификатор, ключевое поле, на которое ссылается таблица Script
Flag	флаг, содержащий дополнительные параметры воспроизведения

Таблица 6. Информация о простых (одиночных) стимулах (Media Files)

Имя поля	Описание
ID	идентификатор, ключевое поле
Type ID	тип стимула, хранит ссылку на таблицу MediaFileTypes (типы медиа-файлов)
Category ID	категория, хранит ссылку на таблицу MediaFileCategories (категории простых стимулов)
Description	Описание

Таблица 7. Информация о типах простых (одиночных) стимулов (MediaFileTypes)

Имя поля	Описание
ID	идентификатор, ключевое поле
Type Name	имя типа
Flag	константа, определяющая тип в программе

Таблица 8. Информация о структуре категорий (MediaFileCategories)

Имя поля	Описание
ID	идентификатор, ключевое поле
Parent ID	ссылка на значение поля ID, указывающая на родительскую категорию в иерархической структуре
Category Name	имя категории
Description	Описание

Главное окно программы предоставляет пользователю возможность перехода к интересующей базе данных. Слева располагается дерево категорий, справа вверху – панель управления, на которой находятся кнопка воспроизведения, кнопка перехода в главное меню, кнопка сохранения изменений и настройки фильтра (рис. 5).

Дерево категорий позволяет осуществлять эффективную навигацию по всей базе данных библиотеки. В программе предусмотрена возможность редактирования категорий, которая производится с помощью контекстного меню и перетаскиванием. Контекстное меню служит и для вызова окна редактирования стимулов.

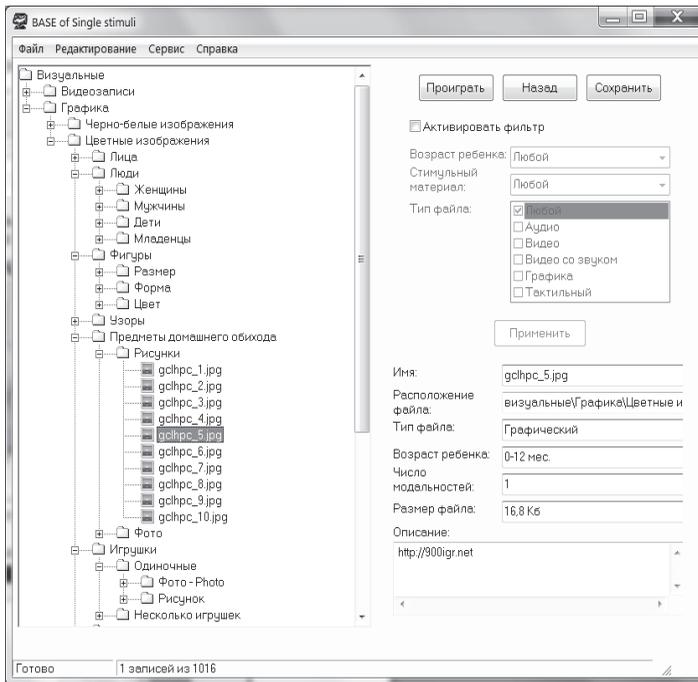


Рис. 5. Диалоговое окно для работы с базой данных

Созданная модель "INFANT.MAVS" представляет собой программный продукт, готовый к установке на персональных компьютерах.

**4. Тестирование.** Тестирование созданной модели проведено на 22 детях (от 1,5 мес до 7 лет), растущих в домашних условиях и развивающихся в соответствии с нормой, и на 84 взрослых (от 19 до 84 лет) испытуемых. Информированное согласие на проведение исследования утверждено Этическим комитетом СПбГУ (№ 02-36 от 16.01.14).

Процедура тестирования: детям с экрана компьютера предъявляли комбинированные стимулы с использованием программы «INFANT.MAVS». Дети находились в состоянии активного бодрствования. Дети в возрасте 0-6 мес - лежали в кроватке, монитор располагался перед лицом младенца на расстоянии около 25 см; дети от 6 до 12 мес сидели на руках у матери перед монитором. Дети 1-7 лет – располагались перед монитором без взрослых. Поведение детей при тестировании записывали на две видеокамеры: одна - фиксировала реакции ребенка и предъявляемые с монитора стимулы, вторая - была сфо-

кусирована только на ребенке. Взрослые испытуемые просматривали стимульный материал, предъявляемый с мультимедийной установки. До начала тестирования испытуемые отмечали в анкете свое состояние и настроение, после просмотра - вызванные ощущения.

Предъявляемый детям стимульный материал условно объединен в три группы: изображения лиц с «материнской речью» (стим\_1); колебательные с разными изображениями (стим\_2), потешки и сказки (стим\_3). Общая длительность предъявления последовательностей стимулов для детей в возрасте 0-6 мес составила – 1-2 мин; для детей 6-12 мес – 2-5 мин; 1-3 года – 2-7 мин; 4-7 лет – 2-10 мин.

Взрослым предъявляли два теста из комбинированных стимулов «Цветы и травы» (5 мин 17 с) и «Настроение» (3 мин 40 с).

Анализ видеозаписей проводили в программе Pinnacle Video Studio. Статистическую обработку данных проводили в программе «SPSS v. 20» «Statistica 8» с использованием критерия Манна-Уитни.

Выделены следующие варианты реагирования ребенка на предъявляемые стимулы (рис. 6): 1. Взгляд к монитору; 2. Взгляд от монитора; 3. Улыбка; 4. Плач, звуки дискомфорта и/или соответствующая мимика. 5. Комфортные звуки; 6. Движение в сторону экрана; 7. Отворачивается от экрана (отвлекается) 8. Засыпает или зевает.

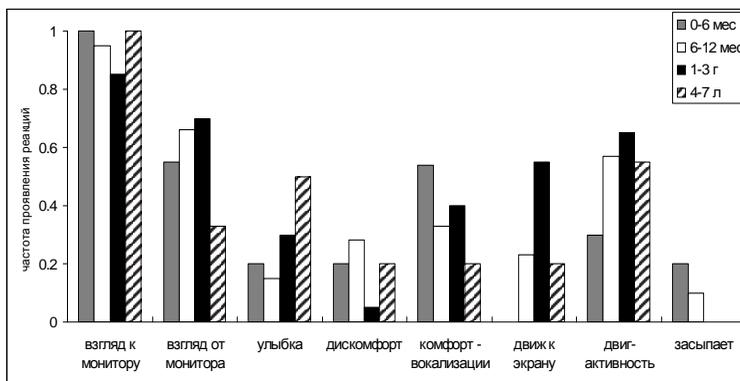


Рис. 6. Частота проявления выделенных реакций у всех детей на все предъявляемые стимулы (по вертикали – частота проявления реакции, по горизонтали – реакция)

Наиболее частыми реакциями у детей всех возрастных групп явились направление взгляда в сторону монитора и от монитора, комфортные вокализации и общая двигательная активность.

У детей первого полугодия жизни предъявление черно-белых изображений лиц детей и взрослых (без звука) вызывало улыбку. От-

ветные комфортные вокализации зарегистрированы на предъявляемые потешки в сочетании с «материнской речью». Дети второго полугодия жизни вокализировали при предъявлении им стимулов, содержащих видео с «материнской речью» и сюжетов сказок под музыку. Движения туловища в сторону экрана, взмахи или хлопки руками, вставание на ноги и подпрыгивание зарегистрированы при прослушивании стимулов с картинками и музыкой. Колыбельные песни вызывали у младенцев снижение двигательной активности, закрывание глаз и зевоту. При предъявлении аудио-последовательностей дети поворачивались к источнику звука, улыбались и вокализировали.

Дети от года до трех лет проявляли большой диапазон разнообразных реакций. На предъявляемые музыкальные последовательности у всех детей зарегистрированы танцевальные движения (до 3-5 мин), цветных картинок животных – произнесение звуков, имитирующих голоса животных. Дети старшей возрастной группы имитировали звуки животных, подпевали, на колыбельные – зевали и закрывали глаза.

Анализ анкет взрослых испытуемых показал, что во всех группах испытуемых при просмотре двух тестов (таблица 9) преобладают позитивные ощущения. Более сложные состояния описаны информантами третьей группы, у которых в сочетании с позитивными ощущениями проявляются негативные. Для испытуемых старшей возрастной группы характерны лаконичные позитивные ответы.

Таблица 9. Описание ощущений взрослых испытуемых, вызванных просмотром двух тестов

№ гр.	Ощущения испытуемого											
	Р	Э	Р+Э	К	К+Г	К+Т	Х	Н	Ра	К+Ра	Т	Х+Ра
Тест «Цветы и травы»												
1	52		8	28			4	2	6			
2	28	4	8	32			28					
3	34		7	10	3	7	7	3	3	10	3	13
4	10		47	43								
Тест «Настроение»												
1	32	9	5.5	9	4.5		23		9	8		
4	45		35	20								

В таблице 9 использованы следующие обозначения: Р - "релаксация", Э - "энергия", Р+Э - "релаксация + энергия", К – "красота", К+Г – "красота + грусть", К+Т – "красота + тревога", Х – "хорошее", Н – "нейтральное", Ра – "раздражение", К+Ра – "красота + раздражение", Т - "тревога", Х+Ра – "хорошо + раздражение". Группы испытуемых: 1 гр. - 19,5±0,5 лет (n=22); 2 гр. - 29,9±0,7 лет (n=25); 3 гр. - 30,1±10 лет (n=30); 4 гр. - 75±8,3 лет (n=7).

**5. Заключение.** Результаты проведенного тестирования показали, что сформированная база стимульного материала соответствует целям, для которых она создавалась. У детей, предъявляемые стимулы вызывали ответные реакции в виде сосредоточения внимания, ответной вокализации, улыбки, имитационной активности. Представляется очень важной возможностью вызова у детей разнообразных реакций не при взаимодействии с взрослым, а на стимульный материал, что позволяет создавать не просто обогащенную сенсорную среду для ребенка, но и получать на нее адекватный отклик. Ранняя стимуляция является необходимым условием дальнейшего нормального когнитивного развития и социальной адаптации детей. К примеру, известно, что способность ребенка к мимической и звуковой имитации является предпосылкой для развития коммуникативных способностей [10]. В соответствии с этими данными, предполагается использование модели в научных исследованиях раннего психофизиологического развития детей, воспитывающихся в условиях семьи и дома ребенка.

Важное направление, поддерживаемое моделью, связано с блоком оригинальных стимулов «Для взрослых пользователей», предназначенных для взрослых, работающих с детьми. Тестирование, проведенное на взрослых испытуемых, показало, что предъявляемые стимулы вызывают у них позитивные ощущения отдыха, релаксации, чувства прекрасного. Мы полагаем, что использование данного блока модели поможет избежать или снизить вероятность профессионального выгорания персонала [2], и будет полезно матерям с неустойчивым психоэмоциональным состоянием [8] или страдающим депрессией.

### **Литература**

1. *Белова Н.Ю., Ермакова А.Г., Огородникова Е.А., Люблинская В.В., Пак С.П., Охарева Н.Г., Королева И.В.* Использование компьютерного тренажера «Учись слушать» для развития слухоречевого восприятия детей с нарушениями слуха и речи в условиях образовательного учреждения // Российская оториноларингология. 2013. №3. С. 15-23.
2. *Борисова М.В.* Психологические детерминанты феномена эмоционального выгорания у педагогов // Вопросы психологии. 2005. №2. С. 96-105.
3. *Карпов А.А., Ронжин А.Л.* Многомодальные интерфейсы в автоматизированных системах управления // Известия высших учебных заведений. Приборостроение. 2005. Т.48. №7. С. 9-14.
4. *Карпов А.А., Цирульник Л.И., Железны М.* Разработка компьютерной системы «Говорящая голова» для аудиовизуального синтеза русской речи по тексту // Информационные технологии. 2010. №8. С. 13-18.
5. *Ляко Е.Е.* Вокально-речевое развитие ребенка в первый год жизни // Российский физиологический журнал. 2003. Т.89. №2. С. 207-218.
6. *Ляко Е.Е.* Речевая имитация в диадах «мать-ребенок» с нормально развивающимися детьми и детьми, имеющими неврологические нарушения: лонгитюдное исследование // Сенсорные системы. 2006. Т.20. №3. С. 204-215.
7. *Огородникова Е.А., Королева И.В., Люблинская В.В., Пак С.П.* Компьютерная тренажерная система для реабилитации слухоречевого восприятия у пациентов после опе-

рации кохлеарной имплантации // Российская оториноларингология. 2008. Приложение №1. С. 342-347.

8. Azak S., Raeder S. Trajectories of parenting behavior and maternal depression // *Infant Behavior and Development*. 2013. vol.36. pp. 391-402.
9. Lyakso E., Kurazhova A., Gajkova J., Frolova O., Ostrouhov A., Soloviev A., Bednaya E., Grigoriev A., Losik G., Erchak H. Model "Virtual Mother" for Orphans' Speech Development // *Proceedings of 13-th International Conference «Speech and Computer» (SPECOM'2009)*. SPb. Russia. 2009. pp. 295-299.
10. Poon K.K., Watson L.R., Baranek G.T., Poe M.D. To what extent do joint attention, imitation, and object play behaviors in infancy predict later communication and intellectual functioning in ASD? // *Journal of Autism and Developmental Disorders*. 2012. vol. 42. no. 6. pp. 1064-1074.
11. Ronzhin A., Karpov A. Russian Voice Interface // *Pattern Recognition and Image Analysis*, Pleiades Publishing. 2007. vol. 17. no. 2. pp. 321-336.
12. Why Cry Baby Analyzer HC-WHYCRY URL: [http://www.harriscomm.com/catalog/product\\_info.php?products\\_id=18239](http://www.harriscomm.com/catalog/product_info.php?products_id=18239) (дата обращения: 26.02.2014).

## References

1. Belova N., Ermakova A., Ogorodnikova E., Lublinskaya V., Pak S., Ohareva N., Koroleva I. [Application of computer-assisted training system «Learn to hear» for development of childish auditory perception in educational institution]. *Rossiyskaya otorinolaringologiya – Russian Otorhinolaryngology*. 2013. no. 3. С. 15-23. (In Russ.).
2. Borisova M.V. [Psychological determinants of emotional burnout among teachers]. *Voprosy psichologii – Questions of psychology*. 2005. no. 2. С. 96-105. (In Russ.).
3. Karpov A.A., Ronzhin A.L. [Multimodal interfaces in automatic control systems]. *Izv. vyssh. uchebn. zavedenij: Priborostroenie – Proceedings of the higher educational institutions: Instrumentation*. 2005. vol. 48. no. 7. С. 9-14. (In Russ.).
4. Karpov A.A., Tsirul'nik L.L., Zhelezny M. [Elaboration of the computer system «Speaking head» for audiovisual synthesis of Russian speech by text]. *Informacionnye tehnologii – Information Technologies*. 2010. no. 8. С. 13-18. (In Russ.).
5. Lyakso E.E. [Vocal-speech development of children during the first year of life]. *Rossiyskij fiziologicheskij zhurnal – Russian Journal of Physiology*. 2003. vol. 89. no. 2. С. 207-218. (In Russ.).
6. Lyakso E.E. [Vocal imitation in mother-child dyads with normally development children and children with neurological disorders: longitudinal study]. *Sensornye Sistemy – Sensory Systems*. 2006. vol. 20. no. 3. С. 204-215. (In Russ.).
7. Ogorodnikova E.A., Koroleva I.V., Lublinskaya V.V., Pak S.P. [Computer-assisted training system for rehabilitation of patients after cochlear implantation]. *Rossiyskaya otorinolaringologiya – Russian Otorhinolaryngology*. 2008. Suppl. no. 1. pp. 342-347. (In Russ.).
8. Azak S., Raeder S. Trajectories of parenting behavior and maternal depression. *Infant Behavior and Development*. 2013. vol. 36. pp. 391-402.
9. Lyakso E., Kurazhova A., Gajkova J., Frolova O., Ostrouhov A., Soloviev A., Bednaya E., Grigoriev A., Losik G., Erchak H. Model "Virtual Mother" for Orphans' Speech Development. *Proceedings of 13-th International Conference «Speech and Computer» (SPECOM'2009)*. SPb. Russia. 2009. pp. 295-299.
10. Poon K.K., Watson L.R., Baranek G.T., Poe M.D. To what extent do joint attention, imitation, and object play behaviors in infancy predict later communication and intellectual functioning in ASD?. *Journal of Autism and Developmental Disorders*. 2012. vol. 42. no. 6. pp. 1064-1074.
11. Ronzhin A., Karpov A. Russian Voice Interface. *Pattern Recognition and Image Analysis*, Pleiades Publishing. 2007. vol. 17. no 2. pp. 321-336.
12. Why Cry Baby Analyzer HC-WHYCRY Available at: [http://www.harriscomm.com/catalog/product\\_info.php?products\\_id=18239](http://www.harriscomm.com/catalog/product_info.php?products_id=18239) (accessed: 26.02.2014).

**Ляксо Елена Евгеньевна** – д-р биол. наук, профессор кафедры высшей нервной деятельности и психофизиологии Биологического факультета СПбГУ, руководитель группы по изучению детской речи. Область научных интересов – акустика детской речи, возрастная психофизиология, психолингвистика. Число научных публикаций - 210. [http://bio.spbu.ru/staff/id99\\_eel.php](http://bio.spbu.ru/staff/id99_eel.php). СПбГУ, Университетская наб. 7- 9, Биологический ф-т, 199034. Санкт-Петербург, +7(812)321-33-61, [lyakso@gmail.com](mailto:lyakso@gmail.com)

**Lyakso Elena Evgenievna** – Ph.D., Dr. Sci., professor, department of higher nervous activity and psychophysiology Biology faculty SPbGU, head of child speech research group. Scientific interest – acoustic of child speech, age psychophysiology, psycholinguistic. The number of publications – 210. [http://bio.spbu.ru/staff/id99\\_eel.php](http://bio.spbu.ru/staff/id99_eel.php). SPbGU, University Emb. 7-9, tel: +7(812)321-33-61, [lyakso@gmail.com](mailto:lyakso@gmail.com)

**Ситдиков Владимир Максудович** – магистр, инженер лаборатории психофизиологии речи Института физиологии им. И.П. Павлова РАН. Область научных интересов – информационные системы, программирование, оптоинформатика. Число научных публикаций - 5. Институт физиологии им. И.П. Павлова РАН, наб. Макарова 6, 199034, Санкт-Петербург, +7(812)328-13-01, [sitvlad@mail.ru](mailto:sitvlad@mail.ru)

**Sitdikov Vladimir Maksudovich** – master's degree, engineer, lab. of psychophysiology of speech, Pavlov Institute of Physiology RAS. Scientific interest – information systems, programming, optoinformatics. The number of publications – 5. Pavlov Institute of Physiology RAS, emb. Makarova 6, 199034, St. Petersburg. +7(812)328-13-01, [sitvlad@mail.ru](mailto:sitvlad@mail.ru)

**Григорьев Алексей Сергеевич** – магистр биологии, младший научный сотрудник кафедры высшей нервной деятельности и психофизиологии Биологического факультета СПбГУ. Область научных интересов – акустика детской речи, психофизиология восприятия, психоакустика. Число научных публикаций - 42 . СПбГУ, Университетская наб. 7-9, Биологический ф-т, 199034. Санкт-Петербург, +7(812)321-33-61, e-mail: [a.s.grigoriev89@gmail.com](mailto:a.s.grigoriev89@gmail.com)

**Grigorev Aleksei Sergeevich** - master of biology, junior research, department of higher nervous activity and psychophysiology Biology faculty SPbGU. Scientific interest – acoustic of child speech, psychophysiology of perception, psychoacoustics. The number of publications – 42. SPbGU, University Emb. 7-9, tel: +7(812)321-33-61, e-mail: [a.s.grigoriev89@gmail.com](mailto:a.s.grigoriev89@gmail.com)

**Огородникова Елена Александровна** – к-т биол. наук, зав. лабораторией психофизиологии речи Института физиологии им. И.П. Павлова РАН. Область научных интересов – психофизиология слуха и речи, коррекция нарушений слуха и речи. Число научных публикаций - 197. [http://www.infran.ru/labs/Ogorodnikova\\_r.htm](http://www.infran.ru/labs/Ogorodnikova_r.htm). Институт физиологии им. И.П. Павлова РАН, наб. Макарова 6, 199034, Санкт-Петербург, +7(812)328-13-01, [elena-ogo@mail.ru](mailto:elena-ogo@mail.ru)

**Ogorodnikova Elena Aleksandrovna** – Ph.D., head of lab. of psychophysiology of speech, Pavlov Institute of Physiology RAS. Scientific interest – psychophysiology of hearing and speech, correction of speech and hearing dysfunctions. The number of publications – 197. [http://www.infran.ru/labs/Ogorodnikova\\_r.htm](http://www.infran.ru/labs/Ogorodnikova_r.htm). Pavlov Institute of Physiology RAS, emb. Makarova 6, 199034, St. Petersburg. +7(812)328-13-01, [elena-ogo@mail.ru](mailto:elena-ogo@mail.ru)

**Поддержка исследований.** Работа выполнена при финансовой поддержке РГНФ (проекты № 11-06-12019в, № 13-06-00041а).

**Acknowledgements.** This work was financially supported by RFG (projects number 11-06-12019в, № 13-06-00041а).

## РЕФЕРАТ

### **Ляко Е.Е., Ситдигов В.М., Григорьев А.С., Огородникова Е.А. Программное обеспечение модели «INFANT.MAVS» для изучения психофизиологического состояния детей первого года жизни.**

Создана модель мультимедийной сенсорной среды «INFANT.MAVS», включающая две базы стимулов разной перцептивной сложности – простых стимулов (визуальных, звуковых, тактильных и графических) и сложных, синтезированных на основе сочетания простых. База простых стимулов (1380 файлов, объемом 1,47 Гб) содержит каталоги визуальных (915 файлов; 732,3 Мб), звуковых (401 файл; 533,4 Мб), тактильных (64 файлов; 229,6 Мб) стимулов. База сложных стимулов включает комплексы разномодальных стимулов (177 сложных стимулов, 15 – комплексных) синтезированных путем использования простых стимулов, и служащих основой для создания аудиодорожки и/или видеодорожки. Видео ряд может быть представлен последовательностью видеороликов, статичных картинок и анимаций заданной продолжительности. Комплексные стимулы синтезируются на основе сложных стимулов. Звуковые и речевые стимулы представлены в формате - WAV, музыка - MP3; видео – MPG. Организация базы сложных стимулов материала подразумевает хранение уже готовых комбинаций стимулов и возможность создания новых стимульных комбинаций пользователем. Модель содержит динамическую систему сенсорной стимуляции, соответствующую меняющимся в зависимости от возраста компетенциям ребенка и дополнительный раздел для взрослых пользователей, предназначенный для релаксации, снятия усталости и формирования положительного настроения у взрослого персонала, работающего с детьми. Созданы разделы стимулов для детей от 0 до 6 месяцев; тактильных стимулов; стимулов для детей от 6 до 12 месяцев; раздел для пользователя. Для работы с базой стимульного материала разработана программа, позволяющая осуществлять: 1. Ввод и хранение стимульного материала в заданных каталогах; 2. Выбор стимула в зависимости от задачи (при активации соответствующих фильтров); 3. Просмотр и прослушивание стимульного материала; 4. Создание сложного стимульного материала на основе соединения простых стимулов. Компонент управления программой разработан на языке Microsoft Visual Basic v6.0 и предназначен для работы под управлением операционных систем семейства Windows (9x, NT, ME, 2000, XP, Vista, 7). Используются внешние оболочки: Microsoft Visual Basic Run-time (интегрирован в пакет установки); видео и аудио кодеки, установленные в операционную систему пользователя, позволяющие работать со следующими стандартами сжатия - для видео: MPEG (2,4), AVC, H.265; для звука - MP3 (MPEG-1 Layer I, II, III), AAC, WMA. Интерфейс программы графический.

Произведено тестирование базы «INFANT.MAVS» на 84 взрослых (от 19 до 84 лет) и 22 детях (в возрасте от 1,5 мес до 7 лет). Результаты проведенного тестирования показали, что сформированная база стимульного материала соответствует целям, для которых она создавалась. У детей, предъявляемые стимулы вызвали ответные реакции в виде сосредоточения внимания, ответной вокализации, улыбки, имитационной активности. Наиболее частыми реакциями у детей всех возрастных групп явились направление взгляда в сторону монитора, от монитора, комфортные вокализации и общая двигательная активность. Во всех группах испытуемых при просмотре двух тестов преобладают позитивные ощущения. Для испытуемых старшей возрастной группы характерны лаконичные позитивные ответы. Полученные результаты тестирования подтверждают, что дополнительная сенсорная стимуляция может быть использована для улучшения и нормализации психофизиологического состояния детей и взрослых, работающих с детьми.

## SUMMARY

*Lyakso E.E., Sitdikov V.M., Grigorev A.S., Ogorodnikova EA.* **Software model «INFANT.MAVS» for studying psychophysiological state of infants.**

A model of a multimedia sensory environment «INFANT.MAVS», which includes two bases of perceptual stimuli of different complexity – base of simple stimuli (visual, audio, tactile and graphical) and base of complex stimuli synthesized on the basis of a combination of simple, is elaborated. The base of simple (1380 file, 1.47 Gb) stimuli contains directories (915 files, 732.3 Mb), audio (401 files 533.4 Mb), tactile (64 files, 229.6 Mb) stimulus. The base complex includes stimulus complexes of different modalities of stimulus (177 complex stimuli, 15 - complex) synthesized by the use of simple stimuli, and provide the basis for creating audio and /or video track. Visuals can be represented as a sequence of videos, static images and animations specified duration. Complex stimuli are synthesized on the basis of complex stimuli. Sound and speech stimuli are presented in the format - WAV, music - MP3, video - MPG. The organization of the material bases of complex stimuli involves storage of ready-made combinations of incentives and the ability to create new combinations of stimulus by the user. The model contains a dynamic system of sensory stimulation, the corresponding change depending on the age of the child and competencies of an additional section for adult users, designed for relaxation, fatigue and creating a positive attitude in the adult staff working with children. Forums created incentives for children from 0 to 6 months of tactile stimuli, incentives for children 6 to 12 months; profile for the user. To work with a stimulus material a program was developed that allows users: 1) input and storage of stimulus material in the directories; 2) selection of the stimulus, depending on the problem (for the activation of the filter); 3) viewing and listening to the stimulus material; 4) creating a complex stimulus based on a compound of simple stimuli. The management component of the program is developed in Microsoft Visual Basic v6. 0 and is designed to run under the operating systems of MS Windows. Outer shells are used: Microsoft Visual Basic Run-time (integrated in the installation package), video and audio codec installed on the user's operating system, allowing to work with the following compression standards: video - MPEG (2,4), AVC, H.265, for the sound - MP3 (MPEG-1 Layer I, II, III), AAC, WMA. The program interface is graphical.

To developed framework «INFANT.MAVS» was tested by 84 adults (19 to 84 years) and 22 children (aged 1.5 months to 7 years). The results of the testing showed that the base formed of the stimulus material corresponds to the purposes for which it was created. The children presented stimuli evoked responses in the form of focus, the response vocalizations, smiles, a simulation activity. The most common reactions in children of all age groups were the direction of gaze in the direction of the monitor, from the monitor, comfortable vocalization and general motor activity. In all groups of subjects when viewing the two tests the positive feelings is dominated. The tests of older age group are characterized by concise, positive answers. The obtained test results confirm that the additional sensory stimulation can be used to improve and normalize the psychophysiological state of the children and adults working with children.

Е.Ю. КОСТЮЧЕНКО, Р.В. МЕЩЕРЯКОВ, Л.Н. БАЛАЦКАЯ, Е.Л. ЧОЙНЗОНОВ  
**СТРУКТУРА И БАЗА ДАННЫХ ПРОГРАММНОГО  
ОБЕСПЕЧЕНИЯ ОЦЕНКИ КАЧЕСТВА И РАЗБОРЧИВОСТИ  
РЕЧИ В ПРОЦЕССЕ РЕАБИЛИТАЦИИ ПОСЛЕ ОПЕРАЦИИ  
ПРИ ЛЕЧЕНИИ РАКА ПОЛОСТИ РТА И РОТОГЛОТКИ,  
ЧЕЛЮСТНОЛИЦЕВОЙ ОБЛАСТИ**

---

*Костюченко Е.Ю., Мещеряков Р.В., Балацкая Л.Н., Чойнзон Е.Л. Структура и база данных программного обеспечения оценки качества и разборчивости речи в процессе реабилитации после операции при лечении рака полости рта и ротоглотки, челюстнолицевой области.*

**Аннотация.** В настоящей статье рассмотрены этапы исследования по оценке разборчивости и качества речи, проводимого совместно НИИ онкологии СО РАМН и Томским государственным университетом систем управления и радиоэлектроники. Рассмотрено программное обеспечение для сбора материала для исследования, база данных для хранения собранного материала, текущее состояние по заполнению базы данных и дальнейшие планы исследования.

**Ключевые слова:** качество речи, разборчивость, речевая реабилитация, рак.

*Kostyuchenko E.Y., Meshcheryakov R.V., Balatskaya L.N., Choinzonov E.L. Structure and database of software for speech quality and intelligibility assessment in the process of rehabilitation after surgery in the treatment of cancers of the oral cavity and oropharynx, maxillofacial area.*

**Abstract.** This article reviews the stages of study to assess the intelligibility and quality of speech, held together Oncology Institute of RAMS and Tomsk State University of Control Systems and Radio Electronics. Software to gather material for research, a database for storing the collected material, the current state of the database to fill and further research plans was considered.

**Keywords:** quality speech, intelligibility, speech rehabilitation, cancer.

---

**1. Введение.** Рак полости рта и ротоглотки, челюстнолицевой области занимает значимую долю среди злокачественных новообразований. Так, по данным [1,8] одним из эффективных методов лечения данной разновидности опухолей является комбинированное лечение, предполагающее одним из этапов оперативное вмешательство. При этом практически всегда страдает речевая функция человека, приводя к существенному снижению качества его жизни. После оперативного вмешательства - резекции гортани (злокачественные новообразования гортани находятся на первом месте по частоте онкологических заболеваний опухолей головы и шеи [2]), для восстановления речевой функции пациенту приходится полностью перестраивать речевой аппарат. Это связано с тем, что легкие не могут служить источником потока воздуха при образовании речи и приходится проходить ряд специальных тренировок для выработки так называемого пищевого голоса.

При раке ротовой полости (второе место по частоте онкологических заболеваний опухолей головы и шеи [2]) также происходит существенное изменение речеобразующего тракта и, как следствие, снижение качества речи. Возникает проблема выработки объективных количественных оценок при анализе качества речи, которые позволили бы в динамике отслеживать влияние различных упражнений, проводимых логопедом, на качество речи, оценивать эффективность таких упражнений. Наличие количественных оценок позволило бы спланировать эффективный курс занятий (в частности добившись сокращения продолжительности курса речевой реабилитации) в зависимости от особенностей произношения различных фонем и, соответственно, положений органов речеобразования.

**2. Программный комплекс для оценки разборчивости речи и сбора материалов для дальнейших исследований.** Отправной точкой для оценки качества речи стал стандарт ГОСТ Р 50840-95 [3]. По сравнению с ГОСТ Р 51061-97, который также использует таблицы из ГОСТ Р 50840-95, стандарт позволяет врачу-логопеду использовать более «понятные» оценки. В рамках исследования оценки были выбраны и реализованы методы оценки слоговой и фразовой разборчивости. В качестве таких оценок приняты доля правильно услышанных слогов и, соответственно, фраз, которые выбираются из специальных таблиц. Эти таблицы сформированы таким образом, чтобы охватить все возможные сочетания фонем, возникающие в реальной речи.

Данная методика может быть использована и для оценки качества речевого источника, если влияние канала связи отсутствует или пренебрежимо мало. Для исключения технического канала связи используется непосредственная оценка услышанных слогов логопедом, проводящим занятие. Реализовано и внедрено программное обеспечение [4], которое позволяет осуществлять оценку качества речи по данной методике [3]. Диалоговое окно разработанного программного обеспечения содержит элементы со следующими данными: 1) номер сеанса; 2) тип сеанса; 3) дата проведения сеанса; 4) номер первой таблицы для прочтения из ГОСТ 50840-95; 5) результаты оценки разборчивости; 6) текущая запись; 7) последняя запись; 8) окно редактирования записей.

Следует отметить, что данное программное обеспечение служит не только для получения конкретных оценок слоговой и фразовой разборчивости, но и позволяет собрать материал для дальнейшего исследования разборчивости и качества речи, в частности для выявления фонем, подверженных наиболее сильному изменению в зависимости от оперативного вмешательства на органах речеобразования.

**3. База данных записей для исследования качества и разборчивости речи.** Для работы данной программы была реализована база данных. Информация, подлежащая сбору, определялась, исходя из требований сотрудников НИИ онкологии РАМН: наличие анамнеза, хранение персональных данных, хранение истории контрольных прозношений, действия логопеда и лечащего врача [5]. В результате этого была выявлена следующая информация, подлежащая хранению:

- фамилия, имя и отчество пациента;
- год рождения пациента;
- объем операции (содержит информацию об органе, на котором проходило оперативное вмешательство и степень изменения органа);
- диагноз (заболевание, лечение от которого происходит);
- логопедический диагноз (описание заболевания с точки зрения нарушения речевой функции);
- информация о сеансе работы с логопедом: дата и тип оцениваемой разборчивости (слоговая или фразовая);
- информация о всех звуковых записях, сделанных во время сеанса (путь к файлу на жестком диске);
- услышанный(ая) логопедом слог или фраза;
- непосредственно оценки разборчивости.

На основе представленных данных была разработана база данных, содержащая следующие таблицы:

- gostslog – список слогов из таблиц ГОСТ 50840-95;
- gostphrase – список фраз из таблиц ГОСТ 50840-95;
- pair – таблица для проведения оценки качества речи методом парных сравнений;
- person – таблица, содержащая данные о пациенте;
- phrase – таблица, содержащая записи и результаты оценки фразовой разборчивости;
- seans – таблица, содержащая данные о сеансе оценки разборчивости и сами итоговые количественные оценки разборчивости в рамках этого сеанса;
- selective – таблица, содержащая данные для оценки качества речи по селективным признакам;
- slog – таблица, содержащая записи и результаты оценки слоговой разборчивости;
- test slog – таблица, описывающая наборы слогов для обучения, как взятые из ГОСТ 50840-95, так и сформированные логопедом исходя из специфики оперативного вмешательства;

- testologconn – таблица, описывающая связи между таблицей слогов gostslog и таблицей наборов слогов для оценки качества речи testolog;
- types – таблица методов оценки качества и разборчивости речи;

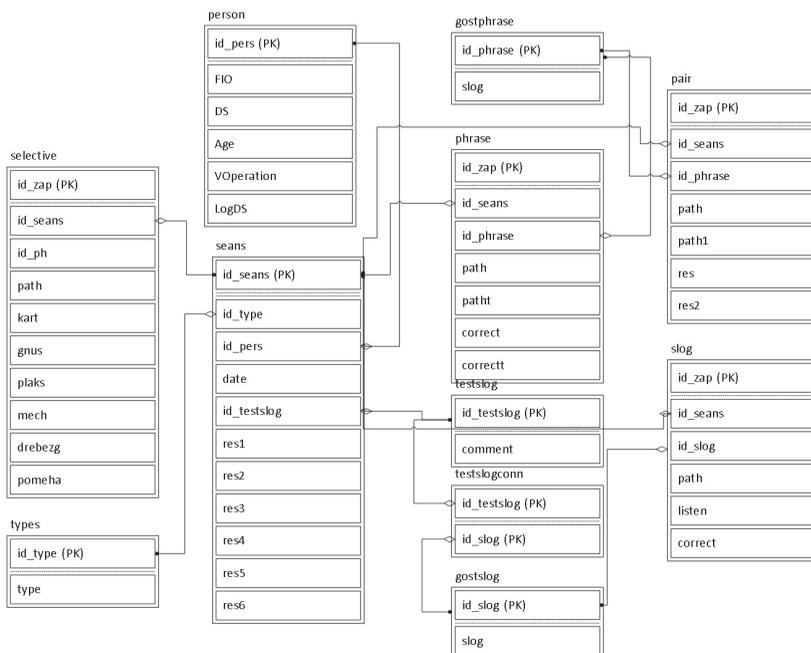


Рис. Структура базы данных для оценки разборчивости.

Структура базы данных представлена на рисунке. Данная база данных учитывает все требования сотрудников НИИ онкологии РАМН и конечных пользователей – логопедов в частности. Она позволяет вести накопление данных по пациентам с учетом специфики заболевания.

**4. Текущее наполнение базы.** В настоящий момент времени в базе собраны данные и речевые сигналы 20 пациентов, для которых произведено 64 сеанса по оценке слоговой и фразовой разборчивости. Из пациентов у 8 поставлен диагноз рак языка, у четверых рак дна полости рта. Логопедический диагноз во всех случаях дизартрия. При этом у всех пациентов присутствуют образцы речи до оперативного вмешательства. Это позволяет получать и использовать параметры речевого сигнала, которые могут быть условно приняты за эталон. Изменения речевого сигнала под воздействием злокачественного новооб-

разования даже на поздних стадиях заболевания незначительны по сравнению с изменениями, возникающими непосредственно при оперативном вмешательстве. Кроме того, имеется группа пациентов, у которых проводилась 2-х и 3-х кратная оценка разборчивости, что позволило внести в базу сигналы для оценки динамики изменения параметров в процессе занятий с логопедом. Однако, общий объем базы и распределение записей в зависимости от диагноза пока не позволяет получать статистически значимые количественные оценки по значениям параметров и их изменениям.

**5. Планы для дальнейшего исследования.** В настоящее время проводится эксперимент, и формируются таблицы фонем, подвергшихся наиболее сильному изменению для двух заболеваний, представленных в имеющейся базе многократно (рак языка и рак дна полости рта). Следующим этапом станет выявление групп фонем, которые претерпевают наиболее сильные изменения, снижающие качество речи, но не влияющие напрямую на разборчивость речи. После этого для таких фонем будет проведен спектральный анализ для получения количественных оценок, позволяющих построить модель речевого сигнала в норме и после оперативного вмешательства. На основе этой модели можно будет судить об эффективности тех или иных упражнений при речевой реабилитации. Предварительный эксперимент, указывающий на существование таких характеристик, уже проведен и описан в [6]. Кроме того, ведется постоянное наполнение базы, позволяющее получить репрезентативную выборку речевых сигналов для конкретных заболеваний, позволяющую определить статистически значимые изменения таких параметров в процессе речевой реабилитации.

**6. Заключение.** В данной работе описана структура базы данных для работы с программным комплексом для оценки разборчивости речи. Разборчивость речи как признак позволяет восстановить социальную коммуникативную функцию пациента в полном объеме [7]. Описанный этап является подготовительным в рамках проведения исследования по выявлению объективных количественных характеристик качества речи в зависимости от заболевания и объема оперативного вмешательства, которые позволят получить оценки эффективности занятий с врачом-логопедом в процессе речевой реабилитации и, в конечном итоге, сократить её время.

### **Литература**

1. Чойнзонов Е.Л., Мухамедов М.Р., Балацкая Л.Н. Рак гортани. Современные аспекты лечения и реабилитации // Томск: Изд-во НТЛ, 2006. 280 с.
2. Чиссов В.А., Старинский В.В., Петрова Г.В. Злокачественные новообразования в России в 2011 году // М.: ФГУ «МНИОИ им. П.А. Герцена» Минздравсоцразвития России. 2012. 288 с.
3. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости.
4. Балацкая Л.Н., Костюченко Е.Ю., Мещеряков Р.В. Программа для оценки качества речи при реабилитации пациентов с нарушениями речеобразующего тракта //

Свидетельство о государственной регистрации программы для ЭВМ. №2013612612. Зарегистрировано в Реестре программ для ЭВМ 6 марта 2013.

5. Meshcheryakov R.V., Balatskaya, L. N., Choinzonov E.L., Chizevskaya S.Yu., Kostyuchenko E.U. Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw // Proceedings of the 15th International Conference, SPECOM 2013, Pilsen, Czech Republic, September 1-5, 2013. pp 294-301.
6. Мецчеряков Р.В., Балацкая Л.Н., Чойнзонов Е.Л. Специализированная информационная система поддержки деятельности медицинского учреждения // Информационно-управляющие системы. 2012. №5(60). С. 51-56.
7. Ronzhin A., Karpov A., Kipyatkova I., Zelezny M. Client and Speech Detection System for Intelligent Infokiosk // Springer-Verlag Berlin Heidelberg, Petr Sojka et al. (Eds.): TSD 2010, LNAI 6231. 2010. pp. 560–567.
8. Балацкая Л.Н. Речевая реабилитация и качество жизни после лечения больных опухолями головы и шеи. Диссертация на соискание ученой степени доктора биологических наук // Томск, 2001.

## References

1. Choinzonov E.L., Muhamedov M.R., Balatskaya L. N. *Rak gortani. Sovremennye aspekty lechenija i reabilitacii* [Cancer of the larynx. Modern aspects of treatment and rehabilitation]. Tomsk: NTL, 2006. 280 p. (In Russ.).
2. Chissov V.A., Starinskiy V.V., Petrova G.V. *Zlokachestvennye novoobrazovanija v Rossii v 2011 godu* [Malignancies in Russia in 2011]. M.: FSI "MNIOI them. PA Herzen "Health Ministry of Russia. 2012. 288 p. (In Russ.).
3. GOST 50840-95. [Speech transmission over communication channels. Techniques for measurements of speech quality, intelligibility and voice identification]. (In Russ.).
4. Balatskaya L.N., Kostyuchenko E.Y., Meshcheryakov R.V. *Programma dlja ocenki kachestva rechi pri reabilitacii pacientov s narushenijami recheobrazujushhego trakta* [Program to assess speech quality in the rehabilitation of patients with impaired speech production tract]. Patent RF. № 2013612612. March 6, 2013. (In Russ.).
5. Meshcheryakov R.V., Balatskaya, L. N., Choinzonov E.L., Chizevskaya S.Yu., Kostyuchenko E.U. Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw. Proceedings of the 15th International Conference, SPECOM 2013, Pilsen, Czech Republic, September 1-5, 2013. pp 294-301.
6. Meshcheryakov R. V., Balatskaya L. N., Choinzonov E.L. [Specialized information system for support of the medical facility] *Informacionno-upravljajushhie sistemy - Information and Control Systems*. 2012. no. 5 (60). pp 51-56. (In Russ.).
7. Ronzhin A., Karpov A., Kipyatkova I., Zelezny M. Client and Speech Detection System for Intelligent Infokiosk. Springer-Verlag Berlin Heidelberg, Petr Sojka et al. (Eds.): TSD 2010, LNAI 6231, 2010. pp. 560–567.
8. Balatskaya L.N. *Rechevaja reabilitacija i kachestvo zhizni posle lechenija bol'nyh opuholjami golovy i shei. Dissertacija na soiskanie uchenoj stepeni doktora biologicheskikh nauk* [Speech rehabilitation and quality of life after treatment of head and neck cancers. Dissertation for the degree of Doctor of Biological Sciences]. Tomsk, 2001. (In Russ.).

**Костюченко Евгений Юрьевич** – к-т техн. наук, доцент кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: искусственный интеллект, анализ речи, идентификация по биометрическим характеристикам, информационная безопасность. Число научных публикаций - 40. [key@keva.tusur.ru](mailto:key@keva.tusur.ru), 634050, г. Томск, пр. Ленина, 40, р.т. +7 (3822) 413-426, факс +7 (3822) 413-426.

**Kostyuchenko Evgeny Yuryevich** – Ph.D., associate, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and

Radioelectronics (TUSUR). Research interests: artificial intelligence, speech analysis, biometric identification, information security. The number of scientific publications - 40. [key@keva.tusur.ru](mailto:key@keva.tusur.ru), KIBEVS Dept.TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; office phone +7(3822)413-426, fax +7(3822)413-426.

**Мешеряков Роман Валерьевич** – д-р техн. наук, профессор кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: системный анализ, анализ и синтез речи информационная безопасность, вопросы обработки информации в интеллектуальных системах, особое внимание уделяется вопросам создания информационно-безопасных систем. Число научных публикаций - 247. [mrv@security.tomsk.ru](mailto:mrv@security.tomsk.ru); 634050, г. Томск, пр. Ленина, 40, ауд. 210; р.т. +7 (3822) 900-111, факс +7 (3822) 900-111.

**Meshcheryakov Roman Valerievich** – Ph.D., Dr. Sci., professor, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech analysis, speech recognition, medical technology, information security. The number of publications - 247, IEEE Senior Member. [mrv@security.tomsk.ru](mailto:mrv@security.tomsk.ru); KIBEVS Dept.TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; office phone +7(3822)413-426, fax +7(3822)900-111.

**Балацкая Лидия Николаевна** – д-р биол. наук, ведущий научный сотрудник отделения опухолей головы и шеи НИИ онкологии СО РАМН. Область научных интересов: новые методические подходы к голосово-речевой реабилитации больных после органосохраняющих и функционально-щадящих операций гортани, органов полости рта, ротоглотки, щитовидной железы, а также изучение качества жизни. Число научных публикаций: более 230. [nii@onkology.tomsk.ru](mailto:nii@onkology.tomsk.ru), 634050, г. Томск, пер. Кооперативный, 5, р.т. +7 (3822) 511-039, факс: +7 (3822) 514-097.

**Balatskaya Lidiya Nikolaevna** – MD, Ph.D., Dr. Sci., a leading researcher of Head and neck cancer department, Tomsk Cancer Research Institute of RAMS. Research interests: new methodological approaches to voice and speech rehabilitation of patients after organ-preserving and functionally-sparing operations of the larynx, oral cavity organs, oropharynx, thyroid, as well as the study of quality of life. The number publications: more than 230. [nii@onkology.tomsk.ru](mailto:nii@onkology.tomsk.ru), 634050, Tomsk, per. Kooperativnyj, 5, office phone +7 (3822)511-039, fax: +7 (3822) 514-097.

**Чойзонов Евгений Лхамцыренович** – академик РАМН, д-р биол. наук, профессор, директор НИИ онкологии РАМН. Область научных интересов: разработка органосохраняющих операций и функционально-щадящих методов лечения опухолей головы и шеи, изучение качества жизни онкологических больных и прогностической роли показателей гомеостаза при опухолевой патологии. Число научных публикаций: более 600. [nii@onkology.tomsk.ru](mailto:nii@onkology.tomsk.ru), 634050, г. Томск, пер. Кооперативный, 5, р.т. +7 (3822)511-039, факс: +7 (3822) 514-097.

**Choinzonov Evgeny Lhamacyrenovich** – MD, Ph.D., Dr. Sci., Academician of RAMS, Director of Tomsk Cancer Research Institute of RAMS. Research interests: the development of organ-preserving operations and do not function-sparing treatments for head and neck tumors, the study of the quality of life of cancer patients and prognostic role of homeostasis in cancer pathology. Number of publications: more than 600. [nii@onkology.tomsk.ru](mailto:nii@onkology.tomsk.ru), 634050, Tomsk, per. Kooperativnyj, 5, office phone +7 (3822) 511-039, fax: +7 (3822) 514-097.

**Поддержка исследований.** Работа поддержана Министерством образования и науки, проект № 1220 «Фундаментальные основы проектирования информационно-безопасных систем»

**Acknowledgements.** This research is supported by the state contract of the Ministry of Education and Science №1220 “Fundamentals of information-security systems designing”.

## РЕФЕРАТ

*Костюченко Е.Ю., Мещеряков Р.В., Балацкая Л.Н., Чойнзонов Е.Л.*  
**Структура и база данных программного обеспечения оценки качества и разборчивости речи в процессе реабилитации после операции при лечении рака полости рта и ротоглотки, челюстнолицевой области.**

Статья посвящена исследованию процесса речевой реабилитации после операции при лечении рака полости рта и ротоглотки, челюстнолицевой области. В статье сформулирована проблема повышения качества речевой реабилитации и приводится описание программного комплекса для проведения исследований. Приведена структура базы данных для хранения информации о работе по реабилитации с пациентами.

В статье приводятся общие сведения о проблеме речевой реабилитации и формулируется задача поиска объективных количественных оценок качества речи в процессе реабилитации.

Далее рассматривается алгоритмическое и программное обеспечение программного комплекса для оценки слоговой и фразовой разборчивости на основе ГОСТ 50840-95 для речевого сигнала пациентов. Приведено описание структуры базы данных для программного комплекса.

Рассматривается текущее наполнение базы данных и приведены проблемы, возникшие в процессе её наполнения. представлены дальнейшие планы по выделению объективных количественных характеристик качества речи в процессе речевой реабилитации.

В результате, в данной работе описана структура базы данных для работы с программным комплексом для оценки разборчивости речи. Разборчивость речи как признак позволяет восстановить социальную коммуникативную функцию пациента в полном объеме. Описанный этап является подготовительным в рамках проведения исследования по выявлению объективных количественных характеристик качества речи в зависимости от заболевания и объема оперативного вмешательства, которые позволят получить оценки эффективности занятий с врачом-логопедом в процессе речевой реабилитации и, в конечном итоге, сократить её время.

## SUMMARY

*Kostyuchenko E.Y., Meshcheryakov R.V., Balatskaya L.N., Choinzonov E.L.*  
**Structure and database of software for speech quality and intelligibility assessment in the process of rehabilitation after surgery in the treatment of cancers of the oral cavity and oropharynx, maxillofacial area.**

The article investigates the process of speech rehabilitation after surgery in the cancers treatment of the oral cavity and oropharynx, maxillofacial area. In the article the problem of improving the speech quality rehabilitation and the developed software system are described. The database structure to store information about the rehabilitation work with patients is presented.

The article provides an overview of speech rehabilitation problem and formulated the task of finding objective quantitative assessments of the speech quality in the rehabilitation process.

The algorithms and software for evaluation of phrase and syllable intelligibility based GOST 50840-95 for speech patients are considered. The description of the database structure for software package is presented.

The current content of the database and the problems encountered in the process of filling are discussed. Further plans will be devoted to allocate objective quantitative characteristics of speech quality in the process of voice rehabilitation.

As a result, this paper describes the structure of the database and software package for evaluating of speech intelligibility. The intelligibility is a sign, which allows a patient to restore the social communicative function in full. A preparatory step in the framework of a study on the identification of objective quantitative speech performance depending on the disease and the extent of surgery that will obtain estimates of efficiency studies with the doctor- speech therapist in voice rehabilitation and, ultimately, to reduce its time is described.

Б.И. КРЮЧКОВ, А.А. КАРПОВ, В.М. УСОВ  
**ПЕРСПЕКТИВНЫЕ ПОДХОДЫ К ПРИМЕНЕНИЮ СЕРВИСНЫХ  
РОБОТОВ В ОБЛАСТИ ПИЛОТИРУЕМОЙ КОСМОНАВТИКИ**

---

*Крючков Б.И., Карпов А.А., Усов В.М. Перспективные подходы к применению сервисных роботов в области пилотируемой космонавтики.*

**Аннотация.** В статье описаны перспективные направления к использованию сервисных роботов (роботов-ассистентов) в области пилотируемой космонавтики. Проанализированы концептуальные подходы к организации внутренней среды сервисных роботов и внешней рабочей среды для совместной деятельности человека и сервисного робота.

**Ключевые слова:** сервисные роботы, роботы-помощники, ассистивные функции, интеллектуальное окружение, многомодальные человеко-машинные интерфейсы.

*Kryuchkov B.I., Karpov A.A., Usov V.M. Promising Approaches for the Use of Service Robots in the Domain of Manned Space Exploration.*

**Abstract.** In this paper, we describe some prospective directions of the use of service robots (robot-assistants) in the high-tech domain of manned space exploration. We analyze conceptual approaches for the organization of an internal environment of service robots and an external work environment for joint functioning a human-operator and a service robot.

**Keywords:** service robots, robot assistants, intellectual environment, multimodal human-computer interfaces.

---

**1. Введение.** В настоящее время в высокоразвитых странах мира наблюдается взрывной рост интереса к робототехническим системам, для которых предлагаются все новые ниши полезного применения, включая те, которые принято относить к социально значимым и при этом требующим соблюдения регламентов и стандартов безопасности для человека и окружающей среды. Тенденция постоянного совершенствования всех типов роботов и расширения их функциональных возможностей связана с несколькими факторами:

- общим прогрессом высоких информационно-коммуникационных технологий (ИКТ);
- успехами ряда смежных научных дисциплин (математики, программирования, материаловедения, электроники, механики и др.);
- изменениями инвестиционного климата в отношении ассистивных технологий, которые расширяют возможности адаптации человека (при ограничениях по функциональным возможностям) к сложным условиям жизнедеятельности.

На первый план динамического развития сервисных роботов (или роботов-помощников человека), предназначенных для функционирования в помещениях и изолированных рабочих зонах как обычного, так и опасного производства (включая герметически замкнутые обитаемые

объекты), выходит такой фактор, как возрастание социально обусловленного спроса на них.

Наиболее значимая цель – создать благоприятные условия труда для высококвалифицированных специалистов и тем самым обусловить востребованность роботов-помощников профессионалами, от которых в наибольшей степени зависит результативность решения целевых задач любых проектов. Этот фактор отчетливо проявляет себя в сфере высоких технологий, в частности, в пилотируемой космонавтике. В этой сфере высочайший профессионализм ключевых фигур, которые находятся на вершине производственной пирамиды, являются конечным звеном в сложно организованном технологическом цикле (применительно к пилотируемой космонавтике – это роль космонавтов, особенно в составе экипажей в длительных экспедициях) сочетается с постоянно возрастающими трудностями рациональной организации рабочей среды. На преодоление такого несоответствия и такого рода трудностей направлено применение роботов-помощников, так как их предназначением является повышение полезной отдачи от использования рабочего времени высококлассных специалистов. Это тем более важно, что со стороны общества наблюдается возрастание требований к результативности и продуктивности труда в указанной области, что находит выражение в увеличении количества ответственных операций в полётном задании и повышение их разнообразия при одновременном усилении требований к безопасности полёта, к эффективному использованию полётного времени при работе с полезной нагрузкой на борту космических аппаратов. Исходя из этого, оценку применения сервисных роботов можно производить с учетом того, насколько успешно удается встраивать сервисные роботы (с разной степенью интеллектуальности, мобильности и автономности) в технологические цепочки, требующие участия персонала, к которому предъявляются самые высокие требования к уровню квалификации и у которых высочайшая стоимость труда в пилотируемом полете. Выявление особенностей применения сервисных роботов в пилотируемой космонавтике позволяет точнее обосновать круг тех требований, которые способствуют повышению востребованности современных технологий робототехники. Представленный далее аналитический обзор является развитием наших концептуальных идей к применению сервисных роботов в области высокотехнологичной медицины и космонавтики [1]. Основной целью данной работы, представленной в статье, является изучение возможностей улучшения коммуникации космонавтов с сервисными роботами в условиях динамически изменяющихся условий применения посредством придания рабочей

среде свойств интерактивности и интеллектуальности с соответствующей настройкой внутренней среды сервисного робота.

**2. Направления внедрения сервисных роботов и принципы их использования в пилотируемой космонавтике.** В современных публикациях прослеживаются обнадеживающие оценки перспектив создания сервисных роботов или роботов-помощников для участия в различных технологических процессах. В частности, обсуждаются не только аспекты узко бытового применения таких роботов (например, роботы для ведения хозяйства) [2, 3], инициировавших в свое время зарождение идей разработки ассистивных технологий, но и возможный эффект их внедрения в высокотехнологические отрасли, в частности в пилотируемую космонавтику [4, 5].

*Информационная поддержка* конкретных целевых контингентов пользователей является наиболее очевидным направлением [5-7]. Из литературы известно, что одним из вариантов информационной поддержки является разработка сервисных роботов по типу «информационного киоска» [8-12]. Эти приложения расценены в качестве полезных и перспективных для пилотируемой космонавтики [5]. В указанных работах отмечается, что состав технологий поддержки диалога и характер контента для информирования пользователя имеют принципиально важное значение не только по удобству, но и по надёжности и оперативности в сложных ситуациях оказания экстренной помощи и поддержки.

В условиях пилотируемого полета в ходе длительной экспедиции бортовой информационный сервисный робот способен предоставить:

- обширный круг сведений в объёме бортовой документации по управлению системами, агрегатами, научной аппаратурой, средствами профилактики неблагоприятного действия факторов космического полёта, использованию защитного снаряжения и средств оказания медицинской помощи в полёте;

- формировать мультимедийные презентации с видеорядом сцен из жизни семьи конкретного члена экипажа для психологической поддержки, а также по индивидуальному выбору – художественные произведения разных жанров.

Применение средств информационного обеспечения космонавта на борту, несомненно, имеет свою специфику, в частности, требуется предусматривать устойчивость процесса коммуникации к разного рода факторам помех, что предполагает дополнительные усилия в области разработки программно-аппаратных средств обеспечения диалога и поддержки интерфейсов.

В работах [7-12] описан состав программно-технических средств и базовых технологий, с помощью которых уже сегодня, опираясь на отечественные разработки, можно реализовать такие способы поддержки диалогового общения пользователя с сервисным роботом, как синтез и распознавание естественной русской речи, мимический и жестовый способы взаимодействия, многомодальные совмещенные и синхронизированные способы передачи информации и др.

Информационные сервисные роботы данного типа могут быть созданы как стационарное устройство с высокой степенью автономности применения, фактически активируемое прямым обращением пользователя, которого робот может вначале идентифицировать, потом выслушать и распознать вопрос, а затем проинформировать пользователя по существу проблемы, если имеет в своей базе данных верное решение, или продолжить диалог для уточнения темы и предмета запроса. Во многих случаях представляет интерес наделение таких роботов способностью к перемещению в рабочем пространстве (мобильностью), следуя за человеком-пользователем по помещениям, в которых могут находиться интересующие человека предметы, аппаратура, агрегаты, назначение самих помещений и пр.

*Обеспечение санитарно-гигиенических требований к состоянию помещений*, мониторинг состояния воздушной среды, в том числе искусственно формируемой, очистке помещений, утилизации биологических отходов и пр. требуют выполнения многочисленных разнообразных вспомогательных операций [13, 14].

В определенной степени для улучшения санитарно-гигиенического состояния помещений пригодны стационарные установки очистки воздушной среды, их использование не исключает, а, напротив, во многих ситуациях удачно сочетается с применением мобильных сервисных средств, если они приспособлены для выполнения таких функций в труднодоступных зонах. Современные технологии позволяют создавать не только носимые, но и миниатюрные сенсоры и приборы детекции и идентификации примесей, дымов, токсинов в воздушной и водной среде, а это обстоятельство коренным образом меняет качество санитарной обработки: выполнение работ сопровождается полноценным контролем со стороны мобильного сервисного робота, оснащенного комплексом носимых приборов диагностики, детекции вредных примесей и патогенов и пр.

В условиях искусственной среды обитания на космической орбитальной станции проблема стоит не менее остро. Фактор микрогравитации способствует "всплыванию" предметов, загрязнению воздушной

среды множеством мелких частиц, которые в процессе производства и приемки продукции не были удалены по каким-то причинам. Кроме того, образование конденсатов воды, жизнедеятельность микроорганизмов, разрушающих биоорганические материалы, разрушение материалов под действием физических факторов и пр. также вносят вклад в загрязнение внутренней среды обитаемого гермообъекта. Хотя сегодня в составе средств обеспечения жизнедеятельности космонавтов существуют эффективные средства очистки искусственной воздушной среды и воды, необходимо дальнейшее совершенствование средств мониторинга состояния модулей и отсеков пилотируемого космического комплекса. В планируемых экспедициях на Луну и Марс фактор загрязнения среды на длительно автономно функционирующем космическом комплексе может стать еще более ощутимым из-за характера почвы, загрязняющей средства защиты организма (скафандры) и инструменты во время выходов космонавтов во внешнюю среду.

В указанных сферах отличительные признаки мобильности и автономности сервисных роботов, несомненно, являются важными для достижения высоких эксплуатационных характеристик при очистке помещений, постоянно подвергающихся загрязнению. Так же предпочтительны для обработки таким способом и помещения, где в определенное время нет персонала. Выбирая этот путь, можно не подвергать людей излишнему риску, правильно планируя расписание работ и рабочее пространство для активности сервисных роботов.

*Поддержание безопасных условий жизнедеятельности* – это наиболее сложное направление, так как системы обеспечения безопасности труда и охраны здоровья являются, как правило, комплексными и обеспечиваются многими компонентами, большую роль играет конкретный учет профессиональных и медицинских факторов риска и принятые нормативные требования охраны здоровья и труда работающих [13-17].

В условиях пилотируемого космического полета сервисные роботы могут контролировать состояние внутренней среды модулей орбитальной станции, в том числе редко посещаемых членами экипажа. Это позволит своевременно обнаруживать отклонения в температурном режиме, появление признаков нештатного функционирования, обнаруживать неработоспособность приборов, которые должны быть включены согласно регламенту работ, обнаруживать критические значения параметров среды обитания и появление опасных концентраций вредных примесей.

В пилотируемой космонавтике автономность и мобильность являются ключевыми требованиями к роботам, так как при недостаточной численности персонала и, напротив, обширности зон контроля необходимо так организовать мониторинг, чтобы обеспечить требования регламента обслуживания всех агрегатов и узлов, своевременно выявлять отказы бортовых систем. В общем случае приоритетная задача для решения средствами робототехники в производственных областях с особыми условиями труда – выявлять признаки опасности для пребывания человека, оповещать диспетчерский состав и/или дежурные бригады специалистов, включать (контактно или дистанционно) устройства связи и оповещения или дежурную сигнализацию и при наличии – соответствующие приборы и устройства для ликвидации угроз (дегазации и очистки среды).

*Доставка и размещение грузов, воды, питания, лекарств, перевозочных и расходных материалов* и т.п. рассматривается как четвертое применение. Для пилотируемой космонавтики фактор перемещения, размещения, поиска и идентификации грузов становится все более значимым, так как постоянно растут размеры орбитальных комплексов, объемы модулей, в том числе грузовых, масса доставляемых грузовыми транспортными кораблями грузов исчисляется тоннами.

*Оповещение персонала о возможных предпосылках к чрезвычайной ситуации (ЧС) и/или при нештатных ситуациях:* возникновении пожара, задымления, затопления, резкого снижения или повышения температуры в помещениях (за пределами нормативных и/или комфортных для человека значений) и др., рассматривается как пятое перспективное направление для применения сервисных роботов. Эти вопросы достаточно широко обсуждаются в литературе в контексте упоминаний об «умном доме». Помимо насыщения помещений датчиками измерений и видеокамерами, необходимо изыскивать средства обследования помещений, безопасные для человека, если там обнаружены признаки неблагополучия, и, возможно, что в этом отношении предпочтительно использовать мобильные сервисные роботы. Аналог этого вида применения сервисных роботов можно найти в практике ликвидации последствий ЧС, когда дистанционно управляемые мобильные роботы первыми проникают в помещения, где высок уровень угроз для здоровья и жизни человека.

*Документирование проводимых бригадами специалистов работ* посредством фото- и видеосъемки, записи звуковой обстановки, переговоров членов рабочей бригады и пр. применяется в тех случаях, когда ситуация не позволяет привлечь дополнительный персонал, в том числе

из-за стесненных и небезопасных условий, но безусловно нужны достоверные материалы о выполненных работах. Эта задача может быть отчасти возложена на сервисного робота, дистанционно управляемого человеком. Такая ситуация типична, например, при испытаниях сложных систем и ликвидации последствий ЧС. В разных областях, где требуется повышенный контроль действий персонала, на практике все шире применяются различные способы документирования, основанные на фиксации хода операций на носителях электронной информации, видеорядов данных, записи речевых комментариев, что способствует более тщательному анализу причин неудач и ошибок, а также может быть использовано в учебном процессе для подготовки специалистов соответствующего профиля. По тем же причинам возможно применение сервисных роботов на борту пилотируемого космического аппарата для улучшения процедур послеполетного анализа работы экспедиций и обеспечивающего космический полет персонала.

*Дистанционное применение манипуляционных сервисных роботов с многомодальными интерфейсами.* В этом случае сервисный робот должен ассистировать человеку, подавая инструмент, освещая рабочую зону, снижая негативные последствия тремора рук при тонких операциях, существенно повышая точность позиционирования инструмента и др., что доказано при применении манипуляционных роботов [18-20].

В пилотируемой космонавтике сегодня, в принципе, не обойтись без манипуляционных робототехнических комплексов (МРТК) при операциях внекорабельной деятельности (ВнеКД), обеспечении погрузочных и транспортных работ и некоторых других полетных операциях [16]. Известными примерами существующих и разрабатываемых космических роботов-помощников (манипуляторов) для ВнеКД космонавтов являются манипуляторы «Декстр» («Dexter» или SPDM), «Канадарм» (и «Канадарм 2»), ERA («European Robotic Arm»), система бортовых манипуляторов «Аист» для корабля «Буран» [4].

Примерами сервисных роботов для поддержки внутрикорабельной деятельности (ВнуКД) космонавтов являются персональный помощник космонавта (робот-шар), «Kirobo» (проект KIBO), «Lemur II», а также телеуправляемые роботы-гуманоиды SAR-401 и «Робонавт-2» («Robonaut-2») [4]. Такие изделия разрабатываются NASA, Европейским космическим агентством, Японским космическим агентством, а также отечественными производителями (в частности, в ЦНИИМаш, ЦНИИ РТК, МГТУ, НПО «Андроидная техника» и др.).

Мобильный манипуляционный сервисный робот, как «третья рука» в помощь космонавту, интересен в плане поддержки различных на-

ладочных и ремонтных операций, когда космонавт вынужден выполнять их в одиночку, не имея возможности оперативно менять рабочий инструмент для ручных манипуляций, не меняя рабочей позы и не отвлекаясь на поиск, извлечение и фиксацию инструмента, применение которого в условиях микрогравитации имеет ряд дополнительных особенностей.

Обобщая доступные литературные данные, можно выделить направления применения сервисных роботов (рис. 1). Приведенный перечень, естественно, не исчерпывает все возможные направления использования сервисных роботов, но и представленного объема работ для потенциального их применения достаточно, чтобы с оптимизмом оценить перспективность этих научно-практических наработок для разных прикладных областей.



Рис. 1. Классификация направлений перспективного применения сервисных роботов в области пилотируемой космонавтики

Далее кратко описаны подходы к организации внутренней среды сервисных роботов и возможности расширения их функциональных возможностей ассистивными средствами рабочей среды.

**3. Концептуальные подходы к организации внутренней среды сервисных роботов.** Рациональная организация внутренней среды робота является ключевым фактором для достижения заявленной разработчиками той или иной функциональности. В данном контексте особую роль играет принцип использования многомодульной конструкции робота, что, помимо экономических преимуществ унификации технических решений [21-23], позволяет создавать семейства взаимодополняемых устройств, настраиваемых на конкретные приложения в зависимости

сти от степени необходимости и приоритетов текущих операций и работ специалистов.

Фактически требования к наличию свойств мобильности, автономности, способности к выполнению определенных типов манипуляций, а также ориентации в сложно организованной среде и др. находят непосредственное отражение в организации внутренней среды робота, а именно в составе и назначении отдельных подсистем, в их структурно-функциональных связях между собой и с устройствами обмена информацией с внешними источниками данных.

Рассмотрим внутреннюю структуру типового сервисного робота (рис. 2), т.е. укрупненные блоки в составе сервисных роботов и их информационные связи между собой и с внешними структурами, которые отражают различные стороны функциональности сервисных роботов. Согласно общепринятым подходам, можно выделить некоторые типовые блоки мехатронного робота (ТБМР), которые, в частности, обеспечивают:

- связь с пультом управления роботом и/или с диспетчерским пунктом для выполнения оператором задач мониторинга активности сервисных роботов;
- поддержание интерфейсов с человеком-оператором (ЧО), управляющим роботом и контролирующим его активность;
- функционирование сенсорных средств робота, предназначенных для распознавания и идентификации предметов во внешней среде, ориентировки в пространстве, локализации, позиционирования и навигации робота, а также измерения параметров среды;
- выполнение предусмотренного назначением робота объема движений, как самим роботом, так и его частями для передвижения в пространстве, для выполнения манипуляций с предметами внешней среды и др.;
- энергетическое обеспечение, поиск источников питания при снижении заряда аккумуляторной батареи ниже допустимого уровня;
- самотестирование робота и самооценка работоспособности его модулей, защита робота от несанкционированного внешнего вмешательства;
- перепрограммирование робота и пополнение его баз данных и баз знаний, включая сведения о рабочей среде (например, цифровую карту помещений и разметку на ней зон с различными правами доступа) и о предписанных правилах поведения в ней;
- обеспечение безопасности для присутствующих в помещении людей и предупреждение столкновения с постоянно расположенными в

помещении агрегатами, приборами и временно присутствующими предметами, обеспечение экстренной остановки и приведения в нейтральное устойчивое состояние с включением предупредительной сигнализации.

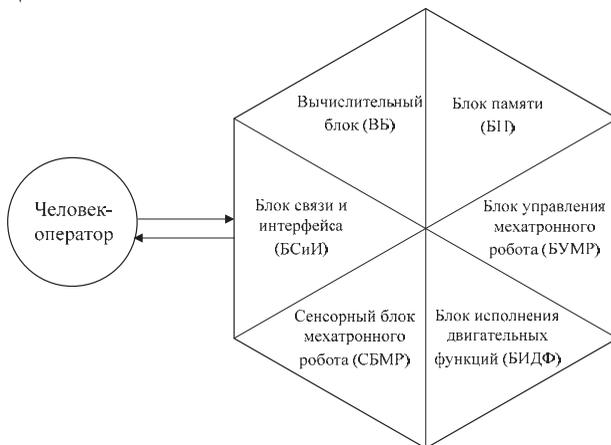


Рис. 2. Общая структура внутренней среды сервисного робота

Необходимо отметить, что многие подсистемы роботов настолько тесно связаны между собой в процессе функционирования, что проектируются только совместно, это отражается на регламентах и алгоритмах применения роботов. Следовательно, на практике существует непростой выбор между вариантами построения универсального робота (как правило, этот путь ведет к конструированию антропоморфных роботов с исполнительными устройствами, с большой точностью моделирующими функциональность рук человека) и вариантами построения специализированных роботов со сменными, но более простыми по своей функциональности, исполнительными механизмами. Выбирая второй вариант, надо отчетливо понимать, что необходимо либо создавать целое семейство однотипных роботов, либо предусматривать переналадку и настройку некоторого базового варианта, комплектуя его сменными исполнительными и сенсорными блоками. Однако такая смена комплектации, даже при достижении простоты наладки, несомненно, влечет за собой существенные изменения в системе управления и построении интерфейсов с ЧО.

Сформулируем назначение отдельных ТБМР. На первое место необходимо поставить блок связи и интерфейса с ЧО, управляющим роботом и контролирующим его активность в соответствии с антропо-

центрическим подходом к построению системы «ЧО – сервисный робот – рабочая среда».

Блок связи и интерфейса (БСиИ) с ЧО предназначен для получения команд от ЧО, работающего за пультом управления (ПУ) как внешним блоком по отношению к мехатронному роботу, при этом следует предусмотреть несколько видов интерфейса. В соответствии с выбранным ЧО типом управления роботом и способом мониторинга его движения в пространстве; важными условиями являются согласованность каналов связи и минимизация задержек прохождения управляющих команд. Дополнительные возможности БСиИ состоят в передаче в память вычислительного блока (ВБ) мехатронного робота обновлений специального программного обеспечения, баз данных, баз знаний, получаемых в порядке настройки и формирования очередного задания. ВБ может быть построен на базе классического ноутбука, использующего стандарты организации локальной сети ЭВМ.

ВБ, функционирующий совместно с блоком памяти (БП) предназначается для выполнения всех необходимых алгоритмов, обеспечивающих работоспособность блоков мехатронного робота. Этот блок может быть построен как специальный программно-аппаратный комплекс, но для центрального вычислителя мехатронного робота допустимо использовать решения, применяемые для современных смартфонов и планшетов. Такое решение может существенно облегчить построение интерфейсов и будет способствовать удешевлению изделия.

Блок исполнения двигательных функций (БИДФ) представляет собой конструкцию (как правило, электромеханического типа или с использованием пневматических приводов), посредством которой отдельные механические звенья могут двигаться друг относительно друга. Характер этого движения определяется законами управления, реализуемыми блоком управления мехатронного робота (БУМР).

Мобильность роботов определяется наличием у них исполнительного механизма выполнения движений (типа движущейся платформы), который позволяет перемещать их, используя точки опоры или сцепления с поверхностью. Для сервисных роботов в земных условиях обычно предусмотрено перемещение по плоской поверхности с небольшими углами наклона, если используются шасси на колесах, и с достаточно сложным рельефом, если используется шагающая конструкция; кроме того, имеются разработки, позволяющие обеспечить перемещение по вертикальным поверхностям и даже переход с одной плоскости на другую [24, 25].

В условиях космоса при действии фактора микрогравитации плоскостное движение внутри герметически замкнутого объекта сопряжено с дополнительными проблемами, так как необходимы специальные устройства для фиксации в ходе передвижения в условиях микрогравитации. Самостоятельная проблема состоит в обеспечении требуемой точности позиционирования и навигации мобильного робота внутри помещений. Как правило, направление движения мехатронного робота определяется по данным навигационных измерений и параметров позиционирования, задаваемых внешними источниками информации, а в самом мобильном роботе предусматривается гироскопическая система, позволяющая вычислять необходимые углы поворота и передавать эту информацию в блок исполнения двигательного акта в составе БУМР, который в наибольшей степени определяет свойства мобильности и способности к выполнению сложных манипуляций. Манипуляции, выполняемые с помощью сервисных роботов, так же как и исполнительные движения составных частей робота для передвижения в пространстве, требуют применения нескольких уровней системы управления; в типовом варианте можно выделить минимум три иерархических уровня.

Для непосредственного приведения в действие механизмов исполнения движений или манипуляций необходимо наличие системы управления единичным двигательным актом, которое можно назвать системой нижнего уровня. В силу своего назначения он подчинен системе управления среднего или верхнего уровня в зависимости от выбранного ЧО способа управления, например супервизорного, копирующего или по автономно реализуемой роботом программе по заранее сформированному сложному алгоритму. Для реализации движения по заранее сформированной программе, состоящего, как правило, из связанной последовательности единичных двигательных актов согласно хранимой в памяти ВБ робота и инициируемой в супервизорном режиме некоторой программы действий, необходимо наличие системы управления среднего уровня.

Для реализации команд на манипуляции и перемещения, получаемых от ЧО с помощью того или иного типа интерфейса, необходимо применять систему верхнего уровня, когда ЧО сам принимает решение об исполнительном действии и затем задает роботу исчерпывающий набор последовательно выполняемых команд или предписаний. Частным случаем является копирующий режим, когда ЧО, пользуясь устройством типа экзоскелетона или его составными частями («очувствленная» перчатка и пр.) выполняет требуемое движение, а мехатронный робот его копирует (повторяет), подчиняясь законам биомехани-

ки [18, 26]. В этом случае необходимо не только подключение каналов связи для дистанционного управления роботом, технического зрения в составе сенсорного блока мехатронного робота (СБМР), но и использование внешней системы контроля правильности исполнения движения роботом и контроля достижения цели двигательного акта. Внешняя система может быть построена на базе видеокамер в составе системы технического зрения, обеспечивающих 3D-изображение, которое затем анализируется как сложная зрительная сцена с вычислением необходимых параметров для коррекции управления. Нами приведено упрощенное изложение принципов построения управления в системе «ЧО – мехатронный робот», детальное изложение можно найти в публикациях [27-29].

Последним по порядку, но не по значимости для сервисных роботов рассмотрим назначение СБМР. Выше уже отмечалось, что для решения задач управления роботом могут применяться видеокамеры в его составе, посредством которых робот может выполнять анализ наблюдаемых трехмерных визуальных сцен и идентифицировать некоторые предметы в окружающей среде и их взаимное расположение с использованием систем технического зрения (СТЗ). С помощью этих же средств робот способен передавать изображение ЧО в разном разрешении и с разных ракурсов для контроля безопасности выполнения операций. Использование видеокамер робота в сочетании с видеоизображением, формируемым с нескольких внешних по отношению к роботу точек наблюдения, может существенно улучшить условия выполнения ЧО точных манипуляций. Более детально эти вопросы освещены в публикациях по проблемам построения СТЗ для роботов [30, 31].

С методической точки зрения можно ввести условное разделение сенсоров и измерительных приборов мобильного робота:

- используемые преимущественно для «собственных нужд» робота, то есть тех, без которых он не в состоянии выполнить базовые функции передвижения, позиционирования, навигации и др.;
- предназначенные для выполнения предписанных роботу заданий в аспектах целенаправленного получения от его применения дополнительных полезных эффектов (согласно представленному выше перечислению функций).

Это разделение исходит из того предположения, что при модульном построении робота, когда можно заменять часть сенсоров другими, для собственных нужд робота применяют несменяемые сенсоры, в то время как некоторые практические задачи и условия применения требуют дополнительных конфигураций набора сенсоров и носи-

мых приборов. Естественно, что экономические и эксплуатационные ограничения, как правило, не позволят часто менять конфигурацию СБМР, но при недостаточно выраженной универсализации робота это один из путей преодоления такого рода ограничений. Подобные соображения могут быть отнесены и к набору исполнительных сменных механизмов исполнения движений в зависимости от состава предписанных роботу рабочих операций. Чем выше будет универсальность исполнительного звена робота (например, какой-то качественный аналог руки человека), его искусственный интеллект, его когнитивные возможности, тем меньше будет потребность в сменных конструкциях.

Для автономного мобильного робота необходимо предусматривать достаточно развитый набор несменяемых сенсоров, так как ориентировка в пространстве, способность следовать по маршрутам, задаваемым на цифровых картах помещений, и обновлять содержимое этих карт при изменениях в рабочем помещении – достаточно сложные на практике вопросы в любой области внедрения сервисных роботов [32-34]. В настоящее время аналогичные вопросы активно исследуются применительно к мобильным устройствам связи и навигации для позиционирования в помещениях в отличие от успешно решенных вопросов навигации на местности [27, 32, 35].

Далее в статье речь пойдет об интеллектуальном пространстве (рабочей среде), которая наделяется элементами искусственного интеллекта, чтобы специально созданные для этой цели устройства могли направленного собирать информацию в интересах человека, управляющего роботом, и передавать ее в пригодном для восприятия виде.

**4. Концептуальные подходы к организации рабочей среды для совместной деятельности человека и сервисного робота.** Сегодня имеются разработки, улучшающие свойства интерактивности и ассистивности окружающей человека среды, чтобы обеспечить ее «дружественный характер» для человека, который в ней обитает и/или работает [16, 18, 35-37].

При использовании специальных технических устройств, построенных по этим принципам, при поддержке со стороны интеллектуального окружения человеку значительно проще решать задачи ориентировки в пространстве, распознавания предназначения предметов, понимания возможной реакции среды на воздействия и пр. Это положение в полной мере относится и к сервисным роботам, им тоже необходима информационная поддержка в тех ситуациях, когда собственных сенсорных возможностей недостаточно для восприятия сложных зрительных образов и извлечения релевантной информации из окружающей среды, в то вре-

мя как ассистивные устройства способны предоставить роботу управляющую информацию уже в пригодном для прямого использования по назначению виде. Выполнение этого условия предполагает наличие прямых и обратных информационных связей между ПУ роботом, СБМР и системой ассистивных средств в рабочей среде, которые способны хранить и передавать данные в стандартизированной форме по каналам связи и в рамках оговоренных интерфейсных соглашений. Информационное взаимодействие человека-оператора (например, космонавта МКС) и сервисного робота, находящихся в рамках интеллектуального окружения, проиллюстрировано на рисунке 3.

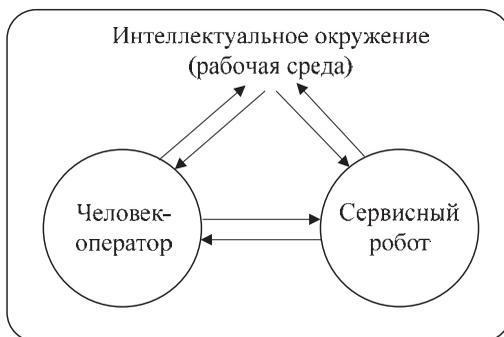


Рис. 3. Взаимодействие сервисного робота и человека-оператора в рамках интеллектуального окружения

Рассмотрим вопрос, какие свойства и качества окружающей среды необходимо разрабатывать и развивать в первую очередь, чтобы с полным основанием говорить о ее интеллектуальном насыщении в интересах организации взаимодействия в системе «космонавт – сервисный робот».

Несомненно, что решение данной задачи приведет к разработке целого ряда новых технологий и технических средств, расширяющих возможности контроля «поведенческой активности» и роботов, и космонавтов. В отношении последних этот процесс должен сопровождаться выполнением требований о защите личной информации и охране прав личности на частную жизнь. Тем не менее, тенденции, которые сегодня наблюдаются на земле в отношении мониторинга всех видов коммуникативной активности, передвижения, поисковых запросов в сети интернет и пр., будут в той или иной степени проявлять себя и в интеллектуальном окружении на пилотируемых космических объектах.

Это связано с тем, что активность космонавтов на борту не сводится исключительно к выполнению полетных операций, есть еще огромный пласт повседневной деятельности, соблюдения режима труда и отдыха, удовлетворения физических и физиологических потребностей и пр. Многое в жизни космонавтов также как и на Земле построено на стереотипах поведения, привычках, наборе достаточно ограниченных основных потребностей в соответствии с теми реально существующими ресурсами, которыми удается обеспечить экипаж. Знание таких стереотипов существенно может облегчить применение сервисных роботов, так как в соответствии с результатами предварительного планирования загрузки на день и анализа ряда предшествующих типовых суток можно составить картину того, какие директивы будут определять активность робота в ближайшее время. Это позволяет сервисному роботу искать в интеллектуальном окружении сведения, какие актуальные потребности человека можно удовлетворить теми или иными ресурсами, где их искать и в какой форме предоставить экипажу сведения такого характера. Сегодня на Земле существуют прототипы программ для мобильных устройств, которые можно рассматривать как прототипы систем активной (контекстно-зависимой) помощи по доступу к ресурсам и услугам для пользователей разных категорий в соответствии с их видами повседневной («привычной») активности.

Из этого концептуального подхода становится понятным принципиально важное требование максимального расширения возможностей устройств, включенных в состав интеллектуальной среды, в отношении поддержания коммуникации. В каком-то смысле, можно говорить о необходимости наделения каждого компонента такой среды «объектными свойствами», позволяющими ему локально, по запросу активного агента, предоставить о себе сведения о своем назначении, состоянии готовности к исполнению предписанной функции, наличных ресурсах и сроках их годности, полномочиях потребителя ресурсов, источники их пополнения и т.д. (рисунок 4).

Сервисные роботы также должны подчиняться законам функционирования такого интеллектуального окружения, чтобы, в свою очередь, использовать ассистивные средства этой среды и данные от тех пользователей, которые могут обмениваться с ней информацией, в том числе, за счет своевременной модификации состояния информационных приборов, маркеров, сигнализаторов в самой физической среде и содержимого баз данных и баз знаний в той вычислительной системе, в которой интегрируется вся информационная картина для выполнения человеком контролируемых функций и функций принятия управленче-

ских решений. Повышение интерактивности среды достигается, прежде всего, на локальном уровне: по стандартным каналам обмена информацией приборы, устройства, агрегаты в рабочей среде становятся способны многое сообщить о себе пользователям, имеющим соответствующие устройства приема-передачи данных, например о своем предназначении, текущем состоянии, планируемой циклограмме работ и пр. Это особенно важно для решения задач позиционирования и навигации мобильных роботов в сложной для ориентировки среде. Существует ряд публикаций [27, 30, 32, 34], демонстрирующих как возможности самих роботов для решения подобных задач на основе систем технического зрения (СТЗ), использования в составе средств навигации роботов лазерных дальнометров, одометров и др., так и возможности сбора этих же данных устройствами, расположенными в помещениях и передающими сведения в систему управления роботом.

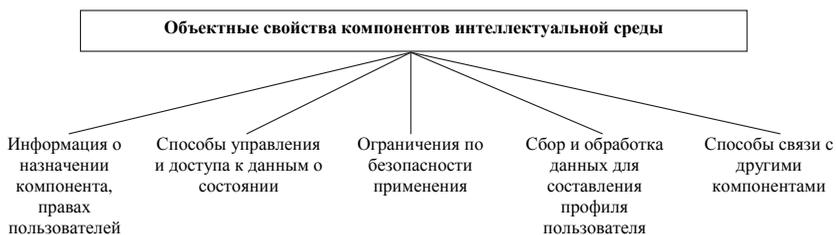


Рис. 4. Объектные свойства компонентов интеллектуальной среды, облегчающие возможности интерактивного взаимодействия с потребителями ресурсов

Допускается вариант и совместного использования сенсоров мобильного робота с возможностями систем контроля перемещений персонала и технических средств внутри помещений на основе системы (радио) меток и считывателей информации с этих меток для обмена данных с системой управления перемещением робота. Одновременно желательно обеспечить передачу этих данных в центральный вычислитель для отображения в интегрированном виде обобщенной информации ЧО, выполняющему диспетчерские функции и функции мониторинга среды.

Когда ставится вопрос о создании интеллектуального окружения на искусственных объектах для длительного обитания человека, то в первую очередь рассматриваются варианты введения в действие (в интересах информирования человека, контроля позиционирования динамически перемещающихся объектов и для облегчения навигации мо-

бильных роботов) систем датчиков и систем сбора информации о текущих параметрах рабочей среды, взаимном расположении предметов, двигательной активности мобильных агентов и т.д. (рисунок 5).

Однако возможен и такой аспект рассмотрения проблемы, как конструктивная перестройка внутренних объемов помещений для снижения помех для функционирования мобильного автономного робота.

В каком-то смысле так и происходит при построении современных космических комплексов, функционирующих в орбитальном поле в условиях невесомости и микрогравитации, где нет необходимости строить среду обитания строго по «земным лекалам», и человек иначе, чем на Земле, строит представления о своей схеме тела и ориентации в пространстве. Это позволяет использовать новые схемы компоновки агрегатов в рабочем пространстве, шире применять «фальшпанели» и прочие защитные конструкции, чтобы снизить риски механического воздействия на те или иные технические системы при передвижении космонавтов в пространстве рабочих модулей орбитальной станции. Очевидно, что этот подход правомерен и для мобильных роботов-помощников космического экипажа.

Карта (план) интеллектуальной среды	Вычислительные средства	Средства взаимодействия с пользователем и роботом	Система управления интеллектуальной среды	База данных и знаний (онтология)
		Средства анализа профиля и активности пользователя		
		Средства контекстного конфигурирования ресурсов		
		Связь с бортовой компьютерной сетью		
		Средства позиционирования и мониторинга активности		
		Сенсоры и датчики		

Рис. 5. Структура интеллектуальной окружающей среды и средств ассистивной поддержки активности космонавта и сервисного робота

Основным направлением исследований является согласование свойств интерактивности и ассистивности интеллектуального окруже-

ния с возможностями восприятия информации человеком, что может быть достигнуто развитием диалоговых средств, включая применение многомодальных интерфейсов в системе «человек – сервисный робот» [7-9]. Наиболее заметная тенденция в построении интерфейсов человека с интеллектуальным окружением и интеллектуальными агентами типа мобильных автономных роботов – использование многомодальных пользовательских интерфейсов (рисунок 6). Широкому применению ЭВМ способствует не только совершенствование их технических параметров, но и улучшение эксплуатационных характеристик, позволяющих любому конечному пользователю практически на интуитивном уровне осуществлять управление ЭВМ и различными процессами для решения собственных задач. Именно по этой причине появились такие формы диалога с ЭВМ, как графический интерфейс, управление голосовыми командами, тактильные интерфейсы [8-12, 17].



Рис. 6. Варианты построения человеко-машинных интерфейсов для рациональной организации интерактивного взаимодействия космонавта с роботом

Наиболее сложной оказывается задача построить эффективный диалог и использовать такое сочетание многомодальных систем, чтобы в каждый момент времени с учетом множества ситуативных обстоятельств ЧО (зашумленное помещение, отсутствие прямой видимости "коммуниканта диалога" и пр.) мог выбрать наиболее простой, доступный и надежный способ общения с интеллектуальным агентом, способным распознавать и синтезировать естественную речь, жестовый язык общения и др. В настоящее время имеются отечественные разработки, которые позволяют с оптимизмом смотреть на внедрение технических

средств поддержки многомодальных интерфейсов на борту пилотируемых космических комплексов [1, 5, 36].

**5. Заключение.** Перспективное использование роботов ассистентов для человека-оператора в различных современных областях науки и техники обусловлено следующими основными причинами:

- высококвалифицированный специалист тратит недопустимо много времени на предварительную подготовку рабочей среды к работе, на организацию взаимодействия с обслуживающим персоналом в ходе решения самих задач и на последующую ликвидацию возможных негативных последствий вмешательства;

- сложившееся функционирование сложной производственной среды предполагает недопустимо высокие по затратам объемы малоквалифицированного рутинного труда, снижающего эффективность технологических операций.

Выполненный анализ тенденций развития робототехники показывает актуальность создания и применения сервисных роботов, предназначенных для оказания помощи и поддержки деятельности космонавтов, работающих внутри пилотируемых космических станций (например, на Международной космической станции).

### Литература

1. *Карпов А.А., Крючков Б.И., Рогаткин Д.А., Усов В.М.* Концептуальные подходы к применению сервисных роботов: общность проблем внедрения (на примерах пилотируемой космонавтики и высокотехнологической медицины) // Биотехносфера. 2013. № 6. С. 48-59.
2. *Prassler E., Ritter A., Schaeffer C., Fiorini P.* A Short History of Cleaning Robots // Autonomous Robots. 2000. vol. 9(3), pp. 211-226.
3. *Rooks B.* Robots reach the home floor // Industrial Robot: An International Journal. 2001. vol. 28(1), pp. 27-28.
4. *Крючков Б.И.* Техническое обслуживание и ремонт в космосе // Учебно-справочное пособие. Звездный городок. 2010. 258 с.
5. *Юсупов Р.М., Крючков Б.И., Карпов А.А., Ронжин А.Л., Усов В.М.* Возможности применения многомодальных интерфейсов на пилотируемом космическом комплексе для поддержания коммуникации космонавтов с мобильным роботом – помощником экипажа // Пилотируемые полеты в космос. 2013. Вып. 3. С.23-34.
6. *Юсупов Р.М., Ронжин А.Л.* От умных приборов к интеллектуальному пространству // Вестник РАН. 2010. Т. 80, № 1. С. 45-51.
7. *Ронжин А.Л., Карпов А.А.* Проектирование интерактивных приложений с многомодальным интерфейсом // Доклады ТУСУР. 2010. № 1 (21), Ч. 1. С. 124-127.
8. *Ронжин А.Л., Карпов А.А., Леонтьева А.Б., Костюченко Б.Е.* Разработка многомодального информационного киоска // Труды СПИИРАН. 2007. Вып. 5. С. 227-245.
9. *Karpov A., Ronzhin A.* Information Enquiry Kiosk with Multimodal User Interface // Pattern Recognition and Image Analysis. 2009. vol. 19(3). pp. 546-558.

10. *Ронжин А.Л., Карпов А.А.* Сравнение методов локализации пользователя много-модальной системы по его речи // Известия вузов. Приборостроение. 2008. Т.51. № 11. С. 41-47.
11. *Карпов А.А., Цирульник Л.И., Железны М.* Разработка компьютерной системы "говорящая голова" для аудиовизуального синтеза русской речи по тексту // Информационные технологии. 2010. № 8. С. 13-18.
12. *Карпов А.А.* Когнитивные исследования ассистивного многомодального интерфейса для бесконтактного человеко-машинного взаимодействия // Информатика и ее применения. 2012. Т. 6. № 2. С. 77-86.
13. *Лапитан Д.Г., Рогаткин Д.А., Куликов Д.А., Рогаткин А.Д.* Целевые функции сервисных медицинских роботов в клинике // Сборник материалов X международной научной конференции «Физика и радиоэлектроника в медицине и экологии» ФРЭМЭ'2012, Владимир: ВГУ. 2012. С. 228-230.
14. *Краевский С.В., Рогаткин Д.А.* Медицинская робототехника: первые шаги медицинских роботов // Технологии живых систем. 2010. Т. 7. № 4. С. 3-14.
15. *Рогаткин Д.А., Лапитан Д.Г., Ланаева Л.Г.* Концепция мобильных автономных сервисных роботов для медицины // Биомедицинская радиоэлектроника. 2013. № 5. С.46-56.
16. *Крючков Б.И., Крикалёв С.К., Салаев А.М., Усов В.М.* Человек и робот на пилотируемом космическом аппарате // Сборник научных статей по материалам докладов на Первом российско-германском семинаре по космической робототехнике, Германия, Штутгарт, 2012, электронные труды.
17. *Куафе Ф.* Взаимодействие робота с внешней средой / Пер. с франц. М.: Мир. 1985. 285 с.
18. *Крючков Б.И., Усов В.М.* Антропоцентрический подход в организации совместной деятельности космонавтов ПКК и робота-помощника андроида типа. // Пилотируемые полёты в космос: Научно-технический журнал. 2012. № 3(5). С.42-57.
19. *Albani J.M.* The Role of Robotics in Surgery: A Review // Missouri Medicine. 2007. vol. 104. no.2. pp. 166-172.
20. *O'Toole M.D., Bouazza-Marouf K., Kerr D., Gooroochurn M., Vloeberghs M.* A methodology for design and appraisal of surgical robotic systems // Robotica. 2010. vol. 28(2). pp. 297-310.
21. *Юревич Е.И.* Основы робототехники. 2-е изд., перераб. и доп. СПб.: БХВ-Петербург. 2005. 416 с.
22. *Лопота В.А., Минаков Е.П., Юревич Е.И.* Современное состояние и перспективы развития отечественной космической робототехники // Материалы Всероссийской конференции «Актуальные проблемы защиты и безопасности», СПб.: Астерион, 2008. Т.5.
23. *Лопота В.А., Юдин В.И., Юревич Е.И.* О системном подходе к развитию экстремальной робототехники // Материалы Всероссийской конференции «Актуальные проблемы защиты и безопасности», СПб.: Астерион, 2005. Т.5. С. 49-54.
24. *Градецкий В., Фомин Л.* Анализ динамики управляемых движений робота вертикального перемещения с многозвенными манипуляторами // Мехатроника, автоматизация, управление. 2007. № 12. С. 9-12.
25. *Градецкий В., Фомин Л.* Обеспечение функционирования интегрированной системы управления мобильным роботом с многозвенным манипулятором // Мехатроника, управление, автоматизация. 2005. № 11. С.11-16.
26. *Тимофеев А.В., Чернакова С.Э., Литвинов М.В., Анчиков А.Д., Полонский Ю.З., Козаченко А.В.* Медицинские аспекты разработки систем человеко-машинного взаимодействия с использованием моделей виртуальной реальности для нейрохирургии // Труды СПИИРАН. 2008. Вып. 6. С. 184-196.

27. *Алферов Г.В., Кулаков Ф.М., Нечаев А.И., Чернакова С.Э.* Информационные системы виртуальной реальности в мехатронике и робототехнике: Учеб. Пособие СПб.: «СОЛЮ», 2006. 146 с.
28. *Кулаков Ф.М., Смирнов Е.Н., Липатов А.Е., Латыпов В.Н.* Информационная технология добавления виртуального объекта в реальный мир. Часть 1 // Труды СПИИРАН. 2004. Вып. 2. Т. 1. С. 236-256.
29. *Кулаков Ф.М., Смирнов Е.Н., Липатов А.Е., Латыпов В.Н.* Информационная технология добавления виртуального объекта в реальный мир. Часть 2 // Труды СПИИРАН. 2005. Вып. 2. Т. 2. С. 223-248.
30. *Миронов С.В., Юдин А.В.* Система технического зрения в задачах навигации мобильных объектов // Программные продукты и системы. 2011. № 1. С. 10-16.
31. *Соколов С.М.* Проблемы машинного зрения в робототехнике и автоматизации производства // Будущее прикладной математики. Лекции для молодых исследователей / Институт прикладной математики им. М.В. Келдыша РАН. М.: URSS, 2004. С. 343-373.
32. *Самарин А.И.* Мультисенсорные навигационные системы для локального позиционирования // Современная электроника. 2006. № 6. С. 10-17.
33. *Аркадьев В.Б., Власенко А.Н., Лапин О.Е.* Мультисенсорная система в составе мобильного робототехнического комплекса для поиска, локализации и идентификации источников гамма-излучения // Труды международной научно-технической конференции «Экстремальная робототехника», СПб: Политехник-сервис, 2011.
34. *Ермишин К.В., Воротников С.А.* Мультиагентная сенсорная система сервисного мобильного робота // Вестник МГТУ им. Н. Э. Баумана. Сер. Приборостроение. 2012. Вып. 6. С. 50-59.
35. *Юсупов Р.М., Карпов А.А., Крючков Б.И., Ронжин А.Л., Сыркин Л.Д., Усов В.М.* Создание «интеллектуального окружения» на пилотируемом космическом комплексе для позиционирования мобильного робота – помощника экипажа // Актуальные проблемы психологии труда, инженерной психологии и эргономики. Вып. 5. / Под ред. А.А. Обознова, А.Л. Журавлева. М.: Институт психологии РАН. 2013. С. 397-422.
36. *Юсупов Р.М., Ронжин А.Л., Прищепа М.В., Ронжин Ал.Л.* Модели и программно-аппаратные решения автоматизированного управления интеллектуальным залом // Автоматика и телемеханика. 2011. № 7. С. 39-49.
37. *Крючков Б.И., Усов В.М.* Создание моделей виртуальной реальности, как способ обучения космонавтов взаимодействию с роботом – помощником экипажа и как условие определение потенциальных областей его полезного применения // Труды международной научно-технической конференции «Экстремальная робототехника», СПб: Политехника-сервис. 2013. С. 230-244.

## References

1. Karpov A.A., Kryuchkov B.I., Rogatkin D.A., Usov V.M. [Conceptual Approaches for Using Service Robots: Common Problems of Implementation (On the Example of Manned Space Exploration and High-Tech)]. *Biotechnosfera – Biotechnosphere*. 2013. no. 6. pp. 48-59. (In Russ.).
2. Prassler E., Ritter A., Schaeffer C., Fiorini P. A Short History of Cleaning Robots. *Autonomous Robots*. 2000. vol. 9(3). pp. 211-226.
3. Rooks B. Robots reach the home floor. *Industrial Robot: An International Journal*. 2001. vol. 28(1). pp. 27-28.
4. Kryuchkov B.I. *Tekhnicheskoe obsluzhivanie i remont v kosmose. Uchebno-spravochnoe posobie* [Technical service and maintenance in the space]. Star City. 2010. 258 p. (In Russ.).

5. Yusupov R.M., Kryuchkov B.I., Karpov A.A., Ronzhin A.L., Usov V.M. [Possibility of Application of Multimodal Interfaces on a Manned Space Complex to Maintain Communication Between Cosmonauts and a Mobile Robotic Assistant]. *Pilotiruemye polety v kosmos – Manned spaceflights*. 2013. no. 3(8). pp. 23-34. (In Russ.).
6. Yusupov R.M., Ronzhin A.L. From Smart Devices to Smart Space. Herald of the Russian Academy of Sciences. 2010. vol. 80, no. 1. pp. 45-51.
7. Ronzhin A.L., Karpov A.A. [Development of interactive applications with multimodal interfaces] // *Doklady TUSUR – Proceedings of Tomsk State University of Control Systems and Radioelectronics*. 2010. no. 1(21). Part 1. pp. 124-127. (In Russ.).
8. Ronzhin A., Karpov A., Leontyeva An., Kostuchenko B. [Development of the multimodal information kiosk]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2007. Issue 5. vol. 1. pp. 227-245. (In Russ.).
9. Karpov A., Ronzhin A. Information Enquiry Kiosk with Multimodal User Interface. Pattern Recognition and Image Analysis. 2009. vol. 19(3). pp. 546-558.
10. Ronzhin A., Karpov A. [Comparison of localization methods of a multimodal system user by his speech]. *Izv. vyssh. uchebn. zavedenij: Priborostroenie – Proceedings of the higher educational institutions: Instrumentation*. 2008. vol. 51, no. 11. pp. 41–47. (In Russ.).
11. Karpov A.A., Tsurulnik L.I., Zelezny M. [Development of a Computer System "Talking Head" for Text-to-Audiovisual-Speech Synthesis]. *Informatsionnye tekhnologii – Information technologies*. 2010. vol. 9. no. 8. pp. 13-18. (In Russ.).
12. Karpov A.A. [Cognitive study of an assistive multimodal user interface for hands-free human computer interaction]. *Informatika i ee primeneniia – Informatics and its applications*. 2012. vol. 6. no. 2. pp. 77-86. (In Russ.).
13. Lapitan D.G., Rogatkin D.A., Kulikov D.A., Rogatkin A.D. [Objective functions of service medical robots in a hospital]. *Sbornik materialov X mezhdunarodnoi nauchnoi konferentsii «Fizika i radioelektronika v meditsine i ekologii»* [Proceedings of 10th International conference "Physics and radioelectronics in medicine and ecology"]. Vladimir. 2012. pp. 228-230. (In Russ.).
14. Kraevsky S.V., Rogatkin D. A. [Medical robotics: the first steps of medical robots]. *Tekhnologii zhivyykh sistem – Technologies of Living Systems*. 2010. Vol. 7. no. 4. pp. 3-14. (In Russ.).
15. Rogatkin D.A., Lapitan D.G., Lapaeva L.G. [Conception of the mobile autonomous service medical robots]. *Biomeditsinskaya radioelektronika – Biomedical Radioelectronics*. 2013. no. 5. pp. 46-56. (In Russ.).
16. Kryuchkov B.I., Krikalev S.K., Salaev A.M., Usov V.M. [Human and a robot in a manned spacecraft]. *Sbornik nauchnykh statei po materialam dokladov na Pervom rossiisko-germanskom seminare po kosmicheskoi robototekhnike* [Proceedings of 1st Russian-German workshop on space robotics]. Stuttgart, 2012.
17. Kuafe F. *Vzaimodejstvie robota s vneshnej sredoj* [Interaction of a robot with an external environment]. M., 1985. 285 p. (In Russ.).
18. Kryuchkov B.I., Usov V.M. [Anthropocentric approach for organization of joint activity of astronauts and a robot-assistant of an android type]. *Pilotiruemye polity v kosmos – Manned spaceflights*. 2012. no. 3(5). pp. 42-57. (In Russ.).
19. Albani J.M. The Role of Robotics in Surgery: A Review. *Missouri Medicine*. 2007. vol. 104. no. 2. pp. 166-172.
20. O'Toole M.D., Bouazza-Marouf K., Kerr D., Gooroochurn M., Vloeberghs M. A methodology for design and appraisal of surgical robotic systems. *Robotica*. 2010. vol. 28(2). pp. 297-310.
21. Yurevich E.I. *Osnovy robototekhniki* [Basics of robotics]. SPb. 2005. 416 p. (In Russ.).

22. Lopota V.A., Minakov E.P., Yurevich E.I. [State-of-the-art and perspectives of development of Russian space robotics]. *Materialy Vserossiiskoi konferentsii «Aktual'nye problemy zashchity i bezopasnosti»* [Proceedings of All-Russian conference “Actual problems of defense and security”]. SPb., 2008. vol. 5. (In Russ.).
23. Lopota V.A., Yudin V.I., Yurevich E.I. [On a system approach to development of extreme robotics]. *Materialy Vserossiiskoi konferentsii «Aktual'nye problemy zashchity i bezopasnosti»* [Proceedings of All-Russian conference “Actual problems of defense and security”]. 2005. SPb., vol. 5, pp. 49-54. (In Russ.).
24. Gradetsky V.G., Fomin L.F. [Analysis of Dynamics of Control Motions of the Wall Climbing Robot with Multilink Manipulators]. *Mekhatronika, avtomatizatsiia, upravlenie – Mechatronics, Automation, Control*. 2007. no. 12. pp. 9-12. (In Russ.).
25. Gradetsky V.G., Fomin L.F. [Providing to Function of the Integral Control System of Mobile Robot with Multilink Manipulator]. *Mekhatronika, avtomatizatsiia, upravlenie – Mechatronics, Automation, Control*. 2005. no. 11. pp. 11-16. (In Russ.).
26. Timofeev A.V., Chernakova S.E., Litvinov M.V., Anichkov A.D., Polonskiy U.Z., Kozachenko A.V. [The medical aspects of development of man-machine interaction systems with using of virtual reality models for neurosurgery]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2008. Issue 6. pp. 184-196. (In Russ.).
27. Alferov G.V., Kulakov F.M., Nechaev A.I., Chernakova S.E. *Informacionnye sistemy virtual'noj real'nosti v mehatronike i robototekhnike* [Information systems of virtual reality in mechatronics and robotics]. SPb., 2006. 146 p. (In Russ.).
28. Kulakov F.M., Smirnov E.N., Lipatov A.E., Latipov V.N. [Information technology of addition of virtual object in the real world. Part 1]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2004. Issue 2, pp. 236-256. (In Russ.).
29. Kulakov F.M., Smirnov E.N., Lipatov A.E., Latipov V.N. [Information technology of addition of virtual object in the real world. Part 2]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2005. Issue 2, pp. 223-248. (In Russ.).
30. Mironov S.V., Yudin A.V. [Machine vision system in mobile object navigation challenges]. *Programmnye produkty i sistemy – Software products and systems*. 2011. no. 1. pp. 10-16. (In Russ.).
31. Sokolov S.M. [Problems of machine vision in robotics and automation of industry]. *Budushchee prikladnoi matematiki. Lektsii dlia molodykh issledovatelei* [Future of applied mathematics. Lectures for young researchers]. M., 2004. pp. 343-373. (In Russ.).
32. Samarin A.I. [Multi-sensor navigation systems for local positioning]. *Sovremennaiia Elektronika – Modern electronics*. 2006. no. 6. pp. 10-17. (In Russ.).
33. Arkadiev V.B., Vlasenko A.N., Lapin O.E. [Multi-sensor system in a mobile robotics complex for search, localization and identification of sources of gamma rays]. *Trudy mezhdunarodnoi nauchno-tekhnicheskoi konferentsii «Ekstremal'naiia robototekhnika»* [Proceedings of the International scientific and technological conference “Extreme robotics”]. SPb. 2011. (In Russ.).
34. Ermishin K.V., Vorotnikov S.A. [Multi-agent sensor system of a service mobile robot]. *Vestnik MGTU im. Baumana. Ser. Priborostroenie – Herald of the Bauman Moscow State Technical University*. 2012. no. 6. pp. 50-59. (In Russ.).
35. Yusupov R.M., Karpov A.A., Kryuchkov B.I., Ronzhin A.L., Syrkin L.D., Usov V.M. [Development of an “intelligent environment” in the manned spacecraft for positioning mobile robot-assistant]. *Aktual'nye problemy psikhologii truda, inzhenernoi psikhologii i ergonomiki – Actual problems of labour psychology, engineering psychology and ergonomics*. M., 2013. Issue 5. pp. 397-422. (In Russ.).
36. Yusupov R.M., Ronzhin An.L., Prischepa M., Ronzhin Al.L. Models and hardware-software solutions of automated control of intelligent meeting room facilities. *Automation and Remote Control*. 2011. vol. 72. no. 7. pp. 39-49.

37. Kryuchkov B.I., Usov V.M. [Creation of virtual reality models as a way of learning astronauts for communication with a robot-assistant of crew and as a condition for detection of potential areas of its application]. *Trudy mezhdunarodnoi nauchno-tekhnicheskoi konferentsii «Ekstremal'naiia robototekhnika»* [Proceedings of International scientific and technological conference “Extreme robotics”]. 2013. pp. 230-244.

**Крючков Борис Иванович** — д-р техн. наук, заместитель начальника по научной работе Научно-исследовательского испытательного центра подготовки космонавтов имени Ю.А.Гагарина. Область научных интересов: пилотируемые полеты в космос, космическая робототехника. Число научных публикаций — 200. bik43@mail.ru; Звездный городок, Московская область, Россия, 141160, РФ; р.т. +7(495)526-3407, факс +7(495)526-2612.

**Kryuchkov Boris Ivanovich** — Ph.D., Dr. Sci., Deputy Director for Research, Federal State Organization “Gagarin Research & Test Cosmonaut Training Center”. Research interests: manned space exploration, space robotics. The number of publications — 200. bik43@mail.ru; Star City, Moscow Region, 141160, Russia; office phone +7(495)526-3407, fax +7(495)526-2612.

**Карпов Алексей Анатольевич** — д-р техн. наук, доцент, ведущий научный сотрудник лаборатории речевых и многомодальных интерфейсов ФГБУН Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: автоматическое распознавание речи, многомодальные интерфейсы, аудиовизуальное распознавание и синтез речи. Число научных публикаций — 180. karpov@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081.

**Karpov Alexey Anatolyevich** — Ph.D., Dr. Sci., Assoc. Professor, Leading researcher, Speech and Multimodal Interfaces Laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: automatic speech recognition, multimodal interfaces, audio-visual speech recognition. The number of publications — 180. karpov@iias.spb.su; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081.

**Усов Виталий Михайлович** — д-р мед. наук, профессор, главный научный сотрудник Научно-исследовательского испытательного центра подготовки космонавтов имени Ю.А.Гагарина. Область научных интересов: пилотируемые полеты в космос, космическая и военная медицина, космическая робототехника. Число научных публикаций — 120. khoper.1946@gmail.com; Звездный городок, Московская область, Россия, 141160, РФ; р.т. +7(495)526-3407, факс +7(495)526-2612.

**Usov Vitali Mikhailovich** — Ph.D., Dr. Sci., Professor, Chief researcher, Federal State Organization “Gagarin Research & Test Cosmonaut Training Center”. Research interests: manned space exploration, space and military medicine, space robotics. The number of publications — 120. khoper.1946@gmail.com; Star City, Moscow Region, 141160, Russia; office phone +7(495)526-3407, fax +7(495)526-2612.

**Поддержка исследований.** Работа выполнена при частичной финансовой поддержке РФФИ (проект № 12-08-01265-а).

**Acknowledgements.** This research is partially supported by RFBR (project № 12-08-01265-а).

## РЕФЕРАТ

*Крючков Б.И., Карпов А.А., Усов В.М.* **Перспективные подходы к применению сервисных роботов в области пилотируемой космонавтики.**

В данной статье описаны перспективные направления к использованию сервисных роботов (роботов-ассистентов) в области пилотируемой космонавтики. Проанализированы концептуальные подходы к организации внутренней среды сервисных роботов и внешней рабочей среды для совместной деятельности человека и сервисного робота.

Прогрессу технологий робототехники способствует успешное решение двух задач: наделение роботов новыми способностями выполнять все более разнообразные и сложные функции, ранее доступные только человеку, и доказательство полезности для повседневной практики этих функций в исполнении роботов.

В статье предложен концептуальный подход к описанию эргатической системы «космонавт – сервисный робот – рабочая среда пилотируемого комплекса» в терминах принципов её организации, построения внутренней среды роботов и создания интеллектуального окружения с набором ассистивных функций, облегчающих космонавту мониторинг активности мобильных автономных сервисных роботов внутри рабочего пространства пилотируемого комплекса.

Также представлены направления внедрения сервисных роботов и принципы их использования в пилотируемой космонавтике, подходы к организации внутренней и внешней среды сервисных роботов. Анализ тенденций развития робототехники показал актуальность создания и применения сервисных роботов, предназначенных для оказания помощи и поддержки деятельности космонавтов, работающих внутри пилотируемых космических станций (например, на Международной космической станции).

## SUMMARY

### *Kryuchkov B.I., Karpov A.A., Usov V.M.* **Promising Approaches for the Use of Service Robots in the Domain of Manned Space Exploration.**

In this paper, we describe some prospective directions of the use of service robots (robot-assistants) in the high-tech domain of manned space exploration. We analyze conceptual approaches for the organization of an internal environment of service robots and an external work environment for joint functioning a human-operator and a service robot.

Progress of technologies in robotics depends on successful solution of two challenges: empowering robots with new abilities to perform more various and complex features, which were available only to a human being, as well as proof of usefulness of these functions performed for everyday practice by robots.

In the paper, we propose a conceptual approach for description of an ergatic system "astronaut – service robot – operating environment of a manned spacecraft" in terms of principles of its organization, building robot's internal environment and creation of an intellectual environment with a set of assistive functions facilitating activity monitoring of mobile autonomous service robots inside the operating environment of a manned spacecraft.

We also discuss directions of service robots implementation, principles of their usage in manned space exploration domain, approaches for organization of the internal environment and external environment of service robots. The analysis of tendencies of robotics development has shown on urgency of development and application of service robots aimed for assistance and information support in activity of astronauts working inside manned spacecrafts (for example, in the International Space Station).

О.О. БАСОВ, И.А. САИТОВ

## КАЧЕСТВО ФУНКЦИОНИРОВАНИЯ И ЭФФЕКТИВНОСТЬ ПОЛИМОДАЛЬНЫХ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ

---

*Басов О.О., Саитов И.А. Качество функционирования и эффективность полимодальных инфокоммуникационных систем.*

**Аннотация.** Отказ от традиционных принципов разделения передаваемой информации на услуги в пользу полимодального представления информации требует разработки новой конструктивной теории построения полимодальных инфокоммуникационных систем. Одним из ее краеугольных камней является количественная оценка степени достижения цели функционирования таких систем и доли участия в этом результате абонентских терминалов. В работе предложен подход к оцениванию эффективности полимодальных систем на основе показателя удельной себестоимости.

**Ключевые слова:** полимодальная инфокоммуникационная система, эффективность, удельная себестоимость, полнота информации.

*Basov O.O., Saitov I.A. Functioning quality and effectiveness of polymodal infocommunicational systems.*

**Abstract.** Abandoning traditional principles of separation of transmitted information on services for polymodal presentation of information requires development of a new constructive theory of building polymodal infocommunication systems. One of its cornerstones is the quantitative research of the degree of the achievement of the purpose of such systems and subscribers' participation share in this result. This paper presents an approach to evaluation of polymodal systems effectiveness on the basis of the prime cost index.

**Keywords:** polymodal infocommunication system, effectiveness, prime cost, completeness of information.

---

**1. Введение.** Расширение поля коммуникативного взаимодействия, а также постоянно растущие психологические нагрузки при принятии управленческих решений, связанные с уменьшением квоты доверия общающихся друг к другу, трансформируют формально-ролевое общение в деловое, при котором наряду с обменом информацией должны учитываться особенности личности абонента, его настроение, физиологическое и психоэмоциональное состояние. Глобальная информатизация процессов управления, возросшая скорость изменения ситуации и увеличения объемов передаваемой информации требуют от абонентов повышения результативности их действий в направлениях скорости принятия решений и адекватности текущей ситуации [1].

В данных условиях инфокоммуникационные системы (ИКС) оказываются функционально ограниченными и не обеспечивают требуемой эффективности коммуникативного взаимодействия абонентов [2].

Причиной существующего положения дел является сохраняющееся в научно-техническом сообществе отношение к инфокоммуни-

кациям как распределенным системам, реализующим функции получения, обработки, передачи и восстановления информации. Системным следствием этого является увеличение пропускной способности (и, как следствие, стоимости всей системы), вызванное последовательным и независимым наращиванием применяемых модальностей (ввод текста с клавиатуры, рукописный ввод, речевой сигнал, изображение) при обработке и передаче информации.

Особенно контрастно выглядят выделенные проблемы практики на фоне возросших возможностей современных вычислительных и информационных технологий. В многомодальных системах человеко-машинного взаимодействия информация от различных (видео, аудио, тактильных) коммуникативных каналов непрерывно отслеживается и обрабатывается, создавая реальное или виртуальное окружение, позволяющее удовлетворить желания пользователя и оперативно адаптировать процессы обработки информации к текущей задаче и другим прикладным аспектам [16]. Внедряются результаты исследований психофизиологического состояния человека на основе эмоциональных признаков речи и лица [7, 24]. Реализуются интеллектуальные системы поддержки принятия решений, основанные на интеграции информации и знаний в контекст, описывающий задачу абонента или требуемую ситуацию и учитывающий динамику окружающей среды [18].

Таким образом, в предметной области, с одной стороны, наблюдается объективная необходимость в повышении эффективности коммуникативного взаимодействия абонентов, с другой – появились технические и технологические предпосылки для решения выделенных проблем практики. В данных условиях целесообразным подходом к решению проблем практики выбран интенсивный путь развития ИКС посредством учета всех сторон общения (коммуникативной, интерактивной и перцептивной) невербальными средствами. Перспективным в этом смысле становится отказ от традиционных принципов разделения передаваемой информации на услуги в пользу ее полимодального представления [2].

Такой подход требует разработки новой строгой, но в то же время конструктивной, теории полимодальных ИКС (ПИКС), одной из основ которой является количественная оценка степени достижения изначально заданных целей ее функционирования.

**2. Общий подход к оценке качества функционирования и эффективности ИКС.** Классический подход [4, 9, 22 и др.] к оценке эффективности ИКС состоит в анализе их качественных характеристик и получении соответствующих частных показателей эффективности (ЧПЭФ). На их основе методами аддитивной или мультипликатив-

ной свертки удается сформировать некий обобщенный показатель эффективности (ОПЭФ) системы. Исследования в предметной области показали, что значение ОПЭФ ИКС будут определяться как архитектурой системы, так и свойствами информационного потока, поступающего от абонентов.

Нормативные документы в области построения ИКС в большинстве случаев рассматривают в качестве приоритетных для отражения степени удовлетворения абонента услугой показатели *QoS* (*Quality of Service – качество обслуживания*) [19]. При этом критерии и параметры *QoS* должны рассматриваться для каждой конкретной услуги и определяться для сквозной связи (рис. 1), конечными точками которой являются точки подсоединения абонентских терминалов (АТ), в терминах, понятных абоненту [11].



Рис. 1. Феноменологическое представление ИКС

В общем случае ИКС предоставляет каждой обслуживаемой паре абонентов либо коммутируемый фиксированный (статическое мультиплексирование), либо постоянный виртуальный (статистическое мультиплексирование) канал, основная задача которого доставлять информацию в виде потока блоков данных от АТ-источника к АТ-адресату. Так как на канальном и физическом уровне сети доступа (СД) и транспортной сети (ТС) формирование обоих типов каналов осуществляется путем последовательного установления соединения ресурсов технически сопряженных элементов направляющих систем электросвязи и элементов запоминающих устройств коммутационного оборудования (КО), то в дальнейшем для них будет использоваться общее модельное понятие "*сети передачи данных*" (СПД).

В целом ряде исследований [6, 8] при оценке эффективности учитывается вложенность и иерархичность элементов ИКС (рис. 2), когда система в целом, и ее подсистемы подвержены воздействиям среды функционирования (СФ).

Тогда исходным состоянием ПИКС будем считать такое, когда полимодальная информация в виде потока блоков данных (сообщений), соответствующего расчетному, поступает в СПД (рис. 1), все элементы СД, ТС и обеспечивающих их подсистем исправны, а дестабилизирующие факторы отсутствуют. Такое состоя-

ние ПИКС характеризуется ее качеством  $Q_0^{\text{ПИКС}}$ . В результате влияния СФ фактическое качество полимодальной ИКС будет изменяться во времени  $Q^{\text{ПИКС}}(t)$ , при этом для абонента важно, чтобы абсолютное отклонение этой величины от расчетной не превышало некоторого заданного значения:

$$\Delta Q^{\text{ПИКС}}(t) = \left| Q_0^{\text{ПИКС}} - Q^{\text{ПИКС}}(t) \right| \leq \Delta Q^{\text{доп}}(t). \quad (1)$$

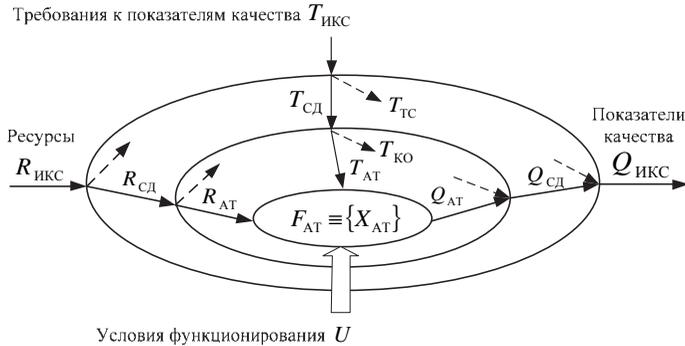


Рис. 2. Обобщенная модель метауровня исследований зависимости внешних функциональных характеристик ИКС от наличия ресурсов и условий функционирования

*Внешняя эффективность* ПИКС будет определяться максимальным значением  $\Delta Q^{\text{ПИКС}}$ , которое она способна обеспечить при ограничениях на приведенные затраты  $C_{\text{пр}}^{\text{ПИКС}}$  на ее построение и эксплуатацию:

$$\mathcal{E}^{\text{ПИКС}} = \Psi_{\text{ВЭ}}(\Delta Q^{\text{ПИКС}}, C_{\text{пр}}^{\text{ПИКС}}), \quad (2)$$

$$C_{\text{пр}}^{\text{ПИКС}} \leq C_{\text{пр}}^{\text{ПИКС доп}}, \quad (3)$$

где  $C_{\text{пр}}^{\text{ПИКС доп}}$  – допустимые приведенные затраты на построение и эксплуатацию ПИКС.

Очевидно, что для оценки внешней эффективности ПИКС необходимо уметь рассчитывать величины  $\Delta Q^{\text{ПИКС}}$  и  $C_{\text{пр}}^{\text{ПИКС}}$ .

**3. Качество функционирования ПИКС.** Интеграция информационных [5] и телекоммуникационных [12–14] систем при формировании системотехнического облика ПИКС обусловила выбор в качестве показателей качества: целостности, своевременности, безопасности, производительности и надежности (рис. 3).



Рис. 3. Показатели качества функционирования информационных, телекоммуникационных и полимодальных систем

При ограничениях на достигаемый уровень безопасности и надежности информации указанные ЧПЭФ могут быть формализованы следующим образом.

*Полнота* – свойство выходной информации отражать состояния всех требуемых объектов учета (ОУ) предметной области ПИКС:

$$P_{\text{ПИКС}} = \frac{1}{M} \sum_{m=1}^M x_m^{\text{ОУ}}, \quad (4)$$

где  $M$  – минимально необходимое для принятия правильного решения число объектов учета, а  $m$ -й показатель полноты определяется как:

$$x_m^{\text{ОУ}} = \begin{cases} 1, & \text{если содержится ОУ;} \\ 0, & \text{в противном случае.} \end{cases}$$

*Точность* – свойство ПИКС обеспечивать достижение согласованных результатов обработки и передачи информации, необходимых для получения достоверной выходной информации:

$$T_{\text{ПИКС}} = \frac{N_{\text{ТОЧН.БД}}}{t} = \frac{N_{\text{ТОЧН.ОБРАБ.БД}}}{t} + \frac{N_{\text{ТОЧН.ДОСТ.БД}}}{t} =$$

$$= \frac{\alpha \cdot N_{\text{СООБЩ.}} - N_{\text{ИСКАЖ.СООБЩ.}}}{t} + \frac{(\alpha-1) \cdot N_{\text{СООБЩ.}} - N_{\text{ПОТЕР.БД}}}{t}, \quad (5)$$

где  $N_{\text{ТОЧН.БД}}$  – общее количество точно обработанных  $N_{\text{ТОЧН.ОБРАБ.БД}}$  и доставленных  $N_{\text{ТОЧН.ДОСТ.БД}}$  БД (сообщений);  $N_{\text{СООБЩ.}}$  – общее количество сообщений, сформированных в результате обработки полимодальной информации и переданных в виде БД;  $\alpha \in (0,1)$  – коэффициент пропорциональности;  $N_{\text{ИСКАЖ.СООБЩ.}}$  – количество искаженных сообщений, не позволяющих получить достоверную информацию;  $N_{\text{ПОТЕР.БД}} = N_{\text{СООБЩ.}} \cdot PLR$  – количество потерянных (доставленных не по назначению) БД;  $PLR$  (*Packet Loss Ratio*) – коэффициент потери пакетов (БД).

*Своевременность* – свойство ПИКС обеспечивать представление выходной информации в сроки, гарантирующие выполнение соответствующей функции согласно целевому назначению системы:

$$C^{\text{ПИКС}} = t_{\text{ОБР}} + (t_3 \pm \Delta t_3), \quad (6)$$

где  $t_{\text{ОБР}}$  – время ввода/вывода и обработки полимодальной информации;  $(t_3 \pm \Delta t_3)$  – задержка передачи сообщений.

*Производительность*  $Pr^{\text{ПИКС}}$  – суммарная скорость передачи данных (БД, сообщений).

Требования к указанным показателям качества [15] представлены ниже в таблице 1.

**4. Удельная себестоимость ПИКС и ее компонентов.** Для оценки эффективности ПИКС предлагается использовать обобщенный показатель, представляющий собой ее гипотетический объем:

$$V^{\text{ПИКС}} = Pr^{\text{ПИКС}} \cdot B^{\text{ПИКС}} \cdot C^{\text{ПИКС}}, \quad (7)$$

где  $B^{\text{ПИКС}} = 1/C^{\text{ПИКС}}$  – быстродействие системы. Данный показатель вводится для обеспечения измерения гипотетического объема ПИКС и ее компонентов в  $\text{с}^{-3}$ .

С одной стороны, данный ОПЭФ включает в себя все основные показатели качества функционирования системы [23], с другой – все ее компоненты взаимосвязаны между собой, но одного или двух из них недостаточно для системного отражения свойств ИКС.

Таблица 1. Целевые показатели качества, воспринимаемого абонентом

Приложение	Типовые скорости передачи данных	$t_3^{\text{доп}}$	$\Delta t_3$	Потери информации***
Телефония	4–64 кбит/с	< 150 мс*; < 400 мс**	< 1 мс	PLR < 3 %
Передача голосовых сообщений	4–32 кбит/с	< 1 с – для воспр. < 2 с – для записи	< 1 мс	PLR < 3 %
Высококачественное потоковое аудио	16–128 кбит/с***	< 10 с	« 1 мс	PLR < 1 %
Видеотелефония	16–384 кбит/с	< 150 мс*; < 400 мс**		PLR < 1 %
Передача видео	16–384 кбит/с	< 10 с		PLR < 1 %
Web-навигация	≈ 10 кБайт/с	< 2 с/страница*; < 4 с/страница**	Не применяется	0
Передача массивов данных	10–10 <sup>4</sup> МБайт/с	< 15 с*; < 60 с**		0
Осуществление транзакций	< 10 кБайт/с	< 2 с*; < 4 с**		0
Команды (управление)	≈ 1 кБайт/с	< 250 мс		0
Неподвижное изображение	< 100 кБайт/с	< 15 с*; < 60 с**		0
Электронная почта (доступ к серверу)	< 10 кБайт/с	< 2 с*; < 4 с**		0
Электронная почта (сервер-сервер)	< 10 кБайт/с	Может достигать нескольких минут		0

Выбор рабочей точки в объеме  $\sup V^{\text{ПИКС}}$  (рис. 4, а) может быть определен исходя из различных соображений. Если пространство  $\Phi$  качества функционирования ПИКС метрическое, то имеет место некоторое отображение:

$$V_i^{\text{ПИКС}} : H_i^{\text{ПИКС}} \rightarrow \Phi, \quad V_i^{\text{ПИКС}} = Pr_i^{\text{ПИКС}} \cdot B_i^{\text{ПИКС}} \cdot C_i^{\text{ПИКС}},$$

где  $V_i^{\text{ПИКС}}$  – неотрицательная счетно-аддитивная функция, представляющая собой меру на пространстве показателей качества функциони-

рования ПИКС, а  $H^{\text{ПИКС}}$  – ее нагрузочная характеристика. Каждой паре  $H_i^{\text{ПИКС}} \in \Phi, H_j^{\text{ПИКС}} \in \Phi$  соотносится вещественное число  $\rho(H_i^{\text{ПИКС}}, H_j^{\text{ПИКС}})$ , т. е. расстояние между элементами пространства, удовлетворяющее условиям теории меры.

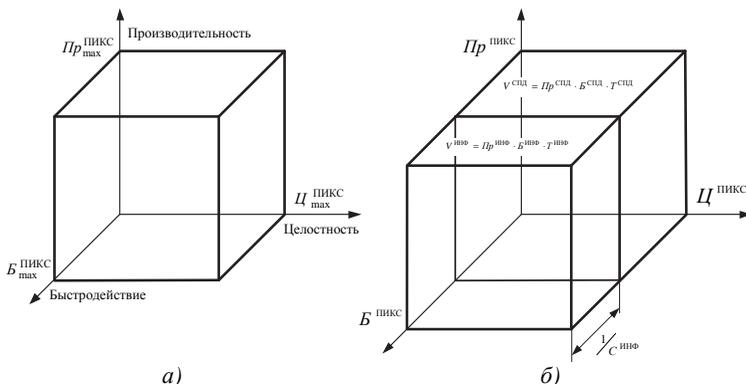


Рис. 4. Эффективность ПИКС (а) и декомпозиция ее показателей (б)

Показатели качества функционирования ПИКС являются компонентами вектора  $R$  в пространстве  $\Phi = (P, B, C)$ . Процесс функционирования в  $\Phi$  ИКС и любого из ее элементов может происходить только в рамках их максимальных  $Pr_{\text{мах}}^{\text{ПИКС}}, B_{\text{мах}}^{\text{ПИКС}}, C_{\text{мах}}^{\text{ПИКС}}$  нагрузочных характеристик  $H^{\text{ПИКС}}$ . Для сравнения различных вариантов построения ПИКС при ограничении (3) предлагается использовать показатель удельной себестоимости [17]:

$$\zeta^{\text{ПИКС}} = \frac{V^{\text{ПИКС}}}{C_{\text{пр}}^{\text{ПИКС}}}, \quad (8)$$

максимальное значение которого определяет лучшее решение.

С учетом принятой иерархии (рис. 2) функциональных характеристик ПИКС показатель (8) может быть декомпозирован на удельную себестоимость информации (9) и удельную себестоимость СПД (10):

$$\zeta^{\text{ИНФ}} = \frac{V^{\text{ИНФ}}}{C_{\text{пр}}^{\text{ИНФ}}}, \quad (9) \quad \zeta^{\text{СПД}} = \frac{V^{\text{СПД}}}{C_{\text{пр}}^{\text{СПД}}}, \quad (10)$$

$$V^{\text{ИНФ}} = Pr^{\text{ИНФ}} \cdot B^{\text{ИНФ}} \cdot C^{\text{ИНФ}}, \quad (11) \quad V^{\text{СПД}} = Pr^{\text{СПД}} \cdot B^{\text{СПД}} \cdot C^{\text{СПД}}, \quad (12)$$

$$B^{\text{ИНФ}} = \frac{1}{C^{\text{ИНФ}}} = \frac{1}{t_{\text{ОБР}}}, \quad (13) \quad B^{\text{СПД}} = \frac{1}{C^{\text{СПД}}} = \frac{1}{(t_3 \pm \Delta t_3)}, \quad (14)$$

$$Pr^{\text{ИНФ}} = \sum_{j=1}^J B_j, \quad (15) \quad Pr^{\text{СПД}} = \sum_{k=1}^K U_k, \quad (16)$$

где  $V^{\text{ИНФ}}$  и  $V^{\text{СПД}}$  – гипотетический объем обрабатываемой информации (сообщений) и СПД соответственно (рис. 4, б);  $Pr^{\text{ИНФ}}$  и  $Pr^{\text{СПД}}$  – производительность источника информации и СПД соответственно;  $C^{\text{ИНФ}}$  и  $C^{\text{СПД}}$  – своевременность обработки и передачи информации (сообщений) соответственно;  $B_j$  – скорость выдачи информации  $j$ -м источником (сигналы различной природы – для традиционных ИКС; модальности – для ПИКС);  $J$  – число таких источников;  $U_k$  – ресурс пропускной способности для сообщений  $k$ -го типа;  $K$  – число одновременно передаваемых сообщений (потоков БД).

Учитывая, что полнота информации, передаваемой в виде БД по СПД,  $P^{\text{СПД}} = 1$ , целостность  $C^{\text{СПД}}$  будет полностью определяться точностью передачи данных (сообщений):

$$T^{\text{СПД}} = \frac{N_{\text{Точн. Дост. БД}}}{t},$$

а гипотетический объем (12) примет вид:

$$V^{\text{СПД}} = Pr^{\text{СПД}} \cdot B^{\text{СПД}} \cdot T^{\text{СПД}}. \quad (17)$$

Тогда задачу по поиску наилучшего варианта построения СПД ПИКС целесообразно осуществлять на основе выбора альтернатив, структур и функциональных характеристик компонентов, обеспечивающих максимизацию удельной себестоимости (10) с учетом выражений (14), (16) и (17). Указанный подход является апробированным инструментарием оценки эффективности сетей связи [10, 17, 20, 23].

При обеспечении требований к производительности  $Pr^{\text{ИНФ}} \leq Pr^{\text{СПД}}$  и своевременности  $C^{\text{ИНФ}} + C^{\text{СПД}} \leq 1/B_{\text{max}}^{\text{ПИКС}}$  ключевым показателем удельной себестоимости информации будет являться целостность информации:

$$C^{\text{ИНФ}} = P^{\text{ИНФ}} \times T^{\text{ИНФ}}, \quad (18)$$

характеризующая полноту и точность отражения требуемых состояний реально существующих ОУ.

Информацию можно считать полной, когда она содержит минимальный, но достаточный для принятия правильного решения набор показателей. Как неполная, так и избыточная информация снижает эффективность принимаемых на основании информации решений.

С учетом выражения (4) целостность полимодальной информации можно определить как:

$$C^{\text{ИНФ}} = \frac{1}{M} \sum_{m=1}^M x_m^{\text{ОУ}} D_m, \quad (19)$$

где  $D_m$  – достоверность оценки состояния  $m$ -го ОУ. При этом под *достоверностью* следует понимать свойство полимодальной информации отражать реальное или оцениваемое состояние объектов и процессов прикладной области (ОУ) со степенью приближения (*точностью*), обеспечивающей эффективное использование этой информации согласно целевому назначению системы.

Следует учесть, что при такой интерпретации размерность показателя  $C^{\text{ИНФ}}$  ( $c^{-1}$ ) при вероятностной оценке достоверности  $D_m$  сохраняется за счет введения соответствующей размерности для показателей  $x_m^{\text{ОУ}}$ .

**5. Сравнительная оценка эффективности ПИКС и ИКС с аддитивным наращиванием модальностей.** Предположим, что при коммуникативном взаимодействии наряду с обменом сообщениями абонентов интересует их истинность (ложность). ОУ в данном случае становятся смысл (с показателем полноты  $x_1^{\text{ОУ}}$ ) и оценка истинности (с показателем  $x_2^{\text{ОУ}}$ ) сообщения. Тогда целостность передаваемой информации (19) можно определить следующим образом:

$$C^{\text{ИНФ}} = \frac{1}{2}(x_1^{\text{ОУ}} D_1 + x_2^{\text{ОУ}} D_2), \quad (20)$$

где  $D_1$  и  $D_2$  – достоверность определения смысла и истинности передаваемых сообщений соответственно.

Будем считать, что при осуществлении сеанса видеотелефонии (табл. 1) абоненты способны полностью определить смысл передаваемых сообщений ( $x_1^{\text{ОУ}} = 1$  при  $D_1 = 1$ ), но не имеют специальной подготовки для определения истинности передаваемой информации ( $x_2^{\text{ОУ}} = 0$ ). Тогда с учетом введенной декомпозиции, а также выражений (9), (11), (13) и (19) для эффективности соответствующей бимодальной системы можно записать:

$$\zeta_1^{\text{ИНФ}} \Big|_{\zeta^{\text{СПД}}=\text{const}} = \frac{Pr^{\text{ИНФ}}}{2 \cdot t_{\text{ОБР}} \cdot C_{\text{пр}}^{\text{ИНФ}}}. \quad (21)$$

Для определения истинности в таких системах необходимо применение дополнительных аппаратно-программных средств (например, полиграфа) [21]. Использование последних неизбежно приведет к увеличению времени обработки  $t_{\text{ОБР}}$  входных сигналов и/или увеличению затрат  $C_{\text{пр}}^{\text{ИНФ}}$  на их реализацию, а в итоге – к уменьшению удельной себестоимости ИКС с аддитивным наращиванием модальностей:

$$\zeta_2^{\text{ИНФ}} \Big|_{\zeta^{\text{СПД}}=\text{const}} < \zeta_1^{\text{ИНФ}} \Big|_{\zeta^{\text{СПД}}=\text{const}}.$$

При полимодальном представлении информации возможна оценка ложности сообщаемой информации по динамике параметров (рис. 5) невербального поведения абонента ( $x_2^{\text{ОУ}} = 1$ ) [3].

Требуемые для реализации такой оценки модальности относятся к визуальному каналу коммуникации [2], характерному для видеотелефонии, а точность оценки их параметров определяет достоверность определения истинности  $D_2$ . Удельная себестоимость соответствующей ПИКС:

$$S_3^{\text{ИНФ}} \Big|_{S^{\text{СПД}} = \text{const}} = \frac{Pr^{\text{ИНФ}} \cdot D_2}{2 \cdot t_{\text{ОБР}} \cdot C_{\text{пр}}^{\text{ИНФ}}}, \quad (22)$$

будет выше ОПЭФ (22).

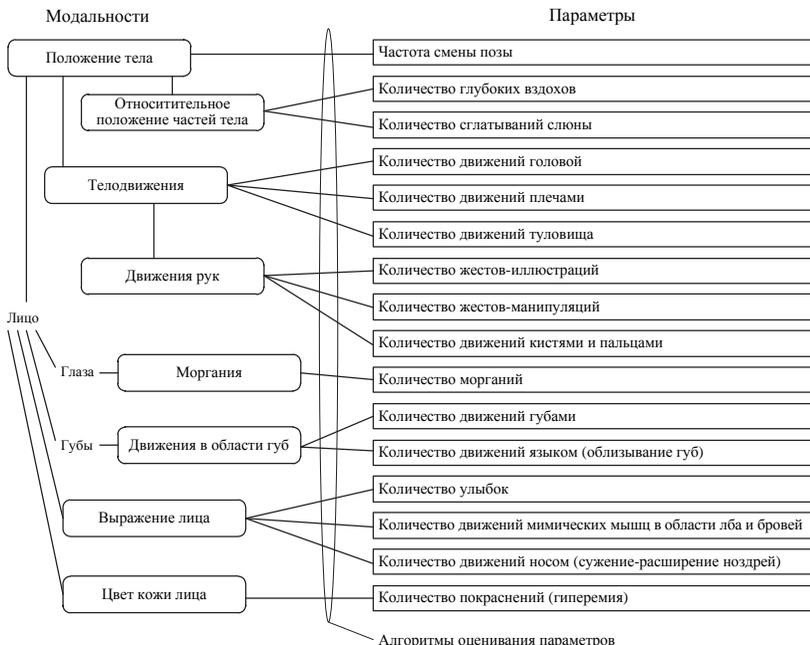


Рис. 5. Параметры невербального поведения человека, используемые для определения истинности (ложности), передаваемой им информации

Таким образом, при заданной удельной себестоимости СПД (10) и ограничении приведенных затрат на обработку информации  $C_{\text{пр}}^{\text{ИНФ}} \leq C_{\text{пр}}^{\text{ИНФ доп}}$  степень обеспечения целостности полимодальной информации (19) будет определять внешнюю эффективность ПИКС.

**6. Внутренняя эффективность ПИКС.** Необходимость согласования производительностей источника информации (АТ) и СПД ( $Pr^{\text{ИНФ}} \leq Pr^{\text{СПД}}$ ) обусловила целесообразность введения единицы канального ресурса (ЕКР) [19]:

$$\kappa = \text{НОД}(B_1, \dots, B_J, U_1, \dots, U_K), \quad (23)$$

где НОД – наименьший общий делитель целочисленных значений скоростей выдачи информации различными источниками (источниками различных модальностей) и требований к ресурсу пропускной способности, необходимому для передачи полимодальной информации.

Очевидно, что такая величина (23) будет меньше любого "шага квантования" производительности, используемого в традиционных задачах синтеза телекоммуникационных систем, что свидетельствует об увеличении числа степеней свободы при решении задач синтеза ПИКС. Результаты моделирования канального ресурса на базе ЕКР являются более прецизионными, а решение задач выбора скорости передачи (выдачи) – более точным.

Следовательно, в соответствии с принципом дополнительности Бора (в расширенной трактовке) ЕКР (23) может быть использована в качестве показателя внутренней эффективности ПИКС.

Однако в традиционных ИКС каждая ЕКР – результат применения конкретной технологии кодирования и передачи (табл.). Следовательно, для «новых услуг» ПИКС [2] должны применяться «свои» ЕКР, что с точки зрения эффективности использования ресурса СПД может оказаться расточительным. Используя результаты исследования сигналов различных модальностей, можно предположить, что в АТ ПИКС есть возможность организовать необходимое число ЕКР, способных в различных ситуациях обеспечить реализацию информационных, телекоммуникационных, перцептивных и интеллектуальных услуг, что требует дополнительных теоретических исследований.

**7. Заключение.** В настоящей работе во взаимосвязи рассмотрены два аспекта оценки эффективности функционирования ИКС: внешний и внутренний. Первый аспект заключается в определении того вклада в достижение цели метасистемы (системы управления), который вносит ПИКС. Во втором случае осуществляется оценка того, насколько близко количественные значения функциональных характеристик ПИКС соответствуют требуемым (или потенциально возможным) значениям.

Внешняя эффективность ПИКС определяется максимальным отклонением фактического качества ее функционирования от расчетного значения при ограничении на приведенные затраты на ее построение и эксплуатацию. При заданной удельной себестоимости СПД и ограничении приведенных затрат на обработку информации внешней эффективностью ПИКС определяет степень обеспечения целостности полимодальной информации.

Внутренняя эффективность ПИКС определяется минимальным числом ЕКР, требуемых для реализации информационных, телеком-

муникационных, перцептивных и интеллектуальных услуг. Использование данного показателя позволяет увеличить вариативность решения задач синтеза ПИКС, однако требует проведения дополнительных теоретических исследований возможности организации необходимого числа ЕКР.

### **Литература**

1. *Басов О.О.* Предпосылки создания полимодальных инф коммуникационных систем // Сборник материалов Международной научно-практической конференции «Неделя науки СПбГПУ». Санкт-Петербургский государственный политехнический университет// Секция «Решение сложных задач в области современных информационных и компьютерных технологий». 3-8 декабря 2012. С. 5–6.
2. *Басов О.О., Саитов И.А.* Основные каналы межличностной коммуникации и их проекция на инфокоммуникационные системы // Труды СПИИРАН. 2013. Вып. 7(30). С. 122–140.
3. *Басов О.О., Саитов И.А.* Способ определения ложности передаваемой информации по динамике параметров невербального поведения человека // патент № 2506048; Российская Федерация: МПК А61В 5/16. – № 2012145655/14; заявл. 25.10.2012; опубл. 10.02.2014, Бюл. № 4. 14 с.: ил.
4. *Вишневецкий В.М.* Теоретические основы проектирования компьютерных сетей // М.: Техносфера, 2003. 512 с.
5. ГОСТ РВ 51987-2002. Информационная технология. Комплекс стандартов на автоматизированные системы типовые требования и показатели качества функционирования информационных систем. Общие положения // М.: Госстандарт России, 2001.
6. *Захаров Г.П.* Методы исследования сетей передачи данных // М.: Радио и связь, 1982. 208 с.
7. *Лукьяница А.А., Шишкин А.Г.* Автоматическое определение изменений эмоционального состояния по речевому сигналу // Речевые технологии. 2009. № 3. С. 60–76.
8. *Одоевский С.М.* Адаптивно-игровая оптимизация функциональных характеристик военных сетей многоканальной радиосвязи // моногр. СПб.: ВАС, 1998. 300 с.
9. *Петухов Г.Б.* Основы теории эффективности целенаправленных процессов // учебник. Ч.1. Методология, методы, модели. Министерство обороны СССР, 1989. 656 с.
10. *Расчесова А.Г.* Нагруженная характеристика в оценке эффективности сети ПД и ее системы технической эксплуатации // Техника средств связи. 1990. № 10. С. 68–79.
11. Рекомендация МСЭ-Т E.802. Принципы и методики определения и применения параметров QoS. Официальный сайт Международного союза электросвязи. URL: [www.itu.int](http://www.itu.int) (дата обращения: 26.10.2013).
12. Рекомендация МСЭ-Т E.800. Определение терминов, относящихся к качеству обслуживания. Официальный сайт Международного союза электросвязи. URL: [www.itu.int](http://www.itu.int) (дата обращения: 26.10.2013).
13. Рекомендация МСЭ-Т E.430. Система качества обслуживания. Официальный сайт Международного союза электросвязи. URL: [www.itu.int](http://www.itu.int) (дата обращения: 26.10.2013).
14. Рекомендация МСЭ-Т I.350. Основные положения качества обслуживания и производительности сети в цифровых сетях, включая ЦСИО. Официальный сайт

Международного союза электросвязи. URL: [www.itu.int](http://www.itu.int) (дата обращения: 26.10.2013).

15. Рекомендация МСЭ-Т G.1010. Категории качества обслуживания конечного пользователя. Официальный сайт Международного союза электросвязи. URL: [www.itu.int](http://www.itu.int) (дата обращения: 26.10.2013).
16. Ронжин А.Л., Карнов А.А., Ли И.В. Речевой и многомодальный интерфейсы // М.: Наука, 2006. 173 с.
17. Саитов И.А. Основы теории построения защищенных мультипротокольных оптических транспортных сетей телекоммуникационных систем // монограф. Орел: Академия ФСО России, 2008. 220 с.
18. Смирнов А.В., Пашкин М.П., Шилов Н.Г., Левашова Т.В. Подход к построению распределенной системы интеллектуальной поддержки принятия решений в открытой информационной среде // Труды СПИИРАН. 2007. Вып. 4. С. 36–49.
19. Степанов С.Н. Основы телеграфика мультисервисных сетей // М.: Эко-Трендз, 2010. 392 с.
20. Тимофеев В.Ф. и др. Концептуально-математическая модель системы эксплуатации сети связи // Информация и космос. 2004. №3. С. 55–60.
21. Фрай О. Детекция лжи и обмана // Пер. с англ. СПб.: Прайм-Еврознак, 2005. 320 с.
22. Царегородцев А.В., Кислицын А.С. Основы синтеза защищенных телекоммуникационных систем // под ред. Е.М. Сухарева. М.: Радиотехника, 2006. 256 с.
23. Цыбизов А.А. Оценка эффективности сетей связи // Сборник. Вестник РГПУ. 2009. Вып. 3(29). С. 19–24.
24. Krinidis S., Buciu I., Pitas I. Facial expression analysis and synthesis: A survey // In 10th International Conference on Human-Computer Interaction (HCI 2003), 2003. pp. 22–27.

## References

1. Basov O.O. [Prerequisites to making polymodal infocommunicational systems]. *Sbornik materialov Mezhdunarodnoj nauchno-prakticheskoj konferencii «Nedelja nauki SPbGPU». Sankt-Peterburgskij gosudarstvennyj politehnicheskij universitet, 3-8 dekabrja Sekcija «Reshenie slozhnyh zadach v oblasti sovremennyh informacionnyh i komp'juternyh tehnologii»*. [Science Week SPbSPU: Collected papers] 2012. pp. 5–6. (In Russ.).
2. Basov O.O., Saitov I.A. [The main channels of interpersonal communication and their projection on the infocommunication systems]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2013. vol. 7(30). pp. 122–140. (In Russ.).
3. Basov O.O., Saitov I.A. *Sposob opredelenija lozhnosti peredavaemoj informacii po dinamike parametrov neverbal'nogo povedenija cheloveka* [A method for determining the falsity of the transmitted information on the dynamics of human nonverbal behavior parameters]. Patent RF №2506048. 2014. (In Russ.).
4. Vishnevskij V.M. *Teoreticheskie osnovy proektirovanija komp'juternyh setej* [Theoretical bases of designing computer networks]. М.: Tehnosfera, 2003. 512 p. (In Russ.).
5. GOST RV 51987-2002. [Information technology. Standards' complex for automated systems standard requirements and indicators of quality of information systems. General provisions]. М.: Gosstandart Rossii, 2001. (In Russ.).
6. Zaharov G.P. *Metody issledovanija setej peredachi dannyh* [Methods of research data networks]. М.: Radio i svjaz', 1982. 208 p. (In Russ.).
7. Luk'janica A.A., Shishkin A.G. [Automatic detection of changes in emotional state on the voice signal]. *Rechevyje tehnologii – Speech technology*. 2009. no. 3 pp. 60–76. (In Russ.).

8. Odoevskij S.M. *Adaptivno-igrovaja optimizacija funkcional'nyh karakteristik voennyh setej mnogokanal'noj radiosvjazi* [Adaptable-playing optimization of the functional characteristics of the military networks multichannel radio traffic]. SPb.: VAS. 1998. 300 p. (In Russ.).
9. Petuhov G.B. *Osnovy teorii jeffektivnosti celenapravlennyh processov* [Fundamentals of the theory of the effectiveness of targeted processes]. Metodologija, metody, modeli. Ministerstvo oborony USSR, 1989. 656 p. (In Russ.).
10. Raschesova A.G. [Load characteristic in assessing the effectiveness of the network of PD and its technical operation of the system]. *Tehnika sredstv svjazi – Inteltech*. 1990. no. 10. pp. 68–79. (In Russ.).
11. Rekomendacija MSJe-T E.802. Principy i metodiki opredelenija i primeneniya parametrov QoS [Principles and methods of determination and application of QoS parameters]. Oficial'nyj sajt Mezhdunarodnogo sojuza jelektrosvjazi [Official web site of International Telecommunication Union]. Available at: [www.itu.int](http://www.itu.int). (accessed 26.10.2013). (In Russ.).
12. Rekomendacija MSJe-T E.800. Opredelenie terminov, odnosjashhhsja k kachestvu obsluzhivaniya [Definition of terms relating to the quality of service]. Oficial'nyj sajt Mezhdunarodnogo sojuza jelektrosvjazi [Official web site of International Telecommunication Union]. Available at: [www.itu.int](http://www.itu.int). (accessed 26.10.2013). (In Russ.).
13. Rekomendacija MSJe-T E.430. Sistema kachestva obsluzhivaniya [System of service quality]. Oficial'nyj sajt Mezhdunarodnogo sojuza jelektrosvjazi [Official web site of International Telecommunication Union]. Available at: [www.itu.int](http://www.itu.int). (accessed 26.10.2013). (In Russ.).
14. Rekomendacija MSJe-T I.350. Osnovnye polozhenija kachestva obsluzhivaniya i proizvoditel'nosti seti v cifrovyh setjah, vkljuchaja CSIO [Main provisions of service quality and network performance in digital networks, including ISDN]. Oficial'nyj sajt Mezhdunarodnogo sojuza jelektrosvjazi [Official web site of International Telecommunication Union]. Available at: [www.itu.int](http://www.itu.int). (accessed 26.10.2013). (In Russ.).
15. Rekomendacija MSJe-T G.1010. Kategorii kachestva obsluzhivaniya konechnogo pol'zovatelja [Categories of service quality of end-user]. Oficial'nyj sajt Mezhdunarodnogo sojuza jelektrosvjazi [Official web site of International Telecommunication Union]. Available at: [www.itu.int](http://www.itu.int). (accessed 26.10.2013). (In Russ.).
16. Ronzhin A.L., Karpov A.A., Li I.V. *Rechevoj i mnogomodal'nyj interfejsy* [Speech and multimodal interfaces]. M.: Nauka, 2006. 173 p. (In Russ.).
17. Saitov I.A. *Osnovy teorii postroenija zashhishhennyh mul'tiprotokol'nyh opticheskikh transportnyh setej telekommunikacionnyh sistem* [Fundamentals of the theory of building secure multiprotocol optical transport networks of telecommunication systems]. Orel: Akademija FSO Rossii, 2008. 220 p. (In Russ.).
18. Smirnov A.V., Pashkin M.P., Shilov N.G., Levashova T.V. [Approach to building distributed systems intelligent decision support in an open information environment]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2007. vol. 4. pp. 36–49. (In Russ.).
19. Stepanov S.N. *Osnovy telegrafika mul'tiservisnyh setej* [Fundamentals of teletraffic of multiservice networks]. M.: Jeko-Trendz, 2010. 392 p. (In Russ.).
20. Timofeev V.F. i dr. [Conceptually mathematical model of the communication network] *Informacija i kosmos – Infocosmo* 2004. no. 3. pp. 55–60. (In Russ.).
21. Fraj O. *Detecting Lies and Deceit*. Wiley-Blackwell. 2000. 276 p. (Russ. ed.: Ershova A., Isakova O., Kulakov A., Mironov N., Smirnov A. *Detekcija lzhi i obmana*. SPb.: Prajm-Evroznak, 2005. 320 p. (In Russ.)).
22. Caregorodcev A.V., Kislicyn A.S. *Osnovy sinteza zashhishhennyh telekommunikacionnyh sistem* [Basics synthesis of protected telecommunication systems] M.: Radiotekhnika, 2006. 256 p. (In Russ.).

23. Cybizov A.A. [Estimation of efficiency of communication networks]. Sbornik. *Vestnik RGRU – Herald of RSREU*. 2009. vol. 3(29). pp. 19–24. (In Russ.).
24. *Krinidis S., Buciu I., Pitas I.* Facial expression analysis and synthesis: A survey. In 10th International Conference on Human-Computer Interaction (HCI 2003). 2003. pp. 22–27.

**Басов Олег Олегович** — к-т техн. наук, докторант Академии ФСО России. Область научных интересов: обработка и кодирование речевых и иконических сигналов, проектирование полимодальных инфокоммуникационных систем. Число научных публикаций — 135. [Oobasov@mail.ru](mailto:Oobasov@mail.ru); Академия ФСО России, Приборостроительная, 35, г. Орел, 302034, РФ; р.т. +79192011897. Научный консультант — И.А. Сaitов.

**Basov Oleg Olegovich** — Ph.D., Doctoral of the Academy of Federal Agency of protection of Russian Federation. Research interests: processing and coding of speech and iconic signals, polymodal infocommunicational systems design. Number of scientific publications — 135. [oobasov@mail.ru](mailto:oobasov@mail.ru); Academy of Federal Agency of protection of Russian Federation, Priborostroitelnaya Street, 35, Orel, 302034, Russia, office phone +79192011897. Scientific adviser – Saitov I.A.

**Сaitов Игорь Акрамович** — д-р техн. наук, профессор, начальник факультета Академии ФСО России. Область научных интересов: анализ и синтез инфокоммуникационных сетей и систем. Число научных публикаций — 160. [Akramovish@mail.ru](mailto:Akramovish@mail.ru); Академия ФСО России, Приборостроительная, 35, г. Орел, 302034, РФ; р.т. +79103034668.

**Saitov Igor Akramovich** — Ph.D., Dr. Sci., professor, dean, Academy of Federal Agency of protection of Russian Federation. Research interests: analysis and synthesis of infocommunicational networks and systems. The number of publications — 160. [Akramovish@mail.ru](mailto:Akramovish@mail.ru); Academy of Federal Agency of protection of Russian Federation, Priborostroitelnaya Street, 35, Orel, 302034, Russia, office phone +79103034668.

## РЕФЕРАТ

### *Басов О.О., Саитов И.А.* **Качество функционирования и эффективность полимодальных инфокоммуникационных систем.**

Основой новой теории полимодальных инфокоммуникационных систем является количественная оценка степени достижения заданных целей ее функционирования. В статье во взаимосвязи рассмотрены два аспекта оценки эффективности функционирования инфокоммуникационных систем: внешний и внутренний.

Первый аспект заключается в определении того вклада в достижение цели метасистемы (системы управления), который вносит полимодальная инфокоммуникационная система. Во втором случае осуществляется оценка того, насколько близко количественные значения функциональных характеристик полимодальных инфокоммуникационных систем соответствуют требуемым (или потенциально возможным) значениям.

Внешняя эффективность определена как максимальное отклонение качества функционирования полимодальной инфокоммуникационной системы от расчетного значения при ограничении на приведенные затраты на ее построение и эксплуатацию. Интеграция информационных и телекоммуникационных систем при формировании системотехнического облика полимодальных инфокоммуникационных систем обусловила выбор в качестве показателей качества: целостности, своевременности и производительности. На основе их формального представления в качестве обобщенного показателя эффективности предложена удельная себестоимость полимодальной инфокоммуникационной системы.

Показано, что при заданной удельной себестоимости транспортной инфраструктуры (сети передачи данных) и ограничении приведенных затрат на обработку информации внешнюю эффективность полимодальных инфокоммуникационных систем будет определять степень обеспечения целостности полимодальной информации.

Необходимость согласования производительностей источника полимодальной информации и пропускной способности сети передачи данных обусловила целесообразность введения единицы канального ресурса и выбора ее в качестве показателя внутренней эффективности. Его использование позволяет увеличить вариативность решения задач синтеза полимодальных инфокоммуникационных систем, однако требует проведения дополнительных теоретических исследований возможности организации необходимого числа единиц канального ресурса.

## SUMMARY

### *Basov O.O., Saitov V.A.* **Functioning quality and effectiveness of polymodal infocommunicational systems.**

The basis of a new theory of polymodal infocommunication systems is the quantitative evaluation of the degree of achievement of intended purposes of its functioning. In reference to each other two aspects of functioning performance evaluation of the infocommunication systems: external and internal are discussed in the article.

The first aspect lies in the fact to determine the contribution to the achievement of purpose of metasystem (control system), which introduces a polymodal infocommunication system. In the second case is an assessment of how close the quantitative values of the functional characteristics of polymodal infocommunication systems meet required (or potentially possible) values.

External efficiency is defined as the maximum deviation of the functioning quality of polymodal infocommunicational system from rated value when restricted to the reduced costs of its construction and operation. Integration of information and communication systems in the formation of circuit analysis appearance infocommunication systems led to selection as indicators of quality: integrity, timeliness and production. On the basis of their formal representation as a generalized indicator of efficiency a polymodal infocommunication system's unit cost is proposed.

It is shown that at a given unit cost of transport infrastructure (data network) and the limitation of reduced expenditures on information processing efficiency of polymodal infocommunication systems will determine the degree of integrity of polymodal information.

Availability of polymodal information source rate coordination and network capacity have required introducing unit channel resource and selecting it as an indicator of internal efficiency. Its use can increase the variability of solving synthesis polymodal infocommunication systems, however, it requires additional theoretical studies of the possibility of organizing the required number of channel provision units.

БАТЕНКОВ К.А.  
**МОДЕЛИРОВАНИЕ НЕПРЕРЫВНЫХ КАНАЛОВ СВЯЗИ В  
ФОРМЕ ОПЕРАТОРОВ ПРЕОБРАЗОВАНИЯ НЕКОТОРЫХ  
ПРОСТРАНСТВ**

---

*Батенков К.А. Моделирование непрерывных каналов связи в форме операторов преобразования некоторых пространств.*

**Аннотация.** Предложена феноменологическая модель непрерывного канала связи. На ее основе разработаны модели линейного непрерывного канала связи в форме оператора преобразования метрических пространств с заданными базисами, а также частный случай, в котором базисом является система координатных функций интегрального канонического представления В.С. Пугачева.

**Ключевые слова:** непрерывный канал связи, оператор, базис пространства, интегральное каноническое разложение В.С. Пугачева, системная характеристика, аддитивная помеха, ряд Вольтерра.

*Batenkov K.A. Continuous channel modeling in shape of some space transformation operators.*

**Abstract.** Continuous channel phenomenological model is suggested. Linear continuous channel models in shape of metric space transformation operators with specified basis and also special case of coordinate function basis of Pugachev integral canonical expansion are designed.

**Keywords:** continuous channel, operator, space basis, Pugachev integral canonical expansion, system characteristic, additive noise, Volterra series.

---

**1. Введение.** Современные информационные системы, выполняющие функции передачи данных, разрабатываются применительно к конкретному типу каналу связи, а точнее с учетом требований, предъявляемых используемой средой передачи. В результате эффективность создаваемой системы напрямую связана с адекватностью представления непрерывного канала связи, содержащего как исходную среду передачи, так и некоторую совокупность конечных устройств (аналоговых интерфейсов). При этом теоретический этап конструирования предусматривает в обязательном порядке синтез математической модели непрерывного канала связи, а процедура испытаний и отладки конечного изделия предполагает наличие имитационной либо физической модели, построенных на базе исходной математической. В итоге постоянное стремление разработчиков к повышению эффективности существующих информационных систем, в том числе с точки зрения использования имеющегося в распоряжении ресурса непрерывного канала связи, приводит к актуализации вопросов, связанных с моделированием непрерывных каналов связи.

В настоящее время наиболее часто применяются модели непрерывных каналов связи в форме конечномерных условных распределений вероятностей [18, 34, 43] либо определенных моментных функций [13, 33, 36]. При этом сложность применяемого аппарата приводит к необходимости использования лишь первых двух моментов распределений, что адекватно исключительно для гауссовских случайных процессов. Для распределений подход основывается на применении Марковских процессов [42, 43], что существенно сужает область применимость разрабатываемых моделей, поскольку значительно ограничивает память в канале. Вышеописанные вероятностные модели довольно хорошо укладываются в понятие косвенного описания непрерывных каналов связи, а значит, требуют наличие исходного прямого описания, например феноменологического, непосредственно связывающего входные и выходные сигналы каналов связи, а не их вероятностные характеристики, и более удобного для моделирования в реальном масштабе времени.

Прямое описание каналов связи базируется на двух подходах – использовании преобразований на основе дифференциальных [13] и интегральных уравнений [18, 38, 37, 39]. При этом интегральные уравнения оказываются более предпочтительными в связи с достаточно устоявшимися терминами, такими как импульсная либо переходная характеристика. Однако большинство работ основываются на понимании модели канала связи как некоторого оператора метрического пространства, для которого заданы базисы разложения как для входных так и выходных сигналов в форме гармонических функций [35, 40, 41]. Подобный подход с одной стороны оказывается весьма универсальным, так как синтезируемые модели опираются на пространства с гармоническим базисом, координатные функции которого являются собственными функциями линейных систем на бесконечном интервале анализа, что гарантирует неизменность его формы как на входе так и на выходе канала связи. В результате, нет необходимости в отыскании преобразований входного базиса. С другой же стороны, предположение о бесконечном интервале анализа является достаточно спорным, поскольку системы связи работают с ограниченными по времени сигналами, а значит гармонические функции уже далеки от собственных, следовательно, возникает необходимость поиске соответствующих трансформаций входного базиса в выходной. Именно поэтому данная статья и посвящена попытке разрешить данное противоречие путем описания подходов, позволяющих моделировать непрерывные каналы

при разработке систем связи с требуемой сложностью их аналитического описания.

**2. Феноменологическая модель непрерывного канала связи на основе интегральных и дифференциальных операторов.** Непрерывный канал связи описывается некоторым оператором  $H$ , отображающим все допустимое множество пространственно-временных сигналов на входе  $x(t, \mathbf{r})$  во множество подобных сигналов на выходе  $x'(t', \mathbf{r}')$ . Несмотря на разнообразие воздействующих на передаваемые сообщения факторов в реальных каналах связи, аналитическую (феноменологическую) модель канала целесообразно представить в виде двух слагаемых (рис.) [4, 5, 13]:

$$x'(t', \mathbf{r}') = H'\{x(t, \mathbf{r})\} + n(t', \mathbf{r}'), \quad (1)$$

где  $H'$  – оператор отображения полезной составляющей принимаемой смеси, зависящего от свойств среды передачи, а также характеристик входных и выходных устройств (согласования);

$n(t', \mathbf{r}')$  – поле аддитивных помех, задающее как внешние помехи, так и внутренние шумы;

$t, \mathbf{r}$  – временная и пространственные координаты соответственно.

Поскольку набор случайных и детерминированных факторов, оказывающих существенное влияние на передачу сигналов, чересчур сложен и разнообразен, то наибольшее распространение получили феноменологические модели каналов связи. При этом для феноменологических моделей канал рассматривается как "черный ящик", внутренняя структура которого при описании не учитывается, а основное внимание уделяется воспроизведению взаимозависимостей между входными и выходными сигналами. Кроме того, в работе акцент делается на аналитической форме описания связи между данными сигналами, поскольку именно она (форма) позволяет получать точные решения задач анализа и синтеза каналов связи без проведения экспериментов и последующей статистической обработки их результатов. В результате рассматриваемая в работе феноменологическая модель укладывается в понятие аналитической модели, так как предполагает получение некоторых теоретических преобразований сигналов на входе канала связи в сигналы на выходе. Следует также заметить, что, в отличие от имитационной модели, имеющей чисто алгоритмическую форму представления, феноменологическая модель в общем случае может быть как алгоритмической, так и аналитической, то есть является более общей.

Физическая же модель, по сути, является некоторой неабстрактной интерпретацией, имеющей физическую основу (например, имитатор канала), исходной математической модели, которая хоть и не имеет строго формального определения, однако большинство специалистов под ней подразумевают описание канала связи в виде математических соотношений (не обязательно аналитических), связывающих требуемые его параметры. Таким образом, рассматриваемая в работе феноменологическая модель принадлежит к довольно широкому классу аналитических моделей, который в свою очередь является подклассом математических моделей.

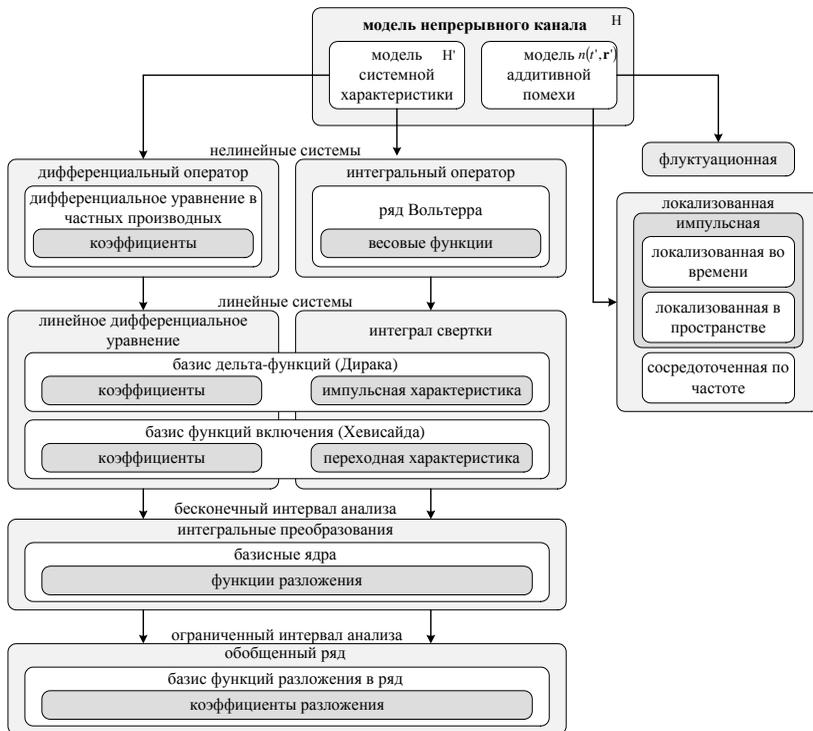


Рис. Структура аналитической (феноменологической) модели непрерывного канала связи

Для подобных моделей оператор  $H'$ , отражающий взаимосвязь входных и выходных сигналов без учета аддитивной помехи, имеет

вид системной характеристики и воспроизводит в общем случае с необходимой точностью статистические характеристики наблюдаемого на выходе сигнала. Исследование системной характеристики канала с постоянными параметрами (детерминированного) сводится к рассмотрению системной характеристики детерминированного типа, которая является частной вариацией стохастической.

Следует также отметить, что при моделировании систем существует два подхода – прямое и косвенное описание [18, 19]. Прямое описание предполагает рассмотрение оператора преобразования множества входных сигналов в выходные, а косвенное – оператора соотношения между их вероятностными мерами. В работе на начальных этапах синтеза математической модели канала используется преимущественно первый подход, поскольку прямые соотношения между случайными величинами более просты и наглядны по сравнению со вторым подходом. К тому же непосредственное косвенное описание модели канала не всегда оказывается возможным, поскольку зачастую не существует прямой взаимосвязи между преобразованиями сигналов в канале связи и их вероятностными характеристиками. На заключительном же этапе синтеза предполагается переход к косвенному описанию на основе формул полной вероятности [18].

Наиболее общими видами моделей системной характеристики  $H'$  являются модели на основе дифференциальных уравнений с частными производными [13, 30]:

$$H'\{x(t, \mathbf{r})\} = \{z(t, \mathbf{r}) \mid \Psi_{t, \mathbf{r}}[x(t, \mathbf{r}), z(t, \mathbf{r}), t, \mathbf{r}] = x(t, \mathbf{r})\}, \quad (2)$$

где  $z(t, \mathbf{r})$  – полезная составляющая принимаемой смеси (без учета аддитивного шума  $n(t', \mathbf{r}')$ );

$\Psi_{t, \mathbf{r}}$  – некоторый оператор, в общем случае нелинейный и содержащий смешанные производные как по времени, так и по пространственным координатам.

Следует отметить, что непосредственно системная характеристика представляет собой коэффициенты этого дифференциального уравнения. Характер решения подобных уравнений, то есть явный вид оператора преобразования  $H'$ , определяется множеством факторов: типом уравнения, формой пространственно-временных сигналов на входе  $x(t, \mathbf{r})$ , начальными условиями. Однако в общем случае аналитических решений для произвольного вида уравнений не существует,

причем даже в квадратурах [17, 30]. Отсутствие обобщенных методов вычисления операторов преобразования  $H'$  полезной составляющей передаваемых сигналов на основе исходных дифференциальных уравнений (2) приводит к необходимости самостоятельного рассмотрения каждого из исследуемых типов уравнений, причем преимущественно на основе численных методов.

Уравнения вида (2) могут быть сведены к широко используемым уравнениям состояния и наблюдения [12, 13]:

$$H'\{x(t, \mathbf{r})\} = \left\{ \Psi'_{\mathbf{r}}[z'(t, \mathbf{r}), t, \mathbf{r}] \mid \frac{\partial z'(t, \mathbf{r})}{\partial t} = \Psi_{\mathbf{r}}[x(t, \mathbf{r}), z'(t, \mathbf{r}), t, \mathbf{r}] \right\}, \quad (3)$$

где  $\Psi_{\mathbf{r}}$ ,  $\Psi'_{\mathbf{r}}$  – некоторые операторы, в общем случае нелинейные и содержащие производные по пространственным координатам;  $z'(t, \mathbf{r})$  – переменные состояния модели канала связи.

В качестве переменных состояния  $z'(t, \mathbf{r})$  могут выступать разнообразнейшие реакции выходного сигнала, например его производные как по времени, так и по пространственным координатам, усиленный или ослабленный сигнал, сигнал, пропущенный через какой-либо фильтр, или же их комбинации. Это свидетельствует о достаточно произвольной интерпретации физического смысла состояний канала связи, которые выбираются исходя из требуемой точности моделирования, простоты и степени абстрактности. Следует также заметить, что в общем случае выражения (3) являются системой дифференциальных уравнений в частных производных (рис.).

Однако для ряда нелинейных систем сведение дифференциальных уравнений с частными производными (2) к уравнениям состояния и наблюдения (3) оказывается невозможным, а зачастую и нецелесообразным [13]. Основными препятствиями при этом являются невыполнимость принципа суперпозиции и условия однородности [19], а также неупорядоченность множества точек пространственных координат (в отличие от времени) [12]. Следует также отметить, что и обратный переход от уравнений состояния и наблюдения (3) к дифференциальным уравнениям (2) не всегда оказывается осуществимым, причем даже для линейных систем [13].

Другой формой описания системной характеристики служит интегральный оператор. Возможность такого представления вытекает из теоремы Фреше [18], которая доказывает возможность аппроксимации непрерывного функционала с любой заданной точностью функцио-

нальным рядом Вольтерра [30]. Для динамических систем с временными и пространственными координатами, к которым (системам) относятся практически все исследуемые модели каналов связи, этот ряд имеет вид:

$$N\{x(t, \mathbf{r})\} = h_0(t', \mathbf{r}') + \sum_{i=1}^n \int \dots \int \left\{ \prod_{j=1}^i x(t_j, \mathbf{r}_j) \right\} \times h_i(t', \mathbf{r}', t_1, \dots, t_i, \mathbf{r}_1, \dots, \mathbf{r}_i) dt_1 \dots dt_i d\mathbf{r}_1 \dots d\mathbf{r}_i, \quad (4)$$

где  $h_i(t', \mathbf{r}', t_1, \dots, t_i, \mathbf{r}_1, \dots, \mathbf{r}_i)$  – некоторые весовые функции (ядра Вольтерра), соответствующие системной характеристике модели канала;  $n$  – число слагаемых аппроксимации по временной и пространственным координатам соответственно;

$h_0(t', \mathbf{r}')$  – начальное значение выходного сигнала (постоянная составляющая).

По сути, ряд Вольтерра является решением исходных дифференциальных уравнений (2) или (3) при условии возможности его представления в квадратурах. Так, для уравнений с частными производными (2) таким условием оказывается аддитивный вид оператора  $\Psi_{t, \mathbf{r}}$ , распадающегося на два слагаемых, каждое из которых зависит лишь от одного типа сигнала (входного или выходного) [30]. В случае невозможности подобно представлению ряд Вольтерра может быть получен на основе аппроксимации исходного дифференциального уравнения в соответствии с требуемым критерием, например минимума ошибки аппроксимации.

По аналогии с уравнениями состояния и наблюдения в дифференциальной форме (3) можно рассматривать такого же рода уравнения в интегральной форме. По аналогии вводятся переменные состояния  $z'(t, \mathbf{r})$  с той лишь разницей, что операторы имеют интегральный вид:

$$N\{x(t, \mathbf{r})\} = \left\{ \Psi'_{t, \mathbf{r}} [z'(t, \mathbf{r}), t, \mathbf{r}] \mid z'(t, \mathbf{r}) = \Psi_{t, \mathbf{r}} [x(t, \mathbf{r}), t, \mathbf{r}] \right\}, \quad (5)$$

где  $\Psi_{t, \mathbf{r}}$ ,  $\Psi'_{t, \mathbf{r}}$  – некоторые интегральные операторы (по пространству и времени), в общем случае нелинейные.

Переход от интегральных уравнений (4) или (5) к дифференциальным в соответствующей форме (2) или (3) возможен путем дифференцирования по пространственным и временным координатам [13].

Таким образом, дифференциальные уравнения позволяют описывать соотношения между входными и выходными сигналами на основе коэффициентов, а ряды Вольтерра – путем учета весовых функций в интегральных преобразованиях. Основное отличие моделей "вход-состояние-выход" (3) и (5) от моделей "вход-выход" (2) и (4) заключается в представлении априорных данных о предыстории процессов. Так, вторые требуют в общем случае знания о всей совокупности значений сигналов до настоящего момента, а первые – состояние системы в текущий момент. Какое из описаний оказывается проще, зависит главным образом от доступности априорной информации, а их применение дополнительно определяются требуемым видом синтезируемых моделей.

Как отмечалось ранее, зачастую каналы связи имеют стохастическую природу, причем случайными свойствами обладает оператор отображения полезной составляющей принимаемой смеси  $N'$ . Подобная ситуация возникает при условии распространения сигналов в среде со случайными неоднородностями [15]. Однако в настоящее время модели непрерывных каналов связи, базирующиеся на теории стохастических операторов, разработаны недостаточно полно [13, 26]. Именно поэтому наибольшее распространение получили структурно-детерминированные модели, характеризующиеся наличием лишь ряда стохастических параметров, а сама структура (вид) оператора преобразования в канале связи оказывается заданной. В результате системные характеристики каналов должны рассматриваться как некоторые случайные поля, в общем случае многомерные, в пространстве восьми параметров, а входные и выходные сигналы – в пространстве четырех параметров соответственно [25].

В этой связи модели каналов связи на основе дифференциальных или интегральных операторов заданной структуры имеют первостепенное значение. Подобные операторы содержат коэффициенты либо весовые функции, обладающие случайными свойствами. В результате описание преобразований осуществляется на основе понятий стохастического интеграла и дифференциала [28].

В строгом математическом смысле первостепенную роль имеют именно интегральные преобразования, определенные в той или иной форме, а дифференциальные являются лишь условными эквивалентами некоторых исходных интегральных, в которых интегралы (Ито,

Стратоновича) имеют неклассический обобщенный смысл [27]. Дополнительно при описании каналов на основе дифференциальных преобразование требуется выполнение условия марковости по тем переменным, по которым осуществляется дифференцирование, что практически возможно путем искусственного введения отношения порядка не только по временным, но и по пространственным координатам. В отличие от обыкновенных дифференциальных уравнений, отражающих физические явления на уровне динамики одной переменной, описание с помощью дифференциальных уравнений в частных производных позволяет рассматривать процессы, зависящие от нескольких переменных. Однако подобный переход оказывается принципиальным усложнением и требует постановок новых задач, а следовательно и разработки методов их обоснования и решений [20].

С другой стороны теоретическое обоснование и методы решения задач, в которых преобразования описываются одномерными интегральными уравнениями, достаточно просто обобщаются на многомерные случаи и уравнения, зависящие от нескольких переменных. Это является следствием того, что в отличие от дифференциальных уравнений, для которых необходимо задавать начальные или граничные условия, интегральные уравнения содержат всю необходимую информации для постановки задачи, что приводит к полной аналогии между одномерными, многомерными и многопеременными задачами.

Кроме того, стохастическое интегрирование, как впрочем и детерминированное, предоставляет возможность учета как равноправности изменений по всем используемым измерениям, так и задание на части из них отношений порядка в случае необходимости.

В результате даже при описании детерминированных каналов связи предпочтительнее оказывается использование моделей системных характеристик на основе интегральных преобразований, поскольку они позволяют рассмотреть наиболее общий вид каналов, а также избавиться от дополнительных искусственных ограничений, способных заметно снизить адекватность синтезируемых моделей.

Оба типа представления системной характеристики существенно упрощаются, если моделирование осуществляется на основе линейных операторов (рис.). Для дифференциальных уравнений вводится требование вещественности и независимости от сигналов коэффициентов дифференциального уравнения [1], что приводит его к линейному виду:

$$\begin{aligned} \mathbf{H}\{x(t, \mathbf{r})\} &= \left\{ z(t, \mathbf{r}) \left| \sum_{k_0, \dots, k_n} a_{k_0, \dots, k_n}(t, \mathbf{r}) \frac{\partial^k z(t, \mathbf{r})}{\partial t^{k_0} \prod_{i=1}^n \partial r_i^{k_i}} = \right. \right. \\ &= \left. \left. \sum_{k'_0, \dots, k'_n} a'_{k'_0, \dots, k'_n}(t, \mathbf{r}) \frac{\partial^{k'} x(t, \mathbf{r})}{\partial t^{k'_0} \prod_{i=1}^{n'} \partial r_i^{k'_i}} \right\}, \end{aligned} \quad (6)$$

где  $n = \overline{1, 3}$ ,  $n' = \overline{1, 3}$  – число рассматриваемых пространственных координат на входе и выходе соответственно;

$a_{k_0, \dots, k_n}(t, \mathbf{r})$ ,  $a'_{k'_0, \dots, k'_n}(t, \mathbf{r})$  – коэффициенты дифференциального уравнения, в общем случае являются функциями времени и пространственных координат;

$k = \sum_{i=0}^n k_i$ ,  $k' = \sum_{i=0}^{n'} k'_i$  – общее количество рассматриваемых слагаемых дифференциального уравнения на входе и выходе соответственно;

$\mathbf{r} = (r_1, \dots, r_n)$ ,  $\mathbf{r}' = (r'_1, \dots, r'_{n'})$  – вектора пространственных координат.

Ряд Вольтерра значительно сокращается, вырождаясь в одно интегральное преобразование (интеграл свертки, или Коши), весовой функцией которого является импульсная характеристика [19]:

$$\mathbf{H}\{x(t, \mathbf{r})\} = h_0(t', \mathbf{r}') + \int \int_{t \mathbf{r}} x(t, \mathbf{r}) h(t', \mathbf{r}', t, \mathbf{r}) dt d\mathbf{r}, \quad (7)$$

где  $h(t', \mathbf{r}', t, \mathbf{r}) = h_1(t', \mathbf{r}', t, \mathbf{r})$  – импульсная характеристика канала связи.

Таким образом, системными характеристиками линейного канала связи являются коэффициенты дифференциального уравнения (6) или импульсная характеристика интегрального преобразования (7). При этом, как отмечалось ранее, у импульсной характеристики есть ясная физическая интерпретация, чего лишены коэффициенты дифференциальных уравнений. Являясь реакцией канала связи на воздействие в виде дельта-импульса, импульсная характеристика оказывается унифицированным описанием любой линейной динамической системы [18]. Кроме того, существует достаточно простая физическая интерпретация интеграла свертки – линейная система проводит операцию взвешенного суммирования всех мгновенных значений входного сигнала.

По аналогии вводятся понятия других системных характеристик, по сути, являющиеся реакциями системы на заданные воздействия. Довольно часто в этом качестве используется переходная характеристика, предполагающая наличие реакции на воздействие в виде единичного скачка. В результате интеграл свертки (7) приобретает следующую форму [1]:

$$H\{x(t, \mathbf{r})\} = h_0(t', \mathbf{r}') + x(0, \mathbf{0}_n)g(t', \mathbf{r}', t', \mathbf{r}') + \iint_{t \mathbf{r}} \frac{\partial^{n+1} x(t, \mathbf{r})}{\partial t \prod_{i=1}^n \partial r_i} g(t', \mathbf{r}', t, \mathbf{r}) dt d\mathbf{r}, \quad (8)$$

где  $\mathbf{0}_n$  – нуль-вектор размерности  $n$ ;

$$g(t', \mathbf{r}', t, \mathbf{r}) = \frac{\partial^{n+1} h(t', \mathbf{r}', t, \mathbf{r})}{\partial t \prod_{i=1}^n \partial r_i} \text{ – переходная характеристика.}$$

**3. Модель линейного непрерывного канала связи как оператора преобразования метрических пространств общего вида.** Вышеописанные представления преобразований в канале связи достаточно хорошо укладываются в рамки теории сигналов [31], согласно которой пространственная, частотная и временная составляющие ресурса канала задают измерения пространств сигналов на входе и выходе, представленных в виде случайных процессов. Поскольку используемые для описания линейных систем операторы преобразований являются также линейными, то они подчиняются принципу суперпозиции и однородности [19]. Данное обстоятельство позволяет интерпретировать сигналы на входе и выходе в рамках некоторых метрических пространств с заданными базисами:

$$x(t, \mathbf{r}) = \iint_{t' \mathbf{r}'} \tilde{x}(t', \mathbf{r}') \varphi(t, \mathbf{r}, t', \mathbf{r}') dt' d\mathbf{r}', \quad (9)$$

$$z(t', \mathbf{r}') = \iint_{t \mathbf{r}} \tilde{z}(t, \mathbf{r}) \varphi'(t', \mathbf{r}', t, \mathbf{r}) dt d\mathbf{r}, \quad (10)$$

где  $\tilde{x}(t', \mathbf{r}')$ ,  $\tilde{z}(t, \mathbf{r})$  – входные и выходные функции разложения соответственно;

$\varphi(t, \mathbf{r}, t', \mathbf{r}')$ ,  $\varphi'(t', \mathbf{r}', t, \mathbf{r})$  – входные и выходные базисные функции соответственно.

При данном представлении возможно взаимнооднозначно сопоставить входным базисным функциям выходные в виде реакций канала связи на первые из них:

$$\varphi'(t', \mathbf{r}', t, \mathbf{r}) = \mathbf{H}'\{\varphi(t, \mathbf{r}, t', \mathbf{r}')\}. \quad (11)$$

Тогда выходной сигнал представим в виде разложения, в котором базисом выступает реакция (11), а функциями разложения – входные функции разложения сигнала, то есть  $\tilde{z}(t, \mathbf{r}) = \tilde{x}(t, \mathbf{r})$ :

$$z(t', \mathbf{r}') = \int \int_{t \mathbf{r}} \tilde{x}(t, \mathbf{r}) \mathbf{H}'\{\varphi(t, \mathbf{r}, t', \mathbf{r}')\} dt d\mathbf{r}. \quad (12)$$

Если же выбор базисов произволен, причем не обязательно входные и выходные базисные функции совпадают, что соответствует ситуации смешанных базисов, то оператор преобразования в канале связи имеет вид [2, 31]:

$$\tilde{z}(t, \mathbf{r}) = \int \int_{t' \mathbf{r}'} K(t, \mathbf{r}, t', \mathbf{r}') \tilde{x}(t', \mathbf{r}') dt' d\mathbf{r}', \quad (13)$$

где ядро интегрального преобразования, по сути являющееся системной характеристикой канала:

$$K(t, \mathbf{r}, t', \mathbf{r}') = \int \int_{t'' \mathbf{r}''} \varphi'^{-1}(t', \mathbf{r}', t'', \mathbf{r}'') \mathbf{H}'\{\varphi(t'', \mathbf{r}'', t', \mathbf{r}')\} dt'' d\mathbf{r}'', \quad (14)$$

определяется на основе сопряженных выходных базисных функций  $\varphi'^{-1}(t', \mathbf{r}', t'', \mathbf{r}'')$ , удовлетворяющих условию ортогональности:

$$\int \int_{t'' \mathbf{r}''} \varphi'(t, \mathbf{r}, t'', \mathbf{r}'') \varphi'^{-1}(t'', \mathbf{r}'', t', \mathbf{r}') dt'' d\mathbf{r}'' = \delta(t - t', \mathbf{r} - \mathbf{r}'). \quad (15)$$

Таким образом, представление как сигналов, так и преобразований в канале связи может быть различно, и задается в первую очередь базисными ядрами метрических пространств в рамках которых рассматриваются исследуемые каналы связи. Кроме того, сравнение выражений (12) и (13) позволяет сделать вывод о возможности представ-

ления сигнала на выходе линейного канала связи в виде свертки системной характеристики канала и входного сигнала вне зависимости от выбора формы базисов, а точнее – на основе функций разложения сигнала, что позволяет рассматривать выходной сигнал в виде разложения в базисе системной характеристики.

Так, для интегрального преобразования (7) в качестве входного используется базис многомерных дельта-функций Дирака (рис.):

$$\varphi(t, \mathbf{r}, t', \mathbf{r}') = \delta(t - t', \mathbf{r} - \mathbf{r}'). \quad (16)$$

В результате входной сигнал представим в виде интегрального преобразования (9), в котором функциями разложения является сам исходный входной сигнал ( $\tilde{x}(t', \mathbf{r}') = x(t', \mathbf{r}')$ ):

$$x(t, \mathbf{r}) = \int \int_{t' \mathbf{r}'} x(t', \mathbf{r}') \delta(t - t', \mathbf{r} - \mathbf{r}') dt' d\mathbf{r}'. \quad (17)$$

Выходной же базис оказывается реакцией (11), представляющей собой импульсную характеристику канала:

$$\varphi(t', \mathbf{r}', t, \mathbf{r}) = h(t', \mathbf{r}', t, \mathbf{r}),$$

что делает состоятельным изображение выходного сигнала в виде преобразования (7) без учета начального состояния канала связи  $h_0(t', \mathbf{r}')$ .

При использовании в качестве входного базиса многомерных функций единичного скачка (включения) Хэвисайда (рис.):

$$\varphi(t, \mathbf{r}, t', \mathbf{r}') = \sigma(t - t', \mathbf{r} - \mathbf{r}'), \quad (18)$$

входной сигнал представляется в виде преобразования (9):

$$x(t, \mathbf{r}) = x(0, \mathbf{0}_n) \sigma(t, \mathbf{r}) + \int \int_{t \mathbf{r}} \frac{\partial^{n+1} x(t', \mathbf{r}')}{\partial t' \prod_{i=1}^n \partial r'_i} \sigma(t - t', \mathbf{r} - \mathbf{r}') dt' d\mathbf{r}'.$$

Таким образом, функциями разложения оказываются смешанные производные  $\tilde{x}(t', \mathbf{r}') = \frac{\partial^{n+1} x(t', \mathbf{r}')}{\partial t' \prod_{i=1}^n \partial r'_i}$ , с учетом их вырожденности для постоянной составляющей (первого слагаемого). Выходной же базис представляет собой реакцию (11) и имеет вид переходной характеристики:

$$\varphi'(t', \mathbf{r}', t, \mathbf{r}) = g(t', \mathbf{r}', t, \mathbf{r}).$$

Следовательно, оператор преобразования полезной составляющей в канале связи  $H'$  совпадает с выражением (8), что соответствует представлению выходного сигнала в базисе переходных характеристик без учета начального состояния канала связи  $h_0(t', \mathbf{r}')$ .

По аналогии подобные представления можно получить и для дифференциальных операторов (6), с той лишь разницей, что выходной сигнал представляется в виде интегрального преобразования, базисом которого будут служить функции, построенные на основе коэффициентов линейного дифференциального уравнения. Однако поскольку данное преобразование, по сути, является решением некоторого дифференциального уравнения, то получить его в явном виде возможно только в ряде частных случаев [13]. Тем не менее, и в этой ситуации целесообразно рассматривать входные и выходные сигналы как элементы метрических пространств с соответствующими базисами, а следовательно характеризовать канал связи с позиции пространственно-частотно-временного ресурса.

В принципе сопоставление входного базисов с выходным возможно и в случае нелинейных операторов, однако подобное сопоставление уже называть взаимнооднозначным нельзя вследствие невыполнимости принципа суперпозиции. Кроме того, в ряде задач синтеза и анализа целесообразно рассматривать входные и выходные базисные функции не как реакции одних на другие. Например, частотное описание входных сигналов и временное выходных (гармонический и базис дельта-функций) является предпочтительным для анализа систем с изменяющимися во времени параметрами [31]. Таким образом, в общем случае необходимо рассматривать смешанные базисы для входных и выходных сигналов, причем закон их соответствия должен являться следствием решения поставленной задачи, а не исходными данными.

**4. Модель линейного непрерывного канала связи как оператора преобразования метрических пространств с заданными базами.** Следует также отметить, что помимо интегральных представлений, использующих в качестве базисных ядер дельта-функции (16) и функции единичного скачка (18), возможны также другие виды преобразований (рис.), например Фурье, Лапласа, Френеля, Меллина и др. Их применение определяется в первую очередь удобством описания преобразований пространственно-временных сигналов линейными системами.

Так, наиболее часто используемым базисом является базис гармонических функций, или базис Фурье [29]. Его применение является следствием достаточной простоты генерирования гармонических колебаний, а также тем, что подобные сигналы оказываются собственными функциями линейных стационарных операторов [1].

Главным их достоинством при этом является возможность описания преобразований в линейном стационарном канале связи не на основе операций интегрирования или дифференцирования, а путем нахождения выходных функций разложения в виде произведения входных функций разложения и системной характеристики, то есть в алгебраической форме. Так, в случае пространственно-временного канала одновременное рассмотрение бесконечных интервалов анализа как по временной, так и по пространственным координатам, позволяет на основе разложения в базисе Фурье представить преобразования в канале связи в виде произведения функций. Базисные функции Фурье имеют вид [23]:

$$\varphi(t, \mathbf{r}, \omega_t, \omega_{\mathbf{r}}) = \varphi'(t, \mathbf{r}, \omega_t, \omega_{\mathbf{r}}) = e^{i\{\omega_t t + \omega_{\mathbf{r}}^T \mathbf{r}\}}, \quad (19)$$

где  $\omega_t, \omega_{\mathbf{r}} = (\omega_{r_1}, \dots, \omega_{r_n})$  – временная частота и вектор пространственных частот соответственно;

$T$  – оператор транспонирования.

Сопряженные выходные базисные функции, удовлетворяющие условию ортогональности (15), являются обратными по отношению к гармоническим функциям (19):

$$\varphi'^{-1}(\omega'_t, \omega'_{\mathbf{r}}, t', \mathbf{r}') = e^{-i\{\omega'_t t' + \omega'_{\mathbf{r}}{}^T \mathbf{r}'\}}.$$

Реакция на входные базисные функции определяется на основе интеграла свертки (7) без учета постоянной составляющей с той лишь разницей, что свойства канала описываются стационарной импульсной характеристикой, зависящей только от разности моментов времени:

$$H[\varphi(t, \mathbf{r}, \omega_t, \omega_{\mathbf{r}})] = \iint_{t \mathbf{r}} e^{i\{\omega_t t + \omega_{\mathbf{r}}^T \mathbf{r}\}} h(t'-t, \mathbf{r}'-\mathbf{r}) dt d\mathbf{r}. \quad (20)$$

Замена переменных в (20) приводит к хорошо известному свойству собственных функций линейных стационарных систем, к которым относятся и гармонические колебания [1] – реакция канала связи на базисную функцию (19) остается неизменной с точностью до некоторого множителя, не зависящего от временных и пространственных координат:

$$H[\varphi(t, \mathbf{r}, \omega_t, \omega_{\mathbf{r}})] = K(\omega_t, \omega_{\mathbf{r}}) \varphi(t', \mathbf{r}', \omega_t, \omega_{\mathbf{r}}), \quad (21)$$

где частотный коэффициент передачи канала связи, по сути являющийся системной характеристикой канала, имеет вид:

$$K(\omega_t, \omega_{\mathbf{r}}) = \iint_{t \mathbf{r}} h(t, \mathbf{r}) e^{-i\{\omega_t t + \omega_{\mathbf{r}}^T \mathbf{r}\}} dt d\mathbf{r}.$$

Подстановка полученного выражения для реакции на входные базисные функции (21) в формулу для ядра интегрального преобразования (14) и использование свойства ортонормированности исходного базиса с сопряженным (15) позволяют представить ядро в вырожденном виде:

$$K(\omega_t, \omega_{\mathbf{r}}, \omega'_t, \omega'_{\mathbf{r}}) = K(\omega_t, \omega_{\mathbf{r}}) \delta(\omega_t - \omega'_t, \omega_{\mathbf{r}} - \omega'_{\mathbf{r}}).$$

В результате на основе (13) выходной сигнал имеет алгебраический вид:

$$\tilde{z}(\omega_t, \omega_{\mathbf{r}}) = K(\omega_t, \omega_{\mathbf{r}}) \tilde{x}(\omega_t, \omega_{\mathbf{r}}), \quad (22)$$

в котором входные и выходные функции разложения  $\tilde{x}(\omega_t, \omega_{\mathbf{r}})$  и  $\tilde{z}(\omega_t, \omega_{\mathbf{r}})$  в базисе Фурье (19) являются по сути спектральным представлением исходных сигналов, а частотный коэффициент передачи – преобразованием Фурье импульсной характеристики.

Дополнительным поводом к использованию именно преобразования Фурье является приближенное описание поведения комплексной амплитуды электромагнитной волны в области Фраунгофера на основе данного преобразования [11], что предполагает рассмотрение модели только плоских волн [21]. Однако данное обстоятельство накладывает серьезные ограничения на практическую реализацию данного преобразования, поскольку требует при малых длинах волн значительных расстояний между модуляторами электромагнитных волн и областью регистрации интерференционной картины [11].

К тому же использование преобразования Фурье не всегда оказывается возможным, поскольку необходимо соблюдение условия абсолютной интегрируемости пространственно-временных сигналов, что не всегда выполнимо. Например, функции единичного скачка либо неограниченные во времени гармонические колебания обладают бесконечной энергией. В этом случае преобразования выполняются путем замены неинтегрируемого сигнала на его более общий вид, зависящий от некоторых параметров и приводящий к выполнимости условия абсолютной интегрируемости [11]. В дальнейшем после выполнения преобразования производится предельный переход (введенные параметры устремляются к пределу) от функции разложения обобщенного сигнала к функции разложения первоначального сигнала. В результате данную операцию можно трактовать как некоторое интегральное преобразование, базисное ядро которого отлично от преобразования Фурье введенными параметрами.

Таковым является преобразование Лапласа, использующее базис экспоненциальных функций с гармоническим заполнением:

$$\varphi(t, \mathbf{r}, \sigma_t, \omega_t, \sigma_{\mathbf{r}}, \omega_{\mathbf{r}}) = e^{\sigma_t t + \sigma_{\mathbf{r}}^T \mathbf{r} + i\{\omega_t t + \omega_{\mathbf{r}}^T \mathbf{r}\}},$$

где  $\sigma_t, \sigma_{\mathbf{r}} = (\sigma_{r_1}, \dots, \sigma_{r_n})$  – коэффициенты затухания по временной и пространственным координатам соответственно.

Данный базис, по сути, оказывается обобщением базиса Фурье (19) при нулевых коэффициентах затухания и представляет функции разложения в пространстве не вещественных, а комплексных час-

тот [1]. Помимо того, что базисные функции Лапласа обладают всеми свойствами собственных функций линейных стационарных систем, они снижают требования к существованию разложений (9) и (10) по сравнению с базисом Фурье, который предусматривает необходимость соблюдения конечной нормы преобразуемых сигналов и системных характеристик [16]. Для экспоненциальных функций с гармоническим заполнением оказывается достаточным условие не более чем экспоненциальной степени роста преобразуемых сигналов и системных характеристик [1].

Использование других преобразований продиктовано также рядом преимуществ при тех или иных условиях. Так, преобразование Френеля достаточно хорошо описывает поведение комплексной амплитуды электромагнитного поля в зоне Френеля, а следовательно предполагает моделирование в виде сферической волны [11]. В результате расстояния между модуляторами электромагнитных волн и областью регистрации интерференционной картины существенно сокращаются по сравнению с реализацией в приближении Фраунгофера, что делает возможным довольно простую реализацию как самого преобразования Френеля, так и производного от него Фурье. Ядро преобразования Френеля имеет вид функции Френеля [14]:

$$\varphi(t, \mathbf{r}, t', \mathbf{r}') = e^{i\{t'(t-t)^2 + (\mathbf{r}-\mathbf{r}')^T(\mathbf{r}-\mathbf{r}')\}}. \quad (23)$$

Несмотря на инвариантность к координатному сдвигу широко используемого преобразования Фурье, оно оказывается чувствительным к изменению масштаба входного пространственно-временного сигнала [32]. Поэтому в случаях разработки систем, инвариантных к масштабированию, применяют преобразование Меллина с ядром в виде степенных функций:

$$\varphi(t, \mathbf{r}, t', \mathbf{r}') = t^{t'-1} \prod_{i=1}^n r_i^{r_i'-1}. \quad (24)$$

**5. Модель линейного непрерывного канала связи как оператора преобразования метрических пространств с базисом в виде координатных функций интегрального канонического представления В.С. Пугачева.** Существуют преобразования, базисные ядра которых имеют форму, определяемую на основе свойств исследуемых пространственно-временных сигналов. Как отмечалось ранее, пространственно-временные сигналы обладают стохастическими свойст-

вами вследствие как случайного характера формируемых сообщений, так и присутствующей неопределенности в канале связи. В результате простейшим с практической точки зрения является представление такого рода сигналов в виде линейной комбинации некоррелированных случайных величин, имеющих нулевые математические ожидания [24]. Таким образом, интегральное преобразование выражает произвольный пространственно-временной сигнал посредством наиболее простого случайного процесса – некоррелированного шума (в общем случае белого неоднородного как по пространству, так и по времени [8, 25]), со следующими характеристиками:

$$M\{\tilde{x}(t', \mathbf{r}')\} = 0, \quad M\{\tilde{x}(t, \mathbf{r})\tilde{x}(t', \mathbf{r}')\} = D_{\tilde{x}}(t, \mathbf{r})\delta(t - t', \mathbf{r} - \mathbf{r}'), \quad (25)$$

где  $M\{x\}$  – математическое ожидание случайной величины  $x$ ;

$D_{\tilde{x}}(t, \mathbf{r})$  – интенсивность нестационарного белого шума (функций разложения  $\tilde{x}(t, \mathbf{r})$ ).

Использование именно подобного представления сигналов продиктовано двумя основными его преимуществами: во-первых, как и модели в виде оператора преобразования метрических пространств с заданными базисами, данные представления позволяют свести выполнение линейных операций (например, интегрирование, дифференцирование) над случайными сигналами на входе канала связи к операциям над неслучайными базисными функциями, от есть к операциям детерминированного анализа, а во-вторых, уже в отличие моделей в виде оператора преобразования метрических пространств с заданными базисами, – заменить исследование прохождения сигналов произвольной стохастической природы через канал связи на анализ прохождения белого шума, являющегося интегральным представлением исходного входного сигнала. Таким образом, вместо рассмотрения преобразований произвольных случайных и детерминированных базисных функций, являющихся представлением сигналов в модели в виде оператора преобразования метрических пространств с заданными базисами, достаточно провести исследование преобразований белого шума и уже несколько иных детерминированных базисных функций, что, несомненно, является более простой задачей.

Полученное преобразование в литературе [10, 22, 24] именуют интегральным каноническим представлением, а базисное ядро – координатными функциями интегрального канонического представления. Оно имеет вид формул (9) и (10), с той лишь разницей, что представ-

ляемый пространственно-временной сигнал  $x(t, \mathbf{r})$  обладает свойством центрированности, т. е. к полученному выражению канонического представления следует прибавить еще функцию математического ожидания исходного нецентрированного процесса.

В итоге базисное ядро интегрального преобразования имеет форму частного корреляционной функции пространственно-временного сигнала с белым шумом и его интенсивности:

$$\varphi(t, \mathbf{r}, t', \mathbf{r}') = \frac{M\{x(t, \mathbf{r})\tilde{x}(t', \mathbf{r}')\}}{D_{\tilde{x}}(t', \mathbf{r}')}, \quad (26)$$

Очевидно, что при различных статистических свойствах пространственно-временных сигналов на входе и выходе канала связи, несмотря на однотипный вид базисных ядер преобразований их форма будет различаться, а следовательно удобство и простота выполнения операций обработки полученных коэффициентов разложения на выходе канала  $\tilde{z}(t, \mathbf{r})$  с лихвой компенсируется трудоемкостью выполнения операций сопоставления входных и выходных базисных ядер, а также выполнения преобразования в канале связи входной функции разложения для получения выходной на основе интегрального оператора (13).

**6. Заключение.** Существует еще довольно большой набор разнообразных базисных ядер, позволяющих осуществлять соответствующие преобразования, пригодные для обработки пространственно-временных сигналов в тех или иных случаях [6, 7]. Кроме того, возможно использование комбинированных преобразований [32], например Фурье-Меллина, Фурье-Френеля и т. п. Однако для них необходимо выполнение основного и в ряде случаев критического допущения – неограниченность интервала анализа по пространственным и временным координатам. При нарушении этого условия (ограничение интервала анализа) неизбежно возникает ошибка аппроксимации, влияние которой на результирующие показатели качества синтезируемых систем, например модулятора и демодулятора, в некоторых случаях оказываются трудно прогнозируемым, а зачастую недопустимым.

В целом же линейные сосредоточенные системы с постоянными параметрами, представляющие собой набор элементов, выполняющих определенную линейную операцию, хорошо описываются на основе дифференциальных уравнений (6) конечного порядка [10]. Однако в результате канал связи рассматривается не в виде феноменологической

модели, то есть некоторого оператора в общем случае стохастического, выполняющего сложные функциональные преобразования, а в виде структурно-физической модели, вскрывающей механизмы детальных преобразований пространственно-временного сигнала в канале и образования помех. Это обстоятельство является достаточно серьезным ограничением при постановке задач синтеза систем передачи информации, поскольку исходный непрерывный канал по своей сути является распределенной системой, системная характеристика которой лишь в некотором приближении может описываться конечным числом элементарных составляющих [13].

Для линейных сосредоточенных систем с постоянными параметрами в принципе достаточен аппарат преобразований Фурье, Лапласа и других, поскольку их ядра являются собственными функциями подобных систем. Однако при рассмотрении систем с переменными параметрами, а тем более распределенных систем на ограниченных промежутках возникают существенные трудности при описании преобразований в данных каналах связи, прежде всего связанные с невозможностью алгебраического описания зависимостей выходных сигналов от входных вида (22). В итоге избавиться от трудно выполнимых операций интегрирования не удастся, а следовательно цель использования преобразований в этих условиях оказывается недостигнутой.

Анализ линейных систем с переменными параметрами на основе интегральных преобразований существенно затрудняется также и вследствие несуществования обратного преобразования для выходного пространственно-временного сигнала или его достаточно сложной формы, требующей дополнительной аппроксимации [20].

Дополнительная трудность представлений входных и выходных сигналов интегральными преобразованиями заключается в континуальном характере исследуемых пространств. Следовательно, для практических целей использовать подобные представления оказывается не совсем разумным, поскольку для физических измерений сигналов или численных расчетов больше подходит использование конечномерных величин [3, 29, 31]. Кроме того, при синтезе систем связи, как упоминалось ранее, одним из этапов является формирование канала дискретного времени, предполагающее исследование конечномерных пространств и преобразований их в бесконечномерные и обратно, что является следствием не только необходимости практической реализации систем передачи, но и требований к моделированию с позиций теории вероятности на основе вероятностных мер [9].

Таким образом, в работе предложена феноменологическая модель непрерывного канала связи, составляющая основу для дальнейшего аналитического описания моделей непрерывных каналов в форме операторов преобразования некоторых пространств. В отличие от наиболее часто используемых (конкретизированных) моделей каналов предлагаемая модель имеет весьма абстрактный операторный вид, который, однако, позволяет представить канал связи на основе его системной характеристики, а также свойств внутренних шумов. Разработанные на основе предложенной феноменологической модели операторные модели обобщают как наиболее известные существующие (например, на основе гармонического базиса и др.) на случай заданных базисных функций, так и позволяют варьировать входной и выходной базисы с целью наиболее удобного с аналитической точки зрения описания преобразований в канале (например, преобразований случайных функций с заданными статистическими свойствами). В результате, предложенный в работе математический аппарат отличается достаточно общим аналитическим описанием каналов связи как некоторых «черных ящиков» на уровне операторной взаимосвязи входных и выходных сигналов, с тем условием, что пространства входных и выходных множеств данных операторов могут задаваться достаточно произвольно исходя из удобства решения поставленных задач анализа и синтеза.

### Литература

1. *Баскаков С.И.* Радиотехнические цепи и сигналы // учебник. М.: Высш. Школа, 1983. 536 с.
2. *Батенков А.А., Батенков К. А.* Дискретизация линейного канала связи с памятью и аддитивным белым гауссовским шумом численным методом // Математическое моделирование. 2009. Т. 21, № 1. С. 53–74.
3. *Батенков К.А.* Математические модели модулятора и демодулятора с заданным порядком нелинейности // Цифровая обработка сигналов. 2013. № 1. С. 14–21.
4. *Батенков К.А.* Математическое моделирование непрерывных многопараметрических каналов связи в операторной форме // Телекоммуникации. 2013. № 10. С. 2–4.
5. *Батенков К.А.* Модели системных характеристик линейных каналов связи на основе интегральных преобразований // Модели, системы, сети в экономике, технике, природе и обществе. 2012. № 3 (4). С. 120–125.
6. *Бейтмен Г., Эрдейи А.* Таблицы интегральных преобразований. Т. 1: Преобразования Фурье, Лапласа, Меллина // М.: Наука, 1968. 344 с.
7. *Бейтмен Г., Эрдейи А.* Таблицы интегральных преобразований. Т. 2: Преобразования Бесселя, интегралы от специальных функций. // М.: Наука, 1970. 328 с.
8. *Вентцель Е.С., Овчаров Л. А.* Теория случайных процессов и ее инженерные приложения. учеб. пособие для вузов. // 2-е изд. стер. М.: Высш. школа, 2000. 383 с.
9. *Возенкрафт Дж.М., Джекобс И.М.* Теоретические основы техники связи: пер. с англ // под ред. Р. Л. Добрушина. М.: Мир, 1969. 640 с.
10. *Драган Я.П.* Модели сигналов в линейных системах // АН УССР: Физ.–мех. ин-т. Киев: Наукова думка, 1972. 302 с.

11. *Залманзон Л.А.* Преобразование Фурье, Уолша, Хаара и их применение в управлении, связи и других областях // М.: Наука. Гл. ред. физ.-мат. лит. 1989. 496 с.
12. *Карташевский В.Г.* Обработка пространственно-временных сигналов в каналах с памятью // М.: Радио и связь, 2000. 272 с.
13. *Кловский Д.Д., Конторович В.Я., Широков С.М.* Модели непрерывных каналов связи на основе стохастических дифференциальных // ред. Д.Д. Кловский. М.: Радио и связь, 1984. 247 с.
14. *Кловский Д.Д., Сойфер В. А.* Обработка пространственно-временных сигналов (в каналах передачи информации) // М.: Связь, 1976. 207 с.
15. *Кловский Д.Д.* Передача дискретных сообщений по радиоканалам // 2-е изд., перераб. и доп. М.: Радио и связь, 1982. 304 с.
16. *Корн Г. Корн К.* Справочник по математике для научных работников и инженеров // М.: 1970. 720 с.
17. *Курант Р.* Курс дифференциального и интегрального исчисления // пер. с нем. и англ. В 2 т. Т. 1. 4-е изд. М.: Наука, 1967. 704 с.
18. *Левин Б.Р., Шварц В.* Вероятностные модели и методы в системах связи и управления // М.: Радио и связь, 1985. 312 с.
19. *Левин Б.Р.* Теоретические основы статистической радиотехники // М.: Радио и связь, 1989. 656 с.
20. Методы классической и современной теории автоматического управления: учебник. В 5 т. Т. 1: Математические модели, динамические характеристики и анализ систем автоматического управления // под ред. К. А. Пупкова, Н. Д. Егупова. 2-е изд. перераб. и доп. М.: Издательство МГТУ им. Н. Э. Баумана, 2004. 656 с.
21. Методы компьютерной оптики : учеб. для вузов // под ред. В. А. Сойфера. 2-е изд., испр. М.: ФИЗМАТЛИТ, 2003. 688 с.
22. *Пугачев В. С. и др.* Основы статистической теории автоматических систем // М.: Машиностроение, 1974. 400 с.
23. *Кремер И. Я., Кремер А. И., Петров В. М. и др.* Пространственно-временная обработка сигналов // под ред. И. Я. Кремера. М.: Радио и связь, 1984. 224 с.
24. *Пугачев В.С.* Теория случайных функций и ее применение к задачам автоматического управления // М.: Физматгиз, 1962. 883 с.
25. *Рытов С.М., Крайцов Ю. А., Татарский В. И.* Введение в статистическую радиофизику // Ч. 2. М.: Наука, 1978. 464 с.
26. *Скороход А.В.* Случайные линейные операторы // Киев: Наукова думка, 1978. 200 с.
27. Справочник по теории автоматического управления // под ред. А. А. Красовского. М.: Наука. Гл. ред. физ.-мат. лит. 1987. 712 с.
28. *Стратонович Р.Л.* Условные марковские процессы и их применение к теории оптимального управления // М.: Изд. МГУ, 1966. 276 с.
29. *Тихонов В.И.* Статистическая радиотехника // М.: Советское радио, 1966. 681 с.
30. *Тихонов В. И., Харисов В. Н.* Статистический анализ и синтез радиотехнических устройств и систем // учеб. пособие для вузов. М.: Радио и связь, 1991. 608 с.
31. *Френкс Л.* Теория сигналов // пер. с англ. М.: Советское радио, 1974. 344 с.
32. *Якушкин Ю.Г.* Теория и расчет оптико-электронных приборов: учебник для студентов приборостроительных специальностей вузов // 3-е изд., перераб. и доп. М.: Машиностроение, 1989. 360 с.
33. *Matz G., et al.* Analysis, Optimization, and Implementation of Low-Interference Wireless Multicarrier Systems // IEEE Transactions on Wireless Communications. vol. 6. no. 5, May 2007. pp. 1921–1931.
34. *Durisi G., Morgenshtern V.I., Bolcskei H.* On the Sensitivity of Continuous-Time Noncoherent Fading Channel Capacity // 2012. Url: <http://arxiv.org/pdf/1107.2527.pdf> (дата обращения: 27.10.2013).

35. Fundamentals of DSL technology // edited by Golden P., Dedieu H., Jacobsen K. NY: Auerbach Publications, 2006. 454 p.
36. *Irshad Ya.* On Some Continuous-time Modeling and Estimation Problems for Control and Communication // dissertation. Karlstad University Studies. 2013. 86 p.
37. *Kaiser T., Zheng F.* Ultra wideband systems with MIMO // Chichester: John Wiley & Sons Ltd, 2010. 254 p.
38. *Kumwilaisak W., Jay Kuo C.-C., Wu D.* Fading channel modeling via variable length Markov chain (VLMC) technique // IEEE Trans. on Vehicular Technology. vol.57. no. 3, May 2008. pp. 1338-1358.
39. *Kuhn V.* Wireless communications over MIMO channels // Chichester: John Wiley & Sons Ltd, 2006. 363 p.
40. *Hanzo L., Akhtman J., Wang L., Jiang M.* MIMO-OFDM for LTE, WiFi, and WiMAX: coherent versus non-coherent and cooperative turbo-transceivers // UK: John Wiley & Sons Ltd, 2011. 658 p.
41. *Yong S. C., et al.* MIMO-OFDM wireless communications with MATLAB // Singapore: John Wiley & Sons (Asia) Pte Ltd, 2010. 544 p.
42. *Zaki Ya.* Future Mobile Communications: LTE Optimization and Mobile Network Virtualization // urn:nbn:de:gbv:urn:nbn:de:gbv:46-00102749-12, University of Bremen Library, 2012.
43. *Zaki Ya, Dhabi A., Weerawardane Th., Li X., Gorg C.* LTE Radio Schedulers Analytical Modeling using Continuous Time Markov Chains // In the 6th Joint IFIP Wireless and Mobile Networking Conference (WMNC) 2013. Dubai, United Arab Emirates (UAE), April 2013.

## References

1. Baskakov S.I. *Radiotekhnicheskie cepi i signaly: uchebnik* [Radiotechnical circuits and signals : text-book]. Moscow: Vyssh. Shkola, 1983. 536 p. (In Russ.).
2. Batenkov A.A., Batenkov K.A. [Discretization of linear channel with memory and additive Gaussian noise by numerical method]. *Matematicheskoe modelirovanie – Mathematical modelling.* 2009. vol. 21, no. 1. pp. 53–74. (In Russ.).
3. Batenkov K.A. [Mathematical models of modulator and demodulator with prescribed order of nonlinearity] *Tsifrovaya obrabotka signalov – Digital signal processing.* 2013. no. 1. pp. 14–21. (In Russ.).
4. Batenkov K. A. [Mathematical modelling of continuous multiparametric channels in operator]. *Telekommunikatsii – Telecommunication.* 2013. no. 10. pp. 2–4. (In Russ.).
5. Batenkov K. A. [System characteristic models of linear channels on base of integral transformations]. *Modeli, sistemy, seti v ekonomike, tehnike, prirode i obschestve – Models, systems, nets in economics, technique, nature and science.* 2012. no. 3 (4). pp. 120–125. (In Russ.).
6. Bateman H., Erdelyi A. *Tablicy integral'nyh preobrazovaniy. T. 1: Preobrazovaniya Fur'e, Laplasy, Mellina* [Tables of integral transforms. Fourier, Laplace, Mellin transforms]. vol. 1. Moscow: Nauka, 1968. 344 p.
7. Bateman H., Erdelyi A. *Tablicy integral'nyh preobrazovaniy. T. 2: Preobrazovaniya Besselja, integraly ot special'nyh funkciy* [Tables of integral transforms. The Bessel transformation, integrals of special functions]. Vol. 2. Moscow: Nauka, 1970. 328 p.
8. Venttsel' E.S., Ovcharov L.A. *Teoriya sluchajnyh processov i ee inzhenernye prilozhenija: ucheb. posobie dlja vtuzov. 2-e izd. ster* [Theory of random processes and its engineering application : tutorial for higher technical school]. Moscow: Vyssh Shkola, 2000. 383 p. (In Russ.).
9. Wozencraft J. M., Jacobs I. M. [Principals of communication engineering]. John Wiley & Sons Inc. 1966. 720 p. (Russ. ed.: Dobrushina R. L. Teoreticheskie osnovy tehniki svyazi. Moscow: Mir, 1969. 640 p.)

10. Dragan Ya.P. *Modeli signalov v linejnyh sistemah* [Signal models in linear systems]. Kiev: Naukova dumka, 1972. 302 p. (In Russ.).
11. Zalmanzon L.A. *Preobrazovanie Fur'e, Uolsha, Haara i ih primenenie v upravlenii, svyazi i drugih oblastjah* [Fourier, Walsh, Haar transforms and its application in control, communication and other fields]. Moscow : Nauka. Gl. Red. fiz.-mat. lit., 1989. 496 p. (In Russ.).
12. Kartashevskii V.G. *Obrabotka prostranstvenno-vremennyh signalov v kanalakh s pamjat'ju* [Space-time signal processing in memory channels]. Moscow : Radio i svyaz', 2000. 272 p. (In Russ.).
13. Klovskaia D.D., Kontorovich V.Ya., Shirokov S.M. *Modeli neprerynykh kanalov svyazi na osnove stokhasticheskikh differentsial'nykh* [Continuous channel models on base of stochastic differential equations]. edited by D. D Klovskaia. Moscow: Radio i svyaz', 1984. 247 p. (In Russ.).
14. Klovskaia D.D., Soifer V.A. *Obrabotka prostranstvenno-vremennyh signalov (v kanalakh peredachi informacii)* [Space-time signal processing (in information circuits)]. Moscow: Svyaz', 1976. 207 p. (In Russ.).
15. Klovskaia D.D. *Peredacha diskretnykh soobshchenij po radiokanalam*. [Discrete message transfer by radio channels]. Moscow: Radio I svyaz', 1982. 304 p. (In Russ.).
16. Korn G., Korn K. *Spravochnik po matematike dlja nauchnykh rabotnikov i inzhenerov* [Mathematical handbook for scientists and engineers]. Moscow, 1970. 720 p. (In Russ.).
17. Courant R. [Course of differential and integral calculus]. Wiley-Interscience. 1954. 616 p. (Russ. ed.: Kurant R. Kurs differentsial'nogo i integral'nogo ischislenija. Vol.1 Moscow : Nauka, 1967. 704 p. (In Russ.)).
18. Levin B.R., Shvarts V. *Verojatnostnye modeli i metody v sistemah svyazi i upravlenija* [Stochastic models and methods in control and communication systems]. Moscow: Radio i svyaz', 1985. 312 p. (In Russ.).
19. Levin B.R. *Teoreticheskie osnovy statisticheskoi radiotekhniki* [Principals of statistical radio engineering]. Moscow: Radio i svyaz', 1989. 656 p. (In Russ.).
20. *Metody klassicheskoi i sovremennoj teorii avtomaticheskogo upravlenija: uchebnik. V 5 t. T. 1: Matematicheskie modeli, dinamicheskie harakteristiki i analiz sistem avtomaticheskogo upravlenija* [Methods of classic and modern automatic control theory: text-book. Mathematical models, dynamic characteristics and analysis of automatic control systems]. edited by Pupkov K. A., Egupov N. D. vol. 1. Moscow: Publishing house MGTU, 2004. 656 p. (In Russ.).
21. *Metody komp'yuternoj optiki : ucheb. dlja vuzov* [Methods of computer optics: text-book for higher schools]. Moscow: Phizmatlit, 2003. 688 p. (In Russ.).
22. Pugachev V.S. et al. *Osnovy statisticheskoi teorii avtomaticheskikh sistem* [Fundamentals of statistic automatic control theory]. Moscow: Mashinostroenie, 1974. 400 p. (In Russ.).
23. Kremer I.Ya. et al. *Prostranstvenno-vremennaja obrabotka signalov* [Space-time signal processing]. edited by. Kremer I.Ya. Moscow: Radio i svyaz', 1984. 224 p. (In Russ.).
24. Pugachev V.S. *Teoriya sluchajnykh funkcij i ee primenenie k zadacham avtomaticheskogo upravlenija* [Theory of random functions and its application to automatic control tasks]. Moscow: Phizmatgiz, 1962. 883 p. (In Russ.).
25. Rytov S.M., Kravtsov U.A., Tatarskii V.I. *Vvedenie v statisticheskuiu radiofiziku* [Preface in statistical radiophysics]. Moscow: Nauka, 1978. vol. 2. 464 p. (In Russ.).
26. Skorohod A.V. *Sluchajnye linejnye operatory* [Stochastic linear operators]. Kiev: Nakova Dumka, 1978. 200 p. (In Russ.).
27. *Spravochnik po teorii avtomaticheskogo upravlenija* [Automatic control theory]. edited by Krasovskii A. A. Moscow: Nauka. Gl. red. fiz.-mat. lit. 1987. 712 p. (In Russ.).
28. Stratonovich R.L. *Uslovnye markovskie processy i ih primenenie k teorii optimal'nogo upravlenija* [Conditioned Markovian processes and its application to optimal control theory]. Moscow: Publishing house MGU, 1966. 276 p. (In Russ.).

29. Tikhonov V.I. *Statisticheskaja radiotekhnika* [Statistical radio engineering]. Moscow: Sovetskoe radio, 1966. 681 p. (In Russ.).
30. Tikhonov V.I., Kharisov V.N. *Statisticheskij analiz i sintez radiotekhnicheskikh ustrojstv i sistem: ucheb. posobie dlja vuzov* [Statistical analysis and syntheses of radio engineering devices and systems: tutorial for higher schools]. Moscow: Radio i svyaz', 1991. 608 p. (In Russ.).
31. Franks L. Signal theory. Prentice-Hall, 1969. 317 p. (Russ. ed.: Vakman D.E. Teorija signalov. Moscow: Sovetskoe radio, 1974. 344 p. (In Russ.)).
32. Yakushenkov U. G. *Teorija i raschet optiko-jelektronnyh priborov: uchebnik dlja studentov priborostroitel'nyh special'nostej vuzov* [Theory and calculation of optical-electronic devices : text-book for instrument making specialities of higher schools]. Moscow : Mashinostroenie, 1989. 360 p. (In Russ.).
33. Matz G., et al. Analysis, Optimization, and Implementation of Low-Interference Wireless Multicarrier Systems. IEEE Transactions on Wireless Communications. vol. 6. no. 5, May 2007. pp. 1921–1931.
34. Durisi G., Morgenshtern V.I., Bolcskei H. On the Sensitivity of Continuous-Time Noncoherent Fading Channel Capacity. Available at: <http://arxiv.org/pdf/1107.2527.pdf>. (accessed: 27.10.2013).
35. Fundamentals of DSL technology. edited by Golden P., Dedieu H., Jacobsen K. NY: Auerbach Publications, 2006. 454 p.
36. Irshad Ya. On Some Continuous-time Modeling and Estimation Problems for Control and Communication: dissertation. Karlstad University Studies. 2013. 86 p.
37. Kaiser T., Zheng F. Ultra wideband systems with MIMO. Chichester: John Wiley & Sons Ltd, 2010. 254 p.
38. Kumwilaisak W. Jay Kuo C.-C., and Wu D. Fading channel modeling via variable length Markov chain (VLMC) technique. IEEE Trans. on Vehicular Technology. vol.57. no. 3, May 2008. pp. 1338-1358.
39. Kuhn V. Wireless communications over MIMO channels. Chichester: John Wiley & Sons Ltd, 2006. 363 p.
40. Hanzo L., Akhtman J., Wang L., Jiang M. MIMO-OFDM for LTE, WiFi, and WiMAX: coherent versus non-coherent and cooperative turbo-transceivers. UK: John Wiley & Sons Ltd, 2011. 658 p.
41. Yong S. C., et al. MIMO-OFDM wireless communications with MATLAB. Singapore: John Wiley & Sons (Asia) Pte Ltd, 2010. 439 p.
42. Zaki Ya. Future Mobile Communications: LTE Optimization and Mobile Network Virtualization. urn:nbn:de:gbv:urn:nbn:de:gbv:46-00102749-12. University of Bremen Library. 2012.
43. Zaki Ya, Dhabhi A., Weerawardane Th., Li X., Gorg C. LTE Radio Schedulers Analytical Modeling using Continuous Time Markov Chains. In the 6th Joint IFIP Wireless and Mobile Networking Conference (WMNC) 2013. Dubai, United Arab Emirates (UAE), April 2013.

**Батенков Кирилл Александрович** – к-т техн. наук, докторант Академии Федеральной службы охраны Российской Федерации. Область научных интересов: статистическая теория связи, модели и методы обработки сигналов. Число научных публикаций – 85. [pustur@yandex.ru](mailto:pustur@yandex.ru); Академия Федеральной службы охраны Российской Федерации, ул. Приборостроительная, д. 35, Орёл, 302034, РФ.

**Batenkov Kirill Aleksandrovich** – Ph.D., doctoral candidate of Academy of Federal Guard Service. Research interests: statistical communication theory, models and methods of signal processing. The number of scientific publications – 85. [pustur@yandex.ru](mailto:pustur@yandex.ru); Academy of Federal Guard Service, Priborostroitel'naya str., 35, Orel, 302034, Russian Federation.

## РЕФЕРАТ

### *Батенков К.А.* **Моделирование непрерывных каналов связи в форме операторов преобразования некоторых пространств.**

Рассматриваемая проблема моделирования непрерывных каналов связи имеет актуальный характер для задач, связанных с проектированием информационных систем, функционально предназначенных для передачи данных.

В работе непрерывный канал связи описывается оператором, отображающим все допустимое множество пространственно-временных сигналов на входе во множество подобных сигналов на выходе.

Предложена феноменологическая модель непрерывного канала, аналитически выражаемая в виде двух слагаемых.

Первое представляет собой системную характеристику в виде оператора отображения полезной составляющей принимаемой смеси, зависящего от свойств среды передачи и параметров входных и выходных устройств, а второе – поле аддитивных помех, задающее как внешние помехи, так и внутренние шумы.

Используемые для описания линейных систем операторы подчиняются принципу суперпозиции и однородности. Данное обстоятельство позволяет интерпретировать сигналы на входе и выходе в рамках некоторых метрических пространств с заданными базисами и представить модель линейного непрерывного канала связи в виде оператора преобразования метрических пространств общего вида.

Описание как сигналов, так и преобразований в канале связи может быть различно, и задается в первую очередь базисными ядрами метрических пространств в рамках которых рассматриваются исследуемые каналы связи. В общем случае необходимо рассматривать смешанные базисы для входных и выходных сигналов, причем закон их соответствия должен являться следствием решения поставленной задачи, а не исходными данными.

Предложена модель линейного непрерывного канала связи в форме оператора преобразования метрических пространств с заданными базисами. Показана возможность использования в качестве базисных ядер дельта-функции, функций единичного скачка, Фурье, Лапласа, Френеля, Меллина. Отмечена относительная произвольность их выбора, определяемая, в первую очередь, удобством описания преобразований пространственно-временных сигналов линейными системами.

Предложена модель линейного непрерывного канала связи в форме оператора преобразования метрических пространств с базисом в виде координатных функций интегрального канонического представления В. С. Пугачева. Главной особенностью подобной модели является задание форм базисных ядер на основе статистических свойств входных и выходных пространственно-временных сигналов, которые задаются в общем случае в виде белого нестационарного как по пространству, так и по времени шума.

## SUMMARY

### *Batenkov K.A.* **Continuous channel modeling in shape of some space transformation operators.**

Considered problem of continuous channel modeling is actual for tasks associated with information system design functional intended for data transmission.

Continuous channel is described in paper by operator mapping total admissible set of input space-time signals to similar output signal set.

Continuous channel phenomenological model analytically expressed in shape of two summand is suggested.

First summand is system characteristic in shape of received composition useful component mapping operator depending on medium nature and parameters of input and output devices. Second summand is additive noise field specifying external disturbances and inner noises.

Operators used for linear system describing are complied superposition and similarity principles. Its case allow to interpret input and output signal in some metric spaces with assigned basis and to express linear continuous channel model in shape of general form metric space mapping operator.

Statement of signals and transforms in channel may be various and specify by metric space basis kernels. In general view it is necessary considered mixed basis for input and output signals. Though matching low of it is assigned task result but not given data.

Linear continuous channel model in shape of some metric spaces with assigned basis is suggested. Availability of delta-function, unit step, Fourier, Laplace, Fresnel as basis kernels is displayed. It choice arbitrariness generated by description convenience of space-time signals in linear systems is signed.

Linear continuous channel model in shape of some metric spaces with coordinate function basis of Pugachev integral canonical expansion is suggested. It principal singularity is basis kernel definition on base of input and output space-time signal statistical property specified in general form by space and time heterogeneous white noise.

Д.К. ЛЕВОНЕВСКИЙ, Р.Р. ФАТКИЕВА  
**ИССЛЕДОВАНИЕ КОМБИНИРОВАННЫХ  
АТАК КЛАССА «ОТКАЗ В ОБСЛУЖИВАНИИ»**

---

*Левоневский Д.К., Фаткиева Р.Р. Исследование комбинированных атак класса «отказ в обслуживании».*

**Аннотация.** DDoS-атаки являются распространённым способом выведения сетевых информационных систем из строя, причём для увеличения эффективности злоумышленники часто используют комбинации из нескольких видов атак. В статье рассматриваются параметры сетевого трафика, позволяющие контролировать состояние системы и отслеживать вторжения. Для этих параметров определены пороговые значения и условия, позволяющие связать поведение параметров с типом атак, которым подвержена система.

**Ключевые слова:** информационная безопасность, обнаружение вторжений, отказ в обслуживании, DDoS, комбинированные DDoS-атаки.

*Levonevskiy D.K., Fatkueva R.R. Research of combined “Denial of Service” attacks.*

**Abstract.** DDoS attacks are a widespread method of making network information systems out of service. Furthermore, the malefactors combine multiple types of attacks in order to increase the intrusion efficiency. This paper considers the network traffic parameters enabling system state monitoring and invasion tracking. There are defined thresholds and conditions that allow linking the parameters' behavior to the type of attacks the system is exposed to.

**Keywords:** information security, intrusion detection, Distributed Denial of Service, DDoS, combined DDoS attacks.

---

**1. Введение.** Сетевые информационные системы могут и зачастую подвергаются компьютерным атакам со стороны злоумышленников. Большое распространение в сети Интернет получили атаки типа «отказ в обслуживании» (Distributed Denial of Service, DDoS), в ходе которых создаются такие условия работы атакуемой системы, при которых обычные, легальные пользователи системы не могут получить доступ к предоставляемым системным ресурсам (серверам), либо этот доступ затруднен [1]. При этом существует множество систем и способов обнаружения атак, позволяющих не только обнаруживать факт несанкционированной активности, но идентифицировать атаку [2-4]. Однако в настоящее время существует тенденция к реализации комбинации атак со стороны злоумышленников. В связи с этим представляет интерес обеспечение возможности обнаружения комбинированных компьютерных атак разных видов.

Определение комбинации атак возможно путем анализа получаемых из сети пакетов по определенным параметрам с вычислением значений этих параметров [5, 6], которые затем сравниваются с эталонными значениями, а факт наличия одиночной или комбинирован-

ной одновременной атаки и определение видов атак определяется по сочетанию установленных условий для параметров [7-13].

Для этого выбираются параметры трафика, инвариантные к изменению величины легального трафика, но чувствительные к появлению DDoS-атак. Все параметры вычисляются в единицу времени. Проведенные исследования [12-14] и дополнительные исследования показали, что в качестве таких параметров целесообразно выбрать следующие (таблица 1).

Таблица 1. Параметры, идентифицирующие DDoS-атаки

Наименование	Значение	Описание
Отношение входящего и исходящего трафика	$R_{IP} = V_{IN} / V_{OUT}$ , где $V_{IN}$ - объем входящего трафика, принятого по протоколу IP; - объем исходящего трафика, отправленного по протоколу IP	Повышение скорости входящего трафика без соразмерного повышения скорости исходящего трафика означает более высокую вероятность наличия атаки
Количество потоков критических приложений	$N_{CR}$	Приложения выполняются в несколько потоков, общее количество которых для операционной системы ограничено вычислительными ресурсами компьютера. При возникновении атаки количество потоков возрастает.
Разность между количеством исходящих и входящих АСК-флагов в ТСП-трафике	$D_{АСК} =  N_{OUT} - N_{IN} $ , где $N_{OUT}$ - количество исходящих АСК-флагов в ТСП-трафике; $N_{IN}$ - количество входящих АСК-флагов в ТСП-трафике	Параметр косвенно характеризует частоту отказа сервером из-за перегрузки ТСП-запросами. Характеризует возможность атак видов SYN-flood и HTTP-flood.
Соотношение объема UDP- и ТСП-трафика	$R_{UDP} = V_{UDP} / V_{TCP}$ , где $V_{UDP}$ - объем входящего UDP-трафика; $V_{TCP}$ - объем входящего ТСП-трафика	Параметр может характеризовать наличие атаки класса UDP-Flood, так как в целом UDP-трафик для HTTP-соединений является нехарактерным.
Соотношение количества входящих UDP- и ТСП-пакетов	$R_{NUD} = N_{UDP} / N_{TCP}$ , где $N_{UDP}$ - количество входящих UDP-пакетов; $N_{TCP}$ - количество входящих ТСП-пакетов	Параметр отражает уровень загрузки канала трафиком, полученным по протоколу UDP.
Доля ICMP-трафика во входящем IP-трафике	$R_{ICM} = V_{ICM} / V_{IN}$ , где $V_{ICM}$ - объем входящего трафика, полученного по протоколу ICMP.	Параметр отражает загруженность трафика служебными пакетами. Превышение этого показателя может рассматриваться как признак ICMP-Flood.

Наименование	Значение	Описание
Относительные доли флагов SYN и PSH во входящих пакетах	$R_{PSH} = N_{PSH} / N_{TCP}$ и $R_{PSH} = N_{PSH} / N_{TCP}$ , где $N_{SYN}$ - количество SYN-флагов во входящих пакетах, переданных по протоколу TCP; $N_{PSH}$ - количество PSH-флагов во входящих пакетах, переданных по протоколу TCP	Частота SYN-флагов определяет долю служебных пакетов этого типа в TCP-трафике. Частота PSH-флагов, напротив, характеризует полезную загрузку канала. При атаке класса SYN-flood субъект атаки намерен перегрузить очередь соединений с помощью служебных пакетов.
Изменение эффективности использования сети	$R_{SP} = R_{SYN} / R_{PSH}$	Значение параметра увеличивается как при увеличении $R_{SYN}$ , так и при уменьшении $R_{PSH}$ , что может свидетельствовать о SYN-flood.
Степень полезной загрузки канала данными прикладных программ	$R_{TCP} = N_{PSH} / (N_{TCP} - N_{PSH})$	Соотношение показывает загрузку канала данными приложений прикладного уровня. Изменение этой величины может быть следствием атаки.
Средняя длина принятого IP-пакета	$L_{AVG} = V_{IN} / N_{IP}$ , где $N_{IP}$ - количество входящих пакетов, передаваемых по протоколу IP.	Параметр позволяет выявить атаки в случаях: если $L_{AVG}$ приближается к минимальному значению, это свидетельствует о перегрузке сервера служебными пакетами, что характерно для SYN-flood и ICMP-flood; если $L_{AVG}$ приближается к максимальному значению, это свидетельствует о попытке перегрузить канал передачи данных сервера и может быть результатом UDP-flood, TCP-flood или HTTP-flood.
Средняя длина принятого TCP-пакета	$L_{TCP} = V_{TCP} / N_{TCP}$	Приближение параметра $L_{TCP}$ к минимальному значению свидетельствует о перегрузке сервера служебными пакетами, что характерно для SYN-flood.

**2. Определение пороговых значений.** Для идентификации атак при использовании рассмотренных параметров важным является определение пороговых (эталонных) значений. Такое определение может быть проведено несколькими методами, в частности, пороговые значения возможно определить в результате адаптации системы, т.е. путем первоначальной работы системы в обычном режиме эксплуатации, но

при заведомом отсутствии атак. Возможно использовать и другой метод, предусматривающий исходное логико-аналитическое определение значений параметров с последующей экспериментальной проверкой. Результаты определения пороговых значений параметров приведены в таблице 2. Повышение вероятности обнаружения атаки достигается также рассмотрением приведенных выше параметров в совокупности. Так как характер влияния каждого вида атаки на параметры известен и отличается от влияния других классов атак, представляется возможным не только определить факт атаки, но и определить ее вид.

Таблица 2. Пороговые значения параметров

Условие для параметра	Виды атак, при которых выполняется условие
$R_{IP} > 2$	Все
$R_{SP} > 10$	SYN
$R_{UDP} > 10$	UDP
$R_{NUP} > 5$	UDP
$R_{TCP} > 10$	TCP
$R_{TCP} < 0.05$	SYN
$D_{ACK} > 100$	SYN, TCP
$L_{AVG} < 65$	SYN, ICMP
$R_{ICM} > 0,5$	ICMP
$N_{CR} > 140$	HTTP
$L_{TCP} < 65$	SYN

Для обнаружения атак используются пороговые значения параметров путем сравнения текущего значения СС (скользящего среднего) параметра с пороговым значением. Результаты выполнения условий для пороговых значений параметров в ходе одиночных атак различного вида приведены в таблице 3 (знак + указывает на выполнение соответствующего условия).

Предложенный набор параметров позволяет определить наличие одиночных атак разных видов. Помимо этого, представленный набор параметров и их пороговых значений позволяет определить наличие одновременных комбинированных атак разных видов и виды применяемых атак. Результаты выполнения условий для пороговых значений параметров в ходе одновременных комбинированных атак различного вида приведены в таблице 4.

Таблица 3. Выполнение условий для пороговых значений параметров в ходе одиночных атак различного вида

Параметр	Виды атак, при которых выполняется условие				
	HTTP flood	TCP flood	ICMP flood	SYN flood	UDP flood
$R_{IP} > 2$	+	+	+	+	+
$R_{SP} > 10$				+	
$R_{UDP} > 10$					+
$R_{NUD} > 5$					+
$R_{TCP} > 10$		+			
$R_{TCP} < 0.05$				+	
$D_{ACK} > 100$				+	
$L_{AVG} < 65$			+	+	
$R_{ICM} > 0,5$			+		
$N_{CR} > 140$	+				
$L_{TCP} < 65$				+	

При этом возможно непосредственно определить наличие и вид одиночной компьютерной атаки по сочетанию рассчитанных значений параметров на основе следующих условий:

если значения параметров  $R_{IP}$  и  $N_{CR}$  превысили пороговое значение, то определяется атака типа HTTP-flood (условие 1);

если значения параметров  $R_{IP}$ ,  $D_{ACK}$  и  $R_{SP}$  превысили пороговое значение, а  $R_{TCP}$  и  $L_{AVG}$  меньше порогового значения, то определяется атака типа SYN-flood (условие 2);

если значения параметров  $R_{IP}$ ,  $R_{TCP}$  и  $D_{ACK}$  превысили пороговое значение, то определяется атака типа TCP-flood (условие 3);

если значения параметров  $R_{IP}$ ,  $R_{UDP}$  и  $R_{NUD}$  превысили пороговое значение, то определяется атака типа UDP-flood (условие 4);

если значения параметров  $R_{IP}$  и  $R_{ICM}$  превысили пороговое значение, а  $L_{AVG}$  меньше порогового значения, то определяется атака типа ICMP-flood (условие 5).

Таблица 4. Выполнение условий для пороговых значений параметров в ходе одновременных комбинированных атак различного вида

Параметр	Виды атак, при которых выполняется условие				
	HTTP+SYN flood	HTTP+TCP flood	HTTP+ICMP flood	SYN+ICMP flood	TCP+ICMP flood
$R_{IP} > 2$	+	+	+	+	+
$R_{SP} > 10$	+	+	+		
$R_{UDP} > 10$	+	+		+	+
$R_{NUD} > 5$			+	+	+
$R_{TCP} > 10$	+			+	
$R_{TCP} < 0.05$		+			+
$D_{ACK} > 100$	+			+	
$L_{AVG} < 65$			+	+	+
$R_{ICM} > 0,5$	+			+	
$N_{CR} > 140$					
$L_{TCP} < 65$					

Наличие комбинированной компьютерной атаки и виды одновременно применяемых атак можно определить по сочетанию рассчитанных значений параметров на основе следующих условий:

если одновременно выполняется условие 1 и условие 2, то определяется комбинированная атака HTTP-flood и SYN-flood;

если одновременно выполняется условие 1 и условие 3, то определяется комбинированная атака HTTP-flood и TCP-flood;

если одновременно выполняется условие 1 и условие 5, то определяется комбинированная атака HTTP-flood и ICMP-flood;

если одновременно выполняется условие 2 и условие 5, то определяется комбинированная атака SYN-flood и ICMP-flood;

если одновременно выполняется условие 3 и условие 5, то определяется комбинированная атака TCP-flood и ICMP-flood.

Таким образом, ряд комбинаций атак может быть обнаружен при совместном выполнении условий отдельных атак.

**3. Заключение.** В результате исследования установлено взаимно однозначное соответствие между комбинацией выполняющихся условий для параметров системы и структурой сетевой атаки, направленной на целевую систему. На основе этих отношений, заданных таб-

лицами 3 и 4, построен ряд правил анализа сетевого трафика, позволяющий в автоматическом режиме контролировать состояние системы, с высокой вероятностью обнаруживать факт и состав атаки и извещать о нём администратора. Так как структура атаки известна, становится возможным принять меры блокировки вредоносного трафика.

### Литература

1. *Котенко И. В., Юсупов Р. М.* Текущее состояние и тенденции развития в области построения безопасных компьютерных систем // Часть 5-й Российской мультиконференции по проблемам управления (МКПУ-2012) - конференция "Информационные технологии в управлении" (ИТУ-2012). 09–11 октября 2012 г. Материалы конференции. СПб, 2012. С. 671-675.
2. *Бабенко Г. В., Белов С. В.* Анализ трафика TCP/IP на основе методики допустимого порога и отклонения как инструмент определения инцидентов информационной безопасности // Технологии технической безопасности. 2011. №5. 9 с.
3. *Schwab T., Wilson B., Thomas R.* Methodologies and metrics for the testing and analysis of distributed denial of service attacks and defenses // USENIX Association Berkeley, CA, USA, 2007.
4. *Петров В. В., Богатырев Е. А.* Статистический анализ сетевого трафика // Радиоэлектроника, электротехника и энергетика: Тез. докл. 10-й Междунар. научно-техн. конференции студентов и аспирантов. Том 1. М: Издательство МЭИ, 2-3 марта 2004.
5. *Lu W., Traore I.* An unsupervised approach for detecting DDoS attacks based on traffic-based metrics. Department of Electrical and Computer Engineering, University of Victoria.
6. *Staterlis C., Maglaris B.* Detecting incoming and outgoing DDoS attacks at the edge using a single set of network characteristics. // National Technical University of Athens, Greece, 2005.
7. *Левоневский Д. К., Пичугин Ю. А., Фаткиева Р. Р.* Исследование компьютерных атак методом сингулярного спектрального разложения сетевого трафика, Тр. СПИИРАН, 2013. № 26. С. 101–114.
8. *Левоневский Д. К., Пичугин Ю. А., Фаткиева Р. Р.* Оценка спектральных характеристик трафика в задаче обнаружения компьютерных атак различного типа, Тр. СПИИРАН, 2013. № 30. С. 56-64.
9. *Левоневский Д. К., Фаткиева Р. Р.* Детектирование атак методом сингулярного спектрального разложения, Труды СПИИРАН. 2013. № 2 (25). С. 135-147.
10. *Фаткиева Р. Р.* Многомасштабный анализ сетевого трафика. Информационная безопасность регионов России (ИБРР-2013). VIII Санкт-Петербургская межрегиональная конференция. Санкт-Петербург, 23-25 октября 2013 г.: Материалы конференции / СПОИСУ. СПб., 2013. 71 с.
11. *Пичугин Ю. А., Фаткиева Р. Р.* Косвенные признаки компьютерной атаки на основе спектрального анализа сетевого трафика в разработке средств информационной защиты // Информационная безопасность регионов России (ИБРР-2013). VIII Санкт-Петербургская межрегиональная конференция. Санкт-Петербург, 23-25 октября 2013 г.: Материалы конференции / СПОИСУ. СПб., 2013. С. 122-123.
12. *Фаткиева Р. Р.* Разработка метрик обнаружения атак на основе анализа сетевого трафика // Вестник Бурятского государственного университета. 2013. № 9. С. 81-86.
13. *Levonevskiy D. K., Fatkueva R. R.* Statistical research of traffic-based metrics for the purpose of DDos attack detection // European Science and Technology: materials of the

IV international research and practice conference, April 10th – 11th, 2013 / publishing office Vela Verlag Waldkraiburg, Munich, Germany. 2013. vol. 1. pp. 259-268.

14. Levonevskiy D. K., Fatkueva R. R. DDoS attack detection method based on the statistical research of the traffic metrics // Материали за 9-а международна научна практическа конференция «Новината за напредна наука». 2013. Том 56. С. 50-55.

## References

1. Kotenko I.V., Yusupov R.M. [The present situation and trends in the area of building of secure computer systems]. *Chast 5 Rossiyskoy multikonferentsii po problemam upravleniya (MKPU-2012) – konferentsiya “Informatsionnye tekhnologii v upravlenii” (ITU-2012)* [The 5<sup>th</sup> part of the Russian multiconference on management problems (MKPU-2012) – conference “Computer technologies in management” (ITU-2012)]. Saint-Petersburg, 2012. pp. 671-675.
2. Babenko G.V., Belov S.V. [TCP/IP traffic analysis on the basis of threshold and deviation methods as an instrument of information security incident detection]. *Tekhnologii tekhnosfernoy bezopasnosti – Technosphere security technologies*, 2011. no. 5. 9 p.
3. Schwab T., Wilson B., Thomas R. Methodologies and metrics for the testing and analysis of distributed denial of service attacks and defenses. USENIX Association Berkeley, CA, USA, 2007.
4. Petrov V.V., Bogatiryov E.A. [Statistical analysis of the network traffic]. *Radioelektronika, elektrotehnika i energetika: Tez. dokl. 10 mezhdunar. nauchno-techn. konferentsii studentov i aspirantov. Tom I.* [Radioelectronics, electrical and power engineering: the 10<sup>th</sup> International scientific and technical conference of students and post-graduates]. Moscow, 2004. vol. 1.
5. Lu W., Traore I. An unsupervised approach for detecting DDoS attacks based on traffic-based metrics. Department of Electrical and Computer Engineering, University of Victoria.
6. Siaterlis C., Maglaris B. Detecting incoming and outgoing DDoS attacks at the edge using a single set of network characteristics. National Technical University of Athens, Greece, 2005.
7. Levonevskiy D.K., Pichugin Y.A., Fatkueva R.R. [Research of computer attacks by means of the singular spectrum analysis of the network traffic]. Saint-Petersburg, *Trudy SPIIRAN – SPIIRAS proceedings*, 2013. no. 26. pp. 101–114.
8. Levonevskiy D.K., Pichugin Y.A., Fatkueva R.R. [Evaluation of traffic spectrum characteristics in the problem of various computer attacks detection]. Saint-Petersburg, *Trudy SPIIRAN – SPIIRAS proceedings*, 2013. no. 30. pp. 56-64.
9. Levonevskiy D.K., Fatkueva R.R. [Attack detection by means of singular spectrum analysis], Saint-Petersburg, *Trudy SPIIRAN – SPIIRAS proceedings*, 2013. no. 2(25). pp. 135-147.
10. Fatkueva R.R. [Multiresolutional network traffic analysis]. *Informatsionnaya bezopasnost' regionov Rossii (IBRR-2013). VIII Sankt-Peterburgskaya mezhtseleynaya konferentsiya* [Information security of regions of Russia (IBRR-2013). The 8<sup>th</sup> Saint-Petersburg interregional conference]. Saint-Petersburg, 2013. p. 71.
11. Pichugin Y.A., Fatkueva R.R. [Indirect computer attack signs on the basis of network traffic spectrum analysis in the information security means development]. *Informatsionnaya bezopasnost' regionov Rossii (IBRR-2013). VIII Sankt-Peterburgskaya mezhtseleynaya konferentsiya* [Information security of regions of Russia (IBRR-2013). The 8<sup>th</sup> Saint-Petersburg interregional conference]. Saint-Petersburg, 2013. pp. 122-123.

12. Fatkueva R.R. [Development of attack detection metrics on the basis of the network traffic analysis]. *Vestnik Buryatskogo gosudarstvennogo universiteta – Bulletin of the Buryat state university*, 2013. no. 9. pp. 81-86.
13. Levonevskiy D. K., Fatkueva R. R. Statistical research of traffic-based metrics for the purpose of DDoS attack detection. European Science and Technology: materials of the IV international research and practice conference, April 10th – 11th, 2013. Publishing office Vela Verlag Waldkraiburg, Munich, Germany, 2013. vol. 1. pp. 259-268
14. Levonevskiy D. K., Fatkueva R. R. DDoS attack detection method based on the statistical research of the traffic metrics. *Materiali za 9 mezhdunarona nauchna practichna konferentsia “Novinata za naprednali nauka”* [Proceedings of the 9th international scientific practical conference “Novinata naprednali for science”]. Sofia, 2013. vol. 56. pp. 50-55.

**Левоневский Дмитрий Константинович** — младший научный сотрудник лаборатории информационно-вычислительных систем Федерального государственного бюджетного учреждения науки Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН). Область научных интересов: исследование DDoS-атак, статистический анализ и моделирование трафика локальных сетей. Число научных публикаций – 5. DLewonewski.8781@gmail.com; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-4369, факс +7(812)328-4450.

**Levonevskiy Dmitriy Konstantinovich** — junior researcher, Laboratory of Computer and Information Systems, St.Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Scientific interests: research of DDoS attacks, statistical analysis and modeling of the network traffic. Number of publications — 5. DLewonewski.8781@gmail.com; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-4369, fax +7(812)328-4450.

**Фаткьева Роза Равильевна** — к-т техн. наук, старший научный сотрудник лаборатории информационно-вычислительных систем Федерального государственного бюджетного учреждения науки Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН). Область научных интересов: моделирование информационных систем. Число научных публикаций — 37. rikki2@yandex.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-4369, факс +7(812)328-4450.

**Fatkueva Rosa Ravilievna**— Ph.D., senior researcher, Laboratory of Computer and Information Systems, St.Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: modeling of information systems. Number of publications — 37. rikki2@yandex.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-4369, fax +7(812)328-4450.

**Поддержка исследований.** В публикации представлены результаты исследований, поддержанные грантом «ИнфоТеКСАкадемия2012», рук. Р.Р. Фаткьева.

**Acknowledgements.** The research is supported by grant «InfoTeKSAcademia2012» headed by R.R. Fatkueva.

## РЕФЕРАТ

### *Левоневский Д.К., Фаткиева Р.Р.* **Исследование комбинированных атак класса «отказ в обслуживании».**

Сетевые информационные системы зачастую подвергаются компьютерным атакам со стороны злоумышленников. Одним из самых распространённых способов вывода серверов из строя является DDoS, причём для повышения эффективности злоумышленники часто используют комбинации из нескольких видов атак, и возникает необходимость обнаружения комбинированных компьютерных атак разных видов.

Определение структуры сетевой атаки возможно путем анализа получаемых из сети пакетов и вычисления определенных параметров, значения которых сравниваются с пороговыми. В качестве параметров используются изменяющиеся во времени функции сетевого трафика, характеризующие различные аспекты загрузки уязвимой системы и эффективности сетевой передачи данных. Кроме того, параметры выбираются с учётом того, что они должны быть инвариантны к изменению величины легального трафика, но чувствительны к появлению DDoS-атак. Все параметры рассчитываются в единицу времени и выстраиваются во временные ряды, к которым для уменьшения погрешности применяется сглаживание методом скользящего среднего.

Для всех приведённых параметров определены пороговые значения и условия, позволяющие связать поведение параметров с типом атак, которым подвержена система. Пороги рассчитываются различными методами: как в результате адаптации системы, так и с помощью логико-аналитического определения допустимых значений параметров с последующей экспериментальной проверкой. Условия представляют собой критерии, состоящие в сравнении очередного значения параметра, полученного из сглаженного временного ряда, с соответствующим порогом.

В результате исследования устанавливается взаимно однозначное соответствие между комбинацией выполняющихся условий для параметров системы и структурой сетевой атаки, направленной на целевую систему.

## SUMMARY

### *Levonevskiy D.K., Fatkueva R.R.* **Research of combined “Denial of Service” attacks.**

Network information systems are often a target of distributed computer attacks. One of the most common ways of making servers and services unavailable is DDoS. Furthermore, malefactors often use a combination of several attack types to improve the efficiency, so there is a necessity of combined computer attack detection.

Determining the network attack structure is possible through incoming packets analysis and calculation of certain parameters. The values of these parameters are compared with the according thresholds. The parameters are functions of network traffic data describing various aspects of the target system load and performance. In addition, the parameters must be derived in view of the need that they must be invariant to legal traffic changes and at the same time sensitive to the DDoS attack appearance. All parameters are calculated per unit of time and form time series that are smoothed by the moving average method in order to increase the detection precision.

For all the parameters there are threshold values and conditions that allow linking the parameters' behavior to the attack types the system is exposed to. The thresholds are calculated using different methods: as a result of system adaptation as well as with help of logical and analytical determination of allowable values followed by experimental verification. Conditions are the rules of comparison of the next item from the time series to a corresponding threshold.

As a result of this research there is a correspondence between the combination of the fulfilled conditions of the system parameters and the structure of the network attack the target system is facing.

Р.Б. ТРЕГУБОВ, Ю.Г. АЛЕКСИКОВ  
**МОДЕЛЬ ПРОЦЕССА МУЛЬТИПЛЕКСИРОВАНИЯ  
ПРОТОКОЛЬНЫХ БЛОКОВ ДАННЫХ  
В КАНАЛЕ СВЯЗИ УПРАВЛЕНИЯ**

---

*Трегубов Р.Б., Алексиков Ю.Г. Модель процесса мультиплексирования протокольных блоков данных в канале связи управления.*

**Аннотация.** Для повышения обоснованности принимаемых решений в ходе синтеза сетей обмена данными управления, обеспечивающих минимизацию затрат ресурсов пропускной способности линий передачи гетерогенных телекоммуникационных систем, необходимо учитывать особенности трафика, формируемого источниками управляющей информации. Предлагаемая в статье модель процесса мультиплексирования протокольных блоков данных в канале связи управления учитывает вариативность интенсивности поступления потока служебных сообщений от источников управляющей информации. Использование модифицированной формулы Энгсета позволяет обеспечить более рациональное распределение канального ресурса, необходимого для организации доставки протокольных блоков данных.

**Ключевые слова:** модель, управление, телекоммуникационная система, сеть обмена данными управления, каналы связи управления, управляющая информация, протокольный блок данных.

*Tregubov R.B., Aleksikov Y.G. Model of process of multiplexing the protocol data units in management channel.*

**Abstract.** For improving the validity of the accepted decision during the synthesis of management-data-exchange network, providing the minimization of resource costs of traffic capacity of transmission links of heterogeneous telecommunication systems, it should be taken into account feature of traffic from sources of controlling information. The model of process of multiplexing the protocol data units in management channel is proposed in the article take into account the variability of intensity of the flow of the official messages from sources of controlling information. Use of the modified formula of Engset allows us to provide more rational distribution of the channel resource necessary for the organization of transmission of protocol data units.

**Keywords:** model, management, telecommunication system, management-data-exchange network, management channels, control information, protocol data unit.

---

**1. Введение.** Адекватное ситуации управление гетерогенными телекоммуникационными сетями (ГТКС) является залогом их эффективного функционирования. С одной стороны, это требует повышения объема циркулирующих управляющих сообщений, с другой, – обуславливает актуальность вопросов минимизации непроизводительных затрат ресурсов пропускной способности линий передачи ГТКС. Качество управления ГТКС во многом определяется полнотой, достоверностью и своевременностью доставки протокольных блоков данных (ПБД) по каналам связи управления (КСУ), которые вместе с узлами коммутации управляющих сообщений (УКУС) образуют сеть

обмена данными управления (СОДУ) [1, 2]. В современных СОДУ реализуется режим коммутации пакетов, при этом ПБД имеют различное назначение, приоритет и длину. Отметим, что увеличение числа услуг связи и количества пользователей ГТКС приводит к закономерному возрастанию затрат на аренду КСУ.

Целью моделирования, представленного в настоящей статье, является формальное описание процесса мультиплексирования потоков ПБД в КСУ для повышения обоснованности принимаемых решений в ходе синтеза СОДУ, обеспечивающего минимизацию затрат ресурсов пропускной способности линий передачи ГТКС.

**1. Модель источника управляющей информации (ИУИ).** С точки зрения проводимого исследования, источники управляющей информации (ИУИ) целесообразно разделить на две группы: с постоянной и с изменяющейся скоростью передачи. В свою очередь ИУИ с изменяющейся скоростью передачи также можно разделить на две группы: источники непрерывного типа и источники стартстопного типа. Поскольку моделью ИУИ с изменяющейся скоростью передачи непрерывного типа можно описать все другие источники, далее рассматривается имеющееся для них решение.

Различают мгновенную  $b(t)$ , максимальную (пиковую)  $\max[b(t)]$ , минимальную  $\min[b(t)]$ , среднюю  $M[b(t)]$  и эффективную  $b_{\text{эф}}(t)$  скорости передачи ИУИ.

Известно, что значения пиковой и средней скоростей передачи ИУИ дают два предельных решения задачи выбора канального ресурса. Использование пиковой скорости  $\max[b(t)]$  обеспечивает нулевую вероятность переполнения буферов УКУС, однако при этом КСУ недогружены. Использование средней скорости  $M[b(t)]$ , напротив, ведет практически к полной загрузке КСУ, однако при этом потери и сетевая задержка ПБД в УКУС достаточно велики. Интуитивно понятно, что потребность в канальном ресурсе должна рассчитываться с использованием значения, лежащего между средней и пиковой скоростями передачи ПБД. В теории телетрафика мультисервисных сетей соответствующая характеристика называется эффективной скоростью передачи.

В работе [3] предлагается для оценки эффективной скорости передачи использовать следующие характеристики: среднюю  $M[b(t)]$  и пиковую  $\max[b(t)]$  скорости передачи, скорость канала связи  $C$  и ограничение на качество передачи в форме доли потерянных блоков данных  $P^{\text{пот. перепр}}$

вследствие перегрузок. Величины  $\psi$ ,  $\xi$  определяются из выражений:

$$\psi = -\frac{2 \cdot \max [b(\tau)]}{C} \cdot \log_{10} (P^{\text{пот. перегр}}),$$

$$\xi = 1 - \frac{1}{50} \cdot \log_{10} (P^{\text{пот. перегр}}).$$
(1)

В свою очередь эффективная скорость передачи  $b_{\text{эф}}(t)$  находится из следующего эмпирического соотношения:

$$b_{\text{эф}}(t) = \begin{cases} \xi \cdot M[b(t)] \cdot \left(1 + 3 \cdot \psi \cdot \left(1 - \frac{M[b(t)]}{\max [b(t)]}\right)\right), & \text{если } 3 \cdot \psi \leq \min \left\{3, \frac{\max [b(t)]}{M[b(t)]}\right\} \\ \xi \cdot M[b(t)] \cdot \left(1 + 3 \cdot \psi^2 \cdot \left(1 - \frac{M[b(t)]}{\max [b(t)]}\right)\right), & \text{если } 3 < 3 \cdot \psi^2 \leq \frac{\max [b(t)]}{M[b(t)]} \\ \max [b(t)] & \text{в остальных случаях.} \end{cases} \quad (2)$$

Отметим, что значение эффективной суммарной скорости передачи независимых потоков ПБД мультиплексируемых в КСУ совпадает с суммой значений эффективных скоростей передачи индивидуальных потоков ПБД. Следовательно, можно утверждать, что если в системе открыто  $n$  однотипных сеансов связи между ИУИ эффективная скорость суммарного потока ПБД может быть найдена как  $n \cdot b_{\text{эф}}(t)$ .

**2. Модель процесса мультиплексирования ПБД в КСУ.** Для того, что бы рассчитать параметры суммарного трафика от нескольких ИУИ в КСУ, рассмотрим механизм статистического мультиплексирования. На рисунке 1 показаны потоки ПБД от отдельных ИУИ и суммарный поток ПБД полученный вследствие их мультиплексирования в КСУ.

Рассмотрим вариант расчета эффективной скорости передачи суммарного потока ПБД в КСУ для соединений "многоточка-точка" (рис. 2). Пусть к УКУС подключено  $R$  однотипных ИУИ, при этом интенсивность потенциальной нагрузки от каждого источника равна  $a$  Эрл. Предположим, что любое соединения допускается в СОДУ без каких-либо ограничений.

Вероятность того, что в произвольный момент времени будет открыто ровно  $n$  сеансов служебной связи можно найти, воспользовавшись биномиальным распределением  $P_n = C_R^n \cdot a^n \cdot (1-a)^{R-n}$ , где  $C_R^n$  –

это число сочетаний, определяемое как:  $C_R^n = \frac{R!}{n!(R-n)!}$ . Среднее число сеансов служебной связи "многоточка-точка" можно рассчитать как:

$$M[n] = \sum_{n=0}^R [C_R^n \cdot a^n \cdot (1-a)^{R-n} \cdot n] = R \cdot a, \quad (3)$$

а среднее значение эффективной скорости суммарного потока ПБД в КСУ соответственно:

$$M[b_{\Sigma}(t)] = b_{\Phi}(t) \cdot R \cdot a. \quad (4)$$

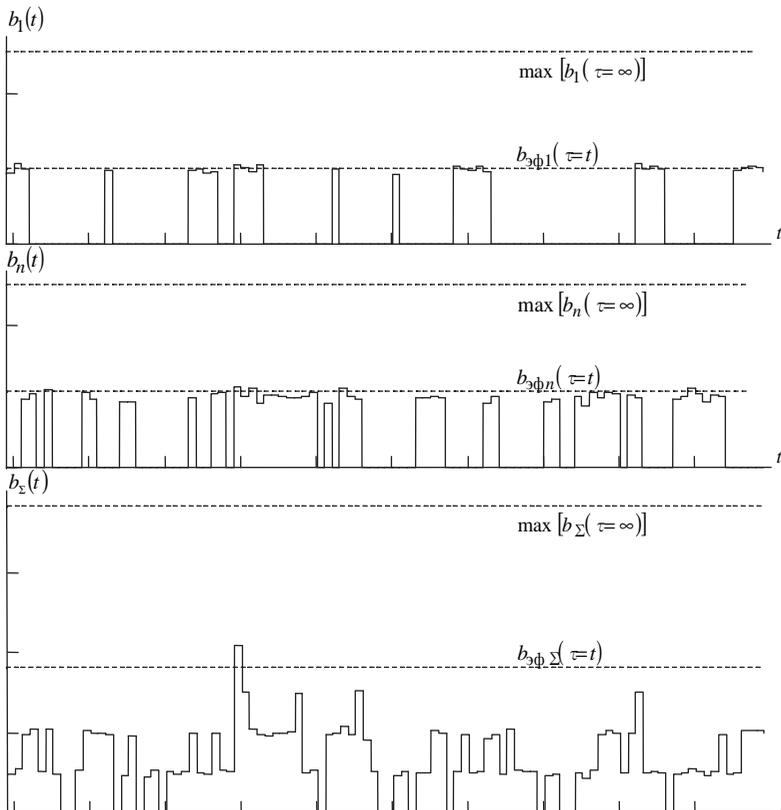


Рис. 1. Скорости передачи потоков ПБД от ИУИ и суммарного потока ПБД в КСУ

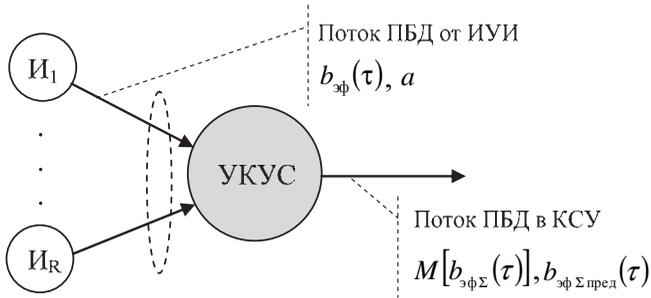


Рис. 2. Суммарный поток ПБД в КСУ для соединений "многоточка-точка"

Задавшись значением вероятности потери соединения  $P^{\text{пот.соед}}$  можно найти предельную эффективную скорость по следующей формуле:

$$b_{\text{эф } \Sigma \text{ пред}}(t) = b_{\text{эф}}(t) \cdot n_{\text{пред}}, \quad (5)$$

где  $n_{\text{пред}}$  – это предельное значение числа сеансов служебной связи "многоточка-точка" организуемых через УКУС, определяемое как:

$$n_{\text{пред}} = \inf \left\{ n \mid \sum_{j=0}^n [C_R^j \cdot a^j \cdot (1-a)^{R-j}] \geq 1 - P^{\text{пот.соед}} \right\} \quad n = 0, 1, \dots, R, \quad (6)$$

где  $\inf \{ \dots \}$  – точная нижняя граница.

Рассмотрим вариант расчета эффективной скорости передачи суммарного потока ПБД в КСУ для соединений "точка-точка" (рисунков 3). Пусть к первому УКУС подключено  $R_1$  однотипных ИУИ, а ко второму –  $R_2$ .

Очевидно, что максимальное число сеансов служебной связи "точка-точка" для фрагмента СОДУ, представленной на рисунке 3, определяется по формуле:

$$n_{\text{max}} = \min \{ R_1, R_2 \}. \quad (7)$$

Определим вероятность того, что в произвольный момент времени будет открыто ровно  $i$  сеансов служебной связи. На рисунке 4 представлена диаграмма интенсивностей переходов марковского про-

цесса, описывающего события открытия и закрытия сеансов служебной связи. Будем считать, что среднее время сеанса служебной связи является условной единицей времени (УЕВ) относительно которой определена интенсивность источника в свободном состоянии  $\alpha$ .

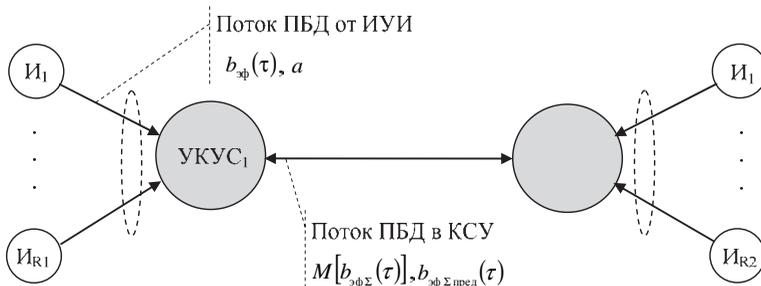


Рис. 3. Суммарный поток ПБД в КСУ для соединений "точка-точка"

В отличие от распределения Энгсета (РЭ), в такой постановке учитывается, что в сеансе служебной связи участвует два ИУИ, и ни один из них, не сможет открыть новый сеанс, не закончив текущий.

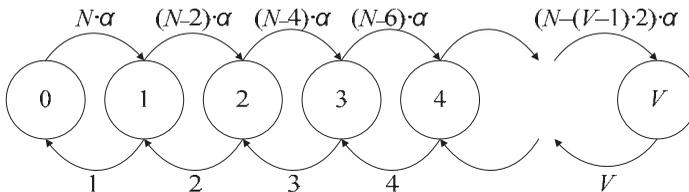


Рис. 4. Диаграмма интенсивностей переходов марковского процесса

Вероятность нахождения системы в  $i$ -м состоянии можно найти как:

$$\begin{aligned}
 P_i &= P_0 \cdot \frac{\prod_{k=0}^{i-1} (N - k \cdot 2) \cdot \alpha}{i!} = P_0 \cdot \frac{\alpha^i \cdot (N - 0 \cdot 2) \cdot (N - 1 \cdot 2) \cdot \dots \cdot (N - (i-1) \cdot 2)}{i!} = \\
 &= P_0 \cdot \frac{\alpha^i \cdot (N - 0 \cdot 2) \cdot (N - 1 \cdot 2) \cdot (N - 2 \cdot 2) \cdot \dots \cdot (N - (i-1) \cdot 2)}{i!} \cdot \frac{2^i}{2^i} =
 \end{aligned}$$

$$\begin{aligned}
&= P_0 \cdot \frac{\alpha^i \cdot 2^i \cdot \left(\frac{N}{2} - 0\right) \cdot \left(\frac{N}{2} - 1\right) \cdot \left(\frac{N}{2} - 2\right) \cdot \dots \cdot \left(\frac{N}{2} - i + 1\right) \cdot \left(\frac{N}{2} - i\right)!}{i! \cdot \left(\frac{N}{2} - i\right)!} = \\
&= P_0 \cdot C_{\lceil N/2 \rceil}^i \cdot \alpha^i \cdot 2^i, \tag{8}
\end{aligned}$$

где  $\lceil \dots \rceil$  – это функция округления до ближайшего целого в большую сторону.

Учитывая, что  $\frac{N!}{i!(N-i)!} = C_N^i$ ,  $V = \min[R_1, R_2]$ ,  $N = (R_1 + R_2)$ ,

$\alpha = a \cdot (1-a)^{-1}$  запишем модифицированное распределение Энгсета (МРЭ) в следующем виде:

$$P_i = \frac{C_{\lceil \frac{R_1+R_2}{2} \rceil}^i \cdot a^i \cdot (1-a)^{-i} \cdot 2^i}{\sum_{j=0}^{\min\{R_1, R_2\}} \left( C_{\lceil \frac{R_1+R_2}{2} \rceil}^j \cdot a^j \cdot (1-a)^{-j} \cdot 2^j \right)}. \tag{9}$$

Среднее число сеансов служебной связи "точка-точка" можно рассчитать по следующей формуле:

$$M[n] = \frac{\sum_{i=0}^{\min\{R_1, R_2\}} \left( C_{\lceil \frac{R_1+R_2}{2} \rceil}^i \cdot a^i \cdot (1-a)^{-i} \cdot 2^i \cdot i \right)}{\sum_{j=0}^{\min\{R_1, R_2\}} \left( C_{\lceil \frac{R_1+R_2}{2} \rceil}^j \cdot a^j \cdot (1-a)^{-j} \cdot 2^j \right)}, \tag{10}$$

а среднее значение эффективной скорости потока ПБД в КСУ соответственно:

$$M[b_{\text{эф}} \Sigma(t)] = b_{\text{эф}}(t) \cdot \frac{\sum_{i=0}^{\min\{R_1, R_2\}} \left( C_{\lceil \frac{R_1+R_2}{2} \rceil}^i \cdot a^i \cdot (1-a)^{-i} \cdot 2^i \cdot i \right)}{\sum_{j=0}^{\min\{R_1, R_2\}} \left( C_{\lceil \frac{R_1+R_2}{2} \rceil}^j \cdot a^j \cdot (1-a)^{-j} \cdot 2^j \right)}. \tag{11}$$

Задавшись значением вероятности потери сеанса служебной связи "точка-точка"  $P^{\text{пот.соед}}$  можно найти предельную эффективную скорость по формуле (5), где  $n_{\text{пред}}$  – это предельное значение одно-временных сеансов служебной связи "точка-точка" организуемых через КСУ (рис. 3), определяемое как:

$$n_{\text{пред}} = \inf \left\{ n \mid \frac{\sum_{i=0}^n \left( C_{\lfloor \frac{R_1+R_2}{2} \rfloor}^i \cdot a^i \cdot (1-a)^{-i} \cdot 2^i \cdot i \right)}{\sum_{j=0}^{\min\{R_1, R_2\}} \left( C_{\lfloor \frac{R_1+R_2}{2} \rfloor}^j \cdot a^j \cdot (1-a)^{-j} \cdot 2^j \right)} \geq 1 - P^{\text{пот.соед}} \right\}, \quad (12)$$

$$n = 0, 1, \dots, \min\{R_1, R_2\}.$$

Анализ графиков, представленных на рисунке 5, позволяет сделать вывод о том, что последствие в потоке заявок на установление сеансов служебной связи в МРЭ более выражено, чем в РЭ [5]. Отметим, что учет выявленной закономерности позволит более рационально распределять каналный ресурс СОДУ.

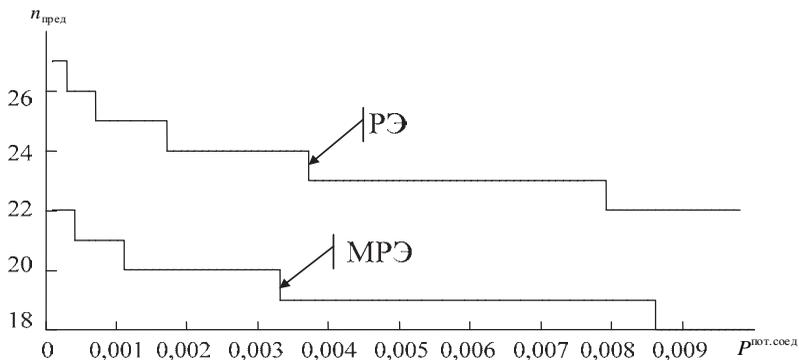


Рис. 5. Зависимость  $n_{\text{пред}}$  от  $P^{\text{пот.соед}}$  для сеансов служебной связи "точка-точка"

Для анализа адекватности разработанной модели процесса мультиплексирования ПБД в КСУ в среде *AnyLogic* была разработана имитационная модель. Сравнение результатов имитационного и аналитического моделирования приведено на рисунке 6.

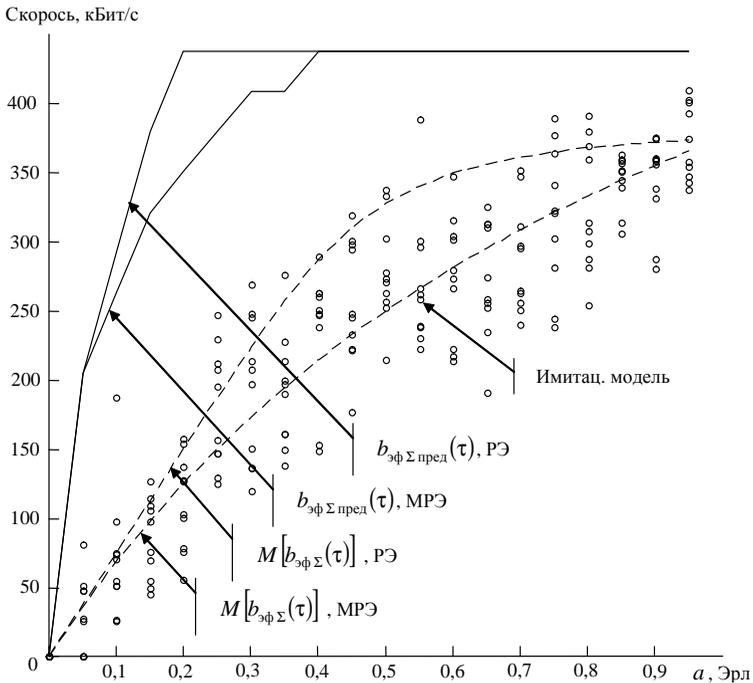


Рис. 6. Сравнение результатов имитационного и аналитического моделирования процесса мультимплексирования ПБД в КСУ

Анализ результатов объективно доказывает, что реальный поток заявок на установление сеансов служебной связи действительно обладает большим последствием, чем это заложено в аналитической модели Энгсета.

**3. Заключение.** Разработанная в работе модель процесса мультимплексирования ПБД в КСУ позволяет определить основные параметры трафика на выходе УКУС для соединений "многоточка-точка" и "точка-точка". МРЭ позволяет точнее оценивать среднее и предельное число сеансов связи "точка-точка", что может способствовать более рациональному распределению канального ресурса СОДУ. Модель соответствует объекту исследования, обладает необходимой чувствительностью к изменению входных параметров и сохраняет устойчивость на всем возможном диапазоне изменения входных параметров, а также при изменении структуры СОДУ. Разработанная модель может быть применена для решения расчетных и исследовательских задач синтеза СОДУ, обеспечивающих минимизацию расхода ресурса линий передачи ГТКС.

## Литература

1. ITU-T Recommendation M.3010. Principles for a telecommunications management network. URL: <http://faculty.wiu.edu/Y-Kim2/M.3010.pdf> (дата обращения: 28.10.2013).
2. Дымарский Я.С., Крутякова Н.П., Яновский Г.Г. Управление сетями связи: принципы, протоколы, прикладные задачи // М.: ИТЦ "Мобильные коммуникации", 2003. 384 с.
3. Степанов С.Н. Основы телетрафика мультисервисных сетей // М.: Эко-Трендз, 2010. 392 с.
4. Kelly F.P. Notes on effective bandwidths // In: F. Kelly, S. Zachary and I. Ziedinis (Eds.), Stochastic Networks: Theory and Applications Telecommunications Networks, Volume 4 of Royal Statistical Society Lecture Notes Series, Oxford University Press. pp. 141–168.
5. Клейнок Л. Теория массового обслуживания // М.: Машиностроение. 1979. 432 с.

## References

1. ITU-T Recommendation M.3010. Principles for a telecommunications management network. Available at: <http://faculty.wiu.edu/Y-Kim2/M.3010.pdf> (accessed 28.10.2013).
2. Dyimarsky Ya.S., Krutyakova N.P., Yanovsky G.G. *Upravlenie setjami svyazi: principy, protokoly, prikladnye zadachi* [Management of telecommunication: principles, protocols, and applied problems]. Moscow: ITC "Mobilnyiekommunikatsii", 2003. 384 p. (In Russ.).
3. Stepanov S.N. *Osnovy tetrafika mul'tiservisyh setej* [Basics of teletraffic of multi-service networks]. Moscow:Eco-Trends, 2010. 392 p. (In Russ.).
4. Kelly, F.P. Notes on effective bandwidths. In: F. Kelly, S. Zachary and I. Ziedinis (Eds.), Stochastic Networks: Theory and Applications Telecommunications Networks, Volume 4 of Royal Statistical Society Lecture Notes Series, Oxford University Press. pp. 141–168.
5. Kleinrock L. *Teoriya massovogo obsluzhivaniya* [Theory of Queuing Systems]. Moscow: Mashinostroenie, 1979. 432 p.(In Russ.).

**Трегубов Роман Борисович** — к-т техн. наук, старший научный сотрудник Академии ФСО России. Область научных интересов: теория массового обслуживания, теория вероятности, применение методов математического моделирования в телекоммуникациях. Число научных публикаций — 60. [treba@list.ru](mailto:treba@list.ru); Академия ФСО России, ул Приборостроительная, д. 35, г. Орел, 302034, РФ; р.т. +7(4862)54-9731.

**Tregubov Roman Borisovich** — Ph.D., Senior researcher of the Academy of Federal Agency of protection of Russian Federation. Research interests: waiting theory, probability theory, application of mathematical model approaches in telecommunications. The number of publications — 60. [treba@list.ru](mailto:treba@list.ru); Academy of Federal Agency of protection of Russian Federation, Priborostroitelnaya Street, 35, Orel, 302034, Russia, office phone +7(4862)54-9731.

**Алексиков Юрий Григорьевич** — старший научный сотрудник Академии ФСО России. Область научных интересов: теория массового обслуживания, теория вероятности, применение методов математического моделирования в телекоммуникациях. Число научных публикаций — 29. [allexic@mail.ru](mailto:allexic@mail.ru); Академия ФСО России, ул Приборостроительная, д. 35, г. Орел, 302034, РФ; р.т. +7(4862)54-9731.

**Aleksikov Jurij Grigor'evich** — Senior researcher of the Academy of Federal Agency of protection of Russian Federation. Research interests: waiting theory, probability theory, application of mathematical model approaches in telecommunications. The number of publications — 29. [allexic@mail.ru](mailto:allexic@mail.ru); Academy of Federal Agency of protection of Russian Federation, Priborostroitelnaya Street, 35, Orel, 302034, Russia, office phone +7(4862)54-9731.

## РЕФЕРАТ

### *Tregubov R.B., Aleksikov Y.G. Модель процесса мультиплексирования протокольных блоков данных в канале связи управления.*

Эффективное управление гетерогенными телекоммуникационными сетями требует обеспечения полноты, своевременности и достоверности процесса доставки управляющей информации. При этом требуется решение вопросов минимизации непроизводительных затрат ресурсов пропускной способности линий передачи телекоммуникационных систем.

Целью моделирования, представленного в настоящей статье, является формальное описание процесса мультиплексирования потоков протокольных блоков данных в каналах связи управления для повышения обоснованности принимаемых решений в ходе синтеза сети обмена данными управления, обеспечивающего минимизацию затрат ресурсов пропускной способности линий передачи телекоммуникационных систем.

Исследованы характеристики трафика одиночного источника управляющей информации, а также потока протокольных блоков данных, образованного путем мультиплексирования данных от группы источников.

В полученном модифицированном распределении Энгсета учитывается, что в сеансе служебной связи участвует два источника управляющей информации, и ни один из них, не сможет открыть новый сеанс, не закончив текущий. Использование модифицированной формулы Энгсета позволяет обеспечить более рациональное распределение канального ресурса, необходимого для организации доставки протокольных блоков данных.

Разработанная модель может быть применена для решения расчетных и исследовательских задач синтеза сетей обмена данными управления, обеспечивающих минимизацию расхода ресурса линий передачи гетерогенных телекоммуникационных систем.

## SUMMARY

### *Tregubov R.B., Aleksikov Y.G. Model of process of multiplexing the protocol data units in management channel.*

Effective management of heterogeneous telecommunication systems requires ensuring of the completeness, timeliness and reliability of delivery process of controlling information. It requires the decision of questions of the minimization of unproductive expenses of transmission links of telecommunication systems.

The goal of modeling is formal description of process of multiplexing stream of the protocol data units in management channel for improving the validity of the accepted decision during the synthesis of management-data-exchange network, providing the minimization of resource costs of traffic capacity of transmission links of the telecommunication systems.

The characteristics of single sources of controlling information and stream of protocol data units, educated by multiplexing the protocol data units from group of sources are investigated.

In the result the modified Engset distribution two sources of controlling information is considered in a service session, and none of them can open new session after finishing the current one. Use of the modified formula of Engset allows us to provide more rational distribution of the channel resource necessary for the organization of transmission of protocol data units.

The developed model can be applied for the solution of settlement and reseach tasks of synthesis of management-data-exchange network, providing the minimization of resource costs of transmission links of telecommunication systems.



Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Труды СПИИРАН» <http://www.proceedings.spiiras.nw.ru>. При регистрации авторам рекомендуется заполнить все предложенные поля данных, так как это значительно ускорит процесс оформления метаданных к новым статьям.

Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше. При подаче материала в редакцию сначала отправляется только статья в формате \*.docx. Для обеспечения требований слепого рецензирования при представлении статьи в журнал авторам необходимо удалить персональные данные, содержащиеся в тексте файла и его свойствах.

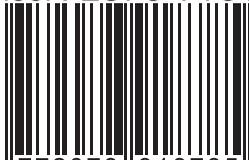
Объем основного текста – от 5 до 20 страниц включительно. Формат страницы документа – А5 (148 мм ширина, 210 мм высота); ориентация – портретная; все поля – 20 мм. Верхний и нижний колонтитулы страницы – пустые. Основной шрифт документа – Times New Roman, основной кегль (размер) шрифта – 10 pt. Переносы разрешены. Абзацный отступ устанавливается размером в 10 мм. Межстрочный интервал – одинарный. Номера страниц не проставляются.

Не допускается использования цветных шрифтов, цветовых выделений и цветных рисунков. Статьи должны быть полностью готовы к черно-белой печати.

Основная часть текста статьи разбивается на разделы, среди которых являются обязательными: введение, хотя бы один «содержательный» раздел и заключение. Допускается также мотивированное содержанием и структурой материала выделение подразделов.

В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

ISSN 2078-9181



9 772078 918785 >