



Речевые

ТЕХНОЛОГИИ

1/2012

Главный редактор Харламов А.А., доктор технических наук

Состав редколлегии:

- Потапова Р.К.*, доктор филологических наук, профессор,
заместитель главного редактора
Ронжин А.Л., доктор технических наук, доцент
Женило В.Р., доктор технических наук, профессор
Жигулёвцев Ю.Н., кандидат технических наук, доцент
Кривнова О.Ф., доктор филологических наук, профессор
Кушнир А.М., кандидат психологических наук
Лобанов Б.М., доктор технических наук (Беларусь)
Максимов Е.М., доктор технических наук
Голенков В.В., доктор технических наук, профессор (Беларусь)
Ромашкин Ю.Н., кандидат технических наук
Петровский А.А., доктор технических наук, профессор (Беларусь)
Хитров М.В., кандидат технических наук
Чучупал В.Я., кандидат физико-математических наук
Шелепов В.Ю., доктор физико-математических наук (Украина)
Кушнир Д.А., кандидат технических наук, ответственный секретарь

Содержание

<i>Голенков В.В., Гулякина Н.А.</i> Графодинамические модели параллельной асинхронной обработки знаний	3
<i>Гусев М.Н., Дегтярёв В.М.</i> Выделение ключевых слов	14
<i>Зулкарнеев М.Ю., Репалов С.А., Шамраев Н.Г.</i> Применение синтаксического анализа в задаче распознавания речи	22
<i>Иванов В.Б.</i> Синтез просодических моделей в языке дари	29

<i>Каганов А.Ш.</i> Взаимосвязь и взаимодополнение составных частей комплексного идентификационного исследования личности по голосу и звучащей речи	46
<i>Потапова Р.К., Потапов В.В.</i> Модификация речевого сигнала как следствие наличия эмоциональных состояний «страх»/«тревожность»	52
<i>Свириденко В.А.</i> Аппаратно программная платформа для ИКН-терминалов	61
<i>Шелепов В.Ю., Ниценко А.В.</i> Новый подход к определению границ речевого сигнала. Проблемы конца сигнала	74
<i>Юрьев Г.А., Куравский Л.С.</i> Технология распознавания и озвучивания текстов для обеспечения учебного процесса людей с нарушениями зрения	79
Отчёт о конференции OSTIS-2012	91
Информация о конференции OSTIS-2013	94

Редакция:

Редактор — Елена Долматова
 Выпускающий редактор — Анастасия Чипенко
 Корректор — Татьяна Денисьева
 Дизайн — Анна Ладанюк
 Вёрстка — Александр Перевозов

Адрес редакции: 109341, Москва, ул. Люблинская, д. 157, корп. 2
Тел.: 8 (495) 979-54-27

Подписано в печать 20.04.2012. Формат 60х90/8. Бумага офсетная. Печать офсетная
 Печ. л. 12,0. Тираж 1000 экз. Заказ № . Издательский дом «Народное образование»
 Отпечатано в ООО «Чебоксарская типография № 1», 428019, г. Чебоксары, пр. И. Яковлева, 15

Графодинамические модели параллельной асинхронной обработки знаний

Голенков В.В., доктор технических наук, профессор

Гулякина Н.А., кандидат физико-математических наук, доцент

В статье рассматриваются принципы построения технологии проектирования интеллектуальных систем, ориентированных на семантическое представление знаний, расширение контингента разработчиков и сокращение сроков проектирования.

• графодинамическая модель обработки информации • графодинамическая память • многоагентная система над графодинамической памятью • графовый язык программирования • семантическая сеть • бинарная семантическая сеть • язык визуализации семантических сетей • унификация семантических сетей • интеграция семантических сетей • семантическая технология проектирования интеллектуальных систем • семантическая технология проектирования баз знаний • семантическая технология проектирования решателей задач • семантическая технология проектирования пользовательских интерфейсов • библиотека семантически совместимых многократно используемых компонентов интеллектуальных систем.

The principles of construction technology of designing intelligent systems which are oriented on semantic representation of knowledge, expansion of the number of developers and shortening time of design are considered.

• graph-dynamic information processing model • graph-dynamic memory Multiagent system over graph-dynamic memory • graph programming language • semantic network • binary semantic network • semantic network visual language • unification of semantic network • semantic network integration • semantic network merging • semantic network fusion • semantic intelligent system design technology • semantic knowledge base design technology • semantic problem solver design technology • semantic user interface design technology • library of reusable semantically compatible components for intelligent systems.

Введение

Мы исходим из того, что основным результатом исследований в области искусственного интеллекта является **разработка технологий**, позволяющих быстро и в большом количестве порождать самые разнообразные интеллектуальные системы, имеющие большую практическую ценность. Очевидно, что составляющие таких технологий:

- формальная теория интеллектуальных систем;
- методы проектирования интеллектуальных систем;
- инструментальные средства (средства автоматизации проектирования интеллектуальных систем);



— средства информационной поддержки (информационного обслуживания) разработчиков интеллектуальных систем;

— средства компьютерной поддержки управления коллективной разработкой интеллектуальных систем.

Современные технологии проектирования интеллектуальных систем имеют ряд недостатков:

— технологии искусственного интеллекта *не ориентированы на широкий круг разработчиков* интеллектуальных систем и, следовательно, не получили массового распространения;

— *велики сроки разработки* и трудоёмкость их сопровождения;

— *высока степень зависимости технологий искусственного интеллекта от платформ*, на которых они реализованы, что приводит к существенным изменениям технологий при переходе на новые платформы;

— для эффективной реализации даже существующих моделей представления знаний и моделей решения трудно формализуемых задач *современные компьютеры оказываются плохо приспособленными*. Это требует разработки принципиально новых компьютеров;

— современное состояние в области проектирования интеллектуальных компьютерных систем представляет собой «вавилонское столпотворение» самых различных моделей, методов, средств, платформ;

— отсутствуют подходы, позволяющие на некоторой универсальной основе *интегрировать научные и практические результаты* в области искусственного интеллекта, что порождает высокую степень дублирования результатов. В частности, высока *трудоёмкость интеграции* различных моделей представления информации, моделей обработки информации, моделей решения задач и, следовательно, различных компьютерных систем.

Искусственный интеллект — междисциплинарная научная дисциплина. Этим обусловлен большой её потенциал, так как на стыках научных дисциплин рождаются сильные результаты. Но этим же обусловлены и большие трудности, так как развитие искусственного интеллекта требует глубокого взаимопонимания и сотрудничества исследователей, имеющих разные стиль мышления, подход к объекту и предмету исследования, менталитет, целевые установки и традиции. Современный этап развития искусственного интеллекта остро нуждается в преодолении указанных трудностей.

Важнейшей задачей искусственного интеллекта в настоящее время является построение общей комплексной теории интеллектуальных систем, в рамках которой сочетались бы разные направления искусственного интеллекта: и теория представления знаний, и теория решения задач, в том числе различных исчислений, эвристик, стратегий, и теория программ (процедурных, декларативных, параллельных, последовательных), и архитектуры интеллектуальных систем, в том числе детализированные до уровня аппаратной поддержки, и теория интеллектуальных пользовательских интерфейсов, и компьютерная лингвистика.

В настоящее время эпицентром развития искусственного интеллекта является не столько разработка отдельных его направлений, сколько их глубокая **семантическая интеграция**, целью которой должна быть не только общая теория интеллектуальных систем, но и общая, доступная технология их комплексного проектирования.

В основе предлагаемого нами подхода к созданию технологии проектирования интеллектуальных систем, направленного на устранение вышеуказанных недостатков, лежат следующие принципы.

Принцип 1. Использовать опыт наиболее продвинутых технологий

В первую очередь, имеется в виду технология проектирования микросхем, которая за последнее время обеспечила сокращение времени и повышение качества разработок, благодаря:

- созданию языковых средств формального описания проектируемых микросхем на разных уровнях детализации;
- чёткому разделению процесса разработки формальных описаний микросхем и процесса их реализации по заданным формальным описаниям;
- созданию мощных и доступных библиотек формальных описаний типовых (многократно используемых) компонентов микросхем.

Для того чтобы аналогичным образом построить технологию проектирования интеллектуальных систем, необходимо:

- создать языковые средства полного унифицированного формального описания интеллектуальных систем;
- чётко отделить разработку полного унифицированного формального описания проектируемой интеллектуальной системы от разработки различных вариантов интерпретации таких формальных описаний интеллектуальных систем;
- создать библиотеки формальных описаний типовых (многократно используемых) компонентов интеллектуальных систем. Но для того чтобы такая технология была создана, необходимо обеспечить **интегрируемость** (семантическую совместимость) указанных компонентов интеллектуальных систем.

Принцип 2. Графодинамические модели

В качестве формальной основы проектируемых интеллектуальных систем использовать **графодинамические модели обработки информации**.

Графодинамическая модель обработки информации трактует процесс обработки информации как процесс преобразования графовой структуры, в ходе которого меняется не только состояние элементов этой графовой структуры, но и её конфигурация (появляются или удаляются её вершины, а также связи между ними). Заметим, что для создания графодинамических моделей обработки информации недостаточно тех видов графовых структур, которые в настоящее время исследуются в теории графов. Нам потребуется не только увеличение числа компонентов инцидентных ребру (т.е. переход от ребра к **гиперребру**), но и увеличение числа компонентов инцидентных дуге (т.е. переход от дуги к **гипердуге**, которая по сути является графовой трактовкой многоместного кортежа). Нам потребуются не только рёбра, гиперрёбра, дуги, гипердуги, компонентами которых являются вершины графовой структуры, но и рёбра, гиперрёбра, дуги, гипердуги, компонентами которых являются другие рёбра, гиперрёбра, дуги и гипердуги. Нам потребуются связующие элементы графовых структур, задающие целые фрагменты (подструктуры) заданной графовой структуры, в состав которых входят соответствующие вершины, рёбра, гиперрёбра, дуги, гипердуги.

Приведём общее определение **графовой структуры**, на основе которого можно строить практически полезные графодинамические модели обработки информации. Графовая структура G задаётся пятёркой $\langle V, C, K, M, I \rangle$, где:

- V — множество **вершин** (первичных элементов, терминальных элементов);
- C — множество **связующих элементов** графовой структуры, каждый из которых задаёт некоторый фрагмент графовой структуры;
- K — множество **ключевых вершин** графовой структуры, каждая из которых задаёт некоторый класс эквивалентных (однотипных) в определённом смысле элементов графовой структуры ($K \subset V$);



M — множество **меток** элементов (алфавит элементов) графовой структуры, каждая из которых задаёт некоторый базовый класс эквивалентных, в определённом смысле, элементов графовой структуры. К таким классам элементов, в частности, относятся:

- класс всех вершин графовой структуры;
- класс всех связующих элементов графовой структуры;
- класс всех ключевых вершин графовой структуры;
- класс всех меток графовой структуры;
- класс всех используемых в графовой структуре отношений инцидентности элементов графовой структуры.

I — множество используемых в графовой структуре **отношений инцидентности** элементов. Все эти отношения инцидентности являются бинарными ориентированными отношениями. Среди этих отношений выделим:

- отношения инцидентности вершин. Примером такого отношения является последовательность символов в строке символов;
- отношения инцидентности, каждая пара которых связывает связующий элемент графовой структуры с элементом (компонентом) того фрагмента, который задаётся этим связующим элементом. Подчеркнём, что компонентами связующего элемента могут быть элементы графовой структуры любого вида (вершины, связующие элементы, метки, отношения инцидентности). Подчеркнём также, что компонентами одного и того же связующего элемента могут выполнять разные роли в рамках соответствующего фрагмента графовой структуры, который задаётся связующим элементом. Указанные роли соответствуют разным отношениям инцидентности, входящим во множество I ;
- отношения инцидентности, каждая пара которых связывает ключевую вершину графовой структуры с тем элементом графовой структуры, который входит в класс элементов, задаваемый этой ключевой вершиной. Подчеркнём, что элементами, инцидентными ключевой вершине, могут быть элементы графовой структуры любого вида;
- отношения инцидентности, каждая пара которых связывает метку графовой структуры с тем элементом графовой структуры, который имеет указанную метку. Подчеркнём, что элементами, инцидентными метке, могут быть элементы графовой структуры любого вида.

Заметим, что связь каждого отношения инцидентности (каждого элемента множества I) с соответствующими парами инцидентности и связь каждой пары инцидентности с элементами графовой структуры, соединяемыми этой парой инцидентности, можно условно считать неявно задаваемыми связями инцидентности более низкого уровня.

Каждую графовую структуру G мы будем трактовать как множество всех элементов, входящих в её состав: $G = (V \cup C \cup K \cup M \cup I)$. Таким образом, в число элементов графовой структуры входят все её вершины (в том числе ключевые), связующие элементы, метки и отношения инцидентности.

Множество связующих элементов графовой структуры можно разбить на:

- множество **связок** (простых связующих элементов);
- множество **подструктур**, каждая из которых задаёт фрагмент графовой структуры. В состав этого фрагмента входят такие связующие элементы, которые связывают элементы графовой структуры, входящие в указанный фрагмент.

В свою очередь, по признаку ориентированности множество связок можно разбить на:

- множество **ориентированных связей**, некоторые компоненты которых выполняют в рамках этих связей *разные роли*;
- множество **неориентированных связей**, все компоненты которых выполняют в рамках этих отношений связей *одинаковые роли*.

Важным частным случаем ориентированности связки является **кортеж**. Он задаёт такое подмножество элементов графовой структуры, в котором роли всех элементов пронумерованы. То есть в рамках указанного подмножества имеется элемент, который в этом подмножестве играет роль первого элемента (первого компонента), имеется элемент, который в этом подмножестве играет роль второго элемента (второго компонента) и т. д.

Если в графовой структуре имеются кортежи, то в число её отношений инцидентности должны входить следующие отношения:

- быть первым компонентом;
- быть вторым компонентом;
- быть третьим компонентом;
- и т. д.

По количеству компонентов множество связей можно разбить на:

- множество **унарных связей** (одноместных, однокомпонентных);
- множество **бинарных связей** (двухместных, двухкомпонентных);
- множество **многокомпонентных связей** (многоместных), имеющих более двух компонентов.

Неориентированные бинарные связки будем называть **рёбрами**, ориентированные — **дугами**. Неориентированные многокомпонентные связки будем называть **гиперрёбрами**, а ориентированные — **гипердугами**.

Множество связующих элементов графовой структуры можно трактовать как подмножество **шкалы множеств**, заданной над множеством

$$(V \cup M \cup I).$$

Шкала множеств (H) над указанным множеством определяется рекурсивно:

$$(1) H \supset (V \cup M \cup I),$$

$$(2) \text{ если } hj_1, hj_2, \dots, hj_n \in H,$$

то $\{hj_1, hj_2, \dots, hj_n\} \in H$.

Т.е. любое множество, состоящее из любых элементов шкалы множеств, само также является одним из элементов шкалы множеств.

$$(3) \text{ если } hj_1, hj_2, \dots, hj_n \in H; ij_1, ij_2, \dots, ij_e \in I,$$

то множество $hj = \{hj_1, hj_2, \dots, hj_n\}$, между которым и множеством $\{ij_1, ij_2, \dots, ij_e\}$ задано произвольное соответствие, определяющее то или иное распределение ролей между элементами множества hj , также является элементом шкалы (т.е. $hj \in H$).

Какие вопросы ассоциируются с рассмотрением графодинамической модели (графодинамической парадигмы) обработки информации:

- Можно ли на основе теории графов построить **универсальную абстрактную модель обработки информации**, которая могла бы конкурировать с абстрактной машиной фон Неймана, лежащей в основе традиционных компьютерных систем?
- Нужно ли это делать, какими преимуществами эта модель обладает по сравнению с абстрактной машиной фон Неймана? Какими преимуществами обладают системы, создаваемые на основе этой модели? Что принципиально нового даёт графодинамическая парадигма обработки информации?
- Есть ли к этому предпосылки?



— Можно ли различные модели решения задач (в том числе различные логические исчисления) формально описать в виде графодинамических моделей обработки информации и можно ли обеспечить совместимость (интегрируемость) таких графодинамических моделей?

Интерес к графодинамическим моделям обработки информации имеет достаточно длительную историю. Для подтверждения этого достаточно отметить:

- предложенное А.Н. Колмогоровым уточнение понятия алгоритм [Колмогоров, 1958];
- работы школы М.А. Айзермана по графодинамике [Айзерман, 1988];
- исследования по графовым грамматикам [Петров, 1987];
- исследования по теории программирования и CASE-технологиям [Касьянов, 2003], [Гостев, 1981а], [Гостев, 1981б];
- разработка параллельных моделей обработки информации [Котов, 1966];
- предложенные В.Б. Борщевым и М.В. Хомяковым клубные системы и вегетативная машина [Борщев, 1983].

Для разработки графодинамических моделей обработки информации необходимо рассматривать графовую структуру с позиций семиотики и трактовать её как **знаковую структуру** (текст), представляющую собой систему взаимосвязанных знаков. Такая трактовка графовых структур позволяет «вдохнуть» семантику в теорию графов.

Действительно, почему тексты обязательно должны быть линейными (т.е. цепочками символов). Но, как только мы введём понятие **графового языка** (языка, текстами которого являются в общем случае графовые структуры различной конфигурации), возникают следующие вопросы:

- В чём преимущество графовых языков по сравнению с традиционными линейными языками, текстами которых являются цепочки (строки) символов?
- Можно ли построить **универсальный графовый язык**, обеспечивающий представление информации (знаний) любого семантического вида?
- Можно ли в универсальном графовом языке сделать так, чтобы множество всех меток, используемых во всех графовых структурах, которые являются текстами универсального графового языка, было конечным?

Говоря о графовых языках, следует подчеркнуть то, что графовые структуры, являющиеся текстами таких языков, представляют собой *абстрактные* математические структуры, не уточняющие (не детализирующие) способ их материального представления (например, способ кодирования в компьютерной памяти, способ графического изображения, ориентированного на человеческое восприятие). То есть графовая структура как абстрактный математический объект и её, например, графическое представление, это принципиально разные вещи. Из этого, в частности, следует, что каждому графовому языку может соответствовать несколько языков, использующих разные способы представления (кодирования, изображения) текстов этого графового языка.

Очевидно, что накопленный опыт развития и применения теории графов и все полученные в ней результаты становятся хорошим математическим фундаментом для разработки различных графовых языков и различных графодинамических моделей обработки информации, а также для создания их теории.

Принцип 3. Графодинамические параллельные асинхронные графодинамические модели

В качестве формальной основы проектируемых интеллектуальных систем использовать графодинамические модели специального вида, ориентированные на **параллельную и асинхронную** обработку информации.

Почему акцентируется внимание на **параллельных** графодинамических моделях? Во-первых, потому что это актуально, так как без организации параллельной обработки информации невозможно рассчитывать на необходимую производительность подавляющего числа практически полезных интеллектуальных систем. Во-вторых, потому что целый ряд исследований [Котов, 1966] показал перспективность создания параллельных моделей обработки информации именно на основе графодинамического подхода.

Почему отдаётся предпочтение **асинхронному** варианту управления обработкой информации? Потому, что асинхронные модели обработки информации являются более гибкими, их легче интегрировать и наращивать новыми функциональными возможностями.

Графодинамическая модель параллельной асинхронной обработки информации, которую будем также называть **графодинамической параллельной асинхронной машиной**, трактуется нами как абстрактная **многоагентная система**, состоящая из:

- абстрактной **графодинамической памяти**, в которой хранятся обрабатываемые графовые структуры;
- **коллектива агентов**, работающих *над общей для них графодинамической памятью* и обменивающихся информацией *только* через эту память (в т. ч. и для координации своих действий).

Графодинамическая память носит реконфигурируемый, структурно перестраиваемый характер, поскольку процесс обработки графовых структур в конечном счёте сводится к генерации и удалению различных элементов графовых структур, а также к генерации и удалению пар инцидентности между этими элементами. Другими словами, процесс обработки информации в графодинамической памяти сводится не только к изменению состояния элементов памяти, но и к изменению конфигурации связей между ними.

Агенты, работающие над общей графодинамической памятью, делятся на три вида:

- внутренние агенты, каждый из которых реагирует на определённого вида ситуации или события в графодинамической памяти и осуществляет изменение состояния графодинамической памяти, соответствующее своему функциональному назначению;
- рецепторные агенты, каждый из которых реагирует на определённые события во внешней среде и осуществляет первичное отражение этих событий в графодинамической памяти;
- эффекторные агенты, каждый из которых реагирует на определённого вида команды, формируемые внутренними агентами в графодинамической памяти, и осуществляет соответствующее изменение материального (физического) состояния интеллектуальной системы, которое определённым образом влияет на изменение её внешней среды.

Агенты могут работать параллельно, если одновременно возникают условия их инициирования.

Асинхронность деятельности внутренних агентов заключается в том, что наличие условия инициирования агента ещё не означает начала его работы. То есть время реакции каждого внутреннего агента в известной мере субъективно и достаточно произвольно. В этом смысле указанные агенты обладают свободой выбора:

- момента начала реакции на условие инициирования;
- последовательности обработки условий инициирования, если в текущий момент возникло несколько таких условий.

Для обеспечения эффективного взаимодействия агентов, работающих над общей графодинамической памятью, наряду с представляемой им свободой, необходима разработка



таких правил их поведения, которые гарантируют безопасность и производительность каждого из них. В конечном счёте, эти правила сводятся к двум положениям:

- позаботься о своей безопасности, точнее, об обеспечении безопасного выполнения своей задачи;
- не навреди другим агентам (помни о том, что ты не один, не создавай для других «аварийных» ситуаций).

Для обеспечения безопасного выполнения своей задачи агент блокирует некоторые элементы графовой структуры, которая хранится в общей графодинамической памяти. Блокировка — это запрет, установленный заданным агентом и адресованный другим агентам, на выполнение тех или иных действий над заданным элементом хранимой графовой структуры. Таким образом, существует несколько видов таких блокировок. Приведём некоторые из них:

- запрет на удаление заданного элемента графовой структуры;
- запрет на удаление *всех* элементов хранимой графовой структуры, инцидентных заданному (блокируемому) элементу;
- запрет на удаление заданного вида элементов хранимой графовой структуры, которые связаны с блокируемым элементом выходящей из него (или входящей в него) парой инцидентности, принадлежащей заданному отношению инцидентности;
- запрет на генерацию заданного вида элементов в хранимой графовой структуре, связанных с блокируемым элементом выходящей из него (или входящей в него) парой принадлежности, которая принадлежит заданному отношению инцидентности.

Приведём некоторые правила поведения агента, работающего в коллективе агентов над общей графодинамической памятью:

- не нарушать блокировочные запреты, сформированные другими агентами;
- самому заблокировать тот фрагмент обрабатываемой графовой структуры, целостность которого необходимо сохранить до завершения своей работы;
- не «жадничать» — не блокировать больше, чем надо;
- снимать свои блокировки как можно быстрее, как только в них отпадает необходимость (т. е. желательно это делать до завершения своей работы);
- удалять сгенерированные для своей работы вспомогательные структуры (информационные «леса») как можно быстрее, как только в них отпадает необходимость (т. е. желательно убирать информационный «мусор» по мере возможности до завершения своей работы);
- поиск фрагментов хранимой графовой структуры, являющихся условиями инициирования агента, осуществлять поэтапно, начиная с поиска тех частей этих условий, которые реже появляются в памяти (это необходимо для того, чтобы скорее установить факт отсутствия условий инициирования).

В случае возникновения **конфликтов** между агентами используются внутренние агенты специального вида — метаагенты-судьи — обеспечивающие разрешение таких конфликтов.

Для организации своей деятельности над графодинамической памятью каждый агент «опирается» на соответствующее ему семейство постоянно присутствующих в памяти (резидентных) элементов хранимой в памяти графовой структуры. Указанные элементы будем называть ключевыми элементами

агентов. Очевидно, что такие элементы соответствуют константам программ, описывающих поведение агентов.

Сложность комплексного перехода на графодинамическую парадигму параллельной асинхронной обработки информации определяется исключительно психологическими обстоятельствами — это непривычно и, следовательно, вызывает чувство боязни. Но накопленный человечеством опыт по созданию компьютерных систем, в частности, интеллектуальных систем, позволяет этот переход сделать достаточно быстро, так как многие проблемы, возникающие при реализации и применении графодинамических моделей, имеют достаточно близкие аналоги в традиционных компьютерных системах, но многие из них могут быть решены значительно проще и элегантнее. Кроме того, в результате перехода к графодинамическим моделям «обнажается» целый ряд проблем, которые ранее были не видны.

Эпицентром такого перехода является формализация семантики и разработка семантически совместимых языковых средств представления различных видов знаний.

Принцип 4. СЕМАНТИЧЕСКИЕ МОДЕЛИ ПРЕДСТАВЛЕНИЯ И ОБРАБОТКИ ЗНАНИЙ

В качестве формальной основы проектируемых интеллектуальных систем, в качестве основы абстрактных логико-семантических моделей интеллектуальных систем использовать графодинамические модели специального вида — семантические модели представления и обработки знаний, в основе которых лежат **семантические сети**.

Фактически, речь идёт о создании формальных средств описания семантики различных видов знаний и формальных средств описания обработки знаний на семантическом уровне.

Семантическая сеть — это **графовая структура**, задаваемая пятёркой $\langle V, C, K, M, I \rangle$ и удовлетворяющая следующим требованиям, которые дополняют свойства множеств V, C, K, M, I , указанные в определении графовой структуры:

- каждая вершина $v_j \in V$ является знаком одного из объектов, описываемых семантической сетью;
- каждая ключевая вершина $k_j \in K$ является знаком соответствующего класса элементов графовой структуры G ;
- каждая метка $m_j \in M$ также является знаком соответствующего класса элементов графовой структуры G ;
- каждая пара инцидентности, принадлежащая любому отношению инцидентности $ij \in I$, является парой принадлежности, связывающей знак некоторого множества элементов семантической сети с одним из этих элементов;
- в семантической сети вершины $v_j, v_k \in V$ могут быть инцидентны друг другу, но только в том случае, если по крайней мере одна из них (например, v_k) является ключевой ($v_k \in K$), а вторая является вершиной, принадлежащей множеству, обозначаемому ключевой вершиной v_k (т.е. $v_j \in v_k$);
- каждый элемент множества I ($ij \in I$) является знаком некоторого подмножества отношения принадлежности, задающего определённую роль, выполняемую соответствующими элементами семантической сети в рамках соответствующих множеств таких элементов. Указанные подмножества отношения принадлежности будем называть ролевыми отношениями;
- каждый связующий элемент $c_j \in C$ является знаком некоторого фрагмента графовой структуры G , а точнее знаком некоторого подмножества множества всех элементов графовой структуры G ;
- среди элементов графовой структуры G нет пар, синонимичных друг другу знаков, т.е. знаков, обозначающих один и тот же объект (одну и ту же сущность). либо один и тот



же внешний описываемый объект, либо одно и то же множество элементов графовой структуры;

— среди элементов графовой структуры G нет омонимичных знаков, которые в разных контекстах, в разных обстоятельствах могут обозначать разные сущности.

Следовательно, все (!) элементы (атомарные фрагменты) семантической сети являются знаками различных сущностей (объектов). Такими сущностями могут быть всевозможные внешние описываемые объекты, а также различные множества, состоящие их элементов (атомарных фрагментов) этой же семантической сети.

Таким образом, семантическая сеть — это абстрактная знаковая конструкция «рафинированного вида», в которой нет ничего кроме знаков и инцидентности этих знаков. В частности, в семантической сети отсутствуют элементарные незначащие фрагменты (символы), имена описываемых объектов, слова, из которых эти имена состоят, всевозможные разделители и ограничители, обеспечивающие структуризацию текста. В отличие от текстов традиционного вида, семантическая сеть имеет в общем случае нелинейный характер, поскольку каждый элемент семантической сети может быть инцидентен более чем двум другим элементам.

Семантическую сеть можно трактовать как абстрактный текст, который является семантическим инвариантом соответствующего максимального множества семантически эквивалентных текстов, принадлежащих всевозможным языкам.

На основе понятия семантической сети вводится понятие языка семантических сетей в заданном алфавите и с заданным набором ключевых узлов.

Семантические сети как модели представления знаний известны давно. Но, в отличие от фреймовых, продукционных и логических моделей, для семантических сетей не были разработаны достаточно удобные и практически используемые языки представления знаний, достаточно удобные языки программирования, специально ориентированные на обработку семантических сетей. И, как следствие этого, не были созданы широко используемые комплексные технологии проектирования интеллектуальных систем, в основе которых лежат семантические сети. Причин тому много. Одна из них — это не совсем привычный, нетрадиционный характер таких моделей и возникший на этой основе миф о сложности их реализации. Если такие графодинамические семантические модели реализовывать «в лоб» и без ориентации на последующую аппаратную поддержку, то, конечно, это будет неэффективно.

Но жизнь берёт своё. И на фоне бурного развития микроэлектронных технологий подобного рода мифы выглядят всё менее и менее убедительными.

Более того, развитие Интернет-технологий привело к необходимости формализации семантики информации, обрабатываемой в сети Интернет, что вызвало бурное развитие целого направления — Semantic Web.

Достоинство семантических сетей и семантических моделей обработки информации:

- Представление знаний в виде семантических сетей позволяет существенно упростить процедуру интеграции знаний и свести эту процедуру к выявлению и склеиванию синонимичных элементов интегрируемых семантических сетей.
- Специфика обработки баз знаний заключается в том, что порождаемые (генерируемые) новые фрагменты знаний необходимо не просто построить,

но и погрузить, интегрировать в текущее состояние базы знаний, так как в этих порождаемых фрагментах знаний могут появиться знаки, синонимичные тем, которые уже присутствуют в текущем состоянии базы знаний. Таким образом, процедура интеграции порождаемых фрагментов обрабатываемой базы знаний является процедурой, постоянно используемой в ходе обработки знаний. Следовательно, представление знаний в идее семантических сетей, благодаря упрощению процедуры интеграции знаний, позволяет упростить не только ввод новых знаний извне, но и интеграцию в состав текущего состояния базы знаний новых знаний, порождаемых в ходе решения задач.

- База знаний интеллектуальной системы, представленная в виде корректно построенной семантической сети, полностью исключает дублирование информации в рамках такой базы знаний — каждая информация, представленная соответствующим фрагментом семантической сети, должна находиться в рамках этой семантической сети там и только там, где она должна находиться, и нигде больше.
- Представление знаний в виде семантических сетей позволяет существенно упростить процедуру ассоциативного доступа к различным видам фрагментов хранимой базы знаний, а также существенно расширить типологию запросов (вопросов) к базе знаний.
- Семантические модели обработки знаний не просто хорошо приспособлены к поддержке параллельной асинхронной обработки информации, но и обеспечивают обмен информацией через общую графодинамическую память между различными параллельно (одновременно) протекающими (выполняемыми) процессами, что может существенно ускорить каждый из этих процессов. Примером такого взаимодействия параллельно протекающих процессов является одновременная реализация разных стратегий и тактик, направленных на поиск пути решения заданной нетривиальной задачи.
- С помощью семантических моделей представления и обработки знаний можно интерпретировать все известные виды моделей представления обработки знаний (фреймовые, продукционные, логические), а также все известные модели решения задач различного вида и все известные модели рассуждений. Это даёт возможность рассматривать перечисленные модели не как альтернативные, а как дополняющие друг друга модели, которые могут сосуществовать в разных сочетаниях в разных интеллектуальных системах.

Семантическая модель обработки знаний представляет собой абстрактную многоагентную систему, состоящую из **абстрактной семантической памяти**, в которой хранятся семантические сети, и из множества агентов, ориентированных на обработку семантических сетей, хранимых в указанной семантической памяти.

Семантическую память можно трактовать как абстрактную семантическую модель памяти интеллектуальной системы.

Семейство абстрактных агентов, работающих над семантической памятью вместе с ней самой, можно трактовать как **семантическую модель решения задач**, используемую в соответствующей компьютерной системе, или как операционную семантику этой компьютерной системы. Подчеркнём, что семантическую модель обработки информации можно построить для любой компьютерной системы (как для интеллектуальной системы, так и для компьютерной системы традиционного вида), обеспечивая, тем самым, семантическую совместимость (на абстрактном уровне) не только интеллектуальных систем, но и компьютерных систем любого уровня интеллектуальности.

Всю семантическую сеть (максимальную семантическую сеть), хранимую в семантической памяти (абстрактной логико-семантической модели интеллектуальной системы), будем называть **абстрактной семантической моделью базы знаний** этой интеллектуальной системы.

База знаний должна содержать в себе всю информацию, необходимую агентам, работающим над семантической памятью, для организации коллективной деятельности по решению задач, с которыми должна справляться интеллектуальная система (сюда, в том



числе, входит и описание блокировок, задаваемых разными процессами в семантической памяти).

Семантическая модель базы знаний интеллектуальной системы — это, образно говоря, формальная трактовка «семантического пространства», в котором «живёт» эта интеллектуальная система, а точнее, такого фрагмента, указанного «семантического пространства», который в текущий момент известен указанной интеллектуальной системе.

В целом логико-семантическая модель интеллектуальной системы включает в себя:

- семантическую модель базы знаний этой интеллектуальной системы;
- семантическую машину обработки знаний этой интеллектуальной системы, которая, в свою очередь, состоит из семантической памяти и коллектива агентов над ней.

***Продолжение рубрики «Семантические представления»
читайте в номере 3***

Сведения об авторах

Голенков В.В. —

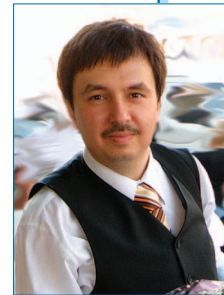
доктор технических наук, профессор, заведующий кафедрой «Искусственного интеллекта» Белорусского государственного университета информатики и радиоэлектроники. Golen@bsuir.ru

Гулякина Н.А. —

кандидат физико-математических наук, доцент, заместитель заведующего кафедрой «Искусственного интеллекта» Белорусского государственного университета информатики и радиоэлектроники. Guliakina@bsuir.ru

Выделение ключевых слов

*Гусев М.Н., кандидат технических наук,
ООО «Вокатив».*



*Дегтярёв В.М., доктор технических наук,
профессор СПбГУТ*



В статье описываются преимущества использования систем поиска ключевых слов для обеспечения безопасности. Рассматриваются различные подходы к построению систем поиска ключевых слов, анализируются их достоинства и недостатки, выбирается оптимальный вариант. Описывается общий алгоритм работы системы, вводятся основные критерии оценки её качества. Приводятся результаты тестирования разработанного решения.

• *распознавание речи* • *поиск ключевых слов.*

The article presents advantages of keyword spotting systems being used for security maintenance. Various approaches to construction of keyword spotting systems are considered, their «pros» and «cons» are analyzed, finally the optimum variant being chosen. The general algorithm is outlined and the basic criteria of its quality estimation are given. Test results of the developed solution are presented.

• *speech recognition* • *key-word spotting.*

Введение

Речь — это идеальное, доступное средство передачи информации, первичный языковой навык и неотъемлемый инструмент общения. Для человека говорить так же естественно, как есть или спать [1]. Чтобы лучше понимать, что происходит в компании, службе безопасности, полезно знать, о чём говорят сотрудники, клиенты и контрагенты.

Отслушка переговоров давно входит в набор средств обеспечения безопасности. Обычно ведётся точечный контроль сотрудников из «группы риска». Регулярная работа создаёт

эффективную среду противодействия рискам. Однако у данного подхода есть ряд недостатков:

- 1) неполнота охвата и недостаточная эффективность;
- 2) высокая вероятность пропустить «подозрительный» разговор;
- 3) отслушка записей переговоров незаконна, так как для неё требуется соответствующее решение суда.

Автоматизация отслушки позволяет сделать охват полным и во много раз повысить оперативность получения данных. Кроме того, решается вопрос с законностью, поскольку на отслушке работает не человек, а машина.

Возможные подходы к построению системы

Поиск ключевых слов работает очень просто: на вход системы подаются записи переговоров и список искомых ключевых слов и фраз (КС), на выходе получается список подозрительных разговоров (см. рис.1).



Рис.1. Схема работы системы

За кажущейся простотой работы системы скрывается реализация сложных математических и лингвистических алгоритмов. Возможны различные подходы к решению задачи поиска КС, обладающие достоинствами и недостатками:

- KWS¹ на основе динамического программирования;
- KWS на фоновой сети;
 - на монофонной сети;
 - на трифонной сети;
- KWS на основе ASR;
 - по словным латтисам;
 - по фоновым латтисам;
- KWS на моделях ключевых слов (КС).

При создании системы KWS на основе принципов динамического программирования КС произносится несколько раз несколькими дикторами. По произнесенным словам строится шаблон слова, который ищется в потоке речи. Это неудобно, так как для каждого искомого слова требуется создавать свой шаблон. Создание нового шаблона и смена списка искомых КС оказывается трудоёмкой и финансово затратной операцией.

Часто для поиска КС используются те же модели, что и для распознавания слитной речи. Сначала обучаются модели отдельных звуков. Затем по моделям звуков строятся фоновые (или фонетические) сети, или модели КС.

¹ От английского key-word spotting – поиск ключевых слов.

На рис. 2 представлен пример структуры данных, используемой в KWS системе, основанной на фоновой сети.

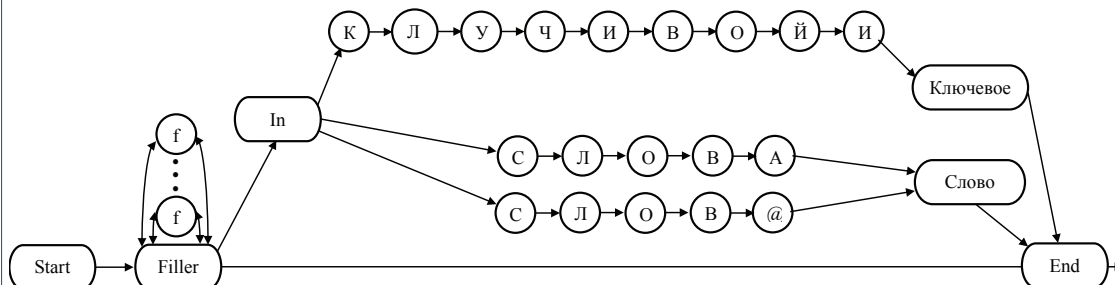


Рис. 2. Пример структуры фоновой сети

В зависимости от используемых моделей звуков фоновые сети разделяются на монофонные и трифонные. В первом случае используются звуковые модели, не учитывающие звуковой контекст и дающие менее точное описание речевого сигнала. Во втором случае точность описания КС значительно повышается, но снижается скорость обработки звука, поскольку увеличивается количество задействованных звуковых моделей. Конечно, возможно применение методов оптимизации фонетической сети, но возможности ускорения упираются в количество и структуру искомых КС.

Во втором, используемом нами, варианте из моделей звуков собираются модели КС. Кроме моделей КС, строятся модели заполнения, описывающие шумы и неречевые сигналы. Также создаются модели речевого мусора или модели усреднённого речевого потока, которые описывают все остальные слова, не являющиеся искомыми.

Для каждого КС строятся свои модели заполнения и усреднённого речевого потока, что позволяет оптимизировать их структуру и увеличить качество поиска. Пример структуры данных, используемой в KWS системе на основе моделей КС, приведён на рис. 3.

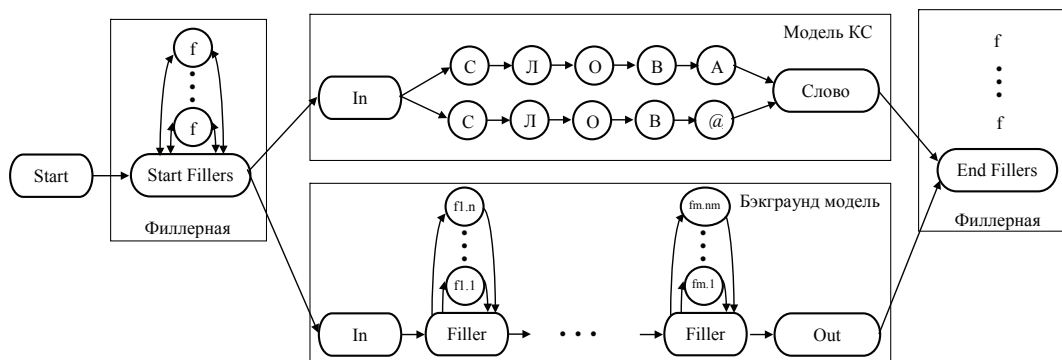


Рис. 3. Пример структуры данных в KWS системе на моделях КС

Ещё один подход к поиску КС основан на распознавании речи в чистом виде. В результате работы системы распознавания речи формируется латтис — направленный связный граф, некоторая сеть слов, содержащая слова кандидаты на распознавание, связи между ними и вероятности переходов. Наилучший вариант пути в графе (имеющий наибольшую вероятность) присутствует в графе, однако графом описывается некоторое количество конкурирующих гипотез. Латтис, в узлах которого находятся слова, называется словным латтисом.



На основании словных латтисов, формируемых системой распознавания с большим словарём, выполняется индексация звукового массива. Далее поиск КС и фраз выполняется по полученным словным латтисам. Достоинством такой системы является высокая скорость поиска КС в индексированных звуковых данных. Проблема такой системы — в принципиальной невозможности нахождения слова, отсутствующего в словаре системы распознавания. Кроме того, такие системы поиска КС оказываются сильно завязанными на качество работы систем распознавания.

Альтернатива словного латтиса — фонемный латтис, в узлах которого находятся не слова, а отдельные звуки речи. Система распознавания аналогично словному латтису строит фонемный латтис, по которому и выполняется поиск КС.

Преимуществом фонемного латтиса является возможность поиска любых КС, так как в системе распознавания не используется словарь и отсутствует привязка к словам. Сложность такой системы заключается в том, что фонемный латтис оказывается широким, качество фонемного распознавания — низким, а пространство поиска — велико.

Основные элементы разработанной системы

В результате анализа достоинств и недостатков различных принципов построения KWS систем было решено создавать систему на основе моделей КС. Основу разрабатываемой системы поиска ключевых слов составляют следующие модули:

- база НММ-моделей звуков речи;
- автоматический транскриптор ключевых слов;
- звуковой препроцессор, выполняющий предварительную обработку звукозаписей и преобразование звука в параметры;
- формирователь альтернативных моделей (моделей усреднённого речевого потока и моделей заполнения);
- декодер параметризованного звука.

Система работает по следующему алгоритму:

- транскриптор формирует возможные варианты произнесения искомым ключевых слов и фраз;
- для всех ключевых слов формируются альтернативные модели;
- полученные структуры данных объединяются в общую сеть поиска — рабочую структуру системы распознавания;
- звуковой поток обрабатывается препроцессором и переводится в пространство признаков;
- звуковой поток разделяется на окна. Каждое окно подаётся на вход декодера. Длины окон и параметры перекрытия определяются исходя из звукового состава искомым слов;
- декодер анализирует параметризованный речевой поток и принимает решение о наличии или отсутствии ключевых слов. Декодирование выполняется с помощью модифицированного алгоритма пересылки маркера [2];
- получаемые результаты распознавания привязываются к звуковому потоку и сохраняются в специальном индексном файле.

Оптимизация структур данных

Для увеличения скорости поиска возможна оптимизация рабочих структур данных распознавателя [3]. Оптимизация заключается в объединении одинаковых цепочек фонов, в рамках вариантов различных транскрипций КС.

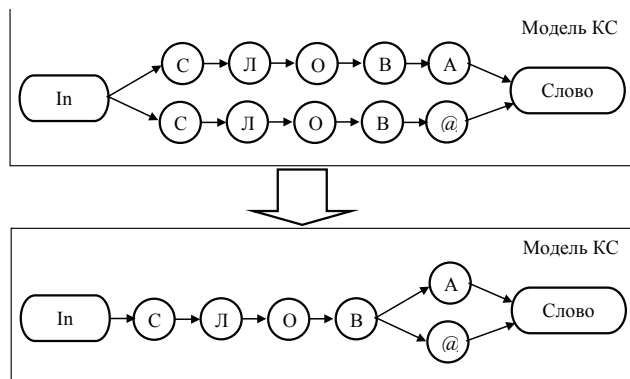


Рис. 4. Пример оптимизации модели КС

Оптимизация может быть продемонстрирована на примере структуры данных, представленной на рис. 3. В результате объединения одинаковых цепочек фонов модель КС примет вид, как показано на рис. 4. Может показаться, что с учётом общего количества звуковых моделей, используемых в моделях заполнения и усреднённого речевого потока, получаемая оптимизация не велика. На самом деле, она позволяет сократить от 5 до 20% всех звуковых моделей. Всё зависит от сложности общей структуры и вариативности произношения искомого КС, его длины и звукового состава. Чем длиннее КС и чем выше возможная вариативность произнесений, тем больше оптимизация.

Критерии оценки качества системы

Одна из важнейших характеристик системы поиска ключевых слов — точность. Под точностью понимается пара значений: DR (Detection Rate) и FA (False Alarm).

Значение DR определяет процент правильно обнаруженных слов и рассчитывается по формуле:

$$DR = 100 * N_{\text{found}} / N_{\text{all}}, \text{ где:}$$

N_{found} — количество правильно найденных реализаций КС в тестовых данных;

N_{all} — общее количество реализаций КС в тестовых данных.

Значение FA определяет количество ложных срабатываний в час и рассчитывается по формуле:

$$FA = (A_{\text{all}} - N_{\text{found}}) / \text{Hrs}, \text{ где:}$$

A_{all} — общее количество всех найденных слов в тестовых данных;

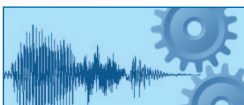
Hrs — длительность звучания тестовых данных в часах.

Результаты тестирования

Тестирование системы проводилось на звуковых файлах общей длительностью звучания чуть больше часа (1 час 2 минуты 37 секунд). Звонки были выполнены с городских телефонных аппаратов. В тестировании приняли участие десять дикторов (шесть мужчин и четыре женщины).

В результате тестирования различных режимов построения моделей заполнения и моделей усреднённой речи были получены значения точности работы системы, представленные в таблице. Тестировались следующие режимы:

«**Тюнингованные**» — в данном режиме параметры поиска и альтернативные модели подбираются вручную для каждого КС на множестве обучающих звуковых данных;



«**Смарт-авто**» — система автоматически формирует параметры поиска и альтернативные модели для каждого слова исходя из их фонетического состава КС и известных параметров фонем;

«**Стандарт**» — используется фиксированный набор параметров поиска и альтернативных моделей;

«**Короткий**» — режим, аналогичный режиму «Стандарт», но с меньшей вариативностью моделей заполнения.

Показатели качества работы системы

Ключевое слово	Тюнингованные		Смарт-авто		Стандарт		Короткий	
	DR	FA2	DR	FA2	DR	FA2	DR	FA2
Кодовое слово	100,00	38,49	100,00	35,00	100,00	400,97	96,00	1350,48
Кредит	96,00	170,00	70,00	32,00	93,00	362,00	96,00	3570,27
Задолженность	100,00	48,11	83,00	6,41	100,00	131,52	83,00	1376,14
Номер карты	83,00	38,49	63,00	0,00	96,00	109,06	76,00	1661,63
Конфиденциально	95,00	79,57	61,00	19,66	88,00	230,67	74,00	982,59
Вакансии	95,00	84,04	77,00	4,47	91,00	167,00	84,00	1957,15
	94,83	76,45	75,67	16,26	94,67	233,54	84,83	1816,38

Видно, что показатели качества работы системы во многом зависят от режима работы системы и от самого искомого КС. Способы построения альтернативных моделей требуют дальнейших исследований и разработки.

Литература

1. Смирнов В., Ермилов С. Технология распознавания речи на службе корпоративных интересов // «Директор по безопасности». 2010. № 11. С. 27–37.
2. Young S.J., Russell N.H., Thornton J.H.S. Token Passing: a Conceptual Model for Connected Speech Recognition Systems // CUED Technical Report F INFENG/TR38, Cambridge University, 1989. (ftp from svr-ftp.eng.cam.ac.uk)
3. Гусев М.Н., Дегтярев В.М. Увеличение производительности системы распознавания речи. «Вопросы радиоэлектроники» (Серия Общетеchnическая), 2010. Вып. 2. С. 115–126.

Сведения об авторах

Дегтярёв Владимир Михайлович — заведующий кафедрой Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, д. т. н., профессор, академик Международной академии информатизации, секция «Аудио-, видеоинформации».

Гусев Михаил Николаевич — Главный инженер-программист ООО «Вокатив», к. т. н. Научные интересы — обработка речевых сигналов и распознавание речи.

Публикации

1. Дегтярёв В.М., Гусев М.Н. Развитие систем распознавания речи // Труды СПбГТУ «Вычислительная техника, автоматика, радиоэлектроника»/ СПбГТУ. СПб, 2006. № 499. С. 119–124.

2. *Valentin Smirnov, Mikhail Gusev.* Objective method of speech signal quality estimation // Proceedings of the 11-th International Conference «Speech and Computer» SPECOM'2006. St.Petersburg, Anatolya Publishers, 2006. Pp. 242–244.
3. *Гусев М.Н., Смирнов В.А., Дегтярев В.М.* Компьютерная статистическая модель русского языка // Труды учебных заведений связи / СПбГУТ. СПб, 2006. № 174. С. 129–135.
4. Патент РФ № 2296377. Способ анализа и синтеза речи. Гусев М.Н., Дегтярёв В.М., Ситников В.В. Официальный Бюллетень Федеральной службы по интеллектуальной собственности, патентам и товарным знакам. Изобретения. Полезные модели, 27.03.2007, № 9(2), 2007.
5. Патент РФ № 2312405. Способ осуществления машинной оценки качества звуковых сигналов, Гусев М.Н., Дегтярёв В.М., Жарков И.В. Официальный Бюллетень Федеральной службы по интеллектуальной собственности, патентам и товарным знакам. Изобретения. Полезные модели, 10.12.2007, № 34(2), 2007.
6. *Гусев М.Н., Дегтярёв В.М., Семёнов Н.Н.* Оптимизация системы распознавания речи с учётом особенностей артикуляции. // Труды учебных заведений связи / СПбГУТ. СПб, 2007. № 177. С. 20–24.
7. *Bolotova Olga, Gusev Michael, Smirnov Valentin.* Speech Recognition System for the Russian Speech // Proceedings of the 12-th International Conference «Speech and Computer» SPECOM'2007. Moscow, 2007. V.II. Pp.475–480.
8. *Гусев М.Н., Дегтярев В.М.* Расчёт и измерение качества речевых сигналов. Санкт-Петербург: «Геликон Плюс», 2008. 276 с.
9. *Смирнов В., Ермилов С.* Технология распознавания речи на службе корпоративных интересов // «Директор по безопасности», 2010. № 11. С. 27–37.
10. *Гусев М.Н., Дегтярёв В.М.* Моделирование длительности звуков в системе распознавания речи. «Вопросы радиоэлектроники» (Серия Общетеchnическая), 2010. Вып. 2. С. 106–115.



Применение синтаксического анализа в задаче распознавания речи

Зулкарнеев М.Ю., кандидат физико-математических наук.

Репалов С.А., кандидат физико-математических наук.

Шамраев Н.Г., аспирант.

Статья посвящена методам повышения точности распознавания речи, основанным на применении синтаксического анализа предложения. Рассмотрены модификации классической N-граммной языковой модели с помощью синтаксической модели предложения, которая строится на основе вероятностных грамматик Хомского. Рассмотрены различные варианты применения алгоритма Коке-Янгера-Касами для синтаксического разбора предложения. Показано преимущество построенной языковой модели перед классической N-граммной моделью.

• *распознавание речи* • *синтаксический анализ* • *алгоритм Коке-Янгера-Касами.*

The article is devoted to methods of speech recognition accuracy improvements, based on the use of parsing sentences. Modification of the classical N — gram language model using syntactic model, based on probabilistic Chomsky grammar, is presented. The different versions of the Cocke-Younger-Kasami algorithm application for sentence parsing are discussed. The advantage of the proposed language model over classical N-gram model is shown.

• *speech recognition* • *syntactical analysis* • *Cocke-Younger-Kasami algorithm.*

Введение

В настоящее время в распознавании речи широко распространён метод, использующий скрытые марковские модели и n-граммные модели языка [1]. Хотя использование этого метода позволило достичь определённых успехов при создании систем распознавания речи, используемые модели обладают многими недостатками. В частности, n-граммная модель языка способна описывать зависимости между словами, расстояние между которыми не превышает некоторого числа.

В данной работе исследуется возможность моделирования языка при помощи синтаксических моделей, представленных в виде вероятностных контекстно-свободных грамматик Хомского (PCFG — probabilistic context-free grammar) и, в частности, использование этих моделей в задаче распознавания речи. Для этого предлагаются и исследуются два метода, позволяющих использовать синтаксическую модель языка, представленную в виде PCFG, в задаче распознавания речи, а также приводятся результаты экспериментов, показывающих преимущество предлагаемого подхода по сравнению с подходом, основанным на n -граммных моделях языка.

Описание алгоритма Коке-Янгера-Касами

Классический алгоритм Коке-Янгера-Касами (СКУ) предназначен для обработки линейных последовательностей слов. Впервые он приводится в работе Янгера [2]. В статье мы рассмотрим реализацию алгоритма для контекстно-свободных грамматик в нормальной форме, т.е. когда каждое правило имеет либо вид $A \rightarrow BC$, либо $A \rightarrow \beta$, где A, B, C — нетерминальные символы, β — терминальный символ.

Задача рассматриваемого алгоритма состоит в ответе на вопрос: может ли данное предложение W быть сгенерировано грамматикой Хомского, заданной фиксированным набором таких правил.

Пусть $\Psi_{ij}(A)$ обозначает некоторую строку α , генерируемую правилом A , начинающуюся с i -го слова и заканчивающуюся на j -м слове в исходной последовательности слов W . $\varphi_{ij}(A)$ — функция стоимости вывода этой последовательности правилом A . Для задачи синтаксического анализа предложения в качестве функции стоимости выступает логарифм вероятности правила перехода нетерминальных классов в терминалы (слова).

В классическом алгоритме начальные значения функций устанавливаются для каждого слова по следующим формулам:

$$\varphi_{ii}(A) = \min_{\{A \rightarrow \alpha\}} \{C[\alpha|t_{i-1,i}]\},$$

$$\psi_{ii}(A) = \alpha = \arg \min_{\{A \rightarrow \alpha\}} \{\varphi_{ii}(A)\}, \text{ для } 1 \leq i \leq |W|.$$

Все остальные значения Ψ и φ не определены в этот момент. Далее уже для пар $1 \leq i < j \leq |W|$ последовательно вычисляются

$$\varphi_{ij}(A) = \min_{\{A \rightarrow BC\}} \{\min_{i \leq l < j} \{\varphi_{il}(B) \cdot \varphi_{l+1,j}(C)\}\}, \quad (1)$$

$$\psi_{ij}(A) = \psi_{il}(B) \otimes \psi_{l+1,j}(C). \quad (2)$$

Т.е. идёт перебор по l -индексу слова-разделителя и всевозможным правилам B и C , таким, что $A \rightarrow BC$. Вычисление начинается с пар, таких, что $|i - j| = 1$, затем для $|i - j| = 2$ и так далее по возрастанию до $|i - j| = |W| - 1$.

Если правило A генерирует подстроку α_{ij} , то для него найдётся тройка оптимальных объектов $(\bar{l}, \bar{B}, \bar{C}) = \arg \min_{\substack{\{A \rightarrow BC\}, \\ i \leq l < j}} \{\varphi_{ij}(A)\}$ — индекс слова разделителя l , и два правила B и C . Если предложение W может быть сгенерировано из стартового символа S , т.е. $S \Rightarrow W$, то алгоритм восстановления бинарного дерева будет следующим.

Рассмотрим тройку для стартового слова $(l, B, C) = \psi_{1N}(S)$, где $N = |W|d$. Используя это разбиение, перейдём к левому поддереву $(l_1, B_1, C_1) = \psi_{l_1}(B)$ и к правому поддереву $(l_2, B_2, C_2) = \psi_{l_2+N}(C)d$.



Далее рекурсивно разбираем каждую тройку, пока не получим последовательность терминальных символов.

Подход на основе линейной последовательности слов

Первый подход к использованию синтаксического анализа для распознавания речи заключается в использовании алгоритма СКУ для линейной последовательности псевдослов, которые получаются на основе решёток слов, генерируемых декодером.

Результатом работы декодера является решётка слов, при этом каждому слову сопоставлены два значения — акустическая и лингвистическая вероятности. Для оценки лингвистической вероятности при построении решётки используется 3-граммная модель языка. Пример решётки приведён на рис. 1.

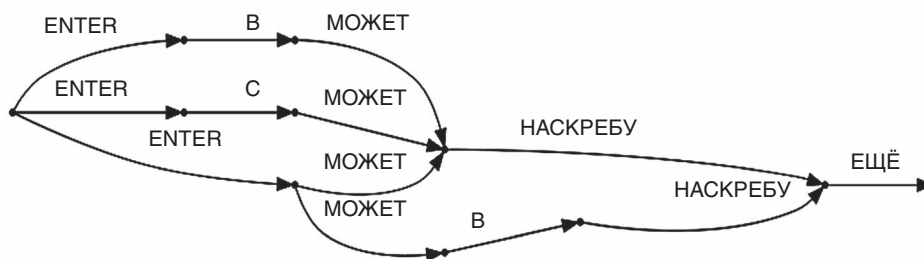


Рис. 1. Пример решётки слов

При помощи алгоритма кластеризации [3] эта решётка преобразуется в линейную сеть (Word Confusion Network, WCN), которая является компактным представлением множества гипотез. Представление гипотез в виде всех возможных предложений и последующий их последовательный анализ не эффективен с точки зрения быстродействия. Если множество гипотез будет слишком маленьким, снижается вероятность найти правильную гипотезу, если множество гипотез слишком большое, очень сильно (экспоненциально) возрастает время, затрачиваемое на синтаксический разбор.

Один из предлагаемых здесь способов ускорения работы и прямого использования метода СКУ состоит в возможности представления множества гипотез в виде линейной сети слов. В такой сети слова разбиты на группы, которые упорядочены во времени. Пример такой сети приведён на рис. 2. На нём цифрами обозначены вероятности, символом eps — пропуск слова.

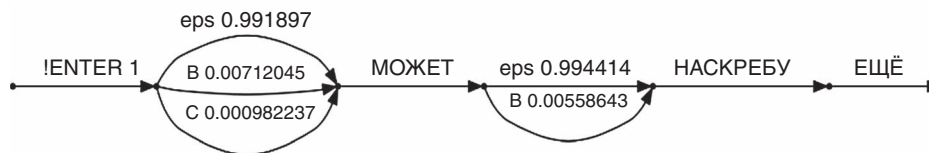


Рис. 2. Пример линейной решётки слов

Предварительная подготовка для обработки гипотез заключается в определении множества правил, записанных в форме контекстно-свободной грамматики Хомского для конкретного языка и вычисление вероятностей всех правил,

нетерминальных и терминальных, входящих в эту грамматику (осуществляется с помощью метода Бейкера для большого массива предложений).

На этапе анализа линейная решётка слов (см. рис. 2) преобразуется в последовательность линейно присоединённых элементов. Каждый элемент соответствует некоторому фрагменту предложения и содержит набор конкурирующих слов. Это позволяет представить решётку как линейную последовательность некоторых формальных символов:

$$W_1, W_2, \dots, W_N \quad (3)$$

Например, формальное слово W_1 может включать в себя слова $word_1, word_2, \dots, word_{k_1}$. Далее все возможные слова, соответствующие W_1 , будут генерироваться некоторым множеством предтерминальных классов. Тогда в общее множество правил добавляются порождающие терминальные правила вида $S_{ij} \rightarrow W_1$. Вероятность каждого правила вычисляется по следующей формуле:

$$\sum_{word_i \in W_1} P_{dic}^{(i)} P_{lat}^{(i)},$$

где $word_i$ — некоторое слово, находящееся в последовательности на первом месте, j — индекс перебора множества предтерминальных классов, генерирующих все слова $word_i$, S_{i1} — соответствующий грамматический (предтерминальный) класс для $word_i$, $P_{dic}^{(i)}$ — вероятность $word_i$ в словаре, $P_{lat}^{(i)}$ — вероятность $word_i$ в линейной решётке. Так как одно и то же слово может генерироваться разными грамматическими классами, то нужно рассматривать все допустимые варианты.

После пополнения таким способом множества правил, для последовательности символов (3) её синтаксический разбор становится возможен. Он осуществляется при помощи вероятностного алгоритма Коке-Янгера-Касами. На выходе получается наиболее вероятное дерево правил, генерирующее последовательность (3). Это дерево представляет собой фактически синтаксический разбор предложения. Таким образом, мы получаем наиболее вероятную последовательность грамматических классов G_1, G_2, \dots, G_N для исходного предложения. Далее, для восстановления оптимальной гипотезы из слов-кандидатов, принадлежащих формальному слову W_i , выбирается слово, обладающее максимальной вероятностью среди всех слов, принадлежащих грамматическому классу G_i .

К сожалению, данный метод синтаксического анализа имеет серьёзные недостатки:

1. Преобразование исходной решётки в линейную последовательность слов приводит к потере части гипотез.
2. Вторым следствием преобразования является фиксированная длина гипотезы предложения. Определённая гибкость сохраняется лишь при переходе псевдослова в «пустой» класс eps, т.е. только для уменьшения фиксированной длины.
3. Фиксация грамматических классов на последнем этапе может привести к гарантированно неоптимальному выбору в случае, если истинное слово принадлежит другому классу.

Недостатки данного метода можно преодолеть, если использовать более гибкий подход, сохраняющий на каждом этапе большую информацию о конкурентных гипотезах. В качестве такого подхода мы предлагаем модификацию метода СКУ для решётки гипотез, рассматриваемую в следующем разделе.

Синтаксический разбор решётки слов

Классический метод Коке-Янгера-Касами предназначен для обработки линейных последовательностей слов. В таком виде он не очень удобен для применения к задаче распознавания речи. Метод прямого перебора гипотез оказывается неэффективным с точки

зрения времени обработки [4]. Метод, основанный на WCN, позволяет эффективно повысить скорость обработки. Однако он является искусственным и обладает рядом недостатков, самый важный из которых — фиксация цепочки грамматических классов и подбор гипотезы в зависимости от этой цепочки. Это делает метод сильно зависящим от ошибок синтаксического разбора.

В работе [3] предлагается обобщение алгоритма Коке-Касами-Янгера, в котором в качестве анализируемых сегментов выступают временные интервалы, а не цепочки слов. Эта модификация формально не меняет расчётные формулы (1), (2), а приводит только к увеличению размеров матрицы $\Psi_{ij}(A)$ (теперь она будет равна $T \times T$, где T — количество дискретных моментов времени).

В данной работе на анализируемые временные интервалы накладывается ограничение в виде решёток слов, которые задают возможные цепочки слов для данного временного интервала. Данному временному интервалу соответствует цепочка слов, если в решётке имеются узлы, соответствующие началу и концу временного интервала, и они связаны переходами. Ограничение приводит к тому, что размер матрицы $\Psi_{ij}(A)$ будет равен количеству узлов в решётке. Результатом работы алгоритма Коке-Касами-Янгера, в этом случае, будет наиболее вероятный, с точки зрения PCFG, путь в решётке. По аналогии с синтаксическим разбором предложения здесь можно говорить о синтаксическом разборе решётки слов.

Рассмотрим работу описанного алгоритма Коке-Касами-Янгера на решётках на примере разбора решётки слов, представленной на рис. 3. Узлы решётки обозначены прямоугольниками и соответствуют времени конца слов, которыми они подписаны. Переходы между узлами фактически соответствуют словам в речевом сообщении. Узлы решётки пронумерованы в соответствии с топологией решётки (номер начального узла любого перехода меньше, чем номер конечного узла этого перехода).



Рис. 3. Решётка слов

В алгоритме используется матрица D , которая хранит информацию о том, какие грамматические классы и с какой вероятностью соответствуют различным частичным путям в решётке. Например, элемент матрицы $D[0,2]$ хранит список грамматических классов, которые могли бы соответствовать цепочке слов «МЕЖДУ ДЕРЕВЬЯМИ».

Работа алгоритма начинается с инициализации элементов матрицы D , которые соответствуют переходам между узлами. В них заносится информация о возможных грамматических классах, которым могут принадлежать слова переходов. Например, элементу матрицы $D[0,1]$ соответствует переход $(0,1)$ (слово «МЕЖДУ»). В этот элемент заносится класс PR (предлог) и класс ADV (наречие) с соответствующими вероятностями.

Далее идёт основная часть алгоритма, которая состоит в следующем. Рассматриваются всевозможные пары узлов по возрастанию разности их номеров, начиная с разности, равной 2, и для каждой такой пары (i,j) рассматриваются всевозможные узлы k , которые могут лежать на частичных путях, соединя-

ющих эти узлы. Для этой пары узлов (i, j) принимается решение: может ли ей соответствовать некоторый грамматический класс, исходя из имеющихся правил и информации, хранящейся в элементах матрицы $D[i, k]$ и $D[k, j]$. Например, при рассмотрении пары узлов $(0, 2)$ принимается решение, что цепочке слов «МЕЖДУ ДЕРЕВЬЯМИ» соответствует грамматический класс *abst* (обстоятельство), поскольку существует правило $abst \rightarrow PR\ predl_mn_sred_tvor_neod$, а в элементах $D[0, 1]$ и $D[1, 2]$ хранится информация о грамматических классах *PR* (предлог) и *predl_mn_sred_tvor_neod* (неодушевлённая предложная фраза множественного числа, среднего рода, творительного падежа) соответственно. Описанная процедура продолжается до тех пор, пока не будет рассмотрена пара самых удалённых узлов $(0, 6)$.

На этом синтаксический разбор решётки завершён, и дальше информация, сохранённая в матрице D , используется для восстановления наиболее вероятной цепочки слов. Также при необходимости информация из D может быть использована для восстановления наиболее вероятного дерева синтаксического разбора.

Полученные результаты

В данной работе алгоритм Коке-Касами-Янгера на решётках слов включён в процесс распознавания, как постобработка решёток слов, которые получены на первом этапе распознавания, направленная на уточнение результатов распознавания. В качестве функции $\Psi_{ij}(A)$ берётся комбинация языковой вероятности, полученной при помощи трёхграммной модели языка и вероятности синтаксических правил. Это позволяет объединять информацию, заложенную в трёхграммной модели и в синтаксической модели и, таким образом, получать более надёжные результаты.

Для проверки предложенных подходов были проведены эксперименты по распознаванию программ на русском языке, записанных из прямого эфира радио «Свобода». В качестве базовой была взята система распознавания русской речи, использующая связанные трифоны и трёхграммную модель языка. Синтаксическая модель русского языка, представленная в виде PCFG, была сгенерирована автоматически на основе синтаксически размеченных предложений Национального корпуса русского языка [5, 6].

Эксперименты показали, что применение метода синтаксического разбора, основанного на предвословах, позволило повысить точность с 72,39 до 75,16%. В основном, повышение точности было достигнуто за счёт согласования падежей прилагательных и существительных, употребления предлогов.

Для проверки подхода, основанного на синтаксическом разборе решёток слов, была создана программная реализация метода Коке-Касами-Янгера синтаксического разбора решёток слов.

Из-за сложности алгоритма Коке-Янгера-Касами, который составляет $O(N^3)$, где N — количество узлов в решётке, реализация не позволила провести полноценные эксперименты для больших решёток. Однако эксперименты для усечённых решёток показали преимущества подхода по сравнению с n -граммными моделями языка. Подход позволяет устанавливать зависимость между значительно удалёнными друг от друга словами в предложении.

Заключение

Использование синтаксического анализа при распознавании речи даёт дополнительный источник информации, который позволяет значительно повысить точность распознавания речи. Обобщение метода Коке-Янгера-Касами на решётки слов — наиболее удобный механизм включения синтаксических правил в систему распознавания речи. Однако его сложность не позволила выполнять анализ полноценных решёток слов.



В качестве дальнейшего развития планируется расширить алгоритм, включив в него дополнительные механизмы отсечки маловероятных гипотез, которые бы значительно ускорили обработку, но при этом не привели к потерям правильных гипотез.

Литература

1. *Rabiner L.R., Juang B.H.* Fundamentals of Speech Recognition. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.
2. *Younger D.* Recognition and parsing of context free languages in time n3. Information and Control. Volume 10. Issue 2. February 1967. Pp. 189–208.
3. *Levenson S.C.* Mathematical models for speech technology. John Wiley & Sons Ltd, NJ, USA, 2005.
4. *Батальщиков А.А., Зулкарнеев М.Ю., Шамраев Н.Г.* Оценка гипотез с использованием синтаксического анализа // Сборник трудов XXII сессии Российского акустического общества и Сессии Научного совета РАН по акустике. Т. 3. М.: ГЕОС, 2010. С. 22–25.
5. Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. 502 с.
6. <http://www.ruscorpora.ru/>

Сведения об авторах

Зулкарнеев Михаил Юрьевич —

старший научный сотрудник ФГНУ НИИ «Спецвузавтоматика», г. Ростов-на-Дону. Область научных интересов — распознавание и анализ устной речи.

Репалов Сергей Анатольевич —

заведующий лабораторией ФГНУ НИИ «Спецвузавтоматика», г. Ростов-на-Дону. Область научных интересов — автоматическая и инструментальная обработка устной речи, цифровая обработка сигналов.

Шамраев Николай Георгиевич —

научный сотрудник ФГНУ «НИИ «Спецвузавтоматика», г. Ростов-на-Дону. Область научных интересов — распознавание и анализ устной речи.

Синтез просодических моделей в языке дари

Иванов В.Б., доктор филологических наук, профессор



Речь на афганском дари имеет систему фонетических отличий от близкородственных языков (персидского и таджикского). В этой системе значительное место принадлежит просодическим моделям — другому типу словесного ударения. Если в персидском языке ударение чисто тоническое, а в таджикском по преимуществу тоническое, то в дари оно по преимуществу квантитативное, т.е. ударный слог, как правило, выделяется большей длительностью (по данным анализа ударный гласный удлиняется примерно на 10%, ударный слог — на 35%). Синтез речи позволил определить количественные границы этого противопоставления (даже 10-процентного удлинения слога достаточно для восприятия его ударным). Контур тона оказался незначим для ударения. Квантитативный компонент ударения может быть нейтрализован 25-процентным приростом тона и/или интенсивности в безударном слоге.

• *Дари* • *иранские языки* • *персидский* • *таджикский* • *фарси* • *синтез* • *просодика* • *ударение*.

Afghan Dari is characterized by a system of phonetic differences from closely related languages Persian and Tajik. In this system significant role is played by prosody and different type of word stress. Word stress in Persian is purely tonic, in Tajik – mainly tonic, but in Dari it is principally quantitative, i.e. stressed syllable is longer, as a rule (analysis shows us a 10% lengthening of stressed vowel and a 35% lengthening of stressed syllable). Speech synthesis allowed us to distinguish the quantitative boundaries of that contrast (even a 10% lengthening of a syllable is sufficient to be perceived as stressed). Tone pattern appeared to be non-significant for the stress. Quantitative component of the word stress can be neutralized by a 25% tone and/or intensity rise in the unstressed syllable.

• *Dari* • *Iranian languages* • *Persian* • *Tajik* • *Farsi* • *synthesis* • *prosody* • *stress* • *accent*.

Введение

Иранская группа языков выделилась из индоиранской ветви индоевропейской семьи на рубеже III–II тыс. до н.э. Наиболее древним представителем этой группы является язык Авесты — религиозного памятника последователей пророка Заратуштры (середина II тыс. до н.э.). Среди других мёртвых языков иранской группы можно отметить скифский (IX–IV вв. до н.э.), мидийский (IX–VI вв. до н.э.) и древнеперсидский языки (VI–IV вв. до н.э.). Последний относится к юго-западной ветви иранских языков. Его потомками были среднеперсидский (III в. до н.э. — VII в. н.э.) и новоперсидский языки (IX–XV вв. н.э.). Современные персидский, таджикский и дари происходят от новопер-

сидского языка (рис.1). Персидский, называемый также языком фарси́, является государственным языком Ирана, таджикский — Таджикистана, дари — один из двух государственных языков Афганистана.

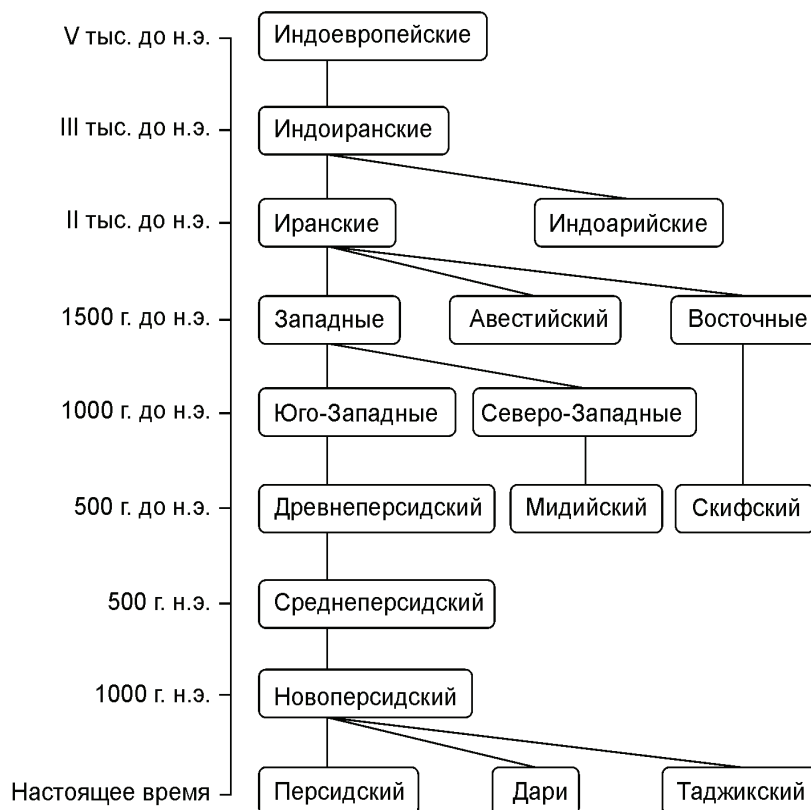


Рис. 1. Происхождение языка дари

Подавляющая часть населения Афганистана, численность которого оценивается в 32 млн. человек, как минимум двуязычна: владеет языками дари и пашто (пушту). В некоторых областях Афганистана, как, например, в Бадахшане, граничащем с таджикским Памиром, люди не говорят на пашто, но помимо дари говорят ещё на одном-двух языках (например, на шугнанском и ваханском). На языке дари, таким образом, можно общаться по всей территории Афганистана и даже за её пределами: в Иране, Таджикистане, в Узбекистане, в Пакистане, в Индии и Китае.

Само официальное название «язык дари» по отношению к современному языку Афганистана появилось исторически недавно, в афганской конституции 1965 года. Ранее этот термин относили к новоперсидскому языку. Поэтому до сих пор в повседневном общении носители языка называют его фарси́, т.е. персидский. Носители трёх близкородственных языков (персидского, таджикского и дари) прекрасно общаются между собой без переводчика. Иранская Академия языка и литературы по этой причине отказывает таджикскому и дари в статусе отдельных языков, считая их диалектами персидского. В российском и мировом востоковедении их считают различными языками, поскольку они требуют отдельного преподавания и обозначены как государственные языки в конституциях соответствующих стран.

Отличия между персидским, таджикским и дари отмечаются на всех уровнях: в фонетике, в морфологии, в лексике, во фразеологии.

Для дари на фонетическом уровне характерно:

- а) восьмифонемный вокализм с долгими [ē], [ō] против шестифонемного в персидском и таджикском языках;
- б) квазидифтонги [ay], [aw] против [ey], [ow] в персидском и [ay], [av] в таджикском;
- в) различие согласных фонем [q], [ɣ], которые также сохранились в таджикском, но объединились в одну фонему [q] в персидском;
- г) наличие краткого [a] в исходе слова, которое сохранилось в таджикском, но перешло в [e] в персидском;
- д) опущение согласного [h], в то время как в таджикском и персидском он сохраняется.

Есть и некоторые другие менее регулярные, но на слух хорошо воспринимаемые отличия языка дари от близкородственных. Вариативности вокализма немало способствовала арабская письменность, которая не фиксирует на письме краткие гласные. Для таджикского и дари характерно «аканье», т.е. вероятность появления в речи краткого [a] очень высока. Для персидского языка характерно «эканье».

Уже по одной формуле приветствия можно весьма точно указать место, из которого прибыл носитель языка. Для определения места проживания можно также попросить его посчитать до восьми (разговорные формы числительных приведены в таблице 1).

Таблица 1

Ряды числительных в персидском, таджикском и дари

	Дари	Персидский	Таджикский
1	yak	yek/ye	yak
2	du	do	du
3	se	se	se
4	čār	čār	čor
5	pañj	pañj	pañj
6	šaš	šeš	šaš
7	aft	haft	haft
8	ašt	hašt	hašt

Структура предложения во всех трёх близкородственных языках относится к категории SOV (субъект — объект — сказуемое). Языки хорошо различимы по интонационному рисунку. На слух европейца персидский язык наиболее певуч. В персидском повествовательном предложении на группе подлежащего происходит самое значительное повышение тона (особенно, в исфаганском диалекте), далее следует ощутимая пауза и на группе сказуемого тон идёт вниз. Глагол, находящийся в самом конце предложения, как правило, самостоятельного синтагматического (иногда словесного) ударения не принимает, т.е. произносится как клитика. В языке дари описанная интонация существует, но она не так ярко выражена. Таджикская повествовательная интонация представляется довольно ровной. Описанная повествовательная интонация в общем случае находится за пределами системы интонационных конструкций Е.А. Брызгуновой [5], так как различны типичные синтаксические структуры славянских и иранских языков (SVO vs. SOV). Данная интонация, особенно в её иранском варианте, близка к тюркской (азербайджанской, турецкой), где структура предложения также SOV. Возможно, она будет противопоставлена и более обширным инвентарям просодических признаков, выделяемых на основе русского языка [14, с. 14], что может быть использовано для автоматического различения языков типа SVO vs. SOV и требует отдельного исследования.



В персидском языке есть уникальная интонация вопросительного предложения без вопросительного слова (переспрос), в которой тон, плавно повышаясь к концу предложения, на последнем слоге падает и вновь повышается до максимума и удерживается на этом уровне. Тонировка последнего слога переспроса сильно напоминает 3-й тон в китайском языке. Эта интонация также отсутствует в известных нам системах интонационных конструкций русского языка. В таджикском и дари этой интонации нет, там тон переспроса более ровный, весьма схожий с русской интонацией.

В среде ираноязычных народностей из поколения в поколение передаётся техника чтения классических стихов, основанная на количественной системе противопоставления слогов (арузе) и уходящая своими корнями в первое тысячелетие н.э. Если попросить носителей прочесть отрывки из Омара Хайяма или Хафиза, то по интонационному рисунку будет практически невозможно определить их языковую принадлежность. Межъязыковые различия в стихотворной речи нивелируются, и для определения языка на первый план выдвигаются упомянутые ранее сегментные особенности. Для распознавания языка и диалекта говорящего более показательна разговорная речь, а литературная и архаичная фразеология, наоборот, затушёвывает его. Поэтому стихотворную речь можно попробовать использовать для различения групп языков, например, славянские vs. иранские.

Данная статья является продолжением исследования, результаты которого можно суммировать следующим образом: три информанта (носители языка дари) произносили различные по лексическому значению, но одинаковые по фонемному составу слова, отличавшиеся лишь местом ударения. Анализ звукозаписей показал, что соотношение длительностей слогов — наиболее информативный критерий места ударения. Тон также маркировал ударный слог, но мог быть нейтрализован вопросительной интонацией. Интенсивность не играла статистически существенной роли. В результате ударение в дари было определено как количественное [12].

Сопоставляя результаты этого исследования с данными других языков, отметим, что в близкородственных иранских языках — в персидском и таджикском — ударение оказалось тоническим. Причём в персидском языке оно чисто тоническое — частота основного тона (F_0) ударного слога всегда выше, чем безударного [8], а в таджикском — оно по преимуществу тоническое, т.е. F_0 в ударном слоге, как правило, выше, чем в безударном [18]. В языке пашто (другом государственном языке Афганистана, принадлежащем юго-восточной ветви иранской группы) ударение оказалось количественным [16]. Как известно, в русском языке ударение количественное [7], а в английском — тоническое [4, р. 95].

Опыт показывает, что аналитические и теоретические исследования не могут охватить всех возможных сторон явления, и даже элементарные попытки его моделирования существенно уточняют и пересматривают наше знание о нём. Обычно явление полагается хорошо изученным, если его удаётся воссоздать и управлять его развитием. Для моделирования просодической структуры слова в языке дари была проведена серия экспериментов, в которых предпринималась попытка синтезировать различные акцентные структуры для проверки гипотезы о количественном ударении в дари и выявления пределов вариации основных просодических параметров.

В первом эксперименте рассматривалось противопоставление двух синтезированных контрастирующих акцентных структур, отмечаемых в ряде

иранских языков. По наиболее типичным реализациям слов *راحت* *rāhat* *твой путь* vs. *راحت* *rāhat* *спокойный* были выбраны параметры синтеза: временные значения, формантные частоты, амплитудные характеристики и тон. По ним был синтезирован исходный стимул (рис. 2).

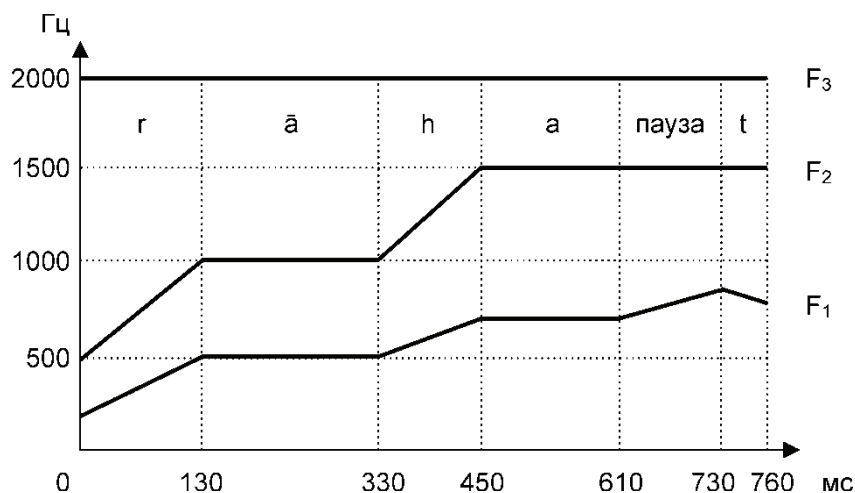


Рис. 2. Формантно-временная структура синтезируемого слова *راحت* *rāhat* *спокойный*.

Жирными линиями обозначено движение формант F₁, F₂, F₃.

Вертикальными пунктирными линиями обозначены границы сегментов

Синтез проводился с помощью программного синтезатора Д. Клатта [2] на компьютере PDP-9 в Массачусетском технологическом институте (Бостон, США) под руководством профессора К.Н. Стивенса. Синтезатор был настроен на моделирование английской речи, поэтому мы и без дополнительной настройки получили удовлетворительные звуки дари, сходные с соответствующими английскими¹. Естественно звучал глайд [h], который близок к соответствующему английскому звуку. Конечный [t] был получен из английского отключением придыхания. Афганцы сочли в дальнейшем звучание почти всех компонентов стимулов весьма естественным. Синтезатор Д. Клатта удобен для экспериментатора возможностью автоматического вычисления амплитуд формант по их частоте и ширине. Причём нам не было необходимости определять ширину формант: вполне приемлемые значения по умолчанию были заданы Д. Клаттом при написании программы. Практически для большинства звуков речи можно ограничиться описанием формантных частот.

Определённые проблемы возникали при моделировании раскатистого [r], которое отсутствует в английском, и в лаборатории не было достаточного опыта для его синтеза. Аудиторы, в целом, правильно распознавали его, но отмечали некоторый церебральный оттенок. Чтобы дать возможность аудитору привыкнуть к звучанию такого [r] и отвлечь его внимание от машинного «акцента», стимулы неоднократно воспроизводились перед началом аудирования. Перцептивная теория говорит о том, что если восприятие идёт в речевом режиме, то, несмотря на помехи, каковыми можно считать машинный «акцент», стимулы доступны лингвистической обработке [6, глава 1]. Во всяком случае, неточности конструирования [r] на сегментном уровне не влияли на общую картину просодики и интонации всего слова в целом.

Основой для конструирования континуума были наблюдения над естественной речью. Усреднённая по трём измерениям темпоральная структура реализаций, на основе которых строились серии стимулов, приведена на рис. 3. Длительность слова *راحت* *rāhat*

¹ В этом случае был проведён так называемый синтез с нуля. То есть в синтезирующую программу вводился ряд параметров в цифровом и графическом виде вручную, на основании прошлого опыта.



составила 707 мс, راحت *rāhāt* — 618 мс. Длительность каждого звука речи представлена в процентах ко всей длительности слова. Перенос ударения на первый слог удлинняет как его слогоноситель (25% → 34%), так и весь слог в целом (18% + 25% → 25% + 34%).

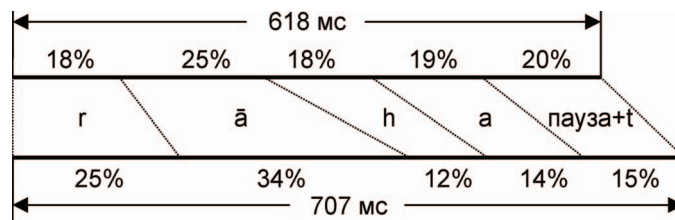


Рис. 3. Временные пропорции реализаций слов *rāhāt* (внизу) vs. *rāhāt* (вверху). Пунктирные линии обозначают границы сегментов

Данный эксперимент планировался тогда, когда по более ранним исследованиям уже было известно, что в родственных языках — персидском и таджикском — ударение тоническое или по преимуществу таковое [8; 18], а ударение в дари ещё не было обследовано экспериментально. Поэтому, прежде всего, проверялась гипотеза о релевантности тона. Для этого в исходном стимуле при постоянных темпоральных и формантных структурах ступенчато варьировалась частота основного тона. Таким путём было получено 12 производных стимулов с различным соотношением тонов, тональная структура первых пяти показана на рис. 4.

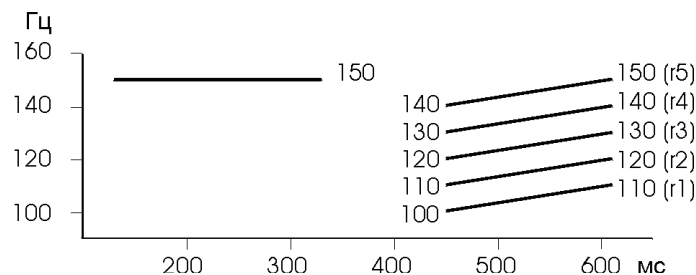


Рис. 4. Тональная структура синтезированных стимулов *r1* — *r5* на основе слова *rāhāt*

В стимулах *r1* — *r5* первые слоги одинаковы по всем параметрам. F_0 на всём их протяжении равно 150 Гц. Тон вторых слогов в каждом стимуле имеет свою собственную начальную частоту, но везде возрастает на 10 Гц.

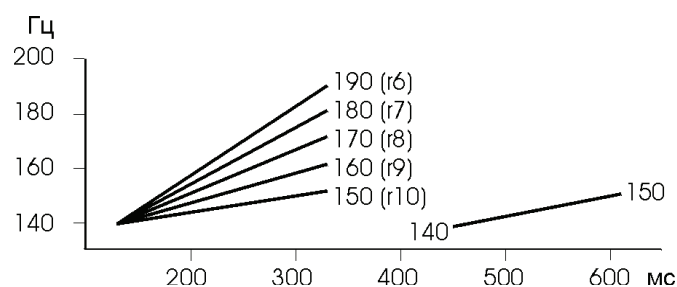


Рис. 5. Тональная структура стимулов *r6* — *r10*

Стимулы r6 — r10 различаются тоном первого слога, который повышается от 140 Гц до указанной на рис. 5 величины. Вторые слоги у них одинаковы по всем параметрам и тождественны второму слогу стимула r5.

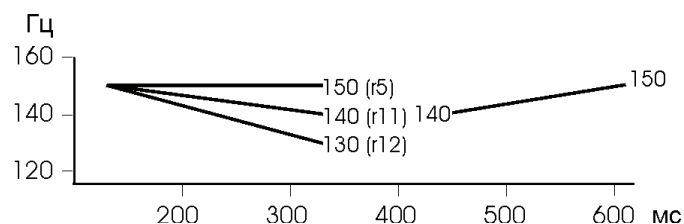


Рис. 6. Тональная структура стимулов r5, r11, r12

Стимулы r5, r11, r12 имеют идентичные вторые слоги (рис. 6) и различаются лишь тоном первого слога.

Стимулы r1 — r12 были скомпонованы в четыре аудиотеста, в каждом из которых они повторялись случайным образом пять раз. После каждого стимула была вставлена трёхсекундная пауза. После каждого десятого стимула добавлялся ещё один трёхсекундный интервал. Итого, после каждой группы из десяти стимулов пауза длилась шесть секунд. С точки зрения интонации аудиотесты интерпретировались как слова, начитанные перечислительной интонацией.

Тест 1 проводился на стимулах r1 — r5. В нём участвовало четыре информанта (все аудиторы после окончания лица в Кабуле были студентами подготовительного факультета МГУ; аудитор А3 — девушка). Перед прослушиванием идентификационного теста аудиторам ставилась задача отметить те стимулы, которые они воспринимают как слово راحت *rāhāt* *спокойный* (т.е. с ударением на последнем слоге). Число утвердительных ответов по каждому стимулу приведено в таблице 2. Дисперсионный анализ тестирования дал неоднозначные результаты.

Таблица 2

Восприятие финального ударения в стимулах r1 — r5

Стимулы/ Аудиторы	r1	r2	r3	r4	r5
A1 (С.З.)	0	0	0	0	2
A2 (Ф.К.)	3	2	1	0	5
A3 (С.Б.)	2	2	3	3	5
A4 (А.М.)	4	4	5	3	0
Итого	9	8	9	6	12

Первые три информанта А1 — А3 проявили высокую чувствительность к изменениям в просодике в зависимости от высоты тона ($p < 0,02$). Различия между аудиторами были высокозначимыми ($p < 0,0001$). Все информанты улавливали перемещение ударения на последний слог по мере повышения в нём тона (F_0), но пороговые значения F_0 , при которых ощущалось это перемещение, для каждого из аудиторов было различным.

Реже других финальное ударение «слышал» информант А1: стимулы r1 — r4, по его мнению, всегда обладали начальным ударением и только r5 в двух случаях создал эффект финального ударения. Чаще других финальную акцентуацию отмечала А3. Стимул r5 всегда относился ей к этой категории, а остальные стимулы — в 40% — 60% случаев. Мы не можем связать различия в разбросе ответов аудиторов с ка-



кими-либо образовательными или диалектными категориями, поэтому с точки зрения эксперимента считаем их случайными или относим к индивидуальным особенностям. Для этой группы аудиторов стимул r5 оказался существенно противопоставленным остальным. В нём оба слога были одинаково тонированы. Понижение тона (F_0) во втором слоге на 6,7%, как это имело место в стимуле r4, приводило к сдвигу ударения на первый слог. Дальнейшее понижение тона (как в стимулах r1 — r3) не влияло на опознавание просодики. В этой части исследования всё указывало на то, что ударение в языке дари связано с высотой тона слога (F_0).

Против ожидания, привлечение к эксперименту информанта А4 не подтвердило этот вывод.

С учётом показателей информанта А4 влияние тона (F_0) на просодику потеряло статистическую значимость². Причём помимо количественных различий обнаружались различия качественные. Вопреки мнению аудиторов А1 — А3 (рис. 7) информант А4 ощущал перемещение ударения на последний слог по мере падения тона в нём, о чём свидетельствует значимое взаимодействие факторов ($p < 0,01$; первый фактор — изменения в тоне, второй — различия в аудиторах).

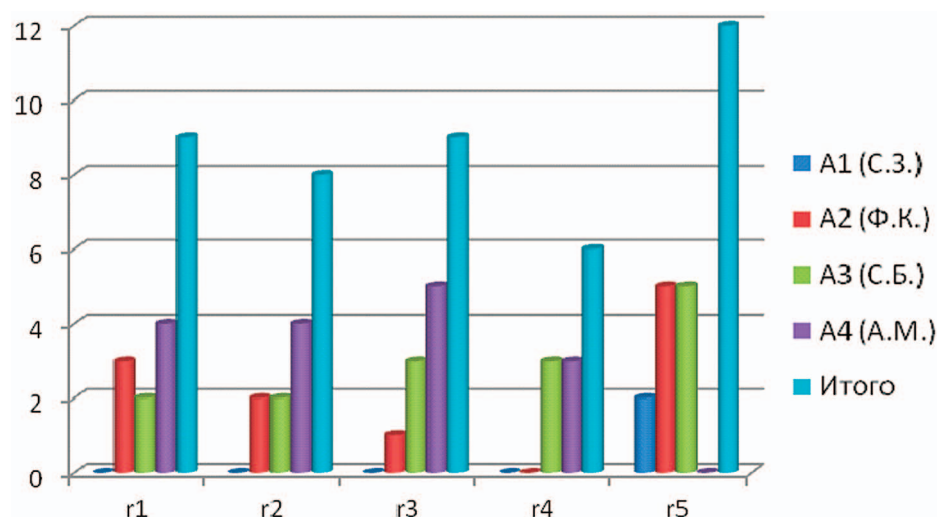


Рис. 7. Восприятие финального ударения в стимулах r1 — r5 аудиторами А1 — А4

Схожие тенденции были обнаружены в **Тесте 2**, где изучалось восприятие стимулов r6 — r10. В них ступенчато менялся тон (F_0) первого слога при постоянстве во втором. Только один аудитор А1 закономерно связывал место ударения с тоном (см. таблица 3 и рис. 8). Разброс суждений остальных, однако, полностью затушевали эту закономерность, сделав её меньше порога значимости ($p > 0,15$). Восприятие стимулов r9 и r10 аналогично стимулам r4 и r5 в **тесте 1** оказалось статистически противопоставленным восприятию остальных стимулов, где тоны имели больший контраст.

² N.B. — редкий случай! Обычно дополнительных информантов привлекают, чтобы усилить статистическую значимость выводов, если «немного не хватает» до критического уровня (как правило, $p = 0,05$).

Таблица 3

Восприятие финального ударения в стимулах r6 – r10

Стимулы/ Аудиторы	r6	r7	r8	r9	r10
A1 (С.З.)	0	0	0	2	5
A2 (Ф.К.)	1	3	2	2	4
A3 (С.Б.)	1	1	3	5	4
A4 (А.М.)	2	2	1	1	1
Итого	4	6	6	10	14

В **Тест 3** были сведены стимулы r5, r10, r11, r12, которые различаются не максимальной высотой тона (F_0), а только его контуром. Здесь проверялась гипотеза о влиянии контура на ударение. Были использованы общие, часто встречающиеся в лингвистических описаниях типы контуров: ровный, восходящий, нисходящий. Поскольку тестирование не обнаружило статистической связи этих контуров с ударением, другие, более сложные типы контуров, не рассматривались.

Тест 4 включал стимулы r14 и r15. Тон второго слога в них в отличие от всех других стимулов был выше первого: в r14 — на 10 Гц, в r15 — на 20 Гц. Половина аудиторов услышала в них финальное ударение, другая половина — начальное. Причём внутри этих подгрупп оценки информантов были достаточно хорошо согласованы, что говорит о неслучайности суждений. Таким образом, и этот тест не подтвердил гипотезу о релевантности тона (F_0) для ударения дари.

На основании результатов **Эксперимента 1** можно сформулировать следующую гипотезу. В языке дари соотношение частот основного тона начального и финального слогов так же, как и контур тона в них, с лингвистической точки зрения несущественны для распознавания просодики ударения. С психологической точки зрения некоторые носители ощущают связь этих факторов с просодикой, но эта связь очень субъективна. Одни слышат ударение на финальном слоге, если тон в нём выше. Другие воспринимают ударение на том же слоге, если тон в нём ниже. Кроме того, носители языка существенно различаются по пороговым величинам, при которых ими ощущается смещение ударения с одного слога на другой.

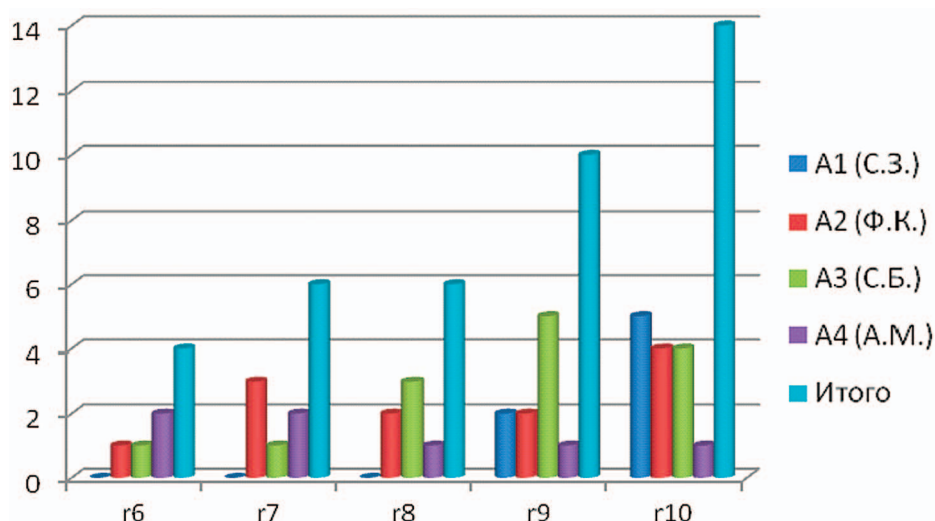


Рис. 8. Восприятие финального ударения в стимулах r6 – r10



Временная структура стимулов была получена по усреднённым данным. Значит, гипотетически возможны случаи, когда темпоральная структура вступает в противоречие с тональной. В первом эксперименте такие случаи специально не рассматривались, поэтому пока трудно судить, какой фактор имеет преимущество. Оценивая эксперимент, нужно иметь в виду, что во время тестирования аудиторы могли допускать грубые ошибки?³. Всё это предполагает дальнейшую проверку гипотез на ином материале и с другими информантами.

В Эксперименте 2 изучался один из наиболее вероятных случаев нефинального ударения в дари — присоединение к именной части речи энклитического безударного артикля — *ē*: например, *Sēb яблоко* — *sēb-ē какое-то яблоко*. Артикль, аналогично другим местоименным энклитикам, входит в состав того же фонетического слова, что и имя, и не образует стыка со следующей морфемой, который мог бы сигнализировать носителю языка о словесной границе. Возможно, это обстоятельство и позволило В.И. Миколайчику считать артикль дистантным суффиксом [15, Введение]. На наш взгляд, однако, верна традиционная точка зрения на артикль как служебную часть речи, так как она согласуется с широко принятым постулатом о непроницаемости слова в иранских языках⁴. Пересмотр этого положения, конечно, возможен с формальной точки зрения, но приведёт к очень громоздкой грамматике. В частности, в случае отказа от принципа непроницаемости пришлось бы вводить целый ряд правил, объясняющих, почему словосочетания и предложения нельзя считать словами.

Исследование ударения на минимальных парах предполагает, что структуре с артиклем должна быть противопоставлена другая идентичная по фонемному составу структура, но с ударным [*ē*] в абсолютном исходе. Проблема заключается в том, что ударное [*ē*] в абсолютном исходе встречается в языке дари крайне редко. Это является ещё одним подтверждением слабой фонемной противопоставленности гласных⁵. Слабость, которая компенсируется закреплённостью гласных за определёнными слоговыми структурами и позициями по отношению к ударению. Односложные слова типа *bē без*, *pē нет*, *šē что*, *sē три* в счёт не идут, так как в них нет безударных слогов и поэтому их акцентуацию можно считать финальной с таким же успехом, как и начальной.

Единственной подходящей для просодического экспериментирования структурой можно считать слово *šanbē суббота* и производные от него *yakšanbē воскресенье*, *došanbē понедельник* и т.д. до *panjšanbē четверг*. К сожалению, для корректности эксперимента в языке отсутствует слово *šanb**, которое в сочетании с артиклем создавало бы контрастную структуру с начальной акцентуацией. Поэтому вновь пришлось обратиться к искусственным словам, и за исходный материал было взято нонсенс-слово

³ Это замечание, вероятно, можно отнести к информанту А4 (его ответы статистически значимо отличались от ответов остальных). Под грубой ошибкой мы понимаем опечатку, когда информант просто не там поставил галочку. В дальнейшем могло произойти закрепление неправильной реакции, что могло дать значительную систематическую ошибку.

⁴ Полагается, что между морфемами одного слова нельзя вставить другое слово, в частности, между корнем и суффиксом. Если же считать артикль суффиксом, то придётся мириться с тем, что между корнем существительного и его суффиксом можно вставить прилагательное. Более подробно об этой дискуссии [см. 9].

⁵ Как было показано в работе [10], все 8 гласных языка дари разбиваются на 3 кластера: [*i-ē-e*], [*a-ā*] и [*o-ō-u*], внутри которых в значительном слое лексики они могут в зависимости от стиля речи переходить друг в друга, не меняя значения слова. В этом и заключается их слабая фонемная противопоставленность.

bēbē*. В этом слове два одинаковых по фонемному составу слога, параметры которых можно варьировать при синтезе независимо друг от друга. Структура с ударением на последнем слоге соответствует именному типу акцентуации, которое должно восприниматься носителями как некое существительное или прилагательное. Структура с нефинальным ударением выглядит как существительное bēb* с артиклем -ē, т.е. bēb-ē*.

Синтез проводился на программном синтезаторе Д. Клатта, реализованном экспериментатором на ЭВМ СМ-4 Филфака МГУ. Для воспроизведения звука использовался комплекс КАМАК с цифро-аналоговым преобразователем, который управлялся специальной программой [17]. Адаптация синтезатора к комплексу СМ-4 КАМАК потребовала существенных изменений в программе Клатта. Кроме того, нам потребовалось написать программу генерации идентификационных тестов, совместимую с использованной ранее американской в **Эксперименте 1**.

Гласные слова bēbē были идентичны по формантному составу: три низшие форманты имели следующие частоты: $F_1 = 450$ Гц, $F_2 = 2200$ Гц, $F_3 = 2800$ Гц. Их ширина соответственно равнялась 60, 150 и 300 Гц. Эти величины были подобраны, исходя из результатов эксперимента по восприятию гласных иранцами и таджиками [11], таким образом, чтобы синтезированный [ē] звучал как нечто среднее между [e] и [i] персидскими⁶.

Длительность первого слога, параметры которого во всех стимулах были одинаковы, составляла 200 мс. Согласный [b] был создан формантными переходами в начале слога без привнесения дополнительных шумовых характеристик. Он занимал 30 мс. За это время частоты формант F_2 и F_3 линейно возрастали на 500 Гц до указанных выше величин, т.е. до квазистационарного участка гласного. Частота первой форманты (F_1) за это время возросла от 150 до 450 Гц, т.е. на 300 Гц.

Амплитуда голосового возбуждения в течение первых 30 мс возрастала от 0 до 60 Дб, затем поддерживалась постоянной в течение 140 мс, а за последние 30 мс звучания слога уменьшалась до 0. Затем следовала 50-миллисекундная пауза, которая разделяла слоги (см. рис. 9).

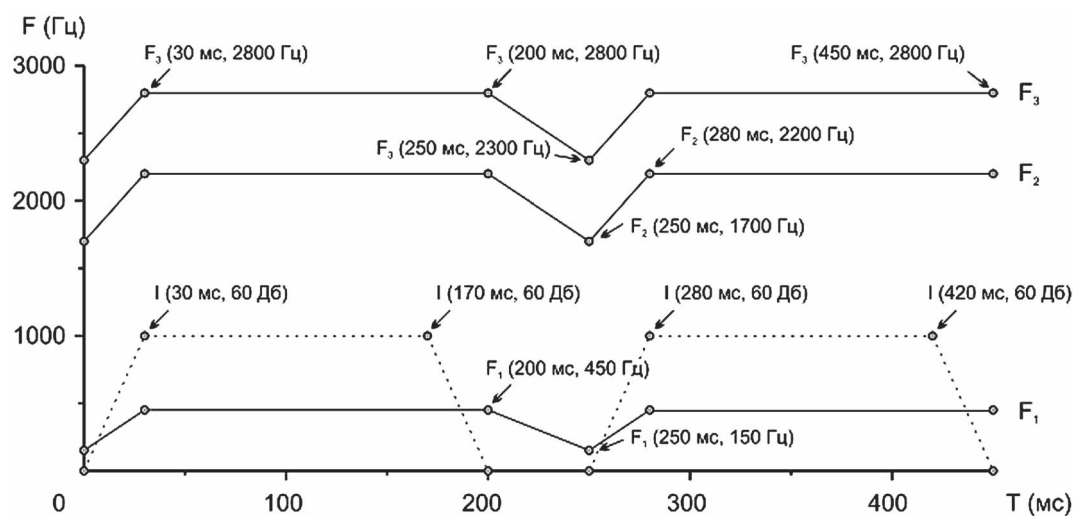


Рис. 9. Формантно-временная структура базового стимула nonsensical-слова bēbē. Контур интенсивности (I) обозначен пунктиром

⁶ Это помещает его в середину кластера [i-ē-e], где он должен теоретически находиться с точки зрения артикуляции.



Работа синтезатора предполагает использование и других параметров помимо описанных, которые были взяты из статьи [2, р. 976]. Параметр GO (общее усиление) в каждом стимуле подбирался экспериментально для устранения искажений при оптимальном соотношении «сигнал/шум». Различия в громкости стимулов не превышали 5 Дб. Синтезатор Д. Клатта моделирует инерционность речевого тракта, поэтому при изменении амплитуды голосового возбуждения интенсивность сигнала меняется не мгновенно, а за время порядка 5–10 мс. Другая особенность — моделирование взаимосвязи интенсивности с частотой основного тона, т.е. при прочих равных условиях стимулы с более высоким тоном получаются более громкими. Эти особенности обуславливают частные отклонения интенсивности стимулов от контура на рис. 9. По оси X откладывается время в мс, по оси Y — частота в Гц. Пунктирный контур (I) демонстрирует амплитудные характеристики, т.е. наличие фонации.

В эксперименте проверялись гипотезы о влиянии интенсивности (I), длительности (T), частоты основного тона (F_0) и его контура на восприятие ударения. Первые три фактора варьировались только во втором слоге, контур менялся в обоих слогах. Итого, варьировалось пять факторов, каждый из которых мог принимать два значения. Количественные факторы — интенсивность (I), длительность (T) и тон (F_0) — могли быть во втором слоге либо на 25% выше, чем в первом, либо на столько же процентов ниже. Уровень в 25% был взят как заведомо более высокий, чем тот, который бывает в естественной речи, чтобы, по выражению специалистов по планированию экспериментов, обязательно накрыть критические точки зависимостей. Качественный фактор — контур тона — мог быть либо падающим, либо восходящим. В том и другом случае приращение составляло 20 Гц на всю длительность слога, что тоже заведомо выше порога чувствительности информантов, который для людей с нормальным (не тренированным) слухом составляет около 5 Гц.

Варьирование пяти факторов на двух уровнях даёт нам 2^5 комбинаций, т.е. 32 стимула. Эти стимулы были сведены в идентификационный тест, в котором они случайным образом повторялись два раза со стандартными паузами (см. описание **Эксперимента 1**). В отличие от экспериментирования со словом راحت *rāhāt* *удобный*, когда перед аудиторами ставилась привычная задача соотносить звучание с написанием, анкета второго эксперимента потребовала более сложной установки. Информантам A5 и A6 ставилась задача распознать слово с артиклем (و حبت ی *wā-ye wahdat*), что предполагает их несколько более высокую лингвистическую образованность, не превышающей, однако, требований лица. В анкете был напечатан ряд нулей и единиц. Аудиторам было объяснено по-персидски, что при опознавании слова с артиклем они должны отметить единицу, в противном случае — ноль.

Для проверки усвоения объяснений перед тестированием было проведено две тренировки. Крайние точки континуума, т.е. стимул с явным ударением на начальном и стимул с явным ударением на финальном слоге, были сведены в отдельный аудиотест, где они повторялись в случайном порядке десять раз. После каждой тренировки анкеты анализировались, обсуждались с аудиторами и в случае ошибок делались соответствующие замечания. Информанты хорошо усваивали поставленную задачу и уже после первой тренировки грубых ошибок не допускали.

Дисперсионный анализ показал, что все три количественных фактора — длительность (T), интенсивность (I) и тон (F_0) — существенно связаны с просодикой ударения (все ниже критического уровня $p = 0,05$). Наиболее зна-

чимым оказался тон ($p = 0,005$), затем следуют интенсивность ($p = 0,01$) и длительность ($p = 0,03$). Применённая методика дисперсионного анализа не позволяла, однако, определить степень значимости различия между факторами. Контур тона в слогах оказался незначимым, что подтвердило выводы предыдущего эксперимента.

Описываемый тест, как и предыдущий эксперимент, выявил определённую склонность информантов к более частому восприятию финального ударения. Во всех стимулах, где три фактора возрастали во втором слоге, слышалось финальное ударение. Но только 75% реализаций с обратным соотношением параметров воспринято с начальным ударением. Мы можем объяснить эту асимметрию большей распространённостью финального ударения в дари, так что требуется дополнительное усиление акустических факторов, чтобы побудить информантов услышать нечто противоречащее их ожиданиям⁷.

Особый интерес представляют случаи, когда факторы действуют противоречиво, т.е. два из них выделяют первый слог, один из них — второй. И здесь эффект финального ударения достигается значительно легче, чем начального. Тонем и интенсивностью по отдельности удаётся превозмочь действие остальных факторов в 58,3% случаев. Противодействие длительности выражено значительно слабее: только в 1/3 предъявлений была воспринята финальная акцентуация.

В противоположной части континуума, там, где первый слог выделялся одним фактором, второй — двумя, наибольшее противодействие восприятию финального ударения оказывала длительность (16,7%). Более высокий тон первого слога только в 8,8% нейтрализовал остальные факторы. Большая интенсивность первого слога ни разу не помешала аудиторам воспринять последний слог ударным.

В целом, результаты **Эксперимента 2** можно интерпретировать следующим образом. При значительном возрастании (около 25%) длительности, интенсивности и тона в одном из слогов каждый фактор становится значимым и связывается информантами с просодикой слова. Для однозначной интерпретации типа акцентуации достаточно согласованного воздействия двух факторов из трёх. Оставшийся фактор может противодействовать первым двум, практически не влияя на ударение.

При прочих равных условиях чаще слышится ударение на финальном слоге. Это хорошо согласуется с нашими предыдущими исследованиями и с общеизвестной гипотезой о более сильном выделении начального слога. В самом деле, обычно в естественной речи первый слог характеризуется более высокими уровнями факторов при изолированном произнесении слова. Когда же оба слога одинаковы и аудиторы не слышат усиления инициала, они полагают, что выделена финаль. Используемый уровень различия слогов $\pm 25\%$, очевидно, не предельно, что предполагает возможность синтета слов с ещё большими межслоговыми различиями.

Эксперимент 3. Ранее обследованные модели просодических оппозиций типа «имя + энклитика», где в качестве энклитики может быть местоимение или артикль, в противопоставлении цельному слову применимы ко многим иранским языкам. В трёх близкородственных языках — персидском, таджикском и дари — эти модели охватывают значительную часть случаев нефинального ударения в любом тексте.

Существует, однако, специфический для дари случай нефинальной акцентуации. Он обнаруживается тогда, когда заимствованное из арабского языка слово оканчивается на «согласный + 'айн». Такие слова даже в полном стиле произносятся так, как будто вместо 'айна звучит гласный [а]. Ударение не падает на него, так как его основной фонологический вариант является согласным. Например, *gôba* из رجب *rob'* четверть, *tâba* из طبع *tab'* печать. При образовании словосочетаний, когда после такого рода слов появляется морфема с гласным началом, гласный [а] исчезает или заменяется

⁷ Это явление аналогично перцептивной защите, которая проявляется, когда частота опознавания незнакомых слов становится существенно меньше, чем приличных.



согласным [ʔ] — перерывом в голосе. Например, طب تاба tába печать, но اول طب tabʔ -e awwá I первое издание.

К сожалению, и в этом случае не удаётся обнаружить чистых словарных оппозиций. Попытка использовать оппозицию قلع qála олово — قلعة qalá крепость не удалась, так как произношение qalá характерно только для неполного стиля. В полном же стиле, в котором информанты привыкли начитывать тексты, это слово произносится qaʔa. Поскольку из-за смешения стилей от информантов не удалось добиться естественного произношения, эта оппозиция не рассматривалась.

В отличие от **Эксперимента 2**, где слоги резко противопоставлялись друг другу (параметры изменялись скачком в $\pm 25\%$), в **Эксперименте 3** были запланированы более тонкие различия. В тесте 1 длительность (Т), интенсивность (I) и тон (F_0) варьировались во втором слоге в пределах $\pm 10\%$. Ранее было взято всего два уровня для каждого фактора, здесь же число уровней было доведено до пяти, в том числе, каждый параметр второго слога мог принимать значение, равное первому слогу (табл. 4). Варьирование трёх параметров на пяти уровнях даёт $5^3 = 125$ комбинаций.

Параметры первого слога во всех стимулах были постоянными и равнялись значениям, указанным в средней (третьей) строке таблицы. Таким образом, каждый параметр второго слога мог принимать значения от -10% до $+10\%$ от значения первого. Шаг в 5% процентов был выбран как весьма близкий к порогу восприятия.

Таблица 4

Уровни факторов во втором слоге нонсенс-слова táta

	Уровень	Интенсивность (Дб)	Тон (Гц)	Длительность (мс)
1	+10%	66	143	220
2	+5%	63	136	210
3	0%	60	130	200
4	-5%	57	124	190
5	-10%	54	117	180

Для афганских аудиторов (в то время студентов ТашГУ) были подготовлены анкеты, напечатанные в привычной для них арабо-персидской графике. В них в несколько колонок были напечатаны два опознаваемых слова: تاتا táta (буква ع ayn на конце означала безударный гласный [a]) и تاته tatá (буква ه hā-ye havváz на конце обозначала ударный гласный [a]). Как и в **Эксперименте 2**, оба предъявляемых слова были нонсенс-словами.

Тест 1 был прослушан двумя аудиторами — А6 и А7. Поскольку тест не выявил влияния длительности на ударение, а влияние интенсивности и тона было близко к критическому уровню, планирование эксперимента было несколько изменено. Уровень варьирования длительности второго слога был повышен до $\pm 18\%$ с шагом в 9%, с тем чтобы подобрать такие значения, которые могли бы нейтрализовать тон. Так был получен **Тест 2**.

Тест 2 был прослушан аудиторами А7 и А8. Аудитор А8 показал очень высокую чувствительность к тону ($p < 0,001$) и интенсивности ($p < 0,0001$). Однако в его восприятии не обнаружилось существенной связи длительности (Т) с просодикой. Совместная обработка данных по обоим аудиторам не выявила значимого влияния ни одного из факторов в отдельности на вос-

приятие ударения. Обнаружились существенные расхождения в оценках аудиторов ($p < 0,001$). Тем не менее, отмечены статистически значимые взаимодействия факторов: длительности (Т) и тона (F_0 , $p < 0,02$), интенсивности (I) и тона (F_0 , $p < 0,05$). Это можно интерпретировать как то, что для создания эффекта ударения на нужном слоге недостаточно каждого фактора в отдельности, необходима их оптимальная комбинация. В тех случаях, когда у обоих слогов все параметры были идентичны, преимущества какого-либо слога выявлено не было: в половине случаев ударным указывался первый слог, в остальных — второй.

В итоге проведения **тестов 1 и 2** не удалось подобрать такого сочетания параметров, при котором длительность слогов стала бы значимой для восприятия ударения. Поэтому был запланирован **Тест 3**. В его стимулах соотношение слогов по длительности (Т) варьировалось более крупными шагами по 15 мс, т.е. по 10% от длительности каждого слога. Тон (F_0) и интенсивность (I) изменялись в более узких пределах. Приращение интенсивности (I) составляло 2 Дб, т.е. 3,3%; приращение тона (F_0) — 4 Гц, т.е. 3%. Для более естественного звучания длительность (Т) каждого слога была уменьшена до 150 мс (исходный уровень). Пауза между слогами составляла 50 мс. Так же, как и в предыдущих текстах, три параметра варьировались в различных сочетаниях на пяти уровнях, что даёт $5^3 = 125$ сочетаний.

Тест 3 был прослушан аудиторами А9 и А10. Самым существенным фактором, определяющим ударение, была длительность (Т) ($p < 0,1$). Влияние тона было далеко от критического ($p < 0,25$). Однако при незначительном увеличении числа информантов вполне возможно усиление этого фактора до значимого⁸. Влияние интенсивности (I) оказалось столь невелико, что его вероятность не поддавалась определению по статистическим таблицам.

Эксперимент 3 показал, что при небольших уровнях варьирования тона и интенсивности на первый план выходит длительность слогов как фактор, определяющий место ударения.

Заключение

Инструментальные исследования естественной речи носителей языка дари показали статистическую значимость длительности слога для восприятия места ударения. Эксперимент с синтезированной речью подтвердил этот вывод. 10-процентное увеличение длительности какого-либо слога по сравнению с другим при прочих равных условиях достаточно для восприятия его ударным.

Синтез речи позволил получить и ряд принципиально новых результатов. Выяснилось, что 20-процентное повышение тона способно нейтрализовать действие темпорального фактора. Повышение интенсивности также может нейтрализовать воздействие темпорального фактора на восприятие, но при ещё более высоких уровнях — от 25% и выше. Причём взаимосвязь тона и интенсивности, с одной стороны, и ударения, с другой, ощущается не всеми носителями дари. Изменение контура тона не влияет на восприятие ударения.

Такое соотношение просодических характеристик может быть объяснено следующими факторами:

1. Ударный слог главным образом маркируется большей длительностью. Однако более высокий тон и интенсивность могут воздействовать на восприятие ударения в каком-либо слоге как общечеловеческие психоакустические признаки затраты больших усилий на произнесение того или иного слога [3, р. 231].

⁸ Существует мнение, что при неограниченном увеличении информантов любой фактор можно представить статистически значимым. Поэтому, как это ни парадоксально, наибольшего доверия вызывают высокозначимые результаты, полученные на малых и средних выборках.



2. Принимая во внимание, что в близкородственных языках — таджикском и персидском — ударение тоническое, можно предположить, что просодическая роль тона в дари была отодвинута на второй план его историческим развитием и контактами со вторым официальным языком пашто.

За последние десятилетия на дари в меньшей степени действовали консервативные факторы в виде средств массовой информации и стандартизированного школьного образования, чем на соседние государственные — персидский и таджикский. Дари отличается повышенной вариативностью, что скорее приводит к расшатыванию произносительной системы, чем к её консолидации. Существенное влияние на дивергенцию просодических моделей персидского и таджикского, с одной стороны, и дари, с другой стороны, могли оказать контакты с языком пашто, ударение в котором имеет преимущественно квантитативный характер.

Литература

1. *Hirst Daniel and Di Cristo Albert. A survey of intonation systems // Intonation Systems: A Survey of Twenty Languages. Cambridge: Cambridge University Press, 1998. Vol. 1: 1.*
2. *Klatt D. H. Software for a cascade/parallel formant synthesizer // Journal of the Acoustical Society of America. 1980. Vol. 67(3).*
3. *Lehiste I. Suprasegmental features of speech. Contemporary issues in experimental phonetics. New York, London: Academic Press, 1976.*
4. *Roach P. English Phonetics and Phonology. A practical course. [s.l.]: Cambridge University Press, 2002.*
5. *Брызгунова Е.А. Звуки и интонация русской речи. Лингафонный курс для иностранцев. М.: [б.н.], 1969.*
6. *Галунов В.И. и Родионов В.Д. Моделирование процессов передачи информации в звуковом диапазоне. Ленинград: Внешторгиздат, 1988.*
7. *Златоустова Л.В. Фонетическая природа русского словесного ударения. Спб.: МГУ, 1953.*
8. *Иванов В.Б. Акустические характеристики персидского словесного ударения. Автореферат кандидатской диссертации. М.: МГИМО, 1976.*
9. *Иванов В.Б. Границы слова и инкапсуляция в персидском, таджикском и дари // Вопросы языкознания. М.: [б.н.], март 1995 г. С. 107–117.*
10. *Иванов В.Б. Динамика вокализма в персидском языке и дари // Исследования по иранской филологии / ред. Островский Б.Я. М.: Кафедра иранской филологии ИСАА при МГУ, 1997. Т. I: III.*
11. *Иванов В.Б. О сопоставлении персидского и таджикского вокализма на основе синтеза речи // Филологические науки. М.: МГУ, 1987(5) г.*
12. *Иванов В.Б. Ударение в дари (этап анализа) // Актуальные проблемы строя иранских языков / ред. Лебедев К.А. и Островский Б.Я. М.: МГУ, 1988.*
13. *Кодзасов С.В. и Кривнова О.Ф. Общая фонетика. М.: РГГУ, 2001. 592 с.*
14. *Кодзасов С.В. Исследования в области русской просодии. М.: Языки славянских культур, 2009. 492 с.*
15. *Миколайчик В.И. Основы теоретической грамматики персидского языка. М.: Военный университет, 1980.*

Иванов В.Б.

Синтез просодических моделей в языке дари

16. *Тарбеева Н.М.* Экспериментальное изучение словесного ударения в языке пашто. М.: ИСАА МГУ, 2010.
17. *Трунин-Донской В.Н. и Чучупал В.А.* Цифровая обработка речевого сигнала // Автоматическое распознавание слуховых образов (АРСО). Киев: [б.н.], 1977.
18. *Хаскашев Т.Н.* Фонетическая природа словесного ударения в современном таджикском литературном языке (экспериментально-фонетическое исследование). Спб.: ЛГУ, 1972.

Сведения об авторе

Иванов Владимир Борисович —

заведует кафедрой Иранской филологии Института стран Азии и Африки (ИСАА) МГУ им. М.В. Ломоносова, а также в качестве совместителя является ведущим научным сотрудником в секторе Иранских языков Института лингвистики РАН, доктор филологических наук, профессор. Преподаёт персидский язык, его теорию и общелингвистические дисциплины. Ведёт полевые и экспериментальные исследования в области иранских языков и диалектов. Опубликовал более 85 работ. Автор учебников по персидскому и таджикскому языкам.



Взаимосвязь и взаимодействие составных частей комплексного идентификационного исследования личности по голосу и звучащей речи

Каганов А.Ш., кандидат технических наук, старший научный сотрудник Российского центра судебно-медицинской экспертизы

Цель статьи — прослеживание взаимосвязи и взаимодействия составных частей комплексного исследования в задаче криминалистической идентификации личности по голосу и звучащей речи. С этой целью анализируются различные этапы такой криминалистической экспертизы с точки зрения вопроса о сферах использования инструментального и слухового анализа.

Статья написана на материале реальных криминалистических экспертиз с соблюдением современных требований, которые предъявляются к экспертным идентификационным исследованиям личности говорящего.

• *голос* • *речь* • *криминалистические исследования* • *судебная экспертиза*.

The purpose of this paper is to establish a relationship of complementarities of the components of complex research in the problem of forensic identification by voice and sounding speech. To this end, analyze the different phases of a forensic examination in terms of the scope of the use of instrumental and auditory analysis in this problem.

The article is written on a material of real forensic examinations with observance of modern requirements which are shown to expert identification researches of the person of the speaking.

• *voice* • *speech* • *criminalistic research* • *forensic examination*.

Анализ звукозаписей, поступающих в последние годы в государственные экспертные учреждения Минюста и Минобороны России для проведения криминалистической идентификации личности по голосу и звучащей речи, показывает тенденцию изменения характеристик, предназначенных для этих исследований фонограмм, в сторону ухудшения их качества.

В ситуации недостаточного объёма, зашумлённости, низкой разборчивости анализируемого речевого материала, различия технических характеристик каналов записи и воспроизведения и др. может возникнуть вопрос о достоверности фундаментальных теоретических выводов, полученных

в хороших условиях записи, близких к лабораторным. Однако современная экспертная практика подтверждает достоверность теории речеобразования и её объяснительные возможности. В то же время указанные выше частотность и многообразие ограничений, накладываемых на материалы реальных звукозаписей, поставили перед экспертами-криминалистами проблему разработки методов преодоления названных трудностей, что можно рассматривать как дальнейшее развитие теории в одном из её прикладных аспектов.

Ограничения, накладываемые на речевой материал реальных экспертиз, в большинстве случаев вызваны следующими причинами:

- низкое качество фонограмм: зашумлённость, наличие на фонограмме помех канала записи и искажений речевого сигнала;
- ограниченность объёма речевого материала, вплоть до отдельных предложений;
- несопоставимость материалов исходных и сравнительных записей, обусловленная различием эмоционального состояния говорящих, неодинаковой ситуацией речевого общения, различием характеристик шумов акустического окружения.

Обсудим те методы, которые применяются в ходе криминалистической идентификации личности по голосу и звучащей речи.

В процессе решения рассматриваемой задачи идентификации используются различные виды анализа речевого сигнала [3]. Характеризуя эту совокупность, необходимо раскрыть содержание и сопоставить области применения в исследовании следующих терминов (и соответственно, видов анализа): акустический, перцептивный, артикуляционный, слуховой.

Термины «акустический» (от греч. *akustikos* — слуховой) и «перцептивный» (от лат. *perceptio* — представление, восприятие) содержат общий смысловой компонент: «то, что слышно и воспринимается на слух». В то же время следует обратить внимание на тот факт, что есть существенное различие в значении и в употреблении этих терминов. Перцептивные признаки описываются с фонологической и психолингвистической точек зрения, а акустические измеряются инструментально. Последнее делает инструментальный анализ одним из центральных в рассматриваемом комплексном идентификационном исследовании.

Любой речевой отрезок можно охарактеризовать в акустических и в артикуляционных терминах. Так, в синтезе речи, в работе над интонацией, в инструментальном анализе используются акустические характеристики (первая форманта, вторая форманта; повышение основного тона, понижение основного тона и др.), а при постановке или анализе произношения звуков — артикуляционные характеристики (щелевой, переднеязычный, губно-губной и др.). Выбор использования акустического или артикуляционного описания зависит от содержания и цели теоретических и прикладных работ. Отметим, что в известных классификациях звуков совмещаются артикуляционные и акустические характеристики: например, переднеязычный, заднеязычный, щелевой, смычный, взрывной, с одной стороны, и глухие, звонкие согласные — с другой.

Далее необходимо обратить внимание на то, что термины «перцептивный» и «слуховой анализ» имеют общий смысловой компонент: опора на слух, на восприятие речи. Но при этом термин «перцептивный» используется в области фонологии (перцептивная функция фонемы, перцептивно сильная позиция, перцептивно слабая позиция гласного), в то время как «слуховой анализ» позволяет охватить большой объём речевого материала, отметить мельчайшие изменения в движении органов речи (особенно тела языка) и хорошо согласуется с данными инструментального анализа, позволяя установить соотношение того, что слышно, и того, что можно наблюдать визуально с помощью программ «видимой речи» [6].

Из всех терминов, описывающих речевой сигнал, однозначным, самобытным является термин «артикуляционный». Другие термины (акустический, перцептивный, слухо-



вой) различаются сферой употребления и особенностями исследуемого материала.

Перейдём теперь к вопросу о сферах использования инструментального и слухового анализа в задаче криминалистической идентификации личности по голосу и звучащей речи.

Инструментальный анализ используется для:

- измерения параметров частоты основного тона (ЧОТ) (средняя, минимальная, максимальная, среднеквадратическое отклонения ЧОТ, относительный диапазон изменения ЧОТ — D и др.);
- изучения формантных характеристик (значений формант, формантных соотношений, в том числе, при разном эмоциональном состоянии говорящего);
- определения средневзвешенной длительности;
- вычисления средневзвешенного среднеквадратического отклонения параметров речевого сигнала;
- исследования непрерывности / прерывности речевого потока.

Возможности слухового анализа также поддаются каталогизации. При этом следует учитывать, что в практике экспертного исследования принято разделение слухового анализа на аудитивную и лингвистическую части, которые проводятся в разных аспектах. Разделение обусловлено тем, что обычно представление идентификационных признаков даётся в противопоставлении таблиц, иллюстрирующих доказательство наличия реализаций того или иного идентификационного признака в речи фигуранта экспертного исследования, с одной стороны, и в речи подозреваемого — с другой. Однако не все идентификационные признаки укладываются в табличное представление. В то же время эти признаки весомы для характеристики фигуранта с разных точек зрения. Поэтому в подавляющем большинстве случаев аудитивный анализ выделяется в самостоятельную часть экспертного исследования.

Аудитивная часть идентификационного исследования включает в себя:

- позиционный анализ фонетического слова, включающий разного вида ассимиляции, степени продвинутости фокусов артикуляции вперёд или назад, скольжение гласного в момент произнесения, степень мышечной напряжённости, способ соединения (примыкания) согласного — гласного и гласного — согласного, характер редукции гласных, возможные отклонения от формулы А.А. Потемби, степень чёткости артикуляционных переходов от слога к слогу, от слова к слову, степень раствора ротовой полости в потоке речи при произнесении отдельных гласных;
- определение типов интонационных конструкций (ИК) в нейтральных и эмоциональных реализациях, определение переходных типов ИК — ИК — 1², ИК — 3², ИК — 4⁶ и др., место центра ИК, стыковки ИК, владение интонационной синонимией, владение интонационно-смысловым членением на синтагмы, степень чёткости произношения постцентровых частей ИК — 1, ИК — 2, ИК — 3, ИК — 7 [1];
- определение тембра голоса с точки зрения его восприятия на слух (обволакивающий тембр, расплывающийся по краям, густой и др.);
- темп и владение разными темпами речи на протяжении высказывания;
- степень владения речевым дыханием, возможность влияния речевого дыхания на длину синтагмы и др.;
- определение индивидуальных особенностей речи в области произношения;

— определение региональных особенностей речи в области звуков, фонетического слова, интонации.

Цель аудитивной части исследования — установление идентификационных признаков данной группы с помощью методов слухового анализа голоса и звучащей речи.

Для проведения такого анализа в настоящее время используются специализированные программные пакеты и звуковые редакторы различных типов, а также аппаратура, предназначенная для воспроизведения фонограмм, усиления и коррекции акустических сигналов. Оценку качественных характеристик звучащей речи следует проводить по критериям ГОСТа Р 50840–95 [3].

Помимо инструментальной и аудитивной частей комплексного исследования в задаче криминалистической идентификации говорящего существенную роль играет лингвистический анализ, который учитывает и обобщает взаимосвязи различных уровней языка и включает:

- остаточные диалектные явления;
- синтаксические особенности речи (степень владения синтаксисом в диалогической и монологической речи: соотношение простых и сложных предложений, наличие или отсутствие грамматических несогласованностей словоформ);
- лексический запас фигуранта экспертизы в исходной и в сравнительной записях;
- стилистические особенности речи (профессионализмы, метафоризация и др.);
- соотношение нейтрального и эмоционально окрашенного в области синтаксиса, лексики, интонации;
- отражение психического и эмоционального состояния фигуранта в его речи;
- сопоставление речевых особенностей фигуранта в исходной и в сравнительной записях;
- частотность употребления слов и выражений, характерных для определённого временного отрезка: два-три года, пять лет («имеет место быть», «ускорение», «ваучеризация»);
- обобщение выявленных идентификационных признаков фигуранта экспертизы (идиолект) и др.

В процессе проведения экспертного исследования, составляющие каждого из приведённых выше перечней могут быть конкретизированы.

Таким образом, параметры слухового анализа сопоставимы с параметрами инструментального и находятся с ними в отношениях взаимодополнения. Инструментальный анализ способствует объективизации исследования, выявляя акустические параметры речевого сигнала. Сферы инструментального и слухового анализа взаимосвязаны и в то же время каждая из них обладает автономностью.

Типичным проявлением автономности является установление различий голосов, имеющих сходство при восприятии на слух. Сходство голосов может быть обусловлено возрастной близостью говорящих, их сходным эмоциональным напряжением (состоянием), общим уровнем владения литературным языком и др. Кроме того, сходство может умышленно имитироваться, в том числе, и с противоправными целями. В такой ситуации эффективность инструментального анализа выше. Он позволяет выявить особенности параметров основного тона и формантного состава, идентифицирующие конкретного говорящего.

В других ситуациях более эффективным оказывается слуховой анализ речевого материала. Именно с его помощью легче установить артикуляционную базу говорящего в области звуков, фонетического слова, интонации, а также в области интонационно-смыслового (синтагматического) членения предложения в живой речи.

Так, например, нередко в распоряжении эксперта оказываются несколько звучащих текстов с одинаковым лексико-грамматическим составом. Возникает вопрос: «Являются



ли эти тексты копиями одного и того же образца или каждый раз они наговаривались заново»? В этом случае эффективнее окажется слуховой анализ с последующей интонационной транскрипцией, в которой указывается тип интонационной конструкции, место интонационного центра и синтагматическое членение каждого из текстов.

В одном из наших экспертно-криминалистических исследований необходимо было провести аудитивно-лингвистический анализ голоса и речи неизвестных дикторов в четырёх звучащих текстах идентичного содержания. Указанный анализ исходных фонограмм 1–4 показал наличие следующих особенностей.

Во всех четырёх исходных фонограммах голос диктора относился к группе мужских голосов, низкого тембра. Громкость речи сохранялась примерно постоянной на большей части каждой из исходных записей. Текст произносился близко к полному стилю, с достаточной разборчивостью, с паузами, увеличенными между фразами, речевое дыхание — в пределах нормы.

На слух создавалось впечатление, что нарушена естественность звучания. Появлялись оттенки гнусавости, которые могут быть присущи говорящим, а могут быть вызваны сознательно или с помощью специализированных технических средств.

Исследование показало, что в исходных звучащих текстах 1–4 использовались разные типы интонации, но с незначительными различиями по степени выраженности акустических компонентов (темпа речи, паузации, громкости и др.). Поэтому в целом интонация во всех записях воспринималась как сдержанно-монотонная: отклонение частоты основного тона от среднего значения составило около 10% при типичной величине для разговорной речи до 20% [4].

Проведённая в процессе исследования сплошная интонационная транскрипция текстов выявила незначительные отличия в записях 1–4 по интонационно-смысловому членению предложений, употреблению типов интонации незавершённости и по смысловым акцентам. Отмеченные различия отразили варьирование смысловых оттенков, свойственных речи каждого говорящего даже при многократном чтении им одного и того же текста [2].

Таким образом, проведённый аудитивно-лингвистический анализ позволил установить, что звучащие тексты 1–4 каждый раз наговаривались заново.

В то же время общими для речи говорящих в записях 1–4 явились выбор и структура интонации завершённости и незавершённости. Интонация завершённости выражается незначительным понижением тона и незначительным усилением словесного ударения¹. Из разных вариантов интонаций незавершённости в анализируемых записях преобладает слабое повышение тона на выделенном слове².

Кроме того, прослушивание исходных записей 1–4 показало совпадение ряда фонетических особенностей речи диктора во всех проанализированных фонограммах.

Таким образом, инструментальный и слуховой (аудитивный и лингвистический) виды анализа в многообразии своих конкретизаций находятся

¹ В специальной транскрипции, принятой в «Русской грамматике» АН СССР 1980 г., это обозначается как ИК — 1².

² В специальной транскрипции это обозначается как ИК — 4⁶ [5].

в отношениях взаимодополнения. Они составляют единый методологический фундамент идентификации личности по голосу и звучащей речи, основанный на концепции «стыка» естественных и гуманитарных наук, научная база которых используется в криминалистической экспертизе звукозаписей (далее — КЭЗ). Юридически данная концепция реализуется в тексте Заключения эксперта-криминалиста в виде синтезирующей части, в которой сводятся воедино результаты аудитивной, лингвистической инструментальной частей комплексного идентификационного исследования и устанавливается индивидуально-конкретное тождество или различие между голосом и речью фигуранта экспертизы на исходной записи и голосом и речью подозреваемого на фонограмме — образце.

В настоящее время требуется развитие этой методологии в связи с усложнением задач КЭЗ на современном этапе её развития. Особенно это касается того круга вопросов, который связан с инструментальной частью исследования.

Литература

1. Брызгунова Е.А. Практическая фонетика и интонация русского языка. М.: Наука, 1963. 120 с.
2. Зиндер Л.Р. Общая фонетика. М.: Высшая школа, 1979. 312 с.
3. Каганов А.Ш. Криминалистическая экспертиза звукозаписей. Монография. М.: Юрлитинформ, 2005. 272 с.
4. Рамишвили Г.С., Чикоидзе Г.Б. Криминалистическое исследование фонограмм речи и идентификация личности говорящего. Тбилиси: Мецниереба, 1991. 265 с.
5. Русская грамматика в 2 т. / Гл. ред. Н.Ю. Шведова. М.: Наука, 1980; М.: ИРЯ РАН, 2005–2006 (репринт издания 1980). 1492 с.
6. Kersta L.G. *Voiceprint identification*. Bell Telephone Laboratories, Inc., Murray Hill, NJ; Nature, vol. 196, no. 4861, pp. 1253–1257, Dec 29, 1962.

Сведения об авторе

Каганов Александр Шлемович —

кандидат технических наук, старший научный сотрудник Российского центра судебно-медицинской экспертизы (РЦ СМЭ) Минздравсоцразвития России; эксперт 111-го ГЦСМиКЭ Министерства обороны Российской Федерации, г. Москва



Модификация речевого сигнала как следствие наличия эмоциональных состояний «страх» / «тревожность»

Потапова Р.К., академик Международной академии информатизации, доктор филологических наук, профессор, заслуженный работник Высшей школы РФ

Потапов В.В., доктор филологических наук, старший научный сотрудник филологического факультета МГУ им. М.В. Ломоносова

В статье представлен поиск релевантных признаков на перцептивно-слуховом уровне, соотносящихся с усилением психического напряжения в связи с изменением ситуации с последующим влиянием на речевую деятельность. В ходе эксперимента по перцептивно-слуховому анализу аудиторам было необходимо определить специфические характеристики голоса и речи каждого диктора в трёх ситуациях общения: тембр (окраска) голоса, манеру речи, высоту голоса и т.д. Полученные результаты могут быть использованы для криминалистической перцептивной идентификации говорящего, а также идентификации эмоциональных состояний «страх/тревожность».

• перцептивно-слуховое восприятие речи • эмоциональные состояния «страх/тревожность» • просодия речи.

This paper investigates whether psychic tension can influence upon speaker emotional state regarding prosodic features of fear and anxiety. As variation values were analysed different situations and different prosodic parameters (e.g. timbre of voice, manner of speaking, pitch etc.). The results can be used in the field of forensic perceptual identification of a speaker and identification of emotional state of fear and anxiety.

• auditory speech perception • emotional states «fear/anxiety» • prosodic of speech.

Эмоциям в речи посвящено огромное количество исследований в области психологии, социологии, лингвистики, психолингвистики. Эмоции и разум в речи неразрывно связаны между собой. Эмоции не сопровождают деятельность человека, а являются неотъемлемой составляющей сложной организации его психических процессов и состояний. Речь следует рассматривать как единство интеллектуального и эмоционального,

так как она является способом формирования и формулирования мысли посредством языка [1].

Являясь формой человеческого поведения, речь подвержена влиянию широкого ряда не полностью изученных факторов [2]. Измерения речевых характеристик говорящего показывают, что на качество голоса могут влиять различные факторы, например, смысл произносимого текста; язык, на котором произносится текст; физическое и психическое состояние говорящего; анатомические изменения органов речи в результате операции и др. Такие внутренние факторы, как усталость, болезнь, наличие алкоголя или наркотиков в крови, могут воздействовать на речевой сигнал как на сегментном, так и на супraseгментном (просодическом) уровнях. На качество голоса также может повлиять переход с одного языка (например, родного) на другой (неродного) и эмоциональное состояние говорящего. Установлено, что различным эмоциональным состояниям свойственны свои просодические характеристики, лексические единицы, словообразовательные средства (например, суффиксы), междометия, частицы и т.д. [3; 4].

Многочисленные исследования позволяют утверждать, что эмоциональные составляющие влияют и на темп речи. Например, в состоянии грусти, злорадства длительность высказывания, по сравнению с нейтральным произнесением, увеличивается, а при гневе и раздражении — уменьшается [5].

Проводились исследования влияния эмоционального состояния и на интонацию. В частности, В.С. Соколова исследовала влияние эмоций на вопросительные интонации во французском языке [6]. По её мнению, эмоции способны разрушать категорию вопросительности. Интонационный контур фразы способен изменяться вплоть до своей противоположности. Там, где в эмоционально относительно нейтральной речи наблюдается повышение тона, в эмоционально окрашенной речи может появиться падение тона и наоборот.

Кроме интонационных характеристик эмоциональная окраска способна оказывать влияние и на артикуляцию. Например, при выражении грусти имеется тенденция к закрытому варианту гласных и к увеличению степени их лабиализации, тогда как открытый вариант и уменьшение лабиализации могут быть характерны для реализации эмоций радости и злорадства, а просто открытый вариант — гнева [2].

В исследованиях эмоционального состояния говорящего применительно к криминалистике большое внимание уделяется, прежде всего, анализу таких эмоциональных состояний, как «страх» и «тревожность».

Ранее нами были выявлены следующие перцептивно-слуховые индикаторы эмоционального состояния «страх» / «тревожность»:

- увеличение числа хезитационных пауз;
- увеличение длительности хезитационных пауз;
- увеличение скорости артикуляции (для гласных);
- темпоральное слоговое скандирование;
- ограниченная громкость высказываний;
- «неуверенный» тембр произнесения гласных;
- наличие компонентов шума в артикуляции гласных;
- нарушение плавности в реализации частоты основного тона;
- появление элементов шепотной речи и др. [2].

Как показали исследования речевой симптоматики психофизиологического состояния индивидуума, эмоциональное состояние «страх» обладает такими характеристиками, как:

- тихий голос;
- сниженная голосовая насыщенность (часто реализация так называемого тонкого голоса);



- в большинстве случаев высота тона выше среднего уровня;
- в ряде случаев хрипотца голоса (следствие «зажатого» голосообразования);
- напряжённая речевая ритмика;
- наличие «скованных» ритмов, а также «рваного» (неравномерного) речевого ритма;
- слабая акцентуация ударных слогов («неуверенность в себе»);
- затруднённая артикуляция [2; 7].

Итак, как правило, ранее основное внимание исследователей ранее было направлено на изучение особенностей паузации, темпа речи, мелодических и динамических характеристик речи под влиянием эмоционального состояния «страх».

В данном случае на первом этапе исследования анализировались такие воспринимаемые при перцептивно-слуховом анализе характеристики речи, как речевое дыхание и манера говорения. Были рассмотрены также высота, сила и окраска голоса.

Следует упомянуть, что многие предыдущие исследования проводились на материале аудиозаписей речи актёров, имитирующих различные эмоциональные состояния. Однако такие эмоциональные состояния, как «страх» и «тревожность», наименее адекватно передаются в речи актёров и при различного рода имитации. А именно эти эмоциональные состояния наиболее важны при идентификации личности по голосу в судебных целях. Впервые наиболее подробно на неимитируемом естественном материале речь в этих состояниях была проанализирована в [8; 9].

Поставленная проблема решалась благодаря использованию в ходе эксперимента записей, сделанных в ситуациях реального речевого общения с участием дикторов — студентов отделения прикладной лингвистики МГЛУ¹. Целью исследования была проверка гипотезы о влиянии эмоционального состояния «страх» / «тревожность» на модификацию речевого сигнала.

Результатом данной работы стало выявление релевантных характеристик на перцептивно-слуховом уровне, соотносящихся с усилением психического напряжения с последующим влиянием на речевую деятельность.

Подготовка экспериментальных данных включала аудиозапись речевых высказываний дикторов (в возрасте от 17 до 20 лет) в трёх ситуациях общения:

- спонтанная речь (описание картинки без предварительной подготовки);
- реализация высказываний на семинаре (квазиспонтанная речь);
- реализация высказываний в условиях экзамена (квазиспонтанная речь).

Для получения фонограмм спонтанной речи информантам предлагалось описать картинку из каталога Н.И. Корнилова [8]. На описание картинки каждый испытуемый затрачивал не более одной минуты. Записи квазиспонтанной речи студентов были получены в ходе семинарских занятий и экзаменов. Запись производилась на магнитную ленту с использованием магнитофона «OLYMPUS Pearlrecorder S711», диктофона «SANYO

¹ Эксперименты проводились под нашим руководством М.М. Зайченко и Е.А. Корсаковой. (См. Collegium Linguisticum. 2008. М.: Рема, 2008.).

TRC-960C» и сотового телефона «Sony Ericsson». После накопления экспериментального материала была проведена его оцифровка с использованием программ «Adobe Audition 1.0» («Adobe System Inc.») и «Zhenilo Signal Workshop». Данные были переведены в формат WAV. Записи, сделанные на сотовый телефон с помощью программы «WAV<>AMR», были переведены в формат WAV. Затем из большого количества полученного материала были отобраны фонограммы девяти дикторов (трёх дикторов мужского пола и шести дикторов женского пола) в трёх ситуациях общения. Таким образом, общее число фонограмм, представленных «экспертам-фонетистам» для определения специфических параметров голоса и речи, составило 27 при общей длительности звучания, равной 24 мин.

В ходе эксперимента по перцептивно-слуховому анализу аудиторам необходимо было определить специфические характеристики голоса и речи каждого из дикторов в трёх ситуациях общения:

- 1) *высоту голоса* — очень низкий, низкий, высота ниже среднего, высота средняя, высота выше среднего, высокий, очень высокий;
- 2) *речевое дыхание* — норма, прерывистое, дискомфортное;
- 3) *силу голоса* — слабый, сила ниже среднего, сила средняя, сила выше среднего, сильный;
- 4) *тембр (окраска) голоса* — звонкий, глухой, бархатный, хриплый, взволнованный, мягкий, назализованный, сиплый, певучий, резкий;
- 5) *манеру речи* — официальная, сдержанная, спокойная, непринуждённая, развязная.

В процессе проведения эксперимента каждый испытуемый работал индивидуально. Фактор обсуждения признаков с другими испытуемыми исключался.

Согласно оценкам аудиторов, полученным в ходе перцептивно-слухового анализа, прослеживается тенденция, согласно которой высота голоса дикторов в ситуации описания картинки — *средняя / ниже среднего*; в ситуации ответа на семинаре — *выше среднего*; в ситуации ответа на экзамене — *выше среднего* (рис. 1).

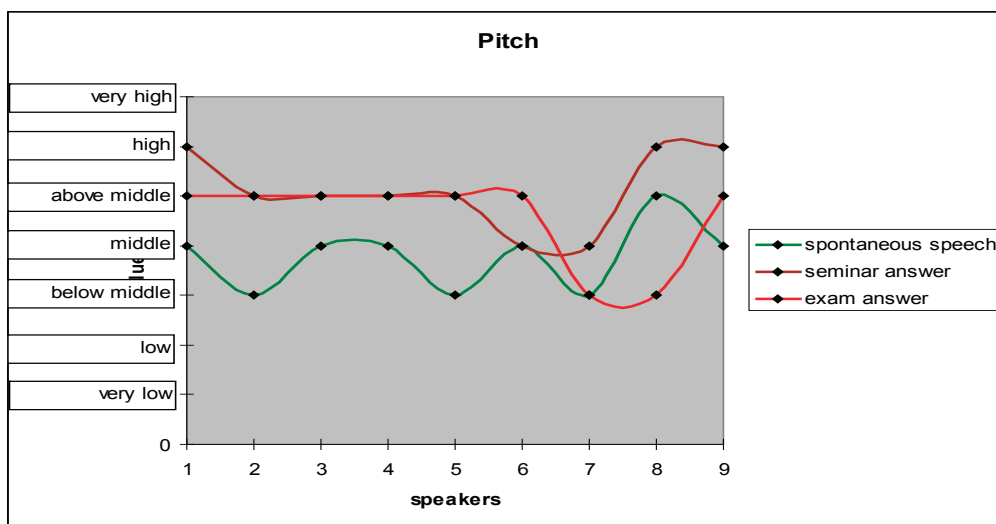


Рис. 1. Изменение высоты голоса

Сила голоса различных дикторов в ситуации описания картинки — *средняя*; в ситуации ответа на семинаре — *средняя / выше среднего*; в ситуации ответа на экзамене — *средняя / ниже среднего* (рис.2).

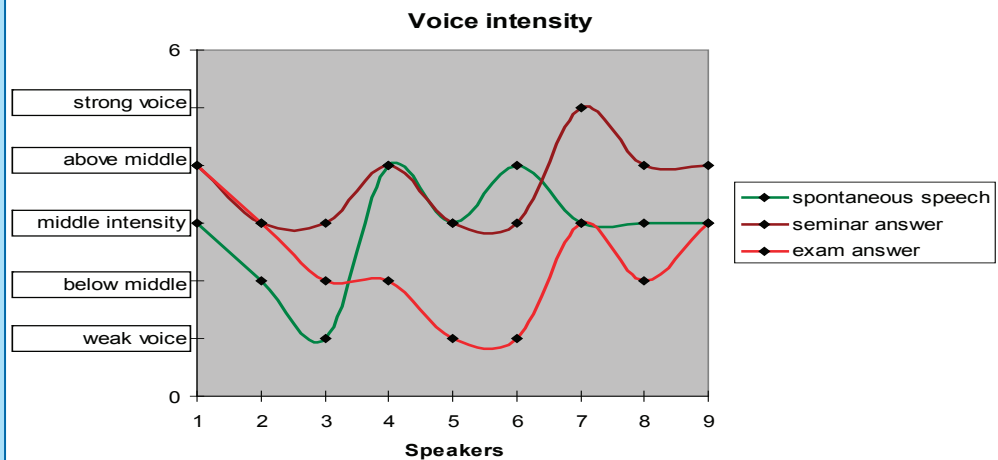


Рис. 2. Изменение силы голоса

Речевое дыхание у различных дикторов в ситуации описания картинки — норма; в ситуации ответа на семинаре — прерывистое; в ситуации ответа на экзамене — прерывистое (рис. 3).

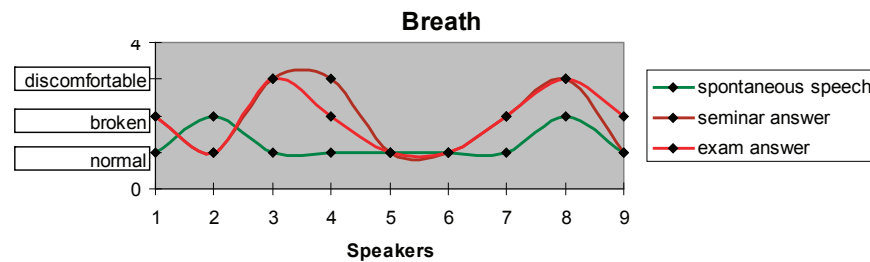


Рис. 3. Изменение характеристик речевого дыхания

Манера речи различных дикторов в ситуации описания картинки (спонтанная речь) — варьируется по дикторам между значениями «непринуждённая — спокойная — сдержанная»; в ситуации ответа на семинаре (квазиспонтанная речь) — «официальная»; в ситуации ответа на экзамене (квазиспонтанная речь) — «официальная» (рис. 4).

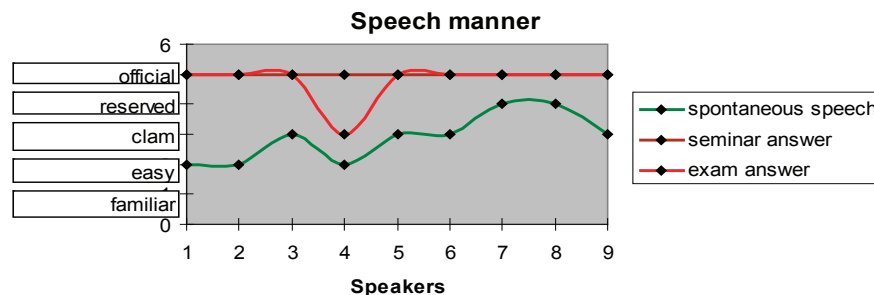


Рис. 4. Изменение манеры речи

По оценкам аудиторов окраска голоса дикторов в различных ситуациях общения имеет следующие характеристики: в нейтральной ситуации общения

(описание картинки) тембры голоса у дикторов друг от друга отличаются; в ситуации ответа на семинаре — у всех тембр *звонкий, взволнованный*; в ситуации ответа на экзамене — *взволнованный, глухой*.

В заключение следует сказать, что при интерпретации данных исключались такие факторы, как тип высшей нервной деятельности испытуемых, уровень воспитания, уровень подготовки при ответах на семинаре и экзамене. Учитывались только речевые характеристики поведения в различных ситуациях общения. В итоге на первом этапе применительно к исследованию были сделаны следующие выводы:

- под влиянием эмоциональных состояний «страх»/«тревожность», которые являются типичными для ситуаций ответа на семинаре и экзамене, высота голоса повышается;
- прослеживается тенденция увеличения силы голоса в ситуации «Семинар» и уменьшения силы голоса в ситуации «Экзамен»;
- речевое дыхание также меняется от нормального до прерывистого и дискомфортного под влиянием эмоциональных состояний «страх»/«тревожность» в ситуациях семинара и экзамена;
- манера речи становится официальной по сравнению с нейтральной ситуацией общения, и голос многих дикторов становится звонким и взволнованным в ситуации семинара; глухим и взволнованным в ситуации экзамена.

Полученные данные могут быть использованы в судебной фонетике при идентификации говорящего по голосу и речи, при идентификации эмоционального состояния говорящего, для дальнейшего сравнения с другими результатами, которые учитывают вербальную реализацию данных эмоций на примере других языков.

Главной задачей исследования на втором этапе был анализ характеристик речи, релевантных для речевой реализации таких эмоций, как «страх»/«тревожность» в русской речи. В ходе работы был проведён перцептивно-слуховой эксперимент, в котором приняли участие те же испытуемые.

На данном этапе аудиторы должны были отмечать особенности речи испытуемых с учётом следующих параметров:

- 1) *темп речи* (очень медленный, медленный, средний, быстрый, очень быстрый);
- 2) *паузы* (очень длинные, длинные, средние, короткие, очень короткие);
- 3) *заполнение пауз* (паузы незаполненные, сонант, слог, неопределённый звук, гласный);
- 4) *акцентуация (ритм)*:
 - сильная (ярко выраженный ритмический рисунок);
 - слабая (слабо выраженный ритмический рисунок).

Для каждой фонограммы (реализации высказываний на экзамене, на семинаре и реализации высказываний без коннотативного эмотивного значения /описание предложенной картинки/) в программе «Soundforge» было выделено по одной минуте. Из этих фонограмм для каждого диктора были сформированы файлы, представляющие собой три фонограммы для трёх речевых ситуаций, следовавших в определённой последовательности: описание картинки — ответ на семинаре — ответ на экзамене, общей продолжительностью 24 минуты.

Каждому аудитору был предъявлен специальный «протокол», где аудиторы отмечали те или иные характеристики дикторов, которых они прослушивали. Следует отметить, что запись можно было прослушивать желаемое количество раз. В процессе проведения эксперимента каждый аудитор работал индивидуально. Фактор обсуждения признаков с другими аудиторами исключался.

В ходе перцептивно-слухового эксперимента было установлено следующее:



Темп речи. В большинстве случаев в ситуации «нейтральная речь» (описание картинки) аудиторы отмечали темп речи как средний. В ситуации «семинар» темп речи обычно быстрый. Темп речи высказываний на экзамене обычно отмечался так же, как быстрый (см. рис. 5). Общая картина временных изменений речи подтвердила результаты в [9; 10; 11].

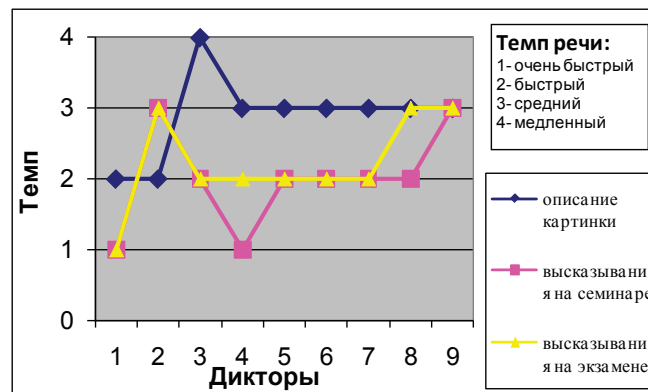


Рис. 5. Изменение темпа речи

Среднее значение темпа речи в трёх речевых ситуациях отмечено большинством аудиторов. Как видно из рис. 5, наблюдается тенденция к увеличению темпа в ситуациях «семинар» и «экзамен» по сравнению с нейтральной ситуацией (описание картинки). Вместе с тем темп речи в ситуациях «семинар» и «экзамен» может быть как одинаковым, так и различаться. В последнем случае темп речи в ситуации «семинар» реализуется чаще быстрее, чем в ситуации «экзамен».

Паузы в речи. Данные протоколов позволили определить, что паузы в «нейтральной речи» (описание картинки) в большинстве случаев оцениваются как средние, в ситуации «семинар» — как короткие или очень короткие. В ситуации «экзамен» паузы обычно оцениваются как короткие (см. рис. 6).

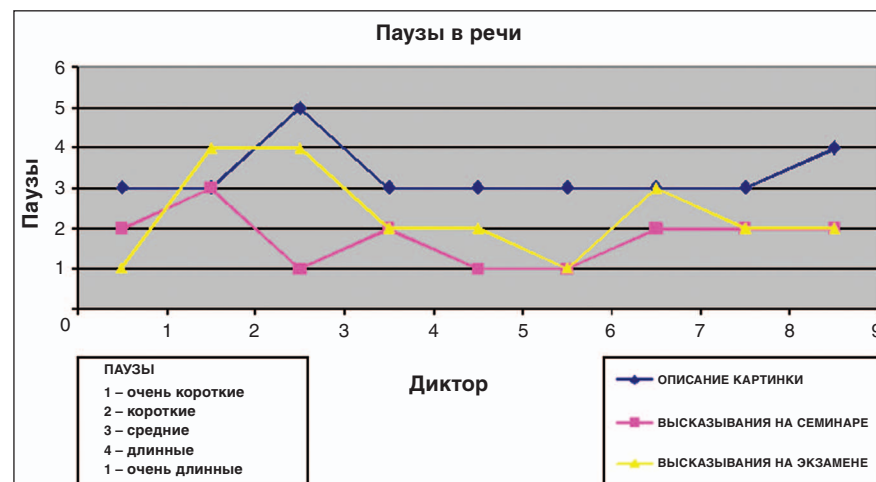


Рис. 6. Изменение паузации

Следует выявить наличие тенденции к уменьшению длительности пауз в ситуациях «семинар» и «экзамен» по сравнению с «нейтральной речью».

Причём паузы в ситуации «семинар» чаще короче, чем паузы в ситуации «экзамен». Иногда паузы при реализации высказывания на экзамене совпадают по длительности с паузами при описании картинки.

Заполнение пауз. В ситуации «нейтральная речь» (описание картинки), как отмечали аудиторы, паузы часто не заполнены. В ситуации «семинар» и «экзамен» паузы заполняются гласными довольно часто. Заполнение пауз сонантами, слогами или неопределённым звуком отмечалось аудиторами во всех ситуациях крайне редко.

Ритмический рисунок. В «нейтральной речи» (описание картинки) аудиторы отмечали как ярко выраженный, так и слабо выраженный ритмический рисунок. В ситуации «семинар» все аудиторы отметили ярко выраженный ритмический рисунок (усиленная акцентуация). В ситуации «экзамен» отмечали как ярко выраженный ритмический рисунок, так и слабо выраженный ритмический рисунок.

Результаты исследования позволяют констатировать наличие только определённых тенденций. Вместе с тем существуют также и отклонения от общего вектора воспринимаемых признаков, что может быть связано с:

- типом высшей нервной деятельности испытуемых;
- уровнем воспитания и культурой речи испытуемых;
- уровнем профессиональной подготовки испытуемых.

Проведённый эксперимент может рассматриваться как поиск релевантных признаков на перцептивно-слуховом уровне, соотносящихся с усилением психического напряжения в связи с изменением ситуации с последующим влиянием на речевую деятельность.

Результаты дальнейших исследований могут быть использованы для достижения следующих целей:

- определения по лингвопросодическим параметрам психического состояния личности в различных ситуациях;
- использования лингвопросодических параметров при слуховой идентификации говорящего;
- определения роли каждой из лингвопросодических характеристик;
- выявления возможностей по определению лингвопросодических признаков, связанных с изменением эмоционального состояния в направлении дихотомии «страх/тревожность» (своего рода слуховой детектор);
- распознавания эмоционального состояния говорящего в целях проведения криминалистической экспертизы;
- составления «портрета говорящего» (по Р.К. Потаповой), определения его паравербальных характеристик как носителя конкретного языка в конкретной ситуации общения;
- диагностирования психического и физического состояния говорящего, например, в медицинских целях.

Литература

1. Витт Н.В. Речь и эмоции: Учебное пособие к спецкурсу по психологии. М.: МГПИИЯ им. М. Тореца, 1984. 74 с.
2. Потапова Р.К., Потапов В.В. Язык, речь, личность. М.: Языки славянской культуры, 2006. 496 с.
3. Потапова Р.К. Коннотативная паралингвистика. М.: Триада, 1997. 2-е изд. 1998. 100 с.
4. Scherer K.R. Profiles of Emotion-antecedent Appraisal: Testing Theoretical Predictions across Cultures // *Cognition & Emotion*. 1997. P. 23–37.
5. Геладзе Н.В. Модификация интонационной структуры повествовательной фразы под влиянием эмоциональной окраски во французском языке // Сб. науч. трудов. Вып. 84. М.: Тр. / МГПИИЯ им. М. Тореца, 1975. 185 с.



6. Соколова В.С. Методические указания и задания по теоретическому курсу фонетики французского языка: Статьи. М.: МГПИИЯ им. М. Тореза, 1983. 20 с.
7. Fähmann R. Die Deutung des Sprachausdrucks. Bonn, 1967. 149 S.
8. Корнилов Н.И. Моя Каракалпакия. М.: Международная ассоциация «Искусство народов мира», Галерея «Никор», 2007. 150 с.
9. Potapova R.K., Potapov V.V. Speech temporal correlates of fear — anxiety (for Russian native speakers) // Proc. of IAFP Conference 2001. Paris, 2001. P. 3–5.
10. Потапова Р.К., Потапов В.В. Временные корреляты эмоций как специфические индивидуальные параметры идентификации говорящего в судебной фонетике // Ежегодник Рос. акустического общества «Акустика речи и прикладная лингвистика». М., 2002. С. 3–13.
11. Potapova R.K., Potapov V.V. Temporal Correlates of Emotions as a Speaker-State Specific Parameters for Forensic Speaker Identification (Speech Temporal Correlates of Fear — Anxiety for Russian Native Speakers) // Proc. of the International Workshop «Speech and Computer» (SPECOM'2003). Moscow, 2003. P. 291–301.

Сведения об авторах

Потапова Родмонга Кондратьевна —

академик Международной академии информатизации, доктор филол. наук, профессор. Заслуженный работник Высшей школы РФ, директор Института прикладной и математической лингвистики, зав. кафедрой прикладной и экспериментальной лингвистики, директор Центра фундаментального и прикладного речеведения Московского государственного лингвистического университета. Специалист в области романо-германского языкознания, общей и прикладной фонетики, теоретической, прикладной, экспериментальной и математической лингвистики. Автор свыше 500 научных и научно-методических публикаций. RKPotapova@yandex.ru

Потапов Всеволод Викторович —

доктор филологических наук, старший научный сотрудник лаборатории обеспеченного компьютером обучения центра новых технологий в гуманитарном образовании филологического факультета Московского государственного университета им. М.В. Ломоносова. Автор свыше 250 научных публикаций. Специалист в области славянского языкознания, компаративистики, теоретической и прикладной лингвистики. RKPotapova@yandex.ru

Аппаратно-программная платформа для ИКН-терминалов

*Свириденко В.А., доктор технических наук,
профессор, технический директор ГК SPIRIT DSP*



С появлением высокопроизводительных микропроцессоров актуальной стала задача разработки многофункциональных терминалов телематических систем на базе интеграции информационных, коммуникационных и навигационных (ИКН) технологий, ориентированных на решение комплекса задач в интересах различных мобильных пользователей и сервисов для них. Эти технологии служат основой существующих и проектируемых телематических систем, включая системы мониторинга и диспетчеризации подвижных объектов, интеллектуальные транспортные системы, планируемая ко вводу система ЭРА Глонасс для экстренного реагирования на аварии (аналог европейской системы eCall), системы обеспечения безопасности автомобилей, системы инфотейнмента (In Vehicle Infotainment — IVI), предоставляющих информационную поддержку и развлечения для водителя и пассажиров транспортного средства (ТС), и другие.

В статье рассматривается аппаратно-программная платформа (АПП) для терминалов ИКН-систем на основе универсальных и специализированных процессоров, выполненных по современным технологическим нормам и способных реализовать большой набор функций в составе оконечного компьютеризированного бортового оборудования ТС, а также концепции программного движка (engine) в составе прикладного программного обеспечения (ПО) АПП.

• телематика • интеллектуальные транспортные системы • инфотейнмент • аппаратно-программная платформа • терминал • программный движок (engine) • речевой интерфейс • SDR • host based принцип • навигация • Intel IVI • интеграция приложений • радиointерфейс • ОС • GPS/Глонасс • мультимедиа • движок речевой связи • навигационный движок • архитектура • клиентское ПО • серверное ПО • кодек IP-MR • приёмник.

High productive general purposed/specific processor based hardware and software platform for multifunctional terminals of telematic systems is proposed in the paper. It integrates information, telecommunication and navigation (ITN) technologies. The technologies serves as a basis for designed systems for mobile objects monitoring and control, intellectual transport systems, planned putting into operation ERA Glonass system (as an analog of European eCall), car security systems, in vehicle infotainment systems for driver and passengers entertainment etc. Different approaches may be used for the terminal implementation. Software engines (voice, video, telecommunication, navigation and others) are described shortly in the paper as a basic for flexible and quick realization of the car board electronic equipment.

• telematics • intellectual transport systems • infotainment • hardware and software platform • terminal • software engine • speech interface • SDR • host based principle • navigation • IVI • integration of applications • radiointerface • OS • GPS/Glonass • multimedia • speech communication engine • navigation engine • architecture • client SW • server SW • IP-MR codec • receiver.



Обеспечение многофункциональности, одновременное решение ряда задач, поддержка разнообразных услуг, доступ в Интернет по радиоканалу — основные тенденции развития инфраструктуры, ориентированной на сервис для широкого круга мобильных пользователей. Системы ИКН призваны удовлетворить требованиям таких пользователей. Одно из решений на уровне терминала — создание и использование унифицированной и достаточно универсальной аппаратно-программной платформы.

Известно, что выбор АПП для ИКН-терминалов и самой конфигурации телематической системы сложен. Он определяется приложениями, которые, в свою очередь, определяют рабочую нагрузку на вычислительные ресурсы и память её процессорных ядер. Но принимать во внимание нужно также другие технические и коммерческие требования.

Рассмотрим основные функции ИКН-терминала и указанные требования к нему. К функциональным требованиям, имея в виду перечисленные приложения, можно отнести следующие [1]:

- информационная поддержка на базе мобильного офиса и доступ в Интернет;
- поддержка двусторонней связи (в частности, видео и речевой);
- коммуникация в ближней зоне (например, Bluetooth, RFID и др.);
- коммуникация в дальней зоне (GSM/GPRS, 3G/4G, WiMax/LTE, спутниковая и КВ-связь и др.);
- речевой интерфейс водителя с бортовым компьютером и речевое управление некоторыми подсистемами транспортного средства;
- регулярная оценка PVT-данных (P — позиция, V — скорость, T — время) для ТС;
- цифровые карты и отображение текущей позиции ТС на цифровой карте;
- сбор информации от датчиков для оценки состояния подвижного объекта;
- регистрация событий;
- регулярная передача собранных данных в центр мониторинга/диспетчеризации;
- охранные функции (передача данных об особых событиях, биометрика, тревожная кнопка и др.);
- реализация Assisted-GNSS, поддержка системы дифференциальной коррекции и мониторинга (СДКМ) и SBAS при решении навигационных задач;
- вспомогательные функции (прокладка маршрута на цифровой карте, информация о пробках);
- проигрывание аудио- и видеофайлов;
- речевая или видеоконференцсвязь в многоточечном режиме (как опция);
- другие (в частности, спецприложения).

Технические требования:

- высокая производительность процессорного ядра;
- поддержка интерфейсов (CAN, USB и др.);
- малые массогабаритные параметры;
- низкое энергопотребление и возможности по его снижению;
- высокая точность навигации (PVT-данные) и малые значения времени старта (TTFF);
- хорошее качество речевой и видеосвязи;
- надёжность для разных условий эксплуатации;

- ремонтпригодность;
- возможности по модернизации;
- простота инсталляции;
- масштабируемость ПО.

Коммерческие требования:

- стоимость приобретения и гарантии/поддержки;
- стоимость эксплуатации и обновления;
- стоимость ремонта;
- стоимость услуг.

С учётом указанных требований и специфики приложений терминал ТС может рассматриваться в качестве клиента информационно-управляющей системы, взаимодействующего с её серверной частью (система «клиент-сервер»). С позиций пользователя такой терминал должен обладать собственными вычислительными ресурсами и большой памятью, чтобы самостоятельно решать ряд задач в его интересах, т.е. функционально это должен быть «толстый клиент».

Реализация такого терминала с высоким уровнем интеллектуализации становится возможной благодаря прогрессу процессорных чипов (более продвинутые технологические нормы (32 нм и менее), развитые архитектуры, многоядерность и др.). Возросшие ресурсы процессоров позволяют существенно расширить функциональность терминала телематической системы, одновременно решать несколько задач, обеспечить предоставление разнообразных услуг в интересах мобильных пользователей. Это также позволяет не только ставить новые задачи в рамках развития телематических систем, способствующих их интеллектуализации, начиная с терминального оборудования, но и менять основные реализационные концепции при их разработке и производстве (например, использовать подход на базе программного радио (Software Defined Radio — SDR) при имплементации разнообразных радиоинтерфейсов или хост (host based) принцип при имплементации телекоммуникационной, навигационной и других функций).

Термин «АПП» различными авторами определяется по-разному. Обычно под ним подразумевают компьютерную платформу с многоуровневой (многослойной) логической организацией, состоящую из аппаратной и программной части. К аппаратной части относятся физические устройства, архитектура процессорного ядра, используемые шины и система команд на нижних уровнях. Программная часть включает связующее (middleware) и системное ПО, а также прикладное ПО вместе с библиотеками, графическим (GUI) и речевым интерфейсами. Предполагается, что аппаратной части АПП соответствуют совместимая с ней программная часть (middleware, операционная система, прикладные программы).

Понятие «платформа» применительно к телематическим, в частности, интеллектуальным транспортным системам, используется широко. При этом различные крупные западные фирмы уже представили свои решения для ИКН-терминала в автомобиле. В их числе NXP ATOP, Microsoft Auto, CSR SIRFprima и SIRFprimaAuto, Intel IVI и другие [2–5].

Платформа ATOP (Automotive Telematics On-board unit Platform) позиционируется компанией NXP как многофункциональное трёхпроцессорное открытое телематическое решение «под ключ» с малым форм-фактором в виде системы в корпусе (SiP) размером 33x33x3 мм, в которой объединены три чипа. Оно может использоваться автономно или же в составе распределённой телематической системы. Решение ориентировано для таких приложений, как eCall, автоматизированное управление транспортом, оплаты проезда на дороге, охрана автомобилей, сервис на основе местоположения (LBS) и др. В качестве микроконтроллера используется RISC-процессор ARM 7, который предоставляет различные интерфейсы для интеграции ATOP с другими компьюте-



ризованными подсистемами автомобиля. Навигационный GPS-приёмник на базе отдельного чипа (в перспективе GNSS-приёмник) входит в состав чипсета.

Фирма SIRF, вошедшая недавно в компанию CSR, известна своими качественными навигационными GPS-чипами, которые широко используются в различных навигаторах и смартфонах с навигационной поддержкой. Последние ИКН-решения CSR для инфотейнмента в виде системы на кристалле (SoC) SIRFprimaAuto (дальнейшее развитие системы SIRFprima) предлагают многофункциональную платформу. Она акцентирована на позиционирование в сложных условиях, радиосвязь внутри автомобиля и связь с внешним миром, разнообразную мультимедийную поддержку, простую интеграцию в компьютеризированную среду автомобиля, позволяющую экономить время и деньги. Платформа SIRFprimaAuto комбинирует навигационный приёмник GPS/Galileo, процессор ARM11 в качестве CPU, аппаратные ускорители для 2D/3D графики (OpenGL ES, Direct3DM и OpenVG), мультимедийные средства и поддержку разнообразной периферии (сенсорные экран, системы хранения данных, видеокамеры, проигрыватели DVD, радиоприёмник и др.) на базе набора интерфейсов, а также коммуникационные компоненты, обеспечивающие таким образом много возможностей для взаимодействия различных подсистем. Микроэлектронные компоненты выполнены по современным технологическим нормам (65 нм в BGA-корпусе), что повышает надёжность электроники в условиях эксплуатации, характерных для автомобиля.

Программная часть этой платформы поддерживает все профайлы Bluetooth (включая A2DP, HFP, PBAP, SPP, AVRCP, DUN, MAP и др.). Встроенные речевые технологии (распознавание речи, синтез речи по тексту, подавление акустического эха, шумоподавление) обеспечивают речевое управление IVI-системой и качественную речевую связь с абонентами радиосетей и Интернета. В структуру платформы включён и видео-движок, обеспечивающий декодирование видеопотоков в соответствии с протоколами H.264/MPEG1/MPEG2/MPEG4/WMV9. Отметим, что ПО может работать под операционными системами Linux, Windows CE и Android.

Навигация — важная функция для подвижного объекта. Спутниковый навигационный приёмник с высокой чувствительностью (–161 дБм в режиме слежения) входит в качестве обязательного компонента в ИКН-терминал. В нём компания CSR использовала свою продвинутую технологию SiRFDRive GPS/dead reckoning. Эта технология сопрягает спутниковый навигационный приёмник, который обеспечивает долговременную точность позиционирования автомобиля, с сенсорами (например, одометрами, акселерометрами или приборами MEMS), не зависящими от каких-либо радиотехнологий позиционирования, но способными обеспечить только кратковременную точность. Это позволяет такой навигационной компоненте работать в сложной обстановке (насыщенной городской среде с затенениями, туннелями, городскими каньонами, эстакадами, подземными парковками и пр.), когда позиционирование при условии прямой видимости достаточного числа спутников обеспечивается спутниковым приёмником, а при отсутствии таких условий (например, в туннелях) позиционирование осуществляется на базе инерциальной компоненты интегрированного навигационного движка (engine). В спутниковом приёмнике используется фильтр промежуточной частоты (ПЧ) полосой 2 МГц с высокой избирательностью по соседнему каналу, чтобы минимизировать влияние радиопомех при работе электрического или элек-

тронного оборудования автомобиля. Ещё одним позитивным фактором технологии SiRFDRive является встроенная в неё технология SiRFInstantFix, обеспечивающая быстрый холодный старт навигации на основе вычисления (синтезирования) эфемерид навигационных спутников, снижающих параметр TTFF до значения пяти секунд.

В настоящее время платформа SiRFprimaAuto является одной из наиболее продвинутых для систем IVI, готовых для встраивания в ТС. Однако, на наш взгляд, её с точки зрения универсальности и функциональности превосходит платформа Intel IVI на базе процессора Atom, предназначенного для различных встроенных систем для мобильных приложений.

АПП Intel IVI реализована в виде пятислойной архитектуры или архитектурного стека. Его нижний слой представлен аппаратной частью, главным элементом (ядром) которой является процессор Atom (CPU). Он произведён по передовым технологическим нормам, которые ставят его вне конкуренции по части вычислительных ресурсов и памяти на рынке мобильных устройств. Это, в свою очередь, позволяет предоставить в распоряжение мобильного пользователя все те информационные, коммуникационные и мультимедийные возможности, которые он мог бы получить только на базе современного ноутбука.

CPU дополняется средствами ввода/вывода (I/O), как специфическими (например, шины MOST, CAN, SPI и др.), так и соответствующими промышленным стандартам (например, PCI Express), что позволяет производителям автоэлектроники реализовать конечные решения со многими опциями без дополнительных издержек и доработок ПО.

Типовая конфигурация аппаратной части Intel IVI — платформы дана в Таблице 1.

Таблица 1

Типовая конфигурация аппаратной части Intel IVI

Аппаратная функция	Краткое описание
Процессор Intel Atom	CPU поддерживает операции с фиксированной и плавающей точкой, технологию Hyper-Threading (Intel HT Technology) и технологию Virtualization (Intel VT)
Контроллер памяти	Поддерживает 1–2 MB DIMM/UDIMM подобно DDR2–533 и DDR2–667
Видео-декодер	Полная аппаратная поддержка декодеров MPEG2, MPEG4, VC1, WMV9, H.264 (базовый и расширенный профайлы уровня 4.1), DivX
Графический движок	Характеристики: скорость 400 мегапиксел/с и 3DMark*05 из 120
HD аудио	Высококачественное (HD) аудио на базе Intel HD Audio
Дисплей	Аппаратная поддержка типа LVDS/DVI/dRGB/TV Out WXGA 1280x800 18 bpp; XGA 1024x768 24 bpp
Ввод/вывод (I/O)	Gen1 PCI Express* x1 Expansion slots и USB 2.0
Блок I/O совместимости	Блок компонент типа PIC, RTC, Timer, GPIO, Power Management, Firmware Hub Interface и LPC, чтобы обеспечить загрузку ОС
Специфические I/O для авто-ОЕМов	MOST, CAN, SPI, Bluetooth, UART, SDIO, Ethernet, Radio Tuner, Video Capture, GPS, GRIО и др.

Над аппаратным слоем располагается уровень ОС. Он предполагает использование различных операционных систем, включая ОС реального времени (например, RTOS: QNX, Windows CE, VxWorks, Embedded Linux) и коммерческие ОС, используемые в PC (например, Linux, MS Windows XP Embedded). Слой ОС включает также драйверы для подключения внешних устройств в автомобиле.

Следующий слой — связующее ПО (Middleware) — включает разнообразный набор компонент и интерфейсов, реализующих для надстроенного над ним прикладного ПО такие функции, как интерфейс Bluetooth с поддержкой различных профайлов, подключение к сетям связи и протокольные стеки CAN и MOST.



Прикладной слой реализует различные приложения, включая те, что характерны для мобильных интернет-устройств (MID) или наладонных устройств (веб-браузеры, календари, управление подсистемами автомобиля, подсистема мультимедиа и т.д.). Прикладное ПО этого слоя отвечает за информационные функции (типа мобильного офиса), доступ в Интернет, навигацию и цифровые карты, мультимедиа (аудио и видео), диагностику, управление компьютеризированными подсистемами автомобиля и др.

Самый верхний слой обеспечивает интерфейс «человек-машина» (HMI). Он определяет специфику взаимодействия пользователя с системой IVI.

Отметим, что эта концепция АПП принята GenIVI — Альянсом производителей автомобилей, поставщиков аппаратно-программных платформ и сервисов для развития IVI-систем, куда входят BMW, GM, Renault, Intel, ARM, Continental, HARMAN и др.

Основная задача, которую ставят перед собой разработчики подобных платформ, — относительная простота интеграции различных приложений от разных поставщиков. Подходы к выбору платформы во многих случаях связаны с тем, на какое процессорное ядро делать ставку и какую операционную систему, вокруг которой строится экосистема, взять за основу. Этот выбор, прежде всего, определяется набором функций, которые должен поддерживать терминал. Главные из этих функций — связь, информация, навигация, мультимедиа, экстренное реагирование на ДТП, охранная сигнализация с мониторингом через Интернет, обеспечение удобства и надёжности управления, интеграция с внутренними подсистемами автомобиля.

Эти функции можно реализовать отдельно, но издержки и стоимость будет велика. Поэтому более эффективен путь интеграции в рамках платформы с многоуровневой структурой. В обобщённом виде она может быть представлена так, как показано на рис. 1 [6].

Здесь предполагается шесть уровней. Нижний из них (физический уровень) выделен из аппаратной части архитектуры платформы, чтобы в явном виде показать радиоинтерфейсы для радиосвязи (GSM/GPRS или TETRA — для обеспечения региональной радиосвязи там, где имеется соответствующее покрытие сетями мобильной связи: WiFi, 3G/4G, WiMax/LTE — для выхода в Интернет; Bluetooth и другие — для связи внутри транспортного средства или в непосредственной близости от него; спутниковая и КВ-связь — для обеспечения глобальной телекоммуникации для спецприложений), а также АЦП и ЦАП (например, для оцифровки сигнала микрофона или вывода на громкоговоритель), физические интерфейсы для подключения к подсистемам автомобиля (к датчикам) и др.

Собственно аппаратная часть включает высокопроизводительный центральный процессор (CPU) и, возможно, дополнительные чипы для телеком-поддержки (например, WiMax/LTE), навигации мобильного объекта и поддержки мультимедийных функций.

Уровень ОС создаёт операционную среду соответствующим системным ПО.

Связующее ПО также выполняет свои функции посредника между системным и прикладным ПО.

На самом верхнем уровне АПП расположен человеко-машинный интерфейс (ЧМИ), обеспечивающий пользователю удобное взаимодействие с терминалом и ИКН-системой в целом. Он также воспринимает биометрическую информацию (например, голос или изображение), которая используется для безопасности и охранных целей.

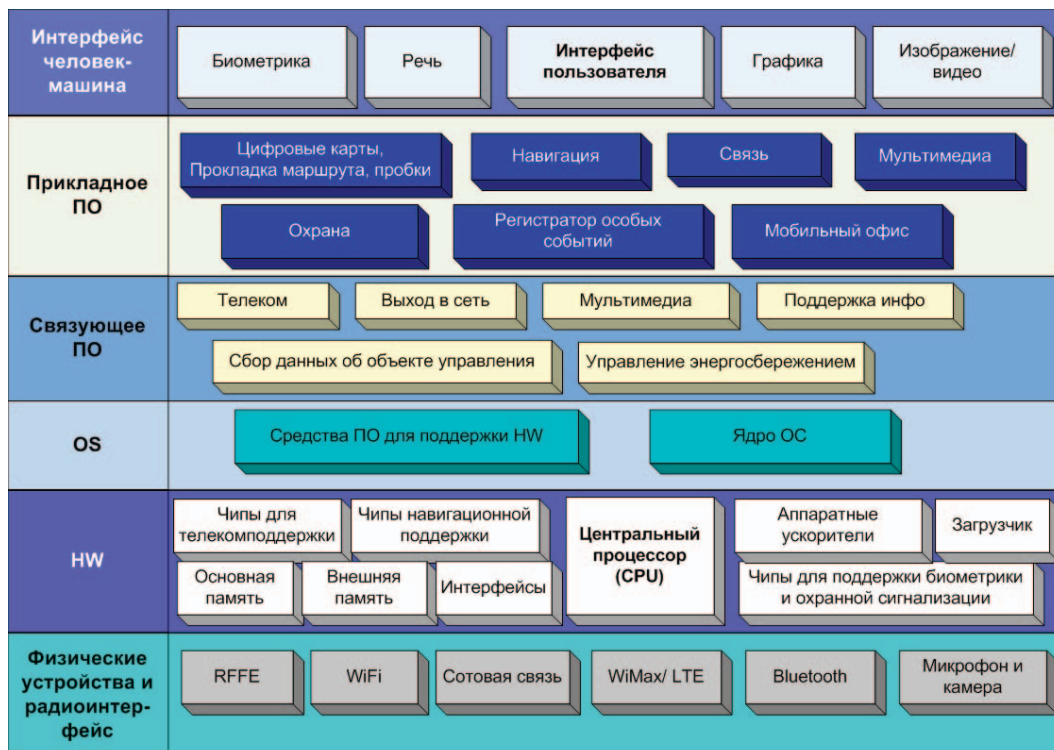


Рис. 1. Многоуровневая организация универсальной платформы ИКН-терминала

Усечённые версии этой универсальной платформы могут быть использованы для решения ограниченного набора задач. Так мониторинг транспортных средств предъявляет достаточно скромные требования к телематическому терминалу. Это — связь по сети сотовой связи и регулярная передача PVT-данных и информации о состоянии ТС, оцениваемого его внутренней системой сбора данных, по каналам GSM/GPRS. При необходимости обеспечивается речевая связь водителя с центром мониторинга и передача сигнала тревоги в экстремальных ситуациях. В основе такого терминала — многочиповое решение на базе микроконтроллера, GPS-приёмника и GSM/GPRS-чипа. Именно такой тип терминала успешно использовался в РФ ранее, но в последнее время GPS-чипы заменили на GPS/Глонасс-модули. В связи с началом выпуска GPS/Глонасс-чипов в 2011 г. началась замена указанных модулей на более компактные с двухсистемными приёмниками, реализованными в виде системы в корпусе (SiP) по технологии 90 нм. Сильную конкуренцию им начали составлять многосистемные навигационные чипы западных вендоров. Однако уже имеются планы увеличения числа функций в подобных терминалах, что определённо приведёт разработчиков к концепции универсального терминала, один из подходов к разработке которых представлен в данной статье.

В настоящее время большое внимание уделяется созданию системы ЭРА Глонасс, которая должна обеспечить быстрый вызов скорой помощи и сотрудников ГИБДД на место аварии, чтобы в «золотой час» повысить вероятность спасения участников дорожно-транспортного происшествия (ДТП) при худших его сценариях и оперативно освободить место ДТП. Очевидно, что требования надёжности фиксации ДТП, его местоположения и постоянной связи с соответствующими службами являются самыми важными. Основные функции ИКН-системы в этом случае таковы:

- двусторонняя речевая связь;
- позиционирование транспортного средства в любых его положениях;



- реализации функции тахографа (хранение логов);
- регистрация особого события (ДТП);
- сбор данных от датчиков (опция);
- автоматическая передача данных в центр спасения при регистрации особого события;
- совместимость с европейской системой eCall.

В этом случае в универсальной платформе можно исключить все функции, которые поддерживают инфотейнмент, кроме, возможно, обеспечения доступа в Интернет. Это удешевляет стоимость самой платформы для таких приложений. Однако к надёжности позиционирования (особо при регистрации ДТП) требования повышаются. Всё это обуславливает важность навигационной составляющей в терминалах системы ЭРА Глонасс.

Основной элемент представленной концепции универсальной платформы — центральный процессор. При его выборе необходимо оценить требуемые вычислительные ресурсы для реализации основных функций. Их грубая оценка в предположении, что соответствующая функция реализуется программно и не требуется соответствующая аппаратная поддержка, такова:

- навигация > 300 МИПС;
- коммуникации — 200 МИПС;
- мультимедиа > 500 МИПС;
- распознавание образов < 150 МИПС;
- система сбора данных — 20 МИПС;
- мобильный офис — 100 МИПС;
- другие приложения — 150 МИПС.

Ресурсы по памяти: основная память $\geq 5...10$ МВ, L2-кэш $\geq 128...512$ КВ, внешняя память $\geq 100...500$ МВ. Этим требованиям в полной мере удовлетворяют процессоры Intel Atom и Intel Core iX. В меньшей степени — RISC-процессоры ARM Cortex Aх, но они более энергоэкономичны, что важно для мобильных устройств. Из отечественных кандидатов на роль CPU можно назвать такие процессоры, как СБИС К1879ХК1Я (НТЦ «Модуль») и NVCom-01 («Ангстрём»), выполненные по нормам 130 нм.

В [1] предложен универсальный вариант АПП, которая адаптируется к конкретной телематической системе на основе концепции программного движка (engine), представляющего собой центральную часть ПО для реализации конкретной прикладной задачи. Такой движок может быть относительно автономным и переносимым, и наиболее часто он представлен в виде библиотеки или SDK. Использование готовых движков позволяет сократить время разработки ПО терминала и компоновать общее решение на базе конструирования архитектуры ПО для разнообразных терминалов. Применительно к ТС такими движками могут быть навигационный движок, движок речевой связи, видеодвижок, движок речевого интерфейса с пользователем, движок биометрики, телекоммуникационный движок, движок регистрации событий, движок сбора данных, движок машинного зрения и т.п. Такой блочный подход может сочетаться с готовыми программными решениями (например, с готовым ПО мобильного офиса, которое обеспечивает информационную составляющую программной платформы). На рис. 2 представлено одно из решений с акцентом на прикладное ПО для архитектуры АПП на базе движков. Оно развивает собой концепцию АПП, показанной на рис.1. Здесь зелёным цветом полностью или частично окрашены компоненты, разработанные компанией SPIRIT.

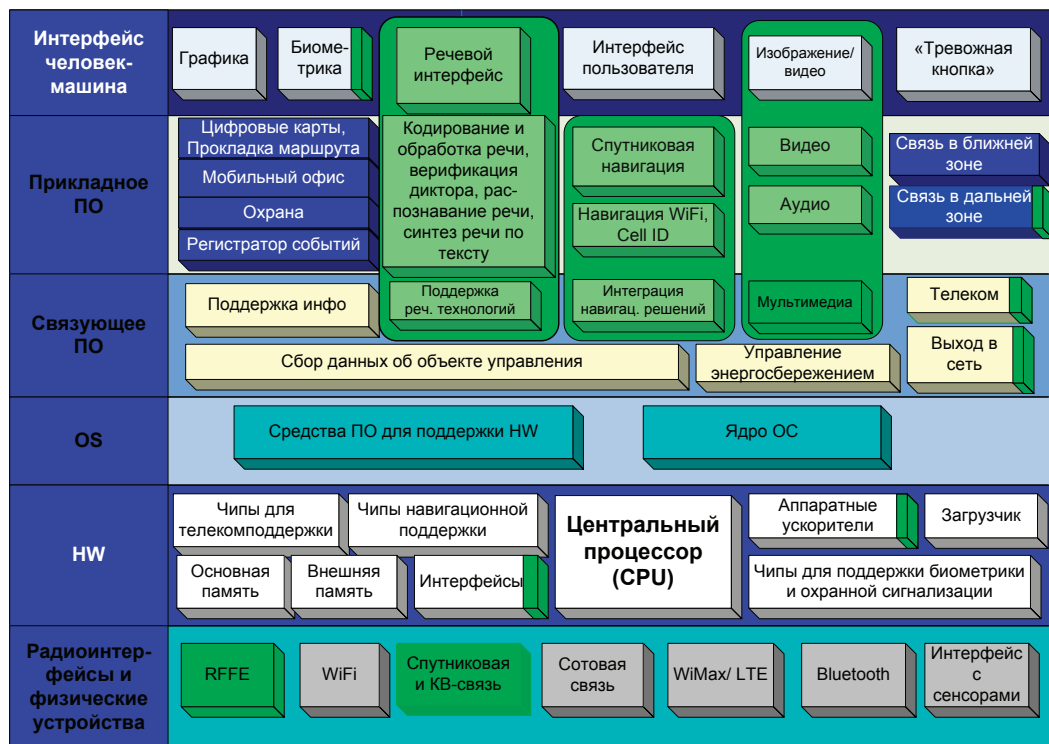


Рис. 2. Архитектура универсальной платформы ИКН-терминала на базе движков

Рассмотрим детально два важных движка в составе прикладного ПО, которые обязательны и весьма специфичны для ИКН-терминала. В их число входит речевой движок, обеспечивающий качественную речевую связь (особенно в режиме громкой связи) в цифровой форме между водителем (или пассажиром) и службами телематической системы, и навигационный движок.

Важность повышения качества речевой связи декларировалась не раз. Для этого необходимо обеспечить не только кодирование речи с высоким качеством (фактор MOS должен быть на уровне 3.8...4.5, что обеспечивает высокий уровень разборчивости и узнаваемости), но и подавления акустического эха и шумов. Примерно так же ставится задача при разработке речевого движка в системах VoIP и видеоконференсинга. Речевой и видеодвижок SPIRIT отвечает указанным требованиям и ориентирован на использование в подобных условиях. Его структура показана на рис. 3 (стр. 70) в составе клиентского и серверного ПО (решение TeamSpirit® [7]). Надстройкой является стек протоколов транспортного уровня, а аппаратной частью средства ввода/вывода звука и видео (Sound IO и Video IO).

Речевой движок реализован на разных платформах, включая и мобильные устройства, и лицензирован многим известными компаниями в мире, предлагающими рынку системы ВКС. Такой движок включает набор стандартных и проприетарных речевых кодеков для разных скоростей передачи (4.8 Кбит/с... 32 Кбит/с), включая специально спроектированный масштабируемый многоскоростной кодек IP-MR для сетей IP, обеспечивающий наивысшее качество передачи речи с учётом реальной текущей пропускной способности виртуального канала в такой сети. В состав речевого движка с API входит также подсистема улучшения качества речи, включающая адаптивный подавитель акустического эха (AEC), нелинейный процессор (NLP), подавитель шума (NS) на базе проприетарной технологии RealDuplex, подсистема нейтрализации таких

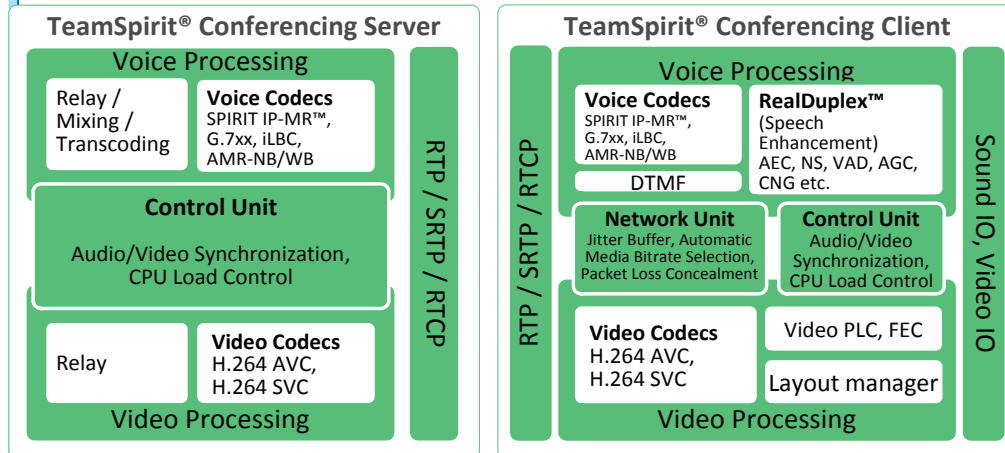


Рис. 3. Клиентское и серверное ПО для передачи речи и видео в сети IP

проблем, как задержки, «дрожания», потери пакетов, неизменно присутствующие в сетях пакетной коммутации, на базе технологии NetJet, а также АРУ для регулировки уровня сигнала, детектор активности речи (VAD), генератор комфортного шума (CNG) и модуль DTX, снижающие нагрузку виртуального канала в речевых паузах. Специальный модуль Wizard и синхронизатор тактовых частот Clock Drift Control на разных концах речевого канала обеспечивают наилучший набор параметров подавителя эха для согласования тактовых частот и создания наилучших акустических условий в режиме громкой связи, что важно для водителя и пассажиров ТС.

Такой движок входит в состав комплексного программного продукта TS Voice & Video Engine в виде SDK, отвечающего за передачу речи и видео в рамках многоточечной системы ВКС компании «ВидеоМост» [8].

Навигационная составляющая используется практически для всех приложений и особо для приложения «ЭРА Глонасс». Навигационный движок может быть в своей основной (baseband) цифровой части реализован программно на основе SDR-принципа или на базе host based подхода, когда наиболее ресурсоёмкая (корреляционная) обработка сигнала выполняется аппаратно [9]. СВЧ-часть движка (радиоинтерфейс) выполняется аппаратно всегда в силу своей специфики.

Надёжность позиционирования «всегда и везде» может быть обеспечена только интеграцией разных технологий, а именно — радиотехнологий (спутниковая навигация, позиционирование по WiFi или по вышкам сети сотовой связи Cell ID) и инерциальных средств (акселерометры, приборы MEMS). В условиях использования транспортных средств на первое место выдвигается спутниковый навигационный приёмник, который должен работать при разных положениях своей антенны, возможного отсутствия прямой видимости спутников, наличии помех. Этого, в первую очередь, можно достигнуть путём существенного повышения чувствительности приёмника для обеспечения работоспособности при сильном ослаблении радионавигационного поля за счёт сложных условий приёма (городской каньон, затенение, приём отражённого сигнала спутника от естественных или искусственных препятствий при опрокидывании автомобиля после ДТП и др.).

Дополнительная проблема при приёме ослабленного радиосигнала — влияние электромагнитных помех (EMI), излучаемых работающим электронным или электрическим оборудованием. Оно особенно заметно при весьма ослабленных полезных радиосигналах от навигационных спутников. Поэтому повышение чувствительности спутникового приёмника и борьба с помехами должны обеспечиваться в комплексе.

Для универсальной платформы с мощным CPU предлагается использовать программный GNSS-приёмник, обеспечивающий работу с сильно ослабленными радиосигналами благодаря его сверхчувствительности [10]. Он включает аппаратную часть с радиочастотным фронт-эндом (RFFE) и интерфейсом с CPU и программную часть, выполняющую главные функции обработки сигнала в основной полосе (baseband) вплоть до вывода PVT-данных. Этот подход базируется на принципах программного радио (Software Defined Radio — SDR), который становится всё более привлекательным по мере роста вычислительной мощности универсальных процессоров.

Идеи повышения чувствительности приёмного устройства хорошо известны. Это параллельная обработка сигнала банком корреляторов, плотно расставленных на плоскости «время-частота», и когерентная или некогерентная интеграция (накопление) принимаемого сигнала. Повышение числа корреляторов (до десятков и сотен тысяч) снижает время поиска сигналов и оставляет больше времени для интеграции, т.е. повышает тем самым чувствительность приёмника. Время поиска может быть дополнительно снижено путём уменьшения окна поиска на указанной плоскости при получении внешней информации о частоте и времени. Такой «аппаратный» подход используется при проектировании современных чипов или чипсетов, реализующих функции GNSS-приёмника.

Накопление — обязательный компонент обработки сигнала в приёмнике. Когерентное накопление даёт лучший эффект в повышении чувствительности приёмника, но более ресурсоёмко, а некогерентное проще реализовать. Можно сочетать когерентное и некогерентное накопление для компромисса между чувствительностью приёмника и сложностью его реализации. Именно такой подход реализован в сверхчувствительном приёмнике (в его baseband-части) программными средствами на CPU Intel Atom или Core iX, которые используют специфику архитектуры процессора. Дополнительно программным способом подавляются электромагнитные помехи. Получаемая таким путём увеличенная чувствительность зависит от времени накопления. При этом RFFE обеспечивает перенос спектра высокочастотного сигнала в область низкой промежуточной частоты (ПЧ) и оцифровку отсчётов ПЧ-сигнала в аналого-цифровом преобразователе (АЦП). Для ввода оцифрованных ПЧ-отсчётов в baseband-часть приёмника его RFFE в настоящее время представлен в виде USB-донгла (ВЧ-адаптера). Альтернативным решением реализации «RFFE+интерфейс» может быть размещение RF-чипа на материнской плате терминала.

Указанная алгоритмическая специфика проявляется приёмником при работе с сильно ослабленным радиосигналом на его входе, что автоматически программно оценивается. При приёме спутникового радиосигнала вне зоны его ослабления нет необходимости поддерживать высокую чувствительность, так как в этом случае обеспечивается прямая видимость спутников. Если уровень мощности принимаемого сигнала достаточно высок, нет необходимости проводить долгое накопление. Это существенно уменьшает требование к динамике подвижного объекта, т.е. в условиях высокой динамики, а такая ситуация характерна для объектов на просторе, вне укрытий, интенсивность радиосигнала относительно велика и время его накопления целесообразно снизить, тем самым обеспечив условия получения PVT-данных для высокой динамики объекта. В этой ситуации можно исключить подавление электромагнитных помех, так как при сильном сигнале их влиянием можно пренебречь. Это позволяет снизить нагрузку на процессор.



Характеристики описываемого приёмника:

- 32 канала и более;
- точность позиционирования: 3 м (СЕР) в автономном и 1 м в дифференциальном режиме;
- чувствительность в режиме холодного старта до — 159 dBm;
- чувствительность в режиме горячего старта (< 1сек.) и в режиме A-GNSS до — 164 dBm;
- чувствительность в режиме поиска сигнала до — 166 dBm;
- чувствительность в режиме слежения до — 170 dBm;
- возможные режимы: ГЛОНАСС+GPS, только GPS, только ГЛОНАСС;
- поддержка Internet aided A-GNSS.
- Возможное обеспечение приёма сигналов Galileo.

Очевидно, что приём навигационных радиосигналов от спутников разных созвездий (GNSS) существенно повышает надёжность навигации и точность позиционирования в условиях сильного ослабления мощности принимаемых сигналов из-за возможности выбора лучшей геометрии по комбинации сопровождаемых спутников. Именно поэтому многосистемность для такого приёмника — важный фактор улучшения его основных технических характеристик.

Для того чтобы обеспечить позиционирование ТС в любых сложных условиях, целесообразно интегрировать различные технологии позиционирования. В полной мере это реализуется в рамках концепции объединения радиотехнологий позиционирования (спутниковая навигация, WiFi, Cell ID и др.) и инерциальных устройств (акселерометров, MEMS, IMU), причём объединяющим центром такого гибридного навигационного движка (Hybrid Location Engine) является программный приёмник.

В заключение укажем, какое готовое прикладное программное обеспечение, помимо рассмотренного выше навигационного ПО, может обеспечить реализацию важных функций универсальной платформы не традиционно аппаратными, а программными средствами. Это телекоммуникационное ПО, разработанное на базе SDR-подхода. Сюда включаются спутниковая и КВ-связь, обеспечивающие глобальное радиопокрытие [11], речевые технологии (распознавание и синтез речи, верификация диктора, подавление акустического эха, одно- и многомикрофонное шумоподавление, АРУ, детекторы речевой активности, линейка проприетарных и стандартных речевых кодеков и др.), аудиопроцессинг (аудиокодеки MP3, AAC, AAC+, OGG, WMA, эквалайзинг и др.), видеопроцессинг (видеокодеки H.263, H.264, MPEG 2, MPEG 4 и др.). Всё это ПО разработано и апробировано его лицензиатами. Оно также использовано в качестве компонент в системах VoIP и видеоконференсинга, включая доставку мультимедийных потоков мобильным пользователям.

Интеграция информационных, коммуникационных и навигационных технологий в рамках единой платформы для транспортных средств и поддержка продвинутых программных решений по видеоконференцсвязи, где некоторые абоненты могут быть мобильными, — это уже ближайшее будущее телематических систем.

Литература

1. Свириденко В.А. , Будник Р.А. Платформа и прикладное ПО для смарт-терминалов телематических систем. Т-Comm. № 2. 2011. С. 30–34.
2. NXP Automotive Telematics On-board unit Platform (ATOP).
3. Microsoft Corporation, MS Auto Platform Overview.
<http://download.microsoft.com/download/6/5/0/6505FA0E-1F39-4A34-BDC9-A655A5D3D2DB/MicrosoftAutoPlatformOverview3%201.pdf>
4. CSR launches SiRFprimaAuto location platform SoC. <http://johndayautomotiveelectronics.com/?p=4199>
5. Suresh Marisetty et al. Low Power Intel Architecture Platform for In-Vehicle Infotainment. Intel Technology Journal, v.13, 2009.
6. Р.А. Будник, В.А. Свириденко. Универсальная компьютерная платформа для информационно-коммуникационно-навигационных терминалов и её программная составляющая для различных приложений. Доклад на конференции «Оборудование спутниковой навигации, модули и электронные компоненты», ChipExpo-2010.
7. <http://www.spiritdsp.com/products/voice-video-engine-mobile/>
8. <http://www.videomost.com/products/>
9. Беркович Г.М. , Будник Р.А. , Свириденко В.А. Программно-аппаратные навигационные решения для мобильных устройств. Сборник трудов МЭС-2010. С. 227–232.
10. http://www.spirit-telecom.ru/super_sensitive_receiver.html
11. В.А. Свириденко, П.В. Смирнов. Навигационно-коммуникационный комплекс для магистральной связи. Первая миля. 2008. № 5. С. 34–39.

Сведения об авторе

Свириденко Владимир Александрович —

доктор технических наук, профессор, технический директор ГК SPIRIT DSP, автор ряда книг, монографий и учебных пособий в области сжатия данных, передачи и распределения информации, моделирования систем, автор и соавтор более 100 статей и более 30 авторских свидетельств и патентов.

При его непосредственном участии инженерное подразделение SPIRIT DSP в течение уже почти двух десятилетий специализируется в области разработки систем голосового (VoIP) и видеоконференсинга, проектирования речевых, аудио- и видеодвижков (engines), включая и наиболее важные их компоненты, речевых технологий (кодирования и обработки речевых сигналов, распознавания речи, верификации диктора, модификации голоса и др.), ЦОС, цифрового радио, но также телекоммуникационных систем и спутниковых навигационных приемников. Большинство продуктов ГК SPIRIT DSP отмечено различными премиями и почти все они лицензированы многими известными IT- и полупроводниковыми компаниями, которые выпускают более 60% смартфонов в мире.



Новый подход к определению границ речевого сигнала. Проблемы конца сигнала

*Шелепов В.Ю., доктор физико-математических наук,
профессор*

Ниценко А.В., программист

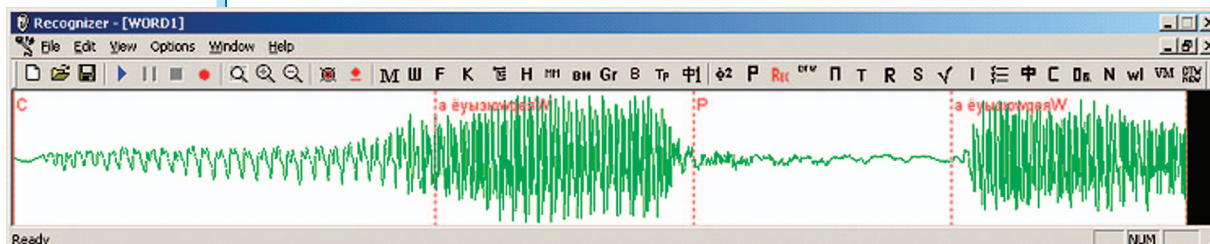
В статье предлагается новый метод записи и выделения глухих взрывных согласных в конце слова, определения наличия звонкого согласного в конце постепенно затухающего сигнала. Инструменты: вариация с переменным верхним пределом по модулю 256, последовательные сглаживания.

• *глухой взрывной* • *звонкий согласный в конце слова* • *вариация с переменным верхним пределом* • *сглаживание.*

The content of the article is new method of recording and detachment toneless stop consonant in the end of the word, presence of voiced consonant in the end of gradually damped signal clarification. Instruments: variable upper boundary variation (module 256), smoothening.

• *toneless stop consonant* • *voiced consonant in the end of the word* • *variable upper boundary variation* • *smoothening.*

При записи речевого сигнала с целью распознавания речи требуется как можно более точное определение начала и конца записываемого речевого отрезка. Недопустимо, например, чтобы участок какого-либо низкоамплитудного звука речи оказался целиком «отрезанным». В связи с этим возникают так называемые проблемы конца сигнала. Первая проблема — запись и выделение в конце сигнала участков, отвечающих глухим взрывным звукам [К], [П], [Т]. Эти звуки являются паузообразными. Поскольку при их произнесении есть момент полного перекрытия голосового тракта, и при этом голосовые связки молчат, в сигнале появляется характерный паузообразный сегмент. В качестве примера приведём визуализацию сигнала, отвечающего слову «лапа» (рис. 1).



Добиться того, чтобы при записи речевого сигнала подобный сегмент не обрезался, если он окажется в конце, — непростая задача.

Вторая проблема конца сигнала возникает из ситуации, иллюстрируемой рис. 2. На нём приведена визуализация записи слова «она».



Рис. 2. Визуализация записи слова «она»

Слово произнесено через некоторое время после включения записи. Она остановлена также через некоторое время после окончания речи. Рисунок отражает тот факт, что речевой сигнал всегда затухает постепенно, при этом значительно медленнее, чем нарастает вначале. Поэтому при автоматической сегментации¹ программа может ошибочно определить в конце наличие **звонкого** согласного звука, которого на самом деле нет. Это побуждало нас долгое время отказываться от распознавания **звонких** согласных в конце слова, вплоть до исключения этих звуков из автоматически создаваемой транскрипции слов распознаваемого словаря.

В данной работе предлагается механизм решения этих проблем. Он основан на использовании величины $W(n)$, введённой в работе [1] при определении понятия вариационной меры. Напомним определение этой величины. Мы работаем с 8-битной записью при частоте дискретизации 22050 Гц, в условиях отсутствия существенного внешнего шума. Пусть x_0, x_1, \dots — последовательные отсчёты записанного сигнала.

Рассматривается численный аналог полной вариации с переменным верхним пределом

$$V(0) = 0, \quad V(n) = \sum_{i=0}^{n-1} |x_{i+1} - x_i|$$

Пусть N_1 — максимальное число такое, что $V(N_1) \leq 255$. Полагаем

$$W(n) = V(n) \quad \text{при } 0 \leq n \leq N_1,$$

$$W(n) = V(n) - 256, \quad \text{при } N_1 + 1 \leq n \leq N_2, \quad \text{где}$$

N_2 — максимальное число такое, что $W(N_2) \leq 255$ и так далее.

Таким образом, $W(n)$ — кусочно-возрастающая в широком смысле слова функция, связанная с $V(n)$ условием

$$W(n) \stackrel{\text{mod}256}{=} V(n). \quad (1)$$

При построении $W(n)$ значения $V(n)$ «сбрасываются» вниз на 256 единиц, как только они начинают превышать 255.

На рис. 3,4 приведена визуализация сигнала, отвечающего слову «сумма» и функции $W(n)$, построенной по этому сигналу.

¹ Сегментация — разбиение сигнала на участки, отвечающие отдельным звукам речи.

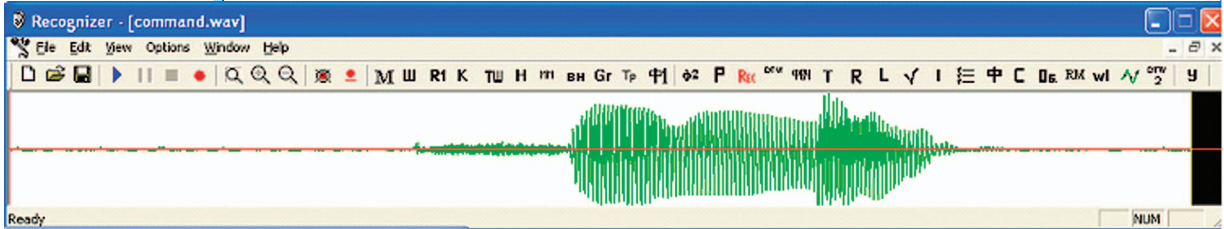
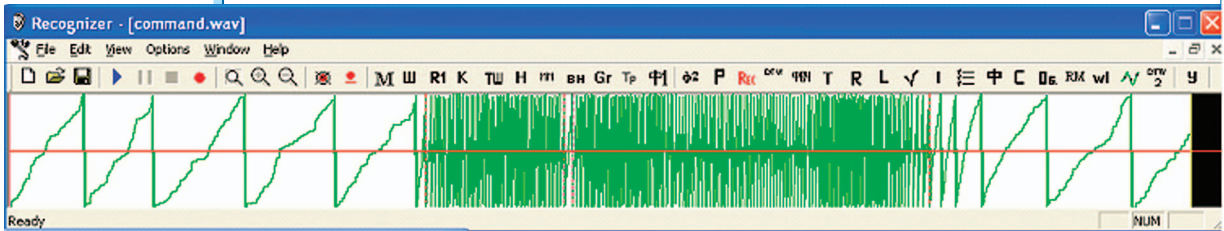


Рис. 3. Визуализация записи слова «сумма»

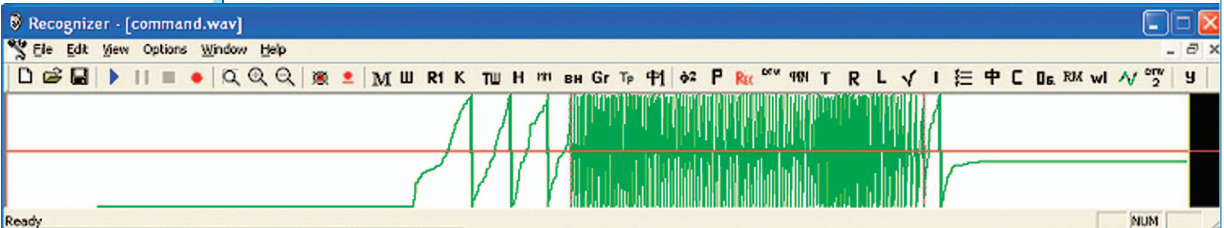
Рис. 4. Визуализация соответствующей функции $W(n)$

Участкам молчания до и после слова отвечают промежутки медленного возрастания $W(n)$ от 0 до 255 за счёт фонового шума звуковой карты.

Мы называем «сглаживанием сигнала» обработку его скользящим трёхточечным фильтром:

$$x_i = \frac{x_{i-1} + x_i + x_{i+1}}{3}.$$

Подвергнем исходный сигнал пятикратному сглаживанию. Тогда упомянутый фоновый шум **нивелируется**, и вместо сигнала на рис. 4 мы получим сигнал на рис. 5:

Рис. 5. Функция $W(n)$ для пятикратно сглаженного сигнала слова «сумма»

Теперь участкам молчания до и после речи соответствуют участки постоянства функции $W(n)$. Момент, когда при движении от начала записи вправо первая из этих постоянных превышает на p_1 единиц, будем считать соответствующим началом речи. Аналогично, двигаясь от конца записи влево и используя порог p_2 , определяем момент окончания речи. Пороги p_1 и p_2 определяются экспериментально в зависимости от уровня шума конкретной звуковой карты и микрофона. В наших условиях мы брали их равными 10. Проведённые многочисленные эксперименты показывают, что это весьма точное определение начала и конца речи. Очевидные ошибки чрезвычайно редки. При этом, если слово заканчивается одним из глухих взрывных звуков $[K]$, $[П]$, $[Т]$, перед концом речевого сигнала появляется достаточно длинный паузообразный участок, наличие которого позволяет

обнаруживать этот звук в конце слова. Например, визуализация записи слова «аббат» (рис. 6):

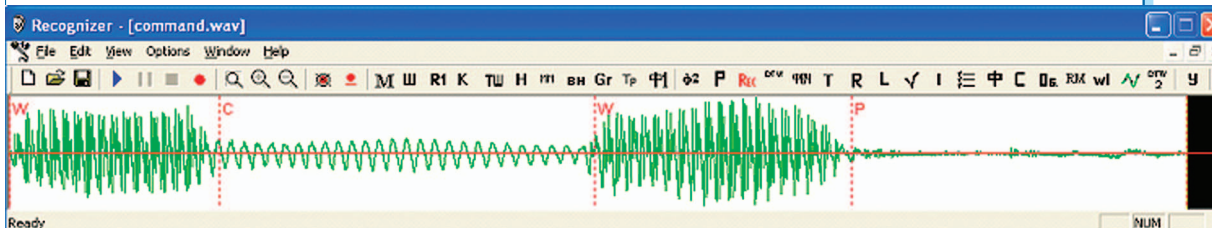


Рис. 6. Визуализация записи слова «аббат»

Всё сказанное о звуках [К], [П], [Т] в конце слова в равной мере относится к звукам Ф и Х, которые, **ввиду малой интенсивности также могут при распознавании восприниматься как паузообразные**. Результаты распространяются и на мягкие варианты всех этих звуков. Но следует учесть, что тогда возможно появление в конце ярко выраженной фрикативной части (как правило, у мягкого звука [Т], см. рис. 7). Однако эта часть существенно короче шипящих и свистящих звуков ([С], [Ш] и т.д.), что позволяет не приписать сигналу этих звуков по ошибке.

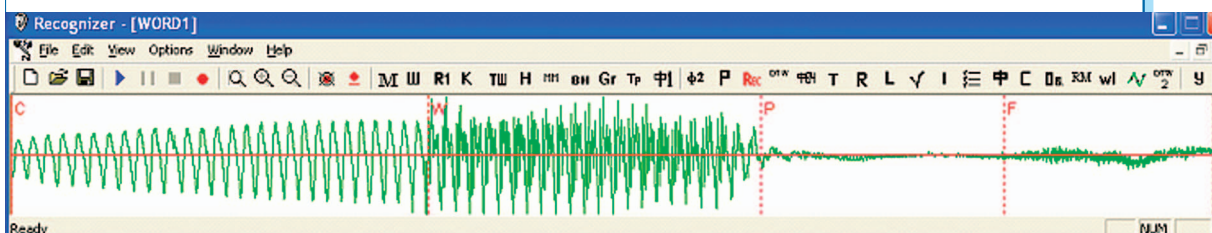


Рис. 7. Визуализация записи слова «мать»

Переходим ко второй вышеупомянутой проблеме конца сигнала. Пусть слово заканчивается голосовым звуком. Применим тот же алгоритм, заменив пятикратное сглаживание 50-кратным. Тогда заключительный участок затухания последнего голосового звука станет паузообразным, и метка конца, полученная с помощью сглаженного сигнала, передвинется влево. Этот заключительный участок, находящийся между новой меткой конца при 50-кратном сглаживании и первоначально полученной меткой конца при пятикратном сглаживании отбрасывается, и в окончательный буфер для визуализации и распознавания заносится укороченный таким образом сигнал. Как показывают эксперименты, это гарантирует при последующей сегментации от ошибочного выделения в конце слова голосового согласного звука. Алгоритм этой «последующей» сегментации предложен в работе [2], описание несколько усовершенствованного алгоритма содержится в [3]. На рис. 8 приведена визуализация записи слова «оса», где заключительный участок затухания сигнала выделен вышеописанным образом.

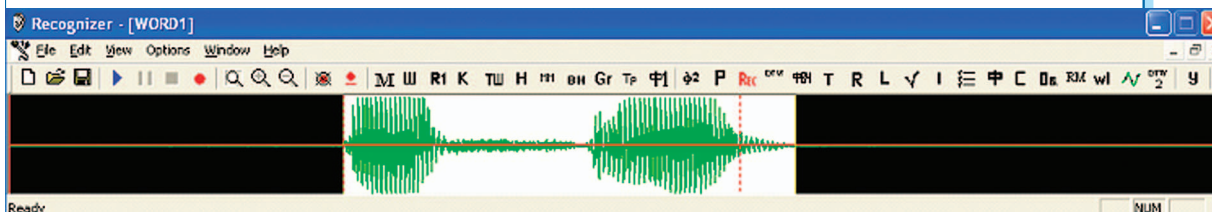


Рис. 8. Выделение заключительного участка затухания в слове «оса»



Все изложенные алгоритмы предполагают наличие сигнала, содержащего речевой отрезок и записанного с некоторым «запасом», т.е. с участками «молчания» слева и справа. Такую предварительную запись можно получить, следуя алгоритмам, изложенным в главе 1 книги [4]. При этом проверка на наличие речи с использованием квазипериодической структуры голосовых элементов позволяет осуществлять запись в автоматическом режиме. После включения пользователем начала записи компьютер будет всё время записывать один за другим слитные речевые отрезки, находясь в режиме непрерывной записи звука и ожидания речи.

Таким образом, достигается весьма эффективное решение обеих вышеописанных проблем конца сигнала.

Литература

1. Шелепов В.Ю., Ниценко А.В., Жук А.В., Азаренко Д.С. О распознавании фонем с помощью анализа речевого сигнала в частотной и временной областях. Приложение к распознаванию синтаксически связанных фраз // Речевые технологии. М., 2008. № 2. С. 43–52.
2. Шелепов В.Ю., Ниценко А.В. Структурная классификация слов русского языка. Новые алгоритмы сегментации речевого сигнала и распознавания некоторых классов фонем // Искусственный интеллект. 2007. № 1. С. 213–224.
3. Шелепов В.Ю., Ниценко А.В., Жук А.В. Построение системы голосового управления компьютером на примере задачи набора математических формул // Искусственный интеллект. 2010. № 3. С. 259–267.
4. Шелепов В.Ю. Лекции о распознавании речи. Донецк: ИПШ «Наука і освіта», 2009. 196 с.

Сведения об авторах

Шелепов Владислав Юрьевич —

доктор физико-математических наук, профессор, автор ряда работ в ведущих математических журналах СССР. С 1993 г. возглавлял отдел распознавания речевых образов Института проблем искусственного интеллекта НАН и МОН Украины. В настоящее время — заведующий кафедрой систем искусственного интеллекта Института информатики и искусственного интеллекта Донецкого национального технического университета.

Ниценко Артём Владимирович —

профессиональный программист, специализирующийся в области распознавания речи, автор многочисленных экспериментальных и прикладных программ.

Технология распознавания и озвучивания текстов для обеспечения учебного процесса людей с нарушениями зрения

Юрьев Г.А., аспирант Московского городского психолого-педагогического университета



Куравский Л.С., доктор технических наук, профессор



В статье рассматривается способ решения проблемы доступа к текстовой информации, возникающей у людей с нарушениями зрения. Текст — основной источник знаний в пространстве, которое принято называть «информационным». Несмотря на огромное количество электронных источников, в настоящее время существует распространённая ситуация, когда доступен только бумажный вариант текста. Далеко не всегда нужную информацию можно получить из книг, набранных Брайлевским шрифтом. В то же время людям с нарушениями зрения остаются недоступными обычные печатные материалы и часть информации на экранах компьютерных мониторов. Поэтому разработка программно-аппаратных средств, обеспечивающих для незрячих людей доступ к этим источникам, является в настоящее время чрезвычайно актуальной.

В этой статье представлена новая технология обработки текстов для людей с нарушениями зрения, интегрирующая средства сканирования, распознавания и озвучивания.

• распознавание образов • нейронные сети • вейвлет-преобразования • сети Хэмминга • восстановление изображений • цепи Маркова.

The technology intended for recognition of text information by visually impaired persons is under consideration. Participation of such specialists in diagnostic works is frequently necessary because of social reasons and entails relevant technical support. The presented technology is intended for transformation of initial text images obtained online with the



aid of a video camera from computer monitors and similar sources to their audio representation. Initial images after their acquisition are converted into corresponding contour shapes by a proper wavelet transform and, then, decomposed into character strings, which are recognized by means of a set of techniques, one of which combines capabilities of wavelet transforms and relaxation neural networks. Recognized character strings are then spoken by a speech synthesizer. Features of the implemented hardware-software system are given.

- *pattern recognition* • *neural networks* • *wavelet transform* • *Hamming network*
- *restoring images*.

Введение

Лёгкий доступ к информации стал неотъемлемой частью жизни. Для большинства людей получение информации не представляет труда. Однако слабовидящие люди лишены возможности пользоваться информацией в полном объёме. Проблема доступа к текстовой информации — одна из наиболее значимых в процессе адаптации [1] людей с нарушениями зрения в современной компьютеризированной среде. В настоящее время большое количество художественной литературы появляется в аудиоварианте. Однако ни публицистика, ни большая часть технической литературы в аудиоформате не выходят. У издательств в этом нет особенной необходимости, так как аудитория подобных проектов была бы слишком мала. Но возможность для слабовидящих, например, в индивидуальном порядке прослушать газету, журнал, техническую инструкцию может быть весьма ценной.

В этой статье представлена новая технология обработки текстов для людей с нарушениями зрения, интегрирующая средства сканирования, распознавания и озвучивания. Она может быть использована для чтения литературы, изданной традиционным печатным способом, и работы за экраном компьютера (здесь и далее подразумевается любой жидкокристаллический дисплей) с обычным программным обеспечением, не предназначенным для незрячих пользователей.

Решение проблемы чтения именно плоскочечатного текста обеспечивает доступ к образовательным ресурсам, повышает эффективность процесса обучения и позволяет реализовать профессиональные навыки. Распространённые традиционные средства доступа, включая азбуку Брайля, мультимедийные книги и программное обеспечение для озвучивания представленных в электронной форме текстов, частично решают указанную проблему. Эти средства требуют значительных усилий, затрачиваемых на предварительную подготовку исходного материала и перевода его в доступную для восприятия брайлевскую или электронную форму представления. Технология обеспечивает возможность считывания текста с бумажных и других носителей при минимальных требованиях к аппаратной поддержке, что повышает мобильность и доступность системы. Разработанное программное обеспечение включает развитые средства устранения ошибок, обусловленных низким качеством обрабатываемого изображения.

Из средств воспроизведения текстов, предназначенных для людей с нарушениями зрения, в настоящее время доступны:

- системы оптического распознавания текстов FineReader [6], CuneiForm [8] и другие, позволяющие озвучивать занесённые в память компьютера отсканированные тексты с бумажных источников;
- системы невидимого доступа JAWS [7], Adriane Knoppix [9] и их аналоги, озвучивающие представленные в электронной форме фрагменты текстов и имеющие развитую систему навигации по ключевым словам;
- звуковые книги, подготовленные в специальных аудиоформатах.

Однако стоит учесть, что:

- система FineReader, предназначенная для работы со сканированной информацией, не обеспечивает нужный темп и гибкость работы с текстами, напечатанными на бумаге, и совершенно не приспособлена к озвучиванию текстов, отображаемых на экране компьютерного монитора и других недоступных сканированию поверхностей;
- система JAWS не работает с печатными текстами и текстами, представленными в графических форматах в виде рисунка;
- звуковые книги требуют предварительной аудиозаписи речи в студии и её последующей трудоёмкой разметки.

Учитывая указанные ограничения, технология озвучивания плоскочечатных и компьютерных текстов, работающая в режиме, задаваемом самим пользователем, поддерживает возможности, не реализованные в настоящее время ни в одной из существующих программных систем.

Из известных аппаратных решений, близких по тематике, на рынке присутствует IRISPen. Это устройство — маркер по форме — подключается к компьютеру через USB интерфейс. С одной стороны устройства расположена камера. Предполагается, что пользователь ведёт маркер по строке, в процессе его движения отсканированный текст транслируется на выбранный язык встроенной системой распознавания и перевода. Компания позиционирует устройство как предмет для перевода цитат информации с визитных карточек, т.е. работа с большими массивами данных не предусматривается. Очевидно, для использования этой системы необходимо видеть строки, поэтому такой вариант неприемлем для людей с нарушениями зрения.

Аналогичным по функциональности является устройство Intel Reader. Однако оно обладает сравнительно высокой стоимостью, на данный момент не поддерживает русский язык и позиционируется, скорее, как разработка, позволяющая реализовать чтение книжек вслух для людей без ограничений по зрению, что в свою очередь сказывается на управлении прибором.

Основные компоненты технологии распознавания

Рассматриваемая технология распознавания и озвучивания текстов представлена на рис. 1. Она содержит три основных компонента:

- предварительную обработку изображения;
- распознавание символов;
- озвучивание распознанного текста.

Предварительная обработка изображения включает:

- ввод изображения в одном из стандартных графических форматов;
- преобразование изображения к монохромному представлению;
- определение контуров рисунка (преобразование к бинарному представлению);
- распознавание границ текстовых строк;
- определение прямоугольных фрагментов изображения, занимаемых символами в строке, включая составление списка их геометрических характеристик.

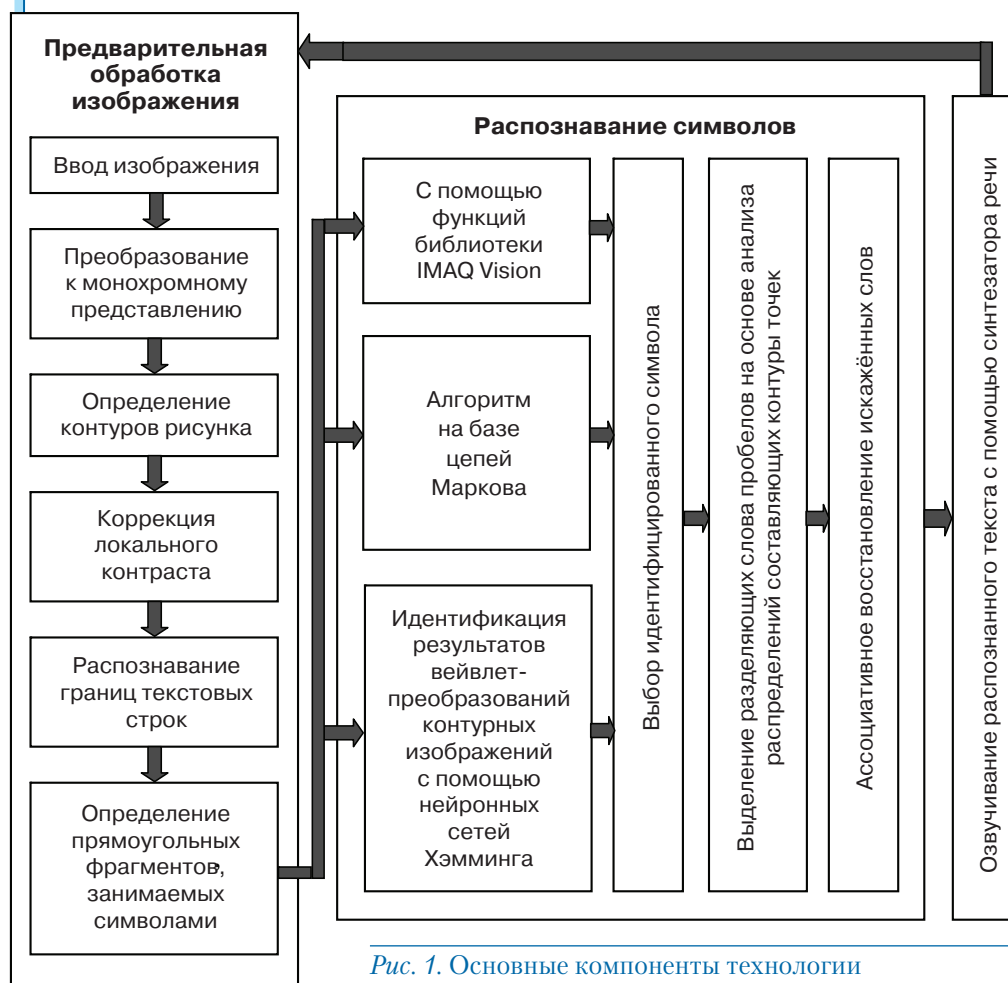


Рис. 1. Основные компоненты технологии распознавания и озвучивания плоскопечатных текстов

Для повышения надёжности распознавание символов производится независимо тремя различными способами:

- с помощью функций для обработки изображений из библиотеки IMAQ Vision, входящей в среду графического программирования LabVIEW [5] (этот встроенный набор функций в основном ориентирован на распознавание символов, включённых в штрихкоды, что делает механизм недостаточно эффективным при использовании на изображениях с низким разрешением и высоким уровнем помех);
- с применением нового алгоритма, использующего возможности цепей Маркова;
- путём идентификации результатов вейвлет-преобразований контурных символьных изображений с помощью релаксационных нейронных сетей Хэмминга [2–4]. Этот метод показал наивысшую эффективность при работе с сильно зашумлёнными изображениями, что связано как с особенностью вейвлет-преобразований, так и с вероятностной природой релаксационных нейронных сетей.

Подобное голосование уменьшает вероятность ложного распознавания. Символы, которые будут признаны нераспознанными, позже восстановит специальный механизм, опирающийся на словарь, что является стандартным методом в подобном случае.

Важными подзадачами, решаемыми в процессе распознавания текстов, являются:

- выделение пробелов;
- ассоциативное восстановление слов.

Выделение разделяющих слова пробелов происходит на основе анализа распределений, составляющих контуры точек. Восстанавливать искажённые слова приходится вследствие ошибок при идентификации символов.

Особенности программной реализации и практического использования

Программная реализация выполнена в среде графического программирования LabVIEW. Использовались стандартные виртуальные инструменты для анализа данных, функции для обработки изображений из библиотеки IMAQ Vision, а также ряд процедур обработки изображений, выполненных в среде Borland Delphi и интегрированных в LabVIEW в форме динамически подключаемых библиотек.

Стандартный вариант применения рассматриваемой технологии предполагает сканирование текста с помощью веб-камеры, инициализация которой происходит в лучшем из доступных режимов¹. Использование веб-камеры в качестве универсального считывающего устройства обусловлено её доступностью, компактностью, простотой и однотипностью в управлении на программном уровне. Для работы со всеми мультимедийными устройствами в Windows системах используется интерфейс программирования приложений DirectShow. Интерфейс USB камер поддерживается на большей части портативных компьютеров. В нашей жизни всё большее распространение получают т.н. субноуты. Они обладают низким электропотреблением, которое обеспечивает большее время работы, и маленькими габаритами, что делает их ещё более мобильными по сравнению с другими портативными компьютерами. Совершенно очевидно, что их технические характеристики полностью удовлетворяют требованиям рассматриваемой системы, что делает доступной простому пользователю.

Для установления связи с веб-камерой используется DLL, работающий через DirectShow. В DirectShow включены функции активации устройства, функция запуска его в предпочтительном режиме, функция захвата изображения и функция завершения сеанса работы, что вполне достаточно для решения задачи. Аналогичная задача может быть решена с использованием компонентов IMAQ Vision, но в силу большей универсальности этот инструмент менее производителен, хотя опирается на абсолютно такие же принципы.

Введённое изображение преобразуется в монохромное чёрно-белое представление. Именно в таком виде наиболее удобны любые операции распознавания и анализа изображений. Для работы релаксационной нейронной сети символ переводится в контурный рисунок с помощью непрореженного вейвлет-преобразования на базе биортогонального вейвлета (рис. 2).

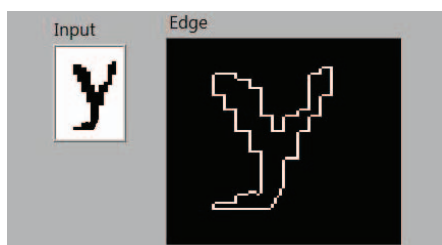


Рис. 2. Преобразование монохромного представления символа в контурный рисунок

¹ Как правило, распознавание проводилось с разрешением 320x240 и глубиной цвета 24bit.



Важной операцией, необходимой для корректной реконструкции связанного текста, является выделение из изображения символьных строк. Задача усложняется тем, что камера, как правило, удерживается в руке и, вследствие этого, строки попадают в кадр повёрнутыми. Для повышения надёжности распознавания при больших углах поворота необходимо восстанавливать горизонтальное положение строк. Для этого на основе анализа плотностей чёрных и белых точек, на нескольких вертикальных срезах изображения определяют местоположение начал и концов строк. Далее вычисляется средняя разность между значениями на срезах. Таким образом, мы получаем среднее расстояние, на которое конец строки смещён относительно начальной позиции. Последний подготовительный шаг — вычисление расстояния в горизонтальной плоскости. Исходя из полученных данных, возможно построить прямоугольный треугольник с известными сторонами и определить углы между катетами и гипотенузой. Угол α на рис. 3 является искомым, после его вычисления изображение поворачивается (рис. 3, 4).

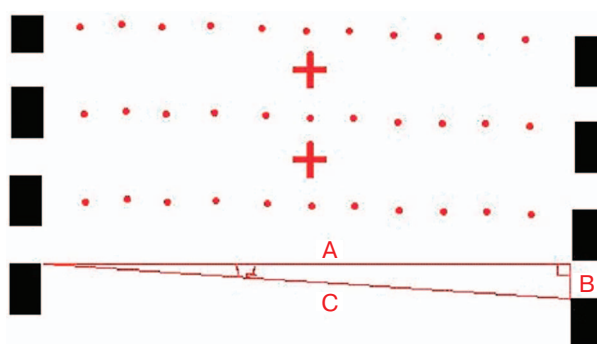


Рис. 3. Оценка расположения границ и угла поворота текстовых строк

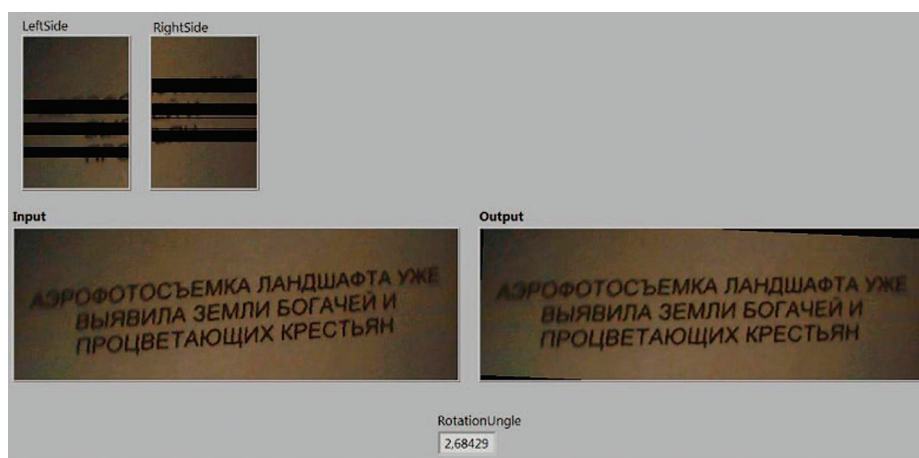


Рис. 4. Поворот тестовых строк

Поскольку для всех вычислений используются средние значения, метод не позволяет сделать строки параллельными границам изображения, но для работы остальных алгоритмов результат вполне приемлем. Ещё одним минусом подобной технологии является то, что в кадр обязательно должно попасть минимум две строки. Если это условие не выполняется,

изображение остаётся неизменным. Стоит отметить, что если строка обрывается на середине, это не влияет на вычисление в целом в силу того, что срезы делаются более чем на двух участках изображения. После такого поворота границы текстовых строк определяются повторно. На основе информации об этих границах фиксируются занимаемые символами прямоугольные фрагменты изображения (рис. 5).

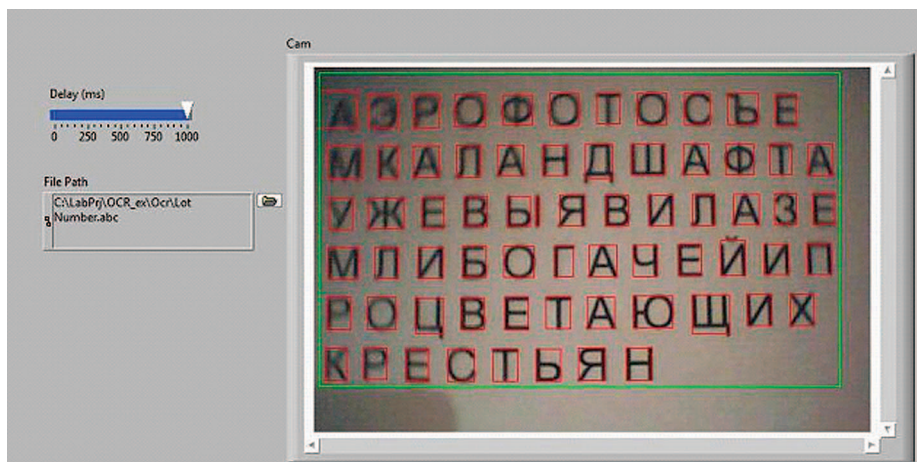


Рис. 5. Фиксация прямоугольных фрагментов изображения, занимаемых символами

По окончании выделения указанных прямоугольных фрагментов для всех символов в строке, соответствующий ей участок вырезается из кадра, просматриваемого веб-камерой, при этом малые фрагменты, не превышающие заданный порог, удаляются из строки. Оставшиеся фрагменты рассматриваются как области, содержащие распознаваемые символы произвольного размера. Указанная процедура повторяется для всех строк из кадра, просматриваемого веб-камерой. Информация о геометрических характеристиках строк и прямоугольных фрагментов изображения, включающих в себя символы, сохраняется для последующего распознавания в порядке расположения фрагментов в кадре.

Ослабить искажения в анализируемой части изображения позволяет алгоритм коррекции локального контраста. Этот приём позволяет уменьшить эффект, возникающий в связи с неравномерным освещением. При переводе в чёрно-белый формат невозможно установить общий уровень контраста для всего изображения, так как либо часть символов распадётся, и распознать их будет невозможно, либо некоторые символы будут закрашены чёрным цветом. Суть метода заключается в следующем: на выделенных прямоугольных фрагментах перебираются все чёрные точки, из числа которых удаляются слабосвязанные (имеющие малое число одноцветных соседей). В то же время области, в которых предполагается исчезновение чёрных точек из-за нарушения освещённости или иных условий сканирования, заливаются чёрным цветом². Указанный приём улучшает качество распознавания примерно на 10%, практически не влияя на скорость работы программы.

Изображение, вводимое с веб-камеры, обычно затемнено с одного или нескольких краёв, и несмотря на все использованные алгоритмы, с таким уровнем затемнения бороться бессмысленно. Эффект обусловлен непараллельностью поверхности текста и линзы, а также оптическими искажениями и недостаточным освещением. В случае значительных искажений анализируется только центральная часть графического представления.

² Примером может служить одна белая точка, все соседи которой в некоторой окрестности являются чёрными.



Один из реализованных способов распознавания символов опирается на возможности функций для обработки изображений, входящих в библиотеку IMAQ Vision среды графического программирования LabVIEW. После предварительного обучения на образцах пользователя он позволяет проводить идентификацию содержимого, сформированной ранее последовательности, прямоугольных фрагментов изображения. Результатом анализа является строка текста без пробелов, с обозначенными нераспознанными позициями.

Для обеспечения надёжности распознавания текста в особенно неблагоприятных условиях, характерных для рассматриваемой задачи, потребовалась реализация двух дополнительных независимых способов распознавания.

Первый опирается на возможности цепей Маркова.

Каждый прямоугольник с символом переводится в контурный рисунок с помощью непрореженного вейвлет-преобразования на базе биортогонального вейвлета, прямоугольные фрагменты с контурными изображениями последовательно преобразуются в два числовых вектора. Один из них описывает распределение точек контура вдоль оси абсцисс, другой — вдоль оси ординат. Таким образом, теряя часть информации, получаем простой инвариант длиной в 64 числа (символы отмасштабированы для окна 32x32). Такой вектор может быть проанализирован с удовлетворяющей для задачи скоростью. Полученный вектор подвергается быстрому вейвлет-преобразованию с использованием вейвлетов Добеши 4-го порядка. Полученные подмножества последовательностей вейвлет-коэффициентов $\mathbf{W}=\{w_{ij}\}_{i=1,\dots,N}$ рассматриваются как представления символов и анализируются с помощью распознающей цепи Маркова, структура которой представлена на рис. 6. Процедура распознавания основана на анализе знаков вейвлет-коэффициентов из указанной последовательности.

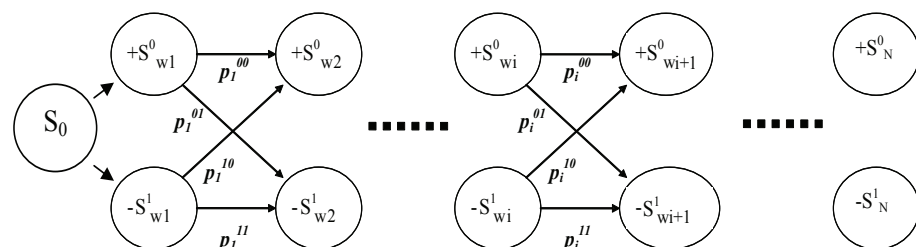


Рис. 6. Структура цепи Маркова

Система, описываемая данной цепью, функционирует N моментов времени, где N — длина анализируемой последовательности вейвлет-коэффициентов. В начальный момент времени система находится в состоянии S_0 . В i -й момент времени, где i изменяется от 1 до N , система может быть или в состоянии $+S_i^0$, или в состоянии $-S_i^1$, а именно: в состоянии $+S_i^0$ — в случае неотрицательности i -го элемента последовательности \mathbf{W} и в состоянии $-S_i^1$ — в случае его отрицательности. Система является нестационарной: изменяется матрица $\mathbf{P}_i = \|p_{ij}\|$ вероятностей переходов между состояниями, такая, что $\mathbf{p}_i = \mathbf{P}_i \mathbf{p}_{i-1}$, где \mathbf{p}_i — вектор-столбец вероятностей пребывания в состояниях цепи в i -й момент времени, причём в указанный момент времени все её элементы равны нулю, кроме вероятностей переходов между парами состояний S_{i-1}^1 и S_i^1 , S_{i-1}^0 и S_i^1 , S_{i-1}^1 и S_i^0 , и S_{i-1}^0 и S_i^0 , которые далее обозначены, соответственно, как p_i^{11} , p_i^{01} , p_i^{10} и p_i^{00} . В каждый из моментов времени i выполняются следующие нормировочные условия: $p_i^{11} + p_i^{10} = 1$ и $p_i^{01} + p_i^{00} = 1$.

Последовательности \mathbf{W}_a , соответствующей заданному символу a , приведённые правила ставят в соответствие определённую последовательность \mathbf{S}_a прохождения состояний рассматриваемой цепи Маркова³.

При обучении цепи Маркова на алфавите $\mathbf{A}=\{a_m\}$ для каждого символа a_m формируется множество его эталонов⁴ a_{mk} , $k=1, \dots, K_m$. Каждому символу a_m ставится в соответствие цепь Маркова, вероятности переходов, между состояниями которой определяются методом максимального правдоподобия. Для этого вычисляется распределение вероятностей переходов, обеспечивающее наибольшую вероятность P появления заданного множества последовательностей состояний, соответствующих эталонам данного символа: $P = \prod_{k=1, \dots, K_m} P_k$, где $P_k = \prod_{i=1, \dots, N} p_i^{\text{sign}W_{i-1,k} \text{ sign}W_{i,k}}$, где k — номер эталона символа, а i — номер состояния, причём полагается, что $\text{sign}w_{0,k}=1$.

Указанная задача может быть решена одним из подходящих численных методов нелинейного программирования⁵. С целью упрощения решения удобно перейти к критерию $\ln P$, монотонно связанному с исходным критерием P , переформулировав исходную постановку в задачу оптимизации для величин $\ln p_i^{\text{sign}W_{i-1,k} \text{ sign}W_{i,k}}$, на которые наложено условие неположительности⁶.

Распознавание поступающих на вход системы новых символов возможно после завершения обучения. При этом для цепи Маркова, соответствующей каждому из символов алфавита \mathbf{A} , вычисляется вероятность прохождения последовательности состояний, определяемой вейвлет-коэффициентами $\{w_j\}_{j=1, \dots, N}$ распознаваемого символа, которая равна $\prod_{i=1, \dots, N} p_i^{\text{sign}W_{i-1} \text{ sign}W_i}$. Тип символа, который соответствует цепи Маркова, имеющей наибольшую вероятность такого рода, выдаётся как результат распознавания.

Оценка вероятности корректного распознавания символов с помощью описанного алгоритма, полученная по результатам 1000 проб на базе асимптотической аппроксимации биномиального распределения, равна 0,791, причём нижняя и верхняя границы 95% доверительного интервала есть 0,76 и 0,83.

Оказалось, что достаточно использовать последовательности \mathbf{W} , состоящие из двенадцати вейвлет-коэффициентов ($N = 12$), начиная с третьего коэффициента разложения распределений контурных точек. Расчёты выявили неинформативность первых двух вейвлет-коэффициентов этого разложения для решения задачи распознавания.

Второй способ реализует новый алгоритм [2–4], использующий возможности вейвлет-преобразований и релаксационных нейронных сетей. Он способен эффективно работать после обучения на ограниченном числе образцов (рис. 7 стр. 88).

В этом методе используется сеть Хэмминга. Она требует значительно меньших затрат по сравнению с сетью Хопфилда, часто используемой в подобных случаях. При решении поставленной задачи нет необходимости восстанавливать образец по входам сети, т.е. нужен только номер предполагаемого образца. Таким образом, вместо классического варианта с ассоциативной памятью выбран алгоритм с релаксационной нейронной сетью. Суть данной структуры заключается в поиске хэммингова расстояния⁷ от представленного образца до образца из обучающей выборки. Образец, до которого такое расстояние окажется наименьшим, признаётся искомым. Так же характерным признаком сети является то, что нейроны её выходного слоя связаны между собой дополнительными отрицательными обратными синаптическими связями. Единственный синапс, с положительной обратной связью для каждого нейрона,

³ Следует заметить, что рассматриваемая цепь Маркова может быть описана и в терминах конечного автомата.

⁴ Эталоны, как правило, соответствуют различным вариантам изображения символов, принятым в разных шрифтах.

⁵ Для этого, в частности, может быть использован модуль Solver электронной таблицы Microsoft Excel.

⁶ Поскольку $0 \leq p_i^{\text{sign}W_{i,k} \text{ sign}W_{i+1,k}} \leq 1$ для всех i и k .

⁷ Расстояние Хэмминга – число отличающихся битов в двух бинарных векторах.

соединён с его же аксоном. Это позволяет сделать пространство поиска более «контрастным» и, как следствие, улучшить распознавание в сложных случаях.

На первом этапе нейросетевого метода распознавания прямоугольные фрагменты с контурными изображениями последовательно преобразуются в два числовых вектора. Один из них описывает распределение точек контура вдоль оси абсцисс, другой — вдоль оси ординат. Таким образом, теряя часть информации, получаем простой инвариант длиной в 64 числа (символы отмасштабированы для окна 32x32). Такой вектор может быть проанализирован с удовлетворяющей для задачи скоростью. Полученный вектор подвергается быстрому вейвлет-преобразованию, результаты которого подаются на вход сети Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации [2–4]. Весовые коэффициенты радиальных элементов этих сетей вычисляются в соответствии с имеющимися эталонными образцами распознаваемых символов. После циклических вычислений нейронная сеть сходится к номеру ближайшего эталона. Таким образом, сходство с одним из заданных эталонных символов определяется нахождением в соответствующей области притяжения в пространстве допустимых представлений входного фрагмента. Последовательность обработки данных в этом методе распознавания представлена на рис. 7.

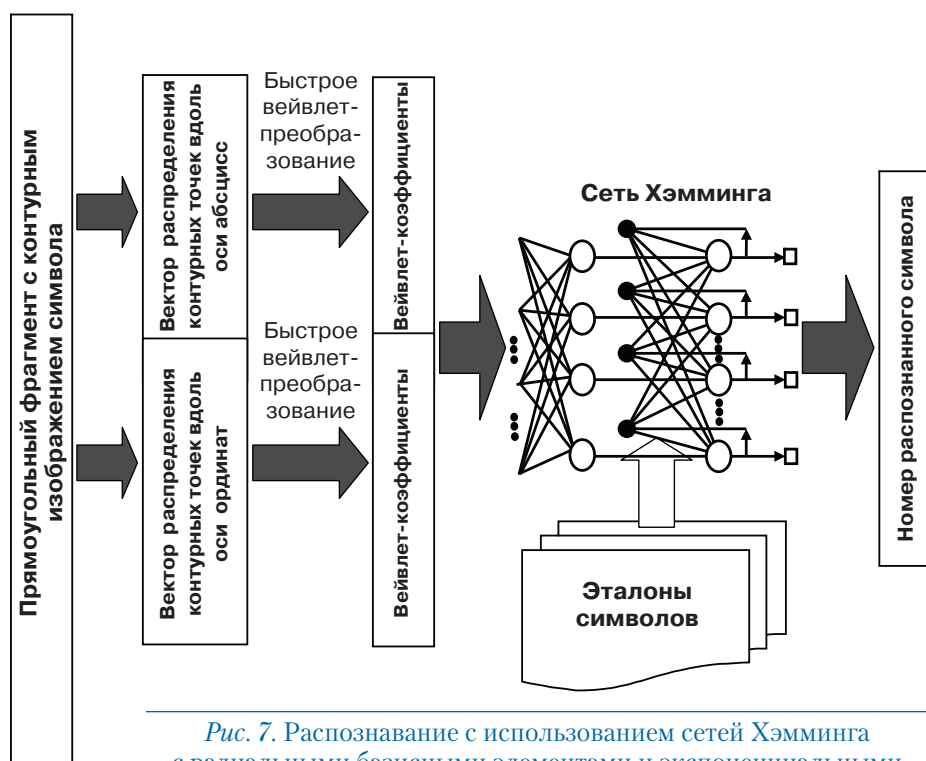


Рис. 7. Распознавание с использованием сетей Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации

Символ считается идентифицированным, если он выдаётся в качестве результата не менее чем двумя используемыми способами.

Если распознанная строка символов не будет содержать информацию о разделяющих слова пробелах, то синтезатор речи не сможет корректно воспроизвести полученный текст. Выявление пробелов в тексте необходимо

для обеспечения работы синтезатора речи. Поиск пробелов происходит на основании частотного анализа.

Расстояния между буквами обычно меньше, чем между словами, в полтора-два раза. Этот факт позволяет говорить о весьма устойчивом разделении слов текста. Данный анализ производится после нахождения прямоугольных фрагментов изображения, содержащих символы, поэтому можно легко определить позиции, после которых следует вставлять пробелы. Пробелы вставляются в текстовую строку перед заключительной коррекцией полученного текста, которая проводится на последнем этапе распознавания. Эта коррекция выполняется путём ассоциативного восстановления слов (искажённых вследствие ошибок при идентификации символов либо содержащих нераспознанные элементы) с помощью встроенного словаря.

Для того чтобы текст был озвучен, необходимо подготовить строку, содержащую последовательно все символы с пробелами. Эта строка передаётся заранее настроенному синтезатору речи. Помимо процедуры инициализации, синтезатор требует указать, какой из возможных вариантов голоса использовать при озвучке. Этот профиль, естественно, тоже можно выбрать в программной оболочке. По умолчанию используется один голос, профиль которого можно перевести как женский взрослый. Также программа осуществляет контроль за тем, завершился процесс воспроизведения или нет. Если процесс не завершился, новая порция текста не передаётся, так как передача вызовет заполнение буфера и воспроизведение длительной последовательности звуков уже после завершения сеанса работы с программой. Для озвучивания распознанного текста используется стандартный синтезатор речи. В среде операционной системы Windows для синтеза речи может быть использован интерфейс прикладного программирования Microsoft SAPI (Speech Application Programming Interface).

Очередной кадр изображения обрабатывается после озвучивания предыдущего. Темп обработки можно регулировать по желанию пользователя. Это вполне разумно, поскольку для восприятия различных по уровню информационной насыщенности текстов требуется, как правило, разное время. Не следует забывать и об индивидуальных различиях в скорости восприятия информации. Установка заниженного темпа обработки может быть полезна на начальных этапах работы с системой, пока пользователь не привык к новому способу восприятия информации, по сути требующему координированной работы рук, слуха и одновременной интерпретации услышанного.

Текущая версия программной реализации эффективно распознаёт и озвучивает произвольный плоскочечатный текст или текст на экране компьютерного монитора при достаточном освещении. Проведённое тестирование позволяет говорить о том, что не существует букв русского алфавита, плохо идентифицируемых алгоритмом.

Основные результаты и выводы

Разработана и программно реализована технология обработки текстов, особенностями которой являются:

- интеграция в одном программном продукте средств сканирования, распознавания и озвучивания;
- наличие развитых средств устранения ошибок, обусловленных низким качеством сканированного изображения;
- применение нового алгоритма распознавания символов, опирающегося на возможности вейвлет-преобразований и релаксационных нейронных сетей и способного эффективно работать после обучения на ограниченном числе образцов;
- разработан и программно реализован новый алгоритм распознавания символов на базе цепей Маркова, даны формальные оценки его эффективности;
- использование веб-камеры для сканирования озвучиваемых изображений.



Разработанная технология имеет значимые преимущества перед существующими в настоящее время средствами озвучивания текстов для людей с нарушениями зрения, связанные с:

- мобильностью аппаратных средств;
- высокой скоростью и гибкостью воспроизведения информации в удобном для пользователя режиме;
- возможностью работы с текстами, представленными не в электронной форме;
- способностью работать с изображениями на экране компьютерного монитора.

Представленные средства могут быть использованы:

- для чтения литературы, изданной традиционным плоскочечатным способом;
- для работы за компьютером с обычным программным обеспечением, не предназначенным для незрячих пользователей.

Литература

1. Богомолов А.М. Личностный адаптационный потенциал в контексте системного анализа. Психологическая наука и образование. 2008. № 1. С. 67–73.
2. Куравский Л.С., Баранов С.Н., Буланова О.Е., Кравчук Т.Е. Нейросетевая технология диагностики патологических состояний по аномалиям электроэнцефалограмм. Нейрокомпьютеры: разработка и применение. 2007. № 4. С. 4–14.
3. L.S. Kuravsky and S.N. Baranov. Wavelet transforms and relaxation neural networks as promising technology components of technical and medical diagnostics and monitoring. — In: Proc. 2nd World Congress on Engineering Asset Management and 4th International Conference on Condition Monitoring, Harrogate, United Kingdom, June 2007. Pp. 1117–1132.
4. L.S. Kuravsky and S.N. Baranov. Technical diagnostics and monitoring based on capabilities of wavelet transforms and relaxation neural networks. Insight. Vol. 50. No 3. March 2008. Pp. 127–132.
5. LabVIEW tutorial for Windows. National Instruments Corp., 2004–2007.
6. FineReader (http://www.abbyy.ru/upload/files/FineReader_9.0_Reviewer%27s_Guide_Russian.pdf).
7. JAWS (<http://www.cardiff.ac.uk/accessibility/technicalinformation/guidetojaws/index.html>).
8. CuneiForm (ftp://ftp.dol.ru/pub/users/cgntv/download/cuneiform/eng/cunei_e.pdf).
9. Adriane Knoppix (<http://www.knopper.net/knoppix-adriane/index-en.html>).

Сведения об авторах

Юрьев Григорий Александрович —

аспирант факультета информационных технологий Московского городского психолого-педагогического университета.
nezdeshni@gmail.com

Куравский Лев Семёнович —

декан факультета информационных технологий, заведующий кафедрой прикладной информатики МГППУ, профессор, доктор технических наук. kuravsky@yahoo.com

Отчёт о конференции OSTIS-2012

16 февраля 2012 г. в Белорусском государственном университете информатики и радиоэлектроники открылась II Международная научно-техническая конференция «Открытые семантические технологии проектирования интеллектуальных систем» — OSTIS-2012.

Конференция OSTIS в этом году собрала преимущественно людей, работающих в области искусственного интеллекта, из стран ближнего и дальнего зарубежья. Интересные и харизматичные личности, которым небезразлично будущее интеллектуальных информационных систем, обсуждали проблемы передовых направлений информатики. На конференцию съехались более 120 участников из Белоруссии, России, Казахстана и Украины, из них 36 профессоров и 34 доцента, а также 39 докторов и 49 кандидатов физико-математических, математических, филологических, педагогических, медицинских и психологических наук. Было заслушано сорок восемь докладов, среди которых доклады профессоров, молодых учёных, студентов и представителей IT-компаний.



Конференция длилась три дня и состояла из шести секций. В рамках каждой секции проводился круглый стол, на котором велось обсуждение докладов и вопросов, связанных с научным направлением секции. В течение конференции действовала выставка, где желающие могли ближе познакомиться с авторами докладов и результатами их работы.

В первый день работы конференции были заслушаны доклады, касающиеся базовых универсальных семантических моделей представления и обработки знаний и их про-



граммной и аппаратной реализации. В тот же день прошла секция, на которой были представлены доклады, связанные с семантическими технологиями проектирования баз знаний, программ и пакетов программ, ориентированных на обработку знаний. Бурную дискуссию вызвал доклад, посвящённый памяти Джона Маккарти — выдающегося информатика, автора термина «искусственный интеллект» (1955), изобретателя языка Лисп (1958), основоположника функционального програм-

мирования и лауреата Премии Тьюринга (1971) за огромный вклад в области исследований искусственного интеллекта, ушедшего из жизни в октябре 2011 года. В рамках доклада делился своими воспоминаниями о встречах с Дж. Маккарти профессор В.Л. Стефанюк (Москва).

Во второй день конференции были заслушаны доклады, связанные с семантическими моделями информационного поиска и решения задач, семантическими технологиями проектирования интеллектуальных информационно-поисковых машин и решателей задач. Затем были озвучены доклады, посвящённые семантическим технологиям проектирования мультимодальных и естественно-языковых интерфейсов.



Последний день конференции был посвящён обсуждению вопросов методологии и менеджмента разработки интеллектуальных систем на основе семантических технологий в рамках Open Source проектов и прикладных интеллектуальных систем, основанных на семантических сетях. Горячая дискуссия на тему затронутых в рамках этих направлений проблем продолжалась довольно долго.



По итогам конференции состоялся круглый стол. Желающих высказаться было очень много, что не позволило выступить абсолютно всем, но все выступающие подчеркнули важность и актуальность затронутых в рамках конференции проблем.

По итогам конференции состоялся круглый стол. Желающих высказаться было

очень много, что не позволило выступить абсолютно всем, но все выступающие подчеркнули важность и актуальность затронутых в рамках конференции проблем.

Завершающим аккордом OSTIS-2012 стало награждение лучших работ и молодых учёных от имени Российской ассоциации искусственного интеллекта, Парка высоких технологий, администрации БГУИР. Представители резидентов парка высоких технологий вручили ценные подарки



докладчикам и активно помогавшим в организации и проведении конференции студентам. Работы «Интеграция логики движения и бинарной модели данных и знаний» (Горкина А.А.), «Достоверность правил в экспертной системе» (Моросанова Н.А.), «Разработка интерфейса интеллектуального помощника проекта» (Боргест Н.М., Чернов Р.В., Шустова Д.В.), «Семантическая модель мультимодальных пользовательских интерфейсов и семантическая технология их проектирования» (Корончик Д.Н.), «Интеллектуальный морфологический анализатор, основанный на семантических сетях» (Шарипбаев Д.Ш., Бекманова Г.Т., Ергеш Б.Ж., Бурибаева А.К., Карабалаева М.Х.), «Комплексная методика проектирования семантических моделей интеллектуальных справочных систем» (Давыденко И.Т.) получили премию Программного комитета конференции и были рекомендованы для публикации в научных изданиях Российской Федерации и Республики Беларусь.

Планируемая дата проведения III Международной научно-технической конференции OSTIS-2013 — 21–23 февраля 2013 года.



Информация о конференции OSTIS-2013



**III МЕЖДУНАРОДНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ
«Открытые семантические технологии проектирования интеллектуальных систем»**

OSTIS-2013

**Open Semantic Technologies
for Intelligent Systems**

**21 – 23 февраля 2013 г. Минск.
Республика Беларусь**

П Е Р В О Е И Н Ф О Р М А Ц И О Н Н О Е С О О Б Щ Е Н И Е

Приглашаем принять участие в III Международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2013). Конференция пройдет в период с **21 по 23 февраля 2013** года в Белорусском государственном университете информатики и радиоэлектроники, г. Минск, Республика Беларусь. Рабочие языки конференции: русский, белорусский, английский

ОСНОВНЫЕ ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ

- Российская ассоциация искусственного интеллекта (РАИИ)
- Белорусский государственный университет информатики и радиоэлектроники (БГУИР)
- Государственное учреждение "Администрация Парка высоких технологий" (Республика Беларусь)

ПРОГРАММНЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

Кузнецов О.П., д.т.н., проф., РФ
Боргест Н.М., к.т.н., доц., РФ
Гаврилова Т.А., д.т.н., проф., РФ
Глоба Л.С., д.т.н., проф., Украина
Голенков В.В., д.т.н., проф., РБ
Грибова В.В., д.т.н., РФ
Гулякина Н.А., к.ф.-м.н., доц., РБ
Еремеев А.П., д.т.н., проф., РФ
Ефименко И.В., к. фил. н., РФ
Заболеева-Зотова А.В., д.т.н., РФ
Загорюлько Ю.А., к.т.н., доц., РФ
Иванюк А.А., д.т.н., доц., РБ
Ижуткин В.С., д.ф.-м.н., проф., РФ
Клещев А.С., д.т.н., проф., РФ
Кобринский Б.А., д.мед.н., РФ
Козлов О.А., д.п.н., проф., РФ
Комарцова Л.Г., д.т.н., РФ
Курейчик В.М., д.т.н., проф., РФ

Лобанов Б.М., д.т.н., проф., РБ
Найденова К.А., к.т.н., РФ
Невзорова О.А., к.т.н., доцент, РФ
Осипов Г.С., д.ф.-м.н., проф., РФ
Плесневич Г.С., к.ф.-м.н., РФ
Родченко В.Г., к.т.н., доц., РБ
Сморозин В.С., д.т.н., РБ
Соловьёв С.Ю., д.ф.-м.н., проф., РФ
Соснин П.И., д.т.н., проф., РФ
Стефанюк В.Л., д.т.н., проф., РФ
Сулейманов Д.Ш., академик АН
Татарстана, РФ
Тарасов В.Б., к.т.н., доц., РФ
Харламов А.А., д.т.н., РФ
Хейдоров И.Э., к.ф.-м.н., доц., РБ
Хорошевский В.Ф., д.т.н., проф., РФ
Шарипбаев А.А., д.т.н., проф., Казахстан

ДОКЛАДЫ НА КОНФЕРЕНЦИЮ БУДУТ ПРИНИМАТЬСЯ ПО СЛЕДУЮЩИМ НАПРАВЛЕНИЯМ:

1. Базовые универсальные семантические модели представления и обработки знаний и их программная и аппаратная реализация.
2. Семантические технологии компонентного проектирования совместимых баз знаний, программ и пакетов программ, ориентированных на обработку знаний.
3. Семантические модели информационного поиска и решения задач. Семантические технологии компонентного проектирования совместимых интеллектуальных информационно-поисковых машин и решателей задач.
4. Семантические технологии компонентного проектирования совместимых мультимодальных и естественно-языковых интерфейсов интеллектуальных систем.
5. Методологии и менеджмент компонентного проектирования семантически совместимых интеллектуальных систем в рамках Open Source проектов.
6. Семантически совместимые прикладные интеллектуальные системы.

ЦЕЛЬ И ФОРМАТ ПРОВЕДЕНИЯ КОНФЕРЕНЦИИ

Целью конференции является обсуждение проблем создания комплексной технологии проектирования семантически совместимых интеллектуальных систем. Этим определяется и формат её проведения, предполагающий (1) последовательное проведение секций, (2) ориентацию на круглые столы, посвящённые обсуждению различных вопросов, (3) обеспечение возможности всем авторам не только выступить с докладами, но и продемонстрировать свои результаты на выставочных стендах. Важнейшей задачей конференции является привлечение к её работе не только учёных и аспирантов, но и студенческой молодёжи, интересующейся проблемами искусственного интеллекта, а также коммерческих организаций, готовых сотрудничать с научными коллективами, работающими над созданием современных технологий проектирования интеллектуальных систем

ПОРЯДОК ПРЕДСТАВЛЕНИЯ МАТЕРИАЛОВ ДОКЛАДОВ

Материалы докладов (только по перечисленным выше направлениям) представляются в готовом для публикации виде. Текст доклада должен быть логически законченным и содержать новые научные и практические результаты. **Крайний** срок получения материалов Оргкомитетом – **1 декабря 2012**. Объем материалов доклада, включая краткий перевод, иллюстрации и список литературы, – 6 или 8 полных страниц. Переписка с авторами будет вестись только по электронной почте. Адрес электронной почты Оргкомитета: ostisconf@gmail.com. Для переписки необходимо зарегистрироваться на сайте конференции <http://conf.ostis.net>. Правила подготовки материалов докладов размещены на сайте конференции.

ФОРМИРОВАНИЕ ПРОГРАММЫ КОНФЕРЕНЦИИ

Программа конференции формируется Программным комитетом по результатам рецензирования представленных материалов докладов и будет опубликована **15 января 2013 г.** на сайте конференции <http://conf.ostis.net>



ПУБЛИКАЦИЯ МАТЕРИАЛОВ КОНФЕРЕНЦИИ

Материалы докладов, включённых в программу конференции, печатаются в сборнике материалов конференции и публикуются на сайте конференции за неделю до начала конференции

УСЛОВИЯ УЧАСТИЯ В КОНФЕРЕНЦИИ

Участие в конференции не предполагает организационного взноса

СВЯЗЬ С ОРГАНИЗАТОРАМИ КОНФЕРЕНЦИИ

Сайт: <http://conf.ostis.net>
E-mail: ostisconf@gmail.com