

*К 50-летию
Объединенного института
проблем информатики*

КИБЕРНЕТИКА И ИНФОРМАТИКА В НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

.....
ОЧЕРКИ РАЗВИТИЯ

УДК [004+681.5]:001.89:005.71(476-25)

ББК 32.81(4Бей)

К38

Печатается по решению редакционной коллегии
Объединенного института проблем информатики Национальной академии наук
Беларуси, Бюро отделения физики, математики, информатики и редакционно-
издательской комиссии Национальной академии наук Беларуси.

Научные редакторы:
академик НАН Беларуси *С.В. Абламейко*,
член-корреспондент НАН Беларуси *А.В. Тузиков*,
кандидат технических наук *О.И. Семенков*

К38 Кибернетика и информатика в Национальной академии наук Беларуси : очерки развития / Объединенный институт проблем информатики Национальной академии наук Беларуси ; науч. ред.: С.В. Абламейко, А.В. Тузиков, О.И. Семенков. – Минск : Тэхналогія, 2015. – 348 с.

ISBN 978-985-458-258-0.

В книге описывается развитие тех научных направлений, которые связаны с историей Объединенного института проблем информатики. Данное изложение материала не претендует на глубину и полноту исследований развития информатики в Национальной академии наук Беларуси и в целом в стране.

За рамками исследования остался целый ряд весьма важных перспективных направлений информатики. Авторы выражают уверенность, что эти направления станут объектом анализа и исследований в последующих публикациях.

Рассчитана на широкий круг читателей, интересующихся историей института и нынешним развитием данного научного направления.

УДК [004+681.5]:001.89:005.71(476-25)

ББК 32.81(4Бей)

ISBN 978-985-458-258-0

© Объединенный институт проблем информатики
Национальной академии наук Беларуси, 2015
© Оформление. НПК «Тэхналогія», 2015

Оглавление

| | |
|--|-----|
| Введение | 6 |
| Глава I | |
| Автоматизация проектно-конструкторских и технологических работ | 13 |
| <i>(А.Г. Гривачевский, Л.В. Губич, О.И. Семенков)</i> | |
| Предпосылки создания САПР | 13 |
| Первые пакеты прикладных программ автоматизированного проектирования | 19 |
| Методология построения САПР | 41 |
| Георгий Константинович Горанский – выдающийся ученый, крупный организатор науки в Республике Беларусь (1912–1999)..... | 63 |
| Глава II | |
| Компьютерная графика. От машин-автоматов к интеллектуальным системам | 67 |
| <i>(Е.В. Днепровский, В.В. Ткаченко, В.К. Ероховец)</i> | |
| Исследования и разработки базовых средств графического взаимодействия человека и вычислительной машины (1960–1970-е гг.) | 68 |
| Новые технологии компьютерной графики (1980-е гг.) | 75 |
| От технических средств машинной графики к технологиям информационного общества (1990-е гг.) | 81 |
| Перспективы развития работ в области компьютерной графики | 90 |
| Глава III | |
| Технические средства и системы автоматизации ввода графических изображений | 101 |
| <i>(Г.И. Алексеев)</i> | |
| Становление белорусской школы автоматизации ввода графики | 101 |
| От единичных лабораторных образцов к серийному промышленному выпуску отечественных планшетных дигитайзеров | 105 |
| Растровый ввод крупноформатных изображений: интерактивные технологии, комплексы, системы | 111 |

| | |
|--|-----|
| Итоги пройденного пути. Некоторые возможные направления будущих исследований проблемы ввода графики | 119 |
| Глава IV | |
| Логическое проектирование | 125 |
| <i>(П.Н. Бибило, Л.Д. Черемисинова)</i> | |
| Начальный этап исследований | 125 |
| Проектирование схем из элементов малой степени интеграции | 127 |
| Проектирование схем большой и сверхбольшой степени интеграции | 130 |
| Современные исследования и разработки | 135 |
| Аркадий Дмитриевич Закревский – научная, научно-организационная, педагогическая и общественная деятельность (1928–2014) | 141 |
| Глава V | |
| Математические модели и методы для систем поддержки принятия решений | 147 |
| <i>(М.Я. Ковалев, Г.М. Левин, Ю.Н. Сотсков, М.Я. Шафранский, А.В. Тузиков)</i> | |
| Параметрическая декомпозиция оптимизационных задач | 149 |
| Табличные методы принятия решений | 155 |
| Теория расписаний | 156 |
| Вячеслав Сергеевич Танаев – этапы жизни и творчества академика (1940–2002) | 168 |
| Глава VI | |
| Технологии синтеза и анализа речевых сигналов | 173 |
| <i>(Б.М. Лобанов)</i> | |
| Начальный этап развития речевых исследований | 174 |
| История «средних» лет (1970–1980-е гг.) | 177 |
| Новейшая история | 179 |
| Современные исследования и разработки | 181 |
| Глава VII | |
| Информационно-коммуникационные технологии в научных исследованиях | 191 |
| <i>(Р.Б. Григянец, В.А. Артамонов)</i> | |
| Программное обеспечение вычислительной техники | 191 |
| <i>(А.С. Метельский, И.В. Гайшун)</i> | |
| Создание и развитие научно-информационной компьютерной сети Национальной академии наук Беларуси и ее информационных ресурсов | 204 |

6

ТЕХНОЛОГИИ СИНТЕЗА И АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ

Из всего живого только человек обладает даром речи, благодаря которому нашим предкам удалось значительно развить свои интеллектуальные способности и, по мнению многих философов, в конечном итоге стать «человеком разумным». Нечто подобное происходит на наших глазах и с компьютером, чьи возможности интенсивно расширяются в различных спектрах речевых технологий: от кодирования телефонных сигналов до синтеза, распознавания и понимания устной речи. Вполне реально надеяться, что дальнейшее развитие человекомашинных систем речевого общения неизбежно будет вести к постоянному совершенствованию языковой «грамотности» компьютеров и, следовательно, к непрерывному приближению их интеллектуальных способностей к человеческим. Приведем уместную в данном контексте цитату из книги академика РАН В.В. Иванова «Лингвистика третьего тысячелетия»: «Мы впервые в истории вида начинаем широко пользоваться техническими говорящими орудиями – иначе говоря, не только изготавливать орудия (чем человек отличается от животных), но и обучать их нашему языку (чем мы начинаем отличаться от всех ранее живших людей)».

Однако на пути создания практически востребованных систем, реализующих речевые технологии, все еще предстоит преодолеть немало трудностей. В этом смысле характерно высказывание вице-президента компании Microsoft Г. Синха: «...Билл Гейтс расстроен тем, насколько долго технологиям распознавания речи приходится „пробивать“ себе дорогу. Ведь корпорация вкладывает деньги в этот проект, по крайней мере, с 1991 года...» (по материалам CNET News).

Современные речевые технологии условно можно разделить на следующие виды:

- распознавание речи (в том числе команд, ключевых слов, цепочек слов, спонтанной речи);

- синтез речи по тексту (в том числе многоголосой, персонализированной, выразительной, мультимодальной);
- голосовая биометрия (идентификация и верификация личности по голосу и речи);
- диагностика личностных характеристик и состояния говорящего;
- медицинская диагностика (диагностика заболеваний по голосу и речи).

В настоящее время сфера применения речевых технологий постоянно расширяется и затрагивает все новые и новые области человеческой деятельности. Так, например, в 2012 г. Международный аэропорт Шереметьево в Москве объявил о запуске голосовой системы информирования о статусе рейса. Сервис работает по номеру +7 (495) 956-46-66. Запуск вызвал закономерный интерес пользователей: удобный дружественный интерфейс выгодно отличает его от иных автоматических сервисов, работающих в отечественных компаниях.

Начальный этап развития речевых исследований

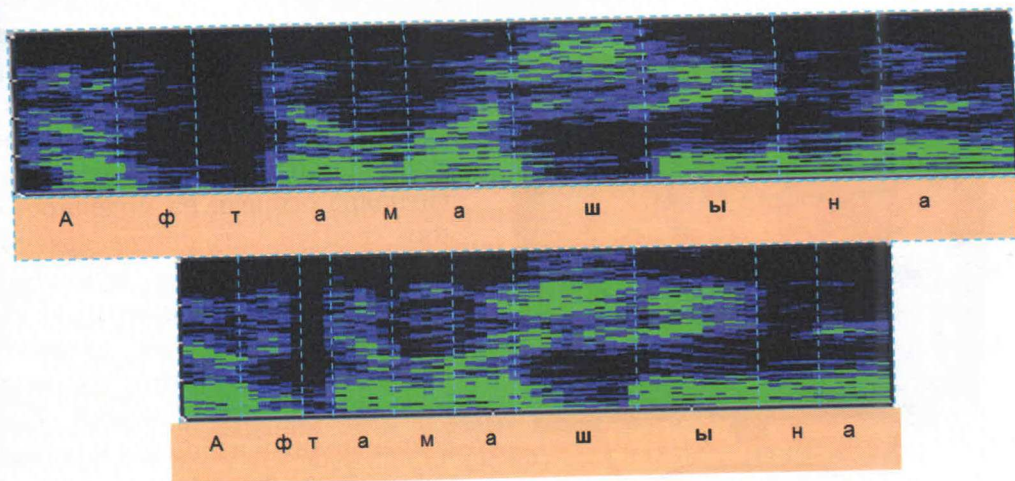
Начало современной истории речевых исследований в СССР датируется серединой 60-х гг. прошлого века, когда впервые начала работать Всесоюзная школа-семинар по автоматическому распознаванию слуховых образов (АРСО), собиравшая в лучшие годы до 300 участников. К этому же времени относится и начальный этап развития речевых исследований в Беларуси. В 1965 г. в научной лаборатории кафедры радиоприемных устройств Минского радиотехнического института под руководством Б.М. Лобанова была организована группа исследования речевых сигналов. В нее входили Н.П. Дегтярев, Б.В. Панченко, М.К. Фатеев и др., которые еще долгое время работали в этом направлении, а некоторые из них работают и по сей день.

Первые исследования группы были связаны с разработкой общих принципов анализа речевых сигналов и выделения информативных признаков, которые позволили бы представить непрерывный речевой сигнал последовательностью фонетических сегментов. Результаты этих исследований были обобщены в диссертации Б.М. Лобанова «Некоторые вопросы анализа речевых сигналов», защищенной в 1968 г. в Гос. НИИ радио (Москва). Наиболее важные результаты работы позднее были опубликованы в авторитетных международных журналах. На базе данных исследований впервые в СССР было разработано относительно простое устройство распознавания речевых команд «Сезам-2», получившее в 1968 г. серебряную медаль ВДНХ СССР. Устройство состояло из двух блоков: анализатора признаков речевого сигнала, таких как «голосовой», «шумный», «гласный» и др., и счетчика количества признаков в речевой команде. Достигнута достаточно высокая

надежность распознавания 20 команд (включая названия цифр) независимо от голоса диктора, громкости и темпа произношения. В тот же период времени были разработаны специализированные приборы для экспериментально-фонетических исследований речи: анализатор динамических спектров и интонограф, с помощью которых в последующие годы проведены многочисленные исследования в фонетических лабораториях Института языкознания АН БССР и Минского института иностранных языков.

Исследования динамических спектров речи дали толчок к развитию нелинейных методов сопоставления распознаваемых слов устной речи с их эталонами. Спектральные изображения речи в отличие от обычных визуальных изображений объектов могут подвергаться неконтролируемым нелинейным искажениям временной оси. Ниже показаны два спектральных изображения слова «автомашина», произнесенных одним и тем же диктором с различным темпом речи.

Из рисунка видно, что при ускоренном темпе (нижняя спектрограмма) при общем сокращении длительности слова на 30 % длительность звуков ударного слога «шы» практически не изменилась, в то время как звуки «т» и «н» в безударных слогах сократились более чем в два раза. Можно сделать вывод, что простого масштабирования спектральных изображений недостаточно для их надежного распознавания. Ситуация, образно говоря, схожа с той, которая могла бы возникнуть в условиях «кривого зеркала» при распознавании зрительных образов. Решение фундаментальной проблемы распознавания речи, связанной с нелинейными искажениями временной оси, было предложено независимо и практически одновременно Г.С. Слуцкером (Гос. НИИ радио) и Т.К. Винцюком (Институт кибернетики АН УССР)



Спектральные изображения речи

во второй половине 1960-х гг. Суть предложенного решения заключалась в нахождении методом динамического программирования (ДП-методом) оптимального пути на матрице локальных расстояний между временными отсчетами векторов распознаваемого и эталонного спектров. В 1969 г. Б.М. Лобановым совместно с сотрудниками Гос. НИИ радио была опубликована статья, в которой обосновывалось дальнейшее развитие ДП-метода для исключительно важного практического случая, когда границы распознаваемого слова неизвестны, т. е. для решения задачи обнаружения и распознавания звукоочетаний в непрерывном речевом сигнале. ДП-метод получил широкое признание зарубежных исследователей и наряду с методом скрытых марковских моделей до сих пор используется в современных системах распознавания речи.

К концу 1960-х гг. относится также начало работ по созданию синтезаторов речи. Стимулом послужило осознание того, что разработка и исследование моделей синтеза речи – это прямой путь к получению более детальных знаний о природе образования и свойствах речевого сигнала, опираясь на которые в дальнейшем можно будет построить более совершенные алгоритмы анализа и распознавания речи. Немаловажную роль в освоении мирового технологического уровня синтеза речи того времени сыграла научная стажировка Б.М. Лобанова в 1970 г. в лаборатории профессора Лоренца Эдинбургского университета Великобритании, где была разработана одна из первых формантных моделей синтеза речевых сигналов. С помощью синтезатора этой лаборатории были впервые получены высококачественные образцы синтезированной русской речи.

Первая и пока еще не вполне совершенная модель синтезатора русской речи по тексту «Фонемофон-1» «заговорила» в начале 1970-х гг., и успех в ее создании связан в первую очередь с разработкой новых методов аппаратной реализации формантного синтеза речевых сигналов. Принцип формантного синтеза речевых сигналов основан на моделировании свойств источников возбуждения (голосового и шумового) и резонансных (формантных) характеристик речевого аппарата человека. В результате экспериментальных исследований был создан полный набор формантных «портретов» фонем, позволивший впервые осуществить синтез рус-



Синтезатор «Фонемофон-1»

ской речи по произвольному тексту. Позднее появилась улучшенная версия синтезатора – «Фонемофон-2» – с дополнительным блоком преобразования «фонема – аллофон».

История «средних» лет (1970–1980-е гг.)

Решающую роль в ориентировке на решение прикладных задач использования речевых технологий сыграло создание в 1974 г. лаборатории обработки речевых сигналов в составе Минского отдела Центрального НИИ связи (МОНИИС). В 1976 г. на Всесоюзном семинаре АРСО-9, проведенном в Минске, был впервые продемонстрирован прототип автоматической телефонной справочной службы с синтезированным речевым ответом, а уже с начала 1980-х гг. в Минске длительное время работала система автоматического обзвона должников за междугородние переговоры (в их число иногда попадали и авторы этой разработки). К середине 1980-х гг. эта система была внедрена во многих городах СССР – от Бреста до Петропавловска-Камчатского. Успешному внедрению синтезаторов речи предшествовала длительная работа по совершенствованию как качественных показателей синтезированной речи, так и технологии их реализации в виде нового класса внешних устройств ЭВМ. Основными недостатками первых моделей синтезаторов «Фонемофон-1», «Фонемофон-2» были невысокие разборчивость и качество синтезированной речи, обусловленные прежде всего использованием предельно упрощенных моделей взаимодействия звуков в процессе речеобразования (эффектов коартикуляции и редукции) и недостаточно проработанной модели интонирования речи по тексту. В следующей модели – «Фонемофон-3» – были введены дополнительные блоки моделирования процессов артикуляции и интонирования речи, что существенно повысило качественные показатели синтезированной речи.

В 1979 г. «Фонемофон-3» демонстрировался на Всемирной выставке «Телеком-79» в Женеве. Известный фантаст Артур Кларк, посетив павильон СССР и ознакомившись с синтезатором речи, записал в книгу отзывов: «Вы предвосхитили мои фантазии из фильма „Космическая Одиссея – 2001“», а швейцарская газета «Обозреватель» опубликовала статью «Теперь русские изучают иностранные языки с помощью компьютера, который говорит».

Важную роль в создании промышленных синтезаторов речи сыграла разработка полностью цифровой модели синтезатора речи «Фонемофон-4», системы распознавания речи «Сезам» и речевого терминала «Марс». Их серийный выпуск впервые в СССР (1983) был налажен на ПО «Кварц» (Калининград) благодаря энтузиазму сотрудников конструкторского отдела,



Речевого терминала «Марс»

возглавляемого В.П. Афанасьевым. Ключевую роль в создании промышленной лаборатории сыграли разработки наших сотрудников Н.П. Дегтярева и В.В. Шатерника. В речевом терминале «Марс» впервые были интегрированы функции распознавания и синтеза речи. В основу алгоритмов распознавания речи положен ДП-метод принятия словесных решений на базе набора формантных признаков речевого сигнала. Опытные образцы систем «Сезам» и «Марс» были выполнены на микропроцессорной основе и по параметрам назначения не уступали лучшим зарубежным аналогам того времени. Оригинальность технических решений, использованных при создании систем «Фонемофон», «Сезам» и «Марс», защищена многочисленными авторскими свидетельствами СССР на изобретения.

К началу 1984 г. относятся окончательная формулировка, теоретическая и экспериментальная разработка единого лингвоакустического подхода к

решению проблемы синтеза речи по тексту, его реализация в виде технических систем и практическое внедрение в составе автоматизированных систем управления и связи. Результаты этих исследований были обобщены в докторской диссертации Б.М. Лобанова «Методы автоматического синтеза русской речи по тексту», защищенной в 1984 г. в Институте электро-

*Thank you for demo of
really good English synthesis.*

*Samuel Brand
G. Fant*

August 1-7 1987
Tallinn
Estonia, U.S.S.R.



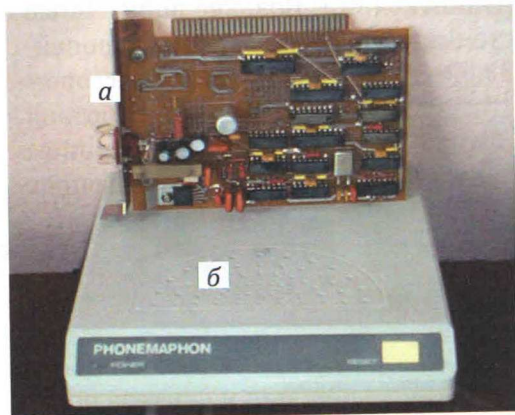
Факсимиле отзыва о демонстрации
профессора Г. Фанта

ники и вычислительной техники АН Латвийской ССР. Позднее полученные результаты были адаптированы для систем синтеза речи на других европейских языках. В частности, к 1987 г. благодаря сотрудничеству с профессором Минского института иностранных языков Е. Карневской была разработана англоязычная версия синтезатора, продемонстрированная на Всемирном конгрессе фонетических наук и получившая высокую оценку англоязычных специалистов.

Новейшая история

В 1988 г. на базе лаборатории МОНИИС в Институте технической кибернетики АН БССР была создана лаборатория распознавания и синтеза речи, на должность заведующего которой дирекцией института был приглашен Б.М. Лобанов. Многие еще хорошо помнят, что конец 1980-х гг. ознаменовался появлением первых персональных компьютеров (ПК), поэтому естественным образом в планах работ лаборатории появилась тематика, связанная с оснащением ПК системой речевого ввода-вывода информации. Формантный метод, который долго играл ключевую роль в системах синтеза речи по тексту, не подходил для этой цели из-за необходимости большого объема вычислений, что было недостижимо для ПК тех лет. В конце 1980-х гг. был предложен новый микроволновой (МВ) метод синтеза речевых сигналов, в котором вместо вычислений формантных колебаний (звуковых волн) использовался подготовленный заранее набор микроволн естественного речевого сигнала. Набор микроволн состоял из отрезков сигнала, равных длительности периода, а их количество, необходимое для генерации любого звука речи, достигало нескольких сотен. МВ-метод воплощен А.Н. Ивановым в синтезаторе «Фонемофон-5» в виде специализированного ПО синтезатора, ориентированного на работу с внутренней звуковой платой либо с автономным устройством, подключаемым к порту RS-232.

Удивительная для многих компактность его ПО (всего 64 Кбайт) позволила оснастить синтезом речи уже первые IBM PC-XT и даже отечественные ПК ЕС-1840. Синтезатор речи был востребован во многих практических приложениях, но особенно широко он использовался (и используется в настоящее время) незрячими пользователями ПК. Более сотни комплектов специализированных аппаратно-программных продуктов для незрячих было создано и распространено сотрудником лаборатории Г.В. Лосиком в России, Украине и Беларуси в первой половине 1990-х гг. До сих пор его вполне разборчивое звучание можно услышать в Интернете или на приобретенном на рынке CD-ROM «Говорящая мышь». В дальнейшем на основе МВ-метода были разработаны версии для чешского и польского языков (примерно за три месяца пребывания в стране по приглашению заказчиков), а также ав-



Микроволновой синтезатор речи:
а) звуковая плата ПК; б) автономное устройство.

тономный одноплатный модуль синтеза речи, украинскоязычная версия которого долгое время работала на линии киевского метро.

Сложная экономическая ситуация, сложившаяся в стране в середине 1990-х гг., заставляла искать источники финансирования исследований за рубежом, в первую очередь в форме совместных международных проектов. Первым из них стал международный проект «Двухязычный синтез речи – немецкий/русский» (1995–1996), который выполнялся совместно с Дрезденским техническим университетом

и финансировался германским научным фондом FTU Karlsruhe.

Следующий проект «Анализ естественного языка и речи» (1996–1997) выполнялся совместно с Саарбрюккенским университетом (Германия), Манчестерским университетом (Великобритания) и Институтом проблем передачи информации РАН и финансировался европейским фондом INTAS. Участие в этом проекте было связано с дальнейшим развитием моделей синтеза речи путем их интеграции в системы обработки естественного языка методами компьютерной лингвистики.

Важную роль в интеграции белорусских исследователей лингвистики и речи в европейское сообщество сыграло участие в Международном проекте «Развитие Европейской компьютерной сети по лингвистике и речи в восточном направлении» (1997–1998), финансировавшемся европейским фондом COPERNICUS. С 1998 г. лаборатория распознавания и синтеза речи института является координационным центром данной сети в Беларуси.

Кроме европейских научных организаций интерес к сотрудничеству с лабораторией проявили в эти годы и некоторые коммерческие организации. В 1996 г. французская фирма «Секстант Авионик» предложила реализовать научный проект «Распознавание речевых команд в условиях шумов в кабине самолета». Проект финансировался фондом Министерства обороны Франции. Несмотря на исключительную сложность поставленной задачи, он был успешно выполнен и в 1997 г. принят заказчиком. Основные научные результаты этого проекта изложены в статье «Continuous Speech Recognizer for Aircraft Application». Следующий коммерческий проект «Интеллектуальный телефонный автосекретарь» выполнялся с 1997 г. по договору с фирмой NovCom NV (США). Суть его заключалась в решении задачи распознавания

произносимых по телефону имен абонентов и другой служебной информации с тем, чтобы система смогла выполнять функции телефонного автосекретаря. Проект был завершен в 1999 г., его основные научные результаты опубликованы в статье «An Intelligent Answering System Using Speech Recognition». В выполнении этих международных проектов ключевую роль сыграли сотрудники лаборатории Т.В. Левковская, А.Н. Иванов и А.В. Кубашин. Работы по проектам явились стимулом к возрождению и дальнейшему развитию алгоритмов распознавания речи, предложенных еще в 1969 г.

Современные исследования и разработки

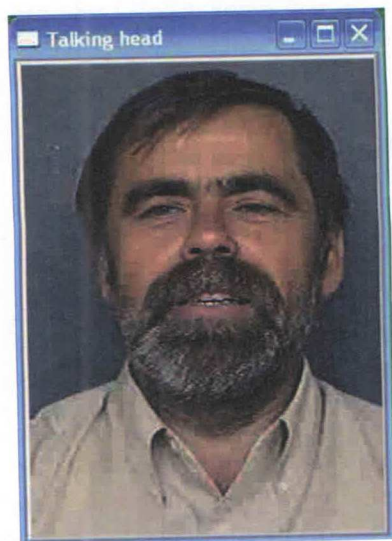
Начиная с 2000 г. научные интересы лаборатории от проблем распознавания вновь переходят к проблемам синтеза речи. Разрабатывается новая мультиволновая модель синтеза речи по тексту, обеспечивающая высокое качество и персонализацию синтезированной речи. Весомый вклад в создание на основе этой модели синтезатора речи «МультиФон» внесли сотрудники лаборатории В.В. Киселев и Д.В. Жадинец. Одновременно теоретически обосновывается и развивается новое научное направление в речевых исследованиях – компьютерное клонирование речи – для решения задач высококачественного синтеза речи по тексту с максимально возможным приближением к голосу и манере чтения конкретного человека. Создаются технология и специальные программные средства клонирования акустических, фонетических и интонационных характеристик речи человека. Существенную роль в создании программных средств клонирования речи «ФоноКлонатор» и «ИнтоКлонатор» сыграли А.Г. Давыдов и Л.И. Цирульник.

Полученные в начале 2000-х гг. теоретические и экспериментальные результаты позволили осуществить ряд совместных прикладных проектов.

В 2005–2007 гг. выполнен Международный INTAS-проект «Разработка многоголосой и многоязычной системы синтеза и распознавания речи (языки: белорусский, польский, русский)». Участники проекта: Беларусь, Германия, Польша, Россия. В результате выполнения проекта были заложены основы создания многоязычной системы автоматического перевода устной речи для трех славянских языков.

В 2006–2007 гг. по заказу ООО «ИнвоСервис» (Минск) разработана система «Чтец», ориентированная на создание и редактирование «говорящих» учебников для незрячих учеников спецшкол Беларуси.

В 2007–2009 гг. в рамках БРФФИ–РФФИ выполнен проект «Разработка мультимедийной системы аудиовизуального синтеза речи», так называемой «Персонализированной говорящей головы», которая позволила существен-



«Говорящая голова» профессора
В.В. Голенкова

но улучшить комфортность восприятия речи в ряде практических приложений, в частности в системах дистанционного обучения.

В 2009–2011 гг. в рамках БРФФИ – РФФИ выполнен проект «Синтез выразительной речи на базе глубокого синтаксического анализа читаемого текста», результатом которого стало создание первой двуязычной системы синтеза выразительной речи на белорусском и русском языках, демонстрировавшейся на нескольких научно-технических выставках. В эти же годы проведен большой объем исследований, направленных на разработку белорусскоязычной модели синтеза речи. Основную роль в создании синтезатора белорусской речи сыграли исследования и разработки заведующего лабораторией распознавания и синтеза речи Ю.С. Гецевича.

В 2011–2012 гг. по заказу Якутского государственного университета была разработана впервые в мире система синтеза якутской речи. Синтезатор якутской речи используется в системах компьютерного обучения якутского языка, а также в системах озвучивания текстов для слепых.

Ниже приведены два примера современных разработок лаборатории.

Компьютерное клонирование персонального голоса и дикции

Многолетние исследования, проведенные в XX в., позволили создать синтезаторы, обеспечивающие качество и разборчивость речи, вполне пригодные для широкого спектра практических приложений. Однако, несмотря на все усилия, синтезированная речь оставалась еще далекой по качеству от натуральной и обладала узнаваемым машинным акцентом. Причиной этому были не столько низкий уровень наших знаний о процессах речеобразования и о фонетике, сколько нехватка вычислительных ресурсов компьютеров того времени. Сейчас мы можем не ограничивать себя ни объемом оперативной и дисковой памяти, ни требуемым объемом вычислений, и это явилось предпосылкой к постановке задачи синтеза речи по тексту с максимально возможным приближением по звучанию к голосу и манере чтения конкретного человека.

Такая постановка задачи, хотя и отдаленно, но напоминает широко известную биологическую проблему клонирования. В нашем случае в отличие от классической задачи клонирования делается попытка создания близкой копии, но не биологической, а компьютерной, и не всего существа в целом

(в данном случае человека), а только одной из его интеллектуальных функций – чтения произвольного орфографического текста. При этом ставится задача максимально полного сохранения персональных акустических особенностей голоса, фонетических особенностей произношения и акцента, а также просодической индивидуальности речи (мелодики, ритмики, динамики).

Синтезатор состоит из четырех процессоров: лингвистического, просодического, фонетического и акустического. Каждый из процессоров использует для осуществляемых им преобразований специализированные базы данных. В этих базах данных заложены как общие языковые правила (лингвистические, просодические, фонетические, акустические), так и правила, связанные с индивидуальными особенностями голоса и речи диктора.

Клонирование акустических характеристик голоса. Персональные акустические характеристики голоса человека обусловлены множеством факторов, таких как анатомические особенности строения и функционирования элементов речевого аппарата (гортани, голосовых связок, глотки, полости рта и др.), динамические особенности взаимодействия колебаний голосовых связок и резонаторов речевого аппарата, а также многое другое. Как известно, попытки имитации персональных характеристик голоса в системах «текст – речь» на основе моделирования физиологических и акустических процессов речеобразования из-за их чрезвычайной сложности до сих пор не привели к ощутимым результатам. В связи с этим наиболее разумным представляется использование отрезков натуральной речевой волны в качестве минимального «генетического материала» для клонирования голоса. В качестве такого отрезка целесообразно выбрать аллофон как наиболее изученную фонетическую субстанцию, причем ограниченный набор аллофонов способен обеспечить порождение устной речи произвольного содержания. При этом звуковая волна содержит в себе все существенные персональные особенности голосообразования, проявляющиеся в данном конкретном аллофоне.

Клонирование персональных фонетических особенностей произношения. В отличие от персональных акустических характеристик голоса, обусловленных в основном статическими параметрами речевого аппарата, фонетические особенности произношения определены главным образом динамикой артикуляторных движений, осуществляемых в процессе речи. Присущие данному индивиду скорость артикуляторных движений, индивидуальные особенности артикуляции того или иного звука (например, [P]), региональный или иностранный акцент обуславливают возникновение своеобразных позиционных и комбинаторных оттенков фонем и создают уникальную систему аллофонов. Таким образом, успешное клонирование персональных



Структурная схема синтезатора персонализированной речи,
на базе которого осуществляется клонирование

фонетических особенностей произношения может быть достигнуто путем имитации особенностей фонемно-аллофонного преобразования, присущего данному человеку в процессе речи.

Клонирование персональных просодических характеристик речи. Комплекс просодических (интонационных) характеристик речи, включающий мелодику, ритмику и энергетику, задается закономерными изменениями во времени частоты основного тона, длительности звуков и амплитуды звуковых сигналов. Характер этих изменений определяется не только конкретным текстом, но и персональной манерой его чтения. Решение задачи клонирования просодических характеристик речи заключается в создании достаточно полного набора персональных интонационных «портретов» речи.

Технология клонирования и ее приложения. Для успешного клонирования персональных характеристик голоса и дикции необходимо создать достаточно полные наборы звуковых волн аллофонов и интонационных «портретов» речи. В случае если клонируемый голос диктора физиче-

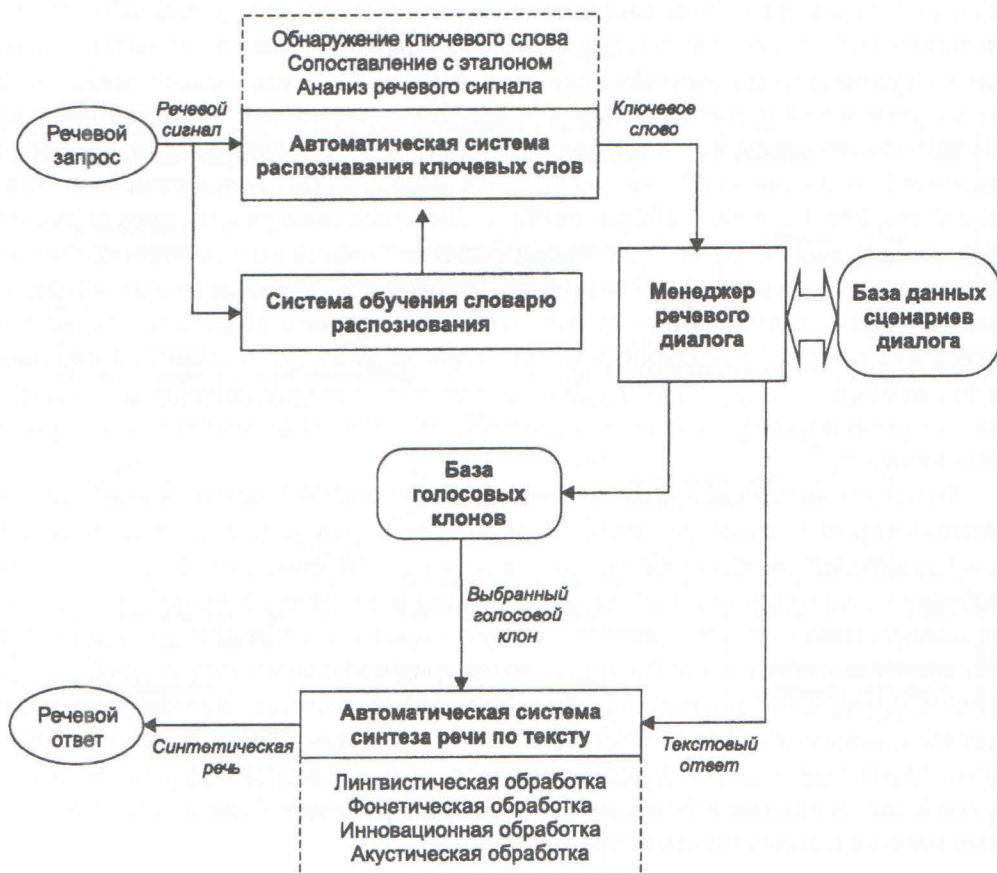
ски доступен, для этой цели используется специально разработанный компактный звуковой массив слов и отрывков текста, начитываемый им в студии или в обычных условиях. Если же диктор недоступен, то используются уже имеющиеся записи его голоса на радио, телевидении и др. Первые результаты по клонированию на примере персонального голоса и дикции Б.М. Лобанова были получены в 2000 г. В 2001 г. получен клон женского голоса, а к концу 2005 г. набор клонов состоял уже из трех мужских и двух женских голосов. В настоящее время создана компьютерная технология клонирования фонетико-акустических и просодических характеристик речи, позволяющая в значительной степени автоматизировать и ускорить процесс создания речевых клонов диктора. Основные научные и практические результаты работ по компьютерному синтезу и клонированию речи отражены в монографии «Компьютерный синтез и клонирование речи».

Отметим некоторые возможные коммерческие и практические аспекты компьютерного клонирования. Наверняка найдется большое количество пользователей, желающих, чтобы персональный компьютер заговорил их собственным голосом или, например, голосом знакомого человека, любимого актера. Интересным может быть также проект оживления голосов давно ушедших великих людей по оставшимся грамофонным или студийным записям. Разработка технологии создания голосовых клонов может оказаться кардинальным средством борьбы с так называемым телефонным терроризмом, обеспечив идентификацию личности по голосу путем автоматического сравнения оперативной записи голоса с содержимым базы данных голосовых клонов потенциальных правонарушителей.

Компьютерная модель речевого виртуального собеседника

Компьютерная модель устно-речевого виртуального собеседника (система РЕВИРС) – новая разработка лаборатории распознавания и синтеза речи, в которой интегрированы оригинальные научно-технические решения, полученные сотрудниками лаборатории в течение последних лет. Система РЕВИРС позволяет создавать сценарии диалогов для разнообразных приложений и осуществлять их посредством устно-речевого человекомашинного общения. Уникальность этой системы заключается в следующем:

- надежном распознавании ключевых слов запроса в непрерывном потоке речи;
- многодикторном распознавании ключевых слов в условиях акустических помех и искажений;
- многоголосом синтезе речи по произвольному тексту;
- возможности «клонирования» голоса личности в процессе синтеза речи;
- дуплексном режиме в реальном времени (возможности прерывания голосового ответа).



Структурная схема системы РЕВИРС

Речевой сигнал поступает на вход системы распознавания речи, которая осуществляет анализ информативных признаков сигнала, их сопоставление с эталонами ключевых слов, обнаружение и принятие решения о произнесенном слове. Если распознающий модуль обнаружил ключевое слово, оно перенаправляется менеджеру речевого диалога, который формирует текстовый ответ. Менеджер речевого диалога выбирает также голосовой клон для синтеза речевого ответа. Выбранный клон и текст ответа поступают на вход системы синтеза речи, которая осуществляет лингвистическую, интонационную, фонетическую и акустическую обработку; в результате текст преобразуется в звучащую речь заданного голосового клона.

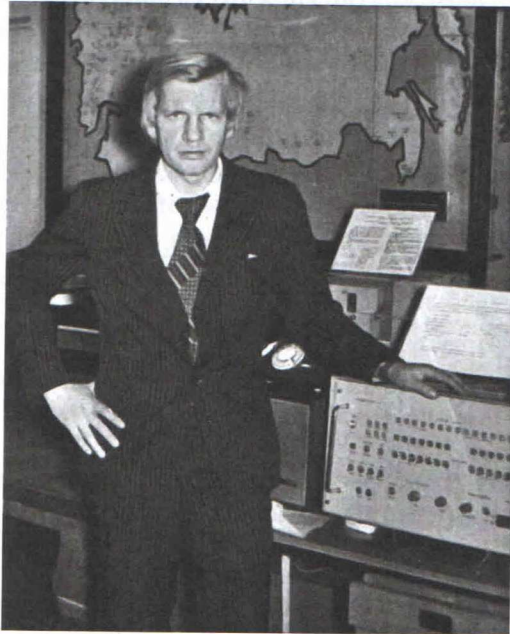
Изложенный в настоящей главе краткий очерк истории создания и развития речевых технологий в Беларуси не претендует, конечно, на полноту освещения всех научно-технических результатов 40-летних исследований.

Не было возможности даже кратко упомянуть работы всех авторов, внесших определенный вклад в развитие этой отрасли знаний (в общей сложности ими опубликовано более 300 научных работ, в том числе 5 монографий).

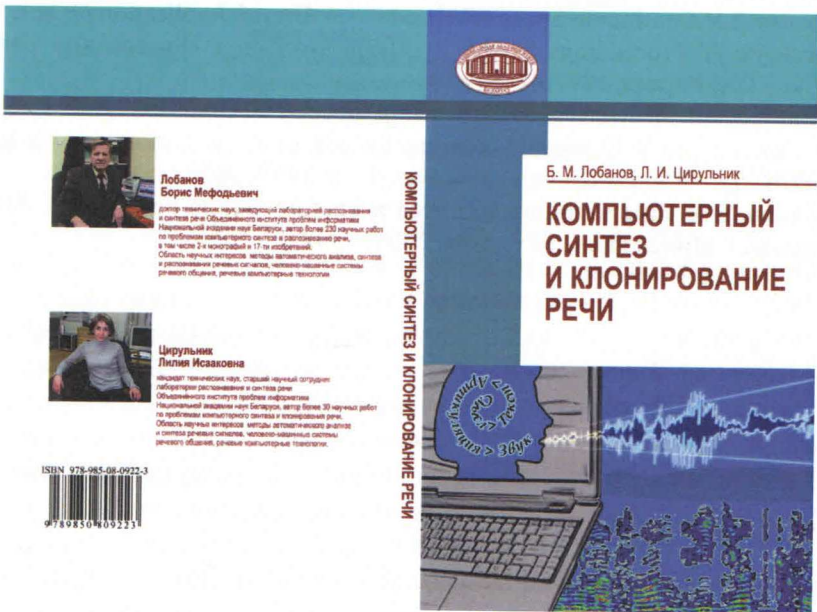
Несмотря на проделанный большой путь и полученные ощутимые научные и практические результаты, история речевых исследований в Беларуси на этом не заканчивается. Как и 40 лет назад, разгадка природы человеческой речи остается притягательной для молодых исследователей, а возможные приложения речевых технологий, как никогда, становятся актуальными.

Литература

1. Лобанов, Б.М. Автоматическое распознавание звукоочетаний в текущем речевом потоке / Б.М. Лобанов, Г.С. Слуцкер, А.П. Тизик // Труды НИИР. – М., 1969. – С. 67–75.
2. Lobanov, B.M. More About Speech Signal and the Main Principles of its Analysis / B.M. Lobanov // IEEE Transactions on Audio and Electroacoustics. – 1970. – № 3. – P. 316–318.
3. Lobanov, B.M. Classification of Russian Vowels Spoken by Different Speakers / B.M. Lobanov // J. of the Acoustical Society of America. – 1971. – № 2 (2). – P. 521–524.
4. Lobanov, B.M. The Phonemophon Text-to-Speech System / B.M. Lobanov // Proc. of the XI Intern. Congr. of Phonetic Sciences, Tallin, 1987. – Tallin, 1987. – P. 100–104.
5. Lobanov, B.M. MW – Speech Synthesis from Text / B.M. Lobanov, E.B. Karnevskaia // Proc. of the XII Intern. Congr. of Phonetic Sciences. – Aixen-Provense, Franse, 1991. – P. 387–391.
6. Lobanov, B.M. Continuous Speech Recognizer for Aircraft Application / B.M. Lobanov, T.V. Levkovskaya // Proceedings of the II Intern. Workshop «Speech and Computer SPECOM'97». – Cluj-Napoca, 1997. – P. 97–102.
7. An Intelligent Answering System Using Speech Recognition / B.M. Lobanov [et al.] // Proc. of the V European Conf. on Speech Communication and Technology (EUROSPЕECH'97). – Rhodes-Greece, 1997. – Vol. 4. – P. 1803–1806.
8. Лобанов, Б.М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник. – Минск : Беларус. наука, 2008. – 342 с.



Б.М. Лобанов со своим изобретением «Фонемофон-3» на выставке в Женеве, 1979 г.



Монография, вышедшая в издательстве «Наука», 2008 г.



Премьер-министр Беларуси *Михаил Мясникович* и вице-президент Российской академии наук *Жорес Алферов* слушают рассказ молодого ученого *Юрася Гецевича* о разработках синтеза белорусской и русской речи (фото: «Рэспубліка» за 16.11.2011 г.)



Лаборатория распознавания и синтеза речи, декабрь, 2013.
Первый ряд (слева направо): *Б.М. Лобанов*, *Ю.С. Гецевич*; второй ряд (слева направо):
Е.Н. Скопинова, *С.И. Лысый*, *Т.И. Окрут*, *Ю.С. Бородина*, *Е.В. Гюнтер*, *Д.А. Покладок*