ИНФОРМАТИКА

2004 **октябрь-декабрь** **№4**

УДК 395.521

Б.М. Лобанов

ИСТОРИЯ СОЗДАНИЯ И РАЗВИТИЯ В БЕЛАРУСИ КОМПЬЮТЕРНЫХ СИСТЕМ РАСПОЗНАВАНИЯ И СИНТЕЗА РЕЧИ

Описывается 40-летняя история создания и развития в Беларуси компьютерных систем распознавания и синтеза речи, включающая начальный этап в рамках группы исследования речи в Минском радиотехническом институте (1965 – 1974), развитие исследований в составе лаборатории обработки речевых сигналов Московского отделения Центрального научно-исследовательского института связи (1974 – 1988) и новейшую историю разработок компьютерных систем распознавания и синтеза речи в лаборатории ИТК (ныне ОИПИ) НАН Беларуси. Описываются последние научные и практические разработки лаборатории, в частности компьютерное клонирование персонального голоса и дикции человека.

Введение

Природа и механизмы речи волновали еще средневековых философов. Первые попытки создания механической говорящей машины относятся к концу XVIII в., когда во времена правления Екатерины II Петербургская академия наук объявила всемирный конкурс на объяснение природы гласных звуков. Победителем конкурса стал профессор Петербургского университета Крантценштейн, который построил систему акустических резонаторов, издававших гласные звуки русской речи при помощи вибрирующих язычков, возбуждаемых воздушным потоком. Несколько позже немецкий механик В. фон Кемпелен разработал более сложную модель генерации связной речи (рис. 1). В ней в роли резонаторов речевого тракта выступала гибкая трубка из кожи, управляемая оператором. Имелись также отверстия для имитации носовых полостей и ручки управления свистками, создававшими фрикативные звуки. Следующим заметным шагом в понимании природы речи стало создание во второй половине XIX в. немецким физиком Г. Гельмгольцем резонансной (формантной) теории звуков речи. В частности, им впервые были изучены формантные характеристики гласных звуков.

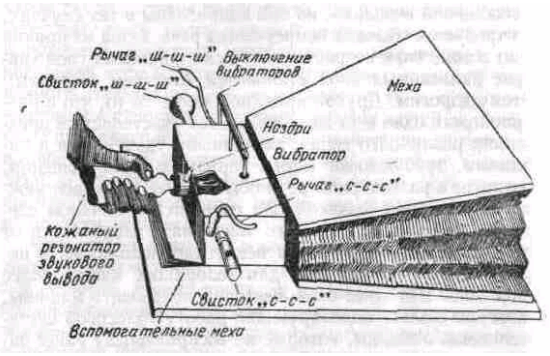

Рис. 1. Синтезатор Кемпелена

Создание первых технических систем обработки речи связано с развитием радиоэлектроники в первой половине XX в. В 1939 г. американский ученый Дадли продемонстрировал вокодер (Voice Coders), который впоследствии стал прототипом электронной говорящей машины и анализатора речи. Следующая заметная попытка синтеза речи была связана с развитием звукового кино и электронной музыки. В московской Студии электронной музыки Музея А.Н. Скрябина ее сотрудник Мурзин сконструировал банк «чистых тонов» на основе стеклянного диска, очень похожего на современный компакт-диск. На его основе был создан синтезатор звуков АНС (от инициалов композитора Скрябина, которому посвятил свое изобретение автор). Первые модели говорящих устройств того времени были очень похожи на музыкальные инструменты, а обучение операторов напоминало обучение музыкантов и требовало немало времени и способностей.

С начала 1940-х гг. появились первые работы, посвященные кодированию и распознаванию речи. В 1943 г. в журнале «Успехи физических наук» была опубликована статья ленинградского профессора Мясникова, в которой впервые описано устройство распознавания гласных фонем русской речи. Сразу после окончания войны крупные инвестиции в эту область

науки были сделаны непосредственно по указанию И.В. Сталина. В Москве была создана организация «НИИ-100», где большой коллектив математиков, физиков и инженеров решал проблемы кодирования речевого сигнала, идентификации личности по голосу и распознавания ключевых слов в потоке речи. Этот период речевых исследований прекрасно описан А. Солженицыным в романе «В круге первом». В конце 1960-х гг. в рамках выполнения одного из договоров автору посчастливилось сотрудничать с некоторыми из прототипов героев этого романа.

1. Начальный этап развития речевых исследований в Беларуси

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 автора этой статьи была организована группа исследования речевых сигналов. В то время в нее входили Н.П. Дегтярев, Б.В. Панченко, М.К. Фатеев и др. Некоторые из них и по сей день работают в этой области.

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 автора «Некоторые вопросы анализа речевых сигналов», защищенной в 1968 г. в Московском государственном НИИ радио. Наиболее важные результаты этой работы позднее были опубликованы в авторитетных международных журналах [1, 2]. На базе этих 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. Впоследствии, ко дню 50-летия комсомола, устройство было подарено Минским радиотехническим институтом Центральному комитету ЛКСМБ и находится ныне в запасниках Белорусского краеведческого музея. В тот же период 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.

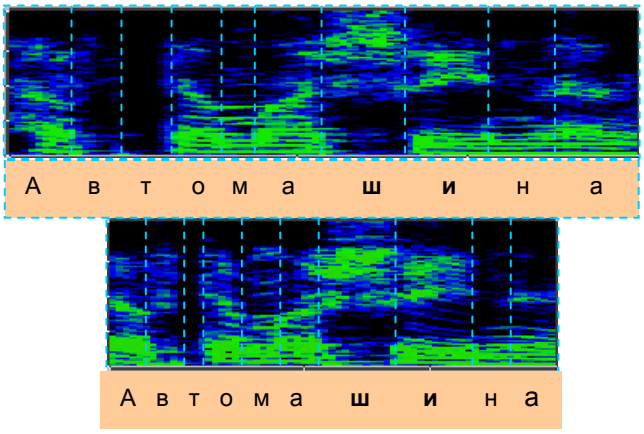

Рис. 2. Спектральные изображения речи

Исследования динамических спектров речи дали толчок 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. На рис. 2 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 практически не изменилась, а некоторых звуков в безударных слогах сократилась более чем в два раза. Из приведенного примера видно, 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 (Московский государственный НИИ радио) и Т. Г. Винцюком (Институт кибернетики АН УССР) во второй половине 1960-х гг. Суть предложенного решения заключалась в нахождении методами динамического программирования оптимального пути на графе локальных расстояний между временными отсчетами векторов распознаваемого и эталонного спектров (ДП-метод). В 1969 г. автором совместно с сотрудниками Московского государственного НИИ радио была опубликована статья [3], в которой дано дальнейшее развитие ДП-метода для исключительно важного практического случая, когда границы распознаваемого слова неизвестны, т. е. для решения задачи обнаружения и распознавания звукосочетаний 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 к выполнению этих работ 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 автора этой статьи в 1970 г. в лаборатории профессора Лоренца (Эдинбургский университет), 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, пока еще не вполне совершенная модель синтезатора русской речи по тексту «Фонемафон-1» (рис. 3) «заговорила» в начале 1970-х гг., и успех в ее создании связан, прежде всего, 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 русском языке 42 фонемы (мельчайшие смыслоразличительные звуковые единицы), каждой из которых соответствует определенный набор формантных параметров (частоты и амплитуды формант). В результате экспериментальных исследований был создан полный набор формантных «портретов» фонем, позволивший впервые осуществить синтез русской речи по произвольному тексту. Позднее появилась улучшенная версия синтезатора – «Фонемафон-2» – с дополнительным блоком преобразования «фонема – аллофон».

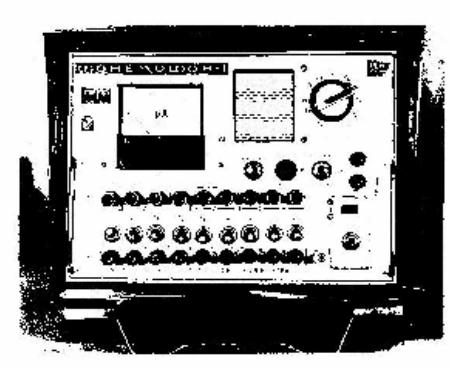

Рис. 3. Синтезатор «Фонемафон-1»

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

Решающую роль в ориентировке на решение прикладных задач сыграло создание в 1974 г. лаборатории обработки речевых сигналов в составе минского отдела Московского отделения Центрального научно-исследовательского института связи (МОЦНИИС). 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 автоматической обзвонки должников за междугородние переговоры. К середине 1980-х гг. эта система была внедрена во многих городах СССР. Успешному внедрению синтезаторов речи предшествовала длительная работа по совершенствованию как качественных показателей синтезированной речи, так и технологии их реализации в качестве нового класса внешних устройств ЭВМ. Основным недос-

татком первых моделей синтезаторов «Фонемафон-1, -2» были недостаточно высокие 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 – «Фонемафон-3» – были введены дополнительные блоки моделирования процессов артикуляции и интонирования речи [4], что существенно повысило качественные показатели синтезированной речи.

Рис. 4. «Фонемафон-3» на выставке в Женеве

В 1979 г. «Фонемафон-3» демонстрировался на Всемирной выставке «Телеком-79» в Женеве (рис. 4). 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 «Фонемафон-4» (1983 г.). Ее серийный выпуск впервые в СССР был налажен на ПО «Кварц» (Калининград) благодаря энтузиазму сотрудников конструкторского отдела, возглавляемого В.П. Афонасьевым. Заметный след в истории речевых технологий 1980-х гг. оставили совместные с ПО «Кварц» опытно-конструкторские разработки и последующее серийное производство системы распознавания речи «Сезам» и речевого терминала «Марс» (рис. 5). Ключевую роль в их создании сыграли разработки сотрудников лаборатории МОЦНИИС 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 упомянутый ранее 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 «Фонемафон», «Сезам» и «Марс», защищена многочисленными авторскими свидетельствами СССР.

К началу 1984 г. относится окончательная формулировка, теоретическая и экспериментальная разработки 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 автора «Методы автоматического синтеза русской речи по тексту», защищенной в 1984 г. в Институте электроники и вычислительной техники Академии наук Латвийской ССР. Позднее полученные результаты были адаптированы для систем синтеза речи на других европейских языках. В частности, благодаря сотрудничеству с профессором Минского института иностранных языков Е.Б. Карневской, к 1987 г. была разработана англоязычная версия синтезатора [5], демонстрировавшаяся на Всемирном конгрессе фонетических наук и получившая высокую оценку англоязычных специалистов. Один

Рис. 5. Речевой терминал «Марс»

из виднейших в мире исследователей речи Г. Фант написал в книге отзывов: «Thank you for demo of really good English synthesis» («Спасибо за демонстрацию очень хорошего синтеза английской речи»).

3. Новейшая история

В 1988 г. на базе лаборатории МОЦНИИС в Институте технической кибернетики АН БССР была создана лаборатория распознавания и синтеза речи, на должность заведующего которой руководством института был приглашен автор данной статьи. Конец 1980-х гг. ознаменовался появлением первых ПК, поэтому в планах работ лаборатории появилась тематика, связанная с оснащением ПК системой речевого ввода-вывода информации. Формантный метод, который долгое время играл ключевую роль в системах синтеза речи по тексту, не подходил для этой цели из-за необходимости большого объема вычислений в реальном времени. В конце 1980-х гг. был предложен новый микроволновой (МВ) метод синтеза речевых сигналов [6], 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 «Фонемафон-5» в виде специализированного ПО синтезатора, ориентированного на работу с внутренней звуковой платой либо с автономным устройством, подключаемым к порту RS-232 (рис. 6). Удивительная для многих компактность его ПО (всего 64 Кб) позволила оснастить синтезом речи уже первые IBM PC-XT и даже отечественные ПК ЕС1840. Синтезатор речи был востребован во многих практических приложениях, но особенно широко он используется незрячими пользователями ПК. (Более сотни комплектов специализированных аппаратно-программных продуктов для незрячих были созданы и распространены Г.В. Лосиком на Украине, в России, и Беларуси в первой половине 1990-х гг.). До сих пор еще его вполне разборчивое звучание можно услышать, приобретя на рынке CD-ROM «Говорящая мышь». В дальнейшем на основе МВ-метода были разработаны версии для чешского и польского языков, а также автономный одноплатный модуль синтеза речи, украинскоязычная версия которого некоторое время работала на линии киевского метро.

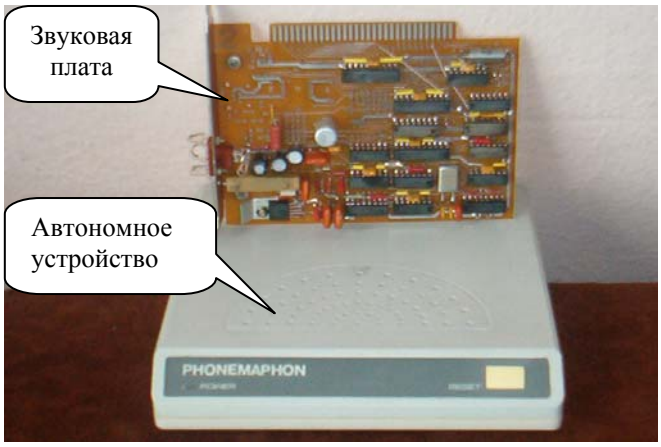

Рис. 6. Микроволновой синтезатор речи

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 «Двуязычный синтез речи – немецкий / русский» (1995 – 1996), выполнявшийся совместно с Дрезденским техническим университетом и финансировавшийся германским научным фондом FTU Karlsruhe. Основные результаты этого проекта изложены в работе [7].

Следующим был проект «Анализ естественного языка и речи» (1996 – 1997), который выполнялся совместно с Саарбрюкенским университетом (Германия), Манчестерским университетом (Великобритания) и Институтом проблем передачи информации (Россия) и финансировался европейским фондом INTAS. Участие в этом проекте было связано с дальнейшим развитием моделей синтеза речи [8] путем их интеграции в системы обработки естественного языка методами компьютерной лингвистики.

Важную роль в интеграции белорусских исследователей в области лингвистики и речи в европейское сообщество сыграло участие в международном проекте «Развитие Европейской компьютерной сети по лингвистике и речи в восточном направлении» (1997 – 1998), финанси-

ровавшемся европейским фондом COPERNICUS. С 1998 г. лаборатория распознавания и синтеза речи ИТК НАН Беларуси является координационным центром этой 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, проект был успешно выполнен и в 1997 г. принят заказчиком. Основные научные результаты этого проекта изложены в работе [9]. Другой коммерческой разработкой стал проект создания интеллектуального телефонного автоответчика, выполнявшийся с 1997 г. по договору с фирмой NovCom NV (США). Суть проекта заключалась в решении задачи распознавания произносимых по телефону имен абонентов и другой служебной информации с тем, чтобы система смогла выполнять функции телефонного автосекретаря. Проект завершен в 1999 г., а его основные научные результаты опубликованы в [10]. В выполнении международных проектов ключевую роль сыграли сотрудники лаборатории Т.В. Левковская, А.Н. Иванов и А.В. Кубашин. Работы по этим проектам явились стимулом к дальнейшему развитию алгоритмов распознавания речи, предложенных еще в 1969 г. [3], но по-прежнему успешно использующихся в современных исследованиях лаборатории.

4. Современные исследования и разработки

В этом разделе описываются разработки последних четырех лет, наиболее интересные как в плане новизны, так и в плане перспектив их практического применения.

4.1. 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 синтезированная речь по качеству оставалась еще далекой от натуральной и обладала узнаваемым машинным акцентом. Причиной этого были не столько уровень наших знаний о процессах речеобразования и о фонетике, сколько нехватка вычислительных ресурсов компьютеров того времени. Сейчас уже можно не ограничивать себя ни объемом оперативной и дисковой памяти, ни требуемым объемом вычислений и приступать к созданию системы синтеза речи по тексту с максимально возможным приближением по звучанию к голосу и манере чтения конкретного диктора.

Такая постановка задачи, хотя и отдаленно, напоминает широко известную биологическую проблему клонирования, когда на основе носителя генетической информации делается попытка воспроизвести копию живого существа. 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). В принципе, в генетике рассматривается и такая возможность, как создание своеобразных «химер» из разнородного генетического материала. В случае клонирования голоса и речи в основу синтеза закладываются, например, акустика голоса одного диктора, фонетические особенности произношения другого, а просодическая индивидуальность речи – третьего.

Аллофонно-волновой синтезатор речи, на базе которого осуществляется клонирование, состоит из четырех процессоров: лингвистического, просодического, фонетического и акустического (рис. 7). Каждый из процессоров использует для осуществляемых им преобразований специализированные базы данных (БД). В этих БД 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 многих других. Известно, что 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.

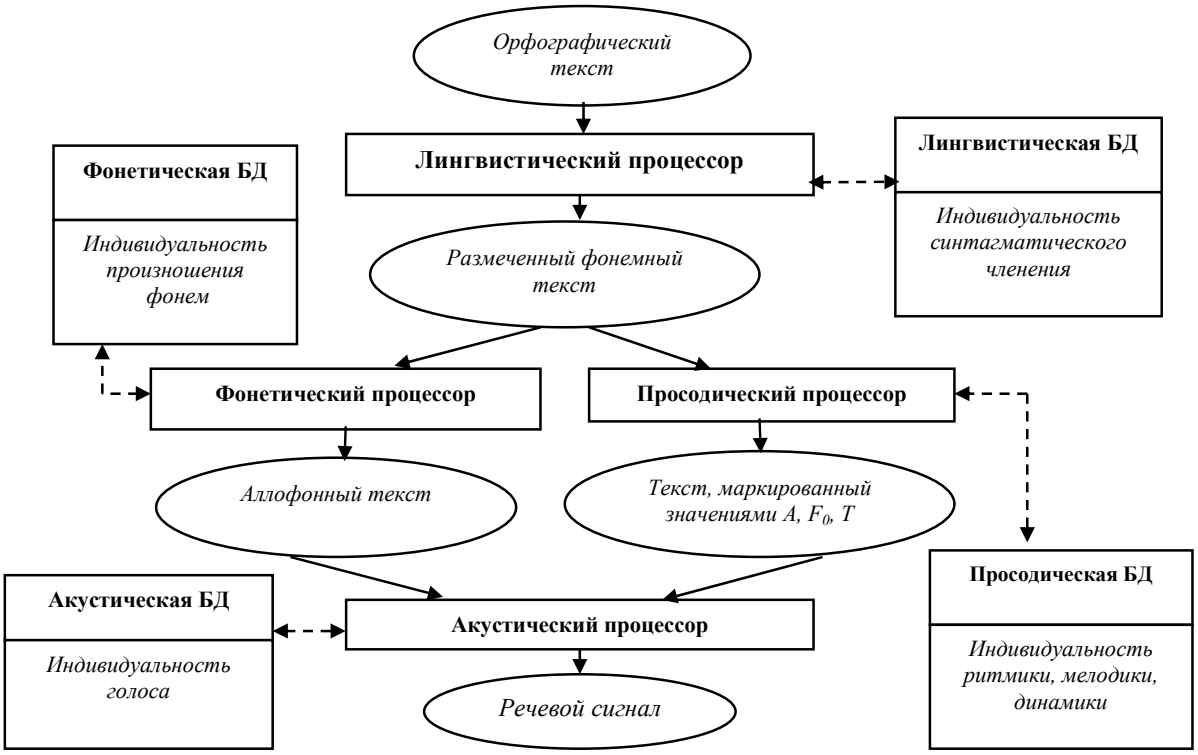

Рис. 7. Общая структурная схема аллофонно-волнового синтезатора речи по тексту

Клонирование персональных фонетических особенностей произношения. В отличие от персональных акустических характеристик голоса, обусловленных в основном статическими параметрами речевого аппарата, фонетические особенности произношения обусловлены главным образом динамикой артикуляторных движений, осуществляемых в процессе речи. Присущие данному индивиду скорость артикуляторных движений, индивидуальные особенности артикуляции того или иного звука (например, /Р/), региональный или иностранный акцент обусловливают возникновение своеобразных позиционных и комбинаторных оттенков фонем и создают уникальную систему аллофонов. Таким образом, успешное клонирование персональных фонетических особенностей произношения, присущих данному человеку в процессе речи, может быть достигнуто путем имитации особенностей фонемно-аллофонного преобразования.

Клонирование персональных просодических характеристик речи. Комплекс просодических (интонационных) характеристик речи, включающий мелодику, ритмику и энергетику, за-

дается 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. Для этой цели используется специально разработанный компактный звуковой массив слов и отрывков текста, начитываемый диктором, голос которого клонируется, в студии или в обычных условиях. Если же диктор физически недоступен, то применяются уже имеющиеся записи его голоса на радио, телевидении и др. Первые результаты по клонированию были получены в 2000 г. и опубликованы в феврале 2001 г. [11]. К концу 2001 г. получен клон женского голоса, а к концу 2003 г. набор клонов состоял уже из трех мужских и двух женских голосов [12].

Отметим некоторые возможные коммерческие и практические аспекты компьютерного клонирования. Вероятно, найдется большое количество пользователей компьютера, желающих, чтобы ПК заговорили их собственными голосами или, например, голосом близкого им человека или любимого актера. Интересным может быть также проект «оживления» голосов давно ушедших от нас великих людей по сохранившимся граммофонным 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 БД голосовых клонов потенциальных правонарушителей.

4.2. 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 (рис. 8) позволяет создавать сценарии диалогов для разнообразных приложений и осуществлять их посредством устно-речевого человеко-машинного общения. Уникальность системы РЕВИРС заключается в том, что в ней реализуется:

- надежное распознавание ключевых слов запроса в непрерывном потоке речи;
- многодикторное распознавание ключевых слов в условиях акустических помех и искажений;
- многоголосый синтез речи по произвольному тексту;
- возможность «клонирования» голоса личности в процессе синтеза речи;
- дуплексный режим в реальном времени (возможность прерывания голосового ответа).

Звуковой сигнал поступает на вход системы распознавания речи, которая осуществляет анализ информативных признаков сигнала и обнаружение слова, его сопоставление с эталонами ключевых слов и принятие решения о произнесенном слове. Если распознающий модуль обнаружил ключевое слово, то 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, осуществляющей лингвистическую, интонационную, фонетическую и акустическую обработки, в результате которых текст преобразуется в звучащую речь заданного голосового клона.

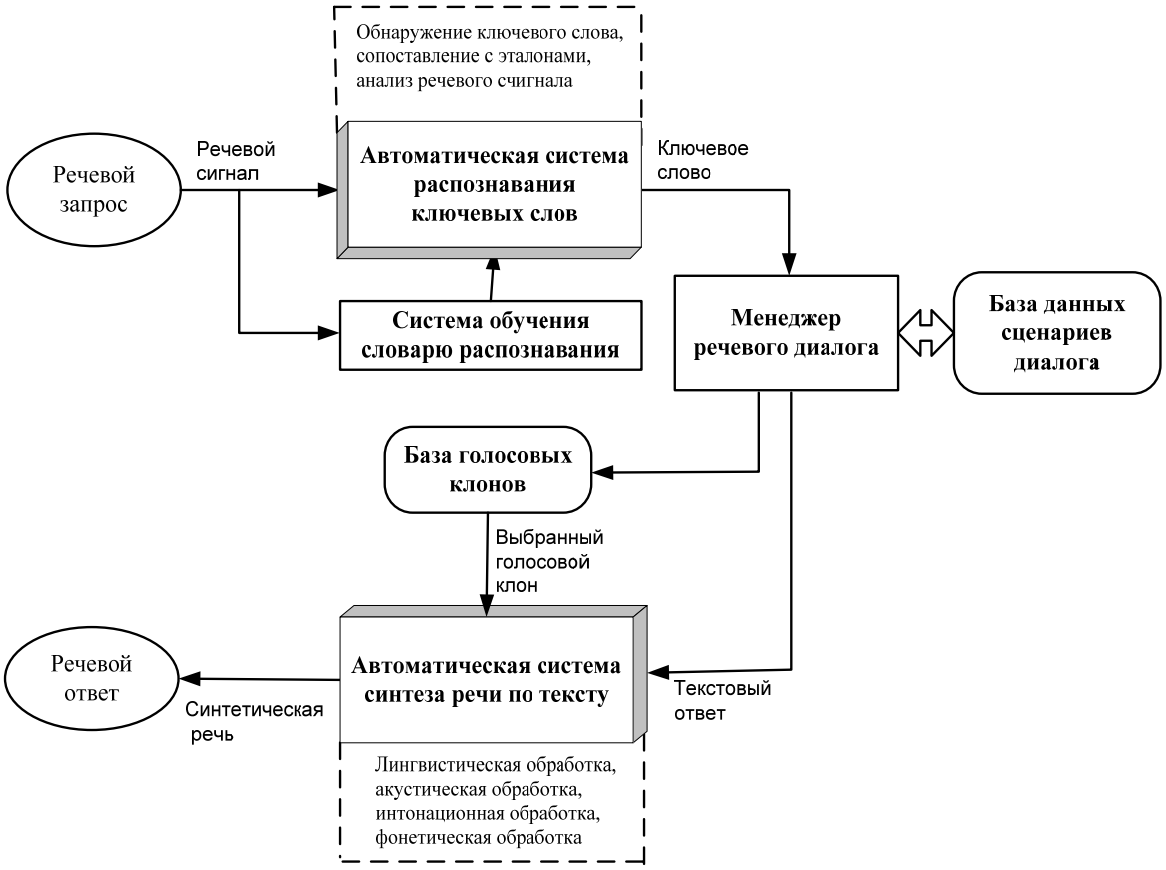

Рис. 8. Структурная схема системы РЕВИРС

Заключение

Изложенный в настоящей статье история создания и развития в Беларуси компьютерных систем распознавания и синтеза речи не претендует на полноту освещения всех научно-технических результатов 40-летних исследований. Не упомянуты работы всех авторов, внесших определенный вклад в развитие этой отрасли знаний в Беларуси (в общей сложности ими опубликовано не менее 300 научных работ, в том числе пять монографий). Несмотря на то, что проделана большая работа и получены ощутимые научные и практические результаты, история речевых исследований в Беларуси на этом не заканчивается. Как и 40 лет назад, разгадка природы человеческой речи увлекательна для молодых исследователей, остаются актуальными и возможные приложения речевых технологий.

Актуальность речевых технологий обусловлена тем, что речевой способ общения с компьютером обладает рядом бесспорных преимуществ:

– удобством, простотой и естественностью процедуры общения, требующей минимума специальной подготовки;

– возможностью использования для связи с компьютером мобильных и обычных телефонов и уже существующих коммуникационных сетей;

– отсутствием ручных манипуляций при вводе информации и уменьшением зрительного напряжения при получении информации;

– существенным ускорением процессов передачи и восприятия информации.

Первые два преимущества смогут обеспечить неограниченное возрастание числа пользователей информационными ресурсами за счет сокращения психологического и физического расстояний между человеком и компьютером. Вторые два обеспечат оперативность и мобильность их взаимодействия. К сожалению, эти преимущества, как и 40 лет назад, осознаются далеко не всеми разработчиками компьютерных систем, продолжающими упорно придерживаться только традиционных средств общения – клавиатуры и дисплея. Хочется надеяться, что ре-

чевые технологии человеко-машинного взаимодействия со временем будут более широко востребованы в Беларуси, в частности в разработках нашего института.

Список литературы

1. Lobanov B.M. More About Speech Signal and the Main Principles of its Analysis // IEEE Transactions on Audio and Electroacoustics. – 1970. – № 3. – P. 316-318.

2. Lobanov B.M. Classification of Russian Vowels Spoken by Different Speakers // The Journal of the Acoustical Society of America. – 1971. – № 2 (2). – P. 521-524.

3. Лобанов Б.М., Слуцкер Г.С., Тизик А.П. Автоматическое распознавание звукосочетаний в текущем речевом потоке // Тр. НИИР. – М., 1969. – С. 67-75.

4. Lobanov B.M. Articulatory-Formant Speech Synthesis from Printed Text // Proceedings of Franco-Sovietique Symposium on Speech. – Paris, 1981. – P. 221-251.

5. Lobanov B.M. The Phonemaphon Text-to-Speech System // Proceedings of the XI ICPhS. – Tallin, 1987. – P. 100-104.

6. Lobanov B.M, Karnevskaya E.B. MW – Speech Synthesis from Text // Proceedings of the XII ICPhS. – Aix-en-Provense, Franse, 1991. – P. 387-391.

7. A Bilingual German/Russian Text-to-Speech System / B.M. Lobanov et al. // Proceedings of the 3nd International Workshop «Speech and Computer» – SPECOM'98. – St.-Petersburg, 1998. – P. 327-330.

8. Generation of Intonation and Accentuation of Synthetic Speech on the Base of Morpho-Syntactic Knowledge / Lobanov B.M. et al. // Proceedings of the International Workshop «Integration of Language and Speech». – Moscow, 1996. – P. 11-28.

9. Lobanov B.M., Levkovskaya T.V. Continuous Speech Recognizer for Aircraft Application // Proceedings of the 2nd International Workshop «Speech and Computer» – SPECOM'97. – Cluj-Napoca, 1997. – P. 97-102.

10. An Intelligent Answering System Using Speech Recognition / Lobanov B.M. et al. // Proceedings of the 5th European Conference on Speech Communication and Technology – EUROSPEECH'97. – Rhodes, Greece, 1997. – V. 4. – P. 1803-1806.

11. Синтезатор персонализированной речи по тексту «ЛобаноФон-2000» / Б.М. Лобанов и др. // Тр. Междунар. конф., посвященной 100-летию российской экспериментальной фонетики, Санкт-Петербург, 1 – 4 февраля 2001 г. – С. 101-104.

12. Lobanov B.M., Tsirulnik L.I. Phonetic-Acoustical Problems of Personal Voice Cloning by TTS // Proceedings of the International Conference «Speech and Computer» – SPECOM'2004, St.-Petersburg, 2004. – P. 17-21.

Поступила 10.11.04

Объединенный институт проблем информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: lobanov@newman.bas-net.by

B.M. Lobanov

THE HISTORY OF CREATION AND DEVELOPMENT SPEECH RECOGNITION AND SYNTHESIS SYSTEMS IN BELARUS

The 40 years history of the creation and development speech recognition and synthesis systems in Belarus is presented. The last scientific and practical results are also described and, in particularly, a new direction in the development of text-to-speech synthesis and automatic speech recognition is pointed out. It is defined as a computer means of personal voice cloning and virtual voice chat systems.