ИНФОРМАТИКА

2006 **апрель-июнь** **№ 2**

УДК 395.521

Л.И. Цирульник

АВТОМАТИЗИРОВАННАЯ СИСТЕМА КЛОНИРОВАНИЯ ФОНЕТИКО-АКУСТИЧЕСКИХ ХАРАКТЕРИСТИК РЕЧИ

Описывается технология клонирования фонетико-акустических характеристик голоса и дикции в системе синтеза речи по тексту. Рассматривается процедура выбора базового набора элементов речи, формирования корпуса текстов и фонограмм записей естественной речи, создания индивидуализированных фонетико-акустических баз данных (БД). Приводится описание разработанной системы клонирования фонетико-акустических характеристик речи. Система осуществляет сегментацию и аллофонную разметку естественного речевого сигнала, выбор фонетико-акустических речевых единиц и их запись в формируемую БД. Дается MOS-оценка правдоподобия синтезированного речевого клона. Показываются области практического применения системы.

Введение

Современное развитие вычислительной техники и систем синтеза речи по тексту позволяет ставить и решать различные задачи, направленные на повышение качества синтезированной речи. Одна из таких задач – максимальное приближение характеристик синтезированной речи к персональным характеристикам естественной речи конкретного диктора – получила название клонирования. Впервые она была поставлена и детализирована в работах [1, 2]. При этом основными характеристиками речи считаются акустические свойства голоса (тембр, высота голоса и др.), фонетические особенности произношения (акцент, особенности дикции и др.) и просодические (интонационные, динамические, ритмические) свойства речи диктора. Для синтеза речи используется аллофонно-волновой метод как один из наиболее подходящих для воспроизведения индивидуальных особенностей речи диктора [3].

Общая структурная схема аллофонно-волнового синтезатора речи описана в работе [4]. Синтезатор состоит из четырех процессоров: лингвистического, просодического, фонетического и акустического. Лингвистический процессор преобразует входной орфографический текст в размеченный фонемный текст, который поступает затем на вход фонетического процессора, осуществляющего преобразование «фонема – аллофон», и просодического процессора, устанавливающего текущие значения амплитуды, частоты и длительности фонем в каждой синтагме текста (под синтагмой понимается слово или группа слов, отделяемых паузами в звучащей речи). Просодически размеченный аллофонный текст подается на вход акустического процессора, который генерирует речевой сигнал.

При синтезе речи по тексту каждый из процессоров использует не только общеязыковые БД и правила, но и индивидуализированные БД и правила, с помощью которых реализуется процесс клонирования индивидуальности голоса и манеры чтения конкретной личности. Для обеспечения высокого качества синтезированной речи общеязыковые БД должны содержать достаточно полный набор фонетических, просодических и акустических элементов речи, в то время как для обеспечения процесса клонирования индивидуализированные БД должны содержать все необходимые характеристики речи конкретной личности.

Процесс создания индивидуализированных БД включает следующие этапы:

– *выбор базового набора фонетико-акустических элементов речи* для синтеза речи по тексту и его уточнение в процессе клонирования голоса конкретной личности;

– *формирование представительного текстового корпуса* (набора текстов) и соответствующих этим текстам фонограмм речи (речевой базы) диктора, для которого реализуется процесс клонирования;

– *обработка созданной речевой базы*, включающая *сегментацию* на элементарные участки и их аллофонную *маркировку* с сохранением полученного набора элементов в индивидуализированных БД.

1. Выбор базового набора фонетико-акустических элементов речи

В качестве базового набора фонетико-акустических единиц могут выступать аллофоны [5], дифоны [6], слоги [7] либо их комбинация – мультифоны [8]. Описываемая автоматизированная система ставит целью создание достаточного набора аллофонов и составляемых из них мультифонов для клонирования фонетико-акустических характеристик речи желаемого диктора.

В русском языке насчитываются 42 фонемы, из них 6 гласных и 36 согласных. В потоке речи фонемы в зависимости от их окружения могут изменять свои артикуляторно-акустические характеристики, что приводит к появлению их модификаций, называемых *аллофонами*. Аллофоны подразделяются на позиционные и комбинаторные. Позиционные аллофоны определяются позицией данной фонемы относительно полноударного гласного и принадлежат к одному из следующих типов: полноударный, частично-ударный, первый предударный, не первый предударный и заударный. Появление комбинаторных аллофонов фонемы Φ_i связано с ее ближайшим окружением, т. е. предшествующей в потоке речи фонемой Φ_{i-1}, а также последующей в потоке речи фонемой Φ_{i+1}.

В процессе синтеза речи генерируются следующие позиционные аллофоны гласных: полноударный (0), частично ударный (1), первый предударный (2), не первый предударный (3), заударный (4). Здесь в скобках указан первый индекс аллофона. С учетом левого контекста генерируются следующие комбинаторные аллофоны гласных: после синтагматической паузы (0), после переднеязычных (1), губных (2) и заднеязычных (3) твердых, после /Л/ (4), после /Р/ (5), большинства мягких (6), после /R'/ (7), после /М'/ (8), после /Н'/ (9), после гласных /У/ (10), /О/(11), /А/ (12), /Э/ (13), /Ы/ (14), /И/ (15); всего 16 левых контекстов. Здесь в скобках указан второй индекс аллофона. С учетом правого контекста генерируются следующие комбинаторные аллофоны гласных: перед синтагматической паузой (0), перед переднеязычными и заднеязычными (1) и перед губными (2) твердыми, перед мягкими (4). Здесь в скобках указан третий индекс аллофона. Например, в слове «генерируется» аллофоны гласных представлены следующим образом: «$G'E_{364}N'E_{294}R'I_{071}RU_{414}J'E_{461}CA_{410}$».

Итого, при этих условиях обеспечивается генерация N_v = 5*16*5*6(гласных фонем) = 2400 гласных аллофонов. Их число, реально используемое в синтезаторе с учетом известных закономерностей, – порядка 1700.

Аллофоны согласных генерируются только с учетом левого и правого контекстов. Левый контекст: после паузы (0), после глухих (1) и звонких (2) согласных, после гласных (3). Здесь в скобках указан первый индекс аллофона. Правый контекст: перед паузой (0), перед глухими (1) и звонкими (2) согласными, перед безударными (3) и ударными (4) гласными. Здесь в скобках указан второй индекс аллофона. Например, в слове «генерируется» аллофоны согласных представлены следующим образом: «$G'_{03}EN'_{33}ER'_{34}IR_{33}UJ'_{33}EC_{33}A$».

Итого, при этих условиях обеспечивается генерация N_c = 4*5*36(согласных фонем) = 720 согласных аллофонов. Их число, реально используемое в синтезаторе с учетом известных закономерностей, – порядка 500.

2. Формирование текстового корпуса фонограмм речи

Для получения набора персональных цифровых портретов голоса личности необходимо создать БД звуковых волн аллофонов речи, опираясь на начитанный диктором компактный звуковой массив специально подобранного текста либо используя имеющийся достаточно большой объем записей его голоса при чтении произвольного текста. Результаты, обсуждаемые в данной работе, получены на основе записи специального звукового массива, включающего набор русских слов в количестве, равном числу используемых аллофонов. Каждое из слов отбиралось исходя из критерия наилучшей репрезентации данного аллофона. Общее количество аллофонов, используемое при синтезе речи, может варьироваться в широких пределах в зависимости от требуемого качества синтезированной речи. Созданы списки слов, включающие минимально-необходимый (до 500) и расширенный (свыше 2000) наборы аллофонов. Используются также два типа специально подобранных фонетически репрезентативных текста [9], в

которых распределение частот фонем и других фонетических единиц близко к теоретическому распределению.

3. Обработка речевой базы: сегментация и аллофонная маркировка

Сегментация и аллофонная маркировка речевой базы являются весьма ответственными и трудоемкими процедурами, выполняемыми до сих пор, зачастую, вручную [10]. Основная идея автоматизации процессов сегментации и аллофонной маркировки заключается в реализации алгоритмов переноса меток начала и конца аллофонов с синтезированного сигнала на естественный речевой сигнал, произнесенный клонируемым голосом. Алгоритм переноса меток с одного сигнала на другой реализуется известными методами динамического временного сопоставления (ДП-методами). Для синтеза сигнала используется многоголосая БД аллофонов, полученная с помощью процедуры «ручного» клонирования [10]. Для автоматического переноса меток выбирается один из синтезированных голосов, наиболее близкий к клонируемому.

Система сегментации и аллофонной маркировки (рис. 1) выполняет следующие функции:

– преобразование исходного орфографического текста (эталонный набор русских слов для клонирования) в аллофонный текст;

– синтез речевого сигнала (РС) и его аллофонную разметку;

– выделение спектральных признаков речевого сигнала;

– автоматический перенос меток аллофонов с синтезированных спектральных параметров на естественный речевой сигнал и автоматическую маркировку аллофонных сигналов.

Алгоритмы работы системы достаточно полно описаны в работе [11].

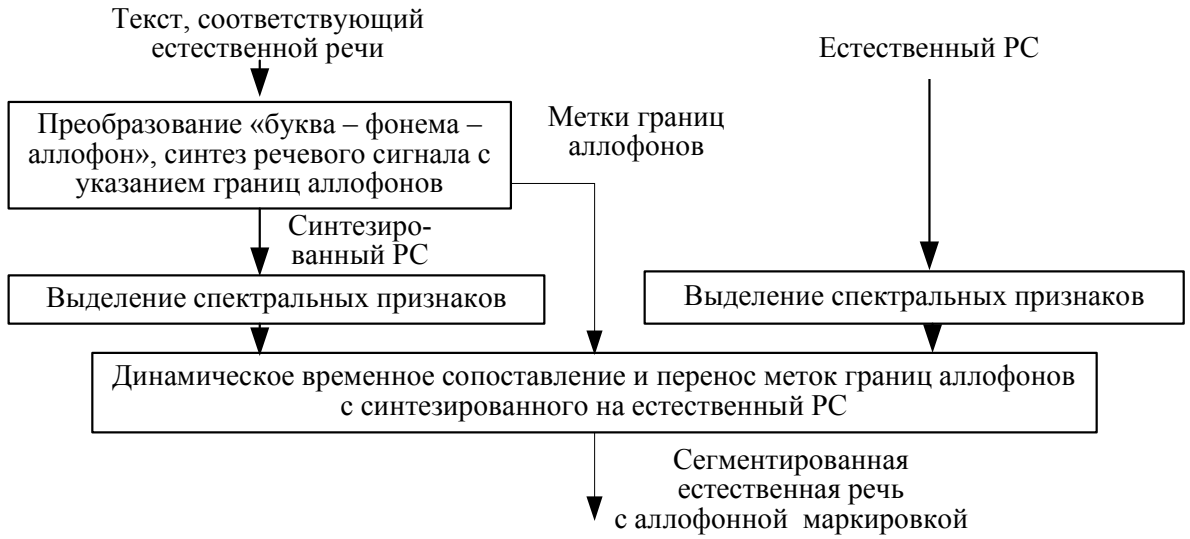

Рис. 1. Структурная схема автоматизированной системы сегментации и аллофонной маркировки

4. Общая схема и принципы работы автоматизированной системы клонирования фонетико-акустических характеристик речи

Описываемая система – программа Clonator – предназначена для автоматизированного создания БД фонетико-акустических характеристик голоса и дикции личности. В системе реализованы описанные выше этапы создания индивидуализированных БД, ответственных за синтез акустических свойств голоса и фонетических особенностей произношения. За рамками данной работы остались вопросы создания индивидуализированных БД, содержащих просодические (интонационные, динамические, ритмические) свойства речи диктора.

Технология создания фонетико-акустической БД с использованием системы, взаимодействие блоков системы, входные и выходные данные показаны на рис. 2.

Входные данные системы:

1. Существующая фонетико-акустическая БД синтезатора (БД звуковых волн аллофонов). Каждый аллофон хранится в виде оцифрованной звуковой волны в отдельном файле в формате WAVE PCM, причем имя файла совпадает с названием соответствующего аллофона.

Рис. 2. Общая функциональная схема автоматизированной системы клонирования

2. Текст, соответствующий естественной речи (текстовая база). Текстовая база содержится в .txt-файле, каждая синтагма записана в отдельной строке файла.

3. Естественная речь (речевая база для клонирования). База содержит набор речевых синтагм, каждая из которых хранится в виде оцифрованной звуковой волны в отдельном файле в формате WAVE PCM.

4. Список сегментов, которые будут помещены в создаваемую фонетико-акустическую БД. Список хранится в отдельном .txt- файле, для каждого сегмента указано, из какой синтагмы он будет получен.

Выходными данными системы является набор речевых сегментов, помещаемых в фонетико-акустическую БД. Каждый из сегментов сохраняется в отдельном файле в формате WAVE PCM с частотой дискретизации 22 кГц и разрядностью квантования 16 бит. Каждый файл сопровождается заголовком, в котором указано имя сегмента.

Создание фонетико-акустической БД, как показано на рис. 2, включает два или более этапа. Последовательность действий, выполняемых на первом этапе, обозначена на рис. 2 сплошными линиями, а на втором и последующих этапах – штриховыми.

На первом этапе в качестве списка сегментов для фонетико-акустической БД используется минимально необходимый или расширенный набор аллофонов, а в качестве речевой базы – звуковой массив, который включает набор русских слов, содержащий все необходимые аллофоны.

Текст, соответствующий звуковому массиву, а также существующая фонетико-акустическая БД являются входными данными для синтеза и аллофонной маркировки речевого сигнала. Аллофонно-размеченный синтезированный речевой сигнал используется для сегментации и аллофонной маркировки естественного РС.

Следующая функция – прослушивание и, при необходимости, ручная корректировка границ аллофонов – реализуется экспертом-фонетистом, который может посмотреть осциллограмму речевого сигнала, прослушать любой его участок, передвинуть метки границ аллофонов. И, наконец, пользователь может прослушать все слова, содержащие указанный речевой сегмент, и установить, из какого именно слова указанный сегмент будет помещен в БД. Результатом первого этапа является фонетико-акустическая БД нового диктора, содержащая минимально необходимый (или расширенный) набор звуковых волн аллофонов.

На втором этапе в синтезаторе речи используется вновь созданная фонетико-акустическая БД. Эксперт-фонетист прослушивает синтезированный по различным текстам речевой сигнал, оценивает его качество, выделяет речевые участки, качество звучания которых неудовлетворительно, и формирует список сегментов для пополнения фонетико-акустической БД. В качестве таких сегментов могут выступать аллофоны, мультифоны (сочетания аллофонов), слоги и другие сегменты речи.

Сформированный список сегментов для пополнения фонетико-акустической БД, дополнительная речевая база, содержащая эти сегменты, и соответствующий текст снова подаются на вход системы. В качестве дополнительных речевых баз используются записи двух специально подобранных фонетически репрезентативных текстов.

Второй этап создания фонетико-акустической БД может повторяться несколько раз, при этом после очередного прослушивания синтезированного РС и оценки его качества эксперт может формировать все новые списки сегментов для пополнения БД, с каждым шагом повышая качество синтезируемой речи.

5. Пользовательский интерфейс системы

Пользовательский интерфейс системы (рис. 3) реализован в среде MS Visual C++ 6.0 на базе MFC-библиотеки. Приложение имеет мультидокументный тип, что дает возможность пользователю варьировать наборы входных данных и делает систему более гибкой и легко настраиваемой.

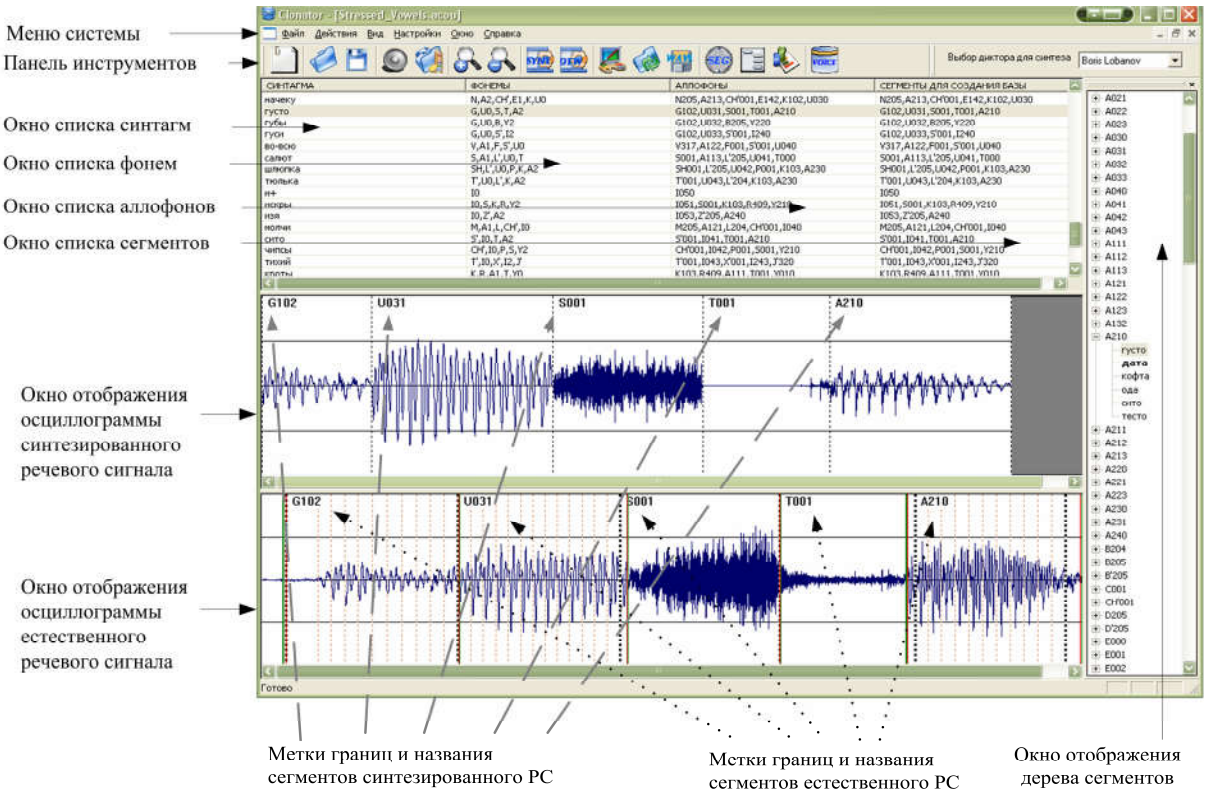

Рис. 3. Общий вид и основные блоки пользовательского интерфейса системы

Пользовательский интерфейс включает следующие основные блоки:

– меню, которое позволяет, в частности, комбинировать различные входные данные системы (звуковые и текстовые файлы), устанавливать параметры вычисления спектральных характеристик сигналов, указывать тип сегментов, помещаемых в фонетико-акустическую БД;

– панель инструментов, содержащую наиболее часто используемые элементы меню;

– блок окон, в котором отображается список синтагм (орфографический текст), соответствующие каждой синтагме фонемный и аллофонный тексты и набор сегментов, помещаемый в фонетико-акустическую БД;

– окна отображения осциллограмм естественного и синтезированного речевых сигналов с указанием меток границ сегментов и названий сегментов;

– окно просмотра дерева сегментов и выбора «активной» синтагмы, из которой будет выделен соответствующий сегмент и помещен в БД.

5.1. Основные функции меню системы

Меню включает, в частности, следующие функционально значимые элементы, отражающие основные действия системы.

Выбор типа сегментов для создаваемой БД. Данный элемент меню позволяет пользователю указать тип сегментов для пополнения фонетико-акустической базы. Система предлагает следующие типы: все аллофоны; различные типы аллофонов гласных (полноударные, частично ударные, первые предударные, не первые предударные и заударные); согласные; различные комбинации диаллофонов (гласный-гласный, согласный-гласный, согласный-согласный); слоги. Выбор сегментов для пополнения фонетико-акустической базы используется на втором этапе работы системы (см. разд. 4). Система при этом просматривает все синтагмы входного списка и находит требуемые сегменты, которые отображает в окне списка сегментов. Все найденные сегменты помещаются также в окно отображения дерева сегментов (см. рис. 3), причем для каждого из них указываются синтагмы, его содержащие.

Статистика сегментов. Окно статистики сегментов (рис. 4), отображаемое при выборе данного элемента меню, используется для просмотра частоты встречаемости каждого сегмента, а также для просмотра «активных» синтагм, сегменты из которых будут помещены в БД. В окне статистики синтагм отображается номер каждого сегмента, его название, количество таких сегментов во входной речевой базе и «активная» синтагма для каждого сегмента. Кроме того, в окне статистики можно просмотреть общее количество сегментов и число сегментов, встретившихся не более или не менее заданного числа раз.

Рис. 4. Окно статистики сегментов

Сохранить сегменты в фонетико-акустической БД. При выборе данного элемента меню открывается окно просмотра каталогов, в котором пользователь указывает путь и имя каталога для записи в формате .wav-файлов всех выбранных речевых сегментов.

Настройки синтеза речевого сигнала. Элемент меню позволяет устанавливать различные параметры синтеза речи, в частности указывать используемую БД звуковых волн аллофонов.

Пользователь выбирает из списка имя диктора, после чего вновь осуществляются синтез речевого сигнала с использованием БД выбранного диктора, сегментация и аллофонная маркировка естественного РС. При этом набор записей естественной речи и текстовый набор синтагм не изменяются. При наличии нескольких БД звуковых волн аллофонов пользователь может подобрать наиболее подходящую БД для данного набора записей естественной речи.

Настройки сегментации речевого сигнала. Элемент меню позволяет устанавливать параметры вычисления спектральных характеристик сигналов и динамического сопоставления естественного и синтезированного РС. Установка параметров влияет на точность сегментации естественного РС.

5.2. Блок окон просмотра списков синтагм, фонем, аллофонов, сегментов

При работе с блоком окон просмотра списков синтагм, фонем, аллофонов, сегментов (рис. 5) необходимо учитывать следующие особенности:

– при изменении набора сегментов, помещаемых в БД, автоматически обновляется содержимое окна просмотра сегментов;

– при выборе пользователем синтагмы из списка система отображает соответствующие осциллограммы естественной и синтезированной речевых синтагм.

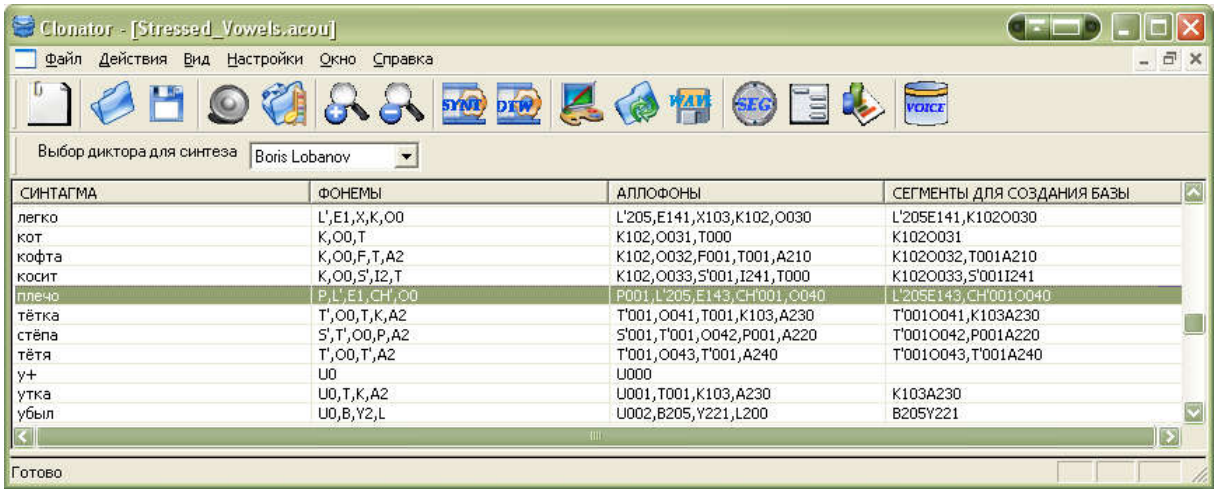

Рис. 5. Блок окон просмотра списков синтагм, фонем, аллофонов, сегментов

5.3. Окно просмотра дерева сегментов

Окно просмотра дерева сегментов (см. рис. 3) предназначено для просмотра сегментов, добавляемых в фонетико-акустическую БД, просмотра соответствующих синтагм и для указания «активной» синтагмы.

Верхний уровень дерева – это набор сегментов для фонетико-акустической БД. Дочерние элементы для каждого из них (элементы второго уровня) – синтагмы, в которых данный сегмент содержится. «Активная» синтагма выделяется жирным шрифтом. Пользователь может прослушать звучание данного сегмента в каждой из синтагм и, используя всплывающее меню, изменить «активную» синтагму.

мента синтагмы;

– при выделении элемента второго уровня (синтагмы) в окнах осциллограмм отображаются естественный и синтезированный сигналы указанной синтагмы.

5.4. Окна отображения осциллограмм речевых сигналов

Окна предназначены для просмотра осциллограмм естественного и синтезированного РС и корректировки, при необходимости, результатов работы блока сегментации и аллофонной разметки естественного РС.

В каждом из окон отображается осциллограмма сигнала, метки границ сегментов и названия сегментов. На рис. 6 отображены фрагменты осциллограмм естественного и синтезированного РС для синтагмы «*лучи заходящего солнца*», в естественном РС выделен слог «*чи*».

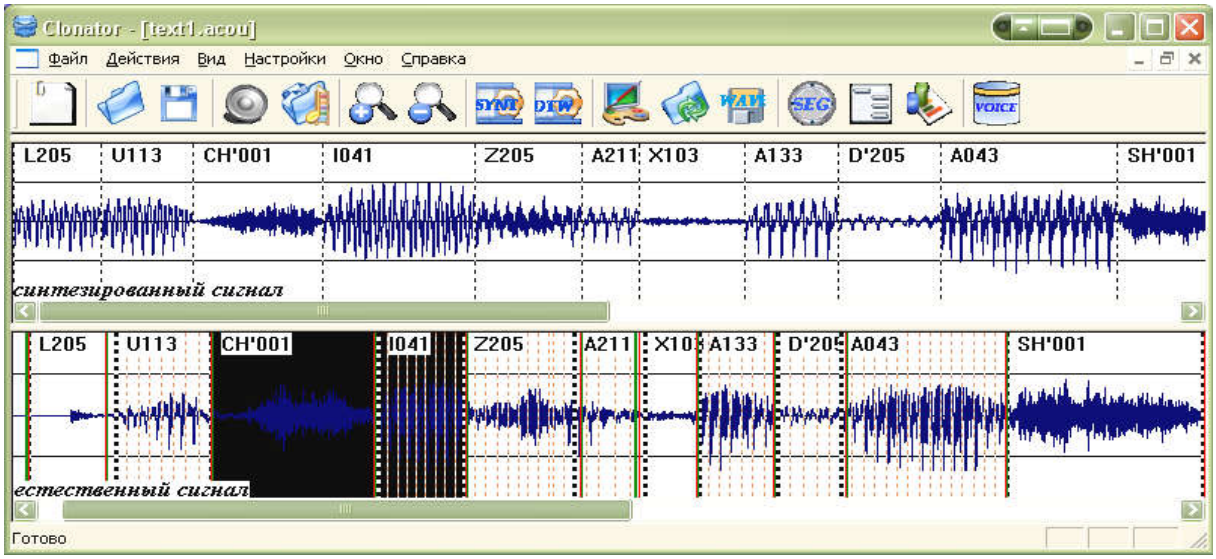

Рис. 6. Отображение осциллограмм речевых сигналов, меток границ сегментов и названий сегментов

Пользователь может выполнить в окнах следующие действия:
– изменить масштаб отображения осциллограммы;
– выделить часть осциллограммы сигнала (рис. 6);
– прослушать выделенную часть сигнала либо весь сигнал;
– изменить положение меток границ сегментов.

6. Оценка степени сходства синтезированного клона с естественной речью

Существует несколько методов оценки качества синтезированной речи [12–15], основанных на расчете корреляции между естественным и синтезированным речевыми сигналами в пространстве различных параметров сигнала. Однако даже лучшие из них не дают результат, приближающийся к результатам субъективной оценки. Поэтому в экспериментах по определению степени сходства синтезированного клона с естественной речью (т. е. правдоподобия речевого клона) предпочтение было отдано оценке субъективного мнения, так называемой MOS-оценке. Методика проведения эксперимента основывалась на Рекомендации P.85 ITU-T «Метод субъективной оценки качества речи устройств речевого вывода» [16], но была адаптирована для данной задачи.

В качестве стимулов использовались пары фраз одинакового текстового содержания. При этом первая фраза в паре являлась записью естественной речи диктора Д1, вторая фраза – либо записью речи того же диктора с искусственно внесенными незначительными мультипликативными искажениями, либо записью его синтезированного клона (клона голоса диктора Д1), полученного в соответствии с описанной выше технологией, либо записью клона другого мужского голоса (клона голоса диктора Д2). Всего использовалось 60 стимулов (по 20 для каждой группы), подаваемых для прослушивания в случайном порядке.

В соответствии с избранной методикой аудиторам предлагалось оценить сходство второго голоса с первым по пятибалльной шкале. В эксперименте по оценке правдоподобия синтезированного речевого клона участвовали восемь аудиторов. Обобщенная MOS-оценка результатов (рис. 7) подтверждает эффективность предложенной технологии клонирования персональных характеристик голоса и дикции.

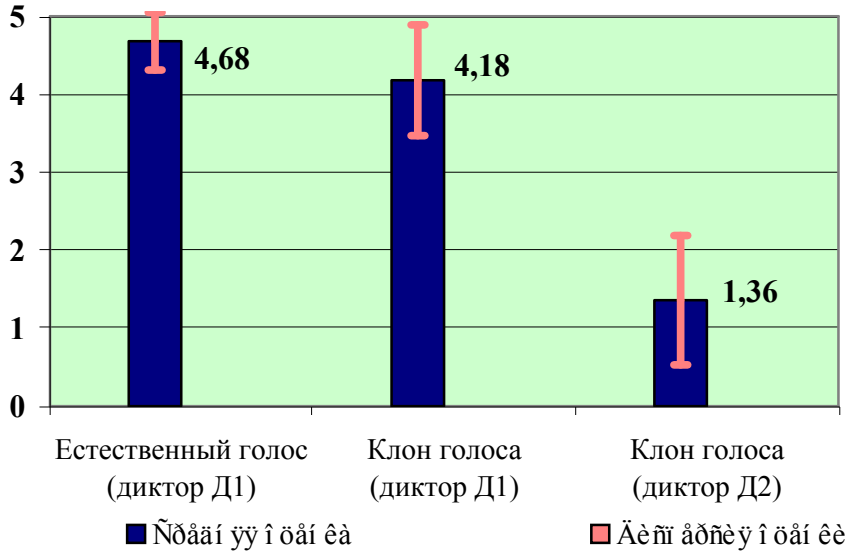

Рис. 7. Обобщенная MOS-оценка правдоподобия речевого клона

Заключение

Разработанная система позволяет автоматизировать трудоемкий процесс создания БД индивидуальных фонетико-акустических особенностей голоса личности по фонограммам речи. Как показал опыт, создание минимально необходимого набора аллофонов опытным фонетистом вручную занимает несколько недель рабочего времени, а расширенного набора – более двух месяцев. При использовании данной системы создание БД, содержащей полный набор звуковых волн аллофонов, занимает не более одного часа машинного времени, при этом качество созданного автоматически речевого клона не уступает качеству речевого клона, полученного вручную.

Кроме очевидного практического применения – создания индивидуализированных речевых БД для последующего высококачественного синтеза речи по тексту с манерой чтения конкретного человека и его голосом, система клонирования фонетико-акустических характеристик речи может использоваться для воссоздания голосов известных личностей по имеющимся фонограммам их речи [17]. Другим важным практическим приложением системы может стать ее использование в криминалистике для создания банка голосовых клонов (цифровых портретов голоса) и оперативной идентификации личности по произвольному отрезку его речи [18].

Проведенное исследование и его практическая реализация были выполнены при поддержке европейского фонда INTAS в рамках проекта «Разработка многоголосовой и многоязыковой системы синтеза и распознавания речи (языки: белорусский, польский, русский)» в соответствии с грантом INTAS № 04-77-7404.

Список литературы

1. Лобанов, Б. М. Компьютерное «клонирование» персонального голоса и речи / Б.М. Лобанов // Новости искусственного интеллекта. – 2002. – № 5(55). – С. 35–39.

2. The AT&T Next-Gen TTS System / M. Beutnagel [et al.] // Proc. of the Joint Meeting of ASA, EAA, and DAGA. – Berlin, Germany, 1999.– P. 41–44.

3. Lobanov, B.M. TTS-Synthesizer as a Computer Means for Personal Voice Cloning (On the example of Russian) / B.M. Lobanov, E.B. Karnevskaya // Phonetics and its Applications. – Stuttgart: Franz Steiner Verlag, 2002. – P. 445–452.

4. Лобанов, Б.М. Синтез речи по тексту / Б.М. Лобанов // Четвертая Междунар. летняя школа-семинар по искусственному интеллекту: сб. науч. тр. – Минск: Изд-во БГУ, 2000. – С. 57–76.

5. Skrelin, P. Allophone-Based Concatenative Speech Synthesis System for Russian / P. Skrelin // Proc. of International Conference TSD '99. – Berlin, 1999. – P. 156–159.

6. Beutnagel, M. Diphone synthesis using unit selection / M. Beutnagel, A. Conkie, A. Syrdal // Proc. of the 3rd International Workshop of Speech Synthesis. – Jenolan Caves, Australia, 1998. – P. 77–80.

7. Law, K. Cantonese Text-To-Speech Synthesis Using Sub-syllable Units / K. Law, T. Lee, W. Lau // Proc. of the International Conference «EuroSpeech'2001». – Aalborg, Denmark, 2001. – Vol. 2. – P. 991–994.

8. Breuer, S. Phoxsy: Multi-phone Segments for Unit Selection Speech Synthesis / S. Breuer, J. Abresch // Proc. of the International Conference «InterSpeech'2004». – Jeju Island, Korea, 2004. – Vol. 2. – P. 983–986.

9. База речевых фрагментов русского языка «ISABASE» / Д.С. Богданов [и др.] // Интеллектуальные технологии ввода и вывода информации. – М., 1998. – С. 20–23.

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

11. Система сегментации речевого сигнала методом анализа через синтез / Б.М. Лобанов [и др.] // Известия Белорусской инженерной академии. – 2004. – № 1(17)/1. – С. 112–114.

12. Thorpe, L. Performance of current perceptual objective speech quality measures / L. Thorpe, W. Yang // Proc. of IEEE Workshop on speech coding. – Berlin, Germany, 1999. – P. 144–146

13. Chen, J.-D. Objective distance measures for Assessing Concatenative Speech Synthesis / J.-D. Chen, N. Campbell // Proc. of the International Conference «EuroSpeech'1999». – Budapest, Hungary, 1999. – Vol. 2. – P. 611–614.

14. Chu, M. An objective measure for estimating MOS of synthesized speech / M. Chu, H. Peng // Proc. the International Conference «EuroSpeech'2001». – Stockholm, Sweden, 2001. – P. 2087–2090.

15. Wouters, J. Perseptual evaluation of Distance Measures for Concatenative Speech Synthesis / J. Wouters, M. A. Magon // Proc. of the International Conference ICSPL'98. – Helsinki, Finland, 1998. – P. 2747–2750.

16. A method for subjective performance assessment of the quality of speech voice output devices. ITU-T Recommendation P. 85. ITU-T, 1994.

17. Лобанов, Б.М. Персональные особенности синтагматического членения речи телеведущего Ю.Сенкевича / Б.М. Лобанов, Л.И. Цирульник // Компьютерная лингвистика и интеллектуальные технологии: тр. Междунар. конф. «Диалог'2004». – М.: Наука, 2004. – С. 419–423.

18. Система экспресс-идентификации голоса личности методом клонирования акустических характеристик речи / Б.М. Лобанов [и др.] // Тез. докл. Междунар. конф. «Теория и практика речевой коммуникации». – М., 2004. – С. 23–28.

Поступила 11.08.05

Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: Liliya_Tsirulnik@ssrlab.com

L.I. Tsirulnik

AUTOMATED SYSTEM FOR INDIVIDUAL PHONETIC-ACOUSTICAL SPEECH PECULIARITIES CLONING

A technology of individual phonetic-acoustical speech peculiarities cloning is outlined. A selection of backbone speech unit set, formation of text corpus and corresponding natural speech phonograms, and creation of personalized phonetic-acoustical database are explored. An automated system for personalized phonetic-acoustical database creation is presented. The system performs phonetic segmentation and labeling of natural speech signal, selection of phonetic-acoustical speech segments and adding the desired segments to the database. The results of MOS evaluation of the synthesized speech clone are described. Applications of the system developed are discussed.