

ИНТЕГРАЦИЯ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ НА ОСНОВЕ РЕЧЕВЫХ ТЕХНОЛОГИЙ И WEB-СЕРВИСОВ

В. А. Жожикашвили, Р. В. Билик, А. Ю. Трощенко, Н. В. Петухова, М. П. Фархадов

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Рассмотрены технологии компьютерного распознавания и синтеза речи и их применение в системах массового обслуживания, а также технология web-сервисов, разработанная для практической реализации концепции сервис-ориентированной архитектуры. Предложена архитектура следующего поколения систем массового обслуживания с интеллектуальными сервисами.

Светлой памяти

Ивери Варламовича Прангишвили посвящается

ВВЕДЕНИЕ

Эволюция информационных систем, движимая все новыми и новыми человеческими потребностями, приводит к существенным изменениям старых систем, к появлению новых систем и к их интеграции. Речь в данной статье пойдет о такой технологии интеграции отдельных систем, которая даст возможность пользователю получать в режиме самообслуживания самую разнообразную информацию и услуги, не подозревая, что в обслуживании его заявки будут вовлечены многие сервисные центры и сети передачи данных. Достижение такой интеграции основано на применении речевых технологий и web-сервисов.

Предоставление информации и услуг пользователю будет осуществляться через контакт-центры нового типа с интеллектуальными услугами.

Определение *контакт-центра как системы массового обслуживания с интеллектуальными услугами* можно сформулировать следующим образом: это совокупность программных и аппаратных средств и алгоритмов, предназначенных для регистрации обращений пользователей (поступающих по телефону или с помощью других электронных каналов коммуникаций), их маршрутизации, контроля решения задач, поиска необходимой информации, выбора наилучшего варианта и выдачи результирующей информации пользователю [1].

В предлагаемой статье затронуты аспекты распознавания речи, рассмотрены возможности web-сервисов и предложена архитектура контакт-центров следующего поколения, позволяющая использовать технологические новшества и автоматизировать качественно новые функции.

Авторы пользуются случаем отметить ту поддержку, которую оказывал И. В. Прангишвили развитию научной проблематики автоматизированных систем массового обслуживания с интеллектуальными сервисами.

1. РАСПОЗНАВАНИЕ И СИНТЕЗ РЕЧИ

Распознавание речи — это преобразование акустического сигнала человеческой речи в набор слов [2]. Полученные на выходе данной процедуры слова могут быть конечным результатом, как в случае распознавания простых команд, или служить входными данными для дальнейшего анализа с целью выявления *смысла* полученных слов.

Синтез речи — это процедура превращения входного орфографического текста в звучащую речь. Помимо преобразования слов в текст, синтез речи должен учитывать ударения, синтаксис и другие особенности устной речи.

1.1. Технология распознавания речи

В типовой системе распознавания речи на первом этапе происходит *параметризация* речевого сигнала, т. е. отображение оцифрованного речевого сигнала (обычно с частотой 8 кГц) в набор векторов (10–20 параметров), значения которых вычисляются каждые 10–20 мс (т. е. с частотой 50–100 Гц). Таким путем достигается уменьшение объема рабочего материала и представление его в более удобном для работы с математической точки зрения виде.

На следующих этапах происходит поиск наиболее вероятных фонем (фонема — наименьший звуковой элемент человеческой речи, в зависимости от языка чис-



ло фоном может варьироваться около 40), групп фоном и далее наиболее вероятного *слова*, удовлетворяющего данному набору параметров входных данных. На алгоритм поиска влияют такие параметры, как акустическая, лексическая и лингвистическая модели. Для настройки параметров этих моделей используются образцы речевых данных.

На выходе такой системы получают с некоторыми вероятностями либо слово, либо набор слов. Интернет-консорциум W3C (Word Wide Web Consortium) разработал стандарт SRGS 1.0 (см. Приложение), формализующий процесс создания грамматик для приложений и позволяющий разработчикам приложений использовать одно и то же приложение с разными платформами распознавания. Помимо этого стандарта, используется также стандарт SISR 1.0 (см. Приложение).

Современные платформы распознавания речи стремятся обеспечить достойную функциональность для разработки приложений с использованием речевых технологий. Среди наиболее популярных функций, реализуемых в платформах, следующие:

- поддержка нескольких языков и возможность одновременной обработки разных языков;
- возможность прерывания голосового меню (barge-in) (функция barge-in — возможность перебивать систему, произнося команды и голосовую информацию, не дожидаясь окончания проигрывания приветствий и/или меню); это позволяет клиентам, уже знакомым с системой, не тратить время на прослушивание всех проигрываемых пунктов меню, а произносить команды немедленно и поэтому с такой системой значительно приятней общаться, не говоря уже о том, что это экономит время звонящего;
- подавление окружающего шума;
- синтаксис для представления грамматических правил (SRGS);
- динамический выбор акустической модели;
- селективный barge-in, когда прерывание происходит только в случае, если фраза соответствует грамматической конструкции;
- возможность переспросить абонента, если вероятность распознанной фразы низка;
- автотранскрипция для фонетического описания отсутствующих в словаре слов;
- дикторнезависимое распознавание;
- распознавание слитной речи.

Лидерами на рынке распознавания речи являются такие американские компании, как «Nuance», «ScanSoft». Русский язык поддерживают всего несколько платформ, среди них ScanSoft (www.scansoft.com) и Sacramento (www.sacramento.com). Наряду с промышленными платформами имеются также некоторые разработки российских ученых, например компания «Стэл — Компьютерные Системы» (www.stel.ru) в сотрудничестве с ведущими специалистами филологического факультета МГУ создает прототип дикторнезависимой системы распознавания русской речи.

На Западе практическое применение распознавания речи началось в самом конце 1990-х гг. В России первые приложения были созданы в Институте проблем управления РАН для получения справок о рейсах, для заказа такси, для управления кошельками в электронных платежных системах [3—5].

1.2. Технология синтеза речи

Существуют различные методы синтеза речи. Выбор того или иного метода определяется связанными с ним ограничениями.

Самый простой метод — воспроизведение предварительно записанной речи (метод компоновки). Он применяется в синтезаторах с ограниченным словарем, в которых фразы строятся путем конкатенации предварительно записанных речевых фрагментов. Очевидный недостаток данного метода — невозможность воспроизводить динамическую информацию.

В синтезаторах с неограниченным словарем элементами речи являются фонемы или слоги, поэтому в них применяется метод синтеза по правилам, а не простая компоновка. Данный метод весьма перспективен, так как обеспечивает работу с любым необходимым словарем, однако качество речи значительно ниже, чем при использовании метода компоновки. При синтезе речи по правилам применяется параметрический метод кодирования на уровне слогов, в соответствии с которым достигается абстрагирование от речевой волны как таковой и она представляется в виде набора параметров. Параметры отражают наиболее характерную информацию либо во временной, либо в частотной области. Например, речевая волна может быть сформирована сложением отдельных гармоник заданной высоты и с заданными спектральными выступлениями на данной частоте.

Для описания особенностей воспроизведения синтезированной речи, таких как фонетическая транскрипция (произношение), громкость, высота, скорость и т. д., был разработан стандарт SSML 1.0 (см. Приложение).

Наиболее известными системами синтеза русской речи являются продукты компании «Scansoft» (RealSpeak), французской «Digalo» и белорусской компании «Sacramento» (Sacramento TTS).

2. WEB-СЕРВИСЫ

Появление концепции web-сервисов было связано со стремлением компаний развивать межкорпоративное взаимодействие на основе возможностей, предоставляемых Интернетом. Web-сервисы являются практической реализацией идеи сервис-ориентированной архитектуры, которая возникла несколько лет назад и с тех пор весьма успешно развивается.

Каждая информационная система призвана обеспечивать работу тех или иных бизнес-процессов, которые, в свою очередь, можно декомпозировать на более мелкие составляющие — бизнес-функции. Бизнес-функция, дающая конкретный измеримый результат, является минимальной сущностью, имеющей ценность для бизнеса, неким «квантом». Именно ее и можно отождествить с сервисом [6]. Теперь, переходя к терминам информационных систем, назовем *сервисом* ресурс, реализующий бизнес-функцию и обладающий следующими свойствами [6]:

- является повторно используемым;
- определяется одним или несколькими явными технологически независимыми интерфейсами;
- слабо связан с другими подобными ресурсами и может быть вызван посредством коммуникационных протоколов, обеспечивающих возможность взаимодействия ресурсов между собой.

Таким образом, разложив бизнес-процессы компании на бизнес-функции, реализовав эти функции в виде сервисов, имеющих стандартный интерфейс и механизм взаимодействия, мы получаем то, что принято теперь называть сервис-ориентированной архитектурой (SOA — Service-Oriented Architecture) построения бизнес-приложений. Или, переходя на более формальный язык, *сервис-ориентированной архитектурой* называется такая архитектура приложений, в рамках которой все функции приложения являются независимыми сервисами с четко определенными интерфейсами, которые можно вызывать в нужном порядке с целью формирования бизнес-процессов [6]. Один из способов реализации такой архитектуры и заключается в построении web-сервисов (см. Web-Services Architecture в Приложении). *Web-сервисом называется программная система, идентифицируемая строкой URI (Universal Resource Identifier), чьи публичные интерфейсы и привязки определены и описаны посредством XML (eXtensible Markup Language). Описание этой программной системы может быть найдено другими программными системами, которые могут взаимодействовать с ней согласно этому описанию посредством сообщений, основанных на XML и передаваемых с помощью Интернет-протоколов.*

Для транспортировки и преобразования данных в программы и обратно web-сервисы требуют применения нескольких смежных XML-технологий [7]:

- протокол SOAP (Simple Object Access Protocol) — открытый протокол, который определяет специализированный формат для связи Интернет-приложений и сервисов на основе XML и для передачи XML-документов по Сети, а также соглашения RPC-взаимодействий (Remote Procedure Call);
- технология UDDI (Universal Description, Discovery and Integration) — реестр web-сервисов и механизм поиска, служит для хранения и упорядочения деловой информации, а также для нахождения указателей на интерфейсы web-сервисов;
- язык описания web-сервисов (Web-Services Description Language, WSDL) — формат XML-схем, определяющий расширенную структуру описания интерфейсов web-сервисов.

Процесс работы с web-сервисами состоит в следующем. Сначала в реестре UDDI находится нужный web-сервис, удовлетворяющий запрашиваемым потребностям. Затем отправляющий компьютер получает WSDL-документ, описывающий формат предоставления данных этого web-сервиса, и SOAP-процессор отправляющего компьютера преобразует данные из собственного («родного») формата в тип данных, predeterminedный в соответствии с содержащейся в WSDL-файле XML-схемой, и отправляет SOAP-запрос на принимающую сторону. Полученный ответ также проходит процесс преобразования обратно в «родной» формат.

Заметим, что web-сервисы в настоящее время не являются мифом или чистой теорией, существует масса примеров уже реализованных web-сервисов. Чтобы убедиться в этом, достаточно зайти в UDDI-реестр <<http://www.uddi.org/find.html>>. Российские примеры web-сервисов можно посмотреть, например, на сайтах ОАО «Аэрофлот — российские авиалинии» или ЦБ РФ:

- расписание рейсов ОАО «Аэрофлот — российские авиалинии» <<http://webservices.aeroflot.ru>>;

- расписание рейсов и заказ билетов <<http://aviabilet.ru/avia>>;
- web-сервисы ЦБ РФ: котировки валют и металлов, ставки и др. <<http://www.cbr.ru/scripts/Root.asp>>.

3. АРХИТЕКТУРА СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ СЛЕДУЮЩЕГО ПОКОЛЕНИЯ

Итак, мы имеем три крупных блока: контакт-центр, инструменты распознавания и синтеза речи и множество web-сервисов. Соединив все три блока в единую систему, мы сможем получить нечто принципиально новое — систему массового обслуживания следующего поколения с интеллектуальными сервисами. В нашем понимании интеллектуальный сервис в контакт-центре — это полностью автоматизированный сервис, способный полностью удовлетворить информационный или иной запрос абонента без привлечения человека-оператора. Приведем примеры таких сервисов.

- Бронирование и заказ билета по телефону. Голосовое взаимодействие системы и человека осуществляется посредством распознавания и синтеза речи. Логика поиска оптимального билета заложена в приложении интеллектуального сервиса, а данные берутся из множества источников в сети Интернет посредством web-сервисов.
- Голосовая поисковая система. Поиск любой информации и предоставление ее по телефону. Например, поиск кинотеатров рядом с определенной станцией метро, в данном случае используются уже существующие приложения, которые реализуют эту функцию в Интернете. Нет необходимости писать отдельное приложение для голосовой системы.
- Анкетирование. Исходящий обзвон и анкетирование граждан. Ведение диалога с помощью распознавания и синтеза речи.

Принцип работы системы можно пояснить на следующем примере.

1. Абонент направляет звонок в контакт-центр компании с целью совершить бронирование билета. Он не знает ни названия авиакомпании, ни расписания рейсов. Единственные входные данные — пункты отправления и назначения и желаемое время отправления или прибытия.

2. Звонок из телефонной сети общего пользования попадает на голосовой шлюз, задачей которого является преобразование голосового потока из традиционной TDM сигнализации (например, E1 ISDN PRI) в IP, с сигнализацией SIP (см. Приложение) или H.323. По этой сигнализации шлюз обращается к голосовому браузеру — устройству, выполняющему функцию, схожие с обычным web-браузером, — интерпретацию полученных от web-сервера данных в воспринимаемый человеком вид, т. е. в нашем случае в голос. Голосовой браузер определяет, на какой и с какого номера пришел звонок (DNIS, ANI), эти данные он получает из H.323 или SIP сигнализации. Далее производится http-запрос (с параметрами ANI и DNIS) на web-сервер.

3. Web-сервер на основании полученных параметров определяет, какую из служб запустить, и отправляет в ответ голосовому браузеру VXML-файл (см. Приложение) приветствия службы бронирования. Голосовой браузер, получив VXML-страницу, интерпретирует ее содержание. В нашем примере страница содержит простое приветствие и просьбу ввести пункты отправления и назначения, дату и время вылета. Для произнесения могут использоваться как предварительно записанные голосовые файлы, так и синтезируемая речь. Для доступа к системе синтеза речи голосовой браузер использует протокол MRCP (см. Приложение).

4. Абонент произносит запрашиваемую информацию, при этом голосовой браузер в процессе сбора этой информации обращается к модулю распознавания речи. Протоколом взаи-



модействия между этими системами также является MRCP (см. Приложение). Наконец, получив данные с помощью системы распознавания уже в виде параметров, голосовой браузер делает следующий http-запрос (с новыми параметрами — пунктами назначения и отправления, датой и временем отправления) на web-сервер.

5. Web-сервер снова запускает интеллектуальный сервис бронирования, который, в свою очередь, совершает запросы в несколько авиакомпаний (используя web-сервисы) и находит наиболее оптимальный вариант для абонента, исходя из времени вылета рейса, параметров класса и цены. Каждая из авиакомпаний предоставляет web-сервис, возвращающий список рейсов с ценами и наличием билетов. Таким образом, логика выбора авиарейса заложена в нашем интеллектуальном приложении, а информация берется из web-сервисов, предоставляемых авиакомпаниями.

6. Наконец, web-сервер возвращает результат — VXML-страницу, которую должен интерпретировать голосовой браузер. В ней находится название авиакомпании, номер рейса, его параметры, а также вопрос о подтверждении бронирования. Для произнесения информации абоненту голосовой браузер обращается к подсистеме синтеза речи.

Если абонент не желает пользоваться автоматизированной службой, он всегда может быть перенаправлен к оператору контакт-центра.

4. ПРИМЕР: ПОЛУЧЕНИЕ СПРАВОК О РЕЙСАХ С ПОМОЩЬЮ WEB-СЕРВИСА КОМПАНИИ ОАО «АЭРОФЛОТ — РОССИЙСКИЕ АВИАЛИНИИ» И РЕЧЕВОГО ИНТЕРФЕЙСА

В лаборатории автоматизированных систем массового обслуживания Института проблем управления РАН создана система автоматического получения справок о рейсах, использующая речевой интерфейс на базе программы распознавания речи Speech Pearl [4, 8]. Прикладное программное обеспечение реализует управление телефонными линиями, организует диалог с клиентом и собирает в ходе этого диалога данные для обращения в информационную систему. Далее возможны два варианта. По первому варианту голосовой сервер может взаимодействовать непосредственно с базой данных и получать информацию из нее. Недостаток этого варианта заключается в уникальности интерфейса с базой данных. Второй вариант, более универсальный и основанный на современных международных стандартах, состоит в использовании web-сервиса, недавно введенного на сайте компании ОАО «Аэрофлот — российские авиалинии» и позволяющего легко интегрировать приложения, независимо от их платформы и языка разработки. Данные клиента помещаются в оговоренную форму и отправляются на сайт, откуда возвращается список структур, состоящих из ряда полей.

Из возвращенных полей выбираются нужные данные, которые далее озвучиваются с помощью метода компоновки в виде естественной речи.

Web-сервис позволяет получать данные из Центра управления полетами ОАО «Аэрофлот — российские авиалинии». Данные обновляются каждые 5 мин. Предлагаемая услуга бесплатная.

Данный вариант был опробован при разработке речевого интерфейса в системе получения справок по расписанию рейсов ОАО «Аэрофлот — российские авиалинии».

ЗАКЛЮЧЕНИЕ

Интеграция систем массового обслуживания на основе речевых технологий и web-сервисов позволяет ав-

томатизировать качественно новые функции. В случае контакт-центра с интеллектуальными сервисами автоматизация затрагивает выяснение сути звонка (посредством распознавания речи), нахождение запрашиваемой информации (используя данные и функции, предоставляемые web-сервисами) и доведение информации до абонента (посредством синтеза речи). Все процессы происходят без участия человека.

Кроме общеизвестных преимуществ, которые дает применение речевых технологий в системах обслуживания, данная архитектура имеет ряд дополнительных достоинств:

- открытые стандарты позволяют использовать модули от разных производителей (например, голосовой шлюз, голосовой браузер, системы синтеза и распознавания речи, контакт-центр);
- современные технологии web-сервисов позволяют гибко использовать существующие в сети ресурсы для организации интеллектуальных сервисов, без необходимости создавать их заново;
- рост популярности web-сервисов и увеличение их числа в сети Интернет, а также простота пользования ими позволят существенно обогатить возможности голосовых сервисов.

Открытость стандартов и возможность использовать модули разных производителей понижает порог для выхода на рынок программных продуктов, а это означает, что программы, созданные российскими производителями, смогут успешно конкурировать с программами западных гигантов в соответствующих областях.

ПРИЛОЖЕНИЕ.

Краткие характеристики некоторых стандартов

SRGS 1.0 — Speech Recognition Grammar Specification.

Определяет синтаксис для представления грамматических правил в системах распознавания речи, иными словами, для описания структуры данных при их вводе средствами голосового приложения. Данный синтаксис позволяет разработчикам задавать слова и шаблоны словарных форм для сообщения их системам распознавания речи. Синтаксис грамматики представлен в двух форматах: в Augmented BNF Form и XML Form. Спецификация определяет отображение между этими представлениями, так что одно представление может быть трансформировано в другое. <<http://www.w3.org/TR/speech-grammar/>>.

SISR 1.0 — Semantic Interpretation for Speech Recognition.

Определяет процессы семантической интерпретации для распознавания речи и синтаксис и семантический смысл тегов, которые могут быть добавлены к грамматике распознавания речи, в частности в стандарте SRGS. Например, SISR позволяет определить связи между грамматическими правилами и семантикой приложения, что позволяет трактовать как эквивалентные варианты употребления одного и того же элемента, например «да», «ага», «правильно». <<http://www.w3.org/TR/semantic-interpretation/>>.

SSML 1.0 — Speech Synthesis Markup Language.

Создан для описания особенностей воспроизведения синтезированной речи, таких как фонетическая транскрипция (произношение), громкость, высота, скорость и т. д., для различных платформ синтеза речи. <<http://www.w3.org/TR/speech-synthesis/>>.

Web Services Architecture.

Описывает архитектуру web-сервисов. Определяются функциональные компоненты и взаимодействия между ними. <<http://www.w3.org/TR/ws-arch/>>.

SIP: Session Initiation Protocol.

Протокол прикладного уровня, предназначенный для установления, изменения и окончания сеансов с одним или не-

сколькими участниками. Сеансы могут включать в себя телефонные звонки и мультимедиа-конференции, а также задействовать другие мультимедийные ресурсы. <<http://www.ietf.org/rfc/rfc3261.txt>>.

VXML 2.0 — Voice Extensible Markup Language.

Разработан для создания голосовых меню, использующих синтез и распознавание речи, тональный набор DTMF, запись разговора из диалога и других телефонных функций. Основная цель стандарта — обеспечить доступ к web-информации в системах интерактивного речевого взаимодействия. <<http://www.w3.org/TR/voicexml20/>>.

MRCP: Media Resource Control Protocol.

Протокол, обеспечивающий интеграцию средств распознавания речи и преобразования текста в речь с голосовой платформой. <<http://www.ietf.org/internet-drafts/draft-shanmugham-mrcp-07.txt>>.

ЛИТЕРАТУРА

1. *Самолубова А. Б.* Call Center на 100 %. — М.: Альпина Бизнес Букс, 2004. — 308 с.
2. *Survey of the State of the Art in Human Language Technology* <<http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>>.

3. *Жожикашвили В. А., Билик П. В., Вертлиб В. А.* и др. Открытые системы массового обслуживания // Проблемы управления. — 2003. — № 4. — С. 55—62.
4. *Жожикашвили В. А., Жожикашвили А. В., Петухова Н. В., Фархадов М. П.* Применение распознавания речи в автоматизированных системах массового обслуживания // Автоматизация и современные технологии. — 2003. — № 11. — С. 22—28.
5. *The first voice recognition applications in Russian language for use in the interactive information systems* // V. A. Zhzhikashvili, M. P. Farkhadov, N. V. Petukhova and A. V. Zhzhikashvili // Proc. of the Ninth International Conf. «Speech and Computer» SPECOM'2004, Saint-Petersburg. — 2004. — P. 304—307.
6. *Долотин И.* Цикл статей «Технологии web-сервисов». <http://www.ubs.ru/ws/ws_basics1.html>.
7. *Ньюкомер Э.* Web-сервисы. XML, WSDL, SOAP и UDDI. Для профессионалов. — СПб: Питер, 2003. — 256 с.
8. <<http://www.ipu.ru/labs/lab17/frame17.htm>>.

☎ (495) 334-87-10, 334-90-60, e-mail: mais@ipu.ru

Статья представлена к публикации членом редколлегии А. Г. Бутковским. □

УДК 621.396

СИСТЕМЫ МНОГОСКОРОСТНОЙ ОБРАБОТКИ МНОГОМЕРНЫХ СИГНАЛОВ. Ч. II¹

М. К. Чобану

Московский энергетический институт (технический университет)

Рассмотрены методы полиномиального синтеза многомерных неразделимых многоскоростных систем. Приведены основные требования, которым должны удовлетворять синтезируемые цифровые фильтры. Предложены три подхода к синтезу, основанные на применении полиномов Бернштейна, метода лифтинга и метода достройки матрицы.

ВВЕДЕНИЕ

В первой части работы рассмотрены приложения, в основе которых лежит применение многомерных (ММ) неразделимых многоскоростных систем [1]. Известно достаточно много различных методов синтеза ММ цифровых фильтров для многоскоростных систем. Среди них можно отметить два подхода к синтезу фильтров — аналитический и численный. Такое деление весьма условное, так как при аналитическом синтезе в полученной структуре могут оставаться свободные параметры, кото-

рые определяются обычно численными методами, путем минимизации того или иного численного критерия.

В настоящей статье описываются методы аналитического синтеза неразделимых цифровых фильтров с конечной импульсной характеристикой (КИХ) для ММ многоскоростных систем.

1. ОСНОВНЫЕ ТРЕБОВАНИЯ, ПРЕДЪЯВЛЯЕМЫЕ К СИНТЕЗИРУЕМЫМ ФИЛЬТРАМ

За 30 лет развития теории и методов многоскоростной обработки сигналов было предложено много подходов к синтезу ММ цифровых фильтров для банков анализа и синтеза.

¹ Работа выполнена при поддержке РФФИ и японского общества JSPS, грант № 06-07-91751-ЯФ_а.