

М.П. ФАРХАДОВ

РАСПОЗНАВАНИЕ РЕЧИ В СИСТЕМАХ МАССОВОГО ОБСЛУЖИВАНИЯ НАСЕЛЕНИЯ

Фархадов М.П. Распознавание речи в системах массового обслуживания населения.

Аннотация. Статья посвящена использованию современных компьютерных речевых технологий в системах массового обслуживания населения. Рассматриваются проблемы разработки эффективного и удобного речевого интерфейса, методы обнаружения и коррекции ошибок распознавания, модели для сравнительного анализа сценариев речевого диалога, математические модели для вычисления характеристик обслуживания заявок и оптимизации функционирования контакт центров с сервисами самообслуживания на базе распознавания речи.

Ключевые слова: распознавание речи, информационное общество, системы массового обслуживания, обнаружение и коррекция ошибок при распознавании речи, речевой интерфейс, моделирование контакт центров с сервисами самообслуживания.

Farkhadov M.P. Speech recognition in the automated queuing service systems for users.

Abstract. This article is devoted to modern computer speech technology in queuing systems for the population. Problems of working out of the effective and convenient speech interface, methods of detection and correction of errors of recognition, model for the comparative analysis of speech dialogue scenarios, mathematical models for calculation of service characteristics of demands and optimization of functioning contact centers with self-service on the basis of speech recognition are considered.

Keywords: speech recognition, information society, queuing system, voice user interface, detection and correction of errors at speech recognition, modeling contact centers with speech recognition self-services.

1. Введение. Актуальность проблемы доступа людей к информации на переходном этапе продвижения страны и мира к информационному обществу не вызывает сомнений. Быстро развивающиеся информационные и телекоммуникационные технологии проникают во все сектора экономики: в производство, в сферу услуг, в образование, в государственное управление, банковскую сферу, в частный бизнес и др. Наблюдается тенденция к социализации информации – нарастанию количества информации, которая необходима людям в их повседневной жизни. Помимо наличия знаний и информации для развития информационного общества необходимо создание условий для того, чтобы члены общества имели доступ к информации и могли ее использовать. Бурное развитие технологий обостряет существующую во всем мире проблему информационного неравенства. Для России сложилась ситуация, когда темпы роста информатизации оказались выше, чем темпы развития компьютеризации и коммуникаций. Это привело к

обострению проблемы неравного доступа граждан к информации, особенно малообеспеченных людей и населения отдаленных регионов. В качестве одного из средств смягчения этой трудной проблемы рассматриваются речевые технологии, достигшие за последние годы достаточного уровня качества и способные обеспечить гражданам дополнительный, а иногда и единственный канал доступа к информационным и сервисным системам. Создание телефонных сервисов самообслуживания, речевых порталов и речевых браузеров повышает эффективность и открытость автоматизированных систем массового обслуживания и решает актуальную народнохозяйственную и социальную задачу облегчения доступа населения к информации и услугам. В статье рассматриваются различные аспекты проблемы применения речевых технологий в сетевых и телефонных системах массового обслуживания.

2. Современное состояние речевых технологий. Речевые технологии в последние годы достигли достаточно высокого уровня развития. На Западе телефонные сервисы самообслуживания на базе речевых технологий получили широкое распространение, и этот рынок успешно развивается. Так, по данным компании BCC Research, специализирующейся на исследовании рынков и прогнозе успешности внедрения новых технологий [<http://www.bccresearch.com>], предполагаемый рынок технологии распознавания речи до 2015 года будет ежегодно расти на 8,8%, объем рынка к 2015 году оценивается в 58,4 миллиардов долларов США. В нашей стране речевые технологии до сих пор не получили широкого распространения. Наиболее существенными причинами для объяснения такого отставания являются: отсутствие на рынке вплоть до последнего времени надежных распознавателей для русского языка и низкое качество передачи речи в телефонных сетях старого поколения. Не последнюю роль играет также недоверие потенциальных потребителей систем с распознаванием речи к новой технологии, которое объяснялось опасением, что ошибки распознавания будут вызывать раздражение клиентов и нанесут вред имиджу компании.

Однако ситуация к настоящему времени существенно изменилась. Появились доступные для приобретения и использования русскоязычные распознаватели компаний Nuance, США, [<http://nuance.com>] и Loguendo, Италия, [<http://loguendo.com>], находятся на стадии тестирования продукты отечественных компаний Vocative [<http://vocative.ru>] и «Речевые технологии» [<http://www.speechpro.ru/techno/recognition>]. Большой вклад в теоретические исследования и практические разра-

ботки сделан специалистами СПИИРАН [1, 2], ВЦ РАН [3, 4] и др. Качество передачи речи по сети также улучшилось благодаря изменениям в сетевой инфраструктуре: происходит модернизация магистральных сетей, внедряются цифровые технологии, оптические кабели, высокоскоростной цифровой абонентский доступ. Надежность распознавателей заметно повышается.

Таким образом, при современных тенденциях в развитии коммуникационной техники и речевых технологий в нашей стране и при возрастании значимости доступа к информационным ресурсам и сервисам имеются условия для использования речевых технологий для доступа населения к системам массового обслуживания населения. Предлагаемые решения предоставят населению возможность более широко использовать телефон для получения информации и услуг. Эти решения могут использоваться как в сетевых информационных системах, дополняя традиционный Интернет браузер голосовым браузером, так и в центрах телефонного обслуживания, значительно расширяя функции IVR. Сервисы самообслуживания на базе распознавания речи дешевле, чем услуги операторов, они могут функционировать круглосуточно и на разных языках. Помимо этого, услуги и данные могут быть предоставлены комплексно, из разных систем, путем организации интеллектуального диалога с сервером самообслуживания и взаимодействия этого сервера с разнообразными прикладными системами, что также выгодно и удобно для пользователей. Одновременно поставщики услуг, контакт центры и контент провайдеры могут повысить свою эффективность путем внедрения у себя сервисов самообслуживания, что позволит им сократить число операторов, сгладить пиковые нагрузки, снизить расходы на организацию инфраструктуры.

Отдельную область применений речевых технологий составляют нужды людей с ограниченными возможностями. Для таких людей речевые технологии позволяют создать сервисы, которые помогут им получать информацию, образование, профессиональные знания и успешно включаться в общественную и трудовую жизнь.

При внедрении речевых технологий приходится сталкиваться с целым рядом все еще нерешенных проблем, суть которых определяется спецификой самих речевых технологий и психофизиологическими особенностями человека, вступающего во взаимодействие с ними. Наиболее общий характер имеют проблемы, относящиеся к построению эффективного речевого человеко-машинного интерфейса и связанные с повышением устойчивости систем с речевым интерфейсом к ошибкам распознавания и задачи

оптимизации сценария диалога при различных вариантах организации процедур выявления и коррекции ошибок. Другую группу новых задач образуют задачи разработки математических моделей для расчета характеристик обслуживания заявок в центрах обслуживания вызовов, где речевые технологии применяются для организации функций самообслуживания, и задачи оптимизации функционирования таких центров.

3. Разработка речевого интерфейса. Речевой интерфейс отличается от графического наличием более серьезных ограничений по взаимодействию как с машинной стороны, так и со стороны человека. В связи с этим при разработке принципов и методов создания эффективного речевого интерфейса первоочередной задачей является исследование проблематики интерфейса как со стороны распознавателей речи, так и со стороны человека.

В [5, 6] опубликованы результаты экспериментального исследования зависимости качества распознавания речи от различных условий. Для исследования была создана система, обеспечивающая доступ к реальным распознавателям через телефон. Система построена на компьютере Pentium 4 (операционная система Windows 2000/XP), дополненном многоканальной интерфейсной платой Dialogic, к которой подключены 4 телефонные линии. Использовались системы распознавания речи разработки компаний Nuance Communications и Philips.

Для выполнения экспериментов был разработан речевой диалог, созданы речевые блоки и записаны аудио файлы для озвучивания текста.

Тестирование производилось различными людьми разного возраста, пола и национальности, что позволило представить достаточно полный набор вариантов произношения.

Выбор грамматик (грамматика – текстовый файл, описывающий множество слов и фраз, которые могут быть произнесены клиентом на данном шаге диалога) осуществлялся таким образом, чтобы исследовать, с одной стороны, специфичные для систем массового обслуживания грамматики (цифры, числа, даты, месяцы, пин коды, города, улицы, телефонные номера, названия валют и др.), и с другой стороны, охватить все наиболее интересные для исследования аспекты проблем распознавания. Грамматики отличались размером, сложностью логики, длиной произносимых фраз, длиной слов.

Помимо лабораторных данных использовались сохраняемые записи реальных вызовов, поступающих в работающую систему «Авто-секретарь», в которой применено распознавание речи. Для исследова-

ния были взяты двухнедельные фрагменты архива, каждый из которых содержал около 3000 диалогов клиентов с системой. Звуковым файлам с записями речи клиентов были сопоставлены соответственно несколько информационных меток: смысловое значение, качество записанного звука, половая принадлежность клиента, темп и громкость речи. Смысловое значение использовалось для определения правильности распознавания имен. Грамматика системы «Автосекретарь» содержит около 1500 слов, представляющих собой имена, фамилии, отчества. Исследование проводилось путём многократной обработки звуковых файлов системой распознавания при различных значениях параметров распознавателя с последующим запоминанием результатов экспериментов.

В качестве критерия качества распознавания было выбрано относительное число правильно распознанных слов WRR (Word Recognition Rate) $WRR = R/N$, где R – число правильно распознанных слов, N – общее число произнесенных слов.

Были получены зависимости WRR от сложности грамматики, от продолжительности высказывания, типа моделей, влияния шума, стиля произношения, темпа речи, громкости голоса, и другие.

Исследовано также влияние на результаты распознавания параметров распознавателей и величины порога уверенности распознавателя в результате. Полученные зависимости приведены в [5].

Проведенные экспериментальные исследования позволили сделать следующие *выводы о поведении машинной части речевого человеко-машинного интерфейса*:

- существуют зависимости качества распознавания от целого ряда факторов, и эти зависимости выявлены и представлены в [5, 6];

- знание этих зависимостей позволяет разработчику управлять в определенных пределах поведением распознавателя, влияя на содержание ответов клиента путем формулирования вопросов в «нужной» форме;

- настройка параметров может повысить, и действительно повышает, качество распознавания.

Свойства человека в контексте речевого интерфейса

Главной проблемой унимодального речевого интерфейса является проблема невидимости: зрение не участвует в процессе взаимодействия с машинной стороной, и восприятие вопросов и ответной информации производится клиентом только лишь на слух. Исключение зрения из процесса взаимодействия с машиной приводит к значительному возрастанию *когнитивной* нагрузки на человека.

Наибольшее внимание в процессе дизайна речевого интерфейса требуется уделять следующим составляющим когнитивной нагрузки: *нагрузке на память, удержанию внимания и понятийной сложности*. В [6, 7] приведены методы снижения когнитивной нагрузки при организации речевого взаимодействия клиента с автоматизированной системой обслуживания.

Снижение нагрузки на память. Для снижения нагрузки на память предложены следующие методы конструирования меню и формулирования промптов (промпт – произносимый системой текст: вопрос, инструкция, пояснение, помощь, подсказка и т.п.):

- меню не должно содержать более 5 пунктов. Оптимальный размер меню – 3-4 пункта;

- наиболее популярные пункты меню должны располагаться в начале списка; это позволит сократить диалог при использовании функции *barge-in* (возможность перебивать систему, не дожидаясь окончания проигрывания ее сообщения);

- концепт, который пользователь должен запомнить лучше всего, необходимо помещать в конец промпта;

- промпт должен сначала объяснять функцию и только потом называть соответствующее ей действие;

- первичные промпты-инструкции должны быть короткими и должны содержать только самую необходимую информацию; дополнительная информация должна быть размещена в промптах помощи, которые могут быть несколько продолжительнее.

Удержание внимания. Проблема удержания внимания клиента становится наиболее актуальной на этапе предоставления ему выходной аудио информации. Крайний случай – вывод всех данных сразу. Ограниченность кратковременной памяти и потеря внимания имеют следствием высокую вероятность необходимости повтора всего вывода. Кроме того, с большой вероятностью может оказаться, что далеко не вся эта информация является одинаково важной для клиента. Деление выводимой информации на фрагменты смягчает эти проблемы, но требует, в свою очередь, дополнительных затрат времени на диалог по поводу вывода этих фрагментов. Оптимальное число фрагментов зависит от контекста конкретного приложения и «делимости» выводимой информации. Следует помнить также, что фрагмент должен включать ограниченное число элементов данных для запоминания (в среднем 3-4 элемента).

Снижение понятийной сложности

Предложены следующие правила для снижения понятийной сложности при взаимодействии пользователя с системой:

- *Унификация действий и терминологии с целью «формирования полезной привычки».*

- *Использование универсальных команд навигации*, всегда доступных пользователю вне зависимости от приложения и контекста.

- *Пояснения и примеры.* Примеры и пояснения позволяют снизить число ошибок при вводе данных, помогая клиентам правильно сформулировать ответы на вопросы системы.

- *Помощь, немедленная обратная связь (эхо-ответы), ориентация.* Предоставление пользователю этих возможностей значительно повышает его уверенность в успехе, способствует сохранению спокойного стиля общения с системой, снижает задержки ответов, что способствует успешному распознаванию системой его ответов и снижению продолжительности диалога.

На основании проведенных исследований в [5,7] приведены рекомендации по проектированию речевого человеко-машинного интерфейса с системами обслуживания.

Для иллюстрации влияния на качество работы системы с распознаванием речи тщательного дизайна интерфейса, правильного согласования промптов и грамматик и оптимальной настройки параметров распознавателя на рис.1 приведены обобщённые результаты работы ненастроенной системы «Автосекретарь», а на рис. 2 показаны результаты работы той же системы после выполнения необходимых корректировок и настройки параметров.

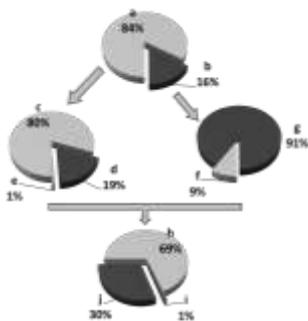


Рис. 1. Обобщенные результаты работы ненастроенной системы «Автосекретарь».

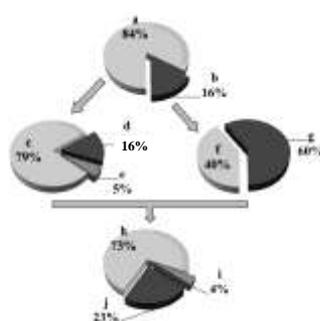


Рис. 2. Обобщенные результаты работы настроенной системы «Автосекретарь».

На схемах использованы следующие обозначения:

- a – правильные высказывания, характеристика исходной выборки;
- b – неправильные высказывания (шумы, посторонние звуки и пр.), характеристика исходной выборки;
- c – распознано правильных высказываний;
- d – неверно распознано правильных высказываний;
- e – не распознано правильных высказываний;
- f – верно отброшено неправильных высказываний;
- g – неверно распознано неправильных высказываний;
- h – верные результаты работы;
- i – нераспознанные высказывания;
- j – неверно распознанные высказывания.

Как видно из приведенных рисунков, грамотно спроектированный интерфейс и оптимальная настройка параметров позволяют улучшить качество работы системы с распознаванием речи.

4. Обнаружение и коррекция ошибок распознавания речи.

Стохастическая природа процессов, лежащих в основе работы распознавателей, обуславливает возможность появления на их выходе ошибочных результатов. Ошибки при распознавании снижают как полезность приложения, так и степень удовлетворенности пользователей. В связи с этим обнаружение и исправление ошибок распознавания является определяющим свойством речевых приложений. В [8, 9] рассмотрены методы выявления ошибок на основе анализа показателя уверенности распознавателя в предлагаемой гипотезе и использования накопленных статистических данных о результатах распознавания.

Метод выявления ошибок на основе анализа показателя уверенности распознавателя в предлагаемой гипотезе состоит в сравнении величины этого показателя с *пороговым значением*, установленным для вызова процедуры подтверждения. Если показатель уверенности распознавателя в предлагаемой им гипотезе выше порогового значения, то считается, что вероятность ошибки невелика и гипотеза принимается, а диалог переходит на следующий этап. Если вероятность гипотезы ниже порога, то диалог развивается по пути активации процедуры подтверждения. Задача состоит в *выборе оптимальной величины порога показателя уверенности в гипотезе*.

Определение оптимальной величины порога показателя уверенности в гипотезе состоит в применении *принципа минимума ожидаемых затрат*, который применительно к данной задаче

формулируется следующим образом: порог уверенности должен быть таким, чтобы суммируемые по всем состояниям ожидаемые затраты для выбранного в соответствии с ним действия были минимальными.

Коррекция ошибок может быть значительно более эффективной, если использовать знания, накопленные в результате сбора и обработки статистических данных об ошибках. В базовых системах распознавания речи предусмотрены средства для сбора статистических данных. Процесс сбора статистики заключается в последовательном прослушивании диалогов, записанных в лог-файлах, и их описании. Описание диалогов состоит в оценке правильности распознавания и внесении дополнительных комментариев. Существующая процедура является трудоемкой и затратной по времени.

В [8] описан *метод автоматизированного накопления статистики о правильности распознавания*. Метод основан на анализе ответов на запросы подтверждения. Из ответа клиента на просьбу системы подтвердить гипотезу s_j , который может быть сформулирован по-разному («Да», «Да, правильно», «Правильно», «Да-да» и т.д.), извлекается смысл («да» или «нет») и фиксируется вместе со значением возвращаемой гипотезы s_j и показателем уверенности в ней. Ответ «да» соответствует успешному распознаванию данного слова или фразы. Ответ «нет» соответствует ошибке, которая относится к разновидности «замена». После ответа «нет» диалог переходит к фазе повторения ввода. Ответ «да» на запрос подтверждения повторного ввода (новой гипотезы s_j) позволяет зафиксировать не только факт имевшей место ошибки распознавания, но и само искажение: распознавание слова s_j как s_j . Ответ «нет» на запрос подтверждения повторного ввода инициирует новую процедуру подтверждения, если это предусмотрено алгоритмом. Он также фиксируется и может свидетельствовать о том, что произносимое клиентом слово или фраза отсутствуют в грамматике или имеют ошибку в фонетической транскрипции, особенно если возвращаемая уверенность в гипотезе невелика.

Собранная таким образом статистика может быть использована для проверки фонетических транскрипций, для выявления часто встречающихся «лишних», т.е. отсутствующих в грамматике, слов и для составления таблиц искажений для коррекции ошибок, связанных с искажением слов.

Пусть система пытается распознать произнесенное слово из n слов s_1, s_2, \dots, s_n , предлагая свой вариант s_0 . Пусть также:

l — математическое ожидание числа переспросов;

- t_j — вероятность того, что система восприняла слово как S_j ;
 p_i — вероятность того, что было произнесено слово S_i ;
 q_{ij} — вероятность того, что слово S_i распознано как S_j ;
 r_{ij} — вероятность того, что произносилось слово S_i при условии, что система приняла его за слово S_j .

В частности, q_{ii} — вероятность того, что слово распознано правильно, а q_{i0} — вероятность того, что слово S_i не было распознано вообще. Вероятности p_i и q_{ij} определяются статистически. Здесь и далее $i, j = 1, 2, \dots, n$, $\sum_{i=1}^n p_i = 1$, $\sum_{j=0}^n q_{ij} = 1$ для всех i .

Нетрудно видеть, что $r_{ij} = \frac{p_i q_{ij}}{t_j}$, где вероятность $t_j = \sum_{k=1}^n p_k q_{kj}$.

Алгоритм, учитывающий статистику искажений, состоит в следующем. Пусть система выдает гипотезу S_j и просит клиента сообщить, правильна ли она. В случае ответа «да» процедура заканчивается. В случае ответа «нет» система выбирает наибольшую вероятность r_{ij} из ряда $r_{1j}, r_{2j}, \dots, r_{nj}$, в котором нет величины r_{jj} , и предлагает слово S_i в качестве новой гипотезы с просьбой подтвердить ее. Если и это не подтверждается, система ищет в ряду $r_{1j}, r_{2j}, \dots, r_{nj}$ вторую по величине вероятность и предполагает, что было произнесено второе слово, и т.д.

Математическое ожидание числа переспросов l_j равно $l_j = r_{1j} + 2r_{2j} + 3r_{3j} + \dots + nr_{nj}$, где r_{ij} — вероятности слов, предлагаемых последовательно системой в соответствии с данным алгоритмом при условии, что на первом шаге система распознала слово как S_j . Среднее число переспросов l вычисляется по формуле:

$$l = \sum_{j=1}^n t_j l_j = \sum_{j=1}^n \left(\sum_{i=1}^n p_i q_{ij} \sum_{i=1}^n i r_{m_j j} \right).$$

Сравнение числа переспросов для алгоритма коррекции ошибок с учетом статистических данных об искажениях слов с числом пере-

спросов для алгоритма без учета вероятностей искажений произведено на примере распознавания произносимых пользователем цифр от 0 до 9. Вероятности q_{ij} были определены экспериментально. Вероятности p_i были приняты равными между собой, т.е. $p_0 = p_1 = \dots = p_9 = 0,1$.

Среднее число переспросов по первому алгоритму составило 1,08, по второму – 1,18, что свидетельствует о преимуществах алгоритма с коррекцией ошибок с учетом статистики искажений. Таким образом, предложенная математическая модель позволяет оценить, насколько можно сократить среднее число переспросов, если использовать накапливаемую статистику о распознавании слов.

Предложенный метод обнаружения и исправления ошибок путем автоматизированного накопления статистических данных об ошибках и использовании знаний об искажениях слов является новым шагом в решении проблемы ошибок распознавания. Разработчики приложений, не имея возможности вмешаться во внутренние модели распознавателей и вынужденные рассматривать их как черный ящик, могут добиться достаточно высоких показателей качества работы приложения путем применения предложенного метода по накоплению и использованию знаний о работе распознавателя.

5. Анализ сценариев диалогов с различными процедурами обнаружения и коррекции ошибок.

Эти задачи возникают на этапе проектирования системы, когда принимаются решения относительно структуры и расположения элементов диалога и процедур подтверждения. На рис. 3 приведена классификационная схема речевого человеко-машинного взаимодействия, в основу которой положены следующие признаки: типы элементов диалога, местоположение процедур выявления и корректировки ошибок, способ реализации процедур выявления и корректировки ошибок.

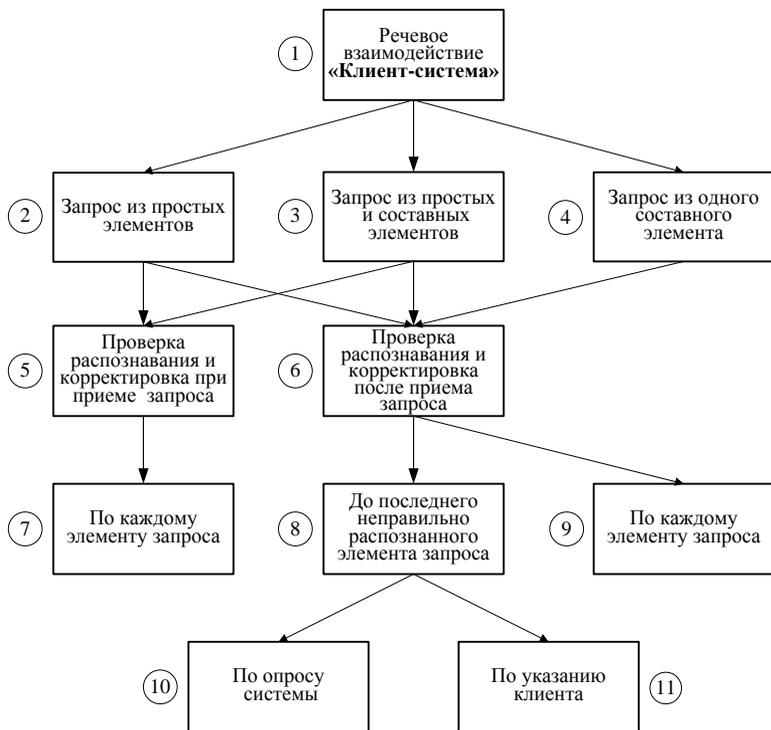


Рис. 3. Классификационная схема речевого человеко-машинного взаимодействия.

Произведенная классификация позволяет охватить самые разные варианты организации взаимодействия клиентов с системой и производить сравнение этих вариантов.

Методика выполнения сравнительного анализа вариантов приведена в [9-11] и состоит в использовании модифицированного варианта применительно к данной задаче принципа *квантификации*, когда варианты сравниваются на множестве одинаковых типовых количественных значений ряда параметров. Например, сравнительная оценка вероятности успешного завершения диалога осуществляется при одинаковых значениях вероятностей правильного распознавания речевых блоков и одинаковом допустимом числе переспросов для сравниваемых вариантов сценариев. Необходимость в таких допущениях диктуется большой размерностью моделей и широким диапазоном изменения

параметров. Квантификация является достаточно популярным методом количественного анализа качества интерфейсов.

В качестве характеристик для сравнения вариантов выбраны два главных количественных критерия оценки качества речевого диалогового человеко-машинного взаимодействия: *вероятность успешного завершения диалога* и *продолжительность диалога*. Поиск оптимального сценария осуществляется на основе минимизации оценок продолжительности диалога с обеспечением заданной вероятности его успешного завершения.

Предложенные в [8-11] модели анализа и формулы вычисления вероятностных и временных характеристик диалогов позволяют:

- 1) определять необходимое число переспросов клиента для получения требуемой достоверности распознавания элементов диалога;
- 2) определять среднее время, затрачиваемое на элемент диалога с учетом рассчитанного для заданной достоверности распознавания числа переспросов;
- 3) вычислять вероятности успешного завершения диалогов различной структуры;
- 4) анализировать и сравнивать между собой различные стратегии управления диалогом по предложенным критериям продолжительности диалога и при обеспечении требуемой вероятности успешного завершения диалогов;
- 5) производить обоснованный выбор сценариев и алгоритмов управления диалогом для конкретных задач на основе разработанной методики.

6. Математические модели для расчета характеристик обслуживания заявок и оптимизации функционирования контакт центров с сервисами самообслуживания.

Для описания функционирования центра обслуживания вызовов с сервисами самообслуживания использована открытая экспоненциальная сеть массового обслуживания с двумя узлами, первый из которых описывает порты самообслуживания, а второй – операторскую группу. Узел i сети представляет собой многолинейную СМО с идентичными приборами, $i = 1, 2$. Поток заявок из внешнего источника является пуассоновским интенсивности λ и поступает в узел 1 или 2 в зависимости от рассматриваемой дисциплины обслуживания. Время обслуживания заявки в узле i распределено экспоненциально с параметром μ_i .

С вероятностью $1 - p$ обслуживание заявки в первом узле является успешным, и в момент завершения обслуживания эта заявка покидает систему. В противном случае с вероятностью p обслуживание является неудачным (в ходе обработки запроса произошла ошибка, и запрос должен быть обработан оператором), и заявка переходит во второй узел (рис. 4).

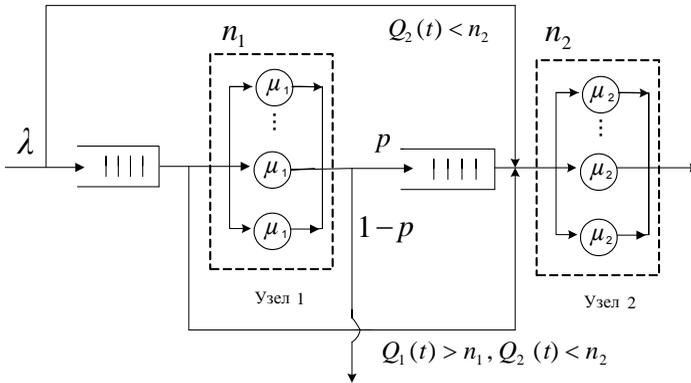


Рис. 4. Схема сети массового обслуживания с двумя узлами.

Задачей является нахождение характеристик обслуживания заявок в системе при различных моделях управления очередью на входе и различном числе мест для ожидания в узлах.

Для решения задачи вводится случайный процесс, описывающий состояние сети в момент времени t :

$$\{X(t)\}_{t \geq 0} = \{Q_1(t), Q_2(t)\}_{t \geq 0},$$

где $Q_i(t)$ – число заявок в i -м узле, $i = 1, 2$ и рассматриваются стационарные вероятности

$$\pi(i, j) = \lim_{t \rightarrow \infty} P[X(t) = (i, j)], i = \overline{0, N}, j \geq 0.$$

Составляются уравнения баланса для стационарных вероятностей системы, вводятся в рассмотрение векторы стационарных вероятностей

$$\pi_j = (\pi(0, j), \pi(1, j), \dots, \pi(N, j)), \quad j \geq 0,$$

где $\pi(i, j) = 0$ для $n_1 < i \leq N$, $0 \leq j < n_2$,

и уравнения баланса преобразуются в векторно-матричную форму. Алгоритм для вычисления стационарных вероятностей основан на матрично-аналитическом подходе к анализу многомерных цепей Маркова, имеющих не более одной счетной компоненты. После вычисления стационарных вероятностей вычисляются характеристики обслуживания.

В [12-14] приведены модели и получены формулы для расчета характеристик обслуживания для различных вариантов организации очередей ожидания в узлах: отсутствие мест для ожидания в первом узле, ограниченные и неограниченные очереди в узлах.

Задача оптимизации числа приборов в узлах

Введена следующая структура штрафов за функционирование системы:

- $c_{0,k}$ – стоимость в единицу времени ожидания в очереди узла k ,
- $c_{u,k}$ – стоимость в единицу времени использования прибора в узле k ,
- $c_{e,k}$ – стоимость в единицу времени простоя прибора в узле k ,
- c_f – фиксированная стоимость включения системы мониторинга состояний.

Задача состоит в минимизации функционала потерь:

$$\bar{V}(n_1, n_2) = \bar{V}(\lambda, \mu_1, \mu_2, p, n_1, n_2) \rightarrow \min_{n_1, n_2},$$

который в данном случае имеет вид

$$\bar{V}(n_1, n_2) = c_{0,1} \bar{Q}_1 + c_{0,2} \bar{Q}_2 + c_{u,1} \bar{C}_1 + c_{u,2} \bar{C}_2 + c_{e,1} \bar{Z}_1 + c_{e,2} \bar{Z}_2 + c_f \frac{1}{B},$$

где $\bar{Q}_1 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} j \pi(i, n_1, j)$ и $\bar{Q}_2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} j \pi(i, j, n_2)$ обозначают среднее число заявок в очереди k -го узла, $\bar{C}_1 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \min\{i, n_1\} \pi(i, j)$ и $\bar{C}_2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \min\{i, n_2\} \pi(i, j)$ обозначают среднее число занятых приборов в k -ом узле,

$\bar{Z}_1 = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} (n_1 - i) \pi(i, j)$ и $\bar{Z}_2 = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} (n_2 - j) \pi(i, j)$ обозначают среднее число свободных приборов в k -ом узле,

$$\bar{B} = \frac{1}{\lambda \pi(0,0)} - \text{средняя длительность цикла между соседними}$$

посещениями состояния $(0, 0)$.

В качестве примера рассмотрена система со следующим набором параметров: $(\lambda, \mu_1, \mu_2, p, c_{0,1}, c_{0,2}, c_{u,1}, c_{u,2}, c_{e,1}, c_{e,2}, c_f) = (0.9, 0.5, 0.7, 0.001, 5.5, 2.5, 0.001, 0.005, 0.001, 0.005, 0.9)$. Для такой системы оптимальное

число приборов равно $(n_1^*, n_2^*) = (3, 4)$ и оптимальное значение функционала потерь $\bar{V}(n_1^*, n_2^*) = 0.247$.

Для случая $(\lambda, \mu_1, \mu_2, p, c_{0,1}, c_{0,2}, c_{u,1}, c_{u,2}, c_{e,1}, c_{e,2}, c_f) = (2.5, 0.5, 0.7, 0.001, 5.5, 2.5, 0.01, 0.05, 0.01, 0.05, 2.9)$ оптимальные значения определяются как: $(n_1^*, n_2^*) = (6, 4)$, $\bar{V}(n_1^*, n_2^*) = 1.011$. Зависимость функции $\bar{V}(n_1, n_2)$ от числа приборов показана на рис. 5.

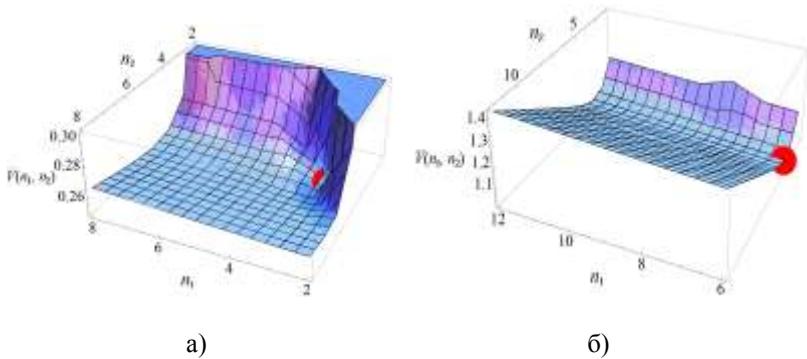


Рис. 5. Значение функции $\bar{V}(n_1, n_2)$ в зависимости от числа приборов:
 а) $(n_1^*, n_2^*) = (3, 4)$; б) $(n_1^*, n_2^*) = (6, 4)$.

Оптимальное управление очередью для центра обслуживания вызовов с быстрыми и медленными приборами.

Для центров обслуживания вызовов с сервисами самообслуживания характерным является разница в скоростях и стоимостях работы приборов первого и второго узлов. Узел 1, моделирующий порты самообслуживания на базе ненадежного распознавателя речи, имеет более низкую скорость обслуживания, но при этом малые затраты на обслуживание. Узел 2, моделирующий работу операторов, является надежным, более быстрым и одновременно более дорогим по сравнению с узлом 1. Задача оптимального управления общей очередью в такой системе принадлежит к области задач управляемых марковских систем массового обслуживания.

Применительно к центрам с функциями самообслуживания задачей является поиск оптимального управления общей очередью с целью минимизации среднего времени пребывания заявок в системе [13], где центр обслуживания вызовов с функциями самообслуживания рас-

сматривается как открытая экспоненциальная сеть массового обслуживания с двумя узлами, каждый из которых представляет собой многолинейную систему массового обслуживания типа $M/M/n_i$ с идентичными приборами. Поток заявок в сеть является простейшим с параметром λ . Заявка, прибывающая в сеть, поступает в узел 1 или 2 в зависимости от числа заявок в узле 2. Если последнее равно или превышает пороговый уровень q_2 ($q_2 > n_2$), то заявка направляется в узел 1, в противном случае – в узел 2.

Как только число заявок в узле 2 уменьшается до $q_2 - 1$, а в узле 1 есть ожидающие заявки, заявка, стоящая в начале очереди узла 1, переходит в конец очереди узла 2. Выбор данного вида управления обусловлен тем фактом, что время обслуживания в узле 1 может превышать время ожидания в узле 2, так как приборы узла 2 имеют более высокую скорость обслуживания.

Для решения задачи вводится случайный процесс $\{X(t)\}_{t \geq 0} = \{Q_1(t), Q_2(t)\}_{t \geq 0}$, обозначающий состояние сети в момент времени t , где $Q_i(t)$ – число заявок в i -м узле сети в момент времени t , $i = 1, 2$. Через $E = \{x = (i, j) : i, j \geq 0\}$ обозначается множество состояний процесса $\{X(t)\}_{t \geq 0}$, где состояние (i, j) случайного процесса означает, что в узле 1 находится i заявок, а в узле 2 – j заявок. Случайный процесс $\{X(t)\}_{t \geq 0}$, описывающий поведение этой системы, является однородным и марковским. В предположении, что существуют стационарные вероятности

$$\pi(i, j) = \lim_{t \rightarrow \infty} P[X(t) = (i, j)],$$

выписывается система уравнений равновесия.

Результатом решения задачи является доказательство существования мультипликативной формы стационарных вероятностей с представлением всех вероятностей системы в виде функции, зависящей от вероятности $\pi(0, q_2 + 1)$, которая, в свою очередь, вычисляется из условия нормировки.

Вычислив стационарное распределение вероятностей состояний сети, можно получить *характеристики обслуживания вызовов*, формулы для которых приведены в [12, 13].

Оптимизация порогового уровня и численные примеры

Для оптимизации порогового уровня q_2 введена следующая структура штрафов:

$c_{0,k}$ – стоимость ожидания заявки в очереди узла k в единицу времени,

$c_{u,k}$ – стоимость работы одного прибора в узле k в единицу времени.

Задача состоит в минимизации функционала потерь

$$\bar{V}(q_2) := \bar{V}(\lambda, \mu_1, \mu_2, p, n_1, n_2, q_2) \rightarrow \min_{q_2},$$

который в данном случае имеет вид

$$\bar{V}(q_2) = \sum_{k=1}^2 c_{u,k} \bar{C}_k + \sum_{k=1}^2 c_{0,k} \bar{Q}_k.$$

Замечание. Из последней формулы следует, что:

1) если $c_{0,k} = c_{u,k} = 1$, $k = \overline{1,2}$, то задача сводится к минимизации среднего числа заявок в системе \bar{N} или времени пребывания \bar{T} .

2) если $c_{0,k} = 1$, $c_{u,k} = 0$, $k = \overline{1,2}$, то задача сводится к минимизации среднего числа заявок в очереди \bar{Q} или времени ожидания \bar{W} .

На рис. 5 показаны оптимальные пороговые уровни для различных значений стоимостей и интенсивностей обслуживания при некоторых заданных значениях параметров сети. На рис. 5 (а) для $\lambda = 0.9$ показано влияние вероятности p ошибки при обслуживании в узле 1, а на рис. 5 (б) – влияние интенсивности λ поступления новых заявок для $p = 0.01$ на функционал потерь $\bar{V}(q_2)$. Очевидно, что увеличение значений параметров p и λ приводит к увеличению оптимального порогового уровня q_2 , так как в этом случае возрастает нагрузка на систему и возникает необходимость более интенсивного использования быстрого, но при этом дорогого узла 2.

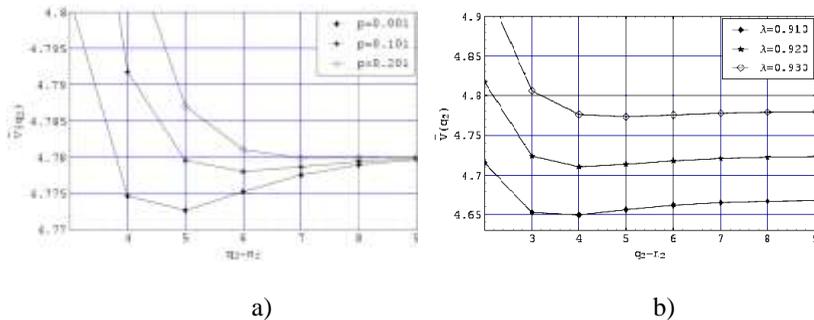


Рис. 5. Значение функции $\bar{V}(q_2)$ для различных значений q_2 , p и λ .

7. Заключение. Применение речевых технологий позволяет существенно расширить круг пользователей систем массового обслуживания, предоставив им удобный дополнительный, а иногда и единственный канал доступа к услугам этих систем. Это обеспечит новый уровень информационных услуг, предоставляемых населению и организациям в режиме самообслуживания, а также повысит эффективность использования систем массового обслуживания. Практическое применение речевых технологий сталкивается с рядом еще нерешенных проблем. Исследование этих проблем, поиск решений и выработка методических рекомендаций для разработчиков систем с речевым интерфейсом является актуальной задачей и имеет важное теоретическое и практическое значение.

Литература

1. *Ронжин А.Л., Карпов А.А., Ли И.В.* Речевой и многомодальный интерфейсы // М.: Наука, 2007. 173 с.
2. *Ронжин А.Л., Ли И.В.* Автоматическое распознавание русской речи // Вестник Российской Академии Наук: научный и общественно-политический журнал, Том 77, Вып.2, 2007. С. 133-138.
3. *Чучупал В.Я., Маковкин К.А., Чичагов А.В.* К вопросу об оптимальном выборе алфавита моделей звуков русской речи для распознавания речи // Искусственный интеллект. 2002. № 2.
4. Модели, методы, алгоритмы и архитектуры систем распознавания речи // Вычислительный Центр им. А.А. Дородницына РАН. М.: 2006. <http://www.ccas.ru/depart/chuchu/publics/sbornikCCAS-2006.pdf>
5. *Абраменков А.Н., Мясоедова З.П., Петухова Н.В., Фархадов М.П.* Исследование влияния параметров настройки распознавателя на качество распознавания речи // Труды 2-ой Российской конференции с международным участием «Технические и программные средства систем управления, контроля и измерения» (УКИ-10). ИПУ РАН. 2010. С. 01-07.
6. *Билик Р.В., Мясоедова З.П., Петухова Н.В., Фархадов М.П.* Речевой интерфейс как разновидность человеко-машинного взаимодействия // Материалы международной научно-практической конференции «Информационные технологии и информационная безопасность в науке, технике и образовании (ИНФОТЕХ-2009)». СевНТУ, 2009. С. 279-282.
7. *Петухова Н.В., Мясоедова З.П., Фархадов М.П.* Исследование особенностей проектирования речевого интерфейса // Труды XIV международной ФАМЭТ'2010 конференции. Под ред. О.Ю. Воробьева. — Красноярск: Крас. гос. торг.-эконом. ин-т, Сиб. фед. ун-т, 2010. — 357 с. С. 251-255.
8. *Фархадов М.П., Жожикашвили А. В.* Математическая модель для расчета среднего числа переспросов при компьютерном распознавании речи // *Проблемы управления.* 2006. №2. С. 38-41.
9. *Билик Р.В., Жожикашвили В.А., Петухова Н.В., Фархадов М.П.* Анализ речевого интерфейса в интерактивных сервисных системах I. / *Автоматика и телемеханика.* 2009. №2. С. 80-89.

10. *Билик Р.В., Жожикашвили В.А., Петухова Н.В., Фархадов М.П.* Анализ речевого интерфейса в интерактивных сервисных системах II. / *Автоматика и телемеханика*. 2009. №3. С. 97-113.
11. *Билик Р.В., Мясооедова З.П., Петухова Н.В., Фархадов М.П.* Под ред. проф. Жожикашвили В.А. Анализ речевого интерфейса при взаимодействии клиента с автоматизированной системой массового обслуживания. М.: МАКС Пресс. 2007. 112 с.
12. *Фархадов М.П., Петухова Н.В., Ефросинин Д.В., Семенова О.В.* Моделирование гибридного центра связи с сервисами самообслуживания и пороговым управлением размещением заявок / *Управление большими системами*. 2010. №30-1. С. 352-370.
13. *Фархадов М.П., Петухова Н.В., Ефросинин Д.В., Семенова О.В.* Двухфазная модель с ограниченными очередями для расчета характеристик и оптимизации речевых порталов самообслуживания / *Проблемы управления*. 2010. № 6. С. 53-57.
14. *Фархадов М.П., Петухова Н.В., Ефросинин Д.В., Семенова О.В.* Математическая модель центра обслуживания вызовов с сервисами самообслуживания // *International Workshop "Distributed Computer and Communication Networks DCCN'2009"*. Sofia. Bulgaria. М.: Издательство ФГУП «Производственно-издательский комбинат ВИНТИ». 2009. С. 86-95.

Фархадов Маис Паша оглы – канд. техн. наук., с.н.с.; заведующий лабораторией автоматизированных систем массового обслуживания Учреждения Российской академии наук Института проблем управления им. В.А. Трапезникова РАН (ИПУ РАН). Область научных интересов: автоматизированные системы массового обслуживания, распознавание речи, голосовое управление и речевой интерфейс, интеллектуальные системы, теория массового обслуживания, жестовый язык, речевые и мультимедийные технологии, теория вероятностей, системный анализ, интернет-технологии, компьютерная лингвистика. Число научных публикаций – 70. mais@ipu.ru, www.ipu.ru; ИПУ РАН, Профсоюзная улица, д. 65, г. Москва, 117997, РФ; р.т. +7(495)334-8710, факс +7(495)334-9340.

Farkhadov Mais Pasha - Ph.D. (candidate of tech. Sciences.), senior researcher, Head of Laboratory of “Automated queuing systems”, Institute of Control Sciences of the Russian Academy of Sciences (ICS RAS). Research interests: automated queuing systems, speech recognition, voice control and voice user interface, intelligent systems, queuing theory, sign language, speech and multimedia technologies, probability theory, systems analysis, internet technology, computational linguistics. The number of publications -70. mais@ipu.ru, www.ipu.ru; Institute of Control Sciences, Profsoyuznaya Ulitsa 65, Moscow 117997, Russia; office phone +7 (495) 334-8710, fax +7 (495) 334-9340.

Рекомендовано лабораторией речевых и многомодальных интерфейсов, заведующий лабораторией Ронжин А.Л, д-р техн. наук, доц.
Статья поступила в редакцию 30.06.2011.

РЕФЕРАТ

Фархадов М.П. Распознавание речи в системах массового обслуживания населения.

Статья посвящена решению различных вопросов, возникающих при применении речевых технологий в системах массового обслуживания. Использование речевых технологий в информационных и сервисных системах позволяет решить актуальную проблему облегчения доступа людей к информации и услугам, повышая эффективность и открытость систем массового обслуживания населения.

При внедрении речевых технологий приходится сталкиваться с целым рядом все еще не решенных проблем, суть которых определяется спецификой самих речевых технологий и особенностями человека. Наиболее общий характер имеют проблемы, относящиеся к построению эффективного речевого человеко-машинного интерфейса, проблемы, связанные с повышением устойчивости систем к ошибкам распознавания и задачи оптимизации сценария диалога при различных вариантах организации процедур выявления и коррекции ошибок. Другую группу новых задач образуют задачи разработки математических моделей для анализа и исследования систем обслуживания заявок, где речевые технологии применяются для организации функций самообслуживания, и задачи оптимизации функционирования таких центров.

В статье рассматриваются проблемы и предлагаются решения следующих задач, связанных с применением речевых технологий:

- экспериментальное исследование показателей качества распознавателей русской речи, ориентированных на применение в системах обслуживания населения,
- разработка речевого интерфейса для взаимодействия населения с автоматизированными системами обслуживания,
- классификация, критерии и методы оценок вариантов диалогов с различными процедурами выявления и коррекции ошибок и методика обоснованного выбора оптимальных сценариев диалогов,
- автоматизированное накопление данных об ошибках распознавания и стратегия коррекции ошибок на основе знаний о результатах распознавания,
- математические модели для расчета вероятностно-временных характеристик обслуживания заявок и оптимизации функционирования сетевых систем массового обслуживания с сервисами самообслуживания на основе распознавания речи.

SUMMARY

Farkhadov M.P. Speech recognition in automated information and services systems for the population.

Paper is devoted to the various issues arising from the use of speech technologies in automated queuing systems. The use of speech technologies in information and service systems allows to solve an actual problem to facilitate people's access to information and services and to improve efficiency and openness of service systems.

In implementing of speech technologies developers have dealing with a number of still unresolved problems, which are determined by the specific features of speech technologies and human factors.

The most common problems are the construction of an effective voice man-machine interface, resistance to errors and the optimization scenario dialogue with different of procedures of error detection and correction. Another group of new tasks is developing of mathematical models for analysis of service systems where speech technologies are used to organize self-service functions, and the task of optimizing of contact centers.

The decision of following problems became the purpose of this article:

- experimental study of quality of Russian speech recognizers oriented to application in public services systems,
- development of a speech interface for people interaction with automated services systems,
- the classification, criteria and methods of evaluation of dialogues with the various procedures of errors identify and correction and methods of optimal scenario dialogues choice,
- the automated accumulation of the data about recognition errors and strategy of correction of errors on the basis of knowledge of results of recognition,
- mathematical models for calculation of the probability-time characteristics of service applications and optimization of network queuing systems with self-services based on speech recognition.