

УДК 004.522

Д.А. Суранова, С.Д. Тиунов, Р.В. Мещеряков

## Численная реализация эффективного доступа в биллинговой системе с использованием голосового интерфейса

Предложены варианты использования голосового интерфейса к биллинговой системе оплаты за услуги ЖКХ с использованием эффективных численных алгоритмов обработки речевого сигнала. Показана возможность оптимизации доступа к данным как на уровне базы данных, так с использованием распараллеливания работы алгоритмов обработки речевого сигнала. Предложены схемы численной реализации в виде алгоритмического и программного обеспечения.

**Ключевые слова:** математическое моделирование, алгоритм, речевой сигнал, база данных, биллинг.

**Постановка задачи.** В настоящее время решения задач находят на стыке двух разных областей. Имея преимущества какой-либо одной области и компенсируя ее недостатки преимуществами другой, можно получить хороший результат. В рамках данной статьи будет представлено решение для графического интерфейса биллинговой системы, использующее голосовой режим там, где такое применение даст эффект. При этом голосовые возможности будут еще более востребованы, если они смогут запоминать и применять контекст, полученный в процессе взаимодействия. Голосовой интерфейс для каждого из режимов будет работать только тогда, когда список доступных речевых команд будет сформирован на основе сбора и анализа фраз, употребляемых для обозначения определенного действия реальными пользователями. Также необходимы анализ режимов работы и реализация голосового взаимодействия для наиболее часто используемых режимов. Разработка будет полезна не только для новичков, но и для более опытных пользователей системы. Такая оптимизация позволит сократить количество настроек (сами настройки не нужно будет показывать на экране, достаточно просто их вызвать голосом), повысить скорость работы (не нужно будет искать пункты меню) и уменьшит вероятность ошибки распознавания (команда будет максимально короткой, но с учетом привычек пользователей). Также к плюсам можно отнести тот факт, что пользоваться системой можно будет не только с ПК, но и с планшетных устройств (ввиду сокращения отображаемого графического функционала).

Подобную технологию можно встретить в решениях для «Умного дома». Произнося команду «Включить свет» или «Поставить хлеб», человек сможет далее сам контролировать процесс. Похожее решение было использовано командой разработчиков из Мичиганского университета – они создали прототип графического редактора, управляемого голосом и жестами. Прототип, названный PixelTone[1], пока может распознавать лишь самые простые операции и жесты. Он понимает простые команды и ориентируется в описательных терминах вроде «верх», «низ», «ярче», «темнее». Кроме того, он может запоминать названия объектов и имена людей на фотографиях, что позволяет использовать команды вроде «Сделай Васю чуть контрастнее».

В России голосовые решения используются в продуктах центра речевых технологий [2], например, таких, как «Потребитель» – система сбора показаний по приборам учета. Позвонив по телефону, человек сообщает номер лицевого счета и величину показания, и после подтверждения корректности информация поступает в систему. Кроме того, необходимо учитывать требования к многомодальному интерфейсу [3, 4, 11].

**Доступ к сведениям биллинговой системы.** В рамках данной статьи применение речевых возможностей будет рассмотрено на примере биллинговой системы учета платы за жилищно-коммунальные ресурсы. Пользователями данной системы являются бухгалтера, которые вводят и изменяют информацию по жильцам дома и тем услугам, которыми жильцы пользуются. Информация вводится периодически (например, данные по счетчикам подаются каждый месяц) либо по мере поступления (обращения жильцов, принятие решений на общедомовых собраниях, изменение законодательства).

В процессе работы была собрана статистика от 200 пользователей биллинговой системы. Чаще всего они работают в следующих режимах:

- Поиск лицевого счета для последующей работы с ним.
- Изменение свойств лицевого счета (изменение площади, адреса, признака приватизации, фамилии ответственного квартиросъемщика и др.).
- Добавление проживающих или периодов проживания в лицевые счета.
- Занесение показаний по приборам учета.
- Формирование справки/выписки по лицевому счету.
- Формирование отчета.
- Формирование запроса по определенным параметрам.

На основе анализа общей статистики и подсчета действий, совершенных пользователями за апрель 2013 года, был сделан вывод, что чаще всего пользователи работают с лицевыми счетами. Процентное соотношение действий, связанных с лицевыми счетами к общему количеству действий системы приведено в табл. 1.

Таблица 1

**Статистика действий пользователей по лицевым счетам, %**

Переход на страницу лицевого счета	Поиск лицевого счета по адресу	Поиск лицевого счета по номеру	Прочие действия
27,5	15,5	10	47

Рассмотрим более подробно этапы, выполняемые оператором для совершения действий по редактированию лицевого счета. Операторы бывают 2 типов: оператор ВЦ и бухгалтер по квартплате. Оператор ВЦ имеет доступ ко всему жилому фонду, бухгалтер по квартплате – только к тем домам, которые он обслуживает. Любой оператор входит в систему путем введения логина и пароля. После проверки и успешной идентификации появляется меню в виде древовидной структуры, через которое можно совершить нужное действие, например попасть в лицевой счет. Для этого необходимо зайти в дом, выбрать квартиру и лицевой счет в ней или открыть непосредственно лицевой счет из списка, если известен его номер. Также можно воспользоваться формой поиска лицевого счета, доступной сразу же при входе; форма идентична для бухгалтера по квартплате и оператора ВЦ. Поиск происходит после ввода в необходимые поля номера лицевого счета либо его адреса.

Для операторов ВЦ и бухгалтеров по квартплате был проведен эксперимент для определения среднего времени, затрачиваемого на совершение действий по поиску лицевого счета. Эксперимент был проведен для нескольких опытных пользователей, каждый выполнял набор действий из следующего списка.

1. Поиск лицевого счета по номеру из дерева объектов для оператора ВЦ (Д1).
2. Поиск лицевого счета по адресу из дерева объектов для оператора ВЦ (Д2).
3. Поиск лицевого счета по номеру из дерева объектов для бухгалтера (Д3).
4. Поиск лицевого счета по адресу из дерева объектов для бухгалтера (Д4).
5. Поиск лицевого счета по номеру из формы быстрого поиска (Д5).
6. Поиск лицевого счета по адресу из формы быстрого поиска (Д6).

Для каждого из действий фиксировалось время в секундах. После эксперимента было подсчитано среднее время на выполнение каждого из действий. Результаты приведены в табл. 2.

Таблица 2

**Среднее время выполнения действий, с**

Категория пользователя	Д1	Д2	Д3	Д4	Д5	Д6
Опытные пользователи	18	42	9	20	7	15

Как видно из таблицы, поиск лицевого счета занимает в среднем от 7 до 42 с в зависимости от режима работы пользователя и типа поиска. Можно попробовать уменьшить время, затрачиваемое на выполнение действий, путем использования голосового ввода команд.

Параметрами для вариантов поиска лицевого счета могут быть: полный номер лицевого счета, короткий номер (если это режим работы бухгалтера), полный адрес или короткий адрес – например номер квартиры (если будет использован контекст предыдущего сеанса).

На основе вышеизложенного было принято решение о целесообразности написания модуля, использующего голосовые возможности, для улучшения процесса взаимодействия при поиске лицевого счета. Для оценки эффективности необходимо сравнить время, затрачиваемое на поиск лицевого счета в голосовом режиме и обычном способом.

**Модули системы.** В рамках исследования был создан модуль, работающий с фразами пользователей, введенными голосом. В связи с тем, что необходимость в применении голосового интерфейса может иногда не возникнуть (нечеткая дикция, болит горло, нет настроения и т.д.), он был реализован в виде сервиса с использованием системы распознавания и синтеза речи google. Для вызова сервиса необходимо нажать на значок микрофона внизу экрана.

Модуль состоит из следующих компонентов: компонент синтеза; компонент распознавания, преобразующий речь человека в текст; компонент анализа и сопоставления текста с набором команд (он также уточняет команду в случае ошибки); компонент применения команды.

**Компонент распознавания.** Для преобразования речи в текст и синтеза использовались голосовые сервисы Google [5]. Компонент активируется нажатием на кнопку микрофона либо при запуске программы. Произнесенные фразы запоминаются, и по окончании ввода распознанный текст передается в разработанный модуль анализа и сопоставления с командами. Остановить работу модуля можно произнеся команду «готово» или нажав на микрофон внизу экрана, а исправить некорректно введенные данные – командой «вместо».

**Компонент синтеза.** Для преобразования речи в текст используются элемент html5 <audio> и сервисы Google.

**Компонент анализа и сопоставления текста с набором команд.** Текст, сформированный на основе фразы пользователя, разбирается на слова, и полученные слова анализируются в произнесенной последовательности, сравниваются с набором синонимов, и происходит поиск команд, соответствующих фразе.

Реализован механизм запоминания контекста. Принцип работы следующий: при первом вводе проанализированные данные сохраняются в памяти и остаются там как контекст. При последующих операциях, если вводятся данные того же типа, что уже существующие в контексте, они и все от них зависимые удаляются. Иначе – используется сохраненный контекст. При необходимости можно очистить контекст командой «заново». Модуль будет использовать права пользователя.

**Компонент применения команды.** Производит поиск и выполнение команды.

Разработанный модуль был использован для ввода и распознавания команд поиска лицевого счета по номеру и по адресу. В качестве формата команды была выбрана фраза «найти лицевой [параметр]», где параметром был 8 или 4-значный номер лицевого счета либо адрес. Ввод команды «Найти лицевой 1245» занял в среднем 5 с, а команды с произнесением 8-значного кода «Найти лицевой 21111245» – 7 с. Для ввода адреса лицевого счета понадобилось чуть больше времени – так, на ввод команды «найти лицевой адрес город Барнаул улица Попова дом 95 квартира 6» ушло в среднем 14 с, а если использовать запоминание контекста адреса (дом) и произносить только номер квартиры, например «найти лицевой квартира 7», то среднее время составит 6 с. Эксперименты проводились для опытных пользователей системы.

**Обработка речевого сигнала.** Вместе с тем обработка речевого сигнала является одной из ресурсоемких операций, требующих значительных вычислительных ресурсов. Для повышения эффективности было решено проводить оптимизацию работы за счет распараллеливания операций.

Одно из наиболее перспективных направлений анализа речи основано на использовании модели слуховой системы человека [6–9]. При таком подходе широкое применение находят методы цифровой фильтрации сигналов. Так, для вычисления частоты основного тона необходимо сделать следующие операции [6]:

- выделение спектра сигнала – вычисление отклика  $K$  фильтров размера  $W_s$ ;
- моделирование одновременной маскировки – вычисление отклика  $K$  фильтров размера  $W_1$ ;
- вычисление частоты основного тона – сопоставление с  $K_2$  шаблонами размера  $W_x$ .

Точность вычисления зависит от величины  $K$  – количества каналов анализа сигналов. Значение  $K$ , как правило, для практических целей выбирается в диапазоне от 250 до 1000. При таких значениях  $K$  обработка речевого сигнала с частотой дискретизации 8 кГц (что является минимально приемлемым для обработки речи) в реальном времени на одном современном процессоре общего назначения практически невозможна.

При обработке речевого сигнала может выполняться множество других операций, таких как: выделение различных параметров спектра и сегментирование сигнала по различным признакам звуков, поэтому общее время обработки одного отсчета сигнала растет и тем более не может быть осуществлено в реальном времени при использовании ресурсов одного процессора. Задача усложняется кратно, если требуется осуществлять потоковую обработку нескольких или многих речевых потоков, например в случае телефонного голосового сервиса.

Одним из средств повышения производительности приложений является применение параллельных вычислений.

**Описание основных операций модели слуха.** Основные операции модели слуховой системы человека:

- выделение спектра сигнала: на входе сигнал – вектор  $X$  размера  $N$ , на выходе спектрограмма – матрица  $Y$  размера  $K \times N$ , при этом каждая строка спектрограммы представляет собой результат фильтрации входного сигнала;

- моделирование одновременной маскировки: на входе спектрограмма – матрица  $Y$  размера  $K \times N$ , на выходе матрица  $Z_1$  одновременной маскировки размера  $K \times N$ , при этом каждая строка матрицы  $Z_1$  представляет собой результат двумерной фильтрации нескольких соответствующих строк матрицы  $Y$ ;

- моделирование последовательной маскировки: на входе спектрограмма – матрица  $Y$  размера  $K \times N$ , на выходе матрица  $Z_2$  последовательной маскировки размера  $K \times N$ , при этом каждая строка матрицы  $Z_2$  представляет собой результат фильтрации соответствующей строки матрицы  $Y$ ;

- вычисление частоты основного тона, на входе матрица  $Z_1$  одновременной маскировки размера  $K \times N$ , на выходе вектор частот основного тона  $T$  размера  $N$ , описание вычисления частоты основного тона приведено ниже.

Здесь  $N$  – размер сигнала, а  $K$  – количество каналов анализа сигнала. Каждому каналу соответствует определенная резонансная частота, и вместе частоты всех каналов образуют шкалу резонансных частот анализа.

Графически описанные операции и типы данных приведены на рис. 1.

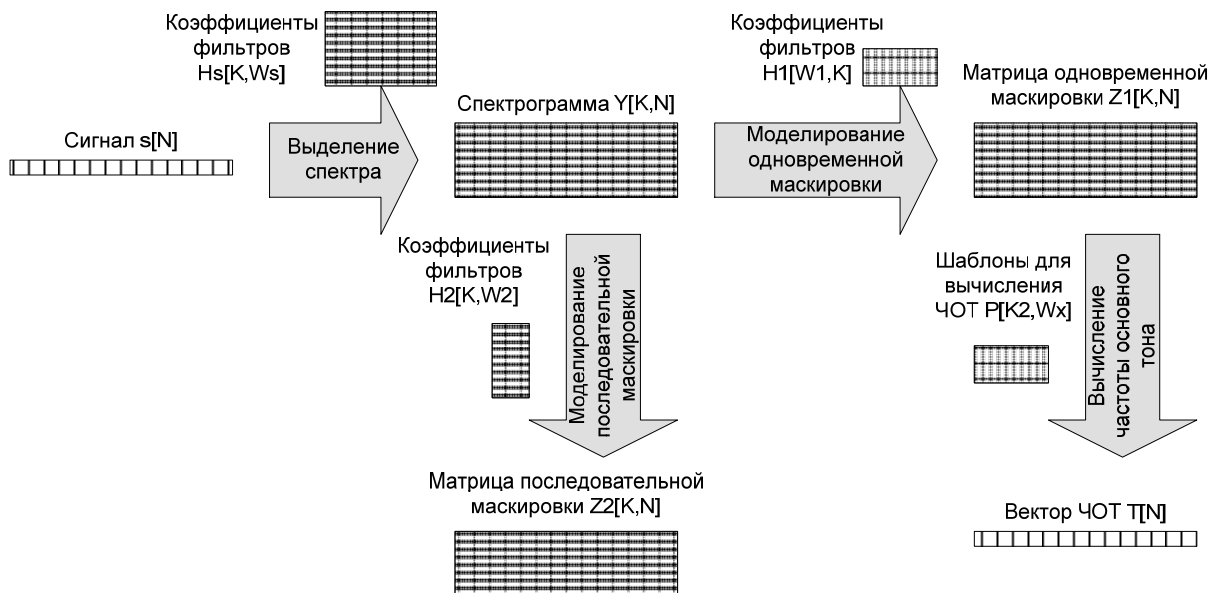


Рис. 1. Рассматриваемые операции модели слуховой системы человека

Приведем анализ возможности параллельных вычислений данных операций. Для этого рассмотрим каждую операцию сначала обособленно, а затем во взаимодействии с другими операциями.

**Приведение к параллельным вычислениям операций.** При фильтрации в  $K$  каналов при выделении спектра, моделировании одновременной и последовательной маскировки, вычисления в каждом из  $K$  каналов являются независимыми друг от друга, поскольку все, что необходимо для расчета отклика  $k$ -го канала, – это некоторые входные данные и коэффициенты фильтра  $k$ -го канала. Таким образом, вычисление отклика каждого из каналов может производиться независимым узлом вычислительной сети.

При вычислении частоты основного тона каждый столбец матрицы одновременной маскировки  $Z_1$  (который соответствует определенному временному отсчету) сопоставляется с множеством шаблонов, каждый из которых соответствует некоторому значению частоты основного тона. Затем выбирается тот шаблон, который подходит больше всего, а соответствующая ему частота канала считается частотой основного тона в данной точке. Вычисления частоты основного тона для каждого из  $N$  отсчетов являются независимыми друг от друга.

Таким образом, каждая из рассматриваемых операций может быть разбита на большое количество параллельных операций: в операциях фильтрации можно выделить  $K$  параллельных потоков в каждый момент времени, в операции выделения частоты основного тона можно выделить отдельный параллельный поток для каждого момента времени.

**Организация взаимодействия операций.** Задачей организации взаимодействия различных операций является выбор таких входных и выходных типов данных операций, которые бы минимизировали зависимости между функционально различными операциями.

Предлагается осуществлять цифровую фильтрацию блоками, для этого существует, например, метод пересечения с накоплением (overlap-save) [10]. Тогда при поступлении очередного блока входных данных может быть инициировано вычисление соответствующих операций над этим блоком.

Например, при поступлении блока сигнала может быть запущено  $K$  операций вычисления спектра. При поступлении блока спектра  $Y$  для  $k$ -го канала может быть инициировано вычисление  $k$ -го блока последовательной маскировки. При поступлении нескольких блоков спектра  $Y$  на определенных (соседних) каналах может быть инициировано вычисление соответствующего блока одновременной маскировки.

Таким образом, можно рассмотреть обработку одного блока входного сигнала определенного размера  $M_1$ , которая графически показана на рис. 2.

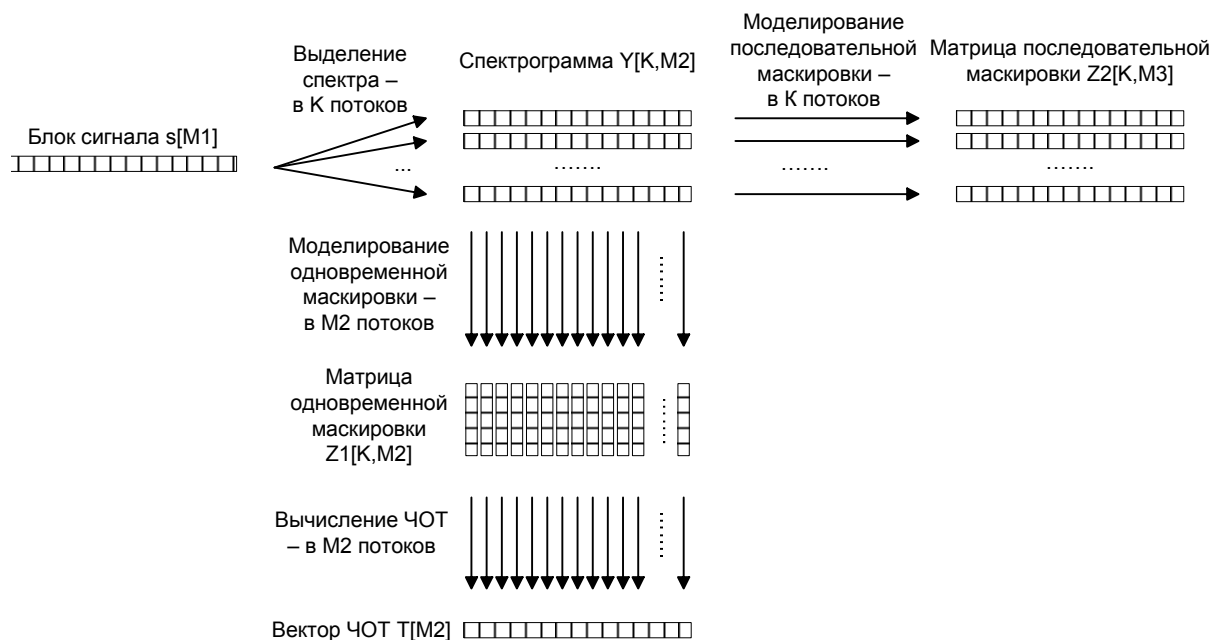


Рис. 2. Схема параллельных вычислений

**Распределение вычислений по функциональным узлам.** Вычисления, приведенные на рис. 2, можно выполнять параллельно. Однако поскольку в кластере различные потоки вычислений происходят на различных узлах, необходимо предусмотреть разделение различных операций по различным узлам.

Вычисление спектра в  $K$  потоков в реальном времени при достаточно больших значениях  $K$  может оказаться непосильным даже для достаточно мощного узла. Поэтому целесообразно распределить эту операцию между несколькими узлами, каждый из которых вычисляет спектр в несколько потоков.

Моделирование одновременной и последовательной маскировок являются более быстрыми операциями, чем вычисление спектра, это связано с размерами окон фильтрации. В случае если один узел кластера не способен выполнять данные операции в реальном времени, то их также мож-

но разделить на несколько узлов. В случае последовательной маскировки это происходит аналогично вычислению спектра. В случае одновременной маскировки необходимо дополнительно разделять блоки спектрограммы на несколько равных частей, поскольку части будут обрабатываться различными узлами.

Вычисление частоты основного тона по сравнению с фильтрацией является более быстрой операцией и может выполняться одним узлом, возможно, тем же узлом, что вычисляет одновременную маскировку.

**Заключение.** Приведенные результаты показывают, что голосовой интерфейс имеет потенциал для использования в биллинговых системах и существенно повысит эффективность использования специализированных баз данных при оптимизации некоторых режимов работы. При этом может быть получено преимущество не только по времени, но и по удобству работы. Применение голосового ввода позволит оптимизировать рабочую область экрана, убрав с нее некоторые элементы.

С другой стороны, операции, заданные в модели слуховой системы, могут быть организованы в виде параллельных вычислений на кластере. При воплощении данной схемы в конкретной вычислительной сети необходимо дополнительно выбрать подходящие размеры блоков и методы обмена информацией.

Работа выполнена в рамках проекта 7.701.2011 (проект 1/12) при поддержке Министерства образования и науки Российской Федерации.

#### *Литература*

1. Gierad Laput, Mira Dontcheva, Gregg Wilensky et al. PixelTone: A Multimodal Interface for Image Editing [Электронный ресурс]. – Режим доступа: <http://www.cond.org/pixeltone.pdf>, свободный (дата обращения: 02.06.2013).
2. Система «потребитель» автоматического приема показаний электросчетчиков для энергосбытовых компаний [Электронный ресурс]. – Режим доступа: <http://www.speechpro.ru/product/recognition/platform/potrebitel>, свободный (дата обращения: 02.06.2013).
3. Ронжин А.Л. Особенности дистанционной записи и обработки речи в автоматах самообслуживания / А.Л. Ронжин, А.А. Карпов, И.А. Кагиров // Информационно-управляющие системы. – 2009. – Вып. 42, т. 5. – С. 32–38.
4. Ронжин Ал.Л. Формирование профиля пользователя на основе аудиовизуального анализа ситуации в интеллектуальном зале совещаний / Ал.Л. Ронжин, В.Ю. Будков, Ан.Л. Ронжин // Труды СПИИРАН. – 2012. – Вып. 23. – С. 482–494.
5. Сухов К. Web Speech Api – html5. Распознавание речи на веб-странице // Системный администратор. – 2013. – № 1–2. С. 108–111.
6. Мещеряков Р.В. Сегментация и параметрическое описание речевого сигнала / Р.В. Мещеряков, В.П. Бондаренко, А.А. Конев // Изв. вузов. Приборостроение. – 2007. – Т. 50, № 10. – С. 3–7.
7. Мещеряков Р.В. Структура систем синтеза и распознавания речи // Изв. ТПУ. – 2009. – № 5, т. 315. – С. 121–126.
8. Бондаренко В.П., Конев А.А., Мещеряков Р.В. Обработка речевых сигналов в задачах идентификации // Изв. вузов. Физика. – 2006. – № 9. – С. 207–210.
9. Мещеряков Р.В. Система оценки качества передаваемой речи // Доклады ТУСУРа. – 2010. – № 2(22). – С. 324–329.
10. Блейхут Р. Быстрые алгоритмы цифровой обработки сигналов: пер. с англ. – М.: Мир, 1989. – 448 с.
11. Мещеряков Р.В. Характеристики надежности распределенных криптографических информационно-телекоммуникационных систем с ограниченными ресурсами / Р.В. Мещеряков, А.А. Шелупанов, Т.Ю. Зырянова // Вычислительные технологии. – 2007. – Т. 12, № S1. – С. 62–67.

---

#### **Суранова Дарья Александровна**

Аспирант каф. теоретической кибернетики и прикладной математики

Алтайского государственного университета

Тел.: 8 (385-2) 36-70-18

Эл. почта: [daria@suranova.ru](mailto:daria@suranova.ru)

**Тиунов Сергей Дмитриевич**

Аспирант каф. комплексной информационной безопасности  
электронно-вычислительных систем ТУСУРа  
Тел.: 8 (382-2) 41-34-26  
Эл. почта: t5d@mail.ru.

**Мещеряков Роман Валерьевич**

Профессор каф. комплексной информационной безопасности электронно-вычислительных систем ТУСУРа  
Тел.: 8 (382-2) 41-34-26  
Эл. почта: mrv@keva.tusur.ru

Suranova D.A., Tiunov S.D., Meshcheryakov R.V.

**Numerical implementation of effective access to the billing system with voice interface**

We offer the variants of using a voice interface to the billing system of payment for utility services with the use of efficient numerical algorithms for speech processing. The ability to optimize data access at the level of the database is shown, as well as the usage of paralleling the algorithms for speech signal processing. The schemes for implementation in the form of numerical algorithms and software are proposed.

**Keywords:** mathematical modeling, algorithm, speech, database, billing.