

УДК 044.03

Д.К. Михнов, Д.А. Шуть

Харьковский национальный университет радиоэлектроники

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ ПОДСИСТЕМЫ ПРОДАЖ СУПЕРМАРКЕТА ЗА СЧЕТ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ АВТОМАТИЧЕСКОГО РЕЧЕВОГО РАСПОЗНАВАНИЯ И БЕСПРОВОДНЫХ ТЕХНОЛОГИЙ

Рассмотрены особенности обработки информации в подсистеме продаж крупного торгового предприятия. Проведен анализ традиционных методов обработки информации в подобных системах. Предложена альтернативная модель доступа к информации, основанная на применении речевых технологий. Описана модель распознавания речевых команд, пригодная для построения речевого диалогового доступа.

торговое предприятие, подсистема продаж, продавец-консультант, оперативная информация, речевая команда, Bluetooth гарнитура, скрытая Марковская модель, кепстр

Введение

Успешное функционирование крупного современного супермаркета (сети супермаркетов) невозможно без применения современных информационных технологий. Характерными особенностями организации розничной торговли являются наличие большого числа поставщиков товаров, широкий ассортимент самих товаров, предоставление различными поставщиками одинаковых групп товаров по различным ценам, высокая конкуренция на рынке розничной торговли, независимо от групп товаров и т.д. [1]. Данные особенности непосредственно влияют на характер бизнес процессов, происходящих в типичном предприятии розничной торговли и тре-

буют применения особых технологий и технических средств для ведения бизнеса и автоматизации бизнес-процессов. Эффективность функционирования крупного торгового предприятия напрямую зависит от гибкости информационной системы и возможностей интеграции в ее состав специализированных технических средств.

Функционирование современного супермаркета в принципе невозможно без внедрения сложной специализированной информационной системы. Это обусловлено большой номенклатурой товаров, которая в крупных супермаркетах превышает 15000 наименований [1]; а также тем, что эта номенклатура постоянно изменяется, часто меняются поставщики одних и тех же групп товаров или даже самих товаров.

Постоянная изменчивость номенклатуры, активная дисконтная политика, проводимая супермаркетами типа Cash&Carry, а также тот факт, что продавцы и менеджеры супермаркета не способны, запоминать полную информацию обо всех товарах, вынуждают персонал активно выполнять те или иные запросы к информационной системе предприятия. В первую очередь, в оперативной информации постоянно нуждаются менеджеры специализированных торговых залов, непосредственно общающиеся с клиентами. Номенклатура товаров в подобных залах довольно обширна и даже хороший специалист зачастую не в состоянии запомнить всю информацию, которую могут затребовать клиенты [1]. Оперативная информация о товаре также может потребоваться продавцу за прилавком или иному сотруднику, не имеющему непосредственного доступа к терминалу информационной системы. К оперативной информации можно отнести следующие типы: наличие определенного товара на складе и его количество, наличие скидок на этот товар; расположение товара в торговом зале, если в информационной системе предусмотрено хранение информации о расположении товара, при условии что все стеллажи и полки имеют порядковые номера; дополнительная информация о товаре, которая зависит от поставщика и т.п. [1].

Для повышения эффективности выполнения продавцом-консультантом специализированного торгового зала супермаркета его функциональных обязанностей необходимо применение специализированных методов доступа к оперативной информации. Традиционные методы доступа к каталогам и справочникам торгового предприятия обладают рядом существенных недостатков. Следовательно, возникает необходимость поиска новых решений, ноу-хау, имеющих определенные преимущества перед традиционными методами. Применение достижений современной науки о распознавании и синтезе речи представляется весьма целесообразным для построения командной системы доступа к оперативной информации торгового предприятия. Применение данных технологий в подсистеме продаж торгового предприятия представляется возможным только в случае использования удобных беспроводных гарнитур для ввода/вывода речевой информации. Гарнитуры, использующие технологию беспроводной связи Bluetooth, удовлетворяют указанным требованиям.

Описание подхода к проектированию комплекса речевого взаимодействия с информационной системой

Существует несколько традиционных способов маркировки товаров и несколько типов специализированного оборудования получения информации, для каждого типа маркировки. Способы маркировки включают в себя артикульные номера, штрихкодирование, другие специальные отметки на этикетках или непосредственная печать какой-либо инфор-

мации на этикетках. Помимо традиционных РС терминалов в розничной торговле используется такое оборудование, как этикет-пистолеты, обычная телефонная связь, POS терминалы сбора данных. Отдельно стоит упомянуть систему бумажных каталогов товаров, в которую входят как каталоги от производителей, так и каталоги супермаркета. Каждый из вышеупомянутых способов маркировки товара и способов получения информации о нем имеет свои преимущества и недостатки. Все эти способы можно комбинировать в том или ином порядке. Но максимального эффекта от их использования можно достичь только посредством наиболее тесной интеграции специализированных технических средств с информационной системой супермаркета. При этом более дорогостоящие средства не всегда являются наиболее эффективными, а эффект от их неумелого внедрения может быть значительно ниже, чем затраты на их покупку. Так, терминалы сбора данных, безусловно, обладают рядом неоспоримых преимуществ. Информация о товарах может храниться непосредственно в данных устройствах. Существуют модели терминалов сбора данных, оснащенные сканерами штрих-кодов. На устройства данного класса, зачастую, устанавливают развитые операционные системы, что позволяет писать для них приложения, расширяющие их функциональность. Однако можно выделить и ряд недостатков подобного оборудования. Главным недостатком является их дороговизна, что является существенным препятствием для оснащения каждого сотрудника, нуждающегося в оперативной информации, подобными устройствами. Еще одним существенным недостатком является относительное неудобство их интерфейса: дешевые модели оснащены только лишь примитивной клавиатурой, хотя более дорогие модели, с сенсорным вводом, более удобны.

Ввиду возрастающей конкуренции в области розничных продаж, а также всего вышесказанного, особый интерес представляет задача поиска новых подходов автоматизации торговых предприятий, а также интеграции этих новых решений в уже имеющиеся системы. По мнению аналитиков многих ведущих IT-компаний мира, в ближайшее десятилетие интерфейс персонального компьютера может претерпеть значительные изменения, ввиду ожидаемого скачка в развитии речевых технологий [2]. Интерес представляет использование систем командного управления и речевого командного интерфейса, а также систем синтеза речи. Данные области науки о моделировании речи являются наиболее развитыми и уже с успехом применяются в коммерческих приложениях. В данной статье будет рассмотрена модель распознавания речевых команд, пригодная для создания речевого интерфейса, обеспечивающего получение оперативной информации о товаре.

Для реализации подсистемы распознавания предлагается использовать скрытую Марковскую модель (СММ) [3]. Предположим, мы имеем словарь

из V слов (речевых команд), которые необходимо распознать, и каждое слово должно быть промоделировано отдельной СММ. Далее, предположим, что для каждого слова в словаре мы имеем обучающее множество из K появлений каждого произнесенного слова (произнесенного одним или несколькими дикторами), причем каждое появление слова представляет собой наблюдаемую последовательность, причем наблюдения являются некоторым представлением (спектральных и/или временных) характеристик слова. Для выполнения распознавания изолированных слов необходимо выполнить следующие операции: для каждого слова v из словаря необходимо построить СММ λ , то есть получить оценку параметров (A, B, π) , которая оптимизирует правдоподобие наблюдаемых векторов из обучающего множества для v -го слова; для каждого неизвестного слова, которое необходимо распознать, должен быть выполнен анализ признаков наблюдаемой последовательности

$$O = \{O_1 O_2 \dots O_T\}, \quad (1)$$

далее для всех возможных моделей должны быть вычислены вероятности правдоподобия

$$P(O|\lambda^v), \quad 1 \leq v \leq V, \quad (2)$$

далее выбирается слово, для которого вероятность правдоподобия максимальна, то есть

$$v^* = \arg \max_{1 \leq v \leq V} [P(O|\lambda^v)]. \quad (3)$$

Этап вычисления вероятностей выполняется в соответствии с алгоритмом Витерби [4]. Для словаря небольшого объема, например, $V = 100$ слов, и модели $N = 5$ состояний, а также $T = 40$ наблюдений

неизвестного слова, для распознавания потребуется порядка 10^5 вычислений.

Одним из способов получения наблюдаемых векторов O из речевых отсчетов s является выполнение спектрального анализа [4]. Будем исходить из того, что мы обрабатываем только речевые отсчеты, соответствующие произнесенному слову, то есть весь фон до и после произнесенного слова был убран соответствующим алгоритмом определения границ слова [5]. Тип спектрального анализа, который часто используется и будет использован нами в распознавателе, называется Линейное Предсказательное Кодирование (КЛП). Введем следующие обозначения:

$$\tilde{S}(n) = S(n) - \alpha S(n-1); \quad (4)$$

$$x_f(n) = \tilde{S}(Mf + n), \quad 0 \leq n \leq N-1, \quad 0 \leq f \leq L-1; \quad (5)$$

$$\tilde{x}_f(n) = x_f(n) \times W(n), \quad 0 \leq n \leq N-1; \quad (6)$$

$$R_f(m) = \sum_{n=0}^{N-m} \tilde{x}_f(n) \tilde{x}_f(n+m), \quad 0 \leq m \leq p, \quad (7)$$

a_f – КЛП коэффициенты, $0 \leq m \leq p$; $c_f(m)$ – кепстральные коэффициенты, $1 \leq m \leq Q$;

$$\mathcal{E}_f(m) = c_f(m) \times W_c(m), \quad 1 \leq m \leq Q; \quad (8)$$

$$\Delta \mathcal{E}_f(m) \cong \frac{\partial \mathcal{E}_f(m)}{\partial t}, \quad 1 \leq m \leq Q. \quad (9)$$

Система в целом строится на блочной модели обработки, в которой обрабатывается кадр из N_A отсчетов и вычисляется вектор признаков O_f . Алгоритм состоит из семи шагов [3]. Процедура выделения вектора характерных признаков речевого сигнала приведена на рис. 1.

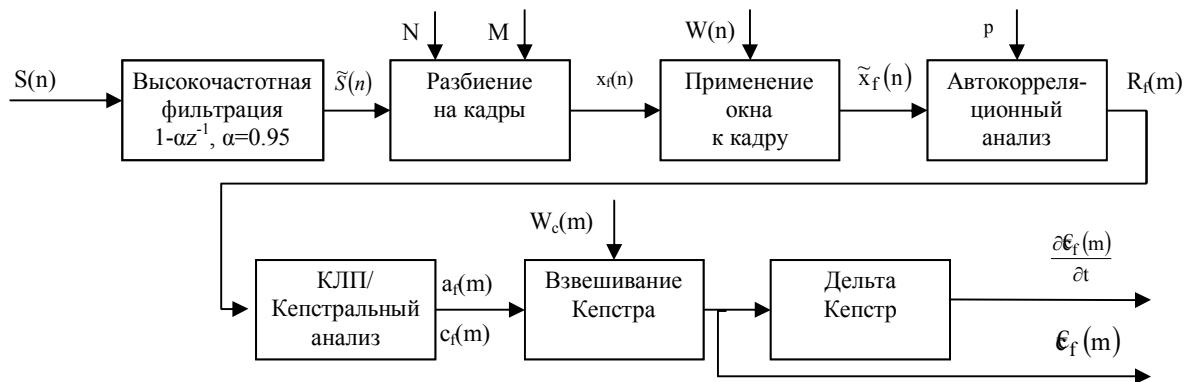


Рис. 1. Процедура выделения вектора признаков из речевого сигнала

Первый шаг алгоритма – предобработка, состоит в фильтрации высокочастотного диапазона входного речевого сигнала, собранного с частотой дискретизации 8 КГц. *Второй шаг* – блокирование в кадры, состоит в следующем: секции из N_A последовательных отсчетов речевого сигнала (будем использовать $N_A = 256$, что соответствует 32 мс сигнала) используются как один кадр. Примыкающие кадры перекрывают друг друга на M_A отсчетов (будем использовать $M_A = 80$, что соответствует 12 мс неперекрывающегося пространства кадра или 20 мс

перекрытия кадров). *Третий шаг* – применение окна, состоит в следующем. Каждый кадр умножается на окно Хэмминга, размером в N_A отсчетов. Окно $w(n)$ применяется для минимизации неблагоприятных эффектов, вызванных выходом N_A -размерной секции за пределы речевого сигнала.

Далее применяется автокорреляционный анализ. Для каждого набора речевых отсчетов строится функция автокорреляции для получения набора из $(p + 1)$ коэффициентов, где p – это порядок желаемого КЛП анализа. Будем использовать $p = 8$.

Далее проводится КЛП/Кепстральный анализ [4, 5]. Для каждого кадра, вектор КЛП коэффициентов вычисляется по автокорреляционному вектору с помощью рекурсивного метода Левинсона или метода Дурбина. Кепстральные коэффициенты, полученные из КЛП, далее объединяются в вектор из Q компонентов, где $Q > p$ и $Q = 12$.

Далее производится применение окна к вектору кепстральных коэффициентов. Q -компонентный вектор $c_f(m)$ для кадра f взвешивается при помощи окна $W_c(m)$, которое имеет вид

$$W_c(m) = 1 + (Q/2)\sin(\pi m/Q), 1 \leq m \leq Q. \quad (10)$$

При этом получается вектор

$$\epsilon_f(m) = c_f(m) \times W_c(m). \quad (11)$$

Далее вычисляется Дельта Кепстр. Производная по времени от последовательности взвешенных кепстральных векторов аппроксимируется полиномиальным окном первого порядка. Размер окна равен $(2K + 1)$, окно центрировано относительно текущего вектора.

Будем использовать $K = 2$, т.е. окно размером 5 для вычисления производной. Кепстральная производная (вектор дельта кепстра) вычисляется как

$$\Delta \epsilon_f(m) = \left[\sum_{k=-K}^K k \epsilon_{f-t-k}(m) \right] \times G, 1 \leq m \leq Q, \quad (12)$$

где G – это множитель, выбранный таким образом, чтобы сделать дисперсии $\epsilon_f(m)$ и $\Delta \epsilon_f(m)$ равными.

Будем использовать $G = 0,375$.

Наблюдаемый вектор O_f , используемый для распознавания и обучения, состоит из взвешенного кепстрального вектора и соответствующего взвешенного вектора дельта кепстра, то есть состоит из 24 коэффициентов:

$$Q_f = \{ \epsilon_f(m), \Delta \epsilon_f(m) \}. \quad (13)$$

Выше был описан сам алгоритм распознавания. Обучение СММ для каждого из слов может быть выполнено классическими статистическими методами, разработанными для этих целей [3]. Для выделения границ слова может быть использован взвешенный суммарный критерий, состоящий из критерия превышения определенного уровня энергии сигнала и критерия пересечения математического ожидания [5]. Рассмотренная модель может быть использована в качестве составной части речевого интерфейса пользователя с информационной системой супермаркета.

Обратная связь с информационной системой может быть реализована с помощью библиотек сторонних разработчиков, отвечающих стандарту Microsoft Speech API. Эффективное использование речевого интерфейса предполагает ряд особенностей, необходимых для реализации в виде программно-аппаратного комплекса. К таким особенностям можно отнести создание эффективной диалоговой модели, создание удобной для произношения маркировки товаров, применение в системе удобных гарнитур. В качестве маркировки товаров удобно

использовать численные артикулы. Последовательность цифр легко распознается речевыми системами. В качестве гарнитуры предлагается использовать гарнитуры Bluetooth и соответственно стандарт беспроводной связи Bluetooth. Стек сетевых протоколов ОС Windows не поддерживает возможность использования речевых каналов Bluetooth, поэтому предлагается реализовать систему речевого распознавания в виде отдельного программно-аппаратного комплекса, тем более, что это достаточно распространенная практика в области речевого распознавания; так реализации систем Speech-To-Text представлены на рынке в частности в виде отдельных портативных устройств. Современная элементная база позволяет создать систему, соответствующую, например, требованиям к системам типа "Command and Control" такого стандарта, как SALT. Данный стандарт предусматривает в качестве минимальных технических характеристик для подобных систем процессор Intel 486 с частотой 66 МГц и память 1Мб [6]. Современные сигнальные процессоры и RISC процессоры архитектуры ARM9 работают на частотах более 150 МГц. Следовательно, на их основе можно построить аппаратные комплексы, способные обрабатывать речевые потоки от трех источников (ограничение пикосети Bluetooth) одновременно.

Выводы

Рассмотренный вариант повышения эффективности функционирования торгового предприятия довольно просто может быть реализован и интегрирован в имеющиеся системы. Помимо непосредственного экономического эффекта, внедрение описанной системы позволит сэкономить на покупке дорогостоящего торгового оборудования. Еще одной особенностью внедрения системы является простота обучения пользователей работе с ней. Предложенное решение является очень гибким и легко может быть интегрировано в любую коммерческую информационную систему, поддерживающую механизм расширений. Экономический эффект от внедрения подобной системы очевиден, так как с улучшением функционирования ИС, улучшается эффективность выполнения бизнес-процессов.

Список литературы

1. Гурьянов И.Б. *Интегрированные системы управления: пути повышения эффективности торгового предприятия [Электр. ресурс]. – Режим доступа: http://www.sprut.vimas.com/doklad_gur.zip.*
2. Gupta A. *The Future of Speech Technologies [Электр. ресурс]. – Режим доступа: <http://www.microsoft.com/presspass/features/2006/08-08speechtek2006.msp>.*
3. Rabiner L. R. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. // Proceedings of the IEEE, 1989.*
4. Furui S. *Digital Speech Processing, Synthesis and Recognition Second Edition Revised and Expanded. // CRC, November, 2000.*

5. Kingston A. *Speech Recognition by Machine*. // Victoria university of Wellington, Technical Report CS-TR-92/2, October 1992.

6. *Speech Application Language Tags (SALT) 1.0 Specification*, July 2002 [Електр. ресурс]. – Режим доступа: <http://www.saltforum.org/saltforum/downloads/SALT1.0.pdf>.

Поступила в редколлегию 17.04.2007

Рецензент: д-р техн. наук, проф. Е.В. Бодянский, Харьковский национальный университет радиоэлектроники, Харьков.