

Речевые технологии в задачах построения информационных и контролирующих систем

Владимир Ковтун, Владимир Бабкин, Артур Ланнэ (Москва)

В статье рассмотрена методика, а также программные и аппаратные средства для построения информационных и контролирующих систем для железнодорожных вокзалов и аэропортов; оценены их практические реализации.

Объявленная в заголовке задача на самом деле является очень широкой и охватывает большое число технических, технологических, гуманитарных и иных систем. Среди них весьма значимыми, практически востребованными, экономически эффективными и в то же время предметно представительными являются информационные и контролирующие системы для вокзалов различного назначения и аэропортов. Рассмотрев их, мы можем определить общие черты, оценить эффективность, определить базовые элементы, разработка и производство которых в тех либо иных комбинациях необходимы для построения и эксплуатации таких систем.

ЦИФРОВАЯ ИНФОРМАЦИОННО-ЗВУКОВАЯ СИСТЕМА «ДИКТОР» ДЛЯ ЖЕЛЕЗНОДОРОЖНЫХ ВОКЗАЛОВ И АЭРОПОРТОВ

Прежде всего, система [3] строится на принципах цифровой обработки речевых сигналов. Представление речи в цифровой форме позволяет использовать вычислительную технику, обеспечивает возможность с малыми материальными и энергетическими затратами сделать систему коммерчески привлекательной, функционально достаточной, удобной для эксплуатации.

Важнейшим элементом системы является сервер, в котором хранится звуковая (речевая) база. Её составление зависит от требований и является неформальным творческим делом. В предлагаемой системе речевая база строится по методу макрокомпилятивного синтеза речи [1], идея которого напоминает идею и технологию мозаики. В результате получается уплотнённое, сжатое представление конкретно-

го, предметно-ориентированного речевого набора. Качество речи получается достаточно хорошим (существенно выше качества синтеза речи по тексту). Чтобы оценить эффективность предлагаемого решения, рассмотрим конкретный практический пример созданной системы «Диктор» для аэропорта Пулково-1 в Санкт-Петербурге. Необходимо было сформировать 100 различных объявлений по каждому рейсу при прилёте и вылете. При наличии записанных 700 аэропортов и 600 авиакомпаний, а также четырёхзначных номеров рейсов получается количество вариантов объявлений $100 \times 700 \times 600 \times 10000 = 420$ млрд. Таковы возможности системы. Для обеспечения подобного количества объявлений при прямой записи каждого (1,5...2 Мб на одно объявление) потребовался бы объём памяти винчестера порядка 600...800 тыс. Тб. А затраты времени (1 мин на объявление) составили бы 800 тыс. лет, что нереально.

На создание речевой базы (словаря) на одном языке методом макрокомпилятивного синтеза потребовалось около трёх месяцев. Это время было затрачено в том числе и на достижение приемлемого качества звучания объявления: приходилось осуществлять запись диктора до трёх раз. Словарь для одного языка получился состоящим из 1600 отрезков фраз и слов и занял объём памяти на винчестере около 80 Мб. Таким образом, применение метода макрокомпилятивного синтеза позволило уплотнить (сжать) необходимую информацию примерно в 10 000 раз. Но это ещё не всё. Подобная запись информации весьма удобна с точки зрения быстрого обновления, коррекции, модификаций.

Программа, синтезирующая объявления и выполняющая сервисное, удобное в эксплуатации средство, занимает от 1 до 2 Мб памяти. Речевой цифровой синтезированный сигнал с выхода ЦАП, установленного в сервере (звуковая карта или специализированный ЦАП), подаётся на звуковые колонки. Параллельно там, где это необходимо, на световом табло (светодиодные, блинкерные и иные типы динамических табло, мониторы, плазменные панели и т.д.) дублируются данные речевой информации с компьютера-сервера и, таким образом, организуется единый центр управления звуковой и визуальной информацией.

МНОГОКАНАЛЬНАЯ СИСТЕМА КОНТРОЛЯ

Многоканальная система контроля [2] переговоров предназначена для записи разговоров с телефонных или микрофонных линий и последующего прослушивания. Её отличительная особенность состоит в доступности для любого выделенного сотрудника. Основой системы является многоканальный регистратор речевой информации, способный вести запись одновременно по 8, 16, 24 или 32 каналам. Если количество каналов больше, то используются несколько регистраторов, помещённых в стойку. При этом гарантируется сохранность всех записанных разговоров до 180 суток по каждому каналу (длительность записи может быть изменена по условиям проекта). Запись ведётся на жёсткий магнитный диск компьютера-регистратора речевой информации с возможностью воспроизведения как на самом регистраторе, так и на удалённом компьютере пользователя по локальной вычислительной сети. Многодневная запись разговоров, большое число каналов и экономное использование локальной сети для её освобождения в целях оперативной передачи информации диктуют использование методов компрессии речи в реальном времени, т.е. вокодеров.

При этом вокодеры должны быть компактными, энергоэкономичными, с высоким качеством синтезированной речи, в перспективе — помехозащищёнными (имеется в виду защита от акустических и линейных помех в тех случаях, когда локальная сеть использует радиоканалы).

Низкоскоростной вокодер

Для сжатия речи в интересах многодневного и многоканального хранения переговоров, экономного использования локальной сети, качественного воспроизведения речевой информации, компактного построения аппаратуры, совместимой с ПК, а также для других задач речевых технологий были разработаны различные (по скоростям, методам преобразования речи, помехоустойчивости, методам реализации (один или несколько вокодеров на кристалле)) варианты вокодеров. Они представлены на сайте www.dsp.sut.ru, где кратко указан метод компрессии, число вокодеров на кристалле, вид рекомендации или оригинальный алгоритм, приведено звучание речи в идеальном канале и в ряде случаев при наличии шума (шум белый или коррелированный с заданным уровнем). Все реализации выполнены на приборах фирмы Texas Instruments. Применительно к рассматриваемой задаче были проведены испытания различных типов вокодеров по совокупности параметров: качество речи, сложность, степень сжатия (скорость) и выбрана рекомендация ITU G 729, которая реализована на процессоре TMS320C54 с оптимизацией кода. Эта рекомендация обеспечивает скорость 8 Кбит/с, что эквивалентно сжатию в восемь раз по сравнению с 64 Кбит/с, а оптимизация кода позволила реализовать на одном кристалле (одном чипе) восемь кодеров (без декодеров) и таким образом сделать платы на 8, 16, 24 и 32 канала. Декодеры же выполнены на ПК в виде программы, которая устанавливается на всех разрешённых персональных компьютерах.

Практические результаты

Посмотреть и наглядно оценить качество и практическую ценность системы в целом и отдельных её подсистем возможно в различных географических местах.

Так, работу системы «Диктор» можно посмотреть и послушать в С.-Пе-

тербурге на Московском, Витебском и Финляндском железнодорожных вокзалах и в аэропортах Пулково-1 и Пулково-2, в Москве на Курском, Казанском, Павелецком, Савёловском железнодорожных вокзалах, в Ростове-на-Дону на Главном и Пригородном железнодорожных вокзалах, в Краснодаре на железнодорожном вокзале Краснодар-1.

За период эксплуатации (в течение от 1 года до 15 лет в зависимости от объекта внедрения) система «Диктор» работала в круглосуточном режиме с надёжностью, определяемой надёжностью ПК, и по своим функциональным возможностям легко адаптировалась к быстроизменяющейся обстановке. Операторы свидетельствуют: ни у кого из них не возникло проблем при освоении работы с системой, а голосовые нагрузки у дикторов снизились в 20...30 раз.

Автоматический вывод информации с системы «Диктор» на динамические табло реализованы в Ростове-на-Дону на Главном железнодорожном вокзале, в Москве на Курском вокзале и в Краснодаре на вокзале Краснодар-1. Система контроля переговоров введена в эксплуатацию на вокзальном комплексе Краснодар-1 (16 объектов контроля), в С.-Петербурге на Московском вокзале (48 объектов контроля), на Витебском вокзале (16 объектов контроля), на Ладужском вокзале (24 объекта контроля), в железнодорожном агентстве (80 объектов контроля), в Москве на Ленинградском вокзале (56 объектов контроля), на железнодорожных вокзалах г. Твери (8 объектов контроля), г. Петрозаводска (16 объектов контроля), г. Мурманска (16 объектов контроля), г. Пскова (8 объектов контроля) и г. Великие Луки (8 объектов контроля). В процессе эксплуатации системы контроля отмечено значительное повышение технологической и трудовой дисциплины работников вокзала. Появилась возможность контролировать и анализировать работу подчинённых в режиме реального времени (прямое прослушивание) или за прошедший период времени (просмотр архива), а также быстро и объективно разбирать конфликтные ситуации по жалобам пассажиров прямо на рабочих местах.

Подобные системы или их составные части могут быть использованы при контроле связи на узлах связи специального назначения, в банковских системах, при контроле техно-

логических процессов, в кабинах управления техническими объектами, в обучающих тренажерах и т.д.

ЛИТЕРАТУРА

1. *Сорокин В.И.* Синтез речи. М.: Радио и связь, 1992.
2. *Goldberg R., Riek L.* A Practical Handbook of Speech Coders. New York: CRC Press.
3. *Бень А.В., Ковтун В.П., Ланнэ А.А.* Информационные и контролирующие системы для ж/д вокзалов и аэропортов. Труды учебных институтов связи. 2007, № 177. С. 32–41.

