УДК: 004.93:[57.087.1];518.22

Возможности по распознанию речи

Силаев А.В. Орёл ГТУ, 12-Р, silaev_12-r@mail.ru

рук. Абашин В.Г.

Актуальность: В связи с активным внедрением электронной техники в жизни людей, появилась необходимость использования человеческой речи для облегчения управления электронными устройствами, бытовой техникой и т.д.

Цель: Изучение технологии распознавания речи, применяемой для широкого использования при создании средств речевого управления электронных устройств.

Распознавание речи - технология, позволяющая использовать естественный для человека речевой интерфейс для взаимодействия с электронной техникой. По мере развития компьютерных систем становится все более очевидным, что использование этих систем намного расширится, если станет возможным использование человеческой речи при работе непосредственно с компьютером, и в частности станет возможным управление машиной обычным голосом в реальном времени, а также ввод и вывод информации в виде обычной человеческой речи.

Существующие технологии распознавания речи не имеют пока достаточных возможностей для их широкого использования, но на данном этапе исследований проводится интенсивный поиск возможностей употребления коротких многозначных слов (процедур) для облегчения понимания. Не так давно распознавание речи нашло реальное применение в жизни, пожалуй, только в тех случаях, когда используемый словарь сокращен до 10 знаков, например при обработке номеров кредитных карт и прочих кодов доступа в базирующихся на компьютерах системах, обрабатывающих передаваемые по телефону данные. Распознавание 20 тысяч слов естественного языка было недостижимой задачей до создания Conversa Web. Conversa Web имеет словарь в 110 000 слов и позволяет отдавать команды в естественной манере. Но все же в технологии распознания речи остаются еще не решенные проблемы.

Для успешного распознавания речи следует решить следующие задачи:

- обработку словаря (фонемный состав);
- обработку синтаксиса;
- сокращение речи (включая возможное использование жестких сценариев);
- выбор диктора (включая возраст, пол, родной язык и диалект);
- тренировку дикторов;
- выбор особенного вида микрофона (принимая во внимание направленность и местоположение микрофона);
- условия работы системы и получения результата с указанием ошибок.

Существующие сегодня системы распознавания речи основываются на сборе всей доступной (порой даже избыточной) информации, необходимой для распознавания слов. Исследователи считают, что таким образом задача распознавания образца речи, основанная на качестве сигнала, подверженного изменениям, будет достаточной для распознавания, но тем неменее в настоящее время даже при распознавании небольших сообщений нормальной речи, пока невозможно после получения разнообразных реальных сигналов осуществить прямую трансформацию в лингвистические символы, что является желаемым результатом. Вместо этого проводится процесс, первым шагом которого является первоначальное трансформирование вводимой информации для сокращения обрабатываемого объема так, чтобы ее можно было бы подвергнуть компьютерному анализу. Примером является "техника сопоставления отрезков", позволяющая сократить вводимую информацию с 50'000 до 800 бит в секунду. Следующим этапом является спектральное представление речи, получившееся путем преобразования Фурье. Результат преобразования Фурье позволяет не только сжать информацию, но и дает возможность сконцентрироваться на важных аспектах речи, которые интенсивно изучались в сфере экспериментальной фонетики. Спектральное представление достигнуто путем использования широко-частотного анализа записи. Хотя спектральное представление речи очень

полезно, необходимо помнить, что изучаемый сигнал весьма разнообразен. Разнообразие возникает по многим причинам, включая:

- различия человеческих голосов;
- уровень речи говорящего;
- вариации в произношении;
- нормальное варьирование движения артикуляторов (языка, губ, челюсти, нёба).

Для устранения негативного эффекта влияния варьирования голосового тракта на процесс распознавания речи было использовано множество методов. Первым делом рассматривалась характеристика пространства траектории артикуляторных органов, включая гласные, используемые говорящим. Наиболее удачные формы трансформации, использованной для сокращения различий, были впервые представлены Сакоя & Чибо и назывались динамичными искажениями (dynamic time warping). Техника динамичного искажения используется для временного вытягивания и сокращения расстояния между искаженным спектральным представлением и шаблоном для говорящего. Использование данной техники дало улучшения точного распознавания (~20-30%). Метод динамичного искажения используют практически все коммерчески доступные системы распознавания, показывающие высокую точность сообщения при использовании. (Рис.1)

Вначале сигнал преобразовывается в спектральное представление, где определяется немногочисленный, но высокоинформативный набор параметров. Затем определяются конечные выходные параметры для варьирования голоса и производится нормализация для составления шкалы параметров, а также для определения ситуационного уровня речи. Вышеописанные измененные параметры используются затем для создания шаблона. Шаблон включается в словарь, который характеризует произнесение звуков при передаче информации говорящим, использующим эту систему. Далее в процессе распознавания новых речевых образцов, эти образцы сравниваются с шаблонами, уже имеющимися в словаре,

используя динамичное искажение и похожие метрические измерения. В настоящее время этот метод изучается и дополняется.

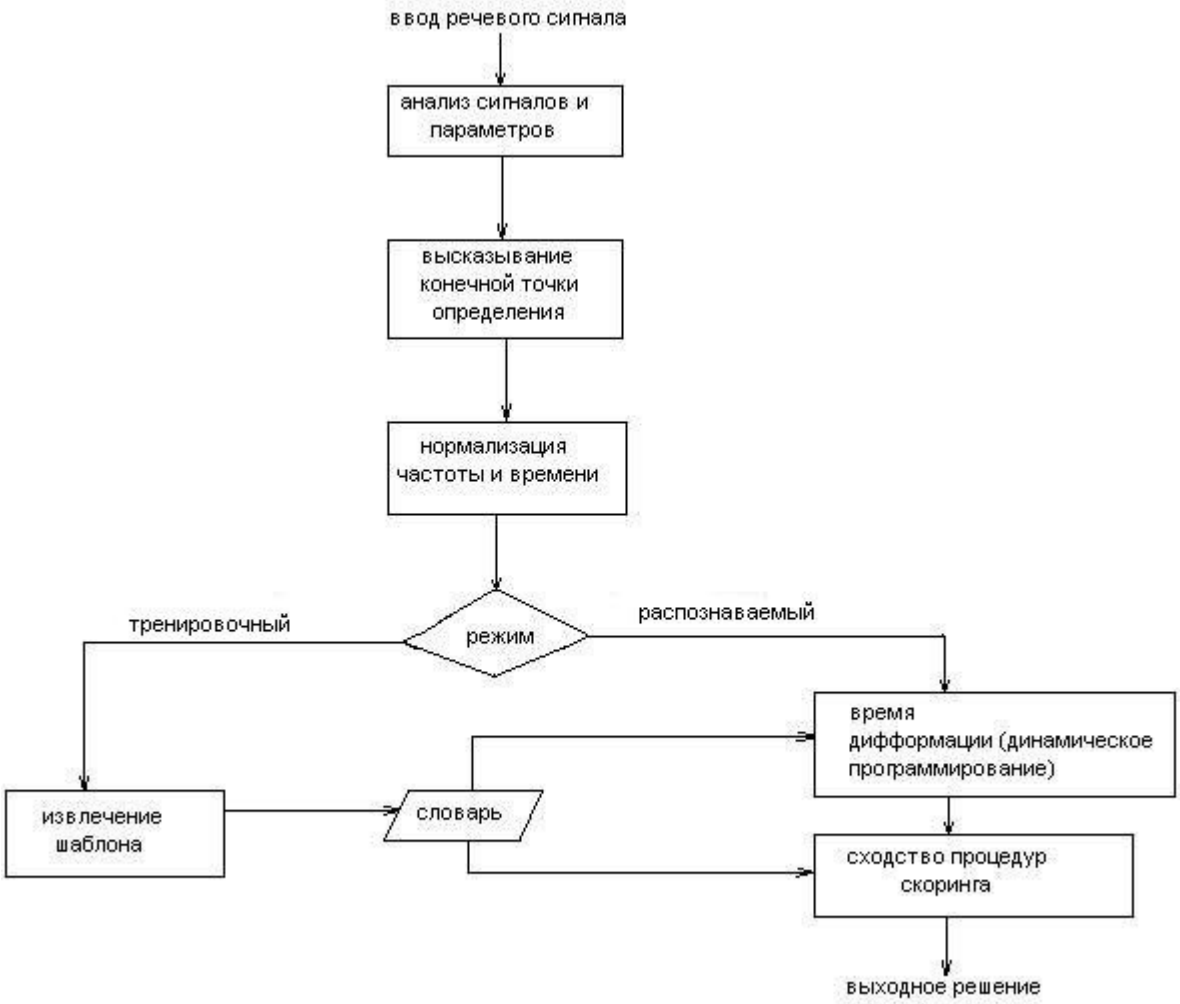

Рис.1. Техника динамичного искажения

Sakrament ASR Engine - разработка компании «Сакрамент», осуществляющая высокоточное распознавание речи на различных платформах. Технология распознавания речи используется при создании средств речевого управления – программ, управляющих действиями компьютера или другого электронного устройства с помощью голосовых команд, а также при организации телефонных справочных и информационных служб. Программа рассчитана на применение в различных аппаратных системах и программных приложениях, использующих технологии распознавания речи, таких как: IVR-системы, мобильные электронные устройства, бытовая техника и другие. Sakrament ASR

Engine может быть легко перенесена на любую существующую программную или аппаратную платформу, а также настроена под конфигурацию любого приложения. Технические характеристики системы распознавания речи Sakrament ASR Engine: Языки: Не зависимая от языка.

- Точность: Точность распознавания достигает 95-98 процентов. Качество распознавания зависит от размера используемых словарей, качества транскрипции, показателя связанности распознаваемых слов, от уровня фонового шума, от параметров используемых каналов связи и характеристик микрофонов.
- Размер Словаря: Виртуально неограничен, распознавание в реальном времени с использованием активных словарей. Возможность создания активных словарей по требованию заказчика, обучение системы с помощью синтезатора речи.
- Дикторонезависмая: Система распознавания поддерживает дикторонезависимый и дикторозависимый режимы работы.
- Поддерживаемые стандарты: MS SAPI 4.1, SAPI 5.1, TAPI 3.0.
- Поддерживаемые платформы: Windows 98, ME, NT, 2000, XP, 2003, Vista.

Система распознавания речи Sakrament ASR Engine полностью совместима с синтезатором речи Sakrament TTS Engine и может быть легко в него интегрирована, что значительно расширяет область применения разработок компании “Сакрамент”. Совместное использование технологий позволяет создавать полнофункциональные онлайновые телефонные информационно-справочные службы и Интернет – сервисы.

Вывод: Внедренные системы распознавания речи облегчат, расширят использование, управление и применение электронных устройств в различных областях жизнедеятельности человека.

Список литературы:

1. Распознавание речи, технология Sakrament ASR Engine [Электронный ресурс] /http://www.sakrament.com/main.php?Lang=ru&TopId=20&Category=2/ (с изм. и доп.) – режим доступа: (дата обращения 29.02.10)

2) Распознавание речи [Электронный ресурс] /http://prof9.narod.ru/library/lib007/doc043.html/ (с изм. и доп.) – режим доступа: (дата обращения 29.02.10)

3) Классификация систем распознавания речи [Электронный ресурс] /http://habrahabr.ru/blogs/artificial_intelligence/64572/ (с изм. и доп.) – режим доступа: (дата обращения 29.02.10)