



Использование современных информационных технологий для обработки речевых материалов

Устюгов С.В., кандидат военных наук

Котов М.А.

Леднов Д.А.

Главатских И.А.

В статье дается краткое описание инновационной отечественной разработки в области речевых технологий компании ООО «Стэл КС», получившей название комплекс программных средств (КПС) «Голос». КПС «Голос» обеспечивает: определение параметров речи; идентификацию языка; распознавание дикторов; поиск ключевых слов; адаптацию, конфигурирование и настройку под требования заказчика. Комплекс реализован на базе промышленного ноутбука (от 4-х вычислительных ядер) с процессором Intel Core-i7 и двумя высокоэффективными графическими ускорителями NVIDIA GTX 780M. Его вес составляет 3,5 кг. Он позволяет быстро и эффективно обрабатывать речевую информацию автономно в любом месте, гибко и просто настраивать процесс обработки и поиска информации, накапливать и сохранять полученную информацию для дальнейшего использования.

• автоматизированная обработка речевых данных • программно-аппаратный инструментарий для обработки речевой информации • комплекс программных средств «Голос»

A brief description of the innovative domestic development of company «Stel CS» in the field of speech technology gives in this article. This development was named complex of software «Voice». «Voice» provides: determination of speech signal parameters; language identification; speaker recognition; keyword search; adaptation, configuration and customer requirements setting. The complex is realized on the basis of industrial laptop (not less than 4 cores) with a processor Intel Core-i7 and two high-performance graphics cards NVIDIA GTX 780M. Its weight is 3.5 kg. It allows you quickly and efficiently autonomously anywhere to process speech information, flexible and easy to configure the processing and search information, accumulate and store the received information for later use.

• automatic processing of speech data • software and hardware tools for speech information processing • complex of software «Voice».

Речь человека сверхинформативна. Помимо смысла она содержит большое количество важной информации, позволяющей определять пол, возраст, уро-

вень образования и культуры персоны, язык сообщения, замысел, мотивы говорящего и т.д.

По речи больного опытный врач может поставить диагноз заболевания. Речь может программировать поведение человека – она может быть управляющим воздействием, мощнейшим стимулом к действию или средством нейтрализации, например, агрессивно настроенного человека. Все перечисленные параметры человек легко определяет на слух, что называется «вручную».

Обрабатывать большие потоки речевой информации на различных языках «вручную», даже при наличии большого количества операторов невозможно. Поэтому для обработки таких массивов речевых данных необходим специальный инструментарий, позволяющий осуществлять их автоматическую обработку и выявлять значимые для потребителя параметры речи. Этим обусловлена актуальность применения современных информационных технологий для автоматизированной обработки речевых данных.

В настоящее время на основе системной интеграции средств акустической регистрации и выборки больших объемов разнородной речевой информации из разных источников, в том числе, на различных языках, а также специализированного математического аппарата и высокопроизводительных вычислительных средств компанией ООО «Стэл КС» создан программно-аппаратный инструментарий для обработки речевой информации на базе серверного центра.

Такой способ реализации речевой технологии, несмотря на то, что он обладает большими возможностями, имеет ко всему еще и значительные массогабаритные характеристики, что не всегда удобно.

Для определенного класса задач существует острая потребность в использовании малогабаритных мобильных высокопроизводительных средств, позволяющих проводить обработку речевого материала с высокой оперативностью в режиме реального времени на удалении от стационарных объектов.

Разработка такого мобильного комплекса программных средств (КПС) также была проведена в компании ООО «Стэл КС».

Разработанный КПС «Голос» обеспечивает высокоскоростную обработку речевого материала. Высокая скорость обработки (в десятки раз быстрее времени звучания) достигается за счет использования многоядерной архитектуры процессоров и видеокарт NVIDIA, что позволяет обрабатывать несколько каналов поступления информации одновременно.

КПС «Голос» обеспечивает решение следующих задач:

1. Определение параметров речи.
2. Идентификация языка.
3. Распознавание дикторов.
4. Поиск ключевых слов.
5. Адаптация, конфигурирование и настройка под требования заказчика.

1. ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ РЕЧИ В ПОТОКЕ РАЗНОРОДНЫХ ДАННЫХ

В рамках реализации указанной задачи обеспечена возможность выделять речь в общем звуковом потоке и игнорировать его неречевую составляющую, определять пол говорящего и качество речевого сигнала, что позволяют ускорять решение различных розыскных задач.



Комплекс позволяет определить:

- длительность речи в сигнале;
- качество речевого сигнала;
- пол диктора;
- возрастную категорию (взрослый, ребенок);
- наличие в сигнале специфических сигналов (шумы, автоинформаторы, гудки, факсимильные и модемные сигналы и другие).

2. ИДЕНТИФИКАЦИИ ЯЗЫКА РЕЧЕВОГО СООБЩЕНИЯ

Система может определить, на каком из известных ей языков произнесено сообщение, или проинформировать, что язык ей неизвестен.

В базовой комплектации комплекс распознает следующие 19 языков: русский, английский, немецкий, французский, испанский, китайский, японский, турецкий, польский, литовский, арабский, корейский, таджикский, чеченский, фарси, дари, грузинский, казахский и узбекский.

При этом для идентификации языка требуется сообщение, содержащее не менее 10 секунд речи. Комплекс обеспечивает высокую точность идентификации языка – более 90%, что является гарантированным результатом распознавания.

Комплекс построен на принципах открытой архитектуры, что позволяет пользователю самостоятельно создать модель нового языка и пополнить или изменить список базовых языков. Для создания модели нового языка требуется выполнить процедуру обучения, для которой требуется корпус речевых данных, состоящий из записей сообщений различных дикторов, произнесенных на требуемом языке, общей длительностью не менее 10 часов.

3. РАСПОЗНАВАНИЕ ДИКТОРА

Для идентификации диктора необходимо провести процедуру обучения на его голос. Для выполнения процедуры обучения необходим образец речи длительностью не менее 30 секунд. Идентификация диктора проводится независимо от текста и языка речевого сообщения. Также система помогает подтвердить, что та или иная звучащая речь произнесена известным диктором (выполнить верификацию).

Система обеспечивает идентификацию и верификацию дикторов в сообщениях длительностью 10 секунд, если количество дикторов содержащихся в базе равно десяти, с точностью распознавания более 95%.

4. ПОИСК КЛЮЧЕВЫХ СЛОВ

Система позволяет следить за появлением оперативно важных ключевых слов и словосочетаний в потоке речевой информации, при этом пользователь имеет возможность задавать для поиска любые слова на 7 языках в текстовом виде. Такая возможность обеспечивается на русском, английском, испанском, французском, арабском, китайском и немецком языках. Каждое найденное ключевое слово выделяется в речевом сигнале метками его начала и конца. При этом система позволяет формировать произвольный словарь ключевых слов. Для русского языка к заданному оператором словарю ключевых слов предусмотрена автоматическая генерация словоформ. Комплекс позволяет обрабатывать более 1000 слов одновременно. Добавление

нового языка возможно, но требует квалифицированного труда лингвистов по подготовке специализированной речевой модели требуемого языка.

5. АДАПТАЦИЯ, КОНФИГУРИРОВАНИЕ И НАСТРОЙКА ПОД ТРЕБОВАНИЯ ЗАКАЗЧИКА

Комплекс обладает следующими возможностями:

Он может быть адаптирован оператором под конкретный канал передачи данных.

1. Комплекс может быть сконфигурирован оператором под конкретный оперативно значимый набор задач.
2. Комплекс позволяет настроить модели для более эффективного решения задач.

Адаптивность позволяет подстроить комплекс к исследуемому материалу для достижения максимальной эффективности. При этом система может быть настроена под конкретный канал передачи данных, а задача поиска ключевых слов (опционально) может решаться в два прохода, что повышает точность нахождения слов.

Свойство конфигурирования системы под конкретную задачу позволяет заказчику настраивать ее под свои требования. Например, комплекс можно настроить так, чтобы он не пропускал ни одного файла с речью целевого диктора. При этом можно допустить достаточно высокий процент ложных срабатываний. Однако задача будет выполнена.

В других условиях необходимо найти в массиве данных хотя бы несколько экземпляров голоса целевого диктора. Задача будет выполнена при минимальном сохранении значимых файлов (не более 1%) из всего материала.

Решение перечисленных задач обеспечивает программно-аппаратный комплекс, который включает:

- комплекс программных средств «Голос»;
- промышленный ноутбук (от 4-х вычислительных ядер) с процессором Intel Core-i7 и двумя высокоэффективными графическими ускорителями NVIDIA GTX 780M.

Комплекс весит 3,5 кг. Он позволяет быстро и эффективно обрабатывать речевую информацию автономно в любом месте, гибко настраивать процесс обработки и поиска информации, накапливать и сохранять полученную информацию для дальнейшего использования.

Оператор комплекса самостоятельно задает схему обработки. Информация, полученная на одном шаге обработки, может быть использована для фильтрации материала перед следующим шагом обработки.

Например, схема, состоящая из шага определения параметров речи и шага определения языка речевого сообщения, в файлах, содержащих речь мужчин длительностью более 10 секунд, позволит сократить входной поток данных и быстро подготовить для переводчика материал, который с большой долей вероятности будет содержать важную информацию. При этом сам переводчик не будет тратить время на отбор материала или на работу с неинформативным материалом.

Если в схему включить задачи определения параметров речи, определения языка, определения личности говорящего и поиска ключевых слов, то из всего потока материалов, можно найти ограниченную часть речевых данных, в



которой целевой диктор сообщает важную информацию, не используя при этом работу переводчиков.

Перечисленные возможности комплекса программных средств «Голос» могут успешно и продуктивно использоваться в таких приложениях ИКТ, как видеоконференцсвязь, дистанционное образование, телемедицина, и, в частности, таком, по мнению авторов, весьма перспективном приложении, как персональный стенограф и др.

В заключении хотелось бы отметить, что созданный комплекс является исключительно отечественной разработкой. Он родился на протяжении ряда лет, и обладает такими продвинутыми возможностями благодаря тесному и плодотворному сотрудничеству компании «Стэл – КС» с ведущими специалистами филологического факультета МГУ им. М.В. Ломоносова, Вычислительного центра РАН, Московского Государственного Лингвистического Университета и ряда других учреждений.

Результаты такого рода сотрудничества лишней раз демонстрируют его перспективный и плодотворный характер, особенно при разработке отечественных программно-аппаратных продуктов.

Считаем необходимым отметить наличие острой потребности в поддержке со стороны государства, особенно в стимулировании спроса на отечественные разработки. Об этом, в частности, шла речь на сентябрьской 2012 года конференции АДЭ в докладе председателя правления компании Spirit Андрея Владимировича Свириденко «Мировой софтверный рынок захвачен монополистами».

Определенный оптимизм здесь внушает недавняя (летом 2013 года) законодательная инициатива заместителя Председателя Государственной Думы, члена комитета Госдумы по информационной политике, информационным технологиям и связи Сергея Владимировича Железняк поддерживать на государственном уровне российскую электронную промышленность и программное обеспечение. Хотелось бы надеяться, что эта инициатива не останется голословной и найдет соответствующий отклик в законопроектной деятельности Минобрнауки и Минкомсвязи России.

Сведения об авторах

Устюгов Сергей Васильевич,

кандидат военных наук, заместитель директора ООО «Стэл КС».

Котов Михаил Андреевич,

руководитель департамента ООО «Стэл – Компьютерные Системы».

Леднов Дмитрий Анатольевич,

кандидат технических наук, старший научный сотрудник, научный консультант научно-технического департамента ООО «Стэл – Компьютерные Системы». Основные научные интересы лежат в областях: моделей обработки данных, случайных процессов, распознавания речи и идентификации дикторов.

Главатских Игорь Александрович,

начальник отдела речевых технологий ООО «Стэл – Компьютерные системы». Область научных интересов: распознавание и синтез речи, идентификация диктора. E-mail: ia_glavatskih@stel.ru