

УДК

Галунов В.И.(1,2), Кутуков Г.П.(1), Матюнин С.Н.(1)

Состояние исследований в области речевых технологий и задачи, выдвигаемые государственными заказчиками.

1. Секция по автоматическому распознаванию и синтезу речи РАН.

2. "ОДИТЕК"

Россия, 194044 С-Петербург, Б.Сампсониевский 37

Тел./факс: (812) 113-36-33; тел. (812) 427-86-13

e-mail: auditech@neva.spb.ru

Настоящий доклад преследует своей целью сформулировать на фоне современного состояния работ в области речевых технологий те задачи, которые ставит государственный Заказчик в России и стимулировать дискуссию о возможностях их решения.

Настоящий доклад преследует своей целью сформулировать на фоне современного состояния работ в области речевых технологий те задачи, которые ставит государственный Заказчик в России и стимулировать дискуссию о возможностях их решения.

Классификация речевых систем.

В настоящее время можно выделить четыре сравнительно изолированных направления в области развития речевых технологий.

Первое - распознавание речи, т.е. преобразование речевого акустического сигнала в цепочку символов, слов. Эти системы могут быть охарактеризованы по ряду параметров. Прежде всего это объем словаря: малые объемы до 20 слов, большие - тысячи и десятки тысяч. Количество дикторов: от одного до произвольного. Стиль произнесения: от изолированных команд до слитной речи и от чтения до спонтанной речи. Коэффициент ветвления, т.е. величина, определяющая количество гипотез на каждом шаге распознавания: от малых величин (<10-15) до больших (~100-200). Отношение сигнал/шум от больших (>30 дБ) до низких (<10 дБ). Качество каналов связи: от высококачественного микрофона до телефонного канала. Качество работы систем распознавания речи обычно характеризуется надежностью распознавания слов, или, что то же самое, процентом ошибок).

Второе направление - определение индивидуальности говорящего. Эти системы прежде всего делятся на 2 класса: верификация говорящего (т.е. подтверждение его личности) и идентификация говорящего (т.е. определение его личности из заранее

ограниченного числа людей). Оба эти класса далее могут быть разделены на текстозависимые и текстонезависимые. Следующий характеристический параметр - объем парольной фразы. Два других (как и в распознавании речи): отношение сигнал/шум и качество канала связи.

Качество работы систем верификации/идентификации говорящего характеризуется 2-мя величинами: вероятностью не опознания "своего" диктора и вероятностью принятия "чужого" диктора за своего.

Третье направление - синтез речи. Практически существует два класса:

1. воспроизведение записанной в той или иной форме ограниченного числа сообщений и

2. синтез речи по тексту.

Характеризуются синтезаторы по следующим параметрам: разборчивость (словесная или слоговая) естественность звучания, помехоустойчивость.

Четвертое направление - компрессия речи. Основной (и единственный) классификационный признак этих систем - степень компрессии: от низкой (32-16 кбит/сек) до высокой (1200-2400 кбит/сек и ниже).

Качество работы систем компрессии речи характеризуются прежде всего разборчивостью компрессированной речи. Дополнительными характеристиками очень важными в ряде приложений являются узнаваемость голоса говорящего и возможность определения уровня стрессованности говорящего.

Современный уровень достижений в развитии речевых технологий.

Автоматическое распознавание речи. В последние годы наблюдается активное развитие этого направления. Основной показатель (процент ошибок в распознавании слов) уменьшается приблизительно в два раза каждые два года. Для распознавания цифр, произносимых слитно произвольными дикторами по телефону, достигнут уровень ~0,3% ошибок. Для диалоговых систем (несколько тысяч слов) для произвольного диктора достигнут уровень ошибки в несколько процентов.

Системы определения индивидуальности говорящего. Здесь наблюдается столь же быстрый процесс, что и для систем распознавания речи. Вероятность суммарной ошибки при верификации говорящего в телефонном канале и по короткой парольной фразе составляет ~1%.

Системы синтеза речи. Уровень синтеза речи, сравнимого с естественной речью, достигнут уже несколько лет назад.

Системы компрессии. В настоящее время можно считать достигнутым уровень компрессии 1200 бит/сек. при удовлетворительной разборчивости и узнаваемости говорящего (хотя следует отметить, что для определения узнаваемости нет стандартных методик).

Состояние разработок речевых технологий в России.

Исследования в области речевых технологий в России (Советском Союзе) имеют длительную историю. Работы по компрессии речи начались в начале 50-х годов, а по автоматическому распознаванию - в конце 50-х. (При этом следует отметить, что первая в мире система автоматического распознавания речи была продемонстрирована в 1939 году в Ленинградском Государственном Университете Л.Л.Мясниковым). В 60-х годах наблюдалось определенное опережение Россией (Советским Союзом) других стран в теоретических и практических разработках в области речевых технологий. Определилось это двумя факторами:

- наличием комплекса прикладных задач, связанных с обороной страны;
- отчетливым осознанием того, что прикладные задачи не могут быть решены без интенсивного проведения фундаментальных исследований, которые и поддерживались в явном и неявном виде.

В 70-х годах в разработке речевых систем начали активно выходить вперед США. Военное ведомство США организовало ряд конкурсных проектов, предоставив для них финансирование и возможность использования вычислительных мощностей военного ведомства. Количество исследователей, работающих в это время в США в области речевых технологий превышало количество речевиков в Советском Союзе минимум в 10 раз. Тем не менее уровень теоретических и прикладных разработок в Советском Союзе и США до середины 80-х годов оставался приблизительно одинаковым.

С середины 80-х годов началось сначала постепенное, а потом обвальное разрушение системы речевых разработок в СССР (России). Причиной этому послужило абсолютное прекращение финансирования этого направления как в прикладном, так и в академических аспектах.

В настоящее время помимо США в области речевых технологий активно и очень успешно работает еще ряд стран (ЕС, Япония, Канада, Австралия).

Возможные перспективы.

Для России представляется недопустимым уход из области речевых технологий. Это невыгодно прежде всего из экономических соображений. Европейский рынок речевых технологий только в телекоммуникационных системах оценивается в несколько миллиардов долларов. В отсутствии собственных разработок российская часть рынка будет полностью отдана зарубежным компаниям. Второе - речевые технологии это значительная часть безопасности страны. Третье - речевые технологии это показатель уровня развития информационных наук в стране, что является индикатором потенциала будущего развития, интеллектуальным стратегическим резервом страны.

Рассматривая возможности развития речевых технологий в России в интересах государственных Заказчиков можно выделить следующие приоритетные направления:

- разработка технологии распознавания слитной речи, включая в себя готовые акустико-фонетические модели фонем речи на иностранных и языках народов СССР, статистическая модель языков;
- создание комплекса программ восстановления искаженных и зашумленных речевых сообщений как на ограниченном (тематическом) словаре, так и смешанном;
- разработка на современных технологиях поиска в информационных сетях речевой информации по заданным ключевым словам или проблематике;

- создание интерпретаторов, верно передающих смысл сильно искаженного речевого сообщения;
- автоматизированное обнаружение в текстах и речевых сообщениях лингвистической информации, значимой для психофизиологической оценки и биометрического контроля.