ритмы распознавания речи являются дикторозависимыми. После настройки на голос одного диктора распознающие системы дают удовлетворительные результаты распознавания для этого типа голоса, но хуже работают на других голосах. Надежность распознавания речи человеком, напротив, не зависит от типа голоса диктора.

Все вышесказанное приводит к тому, что распознавание речи компьютером обладает ограниченной надежностью, существенно повысить которую вероятно не удастся в будущем ни путем совершенствования алгоритмов распознавания, ни путем увеличения вычислительных мощностей компьютера. Постоянно имея в виду это утверждение, можно приступать к анализу достижений в области распознавания речи, классификации стоящих в этой области задач и оценке перспектив их решения.

2. Современное состояние направления распознаванияе речи

Классификацию систем распознавания речи будем производить согласно новому стандарту в области программирования таких систем, принятому сейчас практически всеми известными разработчиками систем распознавания речи - Microsoft Speech API ([1]).

Согласно этому стандарту, системы распознавания речи различают по следующим признакам:

Интервал между отдельными словами. Если система распознает непрерывную речь, пользователь может произносить речевые фразы естественно, не делая паузы между словами. Непрерывное распознавание более предпочтительно, однако оно требует большей вычислительной мощности компьютеров, что приводит пока к малому числу таких систем. В системах, работающих с дискретной речью, пользователь при диктовке должен делать паузу между отдельными словами, обычно составляющую не менее 1/4 частиь секунды. Третьей разновидностью являются системы, выделяющие одно слово из интервала речи, даже если он состоит из нескольких непрерывно произнесенных слов (word-spotting, [1]).

Зависимость от диктора. Системы, обладающие относительной

независимостью от диктора, позволяют пользователю работать с системой без предварительной настройки, однако улучшают надежность распознавания после обучения. Независимость от диктора таких систем обычно достигается за счет хранения звуковых эталонов для всех наиболее типичных голосов носителей данного языка. Это, безусловно, требует в несколько раз большей производительности и объема памяти. Настройка на голос диктора дикторозависимых систем занимает обычно от 30 минут до нескольких часов. Это составляет главное неудобство для пользователя. Обычно дикторозависимые системы позволяют работать с относительной степенью надежности без предварительной настройки на голос конкретного пользователя. Третьей разновидностью систем по этому признаку являются системы, автоматически настраивающиеся на голос диктора по мере их использования. Системы последнего типа обладают двумя особенностями - им нужно знать, сделал ли пользователь ошибку, произнеся конкретное слово (иначе обучение будет неверным); после настройки на одного диктора такие системы перестают надежно работать с другими голосами.

Степень детализации при задании эталонов. Различают алгоритмы, в которых в качестве эталонов используются целые слова, и алгоритмы, использующие эталоны элементов слов. Сравнение целых слов дает большую точность, скорость, однако требует значительно большего объема памяти (пропорционально количеству слов в словаре) и обучения каждого слова. Алгоритмы сравнения элементов слов (фонем, слогов и т.п.) приходится применять в случае больших словарей, т.к. объем требуемой памяти пропорционален количеству этих эталонных элементов слов (например, звуков) и не зависит от объема словаря.

Размер словаря. Системы распознавания речи могут использовать большие или маленькие словари. Размер словаря системы распознавания почти не связан с реальным количеством слов, которые данная система может распознать. Он определяется количеством слов, требуемых для распознавания в данном конкретном состоянии системы. Системы, работающие с маленькими словарями (около 50 слов) позволяет пользователю давать простые команды компьютеру. Для диктовки текстов необходимы большие словари (несколько десятков тысяч слов). Если системы диктовки учитывают контекст для определения активного подсловаря в конкретном состоянии, то фактически они работают со словарями среднего размера (около 1000 слов [1]).

Несмотря на то, что в принципе возможна любая комбинация этих характеристик, в настоящее время наиболее популярными являются системы голосового управления компьютером и системы дискретной диктовки текстов.

Системы голосового управления компьютером ("Command and Control" в терминологии Microsoft Speech API [1]) - это системы дикторонезависимого распознавания непрерывно произносимых команд, составленных из слов ограниченного (до нескольких сотен слов) словаря. Для подобных систем, если пользователь произносит команду, не входящую в список, система либо выдает отказ от распознавания, либо выдает в качестве ответа похожую "на слух" команду. Список команд обычно интуитивно ясен в каждой конкретной ситуации. Согласно стандарту Microsoft Speech API , системы голосового управления должны работать успешно на компьютерах 486/66 МГц с 1 МБ свободной оперативной памяти.

Системы дискретной диктовки текстов ("Discrete Dictation" [1]), т.е. системы дикторозависимого распознавания дискретно произносимых слов из больших по объему (десятки тысяч слов) словарей. Подобные системы обычно требуют процессора Pentium/60 МГц и 8 МБ свободной оперативной памяти.

В сводной таблице 1 приведены характеристики наиболее известных сейчас систем распознавания речи ([5]).

Если оценивать существующие сейчас на рынке системы диктовки и голосового управления компьютером с точки зрения рядового пользователя, подходящего к компьютерным программам с позиций эффективности и удобства, то можно сделать следующие выводы.

Система голосового управления компьютером менее удобна, привычна и проста в обучении, чем клавиатура и мышь. Исключение могут составить лишь пользователи-инвалиды. Применение голоса для управления компьютером станет частью интерфейса лишь тогда, когда появятся принципиально новые мультимедиа-ориентированные операционные системы, изменится архитектура вычислительных машин (заметный шаг в этом направлении - появление процессоров MMX) и будут разработаны надежные системы очистки от стационарных и нестационарных помех.

Системы диктовки текстов являются пока привлекательными для покупателей в силу новизны предоставляющихся для пользователя возмож-