

Литература. 1. В.А. Головкин Нейрокомпьютеры и их применение: книга 4 /; Под редакцией А.И. Галушкина «Нейронные сети: обучение, организация и применение», М.: ИПРЖР 2001. 2. Warren S. Sarle, SAS Institute Inc., Cary, NC, USA Июль 17, 1999. Описание деталей обучения, модифицированных 1 Мая, 2001. Авторское право 1999, 2001 by Warren S. Sarle, Cary, NC, USA. URL: <ftp://ftp.sas.com/pub/neural/dojo/dojo.html>. 3. А.Б.Сергеев «Цифровая обработка сигналов»; СПб.: Питер,2003.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНТЕРФЕЙСОВ ОБМЕНА ЧЕЛОВЕК-КОМПЬЮТЕР

Семченко Д.Е., ПГУ, г. Новополюцк

Для обмена информацией между компьютером и человеком существуют визуальные, акустические и осязательные каналы передачи. Текстовые данные могут быть введены с помощью клавиатуры, произнесены в микрофон или быть оцифрованы из рукописной формы с помощью камеры, сканера или графического планшета. Для представления текстовой, равно как и графической, информации стандартными являются визуальные средства, однако возможности акустического обмена на сегодняшний день используются не в полной мере. Можно сказать, что практически решена задача передачи данных от компьютера человеку: визуальных через монитор, акустических через динамик. Обратная же задача является гораздо более сложной.

Любая оценка преимуществ и недостатков вышеназванных каналов передачи во многом зависит от конкретной предметной области, где они используются. Однако, можно сформировать некоторые универсальные критерии оценки – например, достижимые *надежность* и *скорость* передачи, которые связаны с затратами на используемые технические средства, либо степенью подготовленности пользователя к сеансам обмена и степенью загруженности пользователя процессом обмена.

Распознавание речи – это трансформация представленного речевого высказывания из формы изменяющегося с течением времени сигнала в правильную письменную форму. Можно выделить три основных фактора, оказывающих существенное влияние на используемые технологии распознавания речи – форма представления, объем словарного запаса и степень зависимости от диктора.

Под обработкой произносимой речи понимается манипуляции с цифровым представлением сигнальных форм речевого высказывания, в особенности их

анализ и синтез. В центре интересов стоит, прежде всего, внутреннее содержание сообщения. Проблема анализа состоит в том, чтобы сопоставить представленное речевое высказывание и его адекватное внутримашинное представление. В зависимости от используемого приложения необходимо либо корректно реконструировать текстовое представление произнесенного (автоматическое распознавание речи); либо создать подходящее внутримашинное описание намерений пользователя (автоматическое понимание речи); либо в зависимости от речевого высказывания произвести какие либо действия (командные и диалоговые системы).

Автоматическое распознавание речи является, как и прежде, не полностью решенной проблемой, т.е. результат процесса распознавания в определенной степени обременен ошибками. Частота появления ошибок растет с расширением словарного запаса системы и падает с растущей ригидностью (скованностью) грамматической речевой модели. При работе в многопользовательском режиме она также выше, чем для случаев, когда система рассчитана на одного диктора.

Человек осознает понимание произнесенной фразы как совершенно обычный и само собой разумеющийся навык. По нашим субъективным впечатлениям речевое высказывание состоит из последовательности слов, каждое из которых в свою очередь состоит из последовательности звуков; все же вместе представляется как временно ориентированный поток дискретных единиц, сцепленных между собой подобно отдельным звеньям одной цепи. Каждый звуковой сегмент при этом представляется снабженным инвариантным звуковым свойством, совершенно независимым от данного (текущего) контекста. Однако это впечатление обманчиво; оно следует исключительно из высокосложной нейрофизиологической внутренней обработки речевого сигнала.

Сложности, которые следует преодолеть при машинном распознавании речи, можно разделить на четыре группы. Сюда входят непрерывность, изменчивость, сложность и неопределенность произносимой речи.

Непрерывность. В противоположность вышеописанному впечатлению в речевом сигнале в общем случае не существует никакой видимой дискретности, которая однозначно обозначала бы границы между словами, слогами и звуками. В особенных случаях даже точное обнаружение границы слов гаран-

тировано только тогда, когда определенные границы фразы помечаются диктором интонационно с помощью коротких интервалов «молчания».

Изменчивость. Одни и те же речевые единицы могут быть реализованы акустически многочисленными способами. При этом необыкновенно сложно найти подходящие акустические прототипы для слов и звуков, которые бы (прототипы) представляли различия корректно, несмотря даже на изменчивые сопутствующие факторы. Такой прототип должен выбираться так, чтобы, с одной стороны, охватывать все возможные вариации, с другой стороны, четко разграничивать конкурирующие единицы (прототипы).

Сложность: Автоматическое распознавание речи требует высокой производительности компьютера и больших объемов памяти. Причиной этому являются: скорость передачи данных при вводе речевого сигнала (обычно от 8000 до 20000 квантованных значений сигнала в секунду), обширный инвентарь единиц распознавания и огромная комбинаторика при построении предложения (из X различных слов можно построить Y^X различных последовательностей слов длиной Y).

Неопределенность: Между единицами речи и их реализацией, в общем случае, не имеется взаимнооднозначного соответствия. Например, различные слова или предложения могут быть, при определенных обстоятельствах, произнесены очень похожим образом или вообще одинаково. Такая многозначность встречается на речевых уровнях: как гомофоны («род-рот»), на границе слов, в синтаксической и семантической области.

Несмотря на описанные сложности с внедрением речевого взаимодействия, по совокупности показателей интерфейс взаимодействия человека и компьютера «голос-видео» представляется мне более удобным, чем используемый в настоящее время интерфейс «клавиатура-мышь-видео». Будущее речевого интерфейса зависит от умения разработчиков не только создать технологическую основу речевого ввода, но и от умения предложить варианты практического применения, где интерфейс «голос-видео» будет иметь явные преимущества перед стандартным сегодня интерфейсом «клавиатура-мышь-видео».

Литература: 1. Schukat-Talamazzini E.G. Automatische Spracherkennung.- Vieweg Verlag, 1995. 2. Hunt M. Speech Technology.- 1992. 3. Sagerer G. Automatisches Verstehen gesprochener Sprache.- Bd. 74 von Reihe Informatik, Bibliographisches Institut, Mannheim, 1990.