

Н.И. ЦУПРЕВ

Государственное учреждение высшего профессионального образования
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Могилев, Беларусь

Проблема анализа и распознавания речи как часть общей задачи ввода информации в ЭВМ на естественном языке, является одной из приоритетных проблем, связанных с созданием систем с искусственным интеллектом. Достаточно отметить, что задача создания систем данного класса на ближайшее десятилетие является одной из основных для таких разработчиков и производителей вычислительных средств и информационных технологий как IBM и Microsoft.

Основными направлениями исследований в области речевого общения являются:

- распознавание речевых образов и их интерпретация;
- идентификация и верификация диктора;
- синтез речевых сообщений.

В настоящее время наиболее качественное распознавание речи обеспечивают устройства, работающие с ограниченным словарем, ориентированные на одного диктора.

Вместе с тем, активно проводятся работы по распознаванию слитной речи независимо от диктора. Осуществляются также попытки не только поэлементного распознавания речи, но и анализа смысла речевой реализации. При этом в различных исследованиях в качестве объекта анализа принимается фонема, слог, слово.

Исследования в области синтеза и распознавания речи ведутся в 12 университетах Западной Европы и на них ассигновано более 170 млн. \$ США.

Основными производителями систем распознавания речи являются фирмы IBM, Centigram Corp., Interstate Electronics, Threshold Technology, Verbex Corp., Votan Inc.(США), NEC (Япония), Marconi, Secure Radio Systems (Великобритания).

Распознавание речи является одним из видов распознавания образов, основанным в настоящее время на сравнении входного сигнала с эталонами, заранее произнесенными диктором и хранящимися в запоминающем устройстве.

В большинстве систем речевого общения анализ речевых сигналов с целью их описания осуществляется с помощью оценки фактических характеристик сегментов текущей реализации, расчет которых производится либо с помощью интегральных преобразований (Фурье, Лапласа, Гилберта), либо итерационными методами, либо на основе корреляционных функций.

При этом используются как детерминированные, так и вероятностные характеристики анализируемых сигналов.

Фактором, ограничивающим возможности устройств распознавания речи, является то обстоятельство, что произношение разных людей имеют индивидуальные фонологические характеристики, и смысл произносимого во многом зависит от контекста. Так, диктору трудно все время одинаково произносить даже отдельные слова и в первую очередь соблюдать их длительность. Поэтому, при распознавании предусматривается временная нормализация, т.е. приведение произносимых слов к стандартной длине. При линейной нормализации все части произносимого слова сжимаются или растягиваются во времени одинаково. Нелинейная нормализация предусматривает разное изменение длительности частей слова, как это бывает в естественной человеческой речи. Динамическое программирование – математический метод оптимизации изменения длительности разных частей слова, называемый методом динамической временной деформации.

Однако ни один из методов анализа речевых сигналов не предлагает способа однозначного и конечного кодового описания таких элементов речи как, например, фонема. Особенно неопределенным и зависимым от диктора является интервал временной нормализации.

В результате проведенных исследований разработан способ анализа и метод описания (способ оптимального кодирования) сегментов речевого сигнала путем сравнения его характеристик до и после некоторого преобразования (метод анализа путем оценки реакции характеристик сигнала на воздействие).

В предлагаемом способе оценки распознавание элементов речевых реализаций осуществляется по относительным потерям энергии после параметрического воздействия на переменном интервале анализа.