

Гаюха А.А. — рецензент Киселев Г.Д.

УНК “Институт прикладного системного анализа” НТУУ “КПИ”, Киев, Украина

Способы и особенности организации систем распознавания речи

Распознавание речи – актуальная задача, расширяющая горизонты робототехники. Создание множества систем, принимающих аудиопоток человеческой речи в качестве входных данных и проводящих анализ этого потока, обусловлено стремлением человека научиться общаться с машиной.

Основная задача системы распознавания речи – представление аудиопотока как набора символов. Сами системы можно разделить на два вида:

- Системы, зависимые от диктора;
- Системы, не зависимые от диктора.

Системы первого класса обучаются и настраиваются под речь определенного диктора и требуют полной перестройки при его смене, вторые работают с произвольной речью.

Помимо привязки к диктору, системы можно разделить на два типа, концепции работы которых кардинально отличаются между собой:

- Распознавание голосовых меток;
- Распознавание лексических элементов.

Первый подход позволяет анализировать речь по готовым образцам, заранее хранящимся в базе данных. Второй подход анализирует поток речи, выделяя из него отдельные лексические элементы – фонемы и аллофоны, которые затем объединяются в слоги и морфемы.

Реализация систем распознавания речи подразумевает использование искусственных нейронных сетей, но не ограничивается ими. В качестве модели нейронной сети обычно выбирают сеть Кохонена. Это связано с тем, что она наиболее приспособлена к изменчивости и неоднородности речи. В ней для группы входных сигналов происходит формирование нейронных ансамблей, представляющих данные сигналы. Дополнительным преимуществом использования нейронных сетей является гибкость изменения алгоритма распознавания за счет изменения структуры сети. Задача выбора оптимальной архитектуры нейронной сети решается при помощи использования генетических алгоритмов. При их использовании создается набор правил, который позволяет определять степень выполнения задачи данной нейронной сетью, а также правила модификации сети. В результате формируется нейронная сеть, которая позволяет оптимально решить задачу распознавания голоса для текущих звуковых параметров и тембра голоса.

Для выделения информативных признаков речевого сигнала используется спектральное представление речи, которое можно разделить на два этапа. Первый этап осуществляет получение частотного спектра речевого сигнала, выполняя дискретное преобразование Фурье. Второй этап обрабатывает полученный спектр сигнала, улучшая и очищая его.

Динамические особенности речи и ее изменчивость также учитываются при анализе речевого потока. Используемые для этого параметры представляют собой производные по времени от основных параметров речи, таких как тембр голоса, скорость речи, изменение интонации.

Системы распознавания речи активно внедряются в современную технику и имеют коммерческий интерес. Например: голосовой вызов в телефоне, манипуляции бытовой техникой, автоответчики, способные принимать заказы и заполнять анкеты. Недостатком данного вида услуг является то, что сами системы распознавания голоса еще достаточно ненадежны и нуждаются в дополнительных междисциплинарных исследованиях.

Литература. 1. Фролов А.В., Фролов Г.В. Синтез и распознавание речи. Современные решения., 2003. 2. Москаленко А.М. Использование нейросетей для автоматического распознавания и синтеза речи, 2000. 3. Speech Analysis FAQ: <http://svr-www.eng.cam.ac.uk/~ajr/SA95/SpeechAnalysis.html>. 4. Analysis of Kohonen’s Neural Network with application to speech recognition: <http://www.micai.org/2009/proceedings/complementary/cd/ws-cvpr/138/Analysis%20of%20Kohonen-s%20Neural%20Network.pdf>.