

УДК 621.391:519.72

Моделирование процесса создания эталонов голосовых команд для систем автоматического распознавания речи

Е.К. Левин, К.Е. Левин

Постановка задачи

При автоматическом распознавании ГК происходит сопоставление временной последовательности наборов (векторов) параметров речевого сигнала (РС) с эталонами распознаваемых команд. Эталон, с наибольшей вероятностью соответствующий произнесению ГК, является результатом распознавания.

Распространенным подходом к построению систем автоматического распознавания речи (АРР) является использование эталонов в виде МСМП [1]. В литературе часто упоминается метод построения МСМП, основанный на итерационной процедуре Баума - Уэлча с применением алгоритма Витерби [1,3]. Результат работы процедуры (конечные значения параметров МСМП) во многом определяется начальными значениями параметров модели, то есть «стартовыми» условиями ее создания. Чем ближе «стартовые» параметры к оптимальным, тем точнее формируется эталон и меньше вероятность возникновения ошибки распознавания.

Однако выбор начальных значений параметров МСМП не формализован и подразумевает анализ нескольких вариантов типичного произнесения ГК. Оценка качества каждого варианта требует проведения процедуры контрольного распознавания множества звуковых файлов с последующим анализом результатов распознавания, что является достаточно трудоемким. Кроме того, при анализе требуется дополнительная информация о процессе распознавания, которая не представляется системой АРР при ее эксплуатации.

МСМП характеризуется количеством состояний, матрицей переходных вероятностей, а также законами распределения плотностей вероятностей наблюдаемых векторов параметров РС для каждого состояния. Для получения высокой точности оценок переходных вероятностей и параметров законов распределения требуется большое количество звуковых файлов, используемых для формирования эталона ГК.

Для каждого файла необходимо провести оценку качества его предварительной обработки, в процессе которой удаляются паузы с помехами перед и после произнесения ГК. Выделение информативной части ГК с целью последующего определения параметров РС может сопровождаться ошибками: либо она усекается, либо в ее состав включается часть паузы. Целесообразно провести просмотр «осциллограмм» выделенных информативных участков ГК, что невозможно в рамках эксплуатируемой системы АРР.

Рассматриваются вопросы создания эталонов голосовых команд (ГК) на основе моделей скрытых марковских процессов (МСМП). Приведены данные о разработанной программной модели процесса создания эталонов. Модель позволяет провести анализ причин возникновения ошибок распознавания ГК и снизить трудоемкость создания эталонов.

Создание более точных эталонов ГК требует моделирования всех этапов синтеза и в первую очередь процессов предварительной обработки речевого сигнала и определения собственно параметров МСМП по отдельности.

Моделирование процесса предварительной обработки речевого сигнала

В результате решения первой задачи была разработана программа, которая выводит на экран монитора одновременно «осциллограмму» звукового файла и «траектории» двух выбранных из списка параметров РС. На «траекториях» различным цветом помечаются вокализованные и невокализованные сегменты РС, а также группы сегментов, относящихся к информативной части ГК и к паузам, кроме того, указываются пороговые значения амплитуды РС, которые используются при выделении информативной части ГК. Помеченные участки изображений могут быть просмотрены с «увеличением» в соответствии с выбранным масштабом.

В список параметров РС включены: средняя амплитуда, 10 коэффициентов

$$g_i = (1 - K_i) / (1 + K_i)$$

(где $i = 1 \dots 10$, K_i — i -ый коэффициент частной корреляции) [2], производные по времени этих параметров и частота основного тона. Производные для параметров g_i могут суммироваться в произвольных сочетаниях, образуя новые параметры.

Программа позволяет последовательно анализировать звуковые файлы из заданной директории, автоматически загружая их друг за другом. Кроме того, возможна звукозапись новых файлов с автоматическим присвоением им имен, в которых содержится информация об условиях записи (коды команды, диктора и аппаратуры записи). Такая кодировка обеспечивает автоматическую сортировку файлов.

Каждый файл проверяется на правильность выделения информативной части ГК. Имена файлов с типичным произнесением ГК фиксируются. Затем происходит преобразование полученного множества файлов во множество файлов, содержащих только

выделенную информативную часть ГК. Последнее множество используется непосредственно при формировании МСМП.

Моделирование процесса определения параметров эталона ГК

Структурная схема процесса обработки множества файлов информативных частей ГК в моделирующей программе приведена на рис. 1

Сначала определяются параметры РС из заданного перечня, соответствующие информационным частям ГК. При необходимости создаются текстовые файлы, содержащие векторы параметров. Затем по известному файлу типичного произнесения ГК из соответствующей последовательности векторов параметров РС создается МСМП с начальными значениями числовых характеристик. Далее запускается итерационная процедура Баума — Уэлча и создается итоговая МСМП с конечными значениями числовых характеристик.

После создания всех МСМП для заданного перечня ГК проводится их проверка — проходит этап тестирующего распознавания ГК. Для этого используется еще одно тестирующее множество файлов информативных частей ГК. Результаты распознавания отражаются в окне статистики. Неверно и неуверенно распознанные файлы помечаются и в дальнейшем используются для анализа причин ошибки распознавания.

При анализе для каждого файла определяется (по алгоритму Витерби) наиболее вероятная последовательность состояний каждой МСМП из перечня всех эталонов распознаваемых ГК. Вычисляются также вероятности этих последовательностей.

С целью последующего выбора оптимальной МСМП из полученного их набора каждый вариант эталона ГК хранится в текстовом файле, который может редактироваться стандартными средствами операционной системы Windows 95/98.

На рис.2 представлено изображение окна программы (программа разработана студентом Владимирского государственного университета Д.Горячевым) на этапе тестирующего распознавания.

Справа на рисунке изображено окно статистики и отражены результаты распознавания файла ГК «ноль». Слева указаны названия распознаваемых ГК, а также изображено окно анализа ошибок распознавания (в данной ситуации ошибочно распознанный файл отсутствует в окне анализа).

Верхнее меню позволяет перейти к этапу формирования эталона ГК, загрузить файлы эталонов ГК, а также загрузить тестирующие множества файлов информативных частей ГК или векторов параметров РС.

При разработке программы были введены следующие ограничения на параметры эталона ГК, которые упрощают его создание. Во-первых, корреляционная матрица параметров РС считается диагональной. Во-вторых, распределение плотности

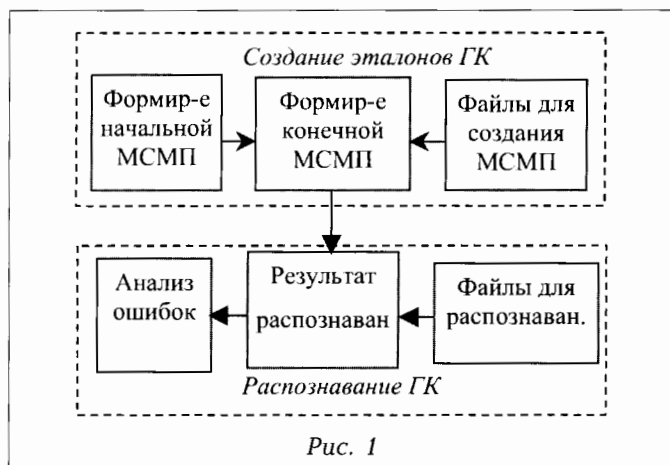


Рис. 1

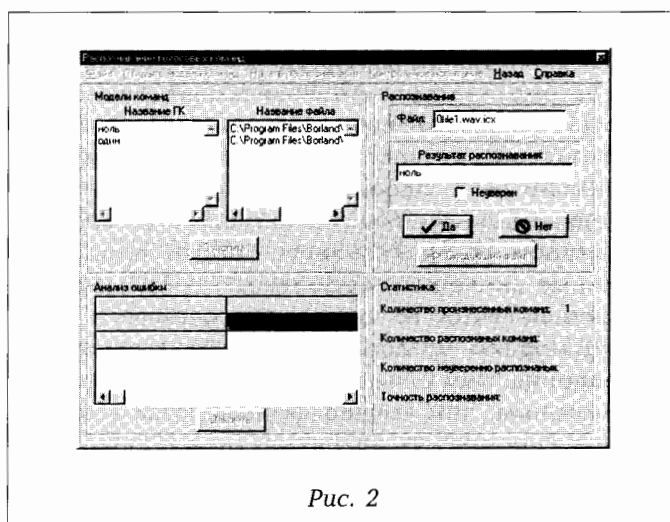


Рис. 2

вероятностей наблюдаемых значений параметров РС для какого либо состояния МСМП считается гауссовым.

Тестирование программы

Так как программа, моделирующая процесс создания МСМП, сложная, то для проверки правильности ее функционирования была разработана специальная тестирующая процедура. Суть ее заключается в следующем. Задается некоторая МСМП, и на ее основе генерируется множество файлов векторов параметров РС. Далее на основе этого множества создается МСМП, которая сравнивается с заданной.

Литература

1. Рабинер Л.Р. Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: Обзор. / ТИИЭР. - М., 1989. - т.77, № 2. - С.86 - 120.
2. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. / Под ред. М.В.Назарова и Ю.Н.Прохорова. — М.: Радио и связь, 1981. - 496с.
3. Елинек Ф. Распознавание непрерывной речи статистическими методами. / ТИИЭР. — М., 1976. т.64, № 4. - С.131 - 160.