УДК 681.142.66

Д.В. Божко, В.А. Грабовая, В.Ю. Шелепов
Институт проблем искусственного интеллекта, г. Донецк

Интерпретатор распознанной цепочки фонем, которая может содержать ошибки

Статья посвящена описанию модуля пофонемного распознавателя, который по полученной цепочке фонем (возможно, содержащей отдельные ошибки) находит наиболее подходящее слово из распознаваемого словаря.

В отделе фундаментальных проблем распознавания речевых образов Института проблем искусственного интеллекта предложена структура пофонемного распознавателя, включающая 3 независимых компоненты:

1. Модуль сегментации записанного речевого сигнала, т.е. автоматического разбиения его на участки, отвечающие отдельным аллофонам. (Отметим, что сама запись представляет собой непростую процедуру, включающую правильное определение начала, а возможно, и конца слова.)
2. Модуль распознавания фонем, соответствующих выделенным участкам.
3. Модуль, который по полученной цепочке фонем (возможно, содержащей отдельные ошибки) находит наиболее подходящее слово из распознаваемого словаря. Назовем его интерпретатором цепочки фонем.

Сегментация речевого сигнала – процедура аналогичная выделению контуров при распознавании зрительных образов. В этом отношении нами предложен ряд алгоритмов: обработка фильтрами, меняющая энергию отдельных участков с последующей классификацией этих участков по амплитуде (наиболее быстрый метод), [1]; выделение переходных участков между фонемами на основе признаков, использующих коэффициенты линейного предсказания; разбиение сигнала на квазипериоды (см. статью настоящего выпуска: Федоров Е.Е., Шелепов В.Ю. «Защита речевых распознавателей от шума и посторонней речи») и вычисление энергии разности последовательных квазипериодов (на переходах между фонемами она максимальна).

В распознавании фонем мы также опробовали различные алгоритмы и системы признаков. Существующие здесь объективные трудности можно представить себе на основании следующей оценки:

В русском алфавите 33 буквы, которые отвечают звукам речи. Качество каждого звука определяется тройкой «предшествующий звук, исследуемый звук, следующий звук». Таким образом, число распознаваемых классов оценивается снизу числом размещений из 33 по 3, которое равно

$33 \cdot 32 \cdot 31$ и, следовательно, больше, чем $30^3 = 27000$.

Нужны алгоритмы, позволяющие объединять различные звуки в большие классы, как это делается при письме. В настоящее время мы считаем наиболее перспективным использование нейросетей, работающих с интегральными спектрами распознаваемых звуков. При этом распознавание ведется по дереву, разработанному нами на основе фонетических сведений. Оно включает последовательное разбиение звуков на шипящие, голосовые и содержащие паузу. Затем из голосовых выделяются звуки «ж» и «з», содержащие шумовую компоненту. Остальные набором соответствующих сетей разбиваются на гласные, согласные и т.д. Для увеличения надежности используется по несколько экземпляров аналогичных, независимо обученных сетей и результат их совместной работы определяется голосованием (схема независимых испытаний Бернулли). В результате, разработана программа, которая позволяет быстро обучать необходимые сети и достаточно надежно распознавать фонемы, не различая при этом ударные и безударные гласные.

Третий модуль – определение слова по цепочке фонем. Остановимся на его описании более подробно. В отделе разработана своя система автоматической транскрипции, которая ориентирована на возможности распознавания, имеющиеся у нас на сегодняшний день [2]. Она, например, отождествляет «б, г, д» (общий транскрипционный знак «д») и «п, к, т» (общий транскрипционный знак «п»). В то же время она легко модифицируется, так что можно отказаться от подобных отождествлений, когда возникнет такая возможность. Например, «б, г, д» можно различать между собой, обучив соответствующие сети на участках перехода от этих согласных к последующим гласным. Далее создается кодовая книга, основанная на некоторой системе признаков [3]. Она содержит кодовые вектора, отвечающие в звучащей речи транскрипционным знакам. На этом этапе нами применялась традиционная система признаков, которая использует относительные частоты длин полных колебаний [4], которая хорошо зарекомендовала себя в разработанных в отделе распознавателях целых слов. Эта кодовая книга создается один раз одним человеком и в последующем используется всеми другими дикторами. Она, конечно, может совершенствоваться, но является дикторонезависимым элементом распознавателя.

Начиная работу с программой, Вы вводите текстовый файл, который содержащий слова, подлежащие распознаванию. При запуске программы машина сама создает файл соответствующих транскрипций. Далее она заменяет каждый транскрипционный знак соответствующим кодовым вектором и автоматически создает эталон каждого слова [4]. После того как в процессе распознавания получена цепочка фонем (точнее, цепочка транскрипционных знаков), машина с помощью кодовой книги создает аналогичное эталону представление полученного «квазислова», и на основании алгоритма DTW [4] сравнивает его с эталонами всех слов распознаваемого словаря. Ближайшее слово объявляется результатом распознавания.

В настоящее время в отделе есть первые версии пофонемного распознавателя, работающие по этой схеме. Далее предстоит работа по их совершенствованию. Однако нам кажется интересным, что наличие интерпретатора цепочки фонем уже сейчас позволяет практически безошибочно различать любую пару не слишком похожих слов. Отметим, что наличие такого

модуля приближает систему распознавания к системе человека, который, как известно, может верно, распознав не более 30% услышанных фонем, восстановить услышанное слово. Правда, при этом ему помогает не только сравнение цепочки фонем с известными ему словами, но также контекстная интерпретация услышанного (ориентация на смысл!).

Вот результат эксперимента, в котором нашей системе предлагалось по десять раз распознать пару слов «молоко» и «далеко» (их транскрипции в нашей системе: «малако» и «dаlипо» соответственно). Строки таблицы (табл. 1), начиная со второй, содержат цепочки транскрипционных знаков, полученные в ходе распознавания.

Таблица 1

Молоко	Далеко
Мамапо	Домипо
Малапо	Domlпо
Мавапо	Дамlпо
Мамапо	Дvмипо
Мавапо	Дамипо
Мавапо	Дадипо
Мамапл	Дамипо
Мамапо	Дalэпо
Мамапо	Доуипо
Мавапо	Дэмипо

При распознавании обоих слов машина только по разу получила цепочки, совпадающие с транскрипциями. Однако во всех случаях, не смотря на ошибки в цепочке фонем, результат распознавания оказывался правильным. Здесь, конечно, можно говорить о некачественном распознавании фонем. Но мы специально выбрали один из наименее удачных примеров распознавания фонем, так как хотим обратить внимание на другое – на роль интерпретатора.

Вот результаты (табл. 2) аналогичного эксперимента с парой слов «шоссе» (транскрипция «шасэ») и «сушка» (транскрипция «сушпа»).

Таблица 2

Шоссе	Сушка
Шасэ	Сушфа
Шаса	Слшпа
Щасэ	Сушпа
Щаса	Слшфа
Шасэ	Сушпа
Шасэ	Сушпа

И здесь оба слова во всех случаях при обработке интерпретатором были распознаны правильно.

Отметим в заключение, что наличие дикторонезависимого интерпретатора, очевидно, должно уменьшать зависимость диктора от результатов работы всей программы распознавания.

Литература

1. Дорохин О.А., Старушко Д.Г., Федоров Е.Е., Шелепов В.Ю. Сегментации речевого сигнала .// Искусственный интеллект.–Донецк.– 2000.– № 3. – С. 450-458.
2. Грабовая В.А., Федоров Е.Е., Шелепов В.Ю. О системе распознавания русской речи с автоматическим построением эталонов //Искусственный интеллект. – 2000.– № 1.
3. Дорохин О.А., Федоров Е.Е., Шелепов В.Ю. Некоторые подходы к пофонемному распознаванию русской речи и распознаванию больших словарей // Искусственный интеллект. – 1999.– № 2. – С. 329-333.
4. Дорохин О.А., Засыпкин А.В., Червин Н.А., Шелепов В.Ю. О некоторых подходах к проблеме компьютерного распознавания устной русской речи // Сборник трудов международной конференции «Знания, диалог, решение».–Т.1.– Ялта, – 1997.– С. 234-240.

Материал поступил в редакцию 25.06.01.