

УДК 004.934

**О РАСПОЗНАВАНИИ СВЕРХБОЛЬШИХ СЛОВАРЕЙ
РУССКИХ СЛОВОФОРМ НА ОСНОВЕ РАСПОЗНАВАНИЯ
КВАЗИОСНОВ
А.В. Ниценко**

Институт проблем искусственного интеллекта

Статья посвящена описанию способа распознавания изолировано произнесенных русских слов, основанного на предварительном распознавании квазиосновы слова. В результате значительно сокращается время работы алгоритма и увеличивается надежность распознавания словоформ. Метод может быть применен к задаче распознавания больших и сверхбольших словарей.

Одной из трудностей применения технологий распознавания речи для русского языка является сложный механизм словообразования. В отличие от, например, английского языка русский язык обладает вариативностью на уровне словоформ. При этом наличие правильных словоформ является неременным условием связной речи, обеспечивающим правильный синтаксис, а значит и правильное понимание. Орфографические словари русского языка содержат около 100 тысяч слов в начальной (словарной) форме. Все они, за исключением служебных слов, имеют наборы словоформ, многие (например, глаголы) – в весьма большом количестве. С учетом этого словарь словоформ русского языка разрастается до размера порядка 3 миллионов. С точки зрения распознавания разные словоформы – это разные классы. С этим связана одна из принципиальных трудностей распознавания русской речи: огромное число классов.

Для решения этой проблемы предлагается распознавать словоформы в два этапа: на первом этапе распознать неизменяющуюся часть слова (квазиоснову), а затем распознавать словоформу целиком на множестве кандидатов, содержащих в себе распознанную квазиоснову. Таким образом, значительно сократится размер распознаваемых словарей и количество распознаваний. Квазиосновы, очевидно, являются общими для словоформ, имеющих общую начальную форму. Если имеется m квазиоснов и

n изменяющихся частей слов (квазифлексий), то их комбинации образуют $m \times n$ словоформ и, распознавая словоформы как целое, мы имели бы словарь для распознавания из $m \times n$ объектов. Распознавая же предварительно квазиосновы, можно сократить количество распознаваний до $m + n$ объектов. В результате время распознавания значительно сокращается, а надежность распознавания увеличивается [1].

Исходя из того, что русский язык является флективным языком (т.е. синтаксическое управление осуществляется с использованием словоформ, образуемых при помощи флексий), слова языка моделируются в виде комбинации постоянной и переменной составляющих:

$$x = C(x) \& F(x)$$

где $C(x)$ – часть лексемы x , которая в процессе словоизменения остается неизменной (квазиоснова), $F(x)$ – ее переменная составляющая (квазифлексия), $\&$ – знак конкатенации.

Рассматривая множество словоформ, их квазиосновы можно объединить в отдельные словари, по которым в дальнейшем будет происходить первичное распознавание.

Так как распознавание будет вестись с использованием эталонов дифонов [2,3], то для каждой квазиосновы используется транскрипция и по ней создается цепочка соответствующих дифонов. Например, для слова «абрикосовый»:

абрикосо → абгикаса → а0-аб-бг-ги-ик-ка-ас-са-а2

Для транскрибирования используется упрощенный автоматический транскриптор, обеспечивающий потребности распознавания. Он реализован как программа, заменяющая одни символы другими в соответствии с правилами, содержащимися в управляющем файле. Каждое из них записано в виде двух частей, соединенных знаком равенства. Слева стоят исходные символы буквенной записи слова, справа – символы которыми они заменяются в транскрипции. Программа, транскрибируя слово, последовательно ищет вхождение левой части очередного правила, и если таковое обнаруживается, заменяет его правой частью.

Перед распознаванием речевого сигнала выполняется его автоматическая фонемная сегментация – разбиение на участки, отвечающие четырем классам широкой фонетической классификации

[3]: гласный (W), звонкий согласный (C), фрикативный (F) и паузообразный (P) (глухой взрывной) или часть аффрикаты (рис.1). Для распознавания применяется алгоритм, известный под названием алгоритма динамической временной деформации (DTW)[4]. При этом используются вектора признаков, связанные с относительными частотами длин полных колебаний на речевых отрезках в 368 отсчетов (см. [3]). Эталоны частей слов распознаваемого словаря формируются из эталонов дифонов, полная база которых в объеме приблизительно полутора тысяч создается для каждого диктора заранее. Под «дифоном» в данной работе, подразумевается участок стандартной длины: 3 окна в 368 отсчетов слева от метки между звуками и 3 таких же окна справа от той же метки, соответствующий межфонемному переходу внутри слова. Эталон дифона – набор 6-ти соответствующих векторов признаков, его имя – пара символов для соответствующих звуков. Кроме того, используется участок в 3 окна в начале слова (имя – символ начального звука с добавлением символа 0) и участок в 3 окна в конце слова (имя – символ конечного звука с добавлением символа 2), условно называя их соответственно начальным и конечным полудифонами (переход от молчания к речи и наоборот). Поскольку глухие взрывные звуки [к], [п], [т] и их мягкие аналоги в начале слова при сегментации не выделяются, используется также начальные полудифоны вида [ка0], [пи0], [тл0] и так далее.

Опираясь на полученную фонемную сегментацию, производится определение квазиосновы по принципу минимума DTW-расстояния [3]. Для этого производится серия последовательных сравнений с помощью алгоритма DTW частей речевого сигнала со всеми эталонами квазиоснов. Вначале сравнивается весь отсегментированный участок сигнала, от первой метки до последней, затем укороченный на один сегмент, затем на два, три и так далее до задаваемого заранее максимального количества отбрасываемых сегментов (рис. 1).

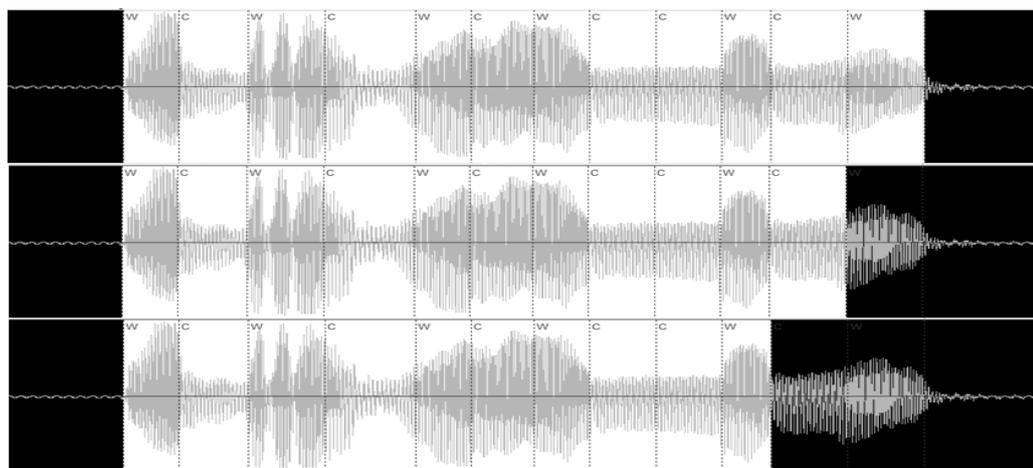


Рисунок 1 – Пример последовательного распознавания участков сигнала

При этом запоминается минимальное значение DTW-расстояния среди всех эталонов и соответствующая этому эталону основа квазиоснова, и далее производится сравнение эталонов со следующим участком сигнала. Таким образом, получается список гипотетических квазиоснов, которые, возможно, содержатся в распознаваемом слове. Чтобы перейти от квазиоснов к словоформам, необходимо отобрать из словаря всех словоформ русского языка те слова, которые содержат одну из распознанных квазиоснов и провести еще одно распознавание по этому сокращенному словарю.

На рис. 2 показан пример такого списка, полученного при распознавании слова «абразионному». В результате четырех последовательных распознаваний по словарю квазиоснов определены две гипотетические квазиосновы – «абразион» и «абдоминальн». Затем для выделенных таким образом квазиоснов происходит поиск соответствующих им словоформ в большом словаре, содержащем около 2 млн. словоформ русского языка. В пределах полученного сокращенного словаря кандидатов осуществляется DTW-распознавание всего речевого сигнала, которое окончательно определяет произнесенную словоформу по принципу минимума расстояния.

абрази\он	14.02		абдоми\альная
абрази\он	15.30		абдоми\альная
абдоминальн	16.66	→	абдоми\альнее
абдоминальн	16.99		абдоми\ального
			абдоми\альное
			абдоми\альной
			абдоми\альном
			абдоми\альному
			абдоми\альнойю
			абдоми\альны
			абдоми\альные
			абдоми\альный
			абдоми\альным
			абдоми\альными
			абдоми\альных
			абрази\онен
			абрази\онна
			абрази\онная
			абрази\оннее
			абрази\онно
			абрази\онного
			абрази\онное
			абрази\онной
			абрази\онном
			абрази\онному
			абрази\онною
			абрази\онною
			абрази\онны
			абрази\онные
			абрази\онный
			абрази\онным
			абрази\онными
			абрази\онных

Рисунок 2 – Список гипотетических квазиоснов и соответствующих им словоформ при распознавании слова «абразионному»

Таким образом, словарь из 2 млн. словоформ сокращается до порядка нескольких десятков или сотен словоформ.

Выводы

В данной работе предложен способ распознавания речи для распознавания словоформ русских слов, основанный на сегментации и алгоритме динамической временной деформации и использующий предварительное распознавание квазиосновы слова. Его перспективность обусловлена трудностью применения существующих технологий распознавания речи для русского языка из-за сложного механизма словообразования. Метод направлен на уменьшение вычислительных затрат при распознавании и увеличение надежности распознавания словоформ, и может быть применен при распознавании изолированно произнесенных слов с большими и сверхбольшими словарями.

Список литературы

1. Nicenko A.V. A «by part» method of Russian word speech recognition / A.V. Nicenko // Eurasian Journal of Mathematical and Computer Applications. – 2014. – Vol.1. – Iss. 2 — P. 102-109.
2. Шелепов В.Ю., Ниценко А.В., Дорохина Г.В. О некоторых вопросах, связанных с дифонным распознаванием и распознаванием слитной речи // Искусственный интеллект. – 2013. – №3 – С. 209-216.
3. Шелепов В. Ю., Ниценко А.В., Дорохина Г.В. О распознавании речи на основе межфонемных переходов // Искусственный интеллект. – 2012. – №1 – С. 132-139.
4. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов // Киев: Наук.думка. – 1987. – 262 с.