

УДК 004.89:004.93

В.Ю. Шелепов, А.В. Ниценко

О РАСПОЗНАВАНИИ СВЕРХБОЛЬШИХ СЛОВАРЕЙ РУССКИХ СЛОВОФОРМ С ИСПОЛЬЗОВАНИЕМ КВАЗИОСНОВ

Предлагается метод распознавания слов сверхбольшого словаря русских словоформ, использующий принадлежащий авторам способ автоматической сегментации речевого сигнала. Сам метод применим как для DTW-распознавания, так и для распознавания с помощью скрытых марковских моделей. Однако при описании практического распознавания речевых единиц имеется в виду развиваемый авторами метод дифонного DTW-распознавания. В связи с этим приводится понятие формализованного дифона. Далее определяются понятия квазиосновы и квазифлексии. Предлагается общий алгоритм построения квазиоснов для заданного списка словоформ. Излагается алгоритм распознавания множества русских причастий с использованием квазиоснов. С целью ускорения распознавания большого словаря квазиоснов предлагается алгоритм распознавания начальных звуков слов или достаточно узких классов, которым они принадлежат. Отметим в заключение следующее. Существует объективная проблема надежного распознавания коротких слов. Поэтому квазиосновы хорошо распознаются, когда они достаточно длинные. В общем же случае уязвимым местом при использовании квазиоснов является то, что приходится распознавать укороченные по сравнению с первоначальными словами речевые отрезки (шагом в направлении преодоления этого затруднения служит вышеупомянутая процедура классификации первого звука). Тем не менее, использование квазиоснов представляется разумным при распознавании больших и сверхбольших словарей русских словоформ.

Сегментация; квазиоснова; словоформы с вхождением; участки первого звука; классы распознавания; переходы в зависимости от промежуточных результатов распознавания.

V.Yu. Shelepov, A.V. Nicenko

ON RECOGNITION OF THE SUPER DIMENSIONED VOCABULARIES OF RUSSIAN WORD FORMS USING QUASI-WORD STEMS

The paper discusses the method of word recognition of the super dimensioned vocabularies of Russian word forms using the authors' automatic segmentation of speech signal system. The method can be applied both for DTW-recognition and recognition by means of the Hidden Markov models. But we mean method of diphone DTW- recognition (developing by authors) when practical recognition of speech units is described. Hence we adduce the conception of formalized diphone. Then the notions of quasi-word stem and quasi-inflection are defined. The general algorithm of quasi-word stems construction for the given list of word forms is suggested. The algorithm of Russian participles recognition using quasi-word stems is expounded. We suggest the recognizing algorithm of the words initial sounds (or sufficiently narrow classes of they belong to) in order to accelerate the recognition of the large vocabulary of quasi-word stems. In conclusion it should be noted that there is an objective problem of robust recognition of short words. Therefore quasi-word stems have good recognition when they are sufficiently long. In general case the vulnerable point of quasi-word stems using is the recognition of shorter speech segments in comparison with the primary words (the step to overcome this difficulty lies in above-mentioned procedure of the first sound classification). Nevertheless using of quasi-word stems seems reasonable in recognition of the super dimensioned vocabularies of Russian word forms.

Segmentation, quasi-word stem, parts of the first sound, recognition classes, transitions depending on intermediate recognition results.

Введение. В последнее время появились реально работающие распознаватели речи с большими словарями, прежде всего в поисковых системах Google и Yandex. Однако работа с ними происходит через сеть Internet. В то же время проблема распознавания на малых локальных компьютерах остается актуальной.

В дальнейшем, говоря о распознавании речевых единиц, мы имеем в виду развиваемый авторами метод дифонного DTW-распознавания, использующий принадлежащий авторам способ автоматической сегментации речевого сигнала с отнесением звуков к шипящим, звонким согласным и т.д. (см. [1]). Сегментации речевых сигналов посвящены также работы [2–6].

При 8-битной записи и частоте дискретизации 22050 Гц, используется разбиение сигнала на окна по 368 отсчетов (удвоенный квазипериод основного тона для мужского голоса средней высоты). На каждом из них вычисляется вектор признаков, связанных с относительными частотами длин полных колебаний. Каждому слову распознаваемого словаря может быть поставлен в соответствие эталон – набор векторов признаков (см. [1]). При дифонном распознавании мы формируем по транскрипциям слов эталоны из заранее созданных эталонов дифонов.

Первоначально дифон определяется как речевой отрезок между серединами двух соседних звуков речи. Занимаясь дифонным распознаванием, мы ввели понятие формализованного дифона (далее называем его просто дифоном) – отрезок в 3 окна слева и 3 окна справа от метки между соседними звуками. Имя дифона – пара символов, соответствующих составляющим звукам. Мы используем также начальный полудифон – 3 окна от начала (его имя снабжается в конце символом 0) и конечный полудифон – 3 окна слева от конца слова. Всего в русской речи (при используемой нами системе автоматической транскрипции) около 1700 дифонов. Использование дифонов вместо чистых звуков предпочтительно ввиду явления коартикуляции.

Из недавних статей, посвященных распознаванию больших словарей, отметим [7–9]. Авторы данной работы располагают электронным словарем русских словоформ размером приблизительно в 2 миллиона единиц (далее словарь (СЛФ)). Разработан также очень быстрый алгоритм поиска в таком словаре, использующий представление его в виде дерева. На сегодняшний день мы имеем распознаватель, успешно работающий со словарями до 100 тысяч слов. Однако даже не приходится говорить о каком-либо приемлемом быстродействии на упомянутом 2-миллионном словаре. Для решения этой проблемы предлагается механизм распознавания с использованием квазиоснов. Его можно применять и при НММ-распознавании. Отметим в связи с этим работы [10–18], имеющие дело с морфемами.

О понятии квазиосновы. Рассмотрим некоторое слово в начальной форме. Множество всех его словоформ называют его парадигмой. Выберем теперь из парадигмы некоторое подмножество словоформ и будем сравнивать все эти словоформы, двигаясь от начала к концу каждой из них. Их максимальную общую часть назовем квазиосновой данного подмножества словоформ. Например, парадигма слова *ЛЕГЕНДАРНЫЙ* состоит из падежных форм *ЛЕГЕНДАРНЫЙ*, *ЛЕГЕНДАРНОГО*, ..., к которым присоединяются формы превосходной степени *ЛЕГЕНДАРНЕЙШИЙ*, *ЛЕГЕНДАРНЕЙШЕГО*, ..., формы сравнительной степени и краткие формы *ЛЕГЕНДАРЕН*, *ЛЕГЕНДАРНА*, *ЛЕГЕНДАРНО*. Общей квазиосновой для всех словоформ, кроме *ЛЕГЕНДАРЕН*, будет *ЛЕГЕНДАРН*. Для словоформы *ЛЕГЕНДАРЕН* квазиосновой будет она сама. Если же рассматривать парадигму целиком, то квазиосновой будет *ЛЕГЕНДАР*.

Часть словоформы, следующую за квазиосновой назовем квазифлексией. Идея использования квазиоснов при распознавании больших и сверхбольших словарей словоформ основана на том, что при подходящем разбиении парадигмы каждая из достаточно больших группы словоформ имеет одну общую квазиоснову. В результате соответствующий словарь квазиоснов оказывается во много раз меньше исходного словаря. Значит, процедуру распознавания разумно сделать пошаговой: вначале провести распознавание на словаре квазиоснов. Затем, ото-

брав из исходного супербольшого словаря (СЛФ) все словоформы, в которых есть вхождения распознанного на первом шаге, провести распознавание на небольшом словаре отобранных словоформ.

Ситуацию можно описать также следующим образом. Рассмотрим множество m квазиоснов и набор квазифлексий в количестве n , с помощью которых образуется mn словоформ. При распознавании результата произнесения одной из этих словоформ, когда она распознается целиком, потребуется провести mn распознаваний. Если же вначале распознать квазиоснову, а затем распознать квазифлексию, то это потребует $m+n$ распознаваний. Последний шаг равносителен распознаванию словоформ с заранее распознанной квазиосновой. При этом отношение $mn/(m+n)$ при $m \rightarrow \infty$ стремится к n . Т.е. число отдельных распознаваний в n раз меньше.

Введенное понятие квазиосновы родственно используемому в лингвистике понятию основы слова, которая при простейшем описании определяется как его неизменяемая часть (приставка+корень+суффикс), то есть является результатом отбрасывания окончания. Мы ввели понятие квазиосновы, поскольку распознавание тем надежнее, чем длиннее распознаваемые речевые отрезки. Поэтому неразумно заменять распознавание коротких словоформ распознаванием еще более короткой их общей части. Рациональнее включать их в число квазиоснов целиком.

Поскольку при распознавании отдельно произносимых словоформ каждая из них произносится целиком, возникает проблема выделения в записанном сигнале части, соответствующей квазиоснове. Эта процедура аналогична процедуре распознавания слитной речи, которую мы реализуем как распознавание последовательно наращиваемых речевых отрезков и выделение нужного отрезка по принципу минимума DTW-расстояния (см. [16]). Однако, работая со словарем квазиоснов, мы предлагаем заранее разбить его на части, для которых длины квазиоснов находятся в заданных пределах.

Алгоритм построения квазиоснов для заданного списка словоформ. Поскольку мы имеем дело с распознаванием устной речи, то длиной слова (или его части) будем называть количество соответствующих отрезков при сегментации результата произнесения. Заметим, что в звуках [ц], [ч], как правило, выделяются паузообразная и фрикативная части, но считается, что каждому из этих звуков соответствует один сегмент. С другой стороны глухие взрывные звуки в начале слова не выделяются при сегментации. Таким образом, длину можно определить также как количество транскрипционных символов, уменьшенное на количество глухих взрывных в начале слова. Так длина слова *РОЛЬ* равняется 3, а длина слова *КАНАЛ* равняется 4.

Далее, говоря о сравнении двух строк, мы будем сравнивать их от начала к концу и называть пересечением соответствующую общую часть. Речь идет о создании для данного списка словоформ соответствующего набора квазиоснов, длины которых не меньше некоторого наперед заданного числа k ; если словоформа короче k , то она включается в число квазиоснов целиком. Отметим, наконец, что мы используем знак ударения только при гласном [о], так как его отсутствие превращает этот звук в безударное [а]. В остальных случаях при нашем подходе к распознаванию ударность или безударность гласных не влияет на результат распознавания слов.

Рассматриваемый список словоформ F упорядочивается по алфавиту с учетом ударений. Далее сравниваются первые две строки. Если длина их пересечения меньше k , то оставляем в списке первую строку и переходим к сравнению второй и третьей строк. Если длина пересечения не меньше k , заменяем их одной строкой – пересечением и переходим к сравнению двух следующих строк. Проводя такую обработку парами, добираемся до конца списка. Потом снова проводим от его на-

чала описанную попарную обработку, и так далее. Этот циклический процесс заканчивается, когда очередная попарная обработка всего списка не меняет его. Обозначим через f_i строку-словоформу с номером i , через k – минимально допустимую длину квазиосновы. Алгоритм формализуется следующим образом.

1. Сохранить копию списка F в списке F_0 . $i=0$. Переход к п.2.
2. Если $i+1$ меньше, чем количество строк в списке F , то переход к п.3, иначе переход к п. 7.
3. Найти пересечение S строк f_i и f_{i+1} . Переход к п.4.
4. Если количество символов в S меньше k , то переход к п. 5, иначе переход к п. 6.
5. Увеличение i на 1. Переход к п. 2.
6. Замена f_i на S , удаление f_{i+1} . Увеличение i на 1 и переход к п. 2.
7. Удаление из F дубликатов строк, расположенных рядом. Переход к 8.
8. Если количество строк в списке F меньше, чем в списке F_0 , то переход к п. 1, иначе конец работы алгоритма. На рис. 1 – блок-схема алгоритма.

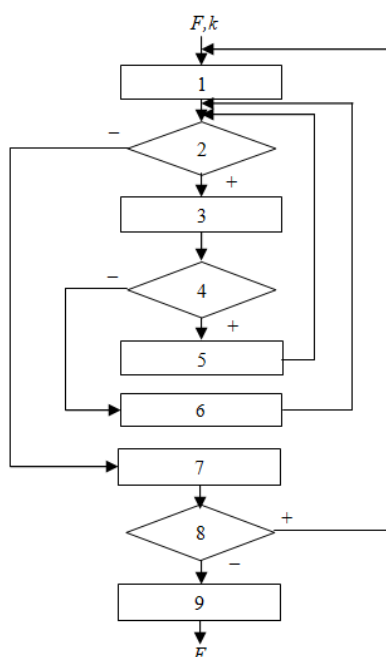


Рис. 1. Блок-схема алгоритма формирования квазиоснов

Если список состоит из парадигм разных слов, то применять описанный алгоритм целесообразно к парадигме каждого слова в отдельности, проходя их одну за другой. Того же можно достигнуть введением в список разделяющих строк (скажем из единиц) между парадигмами. Алгоритм устроен так, что он будет сохранять эти строки до самого конца, после чего их нужно будет удалить. Начальная сортировка по алфавиту должна выполняться между каждыми двумя разделяющими строками с добавлением новой разделяющей строки при появлении, перемещении или исчезновении ударения. Мы используем полный перечень парадигм для всех слов известного словаря А.А. Зализняка [19], полученный с помощью разработанного в Институте проблем искусственного интеллекта морфоанализатора (см. [20–21]).

Далее мы опишем полную процедуру распознавания с использованием квазиоснов на примере причастий. На самом деле использование квазиоснов полных глагольных парадигм, включая причастия, является весьма привлекательным, ибо тогда словарь квазиоснов оказывается сравнительно небольшим. Однако парадигмы причастий, рассматриваемые отдельно, имеют большое число достаточно длинных квазиоснов, с повышенной надежностью распознавания. В связи с этим имеет смысл работать с причастиями, рассматривая в качестве начальных форм именительные падежи мужского рода единственного числа (их несколько, в зависимости от выбора времени и активной или страдательной формы). Это оправдано также тем, что число словоформ для причастия велико: общее число словоформ причастий, соответствующих глаголам словаря Зализняка составляет около 750 тысяч. В то же время процедура образования таких словоформ очень проста. Например, для причастий, оканчивающихся в начальной форме на «ший», при образовании словоформ варьируется только окончание: «его», «ему» и так далее. Таким образом, квазиосновы для причастий можно образовать способом более простым, нежели вышеприведенный алгоритм: они выделяются путём удаления окончаний в словаре начальных форм причастий. Дальнейшее изложение начнем с примера.

Парадигма причастия от глагола *АВАНСИРОВАТЬ* состоит из результатов словоизменения для причастий *АВАНСИРЮЮЩИЙ*, *АВАНСИРОВАВШИЙ*, *АВАНСИРЮЕМЫЙ*, *АВАНСИРОВАННЫЙ* по родам, числам и падежам. К ним добавляются краткие формы: *АВАНСИРЮЕМ*, *АВАНСИРЮЕМА*, *АВАНСИРЮЕМО*, *АВАНСИРЮЕМЫ*, *АВАНСИРОВАН*, *АВАНСИРОВАНА*, *АВАНСИРОВАНО*, *АВАНСИРОВАННЫ*, и возвратные формы, оканчивающиеся на *СЯ*: *АВАНСИРЮЮЩИЙСЯ*, *АВАНСИРОВАВШИЙСЯ* и так далее. Всего парадигма насчитывает 92 словоформы.

Общей квазиосновой является *АВАНСИР*. Однако вышеописанные соображения по поводу надежности распознавания делают целесообразным выделение четырех квазиоснов:

АВАНСИРОВАВШ, *АВАНСИРЮЮЩ*, *АВАНСИРОВАН*, *АВАНСИРЮЕМ*.

Остановившись на этом последнем варианте выделения квазиоснов, приведем следующие цифры для множества всех причастий, соответствующих глаголам из словаря А.А. Зализняка: всего словоформ приблизительно 750 тысяч, квазиоснов – приблизительно 55 тысяч, то есть в 13 с половиной раз меньше.

Алгоритм распознавания причастий с использованием квазиоснов. Все квазиосновы разбиваются на 4 словаря. Они соответственно содержат квазиосновы, оканчивающиеся шипящим (словарь *F.txt*), квазиосновы, оканчивающиеся глухим взрывным звуком (словарь *P.txt*), квазиосновы, оканчивающиеся другими (голосовыми) звуками (словарь *N.txt*). квазиосновы причастий, оканчивающихся на *СЯ* (словарь с именем *СЯ.txt*).

Если произнесенная словоформа заканчивается парой звуков фриктивный плюс голосовой, то эта пара распознается на коротком словаре звукосочетаний $Ends = \{ ся, ший, ший \}$. В случае распознавания *СЯ* далее распознается часть от начала сигнала до конца шипящего, предшествующего *СЯ*, и распознавание ведется по словарю *СЯ.txt*. Для распознанной квазиосновы по словарю (СЛФ) находятся все содержащие ее словоформы. Они образуют в свою очередь небольшой словарь, на котором проводится распознавание всего записанного сигнала (заключительное распознавание). Пусть на упомянутом коротком словаре результат распознавания – не *СЯ*. Тогда распознавание продолжается следующим образом. Если сказанная словоформа после последнего шипящего содержит не более 3-х звуков, то делается заключение, что ее квазиоснова заканчивается шипящим. В этом случае распознается часть сигнала от начала до последнего шипящего включительно и распознавание ведется по словарю *F.txt*. Рис. 2 отражает результат автоматического выделения квазиосновы.

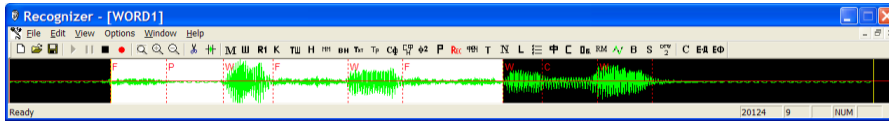


Рис. 2. Автоматическое выделение квазиосновы в слове «спешащего»

Аналогично с использованием словаря *P.txt* происходит распознавание словоформы, содержащей не более 3-х звуков после последнего глухого взрывного. В обоих случаях далее следует поиск вхождений и заключительное распознавание по словарю (СЛФ).

Если сказанная словоформа не обладает ни одним из свойств, описанных в предыдущих абзацах, то дальнейшая работа происходит со словарем *N.txt*. Последний разбивается на словари с именами *n.txt*, где *n* – число звуков от начала до конца квазиосновы. От начала выделяется часть сигнала из *n* отрезков сегментации и распознается на словаре *n.txt*. Эти операции выполняются для всех *n*, при которых соответствующий словарь *n.txt* не пуст. Очевидно, при этом распознавание каждый раз осуществляется на минимальном подходящем словаре квазиоснов. В результате формируется список кандидатов на распознавание квазиосновы. Для каждого кандидата по словарю (СЛФ) находятся все содержащие его словоформы. Все они в совокупности образуют небольшой словарь, на котором затем проводится распознавание всего записанного сигнала.

Выделение начальной части из *n* звуков использует выполненную с самого начала сегментацию. Не исключено, что последняя может содержать ошибки, и тогда обращение к словарю *n.txt* приведет к ошибочному распознаванию квазиосновы. Более надежным является не выделение начальной части сигнала, а отсечение его заключительной части. Но заранее не известно, сколько нужно отсекал. Поэтому мы приходим к последовательному распознаванию на всем словаре *N* сигнала без последнего сегмента, без двух последних сегментов, без трех последних сегментов. Здесь чаще всего не возникает ошибок, если даже мы отсекали заключительную часть сигнала не очень точно. Более того, опыт показывает, что можно выбрать нечто среднее и всегда ограничиваться лишь отсечением двух последних звуков. Тогда мы снова приходим к однократному распознаванию всех квазиоснов словаря *N*. Формализуем описанный алгоритм.

Исходной информацией является распознаваемый сигнал $x(t)$, сегментация *Segm* (поле *let* обозначает метку сегмента, *Start* – начало сегмента, *End* – конец сегмента), количество сегментов *M*, словарь *Ends*, словари *F.txt*, *P.txt*, *СЯ.txt* и *N.txt*, словарь СЛФ.

1. Если $\text{Segm}(M-2).\text{let} = \langle F \rangle$ и $\text{Segm}(M-1).\text{let} \neq \langle F \rangle$ и $\text{Segm}(M-1).\text{let} \neq \langle P \rangle$ то переход к пункту 2, иначе переход к п.9.
2. Распознавание участка сигнала $x(t)$ от $\text{Segm}(M-2).\text{Start}$ до $\text{Segm}(M-1).\text{End}$ на словаре *Ends*, результат в *Res*. Переход к п.3
3. Если *Res* = «ся», то переход к п.4, иначе переход к п. 6.
4. $k = M-2$. Поиск в *Segm* максимального номера сегмента *k*, такого что $\text{Segm}(k).\text{let} = \langle F \rangle$, $k < M-2$. Переход к п. 5.
5. Распознавание участка сигнала $x(t)$ от $\text{Segm}(0).\text{Start}$ до $\text{Segm}(k).\text{End}$ на словаре *СЯ.txt*, результат в *Res*. Переход к п. 10.
6. $k = M-1$. Поиск в *Segm* максимального номера сегмента *k*, такого что $\text{Segm}(k).\text{let} = \langle F \rangle$ или $\text{Segm}(k).\text{let} = \langle P \rangle$, $k < M-1$. Переход к п. 7.
7. Если $(M-1)-k \leq 3$, то переход к п. 8, иначе переход к п. 9.
8. Распознавание участка сигнала $x(t)$ от $\text{Segm}(0).\text{Start}$ до $\text{Segm}(k).\text{End}$ на словаре $[\text{Segm}(k).\text{let}].\text{txt}$, результат в *Res*. Переход к п. 10.

9. Распознавание участка сигнала $x(t)$ от $\text{Segm}(0).\text{Start}$ до $\text{Segm}(M-2).\text{Start}$ на словаре $N.\text{txt}$, результат в Res . Переход к п. 10.
10. Поиск по словарю (СЛФ) множества Forms словоформ, содержащих Res . Переход к п. 11.
11. Распознавание участка сигнала $x(t)$ от $\text{Segm}(0).\text{Start}$ до $\text{Segm}(M-1).\text{End}$ на множестве Forms , результат в Res . Переход к п.
12. Конец работы алгоритма. На рис.2 – блок-схема алгоритма.

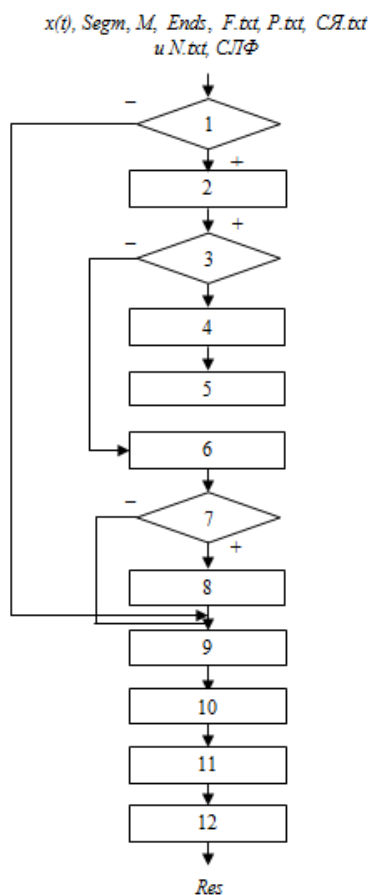


Рис. 3. Блок-схема алгоритма распознавания причастий

С помощью подобных алгоритмов ведется распознавание словоформ и других частей речи. Как отмечалось выше, квазиосновы для причастий выделяются путём удаления окончаний в словаре начальных форм. Таким же образом определяются квазиосновы для прилагательных. Единственный нюанс здесь связан с краткими формами мужского рода. Например, для прилагательного «активный» таковой будет слово «активен». Трудности с квазиосновами для существительных связаны с существованием беглых гласных («сосна-сосен») и чередованиями «е-ё» («платёж-платежей»). Наибольшие проблемы – с глаголами, где в лингвистике с самого начала выделяют две основы – основу инфинитива и основу настоящего-будущего времени. Далее, при работе с глаголами возникает масса проблем с чередованиями. В связи с этим отсылаем читателя к предложенному выше общему алгоритму образования квазиоснов.

Ускорение распознавания путем классификации начальных звуков. Словари квазиоснов все же остаются весьма большими. Для ускорения их распознавания применяется классификация начальных звуков.

В дальнейшем гласные и твердые согласные звуки обозначаются соответствующими русскими буквами в квадратных скобках. Мягкие согласные обозначаются соответствующими латинскими буквами в квадратных скобках. Исключение для мягкого [п'], которое будет обозначаться через [@]. Обозначим через (%) первый сегмент слова, через (*) – отрезок, получаемый из (%) отбрасыванием участка заключительного полудифона, через (0) – отрезок начального полудифона.

Результирующие классами распознавания будут: а, и, о, у, э, ж, ж1, ш, щ, Pa, Pe, Pё, Pi, Po, Py, Pы, Pэ, Pю, Pя, D, j, L, N, R, S, Z, Z1. Каждый из классов а, ..., ж, ж1, ш, щ, j состоит из всевозможных реализаций звука, соответствующего буквенному символу. Класс Pa состоит из всех начальных полудифонов вида Pa0, где на месте P может стоять любой из звуков [к], [п], [т]. Аналогично устроены классы, Po, Py, Pы, Pэ. Класс Pi состоит из полудифонов вида Pi0, где на месте P может стоять любой из звуков [к], [@], [т]. Аналогично устроены классы, Pe, Pё, Pю, Pя. При этом под [е], [я] понимаются только соответствующие ударные звуки. Это связано с тем, что в безударном варианте они произносятся крайне неопределенно. Далее, D– класс звуков [б], [b], [г], [g], [д], [d], [в], [v]; L– класс звуков [л], [l]; N – класс звуков [м], [m], [н], [n]; R– класс звуков [р], [r]; S – класс звуков [с], [s]; Z– класс звуков [з], [z], Z1– класс звуков [з], [z].

В качестве промежуточных выступают классы распознавания:

A – множество всех пар вида (а, Pa), ..., E – множество всех пар вида (е, Pe),

Я – множество всех пар вида (я, Pя). Звуки [е], [я] здесь ударные.

Распознаватели с перечнем классов распознавания:

DTW(*): A, E, Ё, И, О, У, Ы, Э, Ю, Я, ж, ж1, ш, щ, j, R, S, D, L, N, Z, Z1;

DTW_A(0): а, Pa; ...; DTW_Э(0): э, Pэ. (в скобках – отрезок распознавания).

Алгоритм классификации начальных звуков. Начнем с блок-схемы:

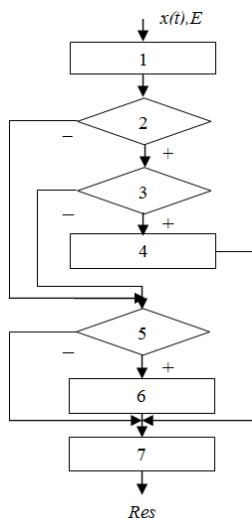


Рис. 4. Блок-схема алгоритма классификации первого звука слова

Приведем пошаговое описание алгоритма. Исходная информация: распознаваемый сигнал $x(t)$ и список эталонов E для классов распознавания.

1. Распознавание с помощью DTW(*): $Res = DTW(*)$. Переход к п.2.
2. Если Res не входит в множество { ж, ж1, ш, щ, j, R, S, D, L, N, Z, Z1 }, то переход к п. 3, иначе переход к п. 5.
3. Если Res входит в множество {А,И,О,У,Э}, то переход к п. 4, иначе переход к п. 5.
4. $Res = DTW_{Res}(0)$. Переход к п. 7.
5. Если Res входит в множество {Ы,Е,Ё,Ю,Я}, то переход к п. 6, иначе переход к п. 7.
6. $Res = \langle P \rangle + Res$. Переход к пункту 7.
7. Конец работы алгоритма.

Тестирование. Тестирование проводилось на полном словаре словоформ причастий. Эксперименты состояли в произнесении 100 случайно отобранных словоформ. При этом зафиксировано не более 10% ошибок в квазиосновах и не более 8 % ошибок в квазифлексиях.

Выводы. Предложен метод распознавания сверхбольшого словаря отдельно произносимых русских словоформ с использованием квазиоснов. Предлагаемый механизм опирается на разработанные авторами априорную сегментацию и диффонное DTW-распознавание. Надежность распознавания выше той, которая приводится в работах [10–15], посвященных аналогичным вопросам для распознавания слитной речи. Правда при этом нужно учитывать, что в нашем случае речь идет о системе, настроенной на конкретного диктора. В целях ускорения распознавания осуществляется распознавание первого звука или содержащего его узкого класса звуков.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бурибаева А.К., Дорохина Г.В., Ниценко А.В., Шелепов В.Ю. Сегментация и диффонное распознавание речевых сигналов // Труды СПИИРАН. – 2013. – Вып. 31. – С. 20-42.
2. Mporas I., Ganchev T. and Fakotakis N. Speech segmentation using regression fusion of boundary predictions // Computer Speech and Language. – 2010. – Vol. 24, No. 2. – P. 273-288.
3. Gómez J.A., Calvo M. Improvements on Automatic Speech Segmentation at the Phonetic Level // Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. – 2011. – Vol. 7042. – P. 557-564.
4. Natarajan V.A., Jothilakshmi S. Segmentation of Continuous Speech into Consonant and Vowel Units using Formant Frequencies // International Journal of Computer Applications. – 2012. – Vol. 56, No. 15. – P. 24-27.
5. Yuan J., Ryant N., Liberman M. [and all]. Automatic phonetic segmentation using boundary models // Proceedings of Interspeech 2013. – 2013. – P. 2306-2310.
6. Patc Z., Mizera P., Pollak P. Phonetic Segmentation Using KALDI and Reduced Pronunciation Detection in Causal Czech Speech // Text, Speech, and Dialogue. – 2015. – Vol. 9302. – P. 433-441.
7. Zhijian Ou, Xiao Ji. A study of large vocabulary speech recognition decoding using finite-state graphs // Chinese Spoken Language Processing (ISCSLP), 7th International Symposium. – 2010. – P. 123-128.
8. Susman D., Kopru S., Yazici A. Turkish Large Vocabulary Continuous Speech Recognition by using limited audio corpus // Signal Processing and Communications Applications Conference. – 2012. – P. 1-4.
9. Saon G. Jen-Tzung Chien. Large-Vocabulary Continuous Speech Recognition Systems: A Look at Some Recent Advances // Signal Processing Magazine. – 2012. – Vol. 29, No 6. – P. 18-33.
10. Byrne W., Hajič J., Ircing P. [and all]. Morpheme Based Language Models for Speech Recognition of Czech // Text, Speech and Dialogue. – 2002. – Vol. 1902. – P. 211-216.
11. Saraswathi S., Geetha T. Morpheme based language model for Tamil speech recognition system // The International Arab Journal of Information Technology. – 2007. – Vol. 4, No. 3. – P. 214-219.
12. Rotovnik T., Maučec M.S., Kačič Z. Large vocabulary continuous speech recognition of an inflected language using stems and endings // Speech Communication. – 2007. – Vol. 49, No. 6. – P. 437-452.

13. *Hong Kai Sze, Tan Tien Ping, Tang Enya Kong, Cheah Yu-N.* Linguistic stem concatenation for malay large vocabulary continuous speech recognition // Research and Development (SCORED), 2010 IEEE Student Conference on. – 2010. – P. 144-148.
14. *Karpov, I. Kipytkova, A. Ronzhin.* Very large vocabulary ASR for spoken Russian with syntactic and morphemic analysis // Proceedings of INTERSPEECH' 2011, Florence. – 2011. – P. 3161-3164, pp. 291-296.
15. *Stas J., Hladek D., Juhar J., Zlacký D.* Analysis of morph-based language modeling and speech recognition in Slovak // Information and communication technologies and services. – Vol. 10, No. 4. – P. 2012, special issue.
16. *Шелепов В.Ю., Ниценко А.В.* К проблеме распознавания слитной речи // Искусственный интеллект. – 2012. – № 4. – С. 272-281.
17. *Шелепов В.Ю., Ниценко А.В.* О некоторых вопросах, связанных с дифонным распознаванием и распознаванием слитной речи // Искусственный интеллект. – 2013. – № 3 – С. 209-216.
18. *Nicenko A.V.* A «by part» method of Russian word speech recognition // Eurasian Journal of Mathematical and Computer Applications. – 2014. – Vol. 1, Issue 2. – P. 102-109.
19. *Зализняк А.А.* Грамматический словарь русского языка. – М.: Русский язык, 1977. – 879 с.
20. *Дорохина Г.В. Павлюкова А.П.* Модуль морфологического анализа слов русского языка // Искусственный интеллект. – 2004. – № 3. – С. 636-642.
21. Патент України № 78806 «Пристрій для збереження і пошуку рядкових величин та спосіб збереження і пошуку рядкових величин» Власник: Інститут проблем штучного інтелекту / Винахідник Дорохіна Г.В. // Промислова власність. – Бюл. № 5. 25.04.2007.
22. *Buribayeva A.K., Sharipbay A.A.* Kazakh Vowel Recognition at the Beginning of Words // Mediterranean Journal of Social Sciences, MCSER Publishing, Rome-Italy. – 2015. – Vol. 6, No 2, S4. – P. 121-127.

REFERENCES

1. *Buribaeva A.K., Dorokhina G.V., Nitsenko A.V., Shelepov V.Yu.* Segmentatsiya i difonnoe raspoznavanie rechevykh signalov [Divonne segmentation and recognition of speech signals], *Trudy SPIIRAN* [Proceedings of SPIIRAS], 2013, Issue 31, pp. 20-42.
2. *Mporas I., Ganchev T. and Fakotakis N.* Speech segmentation using regression fusion of boundary predictions, *Computer Speech and Language*, 2010, Vol. 24, No. 2, pp. 273-288.
3. *Gómez J.A., Calvo M.* Improvements on Automatic Speech Segmentation at the Phonetic Level, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 2011, Vol. 7042, pp. 557-564.
4. *Natarajan V.A., Jothilakshmi S.* Segmentation of Continuous Speech into Consonant and Vowel Units using Formant Frequencies, *International Journal of Computer Applications*, 2012, Vol. 56, No. 15, pp. 24-27.
5. *Yuan J., Ryant N., Liberman M. [and all].* Automatic phonetic segmentation using boundary models, *Proceedings of Interspeech*, 2013, pp. 2306-2310.
6. *Pač Z., Mizera P., Pollak P.* Phonetic Segmentation Using KALDI and Reduced Pronunciation Detection in Causal Czech Speech, *Text, Speech, and Dialogue*, 2015, Vol. 9302, pp. 433-441.
7. *Zhijian Ou, Xiao Ji.* A study of large vocabulary speech recognition decoding using finite-state graphs, *Chinese Spoken Language Processing (ISCSLP), 7th International Symposium*, 2010, pp. 123-128.
8. *Susman D., Kopru S., Yazici A.* Turkish Large Vocabulary Continuous Speech Recognition by using limited audio corpus, *Signal Processing and Communications Applications Conference*, 2012, pp. 1-4.
9. *Saon G. Jen-Tzung Chien.* Large-Vocabulary Continuous Speech Recognition Systems: A Look at Some Recent Advances, *Signal Processing Magazine*, 2012, Vol. 29, No 6, pp. 18-33.
10. *Byrne W., Hajič J., Ircing P. [and all].* Morpheme Based Language Models for Speech Recognition of Czech, *Text, Speech and Dialogue*, 2002, Vol. 1902, pp. 211-216.
11. *Saraswathi S., Geetha T.* Morpheme based language model for Tamil speech recognition system, *The International Arab Journal of Information Technology*, 2007, Vol. 4, No. 3, pp. 214-219.
12. *Rotovnik T., Maučec M.S., Kačič Z.* Large vocabulary continuous speech recognition of an inflected language using stems and endings, *Speech Communication*, 2007, Vol. 49, No. 6, pp. 437-452.
13. *Hong Kai Sze, Tan Tien Ping, Tang Enya Kong, Cheah Yu-N.* Linguistic stem concatenation for malay large vocabulary continuous speech recognition, *Research and Development (SCORED), 2010 IEEE Student Conference on*, 2010, pp. 144-148.

14. Karpov, I. Kipytkova, A. Ronzhin. Very large vocabulary ASR for spoken Russian with syntactic and morphemic analysis, *Proceedings of INTERSPEECH' 2011, Florence, 2011*, pp. 3161-3164, pp. 291-296.
15. Stas J., Hladek D., Juhar J., Zlacky D. Analysis of morph-based language modeling and speech recognition in Slovak, *Information and communication technologies and services*, Vol. 10, No. 4, pp. 2012, special issue.
16. Shelepov V.Yu., Nitsenko A.V. К проблеме распознавания слитной речи [To the problem of continuous speech recognition], *Iskusstvennyy intellekt [Artificial intelligence]*, 2012, No. 4, pp. 272-281.
17. Shelepov V.Yu., Nitsenko A.V. О некоторых вопросах, связанных с дифонным распознаванием и распознаванием слитной речи [On some issues related to difonim, recognized the existence and the continuous speech recognition], *Iskusstvennyy intellekt [Artificial intelligence]*, 2013, No. 3, pp. 209-216.
18. Nicenkov A.V. A «by part» method of Russian word speech recognition, *Eurasian Journal of Mathematical and Computer Applications*, 2014, Vol. 1, Issue 2, pp. 102-109.
19. Zaliznyak A.A. Grammaticheskiy slovar' russkogo yazyka [Grammatical dictionary of the Russian language]. Moscow: Russkiy yazyk, 1977, 879 p.
20. Dorokhina G.V. Pavlyukova A.P. Modul' morfologicheskogo analiza slov russkogo yazyka [The module of morphological analysis of words of the Russian language], *Iskusstvennyy intellekt [Artificial intelligence]*, 2004, No. 3, pp. 636-642.
21. Patent Ukraini № 78806 «Pristriy dlya zberezheniya i poshuku ryadkovikh velichin ta sposib zberezheniya i poshuku ryadkovikh velichin» Vlasnik: Institut problem shtuchnogo intelektu [The patent of Ukraine No. 78806 "Device for storage and retrieval of string values and a way of saving and finding a string value" Owner: Institute of problems of artificial intelligence], *Vinakhidnik Dorokhina G.V., Promislova vlasnist' [Industrial property]*. Byul. No. 5. 25.04.2007.
22. Buribayeva A.K., Sharipbay A.A. Kazakh Vowel Recognition at the Beginning of Words, *Mediterranean Journal of Social Sciences, MCSER Publishing, Rome-Italy*, 2015, Vol. 6, No 2, S4, pp. 121-127.

Статью рекомендовал к опубликованию к.т.н. Н.Н. Сажок.

Шелепов Владислав Юрьевич – Институт проблем искусственного интеллекта; e-mail: vladislav.shelepov2012@yandex.ua; Донецк, Артёма, 118Б; главный научный сотрудник; тел.: +380623113424; д.ф.-м.н., профессор.

Ниценко Артём Владимирович – e-mail: nav_box@mail.ru; научный сотрудник.

Shelepov Vladislav Jurievich – Institute of Artificial Intelligence; e-mail: vladislav.shelepov2012@yandex.ua; 118b, Artyoma street, Donetsk, 83048, DPR; phone: +380623113424; the department of speech recognition; leading researcher; professor.

Nitsenko Artem Vladimirovich – e-mail: nav_box@mail.ru; the department of speech recognition; researcher.

УКД 004.93

И.В. Лошкарёв, Я.М. Демяненко

ОТСЛЕЖИВАНИЕ ОБЛАСТИ ЛИЦА ДЛЯ БЕСКОНТАКТНОГО МОНИТОРИНГА СОСТОЯНИЯ ЧЕЛОВЕКА*

На сегодняшний день системы мониторинга функционального состояния человека являются одним из перспективных направлений развития систем видеонаблюдения. Необходимость такого мониторинга возникает в ситуациях, связанных с длительной монотонной деятельностью лиц, ошибки, в работе которых могут привести к катастрофическим

* Работа выполнена при финансовой поддержке ЮФУ, проект № 213.01 - 07.2014 / 07ПЧВГ.