Ш

УДК 004.89:004.93

В. Ю. Шелепов, А. В. Ниценко
Государственное учреждение «Институт проблем искусственного интеллекта», г. Донецк
83048, г. Донецк, ул. Артема, 118-б

АНТИЭТАЛОНЫ И ПОИСК КЛЮЧЕВЫХ СЛОВ МЕТОДАМИ DTW-РАСПОЗНАВАНИЯ

V. Y. Shelepov, A. V. Nicenko
Public institution «Institute of Problems of Artificial intelligence», Donetsk city
83048, Donetsk, Artema str., 118-b

ANTI-TEMPLATES AND SERCHING OF KEYWORDS WITH DTW-RECOGNITION METHODS

В. Ю. Шелєпов, А. В. Ниценко
Державна установа «Інститут проблем штучного інтелекту», м. Донецьк
83048, м. Донецьк, вул. Артема, 118-б

АНТІЕТАЛОНИ ТА ПОШУК КЛЮЧОВИХ СЛІВ МЕТОДАМИ DTW-РОЗПІЗНАВАННЯ

В работе описываются разработанные авторами методы поиска заранее заданного ключевого слова в произвольной слитно произнесенной русской фразе. Метод состоит в DTW-распознавании всех речевых отрезков записанной фразы, содержащих столько глухих фрагментов, сколько их в ключевом слове. Эталон ключевого слова создается голосом, для неключевых слов предлагается использовать небольшое количество усредненных антиэталонов. За счет малого количества антиэталонов достигается значительное ускорение процесса поиска ключевого слова.
Ключевые слова: усреднение эталонов, антиэталон, сегментация, глухой фрагмент, дифонное DTW-распознавание.

The paper describes authors' methods of keyword searching in a continuously spoken Russian phrase. The method consists of DTW-recognition of all speech segments of the recorded phrase containing as many unvoiced fragments as there are in the keyword. The keyword template is created by voice, for non-key words it is suggested to use a small amount of averaged anti-templates. Due to the small number of anti- templates, a significant acceleration in the keyword searching process is achieved.
Key words: speech template averaging, anti-template, speech segmentation, unvoiced fragment, diphone DTW-recognition.

У роботі описуються розроблені авторами методи пошуку заздалегідь заданого ключового слова в довільній злито вимовленій російській фразі. Метод полягає в DTW-розпізнаванні всіх мовних відрізків записаної фрази, що містять стільки глухих фрагментів, скільки їх в ключовому слові. Еталон ключового слова створюється голосом, для неключових слів пропонується використовувати невелику кількість усереднених антіеталонів. За рахунок малої кількості антіеталонів досягається значне прискорення процесу пошуку ключового слова.
Ключові слова: усереднення еталонів, антіеталон, сегментація, глухий фрагмент, діфонне DTW-розпізнавання.

В статье излагается предлагаемый авторами способ построения антиэталонов и их использование для анализа произвольной слитно произнесенной русской фразы с целью выяснить, содержит ли она наперед заданное слово, которое в этой связи именуется ключевым. Этой проблеме посвящены работы [1-21]. Подчеркнем в этой задаче три момента: 1) должна быть минимизирована так называемая «ошибка пропуска цели», когда при фактическом наличии ключевого слова, программа сообщает, что оно не найдено; 2) должна быть минимизирована так называемая «ошибка ложной тревоги», когда программа ошибочно сообщает о наличии ключевого слова; 3) распознавания всех других слов во фразе не предполагается.

Частным случаем рассматриваемой проблемы является задача определения ключевого слова среди произвольных отдельно произносимых русских слов (используемая словарная база *D* – словарь А. А. Зализняка [22]). Различать неключевые слова при этом не нужно. Поэтому целесообразно использовать для них небольшое количество усредненных эталонов, которые уместно назвать антиэталонами ключевого слова. Ввиду малого количества антиэталонов это заведомо ускорит процедуру.

1 Построение эталона и системы антиэталонов ключевого слова

Термин «антиэталон» введен в работе [23]. Есть S_0 – исходный список слов, из которого выбирается ключевое слово. Для него создается голосовой эталон, который также будем называть ключевым. Все остальные эталоны, упоминаемые ниже, синтезируются из эталонов дифонной базы. В специальном поле задается число *n* неключевых слов от начала списка, усреднение эталонов которых дает первый антиэталон «0» (в нашей программе *n*=10, по поводу усреднения эталонов см. [24], [25]). Для словаря S_0 создается дерево транскрипций. С использованием этого дерева вычисляются DTW-расстояния всех слов списка S_0 до эталонов «0» и ключевого, то есть фактически осуществляется DTW-распознавание слов из S_0 на указанных эталонах. Результаты сохраняются в массиве *T*, состоящем из строк вида: «слово – DTW-расстояние до ключевого эталона – DTW-расстояние до эталона “0”». Затем отбираются те неключевые слова, которые ошибочно распознаны как ключевое слово, (для них ключевой эталон оказался ближе, чем эталон «0»). Полученный список ошибок S_1 упорядочивается по возрастанию DTW-расстояний до ключевого эталона. Путем усреднения *n* первых эталонов слов из списка S_1 создается второй антиэталон «1». Вычисляются DTW-расстояния всех слов списка S_0 до антиэталона «1». Они сопоставляются с таблицей *T* и для каждого слова выбирается минимум из трех расстояний. Далее отбираются те неключевые слова, для которых указанный минимум оказался соответствующим ключевому слову. Получается список ошибок S_2 , который, очевидно, является частью списка S_1. С ним делается то же, что и со списком S_1 и так далее. Эта часть алгоритма заканчивается, когда возникает список S_{k-1} , число слов в котором меньше *n*. Далее алгоритм работает со списком S_{k-1} плюс ключевое слово. Усреднением эталонов слов из списка S_{k-1} создается антиэталон «k-1». С эталонами «k-1» и ключевым осуществляется распознавание всех слов списка S_0 и создается очередной список ошибок S_k. Если количество слов в S_k оказывается меньше, чем в S_{k-} , делаем с ним то же, что с S_{k-1}. Если количество слов в S_k такое же, как в S_{k-1}., то алгоритм заканчивается. Рисунок 1 представляет блок-схему алгоритма построения антиэталонов.

Построение антиэталонов вынесено нами в отдельную подпрограмму. Для проверки результата по кнопке «По эталонам» осуществляется распознавание слов списка S_0 на ключевом эталоне и построенных антиэталонах. Многочисленные эксперименты показывают, что в большинстве случаев все неключевые слова классифицируются правильно, то есть оказываются ближе к одному из антиэталонов. Если

ошибки и возникают, то даже для больших словарей S_0 их количество составляет лишь несколько единиц. При анализе произвольных фраз с целью обнаружения ключевого слова это делает ошибку ложной тревоги малой.

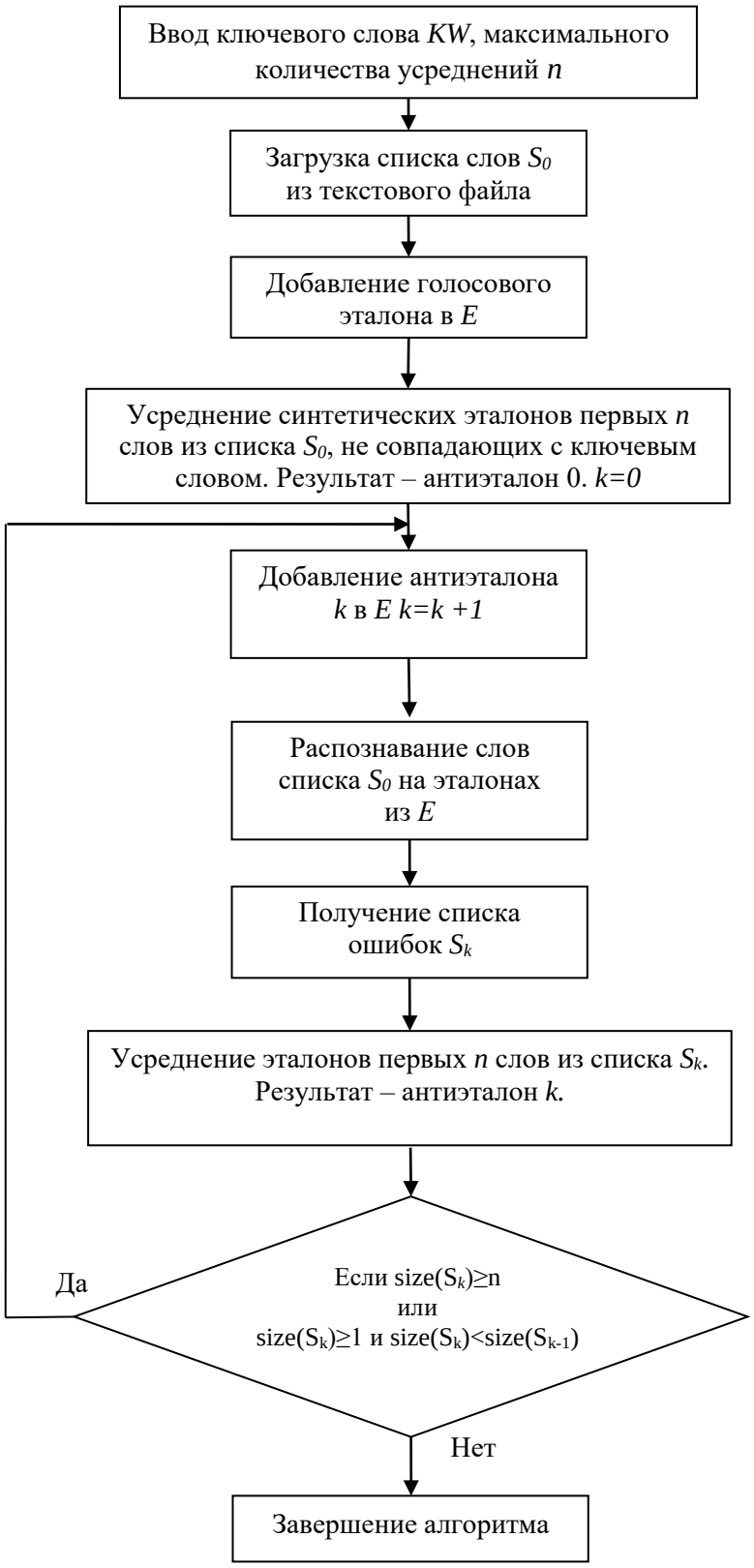

Рисунок 1 – Блок-схема работы алгоритма построения антиэталонов

2 Поиск ключевого слова

По транскрипции определяем количество глухих фрагментов в заданном ключевом слове. Отметим, что глухой фрагмент может состоять из нескольких сегментов сегментации, отвечающих соседним глухим взрывным и фрикативным звукам. Пусть в ключевом слове N глухих фрагментов. При этом если отдельно произносимое слово начинается с глухого взрывного звука, то сегментация не сопоставляет ему отдельный сегмент и количество глухих фрагментов заведомо оказывается на единицу меньше количества глухих звуков. Выбираем из исходной словарной базы D все слова, транскрипции которых обнаруживают N глухих фрагментов. Пусть S_0 – полученный словарь. По нему строим систему антиэталонов, как это описано в п. 1. Пусть E – совокупность эталона и антиэталонов для ключевого слова.

Если в записанном речевом сигнале глухих фрагментов оказалось меньше, чем N, то программа сразу выдает сообщение, что ключевое слово не найдено. В противном случае выделяем от начала записанного сигнала интервал до первого сегмента N-го глухого фрагмента включительно и проводим распознавание этого речевого отрезка на множестве E. Будем теперь последовательно добавлять в конце и убирать в начале упомянутого интервала отрезки сегментации, так чтобы получаемые интервалы по-прежнему содержали ровно N глухих фрагментов. Будем проводить распознавание получаемых интервалов на множестве эталонов E. Когда этот процесс закончен, рассматриваем интервал от конца первого глухого фрагмента до первого сегмента $N+1$-го глухого фрагмента включительно и делаем с ним то же, что и с первоначальным интервалом. Если ключевое слово заканчивается глухим звуком, то, ввиду возможности его озвончения во фразе, нужно повторить процедуру с заменой N на N-1. Процедура заканчивается, когда мы закончили распознавание на всех речевых отрезках, содержащих N (а в только что оговоренном случае и N -1) глухих фрагментов.

Если в ключевом слове нет глухих звуков, то выделяются и распознаются на множестве эталонов E интервалы речевого сигнала, образуемые сегментами, между началом и концом первого голосового фрагмента. При этом, поскольку выполняется дифонное распознавание, то в случае, когда выделенный интервал содержит только один отрезок сегментации (так что дифонов в нем нет), этот интервал при распознавании пропускается. Далее, если не достигнут конец сигнала, выделяется следующий голосовой фрагмент и описанная процедура повторяется. Таким образом, обрабатываются все голосовые фрагменты в сигнале.

Рисунки 2 и 3 представляют блок-схемы алгоритмов распознавания ключевого слова. Первый – в случае, когда в ключевом слове есть глухие звуки, а второй – когда глухих звуков в нем нет. В этих блок-схемах использованы следующие обозначения:

W – список сегментов. End – номер конечной метки сигнала.

N – количество глухих фрагментов в заданном ключевом слове.

$F(n)$ – n-й глухой фрагмент. N_F – количество глухих фрагментов во фразе.

L_b – номер метки начала распознаваемого интервала, L_e – номер метки конца распознаваемого интервала. $W(L_b)$ – сегмент с начальной меткой L_b.

V – голосовой фрагмент из одного или нескольких сегментов.

V_b – номер начальной метки для V. V_e – номер конечной метки для V.

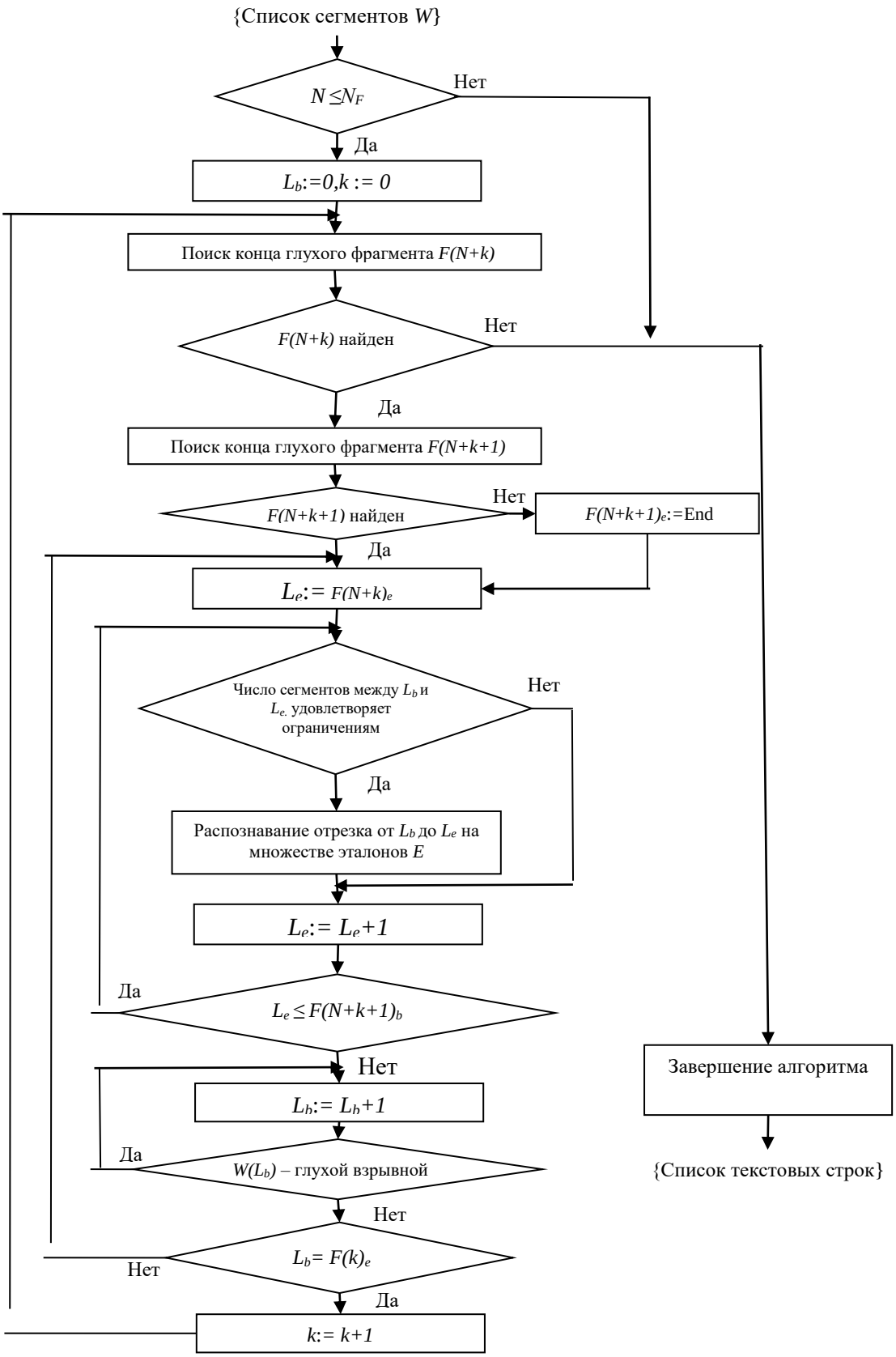

Рисунок 2 – Блок-схема алгоритма поиска ключевого слова для случая, когда в нем есть глухие звуки

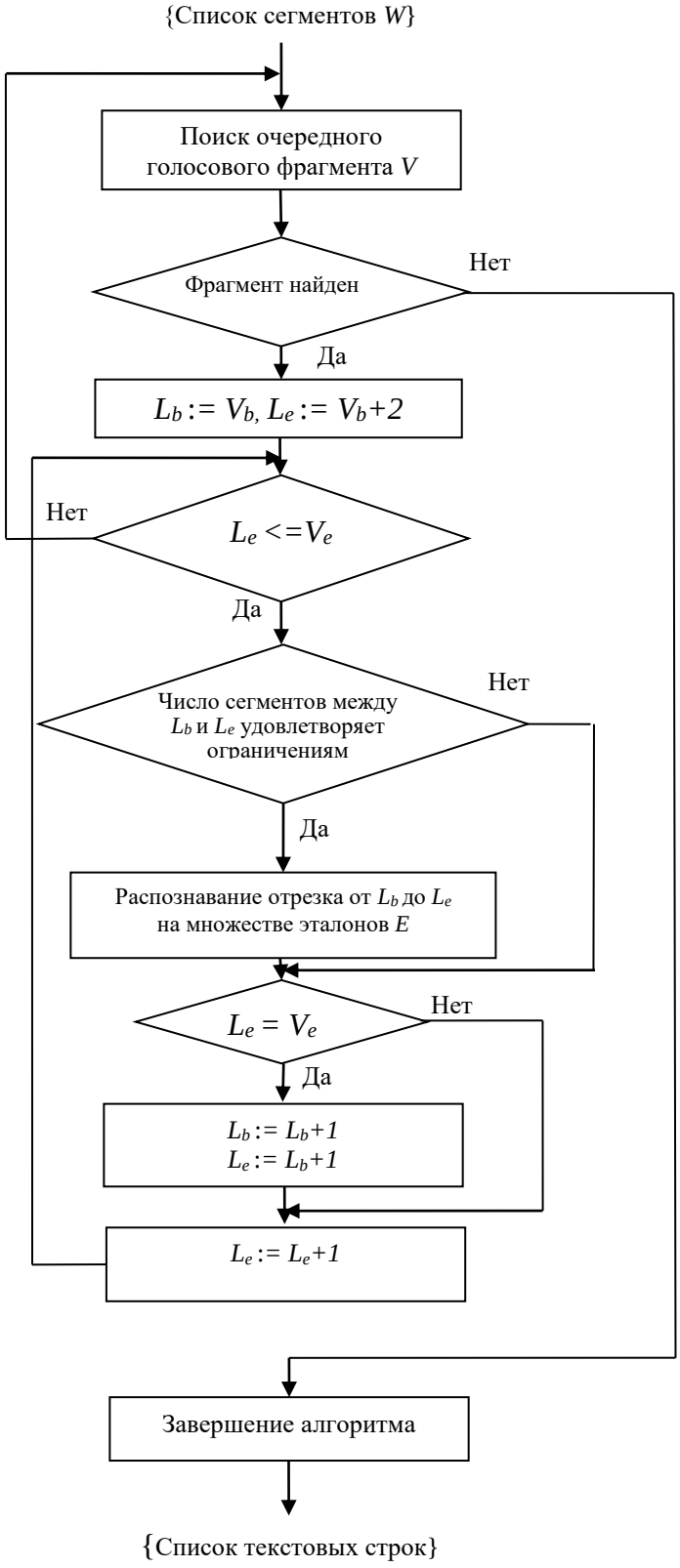

Рисунок 3 – Блок-схема алгоритма поиска ключевого слова для случая, когда в нем нет глухих звуков

В систему добавлена функция учета длины ключевого слова при выборе интервалов распознавания. Программа определяет количество отрезков сегментации в ключевом слове. Описанное выше распознавание проводится только на интервалах анализируемой фразы, содержащих близкое количество отрезков (на случай возможных ошибок сегментации). Числа, указывающие, насколько больше и насколько меньше отрезков может присутствовать, задаются в специальных окнах. В нашей программе это +1 и -4.

Если результатом хотя бы одного из распознаваний является ключевое слово, то это слово считается найденным и программа выдает соответствующее сообщение. В противном случае она выдает сообщение «Ключевое слово не найдено». Малое число ошибок пропуска цели обеспечивается использованием голосового эталона для ключевого слова.

Рисунки 4, 5 иллюстрируют результат поиска ключевого слова «сокурсница». При этом рисунок 5 – общее окно программы распознавания слитной речи.

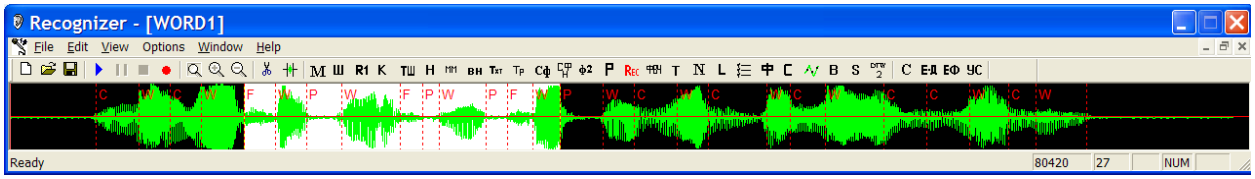

Рисунок 4 – Визуализация фразы «Моя сокурсница уже сдала экзамен»

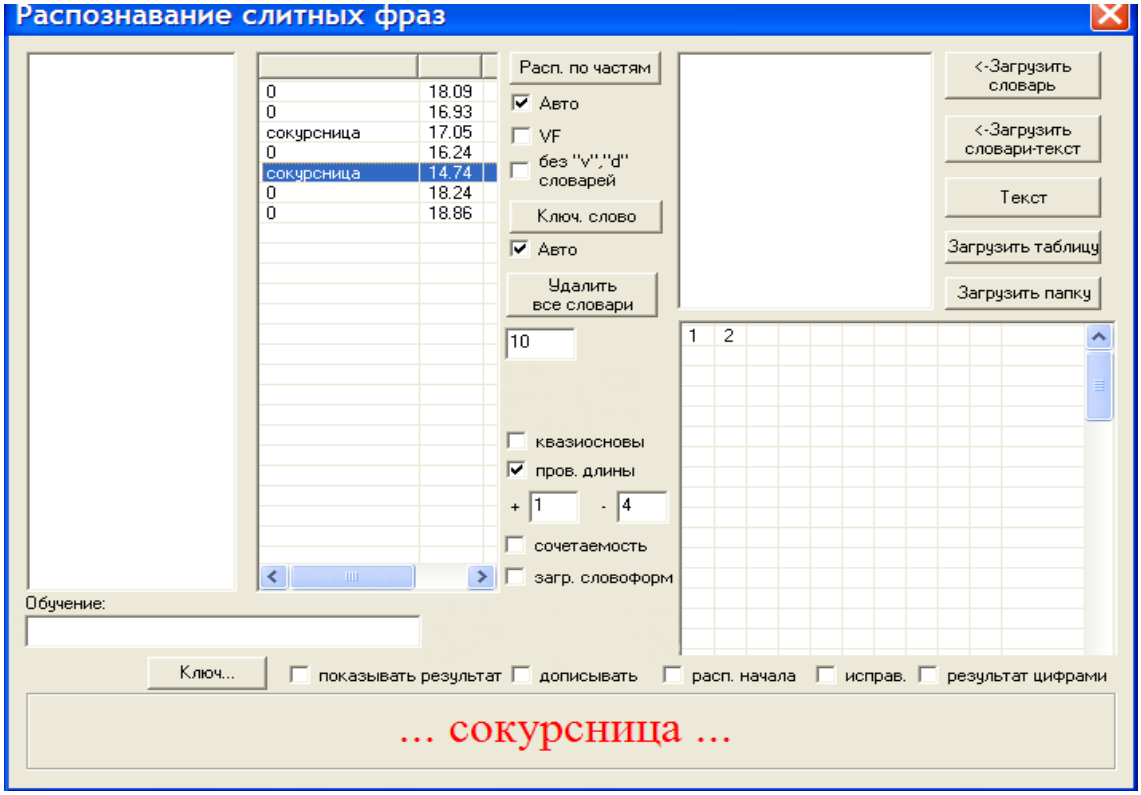

Рисунок 5 – Окно программы с результатом распознавания ключевого слова

Отметим, что в результате выделения в списке кандидатов (среднее вертикальное поле) среди строк со словом «сокурсница» той строки, в которой DTW-расстояние минимально, автоматически выделен отрезок ключевого слова.

Результаты настоящей работы ограничены двумя моментами: 1) среди произносимых слов не должно быть таких, для которых ключевое слово является строгим фонетическим вложением; в противном случае соответствующие части объемлющих слов будут ошибочно восприниматься как ключевое слово; 2) описанная система реализована как однодикторная, хотя авторы имеют примеры успешного распознавания ключевых слов у разных дикторов при одной и той же дифонной базе. Добиться распознавания описанными методами ключевых слов в речи произвольного диктора – предмет отдельного исследования.

Список литературы

1. Automatic recognition of keywords in unconstrained speech using hidden Markov models [Текст] / J.G. Wilpon, L.R. Rabiner, C.H. Lee, E.R. Goldman // IEEE Transactions on Acoustics, speech, and Signal Processing. – 1990. – Vol. 38, no. 11. – P. 1870–1878.
2. James D. A. A fast lattice-based approach to vocabulary independent wordspotting [Текст] / D. A. James, S. J. Young // Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing. – 1994. – Vol. 1. – P. 377–380.
3. Phoneme based acoustics keyword spotting in informal continuous speech [Текст] / I. Szöke, P. Schwarz, P. Matějka, L. Burget, M. Karafiát, J. Černocký // Text, Speech and Dialogue: 8th International Conference, TSD 2005, (Karlovy Vary, Czech Republic, September 12–15, 2005). – Proceedings, vol. 3658 of Lecture Notes in Computer Science. – P. 302–309.
4. Improvement of rejection performance of keyword spotting using anti-keywords derived from large vocabulary considering acoustical similarity to keywords [Текст] / M. Yamada, M. Naito, T. Kato, H. Kawai // Proceedings of the 9th European Conference on Speech Communication and Technology, (Lisbon, Portugal, September 2005). – P. 1445–1448.
5. Харламов А. А. ключевых слов в потоке слитной речи на основе нейросетевых технологий [Текст] / А. А. Харламов, Э. Г. Кнеллер // Нейрокомпьютеры. Разработка и применение. – 2005. – № 8–9. – С. 88–97.
6. Киселёв В. В. Автоматический поиск ключевых слов в непрерывном потоке речи на основе технологии «распознавание через синтез» [Текст] / В. В. Киселёв, А. О. Таланов // Труды международной конференции «Диалог 2006», (Бекасово, 31 мая 4 июня 2006 г). – С. 214–219.
7. Karjigi V. Identification of stop consonants for acoustic keyword spotting in continuous speech [Текст] / V. Karjigi, B. Patel, P. Rao // Proc. of Wireless Personal Multimedia Communications, (Jaipur, India, 2007).
8. Дмитриев В. Т. Дикторонезависимая система автоматического поиска ключевых слов в потоке слитной речи, устойчивая к акустическим шумам [Текст] / В. Т. Дмитриев, И. В. Баландин // Вестник РГРТУ. – 2008. – № 2 (выпуск 24). – С. 15–19.
9. Park A. Unsupervised pattern discovery in speech [Текст] / A. Park, J. Glass // IEEE Trans. ASLP. – 2008. – № 6(1). – P. 1558–1569.
10. Nouza J. Fast keyword spotting in telephone speech [Текст] / J. Nouza, J. Silovsky // Radioengineering. – 2009. – №. 18(4). – P. 665–670.
11. Zhang Y. Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams [Текст] / Y. Zhang, J. Glass. // Proceedings of ASRU. – 2009. – P. 398–403.
12. Jansen A. Point Process Models for Spotting Keywords in Continuous Speech [Текст] / A. Jansen, P. Niyogi. // IEEE transactions on audio, speech, and language processing. – 2009. – Vol. 17, № 8. – P. 1457–1470.
13. Цзинбинь Я. Поиск ключевых слов с использованием решетки слогов [Текст] / Я. Цзинбинь, И. Э. Хейдоров, Р. М. Алиев // Сборник «Компьютерная лингвистика и интеллектуальные технологии». – 2009. – С. 1–5.
14. Гладышев К. К. Система поиска ключевых слов в непрерывном речевом потоке [Текст] / К. К. Гладышев // Естественные и технические науки. – 2009. – № 1. – С. 242–244.
15. Пилипенко В. В. Распознавание ключевых слов в потоке речи при помощи фонетического стенографа [Текст] / В. В. Пилипенко // Искусственный интеллект. – 2009. – № 4. – С. 220–224.
16. Improved keyword spotting system in weighted finite-state transducer framework [Текст] / Y. Guo, Z. Zhang, T. Li, J. Pan, Y. Yan // Journal of Computational Information Systems. – 2013. – vol. 9, № 12. – P. 4947–4954.

17. Баландин И. В. Модификация гибридной модели речевосприятия для алгоритма поиска ключевых слов в потоке слитной речи [Текст] / И. В. Баландин // Вестник РГРТУ. – 2013. – № 4 – 1 (46). – С. 81–84.
18. Hartmann W. Comparing Decoding Strategies for Subword-based Keyword Spotting in Low-Resourced Languages [Текст] / W. Hartmann, Viet-Bac Le, A. Messaoudi. // Interspeech. –2014. – P. 2764–2768.
19. Chen G. Small-footprint keyword spotting using deep neural networks / G. Chen, C. Parada, G. Heigold // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. – 2014. – P. 4087–4091.
20. Tetariy E. Cross-language phoneme mapping for phonetic search keyword spotting in continuous speech of under-resourced languages [Текст] / E. Tetariy,Y. Bar-Yosef, V. Silber-Varod. // Artificial Intelligence Research. – 2015. – vol. 4, № 2. – P. 72–82.
21. A Russian Keyword Spotting System Based on Large Vocabulary Continuous Speech Recognition and Linguistic Knowledge [Текст] / V. Smirnov, D. Ignatov, M. Gusev, M. Farkhadov, N. Rumyantseva, M. Farkhadova // Journal of Electrical and Computer Engineering. – Vol. 2016, Article ID 4062786, 9 pages, 2016.
22. Зализняк А. А. Грамматический словарь русского языка / А. А. Зализняк. – М. : Русский язык, 1980. – 879 с.
23. Федоров Е. Е. Защита речевых распознавателей от шума и посторонней речи [Текст] / Е. Е.Федоров, В. Ю. Шелепов // Искусственный интеллект. – 2001. – № 3. – С. 584–587.
24. Сегментация и дифонное распознавание речевых сигналов [Текст] / А. К. Бурибаева, Г. В. Дорохина, А. В. Ниценко, В. Ю. Шелепов // Труды СПИИРАН. – 2013. – № 31. – С. 20–42.
25. Шелепов В. Ю. Сегментация и дифонное распознавание речи [Текст] / В. Ю. Шелепов, А. В. Ниценко. – Донецк : ГУ ИПИИ, 2015. – 231 с.

References

1. Wilpon J. G., Rabiner L. R., Lee C. H., and Goldman E. R. Automatic recognition of keywords in unconstrained speech using hidden Markov models. *IEEE Transactions on Acoustics, speech, and Signal Processing*, vol. 38, no. 11, pp. 1870–1878, 1990.
2. James D. A., Young S. J. A fast lattice-based approach to vocabulary independent wordspotting. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1994, vol. 1, pp. 377–380.
3. Szöke I., Schwarz P., Matějka P., Burget L., Karafiát M., and Černocký J. Phoneme based acoustics keyword spotting in informal continuous speech. *Text, Speech and Dialogue: 8th International Conference*, TSD 2005, Karlovy Vary, Czech Republic, September 12–15, 2005. Proceedings, vol. 3658 of Lecture Notes in Computer Science, pp. 302–309.
4. Yamada M., Naito M., Kato T., Kawai H. Improvement of rejection performance of keyword spotting using anti-keywords derived from large vocabulary considering acoustical similarity to keywords. *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, September 2005.
5. Harlamov A. A., Kneller G. Raspoznavanie kljuchevyh slov v potoke slitnoj rechi na osnove nejrosetevyh tehnologij [Key words in the flow of solid speech on the basis of neural network technologies]. *Nejrokomp'jutery. Razrabotka i primenenie* [Neurocomputers. Development and application], no. 8–9, 2005, pp. 88–97.
6. Kiseljov V. V., Talanov A. O. Avtomaticheskij poisk kljuchevyh slov v nepreryvnom potoke rechi na osnove tehnologii “raspoznavanie cherez sintez” [Automatic search of key words in a continuous stream of speech on the basis of technology "recognition through synthesis"]. *Trudy mezhdunarodnoj konferencii “Dialog 2006”* [Proceedings of the international conference "Dialogue 2006"], Bekasovo, May 31 – June 4, 2006, pp.214–219.
7. Karjigi V., Patel B., Rao P. Identification of stop consonants for acoustic keyword spotting in continuous speech [Identification of stop consonants for an acoustic keyword spotting in continuous speech]. *Proc. of Wireless Personal Multimedia Communications* [Proc. of Wireless Personal Multimedia Communications], Jaipur, India, 2007.
8. Dmitriev V. T., Balandin I. V. Diktoronezavisimaja sistema avtomaticheskogo poiska kljuchevyh slov v potoke slitnoj rechi, ustojchivaja k akusticheskim shumam [A dictorone-independent system for automatic search for key words in a stream of solid speech, resistant to acoustic noise]. *Vestnik RGRTU* [Vestnik RGRTU]. no. 2 (iss. 24). Rjazan, 2008. pp. 15–19.
9. Park A., Glass J. Unsupervised pattern discovery in speech. *IEEE Trans.* ASLP, 2008, 6(1), pp. 1558–1569.

10. Nouza J., Silovsky J. Fast keyword spotting in telephone speech. *Radioengineering*, 2009, no. 18(4), pp. 665–670.
11. Zhang Y., Glass J. Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams. *Proceedings of ASRU*, 2009, pp. 398–403.
12. Jansen A., Niyogi P. Point Process Models for Spotting Keywords in Continuous Speech. *IEEE transactions on audio, speech, and language processing*, 2009, vol. 17, no. 8, pp.1457–1470.
13. Czinbin J., Hejdorov. I., Aliev R. M. Poisk kljuchevyh slov s ispol'zovaniem reshetki slogov [Search for keywords using the lattice of syllables]. *Sbornik "Komp'juternaja lingvistika i intellektual'nye tehnologii"* [Compilation "Computer Linguistics and Intellectual Technologies"], 2009, pp. 1–5.
14. Gladyshev K. K. Sistema poiska kljuchevyh slov v nepreryvnom rechevom potoke [System of search for key words in a continuous speech flow]. *Estestvennye i tehnicheskie nauki* [Natural and technical sciences], 2009, no. 1, pp. 242–244.
15. Pilipenko V. V. Raspoznavanie kljuchevyh slov v potoke rechi pri pomoschi foneticheskogo stenografa [Recognition of key words in the flow of speech with the help of a phonetic stenographer]. *Iskusstvennyj intellekt* [Artificial Intelligence], 2009, no. 4, pp. 220–224.
16. Guo Y., Zhang Z., Li T., Pan J., Yan Y. Improved keyword spotting system in weighted finite-state transducer framework. *Journal of Computational Information Systems*, 2013, vol. 9, no. 12, pp. 4947–4954.
17. Balandin I. V. Modifikacija gibridnoj modeli rechevosprijatija dlja algoritma poiska kljuchevyh slov v potoke slitnoj rechi [Modification of the hybrid model of speech perception for the algorithm of searching for key words in the flow of solid speech]. *Vestnik RGRTU* [Bulletin of RGRTU], 2013, no. 4–1 (46), pp. 81–84.
18. Hartmann W., Viet-Bac Le, Messaoudi A. Comparing Decoding Strategies for Subword-based Keyword Spotting in Low-Resourced Languages. *Interspeech*, 2014, pp. 2764–2768.
19. Chen G., Parada C., Heigold G. Small-footprint keyword spotting using deep neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2014, pp. 4087–4091.
20. Tetariy E., Bar-Yosef Y., Silber-Varod V. Cross-language phoneme mapping for phonetic search keyword spotting in continuous speech of under-resourced languages. *Artificial Intelligence Research*, 2015, Vol. 4, no. 2, pp. 72–82.
21. Smirnov V., Ignatov D., Gusev M., Farkhadov M., Rumyantseva N., Farkhadova M. A Russian Keyword Spotting System Based on Large Vocabulary Continuous Speech Recognition and Linguistic Knowledge. *Journal of Electrical and Computer Engineering*, vol. 2016, Article ID 4062786, 9 pages, 2016.
22. Zaliznyak A.A. *Grammaticheskiy slovar' russkogo yazyka* [Grammatical dictionary of the Russian language], M., Russkiy yazyk, 1980, 879 p.
23. Fedorov E. E., Shelepov V. Ju. Zaschita rechevyh raspoznavatelej ot shuma i postoronnej rechi [Protection of speech recognizers from noise and extraneous speech]. *Iskusstvennyj intellekt* [Artificial Intelligence], 2001, no. 3, pp. 584–587.
24. Buribaeva A. K., Dorohina G. V., Nicenko A. V., Shelepov V. Ju. Segmentacija i difonnoe raspoznavanie rechevyh signalov [Segmentation and diphone recognition of speech signals]. *Trudy SPIIRAN* [Proceedings of SPIIRAS], 2013, no. 31, pp. 20–42.
25. Shelepov V.Ju., Nicenko A.V. Segmentatsiya i difonnoye raspoznavaniye rechi [Segmentation and duplex speech recognition]. *Problems of Artificial Intelligence*, 2015, no. 0(1), pp. 116 – 122.

RESUME

A. V. Nicenko, V. Ju. Shelepov

Anti-templates and serching of keywords with DTW-recognition methods

Background: The present paper is in line with the methods of diphone DTW-recognition, developed by the authors. It is based on the authors' procedure of a priori segmentation of the speech signal which allows a reliable allocation of voiceless fragments in the signal.

Materials and methods: The paper describes the authors' technique of recognizing (searching) a preset word (key-word) in any continuously spoken Russian phrase. A particular case within the problem is searching a key-word among separately spoken Russian words.

We can restrict the recognition with words containing so many voiceless fragments as keyword contains. Since there is no need to distinguish non-key words it is worthwhile using some average templates. We call them “anti-templates of the key-word”. Owing to a small number of anti-templates the procedure is certain to speed up. The keyword template is voice generated. All other templates in the present text are synthesized from diphone database templates.

Results: The proposed method of constructing anti-templates is described in section 1 of the article. Section 2 specifically deals with the recognition of the key-word using the template and anti-templates. The method is applying of DTW-recognition to every speech segment of a recorded phrase, containing as many voiceless fragments as the keyword does. The proximity of a number of segments to the number of sounds in the keyword is taken into account in selecting recognition intervals.

Conclusion: Due to the small number of anti-templates, a significant acceleration in the keyword searching process is achieved.

РЕЗЮМЕ

А. В. Ниценко, В. Ю.Шелепов

Антиэталоны и поиск ключевых слов методами DTW-распознавания

История вопроса. Настоящая работа лежит в русле развиваемых авторами методов дифонного DTW-распознавания и опирается на принадлежащую авторам процедуру априорной сегментации речевого сигнала, которая с высокой надежностью выделяет в сигнале фрагменты-сочетания глухих звуков.

Материалы и методы. В работе описываются разработанные авторами методы распознавания (поиска наличия) заранее заданного (ключевого) слова в произвольной слитно произнесенной русской фразе. Частным случаем этой проблемы является задача определения ключевого слова среди произвольных отдельно произносимых русских слов (используемая словарная база – словарь А.А. Зализняка). Распознавание можно вести на множестве слов, содержащих столько же глухих фрагментов, сколько их содержит ключевое слово. Различать неключевые слова при этом не нужно. Поэтому целесообразно использовать для них небольшое количество усредненных эталонов, которые уместно назвать антиэталонами ключевого слова. Ввиду малого количества антиэталонов это заведомо ускорит процедуру. Эталон ключевого слова создается голосом. Все остальные эталоны, упоминаемые в тексте, синтезируются из эталонов дифонной базы.

Результаты. Предлагаемый метод построения антиэталонов описан в разделе 1 настоящей статьи. Раздел 2 посвящен собственно распознаванию ключевого слова с использованием эталона и антиэталонов. Метод состоит в DTW-распознавании всех речевых отрезков записанной фразы, содержащих столько глухих фрагментов, сколько их в ключевом слове. При выборе интервалов распознавания учитывается также близость количества отрезков сегментации в них к числу звуков в ключевом слове.

Заключение. За счет малого количества антиэталонов достигается значительное ускорение процесса поиска ключевого слова.

Статья поступила в редакцию 26.04.2018.