

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

О РАСПОЗНАВАНИИ РУССКИХ СЛОВ С ИСПОЛЬЗОВАНИЕМ ОБОБЩЕННОЙ ТРАНСКРИПЦИИ

V. Y. Sheleпов, A. V. Nicenko

Public institution «Institute of Problems of Artificial intelligence», Donetsk city
83048, Donetsk, Artema str., 118-b

ON THE RECOGNITION OF RUSSIAN WORDS USING GENERALIZED TRANSCRIPTION

В. Ю. Шелепов, А. В. Ниценко

Державна установа «Інститут проблем штучного інтелекту», м. Донецьк
83048, м. Донецьк, вул. Артема, 118-б

ПРО РОЗПІЗНАВАННЯ РОСІЙСКИХ СЛІВ ІЗ ЗАСТОСУВАННЯМ УЗАГАЛЬНЕНОЇ ТРАНСКРИПЦІЇ

В работе показано, как, выбрав по одному представителю для гласных, звонких согласных, глухих фрикативных и глухих взрывных звуков, и используя их вместе с соответствующими дифонами, а также априорной сегментацией речевого сигнала, организовать предварительное распознавание русских слов из большого словаря. Результатом является во много раз уменьшенный список слов, из которого сказанное слово быстро определяется с помощью вторичного DTW-распознавания теперь уже с полной базой эталонов звуков и дифонов.

Ключевые слова: априорная сегментация, классы звуков речи, транскрипция и обобщенная транскрипция, дифонное DTW-распознавание.

Let's consider classes of vowels, voiced consonants, unvoiced fricatives and unvoiced plosive sounds. The paper shows how by selecting one representative for each and using them together with the corresponding diphones as well as a priori segmentation of the speech signal to organize preliminary recognition of Russian words from a large vocabulary. The result is a much reduced list of words, from which the spoken word is quickly defined with the help of secondary DTW-recognition, now with a full base of templates of sounds and diphones.

Key words: a priori segmentation, classes of speech sounds, transcription and generalized transcription, diphone DTW-recognition.

В роботі показано, як, вибравши по одному представнику для голосних, дзвінких приголосних, глухих фрикативних і глухих вибухових звуків, і використовуючи їх разом з відповідними дифонами, а також апіорною сегментацією мовного сигналу, організувати попереднє розпізнавання російських слів з великого словника. Результатом є у багато разів зменшений список слів, з якого вимовлене слово швидко визначається за допомогою вторинного DTW-розпізнавання тепер вже з повною базою еталонів звуків і дифонів.

Ключові слова: апіорна сегментація, класи звуків мови, транскрипція і узагальнена транскрипція, дифонне DTW-розпізнавання.

В статье описывается способ распознавания отдельно произносимых русских слов с использованием априорной сегментации, обобщенной транскрипции и алгоритма DTW. Родственные по тематике работы [1–15] перечислены в списке литературы.

1 Обобщенная сегментация

Мы используем авторскую априорную сегментацию речевого сигнала [16], [17], результатом которой является разбиение его на отрезки, отвечающие отдельным звукам речи, с одновременной классификацией в рамках широкой фонетической классификации: W – гласный, С – звонкий согласный, F – глухой фрикативный, P – глухой взрывной. В особом положении находятся аффрикаты [ц] и [ч]. Они, как правило, содержат две части: паузообразную P и следующую за ней фрикативную часть F, но иногда одна из этих частей может отсутствовать. Участок звука [ф] или [х] также может содержать одновременно несколько F-отрезков и P-отрезков. Кроме того, надо отметить следующее. В интересах точности определения границ речевого сигнала мы очищаем запись от шума с помощью методов спектрального вычитания. Но при этом система иногда начинает путать при априорной сегментации F-отрезки и P-отрезки.

Далее мы будем пользоваться обобщением описанной сегментации, при котором соседние W-отрезки объединяются в один W-отрезок и то же делается для С, F, P отрезков, и называть такую сегментацию обобщенной.

Общая структура русского слова определяется чередованием гласных, звонких согласных, глухих фрикативных звуков (и аффрикат) и глухих взрывных (паузообразных) звуков. В идеале она может быть выявлена в результате априорной сегментации. Однако эта сегментация, несмотря на все ее достоинства, иногда бывает не безошибочной. В то же время указанная структура независимым образом проясняется в процессе DTW-распознавания с использованием эталонов всего шести звуков [а], [л], [с], [т], [р], [и] и эталонов соответствующих дифонов *ал*, *ас*, Ниже это распознавание называется распознаванием с обобщенной транскрипцией.

2 Обобщенные транскрипции (обобщенные слова)

Слова распознаваемого словаря транскрибируются обычным образом (см. по этому поводу [17]), а затем символы всех гласных заменяются на “а”, символы всех звонких согласных, кроме “р”, заменяются на “л”, символы всех глухих фрикативных заменяются на “с”, символы всех глухих взрывных заменяются на “т”. Для аффрикат используются три варианта: “тс”, “с”, “т”. В случае повторения нескольких одинаковых символов подряд, они заменяются одним. Полученную для слова цепочку символов будем называть его обобщенной транскрипцией или обобщенным словом. Таким образом, например слово “наши” будет иметь обобщенную транскрипцию *ласа*, слово “решётка” – обобщенную транскрипцию *расата*, а имя “Инна” – обобщенную транскрипцию *ала*.

3 Отбор слов по результату сегментации

Если у нас есть некоторый словарь русских слов, произнесено одно из них и выполнена обобщенная сегментация записанного сигнала, то выпишем цепочку соответствующих символов W, С, F и P. Далее, заменив W на а, С на л, F на с, P на т, получим соответствующую обобщенную транскрипцию и осуществим отбор слов исходного словаря, имеющих данную обобщенную транскрипцию. Отметим, что это действительно всего лишь отбор, а не распознавание, то есть вычислительно значительно более простая процедура. Ввиду сказанного в п. 1 имеет смысл отождествлять в результирующей транскрипции все глухие отрезки, используя при этом вместо символа “т” символ “с”. Таким образом, мы будем иметь дело с классификацией *ГЛАСНЫЙ – ЗВОНКИЙ СОГЛАСНЫЙ – ГЛУХОЙ СОГЛАСНЫЙ*.

4 DTW – распознавание обобщенных слов

Независимо осуществляется DTW – распознавание обобщенных слов по эталонам, использующим лишь эталоны звуков [л], [р], [с], [т], а также эталон для гласного, формируемый как результат усреднения эталонов для двух акустически далеких звуков [а] и [и] (по поводу процедуры усреднения см. [16], [17]). Из этих эталонов, а также эталонов соответствующих дифонов путем конкатенации склеиваются эталоны всех полученных обобщенных слов. Так же, как и в случае с сегментацией, на последнем шаге отождествляем все глухие отрезки и используем символ “с” вместо символа “т”.

Итак, записанному сигналу сопоставляется две обобщенные транскрипции: одна, как следствие обобщенной сегментации, другая – как результат DTW-распознавания на множестве обобщенных слов. По каждой из них выделяется список соответствующих слов исходного словаря. Эти два списка объединяются в один. Отметим, что он составляет малую часть исходного словаря. Если он содержит лишь одно слово, то это окончательный результат распознавания. Если же в указанном списке более одного слова, то на этом списке производится распознавание с эталонами, построенными по полным, а не обобщенным транскрипциям этих слов, с использованием эталонов всех чистых звуков и эталонов всех дифонов.

Таким образом, последовательно работают два распознавателя: предварительный и окончательный. Они обеспечивают во времени, близком к реальному, весьма надежный результат при работе с большими словарями, например, со словарем А. А. Зализняка [18], содержащим около 94 тысяч слов в начальных формах.

5 Примеры

Пример 1. На рис. 1 представлена визуализация сигнала для слова “арифметика” с априорной сегментацией, которая ошибочно не выделила отрезок звука [г] (р мягкое).

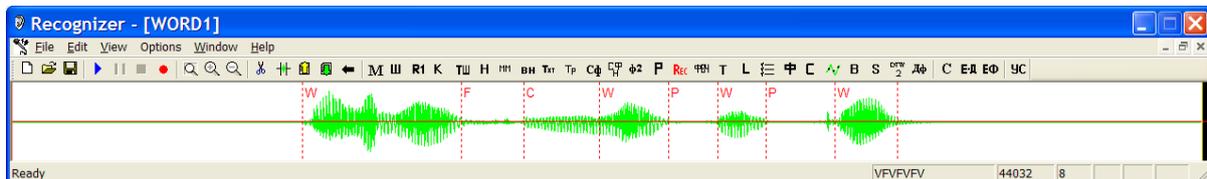


Рисунок 1 – Визуализация сигнала для слова “арифметика” с априорной сегментацией

Таким образом, результат этой сегментации порождает обобщенное слово
асласаса

В то же время результатом распознавания с обобщенной транскрипцией является
арасласаса

и это действительно обобщенная транскрипция слова “арифметика”.

На рис. 2 изображено окно программы-распознавателя с результатами для сигнала, представленного на рис. 1. Обобщенное слово, полученное в результате сегментации, находится в поле, помеченном как “Сегм:”. Аналогичное слово – результат DTW – распознавания с обобщенной транскрипцией по эталонам – в поле, помеченном как “Эталон:”. Результаты распознавания, предварительный и окончательный (с указанием DTW-расстояния до последнего), содержатся в правом нижнем поле.

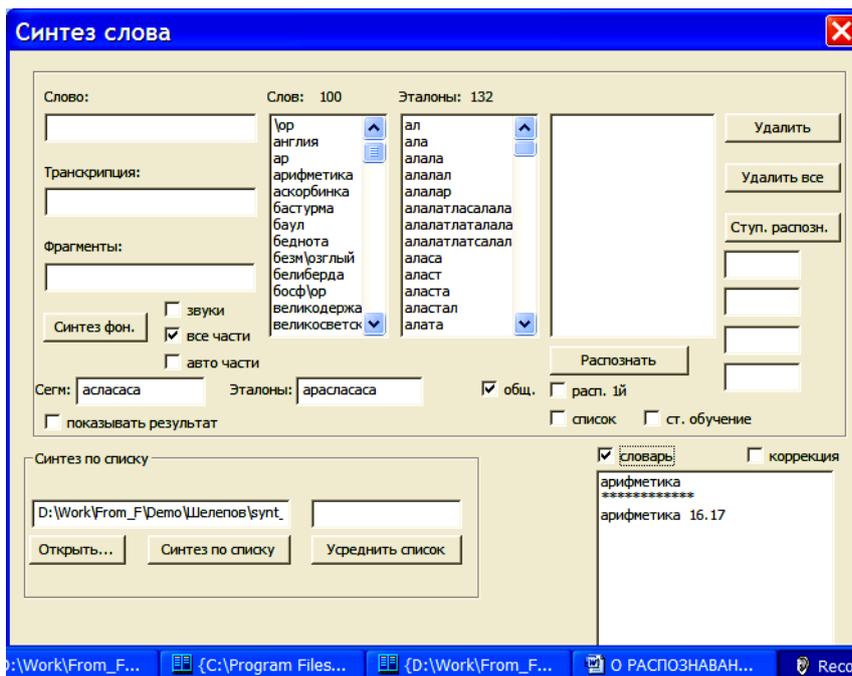


Рисунок 2 – Окно распознавателя с результатами для сигнала на рис. 1

Пример 2. На рис. 3 представлена визуализация сигнала для слова “мышь”.

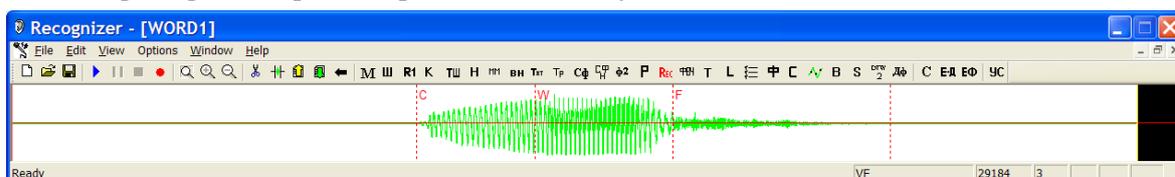


Рисунок 3 – Визуализация сигнала для слова “мышь”

На рис. 4 изображено окно программы-распознавателя с результатами для сигнала, представленного на рис. 3.

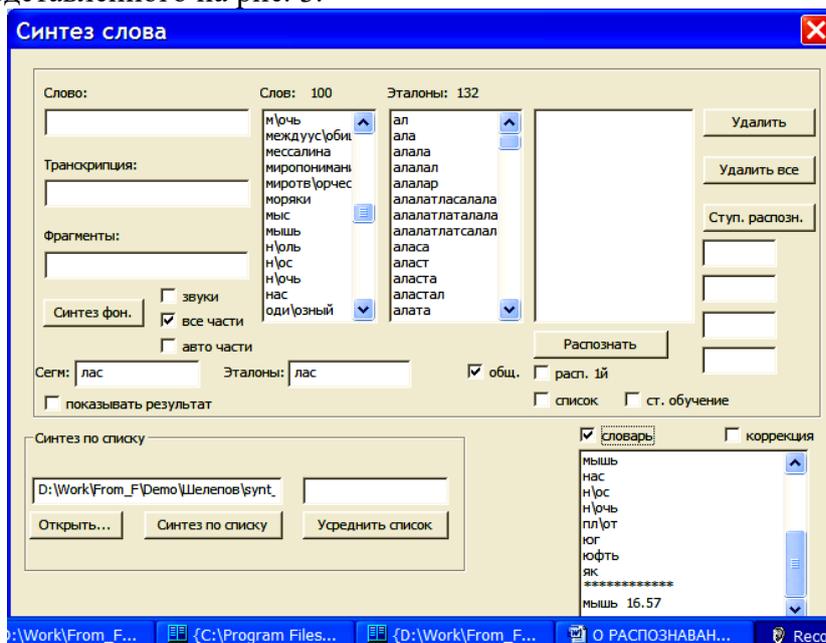


Рисунок 4 – Окно распознавателя с результатами для сигнала на рис. 1

Здесь результаты использования обобщенной сегментации и распознавания с обобщенной транскрипцией совпадают и приводят к списку исходных слов

мышь, нас, нос, ночь, плот, юг, юфть як.

На этом маленьком словаре далее проведено DTW-распознавание с эталонами, построенными по полным транскрипциям при использовании полной базы звуков и дифонов. Оно и определило результат.

Отметим, что для ускорения работы с массивами исходных и обобщенных слов они организуются в виде деревьев.

Выводы

Предложенный метод, использующий предварительное распознавание с обобщенной транскрипцией, позволяет повысить качество и увеличить скорость распознавания больших словарей.

Список литературы

1. Леонтьева А. Б. Моделирование нефонемных речевых элементов и создание альтернативных транскрипций для распознавания спонтанной речи [Текст] / А. Б. Леонтьева, И. С. Кипяткова // Первый междисциплинарный семинар «Анализ разговорной русской речи» (АРЗ - 2007). – 2007. – С. 77–85.
2. Ронжин А. Л. Топологические особенности морфофонемного способа представления словаря для распознавания русской речи [Текст] / А. Л. Ронжин // Вестник компьютерных и информационных технологий. – 2008. – № 9. – С. 12–19.
3. Ронжин А. Л. Сравнительный анализ и оценка моделей словаря для систем распознавания русской речи [Текст] / А. Л. Ронжин // Информационные технологии. – 2009. – № 1. – С. 21–28.
4. Automatic transcription of continuous speech into syllable-like units for Indian languages [Текст] / G. L. Sarada, A. Lakshmi, H. A. Murthy, T. Nagarajan // Sadhana. – 2009. – Vol. 34, Part 2. – P. 221–233.
5. Кипяткова И. С. Эксперименты по распознаванию слитной русской речи с использованием сверхбольшого словаря [Текст] / И. С. Кипяткова, А. А. Карпов // Труды СПИИРАН. – 2010. – № 1 (12). – С. 63–74.
6. Савченко В. В. Методика формирования рабочего словаря в системах автоматического распознавания речи по тематическому файлу в текстовом формате [Текст] / В. В. Савченко, А. В. Савченко // Системы управления и информационные технологии. – 2012. – Т. 48. № 2.2. – С. 284–289.
7. Савченко А. В. Автоматическое построение фонетической транскрипции речи на основе принципа минимума информационного рассогласования [Текст] / А. В. Савченко // Вестник компьютерных и информационных технологий. – 2012. – № 8. – С. 14–19.
8. Efficient DTW-Based Speech Recognition System for Isolated Words of Arabic Language [Текст] / Khalid A. Darabkh, Ala F. Khalifeh, Baraa A. Bathech, and Saed W. Sabah // World Academy of Science, Engineering and Technology International Journal of Electrical and Computer Engineering. – 2013. – Vol. 7, No.5. – P. 586–593.
9. Ду Ц. Отложенная слоговая сегментация при распознавании речи [Текст] / Ц. Ду // Молодежный научно-технический вестник. – 2014. – № 3. – С. 29.
10. Жилияков Е.Г. О сегментации речевых сигналов на однородные отрезки [Текст] / Е. Г. Жилияков, С. П. Белов, А. С. Белов, А. А. Фирсова // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. – 2015. – Т. 34. – № 7-1. – С. 194–199.
11. Shirokova A. Multi-Pronunciation Lexicon for Russian Automatic Speech Recognition [Электронный ресурс] / A. Shirokova, B. Telesnin, V. Rogozhina // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialogue 2016». – Moscow, June 1–4, 2016. – URL : www.dialog-21.ru/media/3425/shirokovaetal.pdf.
12. Червяков Н.И. Проблемы автоматического распознавания слитной речи. Методы обработки исходного речевого сигнала [Текст] / Н. И. Червяков, Н. Н. Кучукова // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. – 2016. – Т. 40. – № 23 (244). – С. 148–155.

13. Шишкина А. Ф. К вопросу об автоматическом распознавании звуков слитной речи [Текст] / А. Ф. Шишкина // Теория. Практика. Инновации. – 2016. – № 4 (4). – С. 18–23.
14. Хеин М. З. Алгоритмы динамического программирования в распознавании речи [Текст] / М. З. Хеин // Auditorium. – 2017. – № 4 (16). – С. 48–52.
15. Худойбердиев Х. А. Модель анализа и сегментации речевого сигнала для послогового распознавания таджикской речи [Текст] / Х. А. Худойбердиев // Вестник Технологического университета Таджикистана. – 2017. – № 4 (31). – С. 85–87.
16. Бурибаева А. К. 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.
17. Шелепов В. Ю. Сегментация и дифонное распознавание речи [Текст] / В. Ю. Шелепов, А. В. Ниценко. – Донецк : ГУ ИПИИ, 2015. – 231 с.
18. Зализняк А.А. Грамматический словарь русского языка [Текст] / А. А. Зализняк. – М. : Русский язык, 1980. – 879 с.
19. Шелепов В. Ю. О распознавании первого звука в слитном речевом отрезке [Текст] / В. Ю. Шелепов, А. В. Ниценко // Проблемы искусственного интеллекта – 2015. – № 0(1). – С. 116–122.

References

1. Leont'yeva A.B., Kipyatkova I.S. Modelirovaniye nefonemnykh rechevykh elementov i sozdaniye al'ternativnykh transkriptsiy dlya raspoznavaniya spontannoy rechi. *Pervyy mezhdistsiplinarnyy seminar «Analiz razgovornoy russkoy rechi»* [The first interdisciplinary seminar "Analysis of spoken Russian speech"] (AR3 - 2007), 2007, pp. 77-85.
2. Ronzhin A.L. Topologicheskkiye osobennosti morfofonemnogo sposoba predstavleniya slovarya dlya raspoznavaniya russkoy rechi. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies]. 2008. no. 9. pp. 12-19.
3. Ronzhin A.L. Sravnitel'nyy analiz i otsenka modeley slovarya dlya sistem raspoznavaniya russkoy rechi. *Informatsionnyye tekhnologii* [Information Technologies], 2009, no. 1, pp. 21-28.
4. Sarada G. L., Lakshmi A., Murthy H. A., Nagarajan T. Automatic transcription of continuous speech into syllable-like units for Indian languages. *Sadhana*, 2009, Vol. 34, Part 2, pp. 221–233.
5. Kipyatkova I.S., Karpov A.A. Eksperimenty po raspoznavaniyu slitnoy russkoy rechi s ispol'zovaniyem sverkhbol'shogo slovarya. *Trudy SPIIRAN* [Proceedings of SPIIRAN], 2010, no. 1 (12), pp. 63-74.
6. Savchenko V.V., Savchenko A.V. Metodika formirovaniya rabocheho slovarya v sistemakh avtomaticheskogo raspoznavaniya rechi po tematicheskomu faylu v tekstovom formate. *Sistemy upravleniya i informatsionnyye tekhnologii* [Control Systems and Information Technology], 2012, T. 48, no. 2.2, pp. 284-289.
7. Savchenko A.V. Avtomaticheskoye postroyeniye foneticheskoy transkriptsii rechi na osnove printsipa minimuma informatsionnogo rassoglasovaniya. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies], 2012, no 8, pp. 14-19.
8. Khalid A. Darabkh, Ala F. Khalifeh, Baraa A. Bathech, and Saed W. Sabah. Efficient DTW-Based Speech Recognition System for Isolated Words of Arabic Language. *World Academy of Science, Engineering and Technology International Journal of Electrical and Computer Engineering*, 2013, Vol. 7, no. 5, pp. 586-593.
9. Du TS. Otlozhennaya slogovaya segmentatsiya pri raspoznavanii rechi. *Molodezhnyy nauchno-tekhnicheskyy vestnik* [Youth scientific and technical bulletin], 2014, no. 3, p. 29.
10. Zhilyakov Ye.G., Belov S.P., Belov A.S., Firsova A.A. O segmentatsii rechevykh signalov na odnorodnyye otrezki. *Nauchnyye vedomosti Belgorodskogo gosudarstvennogo universiteta. Seriya: Ekonomika. Informatika* [Scientific Bulletin of Belgorod State University. Series: The Economy. Computer science], 2015, T. 34, no. 7-1, pp. 194-199.
11. Shirokova A., Telesnin B., Rogozhina V. Multi-Pronunciation Lexicon for Russian Automatic Speech Recognition. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialogue 2016»* Moscow, June 1–4, 2016. URL: www.dialog-21.ru/media/3425/shirokovaetal.pdf.
12. Chervyakov N.I., Kuchukova N.N. Problemy avtomaticheskogo raspoznavaniya slitnoy rechi. Metody obrabotki iskhodnogo rechevogo signala. *Nauchnyye vedomosti Belgorodskogo gosudarstvennogo universiteta. Seriya: Ekonomika. Informatika* [Scientific bulletins of the Belgorod State University. Series: The Economy. Computer science], 2016, T. 40, no. 23 (244), pp. 148-155.
13. Shishkina A.F. K voprosu ob avtomaticheskome raspoznavanii zvukov slitnoy rechi. *Teoriya. Praktika. Innovatsii* [Theory. Practice. Innovation], 2016, no. 4 (4), pp. 18-23.

14. Khein M.Z. Algoritmy dinamicheskogo programmirovaniya v raspoznavanii rechi *Auditorium*, 2017, no. 4 (16), pp. 48-52.
15. Khudoyberdiyev K.H.A. Model' analiza i segmentatsii rechevogo signala dlya poslogovogo raspoznavaniya tadjikskoy rechi. *Vestnik Tekhnologicheskogo universiteta Tadjikistana* [Bulletin of the Technological University of Tajikistan]. 2017, no. 4 (31), pp. 85-87.
16. Buribayeva A.K., Dorokhina G.V., Nitsenko A.V., Shelepov V.YU. Segmentatsiya i difonnoye raspoznavaniye rechevykh signalov. *Trudy SPIIRAN* [Proceedings of SPIIRAS], 2013, no. 31, pp. 20-42.
17. Shelepov V.YU., Nitsenko A.V. *Segmentatsiya i difonnoye raspoznavaniye rechi* [Segmentation and diphone speech recognition], Donetsk: GU IPII, 2015, 231 p.
18. Zaliznyak A.A. *Grammaticheskiy slovar' russkogo yazyka* [Grammatical dictionary of the Russian language], M., Russkiy yazyk, 1980, 879 p.
19. Shelepov V.Ju., Nicenko A.V. O raspoznavanii pervogo zvuka v slitnom rechevom otrezke [On Recognition of the First Sound in Continuous Speech Fragment]. *Problems of Artificial Intelligence*, 2015, no. 0(1), pp. 116 – 122.

RESUME

Shelepov V. Y., Nicenko A. V.

On the recognition of russian words using generalized transcription

Background: At present, the problem of constructing speech recognition systems with large-volume vocabularies working on local computers remains urgent. In particular, we are talking about the speed of such systems.

Materials and methods: The article describes the way of separately pronounced Russian words recognition using a priori segmentation, generalized transcription and the DTW-algorithm. Initially, generalized transcription is recognized, and then the recognition is carried out among the small number of complete transcriptions that correspond to the recognized generalized transcription.

Results: It is shown that the structure of the spoken word can be established with the help of only a few sounds as representatives of vowels, voiced consonants, unvoiced fricatives and unvoiced plosive sounds, if we also use the corresponding diphones and a priori segmentation.

Conclusion: The developed methods and algorithms provide significant improvement in recognition quality and speed when working with large dictionaries and can be used to create computer-assisted speech recognition systems.

Статья поступила в редакцию 26.01.2018.