

И.С. КИПЯТКОВА, А.А. КАРПОВ

ЭКСПЕРИМЕНТЫ ПО РАСПОЗНАВАНИЮ СЛИТНОЙ РУССКОЙ РЕЧИ С ИСПОЛЬЗОВАНИЕМ СВЕРХБОЛЬШОГО СЛОВАРЯ

Кипяткова И.С., Карпов А.А. Эксперименты по распознаванию слитной русской речи с использованием сверхбольшого словаря.

Аннотация. Для распознавания слитной речи со сверхбольшим словарем распознавателю необходима модель языка, описывающая допустимые фразы. В статье представлены результаты экспериментов по распознаванию слитной речи со сверхбольшим (более 100 тыс. слов) словарем с применением n -граммных моделей языка. Проведено количественное сравнение точности распознавания слов, символов и фонем в зависимости от используемой n -граммной модели при изменении n от 0 до 3.

Ключевые слова: распознавание слитной русской речи, сверхбольшой словарь, модели языка.

Kipyatkova I.S., Karpov A.A. Experiments on Continuous Russian Speech Recognition with Usage of an Extra-Large Vocabulary.

Abstract. For extra-large vocabulary continuous speech recognition a language model that describes permissible phrases is needed. In the paper, the results of experiments on extra-large vocabulary (above 100 K words) continuous speech recognition, with usage of n -gram models, are presented. A quantitative comparison of recognition accuracy of words, symbols, and phonemes depending on n -gram model, with n value varying from 0 till 3, was made.

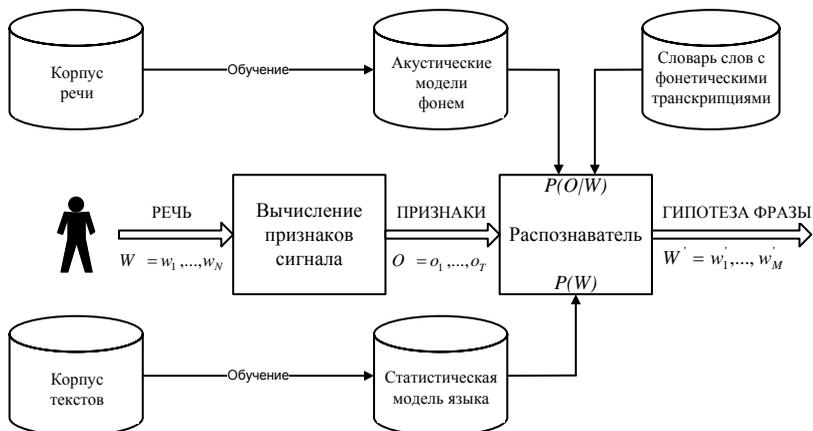
Keywords: continuous Russian speech recognition, extra-large vocabulary, language models.

1. Введение. Автоматическое распознавание русской слитной речи представляет собой очень сложную задачу из-за целого ряда особенностей и трудностей русского языка. В первую очередь русский язык обладает сложной и вариативной системой словообразования, в результате чего в системе распознавания речи необходимо использовать намного больший (по сравнению с английским языком) по объему словарь, что значительно снижает и точность, и скорость автоматического распознавания. В системах распознавания английской речи достаточен словарь, содержащий порядка 60–100 тыс. слов, включая распространенные имена, фамилии и названия. Для русского языка с его богатой морфологией объем фонетического словаря распознавания должен быть в несколько раз больше. Например, грамматический словарь А.А. Зализняка содержит около 150 тыс. употребительных слов русского языка, и при помощи специальной системы обозначений он позволяет построить все словоформы для выбранного слова. При развороте всех словарных статей получается порядка 2 млн. различных

словоформ. Так, например, глаголы могут формировать до нескольких сотен различных словоформ, и их нужно учитывать при создании систем автоматического распознавания русской речи. Кроме того, большинство словоформ одной лексемы различаются только окончаниями, которые произносятся обычно не так четко, как начала слов. Ошибки в окончаниях при распознавании слов приводят к тому, что возникает ошибка в распознавании всей фразы из-за несогласованности слов в предложении. Русская речь также обладает особым фонетическим строем. В международном фонетическом алфавите SAMPA для русского языка, который принято использовать в компьютерных системах, присутствуют 42 фонемы: 36 фонем согласных звуков и 6 фонем гласных звуков. Фонетический алфавит SAMPA для американского варианта английского языка насчитывает 41 фонему: 24 согласных и 17 гласных (включая целый ряд дифтонгов). Однако распознавание согласных звуков гораздо сложнее, чем гласных, из-за того что они менее стабильны, чем гласные и имеют меньшую длительность. Порядок слов в предложениях русского языка не фиксируется жестко правилами грамматики и может сильно варьироваться без потери смысла предложения, а в английском языке используются относительно жесткие грамматические конструкции. Это затрудняет создание моделей языка и грамматик для русского языка и понижает их эффективность. Статистические языковые модели для русского языка не столь эффективны как для английского языка. В целях создания и обучения систем распознавания для многих мировых языков существует несколько больших многодикторных речевых баз данных (микрофонной, телефонной речи и т.п.), в том числе и свободно доступных. Для русского языка такие речевые корпуса только начинают создаваться и, как правило, это недоступные для свободного использования дорогие коммерческие продукты (например, SPEECON или RuSpeech).

В статье представлены результаты распознавания слитной русской речи при использовании сверхбольшого словаря. Сверхбольшим считается словарь объемом более 100 тыс. слов, именно такой объем словаря необходим для создания систем стенографирования устной русской речи.

2. Система автоматического распознавания русской речи. На рисунке показана общая схема распознавателя речи, построенного с использованием аппарата СММ [8]. Человек произносит некоторую фразу, которая представляет собой последовательность слов $W = w_1, \dots, w_N$. Задача системы распознавания речи заключается в том, чтобы правильно распознать эту последовательность слов.



Базовая архитектура системы дикторонезависимого распознавания слитной речи.

Процесс распознавания речи может быть представлен как поиск наиболее вероятной последовательности слов [8]:

$$W = \arg \max_w P(W | A) = \arg \max_w P(A | W) \cdot P(W),$$

где $P(W | A)$ — вероятность появления гипотезы по оценке акустической модели, $P(W)$ — вероятность появления гипотезы по оценке языковой модели.

Первым этапом распознавания речи является параметрическое представление речевого сигнала. Самым распространенным подходом к параметризации речи является спектральный анализ сегментов сигнала с вычислением их кепстральных признаков [5, 10]. Алгоритм включает в себя:

- 1) предварительную обработку цифровых отсчетов, например, предсказывающий фильтр или процедуру весовой обработки окна;
- 2) быстрое преобразование Фурье (БПФ) над сегментами речи длительностью 10–30 мс;
- 3) формирование набора перекрывающихся фильтров, расположенных по нелинейному закону согласно *Mel*-шкале частот [1].

Отсчеты БПФ, включенные в каждый фильтр, пересчитываются с учетом треугольного окна, затем определяется интегральная энергия

на выходе каждого фильтра и далее производится логарифмирование выхода каждого фильтра. Этот набор данных подвергается косинус-преобразованию, что в итоге приводит к получению кепстральных коэффициентов. Кепстральные коэффициенты, полученные с использованием *Mel*-шкалы треугольных фильтров, называются *Mel*-частотными кепстральными коэффициентами (MFCC). Таким образом, в результате процедуры параметрического представления речевой сигнал преобразуется в последовательность векторов-признаков, после чего переходит на более высокий уровень моделирования — звуков и слов.

В качестве акустической модели используются скрытые Марковские модели (СММ), при этом каждая фонема представляется в виде скрытой Марковской модели. Модель фонемы имеет три состояния: первое описывает начало фонемы, второе представляет центральную часть фонемы и третье — окончание фонемы. СММ слова получается путем соединения в цепочку моделей фонем из соответствующего фонетического алфавита. Аналогичным образом соединяются друг с другом модели слов, образуя модели фраз и предложений, соответствующие модели языка.

Для распознавания слитной речи используется модифицированный алгоритм Витерби, называемый методом передачи маркеров (*token passing method*) [11] и реализованный в инструментарии Hidden Markov Model Toolkit. Метод передачи маркеров определяет прохождение возможных путей по состояниям объединенной СММ. В начало каждого слова ставится маркер и применяется итеративный алгоритм оптимизации Витерби, при этом на каждом шаге маркер сдвигается и для него вычисляется акустическая вероятность. При достижении состояния конца некоторого слова в маркер записывается индекс этого слова, а при выходе из каждого состояния маркеры размножаются (путем копирования) по числу дальнейших переходов в модели. При этом в маркер записывается его путь (история) через сеть. Когда маркер переходит от выходного состояния одного слова к входному состоянию другого, переход представляет собой потенциальную границу слов, которая и записывается в историю маркера. В итоге после обработки всей последовательности векторов наблюдений выбирается маркер, имеющий наибольшую вероятность. Когда наилучший маркер достигает конца обрабатываемого сигнала (последовательности наблюдений), то путь, которым он проходит через сеть, известен в виде истории (хранящейся в маркере), и из маркера считывается последовательность пройденных слов, которая и является гипотезой распознавания

фразы. Эта методика распознавания слитной речи эффективно используется в настоящее время для автоматического распознавания речи на многих языках.

В данной работе модель языка и словарь системы распознавания создавались по текстовому корпусу, собранному с Интернет-сайтов электронных газет. Процесс создания модели языка описан в работе [2]. Созданы уни-, би- и триграммная модели языка. Транскрипции для слов из этих моделей языка образованы с помощью программного модуля, позволяющего создавать транскрипции автоматически [3]. На вход этого модуля поступают:

- список слов, которые необходимо транскрибировать;
- словарь словоформ с отметкой ударения;
- используемый фонетический алфавит;
- фонетические правила транскрибирования.

Результатом работы модуля является транскрибированный словарь словоформ. В качестве фонетического алфавита применен модифицированный вариант международного фонетического алфавита SAMPA, в котором используются 48 фонем: 12 — для гласных звуков (с учетом ударных вариантов) и 36 — для согласных (с учетом твердости и мягкости звуков) [4]. Для нульграммной (т.е. при распознавании без модели языка), уни- и биграммной моделей использован одинаковый словарь объемом в 182 тыс. словоформ, для триграммной модели объем словаря — 126 тыс. словоформ. Кроме того, создана униграммная модель с расширенным словарем, объем которого составил 535 тыс. словоформ. Число биграмм составило 3,17 млн, триграмм — 3,54 млн.

Для распознавания слитной русской речи использован разработанный в СПИИРАН декодер SIRIUS [9]. В качестве фонетических единиц при распознавании речи применялись контекстно-зависимые фонемы (трифоны). Запись обучающего и тестового речевого корпуса для системы производилась с частотой дискретизации 44 кГц, 16 бит на отсчет, моно, отношение сигнал/шум больше 35 дБ. Для обучения системы распознавания речи использовано 300 вручную размеченных фраз из речевого корпуса [6]. Система обучена на записях одного диктора и является, таким образом, дикторозависимой. Для тестирования системы использовались 316 слитно произнесенных фраз, состоящих из 1600 слов (8849 символов). В текстовом корпусе, используемом для тестирования, число биграмм, присутствующих в модели языка, составило 71,90 %, триграмм — 24,15 %. Для этого тестового корпуса вычислены энтропия и коэффициент неопределенности (*perplexity*) стати-

стической модели языка [7]. Для униграммной модели коэффициент неопределенности равен 8495,71, энтропия — 13,05 бит/слово, для биграммной модели коэффициент неопределенности равен 2909,03, энтропия — 11,51 бит/слово, для триграммной модели коэффициент неопределенности равен 2074,79, энтропия — 11,02 бит/слово.

3. Предварительные эксперименты по распознаванию слитной русской речи со сверхбольшим словарем. При оценке качества распознавания речи или текста выделяют три типа ошибок: 1) замена одного слова другим, 2) удаление слова, 3) вставка слова [11]. Существует два критерия, по которым определяется качество распознавания: правильность и точность распознавания. Правильность определяется следующим образом [11]:

$$C = \frac{H}{N} \cdot 100\% ,$$

где H — число правильно распознанных слов, N — общее число слов в распознаваемом сообщении.

$$H = N - D - S ,$$

где D и S — число неверно удаленных и замененных слов соответственно.

Этот критерий не учитывает ошибки, связанные со вставкой слов, поэтому обычно используется другой критерий — точность распознавания, который вычисляется по следующей формуле [11]:

$$A = \frac{H - I}{N} \cdot 100\% ,$$

где I — число неверно вставленных слов.

Результаты распознавания слов и символов (под символом понимаются буква и знак пробела) с применением различных моделей языка представлены в табл. 1 и 2. Для экспериментов по распознаванию фонем использовалось 315 фраз, состоящих из 7021 фонем. В табл. 3 представлены результаты по распознаванию фонем при использовании нуль-, уни- и биграммной моделей языка. При использовании нульграммной модели точность распознавания слов оказалась равной 14,25 %, символов — 70,77 %, фонем — 60,93 %. Использование биграммной модели позволило повысить точность распознавания слов до 36,06 %. При применении триграммной модели точность распознавания составила 28,50 %. Стало неожиданным, что наилучшие результаты достигнуты с использованием униграммной модели, где точность распознавания слов составила 46,25 %. А при использовании униграммной модели с расширенным словарем точность возросла до 48,75 %.

Это связано с тем, что достаточно много би- и триграмм из тестового корпуса отсутствуют в обучающем материале, что в целом характерно для русского языка.

Таблица 1. Результаты распознавания слов с использованием различных моделей языка

Тип модели	<i>A</i> , %	<i>C</i> , %	<i>H</i>	<i>D</i>	<i>S</i>	<i>I</i>
0-граммная	14,25	38,56	617	56	927	389
1-граммная	46,25	47,81	765	237	598	25
1-граммная с расширенным словарем	48,75	49,75	796	255	549	16
2-граммная	36,06	58,50	936	44	620	359
3-граммная	29,25	31,94	511	462	627	43

Таблица 2. Результаты распознавания символов с использованием различных моделей языка

Тип модели	<i>A</i> , %	<i>C</i> , %	<i>H</i>	<i>D</i>	<i>S</i>	<i>I</i>
0-граммная	70,77	80,21	7098	716	1035	836
1-граммная	78,36	81,26	7191	879	779	257
1-граммная с расширенным словарем	80,13	82,55	7305	824	720	214
2-граммная	75,71	84,35	7464	512	873	764
3-граммная	66,64	71,39	6317	1340	1192	420

Таблица 3. Результаты распознавания фонем с использованием различных моделей языка

Тип модели	<i>A</i> , %	<i>C</i> , %	<i>H</i>	<i>D</i>	<i>S</i>	<i>I</i>
0-граммная	60,93	79,40	5575	128	1318	1297
1-граммная	80,09	84,72	5948	300	773	325
1-граммная с расширенным словарем	79,92	84,43	5928	338	755	317
2-граммная	75,10	84,05	5901	215	905	628

В табл. 4 представлен пример распознавания предложения «она расчесывала волосы, стоя перед громадным зеркалом». При использовании нуль-граммной модели многие слова оказались поделены на несколько коротких. Из-за того, что число вставленных слов было больше, чем слов во фразе, точность распознавания этой фразы оказалась отрицательной. При использовании биграммной модели число правильно распознанных слов оказалось равно числу вставленных, и в результате точность распознавания была равна нулю. Слово «расчесывала» было правильно распознано только при применении уни-

граммной модели с расширенным словарем, поскольку в других моделях этого слова в словаре не было.

Таблица 4. Пример распознанного предложения

Тип модели	Результат распознавания	Точность распознавания слов во фразе, %	Правильность распознавания слов во фразе, %
0-граммная	Она раж чо со вал волосы сто ы пер эда громадным зеркалом	-14,29	57,14
1-граммная	Она рассказывал волосы ступень громадным зеркалом	57,14	57,14
1-граммная с расширенным словарем	Она расчесывала волосы ступень громадным зеркалом	71,43	71,43
2-граммная	Она рассказывала и волосы в топе и грамотным и зеркалом	0,00	42,86
3-граммная	Она рассказывала волосы ступень громадным зеркалом	57,14	57,14

4. Заключение. Проведены эксперименты по дикторозависимому распознаванию слитно произнесенных фраз с применением нуль-, уни-, би- и триграммной моделей языка. В ходе экспериментов оказалось, что униграммная модель эффективнее, чем би- и триграммная, если сравнивать точность распознавания слов в слитной речи. Это связано с тем, что в тестовом материале было много биграмм и триграмм, которые отсутствуют в обучающем корпусе. Решить данную проблему, возможно, поможет использование метода возврата (*backing-off method*) [7]. Суть данного метода заключается в том, что если некоторая n -грамма отсутствует в обучающем корпусе или частота ее появления очень низкая, тогда вместо нее используется вероятность $(n-1)$ -граммы. Кроме того, в дальнейшей работе мы намереемся модифицировать базовую n -граммную модель с учетом особенностей русского языка. Мы планируем исследовать модели языка, строящиеся на основе начальных форм слов и основах слов, а также модели языка, использующие грамматические показатели.

Литература

1. Загоруйко Н.Г. Распознавание слуховых образов. Новосибирск: Наука, 1970. 340 с.
2. Китяtkова И.С., Карпов А.А. Автоматическая обработка и статистический анализ новостного текстового корпуса для модели языка системы распознавания русской речи // Информационно-управляющие системы. СПб., 2010. С. 2–8.

3. *Кипяткова И.С., Карпов А.А.* Модуль фонематического транскрибирования для системы распознавания разговорной русской речи // Искусственный интеллект. 2008. № 4. С. 747–757.
4. *Ронжин А.Л., Карпов А.А., Лобанов Б.М., Цирульник Л.И., Йюкиш О.* Фонетико-морфологическая разметка речевых корпусов для распознавания и синтеза русской речи // Информационно-управляющие системы. 2006. № 6. С. 24–34.
5. *Сапожков М.А.* Речевой сигнал в кибернетике и связи. М.: Связьиздат, 1963. 452 с.
6. *Jokisch O., Wagner A., Sabo R., Jaeckel R. et al.* Multilingual Speech Data Collection for the Assessment of Pronunciation and Prosody in a Language Learning System // Proc. of 13th Intern. Conf. SPECOM'2009. St. Petersburg. 2009. P. 515–520.
7. *Moore G.L.* Adaptive Statistical Class-based Language Modelling. PhD thesis. Cambridge University, 2001. 193 p.
8. *Rabiner L., Juang B.-H.* Fundamentals of Speech Recognition. Prentice Hall, 1995. 507 p.
9. *Ronzhin A.L., Karpov A.A.* Russian Voice Interface // Pattern Recognition and Image Analysis. (Advances in Mathematical Theory and Applications). 2007. Vol. 17, N 2. С. 321–336.
10. *Strom N.* Continuous Speech Recognition in the WAXHOLM Dialogue System // Stockholm QPSR. 1996. P. 67–95.
11. *Young S. et al.* The HTK Book (for HTK Version 3.4). Cambridge. UK. 2009.

Кипяткова Ирина Сергеевна — м. н. с. лаборатории речевых и многомодальных интерфейсов Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: автоматическое распознавание речи, статистические модели языка. Число научных публикаций — 15. kipyatkova@ias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081. Научный руководитель — канд. техн. наук А.А. Карпов.

Kipyatkova Irina Sergeevna — junior researcher, Laboratory of Speech and Multimodal Interfaces St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: automatic speech recognition statistical language models. The number of publications — 15. kipyatkova@ias.spb.su; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081. Scientific adviser — PhD A.A. Karpov.

Карпов Алексей Анатольевич — канд. техн. наук, старший научный сотрудник лаборатории речевых и многомодальных интерфейсов Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: автоматическое распознавание речи, многомодальные интерфейсы, аудиовизуальное распознавание речи. Число научных публикаций — 100. karpov@ias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081.

Karpov Alexey Anatolyevich — PhD, senior researcher, Laboratory of Speech and Multimodal Interfaces St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: automatic speech recognition, multimodal interfaces, audio-visual speech recognition. The number of publications — 100. karpov@ias.spb.su; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081.

Поддержка исследований. Данное исследование поддержано Министерством образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» (госконтракты № П2579, П2360, П876, 14.740.11.0357) и фондом РФФИ (проекты № 08-08-00128, 09-07-91220-СТ, 10-08-00199).

Рекомендовано лабораторией речевых и многомодальных интерфейсов, заведующий лабораторией д-р техн. наук, доц. А.Л. Ронжин.

Статья поступила в редакцию 16.11.2010.

РЕФЕРАТ

Кипяткова И.С., Карнов А.А. Эксперименты по распознаванию слитной русской речи с использованием сверхбольшого словаря.

Использование сверхбольших словарей (более 100 тыс. слов) необходимо при разработке систем стенографирования русской речи из-за того, что русский язык является флективным языком. В статье представлены результаты экспериментов по распознаванию слитной русской речи со сверхбольшим словарем в зависимости от используемой модели языка. Для распознавания слитной русской речи использовался декодер SIRIUS. Обучение и тестирование системы распознавания осуществлялось на записях одного диктора. В качестве акустических моделей применялись скрытые Марковские модели с тремя состояниями. Для языкового моделирования использовались статистические n -граммные модели со значением n от 0 до 3. Для нуль-, уни- и биграммной моделей использовался одинаковый словарь объемом в 182 тыс. словоформ, для триграммной модели объем словаря был 126 тыс. словоформ. Кроме того, была создана униграммная модель с расширенным словарем, объем которого составил 535 тыс. словоформ. Число биграмм составило 3,17 млн, триграмм — 3,54 млн. Транскрипции для слов из этих моделей языка были созданы автоматически с помощью программного модуля, позволяющего создавать фонематические транскрипции слов. В качестве фонетического алфавита был применен модифицированный вариант международного фонетического алфавита SAMPA, в котором используются 48 фонем: 12 — для гласных звуков и 36 — для согласных.

Для тестирования системы использовались 316 слитно произнесенных фраз, состоящих из 1600 слов (8849 символов). В тестовом корпусе число биграмм, присутствующих в модели языка, составило 71,90 %, триграмм — 24,15 %. Для этого тестового корпуса вычислены величины энтропии и коэффициента неопределенности статистической модели языка. Для униграммной модели коэффициент неопределенности равен 8495,71, энтропия — 13,05 бит/слово, для биграммной коэффициент неопределенности равен 2909,03, энтропия — 11,51 бит/слово, для триграммной коэффициент неопределенности был равен 2074,79, энтропия — 11,02 бит/слово.

В качестве фонетических единиц при распознавании речи применялись трифоны. Была определена точность распознавания слов, символов и фонем. При использовании нульграммной модели точность распознавания слов оказалась равной 14,25 %, символов — 70,77 %, фонем — 60,93 %. Использование биграммной модели позволило повысить точность распознавания слов до 36,06 %. При применении триграммной модели точность распознавания равна 28,50 %. Наилучшие результаты были достигнуты с использованием униграммной модели, где точность распознавания слов составила 46,25 %. Использование униграммной модели с расширенным словарем позволило повысить точность до 48,75 %. Это связано с тем, что достаточно много биграмм и триграмм из тестового корпуса отсутствуют в обучающем материале.

SUMMARY

Kipyatkova I.S., Karpov A.A. **Experiments on Continuous Russian Speech Recognition with Usage of an Extra-Large Vocabulary.**

The usage of extra-large (over 100 K words) vocabularies is necessary for developing of dictation systems for Russian because Russian is an inflective language. In the paper, the results of experiments on continuous Russian speech recognition with the extra-large vocabulary depending on language models are presented. For continuous Russian speech recognition the SIRIUS speech decoder was used. Both training and testing of the recognition system were made in the speaker-dependent mode. Hidden Markov Models with three states were used for acoustical modeling. Statistical n-gram models with n varying from 0 till 3 were used for language modeling. For nullgram, unigram, and bigram models, the vocabulary of 182 K words has been created, for the trigram model the vocabulary has 126 K word-forms. Furthermore, an unigram model with the extended vocabulary with 535 K word-forms has been created. Number of bigrams was 3.17 M, and number of trigrams was 3.54 M. Phonemic transcriptions for the words from these language models have been created automatically with help of a software module. A modified variant of the International phonetic alphabet SAMPA was used as a phonetic alphabet; in this alphabet there are 48 phonemes: 12 phonemes are used for vowels and 36 phonemes — for consonants.

For the system testing, 316 continuously pronounced phrases consisting totally of 1600 words (8849) symbols were used. For this test corpus, the quantity of bigrams, existing in the language model, was 71.90 % and 24.15 % — for trigrams. Entropy and perplexity parameters of the statistical language models have been computed. For the unigram language model the perplexity is 8495.71, the entropy equals 13.05 bit/word, for the bigram model the perplexity is 2909.03, the entropy is 11.51 bit/word, for the trigram model the perplexity is 2074.79, the entropy equals 11.02 bit/word.

Triphones were used as phonetic units during speech recognition. Accuracy of recognition of words, symbols, and phonemes has been determined. When using the nullgram model the word recognition rate was 14.25 %, the symbol recognition rate was 70.77 %, the phoneme recognition rate was 60.93 %. Usage of the bigram model allowed to increase the word recognition rate up to 36.06 %. Word recognition rate equals 28.5 % with using the trigram model. The best results achieved with usage of the unigram model, word recognition rate was 46.25 %. Usage of the unigram model with the extended vocabulary allowed increasing recognition accuracy up to 48.75 %. It is connected with a fact that lots of bigrams and trigrams from the text corpus are missing in the training data.