

7-ая Международная конференция
«Управление развитием крупномасштабных систем»
(MLSD'2013)

Контекстное распознавание слитной речи

Доклад подготовили:
Институт проблем управления РАН
МГТУ им. Н.Э. Баумана
Выхованец В.С., Ду Цзяньмин, Назарова С.И.

Автоматическое распознавание речи

Распознавание речи – область речевых технологий, включающая целый ряд достаточно обособленных направлений, каждое из которых ориентировано на решение конкретных прикладных задач, требующих отдельной теоретической и практической проработки.

Автоматическое преобразование речи – процесс преобразования речевого сигнала в текст.

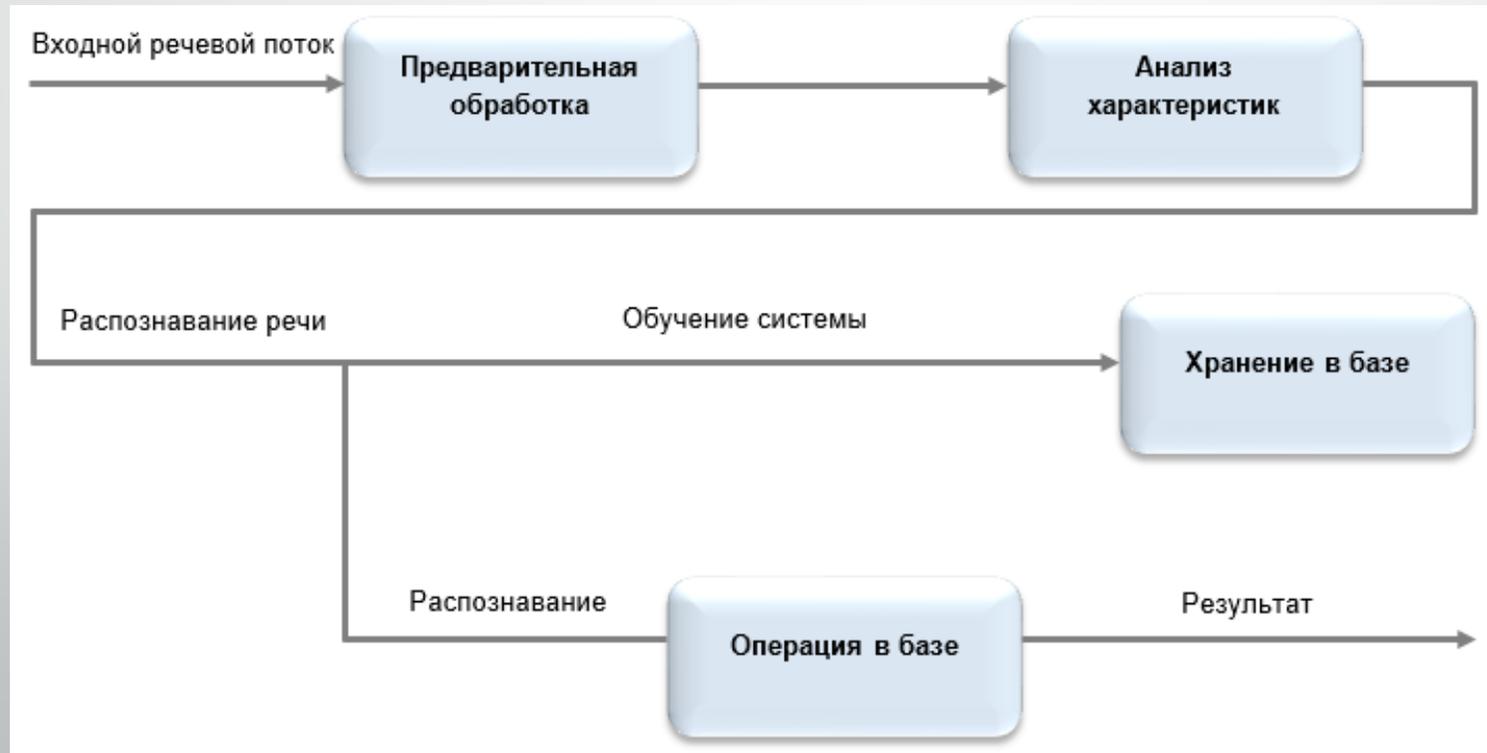


Автоматическое распознавание речи



Автоматическое распознавание речи

Отличительная особенность применяемых для распознавания речи решений – необходимость выделения двух этапов процесса обработки входного речевого потока: этап обучения системы и, собственно, этап распознавания.



Факторы, влияющие на достоверность распознавания



Флективность естественных языков

Флективность – тенденция к словоизменительной аффиксации, свойственная многим синтетическим языкам.

Степень флективности – отношение в отрезке текста числа флексий к числу словоупотреблений.

Кумуляция – постоянное совмещение в одном словоизменительном аффиксе нескольких граммем, принадлежащих разным грамматическим категориям.

Так в русских падежных окончаниях существительных совмещены граммы падежа и числа, а в окончаниях прилагательных – падежа, числа и рода.

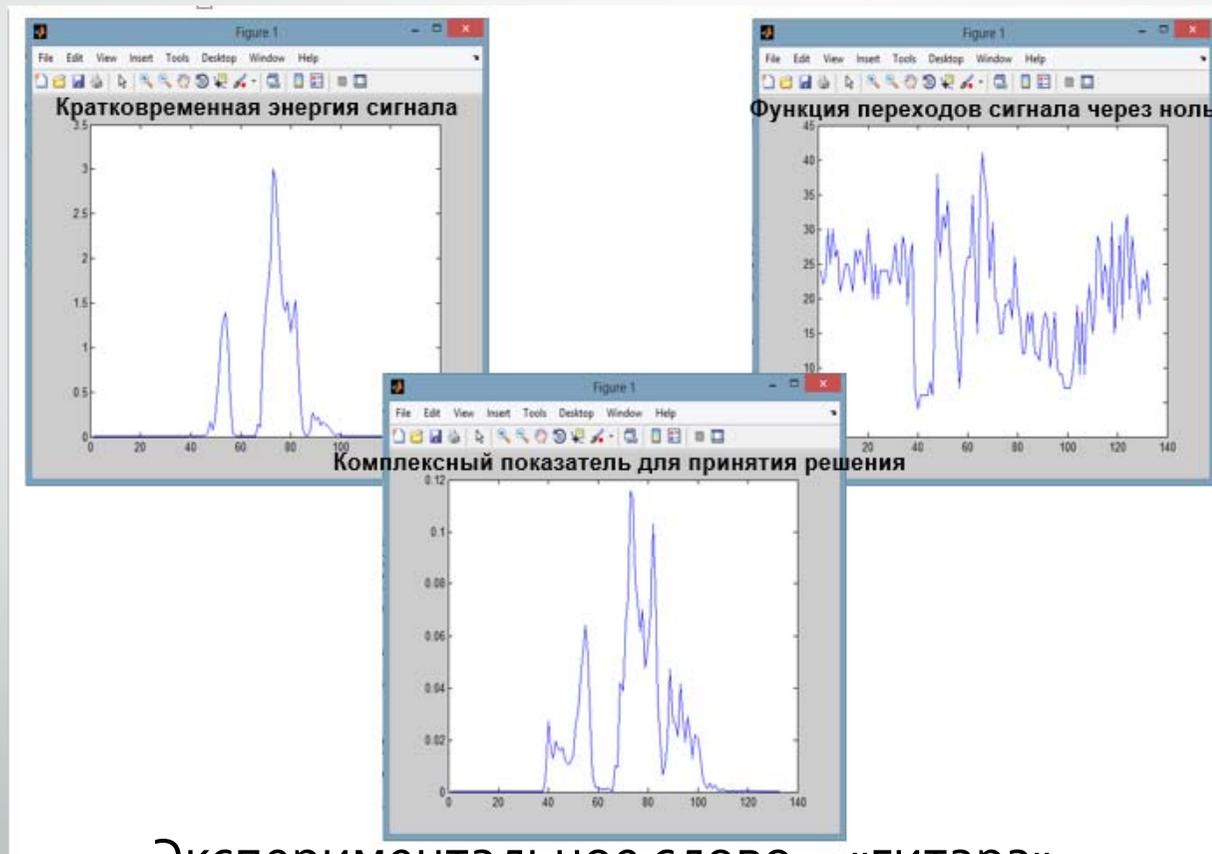
Омосемия словоизменительных аффиксов – наличие ряда параллельных (равнозначных) аффиксов для передачи одного и того же грамматического значения или комплекса значений.

Так в русском языке значение творительного падежа единственного числа у существительных выражается омосемичными флексиями –ом («стол-ом»), -ой («стен-ой»), -ью («кровать-ью»).

Алгоритмы сегментации речи



Метод сегментации речи на основе оценки кратковременной энергии сигнала и числа переходов сигнала через ноль



Экспериментальное слово – «гитара».

Промежуточные выводы

- Было подтверждено, что речевая коммуникация предполагает не только декодирование языковых единиц, но и одновременное раскрытие их смыслов, т.е. для распознавания слитной речи требуется выполнять ее сегментацию одновременно с идентификацией.
- Непрерывная транскрипция речевого потока в текстовую форму признана малоперспективной.
- Наиболее надежной видится сегментация слитной речи на основе выделения в реальном масштабе времени гласных звуков с последующим построением на их основе фонетических слогов с учетом фонетического и грамматического строя речи.

Фонетические слоги

Фонетический слог – гласный или сочетание гласного с одним или несколькими согласными, произносимые одним выдохательным толчком речевого аппарата.

Контекст распознавания фонетического слога – это состояние фонетического анализатора, при котором гласный звук по-разному может быть соотнесен с включающим его слогом. При этом возможно присоединение к гласному звуку как предшествующих участков речи, так и последующих, вплоть до граничащих с ним других гласных звуков.

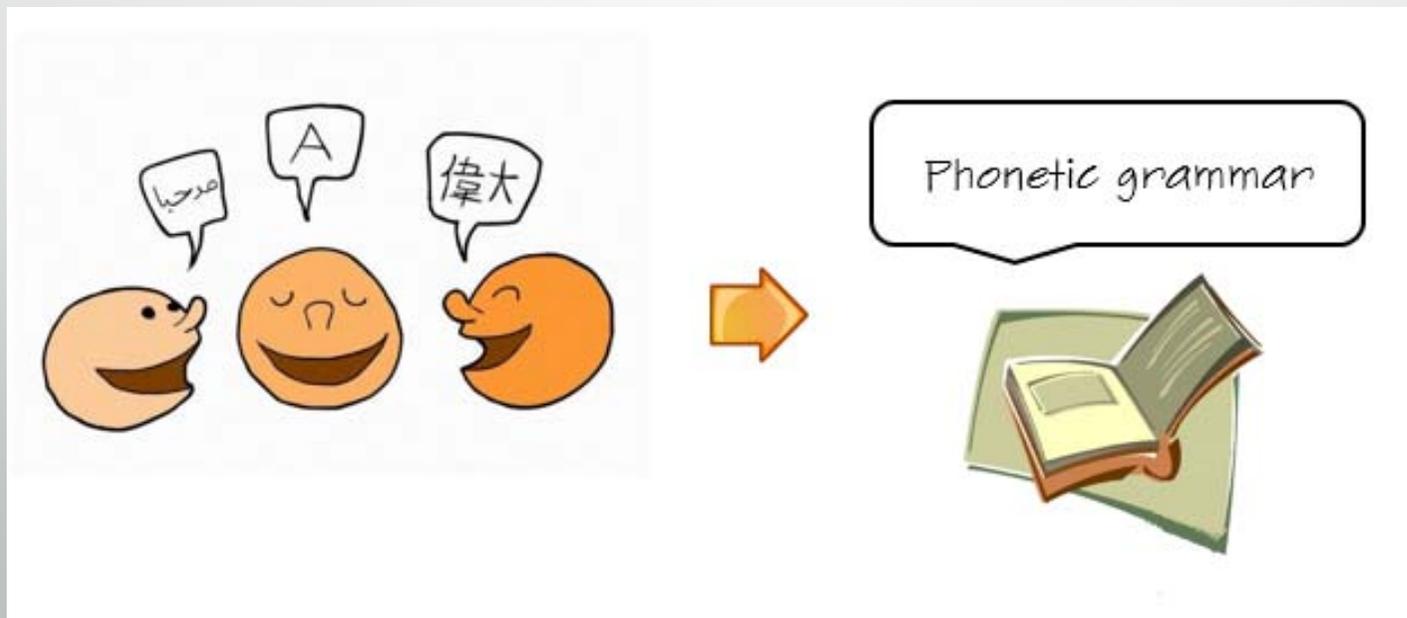
В этом случае принятие решения о слоговой сегментации осуществляется на поздних этапах распознавания – отложенная слоговая сегментация.

Выделение фонетических слогов будет производиться с учетом контекста распознавания, призванного учесть фоновые знания о предметной области дискурса и конкретную ситуацию речевого общения.

Фонетические грамматики

Фонетическая грамматика выражает правила построения правильных речевых отрезков с учетом как синтаксических, так и фонетических правил языка.

Элементарными единицами фонетической грамматики языка являются не слова и словосочетания, а фонетические слоги, задаваемые фрагментами образцовых речевых сигналов.



Терминальные и нетерминальные понятия языка-речи

Слог (терминальный знак) – элементарная **фонетическая единица** языка, состоящая из последовательности звуков, образующих фонетический слог.

Слово (лексема) – элементарная **синтаксическая единица** языка, представленная одним или несколькими слогами, которым уже может быть приписан (сопоставлен) некоторый элементарный смысл.

Понятие (нетерминальный знак) – элементарная **семантическая единица** языка, задаваемая множеством предложений, которым это понятие может выражено в речи.

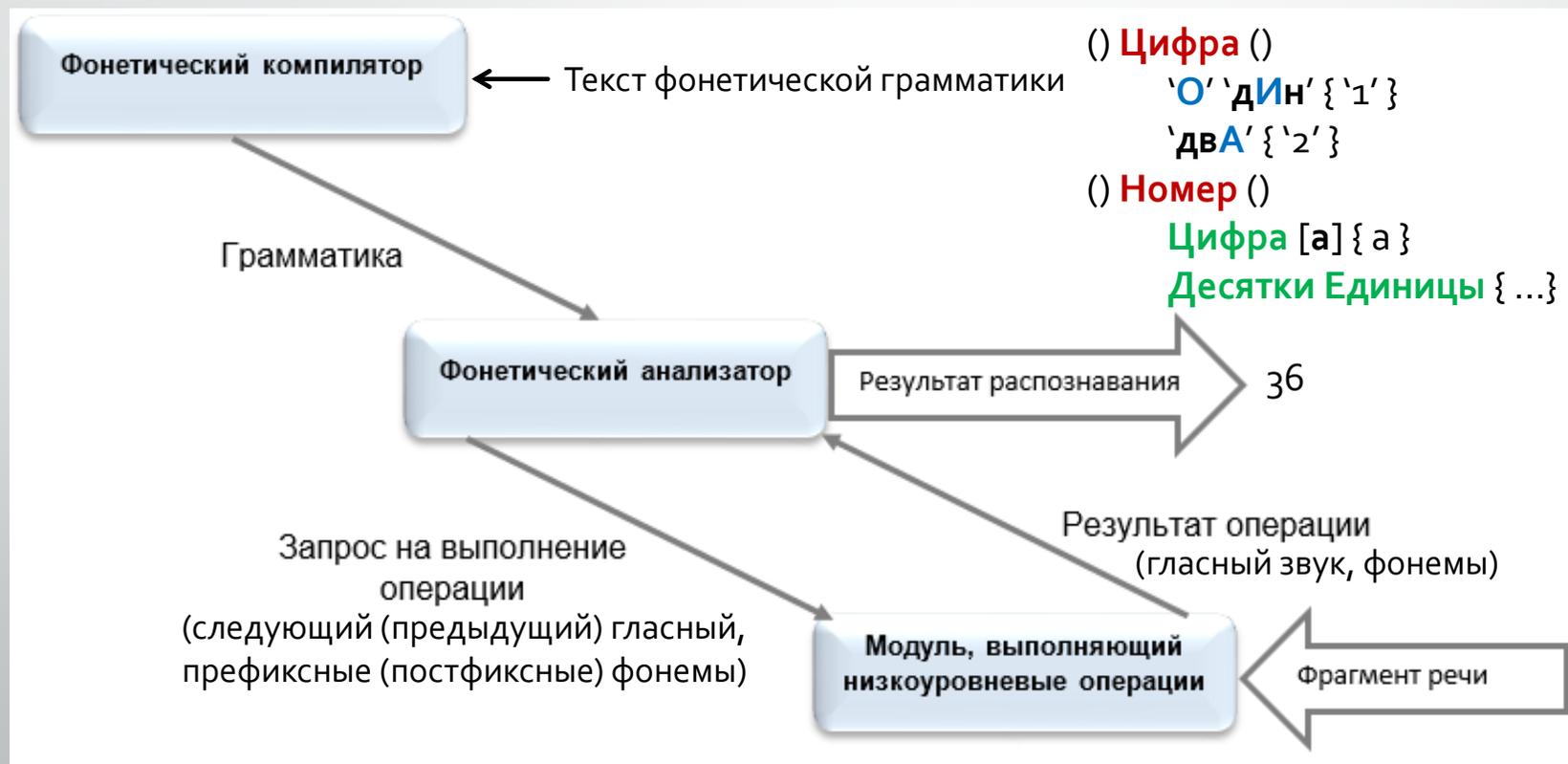
Предложение (продукция) – составная синтаксическая и семантическая единица языка, являющаяся последовательностью слов и понятий, выражающая некоторый **законченный смысл**, или, что то же самое, другое понятие.

Речь (текст) – последовательность предложений, представленная некоторым речевым отрезком и предназначенная для выражения более **сложного смысла**.

Таким образом, каждому нетерминальному знаку фонетической грамматики сопоставлено некоторое нетерминальное понятие описываемого языка-речи.

Тогда продукции грамматики могут быть проинтерпретированы как фонетические и синтаксические правила выражения понятий предметной области дискурса в речи.

Процесс распознавания речи с использованием фонетических грамматик



Заключение

Одной из причин неудач в области автоматического распознавания слитной речи является отсутствие содержательного контекста при распознавании.

Воспринимая речь, человек соотносит сказанное с действительностью, со своими знаниями о ней, со своим опытом, благодаря чему происходит восстановление нераспознанных фрагментов. В процессе восприятия речи человек активен, выдвигает гипотезы относительно содержания фрагментов речи и осуществляет смысловые замены.

В настоящей работе предложен подход к решению проблемы на основе отложенной слоговой сегментации речи, использования фонетической грамматики предметной области и соответствующего этой грамматике проблемного языка. Для задания предметного языка выполняется описание онтологии предметной области, фонетики языка, синтаксиса, семантики и прагматики допустимых языковых конструкций.