

СПОНТАННАЯ РЕЧЬ: ПРОБЛЕМЫ СЕГМЕНТАЦИИ*

Едва ли не важнейшей процедурой в процессе восприятия речи человеком является преобразование непрерывного акустического речевого сигнала в последовательность дискретных лексических единиц, т.е. членение речевого потока на слова.

В качестве одного из возможных вариантов была рассмотрена модель «сегментации через идентификацию» (Kassevich et al., 2000), в которой входная цепочка символов, описывающих акустический речевой сигнал, сравнивалась с элементами перцептивного словаря. Для этого использовались принципы «модели когорты».

Модель строилась из предположения, что носитель языка обладает настолько большим словарем словоформ, что во «входном речевом сигнале» ему не встретятся незнакомые слова. При таком допущении для сегментации произвольного текста с помощью подобной модели достаточно иметь словарь всех словоформ, образующих данный текст.

Работоспособность модели была проверена на примере сегментации беспробельного письменного текста с применением соответствующего словаря словоформ. В дополнение к этому знаки препинания в тексте считались пограничными сигналами, естественным образом определяющими правую (левую) границу слов во «входном речевом сигнале». Предполагалось, что в естественном речевом сигнале знаки препинания, присутствующие в тексте, будут проявляться в виде неких пограничных сигналов, реализуемых теми или иными супraseгментными средствами.

Оказалось, что количество ошибок сегментации не превышает 2-х процентов. При этом характер ошибок таков, что одна часть ошибок сегментации, предположительно, может быть устранена использованием грамматических правил и словаря словоформ, снабженных морфологическими пометами, а другая – за счет правил чередования частей речи в текстах на русском языке: например, невозможность появления двух предлогов подряд, за предлогом не может следовать союз и т.д.

При сплошном фонетическом транскрибировании спонтанной речи ситуация оказалась гораздо более сложной. По существу, единственным надежным пограничным сигналом, разбивающим речевой поток на отрезки с ограниченным лексическим составом и маркирующим начало первого слова и конец последнего, являются паузы разной природы (вдохи и вздохи, молчание, хезитация, ларингализация). Образующиеся при этом интервалы непрерывного «говорения» существенно различаются по длительности и структуре (Слепокурова, Камовкина, 2010),

* Работа выполнена при поддержке гранта РФФИ № 09-06-00244а.

В частности, в интервале между двумя паузами может быть произнесено только одно слово из длинной фразы, а несколько слов, разделенных запятыми, – без каких-либо перерывов. Например, фраза «нам всем очень хочется поспорить про современную российскую, русскую, татарскую, башкирскую книжку.» была произнесена следующим образом (указаны координаты начала и конца речевого отрезка в минутах и секундах, а также его длительность в секундах):

нам всем очень хочется поспорить про	12:58,507	13:00,521	02,014
паус	13:00,521	13:00,898	00,377
современную	13:00,898	13:01,856	00,957
паус	13:01,856	13:02,309	00,452
российскую, русскую, татарскую	13:02,309	13:04,120	01,811
м-э	13:04,120	13:04,335	00,214
башкирскую книжку	13:04,335	13:05,606	01,271
паус	13:05,606	13:06,291	00,685

При этом характер изменения частоты основного тона на прилагательном «современную» соответствует интонации завершенности (рис. 1), т.е. при его изолированном прослушивании носитель языка поставил бы в конце этого слова точку, интерпретировав его, например, как ответ на вопрос.

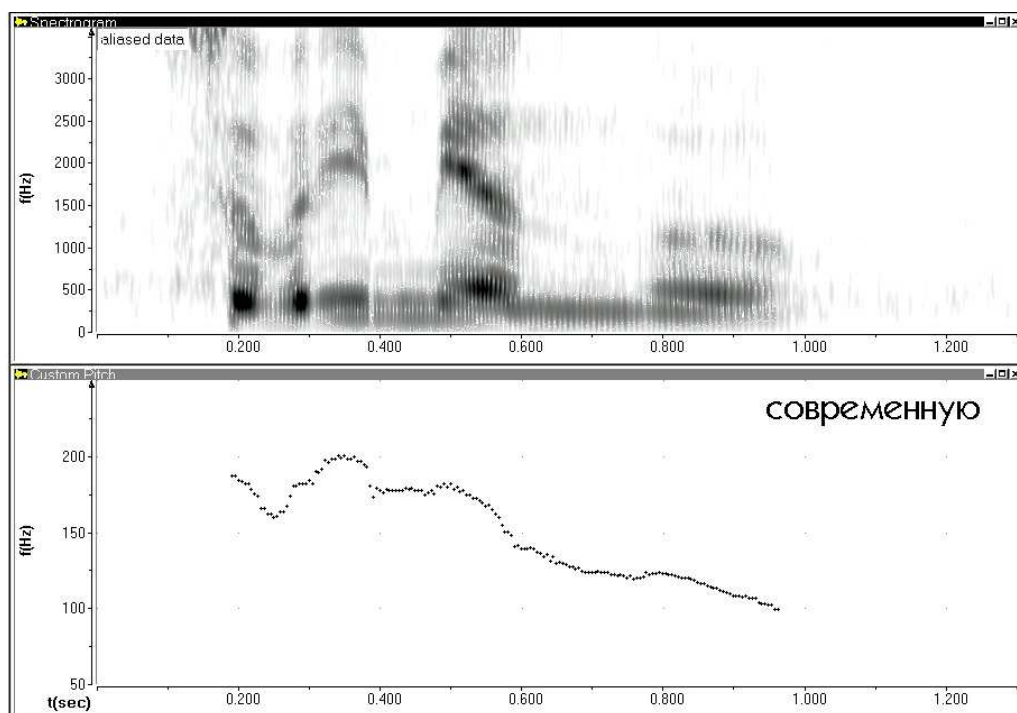


Рис. 1. Динамическая спектрограмма и мелодический контур «фразы» «современную».

В то же время отрезок речевого акустического сигнала между двумя последовательными паузами может содержать несколько синтагм в их общепринятом значении, причем на их стыках не обнаруживается никаких разрывов в мелодическом контуре, которые предположительно могли бы маркировать границы между синтагмами и одновременно являлись бы маркерами границ между словами (рис. 2). Более того, как следует из приведенного рисунка, на стыке двух синтагм может происходить стяжение гласных с полной утратой какой-либо физической границы между ними.

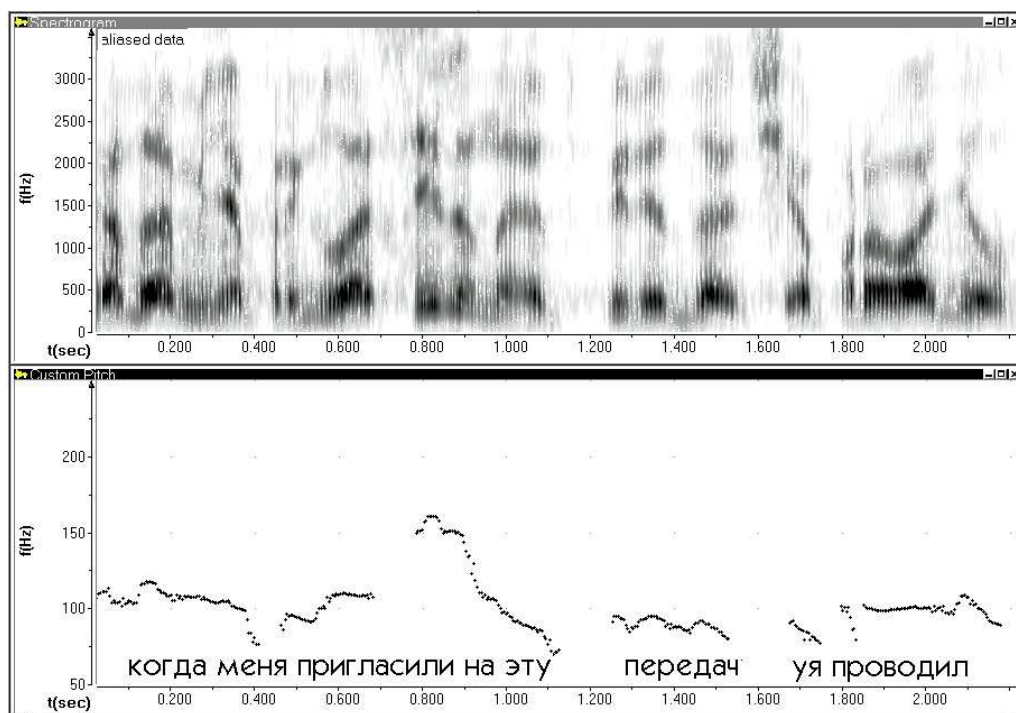


Рис. 2. Динамическая спектрограмма и мелодический контур части межпаузального отрезка спонтанной речи – «когда меня пригласили на эту передачу, я проводил такие мини-опросы общественного мнения, включил всю семью, всех друзей».

При этом следует заметить, что в приведенном примере значимое изменения частоты основного тона маркирует словоформу «пригласили» и, поскольку она оказывается не последней в синтагме, то анализ мелодической кривой не дает возможности определить даже конец синтагмы, не говоря уж о границе самой словоформы.

Таким образом, оказывается, что в подобных случаях анализ (аудиторский или инструментальный) мелодической составляющей просодических признаков не обеспечивает адекватного синтагматического членения. Об этом же говорят и результаты анализа спонтанной речи с целью установить позицию конца предложения: задача оказалась практически невыполнимой (Рыко; Степанова, 2009). Другими словами, в акустическом сигнале представляющем спонтанную речь не существует никаких физических признаков границ между лексическими единицами, кроме пауз.

Отмеченный выше эффект стяжения аллофонов (как гласных, так и согласных) на стыке словоформ является характерной особенностью спонтанной речи. Вот только несколько примеров:

всём_мире [fs'@+m'i+r*] 1
 десять_детей [d'E+s'id'it'E+j] 1
 идёт_со [id'@+ca] 1

 где_они [gd'E+n'i] 2
 если_она [jEsl'&na+] 1
 ему_абсолютно [imo+psal'y+tnQ] 1
 не_обсуждаю_ее [n'E+psuZdaj@] 1
 городе,_оценили [go+red'&sQn'i+l'E] 1
 профессии,_они [praf'E+s'&n'i] 1

Очевидно, что даже при условии, что каждая из «сиамских близнецов» представлена в перцептивном словаре носителя языка, метод прямой идентификации с использованием механизма «когорты» не даст адекватного разделения: либо одна из составляющих словоформ лишится начального (конечного) символа, либо будет начинаться (заканчиваться) «неправильным» гласным. В обоих случаях возникнет ошибка сегментации. И все же это несколько не отражается на способности носителей языка нормально воспринимать речь.

Выше уже говорилось, что для успешной работы алгоритма «сегментации через идентификацию» в перцептивном словаре слушателя должны храниться все словоформы, встречающиеся в распознаваемом тексте. Применительно к звучащему тексту это должны быть словоформы в их фонетической реализации, т.е. в форме, сопоставимой с первичным (аллофонным) описанием акустического речевого сигнала. Однако для спонтанной речи характерно большое разнообразие фонетических реализаций одной и той же лексической единицы.

действительно [d'istv'i+t'iln*] 1
 действительно [d'istv'i+tna] 1
 действительно [d'isv'i+t'il'naj] 1
 действительно [d'ist'i+tn] 1
 действительно [s'i+tna] 1
 человек [Cilav'i+k] 1
 человек [Clae+k] 1
 человек [Cilav'E+g] 1
 человек [\$Elav'E+k] 1

В этих условиях успешное функционирование процедуры сегментации через идентификацию с использованием перцептивного словаря возможно либо при условии, что последний содержит **все** когда-либо встречавшиеся варианты акустической реализации каждой словоформы языка, либо благодаря применению специальных алгоритмов сопоставления практически произвольной спонтанной фонетической реализации с некоторым кодифицированным представлением соответствующей лексической единицы в словаре.

Библиографический список

Комовкина Е.П., Слепокурова Н.А. Анализ межпаузальных интервалов в спонтанном тексте: предварительные результаты // Череповецкие научные чтения-2009: Материалы Всероссийской научно-практической конференции, посвященной Дню города Череповца (2-3 ноября 2009г.) / Ч. 1. Литературоведческие и лингвистические науки в начале XXI в.– Череповец: ГОУ ВПО ЧГУ, 2010.– С. 47–51.

Рыко А.И., Степанова С.Б. Стратегии членения спонтанной речи на синтаксические единицы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.).–Вып. 8 (15).– М.: РГГУ, 2009.–С. 438–443.

Kassevich V.B., Ventsov A.V., Yagounova E.V. The simulation of continuous text perceptual segmentation: A model for automatic segmentation of written text // Language and Language Behavior.– 2000.– V. 3, Pt. II.– SPb.– P. 48–59.