Анатолий Владимирович Венцов
Елена Игоревна Риехакайнен
Наталия Арсеньевна Слепокурова
Россия, Санкт-Петербургский государственный университет

Изучение восприятия устной речи: в поисках оптимального метода[1]

Ключевые слова: восприятие речи, устная речь, русский язык, методы психолингвистических исследований
Key words: spoken word recognition, speech, Russian language, methods of psycholinguistic research

Abstract

In the paper, spoken word recognition is considered as a black box system. The authors formulate basic methodological principles of spoken word recognition research based on the analysis of the experiments in the field and tested on spontaneous Russian.

Одним из способов исследования сложных систем, внутреннее строение которых недоступно прямому наблюдению, является «метод черного ящика». При таком представлении система-«черный ящик» рассматривается как имеющая некий «вход» для ввода информации и «выход» для отображения результатов работы системы, при этом происходящие в системе процессы наблюдателю неизвестны. Предполагается, что состояние выходов функционально зависит от состояния входов.

Очевидно, именно таковой является система восприятия речи человеком. Правда, некоторой информацией о ее внутреннем устройстве и принципах преобразования в ней речевого сигнала мы все же располагаем (строение наружного, среднего и внутреннего уха, преобразование сигнала на базилярной мембране внутреннего уха и в нервной системе) [Чистович и др. 1986]. Имеются также интуитивные представления о структуре и функциях

[1] В статье представлены результаты исследований, поддержанных грантами Президента Российской Федерации № МК-3646.2013.6 и № НШ-1778.2014.6.

ментального лексикона, о некоторых способах обработки лексической информации и правилах принятия решений. В итоге большинство исследователей согласны в том, что система восприятия речи – это интеллектуальная система, способная к постоянному самообучению, и ее реакции на речевой сигнал определяются не только его параметрами, но еще и ситуацией, обусловившей появление данного сигнала и способ реагирования на него, а также внутренними установками испытуемого, т. е. свойствами самой системы.

Если рассматривать когнитивную систему человека как «черный ящик», следует признать, что у нее не один «вход», а множество «внешних» (и речь – только один из них) и неизвестное число «внутренних» (сумма всех знаний слушателя, способность выбирать оптимальную стратегию поведения применительно к конкретной ситуации и др.). Тогда для успешной интерпретации «выхода» важно представлять себе содержание «сигналов» на всех «входах». В частности, необходима исчерпывающая информация о поступающем на вход системы речевом сигнале.

Разумно допустить, что состояние и поведение такой системы в экспериментальных условиях и при естественном речевом общении могут оказаться разными, и без учета вероятных отступлений от естественной речевой ситуации данные, полученные в эксперименте, в модель восприятия естественной речи переносить невозможно.

Изучение системы восприятия устной речи предполагает исследование когнитивных процессов, обеспечивающих переход от акустического речевого сигнала – через лексический, грамматический и семантический анализ – к пониманию смысла речевого сообщения. В перспективе имеется в виду создание функциональной модели восприятия речи, которая по конечному итогу своей «деятельности» была бы сравнима с результатами, обнаруживаемыми у носителя языка.

Несмотря на то, что исследования процессов речевосприятия ведутся во всем мире уже долгие годы, в этой области до сих пор остается множество нерешенных проблем.

В частности, разумное допущение о том, что модель восприятия речи прежде всего должна предусматривать преобразование непрерывного по своей природе акустического речевого сигнала в последовательность дискретных лексических единиц, требует ясного понимания того, что собой представляет преобразуемый сигнал и какие возможные отклонения от «идеальной» реализации лексических единиц языка в нем наблюдаются. Предположим, что это – непрерывная последовательность символов некоторого внутреннего представления акустического сигнала в слуховой системе (для простоты – символов фонетической транскрипции, но, конечно, не фонем и ни при каких условиях не символов орфографии), тогда возникает необходимость анализировать большие объемы звучащих текстов, снабженных акустико-фонетической разметкой.

Главная проблема состоит в том, что ввиду отсутствия технических средств автоматического фонетического транскрибирования получить надежные данные о свойствах речевого сигнала возможно лишь через «ручной» анализ речевого материала. В некоторых работах, впрочем, для этих целей используется инструментарий автоматического распознавания речи, но, по отзывам самих исследователей, результаты его работы настолько ненадежны, что приходится прибегать к дополнительному исправлению результатов «вручную», затрачивая на это массу квалифицированного труда и времени. Как следствие, в большинстве корпусов звучащих текстов отсутствует акустико-фонетическая расшифровка: все ограничивается орфографической аннотацией.

Авторы настоящей работы придерживаются мнения, что наряду с орфографической необходима и сплошная «ручная» акустико-фонетическая расшифровка всех записей. Для минимизации влияния лексических знаний экспертов, анализировать при этом следует отрезки сигнала не длиннее слога, а результаты слухового анализа текущим образом сопоставлять с результатами инструментального анализа в виде динамических спектрограмм. Расшифровки для нашего корпуса устной русской речи доступны на сайте http://www.narusco.ru/ в разделе «Наши ресурсы» → «Поиск по текстам речевого корпуса». Подробности этой части нашей работы можно найти в [Нигматулина, Раева 2015].

На основе полученных расшифровок создан частотный словарь акустических реализаций словоформ, в котором наглядно демонстрируется звуковой «облик» каждой словоформы. Качественный и количественный анализ содержащегося в нем материала позволяет успешнее планировать дальнейшие психолингвистические исследования. Существенную помощь он может оказать и преподавателям русского языка как иностранного.

Результаты анализа имеющегося у нас затранскрибированного речевого материала объемом около 1,5 часов звучания уже позволили обнаружить множество проблем, связанных с функционированием системы речевой перцепции, и главная среди них – проблема сегментации непрерывного речевого потока на последовательность дискретных лексических единиц.

Обсуждаемые в литературе модели лексического поиска обычно включают процедуру соотнесения текущего «символьного» описания речевого сигнала с единицами некоего гипотетического словаря. Присутствие в речевом сигнале сильно редуцированных словоформ и возможность образования стяжений на стыках словоформ вынуждают искать ответы на два вопроса: 1) что собой представляют единицы такого словаря и 2) по каким правилам производится сопоставление текущего речевого сигнала с этими единицами. Получить ответ на эти вопросы можно только путем формулирования соответствующих гипотез и проверки их в психолингвистическом эксперименте.

В свое время при анализе методов психоакустических экспериментов с речевыми и речеподобными сигналами было четко показано, что их результаты

не могут быть использованы для создания модели восприятия речи [Венцов, Касевич 1994]. Стало ясно, что, используя в эксперименте многократно повторяющиеся однотипные стимулы и задавая в инструкции фиксированное число возможных реакций, исследователь «позволяет» испытуемому в ходе эксперимента вырабатывать собственный, необязательно связанный с ожиданиями экспериментатора, «оптимальный» алгоритм поведения *ad hoc* и тем самым разделять стимулы на заданные в инструкции классы по этим «одноразовым» правилам. Однако в перцептивных исследованиях именно возможные алгоритмы поведения испытуемого, т. е. активируемые в экспериментальной ситуации «внутренние входы» «черного ящика», не анализируются, не оценивается и то, насколько они соответствуют ожиданиям исследователя и ситуации естественного речевого поведения.

К сожалению, большинство описываемых в литературе психолингвистических экспериментов, включая самые недавние, воспроизводят эту же ситуацию со всеми упомянутыми выше недостатками и, как следствие, их результаты мало что говорят о реальных процессах восприятия речи.

Главный недостаток этих работ состоит в том, что и при планировании эксперимента, и при обработке и анализе его результатов исследователи не учитывают возможное влияние «сигналов» на «внутренних входах» «черного ящика» – когнитивной системы испытуемого. Тогда как при внимательном анализе зачастую оказывается, что испытуемый действовал по «своим» правилам и к заявленным в работе целям результаты эксперимента не имеют никакого отношения.

В последние годы возросшие мощности вычислительной техники породили у психолингвистов наивную веру в то, что использование громоздких компьютеризованных технических средств позволит достичь небывалых успехов в исследовании механизмов восприятия речи человеком, вплоть до фиксации в реальном времени самого когнитивного процесса (так называемые онлайн методы), а не только его результата.

Весьма показательна работа *Speech reductions change the dynamics in competition during spoken word recognition* [Brouwer и др. 2012], целью которой было исследование процесса обработки «конкурирующих» лексических единиц в процедурах распознавания речи при наличии в речевом сигнале редуцированных словоформ.

В предлагаемых алгоритмах лексической идентификации речевых единиц неизбежно присутствует механизм образования «когорты» словоформ, по ряду признаков совпадающих с анализируемым речевым сигналом. Часть из них оказываются «конкурентами» (*competitors*) и по мере накопления информации устраняются. Точные принципы формирования «когорты» пока неизвестны, но можно с большой долей уверенности утверждать, что образующие ее элементы описаны в терминах внутреннего представления акустического сигнала в слуховой системе.

В указанной работе испытуемым предъявлялись извлеченные из спонтанной речи пары фраз, содержащие одну и ту же словоформу – «каноническую» или редуцированную. Испытуемым на экране компьютера заранее предъявляли четыре варианта ответа в орфографии: полный вариант произнесения исследуемой словоформы, потенциальные «конкуренты» для полного и редуцированного вариантов и отвлекающее слово (*distractor*).

В соответствии с инструкцией, пользуясь указателем компьютерной мыши, испытуемые должны были отметить на экране с л о в о, прозвучавшее в прослушанной ими фразе. Одновременно фиксировались движения глаз испытуемого и время задержки реакции.

Таким образом, свойства одной практически не изученной когнитивной системы (восприятия речи) исследовались через анализ поведения другой мало изученной системы (зрительной) – и при этом без учета возможностей двигательной системы (управление компьютерной мышью) конкретного испытуемого.

Даже без анализа количественных результатов эксперимента ясно, что его результаты не имеют отношения к проблеме «конкурентов» при распознавании (*recognition*) речевого сигнала: по сути, в эксперименте исследовалась «конкуренция» при выборе ответа – испытуемым приходилось сначала «распознавать» все слова в составе предъявляемой фразы, а затем сопоставлять их с заданными на экране компьютера «ответами».

Этот вывод косвенно подтверждают данные о движении глаз. Авторы с удивлением констатируют, что потенциальные «конкуренты» для полного и для редуцированного вариантов произнесения в равной степени привлекают внимание испытуемых в процессе формирования ответной реакции [Brouwer и др. 2012: 554], но игнорируют тот факт, что ровно столько же внимания отдается и отвлекающему слову (*distractor*). Из этого следует, что оба предполагаемых «конкурента» и отвлекающее слово представляли для испытуемого единый тип ответа: они все были отвлекающими. И понятно почему: ни один из них не соответствовал услышанной фразе по семантике.

В целом в эксперименте получен банальный результат: редуцированные словоформы распознавались хуже (больше ошибок) и правильные ответы требовали большего времени реакции (при усреднении по всем испытуемым). Последнее можно принять только с большими оговорками, так как алгоритм обработки результатов измерения в работе не описан и можно предположить, что имело место типичное для психолингвистических работ неграмотное применение статистических методов.

Регистрируемое в эксперименте время реакции складывается из времени лексического поиска (практически мгновенный подсознательный процесс), длительности процесса осознанного выбора и продолжительности двигательной реакции (перемещение мыши и нажатие клавиши). Кроме того, время сенсомоторной реакции может у разных испытуемых значительно варьировать,

а значит, сведение всех данных в одну выборку требует предварительного «нормирования» индивидуальных данных. К тому же, распределение времен реакции существенно отличается от нормального, различие это принципиальное, и при обработке такого материала следует использовать преобразования, приводящие эти распределения к нормальному.

По мнению авторов, трудности идентификации редуцированных форм проявляются в менее частой фиксации на них взгляда по сравнению с каноническими [Brouwer и др. 2012: 554], однако обоснования допустимости такого вывода в работе нет и полностью игнорируется то обстоятельство, что канонические и редуцированные формы предъявлялись разным испытуемым. При этом временная динамика процесса фиксации взгляда позволяет предположить, что при распознавании речевого сигнала она отражает не процесс лексического поиска (*lexical access*), а процесс «чтения» ответа на экране компьютера, и только это, собственно, и позволяет онлайн методика.

Следовательно, подобный метод, возможно, являющийся удачным и перспективным при изучении распознавания письменной речи, не может применяться при исследовании устной речи.

В процессе естественного речевого общения слушатель (испытуемый) имеет дело с речевым сигналом, обладающим грамматической и семантической структурой, и оценки этого сигнала формируются в соответствии с хранящимися у него знаниями (лексическими, грамматическими, семантическими). Очевидно также, что реакция слушателя на услышанное может быть весьма разнообразной: от некоторой поведенческой реакции до ответного высказывания либо принятия к сведению без каких-либо внешних проявлений и т. д. Но грамотный слушатель всегда способен более или менее точно зафиксировать услышанное в письменной форме.

Учитывая все вышесказанное, мы предпочитаем методики, при реализации которых испытуемый находится в условиях, близких к ситуации восприятия естественной речи, когда его выбор определяется лишь свойствами конкретного акустического сигнала и суммой имеющихся у него знаний о речи и языке. Поэтому основным используемым нами методом является эксперимент, предполагающий прослушивание стимулов и их орфографическую фиксацию (*dictation task*) [о применении этого метода см. также: Taft 1984; Ernestus и др. 2002].

Фрагменты, предлагаемые для прослушивания, в зависимости от конкретной задачи могут быть как осмысленными, так и асемантичными. Асемантичные стимулы необходимы, когда исследуется интерпретация носителем языка конкретного акустического признака речевого сигнала на этапе долексического анализа. Подобный подход позволяет получить «объективную» оценку звучания фрагментов устной речи в отвлечении от субъективных интерпретаций исследователя и лексических знаний

испытуемого, а также дает возможность сопоставить ее с результатами инструментального анализа.

Можно отметить некоторые сложности, которые возникают при использовании асемантичных фрагментов:

1) подобные фрагменты часто оказываются очень короткими, и одного предъявления стимула оказывается недостаточно, поэтому в ряде случаев короткие стимулы приходится либо предъявлять два–три раза, либо увеличивать их длительность;
2) при восприятии асемантичного фрагмента речи «нулевой гипотезой […] всегда выступает презумпция осмысленности: любое речевое произведение человек сначала пытается интерпретировать как осмысленное…» [Касевич 2006 (1988): 593]; испытуемый способен делать это даже вопреки предупреждению, что ему будут предъявляться асемантичные фрагменты. В этих случаях приходится дополнительно анализировать результаты: вызван ли осмысленный ответ акустическими характеристиками исследуемого речевого сегмента или он просто отражает лексические ассоциации испытуемого; в любом случае всегда есть возможность не учитывать такого рода реакции испытуемых.

Эксперименты с использованием методики восприятия слов на слух (как на асемантичных, так и на осмысленных фрагментах русской речи) позволили получить новые данные, устанавливающие связь между акустической реализацией звучащей речи на сегментном уровне и интерпретацией этой акустической данности слушателем [Риехакайнен 2010; Нигматулина, Риехакайнен 2011; Раева 2012; Апушкина и др. 2014 и др.].

Таким образом, с нашей точки зрения, не следует искать единственный метод исследования процесса восприятия речи и тем более использовать модный: специалисты, работающие в этой области, должны подбирать (разрабатывать) методику под каждую конкретную исследовательскую задачу, т. е. находиться в постоянном поиске оптимального метода в зависимости от поставленной цели, однако при этом необходимо руководствоваться некоторыми общими методологическими принципами, которые будут отражать специфику области исследования:

– планируя эксперимент, следует прежде всего представить себе, какой процесс исследуется, составив его детальное описание; предположить, как будет вести себя испытуемый в условиях эксперимента и насколько его поведение будет отличаться от поведения в естественных условиях;
– использовать в качестве экспериментального материала записи естественной речи;
– пользоваться методами, при реализации которых испытуемый находится в ситуации, максимально приближенной к ситуации естественного речевого общения;

– разрабатывать методики под каждую конкретную исследовательскую задачу, проверять их в пилотных экспериментах, проводить сопоставительный анализ результатов, полученных в различных экспериментальных условиях;
– при статистической обработке экспериментальных данных первым делом оценивать необходимость и допустимость применения тех или иных критериев для анализа конкретного материала.

Литература

Апушкина И. Е., Венцов А. В., Слепокурова Н. А., 2014, *Альбом динамических спектрограмм безударных двуслогов, выделенных из спонтанной речи, и результатов идентификации носителями русского языка фонемного качества гласных в их составе*, http://www.narusco.ru/ALBUM01 (дата обращения: 27.11.2014).

Венцов А. В., Касевич В. Б., 1994, *Проблемы восприятия речи*, Санкт-Петербург: Издательство Санкт-Петербургского университета.

Касевич В. Б., 2006 (1988), Семантика. Синтаксис. Морфология [в:] Ю. А. Клейнер (ред.), *Труды по языкознанию*, т. 1, Санкт-Петербург: Филологический факультет СПбГУ, с. 373–612.

Нигматулина Ю. О., Раева О. В., 2015, Исследование русской спонтанной речи: новые методы – новые результаты [в настоящем сборнике].

Нигматулина Ю. О., Риехакайнен Е. И., 2011, Сегментация спонтанной речи: восприятие стяжений гласных на стыке словоформ [в:] Е. В. Ерофеева (ред.), *Проблемы социо- и психолингвистики*, вып. 15: *Пермская социопсихолингвистическая школа: идеи трех поколений: К 70-летию Аллы Соломоновны Штерн*, Пермь: [б. и.], с. 31–38.

Раева О. В., 2012, Стратегия распознавания редуцированных вариантов высокочастотных единиц [в:] Е. М. Девяткина (отв. ред.), *Проблемы языка: Сборник научных статей по материалам Первой конференции-школы «Проблемы языка: взгляд молодых ученых» (20–22 сентября 2012 г.)*, Москва: Институт языкознания РАН, с. 241–252.

Риехакайнен Е. И., 2010, Влияние потенциального контекста на распознавание изолированных омофонов, *Вестник Пермского университета*, серия: *Российская и зарубежная филология*, вып. 4 (10), с. 40–45.

Чистович Л. А., Венцов А. В., Люблинская В. В., Столярова Э. И., Чистович И. А., 1986, *Слуховые уровни восприятия речи. Функциональное моделирование* [в:] Л. А. Чистович (ред.), *Акустика речи и слуха: Сборник научных работ*, Ленинград: Наука, с. 97–127.

Brouwer S., Mitterer H., Huettig F., 2012, Speech reductions change the dynamics in competition during spoken word recognition, *Language and Cognitive Processes*, Vol. 27 (4), с. 539–571.

Ernestus M., Baayen H., Schreuder R., 2002, The recognition of reduced word forms, *Brain and Language*, Vol. 81 (1–3), с. 162–173.

Taft M., 1984, Exploring the mental lexicon, *Australian Journal of Psychology*, Vol. 36, с. 35–46.