

От психоакустики к психолингвистике

The widely spread methods of psychoacoustical and psycholinguistical experiments are analyzed. It was shown that results of the experiments could not be used in functional model of human speech perception system, because subject's behavior in the experimental situation differs strikingly from they behavior in natural speech communication.

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 интуитивные представления о структуре и функциях 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, его способность выбирать оптимальную стратегию поведения применительно к конкретной ситуации, неконтролируемые экспериментатором его подсознательные внутренние установки и др.). Тогда для успешной интерпретации «выхода» важно представлять себе содержание «сигналов» на всех «входах». В частности, необходима исчерпывающая информация о попадающем на вход системы речевом сигнале.

Разумно допустить, что состояние и поведение такой системы в экспериментальных условиях и при естественном речевом общении могут оказаться разным, и без учета вероятных отступлений от естественной речевой ситуации окажется невозможным использовать данные, полученные в эксперименте, при создании функциональной модели восприятия естественной речи.

Психоакустика – раздел экспериментальной психологии (психофизики), посвященный исследованию субъективного восприятия человеком звуковых сигналов (как сказали бы сейчас, обработки звуковых сигналов когнитивной системой человека).

В лабораториях, которыми руководили В.А.Кожевников и Л.А.Чистович, такими сигналами были сложные речеподобные сигналы, поскольку ставилась задача создания функциональной модели системы восприятия речи человеком, т.е. такого набора устройств или компьютерных алгоритмов, которые после обработки речевого сигнала выдавали бы результат, сопоставимый с тем, что услышит носитель языка.

В то время перед физиологами ставилась задача исследовать алгоритмы выделения из акустического речевого сигнала признаков, необходимых для последующей «лингвистической» обработки. Предполагалось, что принципы и конкретные правила таковой сформулируют лингвисты, в частности, фонетисты выдадут перечень тех

признаков, которые должны быть извлечены из акустического речевого сигнала для последующей адекватной фонемной, грамматической и семантической интерпретации. В известной степени такой подход стимулировался и практическими потребностями разработчиков систем автоматического распознавания речи (САРР), которые в то время строились на «пофонемном» первичном описании речевого сигнала, основанном на результатах спектрального анализа речевого сигнала. В терминах современных САРР, использовались акустические модели фонемного (аллофонного) уровня.

Ни принципы «лингвистической» обработки, ни структура наборов субъективных признаков, необходимых для адекватного описания непрерывного речевого сигнала, до сих пор не сформулированы. Сейчас можно утверждать, что и не могли быть сформулированы.

Экспериментальная фонетика исследует акустические особенности реализации языковых единиц разного уровня в речи носителей языка. В результате они получают сведения о некоторых усредненных параметрах русского речевого сигнала, но никогда не проверяют, насколько эти параметры могут быть использованы в алгоритмах текущего анализа естественного речевого материала. Да собственно говоря, это не их задача, равно как и разработка самих алгоритмов обработки сигнала при восприятии речи.

Тем не менее, считалось, что задача физиологов – найти принципы обработки речевого сигнала слуховой системой, которые позволяли бы выделять именно эти характеристики.

Поэтому методами психоакустики исследовались принципы субъективного описания спектрально-временной структуры сложных речеподобных сигналов и те характерные признаки, которые предположительно могли бы быть использованы при восприятии естественной речи. Действительно, многое было найдено и даже реализовано в компьютерных моделях, но касалось это в большей степени механизмов преобразования сложных акустических сигналов на периферии слуха, а не механизмов восприятия речи.

Но к великому сожалению, использовать даже эти находки в модели восприятия реального речевого сигнала оказалось невозможным. Прежде всего потому, что не была сформулирована целостная модель восприятия речи и не ясно было, какие именно из параметров сигнала, измеряемых фонетистами, следует использовать в модели и по каким правилам. Кстати, такая модель не сформулирована до сих пор.

Показательна в этой связи история исследования признаков словесного ударения в русском языке. Существует единое мнение, что в русской речи таким признаком является длительность гласного. Вопрос: так ли это и о какой длительности идет речь. Последнее особенно важно, поскольку измерение длительности возможно только при условии, что заранее заданы моменты начала и конца оцениваемого временного промежутка, а до сих пор не ясно, какие моменты времени слуховая система принимает за начало и конец гласного.

Исследование длительностей ударных и безударных гласных (Бондарко и др. 1973) показало, что абсолютные длительности гласных таким признаком быть не могут: длительности гласных изменяются при изменении темпа речи, распределения длительностей ударных и безударных гласных существенно перекрываются и при этом сказывается эффект «собственных» длительностей гласных – ударный [i] может быть короче безударного [a] (рис. 1,А).

В качестве возможной альтернативы рассматриваются относительные длительности и лучший результат дает отношение длительности безударного гласного к длительности ударного в этом же слове (рис.1,Б).

Однако при этом нередки случаи, когда полученное отношение равно или больше единицы, т.е. безударный гласный длительнее ударного, и все еще сохраняется эффект «собственной» длительности гласных, хотя вариативность, связанная с изменением темпа речи, нейтрализуется.

И тем не менее, «... встает вопрос о том, насколько полезным является этот простой и устойчивый признак (отношение длительности безударного гласного к длительности ударного гласного в том же слове – А.В.) при определении места ударения человеком — носителем русского языка.» (Бондарко и др. 1973: 148).

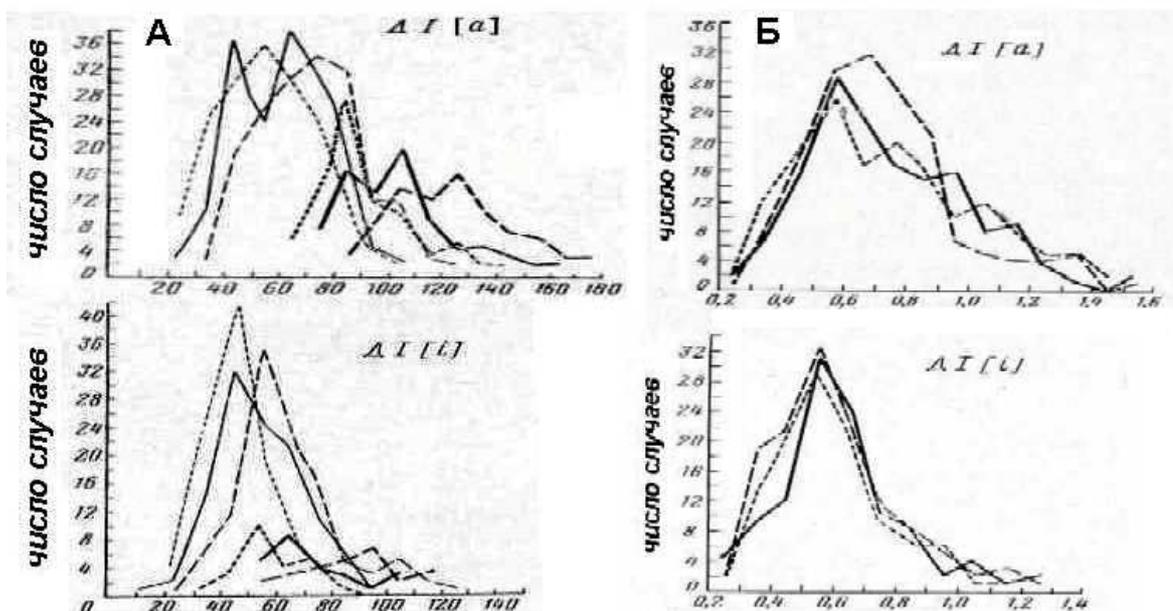


Рис. 1. Результаты инструментального анализа длительностей ударных и безударных гласных. По: Бондарко и др. 1973.

На А: распределение длительностей ударных и безударных гласных в произнесении диктора I; вверху – гласный [а], внизу – гласный [i]: по оси абсцисс – длительность в мс.

На Б: распределение относительной длительности безударных гласных; по оси абсцисс – отношение длительности безударного гласного к длительности ударного в этом же слове; остальные обозначения – как на А.

Заметим прежде всего, что некорректна сама формулировка вопроса: при инструментальных измерениях речевого сигнала исследователь знает заранее позицию ударного гласного (слога), соответственно, не составляет труда оценить длительности ударного и безударного гласных и их отношение; при восприятии речи слушателю позиция ударения априори неизвестна, и обнаружить ее он может только путем текущего анализа сигнала, а уж если он ее обнаружил, последующее сопоставление длительностей гласных становится бессмысленным.

Правда, в проведенном авторами психоакустическом эксперименте использовались знаменательные двусложные слова (му+ку¹ – муку+, пи+ли – пили+ и т.п.), так что, с одной стороны, у испытуемых не возникало проблем с вычислением отношения длительностей (достаточно было сопоставить длительности соседних гласных в предлагаемых стимулах), а с другой, выбирая одну из двух заранее известных позиций, они могли пользоваться не только признаком «длительность», и правила эти были недоступны экспериментатору. При этом сразу же возникает вопрос: что станет делать носитель языка, пытаясь определить место ударения в трехсложном слове? Отношения длительностей каких гласных ему следует использовать в качестве признака? И как быть с «собственными» длительностями гласных? Даже сегодня на такие вопросы нет ответа ни у кого. Естественно, не было их и сорок лет назад.

¹ Символ «+» маркирует ударность предшествующего гласного.

Как правило, в психоакустических экспериментах использовались методики разработанные в экспериментальной психологии, когда испытуемому предъявляется последовательность из многократно повторяющихся стимулов с ограниченным набором свойств, а инструкция предусматривает заранее заданные варианты реакций.

В свое время анализ методов психоакустических экспериментов с речевыми и речеподобными сигналами обнаружил, что полученные результаты не могут быть использованы при создании модели восприятия речи (Венцов, Касевич 1994). Было показано, 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 разделять стимулы на заданные в инструкции классы, пользуясь своими внутренними правилами. Однако в перцептивных исследованиях именно возможные алгоритмы поведения испытуемого, т.е. активируемые в экспериментальной ситуации «внутренние входы» «черного ящика», не анализируются, не оцениваются и то, насколько эти особенности поведения соответствует ожиданиям исследователя и ситуации естественного речевого поведения.

Как следствие, получаемые в таких исследованиях результаты («эталоны», дифференциальные пороги и т.п.) в большей степени отражают условия конкретного эксперимента (количество и диапазон изменения параметров стимулов, тип инструкции, количество испытуемых и пр.), чем ситуацию естественного речевого общения, и оказывается, что их невозможно использовать в работоспособной функциональной модели восприятия естественной речи.

Там же даны некоторые рекомендации по организации психоакустического эксперимента, следуя которым можно было бы приблизить экспериментальные условия к ситуации естественного речевого общения (Венцов, Касевич 1994: 202).

В последние годы стало понятно, что в естественной речи лексические единицы могут подвергаться настолько сильной редукции, что их адекватная идентификация с использованием только акустической информации становится невозможной и приходится обращаться к текущему грамматическому и семантическому анализу высказывания (Раева, Риехакайнен 2014). Соответственно возникает необходимость в новых методах экспериментальных исследований.

Если психоакустика – это экспериментальное исследование субъективных оценок акустических параметров сложного сигнала (в частности, речевого), то психолингвистика, по аналогии, должна заниматься экспериментальным исследованием субъективных оценок носителями языка некоторых лингвистических свойств речевого сигнала.

Судя по литературе, большинство такого рода исследований посвящено попыткам решить проблему преобразования непрерывного речевого потока в последовательность единиц лексического уровня и получить данные о возможной структуре и единицах перцептивного словаря, как части (первого уровня) ментального лексикона.

Содержание же этих публикаций, чаще всего, позволяет отнести их к психолингвистике только потому, что исследованиями вместо физиологов и психологов занимаются лингвисты: по-прежнему исследуются акустические характеристики речевого сигнала, но в качестве оценок вместо позиции ударения, качества гласных, места и способа артикуляции согласных используются типы акцентного выделения, границы словоформ или синтагм и т.п..

При этом создается впечатление, что в своих исследованиях звучащей речи лингвисты (или психолингвисты?) исходят из моделей языка, разработанных при анализе письменных текстов, и в результате пытаются искать в акустическом речевом сигнале то, чего в нем может не быть и без чего носитель языка прекрасно справляется с задачей понять сказанное.

Так, в работе (Schaegis et al. 2005) было обнаружено, что в АВХ–эксперименте с омофонами типа “l’affiche (афиша) – la fiche (лист бумаги), l’amie (подруга) – la mie (хлебный мякиш)” носители французского языка в 75% случаев дают правильные ответы. Из этого делается вывод, что испытуемые «сегментировали» предъявляемые стимулы, отделяя определенный артикль.

Однако, полученный результат может иметь совершенно иное объяснение: испытуемые знали, что первые два сигнала в АВХ–последовательности всегда разные, и могли идентифицировать их различие, пользуясь **любыми** акустическими признаками, даже не прибегая к сегментации, и по этим же признакам оценить третий сигнал в последовательности. Инструментальный анализ, проведенный авторами исследования показал, что использованные ими омофоны действительно различаются акустически.

При этом авторам работы не приходит в голову, что в естественной речи (особенно в спонтанной) этих акустических различий может не быть и в перцептивном словаре носителя языка подобные омофоны могут быть представлены единой «словарной единицей» – фонетическим словом, имеющим несколько значений по аналогии с русскими «заработай – за работой; вначале – в начале; стой – с той», а неоднозначность разрешается либо на уровне контекста, либо на основании соотношения частотностей этих словоформ в языке или в словаре конкретного носителя языка.

Характерный пример приводят сами участники описываемого исследования: “... while setting up for a presentation of the current research, the first author was asked if she had *l’alim* ‘the power cord’ for the laptop computer; missegmenting this as *la lime* ‘the nail file’, she replied ‘*Pourquoi j’aurais besoin d’une lime ?!*’ ‘Why would I need a nail file?!’ We have all encountered similar missegmentations.” (Spinelli et al. 2007: 829).

Нетрудно представить себе, что для женщин-лингвистов частотности понятий «шнур питания к лаптопу» (*l’alim*) и «пилочка для ногтей» (*la lime*) могут существенно различаться и именно этим объясняется обнаружившаяся ошибка в «сегментации».

Очевидно, что к модели поведения носителя языка при восприятии естественной речи результаты подобного исследования не имеют никакого отношения: они просто характеризуют поведение испытуемого в условиях конкретного эксперимента, причем условия эти не имеют ничего общего с условиями естественного речевого общения, когда «сегментация» осознается только грамотным носителем языка, т.е. после овладения им правилами орфографии.

Очевидно также, что любая публикация с результатами психолингвистического эксперимента должна содержать описание того предполагаемого процесса в когнитивной системе человека, который предстоит исследовать, с указанием: что и как измеряется, что с чем и как сравнивается при принятии решения, по каким критериям. И это обязательное требование к процессу планирования любого эксперимента. Практика показывает, что ни в одной работе, посвященной результатам психолингвистического эксперимента, подобного раздела обнаружить не удается.

Соответственно, в подобных работах отсутствует анализ возможных стратегий, используемых испытуемыми в эксперименте, и того, в какой степени полученные результаты объясняются именно особенностями индивидуальных стратегий испытуемых. Тогда как испытуемые способны действовать в соответствии со своей собственно «инструкцией», выработанной в ходе эксперимента с учетом некоторых генеральных требований инструкции, заданной исследователем, а также структуры и свойств конкретного экспериментального набора стимулов.

Таким образом, обнаружив некоторый эффект в эксперименте психолингвисты даже не задаются вопросом, как этот эффект может быть реализован носителем языка при текущем анализе естественного речевого сигнала, а сплошь и рядом выясняется, что к когнитивным процессам при восприятии речи эти результаты не имеют никакого отношения.

Еще одно «несчастье» психолингвистических исследований – стремительное развитие вычислительной техники и сплошная компьютеризация всех исследований.

Лет 30-40 тому назад в публикациях представителей гуманитарных наук чувствовалось, что расчеты произведенные на компьютере обладают для них абсолютной истинностью по-умолчанию. Сегодня, с их точки зрения, экспериментальный материал не заслуживает внимания, если он не обработан статистическим пакетом ANOVA, причем необходимость и допустимость его использования для анализа статистических характеристик конкретных экспериментальных данных обычно не обсуждается и не обосновывается.

Одновременно наблюдается чудовищное усложнение технической организации экспериментов, хотя их принципиальная сущность ничем не отличается от методик психоакустических экспериментов. К примеру, при исследовании особенностей восприятия акустических параметров речевого сигнала в качестве ответной реакции регистрируются еще и движения глаз с помощью компьютеризованного ай-трекера (регистратора движений глаз), т.е. функционирование одной неизученной когнитивной системы изучают через реакции такой же неизученной системы.

В этом смысле весьма показательна работа (Brouwer et al. 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 из хранящихся в ментальном лексиконе слушателя элементов “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.

Авторы упомянутой выше работы предполагали, что в эксперименте, где ответом на прозвучавшую в предъявленной фразе редуцированную словоформу будет выбор одного из нескольких представленных в орфографической форме на экране компьютера слов, регистрация движения глаз позволит в реальном времени зафиксировать динамику подсознательного по своей природе процесса лексического выбора в ментальном словаре.

Однако внимательный анализ использованных речевых стимулов, инструкции испытуемым и самих результатов эксперимента позволяет сделать вывод, что зарегистрированные движения глаз характеризуют процесс поиска и считывания ответа из предлагаемого на экране компьютера набора, но ни коим образом не лексический поиск. Таким образом, в исследованиях процессов восприятия речи совершенно бесполезно использовать сложнейшую технику регистрации движений глаз.

Таким образом, не следует использовать модный или искать единственный метод исследования процесса восприятия речи: психолингвистам следует подбирать (разрабатывать) методику в зависимости от поставленной задачи, однако при этом необходимо руководствоваться некоторыми общими методологическими принципами, которые будут отражать специфику области исследования:

- планируя эксперимент, 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;
- пользоваться методиками, при реализации которых испытуемый находится в ситуации максимально близкой к условиям естественного речевого общения;
- разрабатывать методики под каждую конкретную исследовательскую задачу;

- при статистической обработке экспериментальных данных первым делом оценивать необходимость и допустимость применения тех или иных критериев для анализа конкретного материала.

Литература

Бондарко Л. В., Вербицкая Л. А., Щербакова Л. П. Об определении места ударения в слове // Известия Академии наук СССР. Серия литературы и языка.– 1973.– Т. 32, Вып. 2.– С. 141-153.

Венцов А.В., Касевич В.Б. Проблемы восприятия речи.– СПб.: Изд-во С.-Петербургского ун-та, 1994.– 232 с.

Раева О.В., Риехайнен Е.И. Предсказательная сила контекста: миф или реальность? // Шестая международная конференция по когнитивной науке: Тезисы докладов. Калининград, 23–27 июня 2014 г.– Калининград, 2014.– С. 742–744.

Чистович Л.А., Венцов А.В., Люблинская В.В., Столярова Э.И., Чистович И.А. Слуховые уровни восприятия речи. Функциональное моделирование // Акустика речи и слуха: Сборник научных работ / Ред. Л.А. Чистович.– Л.: Наука, 1986.– С. 97-127.

Brouwer S., Mitterer H., Huettig F. Speech reductions change the dynamics in competition during spoken word recognition // Language and Cognitive Processes.– 2012.– Vol. 27, No 4.– P. 539-571.

Schaegis, A. L., Spinelli, E., Welby, P. Perception of phonemically identical spoken sequences in French // 27th annual meeting of the Cognitive Science Society, July 21–23, 2005, Stresa, Italy.– 2005.– P. 1937-1942.

Spinelli E., Welby P., Schaegis A. L. Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of French elision // Language and Cognitive Processes.– 2007.– Vol. 22, No. 6.– P. 828-859.