

Д.К. Куликов  
ТЕОРИИ ВОСПРИЯТИЯ РЕЧИ В ПСИХОЛОГИИ.

(Психология и безопасность: Материалы научной конференции. – Таганрог: Изд-во ТТИ ЮФУ, 2012. С. 144-151)

(144)

Изучение восприятия речи – это широкое исследовательское поле науки на стыке психологии, физиологии, лингвистики, информатики и пр. Знания в данной области востребованы в технической сфере (например, создание программ распознавания речи), в клинической деятельности (патологии речи и слуха), в информатике (создание «искусственного интеллекта») и др.

Восприятие речи предполагает взаимодействие двух процессов: регистрацию акустического вербального потока и интерпретацию смысла. В живом восприятии разделить эти стороны очень трудно, хотя *чувственная ткань* (звуки) и *смысл речи* – это не одно и то же. Понимание речевого потока – это не только его членение в процессе восприятия, но и извлечение смысла. Иначе говоря, звучащая речь, в отличие от невербальных звуков, несет в себе не только (145) информацию о своем источнике, но и интегрированный в слово смысл. Главная особенность вербального стимула – его потенциальная доступность пониманию. И хотя сам факт перцепции слова не всегда желателен, его произнесение предполагает восприятие.

Восприятие речевого потока требует выделение особой единицы – фонемы или осмысленного звука. Это такой звук, который влияет на наше понимание значений и речи в целом. Фонема – это минимальная единица речевого потока, которая позволяет отличить одно слово от другого или изменить воспринимаемое значение. В качестве примера возьмем акустическое слово *луг*, состоящее из трех фонем. Заменяя одну из них, мы получаем слова с другими значениями: *Буг, лаг, лук*.

В звучащей речи выделяются звуки двух основных типов – гласные и согласные. И те и другие имеют свои важные функции для обеспечения осмысленности речи при ее восприятии. Гласные звуки – низкочастотные и имеют большую длительность. Согласные это высокочастотные звуки, которые имеют большой разброс частот. В работе Френча и Стейнберга [14] показано, что в определенных пределах осмысленное восприятие речи не привязано ни к одному частотному диапазону. Если взять в качестве среднего значения частоту 1900 Гц и *срезать частоты* выше и ниже, т.е. сделать их неслышимыми, то мы обнаружим: при срезании нижних частот мы теряем в смысле немного, опознавая до 70% слов, при срезании частот выше данного значения потерь становится больше, т.к. страдают согласные звуки. Если пропущенные гласные можно достроить, что и делается в беглой речи, то с согласными это сделать трудней. Именно благодаря тому, что основные гласные (И, Э/Е, А, О, У) четко различаются по частоте, они обеспечивают осмысленность.

Фонемы выполняют различительную функцию, поскольку, обладая разными акустическими свойствами, они выборочно усиливают одни признаки

слышимого слова и ослабляют другие. Гласные усиливают тоновую компоненту и ослабляют шумовую, а согласные наоборот. Соотносясь между собой, последовательность фонем образует систему звучащей речи [2, с. 479].

Таким образом, с физической стороны мы имеем речь как артикулированный, т.е. дифференцированный по частоте, акустический ряд. В качестве его физической модели используется так называемая *спектрограмма* – физическая запись звучащей речи. Единица чувственной ткани звучащей речи – *форманта*, или область усиленных частот, образующих такую характеристику речи, как тембр. Форманты могут не совпадать со смысловыми единицами, поэтому разделять речь по формантам, с точки зрения семантики сообщения, ненадежно. По формантам мы членим речь на незнакомом языке, опуская смысловые характеристики. Осмысленную речь мы воспринимаем фонематически. *Фонематический слух* – это способность различать (осуществлять анализ и синтез) фонемы частей звучащей речи. Недоразвитие фонематического слуха у ребенка тормозит обучение чтению, письму, общей речевой способности. Дети часто путают близкие по звучанию фонемы. Поэтому при недоразвитии или нарушении фонематического слуха ребенок может воспринимать именно то, что он услышал, а не то, что ему сказали.

Однако не только явная фонематическая расчлененность речи (146) обеспечивает ее осмысленное восприятие. С точки зрения *теории анализа через синтез*, восприятие речи предполагает спонтанный процесс выдвижения гипотез (синтез) относительно вербального стимула и их верификацию (анализ) различными способами (артикуляционными, лексическими или контекстными). Правда эта теория так и не смогла объяснить механизмы этого синтеза. При моделировании таких систем распознавания не удавалось избежать большого количества ошибок [1, с. 14]. Однако очевидно, что механизмы анализа и синтеза активно задействованы при восприятии речи.

Описан так называемый «*эффект восстановления фонемы*». Он связан с тем обыденным явлением, что мы воспринимаем речь на фоне какого-либо шума, акустических помех, накладывающихся на речевой поток. Эту ситуацию можно создать в лаборатории, что и было сделано в исследованиях Р. Уоррена и коллег [6], [20], [21]. Испытуемому предъявляется фраза, в которой какая-либо фонема заменена чужеродным звуком (например, кашлем). Здесь возможны варианты того, как это повлияет на процесс восприятия акустического потока речи. Допустим, человек слышит фразу: *The state governors met with their respective legislatures convening in the capital city*. В ней кашлем заменена фонема / s / в слове *legi/s/latures*. В этом случае для испытуемого нет альтернативы: в английском языке нет слова, которое бы различалось по значению в зависимости от фонемы в данном месте. Поэтому испытуемый воспринимает слово однозначно. В другом случае используется фраза, в которой кашлем заменена фонема в слове /.../eel. Как происходит восприятие фразы, если возможен целый спектр вариантов: «wheel», «heel», «reel», «meal»? Восприятие будет определяться контекстом предложения *It was found that the /.../eel was on the \_\_\_\_\_*. В зависимости от того, будет ли в конце

предложения стоять слово «axle», «shoe», «orange» или «table» испытуемый будет отчитываться о восприятии той или иной фонемы. Подчеркнем, что в действительности этой фонемы не было в физическом акустическом потоке. Следует обратить внимание также на то, что восприятию пропущенной фонемы способствуют слова, следующие после кашля [5, с. 598]. Характерно, что внутри предложения испытуемый не только реконструирует пропущенную фонему, но и не может точно указать местоположение кашля, т.е. определить, какой именно звук отсутствует. Если мы предложим просто слово с пропущенной фонемой, незамаскированной кашлем, то испытуемый воспримет место пропуска. Это значит, что процессы структурирования речевого потока по формантам и фонемам не совпадают по времени. На уровне фонематического слуха человек «придерживает» неполную информацию, реконструируя контекст. *Шум – это фон, который мы фонематически структурируем при восприятии речи.* Если контекст утрачивается или неясен, то испытуемый будет воспринимать только чувственную ткань, которая может структурироваться им произвольно. Так бывает при прослушивании иностранных песен человеком, который не владеет этим языком.

Кроме этого, установлено, что речь и ее восприятие устойчивы относительно помех [18]. Если с высокой частотой крутить регулятор громкости туда-сюда, то слышимый звук будет то исчезать, то появляться. Если частота такого колебания достаточно высока (до 9 к/с), то потеря в семантике будет (147) всего около 15%.

Есть еще ряд признаков, на основе которых мы реконструируем фонематический строй речи. Известно, что немые легко понимают речь на основе видимого движения губ. Здесь мы наблюдаем важную роль невербальных признаков речи. Известен также «*эффект МакГёрка*», впервые продемонстрированный в 1974 году. Испытуемый видит говорящего и слышит его одновременно. Человек, которого можно видеть, произносит слоги «га-га-га-га». Этот зрительный ряд синхронизирован со звукорядом «ба-ба-ба-ба». Большинство людей слышат «да-да-да-да» [16]. Происходит явление *аудиовизуальной интеграции*. Вариантом этого эффекта является интеграция «па-па-па-па» и «ка-ка-ка-ка» в «та-та-та-та». Считается, что «*эффект МакГёрка*» универсален. По крайней мере, кроме англоязычных, он проверен для говорящих на японском, испанском [15] и французском [9] языках. Явление аудиовизуальной интеграции существует на уровне слогов, но не на уровне речи в целом. Когда мы смотрим дублированный фильм, то несовпадение движения губ и речи не воспринимается.

Человек, который говорит, и человек, который воспринимает речь, имеют общее основание в том, что оба одинаково владеют навыками порождения и понимания речи. *Моторная теория восприятия речи* утверждает, что мы распознаем и понимаем речь, потому что сами умеем ее порождать. Слыша речь, мы ее бессознательно повторяем или бесшумно произносим. Мы владем акустическими единицами, которым соответствуют единицы артикуляционные (действия органов речи, необходимые для четкого произнесения отдельных

фонем). Благодаря этому акустический ряд узнается по артикуляционным единицам. Поясним эту теорию следующим примером. В слогах «ки» и «ку» фонема / к / различна на уровне акустического звука благодаря артикуляции двумя различными гласными. Однако на уровне восприятия эти фонемы воспринимаются как эквивалентные. Моторная теория объясняет, что это возможно как раз благодаря беззвучной и бессознательной артикуляции принимающим субъектом. Хотя физические характеристики фонемы определяются последующей гласной, артикуляционные компоненты для фонемы / к / остаются, как предполагалось, одинаковыми. Проще говоря, понимание воспринимаемой речи происходит через собственный артикуляционный экран, на котором «проявляется» речевой поток. Однако было установлено, что существует значительная вариация индивидуальных особенностей артикуляции, которая, тем не менее, не ведет к различию в понимании речи. С учетом этого моторная теория объясняет понимание речи не через внешне выраженные двигательные эффекты артикуляции, а через мозговые центры, регулирующие соответствующие движения («интендированную фонетическую мимику») [8, p. 518], [10, p. 922]. Эта версия моторной теории сегодня подкрепляется результатами изучения системы «зеркальных нейронов» в мозгу человека. Основной особенностью этих нейрональных центров является совмещенная функция, благодаря которой они одновременно регулируют наблюдение действий и совершение таких же действий. Предположительно они выработались у коллективных приматов в процессе регуляции имитационных схем поведения и были закреплены как основа для коммуникации жестами и (148) лицевой мимикой [19]. Моторные и семантические характеристики речи вырабатывались параллельно, синтезируясь в процессе развития социального общения. Еще А. Мессер в рамках исследования мышления Вюрцбургской школы установил, что восприятие и распознавание акустической формы слова еще не ведет к одновременному пониманию его значения. В зависимости от знакомости слова, его двусмысленности, зашумленности вербального стимула, уровня активации субъекта между двумя уровнями распознавания возникает временной зазор – понимание значения запаздывает [17, s. 73].

В результате мозговых травм могут развиваться тяжелые нарушения моторных функций речи – моторные афазии. При *афферентной афазии*, возникающей в случае поражения нижних отделов постцентральной коры левого (доминантного) полушария, нарушаются механизмы кинестетической афферентации. Сигналы от речедвигательного аппарата не достигают указанных отделов мозга, и больной не может целенаправленно управлять своим речевым аппаратом. *Эфферентная афазия* – это результат поражения нижних отделов премоторной коры левого (доминантного) полушария. В основе лежит нарушение кинетической организации речевого акта, т.е. управление речью при переключении от одной фонемы к другой, от слова к слову. Спонтанное плавное произнесение фразы становится невозможным, хотя по отдельности люди с таким нарушением свободно произносят слова и слоги.

В случае *динамической афазии* затрагивается внутренняя речь, что проявляется в нарушении планирования речи при ее сукцессивной организации в форме высказывания. Это вид возникает при поражении средне- и заднелобных отделов коры левого (доминантного) полушария.

Моторные механизмы речи, которыми обладает индивид, образуют, так сказать, внутренний ресурс, делающий потенциально возможным восприятие любой речевой формы сообщения. Вместе с тем, существует теория, которая придерживается экологического принципа восприятия, разработанного Дж. Гибсоном. Ее называют теорией *прямого реализма*. Как и в случае визуальной перцепции, восприятие речи объясняется принципом «в стимуле есть все», т.е. отказом вводить какие-либо внутренние схемы и средства его репрезентации. Если в теории зрительного восприятия Гибсона объемлющий свет несет информацию, заключенную в его оптическом строе, то в звуковой среде такую информацию несет акустический строй, исходно заданный речевой артикуляцией говорящего [12], [13]. С одной стороны, эта теория сближается с моторной, т.к. утверждает, что интерпретация сигнала обусловлена информацией, заключенной в нем самом (в данном случае в акустическом стимуле и «фонетической мимике»). С другой стороны, она принципиально отлична, т.к. механизмы этого извлечения она понимает как универсальные для восприятия вообще, тогда как моторная теория говорит о специфических механизмах речевосприятия [11, p. 922].

Единство механизма понимания и порождения речи утверждается также довольно старой концепцией, которая сегодня, несмотря на серьезную критику, не считается полностью отброшенной, – *генеративной лингвистикой Н. Хомского*. Критика Хомским теории речевого поведения Б.Ф. Скиннера, внесла вклад в упадок необихевиоризма [7]. Теория оперантного научения (149) Скиннера применительно к речевой деятельности полагала, что семантические процессы, определяющие порождение и понимание человеком речи, можно свести к условным рефлексам. Речевое поведение, тем самым, оказывалось сложной, но так или иначе предсказуемой системой реакций. Суть критики Хомского заключалась в том, что порождение речи – это процесс продуктивный, реализующий глубинные языковые модели независимо от действия внешнего стимула. С одной стороны, количество вариантов высказываний потенциально бесконечно, с другой стороны, конкретный речевой акт не является однозначно предсказуемой реакцией на заданный стимул.

До Хомского процесс порождения речи объяснялся в опоре на принцип языковой закономерности: совершая высказывание, человек подбирает слово к предыдущему с определенной степенью вероятности. Мы понимаем речь с опережением, руководствуясь контекстно-смысловыми признаками. Хомский выдвинул ряд идей, опираясь на рационалистическую психологию и грамматику, развитую в 17 и 18 веке.

1. *Процесс речепорождения не зависит от структуры конкретного языка*. Такая структура – только результат.

2. Речепорождение опирается на *языковую компетенцию*, представляющую некоторую *глубинную структуру*, характеризующую язык как таковой, синтаксис. Процесс речепорождения – это преобразование такой структуры. На глубинном уровне хранятся образцы *ядерных предложений* – моделей построения высказываний. Предложение (фраза) как акт речи реализуется как физический сигнал, вместе к которым в мышлении «образуется система суждений, выражающих значение предложения» [4, с. 39]. И физический сигнал, и система суждений управляются совокупностью операций, называемых *грамматическими трансформациями*.

3. Глубинной структуре противопоставляется *поверхностная структура*, представляющая более конкретное описание синтаксического устройства предложения. Например, предложения «бобры стоят плотины» и «плотины строятся бобрами» являются двумя примерами поверхностной структуры определенного ядерного предложения. Правила грамматической трансформации используются для перехода от глубинной структуры к поверхностной, т.е. для преобразования ядерных структур. Так, предложение «бобры стоят плотины» соответствует ядерной структуре, тогда как предложение «плотины строятся бобрами» имеет ту же структуру, однако в преобразованном виде. Здесь используется правило пассивизации.

Необходимо отметить, что Хомский развивает свою теорию с *позиции нативизма*. Конкретный язык, на котором говорит тот или иной человек, и конкретная система суждений в которых реализуется его мышление, управляются универсальной врожденной системой семантико-синтаксических операций. В конечном счете, мышление сводится Хомским к глубинной языковой структуре, что является главным объектом критики его концепции. Основное направление этой критики сосредоточено вокруг вопросов онтогенеза и филогенеза языка и мышления [3].

Человек существует как участник вербального общения. Объяснение способности человеческого индивида к языковому поведению предполагает **(150)** также и объяснение процесса восприятия речи. Врожденные структуры чувствительности и мозга, «пригодные» для речи, открывают лишь механическую и физическую сторону этого процесса. Природа языка, структурирующего мозговые функции и перцептивные стратегии в направлении конкретной формы языковой компетенции, остается не выявленной в такой трактовке. Онтогенез речи протекает внутри языковой среды, образованной устными и письменными текстами. Все эти тексты завязаны на внеязыковой контекст общения, его предметность и повод, диктующие языку его форму и содержание. Тем самым, речевая компетенция человека генерирована не столько природой самого языка, сколько потребностью отыскать, открыть, усвоить и развить средство высказаться и воспринять сказанное.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Венцов А.В. Касевич В.Б. Проблемы восприятия речи. – М.: УРСС, 2003.
2. Жинкин Н.И. На путях к изучению механизма речи // Психологическая наука в СССР. Т.1. – М.: Изд-во АПН РСФСР, 1959.
3. Лурия А.Р. Научные горизонты и философские тупики в современной лингвистике // А.Р. Лурия. Психологическое наследие. М.: Смысл, 2003.
4. Хомский Н. Язык и мышление. – М.: Изд-во МГУ, 1972.
5. Шиффман Х.Р. Ощущение и восприятие. – 5-е изд. – СПб.: Питер, 2003.
6. Bashford J.A., Riener K.R., Warren R.M. Increasing the intelligibility of speech through multiple phonemic restorations // Attention, Perception & Psychophysics, Vol. 51, № 3. 1992. – P. 211-217.
7. Chomsky N. A Review of B.F. Skinner's Verbal Behavior // Language. Vol. 35, № 1. 1959. – P. 26-58.
8. Cleary M., Pisoni D.B. Speech Perception and Spoken Word Recognition: Research and Theory // Blackwell Handbook of Sensation and Perception. E.B. Goldstein (Ed.), Blackwell Publishing, 2005.
9. Colin C., Radeau M., Deltenre P. Intermodal interactions in speech: A French study // Proceedings of Auditory-Visual Speech Processing № 4-6, 1998. – P. 55-60.
10. Conway C.M., Loebach J.L., Pisoni D.B. Speech Perception // Encyclopedia of Perception. E.B. Goldstein (Ed.). SAGE Publications. 2010.
11. Conway C.M., Loebach J.L., Pisoni D.B. Speech Perception // Encyclopedia of Perception. E.B. Goldstein (Ed.). SAGE Publications. 2010.
12. Fowler C.A. An event approach to the study of speech perception from a direct-realist perspective // Journal of Phonetics. Vol. 14, № 1, 1986. – P. 3-28.
13. Fowler C.A., Rosenblum L.D. The Perception of Phonetic Gestures // Modularity and the motor theory of speech perception: proceedings of a conference to honor Alvin M. Liberman. I.G. Mattingly, M. Studdert-Kennedy (Ed.) Routledge, 1991.
14. French N.R., Steinberg J.C. Factors Governing the Intelligibility of Speech Sounds // Journal of the Acoustical Society of America. Vol. 19, № 1, 1947. – P. 90-119.
15. Massaro D.V.V., Cohen M.M., Gesi A., Heredia R., Tsuzaki M. Bimodal speech perception: an examination across languages // Journal of Phonetics. Vol. 21. № 4, 1993. – P. 445-478.
16. McGurk H., MacDonald J. Hearing Lips and Seeing Voices // Nature. Vol. 264, № 23, 1976. – P. 746-748.
17. Messer A. Experimentell-psychologische Untersuchungen über des Denken // Archive für die gesamte Psychologie. № 8, 1906. – S. 1-223.
18. Miller G.A., Licklider J.C.R. The intelligibility of interrupted speech // Journal of the Acoustical Society of America. № 22, 1950. – P. 167-173.

19. *Rizzolatti G. Arbib M.A.* Language Within Our Grasp // Trends Neuroscience. № 21, 1998. – P. 188-194.

20. *Warren R.M.* Perceptual restoration of missing speech sounds // Science. Vol. 167, №. 3917 (23), 1970. – P. 392-393.

21. *Warren R.M., Warren R.P.* Auditory illusions and confusions // Scientific American, № 225, 1970. – P. 30-36.