

Исследования слухового восприятия речевых сигналов человеком.

Обзор работ, представленных на APCO

Люблинская В.В., кандидат биологических наук



В статье приводится ретроспективный обзор работ, представленных на заседаниях APCO и опубликованных в материалах, посвящённых исследованию принципов и механизмов слуховой обработки речевых сигналов человеком. Работы проводились коллективом сотрудников из лабораторий, руководимых профессорами Института физиологии им. И.П.Павлова Людмилой Андреевной Чистович (1961–1986 гг.) и Валерием Александровичем Кожевниковым (1961–1981 гг.). Они отражают основные теоретические представления о природе восприятия речи человеком и подходы к его изучению, существующие в тот период (от момента основания лабораторий в 1961 г. и до 1989 г., когда был проведён последний семинар APCO, в котором участвовали сотрудники коллектива).

In article is the retrospective review of the works presented at sessions APCO and published in materials, devoted to research the principles and mechanisms of acoustical processing of speech signals by human being. Works were carry out by collective of the laboratories, headed by professors of Pavlov Institute of physiology Lyudmila Andreevna Chistovich (1961 till 1986) and Valery Aleksandrovich Kozhevnikovym (1961 till 1981). The reviewed works reflect the basic theoretical ideas about the nature of perception of speech by human being and the approach to its studying, existing during this period (from the moment of the basis of laboratories in 1961 and till 1989 when the last seminar APCO in which employees of collective laboratories participated).



Введение. Задачи и методы

На первом собрании АРСО (АРСО-1) в Новосибирске в 1966 г. Л.А. Чистович представила лекцию на тему «Психоакустика и вопросы теории восприятия речи» [1, 2]. В ней перечислялись вопросы, из которых складывалась проблема восприятия речи человеком, и давалась оценка методам её исследования. Предлагался систематизированный ряд данных и соображений из различных научных областей: психологии, психоакустики и физиологии, которые должны учитываться теорией восприятия речи и которые могли бы позволить сформулировать вопросы и конкретные задачи исследований.

Во многом содержание лекции не потеряло своей актуальности и в настоящее время. Оно послужило каркасом и основой для составления данного обзора и обсуждения сообщений по теме исследования речи, которые были представлены на последующих собраниях АРСО в течение 24 лет. Мы сочли полезным не ограничиваться публикациями в материалах АРСО, а представить и другие источники, в которых были опубликованы результаты работ по обсуждаемым темам.

В упомянутой лекции Л.А. Чистович были выдвинуты следующие основные допущения, которые в 60-е годы были достаточно новыми, а в настоящее время споров не вызывают и многими исследователями речи приняты за основу.

Процесс восприятия речи представляет собой многоступенчатое отображение внешнего речевого воздействия, при котором на каждой последующей ступени (уровне) участок речевого сообщения описывается более сжато (с помощью меньшего числа параметров), при этом его длительность возрастает с повышением уровня.

Выдвигается предположение о трёх первичных (дограмматических) формах описания слухового сигнала: слуховое ощущение, признаки речевых единиц, признаки речевых последовательностей.

В общем виде основные задачи исследования восприятия речи человеком формулировались как выяснение структуры пространств изображений на каждом уровне и изучение характера преобразований от более низкого уровня к более высокому в процессе слуховой обработки речевых сигналов.

При этом подчёркивался очень важный вопрос о том, какие методы должны использоваться для исследования.

Общеизвестные данные о принципах слухового анализа акустических сигналов были получены на основе традиционных психоакустических методов, которые рассматривали слуховую систему как канал связи, а не как систему автоматического распознавания. При этом достаточным оказывалось использование в качестве стимулов простых звуковых посылок, таких как чистые тоны или шумы.

Очевидно, что подобные методы имеют ограниченные возможности для решения задач изучения восприятия речевых сигналов, имеющих значительно более сложную физическую структуру. Данные психоакустических исследований, так же как и данные нейрофизиологии, дают информацию преимущественно об уровне слухового ощущения, то есть о преобразованиях, осуществляемых на периферии слуха. (В тексте лекции Л.А. Чистович приводится обзор известных психоакустических данных о характеристиках периферического частотного анализа.) Поэтому поиски адекватных мето-



дов исследования восприятия речи имели (и имеют в настоящее время) чрезвычайно важное значение.

Слуховое членение (сегментация) речевого потока. Обработка амплитудной огибающей

При постановке вопроса о характере преобразований от более низкого к более высокому уровню системы следует исходить прежде всего из того, что речевой сигнал — процесс, протекающий во времени. Поэтому любые предположения об устройстве системы его обработки должны учитывать необходимость его текущего перекодирования и сохранения результатов. На эту тему Людмила Андреевна в цитируемом выше докладе предложила схему преобразований на примере самого низкого уровня — уровне слухового ощущения. Слуховое ощущение — это полное описание любого звука (известного и нового), причём оно должно сохраняться после его окончания, чтобы можно было перейти к его обработке на следующем уровне. Этот факт требует введения понятия **слуховой оперативной памяти**.

Таким образом, задачи исследования слухового ощущения формулируются как:

- 1) определение временного окна в оперативной памяти на запись одного отрезка потока;
- 2) определение длительности существования записи (насколько быстро она забывается);
- 3) определение формы (признаков) записи;
- 4) определение характера преобразований сигнала при его записи в оперативную память.

Как можно заметить, такой подход к пониманию временного процесса слуховой обработки речевых сигналов имеет общий характер и может быть применим к любому уровню. Здесь важным моментом является необходимость принятия во внимание оперативной (кратковременной) памяти для записи результатов текущей обработки соответствующего типа. Ко времени, к которому относится лекция, таким образом сформулированные вопросы не только не были разработаны, но вообще не ставились. Забегая вперёд, следует сказать, что последующая история речевых исследований подтвердила плодотворность такого конструктивистского подхода.

Понимание слухового восприятия речи как процесса, протекающего во времени, суть которого состоит в последовательном преобразовании акустического сигнала в иерархически организованные единицы, неизбежно приводит к постановке вопроса о сегментации речевого потока. Сегментация (слуховое членение) речевого потока представляет собой фундаментальную проблему для изучения процесса распознавания речи. В работах коллектива Чистович–Кожевникова ей уделялась большая доля внимания, о чём свидетельствуют публикации, представленные в разных материалах АРСО и в других литературных источниках. Обобщённые данные по этому вопросу можно встретить в монографии [3]. Следует отметить, что постановка задач исследований и начало их разработки были сформулированы уже в первой лекции Л.А. Чистович. Проблема сегментации, иными словами, формулировалась как управление считыванием и записью слуховой информации в оперативную память. Что собой представляет запись информации — отдельный вопрос для исследований. Можно представить, что на разных уровнях иерархической системы распознавания механизмы сегментации управляют выделением и измерением характеристик единиц разной размерности (слоги, слова, синтагмы) и что в каждом случае проявляются свои особенности.

Доклад Жукова на АРСО-6 [4] открывал обсуждение двух гипотез, относящихся к процессу сегментации. Одна гипотеза предполагает периодическое принудительное считывание звуковой информации с определённой частотой (сканирование). Согласно второй гипотезе, моменты считывания синхронизированы с сигналом, в котором присутствуют некоторые события, позволяющие формировать метки сегментации. В качестве послед-



них, например, могут рассматриваться быстрые изменения спектральных составляющих или скачки (возрастание и спад) амплитуды.

Первые данные в поддержку предположения о том, что считывание информации осуществляется синхронно с сигналами, были получены в опытах по имитации изолированных гласных, проведённых ещё в начале 60-х годов. Эти опыты показали (см. лекцию [2, с.72]), что задержки реакций испытуемых строго синхронизированы с началом стимулов. Убедительное подтверждение этому факту было получено позднее (в начале 70-х) в более сложном эксперименте, который заслуживает более подробного рассказа [3, с.57–61].

В опытах испытуемым предъявлялись периодически повторяющиеся двусложные последовательности типа ГСГ или ГССГ (например, *ata, asa, atma, amda* и др.), в ответ на которые они должны были синхронно произносить одну из трёх последовательностей ГСГ (*ata, ada, asa*). Каждый случай ответного произнесения регистрировался на экране осциллографа, что позволяло анализировать моменты начала и конца первого и второго гласного относительно соответствующих моментов гласных стимулов. Анализ флюктуаций сегментных границ ответных звуко сочетаний убедительно показал, что испытуемые выделяли во внешнем речевом стимуле события, служащие индикаторами сегментных границ, и могли определять и воспроизводить длительность гласных и согласных между ними.

Описываемые результаты позволяли предположить, что марками сегментов могут служить моменты быстрых изменений интенсивности звуковых последовательностей (скачкообразные модуляции амплитуды), дающие возможность человеку обнаруживать такие события как начало и конец посылок и, соответственно, измерять временной интервал между ними.

Цикл экспериментальных исследований, связанных с сегментацией, включал работы по восприятию ритмических последовательностей, где измерение длительностей элементов играет ключевую роль, а потому немаловажно было знать временные ограничения процесса выделения сегментных марок. Этому вопросу была посвящена работа Лысенко и Фёдоровой [5], представленная на АРСО-8. Её задачей было определение минимального межимпульсного интервала между тональными посылками, при котором интервал ещё будет воспринят как отдельный элемент последовательности. Было найдено, что образ ритмического рисунка стимулов сохранялся в ответах испытуемых, если период следования предъявляемых звуковых элементов (интервал «тональная посылка плюс пауза») составлял не меньше 110 мс (порог), что примерно соответствует в речи средней длительности слогов.

Отдельным разделом исследований, связанным с восприятием длительности сегментов, была работа Венцова [6], в которой было показано, что длительностью гласного в слове определяется признак ударения в слове, а длительность согласного управляет членением: более длинный согласный рассматривается как граница между двумя словами.

Изучение слухового восприятия скачкообразных модуляций амплитуды показало, что они не только используются для маркировки длительности сегментов, но участвуют в процессе фонемной интерпретации. Об этом свидетельствовали экспериментальные данные, приведённые в докладе Кузьмина и Лисенко [7], опубликованном в материалах АРСО-4. При прослушивании гласноподобных стимулов со скачкообразным изменением амплитуды в момент, находящийся на некотором интервале от начала посылки, стимулы воспринимаются как слоги с сонорным согласным. При положительном скачке слог опознаётся как согласный-гласный (СГ), при отрицательном скачке — как гласный-согласный (ГС).



Восприятие фонетического качества согласного в подобных квазисловах зависит от нескольких параметров. В частности, существенное значение имеет длительность интервала между началом более слабого по амплитуде первого сегмента (согласного) и началом более сильного сегмента (гласного). Первый сегмент воспринимается как звонкий смычный, если этот интервал меньше 30 мс, и как назальный, если он больше 30 мс. В последующих экспериментальных работах неоднократно наблюдался этот эффект [8, 9].

Интересные данные о сканирующей функции марок сегментации в процессе слуховой обработки акустического речевого потока были получены в работе [10], опубликованной в материалах АРСО-8. В экспериментах по слуховому восприятию в качестве стимулов использовались синтезированные гласные, у которых F2 изменялась линейно в диапазоне от 2000 до 800 Гц, что соответствовало сочетанию *iuI*. В середину вставлялась короткая пауза (15 мс), воспринимаемая слушателями как *r*-образный согласный. Наблюдаемый эффект состоял в том, что согласный воспринимался как твёрдый или как мягкий (палатализованный), в зависимости от положения паузы во времени относительно частотной области формантного перехода: согласный идентифицировался как мягкий */rʲ/*, если пауза попадала в точку начала *i*-образного перехода (примерно 1700 Гц), или как твёрдый */r/* — при более низкочастотном положении.

Возможность анализа слуховой системой скачкообразных амплитудных модуляций подтверждается многочисленными нейрофизиологическими данными, показывающими наличие на всех уровнях слухового пути нейронов, реагирующих на быстрые изменения амплитуды звуковых стимулов. Физиологическая литература по данной теме весьма многочисленна: библиографию можно найти, например, в обзорах книги [11], а в материалах АРСО см. статью Бибикова [12].

Как видно, описанные экспериментальные данные свидетельствуют о том, что закономерности слухового восприятия быстрых изменений амплитуды сигналов не только имеют отношение к объяснению процесса сегментации речевого потока, но выполняют более универсальную функцию в процессе распознавания речи. На основании нейрофизиологических и психоакустических данных была разработана функциональная модель обнаружения изменений интенсивности звуковых сигналов (амплитудных неравномерностей), описание которой приводится в статье Столяровой [13].

Природа полезных признаков спектра речевых звуков

В первой лекции Л.А. Чистович [1, 2] обсуждались две существующие к тому времени точки зрения на природу слухового описания речевых звуков, предшествующего фонемным решениям (на котором основывается фонемная идентификация). Одна точка зрения предполагает, что в память записываются полные сведения о сигнале — спектрально-временной рельеф возбуждения нейронов на периферии слуха. Другая точка зрения состоит в том, что в оперативную память записываются только некоторые параметры (признаки) этого рельефа. Приведённые в цитируемом докладе экспериментальные данные, полученные в опытах с применением метода субъективного шкалирования близости между стимулами, позволяли сделать заключение, что на уровне слухового ощущения звуковые образы представлены в памяти в виде полного описания. Переход к полезным субъективным признакам происходит на следующем этапе обработки данных, и естественно предположить, что сами признаки могут быть не врождёнными, а приобретёнными в процессе обучения. По-видимому, к этому типу относятся признаки, определяющие фонетическое качество речевых звуков. Ко времени, к которому принадлежит лекция Л.А. Чистович на первом АРСО, было опубликовано довольно много результатов экспериментальных исследований восприятия синтезированной речи и выделенных отрезков естественной речи. Целью был поиск таких акустических признаков речевых сигналов, которые несут информацию о фонемах. Из этих результатов следует два принципиальных вывода.



1. Практически любые замеченные исследователями физические свойства, которые отличают друг от друга элементы речи, образованные при произнесении разных фонем, используются человеком при идентификации. (Полезных акустических признаков много, и они весьма разнообразны.)
2. Полезная информация, имеющая одно и то же фонематическое значение, распределена на разных участках слога.

Повсеместно признаётся, что наибольшая информация о фонемных признаках содержится в спектре акустических речевых сигналов. Экспериментальные исследования, связанные с выяснением спектральных признаков речевых звуков и проделанные в 60–70-е годы под руководством Л.А. Чистович, определялись двумя основными предпосылками. Во-первых, представлением о том, что слуховая система «работает» как спектральный анализатор с определёнными избирательными характеристиками и что, начиная с периферических отделов, речевой сигнал предстаёт в виде спектрограммы, подобно известным изображениям технических вариантов «видимой речи». Во-вторых, полезными признаками для идентификации вокализованных речевых звуков служат основные максимумы на спектральной огибающей — форманты.

Одной из ранних работ на эту тему, представленных на АРСО-4, была работа Чистович и Мушников [14], содержащая результаты цикла экспериментов по слуховому восприятию гласноподобных стимулов, у которых в процессе опытов варьировали частота или амплитуда формант. В качестве критерия восприятия использовался факт изменения фонемного качества гласного в зависимости от параметров спектральной огибающей. В одной серии работ методом активного поиска определялась фонемная граница между классами [i] и [e] при прослушивании двухформантных синтетических гласных, у которых частота или амплитуда (F_2 , A_2) фиксировалась (варьируя в заданном диапазоне), а частотой первой форманты (F_1) управлял испытуемый. Результат показал, что граница по F_1 всегда устанавливается в одном и том же месте при разных значениях F_2 и A_2 в спектре стимула.

В другой серии опытов выяснялись условия, при которых слух может обнаружить в спектре гласного с одним низкочастотным максимумом (F_1) присутствие второго формантного максимума (F_2), изменяющего исходное фонемное качество стимула. Оказалось, что в этих условиях в качестве спектрального признака может выступать не только хорошо оформленный максимум F_2 , но любая заметная неоднородность спектральной огибающей в данной частотной области. Более подробные сведения об экспериментальных исследованиях восприятия спектральных признаков гласных можно найти в работах [15, 16, 17, 18].

Приведённые результаты давали основания для предположения о том, что такой признак как частота спектрального максимума носит локальный характер и его обнаружение не зависит от других характеристик спектральной огибающей. При этом система обработки слухового спектра должна схожим образом реагировать на максимумы и на неоднородности спектральной огибающей. Добиться подобного эффекта невозможно, если в качестве характеристик спектрального слухового анализа рассматривать только свойства базилярной мембраны. На основании этих фактов делалось предположение о возможном участии в слуховой обработке периферического спектра некоторого механизма обострения (например, механизма латерального торможения). Позднее попытки реализовать механизм обострения были предприняты в моделях двухтонового подавления, описанных Столяровой в разделе моделей [13].



Анализ естественных речевых сигналов по модели слухового анализа, включающей блок двухтонового подавления (см. описание модели), выявил спектральные огибающие гласных, в которых число максимумов, соответствующих неоднородностям, заметно превышает число формант. Особенно это касается низкочастотной области, в которой модель отображает максимумы гармоник основного тона голоса. Такое положение дел навело на мысль искать в слухе дополнительные способы обработки спектральной огибающей, которые давали бы более сглаженное её представление. Поиску способов более сокращённого описания спектральных признаков гласных, например, выделению спектральных «центров тяжести», начиная с 50-х годов было посвящено множество работ (см. обзор [3]).

Одна из работ на эту тему была представлена на APCO-10 [19]. В ней описываются результаты экспериментов, основным методом которых была процедура сравнения по фонетическому качеству двух гласноподобных стимулов, спектр одного из которых был двухформантным (содержал два главных максимума), а второй — одноформантным. Экспериментатор варьировал параметры первого стимула (изменял частоту F_1 , F_2 , амплитуду формант A_1 , A_2 , наклон спектральной огибающей), а испытуемый должен был стараться уравнивать оба стимула, изменяя частоту форманты (F^*) второго из них.

В этих опытах был подтверждён вывод более ранних исследований об объединении информации о близких по частоте формантных максимумах гласного и был определён критический частотный интервал близости, равный примерно 3.5 Барк, при котором происходит объединение. В этом интервале при фиксированных частотах F_1 , F_2 варьирование соотношения амплитуд формант приводило к такому значению F^* , как если бы слушатель устанавливал его в некоторый «центр тяжести» спектрального комплекса первого стимула. При более далёком разнесении F_1 , F_2 изменение амплитуд не приводит к изменению F^* до тех пор, пока один из максимумов не будет подавлен до подпорогового уровня. В этом случае значение F^* устанавливалось близко к частоте более сильного максимума. Обнаруженный пороговый частотный интервал объединения спектра гласных был неоднократно подтверждён в других работах нашего коллектива, а также в работах зарубежных авторов. Более подробное описание экспериментов можно найти в работах [20, 21, 17].

Исследование восприятия гласноподобных сигналов с изменяющимся спектром

Даже беглый взгляд на динамические спектрограммы речевых сигналов позволяет понять, что большую часть из них характеризует спектр, изменяющийся во времени. Исследованию закономерностей восприятия гласных с изменяющимся спектром была посвящена серия экспериментальных работ лаборатории, значительная часть которых докладывалась на семинарах APCO и нашла отражение в печатных материалах. Надо отметить одно обстоятельство, связанное с экспериментальным исследованием восприятия гласных и гласноподобных стимулов (впрочем, не только гласных). Как упоминалось выше, обработка речевых звуков включает, по меньшей мере, два уровня: уровень слухового ощущения и уровень фонемного представления (в виде набора фонетических признаков). Для интерпретации экспериментальных результатов очень важно иметь в виду, с каким описанием работал испытуемый-слушатель в конкретном эксперименте. Преследуя задачу поиска фонетических признаков, можно столкнуться с тем, что восприятие опиралось на признаки слухового ощущения, которые не всегда совпадают с фонемными, и наоборот. (Это в большой мере определяет жёсткие требования к выбору методических средств.)

С подобным положением дел столкнулись авторы работы, направленной на исследование способа слухового описания гласноподобных звуков с переменным спектром [22] в пространстве слуховых ощущений, используя метод оценки близости. В эксперименте использовались синтезированные двухформантные стимулы с характеристиками, пред-



положительно соответствующими одному фонемному классу (в данном случае — /i/), для которых требовалось оценить тембральные различия пар стимулов при варьировании их длительности и типа формантного контура. Предполагалось выяснить, какое из двух описаний более вероятно: (1) — в виде последовательных дискретных отсчётов — или (2) — в виде набора отдельных признаков временного контура частоты «формант».

Вопреки априорным предположениям об однородности фонемного состава выбранных стимулов, данные показали, что при оценке их субъективных состояний испытуемые используют фонемную идентификацию, распределяя ответы в ближайшие классы «психологических фонем» /i/, /y/ и /u/ (последний символ обозначает аллофон русского /u/ после мягких согласных). Два последних из упомянутых классов представляют собой русские дифтонгоиды, которые акустически в простом случае имитируются однонаправленным изменением F2 при неизменной F1=300 Гц.

Полученные результаты иллюстрировали тот факт, что слушатели опирались на фонемные признаки стимулов и что в число признаков входили краевые частоты (или направление изменения частоты) и параметры, определяющие среднее положение контуров на оси частот. Результаты описанной работы показали, что использование стимулов подобного типа может служить удобной моделью для изучения признаков фонемного восприятия сигналов с переменным спектром. Подтверждением этому было найдено в ряде работ, представленных в материалах АРСО [23, 24, 25].

Другая серия работ, связанная с изучением восприятия нестационарных речевых сигналов, была посвящена принципам текущей обработки гласных, спектр которых варьировал случайным образом на протяжении их восприятия. Предметом выяснения были два основных вопроса: происходит ли накопление спектральной информации на протяжении восприятия гласных и какова природа накапливаемой информации. Проверялись две возможные гипотезы. Согласно одной из них, на протяжении восприятия сигнала происходит усреднение периферического спектрального отображения. В альтернативном варианте допускалось, что усредняется не спектр, а предварительные фонемные решения, принимаемые по прошествии коротких интервалов времени (фреймов). Обе гипотезы основываются, с одной стороны, на проявлении эффекта временной суммации, а с другой — на данных о возможности идентификации коротких отрезков гласных длительностью не более 10 мс [26].

Для экспериментального исследования поставленной задачи была применена специальная методика синтеза гласноподобных стимулов, единственным элементом которого был «формантный импульс». Он формировался как упрощённая имитация формы волны отклика одного или двух формантных фильтров за период основного тона. Представление сигнала в виде последовательности формантных импульсов позволяло моделировать закон временного изменения частоты формант и соотношение их амплитуд на протяжении звуков. Также можно было задавать их высотные свойства, управляя изменением межпериодного интервала [27, 28, 29].

Экспериментальные психоакустические исследования восприятия гласных, синтезированных таким способом, составили большую серию работ, опубликованных в материалах АРСО [30, 31, 28] и в разных статьях, например, [32, 33]. Результаты, полученные в этих работах, позволили установить два основных факта:

1) в процессе идентификации гласных действительно осуществляется накопление информации на протяжении их длительности;



2) данные свидетельствуют в пользу гипотезы о текущих фонемных решениях, которые по окончании стимулов объединяются в один образ.

В заключение

На заседании семинара APCO-15 в 1989 г. (последнем, где принимал участие наш коллектив), в пленарном докладе-лекции В.В. Люблинской [34], рассматривались общие представления относительно принципов слуховой обработки речевых сигналов человеком, выработанные специалистами нашего и зарубежных коллективов к концу 80-х годов прошлого века, и приводились основные результаты экспериментальных поисков, вписывающихся в канву общих идей. Надо отметить, что доклады, представленные на APCO-15, относятся к периоду, когда Л.А. Чистович уже не возглавляла лабораторию (она оставила Институт физиологии, уйдя на пенсию). Тем не менее, экспериментальные исследования закономерностей слухового восприятия речевых сигналов и их моделирование проводились коллективом в рамках идей и задач прежних лет, результаты которых составили содержание данного обзора.

Кроме обобщающего доклада, на APCO-15 было представлено пять работ. Две из них были посвящены развитию направления математического моделирования в реализации программного комплекса, включающего модель периферического спектрального анализатора и модель выделения амплитудных неравномерностей, подробный обзор которых приводится в статье Столяровой [13]. Результаты экспериментального изучения закономерностей восприятия «слухового спектра» на следующем уровне обработки составляли содержание остальных докладов. В работах [35, 36] приводятся экспериментальные данные, относящиеся к организации процесса восприятия речевых звуков на фонетическом уровне. В первой цитируемой работе было получено ещё одно свидетельство объединения спектральной информации на протяжении гласного и было показано, что формирование фонемного образа может объединять спектральную информацию с разной временной структурой (при разной частоте основного тона формантных колебаний и при разных источниках возбуждения, таких как тон и шум). Во всех случаях фонемная идентификация стимулов соответствовала суммарному спектру. Во второй работе была сделана попытка выяснить слуховые признаки идентификации источника возбуждения (глухой — звонкий) шумовых звуков, имеющих периодически прерываемую структуру.

Стратегия процесса распознавания речи во времени (лексический анализ) была популярной темой дискуссий и исследований того периода (впрочем, как и в настоящее время), где главным предметом служили подробности различий между активным и пассивным принципами анализа. В работе Залкинда [37], относящейся к данному направлению, были приведены результаты опытов, проведённых по методу быстрой имитации слов, которые свидетельствовали в пользу активного принципа (анализ «сверху вниз»), не исключая в то же время анализа «снизу вверх» (шаг за шагом), используемого в отношении малознакомых слов.

В заключение следует сказать, что доклады, представляемые на APCO, не исчерпывают содержания всего направления исследовательских работ коллектива Чистович — Кожевникова и его последователей. Тем не менее, они отражают в общих чертах идеи, задачи и методы их решения, результаты которых составляют солидный багаж сведений, образующих целое научное направление — экспериментальное изучение слуховой обработки речевых сигналов человеком. Направление, новое в начале 60-х годов, повсеместно признано в настоящее время, доказательством чего служат многочисленные публикации на эту тему (к сожалению, в основном, зарубежные).

Остаётся ответить на вопрос, полезны ли и необходимы ли знания о слуховой обработке речевых сигналов человеком для разработки систем автоматического распознавания и понимания речи. Признание необходимости знаний о человеческих способностях в своё время стимулировало организацию исследований в лабораториях Чистович — Кожев-



никова, а доклады на эту тему были неизменной составляющей заседаний АРСО. Причины, по которым эти знания не нашли конкретного применения в современных системах автоматического распознавания и не произошло желательного реального содружества физиологов и техников, коренятся в истории развития речевых технологий, особенно драматичной в нашей стране. Однако точка зрения современных зарубежных специалистов на положение дел в этом отношении, например, [38, 39, 40], даёт оптимистические прогнозы. Более подробному обзору и обсуждению этой проблемы посвящена публикация автора настоящей статьи [41].

Литература

1. Чистович Л.А. Психоакустика и вопросы теории восприятия речи // Автоматическое распознавание слуховых образов (под ред. Н.Г. Загоруйко и Г.Я. Волошина). «Наука». Новосибирск, 1966. С. 68–169.
2. Чистович Л.А. Психоакустика и вопросы теории восприятия речи // Автоматическое распознавание слуховых образов (под ред. Н.Г. Загоруйко и Г.Я. Волошина). «Наука». Новосибирск, 1970, С. 55–141.
3. Чистович Л.А., Венцов А.В. и др. Физиология речи. Восприятие речи человеком. Руководство по физиологии. Коллективная монография. Л.: Наука, 1976. 388 с.
4. Жуков С.Я. О слуховом членении звукосочетания // Тез. докл. 6-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-6). Таллин, 1971. С. 30–31.
5. Лисенко Д.М., Фёдорова Н.А. К вопросу о временных ограничениях слуховой сегментации // Тез. докл. 8-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-8), часть 2. Львов, 1974. С. 49–51.
6. Венцов А.В. Влияние длительности согласного и гласного на восприятие удара в русском языке // Тез. докл. 8-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-8), часть 3. Львов, 1974. С. 78–80.
7. Кузьмин Ю.И., Лисенко Д.М. Роль изменений интенсивности при фонетической интерпретации речеподобных стимулов // Тез. докл. 4-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-4). Киев-Канев, 1968, С. 229.
8. Жуков С.Я. Люблинская В.В. Столярова Э.И. Обнаружение быстрых амплитудных и частотных изменений в спектре гласноподобных звуков (при слуховом восприятии и на модели) // Тез. докл. Всесоюзного семинара АРСО-14. Каунас, 1986. С. 50.
9. Люблинская В.В., Столярова Э.И., Малинникова Т.Г. Слуховое восприятие частотно-локализованных акустических оп-событий // Физиологический журнал. 1995. № 7. С.31–39.
10. Жуков С.Я., Лисенко Д.М. Марки сегментации и их роль в интерпретации формантного контура // Тез. докл. 8-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-8), часть 2. Львов, 1974. С. 36–40.
11. Слуховая система. Ред. Альтман Я.А. Л. «Наука», 1990. 620 с.
12. Бибииков Н.Г. Некоторые особенности механизмов признакового описания сигнала в слуховой системе // Тез. докл. 13-го Всесоюзного семинара (АРСО-13). Новосибирск, 1984. С. 62–65.



13. Столярова Э.И. Моделирование механизмов слуховой обработки речевых сигналов. (Настоящий номер РТ).
14. Мушников В.Н., Чистович Л.А. Маскировка формант в спектрах стационарных синтетических гласных // Тез. докл. 4-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-4). Киев-Канев, 1968. С. 223–228.
15. Мушников В.Н., Чистович Л.А. Метод экспериментального исследования роли громкостей составляющих гласного в его распознавании // Акуст. журн. 1971. Т. 17. С. 405–411.
14. Chistovich L.A. Auditory processing of speech stimuli evidences from psychoacoustics and neurophysiology. // Proceed. Of the 7th ICA, Budapest, 1971, V.1, P. 27–42.
15. Chistovich L.A. Central auditory processing of peripheral vowel spectra. // Journ. Acoust. Soc. Am. 1985. V. 77. P. 789–805.
16. Чернова Е.И., Чистович И.А. Определение положения форманты в сигналах с треугольной огибающей // Тез. докл. 13-го Всесоюзного семинара (АРСО-13). Новосибирск, 1984. С. 67.
17. Шейкин Р.Л., Люблинская В.В., Чистович Л.А. Слуховая обработка спектральной информации сигнала // Тез. докл. 10-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-10). Тбилиси, 1978. С. 67–69.
18. Бедров Я.Р., Чистович Л.А., Шейкин Р.Л. Частотное положение «центра тяжести» как полезный признак при восприятии гласных // Акустический журнал. 1978. Т. 24. № 4. С. 480–486.
19. Люблинская В.В., Чистович Л.А. Факторы, определяющие субъективную близость гласноподобных стимулов // Физиологический журнал СССР. 1978. Т. 64. № 12. С. 1782–1789.
20. Люблинская В.В. Воспринимаемые параметры формантного контура // Тез. докл. 8-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-8), часть 2. Львов, 1974. С. 74–76.
21. Люблинская В.В., Слепокурова Н.А. Фонемная интерпретация синтетических гласных с переменным во времени спектром сигнала // Тез. докл. 10-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-10). Тбилиси, 1978. С. 99–100.
22. Люблинская В.В. Направление движения частоты формант как фонетический признак гласных-дифтонгоидов // Тез. докл. 12-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-12). Киев-Одесса, 1982. С. 283–286.
23. Люблинская В.В. Обнаружение формантных переходов в синтезированных гласных по фонетическому критерию // Тез. докл. 13-го Всесоюзного семинара (АРСО-13). Новосибирск, 1984. С. 71–72.
24. Огородникова Е.А. Возможность идентификации гласных по одному периоду основного тона // «Речевые исследования». Ред. В.И. Галунов. Л. 1981. С. 137–140.
25. Чистович Л.А., Чихман В.Н., Огородникова Е.А. Подход к исследованию алгоритма временной обработки спектральной информации при фонемной интерпретации гласных человеком // Тез. докл. 11-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-11). Ереван, 1980. С. 190–193.
26. Огородникова Е.А. Высота нестационарного сигнала Тез. докл. 13-го Всесоюзного семинара (АРСО-13). Новосибирск, 1984. С. 66.
27. Чистович Л.А., Чихман В.Н., Огородникова Е.А. Новый подход к определению фонетической близости стимулов и его проверка в автоматизированном эксперименте // Физиологический журнал СССР. 1981. Т. 67. № 5. С. 704–711.
28. Чистович Л.А., Малинникова Т.Г., Огородникова Е.А. Накопление спектральной информации на интервале гласного // Тез. докл. 12-го Всесоюзного семинара «Автоматическое распознавание слуховых образов» (АРСО-12). Киев-Одесса, 1982. С. 291–294.

29. Чистович Л.А., Малинникова Т.Г. Слуховой анализ формы текущего спектра гласного // Тез. докл. 13-го Всесоюзного семинара (АРСО-13). Новосибирск, 1984. С.59–60.
30. Чистович Л.А., Малинникова Т.Г. Центральный анализ периодического слухового спектра гласного // Физиология человека. 1985. Т. 2. С. 716.
31. Малинникова Т.Г., Чистович И.А. Обработка динамического слухового спектра при восприятии гласных // Сенсорные системы. 1988. Т. 2. С. 408–417.
32. Люблинская В.В. Восприятие речи. Общие представления и подходы к исследованию// Тез. докл. Всесоюзного семинара АРСО-15. Таллинн, 1989. С. 32–37.
33. Чистович И.А., Малинникова Т.Г., Жуков С.Я. Объединение спектральной информации при различиях во временной структуре спектральных областей // Тез. докл. Всесоюзного семинара АРСО-15, Таллинн, 1989. С. 241–243.
34. Чернова Е.И. Оценка шумопериодических стимулов как способ исследования принципов восприятия звонкости человеком//Тез. докл. Всесоюзного семинара АРСО-15. Таллинн, 1989. С. 239–241.
35. Залкинд В.И. Возможности метода быстрой имитации вербальных стимулов в опознавании речевых сигналов// Тез. докл. Всесоюзного семинара АРСО-15. Таллинн, 1989. С. 200–202.
36. Morgan M., Bourlard H., Hermansky H. Automatic Speech Recognition: An Auditory Perspective. // Speech Processing in the Auditory System. // Eds. Greenberg S., Ainsworth W.A., Popper A., Fay R.R. New York: Springer-Verlag, 2004. P. 309–337.
37. Lee C-H. Back to Speech Science — Towards a Collaborative ASR Community of the 21st Century. // Dynamics of Speech Production and Perception. // Eds: Divenyi P. et al. IOS Press, 2006. P. 221–244.
38. Szepannek G., Harczos T., Klefenz F., Weihs C. Combining different auditory model based feature extraction principles for feature enrichment in automatic speech recognition. // SPECOM'2009. St. Peterburg, 21–25 June, 2009. P. 205–209.
39. Люблинская В.В. Нужен ли слух для автоматического распознавания речи? // Сенсорные системы. 2009. Т. 23. № 4. С. 275–282.

Сведения об авторах

Люблинская Валентина Владимировна —

кандидат биологических наук. Ведущий научный сотрудник лаборатории психофизиологии речи Института физиологии им. И.П. Павлова РАН, Санкт-Петербург.

Базовое образование — инженер-электроакустик, окончила Ленинградский электротехнический институт в 1959 г.

С 1961 г. работает в Институте физиологии им. И.П. Павлова в лаборатории физиологии речи (в настоящее время — лаборатория психофизиологии речи).

До 1986 г. работала под руководством проф. Л.А. Чистович, с 1986 по 2006 г. руководила сектором исследования речевых сигналов.

Основной научный интерес — закономерности слухового восприятия речевых сигналов человеком. Автор более 150 печатных научных работ на русском и английском языках.