

===== **ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ТЕХНИЧЕСКИХ** =====  
===== **И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ** =====

## **Моторная теория восприятия речи и теория внутренней модели**

**Сорокин В.Н.**

*Институт проблем передачи информации, Российская академия наук, Москва, Россия*

Поступила в редколлегию 11.01.2007

**Аннотация.** Теория внутренней модели объединяет процессы речеобразования и восприятия речи. Внутренняя модель является элементом системы управления артикуляцией, обеспечивая текущий контроль и коррекцию артикуляции при различных нарушениях путем решения обратных задач "проприоцепция - управления" и "акустика - управления". Для успешной деятельности внутренняя модель должна располагать сведениями о механике, аэродинамике и акустике речеобразования, а также о фонетике языка. Основным механизмом действия внутренней модели – поиск экстремума некоторого сложного критерия оптимальности. Способность внутренней модели к решению обратной задачи "от акустики к управлениям" может использоваться и при распознавании речи других людей, дополняя пространство акустических признаков пространством артикуляторных параметров.

Противоречие между наблюдаемым разнообразием акустических параметров и кажущейся устойчивостью восприятия фонетических элементов речи привело к поискам таких способов обработки речевого сигнала, которые обеспечивали бы меньшую изменчивость измеряемых параметров. Эти поиски привели к формулировке различных вариантов моторной теории восприятия речи, предполагающих анализ моторной компоненты речи по речевому сигналу [1, 2, 3]. Представление о том, что при восприятии речи каким-то образом используется информация о свойствах речеобразования, опирается на способность человека к обучению речи. Определенную роль сыграло и явление так называемой внутренней речи, т.е. наблюдающееся иногда проговаривание "про себя" читаемого текста.

Сравнительно долгое время эти представления не получали непосредственного экспериментального подтверждения и моторная теория имела скорее статус гипотезы. Против этой гипотезы было высказано немало возражений, в основном, сводящихся к тому, что ухо получает только акустическую информацию и ничего больше, а больше ничего и не нужно для восприятия речи. Обсуждению моторной теории был посвящен целый номер журнала Американского акустического общества, содержащий, в основном, критические статьи [4, 5, 6, 7, 8].

Тем временем, накапливались наблюдения за компенсацией естественных и искусственных нарушений процесса речеобразования или восприятия. Неврологам и логопедам было давно известно, что при парезе или параличе отдельных лицевых или внутриротных мышц разборчивость речи может и не пострадать. Например, при парезе мышц, управляющих движениями нижней челюсти, артикуляция губных звуков осуществляется за счет большей амплитуды движений губ. Начиная носить зубные протезы с искусственных твердым небом, в ряде случаев люди сохраняли разборчивость своей речи. Иногда больные с удаленной гортанью не только полностью восстанавливали в своей речи различие между звонкими и глухими согласными, но и правильную фразовую интонацию [9], и даже могли петь. Имеются сведения

о том, что замена удаленного языка пластиковым протезом позволила больному сохранить сравнительно разборчивую речь

Для проверки популярной в свое время теории отрицательной обратной связи при управлении артикуляцией были разработаны специальные экспериментальные методики с искусственным нарушением процессов артикуляции. Исследовались эффекты статического возмущения типа байт-блока, т.е. фиксации положения нижней челюсти [10, 11, 12], блока губ с помощью трубочки [13] и искусственного неба [14, 15]. Исследовалась также реакция артикуляторов на неожиданное механическое возмущение движений губ и нижней челюсти [16, 17, 18, 19, 20, 21], изменение формы твердого неба [22], или внезапную электрическую стимуляцию мышц, управляющих артикуляторами [23]. Были также проведены эксперименты с возмущением акустических параметров речевого сигнала – основного тона или формантных частот [24, 25]. Во многих случаях наблюдалась перестройка положений артикуляторов с целью сохранения акустического образа, характерного для произносимого звука или звукосочетания.

Результаты экспериментов такого типа показывают, что компенсация возмущений может происходить достаточно быстро, с задержкой 10 – 40 мс, что не оставляет времени для пробных артикуляторных движений, и сопоставимо с общей задержкой прохождения сигнала от периферии до центральной нервной системы и обратно. Это означает, что сигналы компенсации вычисляются практически мгновенно. Эти эксперименты показали несостоятельность механизма непосредственной обратной связи, т.е. подачи выходного сигнала на вход системы управления артикуляцией. Вообще говоря, этого следовало ожидать, поскольку размерность пространства и физическая природа выходного сигнала (смещение артикуляторных органов или акустические параметры речевого сигнала) не совпадают с размерностью пространства и физической природой команд управления сокращением двигательных единиц мышц. Очевидно, что замыкание обратной связи происходит путем пересчета выходных сигналов во входные с помощью некоторого модуля.

Аналогичная проблема возникает и при исследовании механизма управления неречевыми движениями человека. Решение проблемы ищут путем введения понятия "схемы тела" или "внутренней модели тела" [26, 27, 28]. Предполагается, что внутренняя модель располагает сведениями о механических свойствах управляемых органов и использует их для текущего контроля или компенсации нарушений процесса управления. В поддержку мнения о врожденном механизме формирования внутренней модели тела можно привести сведения о фантомах врожденно-отсутствующих конечностей [29, 30]. Следует заметить, что модель управляемого объекта, включенная в систему обратной связи, является важным элементом в теории автоматического управления. Возможно, что именно идеи из технической области были восприняты в среде исследователей движений человека.

В применении к управлению артикуляцией, понятие внутренней модели было введено в [31], и затем эта гипотеза развивалась в [32, 33, 9, 34, 35, 36, 37]. Эти исследования показали, что внутренняя модель может генерировать сигналы обратной связи и контролировать качество речи в текущем времени только в том случае, если она умеет решать так называемую обратную задачу – от выходного сигнала к управлению. В частном случае, обратная задача решается, когда входом являются акустические параметры речевого сигнала, а выходом – форма речевого тракта, артикуляторные параметры или команды управления. Способность внутренней модели к контролю процесса речеобразования можно попытаться распространить и на процессы восприятия речи другого человека, предполагая, что процесс обучения речи или языку "на слух" происходит путем пересчета акустических параметров чужой речи в артикуляторные параметры собственного речевого тракта и установления соответствия с фонетическими элементами языка. Так усматривается прямая аналогия с основными положениями моторной теории восприятия речи.

Подобная аналогия возникает и при другом подходе. Теоретический анализ потенциальной надежности распознавания слов при различных способах описания фонетических элементов был выполнен в [31] в сопоставлении с реальной разборчивостью

слов. Оказалось, что при хорошем отношении сигнал/шум избыточность на словесном уровне позволяет достичь хорошей распознаваемости с использованием лишь акустических признаков. В то же время ухудшение условий передачи требует привлечения информации о месте артикуляции, что очень трудно определить на акустическом уровне. Результаты этого анализа нашли подтверждение в экспериментах по измерению активности коры головного мозга с использованием функциональной магнито-резонансной томографии. Было установлено, что при восприятии речи в условиях шумов возникает активность в моторной зоне коры головного мозга, тогда как при хороших акустических условиях активизируется только область слуховой коры [38, 39]. Результаты этих экспериментов служат непосредственным доказательством основного положения моторной теории о том, что в распознавании речи человеком могут принимать участие и моторные компоненты.

Как только стало ясно, что для успешной деятельности внутренней модели необходимо решать обратные задачи различного вида, возникло серьезное возражение против предположения о возможности использования механизмов внутренней модели для восприятия речи других дикторов. Дело в том, что система управления артикуляцией располагает избыточностью на многих уровнях. Одно и то же усилие, развиваемое мышцей, может породиться активизацией различных двигательных единиц. Одно и то же смещение артикулятора может достигаться за счет разного сочетания сокращения мышц, управляющих его движениями. Наименьшая площадь поперечного сечения речевого тракта в определенном месте может достигаться при разном смещении артикуляторов. Например, одно и то же расстояние между губами достигается различными сочетаниями смещения верхней и нижней губы, а также нижней челюсти. Наконец, одно и то же сочетание резонансных частот речевого тракта может быть получено при разных его формах. Такая избыточность обеспечивает надежность процесса речеобразования и его устойчивость к различного рода патологиям и возмущениям. Однако это означает, что при попытке определения формы речевого тракта, артикуляторных параметров или команд управления по акустическим параметрам речевого сигнала принципиально отсутствует однозначное решение. В математике такие задачи называются некорректными. Такая кинематическая неоднозначность, казалось бы, делала бессмысленной постановку обратной задачи для речевого тракта и в значительной степени обесценивала концепцию внутренней модели.

В силу некорректности задачи, попытки определения функции площади поперечного сечения речевого тракта путем обращения волнового уравнения оказались безуспешными (обзор таких работ представлен в [40]). Однако, способность человека к обучению языкам все же указывала на возможность решения обратной задачи от акустики к управлению артикуляцией. Надежда на получение приемлемого решения состояла в переходе от абстрактной математической формулировки задачи без ограничений к использованию многочисленных ограничений, присутствующих в речевой обратной задаче. Рассмотрим некоторые из этих ограничений.

1. Ограничение на силу, развиваемую мышцами, определяет максимальную скорость и ускорение артикуляторов.
2. Масса, присоединенная упругость и вязкое сопротивление также ограничивают динамические свойства артикуляторов.
3. Анатомические размеры речевого тракта и артикуляторные параметры находятся в определенных пределах.
4. Функциональный класс преобразований от артикуляторных параметров к площади поперечного сечения речевого тракта фиксирован.
5. Ограничения на форму речевого тракта для разных типов звуков. Например, ограничения на минимальную площадь речевого тракта для гласных требуют, чтобы при этом не возникали турбулентные шумы. Для фрикативных площадь тракта в месте артикуляции должна находиться в определенных (и очень узких) пределах, гарантирующих генерирование турбулентных шумов с заданными характеристиками. Для взрывных должна быть обеспечена нулевая площадь поперечного сечения в месте артикуляции, а площадь прохода в носовую полость не равна нулю только для назальных.

6. Критерии оптимальности, действующие в системе управления артикуляцией, включают в себя энергетические ограничения и ограничения на сложность программирования управлений.

Система управления артикуляцией, безусловно, имеет возможность определения этих ограничений просто путем совершения пробных движений.

Но даже и ограничения на возможные решения речевой обратной задачи вряд ли помогли бы, если бы не развитие теории некорректных задач [41]. В этой теории было установлено, что при определенных условиях может быть получено устойчивое и однозначное псевдо-решение некорректной обратной задачи. Были также найдены условия, при которых погрешность полученного решения не превышает удвоенную погрешность потенциально достижимого решения, хотя эта потенциальная точность остается неизвестной. Следовательно, вопрос о возможности решения речевой обратной задачи сводится к оценке погрешности решения. Если эта погрешность, например, находится в пределах погрешности измерения, то обратную задачу можно считать решенной.

Среди методов решения некорректных обратных задач одним из наиболее мощных оказался вариационный метод, иногда называемый в технических областях методом "анализа через синтез". Этот метод состоит в вариации параметров математической модели инвертируемого процесса с целью достижения минимального (или заданного) отличия (невязки) между измеренными входными параметрами и параметрами, вычисленными с помощью модели. Таким образом, качественные представления о необходимости существования внутренней модели в системе управления артикуляцией получили математическое обоснование в теории некорректных обратных задач, а сам термин "внутренняя модель" оказался точно соответствующим вариационному методу.

Прежде, чем пытаться распространить концепцию внутренней модели на процессы восприятия, следовало бы убедиться в том, что внутренняя модель действительно может решать обратные задачи типа "акустические параметры - форма речевого тракта", "форма речевого тракта - артикуляторные параметры", "артикуляторные параметры - управления" в случае, когда анатомические размеры и динамические параметры артикуляторов известны. С этой целью, во-первых, нужно решить эти задачи с приемлемой точностью, пользуясь только физиологически правдоподобными параметрами. Во-вторых, степень доказательности концепции внутренней модели существенно возросла бы, если бы удалось воспроизвести некоторые явления и эффекты, не находящие объяснения в рамках других подходов к описанию свойств системы управления речеобразованием.

При таком подходе прежде всего необходимо разработать достаточно точные математические модели, описывающие механику, аэродинамику и акустику процессов речеобразования. В этих моделях часть ограничений определяется путем непосредственных измерений речевого тракта, часть находится путем компьютерного моделирования, а часть должна быть найдена в процессе исследования решений обратной задачи. Детальные модели были описаны в [31, 33]. Уточнение этих моделей стало возможным после того, как методы магнитно-резонансной томографии стали доступны для речевых исследований [42, 43]. В основу модели артикуляции была положена структура мышц, управляющих артикуляторами (Рис. 1). В экспериментах по моделированию решения обратных задач использовалось от 16 до 18 артикуляторных параметров.

В исследовании обратных задач для речевого тракта целесообразно отдельно рассмотреть задачу, когда входными сигналами являются измерения формы речевого тракта, и когда входом служат акустические параметры. Решение первой задачи должно ответить на вопрос о том, действительно ли внутренняя модель может использовать сигналы от проприоцепторов для коррекции артикуляции и компенсации возмущений на артикуляторном уровне. Решение второй задачи определяет возможность коррекции и компенсации нарушений на акустическом уровне. Только положительное решение обеих задач позволяет поставить вопрос о возможности использования механизмов внутренней модели для анализа речи другого человека.

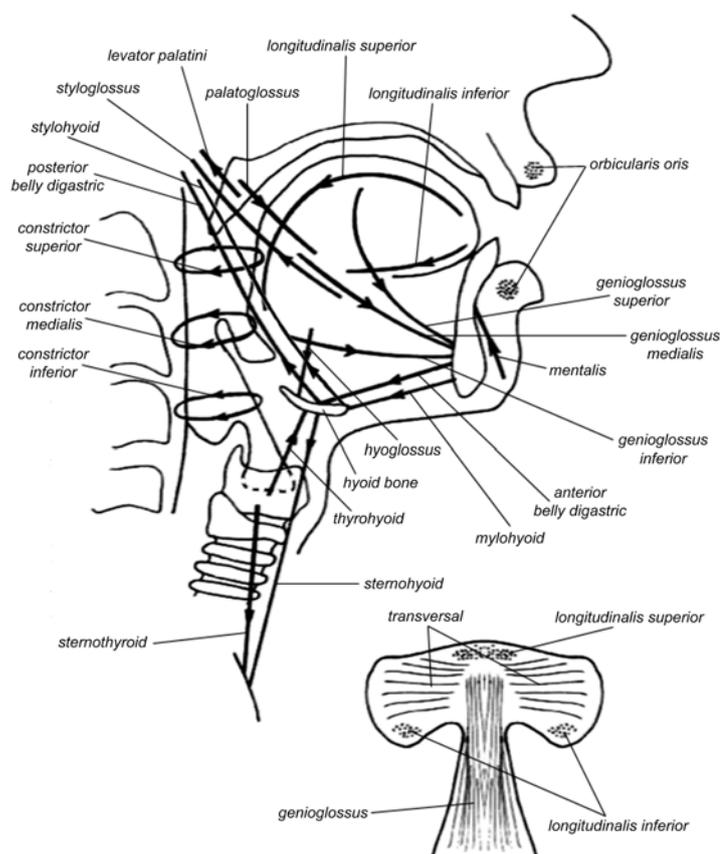


Рис.1. Схема мышц речевого тракта

Задача первого рода - "от формы речевого тракта к управлениям", исследовалась с использованием экспериментальных методик двух видов. В одном из них на кино-рентгенограммах измерялись расстояния между губами, верхними и нижними зубами, кончиком языка и твердым небом, мягким небом и ближайшей точкой на поверхности языка, а также между задней поверхностью речевого тракта и поверхностью языка [34, 36]. В другом исходными данными служили измерения движений артикуляторов на микро-лучевом рентгеноскопе [35].

Решение обратной задачи относительно управлений искалось в виде разрывных кусочно-линейных команд. Экспериментальные оценки амплитудно-частотных характеристик артикуляторов показали, что они ограничены 2 – 7 Гц [31]. Следовательно, исполнение команды, пришедшей меньше, чем через 150 - 500 мс после предыдущей, может быть реализовано только за счет значительного увеличения соответствующих мышечных усилий. По-видимому, это "неблагоприятно" с физиологической точки зрения. Ограничения полосы частот управляемых объектов приводят к тому, что в движениях человека новые команды обычно не регистрируются до определенного времени после предыдущей команды [44]. В этих экспериментах точность аппроксимации измеренных движений после решения обратной задачи оказалась в пределах точности измерений. Было также найдено, что ступенчатые управления могут привести к значительному перерегулированию.

Следующий этап проверки концепции внутренней модели должен состоять в попытке воспроизведения хорошо известных эффектов, приписываемых активности внутренней модели. Рассматривались два таких эффекта – компенсация байт-блока и реорганизация управлений при изменении темпа артикуляции. Выше упоминались эксперименты по ограничению движений

нижней челюсти, в которых была обнаружена перестройка движений остальных артикуляторов. Компьютерная симуляция байт-блока на артикуляторной модели должна была установить наличие такой компенсации, если в систему управления включена внутренняя модель.

Известно, что вариации темпа артикуляции изменяют не только амплитуды ЭМГ потенциалов, но также и временные характеристики и интервалы активизации нейромоторных команд (см. например, [45, 46]). В [47] было найдено, что "партитура" нейромоторных команд, т.е. их амплитуды, длительность и положение во времени существенно реорганизуются в зависимости от темпа артикуляции и положения слога в звукосочетании. Возможные механизмы наблюдаемой реорганизации были не известны.

Оба явления исследовались в [35, 37], где было найдено, что и компенсация байт-блока и реорганизация управлений являются результатом использования внутренней моделью энергетических критериев оптимальности, связанных со смещением от нейтрального положения, скорости и ускорения артикуляторов. Оказалось также, что неречевые и речевые движения артикуляторов управляются разными типами критериев – мгновенными или накопленными на некотором интервале ([35]).

Таким образом, была не только установлена принципиальная возможность решения обратной задачи "от формы речевого тракта к управлению", но и конкретизированы механизмы компенсации нарушений артикуляции и реорганизации управлений при смене темпа речи. Однако, эта возможность может быть реализована только при условии достаточно точного измерения смещения, скорости и ускорения артикуляторов системой проприоцепторов. Тактильные рецепторы дают информацию либо о локальной деформации тела языка, либо о касании им неподвижной поверхности речевого тракта. Мышечные веретена генерируют сигналы, пропорциональные степени сокращения и скорости сокращения мышцы. Вибротактильные рецепторы откликаются на ускорения. Рецепторы Гольджи, расположенные в челюстном суставе также реагируют на ускорение движений нижней челюсти. Однако, проприоцепторы не дают информации непосредственно о расстояниях между артикуляторами и неподвижной поверхностью, за исключением касания. В [31] было показано, что погрешность оценки, например, подъема кончика языка по отклику веретен, достаточно велика. Поэтому можно ожидать, что способность к компенсации нарушений у разных людей различается в зависимости от точности обработки проприоцептивной информации.

Действительно, в экспериментах [13], только один из 11 испытуемых оказался в состоянии компенсировать блок губ после длительной тренировки. Не всем удается компенсировать и байт-блок [48, 49]. Также, не все люди способны восстановить привычное качество речи даже после длительного ношения протеза с искусственным твердым небом [14]. Лишь около 10% больных с удаленной гортанью могут полностью восстановить голосовой источник [9]. Это свидетельствует о том, что обратная задача даже на артикуляторном уровне решается с большим трудом.

Следующий этап исследований состоял в решении обратной задачи, когда входные данные включали в себя акустические параметры речевого сигнала и координаты нескольких опорных точек на губах, нижней челюсти и передней части языка. Траектории 8 точек были получены из базы данных, сформированной с использованием микролучевого рентгеноскопа [50]. Речевой материал содержал все звуки английского языка: гласные, полугласные, назальные, взрывные и фрикативные. Исследовались отдельные звуки, слоги, слова и фразы. Средние погрешности аппроксимации входных данных их аналогами, вычисленными при решении обратной задачи для гласных, составили: для трех первых формант – 3.7%, 3.8% и 2.6%; для характерных частот спектров фрикативных – 8.5%; для координат измеренных точек всех типов звуков - 2.8% [51, 52]. По вычисленной форме речевого тракта были синтезированы отдельные звуки и звукосочетания. Согласно субъективной оценке, исходные и ресинтезированные звукосочетания оказались очень похожи.

Так было установлено, что при наличии акустической информации нет необходимости в измерении всей формы речевого тракта, а достаточно знания о положении губ, нижней челюсти и передней части языка. Тем самым в задаче коррекции артикуляции или компенсации

нарушений ослабляются требования к точности проприоцептивных сигналов. Это повышает вероятность того, что система управления артикуляцией способна решать обратные задачи с целью контроля качества генерируемого речевого сигнала и соответствия его установившимся фонетическим нормам в данном языке.

В процессе этих исследований было также найдено, что для успешного решения обратной задачи необходимо предварительно сегментировать речевой сигнал, по крайней мере, на четыре основных типа: гласноподобные, фрикативные, назальные и смычки. Это связано с различием акустических процессов для этих типов звуков и, соответственно, требует разных способов обработки речевого сигнала и разных критериев оптимальности. Такая сегментация представляет собой не что иное, как распознавание ограниченного набора классов, признаки которых относительно легко вычисляются на акустическом уровне. Следовательно, между решением обратной задачи, и распознаванием речевого сигнала с помощью чисто акустического анализа существует естественная связь, и решение обратной задачи может быть выполнено только при условии распознавания некоторых типов звуков на акустическом уровне.

Возвращаясь к вопросу о том, может ли внутренняя модель в системе управления артикуляцией решать обратную задачу относительно формы речевого тракта, имея на входе лишь акустический сигнал, необходимо провести еще одно модельное исследование. Пусть такая задача решается для речевого тракта с известными анатомическими параметрами, но без использования текущего измерения координат хотя бы нескольких точек внутри речевого тракта. Необходимо найти такие условия, при которых точность решения относительно измеренной формы тракта окажется достаточно высокой. Работы [53, 54, 54, 56, 40], а также контрольные эксперименты, выполненные в [52], показали, что для гласных и фрикативных может быть достигнута точность восстановления формы речевого тракта или координат некоторых точек внутри тракта, сопоставимая с точностью измерения. В этих исследованиях выяснились и условия, при которых может быть получено удовлетворительное решение.

В частности, при решении обратной задачи для гласных из базы данных [50] с использованием только формантных частот, точность аппроксимации координат опорных точек несколько ухудшилась – с 2.8% до 6%, тогда как точность аппроксимации формантных частот даже улучшилась 1.8% для  $F_1$ , 1.6% для  $F_2$ , 1.1% для  $F_3$ . Это объясняется тем, что оптимизатор не принимает во внимание координаты опорных точек, а стремится достичь минимума ошибки только по акустическим параметрам. При решении обратной задачи для фрикативных с информацией об опорных точках и без такой информации, результат несколько иной. Точность аппроксимации опорных точек осталась прежней, тогда как точность аппроксимации характерных частот спектра несколько ухудшилась – до 13.4%. В целом же, разные эксперименты подтверждают возможность более или менее точного восстановления формы речевого тракта (или артикуляторных параметров) на основе анализа только акустических параметров. Пример практически полного совпадения решений приведен на Рис. 2.

При математическом моделировании обратной задачи используются формальные процедуры поиска оптимума некоторой целевой функции. В силу принципиальной неоднозначности решения обратной задачи, наряду с глобальным оптимумом, в целевой функции всегда присутствует и множество локальных экстремумов. Никакой из известных методов оптимизации не гарантирует нахождения глобального экстремума. Один из способов поиска если не глобального, то хотя бы экстремума, доставляющего приемлемую точность решения, состоит в многократном решении задачи оптимизации с использованием различных начальных приближений. Множество таких начальных приближений образует так называемую кодовую книгу. Понятие кодовой книги для речевых задач было введено в [57]. Идея кодовой книги состоит в вычислении соответствия между заданным вектором артикуляторных параметров, и соответствующим ему вектором акустических параметров, используя математическую модель речеобразования. Множество таких пар векторов, оформленное в виде кодовой книги, позволяет для измеренного вектора акустических параметров выбрать множество артикуляторных векторов, которые и используются как начальные приближения.

Помимо применения в формально-математическом исследовании речевых обратных задач, концепция кодовой книги важна при обсуждении физиологических аспектов решения этих задач.

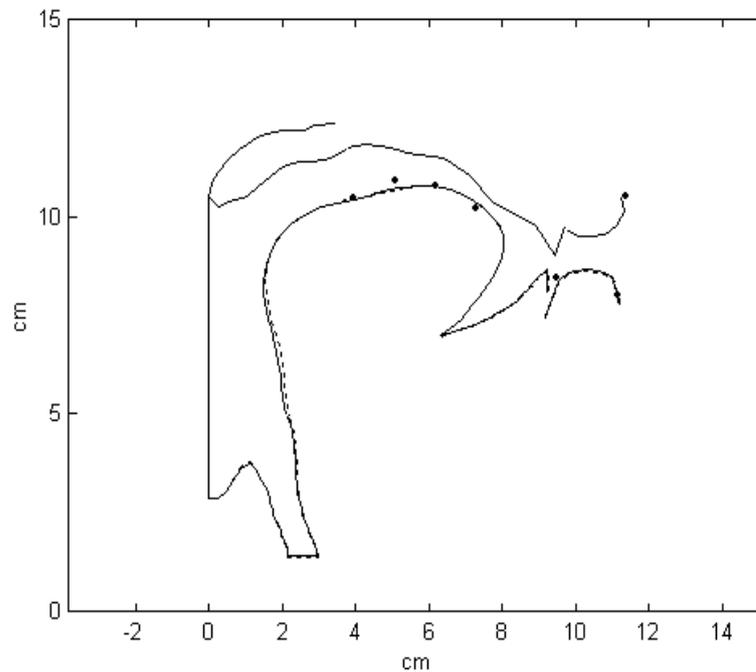


Рис. 2. Форма речевого тракта, полученная решением обратной задачи с использованием и акустических и артикуляторных данных (—), и только для акустических данных (- - -).

Представляется маловероятным, что в процессах управления артикуляцией внутренняя модель действительно решает спектральную задачу для волнового уравнения, вычисляет характеристики источников возбуждения и использует процедуры оптимизации типа градиентного спуска. Скорее всего, организм использует другие средства для решения обратных задач. Кодовая книга может быть одним из таких средств. Решение обратной задачи не обязательно сопровождается процессом оптимизации. Если выбранный из кодовой книги артикуляторный вектор или команда управления удовлетворяет, например, текущим ограничениям на энергетические затраты, то этот вектор и может быть сразу принят в качестве решения. Лишь изменение условий артикуляции вроде искусственного ограничения движений артикулятора потребует запуска процесса оптимизации.

Возможность формирования кодовой книги в системе управления артикуляцией для собственного речевого тракта не вызывает особенных возражений. Можно предположить, что, начиная с периода детского лепета, и пользуясь методом проб и ошибок, внутренняя модель отслеживает процесс изменения анатомических размеров речевого тракта и соответствующим образом корректирует содержание кодовой книги. Примеры восстановления голосового источника после удаления гортани указывают также на удивительную пластичность системы управления рчеобразованием, способную не только отслеживать возрастные изменения параметров речевого тракта, но и менять саму структуру внутренней модели. При этом роль суррогатного голосового источника принимает на себя сфинктер, образованный пищеводом и мышцами-сжимателями глотки, которым и передаются тонкие функции мышц удаленной гортани.

Может оказаться также, что каждое значительное изменение в размерах тракта не приводит к стиранию содержания предыдущего состояния, а включается в состав кодовой книги в виде дополнения. В результате, устанавливается соотношение между акустическими

параметрами и различными размерами речевого тракта, а это может использоваться для решения обратной задачи и для речи других людей с похожими анатомическими размерами. В этом случае речь другого диктора интерпретируется в терминах артикуляторных параметров или формы речевого тракта слушателя.

Кодовые книги для дикторов с другими анатомическими параметрами могут сформироваться и путем наблюдения за их артикуляторной мимикой. Из ежедневной практики хорошо известно, что визуальная информация облегчает восприятие речи других людей, особенно в неблагоприятных акустических условиях или для иностранного языка [58]. Информативность наблюдаемых проявлений артикуляции позволяет общаться глухонемым. Наблюдение за лицом диктора влияет на восприятие речи, и в случае противоречия между видимой артикуляцией и услышанным звуко сочетанием возникают разнообразные эффекты восприятия [58]. Результаты исследований внешних проявлений аудио-визуальных эффектов подтверждаются измерением активности мозга. Например, была обнаружена электрическая активность слуховой зоны коры головного мозга слушателя, наблюдающего за артикуляторными движениями диктора, тогда как неречевая мимика диктора такой активности не вызывала [60, 61].

Таким образом, кодовая книга, созданная для обеспечения управления артикуляцией, может быть дополнена в процессах обучения пониманию речи других людей. То, что этот механизм не исчезает после периода становления речи у детей, подтверждается способностью к усвоению иностранных языков и приспособлению к пониманию речи людей с особенностями произношения.

С тех пор, как были сформулированы положения моторной теории восприятия речи, были получены как прямые, так и косвенные доказательства того, что интуитивные качественные представления о возможной роли моторной компоненты в восприятии речи оказались верны. Об этом свидетельствует далеко не полный список публикаций, приведенный в данной статье. Правда, при этом выяснилось, что в основе механизмов анализа речи лежат гораздо более сложные, чем это казалось ранее, процессы, связанные с решением обратных задач разного вида в специальном модуле системы управления артикуляцией, называемом внутренней моделью. Деятельность внутренней модели разделяется на два вида. Первый из них выполняет контроль за качеством речи и обеспечивает компенсацию патологии и помех речеобразованию, решая обратные задачи с входными данными в виде сигналов от проприоцепторов и от собственной слуховой системы. Способность внутренней модели к обработке акустической информации может использоваться и в задаче другого вида, т.е. в режиме анализа речи других людей.

Концепция внутренней модели и конкретизация понятия некорректных обратных задач в применении к речи создают основу для исследования процесса восприятия речи не только на качественном, но и на количественном уровне, и обеспечивают конструктивный подход к экспериментальной методике и оценке результатов.

## Литература

1. Галунов В.И., Чистович Л.А. (1965). О связи моторной теории с общей проблемой распознавания речи. *Акустический ж.*, т. 11, с. 417-426.
2. Liberman A., Cooper F., Shankweiler D. and Studdert-Kennedy M. (1967). Perception of speech code. *Physiological review*, vol. 74, N 6, pp. 431 - 461.
3. Liberman A. and Mattingly I. (1985). The motor theory of speech perception revised. *Cognition*, vol. 21, pp. 1 - 36.
4. Lindblom B. (1996). Role of articulation in speech perception: Clues from production. *J. Acoust. Soc. Am.*, vol. 99, No. 3, pp. 1683-1692.
5. Ohala J.J. (1996). Speech perception is hearing sound, not tongues. *J. Acoust. Soc. Am.*, vol. 99, No. 3, pp. 1718-1725.

6. O'Shaughnessy D. (1996). Critique: Speech perception: Acoustic or articulatory?. *J. Acoust. Soc. Am.*, vol. 99, No. 3, pp. 1726-1729.
7. Remez R.E. (1996). Critique: Auditory form and gestural topology in the perception of speech. *J. Acoust. Soc. Am.*, vol. 99, No. 3, pp. 1695-1698.
8. Stevens K.N. (1996). Critique: Articulatory-acoustic relations and their role in speech perception. *J. Acoust. Soc. Am.*, vol. 99, No. 3, pp. 1693-1694.
9. Sorokin V., Olshansky V., Kozhanov L. (1998). Internal model in articulatory control: Evidence from speaking without larynx. *Speech Communication*, vol. 25, N 3, pp. 249-268.
10. Lindblom B.E.F., Lubker J. and Gay T. (1979). Formant frequencies of some fixed mandible vowels and a model of speech motor programming by predictive simulations. *J. Phonetics*, vol. 7, pp. 147-161.
11. Fowler C.A., Turvey M.T. (1980). Immediate compensation in bite-block speech. *Phonetica*, vol. 37, pp. 306-326.
12. Gay T.J., Lindblom B., Lubker J. (1981). Production of bite-block vowels : Acoustic equivalence by selective compensation. *J. Acoust. Soc. Am.*, vol. 69, pp. 802-810.
13. Savariaux C., Perrier P. and Orliquet J.P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in the speech production. *J. Acoust. Soc. Am.*, vol. 98, No. 5, pp. 2428-2442.
14. Hamlet S.L., Stone M. (1978). Compensatory alveolar consonant production induced by wearing a dental prosthesis. *J. Phonetics*, vol. 6, pp. 227-248.
15. MacFarland D.H., Baum S.R. (1995). Incomplete compensation to articulatory perturbation. *J. Acoust. Soc. Am.*, vol. 97, N 3, pp. 1865-1873.
16. Folkins J.W., Abbs J.H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *J. Hearing and Speech Res.*, vol. 18, pp. 207-220.
17. Abbs, J.H., Gracco, V.L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech, *J. Neurophysiology*, vol. 51, pp. 705-723.
18. Kelso J.A.S., Tuller B., Vatiokis-Bateson E., Fowler C. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *J. of Experimental Psychology: Human Perception and Performance*, vol. 10, pp. 812-832.
19. Shaiman S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *J. Acoust. Soc. Am.*, vol. 86, pp. 78-87.
20. Kollia H.B., Gracco V.L., Harris K.S. (1992). Functional organization of velar movements following jaw perturbation. *J. Acoust. Soc. Am.*, vol. 91, p. 2474.
21. Munhall K.G., Lofqvist A., Kelso J.A.S. (1994). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *J. Acoust. Soc. Am.*, vol. 95, N 6, pp. 3605-3616.
22. Honda M., Kaburagi T. (2000). Speech compensation to dynamical structural perturbation of the palate shape. In: *Proc. 5<sup>th</sup> Seminar on Speech Production*, Kloster Seeon, Bavaria, pp. 21-24.
23. Folkins J.W., Zimmerman G.N. (1982). Lip and jaw interactions during speech : responses to perturbation of lower-lip movement prior to bilabial closure. *J. Acoust. Soc. Am.*, vol. 71, N 4, pp. 1225 - 1233.
24. Burnett T.A., Freedland M.B., Larson C.R., Hain T.C. (1998). Voice F<sub>0</sub> responses to manipulations in pitch feedback, *J. Acoust. Soc. Am.*, vol. 103, pp. 3153-3161.
25. Jones J.L., Munhall K.G. (2000). Perceptual calibration of F<sub>0</sub> production: Evidence from feedback manipulation. *J. Acoust. Soc. Am.*, vol. 108, pp. 1246-1251.
26. Poeck K., Orgass B. (1971). The concept of body scheme, a critical review and experimental results. *Cortex*, vol. 5, pp. 254 - 277.
27. R.A.Schmidt (1976). The schema as a solution to some persistent problems in motor learning theory, in *Motor Control* (Academic Press, N.Y.), pp. 41-66.
28. Kelso S.J., Stelmach G.E. (1976). Central and peripheral mechanisms in motor control, in *Motor Control. Issues and Trends*, Stelmach, G.E., (Ed.), Academic Press, N.Y., pp. 3-40.
29. Weinstein S., Sersen E.A. (1961). Phantoms in cases of congenital absence of limbs. *Neurology*, vol. 10-11, pp. 905-911.
30. Vetter R.J., Weinstein S. (1967). The history of the phantom in congenital absent limbs. *Neuropsychology*, vol. 5, pp. 335-338.
31. Сорокин В.Н. (1985). Теория речеобразования, Радио и Связь, М.

32. Sorokin V.N. (1987). Coordination of muscles and articulators. In: *Proc. XI-th International Congress of Phonetic Sciences* (Tallinn), Vol. 3, pp. 382-384.
33. Сорокин В.Н. (1992). Синтез речи, Наука, М.
34. Леонов А.С., Сорокин В.Н. (2000). Обратная задача для управления артикуляцией. Доклады Академии Наук, т. 374, № 6, с. 749-753.
35. Леонов А.С., Сорокин В.Н. (2003). Энергетические критерии оптимальности в речевых обратных задачах, Доклады Академии наук, т. 392, № 5, 694-699.
36. Leonov A.S., Sorokin V.N. (2000). Inverse problem for the vocal tract: Identification of control forces from articulatory movements. *Pattern Recognition and Image Analysis*, vol. 10, N 1, pp. 110-126.
37. Leonov A.S., Sorokin V.N. (2004). Controls in the internal model: Score reorganization and compensation. *Pattern Recognition and Image Analysis*, vol.14, N 3, pp. 407-420.
38. Callan D.E., Callan A.M., Kroos Ch., Vatikiotis-Bateson E. (2000). Neural processes underlying perception of audio-visual speech production. Proc. 5<sup>th</sup> Seminar on Speech Production, Kloster Seeon, pp. 273-276.
39. Sekiyama K, Sugita Y. (2002). Auditory-visual speech perception examined by brain imaging and reaction time. Proc. 7<sup>th</sup> Int. Conf. On Spoken Language Processing, Denver, pp. 1693-1696.
40. Sorokin V.N., Leonov A.S. and Trushkin A.V. (2000). Estimation of stability and accuracy of inverse problem solution for the vocal tract. *Speech Communication*, vol. 30, N1, pp. 55-74.
41. Тихонов А.Н., Арсенин В.Я. (1974). Методы решения некорректных задач. М. Наука.
42. Макаров И.С., Сорокин В.Н. (2004). Резонансы разветвленного речевого тракта с податливыми стенками. *Акустический ж.*, т. 50, № 3, с. 389-396.
43. Баден П., Макаров И.С., Сорокин В.Н. (2005). Алгоритм вычисления площадей поперечных сечений речевого тракта. *Акустический ж.*, т. 51, №1, с. 52-58.
44. Schmidt R.A. (1982). More on motor programs. In: *Human Motor Behavior*, Kelso, S., (Ed.) , Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 189-218.
45. Kent R.D., Carney P.J., Severeid L.R. (1974). Velar movement and timing: evaluation of a model for binary control. *J. of Speech and Hearing Research*, vol. 17, pp. 470-488.
46. Gay T., Ushijima T., Hirose H., Cooper F. (1974). Effect of speaking rate on labial and consonant-vowel articulation. *J. Phonetics*, vol. 2, pp. 47-63.
47. Сорокин В.Н. (1980). Влияние скорости артикуляции на нейромоторные процессы лицевых мышц. *Физиология человека*, N 1, с. 40-45.
48. Flege I.E., Fletcher S.G. and Homiedan A. (1988). Compensating for a bite block in /s/ and /t/ production: Palatographic, acoustic, and perceptual data. *J. Acoust. Soc. Am.*, vol. 83, N 1, pp. 212-228.
49. McFarland D.H., Baum S.R., Chabot C. (1996). Speech compensation to structural modifications of the oral cavity. *J. Acoust. Soc. Am.*, vol. 100, pp. 1093-1104.
50. Westbury J. (1994). *X-ray Microbeam Speech Production Database*. User's Handbook, Version 1.,
51. Леонов А.С., Макаров И.С., Сорокин В.Н., Цыплихин А.И. (2003). Артикуляторный ресинтез гласных, Информационные процессы, т. 3, №2, 73-92. <http://www.jip.ru>.
52. Леонов А.С., Макаров И.С., Сорокин В.Н., Цыплихин А.И. (2004). Артикуляторный ресинтез фрикативных, Информационные процессы, т. 4, №2, 141-159. <http://www.jip.ru>.
53. Shirai K., Kobayashi T. (1986). Estimating articulatory motion from speech wave. *Speech Communication*, vol. 5, pp. 159-170.
54. Sorokin V.N. (1992). Determination of vocal tract shape for vowels. *Speech Communication*, vol. 11, N 1, pp. 71 - 85.
55. Sorokin V.N. and Trushkin A.V. (1996). Articulatory-to-acoustic mapping for inverse problem. *Speech Communication*, vol. 19, N 4, pp. 105-118.
56. Hogden J., Lofqvist A., Gracco V., Zlokarnik I., Rubin P., Saltzman E. (1996). Accurate recovery of articulator positions from acoustics: New conclusions based on human data. *J. Acoust. Soc. Am.*, vol. 100, pp. 1819-1834.
57. Atal B.S., Chang J.J., Mathews M.V., Tukey J.W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting technique, *J. Acoust. Soc. Am.*, vol. 63, pp. 1535-1555.

58. Sumbly W.Y., Pollack I. (1954). Visual contribution to speech intelligibility in noise, *J. Acoust. Soc. Am.* vol. 26, pp. 212-215.
59. McGurk H., MacDonald J. (1976). Hearing lips and seeing voices, *Nature*, vol. 264, pp. 746-748.
60. Sams M., Aulanko R., Hamalainen H., Lounasmaa O., Lu S., & Simola J. (1991). Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, vol. 127, pp. 141-145.
61. Calvert G., Brammer M., Bullmore E., Campbell R., Williams S., McGuire P., Voodruff P., Iversen S., & David A. (1997). Activation of auditory cortex during silent lip-reading. *Science*, vol. 276, pp. 593-596.