архив журнала "Звукорежиссер" : 2002 : №1

Часть 17.1
Слух и речь, ч.1

Ирина Алдошина

От автора

Начиная с этого номера журнала вниманию читателей предлагается серия статей, посвященных слуховому восприятию речевых сигналов (включая вокальную речь, т.е. пение). С одной стороны, запись и обработка речи и пения является одной из труднейших задач в звукорежиссуре, с другой стороны, за последние годы в психоакустике появилось много новых результатов по слуховой оценке и восприятию речи (особенно за прошедшее десятилетие, в связи с задачами обучения компьютера пониманию и синтезу речевых сигналов). Поэтому данная информация может оказаться полезной для практической работы.

Название этой серии статей – "Слух и речь" – появилось не случайно. В 1973 г. под таким же названием было издано хорошее учебное пособие проф. Я.Ш. Вахитова (ЛИКИ). По нему училось не одно поколение студентов, но за прошедшие тридцать лет многое изменилось в понимании процессов создания и восприятия речи.

В представленной серии предполагается последовательно рассмотреть вопросы механизмов образования речи, интегральные и статистические характеристики речевых сигналов, методы их оценки и распознавания, и механизмы восприятия речи (разумеется достаточно кратко и на популярном уровне, поскольку проблемы эти чрезвычайно сложны и еще многое в них остается неразгаданным). В России имеется хорошая научная школа, занимающаяся акустикой и восприятием речи, поэтому издано достаточно много русскоязычных книг: труды Л.А. Чистович, Г.А. Вартанян, В.П. Морозова, В.И. Галунова и др. Имеется также огромное количество монографий и учебников на английском, немецком и др. языках. Для тех, кто глубоко заинтересуется этими проблемами, в конце серии статей будет приведена основная библиография.

Основные механизмы звукообразования речи

Речевой сигнал является средством передачи разнообразной информации как вербальной (словесной), так и невербальной (эмоциональной). Для быстрой передачи информации в процессе эволюции был отобран особым образом закодированный и структурированный акустический сигнал. Для создания такого специализированного акустического сигнала используется "голосовой аппарат", совмещенный с физиологическим аппаратом, предназначенным для дыхания и жевания (поскольку речь возникла на поздних стадиях эволюции, то к речеобразованию пришлось приспособить уже имеющиеся органы

Процесс образования и восприятия речевых сигналов, схематически показанный на рисунке 1, включает в себя следующие основные этапы: формулировка сообщения, кодирование в языковые элементы, нейромускульные действия, движения элементов голосового тракта, излучение акустического сигнала, спектральный анализ и выделение акустических признаков в периферической слуховой системе, передача выделенных признаков по нейронным сетям, распознавание языкового кода (лингвистический анализ), понимание смысла сообщения.

Рис. 1
Основные процессы образования и восприятия речи

Голосовой аппарат является, по существу, духовым музыкальным инструментом. Однако среди всех музыкальных инструментов он не имеет себе равных по своей многогранности, разносторонности, возможности передачи малейших оттенков и др. Все способы звукоизвлечения, которые используются в духовых инструментах, используются и в процессе образования речи (в т.ч. вокальной речи), однако все они перестраиваемы (по приказам мозга), и имеют широчайшие возможности, недоступные ни одному инструменту.

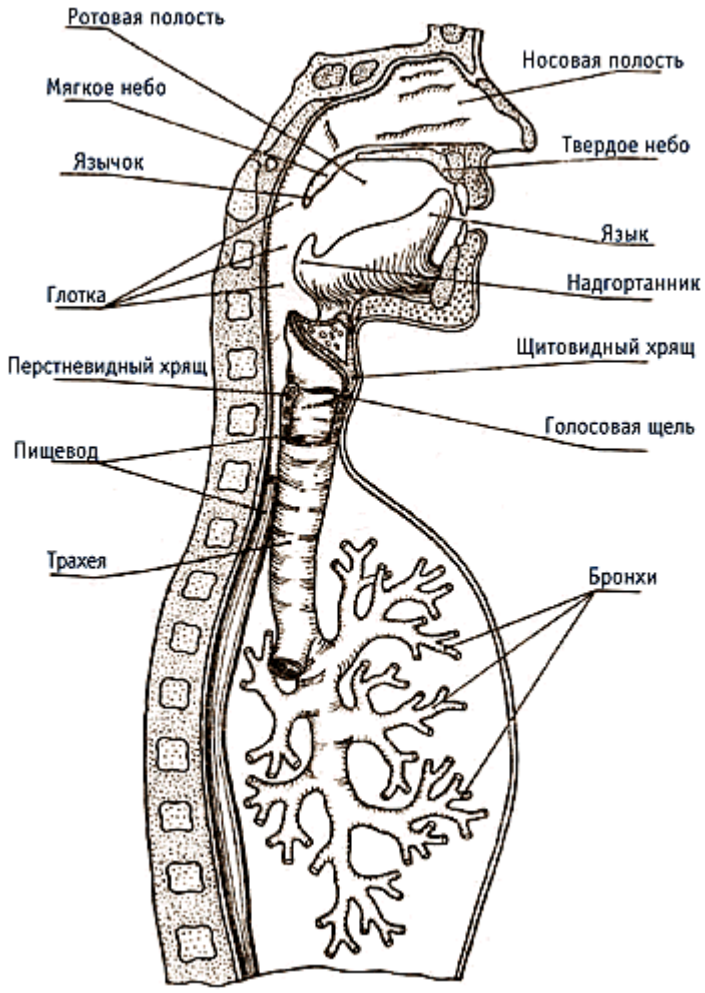

Рис. 2
Структура голосообразующего аппарата

Если рассматривать структуру голосообразующего аппарата как духового музыкального инструмента, он состоит из трех основных частей (рисунок 2):

- **генератора** – дыхательной системы, состоящей из воздушного резервуара (легких), где запасается энергия избыточного давления, мускульной системы и выводного канала (трахеи) со специальным аппаратом (гортанью), где воздушная струя прерывается и модулируется;

- **вибраторов** – голосовых связок, воздушных турбулентных струй (создающих краевые тоны), импульсных источников (взрывов);

- **резонаторов** – разветвленной и перестраиваемой системы резонансных полостей сложной геометрической формы (глотки, ротовой и носовой полости), называемой артикуляционной системой.

Генерация энергии воздушного столба происходит в легких, которые представляют собой своеобразные меха, создающие поток воздуха при вдохе и выдохе за счет разницы атмосферного и внутрилегочного давления. Процесс вдоха и выдоха происходит за счет сжатия и расширения грудной клетки, которые осуществляются обычно с помощью двух групп мышц: межреберных и диафрагмы, при глубоком усиленном дыхании (например, при пении) сокращаются также мышцы брюшного пресса, груди и шеи. При вдохе диафрагма уплощается и опускается вниз, сокращение наружных межреберных мышц поднимает ребра и

отводит их в стороны, а грудину – вперед. Увеличение грудной клетки растягивает легкие, что приводит к падению внутрилегочного давления по отношению к атмосферному, и в этот "вакуум" устремляется воздух. При выдохе мускулы расслабляются, грудная клетка за счет своей тяжести возвращается в исходное состояние, диафрагма поднимается, объем легких уменьшается, внутрилегочное давление растет, воздух устремляется в обратном направлении. Таким образом, вдох – процесс активный, требующий затраты энергии, выдох – процесс пассивный. При обычном дыхании этот процесс происходит примерно 17 раз в минуту, управление этим процессом как при обычном дыхании, так и при речи, происходит бессознательно, но при пении процесс постановки дыхания происходит сознательно и требует длительного обучения.

Количество энергии, которое может быть израсходовано на создание речевых акустических сигналов, зависит от объема запасенного воздуха и соответственно от величины дополнительного давления в легких. Учитывая, что максимальный уровень звукового давления, который может развивать певец (имеется в виду оперный), составляет 100…112 дБ, то очевидно, что голосовой аппарат является не очень эффективным преобразователем акустической энергии, Его КПД составляет порядка 0,2%, как и у большинства духовых инструментов.

Модуляция воздушного потока (за счет вибраций голосовых связок) и создание подглоточного избыточного давления происходит в гортани. Гортань (larynx) – это клапан, (рисунок 3), который находится на конце трахеи (узкой трубки, по которой воздух поднимается из легких). Этот клапан предназначен для предохранения трахеи от попадания посторонних предметов и для поддержания высокого давления при подъеме тяжестей. Именно этот аппарат и используется в качестве голосового источника при речи и пении. Гортань образована из набора хрящей и мышц. Спереди ее охватывает щитовидный хрящ (thyroid), сзади – перстневидный хрящ (cricoid), сзади также располагаются более мелкие парные хрящи: черпаловидные, рожковидные и клиновидные. Сверху гортани расположен еще один хрящ-надгортанник (epiglottis), также типа клапана, который опускается при глотании и закрывает гортань. Все эти хрящи соединены мышцами, от подвижности которых зависит скорость поворота хрящей. С возрастом подвижность мышц уменьшается, хрящи также становятся менее эластичными, поэтому возможности виртуозного владения голосом при пении также уменьшаются.

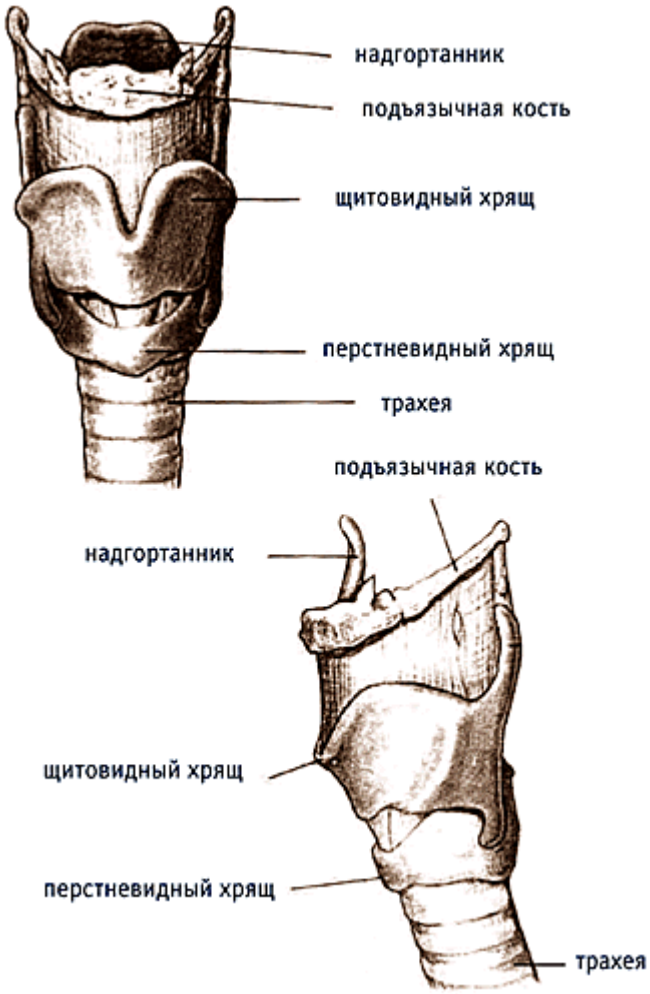

Рис. 3
Строение гортани

Наиболее сложно устроен средний отдел гортани (рисунок 4), в котором расположены парная мышечная перегородка (эластичный конус) и две пары складок. Верхние называются преддверными, или "ложными голосовыми", а нижние – голосовыми. В толще последних лежат голосовые связки, образованные эластическими волокнами, и мышцы (рисунок 5). Промежуток между правой и левой голосовыми складками называется голосовой щелью. Голосовые связки натянуты между щитовидным и черпаловидным хрящами. Размеры голосовой щели в открытом состоянии 2 см в длину и 1 см в ширину.

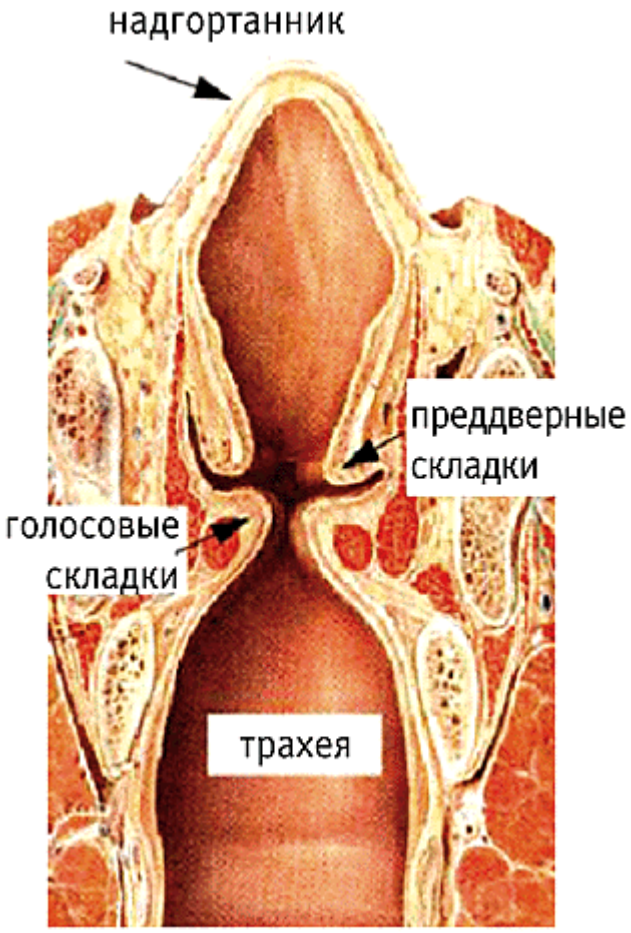

Рис. 4
Сечение трахеи и гортани

Именно **голосовые складки** и являются основным (но не единственным) источником голосообразования (вибратором). Преддверные голосовые складки выделяют специальную слизистую секрецию, которая помогает смазывать голосовые складки и предохраняет их от повреждения при трении во время звукообразования. Обычно они не участвуют в процессе звукообразования, однако при некоторых патологиях истинных связок, они могут участвовать в образовании звука (например, пение Луи Армстронга). *(Хрипота голоса Армстронга была вызвана бородавчатыми образованиями на голосовых связках – это лейкоплакия, проявляющаяся как участки ороговения эпителия. Диагноз "лейкоплакия" был поставлен артисту в зрелом возрасте, но хрипота в голосе присутствует уже на его первых записях, сделанных в возрасте 25 лет. – прим. ред.)*.

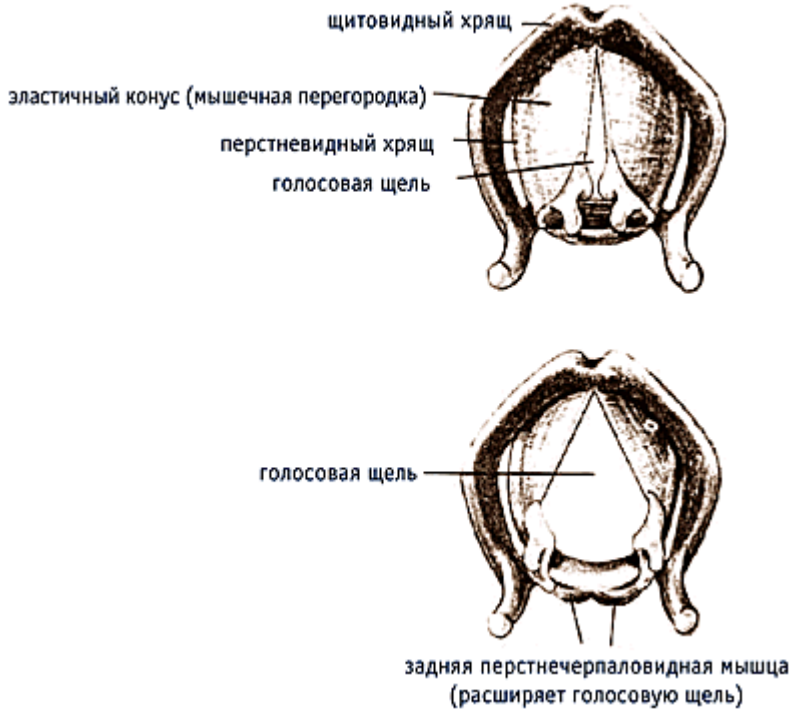

Рис. 5
Голосовая щель и голосовые складки в динамике

Между двумя парами складок находятся небольшие полости (желудочки гортани), которые позволяют беспрепятственно голосовым складкам и играют роль *акустических фильтров*, уменьшая уровень высоких гармоник (скрипучесть голоса), они же играют роль резонаторов для тихих тонов и при пении в фальцете. При движении черпаловидных хрящей голосовые складки могут сдвигаться и раздвигаться, открывая проход воздуха. При поворотах щитовидного и перстневидного хрящей они могут растягиваться и сжиматься, при активации вокальных мышц они могут расслабляться и напрягаться. Процесс образования звуков речи определяется движением (колебаниями) связок, что приводит к модуляции потока воздуха выдыхаемого из легких. Такой процесс называется *фонацией* (существуют и другие механизмы звукообразования, они будут рассмотрены дальше).

Начнем с рассмотрения ***процесса фонации***: перед началом речи голосовые складки должны быть сведены черпаловидными хрящами, что приводит к запиранию потока воздуха и возникновению избыточного подглоточного давления (происходит "предфонационная настройка"). Воздух, который выталкивается легкими из трахеи, накапливается в подскладочном пространстве, и начинает давить на них. Когда избыточное давление повышается до определенной величины, складки размыкаются и воздух устремляется в голосовую щель. В момент максимального открытия щели скорость потока воздуха становится максимальной, давление внутри щели падает (по закону Бернулли), причем скорость протекания воздуха неодинакова – в самой узкой части голосовой щели она максимальна. Внутри голосовой щели образуется зона пониженного давления. Окружающее более высокое давление, а также собственная упругость связок заставляют складки сомкнуться. Этот процесс аналогичен возбуждению колебания тростей в деревянных духовых инструментах. Таким образом, чередование избыточного давления в подскладочном пространстве и отрицательного давления из-за эффекта Бернулли заставляет складки смыкаться-размыкаться, т.е. обеспечивает нормальный режим их колебаний (рисунок 6). При этом происходит модуляция потока воздуха, который порциями (как в духовых инструментах) вталкивается в резонансные полости. Последовательность воздушных толчков, возникающих

в результате колебаний голосовых связок, называется глоттальной волной, обычно она представляется в виде зависимости объемной скорости воздуха от времени (рисунок 7). Как видно из графиков, такой сигнал представляет собой последовательность импульсов, форма которых зависит от соотношения времени открытия складок (скорость потока постепенно нарастает) и времени их закрытия (скорость быстро уменьшается). Период такой волны определяется длительностью общего цикла колебаний связок, т.е. основной частотой колебания. Амплитуда определяется максимальной скоростью потока воздуха, которая, в свою очередь, зависит от величины подскладочного избыточного давления.

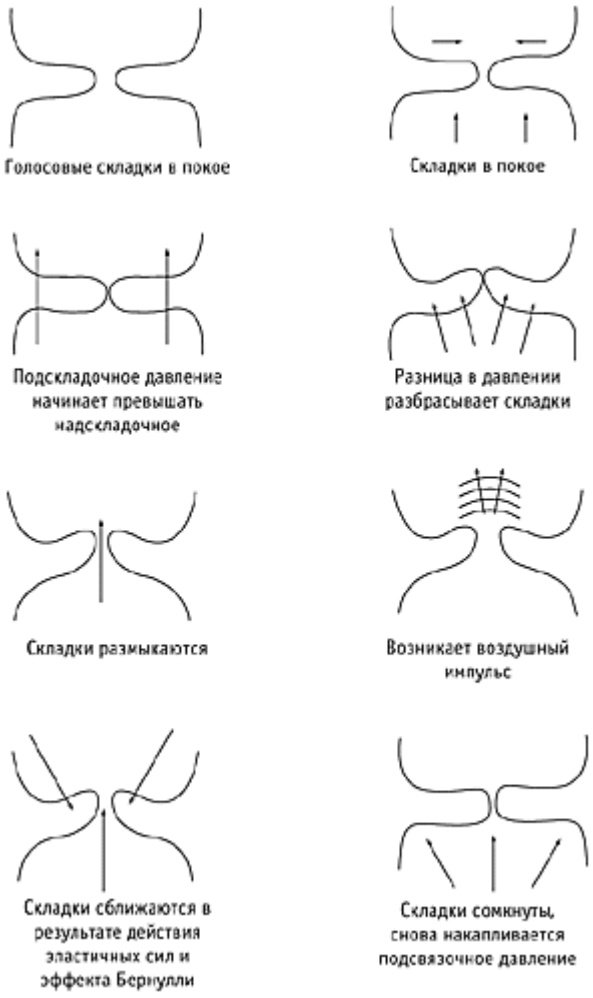

Рис. 6
Процесс колебаний голосовых складок

Частота колебаний складок определяет *высоту голоса* (у мужских голосов при речи она равна в среднем 110 Гц, у женских – 220 Гц), *амплитуда* определяет его *громкость*.

Если записать микрофоном такой звук у самых голосовых складок, то он напоминает гудение или жужжание. Это как бы исходный материал – чтобы получить из него звуки речи, его еще надо обработать в артикуляционном тракте. Спектр такого звука показан на рисунке 7. Поскольку колебания голосовых складок создают периодический сигнал (реальный сигнал не является строго периодическим), то спектр его при нормальной фонации является гармоническим с крутизной убывания 12 дБ/окт. Для увеличении громкости речи необходимо увеличить подскладочное давление (затратить больше энергии), при этом фронты голосовых импульсов становятся более крутыми (складки быстрее открываются). Время, когда щель закрыта, увеличивается от 40…50% при нормальной фонации, до 65…70% – спектр

соответственно изменяется, в нем появляется больше гармоник, что соответственно меняет тембр голоса (делает его ярче).

Способы смыкания складок при фонации могут быть разными. Например, если складки смыкаются не полностью, и между ними имеется щель, то форма импульсов становится почти симметричной, скорость не падает до нуля, в голосе слышен шум (придыхательный голос, шепот). Наоборот, если складки слишком сильно смыкаются (голос становиться зажатым), это также меняет форму импульсов и, соответственно, спектр и тембр голоса.

Все перечисленные характеристики – основная частота колебаний голосовых связок, форма голосовых импульсов, их амплитуда, спектральный состав и форма огибающей спектра – играют существенную роль при слуховом восприятии речи. Особую роль играет частота основного тона: в речевом потоке она определяет высоту голоса, и ее изменение используется также для изменения интонации, логических ударений, а иногда и смысла слов (например, в тональных языках, таких, как китайский). В вокальной речи (пении) частота основного тона может изменяться в широких пределах, обычно одна-две октавы (хотя были уникальные певцы с возможностью изменения высоты основного тона до четырех октав – Има Сумак, Мадо Робен и др.).

Частота основного тона, т.е. число колебаний голосовых связок в секунду, зависит от их длины, массы и натяжения. Приближенно эту связь можно представить, как для струны(хотя они больше похожи на резиновые шнуры) в виде: $f=0,5\sqrt{[T/LM]}$, где T – натяжение (упругость), L – длина, M – поверхностная масса. Таким образом, чем длиннее и тяжелее складки (эти свойства врожденные), тем более низкий тон имеет голос, чем короче и тоньше, – тем голос выше. Масса зависит от длины, толщины и плотности складок. В процессе речи и пения толщина и плотность складок может значительно меняться за счет натяжения.

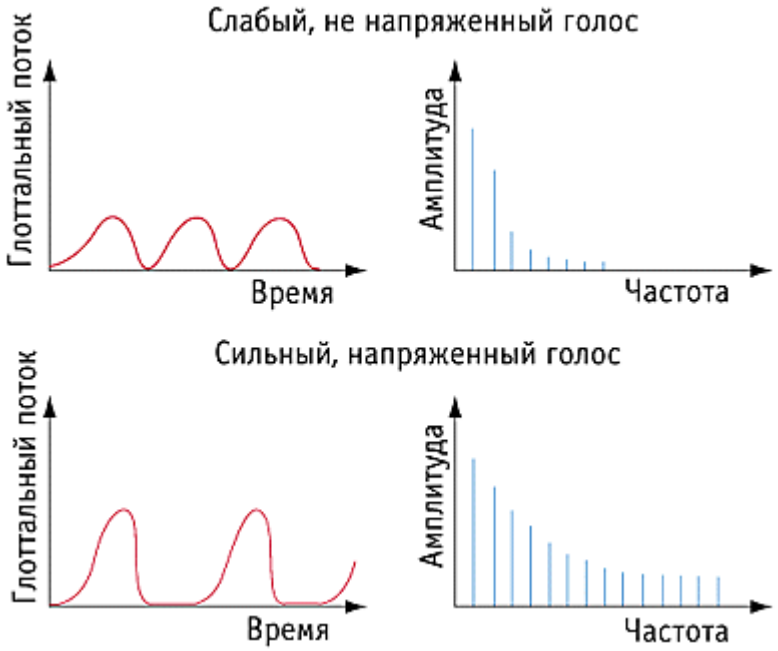

Рис. 7
Форма голосовых импульсов и их спектр

Натяжение обеспечивает повышение высоты голоса, и может осуществляться за счет напряжения внутренних вокальных мускулов (в основном при речи) и поворота щитовидного и перстеневидного хрящей относительно друг друга (в основном при пении). Поскольку при

увеличении громкости голоса растет подскладочное давление, а оно также оказывает некоторое влияние на натяжение складок (мускулы рефлекторно напрягаются), то обычно, при повышении громкости речи растет и высота тона (например, при крике). Только специально обученные певцы могут удерживать высоту тона при увеличении громкости в определенных пределах.

Таким образом, при образовании звуков речи с помощью процесса фонации (т.е. колебания голосовых связок) формируется звуковой сигнал, который затем трансформируется в вокальном тракте, где он превращается из "сырого" материала в последовательность речевых акустических сигналов (другие способы создания источников звука будут рассмотрены позднее).

Таким образом, вокальный тракт выполняет функцию резонатора, т.е. усиливает и фильтрует входной сигнал (аналогично трубам духовых инструментов). Форма труб вокального тракта показана на рисунке 8. Как видно из рисунка, тракт состоит из трех основных резонансных полостей: глотка, ротовая полость, носовая полость. Схематически его вид показан на рисунке 8. Отличия такой системы резонаторов от любых труб в музыкальных инструментах заключаются в следующем:

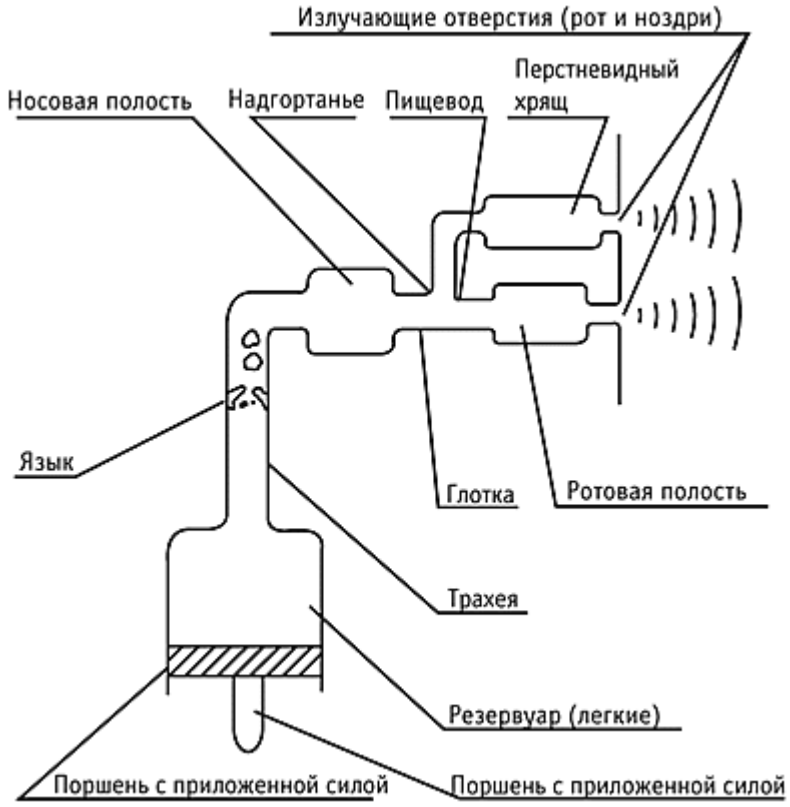

Рис. 8
Схематическая модель вокального тракта

- сложная геометрическая форма: вокальный тракт можно рассматривать как трубу переменного сечения с подключением параллельной трубы (носовой полости, которая может подключаться при опускании заднего мягкого язычка);

- возможность быстрой перестройки формы труб, площади их поперечного сечения, плотности и жесткости стенок, за счет изменения положения языка, мягкого язычка, губ, зубов,

расширения глотки, опускания гортани и др. Возможности перестройки параметров вокального тракта огромны, присущи только человеку, что и позволяет ему произносить все многообразие звуков речи. Этот процесс перестройки называется **артикуляцией**. Каждому звуку речи соответствует либо определенное статическое положение, либо определенная динамика изменения положения языка, челюстей, губ, нёбной занавески, т.е. определенная артикуляция.

Общая длина речевого тракта у взрослого человека (от голосовых складок до губ) около 17 см, длина носовой полости (от нёбной занавески до ноздрей) 12,5 см, площадь переменного сечения тракта в среднем составляет примерно 5…6 $см^2$.

Простейшей моделью вокального тракта можно считать цилиндрическую трубу длиной 17 см, закрытую на одном конце (аналогично трубе кларнета). Собственные моды (формы) колебаний такой трубы показаны на рисунке 9, частоты определяются из соотношений: $L = \lambda/4$; $L = 3\lambda/4$; $L = 5\lambda/4$ и т.д., таким образом частоты равны $f_n = (2n-1)c/4L$, где n-целое число; L-длина трубы; c-скорость звука.

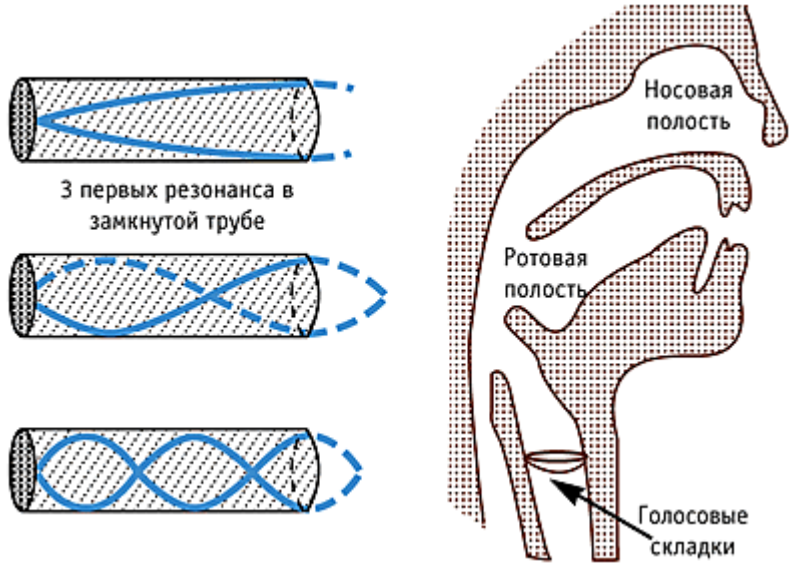

Рис. 9
Формы колебаний для цилиндрической трубы и голосового тракта

В спектре такой трубы присутствуют только нечетные гармоники 1:3:5… Для длины L = 17 см, собственные частоты оказываются равными 500, 1500, 2500 Гц. Если у трубы менять в разных точках площадь поперечного сечения, то положение ее собственных частот будет смещаться. Совершенно аналогичные процессы происходят в вокальном тракте: в нем также имеется свой набор собственных частот с соответствующими модами колебаний, т. е. определенным распределением узлов и пучностей вдоль его длины. Меняя площадь поперечного сечения в вокальном тракте, можно также все время менять положение собственных частот.

Если на вход такой трубы (системы труб) подать сигнал, сформированный при колебаниях голосовых связок (рисунок 7), то на выходе можно записать сигнал, который будет иметь форму, показанную на рисунке 10, т.е. гармоники, совпадающие с собственными частотами тракта, будут усилены за счет резонансов.

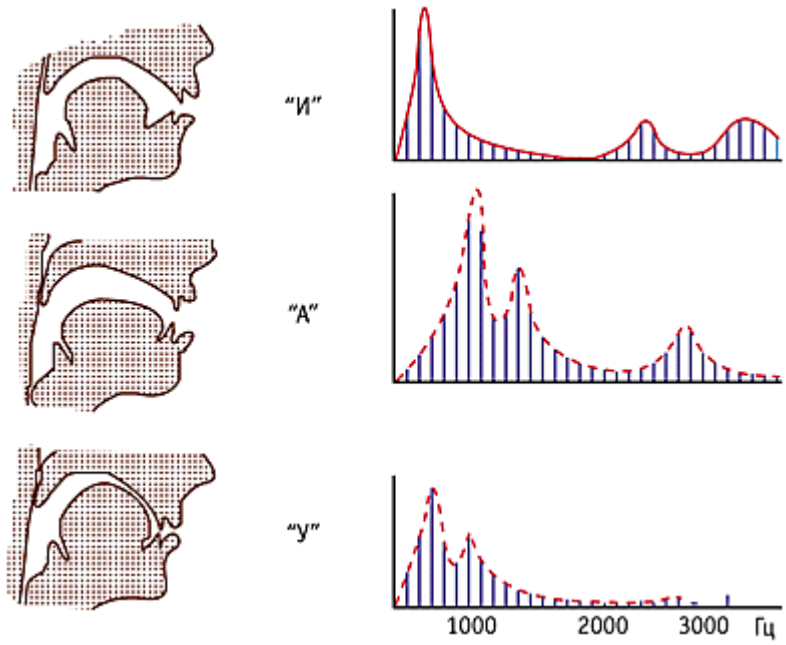

Рис. 10
Положение тракта для разных звуков речи
и вид звукового сигнала с формантами

Области спектральных максимумов, соответствующие резонансным частотам вокального тракта, называются **формантами** (иногда их просто называют резонансами вокального тракта). Каждому звуку речи (простейший звук речи называется фонемой) соответствует своя форма вокального тракта, которая варьируется за счет изменения положения языка, губ, зубов и т.д., и свое положение формант (F-картина). Примеры показаны на рисунке 10.

Существуют некоторые общие закономерности в управлении расположением собственных частот резонаторов: если поперечное сечение трубы уменьшается в области, где форма колебаний (мода), соответствующая данной резонансной частоте (форманте), имеет максимум давления, то частота увеличивается; если в точке, где минимум давления, то частота уменьшается. Изучение движения артикуляционных органов во время речи с помощью рентгенографических съемок показали, что аналогичные закономерности имеют место и в вокальном тракте: при подъеме языка вперед и вверх сужается передняя часть ротовой полости, при этом понижается первая форманта F1 и повышается вторая F2. При сдвиге языка назад сужается поперечное сечение тракта в области глотки, при этом повышается F1 и понижается F2 и т.д. При сдвиге формант по определенным закономерностям происходят изменения в соотношении их амплитуд, что приводит к изменению формы огибающей. Все эти признаки (расположение формант и соотношение их амплитуд) и являются отличительными акустическими признаками гласных звуков речи.

Правда, при беглой речи происходит настолько быстрая перестройка позиции артикуляционных органов (языка, губ и др.), что часто имеет место наложение позиции, соответствующей одному звуку, на позицию другого (обычно гласного на соседний согласный), такое явление называется *коартикуляцией*, и оно очень осложняет восприятие и распознавание речи.

Таким образом, вокальный тракт действует на звуковой сигнал источника как параметрический эквалайзер, при этом существенное значение имеют частоты резонансов, соотношения их амплитуд и ширина резонансных пиков (добротность). Примерные области

расположения первых трех формант для гласных русского языка даны в таблице.

	Частотный диапазон формант (Гц)		**Ширина формант (Гц)**
Тип голоса	Мужской	Женский	
F1	200-800	250-1000	40-70
F2	600-2800	700-3300	50-90
F3	1300-3400	1500-4000	60-180

Распознавание каждой фонемы происходит в основном по положению первых двух формант F1 и F2, более высокие форманты определяют тембральные различия (для пения чрезвычайно существенное значение имеет третья формантная область "певческая форманта"). Расположение формант для гласных английского языка показано на рисунке 11.

Если подходить к процессу образования звуков речи с помощью фонации в терминах передаточных функций, то этот процесс может быть описан следующим образом: $P(\omega)=S(\omega)T(\omega)R(\omega)$, где $S(\omega)$ – передаточная функция входного сигнала, $T(\omega)$ – передаточная функция тракта, $R(\omega)$ – активная составляющая сопротивления излучения, (рисунок 12). Именно эта последовательность операций и реализуется в различных синтезаторах звука. Под передаточной функцией тракта понимается отношение комплексных амплитуд объемной скорости на губах U_0 к объемной скорости у голосовой щели $U_г$: $T(\omega)=U_0/U_г$. Для цилиндрической трубы с одним закрытым концом она вычисляется по формуле: $T(\omega)=1/\cos(2\pi fLc)$. На резонансных частотах, определяемых по формуле $f_n = (2n-1)c/4L$, знаменатель обращается в нуль, и функция имеет максимумы (из-за наличия затухания она имеет конечные значения).

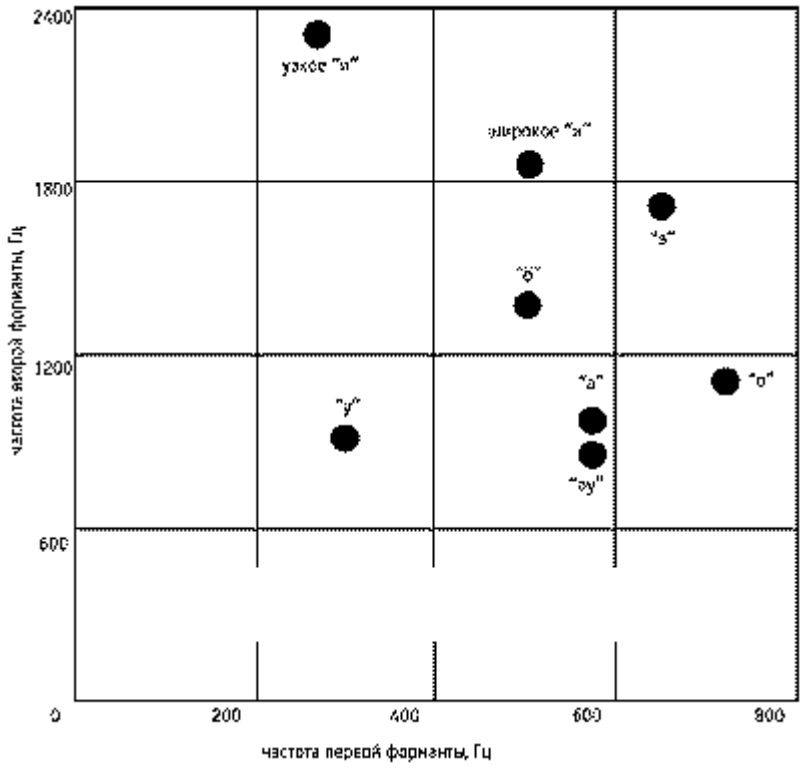

Рис. 11
Форманты для гласных

В реальном голосовом тракте передаточная функция имеет более сложный характер (она может быть вычислена и измерена современными цифровыми методами), но на резонансных частотах тракта, т.е. на формантах, она также имеет максимумы, которые называются полюсами. Таким образом, форманты еще можно определить как полюса передаточной функции.

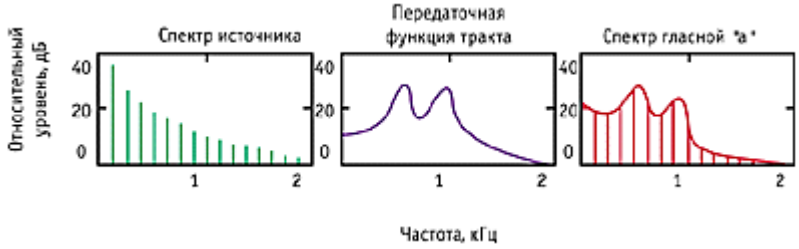

Рис. 12
Передаточная функция голосового тракта

Описанные выше процессы голосообразования относятся в основном к гласным звукам, процессы образования согласных звуков существенно сложнее, и будут рассмотрены в следующей части статьи.