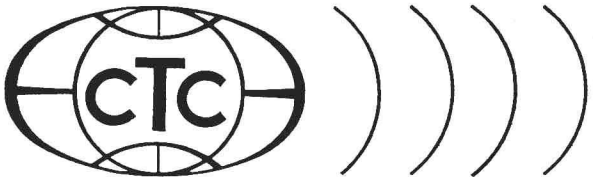
Ю. Н. ПРОХОРОВ

СТАТИСТИЧЕСКИЕ МОДЕЛИ И РЕКУРРЕНТНОЕ ПРЕДСКАЗАНИЕ РЕЧЕВЫХ СИГНАЛОВ



СТАТИСТИЧЕСКАЯ ТЕОРИЯ СВЯЗИ

Ю. Н. ПРОХОРОВ

СТАТИСТИЧЕСКИЕ МОДЕЛИ И РЕКУРРЕНТНОЕ ПРЕДСКАЗАНИЕ РЕЧЕВЫХ СИГНАЛОВ

Выпуск 20

МОСКВА «РАДИО И СВЯЗЬ» 1984

УДК 621.391

Прохоров Ю. Н. Статистические модели и рекуррентное предсказание речевых сигналов. — М.: Радио и связь, 1984. — 240 с., ил. — (Стат. теория связи. Вып. 20).

Синтезированы нелинейные модели авторегрессии речевого сигнала. С помощью теорий марковской фильтрации, регрессионного анализа и метода обновляющего процесса разработаны рекуррентные алгоритмы линейного и нелинейного адаптивного предсказания. Получены упрощенные алгоритмы предсказания, приспособленные к технической реализации на элементах цифровой микросхемотехники. Рассмотрены алгоритмы предсказания и фильтрации речевого сигнала, наблюдаемого в смеси с шумом. Алгоритмы предсказания применяются для цифрового представления сигнала в системах с адаптивной разностной импульсно-кодовой модуляцией и вокодерах. Приведены результаты машинного исследования, выполненного на реальном сигнале.

Для научных работников. Может быть полезна разработчикам цифровых систем передачи, обработки и распознавания речевых сигналов.

Табл. 29, ил. 94, библиогр. 201.

Рецензент Ю. С. ШИНАКОВ

Редакция литературы по радиотехнике

П $\frac{2402020000-026}{046(01)-84}$ — 110—83

ПРЕДИСЛОВИЕ

Речевой сигнал как объект научных исследований уже несколько десятилетий привлекает внимание специалистов разных областей. В основе этого интереса лежит практическая необходимость создания систем передачи информации с повышенной эффективностью и высокой помехоустойчивостью, систем автоматического управления машинами с помощью голоса, информационно-справочных систем с распознаванием речевых сигналов и речевым ответом, систем экономного хранения этих сигналов, систем установления личности говорящего, медицинских систем диагностики и т. д.

К настоящему времени разработаны и частично внедрены в практику системы высококачественной цифровой передачи речевых сигналов со скоростью 24—32 кбит/с. Созданы системы вокодерной передачи. Промышленностью выпускаются простейшие устройства распознавания и синтеза речевых сигналов. Разработаны первые информационно-справочные системы. Эти научные достижения изложены в книгах М. А. Сапожкова, коллектива авторов под общей редакцией А. А. Пирогова, а также в книгах Дж. Маркела и А. Грэя, Л. Рабинера и Р. Шаффера и в др.

Достигнутые успехи выдвинули на первый план задачу устранения ограничений существующих систем обработки и высокоэффективной передачи, к которым прежде всего относят невысокое качество звучания речевых сигналов, восстановленных в низкоскоростных системах передачи, и снижение эффективности обработки и передачи при наличии сопутствующих шумов.

Трудность решения этой задачи традиционными средствами стимулирует развитие методов представления, обработки и передачи речевых сигналов с использованием современной теории случайных процессов и, в частности, хорошо развитых теорий условно-марковских, условно-гауссовских процессов и регрессионного анализа.

Синтез моделей, допускающих применение этих теорий, и разработка рекуррентных алгоритмов оценивания параметров, предсказания и фильтрации, учитывающих структуру речевого сигнала и ориентированных на техническую реализацию с помощью цифровой микросхемотехники, составляют предмет настоящей книги. Именно из соображений простоты технической реализации желательно, чтобы формулы, определяющие алгоритмы обработки сигнала, имели рекуррентный характер. Вместе с тем упрощение алгоритмов за счет рекуррентных соотношений позволяет ре-

шать более сложные, чем ранее, задачи, такие например, как совместное оценивание функции входного возбуждения и параметров модели голосового тракта.

Книга требует предварительного знакомства с основами теории вероятностей и случайных процессов, например, в объеме книг Б. Р. Левина [41], В. И. Тихонова, Н. К. Кульмана [49], Э. Сейджа, Дж. Мелса [101], а также с характеристиками, способами описания, цифровой обработки и передачи речевых сигналов в объеме книги Л. Рабинера, Р. Шаффера [5].

Автор благодарит аспирантов МЭИС Ю. Ю. Гурьева, Ю. И. Журавского, В. И. Ковязина, С. Ф. Лихачева, совместно с которыми получен ряд результатов по исследованию нелинейных моделей и рекуррентных алгоритмов предсказания речевых сигналов.

Особую признательность автор выражает проф. М. В. Назарову за ценные замечания и советы.

1. ПОСТАНОВКА ЗАДАЧИ

1.1. Краткие сведения об образовании и слуховом восприятии речи

Акустическое речевое колебание порождается в результате движения органов артикуляционного аппарата и воспринимается органами слуха человека. Изучение функционирования артикуляционного аппарата в процессе произнесения звуков речи является предметом глубоких физиологических и акустических исследований, результаты которых представлены в [1—3, 5—8, 11, 13]. Не менее пристально изучается слуховое восприятие речевых и других акустических колебаний, а также их обработка на периферии органов слуха и на высших уровнях восприятия, связанных с комбинационной деятельностью мозга [2, 4, 9, 10, 12, 14]. Не останавливаясь подробно на изложении всех «выстроенных» к настоящему времени теоретических положений и выявленных экспериментально фактах, приведем, следуя указанной литературе, краткое описание основных сведений об образовании и восприятии речи человеком, необходимых для правильного толкования последующего материала.

В образовании звуков речи участвуют следующие физиологические органы: рот, нос, язык, небная занавеска, глотка, гортань, голосовые связки, трахея, бронхи, легкие и диафрагма. Выталкиваемый из легких воздух проходит через трахею, гортань, полости глотки, рта и носа. Таким образом, речевое колебание представляет собой акустическую волну, распространяющуюся по речеобразующей системе и в конечном итоге излучаемую через губы и ноздри.

Одну из главных ролей в образовании звуков речи играют голосовые связки, расположенные в гортани. При обычном дыхании голосовые связки разомкнуты и голосовая щель — проход между связками — широко раскрыта. При произнесении некоторых звуков речи связки находятся в сомкнутом исходном состоянии и размыкаются под воздействием давления нагнетаемого из легких воздуха, который, прорываясь через голосовую щель, раздвигает связки в поперечном направлении. Под воздействием суживающих щель мышц, а также благодаря упругости и из-за гидродинамического эффекта Бернулли связки вновь возвращаются в сомкнутое состояние, и далее цикл повторяется. Такие движения голосовых связок характерны произнесению гласных и звонких согласных звуков. В результате этих колебаний проходящий через голосовую щель поток воздуха приобретает им-

пульсный характер и затем поступает в глотку, ротовую и носовую полости. Гортань и ротовую полость обычно называют *голосовым трактом.* Конфигурация голосового тракта в процессе произнесения звуков изменяется во времени. Эти изменения накладываются на проходящий через тракт поток воздуха. Для образования носовых звуков к голосовому тракту через нёбную занавеску подключается носовая полость.

Изменение конфигурации голосового тракта и колебания голосовых связок взаимосвязаны так что вся речеобразующая система функционирует как единый сложный объект, а не как набор автономно функционирующих органов, определенным образом соединенных друг с другом. Одна группа органов — зубы, твердое нёбо, задняя стенка глотки и носовой полости — участвует в артикуляции пассивно, так как остается неподвижной, в то время как другая группа — нижняя челюсть, губы, язык, мягкое нёбо, нёбная занавеска, голосовые связки — является активной, так как осуществляет при артикуляции вполне упорядоченные движения.

Изменение конфигурации тракта вдоль его продольной оси и во времени описывают функцией площади поперечного сечения. В акустике голосовой тракт рассматривают как систему резонаторов, характеристики которых медленно изменяются во времени. Частоты и области резонансов называют соответственно *формантными частотами* и областями. Часто для краткости употребляется термин *форманта.*

При прохождении через речеобразующую систему в воздушном потоке могут возникать турбулентности. Так происходит, например, при обтекании потоком препятствий различной формы, которые создаются активной группой органов при артикуляции. Турбулентности могут возникать и в свободно распространяющемся потоке, если скорость его частиц превышает некоторую критическую скорость, определяемую *числом Рейнольдса,* средней скоростью частиц, свойствами воздуха как газа и геометрическими размерами объекта. Турбулентность порождает случайные (шумовые) составляющие акустического колебания, роль и описание которых слабо представлены в современной акустической теории речеобразования. Шум, возникающий в свободно распространяющемся потоке, иногда называют *вихревым* (он вызывает шипящие звуки), а формирующийся при обтекании потоком препятствий — *краевым тоном* (вызывает «свистящие» звуки). Следует подчеркнуть, что турбулентности могут создавать и колеблющиеся голосовые связки — вопрос, который представляется изученным недостаточно.

Помимо различий в характере турбулентности, формируемой разными факторами, случайные составляющие акустического колебания отличаются и местом образования в речеобразующей системе, которое для разных звуков может быть различным, и, таким образом, место расположения источника шума также активно управляется при артикуляции.

Звуки, при формировании которых голосовые связки осуществляют колебательные движения, называют *вокализованными.* Грубо говоря, все остальные звуки можно отнести к *невокализованным.* Более точно, среди последних по способу образования различают *фрикативные звуки,* возникающие при возбуждении голосового тракта турбулентным широкополосным шумом, *взрывные звуки,* формируемые путем создания в тракте смычки с последующим внезапным высвобождением сжатого за смычкой воздуха, *комбинированные звуки,* образующиеся при сочетании некоторых гласных и согласных звуков и др. [1, 2, 5]. Разделение звуков по способу их образования на вокализованные и невокализованные с акустической точки зрения не имеет четко выраженной границы, так как колебание голосовых связок и образование турбулентных шумов могут сопутствовать друг другу.

При математическом описании речеобразующей системы вводят ряд упрощающих положений, допускающих составление и решение дифференциальных уравнений в частных производных, которым удовлетворяют звуковое давление и объемная скорость воздушного потока, распространяющегося в гипотетической акустической системе. Упрощающими положениями являются предположение о независимой работе голосовых связок и голосового тракта, представление голосового тракта в форме набора секций цилиндрических труб, описание турбулентных составляющих эквивалентным шумом, действующим на входе голосового тракта. В рамках этих положений голосовые связки рассматривают как автономный источник, генерирующий квазипериодическую последовательность импульсов воздушного потока приблизительно треугольной формы. Голосовые связки совершают автоколебательные движения и могут рассматриваться как упругая механическая система, находящаяся под воздействием избыточного давления, обусловленного эффектом Бернулли, в фазе открытой щели. Возможна и другая трактовка, когда значительная роль отводится центральной нервной системе, управляющей ритмическими колебаниями связок, но последние исследования отдают предпочтение автоколебательной теории.

Используя принципы динамических аналогий, механическую систему можно рассматривать как электрическую цепь с сосредоточенными параметрами. Эта цепь описывается нелинейным дифференциальным уравнением первого порядка с переменными коэффициентами, решить которое аналитически трудно. Для понимания последующего материала важно подчеркнуть, что речеобразующая система имеет, по крайней мере, один участок, отображаемый электрическим нелинейным аналогом, в котором порождается квазипериодическое колебание.

Основными характеристиками электрического аналога голосового источника является форма генерируемых импульсов и период (частота) их следования, который обычно называют *периодом* (частотой) *основного тона.*

Голосовой тракт, рассматриваемый как набор секций цилин-

дрических труб, характеризуется функцией площади поперечного сечения, полными сопротивления со сторон источника возбуждения и нагрузки. Подобное описание позволяет перейти к эквивалентной линейной динамической системе с сосредоточенными параметрами в дискретном (или непрерывном) времени, которая может быть представлена передаточной (частотной) характеристикой, или в виде разностного (дифференциального) уравнения, например, линейного уравнения авторегрессии, и, таким образом, изложенные построения могут служить аргументом использования для обработки сигнала аппарата теории динамических систем, временных рядов [15—18] или даже общей теории случайных процессов [19—21].

Речевое колебание воздействует на органы слуха человека, вызывая определенные слуховые ощущения. Изложим кратко общие сведения об устройстве периферийных органов слуха, в которых акустическое речевое колебание превращается в нервное возбуждение.

Первичный акустический преобразователь, используемый человеком при слуховом восприятии, разделяют на три области: наружное, среднее и внутреннее ухо. Наружное ухо состоит из ушной раковины и слухового прохода, который заканчивается тонкой мембраной, называемой барабанной перепонкой. Слуховой проход можно представить как однородную трубу, открытую с одной стороны. Среднее ухо содержит слуховые косточки (молоточек, наковальню и стремечко), обеспечивающие преобразование колебаний барабанной перепонки в смещение объема жидкости во внутреннем ухе. Расположено среднее ухо в полости, заполненной воздухом, сразу за барабанной перепонкой. Преобразование акустических колебаний в среднем ухе описывают передаточной функцией эквивалентной электрической линейной системы, имеющей частотные характеристики фильтра нижних частот. Внутреннее ухо состоит из улитки — органа по форме напоминающего раковину улитки, вестибулярного аппарата и окончаний слухового нерва. В улитке механические колебания преобразуются в нервные возбуждения. Полость улитки заполнена несжимаемой жидкостью. Площадь поперечного сечения улитки со стороны среднего уха (стремечка) составляет примерно 4 мм^2 и уменьшается до 1 мм^2 у противоположного тонкого конца (геликотремы). Внутри улитка разделена продольной перегородкой. Полость с одной стороны перегородки сообщается через овальное окно с косточками среднего уха, которые при движении вызывают смещение жидкости, воздействуя на нее как поршень. Роль поршня выполняет стремечко. Полость с другой стороны перегородки вблизи среднего уха имеет круглое окно, покрытое упругой мембраной. При смещении жидкости ее избыток выходит через круглое окно, так как обе полости сообщаются друг с другом через проход у тонкого конца улитки (геликотремы).

Внутри перегородки имеется канал, называемый *улитковым*

ходом. Одна продольная мягкая стенка улиткового хода называется *мембраной Рейснера,* другая студенистая стенка — *базилярной мембраной,* которая по мере приближения к геликотреме постепенно суживается. На базилярной мембране внутри улиткового хода расположен орган Корти, который содержит чувствительные клетки, связанные с окончанием слухового нерва. Мембрана со стороны среднего уха (у основания) уже, жестче и тоньше, чем у геликотремы и служит дисперсионной средой распространения колебаний. Смещение жидкости, вызываемое движением стремечка, приводит к распространению вдоль базилярной мембраны бегущей волны. Распределенные параметры всей системы таковы, что отражений у геликотремы не происходит и поэтому на мембране не создается стоячих волн. Высокочастотные составляющие по мере приближения волны к геликотреме вследствие дисперсионных свойств мембраны постепенно ослабляются. Механические колебания тела мембраны приводят в движение волоски системы специальных чувствительных клеток органа Корти, при изгибе которых возникают электрические потенциалы в улиточной части нерва. Такова упрощенная схема преобразования механических колебаний базилярной мембраны в электрические потенциалы внутри клеток слухового нерва.

Механизм преобразования смещений мембраны в нервные возбуждения изучен не полно, как, впрочем, и способ кодирования этих смещений электрическими потенциалами в клетках нерва. Тем не менее считается установленным, что локальные смещения мембраны вызывают возбуждение локализованных в этой области нейронов, которые сохраняют упорядоченность и в слуховом нерве. Это обстоятельство может служить поводом для гипотезы о способе кодирования высоты тона — по месту максимального отклонения мембраны или по синхронному с воздействием возбуждению определенной группы нейронов. Гипотеза наводит на мысль о возможности представления базилярной мембраны как частотно-избирательной системы, в которой осуществляется спектральный анализ входного акустического колебания и которую, следовательно, можно описать частотными характеристиками. Под последними понимают зависимости амплитуды и фазы смещения мембраны в некоторой точке от частоты гармонического колебания основания стремечка. Измеренные в соответствии с таким определением частотные характеристики оказались близкими к характеристикам колебательных контуров с примерно одинаковой добротностью. Таким образом, разрешающая способность по частоте оказывается наилучшей у геликотремы, а по времени — у основания улитки. Максимумы частотных характеристик увеличиваются с увеличением частоты примерно до 1000 Гц со скоростью около 5 дБ на октаву. На более высоких частотах максимумы примерно одинаковы. Линейным приращениям координаты вдоль тела мембраны соответствуют приращения частоты резонанса примерно по логарифмическому закону в диапазоне до 1000 Гц.

Математическое описание преобразований акустического колебания во внутреннем ухе основано на предположении, что механическая система в требуемом диапазоне частот пассивна, линейна и инвариантна по отношению к сдвигу времени. В рамках этих ограничений преобразование колебания, распространяющегося от стремечка (вход) к данной точке тела мембраны (выход) описывается дробно-рациональной передаточной функцией и, таким образом, мембрана эквивалентна набору полосовых фильтров, перекрывающих весь диапазон частот, различаемых на слух.

Исследования показали сравнительно высокую чувствительность волокон слухового нерва к частоте стимула, более высокую, чем можно было ожидать исходя из свойств внутреннего уха [22]. Измеренные частотные характеристики волокон [10, 22] оказываются гораздо более избирательными, чем характеристики механической системы уха, что послужило основой для разработки учитывающих этот эффект «обострения» характеристик моделей [4]. Анализ результатов исследований свойств различных групп нейронов позволяет сделать вывод о том, что в нейронной сети слуховой системы осуществляется довольно тонкий анализ, позволяющий отразить в нервных возбуждениях детальное изменение акустического колебания во времени (и частоте) [22]. Способ осуществления этого анализа представляется изученным недостаточно полно, чтобы можно было предложить адекватные математические модели слухового восприятия на акустическом уровне.

Слуховое восприятие обладает двумя важными свойствами, которые сравнительно широко используются в технике: слабая чувствительность слуха к фазовым соотношениям спектральных составляющих сложного акустического колебания, такого как речь, и эффект маскировки одного воздействия другим.

Чувствительности слуха к фазовым соотношениям посвящено много исследований [23—27]. Эти исследования не окончены и проводятся в настоящее время [28, 29]. Тем не менее в инженерной практике принято считать, что некоторая потеря кратковременных фазовых соотношений допустима [28, 30, 31] и, более того, форма акустического речевого колебания может быть изменена, и даже иногда существенно, без заметных на слух потерь в разборчивости [28]. В то же время на качество восприятия, синтезированного в вокодерных системах сигнала, как отмечается, например, в [27, 31], влияет потеря естественных фазовых соотношений гармоник основного тона. Поэтому нечеткий термин «слабая чувствительность», по-видимому, правильно отражает существующее положение.

Эффекты маскировки состоят в изменении порога слышимости одного акустического воздействия при мешающем влиянии другого воздействия. Для каждого типа воспринимаемого и мешающего воздействий могут быть измерены экспериментально свои кривые порога слышимости при маскировке. Различают маскировку гармонического тона шумом, тона тоном и т. п. [9].

Важной характеристикой слухового восприятия является спо-

собность человека отличать на слух тон от колебания со сплошным спектром, сосредоточенным вокруг частоты тона. Диапазон частот спектра речевого сигнала, который без ущерба для слухового восприятия может быть заменен одним эквивалентным тоном, называется критической полоской речи [1]. Это понятие и близкие к нему [1, 9] используются в методах аналитического расчета формантной разборчивости [1, 2, 32].

При детальном анализе свойств слухового восприятия тональных (гармонических) и шумовых воздействий, в частности при маскировке, установлена интересная характерная для многих экспериментов особенность: «... закономерность при восприятии слухом белого шума оказывается намного более простой, чем при восприятии синусоидального тона» [9, с. 109]. Эта особенность может быть интерпретирована так: шумовое широкополосное воздействие является более простым объектом для восприятия, чем тональный сигнал. Сопоставление этого утверждения с рассуждениями, выполненными в [33, с. 55, 56] на основе принципа равной простоты Н. А. Бернштейна, наводит на мысль о том, что шумовое воздействие допустимо рассматривать как элементарное (эталонное), с которым можно сравнивать другие воздействия при их математическом описании. По существу, именно этот принцип лежит в основе многих математических описаний речевых сигналов в рамках теории случайных процессов.

1.2. Статистические модели речевых сигналов и их синтез

Речевым сигналом называется электрическое колебание, наблюдаемое на выходе электрического преобразователя при воздействии на его вход акустической речевой волны. В преобразователь входят микрофон, устройства регулировки уровня и фильтры, формирующие спектр выходного колебания в заданной полосе частот.

Так как смысловое содержание речевой волны априори неизвестно и, кроме того, одному смысловому содержанию в разных экспериментах могут соответствовать различные электрические колебания, отличающиеся расположением во времени, формой, длительностью и т. п., то речевой сигнал можно рассматривать как случайный процесс, реализации которого наблюдаются на выходе преобразователя.

Естественно речевой сигнал представлять как случайную функцию непрерывного времени. Вместе с тем, известно, что его можно подвергнуть операции временной дискретизации с интервалом Δt по теореме В. А. Котельникова и восстановить в первоначальную форму без потерь для слухового восприятия [5, 6, 66]. Поэтому можно представлять речевой сигнал как функцию дискретного времени и рассматривать последовательность отсчетных значений, обозначаемую везде далее символом x_t, $t=..., -1, 0, 1, ...$ В этом случае в преобразователь можно включить и устройство временной дискретизации. Подобное дискретно-аналоговое представление ориентировано на использование для по-

следующей обработки цифровых или дискретно-аналоговых устройств, в которых t выполняет роль укрупненного автоматного времени.

Случайный процесс, как семейство случайных величин (СВ), определяется вероятностными характеристиками — функциями плотности вероятностей, интегральными функциями распределения, моментными функциями или в наиболее общей форме вероятностными мерами. Эти характеристики, если они известны априори, составляют математическую вероятностную модель сигнала.

Априорные сведения о случайном процессе могут быть заданы и другими способами. Широко распространены два из них. Первый состоит в описании каждой реализации частичной суммой обобщенного ряда Фурье. Так получают спектральные модели речевых сигналов [2, 5, 31] и произвольных случайных процессов [193], которые задаются спектральными и корреляционными характеристиками. Второй способ интенсивно развивается в последние десятилетия и основан на представлении случайного процесса в рамках теории динамических систем [49, 70, 101]. Априорные сведения здесь задаются параметрическим описанием динамической системы, на вход которой воздействует белый шум или некоррелированная последовательность СВ, а на выходе формируется процесс с требуемыми характеристиками. Вероятностная модель определяется параметрами и структурой динамической системы.

На практике априорные сведения о случайном процессе часто недостаточно полные, чтобы можно было явно задать его вероятностную модель. В этом случае процесс описывается статистическими характеристиками, которые находятся по опытным данным с помощью оценивания функций плотности вероятностей, интегральных функций распределения вероятностей, спектров, параметров или характеристик динамических систем. Статистические характеристики дают полную информацию о процессе при заданном объеме данных и при неограниченном увеличении объема должны сходиться в том или ином смысле к вероятностным. Это свойство состоятельности позволяет, несколько поступаясь точностью представления процесса, использовать полученные оценки в допредельном режиме, когда объем данных конечен.

Статистические характеристики задают статистические модели сигнала.

Оценки функций плотности вероятностей, спектральной плотности мощности, корреляционной функции речевого сигнала и их приближенное математическое описание приведены в ряде работ [1, 2, 5, 66, 119]. Они получены усреднением за сравнительно большой промежуток времени в предположении стационарности и эргодичности речевого сигнала и вследствие этого не дают полного представления о его мгновенных свойствах. Тем не менее их успешно используют для приближенных расчетов при проектировании систем передачи и обработки [5, 66, 119].

Оценки корреляционных функций и спектральных плотностей мощности, полученные за короткий промежуток времени (примерно 10—30 мс), приведены в [1, 5, 31]. Они в бóльшей мере отражают мгновенные свойства речевого сигнала. Например, оценки корреляционной функции, полученные на участке звучания гласного звука усреднением за промежуток времени примерно 10 мс, имеют максимумы на задержках, кратных периоду основного тона, в то время как оценки, полученные усреднением за большой промежуток времени, сведений об основном тоне не содержат.

Статистические характеристики речевого сигнала как нестационарного случайного процесса, пока глубоко не изучались. Это обусловлено, по-видимому, трудностями проведения обширных экспериментальных исследований.

Для описания нестационарного речевого сигнала обычно применяются статистические спектральные модели и модели в виде динамических систем, параметры которых оцениваются на каждом из примыкающих друг к другу коротких промежутков времени длительностью около 10—30 мс.

Спектральные модели являются традиционным средством описания речевых сигналов, применение которых восходит к таким ранним работам как [194]. Результаты плодотворных исследований в направлении использования методов спектрального анализа при создании систем передачи, автоматического распознавания и синтеза речевых сигналов изложены в [1, 2, 5, 22, 31].

Стимулом к развитию представления речевых сигналов в динамических системах являются, по крайней мере, три обстоятельства.

Во-первых, в отличие от общей ситуации, имеющей место в теории случайных процессов, такие модели оказываются не только удобной формой представления априорных сведений о речевом сигнале, но и хорошей аппроксимацией объекта, порождающего сигнал,— артикуляционного аппарата.

Во-вторых, появляется возможность использования развитого математического аппарата теории марковской фильтрации [48, 49, 101] и примыкающих к нему методов оценивания [42, 129] для решения более широкого, чем ранее, класса задач. Например, выбирая в качестве динамической системы в непрерывном времени последовательно соединенные *RC*-фильтры нижних и верхних частот получают описание речевого сигнала двухкомпонентным марковским процессом, что открывает возможность разработки структуры оптимальных приемников радиотехнических систем передачи [49, 155]. Важной особенностью такого представления является возможность описания сигнала нестационарными процессами, что сложнее выполнить при спектральном подходе.

В-третьих, алгоритмы оценивания, полученные в рамках таких представлений, просты и удобны для реализации в цифровых устройствах, микропроцессорах и микро-ЭВМ, технология изготовления которых бурно развивается в настоящее время.

Модели речевых сигналов в виде динамических систем известны сравнительно давно [1, 3, 34], но наиболее глубокое развитие они получили в последние годы в связи с применением математических методов линейного предсказания [5, 6]. Так как последующий материал основан на таком способе представления, изложим его более подробно.

Рассмотрим вначале формирование невокализованных звуков, когда связки находятся в свободном состоянии. При рапространении воздушной волны, нагнетаемой из легких, в голосовом тракте возникают турбулентные шумы, которые после преобразования из-за артикуляционных движений активной группы органов (состояние связок в этих рассуждениях не учитывается) порождают звуки речи, имеющие шумовой характер. Следуя предположению о линейности преобразований в тракте, эквивалентную динамическую систему в дискретном времени, отражающую процесс речеобразования, можно изобразить в виде схемы рис. 1.1, где ξ_t — последовательность СВ, моделирующая исходный турбулентный шум, *ПС — порождающая* линейная *система*

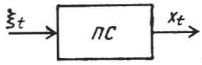


Рис. 1.1. Модель образования невокализованных звуков речи

или эквивалентная модель голосового тракта, в которой производится своеобразная «модуляция» ξ_t посредством медленного изменения ее параметров. Вполне допустимо принять гипотезу о некоррелированности последовательности ξ_t. Это приводит к описанию речевого сигнала в рамках теории случайных процессов, впервые принятому в [34, 35], что вполне согласуется с общей теорией информации, трактующей реальные сообщения как случайные функции, и кроме того, суживает всевозможные стохастические представления сигнала до множества случайных процессов, которые можно сформировать пропусканием последовательности некоррелированных величин через линейный формирующий фильтр.

Упростим представление, для чего зафиксируем параметры линейного фильтра и вероятностные характеристики ξ_t, полагая, что выбранные значения параметров и характеристик соответствуют какому-либо протяжному звуку. Легко видеть, что в результате получилось распространенное представление случайных процессов, известное в литературе как *метод обновляющего процесса*, впервые использованное в [36, 37] для предсказания и широко развитое в [38—43]. В самом простом случае [37, 40] под обновляющим процессом можно понимать последовательность ξ_t, названную в [40, 41] *порождающей*. Сохраним это название для ξ_t, если речь идет об образовании сигнала в *ПС*. Если же рассматривается формирование входного сигнала *ПС* по x_t, то будем использовать термин обновляющий процесс и обозначение $\nu_t^{(x)}$. Символом $\nu_i^{(x)}$ будет обозначаться также мгновенная ошибка предсказания.

Рассмотрим теперь слуховое восприятие x_t. Если в ПС нет никаких преобразований, т. е. $x_t \equiv \xi_t$, то на органы слуха воздействует турбулентное акустическое колебание, вызывающее смещение базилярной мембраны на всем протяжении от стремечка до геликотремы. Приближенная модель периферической слуховой

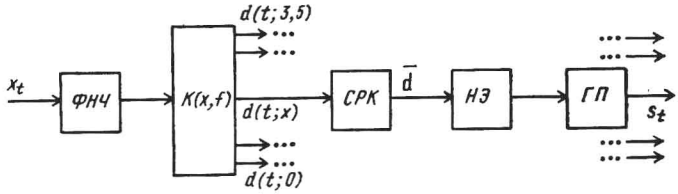


Рис. 1.2. Модель периферической слуховой системы

системы может быть представлена схемой рис. 1.2 [10], где *ФНЧ* — фильтр нижних частот, отражающий преобразование в среднем ухе; $K(x, f)$ — передаточная функция модели базилярной мембраны, зависящая от пространственной координаты x и частоты f; $d(t, x)$ — отклик модели базилярной мембраны; *СРК* — вычислитель среднеквадратического значения сигнала $d(t, x)$ с усреднением за короткий промежуток времени примерно 10 мс; *НЭ* — нелинейный безынерционный элемент; *ГП* — генератор пуассоновского потока нервных импульсов, интенсивность которого пропорциональна значению $\bar{d}$. Для шумовых стимулов, спектр которых сосредоточен в диапазоне средних и высоких частот, математическое ожидание величины $\bar{d}(x)$ будет велико по сравнению с флуктуациями около математического ожидания, так как ширина полосы пропускания модели базилярной мембраны на этих частотах значительно превышает величину, обратную времени усреднения в блоке СРК [10]. Следовательно, интенсивность пуассоновского потока s_t будет определяться в основном среднеквадратическим значением $\bar{d}$, полученным за продолжительный период времени и, так как по предположению $\xi_t \equiv x_t$ — стационарная последовательность, то интенсивность s_t в установившемся режиме будет сравнительно большой для всех x (кроме, возможно, $x \approx 3{,}5$ мм, что соответствует низким частотам). Таким образом, на вход высших уровней восприятия поступает информация о сплошном распределении энергии вдоль координаты x, что на слух воспринимается как ровный шум, не содержащий никаких смысловых сведений. Если в *ПС* осуществляется некоторое линейное преобразование, то выходной сигнал x_t будет отличаться от ξ_t распределением энергии по частотам и возможно одномерным распределением, которое, если принять модель рис. 1.2, слабо влияет на восприятие. Настраивая по-разному *ПС*, можно перераспределять энергию вдоль координаты x, изменяя пространственное распределение интенсивностей пуассоновских потоков и вызывая слуховые ощущения шумовых (в том числе и речеподобных) протяжных звуков. Если перестраивать

$ПС$ упорядоченно, имитируя артикуляционные движения, то получим шепотную речь, в предположении, что движения связок не имеют существенного значения. Можно сделать вывод, что для слухового восприятия важен именно характер изменения параметров $ПС$ во времени, в то время как ξ_t выполняет роль своеобразного переносчика.

Обозначим символом $П$ оператор, устанавливающий соответствие между элементами множеств функций времени на входе $ПС$ и на ее выходе так, что запись $x_t = П\xi_t$ означает преобразование входной последовательности ξ_t в речевой сигнал. В данном случае $П$ — линейный оператор и служит математической моделью $ПС$.

Тогда в схеме на рис. 1.1 основную роль играет состояние $ПС$, описываемое оператором $П$, а относительно последовательности ξ_t необходимо знать лишь функцию корреляции и, возможно, одномерную интегральную функцию распределения.

Также обстоит дело и при представлении системой рис. 1.1 случайных процессов на основе метода обновляющего процесса [40]. Следуя [37], обратим внимание на то, что фазовые соотношения по крайней мере при фиксированных параметрах $ПС$ не влияют на восприятие в рамках модели рис. 1.2 и, таким образом, фазочастотную характеристику $ПС$ можно выбрать «любым способом, согласующимся с условием физической осуществимости» [37, с. 697]. Это условие для кратковременной фазочастотной характеристики можно сохранить и при медленно изменяющихся параметрах.

Обозначим символом $П^{-1}$ оператор, устанавливающий обратное соответствие между элементами множеств функций времени на выходе и на входе $ПС$ так, что запись $\xi_t = П^{-1}x_t$ означает обратное преобразование речевого сигнала в ξ_t. Тогда, пользуясь свободой выбора фазочастотной характеристики, будем считать $ПС$ минимально-фазовой, для которой существует система, осуществляющая обратное преобразование, и которая описывается оператором $П^{-1}$. Модель, правильно отражающая свойства сигнала, обеспечивает возможность предсказания последующих величин x_t по наблюдаемым значениям предыдущих x_s, $s<t$. Оказывается, оператор $П$ однозначно определяет оптимальный в смысле наименьших квадратов предсказывающий фильтр, если последний находить по методу обновляющего процесса [37]. Метод состоит в следующем. Пусть требуется получить оценку (предсказания) $\hat{x}_t$ будущего значения $x_{t+\alpha}$, $\alpha>0$ по наблюдению фрагмента реализации x_s, $s\leqslant t$, если стационарный процесс x_t задан вероятностной моделью рис. 1.1 и ξ_t имеет гауссовское распределение вероятностей. Обозначим через $g(t)$ импульсную характеристику $ПС$ и построим фильтр с импульсной характеристикой

$$g_1(t) = \begin{cases} g(t+\alpha), & t \geqslant 0, \\ 0, & t<0. \end{cases}$$