

НЕПРЕРЫВНАЯ АУТЕНТИФИКАЦИЯ ДИКТОРА ПРИ ВЕДЕНИИ ТЕЛЕФОННЫХ ПЕРЕГОВОРОВ ПО НИЗКОСКОРОСТНЫМ ЦИФРОВЫМ КАНАЛАМ

Афанасьев А.А.¹

В статье рассмотрена задача по разработке системы непрерывной аутентификации диктора. Предложен метод декомпозиции речевого сигнала на взаимосвязанные элементы речи при ее низкоскоростном кодировании на основе линейного предсказания. Использование данного метода при предварительном обучении системы позволяет осуществлять непрерывную аутентификацию абонента при доступе к ресурсу пропускной способности канала связи.

Ключевые слова: речевой сигнал, кодирование речи, линейное предсказание, синтезирующий фильтр, декомпозиция, взаимозависимость.

Введение

При ведении телефонных переговоров с использованием средств связи часто необходима предварительная аутентификация абонента с целью определения его легитимности для дальнейшего доступа к ресурсу пропускной способности инфокоммуникационной сети связи ограниченного доступа. Для выполнения данной процедуры в существующих системах в настоящее время может использоваться однократное подтверждение легитимности пользователя при вхождении в связь, после чего ему предоставляется полный доступ к услугам сети связи. В предлагаемой системе при ведении телефонных переговоров по низкоскоростным каналам связи процесс аутентификации является непрерывным и связанным непосредственно с обработкой речевого сигнала (РС) при применении метода линейного предсказания для его кодирования.

Теоретическая цель работы заключается в обосновании актуальности направления по созданию нового поколения защищенных систем низкоскоростного кодирования речи, основанных на использовании ее индивидуальных статистических и параметрических характеристик, вычисляемых в процессе обработки.

Основная часть

Среди многообразия разработанных методов обработки и кодирования РС одним из наиболее эффективных является метод прямого предсказания речи [1], использующийся в современных стандартах низкоскоростного речевого кодирования (табл. 1).

Анализ таблицы 1 указывает на тот факт, что основная доля информационных ресурсов кадра передачи расходуется на представление информации о сигнале возбуждения. В работе предлагается объективная возможность снижения количества бит, выделяемых на представление сигнала возбуждения при одновременном уменьшении вычислительных затрат на поиск вектора сигнала возбуждения в кодовой книге при векторном квантовании данных.

Современные системы низкоскоростного кодирования речи на основе линейного предсказания в общем случае реализуются с использованием переменной структуры и изменяемыми параметрами в выбранной структуре в процессе обработки. Таким образом, ее можно классифицировать как адаптивную систему с переменной структурой и параметрами. Процессом изменения структуры и параметров устройства обработки

Таблица 1

Распределение информационных ресурсов в стандартизованных липредерах

Стандарт	GSM	G.729	Усредненное распределение
Тип липредера	RPE-LTP-LPC	CS-ACELP	
V_k , кбит/с	13	8	
Коэффициенты модели	0,14	0,23	0,19
Сигнал возбуждения	0,86	0,77	0,81

¹ Афанасьев Андрей Алексеевич, кандидат технических наук, доцент, Академия ФСО России, г. Орёл, fromnet@yandex.ru

в реальном масштабе времени управляет устройством, в котором происходят анализ и выделение параметров РС. В соответствии с полученными параметрами выбирается одна из возможных структур обрабатывающего устройства, которая наиболее адекватно отобразит РС на сегменте [2].

Из статистического анализа характеристик РС известно, что распределение формант и их число для различных звуков речи отличаются друг от друга [3]. Данное свойство РС используется для построения системы компрессии речи с изменяемым числом параметров формирующей модели. В зависимости от полученных статистических характеристик обработанного фрагмента речи принимается решение о передаче характеристических параметров. Ими являются параметры передаточной функции голосового тракта формирующего РС (линейные спектральные пары) и соответствующий сигнал возбуждения, их нахождение реализуется с использованием процедуры анализа через синтез [4]. Классическая функциональная схема кодера, основанного на применении метода линейного предсказания и процедуры анализа через синтез, реализующего данный подход, приведена на рисунке 1. Она соответствует последним практически внедренным разработкам исследователей в данной области и подробно представлена в [5].

Сущность метода линейного предсказания заключается в том, что выборка РС $S(n)$ может быть предсказана линейной комбинацией предшествующих отсчетов этого сигнала:

$$S'(n) = \sum_{i=1}^M a_i S(n-i) + e(n), \quad (1)$$

где $S'(n)$ – предсказанное значение РС; a_i – весовой коэффициент или коэффициент линейного предсказания; M – число коэффициентов или порядок линейного предсказания, $e(n)$ – ошибка предсказания.

Теоретическим основополагающим базисом метода линейного предсказания является авторегрессионная модель, успешно применяемая для решения различных задач цифрового спектрального анализа и предполагающая в общем «идеальном» случае бесконечный порядок формирующей системы при возбуждении ее сигналом в виде дискретного белого гауссовского шума. Ее идентификация связана с решением системы алгебраических матричных уравнений Юла–Уокера [3]. В классической постановке задачи параметрического цифрового спектрального анализа возбуждение формирующего фильтра осуществляется сигналом $u(n)$, представляющим собой реализации белого шума с математическим ожиданием, равным нулю, и единичной дисперсией:

$$\begin{cases} M\{u(n)\} = 0, \\ D\{u(n)\} = \sigma^2\{u(n)\} = 1. \end{cases} \quad (2)$$

Точность идентификации математической модели исследуемого процесса напрямую связана с выбором величины ее порядка M .

В качестве критерия настройки модели в предположении о гауссовском законе распределения исходного процесса используется взвешенная среднеквадратическая ошибка:

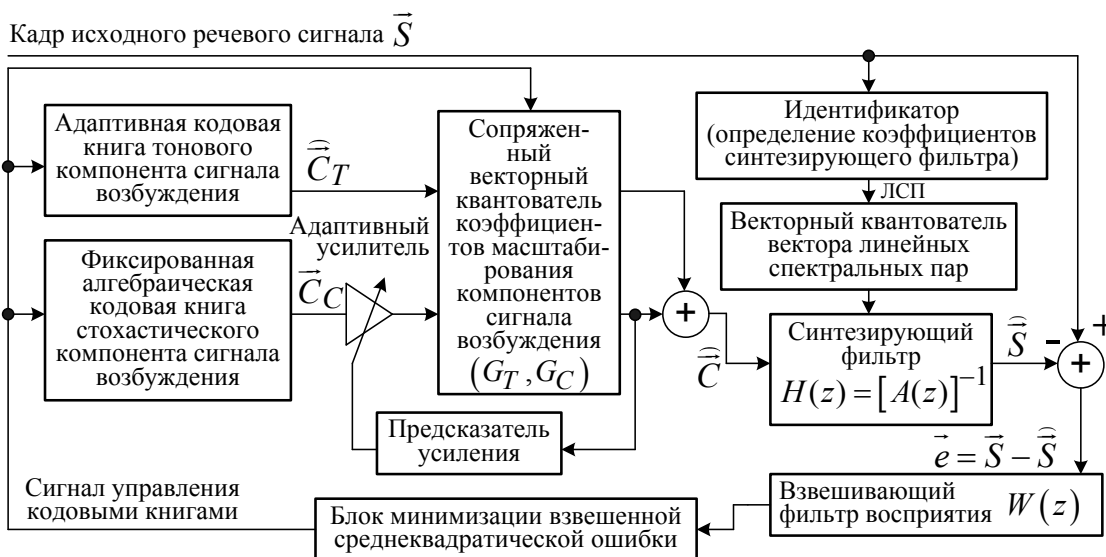


Рис. 1. Функциональная схема классического кодера линейного предсказания речевого сигнала на основе процедуры анализа через синтез

$$e^2(n) = d_2(\vec{S}, \vec{S}') = \frac{1}{N}(\vec{S} - \vec{S}')^T(\vec{S} - \vec{S}') = \frac{1}{N} \sum_{i=1}^N (\vec{S} - \vec{S}')^2, \quad (3)$$

где \vec{S} – вектор оригинального РС, \vec{S}' – вектор синтезированного РС, N – количество отсчетов на сегменте анализа.

Применительно к задаче предсказания речи повышение порядка передаточных функций фильтров анализа и синтеза приводит к «обелению» сигнала остатка предсказания, который является наилучшим сигналом возбуждения.

В классической постановке задачи параметрического цифрового спектрального анализа на основе авторегрессионной модели линейное разностное уравнение формирующего фильтра выглядит следующим образом:

$$y(nT) = -\sum_{m=1}^M a_m y(nT - mT) + u(nT), \quad (4)$$

где $y(nT)$ – выходной сигнал, T – интервал дискретизации, a_m – коэффициенты фильтра.

Его амплитудно-частотная характеристика определяется в виде

$$A(\omega T) = \frac{1}{\sqrt{(1 + \sum_{m=1}^M a_m \cos m\omega T)^2 + (\sum_{m=1}^M a_m \sin m\omega T)^2}} \quad (5)$$

а спектральная плотность мощности

$$G(\omega) = \frac{\sigma^2 \{u(n)\} T}{(1 + \sum_{m=1}^M a_m \cos m\omega T)^2 + (\sum_{m=1}^M a_m \sin m\omega T)^2}, \quad (6)$$

где ω – круговая частота дискретного преобразования Фурье.

Повышение порядка модели в выражениях (4), (5) и (6) приводит к получению более точных оценок относительно анализируемого сигнала \vec{S} . В идеале $e^2(n) \rightarrow 0$ при $M \rightarrow \infty$.

На рисунке 2 представлена спектральная плотность мощности РС и сигнала остатка линейного предсказания на сегменте квазистационарности, анализ которого позволяет утверждать о наличии взаимосвязи данных параметров.

На практике при реализации процедур предсказания значение M всегда ограничено, что приводит к возникновению сигнала ошибки $e(n)$, являющегося сигналом возбуждения фильтра синтеза модели линейного предсказания. Таким образом, сигнал $e(n)$ уже не является реализациями белого шума с математическим ожиданием, равным нулю, и единичной дисперсией, а становится квазидетерминированным относительно множества $\{a_m\}$ и связан с ним соответствующими корреляционными зависимостями.

Такое представление элементов декомпозиции РС в процессе функционирования системы низкоскоростного кодирования при предварительном индивидуальном ее обучении дает возможность ввести термин «непрерывная аутентификация диктора». Предварительное обучение такой систе-

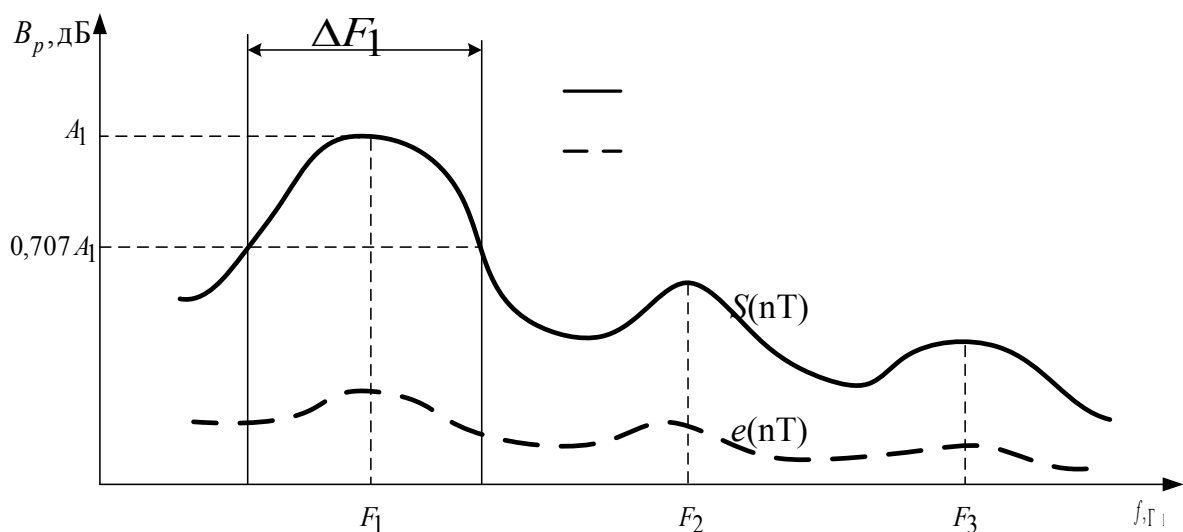


Рис. 2. Спектральная плотность мощности речевого сигнала и сигнала остатка линейного предсказания

мы при котором в процессе выполняется анализ и выделяются индивидуальные параметры диктора, даст возможность осуществлять непрерывную аутентификацию диктора при ведении телефонных переговоров.

Значения (\vec{a}, \vec{e}) , определяющие функционирование модели синтезирующего фильтра, носят векторный характер, при этом они связаны друг с другом и со значениями $S(n)$ некоторыми функциональными ограничениями и зависимостями. Мощности пространств представлений параметров $\{\vec{a}\}$ и $\{\vec{e}\}$ в общем случае различны. Эти ограничения приводятся к следующему виду:

$$(\vec{a}, \vec{e})_v = f_v(s),$$

где v – порядковый номер сегмента анализа.

При этом для получения приемлемых результатов обработки необходимые значения (\vec{a}, \vec{e}) должны быть неизменны на всем интервале времени для v .

Разработка систем подобного класса рассматривается как многокритериальная оптимизационная задача. Подобный класс оптимизационных задач возникает, когда оптимизация ведется по нескольким различным критериям, предъявляемым к системе компрессии. При решении данной задачи целесообразно воспользоваться методами, описанными в [6], которые приводят к получению приемлемых характеристик системы и могут быть использованы при ее синтезе. Однако предложенные методы целесообразно применять в случае унимодальности поставленной оптимизационной задачи, для задач по обработке речи это не всегда возможно.

Основная трудность практической реализации таких систем связана с отсутствием априорной информации о вероятной мощности пространства представлений различных вариантов структур при оптимальном решении этой задачи. Устранение данной неопределенности частично реализуется при рациональном ограничении возможных вариантов и заданном требуемом качестве обработки, так как необходимо передавать информацию о состоянии и структуре системы в процессе ее функционирования. При синтезе систем обработки речевых данных многие решения носят эмпирический характер, так как пока отсутствует четкий математический аппарат, который позволил бы связать воедино паритетный выбор между количеством возможных структур, качеством функционирования, алгоритмом работы и сложностью функционирования.

При обработке РС и использовании байесо-

ва подхода в идеальном виде требуется знание функциональной зависимости ожидаемой ошибки $e(n)$ от принятого решения U и значений $S(n)$ обрабатываемого РС [3]. Для вычисления данного значения необходимо статистическое описание наблюдений $S(n)$, а также параметров состояния синтезирующей системы. При этом полезный объем данных определяет значение ожидаемой ошибки при любом из ограниченного множества возможных решений. Выполнение задачи по синтезу системы в такой постановке осуществляется при известных распределениях вероятности $p(S|\vec{a}, \vec{e})$ и $p(\vec{a}, \vec{e})$, где S – значения отсчетов РС на сегменте анализа, (\vec{a}, \vec{e}) – пара векторов параметров голосового тракта и сигнала возбуждения соответственно, выбранные с использованием процедуры анализа через синтез. Однако на практике полное статистическое описание $S(n)$ и (\vec{a}, \vec{e}) получить невозможно, что связано в данном случае с высокой размерностью решаемой задачи и сложностью статистического анализа речевых данных. При этом задачи по обработке речевых данных сопровождаются большей или меньшей априорной неопределенностью, которая ограничивает полноту статистического описания. Аппроксимация распределений $S(n)$ и (\vec{a}, \vec{e}) с помощью нормального распределения носит достаточно приближенный характер, ограничивающий область применения решений, и привносит ошибки в конечные результаты вычислений.

Таким образом, существуют затруднения в прямом использовании байесова подхода со всеми его преимуществами и возможностями. Возникающая априорная неопределенность относительно свойств источника делает особо важной разработку методов синтеза систем обработки информации и принятия решений при таких условиях.

При наличии данных ограничений общее выражение для функции ошибки определяется в виде $e(\{U\}, \vec{a}, M, S)$ в случае использования функции правдоподобия $p(S|\vec{a}, \vec{e})$. При таком подходе отражаются некоторые статистические характеристики $\{\vec{a}\}$ и $\{\vec{e}\}$, на основании которых вычисляется взаимосвязь между ними. Для задания подробного статистического описания в этом случае необходимы математическое ожидание параметров, дисперсия, а также моменты более высокого порядка. Априорные данные могут быть получены путем задания некоторого математического ожидания многокомпонентной функции $f(\vec{e})$. С теоретической и практической точек зрения большое значение имеет вычисление корреляционной матрицы $R = \|B\|$ вектора \vec{e} . Результаты

таких вычислений используются для управления мощностью пространства представлений сигнала возбуждения.

При этом векторы (\vec{a}, \vec{e}) в процессе обработки могут рассматриваться как векторный скрытый марковский случайный процесс (7), т. е.

$$p(\vec{a}, \vec{e}) = \prod_{i=2}^V p_1[(\vec{a}, \vec{e})_i | (\vec{a}, \vec{e})_{i-1}] p_0(\vec{a}, \vec{e}), \quad (7)$$

где переходная $p_1[(\vec{a}, \vec{e})_i | (\vec{a}, \vec{e})_{i-1}]$ и начальная $p_0(\vec{a}, \vec{e})$ плотности в общем случае полностью неизвестны.

Таким образом, вместо единственного распределения вероятности для параметров (\vec{a}, \vec{e}) с плотностью $p(\vec{a}, \vec{e})$ для решения задач по обработке РС задается класс распределений P_0 , к которому относятся все возможные распределения $p(\vec{a}, \vec{e}) \in P_0$.

При этом наиболее универсальным способом статистического описания априорной неопределенности и учета имеющихся ограниченных сведений применительно к наблюдениям $S(n)$ и к ненаблюдаемым параметрам (\vec{a}, \vec{e}) является параметрический способ.

Двумерное векторное множество (\vec{a}, \vec{e}) дискретно и каждое его значение принимается с определенной вероятностью. При этом если априорное распределение вероятностей не известно, то в качестве неизвестных параметров рассматриваются сами эти вероятности, формальное задание которых с учетом естественных ограничений $p_i(\vec{a}, \vec{e}) \geq 0, \sum_{k=1}^P p_i = 1$ определяет необходимое распределение вероятностей через $P-1$ неизвестный параметр.

Последовательность переходов значений (\vec{a}, \vec{e}) определяет скрытый марковский процесс. Недостатком такого представления является априорное условие распределения параметров (\vec{a}, \vec{e}) по нормальному закону, чего на практике зачастую не выполняется. Тем не менее данный подход является удобным средством для учета имеющихся качественных представлений о статистическом поведении наблюдаемых данных $S(n)$ и параметров (\vec{a}, \vec{e}) в сочетании с незнанием детальных количественных характеристик, точно определяющих это описание. Именно такое сочетание наиболее характерно для большинства прикладных задач обработки речи. Качественные представления, основанные на физической сущности рассматриваемой задачи, дают возможность задать структуру распределений вероятности для $S(n)$ и (\vec{a}, \vec{e}) , при этом если $S(n)$ или

(\vec{a}, \vec{e}) представляют собой конечное множество, то в качестве неизвестных параметров рассматриваются сами вероятности данных значений. Таким образом, параметрическое описание априорной неопределенности является достаточно универсальным средством учета ограниченных априорных сведений.

Данное описание должно удовлетворять двум подчас противоречивым требованиям. Во-первых, оно должно качественно правильно и, по возможности, количественно точно отражать ограниченные априорные знания, чтобы распределения с плотностями $p(\vec{a}, \vec{e})$ при соответствующих переменных действительно представляли возможные в данной задаче распределения. Во-вторых, число параметров не должно быть слишком велико.

Увеличение размерности приводит к ухудшению качества решения основной задачи как из-за сложности технической реализации алгоритмов обработки данных наблюдения $S(n)$, так и из-за утраты некоторой доли входной информации, которую неизбежно приходится затрачивать для определения значений или исключения неизвестных мешающих параметров.

Следовательно, разрабатываемую в данных условиях априорной неопределенности систему дополним подсистемой проверки правильности априорных предположений, положенных в основу принятого в данной задаче параметрического описания. Ее задача – установить, верно ли качественно введенное описание (например при фильтрации процесса $S(n)$ соответствует ли действительности аппроксимация $S(n)$ полиномом заданной степени с вычисленными коэффициентами или эта модель недостаточно адекватна) и достаточно ли число введенных параметров относительно (\vec{a}, \vec{e}) . Такой алгоритм дает возможность при необходимости усложнить параметрическое описание, увеличив число параметров или качественно изменив модель априорной неопределенности, введя статистическую зависимость параметров $\{\vec{a}\}$ и $\{\vec{e}\}$

Наличие такой зависимости объясняется тем, что в стандартах низкоскоростного кодирования речи используется ограниченный порядок анализирующего и синтезирующего фильтров, что определяется возможностью их физической реализации при необходимой и достаточной точности описания передаточной функции голосового тракта человека.

Присутствие взаимозависимостей между параметрами, описывающими передаточную функцию голосового тракта на участке квазистационарно-

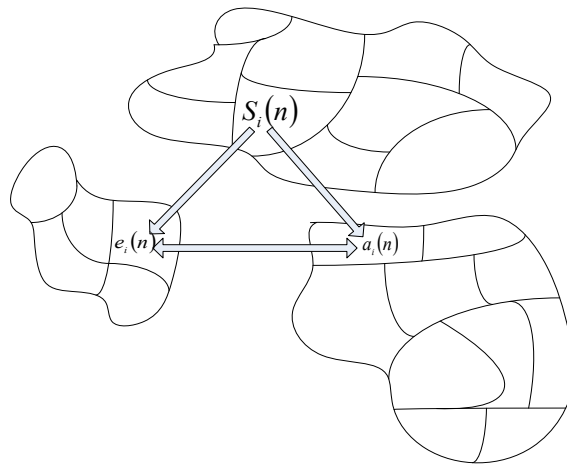


Рис. 3. Классы подпространств представлений соответствий элементов декомпозиции РС

сти, и соответствующим сигналом возбуждения объясняется особенностями постановки и решения обратной задачи цифрового спектрального анализа при фиксированном порядке формирующего фильтра [7].

Точность идентификации математической модели исследуемого процесса напрямую связана с выбором величины ее порядка P . Повышение порядка модели приводит к получению более точных оценок относительно анализируемого сигнала \vec{S} . В идеале $e^2(n) \rightarrow 0$ при $P \rightarrow \infty$. Однако на практике при реализации линейного предсказания значение P всегда ограничено, что приводит к возникновению сигнала $e(n)$, являющегося сигналом возбуждения фильтра синтеза модели линейного предсказания. Таким образом, сигнал $e(n)$ уже не является реализацией белого шума с математическим ожиданием, равным нулю, и единичной дисперсией, а становится квазидетерминированным относительно множества $\{a_m\}$ и связан с ним взаимозависимостями. На рисунке 3 показано, что при формировании ограниченных множеств параметров голосового тракта $\{a_i(n)\}$ и

сигналов возбуждения на основе остатка предсказания в виде кодовых книг данные зависимости вырождаются в соответствующие классы подпространств соответствий между собой и определяют элементы декомпозиции РС.

Таким образом, между пространством параметров формирующей модели и пространством векторов сигнала ошибки линейного предсказания (сигнала возбуждения на приеме) существует взаимосвязь, которую можно использовать с целью непрерывной аутентификации диктора. Для этого необходимо предварительное обучение системы с целью установления соответствия вектора подмножества сигналов возбуждения с соответствующим вектором параметров голосового тракта. Нарушение данных взаимосвязей, соответственно, изменение вероятностного распределения соответствий будет указывать на факт несанкционированного использования средства связи для доступа к услугам сети. Пример установления взаимосвязи для элементов декомпозиции 80-мерного сегмента для векторов параметров голосового тракта представлен в таблице 2.

Таблица 2

Соответствие векторов элементов декомпозиции РС при нейросетевом квантовании

Номера векторов кодовой книги параметров голосового тракта	Номера векторов кодовой книги сигналов возбуждения									
	18	25	34	38	54	69	125	203	250	–
1	18	25	34	38	54	69	125	203	250	–
2	36	42	50	54	56	67	92	–	–	–
3	71	76	80	84	89	93	144	189	226	245
4	5	18	104	200	212	–	–	–	–	–
5	9	32	40	–	–	–	–	–	–	–
...

Приведем доказательство того факта, что сигнал возбуждения не является независимым процессом в виде модели белого шума. Речевой сигнал может быть представлен выражением (8), при этом его оценка на интервале nT есть выражение (9)

$$S(nT) = S'(nT, (\bar{a}, \bar{e})) + n_0(nT), \quad (8)$$

$$\hat{S}'(nT) = M\{S'(nT, (\bar{a}, \bar{e})) | S_0(nT)\} = \sum_{\bar{a}, \bar{e}} S'(nT, (\bar{a}, \bar{e})) p(nT, (\bar{a}, \bar{e})), \quad (9)$$

где $S'(nT, (\bar{a}, \bar{e}))$ – предсказанный РС; $n_0(nT)$ – собственный аппаратный шум, $\hat{S}'(nT)$ – оценка предсказанного РС.

$$p(nT, (\bar{a}, \bar{e})) = p(nT, (\bar{a}, \bar{e}) | S_0(nT)), \quad (10)$$

где $p(nT, (\bar{a}, \bar{e}))$ – апостериорная плотность вероятности параметров его декомпозиции на интервале nT .

Порождающий процесс для сигнала, представленного выражением (8), представим через формулу (11), где $\xi(nT)$ – дискретный белый гауссовский шум (ДБГШ), следовательно, должны выполняться условия:

$$\xi(nT) = S(nT) - \hat{S}'(nT), \quad (11)$$

$$M\{\xi(nT)\} = 0; \quad (12)$$

$$M\{\xi(n_1T)\xi(n_2T)\} = N_0\delta(n_1T - n_2T) / 2,$$

где N_0 – односторонняя спектральная плотность мощности ДБГШ.

$$\begin{aligned} M\{\xi(nT)\} &= M\{S(nT)\} - M\{\hat{S}'(nT)\} = \\ &= M\{S(nT, (\bar{a}, \bar{e})) + n_0(nT)\} - M\{\hat{S}'(nT)\} = \\ &= M\{M\{S'(nT, (\bar{a}, \bar{e})) | S_0(nT)\}_{(\bar{a}, \bar{e})} + n_0(nT)\} - \\ &- M\{\hat{S}'(nT)\} = M\{\hat{S}'(nT)\} - M\{\hat{S}'(nT)\} = 0. \end{aligned} \quad (13)$$

Для доказательства второго положения выражения (12), введем несколько иное задание порождающего процесса через выражения:

$$\xi(nT) = \varepsilon(nT) + n_0(nT), \quad (14)$$

$$\varepsilon(nT) = S'(nT, (\bar{a}, \bar{e})) - \hat{S}'(nT),$$

$$\begin{aligned} M\{\xi(nT)\xi(nT + \Delta T)\} &= \\ &= M\{\xi(nT)[\varepsilon(nT + \Delta T) + n_0(nT + \Delta T)]\} = \\ &= M\{\xi(nT)\varepsilon(nT + \Delta T)\} + M\{\xi(nT)n_0(nT + \Delta T)\} + \\ &+ M\{n_0(nT)n_0(nT + \Delta T)\}. \end{aligned} \quad (15)$$

В выражении (15) первые два слагаемых можно считать равными нулю, что подтверждается в формулах (16) и (17), таким образом выражение (15) трансформируется в формулу (18), что свидетельствует о том, что $\xi(nT)$ является дискретным белым шумом при такой постановке:

$$\begin{aligned} M\{\xi(nT)\varepsilon(nT + \Delta T)\} &= \\ &= M\{M\{\xi(nT)\varepsilon(nT + \Delta T) / S_0(nT + \Delta T)\}_{(\bar{a}, \bar{e})} \}_{S_0(nT + \Delta T)} = \\ &= M\{M\{\varepsilon(nT + \Delta T) / S_0(nT + \Delta T)\}_{(\bar{a}, \bar{e})} \xi(nT)\}_{S_0(nT)} = \\ &= 0 \text{ при } \Delta T > 0 \end{aligned} \quad (16)$$

При $\hat{S}'(nT + \Delta T)$ справедливо $M\{\varepsilon(nT + \Delta T) / S_0(nT + \Delta T)\} = 0$

$$\begin{aligned} M\{\varepsilon(nT)n_0(nT + \Delta T)\} &= \\ &= M\{[S'(nT, (\bar{a}, \bar{e})) - \hat{S}'(nT)]n_0(nT + \Delta T)\} = \\ &= M\{S'(nT, (\bar{a}, \bar{e}))\}M\{n_0(nT + \Delta T)\} - \\ &- M\{\hat{S}'(nT)\}M\{n_0(nT + \Delta T)\} = 0. \end{aligned} \quad (17)$$

$$\begin{aligned} M\{\xi(nT)\xi(nT + \Delta T)\} &= M\{\xi(nT)\xi(nT + \Delta T)\} = \\ &= M\{n_0(nT)n_0(nT + \Delta T)\} = N_0\delta(\Delta T) \end{aligned} \quad (18)$$

Порождающий процесс $\xi(nT)$ представлен выражением (14) ($\varepsilon(nT)$ – ошибка оценки $S'(nT, (\bar{a}, \bar{e}))$ с распределением $p_\varepsilon(\varepsilon, nT)$) при выполнении условия (19), следствием из которого является формула (18), так как $\varepsilon(nT)$ и $n_0(nT)$ некоррелированы при $\Delta T > 0$, и плотность вероятности их суммы равна свертке «широкой» нормальной плотности вероятности и «узкой» плотности вероятности оценки процесса [8]:

$$D\{n_0(nT)\} \gg D\{\varepsilon(nT)\}, \quad (19)$$

$$p_\xi(n_\xi, nT) = \sum_\varepsilon p_0(\xi - \varepsilon, nT) p_\varepsilon(\varepsilon, nT) \approx p_0(\xi, nT). \quad (20)$$

Из формулы (20) следует, что $\xi(nT)$ должен быть случайным процессом вида ДБГШ, однако анализ сигнала $e(n)$ (рис. 2) позволяет утверждать, что он не является процессом вида ДБГШ.

Представленный подход статистического описания $S(n)$ и (\bar{a}, \bar{e}) соответствуют тому, что в каждой конкретной практической задаче создается некоторая аналитическая модель для описания статистических свойств $S(n)$ и (\bar{a}, \bar{e}) с той степенью полноты и подробности, которая соответствует имеющимся знаниям о закономерностях их поведения, физических свойствах, взаимосвязи.

Подобная модель является наиболее сжатым описанием имеющегося опыта, содержащего как результаты изучения данных закономерностей, так и эмпирические данные относительно $S(n)$ и (\vec{a}, \vec{e})

Однако на практике получить статистическое описание параметров (\vec{a}, \vec{e}) отдельного диктора можно для реализации задачи по аутентификации в процессе кодирования данных, при этом следует учитывать тот факт, что аутентификация в процессе функционирования такой системы обработки становится непрерывной. Для более широкого класса задач по обработке речи (кодирование) такое представление достаточно проблематично, что связано с отсутствием априорного знания всех возможных голосов дикторов, подвергаемых обработке. Тем не менее на заранее сформированных универсальных кодовых книгах параметров декомпозиции сформировать классы взаимосвязей подобных параметров для отдельного диктора не составляет труда.

Следует также учитывать тот факт, что заранее до обучения нет возможности указать все классы взаимосвязей между соответствующими подпространствами векторов декомпозиции и, соответственно, выйти на конкретные значения скоростей передачи. Фиксированную информационную скорость можно реализовать на основе использования заранее сформированных пространств векторов декомпозиции конечной мощности, положение центроидов в которых будет определено в процессе обучения. В итоге создаются кодовые книги, содержащие векторы (образцы) параметров, описывающих передаточную функцию голосового тракта и соответствующего им сигнала возбуждения. Это позволяет перейти от учета статистики взаимосвязи между отдельными значениями данных векторов, вычисляемыми при обработке, к статистике взаимосвязи векторов между соответствующими кодовыми книгами.

На практике бывает и так, что кроме эмпирических данных всякая иная априорная информация относительно $S(n)$ и (\vec{a}, \vec{e}) отсутствует. Однако если эти данные получены в обстановке, статистически однородной или хотя бы статистически связанной с той, в которой принимается решение по данным наблюдения $S(n)$, то они являются в определенной степени статистическим эквивалентом аналитических моделей для распределений вероятности $S(n)$ и (\vec{a}, \vec{e}) необходимых для нахождения оптимального правила принятия решения.

Степень этой эквивалентности, естественно, зависит от объема имеющихся эмпирических данных, которые, в свою очередь, могут быть использованы по-разному: непосредственно для нахождения недостающих распределений вероятности; для оценки функциональной зависимости апостериорного риска от $S(n)$ и решения U ; для уточнения структуры и параметров решающего правила, т. е. алгоритма обработки данных наблюдения $S(n)$.

Применение выявленных взаимозависимости между элементами декомпозиции РС приводит к возможности значительного сокращения мощностей подпространств представления сигналов возбуждения синтезирующего фильтра, при этом возрастает требование по наличию достаточного объема запоминающих устройств.

Одновременное использование процедур интерполяции параметров передаточной функции на сегментах одинаковой природы образования РС дает возможность значительного снижения скорости передачи при сохранении качественных показателей.

Представленные подходы могут быть применены в задачах по низкоскоростной передаче речи при ограниченном ресурсе пропускной способности канала связи.

Заключение

Таким образом, в данной работе представлено обоснование перехода к построению систем обработки РС, осуществляющих непрерывную аутентификацию диктора при доступе к ресурсу инфокоммуникационных систем, определены возможные пути создания подобных устройств, основанных на использовании объективно существующих зависимостей между элементами декомпозиции РС. Учет и применение данных зависимостей на этапе обучения системы кодирования дают возможность снизить скорость передачи РС по каналам связи и при этом осуществить непрерывную аутентификацию диктора при ведении телефонных переговоров по низкоскоростным цифровым каналам передачи. Актуальность исследований и значимость полученных результатов подтверждаются существующими объективными требованиями, предъявляемыми к разрабатываемым защищенным системам обработки РС, в частном случае при низкоскоростном кодировании РС, сформированными с учетом тенденций развития систем инфокоммуникаций.

Рецензент: Батенков Кирилл Александрович, кандидат технических наук, сотрудник Академии ФСО России, г.Орел, pustur@yandex.ru

Литература:

1. Маркел Дж.Д. Линейное предсказание речи / Д. Маркел, А.Х. Грей. пер. с англ. / под ред. Ю.Н. Прохорова и В.С. Звездина. – М.: Связь, 1980. – 308 с.: ил.
2. Бухалев В.А. Распознавание, оценивание и управление в системах со случайной скачкообразной структурой. – М.: Наука, Физматлит, 1996. – 288с.
3. Афанасьев А.А., Рыжков А.П. Использование взаимозависимостей параметров линейного предсказания при реализации процедур обработки речевых данных // Телекоммуникации. 2012. № 13. С. 36-39.
4. Быков С.Ф. Цифровая телефония. Учеб. пособие для вузов / С.Ф. Быков, В.И. Журавлев, И.А. Шалимов. – М.: Радио и связь, 2003. – 144 с.
5. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. – М.: Радио и связь, 2000. – 456 с.: ил.
6. Справочник по теории автоматического управления / под ред. А.А. Красовского. – М.: Наука, 1987. – 712 с.
7. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.

CONTINUOUS SPEAKER AUTHENTICATION AT CONDUCTING TELEPHONE NEGOTIATIONS ACROSS LOW BIT RATE DIGITAL CHANNELS

Afanasev A.A.²

The article considers the problem of working out continuous speaker authentication system. The method of a speech signal decomposition on the interconnected speech elements at its low bit rate coding based on a linear prediction is offered. Use of the given method at preliminary system training allows to carry out continuous subscriber authentication at access to communication channel throughput resource.

Keywords: *speech signal, speech coding, linear prediction, synthesizer filter, decomposing, interdependence.*

References:

1. Markel Dzh.D. Lineynoe predskazanie rechi, D. Markel, A.Kh. Grey. per. s angl. Pod red. Yu.N. Prokhorova i V.S. Zvezdina. – М.: Svyaz', 1980. – 308 P.: il.
2. Bukhalev V.A. Raspoznavanie, otsenivanie i upravlenie v sistemakh so sluchaynoy skachkoobraznoy strukturoy. – М.: Nauka, Fizmatlit, 1996. – 288P.
3. Afanas'yev A.A., Ryzhkov A.P. Ispol'zovanie vzaimozavisimostey parametrov lineynogo predskazaniya pri realizatsii protsedur obrabotki rechevykh dannykh, Telekommunikatsii. 2012. No 13, pp. 36-39.
4. Bykov S.F. Tsifrovaya telefoniya. Ucheb. posobie dlya vuzov, S.F. Bykov, V.I. Zhuravlev, I.A. Shalimov. – М.: Radio i svyaz', 2003. – 144 P.
5. Shelukhin O.I., Luk'yantsev N.F. Tsifrovaya obrabotka i peredacha rechi. – М.: Radio i svyaz', 2000. – 456 P.: il.
6. Spravochnik po teorii avtomaticheskogo upravleniya / pod red. A.A. Krasovskogo. – М.: Nauka, 1987. – 712 P.
7. Marpl-ml. S.L. Tsifrovoy spektral'nyy analiz i ego prilozheniya. – М.: Mir, 1990. – 584 s.



² Andrey Afanasev, Ph.D., Assistant Professor. The Academy of Federal Security Guard Service of the Russian Federation, Orel, from-net@yandex.ru