МЕТОД ЭКСТРАПОЛЯЦИИ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ЧАСТОТНЫХ ПРЕДСТАВЛЕНИЙ

Е.Г. ЖИЛЯКОВ
А.А. ЧЕРНОМОРЕЦ
В.А. ГОЛОЩАПОВА

Белгородский государственный университет

e-mail: zhilyakov@bsu.edu.ru

В статье рассматривается проблема экстраполяции речевых сигналов. Для вычисления прогнозного значения речевого сигнала на основе небольшого количества известных его предыдущих значений разработан нетрадиционный подход с использованием понятия частотной концентрации энергии отрезка.

Ключевые слова: речевые сигналы, эмпирические данные, частотные представления, аппроксимация, инвариант, экстраполяция.

Введение

Под эмпирическими данными обычно понимаются результаты регистрации количественных значений некоторого параметра, характеризующего поведение исследуемого объекта, находящегося под воздействием различных факторов, например, с течением времени (например, речевые сигналы). Техническая реализация процедур регистрации возможна только при дискретизации областей определения и значений исследуемых функций, так что эмпирические данные для речевых сигналов представляют собой наборы чисел $x_k=x(k\Delta t)$, $k=1,\ldots,N$, где $x(k\Delta t)$ – эквидистантные отсчёты (с шагом Δt) значений регистрируемого параметра $x(t)$; t-время.

Регистрация эмпирических данных осуществляется с целью получения возможности предсказании поведении объекта при иных, чем в реализованном опыте значениях воздействий. В частности, одной из важнейших задач является предсказание будущих значений речевых сигналов на основе уже зарегистрированных эмпирических данных.

В виду отсутствия априорных сведений о количественной взаимосвязи между временем и значениями регистрируемого параметра (точные количественные описания исследуемых функциональных зависимостей отсутствуют) проблема экстраполяции речевых сигналов является одной из самых сложных в их обработке. Её решение чаще всего сводится к поиску инвариантов, то есть некоторых характеристик речевых сигналов, которые остаются в том или ином смысле неизменными, по крайней мере, на ближайшем отрезке времени (краткосрочное предсказание). Предсказание значений речевых сигналов на основе выявления содержащихся в эмпирических данных инвариантов естественно называть экстраполяцией.

Основным подходом к поиску инвариантов в настоящее время служит аппроксимация на основе обработки эмпирических данных неизвестных функциональных зависимостей регистрируемых параметров от времени с помощью некоторых соотношений из того или иного класса математических моделей (идентификация речевых сигналов). При этом чаще всего в явном виде задаются пригодные для экстраполяции формулы, например, вероятностные распределения или модели генерации значений речевых сигналов в виде разностных стохастических уравнений.

Эмпирические данные тогда используются для оценивания некоторых априори неизвестных параметров предполагаемой функциональной зависимости (подгонка модели генерации речевого сигнала), которые, по крайней мере, в ближайшие будущие моменты времени предполагаются почти неизменными (локальная стационарность), и в этом смысле речь также идёт об установлении инвариантов.

Очевидно, что использование такого рода инвариантов равносильно априорному заданию в явном виде законов изменений регистрируемых параметров (законов природы) с чем не всегда можно согласиться. Кроме того, для оценивания значений неизвестных параметров постулируемых априори моделей генерации значений речевых сигналов может потребоваться значительное количество эмпирических данных, что также может оказаться нереализуемыми в виду нарушения условий стационарности (неизменности параметров).

Поэтому представляется актуальной разработка таких подходов, когда инварианты определяются на основе самых общих представлений и принципов использования моделей генерации значений речевых сигналов, которые, вместе с тем, адекватно отражают физическое содержание проблемы экстраполяции по конечному отрезку зарегистрированных эмпирических данных.

Решению этой задачи и посвящена настоящая работа.

1. Модели речевых сигналов на основе частотных представлений

Частотные представления вида

$$x_k = \int_{-\pi}^{\pi} X(\omega)\exp(j\omega(k-1))d\omega/2\pi, k=1,..,N \ , \tag{1}$$

являются одной из наиболее общих форм моделей, в которых генерация значений речевых сигналов определяется как суперпозиция бесконечно большого числа экспонент с весовой в общем случае комплексной функцией

$$X(\omega) = \operatorname{Re} X(\omega) - j\operatorname{Im} X(\omega)\,, \tag{2}$$

круговой нормированной (безразмерной) частоты *ω, j* – мнимая единица. Компоненты *ReX(ω),ImX(ω)* в правой соотношения части (2) являются вещественнозначными функциями, удовлетворяющими условиям

$$\begin{aligned} \operatorname{Re} X(-\omega) &= \operatorname{Re} X(\omega) \\ \operatorname{Im} X(-\omega) &= -\operatorname{Im} X(\omega) \end{aligned}. \tag{3}$$

Существенное значение имеет то, что при конечном наборе значений речевого сигнала этих условий и соотношений вида (1) (интегрального уравнения) недостаточно для однозначного определения весовой функции *X(ω)*. В частности им будет удовлетворять любая из следующих весовых функций

$$X(\omega) \equiv X_N(\omega) = \sum_{k=1}^{N} x_k \exp(-j\omega(k-1))\,, \tag{4}$$

$$X(\omega) \equiv XE(\omega) = X_N(\omega) + a * \exp(-j\omega N)\,, \tag{5}$$

в последнем соотношении *a* – любое вещественное ограниченное по абсолютному значению число.

Отметим, что сумма в правой части соотношения (4) является определением так называемых трансформант (преобразований) Фурье отрезков речевых сигналов длительности N [1]. В виду ортогональности набора экспонент *exp(jω(k-1)), k=1,…,N*, в области определения *|ω|≤π,* для трансформант Фурье вида (4) нетрудно получить равенство Парсеваля

$$\int_{-\pi}^{\pi} |\,X_N(\omega)\,|^2 \; d\omega/2\pi = \sum_{k=1}^{N} x_k^2 = \|\,\vec{x}_N\,\|^2\,, \tag{6}$$

Здесь и в дальнейшем стрелками сверху отмечаются векторы соответствующей размерности, то есть

$$\vec{z}_M = (z_1,..,z_M)^T\,,$$

где T – знак транспонирования.

В связи с этим отметим, что квадрат модуля трансформанты Фурье определяет распределение энергии отрезка речевого сигнала вдоль оси частот (спектральная плотность). Именно это обстоятельство определяет физическое содержание частотных представлений с весовыми функциями вида (4).

Введём функцию

$$f_N(\omega) = |\,X_N(\omega)\,|^2 \;/\, 2\pi \,/\, \|\,\vec{x}_N\,\|^2\,, \tag{7}$$

которая, очевидно, является неотрицательной

$$f_N(\omega) \ge 0, |\,\omega\,| \le \pi \ , \tag{8}$$

и нормированной к единице в смысле выполнения равенства

$$\int_{-\pi}^{\pi} f_N(\omega)d\omega = 1. \quad (9)$$

Легко поэтому понять, что интеграл

$$P_{rN} = \int_{\omega\in\Omega_r} f_N(\omega)d\omega, \quad (10)$$

определяет долю энергии отрезка речевого сигнала, попадающую в частотный интервал вида

$$\Omega_r = [-V_r, V_{r-1}) \cup [V_{r-1}, V_r), V_{r-1} < V_r; V_0 = 0; V_r \le \pi. \quad (11)$$

Для простоты изложения в дальнейшем предполагается, что вся область определения трансформант Фурье делится на R частотных интервалов одинаковых размеров, то есть

$$V_0 = 0; V_r - V_{r-1} = \pi / R. \quad (12)$$

Ясно, что из соотношений (9) и (10) следует равенство

$$\sum_{r=1}^{R} P_{rN} = 1. \quad (13)$$

В свою очередь с учётом определений (4), (11) и (12) на основе соотношения (1) нетрудно получить следующую форму частотного представления для отсчётов речевого сигнала

$$x_k = \sum_{r=1}^{R} x_{rk}, k = 1,.., N, \quad (14)$$

где x_{rk}– компоненты, определяемые интегралами вида

$$x_{rk} = \int_{\omega\in\Omega_r} X_N(\omega)\exp(j\omega(k-1))d\omega / 2\pi. \quad (15)$$

В дальнейшее они называются частотными компонентами исходного речевого сигнала. Непосредственно из соотношения (15) следует, что они полностью определяются отрезками трансформант Фурье в соответствующих частотных интервалах. Непосредственно из(15) нетрудно получить соотношения для трансформант Фурье частотных компонент

$$X_{rN}(\omega) = \int_{\omega\in\Omega_r} X_N(\omega_1)\exp(j(\omega-\omega_1)(N-1)/2)D_N(\omega-\omega_1))d\omega_1 / 2\pi, \quad (16)$$

где D_N– известная из теории рядов Фурье функция Дирихле

$$D_N(\nu) = \sin(N\nu/2)/\sin(\nu/2). \quad (17)$$

Из соотношений (16) и (17) следует, что энергии частотных компонент сосредоточены не только внутри выделенного частотного интервала (просачивание энергии), однако нетрудно получить равенство

$$S_{rN}^2 = \sum_{k=1}^{N} x_{rk}^2 = \iint_{\omega_1,\omega_2\in\Omega_r} X_N(\omega_1)X_N^*(\omega_2)D_N(\omega_2-\omega_1))d\omega_1 d\omega_2 / 4\pi^2, \quad (18)$$

которое определяет их через соответствующие отрезки трансформанты Фурье анализируемого речевого сигнала (здесь верхний индекс в виде звёздочки означает комплексное сопряжение).

Ясно, что если некоторые из долей энергий P_r малы по сравнению с другими, то мал будет вклад в сумму (14) и соответствующих частотных компонент. С другой стороны существенное превышение доли энергии в одном из частотных интервалов (либо суммы их небольшого количества) над суммой остальных долей свидетельствует о наличии в речевом сигнале квазициклических повторяющихся частотных компонент, которые естественно именовать инвариантами. Такое положение можно описать с помощью следующего соотношения

$$\sum_{r\in R_m} P_{rN} > \sum_{r\notin R_m} P_{rN}, \quad (16)$$

где R_m– множество частотных интервалов, в которых сосредоточена существенно большая (в том числе и подавляющая) доля энергии отрезка речевого сигнала, количество которых (мощность множества R_m) M удовлетворяет условию

$$w = M / R < 0{,}5 . \qquad (17)$$

Ясно, что если с хорошей точностью выполняются приближённые равенства долей энергии во всех частотных интервалах (отрезок так называемого белого шума)

$$P_{rN} \approx 1/R, r = 1,.., R , \qquad (18)$$

то это свидетельствует о невозможности в анализируемом отрезке речевого сигнала выделить искомые инварианты. Возможно, что они проявятся более отчётливо, если увеличить количество анализируемых отсчётов или уменьшить количество частотных интервалов.

Таким образом, характеристика отрезка речевого сигнала в виде распределения долей его энергии (10) по частотным интервалам (11) при выполнении (12) адекватно отражает наличие в нем квазициклических компонент, которые и могут служить пригодными для экстраполяции инвариантами.

2. Метод вычислений долей энергии отрезка речевого сигнала в заданных частотных интервалах

В соответствии с определением (7) представление (10) нетрудно преобразовать к виду

$$P_{rN} = \int\limits_{\omega \in \Omega_r} | X_N(\omega) |^2 \, d\omega / 2\pi / \| \vec{x}_N \|^2 . \qquad (19)$$

Подстановка сюда правой части представления (4) после несложных преобразований с учётом определения частотного интервала (11) даёт искомое вычислительное соотношение

$$P_{rN} = \vec{x}_N A_N^r \vec{x}_N / \| \vec{x}_N \|^2 , \qquad (20)$$

где $A^r{}_N=\{a^r{}_N\}, i, k=1,...,N$– матрица с элементами

$$a_{ik}^r = \int\limits_{\omega \in \Omega_r} \exp(-j\omega(i-k)) d\omega / 2\pi = [\sin(V_r (i-k)) - \sin(V_{r-1}(i-k))] / \pi(i-k) . \qquad (21)$$

Отсюда, имея в виду равенство (12), в результате предельного перехода нетрудно получить выражение для диагональных элементов

$$a_{ii}^r = [V_r - V_{r-1}] / \pi = 1/R . \qquad (22)$$

Частотное представление (21) элементов матрицы, позволяет именовать её субполосной.

Отметим, что соотношение (22) является точным и позволяет осуществлять необходимые вычисления без перехода в частотную область. Очевидно, что это обстоятельство делает реалистичным использование частотных представлений для поиска инвариантов и формулирования принципов их оптимального использования в задаче экстраполяции.

Вместе с тем непосредственное использование квадратичной формы

$$S_{rN}^2 = \vec{x}_N A_N^r \vec{x}_N , \qquad (23)$$

требует неоправданно больших вычислительных затрат, уменьшить которые можно на основе учёта специальных свойств субполосной матрицы.

В самом деле, из определения (21) следует, что эта матрица является положительно определённой и симметричной. Поэтому [2] она может быть представлена в виде

$$A_N^r = G_{rN} L_{rN} G_{rN}^T , \qquad (24)$$

где G_{rN}, L_{rN}– матрица ортонормальных собственных векторов и диагональная матрица неотрицательных собственных чисел, так что выполняются равенства

$$A_N^r G_{rN} = G_{rN} L_{rN} ; \qquad (25)$$

$$L_{rN} = diag(\lambda_{1r},.., \lambda_{Nr}) ;$$

$$\lambda_{r1} > \lambda_{r2} > .. > \lambda_{rN} > 0 ; \qquad (26)$$

$$G_{rN} = \left(\vec{q}_1^{\,r} ... \vec{q}_N^{\,r} \right) ;$$

$$G_{rN}G_{rN}^T = G_{rN}^T G_{rN} = I_N\,. \quad (27)$$

Здесь и в дальнейшем символом I_N обозначается единичная матрица соответствующей размерности.

На основе определения (25) с учётом (21) нетрудно получить частотное представление для компонент собственных векторов субполосных матриц

$$\lambda_{rn} q_{kn}^r = \int\limits_{\omega\in\Omega_r} Q_{nN}^r(\omega)\exp(j\omega(k-1))d\omega/2\pi\,, \quad (28)$$

где $Q^r_{nN}(\omega)$ – трансформанта Фурье соответствующего собственного вектора

$$Q_{nN}^r(\omega) = \sum_{k=1}^{N} q_{kn}^r \exp(-j\omega(k-1))\,. \quad (29)$$

Для скалярных произведений собственных векторов справедливо равенство Парсеваля

$$\sum_{k=1}^{N} q_{kn}^r q_{km}^r = \int\limits_{-\pi}^{\pi} Q_{nN}^r(\omega) Q_{mN}^{r*}(\omega) d\omega/2\pi\,. \quad (30)$$

С другой стороны из соотношения (28) для скалярных произведений тех собственных векторов нетрудно получить и другое равенство

$$\lambda_{rn}\sum_{k=1}^{N} q_{kn}^r q_{km}^r = \int\limits_{\omega\in\Omega_r} Q_{nN}^r(\omega) Q_{mN}^{r*}(\omega) d\omega/2\pi\,, \quad (31)$$

Имея в виду свойство ортонормальности (27) собственных векторов, из соотношений (30) и (31) нетрудно получить равенства

$$\int\limits_{-\pi}^{\pi} Q_{nN}^r(\omega) Q_{mN}^{r*}(\omega) d\omega/2\pi = \begin{cases} 0, n\neq m \\ 1, n=m \end{cases}; \quad (32)$$

$$\int\limits_{\omega\in\Omega_r} Q_{nN}^r(\omega) Q_{mN}^{r*}(\omega) d\omega/2\pi = m\neq n\,; \quad (33)$$

$$\lambda_{rn} = \int\limits_{\omega\in\Omega_r} |Q_{nN}^r(\omega)|^2\, d\omega/2\pi\,. \quad (34)$$

Таким образом, трансформанты Фурье собственных векторов ортогональны не только на всей оси частот, но и на соответствующем частотном интервале (свойство двойной ортогональности).

Из соотношения (34) следует, что собственное число равно доле энергии соответствующего собственного вектора в рассматриваемом частотном интервале.

Сопоставление равенств (32) и (34) с учётом (26) позволяет получить неравенство для значений собственных чисел

$$0 < \lambda_{rn} < 1\,. \quad (35)$$

По определению (24) матрицы $A^r{}_N, L_{rN}$ являются подобными, а, следовательно, их следы должны быть [2]

$$trecA_N^r = \sum_{i=1}^{N} a_{ii}^r = \sum_{n=1}^{N} \lambda_{rn}\,, \quad (36)$$

что в виду (22) позволяет получить следующие соотношения

$$\sum_{n=1}^{N} \lambda_{rn} = N/R\,; \quad (37)$$

$$\overline{\lambda}_r = \sum_{n=1}^{N} \lambda_{rn}/N = 1/R = a_{ii}^r, \forall i\,. \quad (38)$$

С другой стороны для определителя положительно определённой матрицы справедливо неравенство [4]

$$\det A_N^r = \prod_{n=1}^{N} \lambda_{rn} \leq \prod_{i=1}^{N} a_{ii}^r = 1/R^N\,. \quad (39)$$

откуда с учётом (35)и (38) можно сделать вывод о том, что величина некоторых из собственных чисел субполосной матрицы может быть очень мала.

Проведенные нами вычисления показывают, что значения собственных чисел во многом определяются величиной отношения в правой части равенства (37). Причём, если имеет место неравенство

$$n > J = N/R + 2, \tag{40}$$

то с хорошей точностью выполняются приближённые равенства

$$\lambda_{rn} \approx 0. \tag{41}$$

Поэтому подстановка разложения (24) в представление (23) позволяет получить достаточно точное вычислительное соотношение

$$S_{rN}^2 = \sum_{n=1}^{J} \lambda_{rn} \alpha_{rn}^2, \tag{42}$$

где α_{rn}– проекции исходного отрезка речевого сигнала на собственные векторы, то есть значения скалярных произведений вида

$$\alpha_{rn} = (\vec{x}_N, \vec{q}_n^{\,r}) = \sum_{k=1}^{N} x_k q_{kn}^r. \tag{43}$$

Легко понять, что использование соотношений (42) и (43) позволяет по сравнению с прямыми вычислениями на основе (23) уменьшить трудоёмкость почти в R раз.

3. Анализ динамики речевых сигналов на основе частотных представлений

В данном разделе речь идёт об анализе реакции долей энергии отрезков речевых сигналов в частотных интервалах на значения следующих отсчётов, то есть тех которые будут зарегистрированы после N-го отсчёта. Это позволяет описать некоторые характерные особенности динамики долей энергий и в частности выявить инварианты, которые полезны для задачи экстраполяции речевых сигналов на основе частотных представлений.

Возможность осуществления аналитических исследований указанной динамики представляют полученные выше соотношения для частей энергии отрезка речевого сигнала, относящихся к соответствующим частотным интервалам. В их основе используется следующее блочное представление субполосных матриц размерности $N+1$

$$A_{N+1}^r = \begin{pmatrix} A_N^r \vec{a}_{N+1}^{\,r} \\ \vec{a}_{N+1}^{\,rT} a_{N+1,N+1}^r \end{pmatrix}, \tag{44}$$

где $\bar{a}^r{}_N$ – вектор–столбец, состоящий из компонент

$$a_{N+1,k}^r = [\sin(V_r(N+1-k)) - \sin(V_{r-1}(N+1-k))]/\pi(N+1-k), k = 1,..,N,$$

а для нижнего диагонального элемента справедливо общее равенство вида (22)

$$a_{N+1,N+1}^r = 1/R. \tag{45}$$

Поэтому, имея в виду определение (20), долю энергии отрезка речевого сигнала соответствующей размерности нетрудно представить в виде удобном для анализа влияния будущей компоненты на её величину

$$P_{r,N+1} = [S_{rN}^2 + 2 * x_{N+1} z_{rN} + x_{N+1}^2 / R]/[\|\vec{x}_N\|^2 + x_{N+1}^2], \tag{46}$$

где z_{rN} -скалярное произведение вектора-столбца с компонентами (45) и вектора первых N отсчётов речевого сигнала

$$z_{rN} = (\vec{x}_N, \vec{a}_{N+1}^{\,r}) = \sum_{k=1}^{N} x_k a_{N+1,k}^r. \tag{47}$$

Очевидно, что в соответствии с определением (23) выполняется условие

$$\sum_{r=1}^{R} S_{r,N+1}^2 = \sum_{r=1}^{R} S_{rN}^2 + 2 * x_{N+1} \sum_{r=1}^{R} z_{rN} + x_{N+1}^2 = \|\vec{x}_{N+1}\|^2 = \|\vec{x}\|^2 + x_{N+1}^2, \tag{48}$$

из которого следует равенство

$$\sum_{r=1}^{R} z_{rN} = 0. \tag{49}$$

Таким образом, среди скалярных произведений вида (47) будут как положительные, так и отрицательные.

Отметим, что равенство (49) является инвариантным относительно размерности зарегистрированных значений.

Заметим также, что неизменность (инвариантность) попадающей в выбранный частотный интервал части энергии, то есть когда выполняется условие

$$S^2_{r,N+1} = S^2_{rN}\,, \tag{50}$$

будет наблюдаться при следующих значениях будущих отсчётов речевого сигнала

$$x_{1,r,N+1} = 0\,.$$

$$x_{2,r,N+1} = -2R * z_{rN}\ . \tag{51}$$

Отметим, что равенство будущего отсчёта нулю не приносит никаких изменений не только в значения частей энергий для всех интервалов, но и в их доли (отношения вида (46)), и поэтому такая инвариантность представляет мало интереса с точки зрения экстраполяции.

В свою очередь выполнение равенства (51) приводит к уменьшению доли энергии, по крайней мере, в заданном частотном интервале, то есть будет иметь место неравенство

$$P_{r,N+1} = S^2_{rN} /[\|\,\vec{x}_N\,\|^2 + 4R^2 z^2_{rN}\,] \le P_{rN} = S^2_{rN} \,/\, \|\,\vec{x}_N\,\|^2\,. \tag{52}$$

Здесь учтено, что скалярное произведение в правой части (51) может быть равно нулю.

Нетрудно понять, что при одном и том же значении $|x_{N+1}|$ совпадение его знака со знаком z_{rN}, то есть когда

$$sign(x_{N+1}) = sign(z_{rN})\,, \tag{53}$$

правая часть (46) будет больше чем, когда условие (53) не выполняется.

Очевидно, что правая часть представления (46) является бесконечно раз дифференцируемой функцией отсчёта x_{N+1}, что позволяет исследовать её динамические свойства на основе производных. В частности дифференцирование (46) даёт следующее соотношения для первой частной производной

$$\partial P_{r,N+1} \,/\, \partial x_{N+1} = 2[z_{rN}\,\|\,\vec{x}_N\,\|^2 - x_{N+1} h_{rN} - z_{rN} x^2_{N+1}\,]/[\|\,\vec{x}_N\,\|^2 + x^2_{N+1}\,]^2\,, \tag{54}$$

где

$$h_{rN} = S^2_{rN} - \|\,\vec{x}_N\,\|^2 \,/\, R\ . \tag{55}$$

Непосредственно из определения $S^2{}_{rN}$ следует справедливость равенства

$$\|\,\vec{x}_N\,\|^2 \,/\, R = \sum_{r=1}^{R} S^2_{rN} \,/\, R\,,$$

так что коэффициент параметр (55) в правой части (54) представляет собой разность между средним значением всех частей энергии и анализируемой частью. Следовательно, он может быть как положительным, если

$$\|\,\vec{x}_N\,\|^2 \,/\, R < S^2_{rN}\,, \tag{56}$$

так и отрицательным, когда анализируемая часть меньше средней величины

$$\|\,\vec{x}_N\,\|^2 \,/\, R > S^2_{rN}\,. \tag{57}$$

В свою очередь, если доли энергий практически равны, то есть, если справедливы приближённые равенства вида (18), то будет выполняться

$$h_{rN} \approx 0\,, \tag{58}$$

а правая часть (54) преобразуется к виду

$$\partial P_{r,N+1} \,/\, \partial x_{N+1} = 2 z_{rN} [\|\,\vec{x}_N\,\|^2 - x^2_{N+1}\,]/[\|\,\vec{x}_N\,\|^2 + x^2_{N+1}\,]^2\,. \tag{59}$$

При естественной однородности величин отсчетов в смысле выполнения неравенства

$$\|\,\vec{x}_N\,\|^2 > x^2_{N+1}\,, \tag{60}$$

знак правой части (59) будет совпадать со знаком скалярного произведения (47), то есть будет иметь место равенство

$$sign(P_{r,N+1}) = sign(z_{rN})\,. \quad (61)$$

Непосредственно из соотношения (54) нетрудно получить равенство для первой производной в начальной точке

$$\partial P_{r,N+1} / \partial x_{N+1} \,|_{x_{N+1}=0} = 2z_{rN} \,/\, \| \vec{x}_N \|^2 \,, \quad (62)$$

которое показывает, что положительный знак z_{rN} будет определять монотонный рост доли энергии в заданном частотном интервале с ростом положительного x_{N+1} или, наоборот, её убывание при отрицательном z_{rN}.

Представляют интерес и соотношения для частных производных более высоких порядков, которые нетрудно получить, последовательно дифференцируя правую часть представления (54)

$$\partial^2 P_{r,N+1} / \partial x_{N+1}^2 = 2z_{rN}[2x_{N+1}^3 + 3b_{rN}x_{N+1}^2 - 6\| \vec{x}_N \|^2 \, x_{N+1} - b_{rN} \| \vec{x}_N \|^2]/[\| \vec{x}_N \|^2 + x_{N+1}^2]^3 \,, \quad (63)$$

$$\partial^3 P_{r,N+1} / \partial x_{N+1}^3 = F(x_{N+1}) / [\| \vec{x}_N \|^2 + x_{N+1}^2]^4 \,, \quad (64)$$

$$F(x_{N+1}) = -12z_{rN}[x_{N+1}^4 + 2b_{rN}x_{N+1}^3 - 6\| \vec{x}_N \|^2 \, x_{N+1}^2 - 2\| \vec{x}_N \|^2 \, b_{rN}x_{N+1} + \| \vec{x}_N \|^4]\,; \quad (65)$$

$$b_{rN} = h_{rN} \,/\, z_{rN}\,. \quad (66)$$

Здесь предполагается, что знаменатель не равен нулю, в противном случае необходимо множитель перед скобкой вносить в неё.

Тогда получаем, что в точке

$$x_{r,N+1} = 0 \quad (67)$$

будут иметь место равенства

$$\partial^2 P_{r,N+1} / \partial x_{N+1}^2 \,|_{x_{N+1}=0} = -2h_{rN} \,/\, \| \vec{x}_N \|^4 \,, \quad (68)$$

$$\partial^3 P_{r,N+1} / \partial x_{N+1}^3 \,|_{x_{N+1}=0} = -12z_{rN} \,/\, \| \vec{x}_N \|^4 \,. \quad (69)$$

Здесь учтено определение (55). Легко также выразить правую часть (69) через правую часть (62).

Из условия

$$\partial P_{r,N+1} / \partial x_{N+1} = 0 \quad (70)$$

нетрудно определить, как корни соответствующего (54) квадратного уравнения, значения будущего отсчёта, при которых правая часть (46) достигает экстремумов

$$x_{r,N+1}^1 = b_{rN}[(1 + 4\| \vec{x}_N \|^2 \,/\, b_{rN}^2)^{1/2} + 1]/2\,, \quad (71)$$

$$x_{r,N+!}^2 = b_{rN}[-(1 + 4\| \vec{x}_N \|^2 \,/\, b_{rN}^2)^{1/2} + 1]/2\,. \quad (72)$$

С учётом определения (66) непосредственной подстановкой корней (71) и (72) в правую часть представления (46) легко убедиться, что тип экстремума последней будет определяться знаком параметра (55). При этом, если параметр положителен, то есть имеет место неравенство

$$\| \vec{x}_N \|^2 \,/\, R < S_{rN}^2 \,, \quad (73)$$

то экстремум типа максимума будет достигаться на корне (71) и наоборот, максимум будет достигаться на корне (72), когда неравенство (73) не выполняется.

Отметим, что левая часть соотношения (73) является средним значением частей энергий в выбранных частотных интервалах. Таким образом, это условие означает превосходство анализируемой части над средним значением всех, что, возможно, соответствует определённой тенденции в поведении речевого сигнала.

Ясно, что правая часть соотношения (62) определяет градиент изменений доли энергии. При этом, как легко видеть, имеют место равенства

$$\max |\, \partial P_{r,N+1} / \partial x_{N+1} \,|_{x_{N+1}=0}| = 2\max |\, z_{rN} \,| \,/\, \| \vec{x}_N \|^2, r = 1,..,R\,, \quad (74)$$

$$\min |\, \partial P_{r,N+1} / \partial x_{N+1} \,|_{x_{N+1}=0}| = 2\min |\, z_{rN} \,| \,/\, \| \vec{x}_N \|^2, r = 1,..,R\,. \quad (75)$$

Таким образом, всегда можно определить частотный интервал, в котором ожидаются максимальные или минимальные изменения долей энергий.

4. Экстраполяция речевых сигналов на основе требований к долям энергий будущего отрезка отсчётов в частотных интервалах

Выше были получены соотношения, позволяющие прогнозировать будущие значения долей энергии отрезка речевого сигнала в частотных интервалах в зависимости от значения будущего отсчёта x_{N+1} С точки зрения экстраполяции представляется естественным в качестве прогнозируемого значения этого отсчёта выбирать такое, которое сохраняет, по крайней мере, качественную картину распределения энергий предшествующего отрезка отсчётов меньшей длительности. При этом в качестве основы целесообразно использовать неравенство вида (16).

Соотношение (46) нетрудно преобразовать к рекуррентному виду

$$P_{r,N+1} = P_{r,N} + [2 * x_{N+1} z_{rN} - x_{N+1}^2 \varphi_{rN}]/[\| \vec{x}_N \|^2 + x_{N+1}^2], \quad (76)$$

где

$$\varphi_{rN} = P_{rN} - 1/R. \quad (77)$$

На основе свойства (13) легко получить равенство

$$\sum_{r=1}^{R} \varphi_{rN} = 0, \quad (78)$$

так что часть этих характеристик будет отрицательной., а другая часть– нет (когда доли энергии больше их среднего значения по всем частотным интервалам).

Пусть далее символ W_1 означает множество частотных интервалов, удовлетворяющих условию

$$\varphi_{rN} \geq 0, r \in W_1, \quad (79)$$

а R_1– количество этих интервалов (мощность множества). Положим

$$\Phi_{1N} = \sum_{r \in W_1} \varphi_{rN}; \Phi_{1,N+1} = \sum_{r \in W_1} \varphi_{r,N+1}; \quad (80)$$

$$Z_{1N} = \sum_{r \in W_1} z_{rN}. \quad (81)$$

Тогда на основе соотношения (76) нетрудно получить

$$\Phi_{r,N+1} - \Phi_{r,N} = [2 * x_{N+1} Z_{1N} - x_{N+1}^2 \Phi_{1N}]/[\| \vec{x}_N \|^2 + x_{N+1}^2], \quad (82)$$

причём в виду условия (79) выполняется неравенствj

$$\Phi_{1N} \geq 0. \quad (83)$$

Нетрудно показать, что числитель в правой части соотношения (82) достигает максимума, когда будущее значение речевого сигнала равно

$$\hat{x}_{N+1} = Z_{1N} / \Phi_{1N}. \quad (84)$$

Здесь предполагается, что в (83) имеет место строгое неравенство.

Подстановка этого равенства в правую часть (82) даёт значение разницы между исходным и получаемым значениями характеристики (80)

$$\Phi_{r,N+1} - \Phi_{r,N} = Z_{1N}^2 / \Phi_{1N} /[\| \vec{x}_N \|^2 + Z_{1N}^2 / \Phi_{1N}^2]. \quad (85)$$

Или

$$\Phi_{r,N+1} / \Phi_{r,N} = 1 + \Theta_{1N}^2 /[1 + \Theta_{1N}^2], \quad (86)$$

здесь

$$\Theta_{1N}^2 = Z_{1N}^2 / \Phi_{1N}^2 / \| \vec{x}_N \|^2. \quad (87)$$

Ясно, что будет иметь место

$$\Phi_{r,N+1} \geq \Phi_{r,N}. \quad (86)$$

Это соответствует идеологии неубывания долей энергии в частотных интервалах с большей её концентрацией на предыдущем шаге. Поэтому представление (84) может быть использовано для экстраполяции значений речевых сигналов.

Литература

1. Рабинер, Л. Теория и применение цифровой обработки сигналов [Текст] / Л. Рабинер, Г. Голд. – М.: Мир, 1988. – 512 с.
2. Гантмахер, Ф.Р. Теория матриц [Текст] / Ф.Р. Гантмахер. – М.: Физматлит, 2004. – 560с.
3. Хургин, Я.И. Финитные функции в физике и технике [Текст] / Я.И. Хургин, В.П. Яковлев. – М.:

Наука, 1971. – 408 с.: ил.
4. Крамер, Г. Математические методы статистики [Текст] / Под ред. акад. А.Н. Колмогорова. – М.: Мир, 1975. – 648 с.

SPEECH SIGNAL EXTRAPOLATION METHOD ON THE BASIS OF FREQUENCY REPRESENTATIONS

E.G. ZHILYAKOV
A.A. CHERNOMORETS
V.A. GOLOSCHAPOVA

Belgorod State University

e-mail: zhilyakov@bsu.edu.ru

Speech signal extrapolation problem is discussed in the work. The nontraditional approach using frequency concentration of signal segment energy concept is developed for calculation of speech signal prognostic value on basis of its previous values element.

Key words: speech signals, empirical data, frequency representations, approximation, invariant, extrapolation.