ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1978 год Выпуск 76

УДК 519.226:534.784

ОЦЕНИВАНИЕ ПАРАМЕТРОВ РЕЧЕВОГО ТРАКТА В КЛАССЕ МОДЕЛЕЙ АВТОРЕГРЕССИИ СО СТАЦИОНАРНОЙ СЕЗОННОЙ РАЗНОСТЬЮ ПЕРВОГО ПОРЯДКА И СТАЦИОНАРНОЙ РАЗНОСТЬЮ ПЕРВОГО ПОРЯДКА

А.В. Кальманов

Целью данной работы является исследование возможности оценивания параметров речевого тракта по акустическому сигналу для вокализованных звуков речи в классе моделей авторегрессии со стационарной сезонной разностью первого порядка и стационарной разностью первого порядка.

Как известно, при анализе речевых сигналов возникает задача нахождения параметров речевого тракта. При ее решении предполагают, что процесс речеобразования описывается некоторой физической моделью, известной приближенно, а исследователь наблюдает сигнал только на выходе этой модели. Естественный способ решения такой задачи - построение математической модели по известному сигналу с последующим установлением соответствия между построенной и физической моделями. После установления такого соответствия выводы о физической модели можно строить в терминах математической.

В[I] показано, что модель авторегрессии неадекватно описывает голосовые звуки речи при асинхронном анализе, что говорит о невозможности получения точного описания речеобразования в рамках такой модели и необходимости построения другой модели. Данную работу следует рассматривать как попытку построения новой модели и установления соответствия между ее параметрами и параметрами физической модели.

§ I. Постановка задачи

Основные предположения (являющиеся физической моделью) о процессе речеобразования голосовых звуков речи не выходят за рамки общепринятой теории [2] и состоят в следующем.

I. Речевой сигнал является случайным процессом со спектром, ограниченным диапазоном частот $F_д/2$ так, что его можно квантовать с частотой $F_д$ и задавать в виде временного ряда $\{a_n\}$.

2. Речь образуется в результате свертки функции возбуждения y_n с импульсной реакцией h_n речевого тракта. Амплитудно-частотная характеристика излучателя ($20 \log|H_{изл.}|$) имеет наклон +6 дб/октаву.

3. В первом приближении речевой тракт можно рассматривать как линейную динамическую многорезонансную систему конечного порядка p_T с сосредоточенными параметрами. Передаточная функция речевого тракта содержит только полюса

$$H_{pT}(z^{-1}) = \frac{1}{E(z^{-1})} = \frac{1}{1 - \sum_{i=1}^{p_T} e_i z^{-i}}, \qquad (I)$$

а каждый отсчет h_n может быть предсказан по p_T предшествующим отсчетам с ошибкой $\varepsilon \sim N[0, \sigma_\varepsilon]$ так, что

$$h_n = \sum_{i=1}^{p_T} e_i h_{n-i} + \varepsilon_n. \qquad (2)$$

Каждой паре комплексно-сопряженных полюсов $H_{pT}(z^{-1})$ $z_j = \alpha_j \pm i\beta_j$, $j=1,2,\ldots,p_T/2$, соответствует форманта F_j, частота которой F_j и полоса ΔF_j задаются в виде:

$$F_j = \frac{1}{2\pi T} \operatorname{Im}[\ln z_j], \quad \Delta F_j = \frac{1}{\pi T} \operatorname{Re}[\ln z_j], \qquad (3)$$

где $T = 1/F_д$ – интервал квантования сигнала.

4. Функция возбуждения y_n для вокализованных участков речи представляет собой импульсный случайный процесс с детерминированным тактовым интервалом T_0 (периодом основного тона) так, что момент возникновения n-го импульса $t_n = nT_0 + \nu$, где ν случайная величина с математическим ожиданием $M[\nu] = 0$, $|\nu| = T_0/c$, $c \geq 2$, а $T_{min} \leq t_n - t_{n-1} \leq T_{max}$, где $T_{min} = \frac{c-1}{c} T_0$, $T_{max} = \frac{c+1}{c} T_0$. Форма импульсов

возбуждения известна приближенно и изменяется достаточно мало на коротких участках речи (интервалах анализа) длительностью $T_a = \lambda T_0$ при $2 \le \lambda \le 4$, а наклон спектра источника ($20 \log |H_{ги}|$) может изменяться от одного интервала анализа к другому в пределах от -5 дб/октаву до -15 дб/октаву.

Учитывая предположения, задачу нахождения параметров речевого тракта можно сформулировать следующим образом: по выборке $\{s_n\}$ объема $N = T_a F_д + 1$ определить параметры передаточной функции речевого тракта a_i, $i = 1,2,\ldots,p_т$, или формантные параметры $\langle F_j, \Delta F_j \rangle$, $j = 1,2,\ldots,p_т/2$.

§ 2. Решение задачи

Нетрудно убедиться в том, что задача относится к классу некорректно поставленных задач [3] и не может быть строго решена без учета информации об источнике возбуждения.

Задача определения параметров речевого тракта в классе моделей авторегрессии рассматривалась в работах [3-7]. Модель авторегрессии имеет вид [8]:

$$A(z^{-1})\, w_n = \alpha_n , \tag{4}$$

$$w_n = \nabla^d s_n , \tag{5}$$

$$A(z^{-1}) = 1 - \sum_{i=1}^{p} a_i z^{-i} \tag{6}$$

– стационарный оператор авторегрессии p-го порядка ($AP(p)$); $\nabla = 1 - z^{-1}$ – разностный оператор со сдвигом назад; z^{-1} – оператор сдвига назад такой, что $z^{-1} w_n = w_{n-1}$, $i = 1,2,\ldots,p$; $\{s_n\}$ – временной ряд; $\{\alpha_n\}$ – белый гауссовский шум с дисперсией σ_α^2 ; n – номер отсчета. При аппроксимации, как правило, используют $d = 0$ или $d = 1$, предполагая, что с учетом (1) и (2) уравнение речеобразования имеет вид

$$s_n = H(z^{-1}) \varepsilon_n = H_{ги}(z^{-1}) H_{рт}(z^{-1}) H_{изл}(z^{-1}) \varepsilon_n , \tag{7}$$

а оценку $H(z^{-1})$, учитывая (1) и (4)-(7), ищут в виде:

$$H_{ги}(z^{-1}) H_{изл}(z^{-1}) R^{-1}(z^{-1}) = \nabla^{-d} A^{-1}(z^{-1}) . \tag{8}$$

Отсюда следует, что если $H_{ги}(z^{-1}) H_{изл}(z^{-1}) \ne \nabla^{-d}$, то $p > p_т$. В самом деле, $H_{ги}(z^{-1})$ имеет нули, каждый из которых можно ап –

проксимировать бесконечным числом полюсов. Это следует из того, что процесс скользящего среднего первого порядка можно представить в виде процесса авторегрессии бесконечного порядка [8]. Поэтому, сравнивая левую и правую части (8), получаем $p > p_T$. Следовательно, $A(z^{-1}) = E(z^{-1})E_0(z^{-1})$, где $E_0(z^{-1})$ - стационарный оператор $AP(p-p_T)$. На практике полагают, что $p - p_T \le 5$. При этом по $A(z^{-1})$ нельзя найти $E(z^{-1})$, поскольку оценить можно только $A(z^{-1})$. Тем не менее, вычислив корни $z_{ai}=\alpha_{ai}\pm i\beta_{ai}$ ($i = 1,2, \ldots,p/2$) полинома $A(z^{-1})$, их частоты и полосы $\langle F_{ai}, \Delta F_{ai}\rangle$ в соответствии с (3), использовав априорную информацию о частотах и полосах $\langle F_j, \Delta F_j\rangle$ ($j = 1,2,\ldots,p_T/2$) формант для конкретных звуков, можно отождествить некоторые из вычисленных частот F_{ai} с формантными частотами F_j. Но строгий алгоритм отождествления построить достаточно трудно. К тому же при совпадении нулей $H_{гл}(z^{-1})$ с полюсами $H_{рт}(z^{-1})$ некоторые из формант могут быть утеряны.

Применение оператора ∇ к речевому сигналу приводит к подъему спектра его верхних частот и удалению из сигнала постоянной составляющей, что соответствует совместному учету наклона спектра $H_{гл}(z^{-1})$ и $H_{изл}(z^{-1})$, который предполагают равным -6 дб/октаву. В действительности этот наклон изменяется во времени, поэтому используют адаптивный оператор $\nabla_a = 1-az^{-1}$, где $a = R_1/R_0$, а R_0, R_1 - автоковариации ряда $\{s_n\}$ для задержек 0 и 1. Однако, как показывают проведенные эксперименты, этого недостаточно для исключения влияния источника возбуждения и излучения.

Для устранения этого недостатка оценивание параметров проводят на интервалах полного смыкания голосовых связок [3,4]. Поиск таких интервалов является совсем непростой задачей. Если к тому же учесть, что у ряда дикторов эти интервалы отсутствуют, то желаемый результат не будет достигнут.

В данной работе предлагается иной подход для устранения влияния источника возбуждения и излучения без поиска интервалов полного смыкания. А именно параметры речевого тракта предлагается оценивать в классе моделей авторегрессии со стационарной сезонной разностью первого порядка и стационарной разностью первого порядка

$$A(z^{-1})\nabla\nabla_k s_n = \alpha_n, \qquad (9)$$

где ∇_k - оператор сезонной разности первого порядка такой, что $\nabla_k s_n = s_n - s_{n-k} = (1-z^{-k})\, s_n$, т.е. оценку $H(z^{-1})$ предлагается искать в виде:

$$H_{\text{гп}}(z^{-1})H_{\text{изд}}(z^{-1})E^{-1}(z^{-1}) = \nabla^{-1}\nabla_k^{-1}\, A^{-1}(z^{-1}), \qquad (10)$$

где вместо ∇ возможно использование ∇_a. Из (10) следует, что при $A(z^{-1}) = E(z^{-1})$

$$H_{\text{гп}}(z^{-1})H_{\text{изд}}(z^{-1}) = \nabla^{-1}\,\nabla_k^{-1}\,. \qquad (11)$$

Наоборот, если (11) верно, то $A(z^{-1}) = E(z^{-1})$, и отождествлять корни $A(z^{-1})$ и $E(z^{-1})$ не нужно. Проведенные эксперименты показали, что (10) и (11) достаточно хорошо выполняются.

Нетрудно показать, что максимальное устранение влияния источника возбуждения достигается при $k = N_{\text{от}} = T_0 F_д + 1$ ($N_{\text{от}}$ - число отсчетов в тактовом интервале). Поэтому оценивание параметра k эквивалентно определению интервала T_0, который можно найти, например, методом обратной фильтрации [6].

Обозначим $w_n = \nabla\nabla_k s_n$, тогда (9) можно переписать в виде: $A(z^{-1})w_n = \alpha_n$. Оценивание параметров оператора $A(z^{-1})$ сводится к решению системы нормальных уравнений:

$$\sum_{i=1}^{p} a_i R_{j-i} = R_j, \quad 1 \le j \le p,$$

где

$$R_j = \sum_{n=1}^{N_1-j} w'_n w'_{n-j}, \quad N_1 = N-k-1, \; 0 \le j \le p,$$

$$w'_n = w_n \cdot v_n, \; v_n = 0{,}54 + 0{,}46\cos\left[\pi\frac{2(n-1)-N_1}{N_1}\right].$$

Для нахождения корней уравнения $A(z^{-1}) = 0$ первоначально использовались: процедура Мюллера, предложенная в [9] и переведенная на ФОРТРАН для ЭВМ "Минск-32", а также стандартная процедура Берстоу. Однако при степени полинома выше семи эти программы давали существенную ошибку. Поэтому была написана программа, реализующая метод Берстоу [10] с некоторыми модификациями, позволившая значительно уменьшить ошибку.

§ 3. Описание эксперимента и результаты

Оценивание формантных параметров проводилось при помощи шести моделей, которые можно условно обозначить: $A(z^{-1})s_n$, $A(z^{-1})\nabla s_n$, $A(z^{-1})\nabla_a s_n$, $A(z^{-1})\nabla_k s_n$, $A(z^{-1})\nabla\nabla_k s_n$, $A(z^{-1})\nabla_a\nabla_k s_n$ (название моделей ясно из обозначения). Первые три, как отмечалось, уже применялись для нахождения параметров речевого тракта, остальные – предлагаются в данной работе. Эксперименты проводились на синтезированном и реальном сигналах.

Исследовались два вида возбуждения: 1) синусоидальное, 2) возбуждение δ-импульсами. В первом случае импульсы голосовых связок моделировались в виде:

$$y_i=\begin{cases} c\sin\dfrac{\pi(i-1)}{2\alpha(N_{OT}-1)}, & i=1,2,\ldots,]\alpha N_{OT}[,\\ \dfrac{c}{2}\left\{1+\sin\dfrac{\pi}{2\beta}\left[\beta-\alpha+\dfrac{i-1}{N_{OT}-1}\right]\right\}, & i=]\alpha N_{OT}[+1,\ldots,](\alpha+2\beta)N_{OT}[,\\ 0, & i=](\alpha+2\beta)N_{OT}[+1,\ldots,N_{OT}. \end{cases}$$

Параметры α и β позволяют варьировать длительность интервалов открытия и закрытия голосовой щели и наклон спектра источника, который, например, при $\alpha = 0{,}23$ и $\beta = 0{,}25$ равен примерно -12 дб/октаву. При δ-возбуждении $y_i = 1$ для $i = 1$ и $y_i = 0$ для $i = 2,\ldots, N_{OT}$. Длина синтезируемого ряда $N =]\lambda N_{OT}[$ (λ – необязательно целое). К импульсам возбуждения добавлялся гауссовский белый шум: $\tilde{y}_n = y_n + \eta$, $n = 1,2,\ldots, N$, где $\eta \sim N(0,\sigma_\eta)$. В экспериментах $\sigma_\eta = 0{,}01$ при $c = 1$. Наконец, сигнал центрировался:

$$y_n^* = \tilde{y}_n - \frac{1}{N}\sum_{n=1}^{N}\tilde{y}_n.$$

Для моделирования речевого тракта использовалось пять формант. Каждое из пяти формантных колебаний задавалось моделью АР (2), которой соответствует фильтр с передаточной функцией

$$H_i(z^{-1}) = 1/(1-b_{1i}z^{-1}-b_{2i}z^{-2}).$$

Параметры b_{1i} и b_{2i} находятся по частоте F_i и полосе ΔF_i i-й форманты из соотношений: $b_{1i} = 2\rho_i\cos\omega_i T$, $b_{2i} = -2\rho_i^2$, $\omega_i = 2\pi F_i$, $\rho_i = e^{-\Delta F_i \pi T}$. Поэтому

$$H_{рт}(z^{-1}) = \prod_{i=1}^{5} H_i(z^{-1}) = 1/(1-\sum_{i=1}^{10} e_i z^{-i}) ,$$

а параметры e_i нетрудно найти из произведения пяти полиномов второго порядка. Таким образом, модель речевого тракта является системой десятого порядка.

Значения частот и полос формант, используемых при моделировании шести русских гласных, были взяты из [2]. Синтез сигнала осуществлялся по уравнению

$$s_n = \sum_{i=1}^{10} e_i s_{n-i} + y_n^e , \quad 1 \le n \le N. \qquad (12)$$

Для исследования надежности метода в условиях аналого-цифрового преобразования предусматривалась возможность добавления к s_n равномерного на интервале $(-\xi, \xi)$ шума u, т.е. $s_n^* = s_n + u_n$. Семиразрядный преобразователь перекрывает динамический диапазон примерно 42 дб, а его ошибка квантования $\xi = 2^{-6}$. Поэтому $\sigma_u = \xi/2\sqrt{3} = 0{,}002255$. При моделировании использовалось значение $\sigma_u = 0{,}003$.

Пусть F_{ij} – заданное при синтезе значение частоты i-й форманты для j-й гласной, $\hat{F}_{ij}$ – ее оценка, тогда $\delta_{ij} = |F_{ij} - \hat{F}_{ij}|/F_{ij}$ – относительная ошибка оценивания i-й форманты для j-й гласной, $\delta_{\cdot j} = \sum_{i=1}^{5} \delta_{ij}/5$ – средняя ошибка оценивания частоты форманты для каждой из шести гласных, $\delta_{i\cdot} = \sum_{j=1} \delta_{ij}/6$ – средняя ошибка оценивания каждой из пяти формантных частот, $\delta = \sum_{i=1}^{5} \sum_{j=1}^{6} \delta_{ij}/30$ – ошибка оценивания частоты форманты, усредненная по шести гласным.

Если $\{y_n\}$ – последовательность δ-импульсов, то $\{s_n\}$ – последовательность импульсных откликов модели речевого тракта, полученная из (12). Поэтому можно определить ошибку оценивания формантных частот по импульсному отклику в классе моделей АР(10). Очевидно, эта ошибка будет совпадать с ошибкой оценивания максимумов спектра сигнала $\{s_n\}$, содержащего только полюса. В табл. 1 и 2 приведены значения $\delta_{i\cdot}$, $\delta_{\cdot j}$ и δ для различных σ_u и σ_η при δ-возбуждении и порядке аппроксимирующей модели $p = 10$. Из таблиц видно, что при одном и том же σ_η и различных σ_u ошибки оценивания совпадают. Следовательно, ошибки квантования практически

Таблица I

Средняя ошибка (в %) оценивания частоты форманты для шести гласных в классе моделей $A(z^{-1})s_n$ при δ-возбуждении; p = I0

№ п/п	σ_μ	σ_η	|а| δ_1	|о| δ_2	|у| δ_3	|э| δ_4	|ы| δ_5	|и| δ_6	Средняя ошибка
I	0	0	0,84	I,62	2,74	0,7I	2,49	I,33	I,62
2	0	0,0I	0,9	I,66	3,0I	0,49	2,47	I,43	I,66
3	0,003	0	0,84	I,62	2,74	0,7I	2,49	I,33	I,62
4	0,003	0,0I	0,9	I,66	3,0I	0,49	2,47	I,43	I,66

Таблица 2

Средняя ошибка (в %) оценивания каждой из пяти формантных частот в классе моделей $A(z^{-1})s_n$ при δ-возбуждении; p = I0

№ п/п	σ_μ	σ_η	F_1 δ_1	F_2 δ_2	F_3 δ_3	F_4 δ_4	F_5 δ_5	Средняя ошибка
I	0	0	6,5	0,69	0,56	0,29	0,08	I,62
2	0	0,0I	6,63	0,95	0,46	0,I8	0,09	I,66
3	0,003	0	6,5	0,69	0,56	0,29	0,08	I,62
4	0,003	0,0I	6,63	0,95	0,46	0,I8	0,09	I,66

не влияют на ошибки оценивания. Средняя ошибка δ при $\delta_\eta = 0{,}01$ несколько больше, чем при $\delta_\eta = 0$. Это объясняется тем, что при $\delta_\eta = 0{,}01$ последовательность $\{s_n\}$ уже не является последовательностью импульсных откликов, поскольку $\{y_n\}$ – "зашумленные" δ-импульсы. В результате вместо формантных частот оцениваются максимумы спектра сигнала, что и приводит к увеличению ошибки, которая будет тем больше, чем больше сигнал возбуждения отличается от δ-импульса.

В табл. 3 и 4 приведены ошибки оценивания для шести моделей ($I - A(z^{-1})s_n$; $2 - A(z^{-1})\nabla s_n$; $3 - A(z^{-1})\nabla_a s_n$; $4 - A(z^{-1})\nabla_k s_n$; $5 - A(z^{-1})\nabla\nabla_k s_n$; $6 - A(z^{-1})\nabla_a\nabla_k s_n$) при синусоидальном возбуждении и различных σ_μ и σ_η (p = I0). Из таблиц видно, что ошибки квантования практически не влияют на ошибки оценивания. Значения ошибок δ примерно одинаковы для одного и того же класса моделей при различных операторах ∇ и ∇_a. Это объясняется тем, что для гласных звуков $\nabla \approx \nabla_a$, так как $a \approx 1$.

Таблица 3

Средняя ошибка (в %) оценивания частоты форманты для шести русских гласных в классе моделей: $A(z^{-1})s_n(I)$;

$A(z^{-1})\nabla s_n(2)$; $A(z^{-1})\nabla_a s_n(3)$; $A(z^{-1})\nabla_k s_n(4)$;

$A(z^{-1})\nabla\nabla_k s_n(5)$; $A(z^{-1})\nabla_a\nabla_k s_n(6)$ при синусоидальном возбуждении; p = I0

σ_η	$\sigma_\varkappa$	Модель	|а| $\delta_{\cdot 1}$	|о| $\delta_{\cdot 2}$	|у| $\delta_{\cdot 3}$	|э| $\delta_{\cdot 4}$	|ы| $\delta_{\cdot 5}$	|и| $\delta_{\cdot 6}$	Среднее δ
0	0	I	26,59	26,46	33,02	29,30	33,46	26,I2	29,I6
		2	24,48	26,43	8,II	27,30	9,50	26,70	20,42
		3	24,33	27,96	8,08	27,3	9,5I	26,58	20,63
		4	25,66	25,49	4,9I	24,26	25,I7	23,08	2I,43
		5	0,46	I,3I	2,70	0,6I	I,49	I,7I	I,37
		6	0,97	I,23	2,57	0,70	I,48	I,82	I,46
0	0,003	I	26,59	26,48	33,I5	29,30	33,30	26,I2	29,I6
		2	24,52	26,43	8,09	27,28	9,9I	26,53	20,46
		3	24,38	26,80	8,08	27,3I	9,93	26,53	20,5I
		4	25,67	25,5I	4,24	24,26	25,23	23,07	2I,33
		5	0,45	I,3I	2,68	0,63	I,47	I,70	I,37
		6	0,97	I,23	2,57	0,70	I,47	I,83	I,46
0,0I	0	I	26,35	27,20	33,65	28,I5	30,53	25,34	28,54
		2	5,88	6,87	8,9I	27,46	7,88	26,73	I3,95
		3	6,72	7,27	8,88	27,58	7,89	26,74	I4,I8
		4	8,46	5,88	5,82	23,90	II,35	22,95	I3,06
		5	I,43	I,3I	3,I7	I,I3	0,85	I,82	I,62
		6	I,07	I,23	3,05	I,I9	0,85	2,09	I,58
0,0I	0,003	I	26,35	27,I3	33,72	28,I5	30,58	25,34	28,55
		2	5,89	6,9I	8,55	27,46	7,87	26,33	I3,95
		3	6,74	7,25	8,84	27,59	7,88	26,73	I4,I7
		4	8,45	5,90	5,82	23,90	II,37	22,94	I3,06
		5	I,44	I,3I	3,20	I,I4	0,79	I,82	I,62
		6	I,07	I,23	3,08	I,20	0,79	2,09	I,58

Средняя ошибка оценивания частоты форманты δ при различных $\sigma_\varkappa$ и σ_η составляет величину: в классе моделей AP(10) примерно 30%, в классе моделей AP(10) со стационарной разностью перво-

Таблица 4

Ошибка (в %) оценивания пяти формантных частот, усредненная по шести русским гласным, в классе моделей: $A(z^{-1})s_n(1)$; $A(z^{-1})\nabla s_n$ (2); $A(z^{-1})\nabla_a s_n$ (3); $A(z^{-1})\nabla_k s_n(4)$; $A(z^{-1})\nabla\nabla_k s_n(5)$; $A(z^{-1})\nabla_a\nabla_k s_n(6)$ при синусоидальном возбуждении; p = 10

σ_η	σ_u	Модель	F_1 / δ_1, %	F_2 / δ_2, %	F_3 / δ_3, %	F_4 / δ_4, %	F_5 / δ_5, %	Среднее / δ, %
0	0	1	18,75	15,72	6,34	100,00	4,97	29,16
		2	13,79	11,60	5,24	67,33	4,17	20,42
		3	15,93	10,64	5,2	67,32	4,16	20,63
		4	4,88	21,55	6,98	69,3	4,44	21,43
		5	2,18	1,20	1,12	0,87	1,46	1,37
		6	2,60	0,94	1,23	0,84	1,68	1,46
0	0,003	1	18,81	15,64	7,04	100,00	4.97	29,16
		2	13,79	11,47	5,27	67,33	4,44	20,46
		3	13,80	11,61	5,33	67,34	4,45	20,51
		4	4,88	21,54	7,10	68,74	4,40	21,33
		5	2,07	1,05	1,34	1,04	1,41	1,37
		6	2,62	0,86	1,38	1,07	1,35	1,46
0,01	0	1	17,64	13,92	7,60	100,00	3,53	28,54
		2	15,85	11,31	4,37	35,23	3,07	13,95
		3	16,44	11,62	4,65	35,12	3,08	14,18
		4	7,11	10,28	4,39	40,07	3,42	13.06
		5	2,19	1,27	1,17	2,46	1,02	1,62
		6	2,21	1,15	1,12	2,47	0,95	1,58
0,01	0,003	1	17,61	13,9	7,63	100,00	3,59	28,55
		2	15,85	11,31	4,33	35,20	3,08	13,95
		3	16,44	11,62	4,61	35,10	3,09	14,17
		4	7,26	10,5	4,34	40,04	3,83	13,06
		5	2,21	1,27	1,20	2,39	1,01	1,62
		6	2,18	1,15	1,10	2,50	0,95	1,58

го порядка 14–21%, в классе моделей AP(10) со стационарной сезонной разностью первого порядка 13–21%, а в классе моделей AP(10) со стационарной сезонной разностью первого порядка и стационар –

ной разностью первого порядка 1,4–1,7%. Взятие сезонной разности позволяет уменьшить ошибку в 1,4–2,3 раза. Примерно такой же результат дает взятие первой разности. Совместное же их использование понижает ошибку оценивания в 17,6–21,4 раза.

Из табл.4 видно, что ошибка оценивания частоты каждой форманты для всех гласных лежит в пределах 0,84–2,6% при использовании предлагаемых моделей. В то же время она может достигать 100% для других моделей, что говорит о потере форманты (в экспериментах при потере форманты ошибка полагалась равной 100%). Поэтому выполнение неравенства $\delta_{ij} \geq 20i$ для табл.3 означает утерю i форманты для j-й гласной. Аналогично выполнение неравенства $\delta_{ij} \geq 50j/3$ для табл.4 означает утерю i-й форманты для j гласных.

Табл. 5 и рис. 1 и 2 дают наглядное представление о том, чего позволяют достигнуть все шесть моделей при оценивании параметров авторегрессии (a_i) автокорреляций импульсного отклика ($r_i = R_i/R_0$) и логарифмической амплитудно-частотной характеристики модели речевого тракта для гласной /а/ . В первой строке табл.5 приведены заданные при синтезе параметры, во второй – их оценки, полученные по импульсному отклику (при δ-возбуждении). В строках 3–8 приведены оценки параметров для различных моделей (p = = 10) при синусоидальном возбуждении. Наилучшие оценки в последних двух строках.

Рис. 2 иллюстрирует влияние источника на оценивание частотной характеристики. Хорошо видно, что F4 и F5 различимы только для предлагаемых моделей, а сходство оценки частотной характеристики для этих моделей с частотной характеристикой, полученной по импульсному отклику (рис.1), очевидно.

В табл.6–9 приведены ошибки оценивания для различных моделей при p = 12,14. Во всех случаях ошибка оценивания в классе предлагаемых моделей меньше, чем в классе применявшихся ранее. Кроме того, ошибка оценивания в классе предлагаемых моделей при p = 10 меньше, чем ошибка оценивания в классе ранее использовавшихся при p = 12,14. Как уже упоминалось, при $p > R_T = 10$ возникает задача отождествления. Эта задача решалась вручную.

Описанный метод проверялся на реальном речевом сигнале. В ЭВМ "Минск-32" через семиразрядный аналого-цифровой преобразователь вводилась фраза "Вода в луже медленно убывала" при $F_д$=10кгц. Длительность интервала анализа T_a составила 25,6 мсек, а сдвиг

от одного интервала к другому был равен 16 мсек. Для примера на рис.3 приведена картинка видимой речи, полученная по тестовой фразе.

Параметр k находился при помощи метода обратной фильтрации. Значение частот полюсов аппроксимирующих моделей при p = 10,12, 14 выводилось в виде графика для получения траекторий формантных частот. На рис.4 приведены траектории пяти формантных частот, полученных по параметрам модели $A(z^{-1})\nabla\nabla_k s_n$ при p = 10. Их значения хорошо согласуются с данными, имеющимися в литературе не только для гласных, но и для звонких согласных. Для модели $A(z^{-1})\nabla_a\nabla_k$ результаты аналогичны.

Для p > 10 формантные частоты по полюсам моделей определяются неоднозначно (так как при $F_д$ = 10 кгц имеем τ_T = 1с [?]). Неоднозначность устраняется при p = 10, однако для моделей $A(z^{-1})s_n$, $A(z^{-1})\nabla s_n$, $A(z^{-1})\nabla_a s_n$, $A(z^{-1})\nabla_k s_n$, в отличие от предлагаемых, происходит утеря формант под влиянием источника возбуждения. Наибольшее число потерь формант происходит при использовании модели $A(z^{-1})s_n$. Модели $A(z^{-1})\nabla s_n$, $A(z^{-1})\nabla_a s_n$ и $A(z^{-1})\nabla_k s_n$ дают примерно одинаковые, но лучшие, чем модель $A(z^{-1})s_n$, результаты.

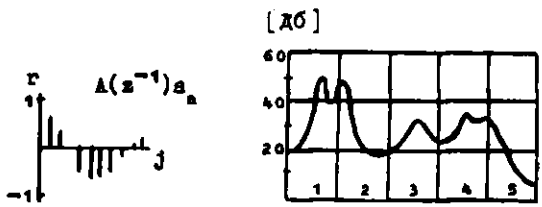

Рис.1. Автокорреляционная функция импульсного отклика и оценка логарифмической амплитудно-частотной характеристики модели речевого тракта, полученная при помощи модели $A(z^{-1})s_n$. Гласная /а/; p = 10; $\sigma_к$ = 0; $\sigma_п$ = 0; $\sigma_п$ = 0; δ-возбуждение.

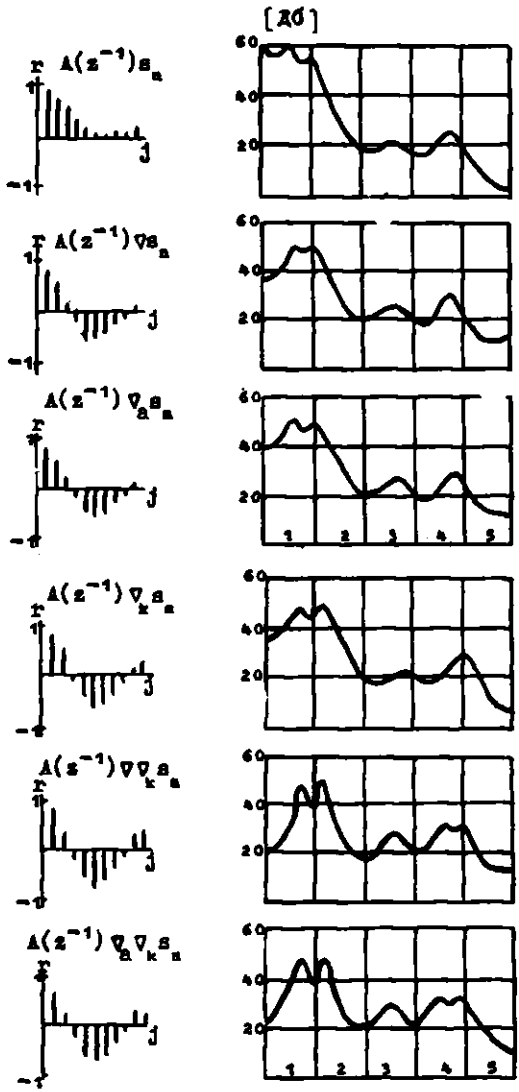

Рис.2. Автокорреляционные функции оценок импульсного отклика и оценки амплитудно-частотной характеристики речевого тракта, полученные при помощи шести моделей для гласной /а/; $p = 10$, $\sigma_и = 0$, $\sigma_\eta = 0$; синусоидальное возбуждение.

Таблица 5

Параметры авторегрессии, заданные при синтезе, и их оценки, полученные при помощи различных моделей при синусоидальном и δ-возбуждениях; $p = 10$, $\sigma_\eta = 0$, $\sigma_\mu = 0$

Модель	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	Вид возбуждения
Истинная, $B(z^{-1})$	0,7388	-0,4892	0,9084	-1,3633	0,6494	-1,1881	0,8081	-0,3273	0,3706	-0,6714	
$A(z^{-1})s_n$	0,7350	-0,5062	0,9017	-1,3876	0,6164	-1,1791	0,8190	-0,3434	0,3325	-0,6830	δ
$A(z^{-1})s_n$	2,7276	-3,0524	2,4581	-2,8246	3,1146	-2,5205	2,0774	-1,5590	0,6114	-0,0478	sin
$A(z^{-1})\nabla s_n$	1,5897	-0,8673	0,5382	-1,1036	0,7763	-0,4303	0,4777	-0,0848	0,3763	0,1669	
$A(z^{-1})\nabla_a s_n$	1,6233	-0,8785	0,5186	-1,0896	0,7791	-0,3967	0,4624	-0,0745	0,4070	0,2029	
$A(z^{-1})\nabla_k s_n$	1,7648	-1,2704	0,9956	-1,5926	1,3451	-0,8925	0,8545	-0,4621	-0,108	0,0907	
$A(z^{-1})\nabla\nabla_k s_n$	0,7264	-0,3745	0,7070	-1,2140	0,4846	-0,9642	0,5642	-0,2229	0,2613	-0,6168	
$A(z^{-1})\nabla_a\nabla_k s_n$	0,7755	-1,3761	0,7036	-1,2028	0,5073	-0,9172	0,6584	-0,2056	0,2409	-0,5690	

Таблица 6

Средняя ошибка (в %) оценивания частоты форманты для шести русских гласных в классе моделей: $A(z^{-1})s_n(1)$, $A(z^{-1})\nabla s_n(2)$, $A(z^{-1})\nabla_a s_n(3)$, $A(z^{-1})\nabla_k s_n(4)$, $A(z^{-1})\nabla\nabla_k s_n(5)$, $A(z^{-1})\nabla_a\nabla_k s_n(6)$ при синусоидальном возбуждении и различных значениях σ_η и σ_μ; $p = 12$

σ_η	σ_μ	Модель	/а/ $\delta_{\cdot 1}$	/о/ $\delta_{\cdot 2}$	/у/ $\delta_{\cdot 3}$	/э/ $\delta_{\cdot 4}$	/ы/ $\delta_{\cdot 5}$	/и/ $\delta_{\cdot 6}$	Среднее δ
0	0	1	0,81	3,61	12,26	1,27	6,04	10,62	5,77
		2	1,05	1,64	4,66	1,66	2,09	3,88	2,50
		3	1,18	1,78	4,56	1,71	2,09	3,88	2,53
		4	0,41	1,90	4,80	0,67	1,28	2,49	1,91
		5	0,92	1,41	2,94	1,12	0,8	1,44	1,44
		6	0,97	1,28	2,90	1,10	0,8	1,41	1,41
0	0,003	1	0,81	3,36	12,42	1,27	6,06	10,63	5,76
		2	1,05	1,63	4,51	1,66	2,08	3,88	2,49
		3	1,18	1,77	4,45	1,70	2,08	3,88	2,52
		4	0,41	1,87	4,74	0,66	1,24	2,49	1,91
		5	0,92	1,40	2,95	1,12	0,79	1,44	1,44
		6	0,96	1,29	2,93	1,10	0,79	1,41	1,41
0,01	0	1	1,54	4,23	13,45	1,69	5,48	10,20	6,10
		2	1,00	1,43	4,9	1,87	1,67	4,17	2,51
		3	1,03	1,45	4,78	1,88	1,66	4,17	2,50
		4	1,28	2,21	3,04	2,02	0,80	2,77	2,02
		5	1,74	2,11	2,43	0,99	1,11	1,66	1,67
		6	1,48	2,04	2,38	1,06	1,12	1,88	1,66
0,01	0,003	1	1,54	4,52	13,46	1,69	5,48	10,20	6,15
		2	1,00	1,43	4,90	1,87	1,66	4,17	2,51
		3	1,03	1,44	4,76	1,87	1,66	4,17	2,49
		4	1,28	2,22	3,10	2,03	0,81	2,76	2,03
		5	1,75	2,10	2,48	0,98	1,11	1,66	1,68
		6	1,50	2,03	2,45	1,06	1,11	1,88	1,67

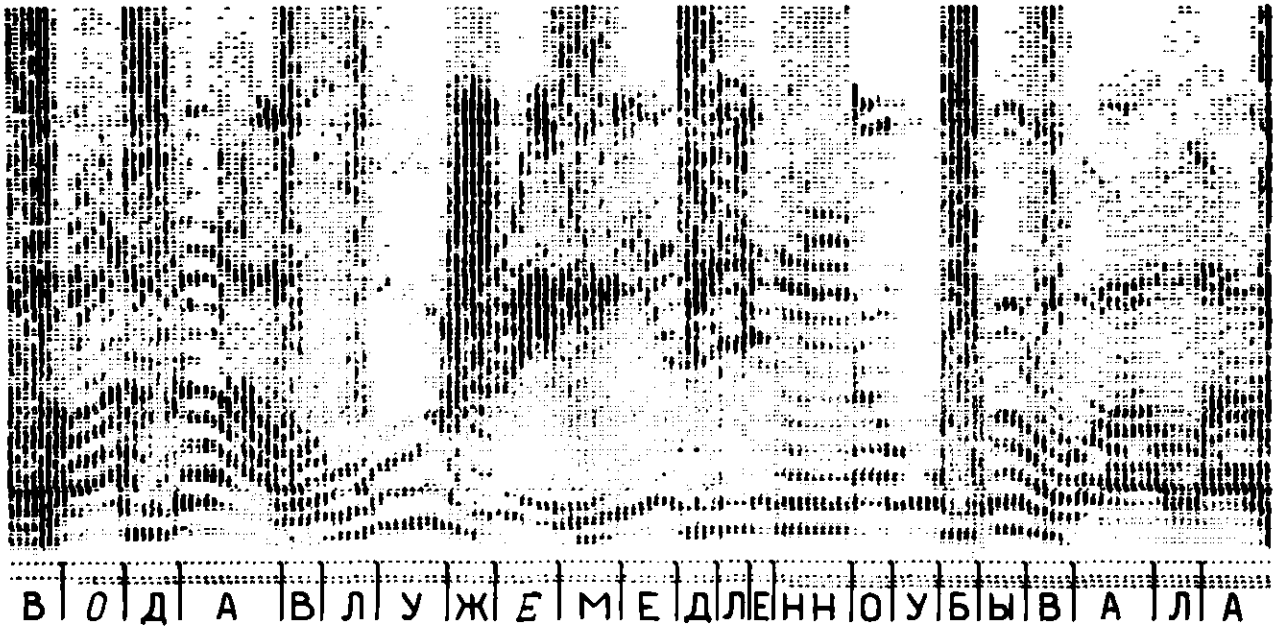

Рис.3. Видимая речь для фразы "Зода в луже медленно убывала".

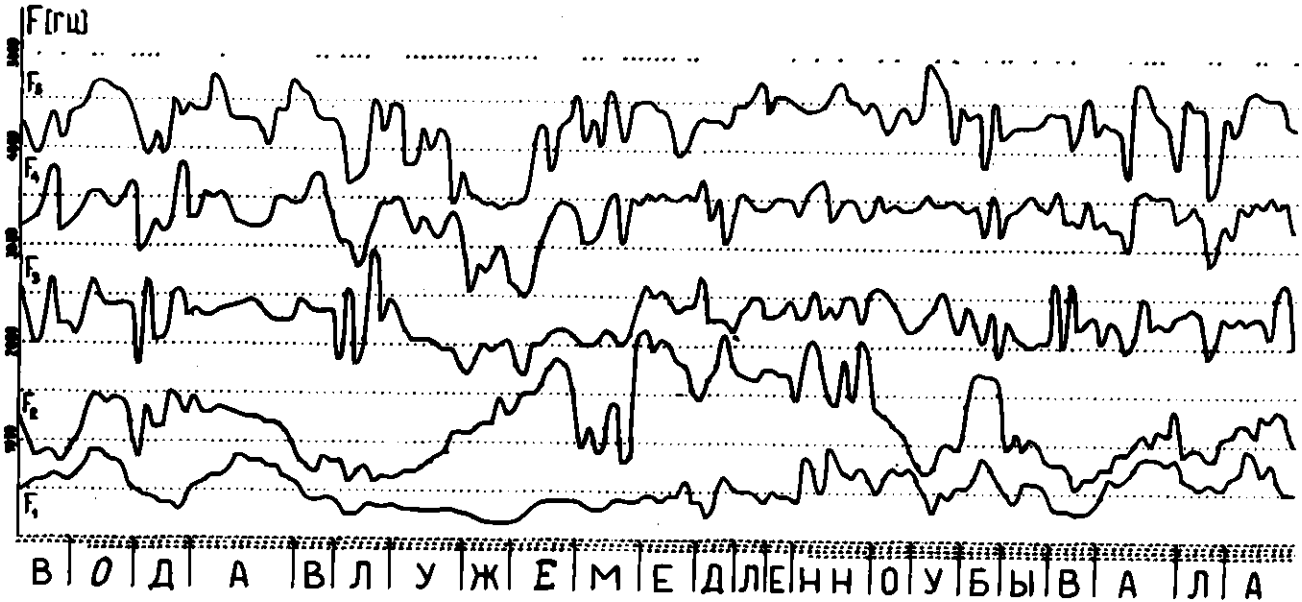

Рис.4. Траектории формантных частот, полученные по параметрам модели $A(z^{-1})vv_k s_n$ при p = IO.

Таблица 7

Ошибка (в %) оценивания формантных частот, усредненная по шести русским гласным, в классе моделей: $A(z^{-1})s_n(I)$; $A(z^{-1})\nabla s_n(2)$; $A(z^{-1})\nabla_a s_n(3)$; $A(z^{-1})\nabla_k s_n(4)$; $A(z^{-1})\nabla\nabla_k s_n(5)$; $A(z^{-1})\nabla_a\nabla_k s_n(6)$ при синусоидальном возбуждении и различных значениях σ_η и σ_k; при $p = 12$

σ_η	σ_k	Модель	F_1 / δ_1,	F_2 / δ_2,	F_3 / δ_3,	F_4 / δ_4,	F_5 / δ_5,	Среднее / δ
0	0	I	18,76	5,19	1,40	1,27	2,27	5,77
		2	5.94	3,50	1,20	0,99	0,86	2,50
		3	6,02	3,48	1,18	1,05	0,93	2,53
		4	1,85	2,60	2,67	1,46	1,06	1,91
		5	2,71	0,96	1,68	1,38	0,44	1,44
		6	2,63	0,93	1,63	1,41	0,45	1,41
0	0,003	I	18,76	5,16	1,38	1,25	2,24	5,76
		2	5,83	3,47	1,18	0,98	0,88	2,49
		3	5,96	3,45	1,14	1,05	0,94	2,52
		4	1,80	2,58	2,71	1,41	1,03	1,91
		5	2,71	0,96	1,71	1,37	0,43	1,44
		6	2,63	0,93	1,66	1,41	0,44	1,41
0,01	0	I	17,21	6,80	1,59	2,28	2,62	6,10
		2	6,15	3,77	1,48	0,50	0,63	2,51
		3	6,14	3,73	1,47	0,50	0,63	2,50
		4	2,07	1,95	3,22	2,27	0,59	2,02
		5	2,00	1,67	2,32	1,85	0,54	1,67
		6	1,98	1,71	2,22	1,76	0,62	1,66
0,01	0,003	I	17,21	6,95	1,62	2,30	2,67	6,17
		2	6,15	3,77	1,47	0,51	0,62	2,51
		3	6,14	3,70	1,46	0,50	0,64	2,49
		4	2,07	1,95	3,25	2,27	0,63	2,03
		5	2,00	1,65	2,35	1,85	0,56	1,68
		6	1,98	1,73	2,25	1,76	0,63	1,67

Таблица 8

Средняя ошибка (в %) оценивания частоты форманты для шести русских гласных в классе моделей: $A(z^{-1})s_n(1)$, $A(z^{-1})\nabla s_n(2)$, $A(z^{-1})\nabla_a s_n(3)$, $A(z^{-1})\nabla_k s_n(4)$, $A(z^{-1})\nabla\nabla_k s_n(5)$, $A(z^{-1})\nabla_a\nabla_k s_n(6)$ при синусоидальном возбуждении и различных значениях σ_η и σ_μ; $p = 14$.

σ_η	σ_μ	Модель	/а/ $\delta_{\cdot 1}$	/о/ $\delta_{\cdot 2}$	/у/ $\delta_{\cdot 3}$	/э/ $\delta_{\cdot 4}$	/ы/ $\delta_{\cdot 5}$	/и/ $\delta_{\cdot 6}$	Среднее δ
0	0	1	0,92	1,44	12,91	2,08	6,09	5,11	4,76
		2	2,02	1,33	3,96	2,98	0,83	4,19	2,55
		3	1,20	1,37	4,00	2,98	0,88	4,18	2,44
		4	1,70	1,30	1,83	1,23	0,86	0,87	1,30
		5	1,14	1,05	1,63	1,15	1,24	1,11	1,22
		6	1,21	1,04	1,65	1,09	1,24	1,18	1,23
0	0,003	1	0,93	1,56	12,94	2,08	6,09	5,10	4,78
		2	2,02	1,32	3,93	2,98	0,84	4,18	2,55
		3	1,20	1,39	3,93	2,98	0,88	4,18	2,43
		4	1,70	1,31	1,86	1,24	0,86	0,87	1,31
		5	1,13	1,04	1,62	1,16	1,24	1,11	1,22
		6	1,21	1,02	1,61	1,11	1,24	1,18	1,23
0,01	0	1	2,42	17,62	11,93	9,54	10,77	4,98	9,54
		2	1,43	1,49	5,36	2,99	0,98	3,99	2,71
		3	1,51	1,53	5,36	2,99	0,97	4,08	2,74
		4	2,18	1,73	1,32	1,54	1,19	1,88	1,64
		5	1,85	1,86	1,42	1,43	0,99	1,84	1,57
		6	2,06	1,86	1,42	1,44	0,99	2,01	1,63
0,01	0,003	1	1,33	17,62	11,95	9,33	10,77	4,98	9,33
		2	1,43	1,50	5,36	2,98	0,96	4,00	2,71
		3	1,51	1,53	5,36	2,99	0,97	4,08	2,74
		4	2,17	1,73	1,32	1,53	1,19	1,88	1,64
		5	1,85	1,86	1,42	1,41	0,99	1,84	1,56
		6	2,06	1,86	1,42	1,44	0,99	2,01	1,63

Таблица 9

Ошибка (в %) оценивания пяти формантных частот, усредненная по шести русским гласным, в классе моделей: $A(z^{-1})s_n(1)$; $A(z^{-1})\nabla s_n(2)$; $A(z^{-1})\nabla_g s_n(3)$; $A(z^{-1})\nabla_k s_n(4)$; $A(z^{-1})\nabla\nabla_k s_n(5)$; $A(z^{-1})\nabla_g\nabla_k s_n(6)$ при синусоидальном возбуждении и различных значениях σ_μ и σ_η; при p = 14.

σ_η	σ_μ	Модель	F_1 / δ_1,	F_2 / δ_2,	F_3 / δ_3,	F_4 / δ_4,	F_5 / δ_5,	Среднее δ
0	0	1	14,81	4,68	1,90	1,60	0,81	4,76
		2	6,39	4,01	0,46	1,00	0,90	2,55
		3	6,10	3,68	0,49	1,03	0,86	2,44
		4	1,89	1,96	0,82	0,91	0,91	1,3
		5	2,04	1,60	0,52	1,01	0,96	1,22
		6	2,12	1,60	0,51	0,97	0,96	1,23
0	0,003	1	14,81	4,67	1,96	1,66	0,83	4,78
		2	6,39	3,98	0,47	0,98	0,93	2,55
		3	6,04	3,68	0,50	1,00	0,87	2,43
		4	1,96	1,93	0,83	0,91	0,90	1,31
		5	2,04	1,60	0,50	0,97	0,97	1,22
		6	2,12	1,59	0,50	0,95	0,98	1,23
0,01	0	1	16,23	19,32	6,84	4,33	0,99	9,54
		2	6,61	4,13	1,16	0,81	0,83	2,71
		3	6,71	4,15	1,16	0,82	0,85	2,74
		4	1,33	2,34	1,54	1,82	1,18	1,64
		5	1,75	1,78	1,71	1,41	1,18	1,57
		6	2,04	1,87	1,76	1,34	1,15	1,63
0,01	0,003	1	15,17	19,28	6,81	4,39	0,99	9,33
		2	6,61	4,12	1,15	0,82	0,83	2,71
		3	6,71	4,15	1,16	0,83	0,85	2,74
		4	1,33	2,34	1,54	1,81	1,18	1,64
		5	1,75	1,78	1,70	1,40	1,18	1,56
		6	2,04	1,87	1,76	1,32	1,16	1,63

Таким образом,

I. оценивание формантных частот в классе моделей авторегрессии со стационарной сезонной разностью первого порядка и стационарной разностью первого порядка позволяет по сравнению с оцениванием:

а) в классе моделей авторегрессии при одинаковом порядке моделей понизить ошибку оценивания в среднем в I8–2I раз при порядке, равном I0; в 4 раза, при порядке, равном I2; и в 4–6 раз при порядке, равном I4;

б) в классе моделей авторегрессии со стационарной разностью первого порядка при одинаковом порядке моделей понизить ошибку оценивания в среднем в 9–I5 раз при порядке, равном I0, и в 2 раза при порядке, равном I2 и I4;

в) в классах моделей авторегрессии и авторегрессии со стационарной разностью первого порядка при одинаковой ошибке получить описание, экономичнее более чем в I,4 раза, и избежать решения задачи отождествления полюсов модели с полюсами передаточной функции речевого тракта.

2. Ошибка квантования семиразрядного аналого-цифрового преобразователя практически не влияет на ошибку оценивания формантных частот.

3. Для выделения параметров речевого тракта рекомендуется использовать модель десятого порядка при частоте квантования сигнала I0 кгц.

4. Предложенный метод позволяет определять резонансные частоты речевого тракта с ошибкой, равной примерно I,4–I,7%.

Литература

I. КЕЛЬМАНОВ А.В. Алгоритм классификации тон/шум, основанный на критерии адекватности модели авторегрессии. – В кн.: Методы обработки информации. (Вычислительные системы, вып.74.) Новосибирск, I978, с.I29–I48.

2. ФАНТ Г. Акустическая теория речеобразования. М., "Наука", I964.

3. ЛОБОВСКИЙ В.С. Аппроксимация отклика системы в z-плоскости и формантный анализ речи. – В кн.: Вычислительные системы, Вып.37. Новосибирск, I969, с.22–37.

4. СОБАКИН А.Н. Об определении формантных параметров голосового тракта по речевому сигналу с помощью ЦВМ. – "Акустический журнал", I972, т.XVII, вып.I, с. I06–II4.

5. ATAL B.S., HANAUER S.L. Speech analysis and synthesis by linear prediction of the speech wave. – "J.Acoust.Soc.Amer.", 1971, v.50, N 2, pt 2, p.637–655.

6. MARKEL J.D. Application of digital invers filter for automatic formant and F_0 analysis.-"IEEE Trans.Audio Electroacoust.", 1973, v.AU-21, N 3, p.154-160.

7. WAKITA H. Direct estimation of the vocal tract shape by invers filtering of acoustic speech waveforms. - "IEEE Trans.Audio Electroacoust.",1973,v.AU-21, N 5,p.417-427.

8. БОКС Дж., ДЖЕНКИНС Г. Анализ временных рядов, прогноз и управление. Том I. М.,"Мир",1974.

9. ДОЗОВСКИЙ В.С. Процедура решения полиномиальных уравнений методом Мюллера. - В кн.: Вычислительные системы. Вып.44. Новосибирск,1971, с. 155-161.

10. КУНЦ К.С. Численный анализ. Киев, "Техника",1964.

Поступила в ред.-изд.отд.
14 апреля 1978 года