



**СТАТИСТИЧЕСКИЙ СИНТЕЗ
АЛГОРИТМОВ ОЦЕНИВАНИЯ
ПЕРИОДА ОСНОВНОГО ТОНА
РЕЧЕВЫХ СИГНАЛОВ**

ОМЕЛЬЧЕНКО А. В., ПРЕСНЯКОВ А. И.

Разрабатывается математическая модель речевого сигнала в виде процесса авторегрессии с нестационарным порождающим процессом и на ее основе выполняется синтез алгоритмов оценивания периода основного тона речи, учитывающих динамику его изменения. Для синтезированного алгоритма исследуются свойства оценок периода основного тона.

Одной из основных задач обработки речи является оценивание частоты основного тона (ОТ). Существует много алгоритмов оценивания частоты (периода) ОТ, каждый из которых обладает определенными достоинствами и недостатками [1-3]. Несмотря на большое количество работ по данной проблеме, отыскание эффективных ее решений является актуальной задачей.

Настоящая статья посвящена статистическому синтезу алгоритмов оценивания периода ОТ речевых сигналов. Задача решается в два этапа: на первом этапе синтезируются алгоритмы оценивания моментов возникновения голосовых импульсов, а на втором – алгоритмы калмановской фильтрации периода ОТ.

1. Модель речевого сигнала

В большинстве алгоритмов оценивания периода ОТ используется свойство периодичности вокализованных участков речи на коротких временных интервалах длительностью $T \leq 0,02$ с. При обработке речевых сигналов широко используются алгоритмы оценивания периода ОТ по автокорреляционной функции сигналов [1]. Недостатком автокорреляционного алгоритма, а также большинства других известных алгоритмов является наличие аномальных ошибок оценивания, обусловленных особенностями "тонкой структуры" речевых сигналов и изменением периода ОТ на интервале анализа.

Цель настоящей работы – разработка алгоритмов оценивания периода ОТ, которые учитывали бы динамику его изменения (мелодику речи).

В основе большинства методов обработки речи лежит модель речеобразующей системы в виде совокупности генератора возбуждения и голосового тракта, которые рассматриваются независимо друг от друга. Генератор формирует возбуждающий сигнал либо в виде последовательности импульсов для вокализованных звуков, либо в форме шумоподобного процесса для невокализованных звуков. Голо-

совой тракт рассматривается как линейная динамическая система. Его передаточная функция характеризуется резонансными (формантными) частотами. Для формирования звуков основную роль играют первые три форманты, которые лежат в диапазоне от 200 до 4000 Гц. В процессе речи все формантные частоты непрерывно изменяются в соответствии с особенностями произносимых звуков.

Опишем математическую модель речевых сигналов, выбранную для решения задачи оценивания периода ОТ речи.

Будем полагать, что речевой сигнал дискретизирован в соответствии с теоремой Котельникова и представлен своими отсчетами, взятыми с интервалом Δt . Рассматривая речевой тракт как линейный фильтр, запишем сигнал на его выходе в виде стохастического уравнения [1]

$$x(\lambda) = \sum_{k=1}^p \alpha_k x(\lambda-k) + u(\lambda), \quad \lambda = 0, 1, 2, \dots, \quad (1)$$

где $u(\lambda)$ – сигнал возбуждения; $\alpha_k, k = \overline{1, p}$ – коэффициенты фильтра.

Следуя работе [3], на интервалах вокализованной речи возбуждающий сигнал $u(\lambda)$ будем рассматривать как нестационарный гауссовский белый шум с нулевым средним и дисперсией, изменяющейся по закону

$$M[u^2(\lambda)] = \sum_i a^2(\Delta t \lambda - \tau_i) + \sigma^2, \quad \lambda = 0, 1, 2, \dots, \quad (2)$$

здесь σ^2 – постоянная величина; $a^2(t)$ – функция времени, задающая изменение дисперсии возбуждения на периоде ОТ. Моменты возникновения импульсов возбуждения τ_i опишем уравнением

$$\tau_i = \tau_{i-1} + T_0 + T_{c_i}, \quad (3)$$

где T_0 – среднее значение периода ОТ; T_{c_i} – случайная составляющая периода ОТ.

Выполненное исследование речевых сигналов показало, что для функции возбуждения $a(t)$ в (2) может быть использована следующая аппроксимация:

$$a(t) = \begin{cases} A_0 \exp\left(-\frac{t}{t_0}\right), & \text{если } t \geq 0; \\ 0, & \text{если } t < 0. \end{cases} \quad (4)$$

Совокупность выражений (1) – (4) задает принятую в данной работе математическую модель речевого сигнала. В частном случае, когда в (3) случайная составляющая периода ОТ равна 0 (с вероятностью 1), сигнал $X(\lambda)$ будет принадлежать классу периодически коррелированных случайных процессов [3, 6].

2. Синтез алгоритма оценивания моментов возникновения импульсов голосового возбуждения

Сформулируем задачу оценивания моментов возникновения импульсов голосового возбуждения.

Будем полагать, что наблюдаются отсчеты речевого сигнала $x(\lambda)$, $\lambda = \overline{0, N-1}$, взятые с интервалом дискретизации Δt , и для последовательности отсчетов сигнала $X(\lambda)$ справедлива модель, задаваемая выражениями (1), (2). Будем также считать, что все параметры в выражениях (1), (2) априори известны,

за исключением моментов возникновения импульсов голосового возбуждения τ_i , $i = \overline{1, k}$. Импульсы возбуждения будем считать неизвестными величинами, удовлетворяющими условиям

$$\tau_i = v_i \Delta t; v_i - v_{i-1} > L_m; v_i \in \{0, 1, \dots, N-1\}, \quad (5)$$

где натуральное число $L_m \gg t_0 / \Delta t$.

Необходимо по наблюдаемой последовательности отсчетов речевого сигнала оценить количество импульсов голосового возбуждения k и моменты их возникновения $\{\tau_i = v_i \Delta t, i = \overline{1, k}\}$.

Сформулированная задача относится к классу задач разрешения [4] и состоит в определении числа компонент сигнала в смеси с одновременным оцениванием их параметров.

Для решения поставленной задачи воспользуемся методом максимального правдоподобия (МП). В работе [4] показано, что алгоритм МП разрешения сигналов имеет вид

$$\left(\hat{k}, \hat{v}_1, \dots, \hat{v}_k \right) = \arg \max_{(k, v_1, \dots, v_k)} \left[\lambda \left(\hat{x} / v_1, \dots, v_k \right) \right], \quad (6)$$

где $\lambda \left(\hat{x} / v_1, \dots, v_k \right)$ – логарифм функции отношения правдоподобия наблюдаемой последовательности отсчетов сигнала; $\hat{x} = (x(0), \dots, x(N-1))$ – вектор временных отсчетов сигнала.

Несложно показать, что для выбранной модели речевых сигналов (1), (2) логарифм функции отношения правдоподобия с точностью до несущественных слагаемых равен

$$\lambda \left(\hat{x} / v_1, \dots, v_k \right) = \frac{1}{2} \sum_{\lambda=p}^{N-1} \left[\frac{\zeta^2(\lambda)}{\sigma^2} - \frac{\zeta^2(\lambda)}{\sum_{i=1}^k a^2((\lambda - v_i)\Delta t) + \sigma^2} \right] + \frac{1}{2} \sum_{\lambda=p}^{N-1} \left[\ln \sigma^2 - \ln \left[\sum_{i=1}^k a^2((\lambda - v_i)\Delta t) + \sigma^2 \right] \right], \quad (7)$$

где

$$\zeta(\lambda) = x(\lambda) - \sum_{s=1}^p \alpha_s x(\lambda - s), \quad \lambda = \overline{p, N-1} \quad (8)$$

сигнал на выходе обеляющего фильтра.

В выражениях (7), (8) использованы обозначения, введенные в разделе 1: $a^2(\lambda \Delta t)$ – отсчеты функции, задающей изменение дисперсии на периоде ОТ; σ^2 – постоянная составляющая дисперсии ОТ в выражении (2).

Учитывая условие (5), приводим выражение (7) к виду

$$\lambda \left(\hat{x} / v_1, \dots, v_k \right) = \frac{1}{2} \sum_{i=1}^k \left\{ \sum_{\lambda=v_i}^{v_i + L_m} \zeta^2(\lambda) b(\lambda - v_i) - \sigma^2 d \right\}, \quad (9)$$

где

$$b(\lambda) = \begin{cases} a^2(\lambda \Delta t) & 0 \leq \lambda \leq L_m; \\ 0, & \lambda > L_m \end{cases} \quad (10)$$

$$d = \sum_{\lambda=0}^{L_m} \ln \left[1 + \frac{a^2(\lambda \Delta t)}{\sigma^2} \right]. \quad (11)$$

Подставив выражение (9) в алгоритм (6), получим следующую модификацию алгоритма МП:

$$\left(\hat{k}, \hat{v}_1, \dots, \hat{v}_k \right) = \arg \max_{(k, v_1, \dots, v_k)} \left\{ \sum_{i=1}^k S(v_i) \right\}, \quad (12)$$

где решающая функция

$$S(v) = \sum_{\lambda=v}^{v+L_m} \zeta^2(\lambda) b(\lambda - v) - \sigma^2 d, \quad v = \overline{p, N-1}, \quad (13)$$

а максимум в (12) ищется среди совокупности v_1, \dots, v_k , удовлетворяющей условию $|v_i - v_j| > L_m$. Поиск максимума в (12) может быть выполнен методами целочисленного программирования.

Для уменьшения вычислительных затрат вместо алгоритма МП (12) можно использовать следующий итеративный алгоритм:

$$\hat{v}_i = \arg \max_{v_i \in R_i} \left\{ \sum_{i=1}^k S(v_i) \right\}, \quad i = 1, 2, \dots, \quad (14)$$

здесь функция $S(v)$ определяется согласно (13);

$$R_0 = \{v \mid S(v) > 0\}; R_i = R_{i+1} \setminus \{\hat{v}_{i-1}, \dots, \hat{v}_{i-1} + L_m\}. \quad (15)$$

Итерации прекращаются, если $R_i = \emptyset$, где \emptyset – пустое множество.

Таким образом, синтезированный алгоритм оценивания моментов возникновения импульсов гортанного возбуждения включает шаги: обеление речевого сигнала в соответствии с выражением (8); вычисление значений решающей функции согласно (13); нахождение оценок моментов возникновения импульсов голосового возбуждения и их общего количества в соответствии с итеративной процедурой (14), (15).

Для экспоненциальной модели изменения дисперсии возбуждающего сигнала (4) найдем аналитическое выражение для временного окна $b(\lambda)$ в алгоритме разрешения (12), (13). Используя (4) в выражении (10), получаем

$$b(\lambda) = \begin{cases} \frac{1}{1 + q^{-1} \exp\left(\frac{2\lambda \Delta t}{t_0}\right)}, & 0 \leq \lambda \leq L_m; \\ 0, & \lambda \geq L_m \end{cases} \quad (16)$$

где $q = A_0^2 / \sigma^2$ – коэффициент, характеризующий изменение мощности речевого сигнала на периоде ОТ.

Форма функций $b(\lambda)$ для случая $t_0 = 10\Delta t$ показана на рис. 1. Здесь значение параметра q приведено над каждым из графиков.

3. Синтез алгоритма фильтрации периода основного тона

Алгоритм разрешения импульсных сигналов (14), (15) позволяет по реализации смеси речевого сигнала и помехи найти оценки максимального правдоподобия моментов возникновения импульсов гортанного возбуждения:

$$\left(\hat{t}_1, \hat{t}_2, \dots, \hat{t}_k \right), \quad \text{где } \hat{t}_i = \hat{v}_i \Delta t, \quad i = \overline{1, k}.$$

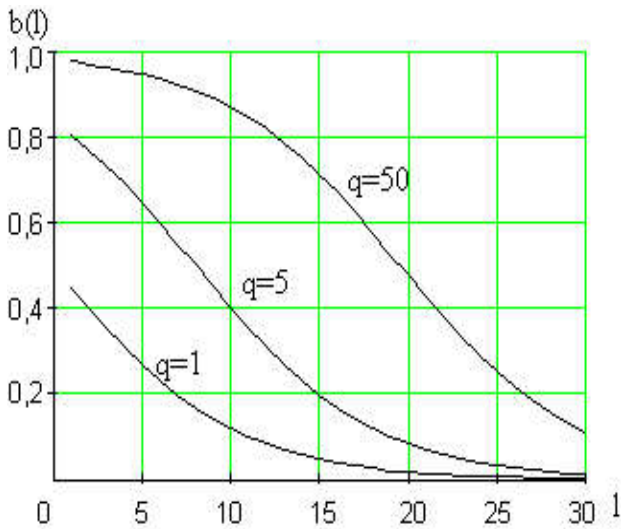


Рис. 1. Форма функций временного окна

Для оценивания текущего значения периода ОТ, которое мы определим как $T_i = t_i - t_{i-1}$, можно воспользоваться алгоритмом [3]

$$\bar{T}_i = \bar{t}_i - \bar{t}_{i-1}, \quad i = \overline{2, \bar{k}}. \quad (17)$$

В алгоритме (17) информация о динамике оцениваемого параметра не используется. Для устранения этого недостатка введем дополнительные предположения, отражающие свойства речевых сигналов.

1. На временных интервалах длительностью $T_c \leq 0,02$ с (участках квазистационарности) будем полагать, что период ОТ неизменен.

2. От одного участка квазистационарности к другому текущее значение периода ОТ изменяется по закону

$$T_j = T_0 + T_{c_j}, \quad j = 1, 2, \dots, \quad (18)$$

где T_0 – среднее значение периода ОТ;

$$T_{c_j} = \beta T_{c_{j-1}} + n_{0j}, \quad j = 1, 2, \dots \quad (19)$$

– процесс авторегрессии, описывающий изменение случайной составляющей периода ОТ; β – коэффициент авторегрессии; n_{0j} – порождающий процесс в виде гауссовского белого шума с дисперсией σ_0^2 .

Предположение 1 позволяет весь интервал наблюдения разбить на участки квазистационарности, состоящие из $L = [T_c / \Delta t]$ временных отсчетов, где $[x]$ – целая часть числа $x \in \mathbb{R}$. Каждому из таких участков квазистационарности поставим в соответствие оценку периода основного тона:

$$\bar{T}_j = \frac{\max_{\bar{v}_i \in Z_j} \bar{v}_i - \min_{\bar{v}_i \in Z_j} \bar{v}_i}{\bar{k}_j - 1} \Delta t, \quad j = 1, 2, \dots, \quad (20)$$

где $Z_j = \{L(j-1)+1, \dots, Lj\}$ – множество отсчетов j -го интервала квазистационарности; \bar{k}_j – количество моментов возникновения импульсов, найденных согласно (14), (15) и принимающих значения из множества Z_j .

Сформулируем постановку задачи фильтрации последовательности оценок частоты основного тона, найденных согласно (14), (15), (20), с учетом динамики изменения периода ОТ.

Будем полагать, что наблюдается последовательность оценок частоты ОТ:

$$\bar{T}_j = h_j T_{c_j} + T_0 + n_j, \quad (21)$$

где T_{c_j} – истинное значение случайной составляющей ОТ на j -м временном интервале; n_j – погрешность оценивания частоты ОТ, имеющая нормальный закон распределения с нулевым средним и дисперсией σ^2 ; $h_j \in \{0, 1\}$ – известная функция временного параметра j , принимающая нулевые значения на участках молчания, а также участках невокализованной речи.

Выражение (21) задает уравнение наблюдения и может быть представлено в следующем виде:

$$\bar{T}_j - T_0 = h_j T_{c_j} + n_j, \quad j = 1, 2, \dots \quad (22)$$

Уравнение состояния описывается процессом авторегрессии (19).

Необходимо синтезировать алгоритм калмановской фильтрации периода ОТ, оптимальный по критерию максимума апостериорной вероятности.

Применяя известную методику синтеза калмановских фильтров [5] к уравнениям (19) и (22), получаем алгоритм фильтрации периода ОТ:

$$\bar{T}_{c_j} = \beta \bar{T}_{c_{j-1}} + h_j \left(\frac{R_j}{\sigma^2} \right) (\bar{T}_j - T_0 - \beta \bar{T}_{c_{j-1}}), \quad (23)$$

$$R_j = \left[\frac{1}{\beta^2 R_{j-1} + \sigma_0^2} + \frac{h_j}{\sigma^2} \right]^{-1} \quad (24)$$

с начальными условиями

$$\bar{T}_{c_0} = 0, \quad R_0 = \frac{\sigma_0^2}{1 - \beta^2}. \quad (25)$$

Уравнения (23), (24) задают рекуррентную процедуру вычисления оценок случайной составляющей периода ОТ \bar{T}_{c_j} и ее апостериорной дисперсии R_j , $j = 1, 2, \dots$. Полное значение оценки периода ОТ равно

$$\bar{T}_{n_j} = \bar{T}_{c_j} + T_0, \quad j = 1, 2, \dots, \quad (26)$$

В уравнении наблюдения (21) не отражены отличия статистических характеристик оценок T_j при

$h_j = 0$, т.е. для интервалов молчания или невокализованной речи, от характеристик на интервалах вокализованного возбуждения при $h_j = 1$. Однако такое отличие несущественно, поскольку синтезированный в рамках принятых допущений алгоритм фильтрации (23), (24) учитывает значения оценок периода ОТ лишь на участках вокализованной речи.

Фильтр (23), (24) является нестационарным с изменяющейся во времени апостериорной дисперсией R_j . Если в течение длительного промежутка времени $h_j = 0$, то $R_j \rightarrow R_0$, если же в течение

длительного времени $h_j = 1$, то $R_j \rightarrow R_\infty$, где R_∞ – апостериорная дисперсия оценок для стационарного фильтра, удовлетворяющая уравнению

$$\frac{1}{R_\infty} = \frac{1}{\beta^2 R_\infty + \sigma_0^2} + \frac{1}{\sigma^2}. \quad (27)$$

Очевидно, что $R_\infty < R_0$. В нестационарном фильтре (23), (24) $R_\infty \leq R_j \leq R_0$, причем $R_j > R_{j-1}$ при $h_j = 0$ и $R_j < R_{j-1}$ при $h_j = 1$.

Динамика изменения апостериорной дисперсии оценок периода основного тона согласно уравнению (24) показана на рис. 2-4. На этих рисунках по оси абсцисс отложен номер участка квазистационарности речевого сигнала j , а по оси ординат – квадратный корень из апостериорной дисперсии оценок периода $OT - \sqrt{R_j}$. При этом продолжительность участка квазистационарности была принята равной 20 мс, а параметр уравнения наблюдения (21) h_j изменялся следующим образом: на интервале вокализованной речи $h_j = 1$ (при $j = \overline{1,20}$) и на интервале невокализованной речи $h_j = 0$ (при $j = \overline{21,40}$). Во всех случаях полагалось, что $\sqrt{R_0} = 0,6$ мс.

На рис. 2-4 отображены графики зависимости $\sqrt{R_j}$, $j = \overline{0,40}$ для различных значений параметра $\sigma' = \sigma/\sqrt{R_0}$: $\sigma' = 1$ (рис. 2); $\sigma' = 0,5$ (рис. 3); $\sigma' = 0,25$ (рис. 4), где σ – среднеквадратическая погрешность оценок периода OT на входе фильтра. На рис. 2-4 зависимость 1 соответствует значению параметра $\beta = 0,01$; 2 – $\beta = 0,9$; 3 – $\beta = 0,95$; 4 – $\beta = 0,98$.

Из анализа графиков, представленных на рис. 2-4, следует, что выигрыш в точности оценивания периода OT , получаемый в результате использования фильтра (23), (24), возрастает с увеличением параметра β и отношения $\sigma/\sqrt{R_0}$. При высокой точности предварительных оценок периода OT $\sigma \leq 0,25\sqrt{R_0}$ применение алгоритма калмановской фильтрации (23), (24) не дает ощутимого выигрыша в точности оценивания и может быть полезно лишь для интерполяции значений периода OT между участками вокализованной речи.

Таким образом, в настоящей работе предложена модель речевого сигнала в виде процесса авторегрессии с нестационарным порождающим процессом и на ее основе выполнен статистический синтез алгоритма оценивания периода OT , учитывающего динамику изменения периода голосовых импульсов. Синтезированный алгоритм оценивания периода OT состоит в последовательном применении алгоритма оценивания моментов возникновения импульсов голосового возбуждения (14), (15) и калмановской фильтрации периода OT в соответствии с выражениями (20), (23), (24). Использование такого алгоритма для решения задач верификации дикторов показало его высокую эффективность [7].

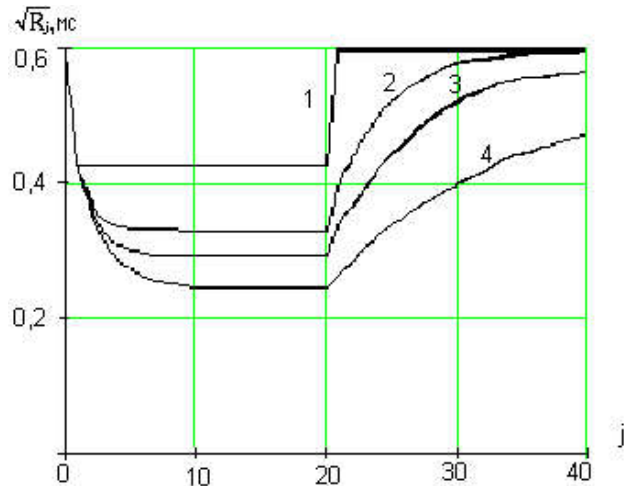


Рис. 2. Динамика изменения апостериорной дисперсии оценок периода OT для случая $\sigma' = 1$

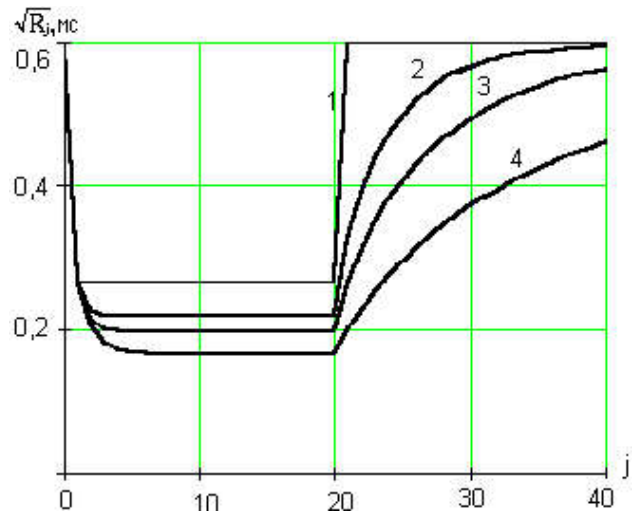


Рис. 3. Динамика изменения апостериорной дисперсии оценок периода OT для случая $\sigma' = 0,5$

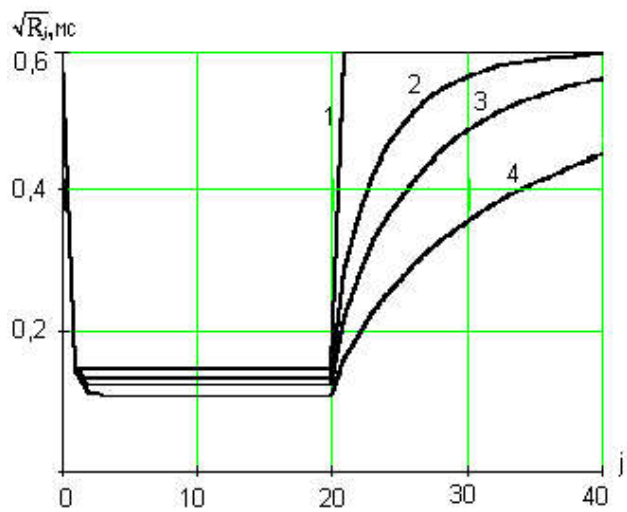


Рис. 4. Динамика изменения апостериорной дисперсии оценок периода OT для случая $\sigma' = 0,25$

Литература. 1. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов: Пер. с англ. / Под ред. М. В. Назарова и Ю. Н. Прохорова. М.: Радио и связь, 1981. 496 с. 2. Методы автоматического распознавания речи: В двух книгах: Пер. с англ. / Под ред. У. Ли. М.: Мир, 1983. Кн. 1. 328 с. 3. Андре-Обрект Р. Сегментация речевых сигналов в реальном времени без предварительного распознавания. В кн. "Обнаружение изменения свойств сигналов и динамических систем": Пер. с англ. / Под ред. М. Бассвилль, А. Баквениста. М.: Мир, 1989. 226–251 с. 4. Трифонов А. П., Шмаков Ю. С. Совместное различение сигналов и оценка их параметров на фоне помех. М.: Радио и связь, 1986. 264 с. 5. Тихонов В. И. Оптимальный прием сигналов. М.: Радио и связь, 1983. 320 с. 6. Драган Я. П., Приймак Н. В. Линейные периодически коррелированные случайные процессы. Львов: Изд. ФМИ, 1986. 30 с. 7. Омельченко А. В., Пресняков А. И. Алгоритмы верификации дикторов для дистанционного обучения // 36. наукових праць. Використання комп'ютерних технологій у навчальному процесі ХТУРЕ, Харків, 1998. С. 41–45.

Поступила в редколлегия 23.03.99

Рецензент: д-р техн. наук Руденко О. Г.

Омельченко Анатолий Васильевич, канд. техн. наук, доцент кафедры ПОС ХТУРЭ. Научные интересы: методы обработки сигналов и распознавание образов. Адрес: Украина, 310115, Харьков, ул. 17 Партсъезда, 8, кв. 33, тел. 40-94-29.

Пресняков Андрей Игоревич, ассистент кафедры ПОС ХТУРЭ. Адрес: Украина, 310202, Харьков, пр. Победы 54-б, кв. 44, тел. 40-94-29.