

4. Волчков В. П. Параметрическое спектральное оценивание случайных сигналов с использованием рекуррентных  $m$ -моделей // Радиотехника и электроника. - 1998. - Т. 43, 4. - С. 421--437.
5. Робинзон, Гренджер. Расчет нерекурсивных цифровых фильтров Уолша // Зарубежная радиоэлектроника. - 1973. - 4 - С. 12.

## SYNTHESIS OF THE RECURRENCE M-FILTERS WITH A GIVEN AMPLITUDE FREQUENCY CHARACTERISTIC

*V.P. Volchkov*

The approach of synthesis of fast digital  $m$ -filters with a given Amplitude Frequency Characteristics is proposed. This method we can regarded of further development and generalization FIC filtration when impulse response is invariant over  $m$ -shifts group and interval of registration is more then interval of  $m$ - representation.

УДК 621.391

## ЧАСТОТНЫЙ АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ

*Е.Г. Жияков<sup>1</sup>, Е.И. Прохоренко<sup>1</sup>*

1 - Белгородский государственный университет Российская Федерация, 308015, г. Белгород, ул. Победы, 85

Рассматривается новый оптимальный по критерию среднеквадратической погрешности метод частотного анализа и фильтрации, включая субполосный анализ, речевых сигналов.

### ВВЕДЕНИЕ

В задачах анализа звуковых сигналов принято использовать так называемую психоакустическую модель, которая описывает механизм восприятия звука человеческим ухом. Согласно этой модели реакция слуха избирательна к спектральному составу воздействующего сигнала, поэтому возникает целесообразность отдельной обработки данных, отражающих спектральные свойства сигналов в разных частотных полосах, в том числе при неравномерном разбиении на частотные интервалы, что отмечается, например, в рекомендациях MPEG [1].

В настоящее время для выделения составляющих сигнала, соответствующих определенным полосам частот, применяются фильтры с конечной импульсной характеристикой (КИХ-фильтры) [2]. Основным недостатком такого подхода обусловлен неконтролируемым эйлазингом (влияние наложения частот), который может существенно исказить спектр. Кроме того, для обеспечения возможности восстановления исходных данных на основе прореживания с использованием КИХ-фильтров допустимо только разбиение оси частот на равные полосы и необходимо выполнить определенные требования, которым должны удовлетворять формы АЧХ этих фильтров. Это может привести к еще большему эйлазингу, а промежуточные нелинейные преобразования, в частности, квантование по уровню, не могут быть скомпенсированы и оценка их влияния практически невозможна.

Поэтому представляется важной разработка и исследование оптимальных по критерию среднеквадратической погрешности методов частотного субполосного анализа речевых сигналов.

### 1. ВЫЧИСЛЕНИЕ ЭНЕРГЕТИЧЕСКОГО СПЕКТРА

Для анализа звуковых сигналов достаточно широкое распространение получил подход на основе так называемого кратковременного спектра [3].

Вычисление мгновенного энергетического спектра, зависящего от времени, позволяет получить представление о распределении энергии отрезка сигнала по частотным интервалам, что является важным, например, в задачах распознавания речи и уменьшения объема звуковых данных за счет удаления малоэнергетических частотных составляющих.

В настоящее время процедура анализа мгновенных спектральных характеристик реализована на основе алгоритма быстрого преобразования Фурье (БПФ) с равноразнесенными частотами анализа [4].

Нами предлагается метод, позволяющий вычислять точное распределение энергии, причем при любом желаемом разбиении на частотные интервалы (в том числе неравномерном).

Пусть компоненты вектора  $\vec{x} = (x_1, \dots, x_N)^T$  представляют собой значения некоторого сигнала (функции времени), которые соответствуют значениям аргумента  $i\Delta t$ , т.е.

$$x_i = x(i\Delta t), \quad i = 1, \dots, N,$$

где  $\Delta t$  - интервал дискретизации по времени.

Положим далее

$$X(\nu) = \sum_{k=1}^N x_k e^{-j(k-1)\nu}, \quad (1)$$

т.е.  $X(\nu)$  представляет собой трансформанту Фурье отрезка отсчетов сигнала, в качестве области определения которой естественно рассматривать

$$-\pi \leq \nu \leq \pi,$$

так что имеет место обратное преобразование

$$x_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\nu) e^{j(i-1)\nu} d\nu.$$

Отсюда не трудно получить равенство Парсеваля

$$\|\vec{x}\|^2 = \sum_{k=1}^N x_k^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\nu)|^2 d\nu,$$

так что

$$P_V(\vec{x}) = \frac{1}{2\pi} \int_{\nu \in V} |X(\nu)|^2 d\nu \quad (2)$$

представляет собой долю энергии отрезка сигнала, соответствующую частотному интервалу

$$V = [-\nu_2, -\nu_1) \cup [\nu_1, \nu_2).$$

Если в правую часть соотношения (2) подставить определение (1), то после очевидных преобразований [5] нетрудно получить представление

$$P_V(\vec{x}) = \vec{x}^T A_V \vec{x}, \quad (3)$$

где  $A_V$  - симметричная матрица

$$A_V = \{a_{ik}^V\};$$

$$a_{ik}^V = \begin{cases} \frac{\sin[\nu_2(i-k)] - \sin[\nu_1(i-k)]}{\pi(i-k)}, & i \neq k \\ \frac{\nu_2 - \nu_1}{\pi}, & i = k. \end{cases} \quad (4)$$

Таким образом, долю энергий отрезка сигнала в любом частотном интервале можно вычислить на основе представления (3), не вычисляя при этом, согласно (1), соответствующую трансформанту Фурье.

Согласно определения (2) правая часть (3) не может быть равна нулю ни в каком частотном интервале в виду того, что трансформанта Фурье вида (1), очевидно, является аналитической функцией, и поэтому не может быть тождественно равна нулю в любом интервале частот конечной длительности.

Таким образом, матрица с элементами вида (4) является положительно определенной и в виду симметрии обладает полным набором ортонормальных собственных векторов, так что [6] справедливо представление

$$A_V = \sum_{k=1}^N \lambda_k^V \bar{q}_k^V \bar{q}_k^{V^T}, \quad (5)$$

где  $\bar{q}_k^V$  - собственный вектор  $A_V$ , соответствующий собственному числу  $\lambda_k$ , т.е. имеет место равенство

$$\lambda_k^V \bar{q}_k^V = A_V \bar{q}_k^V, \quad k = 1, \dots, N. \quad (6)$$

В дальнейшем для простоты предполагается упорядоченность собственных чисел по убыванию, то есть

$$\lambda_1^V > \lambda_2^V > \lambda_3^V > \dots > \lambda_N^V > 0.$$

Используя определения (6) и (4), нетрудно получить соотношение

$$\lambda_k^V q_{rk}^V = \frac{1}{2\pi} \int_{v \in V} Q(v) e^{j(r-1)v} dv,$$

где  $Q_k$  – трансформанта Фурье соответствующего собственного вектора. Отсюда, имея в виду условие

$$(\bar{q}_k^V, \bar{q}_i^V) = \sum_{r=1}^N q_{ri}^V q_{rk}^V = \begin{cases} \sum_{r=1}^N q_{rk}^{V^2} = \|\bar{q}_k^V\|^2 = 1, & i = k \\ 0, & i \neq k \end{cases} \quad (7)$$

Нетрудно получить соотношения

$$\lambda_k^V = \frac{1}{2\pi} \int_{v \in V} |Q_k^V(v)|^2 dv = P_V(\bar{q}_k), \quad (8)$$

$$\int_{v \in V} Q_k^{*V}(v) Q_i^V(v) dv = \int_{-\pi}^{\pi} Q_k^{*V}(v) Q_i^V(v) dv = 0.$$

Таким образом, спектры собственных векторов обладают свойством двойной ортогональности (в заданном частотном интервале и во всей области определения).

Соотношение (8) показывает, что величина собственного числа равна доли энергии соответствующего собственного вектора, приходящейся на заданный интервал.

Поэтому, в виду первого из равенств (7) и аналитичности трансформант Фурье справедливо неравенство

$$\lambda_k^V \leq 1.$$

С другой стороны, собственные векторы матрицы  $A_V$  образуют полную систему, так что любой вектор  $\bar{y}$  размерности  $N$  может быть представлен в виде линейной комбинации

$$\bar{y} = \sum_{k=1}^N \alpha_k \bar{q}_k,$$

вектор коэффициентов которой

$$\bar{\alpha} = (\alpha_1, \dots, \alpha_N)^T$$

определяется однозначно.

Вычисления показывают, что некоторые из собственных чисел матрицы  $A_V$  могут быть близки к нулю.

Пусть в дальнейшем область определения спектра  $[-\pi, \pi]$  разбивается на ряд одинаковых частотных интервалов

$$V_i = [-\nu_{2i}, -\nu_{1i}) \cup [\nu_{1i}, \nu_{2i}) \quad (9)$$

таких что  $\nu_{2i} - \nu_{1i} = \Delta\nu = \text{const}$

$$R_i \frac{\nu_{2i} - \nu_{1i}}{\pi} = \text{const} = R, \quad (10)$$

причем  $R$  является целым числом  $M$ ;

$$\nu_{2i-1} = \nu_{1i} = (i-1)\Delta\nu, \quad i = 1, 2, \dots, R.$$

$$M = 2 \left\lfloor \frac{N}{2R} \right\rfloor,$$

где квадратная скобка означает операцию взятия целой части содержимого,

$$J = M + 2.$$

При выполнении неравенства

$$M \geq 4$$

собственные числа обладают следующими свойствами

$$\lambda_1 \approx \lambda_2 \approx \lambda_3 \approx \lambda_4 \approx 1;$$

$$\lambda_{J+k} \approx 0, \quad k = 1, 2, \dots$$

Таким образом, в соответствии с (5), с достаточной степенью обоснованности для матрицы  $A_V$  можно использовать аппроксимации

$$A_V \cong \sum_{k=1}^J \lambda_k \bar{q}_k \bar{q}_k^T,$$

$$A \cong Q_1 L_1 Q_1^T,$$

где

$$Q_1 = (\bar{q}_1, \dots, \bar{q}_J),$$

$$L_1 = \text{diag}(\lambda_1, \dots, \lambda_J)$$

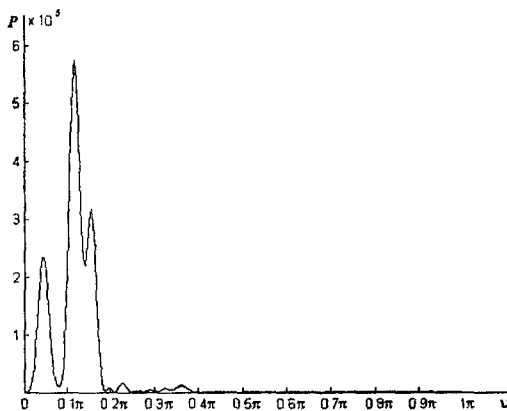
При этом представление (3) принимает вид

$$P_V(\vec{x}) \cong \hat{P}_V(\vec{x}) = \sum_{k=1}^J \lambda_k \alpha_k^2. \quad (11)$$

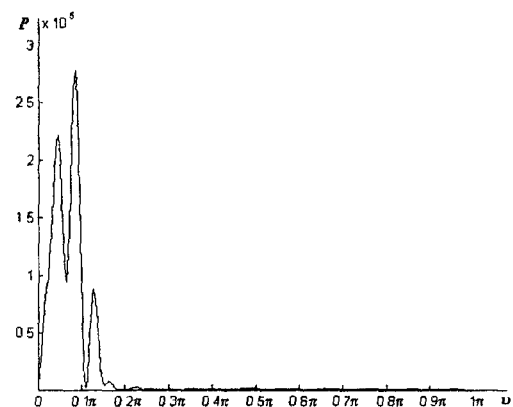
Очевидно, что использование представления (11) позволяет уменьшить объем вычислений по сравнению с вычислением непосредственно квадратичной формы вида (3).

Соотношения (3) и (11) представляют собой новый инструмент, позволяющий анализировать спектральные плотности отрезков звуковых сигналов.

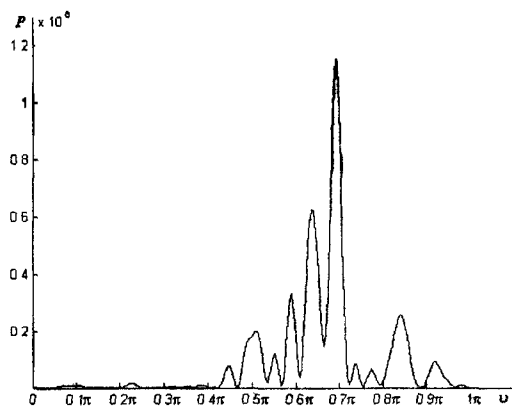
На рисунке 1 приведены энергетические спектры фрагментов речевых сигналов полученные с помощью предлагаемого метода.



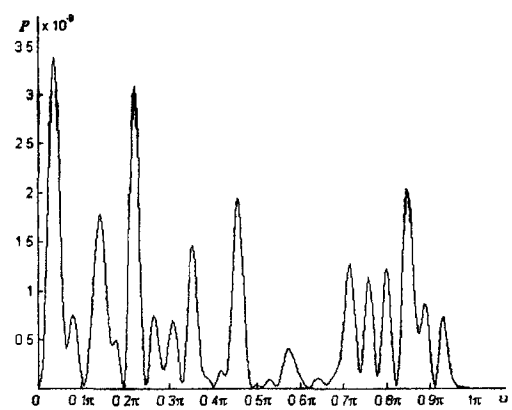
а) Энергетический спектр звука «а»



б) Энергетический спектр звука «и»



в) Энергетический спектр звука «ч»



г) Энергетический спектр шума в паузе речи

Рис. 1

## 2. СУБПОЛОСНОЕ КОДИРОВАНИЕ ЗВУКОВЫХ СИГНАЛОВ

В настоящее время все большее распространение получает подход на основе так называемого субполосного кодирования [8], когда исходный вектор значений некоторого отрезка сигнала  $\vec{x} = (x_1, \dots, x_N)^T$  заменяется вектором такой же размерности

$$\vec{w} = (w_1, \dots, w_N)^T,$$

который определяется следующим образом

$$w_k = \begin{cases} w_{h+1}, & \text{если } h = \frac{k-1}{R} \text{ целое число,} \\ 0, & \text{в противном случае.} \end{cases} \quad (12)$$

Отметим, что тогда определенный в (11) вектор имеет структуру

$$\vec{w} = (w_1, 0, \dots, 0, w_2, 0 \dots 0, w_M, 0, \dots, 0)^T \quad (13)$$

Вектор

$$\vec{w}_V = (w_1, \dots, w_M)^T$$

будем называть субполосным представлением (преобразованием) некоторого отрезка сигнала  $\vec{x} = (x_1, \dots, x_N)^T$ , соответствующего частотному интервалу  $(\nu_1, \nu_2)$ . Для его вычисления предлагается использовать вариационный принцип

$$S_V^2(\vec{w}) = \int_{\nu \in I} |X(\nu) - \mathcal{F}(\nu)|^2 d\nu = \min_{\vec{w}} \quad (14)$$

$$\text{где } \mathcal{F}(\nu) = \sum_{k=1}^N w_k e^{-j(k-1)\nu}.$$

Очевидно, что

$$X(\nu) - \mathcal{F}(\nu) = \sum_{k=1}^N (x_k - w_k) e^{-j(k-1)\nu}.$$

Поэтому, в соответствии с (3) условие (14) можно преобразовать к виду

$$\vec{x}^T A_V \vec{x} - 2 \vec{w}^T A_V \vec{x} + \vec{w}^T A_V \vec{w} = \min_{\vec{w}}.$$

С учетом структуры (13), нетрудно получить представление для оптимального в смысле критерия (14) вектора

$$\vec{w}_V = A A_V \vec{x} / a_{11}$$

где  $AA_{V_i}$  матрица, составленная из строк матрицы вида (4) в заданных частотных полосах (9)

$$AA_{V_i} = \begin{pmatrix} \bar{a}_1^{V_i} \\ \bar{a}_{R+1}^{V_i} \\ \bar{a}_{2R+1}^{V_i} \\ \dots \\ \bar{a}_N^{V_i} \end{pmatrix} \quad (15)$$

Тогда вектор

$$\vec{w} = AA\vec{x} = (\bar{w}_{V_1}^T, \bar{w}_{V_2}^T, \dots, \bar{w}_{V_R}^T)^T$$

будет оптимальным в смысле того, что каждый из его подвекторов  $\bar{w}_{V_i}$  удовлетворяет вариационному принципу вида (14).

### 3. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Для сравнения существующего и разрабатываемого методов субполосного преобразования, нами были проведены вычислительные эксперименты.

В одном из экспериментов осуществлялась фильтрация с прореживанием исходного сигнала при помощи КИХ-фильтров с длинами импульсных характеристик равными 256, 512 и 1024 значений и вычисление оптимальных подвекторов  $\bar{w}_{V_i}$ .

КИХ-фильтрация осуществлялась путем линейной свертки импульсных характеристик фильтров с исходным сигналом. Субполосное преобразование на основе предлагаемого метода представляло собой умножение исходного вектора на соответствующую выделяемому частотному диапазону матрицу с элементами вида (15), где  $R=10$ ,  $N=101$ , 501, 1001 так что  $M=11$ , 51, 101.

Далее осуществлялось вычисление относительных погрешностей аппроксимации отрезков спектров исходных последовательностей в заданных частотных полосах (9).

Наиболее типичные результаты приведены в таблице. Здесь:

$\delta_{opt}$  - среднеквадратическая относительная погрешность аппроксимации отрезков спектра исходной последовательности с помощью оптимальных подвекторов;

$\delta_{256}$ ,  $\delta_{512}$ ,  $\delta_{1024}$  - среднеквадратическая относительная погрешность аппроксимации с помощью прореженных выходных последовательностей КИХ-фильтров с длиной импульсной характеристики равной соответственно 256, 512 и 1024 значениям.

Таблица

Границы частотного интервала	При $N=101$			
	$\delta_{opt}$	$\delta_{256}$	$\delta_{512}$	$\delta_{1024}$
$\nu_1=0; \nu_2=\pi/10$	0,74159	0,54417	0,54056	0,5405
$\nu_1=\pi/10; \nu_2=2\pi/10$	0,16301	0,22581	0,22581	0,22586
$\nu_1=2\pi/10; \nu_2=3\pi/10$	0,40094	0,47917	0,47919	0,47917
$\nu_1=3\pi/10; \nu_2=4\pi/10$	0,30412	0,37807	0,37805	0,37808
$\nu_1=4\pi/10; \nu_2=5\pi/10$	0,32314	0,40571	0,40572	0,40576
$\nu_1=5\pi/10; \nu_2=6\pi/10$	0,30457	0,3991	0,39911	0,39916
$\nu_1=6\pi/10; \nu_2=7\pi/10$	0,26769	0,37328	0,37331	0,37336
$\nu_1=7\pi/10; \nu_2=8\pi/10$	0,26766	0,3999	0,39991	0,39998
$\nu_1=8\pi/10; \nu_2=9\pi/10$	0,25674	0,45352	0,45344	0,45353
$\nu_1=9\pi/10; \nu_2=10\pi/10$	0,2443	0,65221	0,64997	0,65126

Границы частотного интервала	При $N=501$			
	$\delta_{opt}$	$\delta_{256}$	$\delta_{512}$	$\delta_{1024}$
$\nu_1=0; \nu_2=\pi/10$	0,40802	0,97669	0,49158	0,46694
$\nu_1=\pi/10; \nu_2=2\pi/10$	0,077337	0,1897	0,1323	0,12678
$\nu_1=2\pi/10; \nu_2=3\pi/10$	0,22994	0,61339	0,35633	0,33494
$\nu_1=3\pi/10; \nu_2=4\pi/10$	0,15726	0,54423	0,27573	0,24842
$\nu_1=4\pi/10; \nu_2=5\pi/10$	0,1764	0,7624	0,32725	0,27486
$\nu_1=5\pi/10; \nu_2=6\pi/10$	0,1658	0,82956	0,32751	0,26107
$\nu_1=6\pi/10; \nu_2=7\pi/10$	0,14396	0,87493	0,32371	0,24244
$\nu_1=7\pi/10; \nu_2=8\pi/10$	0,15084	1,1714	0,38652	0,26355
$\nu_1=8\pi/10; \nu_2=9\pi/10$	0,15177	1,9089	0,56868	0,31165
$\nu_1=9\pi/10; \nu_2=10\pi/10$	0,17015	11,594	2,1374	0,61242
Границы частотного интервала	При $N=1001$			
	$\delta_{opt}$	$\delta_{256}$	$\delta_{512}$	$\delta_{1024}$
$\nu_1=0; \nu_2=\pi/10$	0,30477	1,1608	0,52115	0,47472
$\nu_1=\pi/10; \nu_2=2\pi/10$	0,050537	0,16214	0,10527	0,099203
$\nu_1=2\pi/10; \nu_2=3\pi/10$	0,16651	0,60462	0,32298	0,29309
$\nu_1=3\pi/10; \nu_2=4\pi/10$	0,11507	0,51585	0,24879	0,2153
$\nu_1=4\pi/10; \nu_2=5\pi/10$	0,12971	0,73188	0,30535	0,24242
$\nu_1=5\pi/10; \nu_2=6\pi/10$	0,12151	0,83927	0,32482	0,23959
$\nu_1=6\pi/10; \nu_2=7\pi/10$	0,10694	0,86573	0,32021	0,22141
$\nu_1=7\pi/10; \nu_2=8\pi/10$	0,11253	1,1866	0,3897	0,23581
$\nu_1=8\pi/10; \nu_2=9\pi/10$	0,1127	1,9867	0,61298	0,27591
$\nu_1=9\pi/10; \nu_2=10\pi/10$	0,14233	18,115	3,6824	0,68402

При возрастании  $N$  среднеквадратическая относительная погрешность аппроксимации отрезков спектра исходной последовательности с помощью оптимальных подвекторов уменьшается (см. рисунок 2).

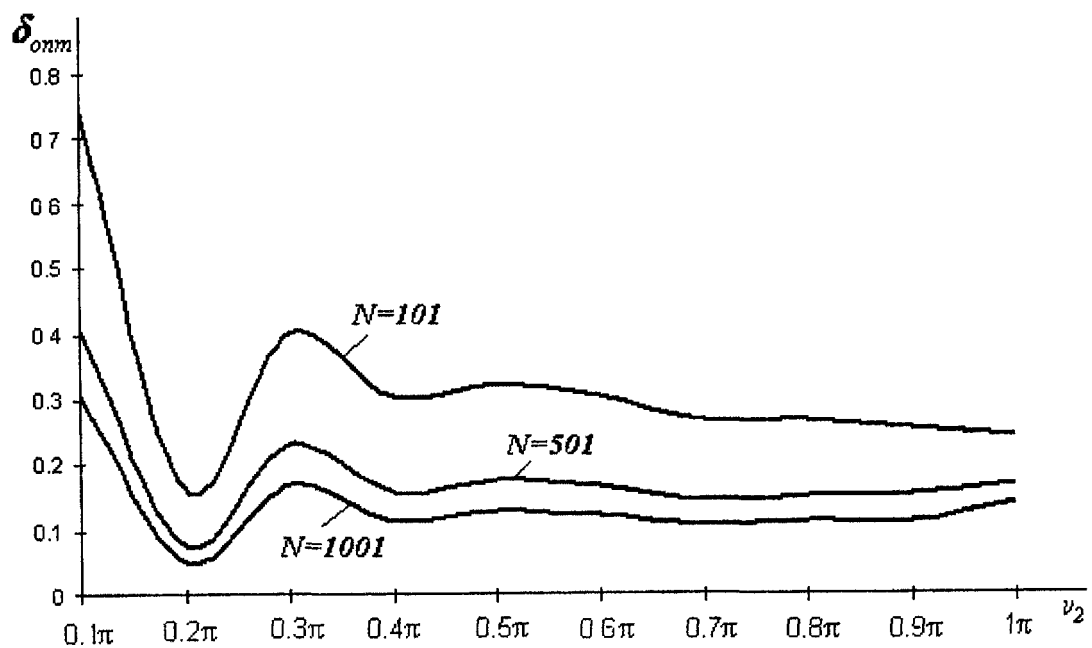


Рис. 2

Анализ результатов вычислительных экспериментов показывает, что по сравнению с субполосным представлением на основе КИХ-фильтров, даже при длине импульсной характеристики в 1024 значения, применение предлагаемого субполосного преобразования позволяет существенно уменьшить погрешность аппроксимации спектра исходных данных.

#### Библиографический список

1. Вологдин Э.И. Слух и восприятие звука: учебное пособие по курсу Аудиотехника. – СПб.: СТ Факультет ДВО, 2004. - 125 с
2. Воробев В.И., Грибунин В.Г. Теория и практика вейвлет-преобразования. - ВУС, 1999. 204 с.
3. 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.
4. Сергиенко А.Б. Цифровая обработка сигналов. – СПб.: Питер, 2003. – 604 с.
5. Жилияков Е.Г., Попов И.Г., Чижов И.И. О субполосном кодировании сигнала. - Вестник национального технического университета "Харьковский политехнический институт", №46, 2004.
6. Гантмахер Ф.Р. Теория матриц - М : Физматлит, 2004. – 560 с
7. Ковалгин Ю.А., Вологодин Э.И. Цифровое кодирование звуковых сигналов. СПб.: Корона, 2004. – 231 с.
8. Рабинер Л., Голд Г. Теория и применение цифровой обработки сигналов. М.: Мир, 1988. – 848 с.

### FREQUENCY ANALYSIS OF SPEECH SIGNALS.

*E.G.Ziliakov, E.I.Prokhorenko*

The authors of the article consider a new optimal method (root-mean-square error serves as a criterion) of frequency analyses and filtering of speech signals, including subband analyses.