

УДК 621.391; 519.21; 519.245

© А. В. Меркушева

ПРИМЕНЕНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ ТЕКУЩЕГО АНАЛИЗА НЕСТАЦИОНАРНОГО СИГНАЛА (РЕЧИ), ПРЕДСТАВЛЕННОГО ЕГО ВЕЙВЛЕТ-ОТОБРАЖЕНИЕМ.

I. ОСНОВНЫЕ ПРИНЦИПЫ

Представлен метод детектирования изменения свойств нестационарного сигнала в области вейвлет-разложения на основе нейронной сети. Дан метод вейвлет-преобразования речевого сигнала с использованием вейвлет-пакета, структура которого согласована с перцептуальной моделью.

ВВЕДЕНИЕ

Детектирование момента изменения свойств нестационарного сигнала является необходимым условием фильтрации с адаптивной подстройкой порога, величина которого зависит от уровня шума. Такая возможность предоставляется в информационно-измерительной сети (ИИС), где зашумленный сигнал, несущий полезную информацию, регистрируется на временных интервалах, которые перемежаются с интервалами, свободными от сигнала. Типичный случай этого варианта контроля и обработки реализуется в информационной системе анализа речевого сигнала. Система служит моделью достаточной общности для нестационарных сигналов в ИИС со спорадически появляющимися информационно значимыми интервалами различной длительности. Другим примером системы такого вида является ИИС, предназначенная для контроля гетерогенного потока товарной нефти [1], где выполняется выделение временных сегментов, на которых отсутствует свободный газ, определяется средняя плотность потока при наличии и отсутствии газа, а затем на основе этих данных вычисляется средняя относительная доля свободного газа.

Метод детектирования изменения свойств нестационарного сигнала основан на вейвлет-преобразовании и нейросетевом алгоритме и рассматривается применительно к информационной системе многоцелевой обработки речевого сигнала. Адаптивная фильтрация шума реализуется путем динамического изменения порога, вычисляемого на интервалах (микропаузах) отсутствия речевого сигнала, поэтому правильное определение наличия или отсутствия речевого сигнала на анализируемом интервале оказывает существенное влияние на качество фильтрации. Эта задача является актуальной в различных системах обработки речевого сигнала:

— в цифровой телефонии для удаления пауз при сжатии речевых сигналов;

— в криптографии, где удаление пауз сокращает избыточность, уменьшающую криптостойкость алгоритма шифрования;

— в устройствах фильтрации речевого сигнала с адаптацией к виду и уровню шума, где во время пауз должны быть получены кратковременные характеристики шума.

Таким образом, в описанных ИИС и других приложениях необходимо детектирование (отслеживание и идентификация) изменения свойств нестационарных процессов, а также адаптация к изменению уровня шума. Это особенно важно для обработки сигнала в реальном времени, при которой алгоритм обработки должен адаптироваться к состоянию процесса.

В разработанном методе использовано вейвлет-разложение речевого сигнала по биортогональному базису Добеши [2], а для принятия решения о типе интервала (сегмента) сигнала — нейронная сеть на многослойном перцептроне. Сравнительно с традиционными спектральными методами вейвлет-преобразование дает более точную локализацию сигнала по времени и по частоте (в субполосах разложения), имеет быстрый алгоритм реализации. Биортогональный базис сохраняет фазовые соотношения частотных компонент сигнала после его восстановления обратным вейвлет-преобразованием. Свойство наличия у базиса преобразования конечного числа нулевых моментов гарантирует точную аппроксимацию полиномиальной части сигнала, что ведет к дополнительному снижению объема данных [3].

ВЕЙВЛЕТ-РАЗЛОЖЕНИЕ РЕЧЕВОГО СИГНАЛА В СООТВЕТСТВИИ С ПЕРСЕПТУАЛЬНОЙ МОДЕЛЮ

Изменение свойств речевого сигнала на интервалах речь/пауза определяется изменением спек-

тральных характеристик. Алгоритм Рабинера (L. Rabiner) [4] для разделения речи и пауз основан на анализе энергии сигнала и числе переходов через нуль. Алгоритм не учитывает особенностей частотного спектра сигнала и шума и поэтому не всегда позволяет правильно классифицировать интервалы, особенно в условиях достаточно мощного узкополосного шума или музыкального фона.

Применение спектрального анализа для детектирования свойств сигнала также не является эффективным, поскольку получение спектра Фурье связано с использованием всего интервала анализируемого сигнала, а кратковременный спектр Фурье имеет либо плохое разрешение по частоте, либо занимает большой временной интервал [5, 6]. Метод Фурье недостаточен также в связи с тем,

что после фильтрации восстановление обратным преобразованием может привести к комплексной форме сигнала. Поэтому в решении задачи распознавания речи и пауз предложен подход на основе вейвлет-преобразования и нейронной сети [7, 8]. Вейвлет-преобразование позволяет более точно локализовать частотные свойства сигнала во времени. Биортогональные вейвлеты Добеши позволяют уменьшить объем вычислений при разложении за счет использования коротких фильтров. Вейвлет-функция и масштабирующая (скэйлинг) функция, используемые для разложения сигнала, представлены на рис. 1. Вейвлет-разложение не приводит к увеличению объема данных при переходе от временного представления сигнала к его представлению в вейвлет-области [9].

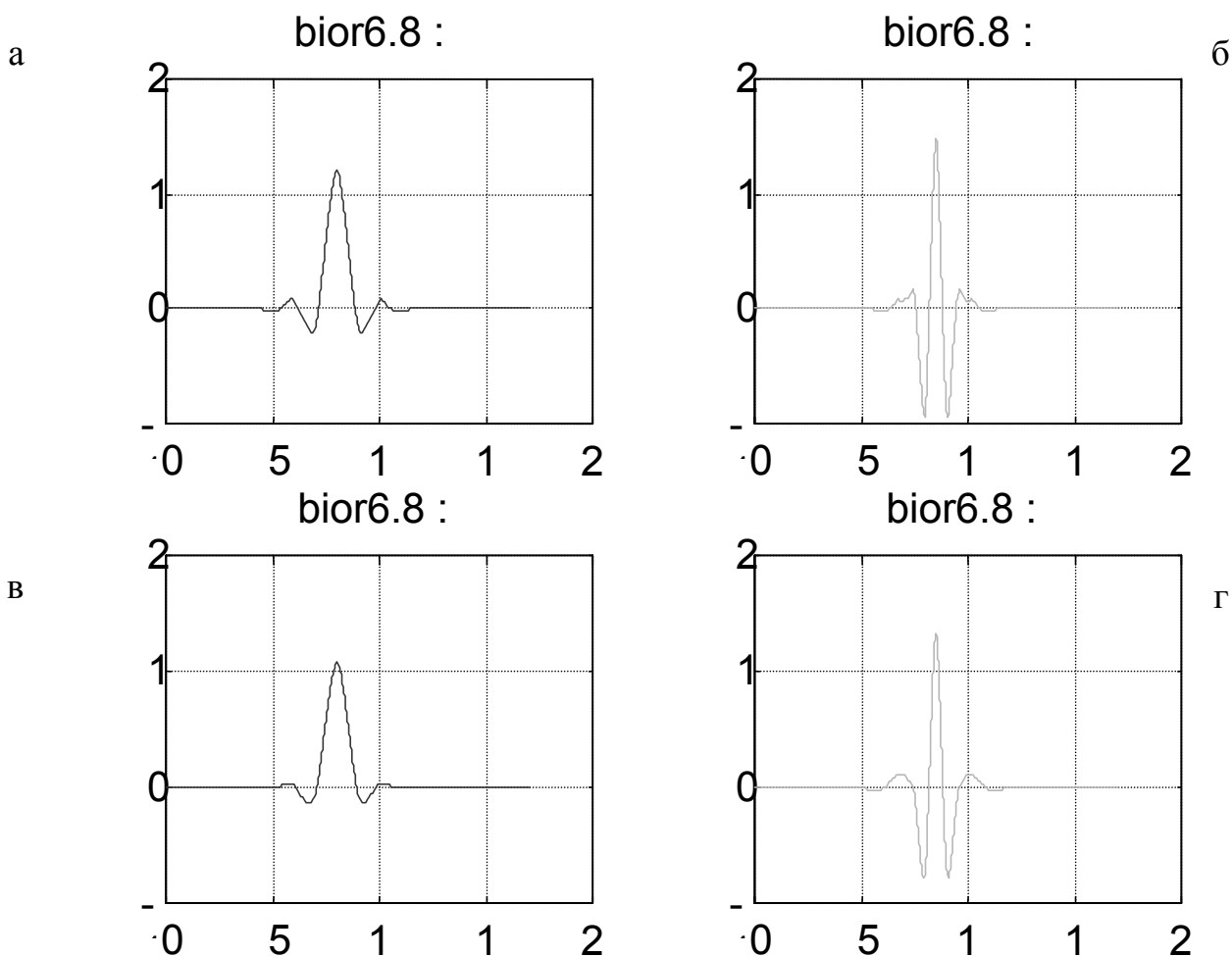


Рис. 1. Базисные функции, используемые при разложении сигнала с помощью вейвлет-пакетов: а — масштабирующая функция анализа; б — вейвлет-функция анализа; в — масштабирующая функция синтеза; г — вейвлет-функция синтеза

Алгоритм разделения речи и пауз построен с учетом особенностей восприятия звука человеком [10, 11, 12], которые описываются перцептуальной моделью [13, 14, 15]. Модель разделяет спектр речевого сигнала на частотные полосы, называемые критическими. Каждая полоса частотного диапазона речи по компоненте шума в ней воспринимается как единое целое, и по слуховому ощущению важна лишь мощность шума в полосе [16]. Перцептуальная модель успешно используется в алгоритмах MPEG (MPEG — Moving Picture Expert Group — международная рабочая группа по стандартизации аудио- и видео-информации) для кодирования акустических сигналов.

По перцептуальной модели частотному диапазону 100–12 000 Гц соответствует 22 критические полосы, ширина которых постепенно возрастает от 100 до 2500 Гц (табл. 1). Диапазон речевого сигнала ограничен частотой 8 кГц, однако в компьютерных системах обработки речи для уменьшения шума, связанного с дискретизацией и квантованием, используют повышенную частоту дискретизации, в частности 22 кГц. Даже при 10%-й неточности оценки частотного диапазона речи, начиная от самых низких частот в пределах 50–9500 Гц, его полностью перекрывают 22 полосы перцептуальной модели. Поэтому в системе обработки речевого сигнала следует использовать 22 полосы перцептуальной модели.

Для определения речевой активности предложен метод [17], который включает:

- получение сегмента речевого сигнала длительностью 10–20 мс;

- использование вейвлет-пакетов для получения разложения сегмента в соответствии с перцептуальной моделью;

- вычисление мощности вейвлет-коэффициентов в каждой области разложения;

- нормирование вектора мощности вейвлет-коэффициентов, полученного на сегменте речевого сигнала;

- использование нормированного вектора в качестве входного вектора нейронной сети.

Вейвлет-преобразование сигнала производит октавополосное разбиение спектра, которое подходит для большинства, но не для всех задач обработки сигнала. В частности, для получения перцептуальной модели необходимо изменить структуру разбиения частотно-временной плоскости. Каскадное соединение блоков вейвлет-фильтров позволяет достичь гибкого разбиения время—частотной плоскости. Метод вейвлет-пакетов позволяет на каждом уровне разложения производить наиболее выгодное разбиение [18–23].

Разложение речевого сигнала в соответствии с перцептуальной моделью обеспечивается специальным выбором двоичного дерева, которое опре-

Табл. 1. Частотные диапазоны перцептуальной модели

Индекс диапазона	Граничные частоты (Гц)	Средняя частота (Гц)	Ширина диапазона (Гц)
1	0–100	50	100
2	100–200	150	100
3	200–300	250	100
4	300–400	350	100
5	400–510	455	110
6	510–630	570	120
7	630–770	700	140
8	770–920	845	150
9	920–1080	1000	160
10	1080–1270	1175	190
11	1270–1480	1355	210
12	1480–1720	1600	240
13	1720–2000	1860	280
14	2000–2320	2160	320
15	2320–2700	2510	380
16	2700–3150	2925	450
17	3150–3700	3425	550
18	3700–4400	4050	700
19	4400–5300	4850	900
20	5300–6400	5850	1100
21	6400–7700	7050	1300
22	7700–9500	8750	1800
23	9500–12000	10750	2500
24	12000–15500	13750	3500
25	15500–23500	19500	8000

деляет структуру вейвлет-преобразования. При формировании такого дерева с использованием вейвлет-пакета получена форма разложения, показанная на рис. 2. Критерием качества аппроксимации перцептуальной модели является показатель среднеквадратичной относительной погрешности соответствия границ критических полос и границ

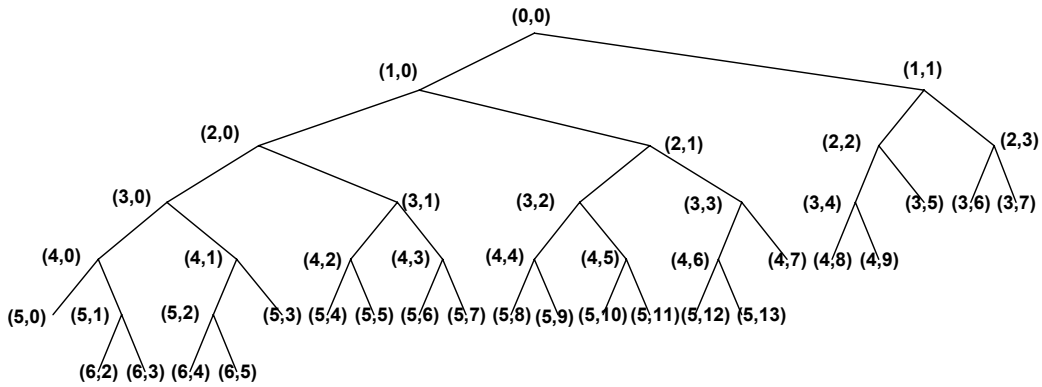


Рис. 2. Аппроксимация перцептуальной модели с помощью вейвлет-пакетов

субполос вейвлет-пакета [7]. Этот показатель (Q) представлен соотношением (1):

$$Q = \sqrt{\frac{1}{p} \sum_{i=1}^p \frac{1}{2} \left\{ \left(\frac{\hat{f}_i^1 - f_i^1}{\hat{f}_i} \right)^2 + \left(\frac{\hat{f}_i^2 - f_i^2}{\hat{f}_i} \right)^2 \right\}}, \quad (1)$$

где f_i^1, f_i^2 — нижняя и верхняя границы частотной полосы в перцептуальной модели, \hat{f}_i — положение центра полосы; \hat{f}_i^1, \hat{f}_i^2 — то же для субполос сформированного вейвлет-пакета; p — количество частотных диапазонов модели, $p = 22$. Несмотря на то что бинарная структура формирования дерева дает ограниченное количество вариантов, полученное разложение удовлетворительно аппроксимирует частотные диапазоны перцептуальной модели. При первоначальной оценке показателя $Q = 0.08$ небольшое укрупнение трех полос в частотных диапазонах перцептуальной модели (табл. 2) улучшает общее соответствие представления речевого сигнала вейвлет-пакетом. При этом показатель качества Q снижается до 0.03.

Само вейвлет-преобразование не дает требуемого снижения размерности (полное число коэффициентов разложения равно числу отсчетов сигнала), поэтому для нейросетевого алгоритма использован вектор с компонентами субполосной мощности сигнала $\mathbf{P} = (\sigma_1^2, \sigma_2^2 \dots \sigma_p^2)$, где σ_i^2 — сумма квадратов коэффициентов, принадлежащих i -й полосе разложения.

Таким образом, с использованием вейвлет-пакетов для каждого сегмента речевого сигнала могут быть получены векторы мощности вейвлет-коэффициентов в субполосах, соответствующих перцептуальной модели. Полученные векторы от-

ражают спектральные свойства сигнала приблизительно с теми же градациями в частотной области, что и слуховой анализатор человека.

Нормирование векторов $\mathbf{p} = \mathbf{P} / \|\mathbf{P}\|$ позволяет устранить влияние общей мощности сигнала на возможные решения о его свойствах и тем самым устранить основной недостаток алгоритмов детектирования речевого сигнала, ориентирующихся в значительной степени на силу сигнала.

ПРИМЕНЕНИЕ ПЕРСЕПТРОНА ДЛЯ ОБНАРУЖЕНИЯ СВОЙСТВ НЕСТАЦИОНАРНЫХ СИГНАЛОВ

Для принятия решения о наличии или отсутствии речевого сигнала на анализируемом сегменте предложено применение нейронной сети [7, 8]. Входным вектором сети является нормированный вектор мощности вейвлет-коэффициентов. Такое решение имеет следующие преимущества:

- система, использующая нейронную сеть, является самонастраивающейся адаптивной системой, для которой нет необходимости предварительно выполнять точный расчет параметров;
- система имеет этап обучения, позволяющий выполнить адаптацию параметров для настройки на конкретные условия сигнала и шума;
- в области вейвлет-разложения не обязательно иметь линейную функцию для разделения вейвлет-образов сигнала и шума, иными словами свойство линейной делимости образов не является обязательным;
- нелинейная активационная функция нейрона позволяет принимать нежесткие решения в каждом нейроне сети;

Табл. 2. Сопоставление частотных субполос вейвлет-представления речевого сигнала с персептуальной моделью

Верхние границы субполос и индексы частотных диапазонов вейвлет-пакета и персептуальной модели (Гц)			
Вейвлет-представление пакетом		Персептуальная модель	
Верхние границы субполос вейвлет-пакета	Индексы вершин дерева вейвлет-пакета ¹⁾	Верхние границы полос модели	Индексы полос ²⁾
250	1	300	1–3
375	2	400	4
500	3	510	5
625	4	630	6
750	5	770	7
1000	6	1080	8–9
1250	7	1270	10
1750	8–9	1720	11–12
2000	10	2000	13
2250	11	2320	14
2750	12–13	2700	15
3000	14	3150	16
3500	15–16	3700	17
4500	17–18	4400	18
5500	19	5300	19
6500	20	6400	20
7500	21	7700	21
9750	22	9500	22
12025	23	12000	23

¹⁾ Индексы 8–9, 12–13, 15–16 и 17–18 указывают на объединение субполос пакета на вершинах дерева разложения сигнала. Верхняя граница субполосы в первом столбце соответствует второму индексу из указанных пар.

²⁾ Индексы 1–3, 8–9, 11–12 соответствуют сгруппированным полосам персептуальной модели. Верхние границы полос соответствуют вторым индексам из указанных пар.

— структура нейронной сети может быть адаптирована к задаче разделения сегментов сигнала и шума.

Нейронной сетью достаточно простой структуры является персептрон [24–27]. В простейшем случае персептрон состоит из одного нейрона (рис. 3). Суммирующий узел нейрона вычисляет линейную комбинацию входов x_i , приложенных к его синапсам w_i , а также учитывает внешний порог θ . К выходу суммирующего узла v

$$v = \sum_{i=1}^p w_i x_i - \theta$$

применяется нелинейная функция $y = \varphi(v)$.

Элементарный персептрон с одним слоем не имеет скрытых нейронов и не может классифицировать входные векторы, которые не являются ли-

нейно разделимыми. Эта проблема решается введением скрытых слоев в многослойном персептроне (МСП). МСП состоит из входных узлов, составляющих входной слой, одного или более скрытых слоев и выходного слоя вычисляющих узлов. Входной сигнал распространяется через сеть слой за слоем. Каждый нейрон сети содержит нелинейность на выходе, которая является гладкой и определяется логистической функцией

$$y_j = \varphi(v_j) = \frac{1}{1 + \exp(-v_j)}.$$

Персептрон выполняет отображение входных p -мерных векторов $\mathbf{P} = (\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$ на выходное q -мерное пространство. В задаче обнаружения речевого сигнала $q = 1$.

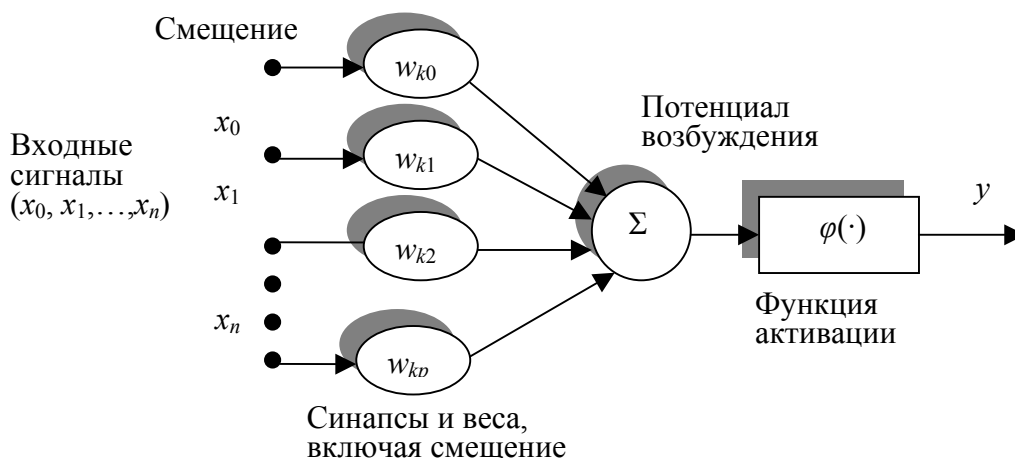


Рис. 3. Структурная схема элементарного персептрона

Согласно Сайбенко—Фунахаши [28, 29], персептрон способен аппроксимировать с любой точностью желаемое отображение вход—выход, т. е. обучаться сопоставлять множеству входных p -мерных векторов заданное множество q -мерных выходных векторов. Т. е. существует ε -аппроксимация F любого непрерывного отображения вход—выход f , которую можно представить в математической форме (2), и эта форма эквивалентна двухслойному персептрону (3) с одним выходным нейроном и одним скрытым слоем из M нейронов (M зависит от величины ε):

$$|F(x_1, x_2, \dots, x_p) - f(x_1, x_2, \dots, x_p)| \leq \varepsilon, \quad (2)$$

$$F(x_1, x_2, \dots, x_p) = \sum_{j=0}^M a_j \varphi \left(\sum_{i=1}^p w_{ji} x_i - \theta_j \right). \quad (3)$$

Здесь a_j — коэффициенты аппроксимации отображения f обобщенным рядом Фурье. В персептроне эти коэффициенты представлены синаптическими связями (весами) выходного нейрона. w_{ji} — веса от i -го к j -му нейрону, структурно связывающие входные узлы с нейронами скрытого слоя; θ_j — потенциал смещения (порог чувствительности) j -го нейрона; φ — логистическая функция преобразования потенциала возбуждения ней-

рона $v_j = \left(\sum_{i=1}^p w_{ji} x_i - \theta_j \right)$ в его выход y_j .

Изложенное подтверждает правильность выбо-

ра типа сети в целях разработки метода для детектирования изменения свойств нестационарного сигнала.

ЗАКЛЮЧЕНИЕ

1. Предложенный метод и модель анализа речевого сигнала дают основу для решения ряда задач детектирования изменения свойств нестационарного сигнала.

2. Для анализа нестационарного сигнала (речи) целесообразно использование нейросетевых алгоритмов на вейвлет-отображении сигнала.

3. Вейвлет-разложение может быть адаптировано к частотной структуре сигнала на основе вейвлет-пакета. Для речевого сигнала выбор вида пакета определен персептуальной моделью восприятия звука.

4. Для реализации нейросетевого алгоритма решен ряд вопросов:

- обоснован метод обучения, который имеет лучшую скорость сходимости для векторов аудиосигнала и позволяет получить меньшие по величине ошибки распознавания свойств сигнала;
- получено ограничение на требуемое время обучения алгоритма;
- обоснована оптимальная структура персептрона, которая не является избыточной и позволяет эффективно решать задачи распознавания.

Этим вопросам посвящена вторая статья данной серии.

СПИСОК ЛИТЕРАТУРЫ

1. *Кратиров Д.В., Меркушева А.В.* Алгоритм, основанный на вейвлет-преобразовании и нейронной сети, для бесконтактного измерения параметров газожидкостного потока // Сборник докладов Международной конференции "Датчики и системы". СПб.: Изд-во СПбГТУ, 2002. Т. 3. С.51–55.
2. *Daubechies I.* Painless No orthogonal Expansions // *Journal of Mathematical. Physics.* 1986. V. 27. P. 1271–1283.
3. *Villemois L.F.* Energy Moments in Time and Frequency for 2-Scale Equations and Wavelets // *SIAM Journal of Mathematical Analysis.* 1992. V. 23. P. 1119–1153.
4. *Рабинер Л.Р., Шафер Р.В.* Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. 492 с.
5. *Allen J.B., Rabiner L.R.* A Unified Approach to Short-Time Fourier Analyses and Synthesis // *Proceedings of IEEE.* 1977. V. 65, N 11. P. 1558.
6. *Portnoff M.R.* Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analyses // *IEEE Transactions on Signal Processing.* 1980. V. 28, N 2. P. 55.
7. *Исмаилов Ш.Ю., Меркушева А.В.* Нейросетевой алгоритм на вейвлет-преобразовании нестационарного сигнала в ИИС // Сб. докладов Международной конференции по мягким вычислениям и измерениям SCM-2001. СПб., 2001. Т. 1. С. 251–256.
8. *Малыхина Г.Ф., Меркушева А.В.* Вейвлет-фильтрация нестационарного сигнала с адаптацией на основе нейронной сети // Сб. докладов Международной конференции по мягким вычислениям и измерениям SCM-2001. СПб., 2001. Т. 1. С. 239–242.
9. *Lang M., Guo H.* Noise Reduction Using Undecimated Discrete Wavelet Transform // *IEEE Signal Processing Letters.* 1996. V. 3, N 1. P. 8.
10. *Atal B.S.* Optimal Noise Specter Form // *IEEE Transactions on Signal Processing.* 1979. N 6. P. 247–254.
11. *Fletcher N.* Auditory Patterns // *Review of Modern Physics.* 1940. P. 47–65.
12. *Zwicker E., Fastl H.* Psychoacoustics, Facts and Models. Berlin: Springer—Verlag, 1984. 420 с.
13. *Scharf B.* Critical Bands // *Foundation of Modern Auditory Theory / Ed. Tobias J.* N.Y.: Acad. Press, 1970. P. 159–222.
14. *Schroeder M.R., Atlas B.C., Hall J.L.* Optimizing Digital Speech Coders by Exploiting Masking Property // *Journal of Audio Engineering Society.* 1995. V. 43, N 11. P. 914.
15. *Zelinski R., Noll P.* Adaptive Bit Allocation with Thin Structure (on Formants) of Optimal Distortion or Noise Specter // *IEEE Transactions on Signal Processing.* 1977. N 8. P. 299–309.
16. *Рабинер Л., Гоулд Б.* Теория и применение цифровой обработки сигналов (перев. с англ.). М.: Мир, 1978. 848 с.
17. *Малыхина Г.Ф., Меркушева А.В.* Детектирование речевого сигнала и фильтрация с адаптивным порогом // Сборник трудов факультета техн. киберн. СПбГТУ: Микропроцессорные средства измерений. СПб., 2001. Вып. 2. С. 26–35.
18. *Berger J, Coifman R.R., Goldberg M.J.* Removing of Noise from Music Using Local Trigonometric Bases and Wavelet-Packets // *Journal of Audio Eng. Society.* 1994. V. 42, N 9. P. 808.
19. *Chui C.K., Li C.* Nonorthogonal Wavelet Packets // *SIAM Journal of Mathematical Analysis.* 1993. V. 24. P. 712–738.
20. *Coifman R.R., Wickerhauser M.L.* Entropy Based Algorithms for Best Bases Selection // *IEEE Transactions on Information Theory.* 1992. V. 38. P. 713–718.
21. *Devis G.M., Mallat S., Zhang Z.* Adaptive time-Frequency Decompositions // *Optical Engineering.* 1994. V. 33, N 7. P. 2183.
22. *Feichtinger H.G.* Irregular Sampling Theories and Series Expansions of Band- Limited Functions // *SIAM Journal of Mathematical Analysis.* 1992. V. 23. P. 530.
23. *Ramchadran K., Vetterli M., Herley C.* Wavelets, subband coding and best bases // *Proceedings of IEEE.* 1996. V. 84, N 4. P. 353.
24. *Уоссермен Ф.* Нейрокомпьютер и его применение (перев. с англ. под ред. А.И. Галушкина). М.: Изд-во Мир, 1992. 236 с.
25. *Цыганков В.Д.* Нейрокомпьютер и его применение. М., 1993. 117 с.
26. *Drodlie K.W.* Unconstrained Optimization // *Numerical Analyses / Ed. D. Jackobs.* London: Academic Press, 1977. P. 229–388.
27. *Russo A.R.* Tutorial N 8 // *IEEE Conference on Neural Networks.* Washington, 1990. P. 12.
28. *Cybenko G.* Approximation by Superposition of a Sigmoidal Function // *Mathematics of Control, signal and Systems.* 1989. N 2. P. 303–314.
29. *Funahashi K.* On the Approximate Realization of Continuous Mapping by Neural Network // *Neural Networks.* 1989. N 2. P. 183–192.

Санкт-Петербург

Материал поступил в редакцию 10.11.2002.

**APPLICATION OF A NEURAL NETWORK
TO ON-LINE ANALYSIS OF NON-STATIONARY (SPEECH)
SIGNALS REPRESENTED BY THEIR
WAVELET TRANSFORM.
I. BASIC PRINCIPLES**

A. V. Merkusheva

Saint-Petersburg

A property-modification detection method for non-stationary signals is suggested, that uses wavelet decomposition and a neural network. A wavelet transformation of speech signal is given using a wavelet-packet whose structure is determined by a perceptual model.