АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
(Вычислительные системы)
1991 год Выпуск 141

УДК 621.31:534.4

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ЗАВИСИМОСТИ МЕЖДУ ПЕРВЫМИ И ВТОРЫМИ МОМЕНТАМИ РЕЧЕВОГО СИГНАЛА ДО И ПОСЛЕ НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ

А.В.Кельманов, С.А.Хамидуллин

1. Введение. Постановка задачи

В существующих трактах связи передаваемые речевые сигналы могут подвергаться нелинейным искажениям [1,2]. Чаще всего эти искажения необратимы [3], поэтому восстановить исходный сигнал, т.е. решить обратную задачу, или (что то же самое) компенсировать искажения напрямую не представляется возможным.

Известно [4], что оцифрованный речевой сигнал вполне адекватно описывается небольшим числом отсчетов его текущей кова - риационной или корреляционной функции. В частности, при частоте квантования 10 кГц достаточно 10-16 текущих отсчетов корреляций (не считая дисперсии и частоты основного тона) или такого же числа параметров авторегрессии (находящихся по корреля - циям), чтобы синтезировать речеподобный сигнал, на слух прак - тически не отличающийся от исходного. Указанного объема информации бывает достаточно для надежного распознавания речи [5], поэтому при синтезе и распознавании речи, а также ее коррекции после нелинейных амплитудных искажений можно ограничиться восстановлением лишь ковариационной (корреляционной) функции речевого сигнала или ее оценок. Иными словами, при компенсации

помех вместо (не всегда возможного) обращения функции,искажающей речевой сигнал, можно перейти к оцениванию (восстановлению) ковариаций по косвенным или искаженным наблюдениям. Очевидно, что подобный переход возможен лишь в том случае, если установлена (теоретически или экспериментально) зависимость ковариаций неискаженного сигнала с какими-либо характеристиками искаженного сигнала.

Как с теоретической, так и с практической стороны в качестве этих характеристик удобнее всего использовать ковариации и математическое ожидание искаженного сигнала (или первый и второй моменты). Дело в том, что к настоящему времени теоретически достаточно полно проработаны вопросы нахождения ковариационных функций на выходе нелинейных систем по ковариационным функциям на их входе при условии, что вид искажения известен и на вход системы поступает стационарный гауссовский процесс [6]. Нахождение математического ожидания после искажения в этих условиях не представляет особого труда.

В том случае, когда функция, связывающая моменты искаженного сигнала с ковариациями неискаженного, неизвестна, перед компенсацией необходимо решить задачу восстановления или оценивания этой функции или зависимости. И если эта зависимость однозначна, в дальнейшем при обработке сигнала можно по ней и по оценкам моментов искаженного процесса получить оценки ковариаций неискаженного и, таким образом, скомпенсировать искажения. В этом плане решение задачи оценивания указанной зависимости представляется актуальным.

Одна из главных проблем, возникающих при решении задачи восстановления зависимостей между моментами, связана с тем,что речевой сигнал по своей природе не является стационарным процессом. Однако относительно короткие участки речевого сигнала (длительностью до 20-40 мс) можно рассматривать как реализации стационарного случайного процесса. Ограничение по длительности

окна анализа вместе с необходимостью использовать возможно меньшую частоту дискретизации сигнала приводит к ограниченности объема выборки или числа отсчетов обрабатываемого сегмента. Поскольку теоретические результаты, касающиеся нелинейных преобразований случайных процессов, носят асимптотический характер, необходима проверка на достаточность того объема выборки, который определяется условиями стационарности сигнала.

Другая проблема состоит в том, что из-за отличия функции распределения речевого сигнала от гауссовской теоретические зависимости будут отличны от реальных. Возникают вопросы, насколько значительны эти различия и возможно ли использование теоретических результатов на практике. Если эти различия значительны, то даже при известной искажающей функции потребуется решать задачу восстановления зависимости между моментами речевого сигнала до и после искажений. Если же эти отличия несущественны, то практически без потери точности можно будет пользоваться теоретическими результатами.

Цель данной работы состоит в получении эмпирических оценок для функции, связывающей корреляции речевого сигнала до и после нелинейных искажений, а также для функции, описывающей зависимость дисперсии неискаженного речевого сигнала от математического ожидания искаженного. При этом проводится сравнение полученных эмпирических зависимостей для речевого сигнала с функциями, справедливыми для стационарных гауссовских процессов. Результаты исследований увязываются с проблемой ограниченности выборки.

Поскольку множество искажающих функций, в принципе, необозримо, ограничимся рассмотрением двух типов искажений, которые, с одной стороны, довольно часто самопроизвольно появляются в каналах связи, а с другой - могут вводиться искусственно с целью ускорения вычислений при обработке речевых сигналов [3, 7].

Пусть $x(t)$, $-\infty < t < \infty$, - непрерывный речевой сигнал (случайный процесс), а $x_n = x(nT)$, $n = 0,\pm1,\pm2,\ldots$, - дискретные значения (отсчеты) этого сигнала, взятые через равные промежутки времени T. Будем считать, что на коротких участках длительностью T_a, содержащих N отсчетов, сигнал является реализацией стационарного случайного процесса, который полностью описывается своим одномерным распределением P_x (отличным от гауссовского) и ковариационной функцией. Под искажением будем понимать нелинейное преобразование $y(t) = f[x(t)]$ или $y_n = f(x_n)$.

Пусть $\sigma_y(m)$, $m = 0, \pm1, \pm2, \ldots$, - ковариационная, а $\rho_y(m) = \sigma_y(m)/\sigma_y(0)$ - корреляционная функция сигнала $\{y_n\}$; $M_y = \mathcal{M}y_n$ - его математическое ожидание. Символами $c_y(m)$, $r_y(m)$ обозначим классические оценки ковариаций и корреляций процесса $\{y_n\}$:

$$\left.\begin{aligned} c_y(m) &= \frac{1}{N}\sum_{n=1}^{N-m}(y_n - M_y)(y_{n+m} - M_y), \\ r_y(m) &= c_y(m)/c_y(0), \quad m = 0,1,\ldots,N-1. \end{aligned}\right\} \qquad (1)$$

Для процесса $\{x_n\}$ формулы аналогичны. Оценки вида (1) будем называть прямыми.

Предположим, что искажение f таково, что корреляции искаженного и неискаженного процессов связаны однозначной функцией $\rho_x(m) = \theta_1[\rho_y(m)]$. Первая рассматриваемая задача состоит в оценивании функции θ_1 по выборочным корреляциям $r_x(m)$ и $r_y(m)$. Во второй задаче, в отличие от первой, предполагается, что для искажения f имеется однозначная зависимость $\sigma_x(0) = \theta_2(M_y)$. Требуется оценить функцию θ_2 по выборочному среднему $\bar{y}$ и выборочной дисперсии $c_x(0)$. Вид искажения конкретизируем ниже.

2. Оценивание зависимости между корреляциями

Решение первой задачи - оценивание зависимости между корреляциями - проведем на примере нелинейного искажения, сводящегося к клиппированию сигнала. Сначала рассмотрим, насколько хорошо можно оценить известную теоретическую зависимость для гауссовских процессов, а затем перейдем к оцениванию неизвестной зависимости для речевых сигналов.

Под клиппированием обычно понимается следующее преобразование:

$$y_n = f(x_n) = \begin{cases} a, & \text{если } x_n \geq 0, \\ -a, & \text{если } x_n < 0, \end{cases} \quad n = 0, \pm 1, \pm 2, \ldots \tag{2}$$

Если клиппированию подвергается стационарный нормальный процесс $\{x_n\}$, имеющий $M_x = 0$ и ковариационную функцию $\sigma_x(m)$, то ковариационная функция искаженного процесса $\{y_n\}$ может быть найдена по формуле [6]:

$$\sigma_y(m) = \sigma_y(0) \frac{2}{\pi} \arcsin \rho_x(m). \tag{3}$$

После обращения (3) для корреляционной функции исходного процесса $\{x_n\}$ имеем:

$$\rho_x(m) = \theta_1[\rho_y(m)] = \sin \frac{\pi}{2} \rho_y(m). \tag{4}$$

Известно [8], что $r_x(m)$ и $r_y(m)$ являются состоятельными оценками теоретических корреляций $\rho_x(m)$ и $\rho_y(m)$. Поэтому, учитывая непрерывность θ_1, при больших N эмпирическое оценивание зависимости между $\rho_x(m)$ и $\rho_y(m)$ можно проводить по оценкам $r_x(m)$ и $r_y(m)$. В первом приближении для этого достаточно по выборочным значениям про-

цессов $\{x_n\}$ и $\{y_n\}$ получить множество точек (r_x, r_y) при фиксированных задержках $m < N$ и нанести их на плоскость. В силу известной ограниченности модулей корреляционных функций и ее оценок геометрическое место точек (r_x, r_y) будет лежать в единичном квадрате: $|r_x| \le 1$, $|r_y| \le 1$.

Указанная процедура была проделана для гауссовского процесса авторегрессии второго порядка следующего вида:

$$x_n = \alpha x_{n-1} - \alpha^2 x_{n-2} + \sigma\varepsilon_n, \quad n = 1,2,... \qquad (5)$$

Здесь ε_n - последовательность независимых одинаково распределенных нормальных случайных величин с нулевым математическим ожиданием и единичной дисперсией. На рис. 1 приведено облако точек-оценок (r_x, r_y), полученное по 150 выборкам процесса (5) при $\alpha = 0,9$ и $\sigma = 128$. Каждая из выборок содержала по $N = 256$ значений. Величина задержки m при оценивании изменялась от 1 до 32, так что облако содержит 32x150 точек. На этом же рисунке приведена теоретическая зависимость (4). По рисунку хорошо видно, что точки группируются около теоретической кривой.

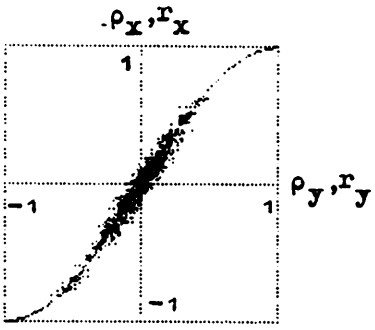


Рис. 1

Реальную функцию, связывающую корреляции двух процессов, можно построить, используя, например, метод наименьших квадратов. Однако в этом нет необходимости, так как ясно, что с ростом объема выборки эмпирическая кривая будет сходиться к теоретической. С другой стороны, проверить теорию и убедиться в сходимости можно путем подстановки в (4) прямой оценки

$r_y(m)$, что даст обратную оценку

$$\tilde{r}_x(m) = e_1[r_y(m)] = \sin\frac{\pi}{2} r_y(m), \qquad (6)$$

и затем сравнить прямую оценку $r_x(m)$ с обратной $\tilde{r}_x(m)$. На рис.2 приведены графики оценок $r_x(m)$ и $r_y(m)$, а на рис.3 совмещены графики $r_x(m)$ и $\tilde{r}_x(m)$. Они иллюстрируют характер искажений корреляционной функции (рис.2) и степень компенсации этих искажений (рис.3), что подтверждает теоретическую зависимость (4). Этот же прием используется ниже при анализе результатов обработки речевых сигналов.

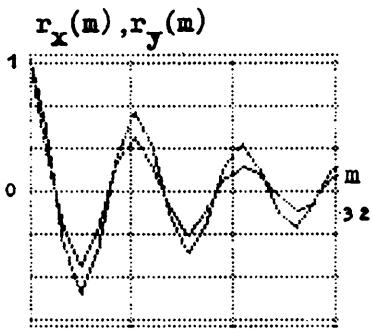


Рис. 2

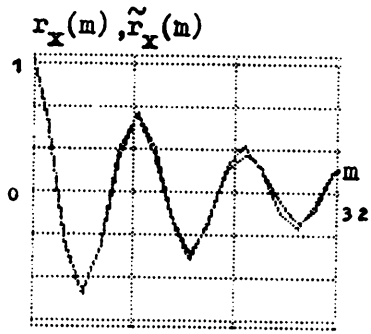
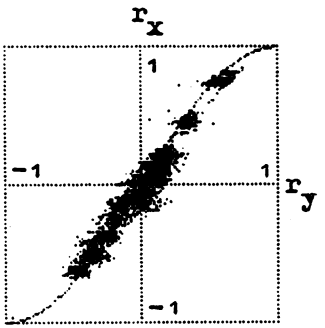
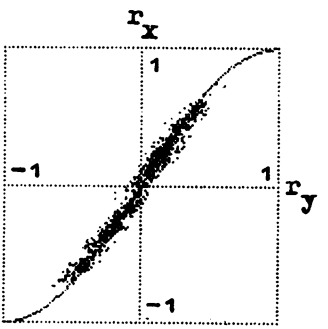


Рис. 3

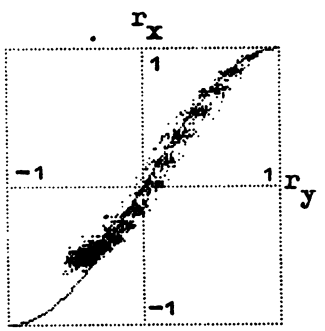
Если клиппированию по формуле (2) подвергается речевой сигнал, то соотношение (4) может не выполняться, так как функция распределения для речевого сигнала отлична от гауссовской. Тем не менее, можно попытаться установить зависимость $r_x(m) = \hat{\theta}_1[r_y(m)]$ аналогично тому, как это было сделано для гауссовского процесса. На рис.4 приведены эмпирические оценки функции θ_1 в виде облака точек (r_x, r_y) для различных фонем



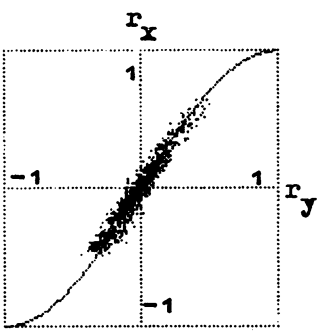
фонема /о/
диктор Д1



фонема /х/
диктор Д1



фонема /у/
диктор Д2



фонема /ш/
диктор Д2

Рис. 4. Зависимость между корреляционными функциями речевого сигнала до и после клиппирования для отдельных фонем

русского языка, полученные по сигналам двух дикторов-мужчин (Д1 и Д2). Как и в случае гауссовского процесса, точки облака тяготеют к кривой, описываемой формулой (4). На рис. 5 приведены аналогичные данные, рассчитанные по фразе "Вода в луже медленно убывала" (диктор Д2).

Ввод речевого сигнала в ЭВМ осуществлялся при помощи восьмиразрядного аналого-цифрового преобразователя при частоте дискретизации 10 кГц. Производилась последовательная обработка сигнала при окне анализа, включающем $N = 256$ отсчетов. Величина задержки m изменялась от 1 до 16. Каждое облако на графиках содержит от 1600 до 2500 точек.

Полученные данные свидетельствуют о том, что и для речевого сигнала связь между его корреляционными функциями до и после клиппирования адекватно описывается формулой (4). Для большей наглядности на рис.6 приведены графики прямых $(r_x(m), r_y(m))$ и обратной $(\tilde{r}_x(m))$ оценок корреляций. Рис. 6а фиксирует заметное отличие оценок корреляций исходного и искаженного сигналов; рис.6б показывает, что для речевого сигнала компенсация искажения может производиться по формуле (4).

3. Оценивание зависимости между дисперсией и математическим ожиданием

Дисперсия сигнала является одной из ключевых характеристик в задачах автоматической обработки речи. Так, при синтезе она используется для управления громкостью речи, а при распознавании набор дисперсий сигнала, измеренных на выходах полосовых фильтров, составляет весьма распространенное информативное первичное описание.

Оценивание зависимости между дисперсией речевого сигнала и математическим ожиданием искаженного проведем на примере преобразования, которое в радиотехнике называют двухполупериодным линейным детектированием:

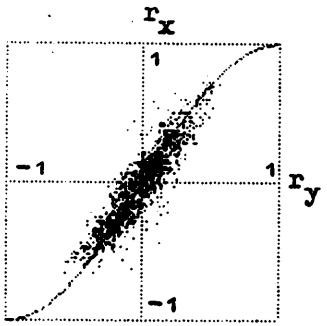
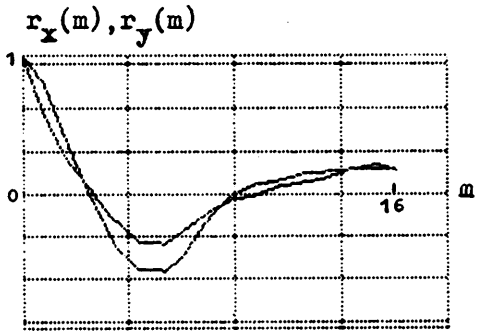


Рис. 5. Зависимость между корреляционными функциями речевого сигнала до и после клиппирования для фразы "Вода в луже медленно убывала"; диктор Д2

а)



б)

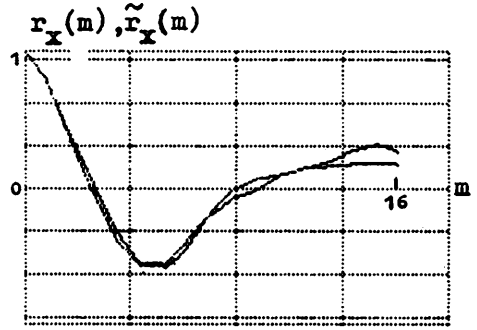


Рис. 6. Прямые и обратные оценки корреляций; диктор Д1, фонема /а/

$$y_n = f(x_n) = |x_n|, \quad n = 0, \pm 1, \pm 2, \ldots \qquad (7)$$

В то время как клиппирование может выступать и как искажение, и как искусственное преобразование речевого сигнала, нелиней - ное преобразование (7) вводится при цифровой обработке речи, как правило, искусственно. Если этому преобразованию подвергается стационарный нормальный процесс $\{x_n\}$, имеющий $M_x = 0$ и ковариационную функцию $\sigma_x(m) = \sigma_x(0)\,\rho_x(m)$, то математическое ожидание искаженного процесса $\{y_n\}$ и дисперсия неискаженного будут связаны формулой [6]:

$$\sigma_x(0) = \theta_2(M_y) = \frac{\pi}{2}(M_y)^2 . \qquad (8)$$

Поскольку $c_x(0)$ (см. формулу (1)) и выборочное среднее $\bar{y}$ являются состоятельными оценками дисперсии и математического ожидания, а θ_2 непрерывна, оценивание зависимости между $\sigma_x(0)$ и M_y^2 при больших N можно проводить по их оценкам. В силу линейности зависимости (8) между дисперсией исходного процесса и квадратом математического ожидания искаженного достаточно оценить отношение

$$\gamma = \frac{c_x(0)}{\bar{y}^2} , \qquad (9)$$

которое для гауссовских процессов должно быть близко к $\pi/2$.

При помощи гауссовского датчика случайных чисел генерировался ряд, содержащий $N = 1024$ отсчетов. Вычисленная по формуле (9) выборочная величина $\gamma = 1.566$ близка к $\pi/2$. Поэтому оценку дисперсии неискаженного гауссовского процесса $\{x_n\}$ можно искать в виде обратной оценки по математическому ожиданию искаженного $\{y_n\}$:

$$\tilde{c}_x(0) = \theta_2(\bar{y}) = \frac{\pi}{2}(\bar{y})^2 . \qquad (10)$$

Вопрос лишь в том, достаточно ли традиционного объема выборки (200-300 отсчетов), использующегося при обработке речевых сигналов, для удовлетворительного оценивания. Для ответа на этот вопрос генерировалось 10000 выборок гауссовского процесса авторегрессии второго порядка (5), содержащих по 256 отсчетов. По каждой выборке вычислялась оценка величины γ. Среднее значение (по всем выборкам) $\overline{\gamma} = 1.547$ примерно на 1.5% отличается от теоретического значения, что с большим запасом подходит для обработки речи.

Оценивание коэффициента γ для речевых сигналов проводилось на выборках, содержащих по $N = 256$ отсчетов. В эксперименте участвовал один диктор-мужчина. Для гласных звуков средняя величина $\overline{\gamma}$, подсчитанная по 270 выборкам или сегментам, равна 1.697 (стандартное отклонение равно 0.104); для звонких согласных (на 350 выборках) оценки величин γ и стандартного отклонения равны 1.937 и 0.166, а для глухих согласных (на 400 выборках) - 1.619 и 0.114 соответственно. В среднем для всех звуков (фонем) $\overline{\gamma} = 1.686$ при стандартном отклонении 0.123. Полученная оценка γ отличается от $\pi/2$. Поэтому при необходимости в (10) вместо коэффициента $\pi/2$ следует подставлять оцененную величину. Однако на практике (для многих приложений) этого не требуется.

При обработке слитной речи (39 тестовых фраз, содержащих около 2500 выборок) полученные оценки $\overline{\gamma}$ и стандартного отклонения равны соответственно 2.343 и 1.118. Значительное отклонение эмпирической величины $\overline{\gamma}$ от теоретической происходит потому, что слитная речь характеризуется наличием участков (пауз), на которых средняя мощность сигнала близка к нулю. Для этих участков вычисления по формуле (9) происходят с заметной потерей точности из-за эффекта типа деления ноль на ноль. Поэтому можно предположить, что формула (9) будет полезной при автоматическом поиске пауз в потоке речи.

4. Обсуждение результатов. Выводы

Полученные результаты свидетельствуют о том, что, несмотря на отличие функции распределения речевого сигнала от гаусовской, эмпирические зависимости между корреляционными функциями, а также между дисперсией и математическим ожиданием речевого сигнала до и после рассмотренных нелинейных искажений могут с достаточной для практики точностью аппроксимироваться теоретическими зависимостями для гауссовского случая. Поскольку аппроксимация негауссовского речевого сигнала гауссовским случайным процессом сохраняет разборчивость речи и узнаваемость диктора [3], можно заключить, что и для других типов нелинейных амплитудных искажений теоретические результаты будут вполне пригодны для практического использования. Таким образом, если вид амплитудного искажения известен и оно необратимо, то при получении оценок ковариационной функции неискаженного речевого сигнала решение задачи оценивания зависимостей между моментами можно опустить.

В том случае, когда имеются наблюдения искаженного и неискаженного сигналов, но вид амплитудного искажения заранее не известен, перед ковариационной компенсацией (т.е. перед получением оценок ковариаций неискаженного сигнала) необходимо оценить (аппроксимировать и, если необходимо, затабулировать) эмпирическую зависимость между моментами. Это без особого труда может быть сделано, например, при помощи метода наименьших квадратов.

При цифровой обработке речи обычно используется скользящее окно анализа, содержащее 200-300 значений сигнала при частоте квантования 10 кГц. Как показывают результаты экспериментов, подобного объема выборки вполне достаточно для удовлетворительного оценивания эмпирических зависимостей между первыми и вторыми моментами сигнала до и после искажений. Здесь следует отметить, что при обработке стационарных гауссовских про-

цессов повысить точность оценивания можно путем увеличения длины выборки ряда. При обработке же речи поступать таким образом неправомерно, так как в этом случае сигнал нельзя считать стационарным.

Нетрудно понять, что описанная процедура оценивания зависимости между корреляциями речевого сигнала может без изменений использоваться в системах распознавания речи для адаптивной подстройки под голос диктора путем коррекции оценок корреляций. В этом случае подстройка под голос нового диктора, не обучавшего систему распознавания, будет заключаться в пересчете эталонных корреляций по косвенным или искаженным наблюдениям сигнала другого диктора.

Литература

1. САПОЖКОВ М.А. Речевой сигнал в кибернетике и связи. - М.: Связьиздат, 1963. - 450 с.

2. РЕПИНА О.И. Искажения в телефонном тракте. - М.:Связь, 1978. - 174 с.

3. КЕЛЬМАНОВ А.В., ХАЙРЕТДИНОВА А.Г. Исследование свойств искаженных речевых сигналов //Анализ данных и знаний в экспертных системах. - Новосибирск, 1990. - Вып. 134: Вычислительные системы. - С. 140-160.

4. МАРКЕЛ Дж.Д., ГРЭЙ А.Х. Линейное предсказание речи: Пер. с англ. /Под ред. Ю.Н.Прохорова, В.С.Звездина.-М.: Связь, 1980. - 308 с.

5. КЕЛЬМАНОВ А.В. Сравнение систем признаков, основанных на частной автокорреляционной функции, при решении задачи распознавания изолированных слов //Эмпирическое предсказание и распознавание образов. - Новосибирск, 1980. - Вып. 83: Вычислительные системы. - С. 74-97.

6. ЛЕВИН Б.Р. Теоретические основы статистической радиотехники. Т.1. - М.: Сов. радио, 1974. - 552 с.

7. КЕЛЬМАНОВ А.В. Метод обратных оценок в задачах первичной обработки речевых сигналов //Автоматическое распознавание и синтез речевых сигналов. - Киев, 1989. - С. 20-22. (Сб. Тр. Ин-та кибернетики АН УССР).

8. АНДЕРСОН Т. Статистический анализ временных рядов.: Пер. с англ./ Под ред. Ю.К.Беляева. - М.: Мир, 1976. - 755 с.

Поступила в ред.-изд.отд.
22 февраля 1991 года