

УДК 621.391

О МЕТОДАХ ОЦЕНИВАНИЯ ПАРАМЕТРОВ СИГНАЛА

Алюнов Д.Ю.

ФГБОУ ВПО «Чувашский государственный университет им. И.Н. Ульянова», Чебоксары, Россия (428015, г. Чебоксары, Московский проспект, д.15), e-mail: aldmity89@gmail.com

Приведены способы оценивания спектральной плотности мощности сигналов, применение их для оценивания параметров речи. Показаны недостатки классических методов спектрального оценивания – периодограммного и коррелограммного, их зависимость от длины анализируемого сигнала, эффекты растекания спектра, свойство обмена частотного разрешения на гладкость оценки, применение окон для сглаживания спектра, особенности окон, зависимость качества оценивания от процента перекрытий окон. Продемонстрированы преимущества и недостатки параметрических методов (метод Берга, ковариационный метод, модифицированный ковариационный метод), их особенности – эффекты просачивания, маскировки слабого сигнала сильным, их преимущества – возможность выделять нужные компоненты сигнала на более коротких отрезках, по сравнению с классическими методами. Описан подход определения сегментов слов в слитной речи, позволяющий с высокой степенью точности определять слова с высоким уровнем шума и фоновых звуков.

Ключевые слова: спектр, спектральная плотность мощности, анализ речи.

ABOUT METHODS ESTIMATING THE PARAMETERS OF SIGNALS

Alyunov D.Y.

Chuvash State University n.a. I.N. Ulianov, Cheboksary, Russia(428015, Cheboksary, Moscow Prospect,15), e-mail: aldmity89@gmail.com

There are ways of estimating the power spectral density of the signals, their use for the evaluation of speech parameters. The drawbacks of classical methods of spectral estimation – periodogram and korrelogrammno, their dependence on the length of the analyzed signal, the effects of the spreading of the spectrum, a property exchange frequency resolution on the smoothness of assessment, the use of windows for smoothing the spectrum, especially the windows, the dependence of the quality of estimation of the percentage of overlapping windows. The advantages and disadvantages of parametric methods (Berg, covariance method, the modified covariance method), their features – trickle-down effect, masking weak signal stronger, their advantages – the ability to allocate the necessary components of the signal at shorter intervals, compared to the classical methods. The approach segment definitions of words in continuous speech, which allows a high degree of accuracy to identify words with a high level of noise and background.

Keywords: spectrum, power spectral density, speech analysis.

Распознавание речи с каждым годом находит все большее применение в нашей жизни (цифровая передача и хранение данных, синтез речи, идентификация диктора, устранение дефектов речи, улучшение параметров речевого сигнала), разрабатываются новые алгоритмы. Разработаны модели, описывающие речь ее информационным содержанием, модели, представляющие речь сигналом и т.п. Обработка сигнала представляет собой формирование описания на основе некоторой модели с последующим преобразованием полученного представления. Заключительным шагом является выделение необходимых параметров сигнала, их анализ и работа с ними.

Исторически распространенными классическими методами спектрального оценивания являются периодограммный и коррелограммный методы оценивания спектральной плотности мощности (СПМ) сигнала.

Спектральная плотность мощности (СПМ) стационарного случайного процесса есть дискретно-временное преобразование Фурье (ДВПФ) автокорреляционной последовательности (r_{xx}) [2,3].

$$P_{xx}(f) = T \sum_{m=-\infty}^{\infty} r_{xx}[m] e^{-j2\pi f m T} \quad (1)$$

Если допустить, что процесс является эргодическим, то:

$$P_{xx}(f) = \lim_{M \rightarrow \infty} \left(\frac{1}{(2M+1)T} \left| T \sum_{n=-M}^M x[n] e^{-j2\pi f n T} \right|^2 \right) \quad (2)$$

Если определять спектральную плотность мощности исходя из автокорреляционной последовательности, получится коррелограммный метод, поскольку случайный процесс непосредственно не используется для оценки СПМ. В том случае, если использовать саму числовую последовательность для оценки СПМ, получится периодограммный метод.

Оценка СПМ, получаемая на основе коррелограммного метода, принимает форму:

$$P_{xx}(f) = T \sum_{m=-L}^L w[m] \hat{r}_{xx}[m] e^{-j2\pi f m T} \quad (3)$$

Среднее значение этой оценки будет сверткой истинного спектра и спектра окна $W(f)$:

$$\langle \hat{P}(f) \rangle = T \sum_{m=-L}^L w[m] r_{xx}[m] e^{-j2\pi f m T} = P_{xx}(f) * W(f) \quad (4)$$

Правильный выбор окна позволит уменьшить растекание спектра и его смещение.

При наличии конечного множества данных $x(n)$, $0 \leq n \leq N-1$ и единственной реализации, это соотношение преобразуется в СПМ выборки или периодограмму [2,3]:

$$\tilde{P}_{xx}(f) = \frac{1}{NT} \left| T \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n T} \right|^2 = \frac{T}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n T} \right|^2 \quad (5)$$

На рисунке (1) представлены оценки СПМ периодограммным методом Уэлча без усреднения и с усреднением по 18 сегментам соответственно. Усредненная оценка имеет гораздо меньшую дисперсию, однако пришлось пожертвовать разрешением спектральных компонент. На графике слева изображена оценка СПМ суммы двух комплексных экспонент с относительными частотами 0.35 и 0.36 периодограммой Уэлча, 1 сегмент – длина окна 1000 отсчетов, SNR=6. Справа – периодограмма Уэлча, 18 сегментов – длина окна 100 отсчетов, сдвиг сегмента 50 отсчетов, SNR=6.

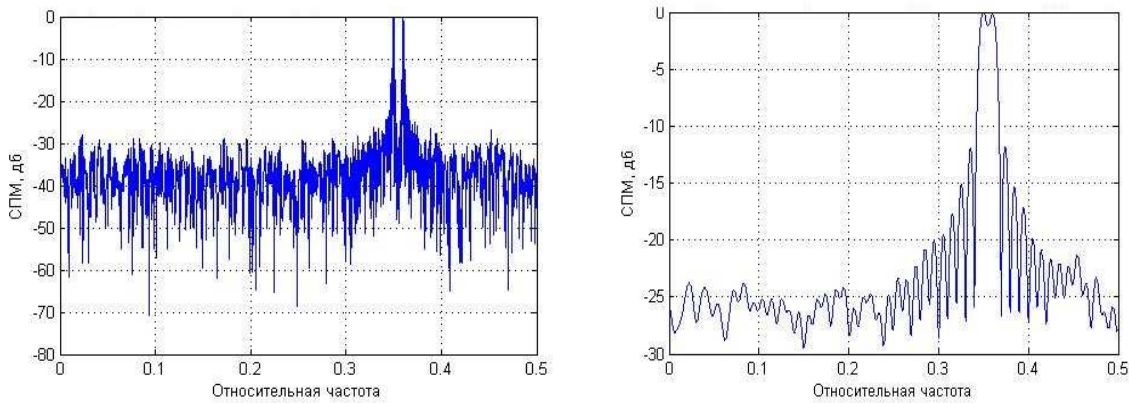


Рис.1. Влияние усреднения на качество периодограммного метода оценивания СПМ

В данном примере использовалось прямоугольное окно. Ширина его главного лепестка уже, чем, например, у окна Хэмминга, поэтому и разрешение спектральных компонент эффективнее. Применение окна уменьшает эффекты просачивания и маскировки вследствие того, что уширение главного лепестка спектра окна происходит за счет уменьшения уровней боковых лепестков. Качество разрешения зависит от количества взятых отсчетов. Аналогичная ситуация и в коррелограммных оценках СПМ близких гармоник. Но имеются некоторые особенности.

Многие случайные процессы дискретного времени описываются следующей моделью [4]:

$$x[n] = -\sum_{k=1}^p a[k] x[n-k] + \sum_{k=0}^q b[k] u[n-k] = \sum_{k=0}^{\infty} h[k] u[n-k] \quad (6)$$

Это можно представить в виде выхода фильтра, где $x[n]$ – входная последовательность каузального фильтра, $u[n]$ – входная возбуждающая последовательность (белый шум с нулевым средним и дисперсией ρ_w), $h[n]$ – импульсная характеристика фильтра, $a[k]$ – коэффициент авторегрессии, $b[k]$ – коэффициент скользящего среднего.

Процесс на выходе фильтра (8) соответствует модели авторегрессии – скользящего среднего (АРСС), где параметры $a[k]$ характеризуют авторегрессионную часть этой модели порядка p , а параметры $b[k]$ – ее часть, соответствующую скользящему среднему порядка q [4].

Спектральная плотность мощности для АРСС процесса имеет вид

$$P_{\text{АРСС}}(f) = T \rho_w |B(f)A(f)|^2 \quad (7)$$

$$A(f) = 1 + \sum_{k=1}^p a[k] \exp(-j2\pi kTf), \quad (8)$$

$$B(f) = 1 + \sum_{k=1}^q b[k] \exp(-j2\pi kTf), \quad (9)$$

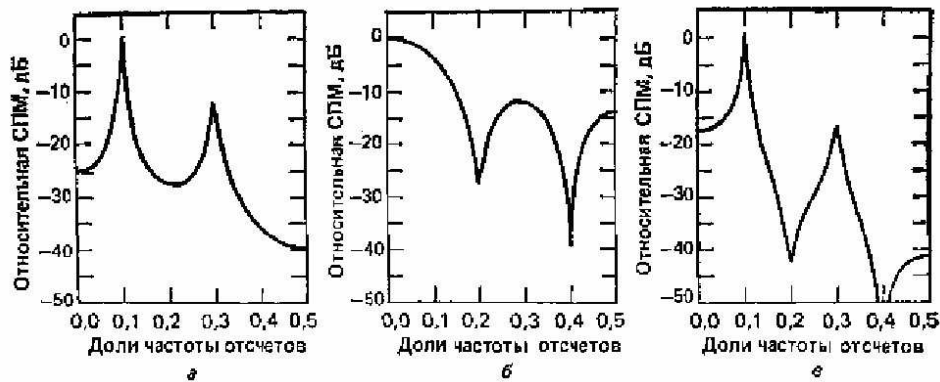


Рис. 2. Примеры оценивания СПМ сигнала с использованием параметрических методов

Первый из параметрических методов – метод Берга, гораздо точнее оценивал результаты. Но он был не без недостатков: расщепление спектральных, эффект смещения спектральных, зависящий от начальной фазы гармоник. В ковариационных методах эти недостатки были устранены (рис. 3 слева – сигнал смеси 2-х синусоид с относительными частотами 0.3 и 0.34; SNR=30). Данные параметрические методы получили широкое распространение в ЦОС, поскольку они позволяют получать высокое разрешение и острые спектральные пики спектральных компонент. При низком порядке модели получаются более сглаженные спектральные оценки, при излишне высоком – увеличивается разрешение, но в спектре появляются ложные пики.

Авторегрессионные методы имеют гораздо лучшее разрешение спектральных компонент по сравнению с классическими методами спектрального оценивания. На рисунке (3) справа очевидно превосходство параметрических алгоритмов ЦСА. Красной линией изображена периодограмма, а синей линией – модифицированный ковариационный метод. Длина последовательности выбрана в 30 отсчетов. Спектр сигнала состоит из 2 спектральных пиков на частотах 0.3 и 0.35.

При использовании классических алгоритмов спектрального оценивания следует учитывать произведение «устойчивость*длительность*ширина полосы»; имеет место свойство обмена частотного разрешения на гладкость оценки; характерны эффекты маскирования и растекания спектра.

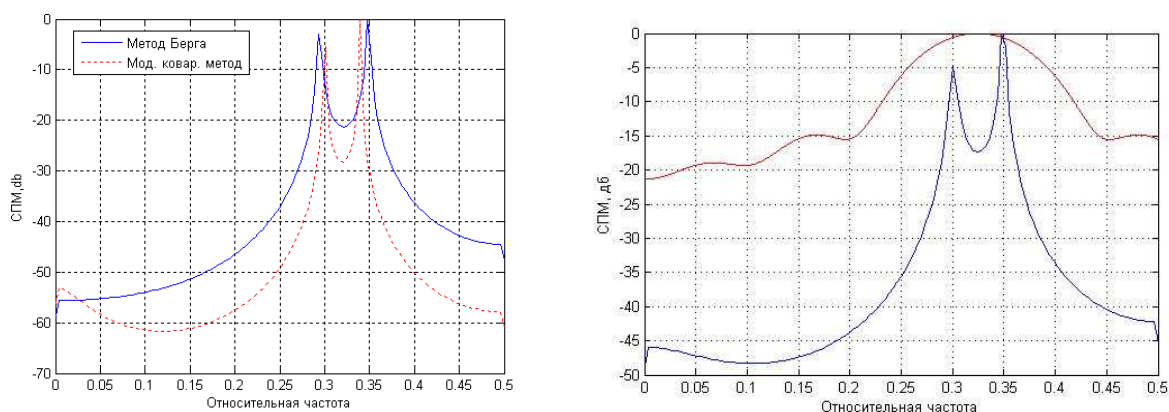


Рис. 3. Сравнение работы модифицированного ковариационного метода и метода Берга – слева. Сравнение качества оценивания двух спектральных компонент – справа

Увеличение порядка АР-модели сопровождается улучшением частотного разрешения, однако при избыточном порядке модели возникают ложные спектральные пики; для всех анализируемых методов характерно следующее свойство: при увеличении числа анализируемых отсчетов сигнала или порядка модели частотное разрешение повышается, однако дисперсия оценки СПМ увеличивается.

Интересным выглядит изучение информационной энтропии спектра сигнала. Как известно, энтропия шумового сигнала и речевого сигнала отличается, и, что можно отнести к преимуществам данного метода, энтропия мало чувствительна к амплитуде сигнала [5,6,7]. Принцип работы изображен на рисунке (4).

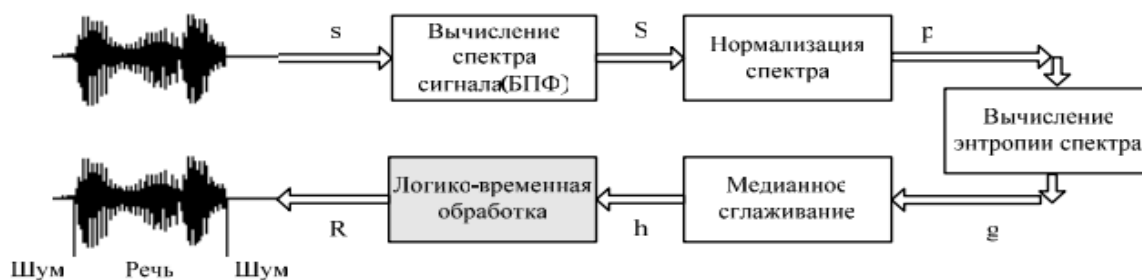


Рис. 4. Анализ речевого сигнала с использованием энтропии спектра. Алгоритм Как происходит обработка сигнала: сигнал дискретизируется, затем делится на сегменты по 256 цифровых отсчетов, перекрытие сегментов сделаем немногим более 25 % для устранения краевых сегментов. Мгновенные спектры мощности сигнала рассчитываем по следующей формуле:

$$S_i(k) = \left| \sum_{m=0}^{M-1} s_i(m) e^{-j \frac{\pi}{M} mk} \right|^2, \quad 0 \leq k \leq M \quad (10)$$

M – размер сегмента сигнала.

Дальше нормализуем спектр по всем частотным компонентам:

$$p_i = \frac{S(f_i)}{\sum_{k=1}^N S(f_k)} \quad (11)$$

Таким образом, мы получили плотность вероятности спектра. Мы ее ограничиваем: верхним и нижним пределом. Если есть равномерное распределение частотных компонент – это белый шум, а также мы исключаем шумы в узкой частотной области.

$$p_i = \begin{cases} 0, & p_i > \delta_1 \\ 0, & p_i < \delta_2 \\ p_i, & \text{иначе} \end{cases} \quad (12)$$

δ_1 и δ_2 – верхний и нижний пределы.

Из экспериментальных расчетов можно оценить порядок пределов как 0,01 и 0,3 при p от 0 до 1. Кроме того, в дополнение мы можем использовать и другие способы выделения сигнала из шума: методы спектрального вычитания и адаптивный фильтр Кальмана.

Мы рассматриваем энтропию как меру беспорядка в распределении, рассчитывая по следующей формуле:

$$H = -\sum_{k=1}^N p_k \log p_k \quad (13)$$

Полученную функцию необходимо сгладить, используя медианное сглаживание. Данный тип сглаживания является наиболее устойчивым по отношению к случайным выбросам. Для этого берется какой-либо интервал $[t-q, t+q]$ и вычисляется скользящая медиана в точке t . Медиана ряда в интервале определяется как центральный член последовательности значений ряда, входящих в этот временной интервал, упорядоченной по возрастанию. Как показывают эксперименты, наиболее точное вычисление медианы происходит при окне величиной в 5 сегментов. В том случае, когда мы вычисляем медиану в точках, близких к краю интервала и меньших, чем размер окна, приходится уменьшать окно [1].

Затем мы вычисляем порог для определения границ речевого сигнала.

$$r = \left(\frac{\max(h) - \min(h)}{2} + \min(h) \right) * \mu \quad (14)$$

, μ – коэффициент зашумленности.

Он подбирается экспериментально, зависит от параметров шума. Данный коэффициент может принимать значения от 0,8 до 1,1 в зависимости от уровня шума. На основе вычисленного значения r выбираются акустические сегменты речевого сигнала.

Заключительным этапом является логическая временная обработка полученной энтропии спектра, используя допустимые на практике длительности речевых и неречевых сигналов, вычисленных ранее. Это необходимо, поскольку зачастую различные звуковые эффекты (кашель и прочее) принимают за речь, а некоторые участки речи за межречевой интервал. Используя адаптивный порог, можно определить сегменты речи на основе вычисления максимальной длительности межречевого участка S и минимальной длительности участка речи (рисунок 2).

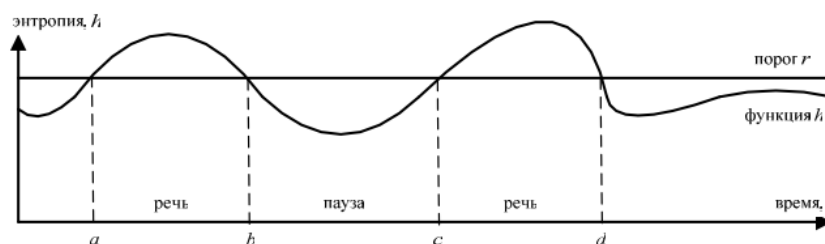


Рис. 5. Энтропия и порог обнаружения

Поскольку человек чисто физически не может произносить речевые фрагменты короче определенного значения, и так как всегда присутствуют паузы, можно экспериментально определить значения параметров R и S . Анализируем полученные результаты следующим образом:

$$\text{речь} = \begin{cases} ad, (ab \geq R) \wedge (cd \geq R) \wedge (bc \leq S) \\ ab, (ab > cd) \wedge (ab \geq R) \\ cd, (ab < cd) \wedge (cd > R) \\ \emptyset, \text{иначе} \end{cases} \quad (20)$$

Если участок сигнала без речи содержит не более S сегментов, а участки, содержащие речь, составляют не менее R сегментов, то образуется сплошной речевой участок.

Экспериментальные расчеты показывают, что наименьший процент ошибок получается в случае узкополосного или белого шума – порядка 1,5 %. Наихудший результат – 16 % ошибок в случае розового шума, наиболее приближенного к реальной речи. К минусам данного метода можно отнести тот факт, что фоновая речь, пусть даже слабая, может быть принята за полезный сигнал. В целом данный метод довольно неплохо определяет речь в сигналах с высоким уровнем шумов и нестационарности.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 14-07-00143.

Список литературы

1. Бияков, О.А. Медианное сглаживание временных рядов / О.А. Бияков // Вестник КузГТУ. – 1999. – № 3. – С. 55-56.
2. Гоноровский И.С. Радиотехнические цепи и сигналы: учебник для вузов. – 4-е изд., перераб. и доп. – М.: Радио и связь, 1986.
3. Кривошеев В.И. Современные методы цифровой обработки сигналов (цифровой спектральный анализ): учебно-методические материалы. – Н. Новгород: ННГУ, 2006. – 117 с. <http://www.unn.ru/pages/e-library/aids/2006/7.pdf>
4. Кэй С.М., Марпл С.Л. Современные методы спектрального анализа: обзор // ТИИЭР. – 1981. – Т. 69. – № 11. – С. 5–50.
5. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 496 с.
6. Цымбал, В.П. Теория информации и кодирование / В.П. Цымбал. – Киев: Высшая школа, 1977. – 288 с.
7. Waheed, K. A robust algorithm for detecting speech segments using an entropy contrast / K. Waheed, K. Weaver, F. Salam // Proceedings of MWSCAS'2002, Oklahoma, USA, 2002.

Рецензенты:

Артемьев И.Т., д.ф.-м.н., профессор, зав. кафедрой математического и аппаратного обеспечения информационных систем (МиАОИС) ФГБОУ ВПО «Чувашский государственный университет им. И.Н. Ульянова», г. Чебоксары;

Охоткин Г.П., д.т.н., профессор, декан факультета радиоэлектроники и автоматики (ФРЭА), ФГБОУ ВПО «Чувашский государственный университет им. И.Н. Ульянова», г. Чебоксары.