

УДК 621.391

УМЕНЬШЕНИЕ ОБЪЕМА ДАННЫХ ПРИ ШИРОКОПОЛОСНОМ КОДИРОВАНИИ РЕЧЕВОГО СИГНАЛА

Афанасьев А.А., Академия ФСО России, 8-(4862)-41-99-47

Трубицын В.Г., Академия ФСО России, 8-(4862)-41-99-32

Ключевые слова: речевой сигнал, кодирование, поток данных, скорость передачи, децимация, алгоритм, спектральная плотность мощности, корреляция.

Для повышения естественности и узнаваемости звучания синтезированного речевого сигнала в последние годы наблюдается тенденция к расширению спектральной области кодируемого сигнала до 7 кГц. Переход к кодированию широкополосного речевого сигнала определен тем, что ранее в телефонных сетях ограничивали полосу частот речевого сигнала в диапазоне 300-3400 Гц. Это ограничение ухудшало качество речи в части естественности, разборчивости и узнаваемости диктора. Однако переход к цифровым сетям связи позволяет увеличить анализируемую полосу частот речевых сигналов. В настоящее время к кодированию и передаче широкополосной речи с диапазоном частот 50-7000 Гц возникает достаточно большой интерес, так как расширение низкочастотного диапазона вниз до 50 Гц способствует большей естественности, а высокочастотный диапазон 3400-7000 Гц улучшает разборчивость. Кодирование широкополосной речи находит применение в конференцсвязи на основе протоколов IP-телефонии и видеоконференцсвязи, но постепенно оно начинает использоваться в большинстве наиболее важных приложений систем телекоммуникаций, связанных с кодированием и передачей речевых сигналов [1]. При кодировании широкополосной речи частота дискретизации выбирается обычно 16000 Гц, в зависимости от реализации алгоритма кодирования [2].

Известны различные алгоритмы кодирования широкополосного речевого сигнала. В устройствах, реализующих данные способы, осуществляется анализ широкополосного речевого сигнала и его эффективное кодирование, при этом частота дискретизации всегда остается постоянной, что приводит к фиксированной, зачастую избыточной полосе частот [3].

Данный факт является недостатком данного класса устройств, так как анализируется достаточно большой объем данных, что связано с высокой фиксированной частотой дискретизации речевого сигнала при широкополосном кодировании. Использование такой частоты дискретизации не всегда является необходимым условием получения высокого качества синтезированной речи при кодировании, а ее фиксация на достаточно высоком уровне приводит к избыточному объему данных и соответственно достаточно высокой скорости передачи и высокой вычислительной сложности алгоритмов кодирования речевого сигнала.

Ставится и решается задача уменьшения скорости потока данных при широкополосном кодировании речи путем оценки ширины его спектральной плотности мощности на отрезках квазистационарности и использования последующей децимации.

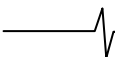
В настоящее время большинство вокодеров работают на фиксированной скорости, вне зависимости от характеристик входного сигнала. Так как вокодеры используют канал или сеть совместно с другими информационными потоками это отражается на качестве речевого обмена. Качество речи, предоставляемое вокодерами, является функцией скорости передачи, производительности коммутационного оборудования и задержек при обработке и предоставления трафика для речевого обмена [3]. В зависимости от предоставляемого трафика существует возможность управления скоростью кодирования для обеспечения переговоров требуемого качества. Основной проблемой цифрового представления широкополосного речевого сигнала является задача качественного и компактного кодирования данных для их передачи по цифровым каналам связи. Решение этой проблемы позволит в условиях заданного критерия качества связи увеличить пропускную способность линейных трактов и каналов передачи. Часто в некоторых задачах кодирования речевого сигнала предполагается снизить скорость передачи при сохранении качественных показателей ее восприятия. В кодеках речевых сигналов с переменной скоростью передачи, ориентированных на использование в системах связи, которые основаны на принципе коммутации пакетов, уместно говорить о снижении средней скорости передачи при сохранении качественных показателей синтезированного речевого сигнала.

Одной из ключевых характеристик при кодировании широкополосного речевого сигнала является понятие участка квазистационарности. Речевой сигнал не является стационарным на длительных участках времени, но возможно допущение о его стационарности на коротких промежутках времени.

Рассмотрим взаимосвязь между функцией автокорреляции $B(\tau)$ речевого сигнала на участке квазистационарности и его спектральной плотностью мощности $\sigma(\omega)$:

$$B(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sigma(\omega) \cdot e^{j\omega\tau} d\omega \quad (1)$$

$$\sigma(\omega) = \int_{-\infty}^{\infty} B(\tau) \cdot e^{-j\omega\tau} d\tau \quad (2)$$



Представленная пара преобразований Фурье позволяет определить спектральную плотность мощности по автокорреляционной функции случайного речевого сигнала на основании теоремы А.Я. Хинчина и Н. Винера. При этом необходимо учитывать требования стационарности для случайного речевого сигнала [4]. В существующих стандартах речевого кодирования с использованием линейного предсказания эти промежутки варьируются от 5 до 20 мс. Длительности звуков русской речи находятся в пределах от 80 до 200 мс для гласных и от 20 до 240 мс для согласных звуков [5].

В соответствии с (1) и (2) можно определить другие важные числовые показатели, характеризующие случайный речевой сигнал. К ним можно отнести интервал корреляции $\tau_{кор}$ и эффективную полосу частот $F_{эф}$, определяющую полосу в которой сосредоточено 90 – 95% мощности.

Между показателями $\tau_{кор}$ и $F_{эф}$ существует взаимосвязь вида

$$\tau_{кор} \cdot F_{эф} \approx K.$$

Значение константы K для различных моделей случайных сигналов имеет различные величины. Значение $\tau_{кор}$ определяется точкой первого перехода автокорреляционной функции через ноль. При анализе речевых сигналов принято считать $K = 0,5$, следовательно, $F_{эф} = 0,5 / \tau_{кор}$.

В упрощенном виде статистическая модель речевого сигнала может быть представлена аналитическим выражением автокорреляционной функции вида [4]:

$$B(\tau) = 2N_0 F_{max} \frac{\sin 2\pi F_{max} \tau}{2\pi F_{max} \tau},$$

где F_{max} - максимальное значение частоты спектра речевого сигнала, N_0 - спектральная плотность мощности.

Для принятой модели характерна идеализация реального процесса, состоящего из множества спектральных составляющих с одинаковой энергией. Значения автокорреляционной функции равны нулю в точках интервала корреляции кратных $\tau = 1 / 2F_{max}$. Это значит, что сечения такого процесса в этих точках не коррелированы, а интервал времени от нуля до $1 / 2F_{max}$ показывает сосредоточение энергии взаимосвязанного процесса, через поведение автокорреляционной функции (1). В рамках представленных рассуждений можно утверждать о соотношении $\tau_{кор} \cdot F_{эф} \approx 0,5$, при этом считать $F_{max} = F_{эф}$ [3, 4].

Принятая математическая модель речевого сигнала на участках его квазистационарности и установленные соотношения между $\tau_{кор}$ и $F_{эф}$ позволяют использовать их для аппроксимации реальных речевых сигналов, при этом правомерность применения данных моделей и параметров подтверждают результаты эксперимента.

На рис.1 и рис.2 представлены: временное представление, автокорреляционная функция и спектральное представление звуков "А" и "Ш", соответственно. Аналого-цифровое преобразование речевого сигнала реализовано на временных интервалах длительностью в 24 мс с частотой дискретизации 22050 Гц с использованием 16-битного квантователя.

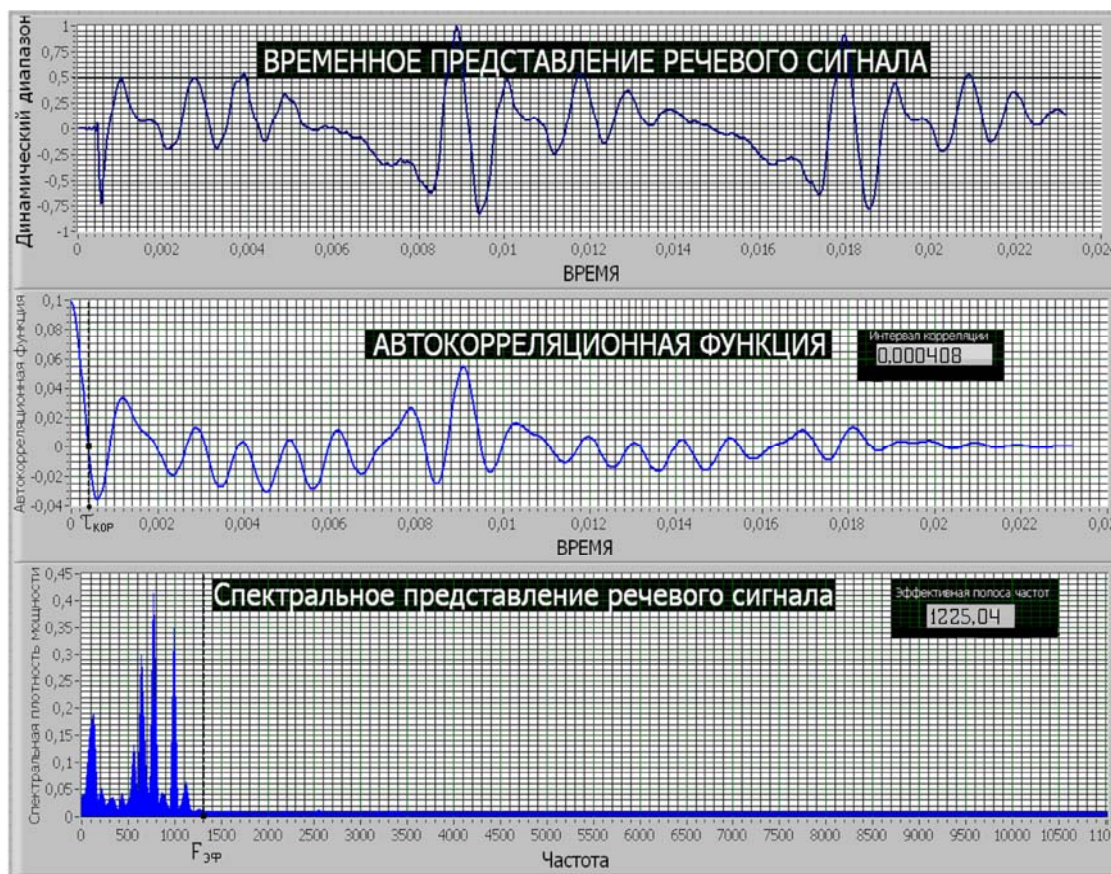


Рис.1. Временное представление, автокорреляционная функция и спектральная плотность мощности речевого сигнала при произнесении звука "А"

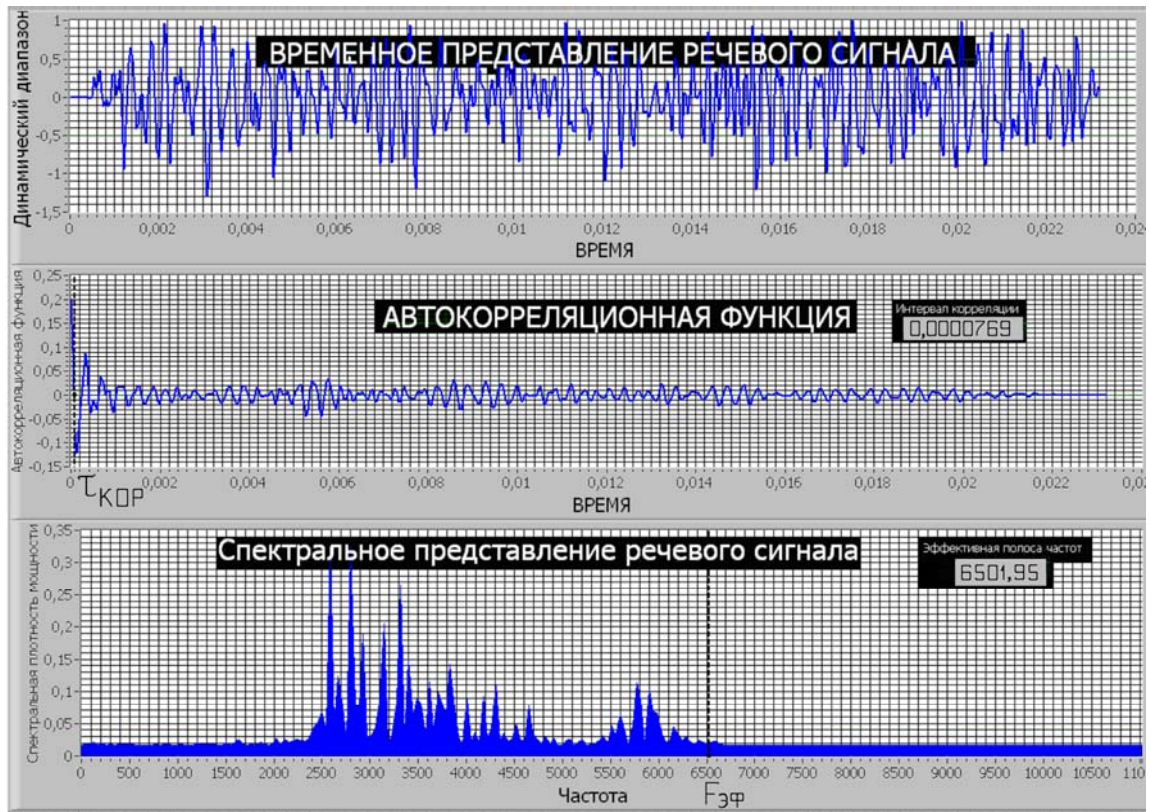


Рис.2. Временное представление, автокорреляционная функция и спектральная плотность мощности речевого сигнала при произнесении звука "Ш"

Анализ графиков кратковременной спектральной плотности мощности речевого сигнала показывает, что в полосе частот от 3 до 7 кГц появляются всплески спектральной энергии только при произнесении некоторой части согласных звуков (например, "ш", "с", "ф", "х"). Данный факт указывает на возможность адаптивного изменения границ спектральной области кодируемого широкополосного речевого сигнала для минимизации объема обрабатываемых данных при реализации процедуры сжатия и сохранении качественных показателей речи не ниже требуемых.

Изменение границ спектральной области кодируемого речевого сигнала можно реализовать с использованием процедуры децимации с дробными коэффициентами.

Соответствующие преобразования исходного дискретного сигнала с частотой дискретизации $f_D = 1/T$, T - интервал дискретизации исходного сигнала, в дискретный сигнал с частотой дискретизации $f_{DY} = 1/T_Y$, T_Y - интервал дискретизации преобразованного дискретного сигнала, осуществляются системами интерполяции и децимации, характеризующимися коэффициентами интерполяции $L = f_{DY} / f_D = T / T_Y$ и децимации $M = f_D / f_{DY} = T_Y / T$ (в простейшем случае – целочисленными) соответственно.

Помимо систем интерполяции и децимации с целочисленными коэффициентами L и M существуют системы преобразования частоты дискретизации с рациональным коэффициентом L/M . Они формируются по принципу последовательного выполнения интерполяции и децимации, а следовательно, являются многократными многоскоростными системами. Например, понижение частоты дискретизации в 1,25 раза может быть

реализовано последовательным соединением системы интерполяции с коэффициентом $L = 4$ и системы децимации с коэффициентом $M = 5$, таким образом, реализуется система децимации с коэффициентом $M=1,25$. Процедуры многоскоростной обработки интерполяция и децимация достаточно подробно описаны в [6].

К достоинствам способа следует отнести тот факт, что уменьшение объема данных при широкополосном кодировании речи приведет к снижению средней скорости передачи речевого сигнала по каналам цифровой связи, а также уменьшит количество требуемых вычислений при реализации процедуры кодирования.

Были проведены экспериментальные исследования для выявления возможности применения предлагаемого способа, которые показали уменьшение объема анализируемых данных речевого сигнала при широкополосном кодировании в среднем на 40%, при этом ошибки в представлении спектральной области речевого сигнала незначительны, что показано на рис.3.

Полученные результаты эксперимента позволяют утверждать, что предлагаемые модели описания случайных речевых сигналов через корреляционные и спектральные характеристики дают возможность адекватно оценивать количественные и качественные показатели параметров, характеризующих источник и огибающую спектра случайного речевого сигнала.

В зависимости от предлагаемого соотношения между параметрами $\tau_{кор}$ и $F_{эф}$ возникает возможность управления последующей обработкой речевого сигнала с целью снижения временных и вычислительных затрат по кодированию.

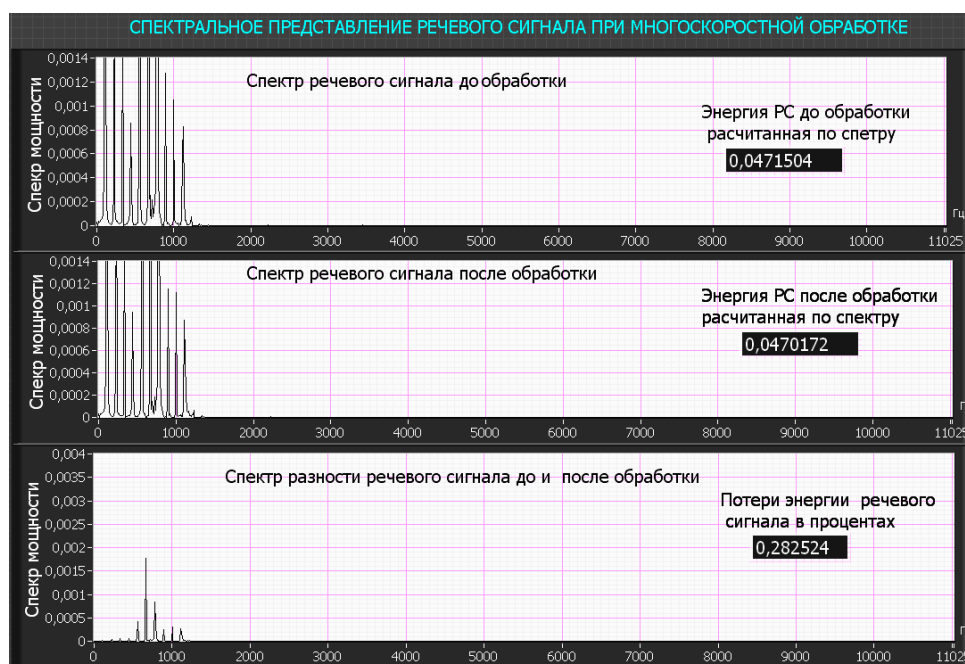


Рис.3 Спектральное представление речевого сигнала при многоскоростной обработке

Заключение

Таким образом, поставленная цель – уменьшение объема передаваемых данных, достигается тем, что при кодировании широкополосного речевого сигнала на выходе аналого-цифрового преобразователя речевой сигнал разделяют на участки квазистационарности, на каждом из которых вычисляют автокорреляционную функцию сигнала, после чего определяют значение интервала корреляции и вычисляют эффективную полосу частот. Далее рассчитывают коэффициент децимации и подвергают анализируемый речевой сигнал децимации, получая на выходе системы децимации необходимое и достаточное количество отсчетов для представления речевого сигнала при широкополосном кодировании.

Литература

1. М.З. Лившиц, М.Парфенюк, А.А. Петровский Широкополосный CELP-кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой структурой // Цифровая обработка сигналов – 2005 -№2. с.20-35.
2. Peter Noll Speech and audio coding for multimedia communications/ Proceedings International Cost 254 workshop on intelligent communication technologies and applications, Neuchatel, Schweiz, In print, 1999.
3. О.И. Шелухин, Н.Ф. Лукьянцев. Цифровая обработка и передача речи / Под ред. О.И. Шелухина. – М.: Радио и связь, 2000. 456 с.
4. А.Г. Зюко, Д.Д. Кловский, В.И. Коржик, М.В. Назаров. Теория электрической связи: Учебник для вузов / Под ред. Д.Д. Кловского. – М.: Радио и связь, 1998. – 432с.
5. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи. – М.: Радио и связь, 1987. – 168 с.
6. Солонина А. И., Улахович Д. А., Арбузов С. М., Соловьева Е. Б., Основы цифровой обработки сигналов: Курс лекций. Изд. 2-е испр. и перераб.– СПб.: БХВ – Петербург, 2005. – с.587-620).

7. С.Ф. Быков, В.И. Журавлев, И.А. Шалимов. Цифровая телефония: Учеб. пособие для вузов. – М.: Радио и связь, 2003. – 144с.

BROADBAND CODING OF THE SPEECH SIGNAL

Afnasjev A.,A., Trubicin V. G.

Materials of given article can be used in systems of teleinformation communications at coding of a wideband speech signal. A main objective of the presented work is reduction of data volume at coding of a wideband speech signal. The object in view is reached by that at coding of a wideband speech signal on an exit of the analogue-digital converter divide a speech signal into near stationary segments, on each of which calculate autocorrelation function of a signal after that define value of an interval of correlation and calculate an effective band of frequencies. Further count factor of decimation and subject an analyzed speech signal to decimation, receiving on a decimation system exit necessary and enough of readout for representation of a speech signal at wideband coding.

The received results allow to assert, that offered models of the description of speech signals through correlation both spectral characteristics and the established parity between an autocorrelation interval $\tau_{кор}$ and an effective band of frequencies $F_{эф}$ allows adequately to estimate quantitative and quality indicators of the parametres characterising a source and bending around spectrum of a casual speech signal.

Depending on an offered parity between $\tau_{кор}$ and $F_{эф}$ is arises possibility of management of the speech signal subsequent processing for the purpose of decrease in time and computing expenses on coding.