

Академия ФАПСИ
302034, Орел, ул. Игнатова, E-mail: aaa@academ.rfnet.ru

Настоящий доклад посвящен одному из новых подходов к обработке и сжатию речевых сигналов – кодированию речи на основе систем с переменной структурой. В нем даны основные положения данного подхода, предлагаются адаптивные алгоритмы обработки с учетом статистических и параметрических характеристик речи. Показаны основные пути уменьшения скорости передачи речевых сигналов и повышения качества восприятия кодированной речи.

Важнейшей тенденцией развития телекоммуникационных систем является широкое внедрение цифровой обработки и передачи сигналов, а также проникновение цифровых технологий в различные области жизни общества. Разработка и совершенствование систем речевого взаимодействия человека с человеком через цифровую технику или человека с компьютером является очень важной задачей. Разработка адаптивных алгоритмов и устройств аналого-цифрового и цифро-аналогового преобразования речевых сигналов, обеспечивающих передачу речи на заданной скорости с качеством, задаваемым потребителями, представляет собой достаточно сложную задачу.

В ряде исследований показано, что среднеквадратическая ошибка восстановленного сигнала полученная в известных кодеках близка к теоретически предельной [1], поэтому в ходе исследования предполагается использовать новый подход, суть которого состоит в построении математической модели и разработке алгоритмов кодирования речи на основе систем с переменной структурой, то есть предполагается изменять не только параметры, но и структуру кодирующего устройства. Методы теории систем с переменной структурой применяются, к примеру, в системах автоматического управления, но в кодеках речи практически не применяются.

Анализ методов, алгоритмов компрессии речевых сигналов привел к идее создания нового поколения систем обработки речевых сигналов, основанных на исследовании статистических и параметрических характеристик распределения параметров речевых сигналов и изменении в соответствии с ними структуры и параметров кодирующего устройства.

Под системами с переменной структурой будем понимать системы, в которых связи между функциональными элементами, количество элементов и их расположение меняются тем или иным образом в зависимости от свойств анализируемого кадра сигнала.

Методы и алгоритмы компрессии должны более точно учитывать свойства речевого сигнала. Необходимо адаптировать структуру кодирующего устройства на всех этапах обработки речевых сигналов.

Процессом изменения структуры и параметров кодирующего устройства в реальном масштабе времени должно управлять устройство, в котором будет происходить анализ и выделение параметров речевого сигнала. В соответствии с полученными характеристиками и сравнения их с эталонами будет выбрана одна из возможных структур кодирующего устройства, которая наиболее адекватно отобразит речевой сигнал на кадре анализа. Следующим шагом оптимизации структуры кодека будет количественный и качественный выбор числа параметров необходимых для кодирования речевого сигнала с заданным качеством.

Рассмотрим возможные подходы построения управляющего устройства на основе которого будет происходить изменение структуры кодирующего устройства. В [1] и [5] описываются четыре основные функции плотности вероятности по законам которых распределяются отсчеты речевого сигнала во временной области. Для различных законов распределения среднеквадратическая ошибка квантованного сигнала при применении одной и той же системы компрессии изменяется. Таким образом, если провести статистический анализ кадров речевых сигналов, выделив при этом функцию плотности вероятности, энергию сигнала на участке анализа (формантные области), оценку основного тона, то можно изменить систему компрессии в соответствии со структурой речевого сигнала. Такой классификатор строится на основе критерия согласия Колмогорова-Смирнова (1) с возможными аналитическими моделями функций плотности вероятности: Гауссова, равномерная, Лапласова, Гамма – распределений.

$$I(\tilde{w}(x)) = \inf \sup |w(x) - w'(x)|, \quad (1)$$

где $w(x)$ - функция плотности распределения случайного процесса;

$w'(x)$ - функция плотности известного закона распределения случайной величины.

В соответствии с полученными статистическими и параметрическими характеристиками выбирается та или иная структура кодирующего устройства, наиболее приспособленная для обработки данного участка речевого сигнала [2].

Для шумовых кадров речи основным методом кодирования было выбрано векторное квантования. Размер и структура построения кодовых книг изменятся в зависимости от статистических характеристик речи. Предлагается несколько алгоритмов построения кодовых книг:

1. Стандартный алгоритм К-средних.
2. Алгоритм максимального покрытия.
3. Алгоритм максимального покрытия с усреднением.
4. Рандомизированный алгоритм.

Первый алгоритм подробно описан в [1]. Остановимся на трех последних алгоритмах.

Алгоритм максимального покрытия предполагает нахождение и запись в кодовую книгу вектора максимально удаленного в смысле среднеквадратической ошибки от всех других векторов кодовой книги. Таким образом, осуществляется наиболее полный охват всего сигнального пространства. Алгоритм максимального покрытия с усреднением предполагает усреднение векторов кодовой книги в соответствии с алгоритмом К-средних. Рандомизированный алгоритм предполагает случайный выбор векторов из обучающей последовательности и их дальнейшее усреднение на основе обучающей выборки.

Для создания всех кодовых книг обучающей последовательностью используется речь дикторов мужского и женского пола длительностью не менее 15 мин. Основной тон на кадре анализа рассчитывается на основе кепстрального метода, кодируется и передается на прием. На основе оценки основного тона из памяти кодера выбираются кодовые книги предназначенные для работы либо на женских, либо на мужских голосах. Для тоновых и переходных кадров речи основным методом кодирования является адаптивное линейное предсказание.

При применении линейного предсказания речевых сигналов одним из возможных подходов является использование липредеров с ограниченным числом коэффициентов предсказания, предназначенных для передачи на участках речи с высокими корреляционными связями. Остальные коэффициенты полностью будут вычисляться на приеме.

В зависимости от полученных статистических характеристик принимается решение о передаче параметров характеризующих спектральную составляющую речевого сигнала. Такими параметрами являются линейные спектральные частоты (LSP) [3]. При произнесении гласных звуков статистическое распределение на нескольких кадрах анализа не изменяется, следовательно нет необходимости от кадра к кадру передавать заданное полное количество LSP, передается только разность между первыми - наиболее значимыми коэффициентами линейного предсказания, на приеме в синтезирующем фильтре происходит подстройка коэффициентов на каждом кадре. Возбуждение синтезирующего фильтра производится на основе векторного квантования остатка линейного предсказания [4], для чего используются четыре кодовые книги образцов векторов возбуждения.

Одним из недостатков предложенных алгоритмов компрессии речевых сигналов, является расчет параметров основного тона для всего кадра анализа. Большую возможность в плане повышения качества передачи РС может принести адаптивное слежение за параметром основного тона и формирование границы кадра анализа при инвертировании бита сигнала тон/шум. Такой подход позволит анализировать и кодировать кадры речи с помощью алгоритмов ориентированных непосредственно либо на шумовые, либо на тоновые звуки, либо на переходные звуки.

При разработке адаптивного кодера речевых сигналов необходимы два адаптивных контура управления: первый производит оптимизацию структуры кодирующего устройства, а второй - оптимизацию передаваемых параметров для выбранной системы компрессии в соответствии со статистическими и параметрическими характеристиками речевых сигналов. Такой способ обработки дает возможность при сокращении скорости передачи информации повысить ее качественные характеристики. Одним из перспективных направлений является формирование адаптивного кадра анализа на основе слежения за параметром основного тона и формирование границы кадра анализа при изменении параметрических характеристик речевого сигнала.

Литература:

1. Vector Quantization in Speech Coding. J.Makhoul, S.Roucos, H.Gish/IEEE v.73, №11. 1985.
2. Bernard Gold Digital Speech Networks: IEEE v.65, №12. 1977, pp.11-18.
3. Коротаев Г.А. Анализ и синтез речевого сигнала методом линейного предсказания. Зарубежная радиоэлектроника №3, 1990.
4. Коротаев Г.А. Некоторые аспекты линейного предсказания при анализе и синтезе речевого сигнала. Зарубежная радиоэлектроника №7.1991
5. О.И. Шелухин, Н.Ф. Лукьянцев Цифровая обработка и передача речи, М., Радио и связь, 2000г.

ADAPTIVE CODER FOR SPEECH SIGNALS BASED ON THE SYSTEMS WITH VARIABLE STRUCTURE

Afanasjev A.

Department of Radiotechnics and Electronics, Academy of Federal Agency for Government
Communication and Information for President of Russian Federation, Oryol, Russia, 302034.
E-mail: aaa@academ.rfnet.ru

The major tendency of telecommunication systems development is the wide introduction of digital processing and transfers of signals, and also penetration of digital technologies into various areas of a society life. The important feature of modern telecommunication systems status is that fact that the essential part of the world information traffic is created by telephone systems. A working out of adaptive algorithms both devices of analog-to-digital and digital-to-analog transformation of the speech signals providing a transfer of speech on given speed and quality set by the consumers is a enough complex task.

Some researches show us that mean squared mistake of a restored signal received in known coder / decoder is close to theoretically limit, therefore in our research it is offered to use a new approach. Essence is consists the construction of mathematical model and development of coding speech algorithms based on the systems with variable structure. Thus, it is necessary to change not only parameters, but also structure of the coding device. The methods of this approach are applied, for example, in systems of automatic control, but in speech coding devices are not applied practically.

The process of the coding device structure and parameters change in real time should be operated by the device in which the analysis and allocation of a speech signal parameters occurs. In accordance to the received characteristics and comparison them with the standards one from possible structures of the coding device will be chosen to display a speech signal on the frame of the analysis more adequately. The following step of optimization of coding device structure will be a quantitative and qualitative choice of variable number of parameters necessary for coding a speech signal with the given quality.

Thus by developing of the coder / decoder of speech signals two adaptive contours of management are necessary: the first makes optimization of the coding device structure, and the second optimization of transmitted parameters for the chosen system of speech signals compression.

Literature

1. John Makhoul, Salim Roucos, Herbert Gish Vector Quantization in Speech Coding: IEEE v.73, №11. 1985.
2. Bernard Gold Digital Speech Networks: IEEE v.65, №12. 1977, pp.11-18.
3. Korotaev G.A. Analysis and synthesis of a speech signal by a method of a linear prediction. Foreign radioelectronics №3, 1990, pp. 31-50.
4. Korotaev G.A. Some aspects of a linear prediction at the analysis and synthesis of a speech signal: Foreign radioelectronics №7, 1991, pp. 3-24.
5. O.I.Sheluchin, N.F.Lukjiancev Digital processing and transfer of speech, Moscow: Radio and Communication, 2000r, pp.66-91, 102-166.