

# СИСТЕМА АДАПТИВНОГО КОДИРОВАНИЯ РЕЧЕВЫХ СИГНАЛОВ

Афанасьев А.А., Пищальников А.В.

Россия, Орел, Академия ФАПСИ

E-mail: [aaa@academ.fagci.ru](mailto:aaa@academ.fagci.ru), [andreycka@mail.ru](mailto:andreycka@mail.ru)

В работе рассматривается задача синтеза кодеков речевых сигналов на основе системы с переменной структурой, которая, как правило, формулируется в двух вариантах.

Первый вариант задачи кодирования речи может заключаться в построении системы, минимизирующей скорость передачи ( $V$ ) при сохранении заданного качества речи ( $A_0$ ). Во втором случае, ставится задача построения системы, обеспечивающей максимально высокое качество передачи речи при заданной скорости передачи и других ограничениях.

В обоих случаях, как правило, вводятся дополнительные ограничения на стоимость системы ( $S$ ), массо-габаритные показатели, устойчивость к сбоям в канале связи  $A_n$ , величину задержки сигнала в системе  $\tau$  и т.д.

Задачи синтеза кодеков речи могут быть формализованы математически в следующем виде:

$$\begin{array}{ll} A \rightarrow \max; & \\ V_m \rightarrow \min; & A \geq A_0; \\ A \leq A_0; & V_m = \text{const}; \\ \tau \leq \tau_0; & \tau \leq \tau_0; \\ A_n \geq A_{0n}(P_c); & A_n \geq A_{0n}(P_c); \\ 1. S \leq S_0; & 2. S \leq S_0; \end{array}$$

Основное отличие рассматриваемого решения задачи синтеза кодеков речи заключается в том, что предлагается создать систему обработки речевых сигналов, у которой параметры и структура изменяются адаптивно, то есть в зависимости от свойств обрабатываемого фрагмента сигнала. Основными элементами такого адаптивного кодека являются: устройство идентификации структуры (классификатор), устройство идентификации параметров, наборы кодирующих устройств.

Для кодирования кадров речи предполагается использовать наборы линейных предсказателей и векторных квантователей.

Рассмотрим предложенную в соответствии с замыслом работы систему компрессии.

Входной речевой сигнал, оцифрованный с частотой дискретизации 8 кГц и 256 уровнями квантования делится на субкадры по 20 мс. Далее кадр центрируется посредством вычитания из всех отсчетов речевого сигнала математического ожидания при условии равновероятности всех отсчетов в кадре. Затем сигнал подается на многоуровневый классификатор структуры кодека. На первом этапе классификации сигнал делится на активные участки и паузы, критерием принятия решения в данном случае служат мощностные характеристики сигнала (1).

$$P = \frac{1}{N} \sum_{i=1}^N s_i^2; \quad P \leq P_0 \begin{cases} \text{да, } S - \text{пауза}; \\ \text{нет, } S - \text{речь}; \end{cases} \quad (1)$$

Для точного кодирования активных кадров речи происходит деление на 4 типа активных кадров, описанное ниже.

Наиболее явным признаком классификации анализируемых кадров на сегментах речевой активности представляется вокализованность и оценка частоты основного тона. Полезность данного признака подтверждается значительной долей информационных ресурсов ( $\approx 25\%$ ), выделяемых для кодирования результата данных оценок. На втором этапе классификации вычисляется оценка тон/шум и выделяется частота основного тона (ОТ). Данные вычисления производятся совместно на основе анализа автокорреляционной функции (АКФ) сигнала и метода Итакуры-Саито, описанных в [2]. Эти два метода в совокупности уменьшают вероятность ошибки в оценке данных параметров.

Целесообразность выбора данных методов для выделения сигнала тон-шум (СТШ) и частоты ОТ обосновывается уменьшением общей вычислительной сложности алгоритма кодирования за счет использования результатов, полученных в ходе вычисления данных параметров, на последующих этапах обработки.

Процесс нахождения СТШ, реализуется путем анализа АКФ данного кадра, при этом вводится следующее правило принятия решения (2), если

$$B(\tau_x) \geq kB(0) \begin{cases} \text{да, } S - \text{тон} \\ \text{нет, } S - \text{шум} \end{cases}, \text{ где } x = 20..160; \quad (2)$$

В данном случае  $k \approx 0,5$ , при увеличении  $k$  уменьшается ошибка 2-го рода, но соответственно увеличивается ошибка 1-го рода, при уменьшении  $k$  наоборот. Для оценки СТШ и частоты ОТ методом Итакуры-Саито вычисляются коэффициенты предсказания с помощью рекурсии Левинсона-Дарбина [4,5]. Далее реализуется процедура линейного предсказания, полученный остаток фильтруется рекурсивным фильтром ФНЧ 6-го порядка (прототип Баттерворта с частотой среза 600 Гц). Данный фильтр выбран из соображений линейности АЧХ и ФЧХ в полосе пропускания, а также приемлемой крутизны АЧХ в полосе расфилтровки. Фильтрация необходима для уменьшения влияния второй форманты на результаты вычислений. После фильтрации остатка предсказания вычисляется его АКФ. Дальнейшие вычисления параметра СТШ и частоты ОТ аналогичны методу на основе анализа АКФ.

Вводится граничное варьируемое значение частоты ОТ в зависимости, от которого осуществляется выбор "мужских" и "женских" кодовых книг для кодирования шумовых и переходных кадров.

К переходным кадрам 1 типа отнесем кадры, в которых относительно ОТ:

1. АКФ - false;

2. Итакуры саито - true;

К переходным кадрам 2 типа отнесем кадры, в которых относительно ОТ:

1. АКФ - true;

2. Итакуры-Саито - false;

В зависимости от выбранного варианта кодека имеется возможность подвергать переходные кадры, как векторному квантованию, так и линейному предсказанию.

При моделировании работы линейного предсказателя было отмечено, что шумовые кадры речи восстанавливаются гораздо хуже, нежели тоновые. Поэтому в качестве базового метода кодирования шумовых кадров используется векторное квантование. Однако следует отметить тот факт, что в речи нет чисто тоновых и чисто шумовых кадров, ибо любой шумовой кадр может нести в себе информацию об основном тоне говорящего. Учитывая данную особенность, производится деление "шумовых" и "переходных" кадров на «мужские» и «женские», посредством введения граничного варьируемого значения частоты основного тона, причем предварительно происходит третий этап классификация "шумовых" кадров речи по коэффициенту огибающей (3) и мощностной характеристике сигнала (1).

$$\eta = \frac{1}{N-1} \sum_{i=1}^{N-1} |S(n+1) - S(n)|; \quad (3)$$

Кодовые книги шумовых кадров речи формируются на основе алгоритма К-средних [1]. Обучающие выборки формируются с учетом частоты основного тона невокализованных кадров речи и статистических характеристик анализируемого шумового кадра.

В соответствии с полученными оценками коэффициента огибающей, мощностной характеристикой и частоты основного тона выбирается кодовая книга, обучение которой производилось на сигналах соответствующего класса. Исходный кадр речи (20 мс.) делится на субкадры по 2,5 мс., которые в свою очередь кодируются независимо друг от друга. На стыках субкадров в декодере используется интерполяция отсчетов.

Для кодирования тоновых кадров используется линейное предсказание с возбуждением от остатка линейного предсказания [3]. Для хранения допустимых векторов возбуждения создается кодовая книга, формируемая на основе анализа остатка линейного предсказания.

Используются две различные книги векторов возбуждения (для мужских и женских голосов). Критерием для выбора номера книги векторов возбуждения служит граничное варьируемое значение частоты ОТ.

Однако в ходе исследований был выявлен недостаток данного варианта алгоритма кодирования, который заключался в том, что при кодировании тоновых кадров фильтр-синтезатор находился в нулевых начальных условиях (ННУ), что в свою очередь приводило к вибрации речи в начале кадра и потери разборчивости. Если же фильтр синтеза не находился к моменту синтеза в ННУ, то это приводило к искажениям в синтезированной речи и потери узнаваемости за счет ошибочного начального заполнения фильтра-синтезатора на основе информации от предшествующих кадров ОТ, а не от предыдущего кадра.

Было предложено следующее решение данной проблемы. При кодировании шумовых и переходных кадров с помощью векторного квантования на приемном конце вычисляются коэффициенты линейного предсказания. Состояние регистров фильтра синтеза постоянно обновляется таким образом, чтобы к моменту кодирования тонового кадра начальные условия фильтра соответствовали информации о предыдущем кадре, который подвергался линейному предсказанию RELP, с той лишь разницей, что эти вычисления производятся на приемном конце.

Разработанный алгоритм функционирования аналого-цифрового и цифро-аналогового преобразования речевых сигналов был промоделирован на ПЭВМ, в результате которого показана его работоспособность. Качество передачи речи в кодеке по сравнению с неадаптивной структурой повысилась, что подтверждается как субъективными оценками, так и расчетом среднеквадратической ошибки.

### Литература

1. Vector Quantization in Speech Coding. John Makhoul, Salim Roucos, Herbert Gish / IEEE v.73, №11. 1985.
2. Маркел Дж. Д., А.Х. Грэй Линейное предсказание речи – М.: Связь, 1980. – 308с.
3. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981.
4. Коротаев Г. А. Некоторые аспекты линейного предсказания при анализе и синтезе речевого сигнала. Зарубежная радиоэлектроника №7.1991
5. О.И.Шелухин, Н.Ф.Лукиянцев Цифровая обработка и передача речи, М., Радио и связь, 2000г.



### SYSTEM OF ADAPTIVE SPEECH SIGNALS CODING

Afanasjev A., Pischalnikov A.

Russia, Oryol, Academy of FAGCI  
E-mail: [aaa@academ.fagci.ru](mailto:aaa@academ.fagci.ru), [andreycka@mail.ru](mailto:andreycka@mail.ru)

Given article has been devoted to one of the up-to-date approaches of processing and compression of speech signals - coding of speech based on the systems with variable structure. Thus, it is necessary to change not only parameters, but also structure of the coding device. The basic trends of this approach has been presented in this article. Taking into account the statistical and parametrical characteristics of speech the adaptive algorithms has been designed. The main ways of reducing of speech coding rate and increasing of quality of perception of coded speech has been described.

The basic difference of the considered decision is in offering of creating system of speech signals processing , where the parameters and structure are changing adaptively, depending on properties of a processable signal frame. Main elements of the same adaptive compression system are the following: the device of identification of structure (qualifier), device of identification of parameters, systems of settings of coding devices.

For coding the speech frame, using of systems of settings of linear predictors and vector quantizes have been offered.

The designed operating algorithm of analog-to-digital and digital-to-analog transformation of speech signals has been tested on computer, the results of its functionality has been shown. Quality of transmitting of speech coding in compare with non-adaptive structure has been improved, it's been proved transmitting as for personality and as for measurements meansquare of errors.

### References

1. Vector Quantization in Speech Coding. John Makhoul, Salim Roucos, Herbert Gish / IEEE v.73, №11. 1985.
2. Markel J.D., Gray A.H.Jr. Linear prediction of speech – М.: Communication, 1980. – 308p.
3. Rabiner L.R., Schafer R.W. Digital processing of speech signals. – М.: Radio and Communication, 1981.
4. Korotaev G. A. Some aspects of linear prediction for analysis and synthesis of speech signals. Foreign radioelectronics №7,1991.
5. O.I.Sheluchin, N.F.Lukjancev Digital processing and speech transmittion, М., Radio and communication, 2000.