Моделирование, оптимизация и информационные технологии.
Научный журнал №3(14) http://moit.vivt.ru/ **2016**

УДК 004.934.2

К.П. Плетнев, Д.Е. Прозоров

МАРКОВСКИЕ МОДЕЛИ КВАНТОВАННЫХ РЕЧЕВЫХ СИГНАЛОВ

Вятский государственный университет

В статье рассмотрены марковские модели квантованных речевых сигналов (КРС). Показано, что корреляционные свойства битовых последовательностей, получаемых квантованием речевых сигналов, позволяют применить математический аппарат простых или сложных (многосвязных) конечных цепей Маркова для статистического описания КРС. Метод параметризации, основанный на марковских моделях КРС, снижает требования к вычислительным ресурсам систем автоматического распознавания речи.

Ключевые слова: модели речевых сигналов, цепи Маркова, параметризация.

Введение

Интерес к интеллектуальным человеко-машинным интерфейсам на основе естественных языков обусловил стремительный рост количества исследований в области цифровой обработки речевых сигналов. Основополагающим в данной области является спектр направлений, связанных с проблемой автоматического распознавания речи (АРР).

Одним из первых этапов распознавания речи, оказывающий существенное влияние на эффективность работы АРР является этап параметризации речевых сигналов. Рассматриваемая задача давно известна и имеет ряд решений, классические подходы к параметризации речевого сигнала изложены в работах А.В. Оппенгейма [1], Л.Р. Рабинера и Б. Гоулда [2], С. Саито и Ф. Итакуры [3], Б.С. Атала и М.Р. Шредера [4], Дж.Д. Маркела [5], П.Е. Папамихалис [6] и др., среди которых наиболее распространены методы, использующие кепстральные преобразования в частотной или временной области.

Современная тенденция повсеместного внедрения встраиваемых вычислительных систем, в том числе систем АРР диктует ограничения на вычислительную сложность методов и алгоритмов, применяемых на этом этапе. В связи с указанными ограничениями возникает необходимость анализа известных и разработки новых методов параметризации речевых сигналов, позволяющих снизить потребление вычислительных ресурсов систем АРР при сохранении приемлемого уровня распознавания речевых сигналов. В данной работе показано, что корреляционные свойства квантованных речевых сигналов (КРС) позволяют использовать модели простых или сложных цепей Маркова для разработки таких методов параметризации.

Квантование речевых сигналов

Наиболее общей формой цифрового представления речевого сигнала можно считать двоичную N – разрядную последовательность, полученную в результате дискретизации с частотой f_∂ и равномерного квантования исходного сигнала:

$$\mathbf{s} = \left\{ s^{(1)} s^{(2)} \ldots s^{(L)} \right\}, \tag{1}$$

где $s^{(k)}$ – k-я двоичная выборка вида

$$s^{(k)} = \sum_{b=0}^{N-1} \left({}^{b}s^{(k)} \cdot 2^{N-b-1} \right),\; k = \overline{1, L},\; {}^{b}s^{(k)} = \overline{0,1}, \tag{2}$$

L – длина фрагмента.

Фрагмент (1) можно рассматривать как суперпозицию битовых последовательностей ${}^{b}\mathbf{s}$ длины L:

$$\mathbf{s} = \sum_{b=0}^{N-1} \left({}^{b}\mathbf{s} \cdot 2^{N-b-1} \right), \tag{3}$$

где ${}^{b}\mathbf{s}$ – бинарная последовательность, образованная b-м битом квантованного фрагмента речевого сигнала.

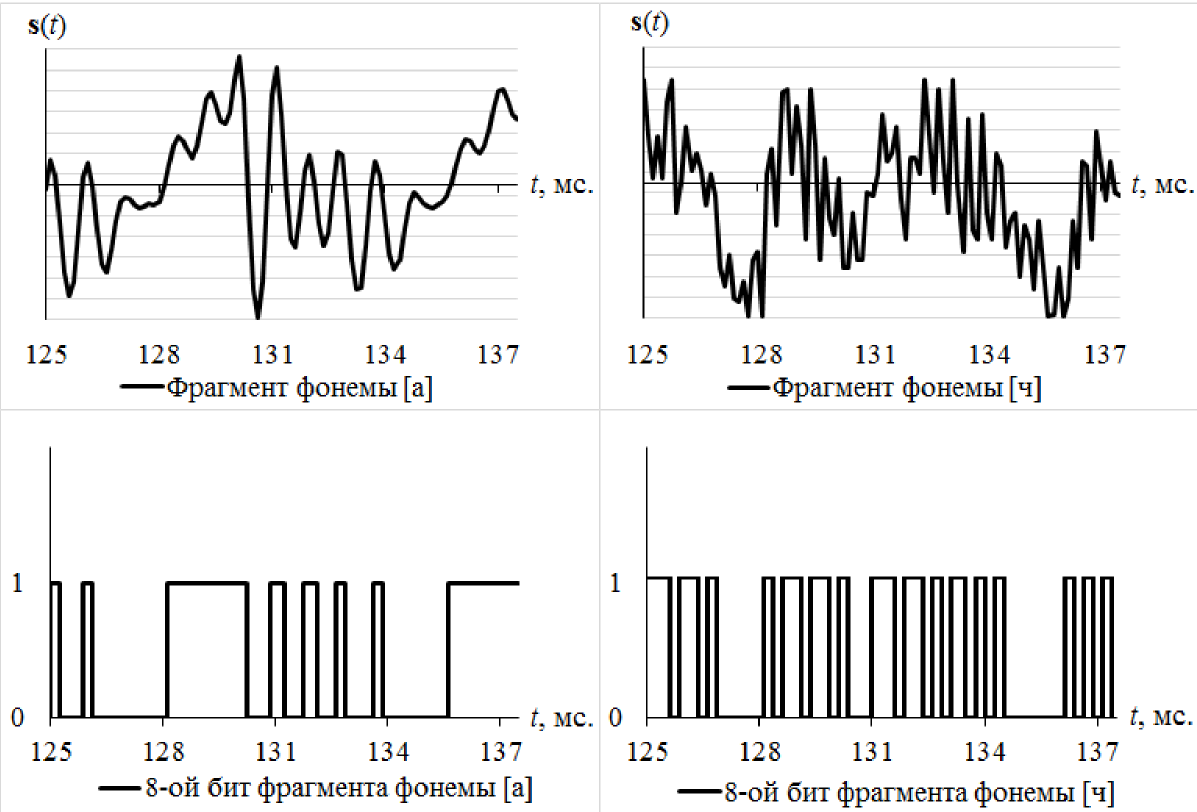

Рисунок 1. Фрагменты выборочных реализаций фонем [а] и [ч]

Наибольший интерес для анализа корреляционных свойств представляют звуки, реализация которых связана с различными механизмами произношения. Примерами таких звуков являются: гласная [а] получаемая при нижнем положении языка и среднем рядом относительно границ ротовой полости, [ч] – аффрикативная согласная, получаемая сочетанием смычного согласного [ть] с фрикативным [щ]. Эпюры речевых сигналов, являющихся реализациями фонем [а] и [ч] и битовых последовательностей ${}^{b}\mathbf{s}$, полученных квантованием этих фрагментов представлены на Рисунке 1.

На Рисунке 2 представлены графики автокорреляционных функций (АКФ) речевых сигналов $\mathbf{s}$ и АКФ битовых последовательностей ${}^{b}\mathbf{s}$, полученных квантованием этих речевых сигналов.

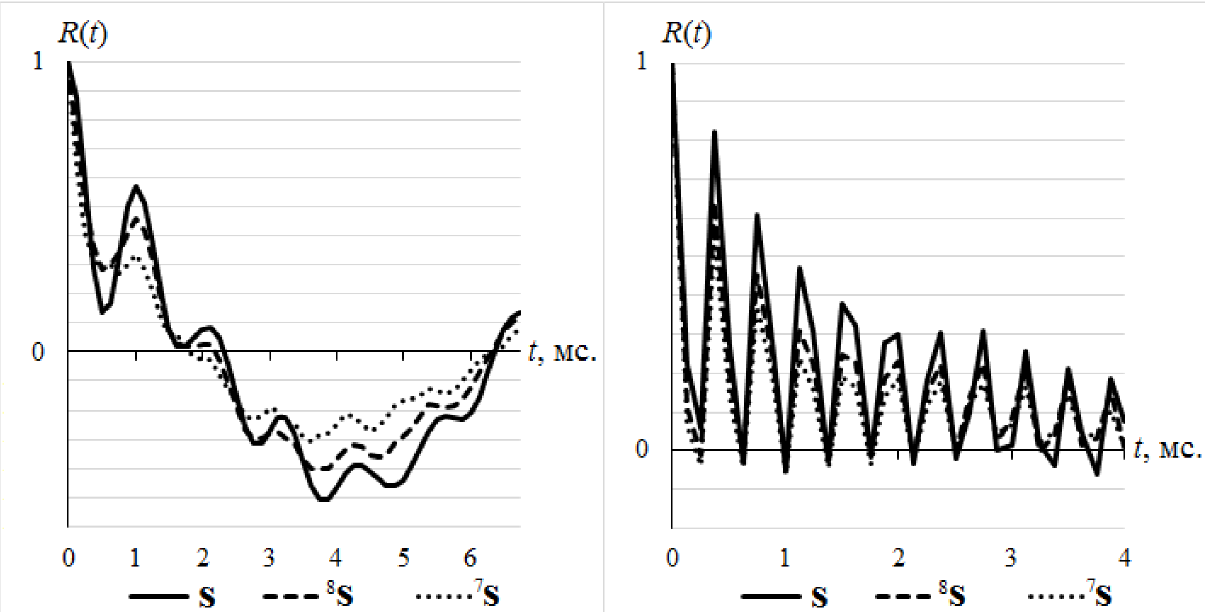

Рисунок 2. Автокорреляционные функции: а) выборочной реализации фонемы [а]; б) выборочной реализации фонемы [ч] (частота дискретизации f_s = 8 кГц)

По графикам Рисунка 2 видно, что АКФ выборочной реализации согласной [ч] имеет более изрезанный вид по сравнению с АКФ гласной [а]. Причем, на коротких интервалах времени (для фонемы [а] – [0;0,5) мс., а для фонемы [ч] – [0;0,25) мс) представленные на рис.2 функции демонстрируют практически монотонный характер убывания.

Таким образом, квантованный речевой сигнал на ограниченном интервале времени обладает свойством монотонного ослабления корреляционных связей. Данный факт позволяет использовать для описания и анализа коротких фрагментов КРС математический аппарат простых или сложных цепей Маркова [8-10].

Моделирование речевого сигнала цепью Маркова

Модель конечной однородной цепи Маркова *n*-го порядка в общем случае задается вектором начальных вероятностей [10]

$$\left[p(s_1),\ldots,p(s_K)\right]^{\mathrm{T}} \tag{4}$$

и матрицей вероятностей переходов (МВП)

$$\Pi^* = \begin{bmatrix} \pi_{1\ldots1,1} & \ldots & \pi_{1\ldots1,K} \\ \pi_{1\ldots2,1} & \ldots & \pi_{1\ldots2,K} \\ \ldots & \ldots & \ldots \\ \pi_{2\ldots1,1} & \ldots & \pi_{2\ldots1,K} \\ \ldots & \ldots & \ldots \\ \pi_{K\ldots K,1} & \ldots & \pi_{K\ldots K,K} \end{bmatrix}, \tag{5}$$

где $\pi_{i\ldots r,l} \equiv p\left(s_l^{(k)} \middle| s_i^{(k-1)} s_j^{(k-2)} \ldots s_l^{(k-n)}\right)$, $i,j,\ldots,l=\overline{1,K}$, K – количество состояний модели.

В работе [7] представлен метод преобразования таких матриц к «квадратному» виду.

Сформируем векторы

$$\vec{s}^{\,(k-1)} = \left(s_l^{(k-1)} s_j^{(k-2)} \ldots s_l^{(k-n)}\right) \tag{6}$$

длины n. Количество различающихся векторов $\vec{s}^{\,(k)}$ равно K^n. Преобразование (6) порождает простую цепь Маркова – векторную последовательность $\left\{\vec{s}^{\,(1)},\vec{s}^{\,(2)},\ldots\right\}$ с матрицей вероятностей переходов размера $K^n \times K^n$ [26].

При известных условных вероятностях (5) можно определить вероятности перехода полученной простой цепи Маркова:

$$p\left(\vec{s}^{\,(k)} \middle| \vec{s}^{\,(k-1)}\right) = \frac{p\left(\left\{s_l^{(k)} \ldots s_v^{(k-n)}\right\}, \left\{s_i^{(k-1)} \ldots s_r^{(k-n)}\right\}\right)}{p\left(s_i^{(k-1)} \ldots s_r^{(k-n)}\right)} =$$

$$= \frac{p\left(s_l^{(k)} s_i^{(k-1)} \ldots s_r^{(k-n)}\right)}{p\left(s_i^{(k-1)} \ldots s_r^{(k-n)}\right)} = p\left(s_l^{(k)} \middle| s_i^{(k-1)} \ldots s_r^{(k-n)}\right), \tag{7}$$

где $l,i,j,\ldots,r,v = \overline{1,K}$.

В результате преобразования (7) формируется простая цепь Маркова с количеством состояний $m = K^n$ и МВП вида

$$\Pi' = \begin{bmatrix} \pi_{1,1} & \pi_{1,2} & \ldots & \pi_{1,m} \\ \pi_{2,1} & \pi_{2,2} & \ldots & \pi_{2,m} \\ \ldots & \ldots & \ldots & \ldots \\ \pi_{m,1} & \pi_{m,2} & \ldots & \pi_{m,m} \end{bmatrix}. \tag{8}$$

МВП (8) удовлетворяют условиям нормировки

$$\sum_{j=1}^{K^n} \pi_{ij} = 1, \quad i, j = \overline{1, K^n}.$$

Рассмотрим примеры для моделей первого (простая цепь) и второго (сложная цепь) порядка, представляющие битовые последовательности ${}^b\mathbf{s}$ (3).

Пример 1 (n = 1, K = 2).

Для модели первого порядка вектор начальных вероятностей (4) и МВП (8) имеют вид

$$ {}^b\mathbf{P} = \begin{bmatrix} {}^b p_1 \\ {}^b p_2 \end{bmatrix}, \tag{9}$$

$$ p\left({}^b s_j^{(k+1)} \middle| {}^b s_i^{(k)}\right) = {}^b\prod = \begin{bmatrix} {}^b\pi_{11} & {}^b\pi_{12} \\ {}^b\pi_{21} & {}^b\pi_{22} \end{bmatrix}, \ i, j = \overline{1,2}. \tag{10}$$

Элементы ${}^b\pi_{ij}$ матрицы вероятностей перехода (10) положительны и удовлетворяют условиям нормировки

$$\sum_{j=1}^{2} {}^b\pi_{ij} = 1, \ i = \overline{1,2}.$$

Пример 2 (n = 2, K = 2).

Вектор начальных вероятностей (4) и МВП (5) для модели второго порядка имеют вид

$$ {}^b\mathbf{P}^* = \begin{bmatrix} {}^b p_1 \\ {}^b p_2 \end{bmatrix},$$

$$ {}^b\prod{}^* = \begin{bmatrix} {}^b\pi_{11,1} & {}^b\pi_{11,2} \\ {}^b\pi_{12,1} & {}^b\pi_{12,2} \\ {}^b\pi_{21,1} & {}^b\pi_{21,2} \\ {}^b\pi_{22,1} & {}^b\pi_{22,2} \end{bmatrix}.$$

Учитывая, что $m = K^n = 4$, после перехода к «квадратной» матрице вида (8) получаем МВП

$$ {}^b\prod = \begin{bmatrix} {}^b\pi_{111} & {}^b\pi_{112} & 0 & 0 \\ 0 & 0 & {}^b\pi_{121} & {}^b\pi_{122} \\ {}^b\pi_{211} & {}^b\pi_{212} & 0 & 0 \\ 0 & 0 & {}^b\pi_{221} & {}^b\pi_{222} \end{bmatrix} \tag{11}$$

и вектор начальных вероятностей

$$ {}^b\mathbf{P} = \left[{}^b p_1 \quad {}^b p_2 \quad {}^b p_3 \quad {}^b p_4\right]^{\mathrm{T}}. \tag{12}$$

Элементы ${}^b\pi_{ij}$ матрицы вероятностей перехода (11) положительны и удовлетворяют условиям нормировки

$$\sum_{j=1}^{4} {}^{b}\pi_{ij} = 1,\ i = \overline{1,4}.$$

Эксперимент по распознаванию речевых сигналов проведенный в работе [11] показал, что представленные марковские модели КРС позволяют снизить время параметризации речевых сигналов в 16-19 раз при сопутствующем снижении вероятности распознавания изолированных слов примерно на 10% при использовании модели простой цепи Маркова и на ~5% при использовании модели сложной (многосвязной) цепи Маркова.

Полученные результаты свидетельствуют о возможности применения марковских моделей для решения задачи параметризации квантованных речевых сигналов. Оценки скорости обработки речевых сигналов и эффективности алгоритмов марковской параметризации позволяют рекомендовать применение подобных моделей и методов марковской параметризации для встраиваемых аппаратно-программных систем автоматического распознавания речи с ограниченными вычислительными ресурсами.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-07-00342-а.

ЛИТЕРАТУРА

1. Оппенгейм, А.В. Цифровая обработка сигналов / пер. с англ. под ред. С.Я. Шаца / А.В. Оппенгейм, Р.В. Шафер // – М.: Связь, 1979. – 416 с.
2. Рабинер, Л. Теория и применение цифровой обработки сигналов / Пер. с англ. под ред. Ю. И. Александрова / Л. Рабинер, Б. Гоулд // – М.: Мир. 1978. 835 с.
3. Saito, S. The Theoretical Consideration of Statistically Optimum Methods for Speech Spectral Density / S. Saito, F. Itakura // Report No. 3107, Electrical Communication Laboratory, N. T. T., Tokyo. 1966.
4. Atal, B.S. Predictive Coding of Speech Signals / B.S. Atal, M.R. Schroeder // Proc. 1967 Conf. Commun. and Proccess. P. 360-361.
5. Маркел, Дж. Линейное предсказание речи / пер. с англ. Ю.Н. Прохорова, В.С. Звездина / Дж. Маркел, А.Х. Грей. – М. Связь. 1980. 308 с.
6. Papamichalis, P.E. Practical Approaches to Speech Coding / Papamichalis P.E. // Prentice Hall, Englewood Cliffs. – NJ. 1987.
7. Яншин В.В. Многосвязные цепи Маркова и их свойства // Радиотехника и электроника. 1993. Том 38. № 6. С. 1081-1091.
8. Прозоров Д.Е., Плетнев К.В. Численное моделирование многосвязной бинарной цепи Маркова // Инфокоммуникационные технологии. 2014. №2. С. 8-11.

9. Венедиктов, М.Д. Дельта-модуляция. Теория и применение / М.Д. Венедиктов, Ю.П. Женевский, В.В. Марков // М.: Связь. 1976. С. 104-114.
10. Кемени, Дж. Конечные цепи Маркова. / Дж. Кемени, Дж. Снелл // Перевод с англ. Под ред. А. А. Юшкевича. – М., «Наука». 1970. 272 с.
11. Прозоров Д.Е. Марковские модели в задачах параметризации речевых сигналов / Д.Е. Прозоров, К.В. Плетнев // Журнал радиоэлектроники: электронный журнал. 2014. №6. URL: http://jre.cplire.ru/jre/jun14/3/text.pdf (дата обращения: 01.12.2015).

K.P. Pletnyov, D.E. Prozorov

MARKOV MODELS OF THE QUANTIZED SPEECH SIGNALS

Vyatka State University

The article considers Markov models of the quantized speech signals (QSS). It is shown that correlation properties of the bit sequences obtained by the quantization of speech signals, allow it to apply the mathematical apparatus of simple or complex (multiply connected) finite Markov chains for the statistical description of the QSS. The method of parameterization based on Markov models of QSS reduces the performance requirements of systems of automatic speech recognition.

Keywords: Markov models of speech signals, Markov chains, parametrization.

REFERENCES

1. Oppengeym, A.V. Tsifrovaya obrabotka signalov / per. s angl. pod red. S.Ya. Shatsa / A.V. Oppengeym, R.V. Shafer // – M.: Svyaz', 1979. – 416 p.
2. Rabiner, L. Teoriya i primenenie tsifrovoy obrabotki signalov / Per. s angl. pod red. Yu. I. Aleksandrova / L. Rabiner, B. Gould // – M.: Mir. 1978. 835 p.
3. Saito, S. The Theoretical Consideration of Statistically Optimum Methods for Speech Spectral Density / S. Saito, F. Itakura // Report No. 3107, Electrical Communication Laboratory, N. T. T., Tokyo. 1966.
4. Atal, B.S. Predictive Coding of Speech Signals / B.S. Atal, M.R. Schroeder // Proc. 1967 Conf. Commun. and Proccess. pp. 360-361.
5. Markel, Dzh. Lineynoe predskazanie rechi / per. s angl. Yu.N. Prokhorova, V.S. Zvezdina / Dzh. Markel, A.Kh. Grey. – M. Svyaz'. 1980. 308 p.
6. Papamichalis, P.E. Practical Approaches to Speech Coding / Papamichalis P.E. // Prentice Hall, Englewood Cliffs. – NJ. 1987.
7. Yanshin V.V. Mnogosvyaznye tsepi Markova i ikh svoystva // Radiotekhnika i elektronika. 1993. Vol. 38. No. 6. pp. 1081-1091.
8. Prozorov D.E., Pletnev K.V. Chislennoe modelirovanie mnogosvyaznoy binarnoy tsepi Markova // Infokommunikatsionnye tekhnologii. 2014. No.2. pp. 8-11.

9. Venediktov, M.D. Del'ta-modulyatsiya. Teoriya i primenenie / M.D. Venediktov, Yu.P. Zhenevskiy, V.V. Markov // M.: Svyaz'. 1976. pp. 104-114.
10. Kemeni, Dzh. Konechnye tsepi Markova. / Dzh. Kemeni, Dzh. Snell // Perevod s angl. Pod red. A. A. Yushkevicha. – M., «Nauka». 1970. 272 p.
11. Prozorov D.E. Markovskie modeli v zadachakh parametrizatsii rechevykh signalov / D.E. Prozorov, K.V. Pletnev // Zhurnal radioelektroniki: elektronnyy zhurnal. 2014. No.6.
URL: http://jre.cplire.ru/jre/jun14/3/text.pdf (data obrashcheniya: 01.12.2015).