УДК 621.371.39

КОМПРЕССИЯ РЕЧИ НА ОСНОВЕ ДОМИНИРУЮЩИХ СПЕКТРАЛЬНЫХ КОМПОНЕНТОВ

канд. техн. наук, доц. С.В. МАЛЬЦЕВ, В.М. ЧЕРТКОВ
(Полоцкий государственный университет)

Рассмотрены особенности сжатия речевого потока без ухудшения качества восстановленной речи на основе отбора доминирующих спектральных компонентов. Отбор доминирующих спектральных компонентов производится на основе критериев психоакустических принципов. Представлена схема компрессии речи и ее математическая модель на основе спектрального сжатия. Для объективной оценки качества восстановленной речи рассчитано математическое ожидание, дисперсия и среднеквадратическое отклонение уровня отличия восстановленного сигнала от исходного. Рассмотрена схема синтеза речевого сигнала в реальном времени при использовании метода наложения со сложением, реализованная на основе набора аппаратных и программных средств DSP на базе микропроцессорного модуля TMS320VC5510.

Введение. Интенсивное развитие информационных технологий в современном мире обостряет проблему быстрой и качественной передачи различного рода информации по цифровым линиям связи. Несмотря на разнообразие применяемых средств связи, основным видом коммуникации между людьми остается обмен информацией посредством речи. В связи с этим возникает необходимость в развитии и совершенствовании методов цифровой обработки и передачи речевых сообщений.

Появление широкодоступных процессоров цифровой обработки сигналов и соответствующих инструментальных сред [1] создало возможности для развития современных систем передачи и обработки речевых сообщений. При этом большое значение приобретает решение проблемы минимизации числа бит, необходимых для передачи сигнала, т.е. повышения скорости передачи без ухудшения качества восстановленной речи.

Компрессии речи на основе отбора доминирующих спектральных компонентов

Существует множество различных подходов и принципов компрессии и кодирования речи. Одним из самых эффективных является подход на основе спектрального представления речи, применение которого совместно с принципами антропоморфической обработки сигнала позволит значительно уменьшить количество компонент, из которых синтезируется речь. Очевидно, что в системах низкоскоростной компрессии речевой информации на основе спектральной модели для уменьшения количества кодируемой информации без ухудшения качества восстановленной речи необходимо использовать специальные критерии отбора наиболее важной для слуха человека информации [2], которые хорошо согласуются с особенностями восприятия [3]. Одним из вариантов компрессии речи на основе спектральной модели речевого сигнала совместно с критериями психоакустических принципов отбора доминирующих частотных пиков является схема, представленная на рисунке 1.

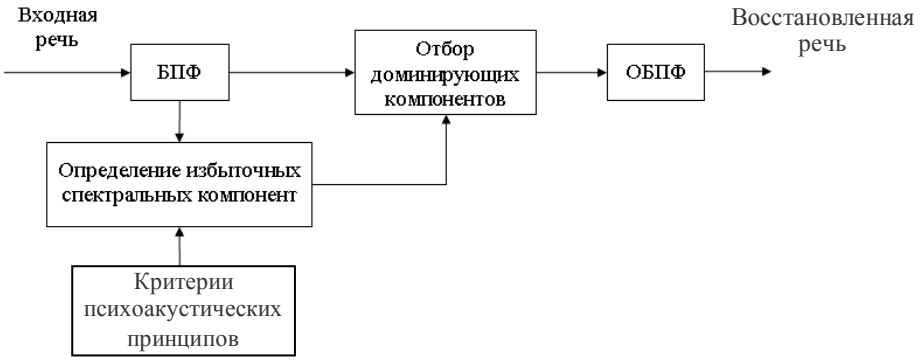


Рис. 1. Схема компрессии речи
на основе отбора доминирующих спектральных компонентов

Представленная схема осуществляет предварительную обработку звука и обеспечивает:

- разложение на спектральные составляющие (используется быстрое преобразование Фурье);
- определение избыточных спектральных составляющих с учетом принципов психоакустики (рассчитываются пороги маскирования и абсолютный порог слышимости для критических полос);

- отбор доминирующих спектральных компонент (на основе сравнения принадлежности к избыточным компонентам);
- синтез речи на основе отобранных спектральных составляющих (используется обратное быстрое преобразование Фурье).

Математическая модель спектрального сжатия речи

Алгоритм математической модели спектрального сжатия речи включает в себя:

Этап 1. *Установка начальных условий*

- Задаются частота дискретизации f_s = 8000 и длина окна N_f = 256 быстрого преобразования Фурье (БПФ).
- Разбиение на критические полосы.
- Определение центров критических полос.

Этап 2. *Выделение из исходной последовательности отсчетов речи очередного отрезка – фрейма.*

Этап 3. *Вычисление ДПФ для текущего фрейма.*

Этап 4. *Нахождение тональных маскеров*

Для каждого отсчета ДПФ из *r*-го фрейма оценивается его энергия в дБ и определяется его принадлежность к тональному маркеру исходя из выражения [3]:

$$\begin{aligned} &P_Y(i) > P_Y(i \pm 1), \\ &P_Y(i) > P_Y(i \pm \Delta_i) + 7 \text{ дБ}, \end{aligned} \tag{1}$$

где $\Delta_i = \begin{cases} 2 & \text{при } 0{,}17 \text{ кГц} < \omega_i < 5{,}5 \text{ кГц}; \\ [2;3] & \text{при } 5{,}5 \text{ кГц} \le \omega_i < 11 \text{ кГц}; \\ [2;6] & \text{при } 11 \text{ кГц} \le \omega_i < 20 \text{ кГц}. \end{cases}$

Здесь ω_i – частота, соответствующая *i*-му элементу частотной характеристики.

На рисунке 2 показан результат расчета энергии тонального маскирования для *r*-го фрейма (линия 1).

Этап 5. *Нахождение шумовых маскеров*

Рассчитывается энергия шума в пределах найденных речевых критических полос согласно выражению

$$P_{NM}(i) = \sum_{j=k_l}^{k_h} P_Y(j) \text{ (дБ)}, \tag{2}$$

где i – номер критической полосы; k_l и k_h – номера первого и последнего элемента *i*-той критической полосы.

Результат расчета энергии шумовых маскеров иллюстрирует линия 2 на рисунке 2.

Этап 6. *Прореживание маскеров*

Анализируются все найденные маскеры как тональные, так и шумовые. Для каждого из двух маскеров, отстоящих друг от друга на удалении не более 0,5 барка, оставляется только тот, который имеет большую энергию.

Этап 7. *Расчет абсолютного порога маскирования*

Для каждого оставшегося маскера рассчитывается абсолютный порог маскирования, исходя из его функции распространения во всем диапазоне частот БПФ (рис. 2). Функция распространения $SF(\Delta_Z)$, имеющая форму треугольника с вершиной, расположенной в центре тонального или шумового маскера, вычисляется по формуле [3]:

$$SF(i,j) = \begin{cases} 17\Delta_Z - 0{,}4P_{NM}(j) + 11 & \text{при } -3 \le \Delta_Z < -1; \\ (0{,}4P_{NM}(j) + 6) \cdot \Delta_Z & \text{при } -1 \le \Delta_Z < 0; \\ -17\Delta_Z & \text{при } 0 \le \Delta_Z < 1; \\ (0{,}15P_{NM}(j) - 17) \cdot \Delta_Z - 0{,}15P_{NM}(j) & \text{при } 1 \le \Delta_Z < 8, \end{cases} \text{ (дБ)} \tag{3}$$

где i – номер отсчета, для которого вычисляется индивидуальный порог маскирования; j – номер отсчета, соответствующий маскеру; $P_{NM,TM}$ – шумовой или тональный маскер; $\Delta_Z = z(i) - z(j)$ – разность частот маскирующего и маскируемого сигналов выраженных в барках.

На рисунке 2 также представлен набор из всех маскеров (тональных и шумовых), прошедших прореживание, для каждого маскера рассчитан порог маскирования в соответствии с выражением (3). Совокупность всех порогов маскирования составляет абсолютный порог маскирования, изображенный на рисунке линией 3.

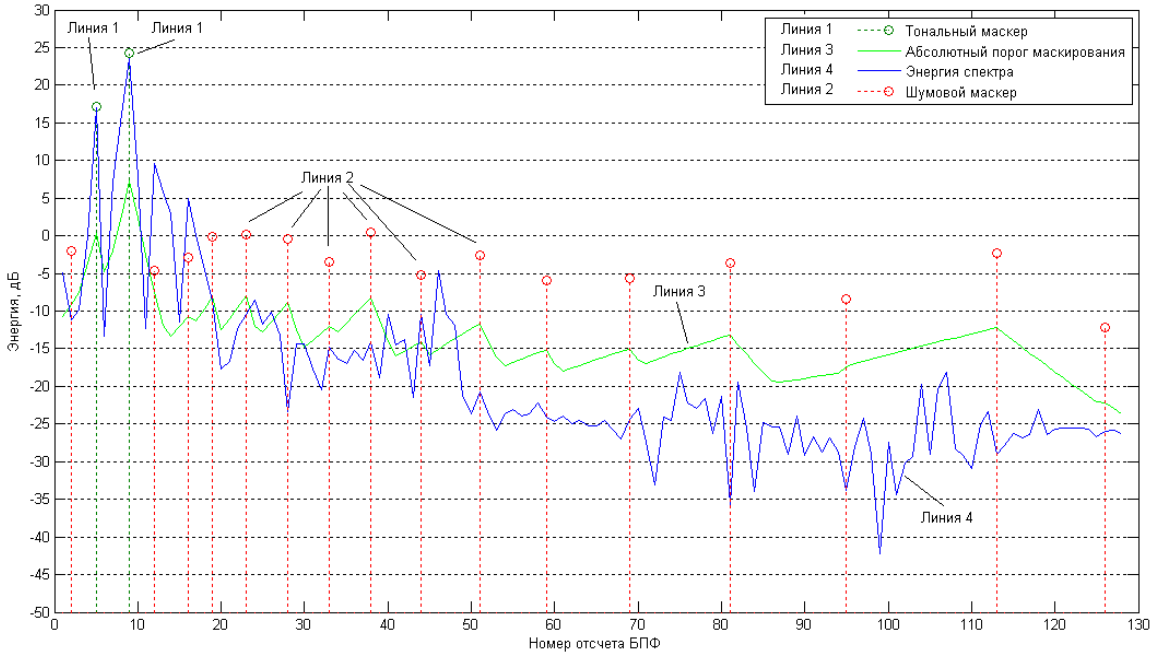


Рис. 2. Абсолютный порог маскирования для всех маскеров:
линия 3 – абсолютный порог маскирования; линия 1 – найденные тональные маскеры, которые соответствуют выражению (1); линия 2 – найденные шумовые маскеры, которые соответствуют выражению (2)

Этап 8. *Расчет абсолютного порога слышимости осуществляется исходя из выражения* [3]:

$$T_q(f) = 3{,}64 \cdot (f/1000)^{-0{,}8} - 6{,}5 \cdot e^{-0{,}6\cdot(f/1000-3{,}3)^2} + 10^{-3} \cdot (f/1000)^4 \quad \text{(дБ)},$$

где f – частота в Гц.

Этап 9. *Отбор доминирующих спектральных компонентов*

Происходит выбор максимального значения из рассчитанных порогов и энергии для каждого отсчета БПФ. Если энергия гармоники превышает абсолютный порог маскирования и абсолютный порог слышимости, гармоника проходит отбор. Из рисунка 3 видно, что будут отобраны всего 4 гармоники, которые расположены выше абсолютного порога слышимости (линия 5). В результате спектральные компоненты, оставшиеся после прореживания будут иметь вид, представленный на рисунке 4.

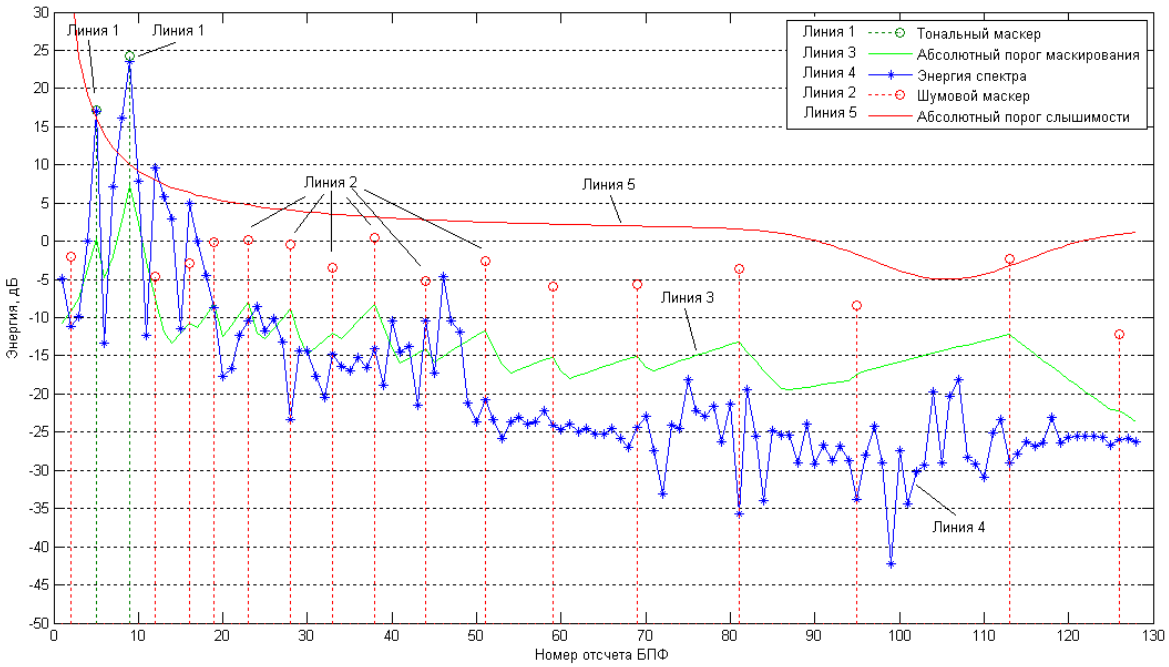
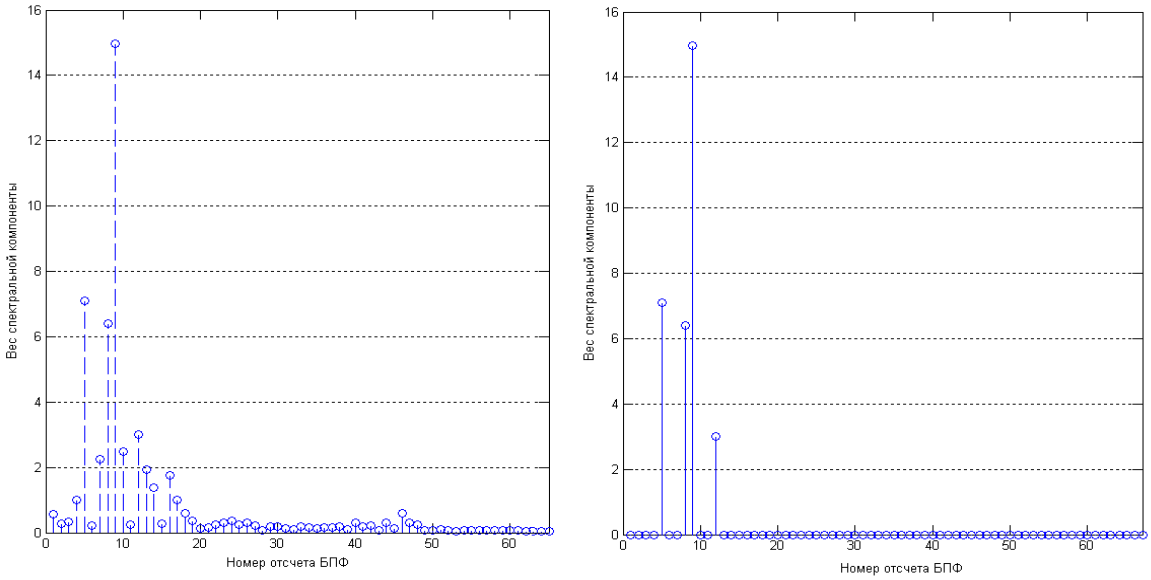


Рис. 3. Выбор спектральных составляющих методом сравнения



а) б)

Рис. 4. Спектральные компоненты, первоначальный амплитудный спектр *r*-го фрейма (а), отобранный амплитудный спектр *r*-го фрейма для восстановления (б)

Этап 10. Из отобранных отсчетов восстанавливается *r*-фрейм (рис. 5) на основе обратного быстрого преобразования Фурье.

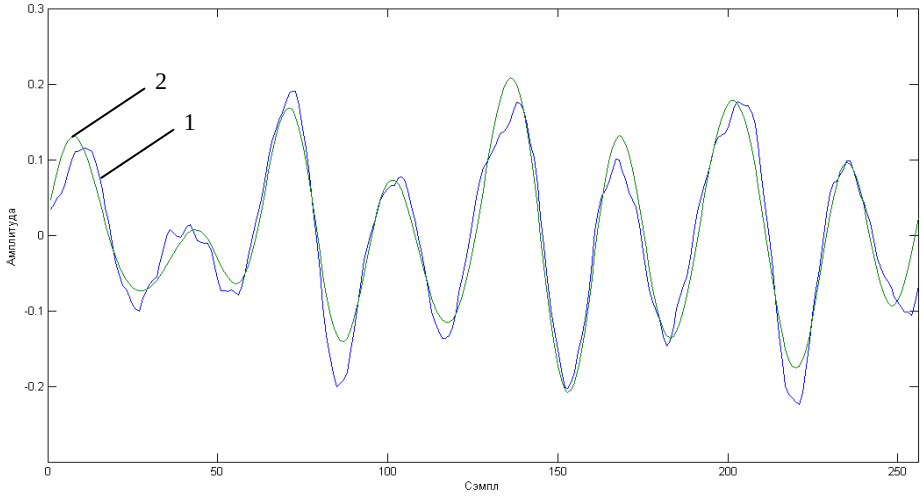


Рис. 5. Исходный и восстановленный сигналы:
линия 1 – первоначальный сигнал; линия 2 – восстановленный

Для объективной оценки качества восстановленной речи рассчитывается математическое ожидание, дисперсия и среднеквадратическое отклонение уровня отличия восстановленного сигнала от исходного.

- математическое ожидание $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i = 0{,}0223$;

- дисперсия $\sigma_x^2 = D(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^2 = 0{,}0002647$;

- среднеквадратическое отклонение $\sigma_X = \sqrt{D(X)} = 0{,}01627$.

Аппаратная реализация предварительной обработки звука на основе DSP TMS320VC5510 и результаты эксперимента

Для реализации предложенной схемы компрессии звукового сигнала использовался набор аппаратных и программных средств DSP на базе TMS320VC5510.

Характеристики средств DSP:

1) ЦПОС TMS320VC5510 с частотой 150 МГц;
2) 16 разрядный АЦП;
3) внешняя память 16MB SDRAM, 128KB Flash ROM;
4) наличие световых индикаторов;
5) связь с компьютером через USB порт.

Самый простой и надежный метод аппаратной реализации заключается в преобразовании входных отсчетов в частотной области на основе БПФ и реализация ОБПФ в качестве заключительной процедуры. Однако следует заметить, что БПФ рассчитывается только для отдельно взятых фреймов и отображает ограниченный набор синусоид с определенными частотами, амплитудами и фазами, что приводит к скачкообразному изменению синтезированного сигнала на границах синтезируемых фреймов. Это приводит к возникновению хриплости и появлению различных искажений в синтезируемой речи. Для «склеивания» соседних фреймов и устранения разрывов на их границах применяется метод наложения со сложением [4].

Процесс синтеза речевого сигнала при использовании метода наложения со сложением показан схематично на рисунке 6.

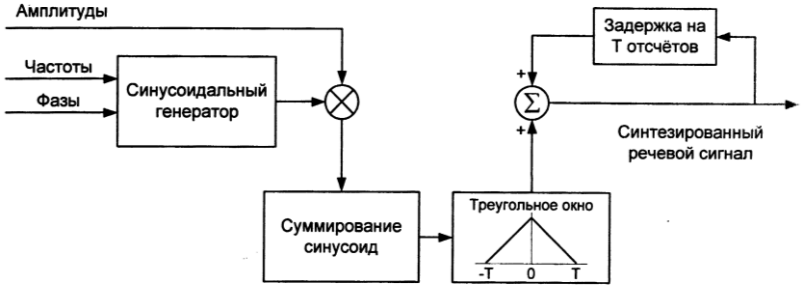
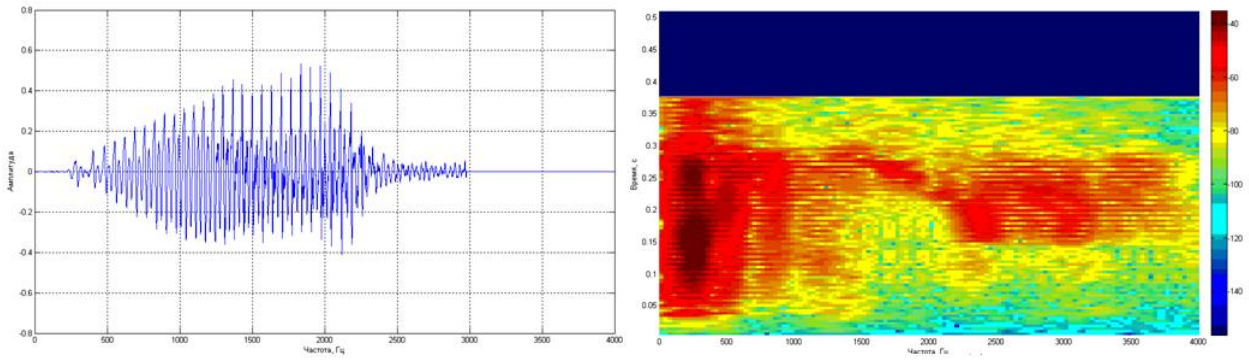


Рис. 6. Схема синтеза речевого сигнала при использовании метода наложения со сложением

Несомненным достоинством данного метода синтеза речевого сигнала является простота реализации и низкая алгоритмическая и вычислительная сложность [5].

В качестве примера рассмотрим процесс предварительной обработки речевого сигнала, которому соответствует слово «нет», произнесенное мужским голосом на русском языке. Частота дискретизации Fs = 8000 Гц – 16 бит на отсчет, количество каналов – 1. Длина окна преобразования Фурье Nf = 256.

На рисунке 7, а отображен отрезок данного речевого сигнала во временной области с нормированной амплитудой от –1 до 1 и размером 4096 отсчетов либо длительностью $4096 \cdot 1/Fs = 4096 \cdot 1/8000 = 0{,}512$ с. На рисунке 7, б отображена спектрограмма данного отрезка речевого сигнала.

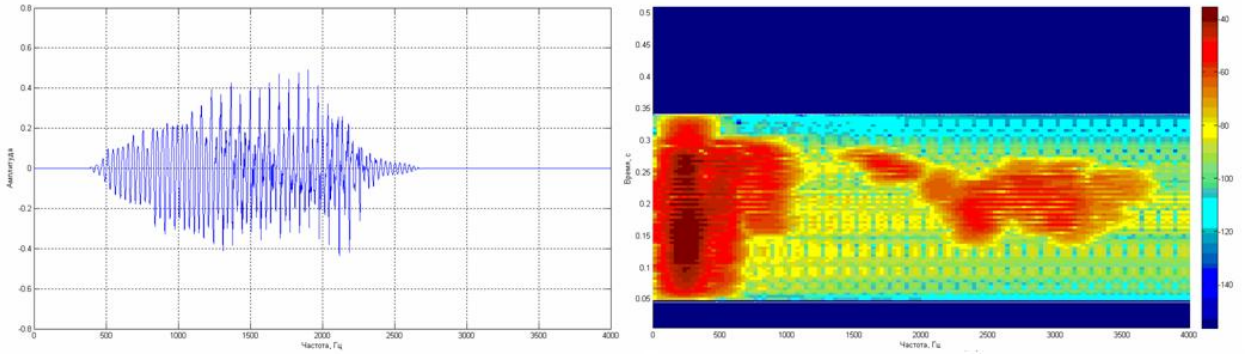


а) б)

Рис. 7. Исходный сигнал:
а – отрезок исходного сигнала; б – спектрограмма исходного сигнала

На рисунках 8, а и 8, б изображен отрезок синтезированного речевого сигнала, оригинал которого и его спектрограмму иллюстрирует рисунок 7.

Экспериментальные результаты показывают, что речь отличается высокой степенью разборчивости и хорошей узнаваемостью диктора даже при ограниченном числе спектральных компонент.



а) б)

Рис. 8. Восстановленный речевой сигнал:
а – отрезок восстановленного сигнала; б – спектрограмма восстановленного сигнала

Заключение. На основании существующего подхода спектрального кодирования речевых сигналов и с учетом антропоморфической обработки применительно к системам низкоскоростной компрессии, реализована система компрессии речевого сигнала для модели с разложением на спектральные компоненты.

Построена математическая модель устройства предварительной обработки звука в MATLAB. Наглядно продемонстрировано разбиение звукового сигнала на фреймы и его восстановление. Продемонстрирован расчет абсолютного порога слышимости, абсолютного порога маскирования и разбиения на критические полосы. Осуществлена аппаратная реализация предварительной обработки звука для спектрального кодера. Речь отличается довольно высокой степенью разборчивости и хорошей узнаваемостью диктора даже при ограниченном числе спектральных компонент. В ходе экспериментов установлено, что число спектральных компонент можно уменьшить в 2 и более раз, что приводит к уменьшению разрядности размера кодовой книги при кодировании речи как минимум на 1 бит.

ЛИТЕРАТУРА

1. Chassaing, R. DSP Applications Using C and the TMS320C6x DSK / R. Chassaing; A Wiley-Interscience Publication. – John Wiley & Sons, Inc. – 2002.
2. Edler, B. Technical description of the MPEG-4 audio coding proposal from University of Hannover and Deutsche Bundespost Telekom / B. Edler // ISO/IEC, JTC1/SC29/WG11 MPEG95/M0414, Oct. 1995. – P. 3617 – 3620.
3. Речевые интерфейсы ЭВС: метод. пособие для студ. спец. 40 02 02 «Электронные вычислительные средства» / А.А. Петровский, Ал.А. Петровский, Д.С. Лихачев. – Минск: БГУИР, 2004. – 66 с.
4. Чертков, В.М. Исследование возможностей предварительной обработки звука на основе анализа методов кодирования аудио данных в IP-телефонии: дис. … магистра техн. наук: 1 45 80 01 – Системы, сети и устройства телекоммуникаций / В.М. Чертков; Полоц. гос. ун-т. – 2009. – 84 с.
5. Лихачев, Д.С. Особенности аппаратной реализации системы компрессии речевой информации на основе слуховой модели человека / Д.С. Лихачев // Изв. Белорус. инж. акад. – 2004. – № 1(17)/3. – С. 122 – 125.

Поступила 01.06.2010

COMPRESSION OF SPEECH ON THE BASIS OF DOMINATING SPECTRAL COMPONENTS

S. MALTSEV, V. CHERTKOV

Present paper is considering the features of compression of a speech stream without deterioration of the restored speech on the basis of selection of dominating spectral components. Selection of dominating spectral components is made on the basis of criteria of psychoaudio principles. The circuit of a compression of speech and its mathematical model on the basis of spectral compression is presented. For an objective estimation of quality of the restored speech the population mean, a dispersion and среднеквадратическое deviation of level of difference of the restored signal from the initial is calculated. The circuit of synthesis of a speech waveform in real time is considered at usage of a method of imposing with the addition, realised on the basis of a set of equipment rooms and software DSP on the basis of microprocessor TMS320VC5510 unit.