

Центр Цифровой Обработки Сигналов,
Санкт-Петербургский Государственный Университет Телекоммуникаций
им. проф. М.А. Бонч-Бруевича, <http://www.dsp-sut.spb.ru>
193232, Санкт-Петербург, пр. Большевиков д.24, тел. 589-82-43
E-mail: barabashka@fnmail.com

Реферат. В статье представлен LPC вокодер со скоростью передачи параметров 1000 и 1200 бит/с, ориентированный на применение в системах служебной связи. Вокодер построен по схеме линейного предсказателя с попеременным возбуждением шумовыми или периодическими сигналами. Рассмотрены особенности построения функциональной схемы вокодера и реализации ее отдельных блоков. Модель вокодера работает в реальном масштабе времени на ПК. Приводятся результаты тестирования и оценки реализации на DSP.

1. Введение

Низкоскоростные вокодеры со скоростью передачи параметров менее 2000 бит/с отличаются синтетическим качеством речи и занимают нишу служебной связи. Для таких вокодеров одним из основных качественных показателей является слоговая разборчивость синтезированной речи [1], которая для устройств второго класса качества должна быть не хуже 75% [2]. Причем хорошая разборчивость речи должна сохраняться во всем диапазоне входных уровней речевого сигнала, при работе в условиях повышенных окружающих шумов, а так же при ограничении частотного диапазона входных сигналов телефонной полосой 300-3600 Гц.

В статье представлен вокодер со скоростью передачи параметров 1000 и 1200 бит/с. Вокодер построен на основе классической модели речеобразования, подробно рассмотренной в [3], по схеме линейного предсказателя с попеременным возбуждением шумовыми или периодическими сигналами. Скорость передачи зависит от длины окна анализа и способа упаковки выходного кадра. Модель вокодера реализована на персональном компьютере (ПК) на языке Си и работает в реальном масштабе времени. Это позволяет проводить испытания вокодера согласно методике [2] еще на этапе моделирования алгоритмов на ПК, до реализации вокодера на DSP, и тем самым снизить общее время разработки проекта.

Статья построена следующим образом. Краткое описание схемы анализатора дано в п.2. В п.3 рассматриваются задачи, связанные с вычислением и квантованием спектральных параметров. В п.4 рассмотрен метод выделения основного тона речи и признака тон/шум. В п.5 изложены вопросы квантования энергии и упаковки выходного кадра. В п.6 кратко описан синтезатор и в п.7 приводятся результаты моделирования алгоритма в реальном времени на ПК и оценки параметров реализации вокодера на DSP.

2. Анализатор

Функциональная блок-схема анализатора приведена на рис. 1. Анализ производится кадрами длительностью 30 или 22.5 мс. На стороне анализатора оцениваются и передаются в канал связи следующие параметры сигнала: энергия кадра, огибающая спектра, полученная на основе модели линейного предсказания (ЛП), частота основного тона речи и признак тон/шум.

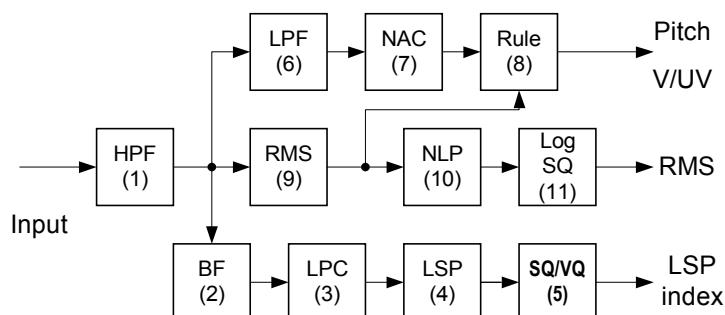


Рис. 1

Основным недостатком низкоскоростных LPC вокодеров данного типа является высокая чувствительность субъективного качества синтезированной речи к ошибкам в расчете параметров возбуждения синтезирующего фильтра, а так же к амплитудно-частотным искажениям спектра входной речи. Это является следствием несовершенства функциональной модели вокодера и ограниченного объема передаваемой информации. Прямая оценка параметров возбуждения, без коррекции на основе сравнения синтезированного сигнала с исходным сигналом, приводит к тому, что даже единичные ошибки в оценке спектральной огибающей сигнала, частоты основного тона, границ озвученных участков и. т. д. хорошо заметны на слух. Низкоскоростные LPC вокодеры, в отличие от CELP вокодеров, не сохраняют форму волны передаваемого сигнала, лишая возможности введения простого критерия оценки ошибки во временной области в виде энергии взвешенного разностного сигнала. При оценке ошибки синтеза в частотной области

так же требуется передача дополнительной информации об амплитудах гармоник основного тона, как это сделано в синусоидальных и MELP вокодерах, или о степени вокализации частотных полос в MELP [4].

Поэтому, при разработке данного вокодера уделялось повышенное внимание к исходной точности выделения и квантования спектральных параметров речевого сигнала, а так же к точности определения траектории основного тона речи и выделения признака тон/шум.

3. ЛП анализ

Исходный речевой сигнал с частотой дискретизации 8 кГц проходит через ФВЧ 4 порядка Чебышева (тип 2) с частотой среза 100 Гц и уровнем подавления на 50 Гц большим -60 дБ. Далее сигнал взвешивается корректирующим фильтром характеристикой $1 - 0.93 \cdot z^{-1}$, имеющим подъем АЧХ в области ВЧ, для улучшения передачи более слабых высокочастотных компонент речевого сигнала, определяющих в основном его разборчивость [1].

Для описания формантной огибающей спектра используется модель ЛП 10 порядка. Анализ осуществляется на длине кадра 180 или 240 отсчетов с предварительным взвешиванием окном Хемминга. Параметры модели ЛП оцениваются автокорреляционным методом с использованием рекурсии Леру-Гэгана. Полученные LPC коэффициенты преобразуются в линейные спектральные пары (LSP) по методу [5]. Далее производится квантование LSP по кодовым книгам, а полученный индекс передается в канал связи.

Сравнивалось два способа квантования – скалярное (SQ) с размером кодовой книги 34 бит и векторное (VQ) 23 бит. В SQ каждый из 10 коэффициентов LSP квантуется независимо с использованием своей кодовой книги, размерностью 3, 4, 4, 4, 4, 3, 3, 3, 3, 3 бита соответственно.

В VQ исходный LSP вектор разбивается на три подвектора размерностью 3, 3, 4 коэффициента, каждый из которых квантуется независимо с использованием кодовых книг размерностью 7, 7, 9 бит соответственно, организованных в целях быстрого поиска в виде бинарных деревьев. При поиске в кодовой книге ближайшего квантованного значения используется взвешенная евклидова метрика. Кодовая книга строилась методом центроидов ("K – средних") на основе выборки русской речи порядка 30 минут с удаленными паузами, записанной различными дикторами, как мужчинами, так и женщинами.

В результате проведенного сравнения квантователей было получено в среднем одинаковое качество звучания синтезированной речи. Однако, скалярный квантователь имеет меньшую максимальную ошибку в передаче спектральных характеристик чем векторный и поэтому реже допускает единичные дефекты звучания, особенно заметные в LPC вокодере, где спектральные ошибки квантования не исправляются в дальнейшем сигналом возбуждения, как например в CELP вокодерах.

4. Выделение основного тона и признака вокализации кадра

Как было подчеркнуто выше, для LPC вокодера особенно важно правильно определить величину основного тона речи (ОТ) и выделить границы озвученных участков речи.

$$A(l) = \frac{\left(\sum_{n=0}^N s(n) \times s(n-l) \right)^2}{\left(\sum_{n=0}^N s^2(n) \right) \times \left(\sum_{n=0}^N s^2(n-l) \right)}, l = PitchMin, PitchMax, \text{ для } \left(\sum_{n=0}^N s(n) \times s(n-l) \right) \geq 0 \quad (1)$$

Величина ОТ оценивается корреляционным методом на базе функции квадрата нормализованной автокорреляции (1). Для повышения помехоустойчивости обнаружения ОТ, анализ ведется в полосе частот до 1.5 кГц. Расчет A(L) осуществляется два раза за кадр. Положение центра окна анализа для различных значений величины сдвига L фиксировано относительно границ кадра. После расчета значений функции A(L) для всех возможных значений L ищутся ее локальные максимумы, образующие набор претендентов {Lp}. Отбор из числа претендентов осуществляется исходя из степени превышения значения A(Lp) порогов, взаимного отношения A(Lp) разных претендентов, степени отклонения величины ОТ претендента от траектории ОТ прошлых озвученных кадров. Особое внимание уделяется проверке претендентов Lp на удвоенное и половинное значение ОТ. Переход на удвоенное величины ОТ возникает, если профиль гласного звука имеет разную энергию четных и нечетных импульсов из-за неравномерного возбуждения голосовой щели [1]. Переход на половинное значение величины ОТ (соответствующее 2 гармонике частоты ОТ) возникает как из-за влияния близости первой форманты спектральной огибающей вокализированных звуков, подчеркивающей вторую гармонику, так и при ограничении спектра входной речи снизу (до 300 Гц), подавляющем основную гармонику ОТ.

Окончательное решение о величине ОТ принимается после медианного сглаживания с использованием значений ОТ для прошлого и будущего кадров. При этом в канал передачи вносится дополнительная задержка на кадр.

Использование функции квадрата нормализованной автокорреляции (1) делает отображение A(L) инвариантным как к общему уровню энергии входного сигнала, так и к различным уровням энергии сравниваемых участков, поэтому построенный на его основе выделитель ОТ более устойчив к быстрым вариациям амплитуды гласных звуков. Значения A(L) всегда лежат в диапазоне от 0 до 1. Функция A(L) выражает степень подобия двух смежных участков речи s(n) и s(n-L) на интервале N. Она равна квадрату

отношения энергии вектора ошибки предсказания $e(n)$, к энергии вектора $s(n)$, предсказываемого на участке N с помощью вектора $s(n-L)$, взятого с оптимальным усилением G : $e(n) = s(n) - G * s(n-L)$.

Решение о вокализации кадра принимается для сигнала, взятого в широкой полосе частот, на основе сравнения величины максимума функции (1), найденного в окрестности найденного значения OT , с порогом равным 0.3. Анализ в широкой полосе частот проводится с целью принятия совокупного решения о степени вокализации кадра в полосе последующего синтеза.

При энергии кадра менее -50 дБ он всегда считается шумовым.

5. Квантование энергии и упаковка кадра

Расчет энергии кадра осуществляется усреднением энергии сигнала на длине кадра. Далее эффективная амплитуда сигнала (RMS) квантуется логарифмически с шагом в 2 дБ с использованием кодовой книги на 5 бит. Для повышения комфорта восприятия пауз, при передаче кадров с уровнем сигнала меньшим -40 дБ осуществляется небольшое плавное расширение динамического диапазона сигналов по линейному закону в логарифмической области. (На -10 дБ для уровня -45 дБ.) Это позволяет уменьшить уровень шума на тихих паузах и снизить эффекты, связанные с проникновением фоновой речи, без заметного влияния на качество речи нормального уровня. Примененное решение плавно понижает уровень шума паузы в синтезаторе без переходных эффектов. Исследовалось также выделение пауз по признакам с принятием порогового решения речь/пауза. Такая классификация позволяет сделать вокодер с переменной скоростью и понижать скорость выходного цифрового потока на паузах в несколько раз. Однако, в целях шумопонижения этот вариант не применяется, так как, в случае ошибок, дает неприятные на слух всплески уровня шума в паузах при работе в условиях сильных окружающих помех.

Упаковка кадра в зависимости от выходной скорости и длительности окна анализа приведена в таблице. Для вокодеров (A) и (C) каждый второй кадр значения LSP прореживаются, а величина OT передается дифференциально.

Параметры	Тип вокодера					
	1200 (A)		1200 (B)		1000 (C)	
Скорость (бит/с)						
Длина окна анализа	22.5 мс	22.5 мс	30 мс	30 мс	22.5 мс	22.5 мс
LSP (бит)	15	19 (ост.)	23	23	10	13 (ост.)
OT и тон/шум (бит)	6	3	6	6	6	3
Энергия (бит)	5	5	5	5	5	5
Синхронизация (бит)	1	-	1	1	1	-
Резерв (бит)	-		2		2	
Итого (бит)	27	27	36	36	22	23

6. Синтезатор

Блок схема синтезатора приведена на рис 2. После распаковки пришедшего кадра осуществляется поиск квантованных значений параметров в кодовых книгах по индексам. Полученные значения LSP проверяются на устойчивость синтезирующего фильтра.

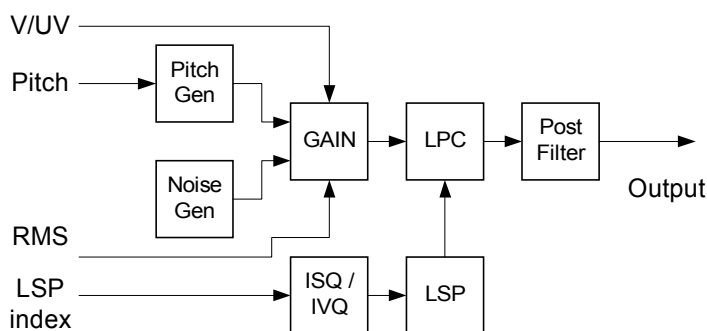


Рис.2

Синтез осуществляется по периодам основного тона для вокализованных участков и до заполнения текущего кадра для шумовых. Параметры синтеза линейно интерполируются между смежными значениями OT .

Тип возбуждения также изменяется синхронно с периодами OT . В качестве сигнала возбуждения на невокализованных фрагментах используется белый шум, состоящий из импульсов с одинаковой амплитудой и случайной фазой, а на вокализованных – импульсный отклик фазового звена, дополненный одиночными импульсами.

Коэффициент усиления сигнала возбуждения выбирается так, чтобы энергия синтезированного фрагмента была равна энергии кадра. При этом учитывается энергия затухающего отклика синтезирующего фильтра на сигнал возбуждения прошлого кадра.

Синтезирующий фильтр ЛП и постфильтр реализованы в едином БИХ фильтре в прямой форме. Интерполяция параметров синтезирующего фильтра осуществляется на уровне LSP. АЧХ постфильтра выбиралась от обратной к АЧХ взвешивающего фильтра до линейной. В результате прослушивания предпочтение было отдано неполной компенсации предискажений с небольшим подъемом ВЧ.

7. Полученные результаты

Все варианты вокодеров реализованы на ПК и работают в режиме реального времени под управлением ОС WINDOWS-95/NT на базе стандартной полнодуплексной звуковой платы ввода-вывода Sound Blaster PCI-128 фирмы Creative Labs. Вокодер написан на языке Си и при работе на ПК с процессором Intel P-II-400 MHz занимает порядка 10...15% процессорного времени. После перевода модели в фиксированную точку реализация на DSP семейства TMS320c54x потребует по нашей оценке не более 10 MIPS.

При сравнении вариантов (А) и (В) методом прослушивания оказалось, что вокодер с величиной кадра 22.5 мс имеет небольшое преимущество в разборчивости, т.к. временная огибающая энергии сигнала передается более точно, чем при величине кадра 30 мс, а это важно для передачи согласных, имеющих взрывной характер. Вокодер (В) имеет небольшое преимущество в естественности тембра, т.к. использует передачу параметров LSP каждый кадр, а вокодер (А) интерполирует параметры спектральной огибающей через кадр.

Экспресс испытания, проведенные по методике [2], для вокодера 1200 бит/с (А) дали оценку слоговой разборчивости на уровне 77 %, а для вокодера 1000 бит/с (С) – около 75%. В эксперименте участвовало два диктора и трое слушателей, принималось по 20 слоговых таблиц. Использовался микрофон ДЭМШ-1А и телефон ТДМ-1Э.

Библиография

1. Сапожков М.А., Михайлов В.Г. Вокодерная связь. – М.: Радио и связь, 1983. – 248 с., ил.
2. ГОСТ В 20775 – 75. Передача речи по трактам связи, оборудованным аппаратурой засекречивания. Требования к разборчивости речи. Метод измерения.
3. Deller J., Proakis J., Hansen J. Discrete-Time Processing of Speech Signals. – IEEE press, 2000.
4. McCree A., Truong K., George E. and others. A 2.4 kbit/s MELP Coder Candidate for the New U. S. Federal Standard. // Proceedings of ICASSP-96, pp. 200-203.
5. Иванов В.Н. Вычисление линейных спектральных частот // Электросвязь. – 1997. – №6. – С. 25-27.



DSP Department
Saint Petersburg State Telecommunication University
named after professor M.A. Bonch-Bruevich, <http://www.dsp-sut.spb.ru>
Bolshevikov av. 24, Saint Petersburg, 193232, Russia,
phone: 589-82-43, e-mail: barabashka@fnmail.com

Abstract. This paper describes the low bit rate LPC speech coder (vocoder) operating at 1000 and 1200 bit/s. The vocoder was designed for using in a special telecommunication area. The functional model of vocoder and its design features are presented. Two ways of an output frame packing are discussed. The real-time implementation of the vocoder model on a PC allows us to get the intelligibility test results during the algorithm development phase. The estimated DSP implementation performance is about 10 MIPS on the TMS320C54x processor family.

1. Introduction

Low bit rate speech coders at speed below 2000 bit/s has synthetic quality of output speech and using in special telecommunication areas. One of important quality measures for these coders is a subjective intelligibility in form of a score of the right receiving meaningless syllables [1]. For a good quality level the score should be no less than 75% [2]. Good intelligibility must be remained for all input speech levels, in presence of background noise and for input speech in 300-3600 Hz frequency band.

This paper describes the low bit rate speech coder (vocoder) operating at 1000 and 1200 bit/s. The vocoder is based on the traditional speech production model [3]. It uses a linear prediction (LP) model with periodic and noise excitation alternately. A transmitting speed depends from the frame length and the bit allocation method. The vocoder model is implemented on a personal computer (PC) using C language and this implementation works in real-time. It allows us to get the intelligibility test results using method [2] at a step of algorithm development on PC before the DSP implementation. It is important for decreasing a development time of the entire project.

This paper is organized as follows. A brief encoder description is done in section 2. The methods of LPC analyses and LSP quantization are considered in section 3. The questions of pitch estimation and tracking are discussed in section 4. The energy quantization and the encoder output bit allocation are viewed in section 5. In section 6 the decoder is explained. The real-time implementation details and the subjective test results are reported in section 7.

2. Encoder

A functional block diagram of encoder is shown in Figure 1. The analysis window (frame) length is 22.5 or 30 ms long. The encoder calculates and transmits into channel these speech parameters: a frame energy, a spectrum envelope based on a LP model parameters estimation, a pitch estimation and a voiced/unvoiced (V/U) decision.

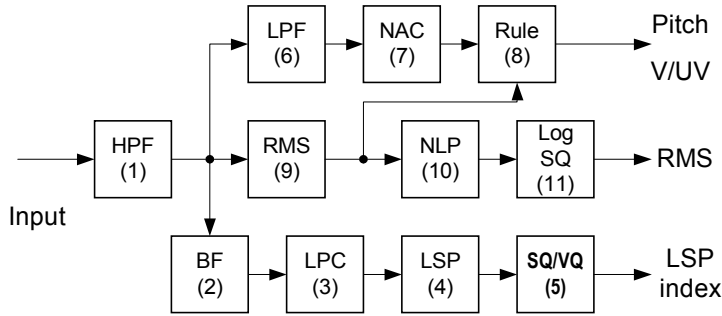


Figure 1. Encoder block diagram

The main drawback of traditional LPC vocoders is a high sensitivity of the synthetic speech quality from the excitation parameters estimation errors and the input speech spectrum distortions. A direct parameters estimation without an analysis by synthesis procedure leads to a rough output speech even for single errors in pitch determination, spectral envelope quantization or V/UV decision. The limit of the output bit rate doesn't allow them to improve speech quality too. In addition, the LPC vocoders don't transmit a phase information about input speech, so they can't use a simple time domain error criterion based on the comparing of input and synthetic speech waveform as the CELP coders do. Frequency domain methods of error estimation are more complicated and need transmitting additional information bits too. Like it is done for transmitting the pitch harmonics magnitudes in the STC and MELP [4] coders or the band voicing decisions in the MELP coder.

Therefore in this vocoder a special care was attended to the initial parameters estimation accuracy and to the parameters quantization quality.

3. Linear Prediction Analysis

An input signal with the 8 kHz sampling rate is filtered out by use of a Chebishev type 2 high pass filter of the 4th order with a cut off frequency of 100 Hz and a rejection about -60 dB at 50 Hz. Then the signal is weighted by a preemphasis filter with characteristic of $H(z)=1 - 0.93z^{-1}$ that increases the relative energy of the high-frequency speech spectrum and improves the overall intelligibility [1].

A 10th order linear prediction analysis of preemphased input speech is performed using the 180 or 240 point Hamming window. The LP coefficients are calculated using the autocorrelation method and the Leru-Guegan recursion. On the base of obtained LP coefficients the linear spectral pairs (LSP) are calculated using method [5]. Then all LSP are quantized using code books and obtained indexes are transmitted trough the channel.

Two methods of the LSP quantization are compared. A scalar quantizer (SQ) uses a 34 bit size code book (CB) and a vector quantizer (VQ) uses a 23 bit size code book. In the SQ every LSP coefficient is processed independently using the own CB with dimension 3, 4, 4, 4, 3, 3, 3, 3, 3 bits correspondingly. In the VQ the initial LSP vector is split up into three subvectors with dimension 3, 3, 4 coefficients. Then each subvector is processed independently using the own CB with dimension 7, 7, 9 bits. Each CB is organized in form of a binary tree for quick searching. A weighted Euclidean metric is used for searching the nearest quantized LSP value. All CB are built using the K-means method. A CB training sequence was about 30 minutes of Russian speech with deleted pauses that was recorded by the different male and female speakers.

Both quantizers have a near the same synthetic speech quality. But the SQ has a less maximum of the spectral error than the VQ. So the vocoder with the SQ generates less audible single artefacts of a synthesized speech. These single spectral errors are not corrected by an excitation signal in the LPC vocoders as the CELP coders do.

4. Pitch determination and V/UV decision

The reliable pitch determination and tracking are very important for the LPC vocoders.

$$A(l) = \frac{\left(\sum_{n=0}^N s(n) \times s(n-l) \right)^2}{\left(\sum_{n=0}^N s^2(n) \right) \times \left(\sum_{n=0}^N s^2(n-l) \right)}, l = PitchMin, PitchMax, \text{ for } \left(\sum_{n=0}^N s(n) \times s(n-l) \right) \geq 0 \quad (1)$$

Input speech is filtered using a low pass Batterworth filter with a cut off frequency of 1.5 kHz. Then a squared normalized autocorrelation function (1) is calculated for lags between 18 and 145 samples two times for every frame. An analysis window is centered for the different lags. Then all local maximums of the A(l) are searched. They

are sorted by value and all pitch candidates {Lp} are chosen among them. A selection procedure within candidates is based on their values comparison and on a memory about the pitch trajectory of the past voiced frames. A special care is paid to pitch doubling and pitch halving effects. A pitch doubling occurs if a magnitude envelope of speech has different energy of the odd and even glottal impulses [1]. A pitch halving (corresponding the second harmonic of pitch frequency) occurs as result of the first formant influencing or the phone band filtering with a low frequency cut off up to 300 Hz. The final decision about the pitch is performed after a median smoothing procedure involving the past and future pitch estimation values.

The squared normalized autocorrelation function (1) is invariant to signal energy, therefore the pitch estimation procedure based on (1) is more reliable in cases of speed varying amplitudes of the vowel sounds. All values of the A(L) are placed always between 0 and 1. The function (1) is a similarity measure between two speech vectors s(n) and s(n-L) on the interval with length of N samples. It expresses the ratio between the energy of prediction error e(n) and the energy of predicted signal s(n) when the signal s(n) is predicted by the signal s(n-L) with optimal gain G: $e(n) = s(n) - G * s(n-L)$.

A V/UV decision is calculated for the wide frequency band speech by comparing a maximum of the function (1) with threshold level of 0.3. The maximum of (1) is searched around the previously found pitch value. The wide frequency band analysis was chosen for making a V/UV decision in the band of a following synthesis. If the current frame energy is less than -50 dB the frame is always considered as unvoiced.

5. Energy quantization and bit allocation

After a frame energy calculation the averaged frame amplitude (RMS) is transformed to a logarithm value and quantized with step of 2 dB using a 5 bit CB. For decreasing of a pause noise level the dynamic range expansion is performed for the frames with energy less than -40 dB. It is a linear function in a log area with the knee point at -40 dB and the depth of -10 dB for the input level of -45 dB. This “soft” noise cancellation decision allows the encoder to decrease also a background speech problem during pauses without the quality deterioration for the normal level speech. A binary speech/pause decision is also considered. It allows the encoder to decrease the output bit rate on pauses but it is not used for the pause noise cancellation because of the annoying noise spikes that arise on the pauses with strong background noise.

A bit allocation depends on the output bit rate and the frame length. The possible allocations are shown in Table. For the (A) and (C) vocoders the LSP index is transmitted every odd frame only. The pitch value for them is coded differentially every second frame too.

Parameters	Vocoder type					
	1200 (A)		1200 (B)		1000 (C)	
Bit rate (bit/s)	22.5 mc		30 mc		22.5 mc	
Frame length	22.5 mc	22.5 mc	30 mc	30 mc	22.5 mc	22.5 mc
LSP index (bit)	15	19 (old)	23	23	10	13 (old)
Pitch & V/UV (bit)	6	3	6	6	6	3
RMS (bit)	5	5	5	5	5	5
Synchronization (bit)	1	-	1	1	1	-
Reserved (bit)	-		2		2	
Total bits per frame	27	27	36	36	22	23

6. Decoder

A functional block diagram of decoder is shown in Figure 2. After the bit stream unpacking all parameters values are extracted from the code books. All dequantized LSP values are checked on the stability of a synthesis filter.

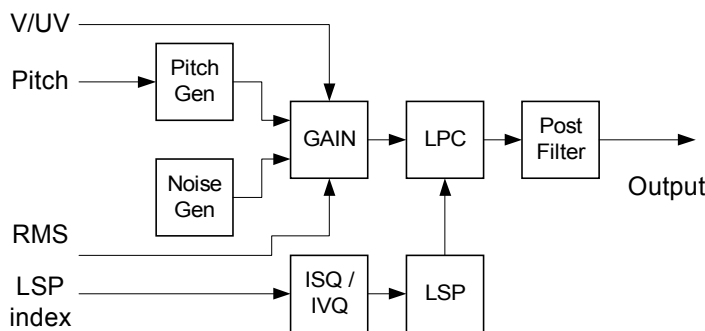


Figure 2. Decoder block diagram

A synthesis procedure is performed on the pitch length base for the voiced frames and on the frame length base for the unvoiced. All synthesis parameters include the pitch are interpolated linearity between the contiguous pitch length sections.

An excitation signal type is also changed synchronically with the crossings of pitch length section boundaries. For the unvoiced frames the white noise excitation is performed using the unit amplitude impulses with a random sign. For the voiced frames the impulse response of an all pass phase filter is used with adding a signal in form of single impulses.

An excitation signal gain is calculated so that the synthesized frame energy to be equal the received energy value. The non-zero memory of the synthesizer LP filter is always taken in account.

The synthesizer LP filter and the postfilter are realized as a single direct form IIR filter. The LP filter coefficients are interpolated at the LSP stage. A frequency response of the postfilter is experimentally chosen from the linear characteristic up to the inverting characteristic of the preemphasis filter. After the subjective listening tests a preference was given to the partial compensation of preemphasis.

7. Results

All vocoders have been implemented on a PC using the float point arithmetic and the C language. They work in real time under control of the WINDOWS-95/NT operating system using a full duplex sound card Sound Blaster PCI-128 from Creative Labs. This PC based implementation takes about 10...15% of the Intel P-II-400 MHz processor load. After code converting into the fixed point arithmetic a DSP implementation will take about 10 MIPS for the TMS320c54x processor family.

During the subjective listening tests it has been found that the vocoder (A) with 22.5 ms frame length has a little more intelligibility than the 30 ms one (B), because a time envelope of the signal energy is transmitted more precisely and that is important for the stop consonants transmitting. But the vocoder (B) has more natural and mellow timbre because its LSP parameters are transmitted every frame without the every second frame interpolation as the vocoders (A) and (C) do.

The fast tests of subjective intelligibility of meaningless syllables were taken using the method [2]. For the vocoder operating at 1200 bps (A) the score of right receiving syllables is estimated at the level of 78 %, for the vocoder at 1000 bps (C) – at 75%. Two speakers and three listeners took part in the experiment. They received 20 tables with 50 syllables each using the microphone ДЭМШ-1А and the ear phones ТДМ-1Э.

Bibliography

- 1. Сапожков М.А., Михайлов В.Г.** Вокодерная связь. – М.: Радио и связь, 1983. – 248 с., ил.
- 2. ГОСТ В 20775 – 75.** Передача речи по трактам связи, оборудованным аппаратурой засекречивания. Требования к разборчивости речи. Метод измерения.
- 3. Deller J., Proakis J., Hansen J.** Discrete-Time Processing of Speech Signals. – IEEE press, 2000.
- 4. McCree A., Truong K., George E. and others.** A 2.4 kbit/s MELP Coder Candidate for the New U. S. Federal Standard. // Proceedings of ICASSP-96, pp. 200-203.
- 5. Иванов В.Н.** Вычисление линейных спектральных частот // Электросвязь. – 1997. – №6. – С. 25-27.