Solution Substituting in Eq. (5.56),

(a) $f = 400$, $f_s = 8000$, $SNR|_{max} = 15.1892 = 11.8179$ dB

(b) $f = 400$, $f_s = 16000$, $SNR|_{max} = 60.7927 = 13.8385$ dB

(c) Use of post reconstruction low pass filter effectively reduces quantization noise. Since, noise is distributed uniformly, use of filter reduces it by a factor of (f_c/f_s) where f_c is cut-off frequency of the LPF. Also noise goes to denominator of Eq. (5.56) which, we can rewrite for this case as

$$SNR|_{max} = \frac{3f_s^2}{8\pi^2 f^2}\left(\frac{f_s}{f_c}\right) = \frac{3f_s^3}{8\pi^2 f^2 f_c} \quad (5.58)$$

(d) Subsituting, for $f = 400$, $f_s = 8000$, $SNR|_{max} = 121.5854 = 20.8488$ dB

for $f = 400$, $f_s = 16000$, $SNR|_{max} = 972.6834 = 29.8797$ dB

Self-test Questions

13. The objective behind DPCM coding is to reduce the variance of the signal to be encoded. Is that correct?
14. Is it true that an FIR digital filter is always stable because its impulse response is finite?
15. The slope overload error in DM can be reduced by increasing (a) step size, (b) sampling frequency, (c) both. Which one is correct?
16. Is the step size in ADM variable?

5.7 VOICE CODERS (VOCODERS)

We listen, in turn, to one person after another pronounce a word or a sequence of words. One has a high-pitched voice, another a low pitch. One speaks with a foreign accent, another has no accent. One enunciates clearly, another slurs words and so on. Yet we understand each speaker. These considerations make it clear that there is a variability (within limits) that can be allowed in the waveform that must impinge on a listener's ear before there appears a loss of recognition of the spoken word. It therefore occurs to us that to transmit speech we need not transmit the precise waveform generated by the speaker. Rather we can transmit information from which a waveform can be reconstructed at the receiver which is only similar to, rather than identical to, the waveform generated by the speaker. We anticipate that by allowing ourselves this lattitude we can operate a digital source coded transmission system at a lower bit rate. This expectation is indeed borne out. The source coders employed are called *vocoders* (voice coders) and they operate at a significantly lower bit rate than even ADM. Typically, vocoder bit rates are in the range 1.2 to 2.4 kb/s. However, the resulting reproduced voice has a synthetic-sounding and a somewhat artificial quality. As a result, vocoders are employed for special applications where it is acceptable to trade speech quality for the advantage of low bit rate. Applications are found in military communications, operator recorded messages, etc.

The people who developed vocoders studied and took account of the physiology of the vocal cords, the larynx, the throat, the mouth and the nasal passages, all of which have a bearing on speech generation. They also studied the physiology of the ear and the psychology associated with the manner in which the brain interprets sounds heard. A discussion of these matters is, however, beyond the scope of this text and the interested reader should see the references[5].

Voice Model It appears that speech consists of, or, at least, can be well approximated by a sequence of *voiced* and *unvoiced* sounds that are passed through a filter. The voiced sounds are sounds that are generated by the vibrations of the vocal cords. The unvoiced sounds are generated when a speaker pronounces such letters as; "s", "f", "p", etc. In this latter case the sounds are formed by expelling air through lips and teeth. A generalized representation of a vocoder is shown in Fig. 5.46. The filter represents the effect, on the generated sounds, of the mouth, throat and nasal passages of the speaker. In the vocoder, the voiced sounds are simulated by an impulse generator whose frequency is the fundamental frequency of vibration of the vocal cords. The unvoiced sounds are simulated by a noise source. Altogether, all vocoders employ the scheme shown in Fig. 5.46 to generate a synthesized approximation to a speech waveform. They differ only in the techniques employed to generate the voiced and unvoiced sounds and in the characteristics, and design of the filter.

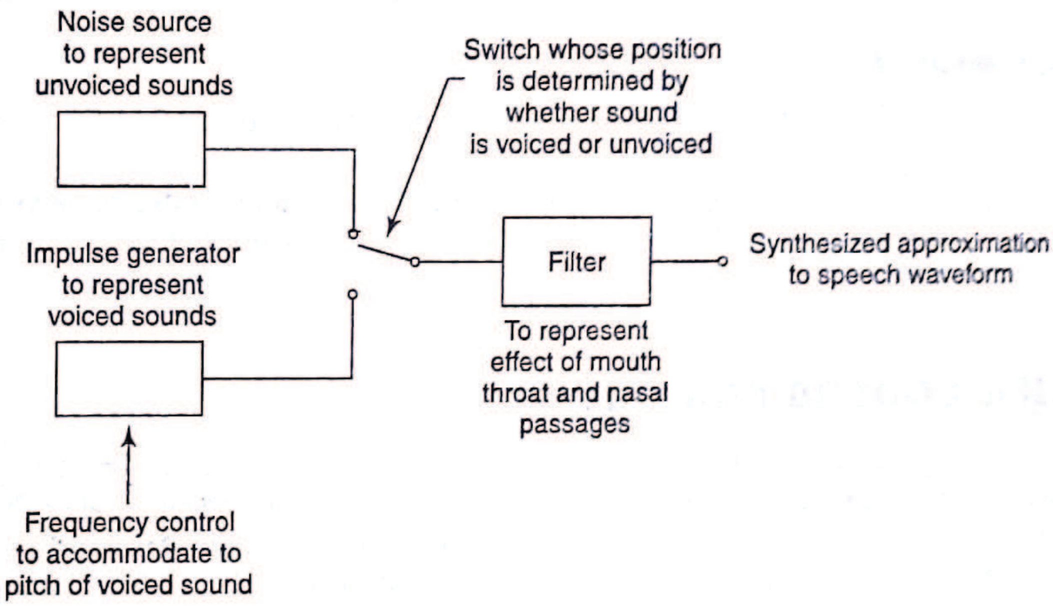

Fig. 5.46 Speech model used in vocoders.

5.7.1 Channel Vocoder

One of the many vocoder systems is the *channel vocoder* shown in Fig. 5.47. In this encoding system, the spectrum of the input speech is divided into 15 frequency ranges each of bandwidth 200 Hz. If the output of one of the bandpass filters were a sinusoidal waveform of fixed amplitude then the cascaded rectifier and 20 Hz low-pass filter would provide a dc voltage of magnitude proportional to that amplitude. In the real case represented in Fig. 5.47 each 20 Hz low-pass filter will provide instead a voltage which is proportional to the amplitude at the output of its associated 200 Hz bandpass filter.

Additionally, the input speech is applied to a frequency discriminator followed by a 20 Hz low-pass filter. When the signal is voiced, the output of the filter provides a voltage which is proportional to the voice frequency. This frequency is the *pitch* of the voice. The discriminator-filter combination is special in that when the speech is unvoiced the output of the filter is a smaller voltage than we encounter for voiced speech. Using a detector we can then determine, by noting the output of the discriminator-filter combination, whether the speech is voiced or unvoiced, and if voiced, the voltage detected is determined by the pitch.

The outputs of the 16, 20 Hz low-pass filters are sampled, multiplexed and A/D converted. If the sampling is at the Nyquist rate of 40 samples/s (corresponding to signals of 20 Hz bandwidth) and if we use 3 bits/sample to represent each voltage sample, the bit rate is

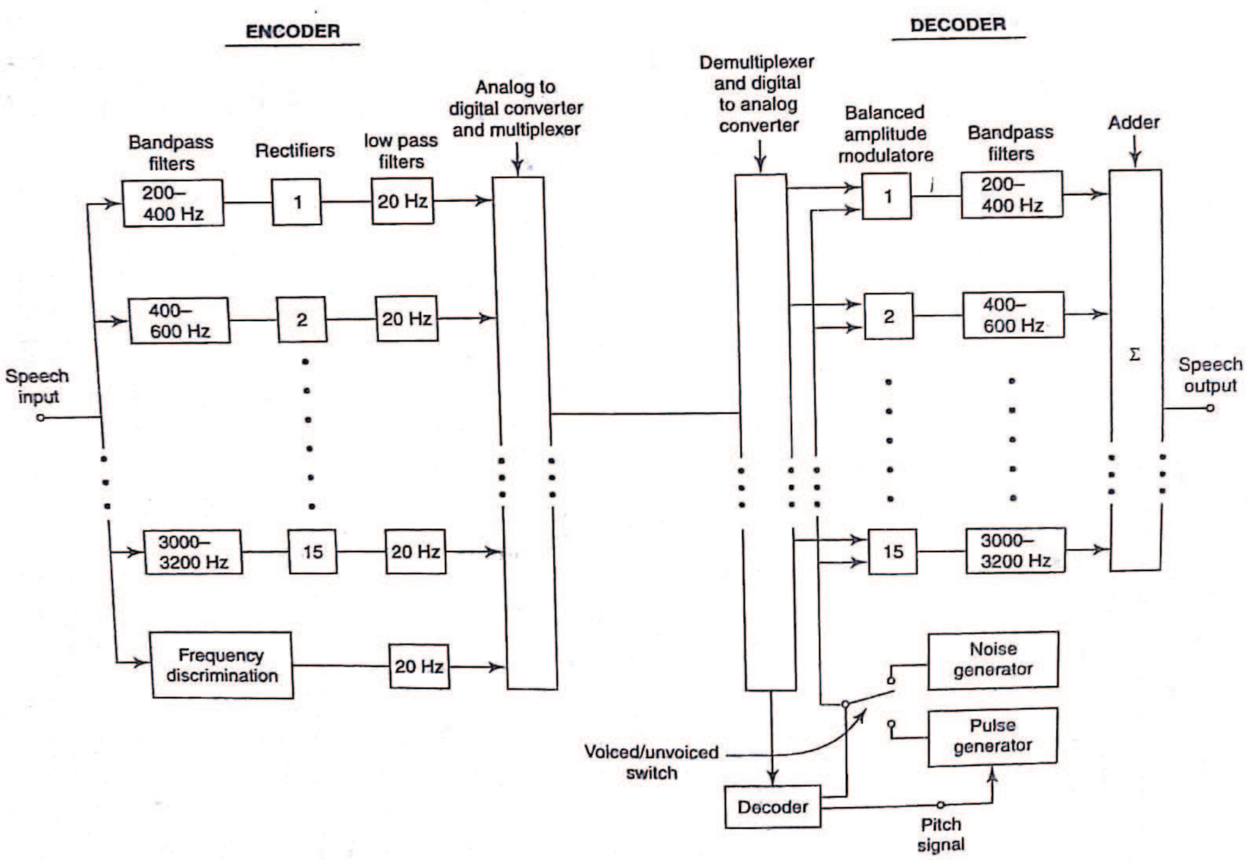

Fig. 5.47. Vocoder.

$$R = 40\,\frac{\text{samples/s}}{\text{filter}} \times 16 \text{ filters} \times 3 \text{ bits/sample}$$
$$= 1.9 \times 10^3 \text{ bits/s} \qquad (5.59)$$

In any particular case the bit rate will depend on the number of low-pass filters, their passband and the number of bits allowed per sample. Typical bit rates vary from 1.2×10^3 to 2.4×10^3 b/s

At the vocoder receiver, as is to be seen in Fig. 5.47, the signal is demultiplexed and decoded that is, converted back to analog form. Corresponding to each filter-rectifier combination at th encoder, there is provided, at the decoder, a balanced amplitude modulator and a bandpass filte with identical pass band. It will be recalled that a balanced modulator is a modulator that provide zero carrier output at zero modulation input. The carrier input to each modulator is the noise o pulse generator waveform. The modulation input is the amplitude signal (of each of the 15 recti fier-filters) provided by the encoder. At each sampling, the amplitude information is updated as i the information about whether the speech waveform is voiced or unvoiced and, if voiced, the pitc is provided by the discriminator.

Suppose, now, that in some sampling interval it happens that the speech is voiced and that th fundamental frequency of the generated sound is 450 Hz. This sound will generate an amplitud signal at the output associated with rectifier 2. The sound will have harmonics, so that there ma be amplitude outputs at frequencies $2 \times 450 = 900$ Hz, $3 \times 450 = 1350$ Hz, etc., which will be see at the outputs of rectifiers 4, 6 etc. The decoder will receive the information that the speech i voiced. Hence the decoder switch will connect to the pulse generator and the pulse generato frequency will be set to the pitch of the voice. The pulse generator waveform then constitutes carrier input to all modulators. However, only modulators 2, 4, 6 etc. will generate outputs. The outputs are amplitude modulated reproductions of the pulse-generator waveform and hence a comprised of a fundamental and a succession of harmonics. However, the filter following modula tor 2 will suppress all except the fundamental, the filter following modulator 4 will suppress a except the second harmonic, etc. All these waveforms are added and the resultant output wavefor consists of a combination of a fundamental and harmonics with the same relative amplitudes existed in the speech input. When the speech input is unvoiced the sound is noise-like and i spectrum extends through the entire speech-frequency range although not necessarily in a unifor manner. In this case we can expect that all paths in the encoder will provide an output as will a paths in the decoder.

It is interesting to note that a vocoder can also be designed which transmits only the large envelope detector outputs, completely ignoring the remaining outputs. For example if the large three outputs are sent, 4 bits/filter $\times$ 3 filters = 12 additional bits would be transmitted to tell th receiver which three filter outputs are sent. In this case, the bit rate becomes

$$R = 40\,\frac{\text{samples/s}}{\text{filter}} \times 3 \text{ filters} \times 3 \text{ bits/sample} + 40 \text{ samples/s} \times \frac{12 \text{ bits}}{\text{sample}}$$
$$= 840 \text{ bits/s} \qquad (5.60)$$

Unfortunately, the quality of the signal is severely degraded. However, the bit rate is significantl