改善된 勵起信号의 4800BPS LPC 보코우터

A 4800 BPS LPS Vocoder With Improved Exitation.

*은 종 관 (Chong Kwan Un)
**성 원 용 (Won Yong Sung)

ABSTRACT

We present an improved 4800 bps LPC vocoder system that virtually eliminates the buzzy effect from synthetic speech. Excitation signal in the new system is formed by adding high-pass filtered pitch pulses or random noise to a baseband residual signal (0 - 600 Hz) that has been coded by pitch predictive PCM. Since the baseband residual is used as a part of excitation, the system is also robust to V/UV and pitch errors. According to our informal listening tests, the synthetic speech of the new system does not have the buzzy effect. As a result the vocoder speech quality is more natural than that of a conventional LPC vocoder.

I. INTRODUCTION

Linear predictive coding (LPC) is presently known to be the most effective speech compression technique for a narrow band communication system. In a low rate (2.4 kbits/s) LPC vocoder, the vocal tract is modeled by a set of LPC coefficients and the excitation signal is represented by a sequence of pulses for voiced sound and random noise for unvoiced sound. Although the speech quality of this system is reasonably good at 2.4 kbits/s, one serious problem is that it has some discernible buzziness that makes synthetic speech unnatural. This undesirable effect is known to be primarily due to the monotonic characteristics of the excitation signal.

To remedy this problem, several researchers have studied different approaches [1-3]. Sambur et al. [2] attributed the lack of naturalness to the high peak factor and the monotoneness of synthetic speech, and proposed the use of non-impulsive source excitation. Makhoul et al. [3] suggested a frequency selective mixed source model for partially unvoiced speech. It has been reported that these methods resulted in some success in improving the speech quality over the conventional pulse excited LPC vocoder.

As an alternative approach to remove the monotoneness of excitation, an improved form of excitation signal that is non-stationary and looks like LPC residual signal is proposed. We propose to use a hybrid excitation signal that may be generated by adding the baseband residual (0 - 600 Hz) to highpass filtered

*正会員, 韓国科学技術院 教授
**正会員, 金星社 中央研究所

pitch pulses or random noise. Since the residual is time varying, the combined signal is not monotonic. Consequently, the hybrid signal is better suited for as an excitation signal than the monotonic pulses.

Furthermore, the use of the baseband residual as a part of excitation signal masks pitch errors including V/UV decision errors. Accordingly, the system is robust to V/UV or pitch errors. This aspect is another significant advantage of using hybrid excitation.

Detailed description of the vocoder system and discussion on computer simulation results follow.

II. DESCRIPTION OF THE VOCODER SYSTEM

The overall system block diagram is shown in Fig. 1. Input speech is sampled at 8 KHz after low-pass filtering (0 - 3.4 KHz), and then preemphasized. LPC analysis in the present system is done in the same way as in a conventional LPC vocoder. The autocorrelation method of LPC analysis is used to model the vocal tract. The baseband residual that is to be transmitted and used as a part of the excitation signal is generated by LPC inverse filtering.

Pitch is extracted from the baseband residual by the average magnitude difference function (AMDF) method [4]. Pitch dependent redundancy of the baseband residual is eliminated by a pitch prediction loop. The resultant signal is down-sampled (6 to 1) and coded by 2-bit PCM. This information together with the normal LPC parameter set (10 reflection coefficients, pitch, and gain) are transmitted at the transmission rate of

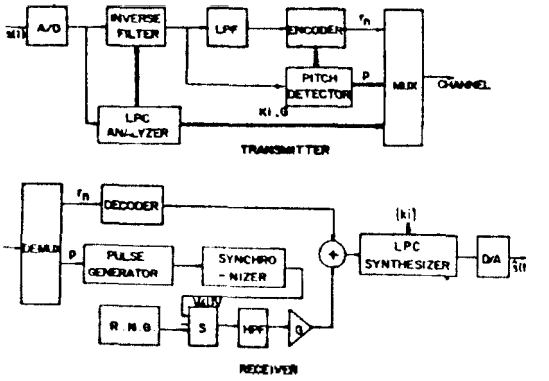

Figure 1. Block diagram of 4800 bps LPC vocoder with hybrid excitation

4.8 kbits/s.

In the receiver, the excitation signal is generated by adding the high-pass filtered pulses or random noise to the decoded baseband residual. In adding the two signals, they must be synchronized. Otherwise, the synthetic speech quality would be degraded. In our system peak detection and coincidence measurement are used for synchronization. LPC synthesis is done in the same way as in a conventional LPC vocoder using the combined excitation signal. Residual coding and synchronization methods are described in detail in the following subsections.

A. Residual Coding

As the present LPC vocoder is a relatively low rate system, it is important to code the baseband residual efficiently. For a 4.8 kbits/s LPC vocoder, we can use only about 2.6 kbits/s for residual encoding, if we assume that 2.2 kbits/s is used for coding LPC coefficients, pitch, and gain. Since the residual is used to remove the monotoneness of excitation signal, its bandwidth needs not be large. For our system, the residual signal is

bandlimited to 600 Hz by a low pass filter with its stop band frequency at 667 Hz.

For coding the baseband residual, conventional waveform coding schemes such as PCM, adaptive differential PCM (ADPCM) and adaptive delta modulation (ADM), have been studied, but none of them has given a satisfactory result. Because the residual signal produced by inverse filtering has very little correlation between the samples, ADM and ADPCM yield little prediction gain. However, one can make use of the fact that the residual signal has strong correlation between the samples delayed by a pitch period. Hence, the use of a pitch prediction loop prior to actual coding would eliminate the redundancy due to pitch, thus giving a significant bit rate reduction or SQNR gain when the output signal is coded.

A block diagram of the baseband residual coder is shown in Fig. 2. The baseband residual signal is first passed through a pitch prediction loop. Since pitch extraction is done in this system, the use of the pitch loop would not add any significant complexity or additional computations. The pitch predictor uses a fixed prediction coefficient. In an unvoiced block, the prediction coefficient is simply set to zero. The output of the pitch prediction loop is down-sampled and encoded by a 2-bit PCM quantizer. An adaptive quantizer with its quantization step size proportional to the root mean square (RMS) value of the baseband residual is used.

In the receiver, the signal is decoded, upsampled, low pass filtered, and then fed to the inverse pitch predictor. In this residual decoding system, a finite impulse response (FIR) digital filter is used. Hence, the number

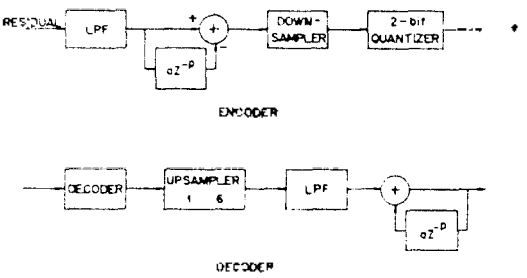

Figure 2. Block diagram of baseband residual coder

of multiplications can be reduced significantly since most of the upsampled signal values are zero. The measured SQNR of this coder is about 11 dB, which is about 4 dB higher than SQNR obtained when ADM coding is used. When the present coding scheme is used, the resulting synthetic speech is found out to be as good as the reference synthetic speech that has been produced using unquantized baseband residual.

B. Generation of Hybrid Excitation Signal

For generation of excitation signal at the synthesizer, the decoded baseband residual is added synchronously to pitch pulses in a voiced block or to random noise in an unvoiced block. As the baseband residual is asynchronous with pitch pulses, synchronization of pulses with the baseband residual must be done. If the residual were an uncoded wide band signal, synchronization can be done by simple peak detection. But, for the baseband residual signal that has been distorted as a result of quantization, one needs to have a sophisticated synchronization scheme.

In our system, synchronization is done as follows. Locations where pitch pulses are to be planted are found by peak detection and pitch based coincidence measurement.

We use a modified "blanking and run-down" method originally used by Gold and Rabiner for detection of peaks including a large peak in the blanking region [5]. As seen in Fig. 3, after each detected peak, there is a blanking interval followed by a run-down period. Whenever a peak exceeds the level of the blanking region or exceeds the level of the run-down region, a peak is detected and the operation is restarted. The length of the blanking and run-down region is proportional to the pitch period. The blanking time and the rundown time constant are 0.4 P and P [P is a pitch period.], respectively. Once major peaks are detected, peaks on which pitch pulses are to be planted are determined by the following coincidence measuring algorithm. The peak with the largest amplitude of the current block is first chosen as a reference peak. Next, pitch peaks are found by having a blanking window followed by a search window with their lengthes proportional to a pitch period. The peaks in the blanking window interval are not detected, but the peak in the search window interval is detected as the one on which a pitch

Figure 3. "Blanking and run-down" operation in peak detection

pulse is to be planted. If there are two or three peaks in one search window interval, the peak whose distance from the previous pitch peak is closer to a pitch period is selected. If there is no peak in the search window interval, the center of the search window interval is selected as the point where a pitch pulse is to be added. The blanking and search window intervals are 0.9 P and 0.2 P, respectively. After selecting a new pitch peak, the new peak is selected following the same procedure.

III. COMPUTER SIMULATION AND DISCUSSION

The new system has been simulated on a computer, and its performance has been tested with male and female speech. Fig. 4 shows the waveforms of original speech, synthetic speech of the proposed system, and synthetic speech of the conventional LPC vocoder. Also, various excitation waveforms are shown. The advantage of the proposed system can be seen through the comparison of the waveforms. According to our informal listening tests, the buzzy tone that exists particularly in a low pitched male voiced sound of the conventional LPC vocoder no longer exists in our synthetic speech. Yet, there still exists some intelligible roughness as compared with the original speech. Nevertheless, the synthetic speech of the new vocoder sounds more natural than that of the conventional system.

The possible sources of roughness introduced to this synthetic speech are believed to be due to the limited residual signal bandwidth, and the quantization noise of the baseband residual. To determine the source of roughness, several experiments have been done. A synthetic speech with the residual bandwidth of 800 Hz was compared with that having the residual bandwidth of 600 Hz. The former appeared closer to natural

speech, but the difference between the two synthetic speech was hardly perceptible. Also, synthetic speech generated by the unquantized residual was compared with that generated by the quantized residual. The difference between them was readily perceptible when ADPCM or ADM was used for residual encoding. But, the difference was reduced significantly when pitch predictive PCM was used.

One can note that the new vocoder system is similar to the residual excited linear prediction (RELP) vocoder [6] in that both systems use the baseband residual signal for generation of excitation signal. However, the use of the baseband residual in each system is conceptually different. In RELP vocoder the baseband residual that is of normally larger

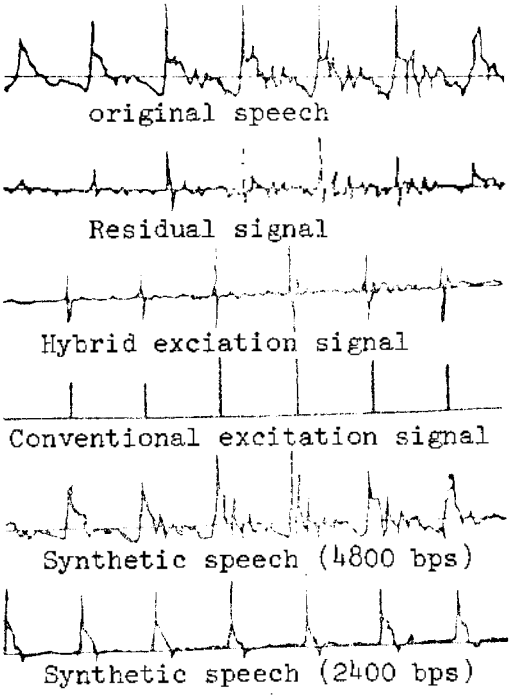

Figure 4. Comparison of various speech and excitation waveforms

bandwidth (0 - 800 Hz or more) than that of the present system is used as the sole excitation signal after non-linear distortion and spectral flattening. In the new vocoder system the residual signal is used as a supplement to pitch pulses to make the resulting excitation signal look like unfiltered LPC residual. Hence the requirement of accurate coding of the baseband residual is less severe in the present system. Also, the bandwidth of the residual needs not be as large as that of RELP system. Accordingly, coding of the baseband (0 - 600 Hz) residual requires only about 2667 bits/s.

IV. CONCLUSION

We have presented a new 4800 bits/s LPC vocoder system with improved excitation. The new vocoder system that uses hybrid excitation signal has the following advantages:

(1) Because the baseband residual signal is used as a part of excitation, the excitation signal is not monotonic. Consequently, the synthetic speech is virtually free from the unnatural buzzy effect.

(2) The use of the residual signal masks V/UV and pitch errors, and thus minimizes the degradation of synthetic speech quality resulting from those errors.

These advantages have been confirmed by computer simulation. The price for the significant improvement of speech quality is a modest increase (about 2400 bits/s) of transmission rate. The proposed system can be implemented easily by adding a residual coder and synchronizer to any type of the conventional LPC vocoder.

REFERENCES

1. B. S. Atal and N. David, "On Synthesizing Natural-sounding Speech by Linear Prediction," IEEE ICASSP '79 RECORD, pp. 44-47, Apr. 1979.
2. M. R. Sambur, A. E. Rosenberg, L. R. Rabiner, and C. A. McGonegal, 'On Reducing the Buzz in LPC Synthesis," J. Acoust. Soc. Amer., Vol. 63, No. 3, Mar. 1978.
3. J. Makhoul, R. Viswanathan, R. Schwartz, and A. W. F. Huggins, "A Mixed Source Model for Speech Compression and Synthesis," J. Acoust. Soc. Amer., Vol. 64, Dec. 1978.
4. C. K. Un and S. C. Yang, "A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF," IEEE Trans. on ASSP., Vol. ASSP-25, No. 6, Nov. 1977.
5 B. Gold and L. R. Rabiner, "Parallel Processing Technique for Estimating Pitch Period of Speech in the Time Domain," J. Acoust. Soc. Amer., Vol. 46, pp. 442-448, Aug. 1969.
6. C. K. Un and D. T. Magill, "The Residual-Excited Linear Prediction Vocoder with Transmission Rate Below 9.6 kbits/s," IEEE Trans. on Com., Vol. COM-23, pp. 1466-1473, Dec. 1975.