# QUALITY ENHANCEMENT OF NARROWBAND CELP-CODED SPEECH VIA WIDEBAND HARMONIC RE-SYNTHESIS

*Cheung-Fat Chan and Wai-Kwong Hui*

Department of Electronic Engineering, City University of Hong Kong
83, Tat Chee Avenue Kowloon, HONG KONG
Email: eecfchan@cityu.edu.hk

## ABSTRACT

Results for improving the quality of narrowband CELP-coded speech by enhancing the pitch periodicity and by regenerating the highband components of speech spectra are reported. Multiband excitation (MBE) analysis is applied to enhance the pitch periodicity by re-synthesizing the speech signal using a harmonic synthesizer. The highband magnitude spectra are regenerated by matching to lowband spectra using a trained wideband spectral codebook. Information about the voiced/unvoiced (V/UV) excitation in the highband are derived from a training procedure and recovered by using the matched lowband index. Simulation results indicate that the quality of the wideband enhanced speech is significantly improved over the narrowband CELP-coded speech.

## 1. INTRODUCTION

Code excited linear predictive (CELP) coding is one of the widely used low-bit-rate speech coding techniques[1]. It is well-known that speech produced by CELP coders suffers from quality degradation which are generally described as muffing with hoarse or noisy characteristics. The muffing characteristics are mainly due to the lack of high frequency components because these low-bit-rate speech coders were designed to operate in narrowband (0-4 kHz) with 8 kHz sampling frequency and the hoarse characteristics are mainly due to the use of noisy stochastic excitation. Because there is a wide installation base of CELP coders in the commercial world, for examples, the Federal Standard FS1016 coder[3], the VSELP coder of EIA/TIA IS54[4] and the half-rate GSM coder[5], there is an urgent need to further improve the quality of CELP-coded speech while keeping their encoding format intact. This research is to improve the quality of CELP-coded speech by regenerating the high-frequency components (4-8kHz) at the decoder and also by reducing the coding noise inherent in voiced harmonics. The techniques investigated are based on wideband re-synthesis of CELP-coded speech through the use of multiband excitation (MBE) model.

## 2. ENHANCEMENT SYSTEM

Fig. 1 shows the block diagram of the proposed enhancement system. In this enhancement system, the lowband information are obtained from narrowband CELP-coded speech using MBE analysis and passed to the wideband MBE synthesizer for synthesis. The lowband information include the V/UV decisions, the magnitudes and phases of the voiced harmonics, and the signal spectrum declared as unvoiced. The enhancement system, therefore, needs to estimate the highband information from all the information available in the lowband. These include the V/UV decisions, the magnitudes and phases of the voiced harmonics, and the magnitude spectrum of the unvoiced signal for the highband.

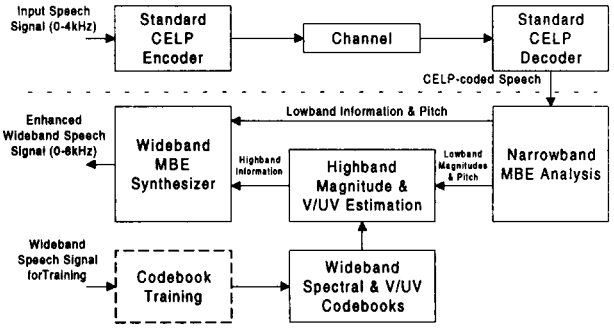


**Fig. 1:** Block Diagram of the Enhancement System

### 2.1. Noise Reduction in CELP-Coded Speech

It is well-known that the background noise in CELP-coded speech is high because noisy stochastic excitation is used. This background noise in between the pitch harmonics can be easily observed by comparing the spectral plots of the original speech (solid curve) and the CELP-coded speech (doted curve) shown in Fig. 2. In this work, a harmonic synthesizer based on MBE model is proposed to "clean" up the background noise. MBE model is capable of producing high quality speech because it allows the flexibility of mixing voiced and unvoiced energies in the frequency domain[6]. In MBE model, speech spectrum is divided into a number of signal bands which are centered on the pitch harmonics. Each band can be individually declared as voiced or unvoiced. It is well known that speech produced by MBE coders is perceptually less nosier than those produced by CELP-based coders because a smooth harmonic excitation rather than a noisy stochastic excitation is used. In this research, the idea for reducing the coding noise is to perform the MBE analysis on the reproduction speech and then to replace the voiced portions of speech spectrum by the corresponding harmonic spectrum synthesized by the MBE synthesizer. Note that only the voiced

portions are replaced, the unvoiced portions, however, are passed to the MBE synthesizer without modification. In this experiment, voiced harmonics are synthesized using sinusoidal oscillators with quadratic phase interpolation using the measured phases[7]. The unvoiced spectrum is converted back to time domain via inverse FFT and added to the voiced signal to obtain the output signal. With this enhancement in voiced spectra, the re-synthesized speech was shown by listening tests to be less noisier than the ordinary CELP-coded speech. Fig. 2 also shows a spectral plot of the re-synthesized speech (dashed curve) which clearly indicates the capability of this harmonic re-synthesis technique for cleaning-up the coding noise. Pair-wise comparison tests have also confirmed that the CELP-coded speech with pitch enhancement is preferred over the CELP-coded speech without enhancement.

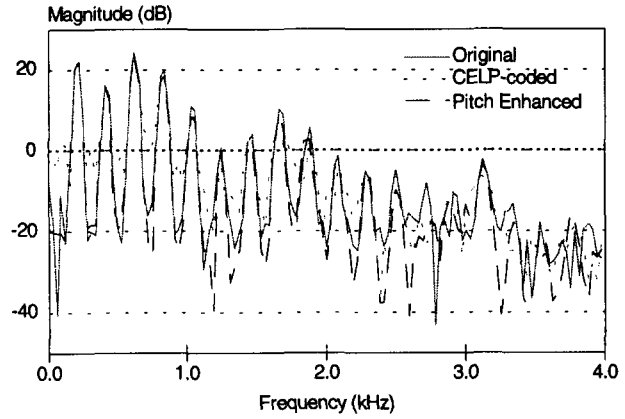


**Fig. 2:** Spectra of the Original, Coded and Pitch Enhanced Speech

## 2.2. Estimation of Highband Spectrum Envelope from Lowband Information

The highband spectrum can only be estimated from the lowband information available in the decoder. In this research, a straightforward classification technique was developed to extract the correlation between the lowband and highband spectra by using a codebook training approach. Specifically, a 20 minutes wideband (16 kHz sampling frequency) speech database contributed by many speakers was first designed and used to train a spectral codebook which consists of 1024 line spectral pair (LSP) vectors, each has a dimension of 18. The wideband codebook was generated by using the well-known generalized Lloyd algorithm with splitting as initialization. However, the training vectors are clustered based on minimizing the spectral distortion in the lowband portion of the wideband LPC spectrum only. The lowband distortion measure is defined as:

$$SD_{AV} = \left[\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\left(10\log\left|H_n(\omega)\right| - 10\log\left|\hat{H}_n(\omega)\right|\right)^2 d\omega\right]^{\frac{1}{2}} \quad (1)$$

where $H(\omega)$ is the wideband LPC spectrum from the training set, and $\hat{H}(\omega)$ is the LPC spectrum of the codeword determined by using minimum distortion rule which is based on lowband distortion only. During codebook training, the centriod in each cluster, i.e., a wideband LSP vector, is calculated by averaging the wideband training vectors that mapped to the cluster. Since the highband distortion is not used in training the wideband codebook, the average spectral distortion in the highband and the percentage outlier will be good indicators for the degree of correlation between the lowband and highband spectra. Table 1. shows the performance in terms of average spectral distortion and percentage outlier for the trained codebook with 1024 LSP vectors. The evaluation were performed using the same training set for generating the codebook. The result indicates that $SD_{AV}$ in the highband and lowband are comparable and, therefore, their correlation is high. Also, the outlier is small, that means the chances to have well-matched lowbands but highly distorted highbands are slim.

| | $SD_{AV}$ (dB) | Outlier % (> 4 dB) |
|---|---|---|
| Lowband | 2.4 | 5.3 |
| Highband | 2.8 | 7.2 |

**Table 1.** Performance of the Trained Codebook

During the estimation of highband envelope for the synthesis stage, the narrowband LPC spectrum of the CELP-coded speech is matched against the lowband portion of the wideband LPC spectrum characterized by the 18-dimension LSP vector in the codebook. The codevector that achieved the smallest lowband spectral distortion is selected. The highband portions of the optimum LPC spectrum is then sampled (at pitch harmonics) and the band magnitudes in the high-frequency portion of speech spectrum are determined as the sampled magnitudes for MBE synthesis. Note that the band magnitudes in the lowband portion of the re-synthesized speech are directly derived from MBE analysis of CELP-coded speech. As the spectrum envelope in the highband are obtained from the wideband LPC envelope, the speech energy in the highband is still needed to be normalized. In this work, a normalization factor is determined based on equalizing the lowband energies of CELP-coded speech and the re-synthesized speech. Specifically, the band magnitudes $A_m$ obtained from MBE analysis of narrowband CELP-coded speech are matched against the sampled band magnitudes $\hat{A}_m$ of lowband portion of the optimum wideband LPC spectrum, i.e., $\hat{A}_m$ are derived from sampling the LPC spectrum at the pitch harmonics as $\hat{A}_m = g\left|H(m\omega_0)\right|$, where $\omega_0$ is the pitch frequency. The normalization factor is then computed as:

$$g = \frac{\sum_{m=1}^{M/2} A_m\left|H(m\omega_0)\right|}{\sum_{m=1}^{M/2}\left|H(m\omega_0)\right|^2} \quad (2)$$

The band magnitudes in the highband are then evaluated as $A_m = g\left|H(m\omega_0)\right|$ for $M/2 + 1 \le m \le M$ where $M$ is the total number of bands in the wideband spectrum. In order to reduce the sporadic peaks in the estimated highband spectrum, a time-domain smoothing technique is applied. An exponential decaying function weights the highband magnitudes from the previous frames so as to make the high-frequency spectra to slowly evolve.

Fig. 3 shows the LPC spectra of the CELP-coded, the wideband estimated speech and the original wideband speech. Table 2 shows the performance of the proposed estimation algorithm.

| | $SD_{AV}$ (dB) | Outlier % (> 4 dB) |
|---|---|---|
| Lowband | 1.8 | 3.5 |
| Highband | 2.6 | 9.2 |

**Table 2.** Performance of the Estimation Algorithm

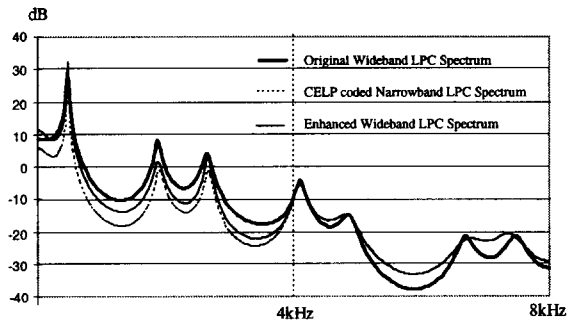


**Fig. 3:** LPC Spectra of the Original, Coded and Enhanced Speech

## 2.3. Estimation of V/UV Information for the Highband

It is necessary to estimate the voiced/unvoiced (V/UV) information in the highband. Many wideband enhancement methods typically assume that the high-frequency spectra consist of entirely unvoiced signals which is generally not correct[8]. In this research, the highband V/UV information is estimated from the spectrum envelope based on an assumption that the spectrum envelope and the distribution of V/UV energies in speech signals are correlated. This fact can be readily shown by observing that high-energy formant regions contain mostly voiced energies while low-energy high-frequency regions of speech spectra largely contain unvoiced energies. In this work, the same wideband speech database was used to design a separate V/UV codebook. Each V/UV codeword in the codebook is derived from a clustered set of V/UV mixture functions obtained as by-products of MBE analysis[9]. This V/UV codebook has a one-to-one correspondence to the wideband spectrum codebook derived earlier.

During wideband re-synthesis when the best three codewords in the spectrum codebook are determined, the corresponding V/UV mixture functions from the V/UV codebook are also extracted. These V/UV mixture functions are weighted and then smoothed over several frames using similar technique described previously. The V/UV decisions in the highband are then assigned according to the derived V/UV information. Table 3 shows the result in terms of percentage error in V/UV decisions. The result is derived from comparing the V/UV decisions obtained from wideband MBE analysis of original speech and the recovering process through CELP coding, narrowband MBE analysis and highband regeneration. It is surprising from this result that V/UV decision error in the highband is smaller than those in the lowband. This may attribute to the fact that some of the voice harmonics in the lowband may have been corrupted by the CELP-coding process and subsequent MBE analysis may incorrectly identify them as unvoiced bands.

| | V/UV Decision Errors (%) |
|---|---|
| Lowband | 16.67 |
| Highband | 11.18 |

**Table 3:** Percentage V/UV Decision Errors

## 2.4. Wideband MBE Synthesis

With all the necessary information available, synthesis of the wideband signals is rather easy. Note that, for ordinary CELP decoders, if a LPC frame size of, say, 20 ms is employed, 160 samples of narrowband speech are needed to be synthesized. For synthesis of wideband signal, 320 samples are needed. In this work, a MBE synthesizer which generates voiced speech in time domain and unvoiced speech in frequency domain is employed. For voiced speech, the phase information for the lowband band magnitudes are extracted from the CELP-coded speech and utilized to control the phases of harmonic oscillators at frame boundaries. All band magnitudes are linear interpolated between frames and their phase functions are quadratic functions. Since phase information for highband are unavailable, phases for voiced harmonics in highband are made to slowly evolve. Unvoiced spectrum in lowband are extracted from the original CELP-coded speech spectrum unmodified. Unvoiced spectrum in highband are constructed by multiplying the estimated highband spectrum envelop with an unity energy white noise spectrum. The wideband unvoiced spectrum is then converted to time domain by using a 512-point IFFT and 320 samples of unvoiced signal are obtained with a weighted overlap-add procedure. Finally, the voiced and unvoiced signals are added to generate the synthetic speech.

# 3. SIMULATION RESULTS

In this simulation, all speech signals were initially band-limited and sampled at 16 kHz (This signal is denoted as wideband input signal). The wideband input signal is then band-limited to 4 kHz using a lowpass digital Butterworth filter and sub-sampled by a factor of two to obtain the narrowband input signal. The narrowband signal is then coded by a FS1016 CELP coder for evaluation. The re-synthesis algorithm was applied directly on the reproduction speech from the CELP decoder. The enhanced speech was then played back through a D/A converter operating at 16 kHz sampling rate. Pair-wise listening tests were performed to compare the qualities of the narrowband CELP-coded speech and the wideband enhanced speech. Listeners in these tests felt that the wideband re-synthesized speech is clean with crispy high-frequency characteristics and they all overwhelmingly agreed that the enhanced speech is more pleasant to listen to than the narrowband CELP-coded speech. The FFT spectra of a wideband original speech, its CELP-coded speech and the corresponding wideband enhanced speech are shown in Fig. 4. Note that the coding noise shown in CELP-coded speech spectrum in (b) was cleaned-up after the enhancement process. The spectrograms for an 1s speech segment are also shown in Fig. 5. These

spectrograms clearly show that the highband estimation algorithm is very accurate.

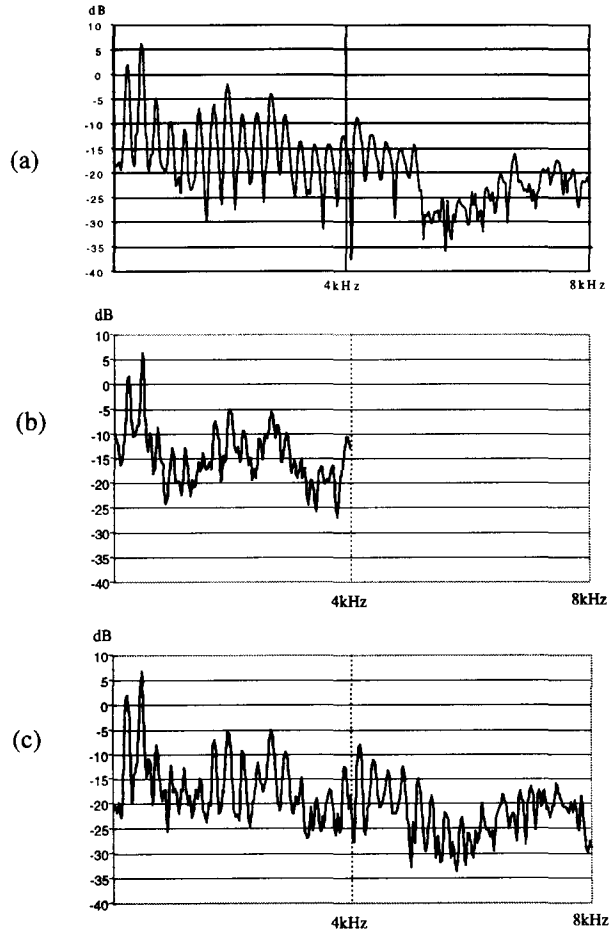


**Fig. 4** FFT Spectra of (a) wideband original, (b) narrowband CELP-coded, (c) wideband enhanced, speech

## 4. CONCLUSION

Wideband harmonic re-synthesis of narrowband coded speech was shown to be capable of improving the quality of CELP-coded speech without altering their encoding format. This was achieved by reducing the coding noise in between the voiced harmonics and regenerating the highband information from the lowband information available in the decoder.

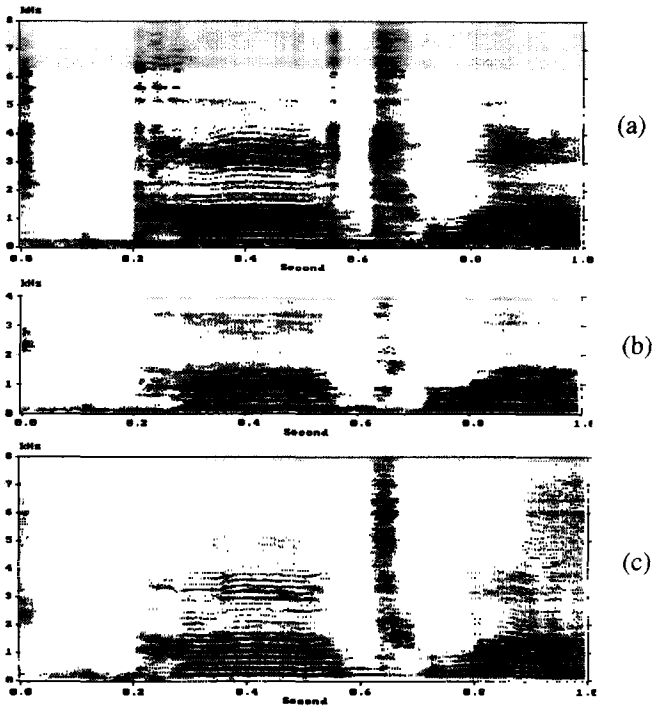


**Fig. 5:** Spectrograms of (a) wideband original, (b) narrowband CELP-coded, (c) wideband enhanced, speech

## 5. REFERENCES


1. Schoreder, M.R., and Atal, B.S., "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates," ICASSP, pp.937-940, 1985.
2. Kroon, P., and Atal, B.S., "Pitch Predictors with High Temporal Resolution," ICASSP, pp.661-664, 1990.
3. Campell, J.R., Jr., Tremain, T.E., and Welch, V.C., "The Federal Standard (FED-STD) 1016 4800 bps CELP Voice Coder," Digital Signal Processing I, pp.145-155, 1991.
4. Gerson, I.A., and Jasiuk, M.A., "Vector Sum Excited Linear Predictive (VSELP) Speech Coding at 8 kbps," ICASSP, pp.461-464, 1990.
5. Gerson, I.A., and Jasiuk, M.A., "A 5600 bps VSELP Speech Coder Candidate for Half-Rate GSM," IEEE Workshop on Speech Coding for Telecommunications, pp.43-44, 1993
6. Griffin, D.W., and Lim, J.S., "Multi-band Excitation Vocoder," IEEE Trans. On Acoustics, Speech, and Signal Processing, Vol. ASSP-36, No. 8, pp.1223-1235, August, 1988.
7. Digital Voice Systems, "IMMARSAT M Voice Codec, Version 2," IMMARSAT-M Specification, IMMARSAT, Feb. 1991.
8. Carl, H. and Heute, U., "Bandwidth enhancement of narrow-band speech signals", EUSIPCO VII. pp.1178-1181, 1994.
9. Chan, C.F., "High-Quality Synthesis of LPC Speech Using Multiband Excitation Model," European Conference on Speech Communication and Technology, pp.535-538, 1993.