

original MELP method and the synthesized speech using improved MELP method are compared and analyzed. In the synthesized speech with improved method, unnaturalness caused by mutation between syllables is reduced, and subjective hearing experience is improved.

A U-V transition frame and a V-U transition frame are chosen from the speech sample, whose waveform comparisons are shown in Fig.4 and Fig.5.

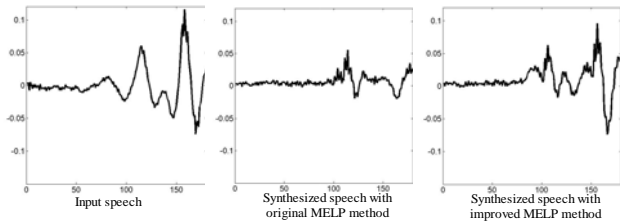


Figure 4. Simulation results of a U-V transition frame.

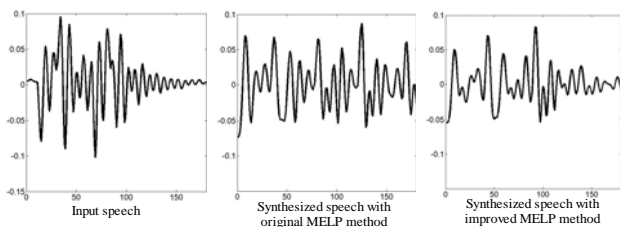


Figure 5. Simulation results of a V-U transition frame.

Viewing from the time domain waveforms, transition frames are synthesized as unvoiced or voiced frames in original MELP algorithm, leading to obvious waveform distortions. The improved synthesized waveform matches better with the speech sample, so the quality of synthesized speech gets better, too.

The test result shows that, the improved algorithm performs better in transition frames. Besides, with the addition of new algorithm, only 180 times of multiplications and additions are added to the computation of every frame. The processing time is almost the same.

B. PESQ test

PESQ (perceptual evaluation of speech quality) is an objective measurement for estimating subjective quality obtained in listening-only tests. We evaluate the performance of the improved algorithm through PESQ tests, using various

voice samples including Chinese male, Chinese female, English male, and English female. The PESQ-MOS scores are shown in Table 1.

TABLE I. PESQ TEST RESULTS

Speech samples	Original MELP	Improved MELP
Male, Chinese	2.53	2.64
Female, Chinese	2.47	2.61
Male, English	2.93	2.99
Female, English	2.88	2.95

The test results show an increase in PESQ-MOS scores, indicating that the speech quality of the improved algorithm is better than the original one.

V. CONCLUSIONS

This paper proposes an improved frame type decision algorithm in MELP which classifies speech frames into three types: the unvoiced, the voiced, and the transition frame. Based on the original MELP algorithm, the improved method determines transition frame and its transition type by peakiness detection of LP residual and energy calculation of sub-frames. Test results show that the improved algorithm reduces distortions caused by simple U/V decision and inaccurate parameters of transition frames, and improves the quality of synthesized speech with a small additional of computation. It has a certain practicality in researches and engineering applications.

REFERENCES

- [1] U.S. Department of Defense, Analog to digital conversion of voice by 2,400 bit/second mixed excitation linear prediction, 1998.
- [2] McCree A, Kwan T, George E B, and Viswanathan V, "A 2.4 kbit/s MELP coder candidate for the new U.S. Federal Standard," Acoustics, Speech, and Signal Processing, IEEE International Conference, vol.1, pp.200-203, 1996.
- [3] Supplee L M, Cohn R P, Collura J S and McCree A, "MELP: the new Federal Standard at 2400 bps," Acoustics, Speech, and Signal Processing, IEEE International Conference, vol.2, pp.1591-1594, 1997.
- [4] McCree A, Kwan T, George E B, Barnwell T P and Viswanathan V, "An Enhanced 2.4 Kbit/s Melp Coder", Speech Coding for Telecommunications, IEEE Workshop, pp.101-102, 1995.
- [5] McCree A V and Barnwell T P, "A mixed excitation LPC vocoder model for low bit rate speech coding", Speech and Audio Processing, IEEE Transactions, vol.3, pp. 242-250, 1995.