4. Experimental results

According to the proposed method, we present four experiments, including accuracy, recognition rate, speed and parameter estimation.

4.1 Accuracy

We use the same dataset listed in Table 2 as testing data. Each data contains 199 frames whose frame size is 20ms. We categorize errors as “Step error” and “Accumulative error”, which are listed in the Table 6. “Step error” denotes the error derived in a single step, and “Accumulative error” denotes the error accumulated so far from the first step. Each of them contains the four errors described in Sec. 3.2.1.

Table 6: Statistic of errors

“Step error” “Accumulative error”	Max		Average	
	Absolute error	Relative error	Absolute error	Relative error
Pre-emphasis	1.6016 1.6016	6.2535% 6.2535%	0.4815 0.4815	0.0388% 0.0388%
Windowing	4.5742 5.5742	48.9457% 52.1051%	0.5269 0.7882	0.0786% 0.1174%
FFT	475262.2812 475256.2812	3170883.4949% 3199194.0618%	2598.6019 25973.2001	34.5058% 34.4888%
Triangular filter	614.625 469407.875	5.5027% 744.926%	104.0894 13407.6643	0.092% 11.8522%
Log	2.4587 1914.0642	0.0221% 20.0826%	0.6782 113.4216	0.0071% 1.1844%
DCT	158.7221 1527.7252	2217.4588% 9146.8397%	21.6741 187.6255	2.0115% 17.4136%
Delta	1.2732 2585.6598	46.3005% 55244.371%	0.4948 313.5883	0.0556% 35.4507%
Acceleration	1.4165 1072.3618	55.8275% 144851.1159%	0.6009 132.0415	0.1679% 37.5213%
Energy	0.9423	0.0046%	0.4057	0.0019%

The overall accumulated error is about 29.682%. In the table, the errors of energy are “Step error” because there is no accumulated data before the energy step. Most errors are derived from the FFT step because the computational complexity of FFT is a lot more than other steps. We can see that the error is diminished after the Log step and this is due to the property of log function, as shown in Fig.8.

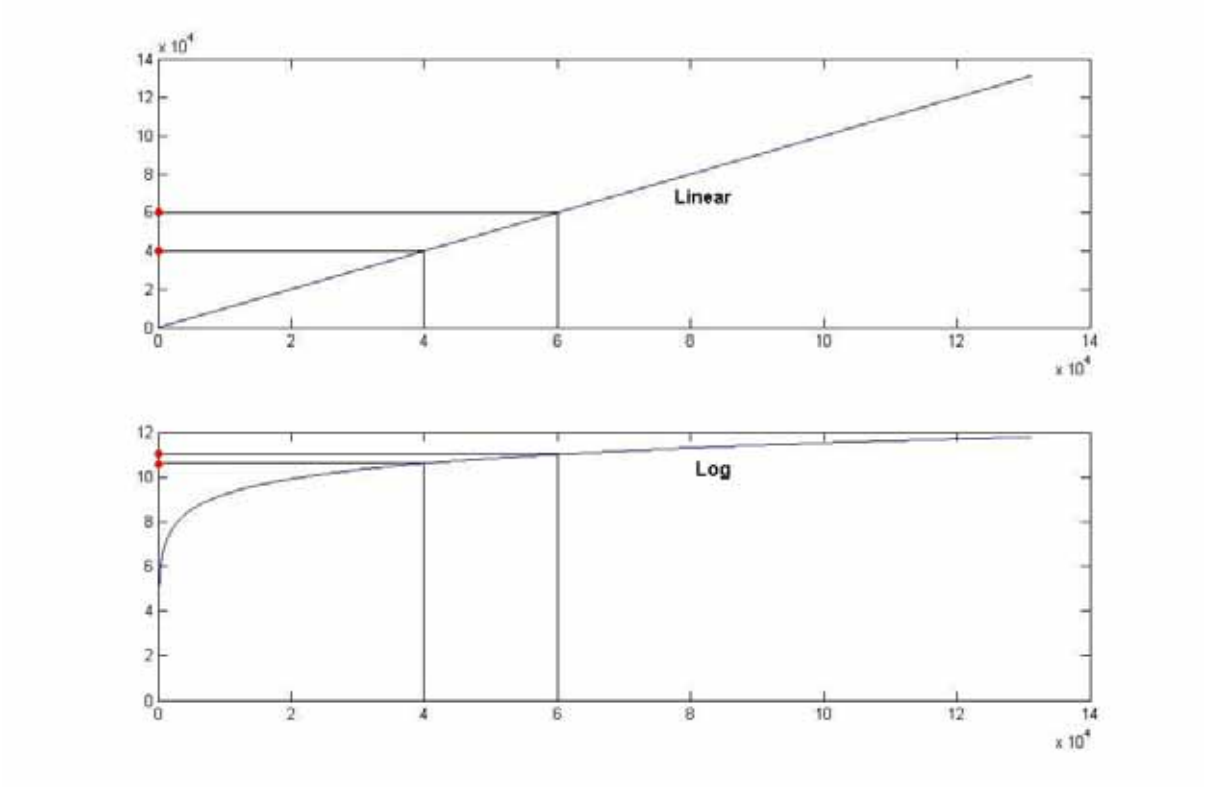

Fig. 8 The linear and log curve.

The original error =(60000-40000)/60000 = 33.33%. In the log domain, the error =(log(60000)-log(40000))/log(60000) = 3.68%

4.2 Recognition rate

We compare the recognition rate of this system with floating-point version. The training and testing datasets are displayed in Table 7. The recognition result is listed in Table 8.

Table 7: Experimental dataset

	Training data	Testing data
Content	TCC300	Tang poem
Speaker	150 males and 150 females	9 males and 1 female
Sampling rate	16 kHz	16 kHz
Bits per sample	16 bits	16 bits
Total	8917 files = 26.2 hours	3211 files = 4.5 hours
Macro	Bi-phone	
Pruning beam-width		Fixed point: 76800 Floating point: 300

Table 8: Comparison of recognition rate

	Fixed point	Floating point
Correct	2987	3082
Wrong	224	129
Recognition rate	93.02%	95.98%

4.3 Speed

This experiment is separated into two parts, feature extraction and Viterbi decoding. We use ten data of Tang poems described in Table 7 to test the speed and compare this system with floating-point version. The result is shown in Table 9.

Table 9: Speed test

	Feature extraction			Recognizer decoding		
	Fixed-point (sec.)	Floating-point (sec.)	Speedup factor	Fixed-point (sec.)	Floating-point (sec.)	Speedup factor
1	2.21	13.66	6.18	5.76	23.14	4.02
2	2.05	13.66	6.66	8.89	34.72	3.91
3	2.05	13.66	6.66	6.91	26.66	3.86
4	2.02	13.66	6.76	6.59	25.79	3.91
5	2.05	13.69	6.68	8.51	33.28	3.91
6	2.05	13.69	6.68	7.26	29.70	4.09
7	2.02	13.66	6.76	5.50	23.26	4.23
8	2.02	13.66	6.76	8.70	37.92	4.36
9	2.05	13.66	6.66	5.86	26.11	4.46
10	2.05	13.69	6.68	6.01	26.43	4.39
Total	20.57	136.69	6.65	69.99	287.01	4.1

We could see that this system totally speed up about 4.68 times comparing with floating- point version.

4.4 Parameter estimation

In Sec. 3.2.3, we mention that pruning beam-width would influence the recognition rate and speed. Thus, we try to estimate the parameter and display its affections in Table 10 and Fig. 9. The training and testing dataset are same as listed in Table 7. The scaling factor is 896 while adjusting the pruning beam-width. The recognizer decoding time is the average time of decoding 10 testing data.

Table 10: The influence of pruning beam-width

Pruning beam-width	Recognition rate	Recognizer decoding time
12800	36.94%	3.11 sec.
25600	71.44%	4.61 sec.
51200	90.38%	6.19 sec.
76800	93.02%	6.99 sec.
102400	93.58%	7.40 sec.

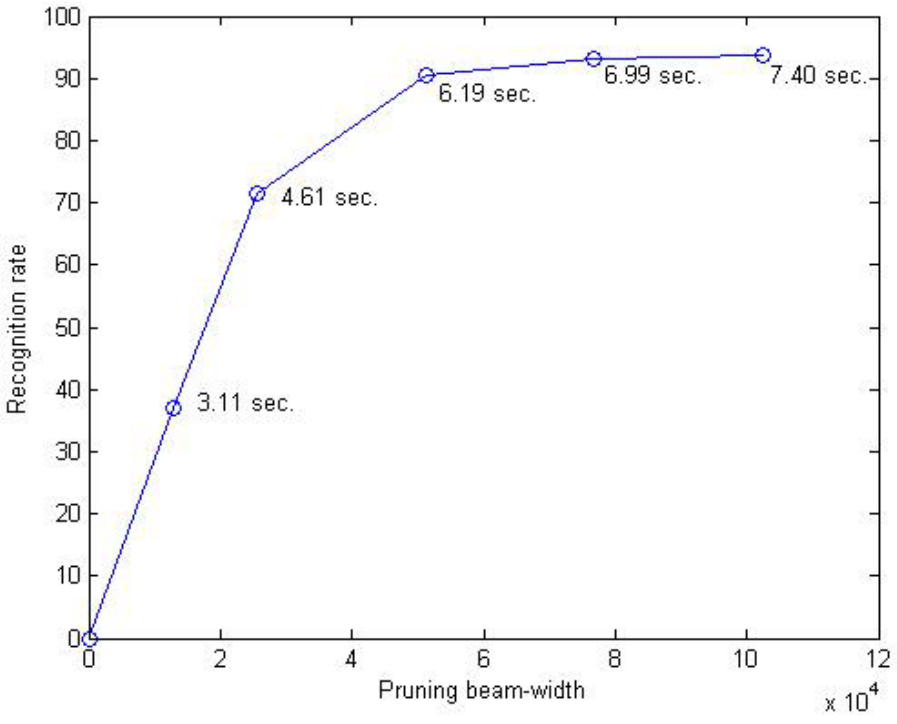

Fig. 9 The influence of pruning beam-width