CHAPTER 5 RESULT AND EVALUATION

To evaluate the performance of a low complextity tone recognition for automatic tonal speech recognizer (ATSR), we evaluate our architecture by three categories:

1. Recognition accuracy
2. Calculation complexity
3. Processing time

We evaluate tone recognition accuracy by comparing autocorrelation method with proposed vowel-magnitude different function (V_{MDF}). We also evaluate recognition accuracy of automatic tonal speech recognizer (ATSR). We compare recognition accuracy between ATSR with tone recognition and ATSR without tone recognition. Processing time of tone recognition is tested with autocorrelation, series V_{MDF}, and parallel V_{MDF}. Finally, power consumption is evaluated by power consumption between autocorrelation, series V_{MDF}, and parallel V_{MDF}.

The evaluation is set by using the Thai voice command corpus database from the National Electronics and Computer Technology Center (NECTEC), Thailand [62].

5.1 Recognition Accuracy

We test the tone recognition with 110 Thai words comprising 187 syllables selected from voice activation for GPS systems and phone dialing options. The 110 test words and its tone results, and tables of international phonetic alphabets are shown in Appendix B. Each word is sampled at 11.025 kHz and 16-bit quantization. The frame length and frame shift are 23.2 ms and 11.6 ms, respectively. The consonant and vowel phonemes of the test words are illustrated by phonemic transcription [70].
In addition, we test the recognition accuracy of an automatic tonal speech recognizer with the proposed tone modeling and without the tone modeling by the word hidden Markov's models (HMMs). The models are evaluated by 45 speech data comprising 35 words having the same phoneme but different tones and 10 words selected from voice activation for GPS systems and phone dialing options. The models are trained on a speech corpus of 2,250 words collected from 5 native Thai speakers.

The word models are set from 32-state HMMs. The feature vectors consist of 12 MFCCs, 12 delta MFCCs, 12 delta-delta MFCCs, delta log energy, and delta-delta log energy. The total processing time and the power consumption are tested by 12 mW at 50 MHz.

The experimental results present the percentage improvement in frame reduction, recognition accuracy, and power consumption. The percentage improvement is calculated from

$$Percentage\ Improvement = \frac{V1 - V2}{V1} \times 100\%$$

where $V1$ is the value given from the conventional method, and $V2$ is the value from the proposed method.

The recognition accuracy is defined by:

$$Accuracy = \frac{number\ of\ correctly\ classified\ words}{total\ classified\ words} \times 100\%$$

To evaluate the tone recognition accuracy of autocorrelation (*AC*) and the proposed V_{MDF}, the vowel signals are used as the input of the system. We evaluate tone recognition accuracy with 110 test words. All F_0 trajectories are presented in Appendix C.

The recognition accuracy in percentage is shown in Figure 5.1 – Figure 5.6. Table 5.1 compares recognition accuracy between autocorrelation (*AC*) and the proposed V_{MDF}. The result indicates 4.4% recognition accuracy improvement. *AC* generally uses the whole input speech as an input signal. Since we use vowel signal as an input of tone recognition, the *AC* accuracy reduces.

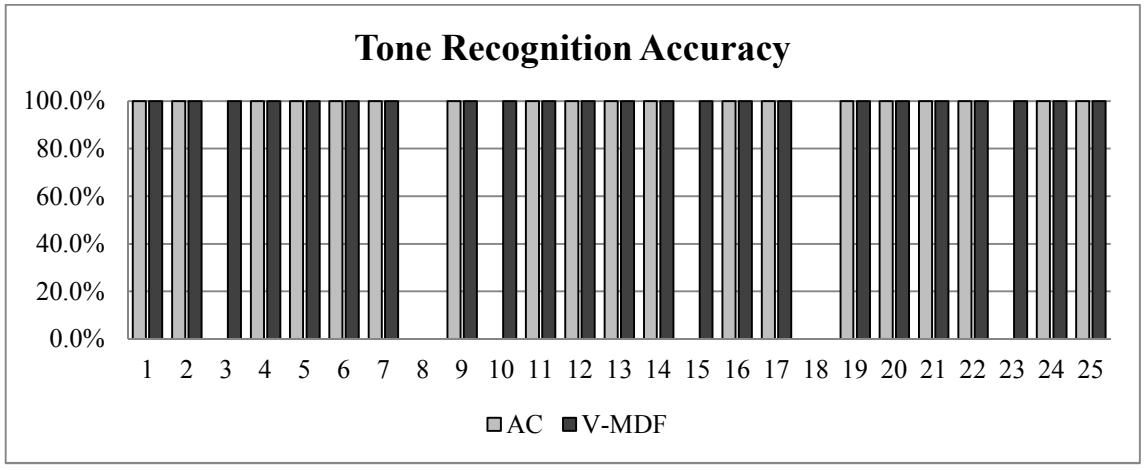

Figure 5.1 Comparison of tone recognition accuracy of the word 1-25

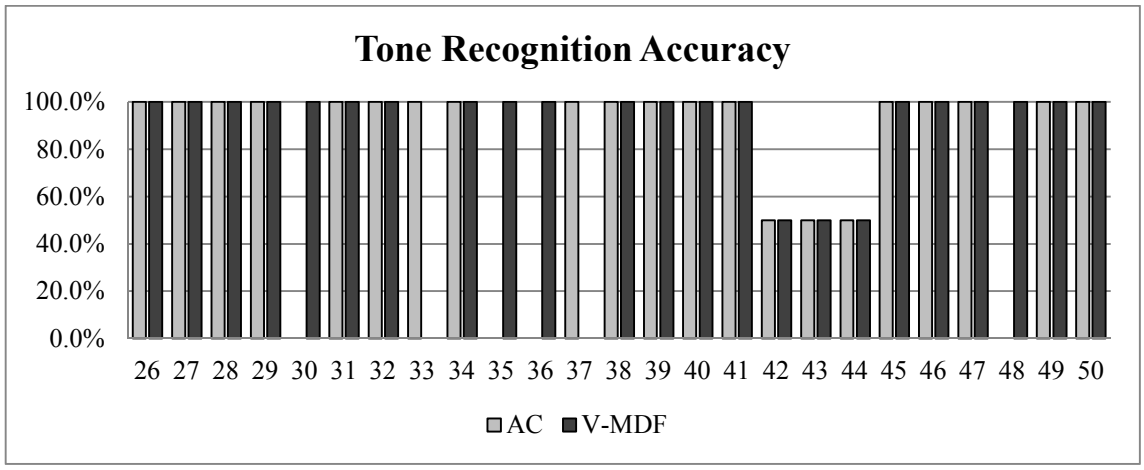

Figure 5.2 Comparison of tone recognition accuracy of the word 26-50

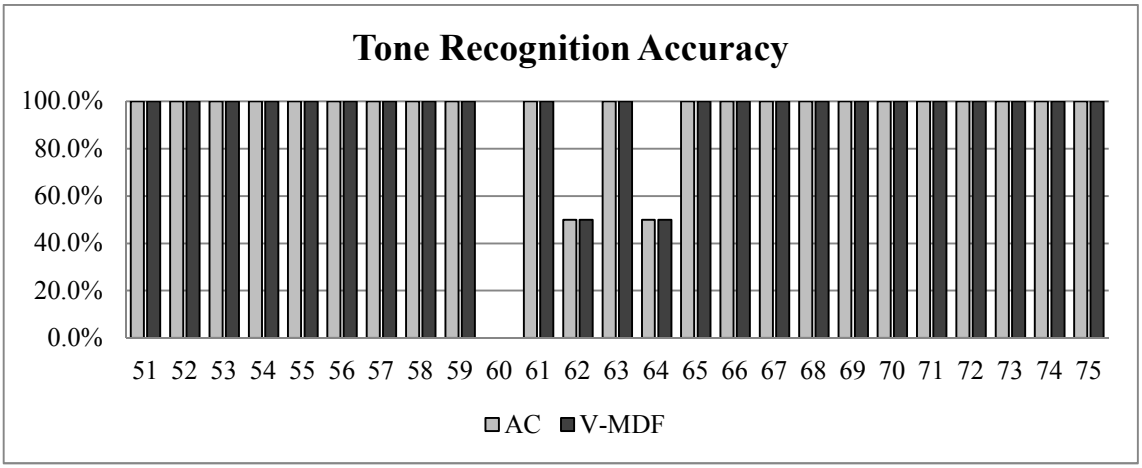

Figure 5.3 Comparison of tone recognition accuracy of the word 51-75

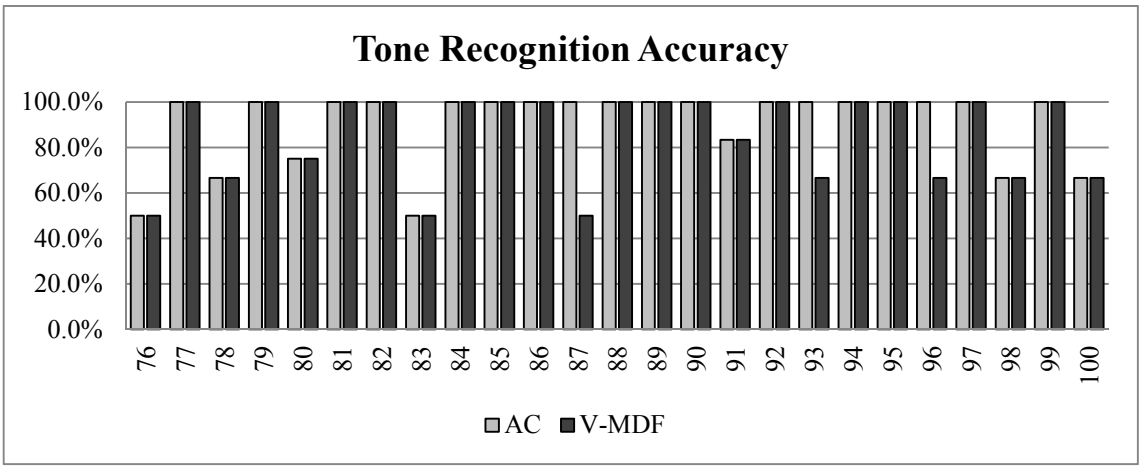

Figure 5.4 Comparison of tone recognition accuracy of the word 76-100

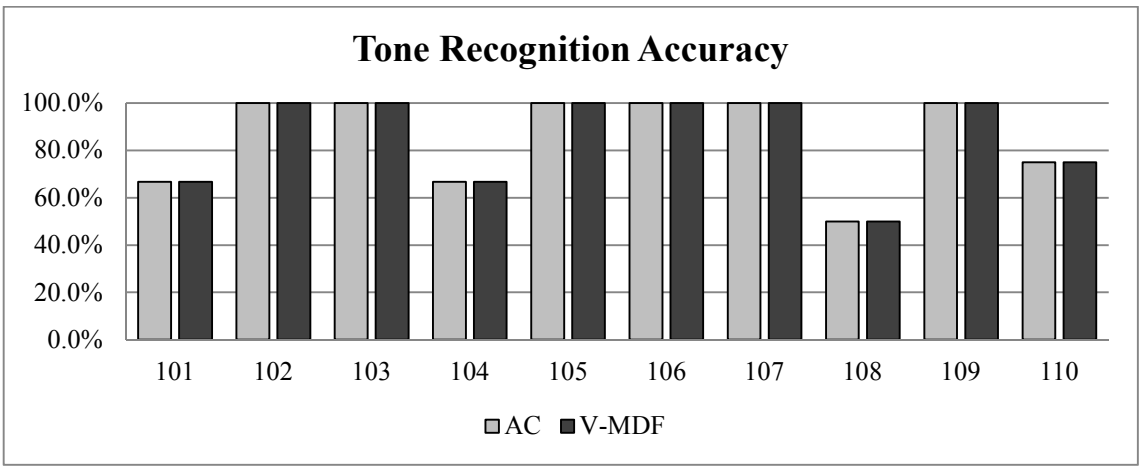

Figure 5.5 Comparison of tone recognition accuracy of the word 101-110

Table 5.1 Percentage improvement of tone recognition accuracy

Tone Recognition Method	Recognition Accuracy
Autocorrelation	84.2%
Proposed V_{MDF}	88.6%
Percentage Improvement	4.4%

Figure 5.6 and Figure 5.7 show speech recognition accuracy results comparing between automatic tonal speech recognizer (ATSR) with tone recognition and without tone recognition Table 5.2 shows the percentage improvement of recognition accuracy between ATSR with tone recognition and without tone recognition. The results show 26.3% recognition accuracy improvement. The recognition accuracy of ATSR with tone recognition is improved especially in the case of words having the same phoneme but different tones.

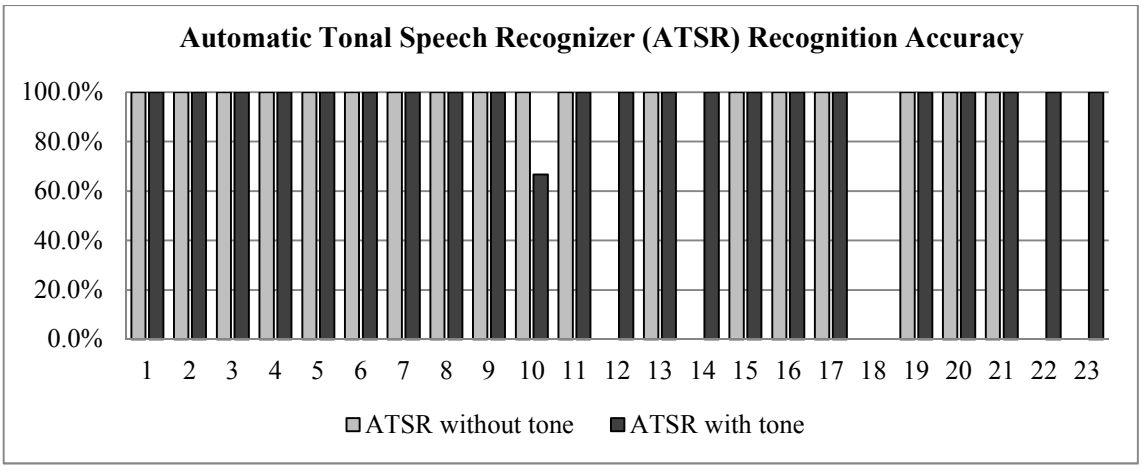

Figure 5.6 Comparison of ATSR recognition accuracy of the word 1-23

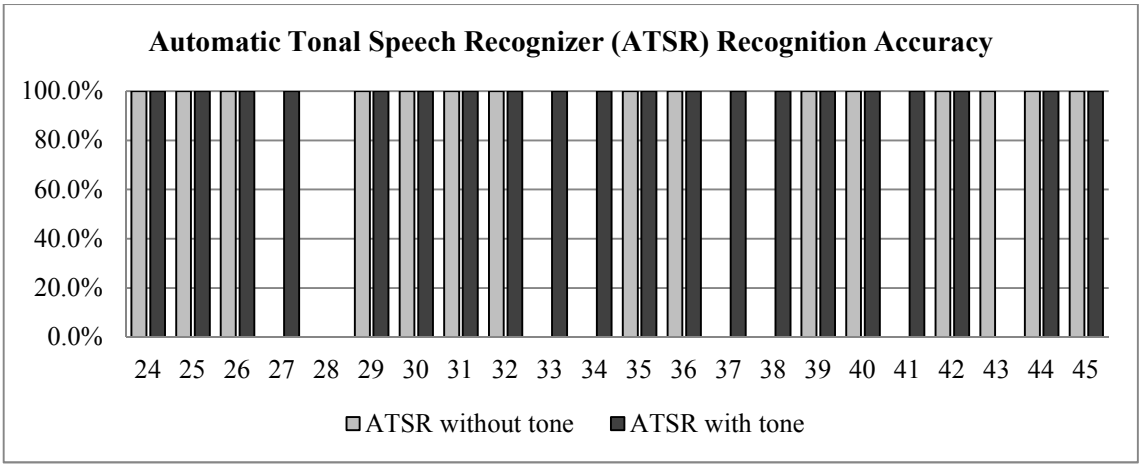

Figure 5.7 Comparison of ATSR recognition accuracy of the word 24-45

Table 5.2 Percentage improvement of speech recognition accuracy

Tone Recognition Method	Recognition Accuracy
ATSR without tone recognition	73.3%
ATSR with tone recognition	92.6%
Percentage Improvement	26.3%

5.2 Calculation complexity

Table 5.3 compares the number of frames between the autocorrelation method and the proposed vowel-MDF (V_{MDF}) method. Autocorrelation finds F_0 features by using the entire input syllable. The V_{MDF} calculates F_0 features by using only vowel signals. This causes the V_{MDF} method to decrease the number of frames by 44.6%, compared with using the entire input syllable.

Since our architecture reduces the number input frame by using V_{MDF} method, we indicate the number of frame of the proposed method compared with using the entire syllable in Table 5.3.

Table 5.3 Number of frames

No.	Test Word	Thai Alphabet	Number of Frames		%Reduction
			Entire Syllable	V_{MDF}	
1	*māa*	มา	69	32	53.6%
2	*màa*	หม่า	117	47	59.8%
3	*mâa*	ม่า	74	27	63.5%
4	*ma'a*	ม๊า	85	39	54.1%
5	*mă̄a*	หมา	58	53	8.6%
6	*mīi*	มี	84	46	45.2%
7	*mìi*	หมี่	84	56	33.3%
8	*mîi*	มี่	63	34	46.0%
9	*mi'i*	มี๊	83	40	51.8%
10	*mĭ̄i*	หมี	74	42	43.2%
11	*māy*	ไม	74	50	32.4%

Table 5.3 Number of frames (cont.)

No.	Test Word	Thai Alphabet	Number of Frames		%Reduction
			Entire Syllable	V_{MDF}	
12	*m`ay*	ไหม่	59	50	15.3%
13	*mây*	ไหม้	51	27	47.1%
14	*ma'y*	ม้าย	74	33	55.4%
15	*măy*	ไหม	68	50	26.5%
16	*khāy*	ใคร	67	16	76.1%
17	*kh`ay*	ไข่	89	39	56.2%
18	*khây*	ไข้	53	11	79.2%
19	*kha'y*	ไค้	83	17	79.5%
20	*khăy*	ไข	82	25	69.5%
21	*kāy*	ไกล	68	37	45.6%
22	*k`ay*	ไก่	70	47	32.9%
23	*kây*	ไก้	54	41	24.1%
24	*ka'y*	ไก้	57	46	19.3%
25	*kăy*	ไก๋	69	46	33.3%
26	*khāa*	คา	80	21	73.8%
27	*kh`aa*	ข่า	85	20	76.5%
28	*khâa*	ฆ่า,ค่า	78	42	46.2%
29	*kha'a*	ค้า	83	48	42.2%
30	*khăa*	ขา	89	37	58.4%
31	*nāa*	นา	72	50	30.6%
32	*n`aa*	หน่า	78	37	52.6%
33	*nâa*	หน้า	60	36	40.0%
34	*na'a*	น้า	71	44	38.0%
35	*năa*	หนา	67	41	38.8%
36	*chûu*	ชื่อ	72	38	47.2%
37	*m`uu*	หมู่	76	46	39.5%

Table 5.3 Number of frames (cont.)

No.	Test Word	Thai Alphabet	Number of Frames		%Reduction
			Entire Syllable	V_{MDF}	
38	*sɔ̄ɔy*	ซอย	88	44	50.0%
39	*kh`eet*	เขต	65	30	53.8%
40	*lām/phōoŋ*	ลำโพง	92	73	20.7%
41	*hɛ̄ɛn / frīi*	แฮนด์ฟรี	90	62	31.1%
42	*sĭaŋ / rîak*	เสียงเรียก	75	57	24.0%
43	*pɔ̀ɔt/sĭaŋ*	เปิดเสียง	78	60	23.1%
44	*pīt/sĭaŋ*	ปิดเสียง	82	49	40.2%
45	*tām / bōn*	ตำบล	67	60	10.4%
46	*ɔɔām / phɤ̄ɤ*	อำเภอ	83	48	42.2%
47	*lōŋ*	ลง	59	44	25.4%
48	*khwăa*	ขวา	83	46	44.6%
49	*th`at / pāy*	ถัดไป	55	32	41.8%
50	*hmɨ̆an*	เหมือน	78	35	55.1%
51	*thɛ̄ɛw*	แถว	24	6	75.0%
52	*lɨ̂ak*	เลือก	26	19	26.9%
53	*khâam*	ข้าม	70	44	37.1%
54	*t`at*	ตัด	77	47	39.0%
55	*pɛ̀*	แปะ	41	26	36.6%
56	*ɔɔāw/ɔɔɔ̄k*	เอาออก	54	37	31.5%
57	*t`ok / lō ŋ*	ตกลง	26	13	50.0%
58	*sŭu ŋ*	สูง	79	27	65.8%
59	*t`am*	ต่ำ	76	56	35.7%
60	*lôt*	ลด	44	28	36.4%
61	*chɔ̂ɔŋ*	ช่อง	60	34	43.3%
62	*tɛ̄ɛn/tîi*	แทนที่	41	14	65.9%
63	*māak / s`ut*	มากสุด	74	50	32.4%

Table 5.3 Number of frames (cont.)

No	Test Word	Thai Alphabet	Number of Frames		%Reduction
			Entire Syllable	V_{MDF}	
64	*nɔ́ɔy/sùt*	น้อยสุด	98	74	24.5%
65	*rā/dāp/sǐaŋ*	ระดับเสียง	88	48	45.5%
66	*rûup/bèɛp*	รูปแบบ	83	36	56.6%
67	*rāay/kāan*	รายการ	88	55	37.5%
68	*sāt/s`uan*	สัดส่วน	88	62	29.5%
69	*hy`ut/phāp*	หยุดภาพ	59	35	40.7%
70	*tām/sa'm*	ทำซ้ำ	102	77	24.5%
71	*phɛ̄ɛn/tîi*	แผนที่	55	32	41.8%
72	*cīi/phīi/ɔɔ ēt*	จีพีเอส	78	35	55.1%
73	*plāay/thāaŋ*	ปลายทาง	24	6	75.0%
74	*nām/thāa ŋ*	นำทาง	98	74	24.5%
75	*pāy/tîi*	ไปที่	88	48	45.5%
76	*sêen/thāaŋ*	เส้นทาง	88	44	50.0%
77	*rōoŋ/rɛ̄ɛm*	โรงแรม	85	56	34.1%
78	*cūt/phâk/phɔ̀ɔn*	จุดพักผ่อน	85	66	22.4%
79	*phāt/tāa/khāan*	ภัตตาคาร	98	40	59.2%
80	*khrûaŋ/pràp/āa/kàat*	เครื่องปรับอากาศ	56	16	71.4%
81	*c`ut/t`at*	จุดตัด	92	57	38.0%
82	*cùt/chōm/wīw*	จุดชมวิว	81	30	63.0%
83	*rót/tìt*	รถติด	29	9	69.0%
84	*thōo/rā/sàp*	โทรศัพท์	71	24	66.2%
85	*thōo/dùan*	โทรด่วน	40	21	47.5%
86	*rāay/kāan/thōo*	รายการโทร	51	23	54.9%
87	*wāaŋ/săay*	วางสาย	77	33	57.1%
88	*thōo*	โทร	99	63	36.4%

Table 5.3 Number of frames (cont.)

No.	Test Word	Thai Alphabet	Number of Frames		%Reduction
			Entire Syllable	V_{MDF}	
89	*srâa ŋ*	สร้าง	90	71	21.1%
90	*pīi*	ปี	81	43	46.9%
91	*prā/chūm/tāaŋ/thōo/ rā/sàp*	ประชุมทางโทรศัพท์	37	24	35.1%
92	*rāp/sāay*	รับสาย	50	8	84.0%
93	*mây/rāp/sāay*	ไม่รับสาย	99	48	51.5%
94	*wēe/lāa/thōo*	เวลาโทร	72	51	29.2%
95	*pāan / klāa ŋ*	ปานกลาง	60	31	48.3%
96	*cùt /rɔ̄ɔy/tɔ̀ɔ*	จุดรอยต่อ	107	62	42.1%
97	*nāy/mīaŋ*	ในเมือง	75	58	22.7%
98	*khɔ̂ɔ/mūun/kāan/cā/ rāa/cɔ̄ɔn*	ข้อมูลการจราจร	90	42	53.3%
99	*dùan*	ด่วน	98	68	30.6%
100	*tū / rā / kīt*	ธุรกิจ	180	104	42.2%
101	*thĭw/sāay/rɔ̄ɔ*	ถือสายรอ	64	15	76.6%
102	*sùan/tūaw*	ส่วนตัว	62	32	48.4%
103	*dīan*	เดือน	77	54	29.9%
104	*pā/tì/thīn*	ปฏิทิน	65	32	50.8%
103	*dīan*	เดือน	67	14	79.1%
104	*pā/tì/thīn*	ปฏิทิน	59	22	62.7%
105	*tīan/khāam/cām*	เตือนความจำ	91	30	67.0%
100	*tū / rā / kīt*	ธุรกิจ	77	57	26.0%
101	*thĭw/sāay/rɔ̄ɔ*	ถือสายรอ	62	18	71.0%
102	*sùan/tūaw*	ส่วนตัว	61	31	49.2%
103	*dīan*	เดือน	63	45	28.6%
104	*pā/tì/thīn*	ปฏิทิน	57	25	56.1%
105	*tīan/khāam/cām*	เตือนความจำ	62	32	48.4%

Table 5.3 Number of frames (cont.)

No.	Test Word	Thai Alphabet	Number of Frames		%Reduction
			Entire Syllable	V_{MDF}	
106	*ŋāan*	งาน	77	54	29.9%
107	*wān/kɔ̀ɔt*	วันเกิด	65	32	50.8%
108	*sāp/dāa*	สัปดาห์	67	14	79.1%
109	*gēem*	เกมส์	59	22	62.7%
110	*sā/mùt/thōo/sàp*	สมุดโทรศัพท์	91	30	67.0%
Average					**44.6%**

5.3 Processing Time

We compare total processing time of the proposed method in data allocation (DA) of PE1 and processing time in PE1 to PE3. Figure 5.8 indicates the reduction in processing time between using the general data arrangement and the proposed data allocation (DA). The proposed data allocation reduces the number of clock cycles by 50.2%. This decreases the amount of processing time when the tone recognition requires input signals.

The proposed tone recognition reduces the number of frames by using V_{MDF}. In addition, the tone recognition executes the whole system by using pipeline and parallel architecture. Therefore, this recognition system decreases total processing time and consumes low power. Figure 5.9 shows the experimental result of total processing time Note that 183 is the maximum frame of this paper. According to the results, the proposed tone recognition is suitable for portable electronic equipment.

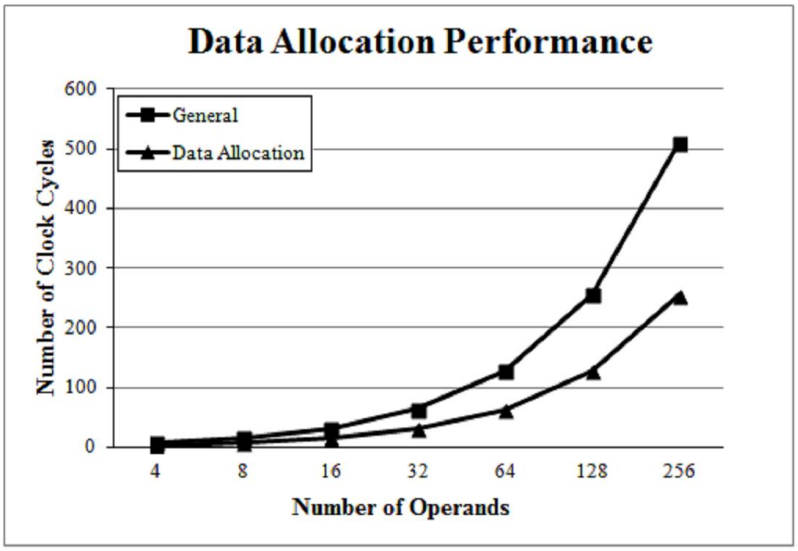

Figure 5.8 Data allocation performance

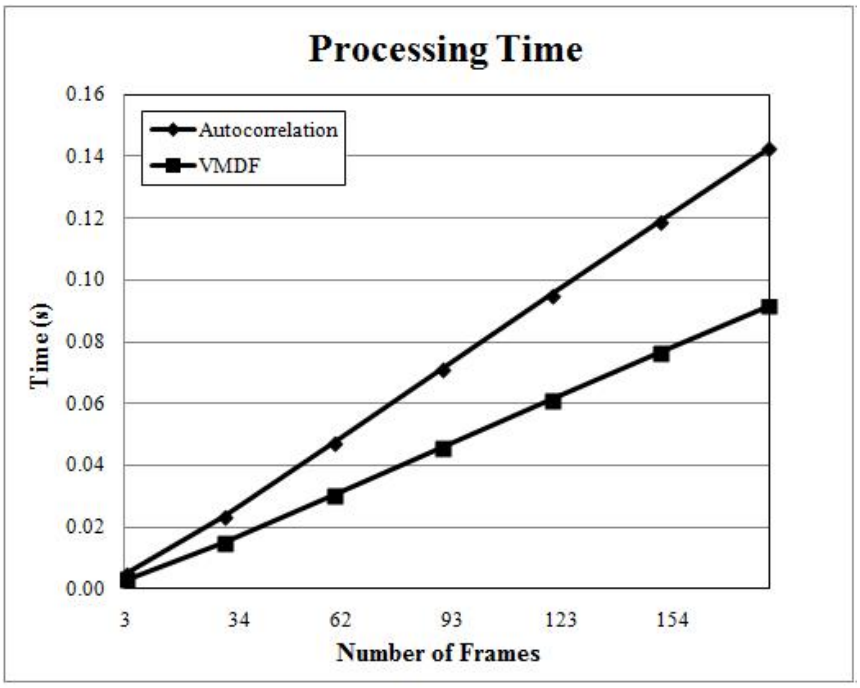

Figure 5.9 Total processing time