บทที่ 3

การออกแบบระบบ

การรู้จำตัวโน้ต คือ การสามารถระบุตัวโน้ตที่ได้ยินจากเสียงเครื่องดนตรีได้ว่าเป็นโน้ตตัวอะไร ซึ่งเป็นงานที่ไม่ยากนักสำหรับนักดนตรีที่ได้รับการฝึกฝนมาจนชำนาญ แต่จะเป็นงานที่ค่อนข้างยากสำหรับระบบคอมพิวเตอร์ในการรู้จำและระบุโน้ตดนตรี สำหรับโน้ตดนตรีที่ใช้ในการรู้จำสามารถแบ่งออกได้เป็นสองกลุ่มคือ กลุ่มของโน้ตดนตรีที่เล่นเพียงโน้ตเดียวในเวลาใดเวลาหนึ่งและกลุ่มของโน้ตดนตรีที่เล่นตั้งแต่สองตัวโน้ตขึ้นไปพร้อมกัน

เนื้อหาของบทนี้ได้นำเสนอการออกแบบและขั้นตอนวิธีการรู้จำตัวโน้ตจากเสียงเครื่องดนตรีไทยด้วยวิธีการใช้ตัวกรองความถี่ โดยมุ่งเน้นไปที่เครื่องดนตรีไทยคือระนาดเอก ซึ่งเป็นเครื่องดนตรีที่เล่นโน้ตได้ทั้งแบบโมโนโฟนิกและแบบโพลีโฟนิก โดยการอธิบายถึงรูปแบบโดยรวมของการออกแบบระบบจะได้อธิบายในหัวข้อ 3.1 และได้อธิบายขั้นตอนวิธีการรู้จำตัวโน้ตอย่างละเอียดในส่วนต่าง ๆ คือ การประมวลผลเบื้องต้น (Preprocessing) ในหัวข้อ 3.2 อธิบายวิธีการออกแบบและการประยุกต์ใช้งานตัวกรองผ่านแถบความถี่ (Bandpass Filtering) ในหัวข้อ 3.3 เนื่องจากโน้ตเสียงระนาดแต่ละตัวมีความเด่นชัดไม่เท่ากันจึงได้มีการเสนอวิธีการชดเชยความเด่นชัดของตัวโน้ตในหัวข้อ 3.4 สำหรับวิธีการตัดสินใจในการระบุตัวโน้ต (Notes Selection) และการทำโพสโปรเซสซิ่ง (Post-processing) เพื่อให้ได้ผลลัพธ์ของการระบุตัวโน้ตจะได้กล่าวในหัวข้อ 3.5 และหัวข้อ 3.6 และสำหรับวิธีการแบ่งช่วงเสียงโน้ตต่อเนื่องที่ใช้ในระบบที่นำเสนอจะได้อธิบายในหัวข้อ 3.7 และสุดท้ายคือหัวข้อ 3.8 เป็นการสรุปเนื้อหาของบทนี้

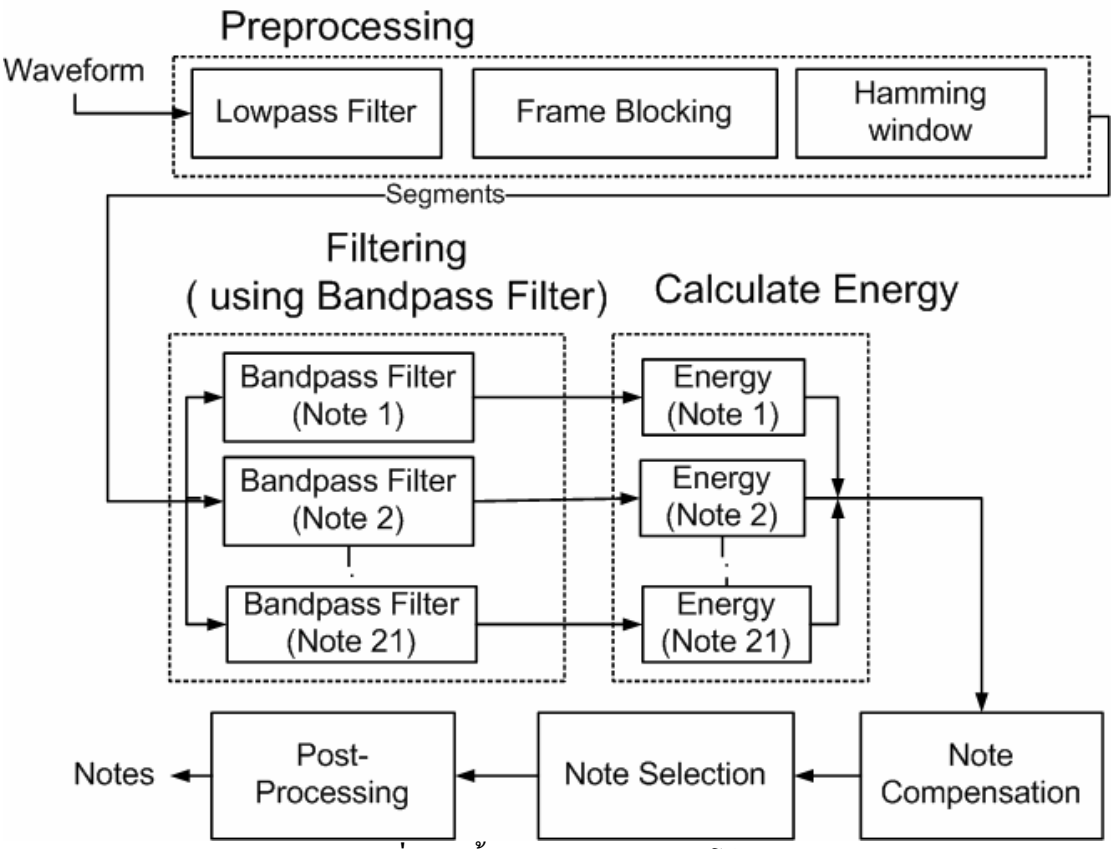
3.1 ภาพรวมของการออกแบบระบบ

เนื่องด้วยข้อจำกัดจากวิธีการเล่นระนาดเอก ทำให้กลุ่มของตัวโน้ตที่ใช้ในการรู้จำสำหรับระนาดเอกมีทั้งที่เป็นแบบ Monophonic และแบบ Polyphonic ที่มีไม่เกินสองโน้ต

ภาพรวมของระบบจะแบ่งการทำงานหลักได้เป็น 2 ส่วนด้วยกันคือ ส่วนของระบุตัวโน้ต (Note Identifier) และ ส่วนของการแบ่งช่วงเสียงโน้ตต่อเนื่อง (Note Boundary Detector) ดังที่เคยได้แสดงในรูปที่ 1.3 สำหรับผลลัพธ์ที่ได้จากการแบ่งช่วงโน้ตเสียงต่อเนื่องจะถูกนำมาป้อนเป็นข้อมูลให้กับส่วนของการระบุตัวโน้ตเพื่อนำไปใช้งานต่อไป

สำหรับขั้นตอนที่ใช้ในการระบุตัวโน้ต (Note Identifier) แบ่งได้เป็น 5 ขั้นตอน คือ การประมวลผลเบื้องต้น (Preprocessing), การกรองผ่านแถบความถี่ (Bandpass Filtering), การชดเชยค่าความเด่นชัดของตัวโน้ต (Note Compensation), กระบวนการตัดสินใจเลือกโน้ตใน

แต่ละส่วนย่อย (Note Selection) และกระบวนการทำโพสโปรเซสซิ่ง (Post-processing) ดังแสดงในรูปที่ 3.1



รูปที่ 3.1 ขั้นตอนวิธีการระบุตัวโน้ต

3.2 การประมวลผลเบื้องต้น (Preprocessing)

การประมวลผลเบื้องต้นเป็นการปรับข้อมูลของเสียงให้เหมาะสมสำหรับการประมวลผลในขั้นต่อไป สำหรับขั้นตอนการประมวลผลเบื้องต้นในงานวิจัยนี้ประกอบไปด้วยการกรองผ่านความถี่ต่ำ (Lowpass Filter) การแบ่งข้อมูลสัญญาณเสียงออกเป็นส่วนย่อย (Frame Blocking) และ การลดขอบด้วยฟังก์ชันหน้าต่าง (Window Function)

3.2.1 การกรองผ่านความถี่ต่ำ

การกรองทางความถี่เป็นการกรองความถี่ในช่วงที่ไม่ต้องการออกโดยในวิทยานิพนธ์นี้ได้อาศัยตัวกรองแบบดิจิตอล ในขั้นตอนนี้ได้ใช้ตัวกรองผ่านความถี่ต่ำเพื่อใช้ในการลดสัญญาณรบกวนที่อาจจะเกิดขึ้น เนื่องจากความถี่ของโน้ตระนาดที่มีค่าสูงสุดไม่เกิน 1,400 เฮิรตซ์ ดังนั้นจึงกำหนดช่วงของความถี่คัทออฟ (cutoff frequency) เท่ากับ 1,400 เฮิรตซ์

3.2.2 การแบ่งเป็นส่วนย่อย

การแบ่งเป็นส่วนย่อย หรือ การแบ่งเป็นเฟรม (Frame) เป็นการแบ่งสัญญาณเสียงออกเป็นช่วงเวลาสั้น ๆ ซึ่งทำให้คุณสมบัติของสัญญาณเสียงไม่มีการเปลี่ยนแปลงหรือมีการเปลี่ยนแปลงตามเวลาน้อยมาก ในการทดลองนี้ได้กำหนดความยาวของเฟรม (Frame Size) เท่ากับ 100 มิลลิวินาทีและมีการขยับเฟรม (Frame Rate) เท่ากับ 10 มิลลิวินาที

3.2.3 การลดขอบด้วยฟังก์ชันหน้าต่าง

เนื่องจากการแบ่งข้อมูลออกเป็นเฟรมย่อยจะทำให้เกิดความไม่ต่อเนื่องตรงช่วงรอยต่อหรือขอบของแต่ละเฟรม ดังนั้นจึงทำการลดขอบด้วยฟังก์ชันหน้าต่าง (Window Function) เพื่อเป็นการปรับสัญญาณให้ราบเรียบและลดผลกระทบที่เกิดจากการแบ่งเป็นเฟรมย่อย สำหรับในการทดลองนี้ได้ใช้ฟังก์ชั่นหน้าต่างแบบ Symmetric Hamming

3.3 การกรองผ่านแถบความถี่ (Bandpass Filtering)

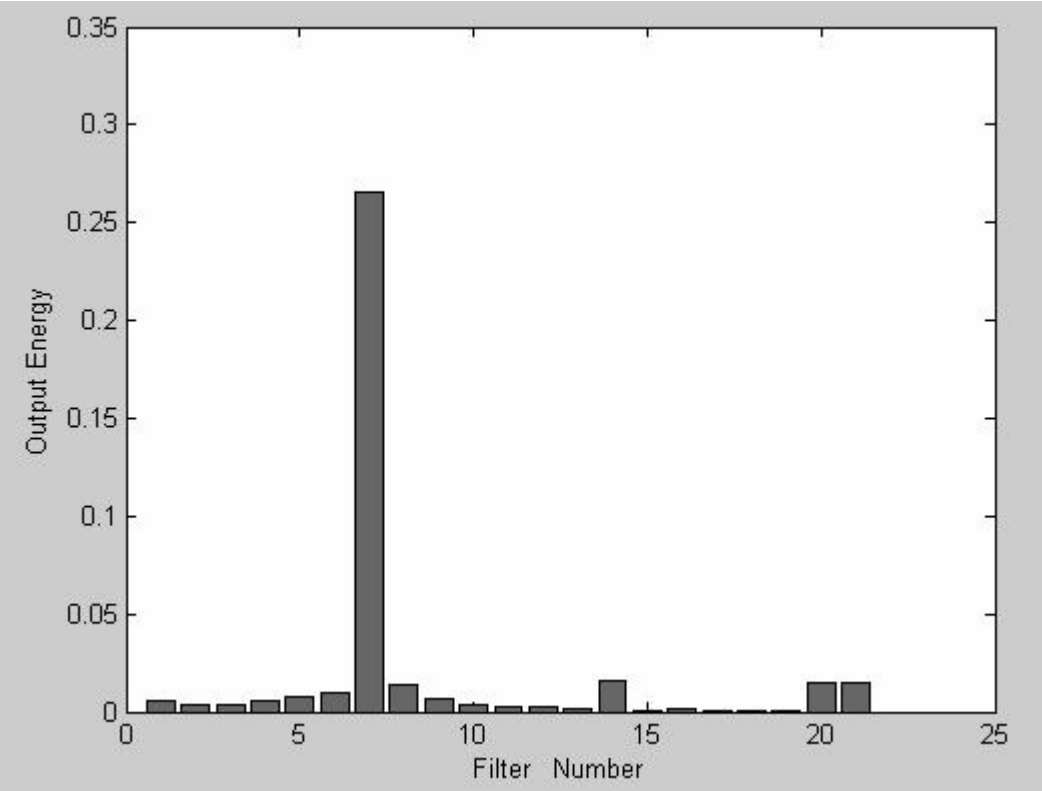
การกรองผ่านแถบความถี่จะยอมให้ข้อมูลสัญญาณความถี่ที่อยู่ในช่วงที่ต้องการผ่านและกำจัดสัญญาณความถี่ที่ไม่ต้องการออกไป ในขั้นตอนนี้ได้ออกแบบตัวกรองผ่านแถบความถี่ทั้งหมด 21 ตัวตามจำนวนของตัวโน้ตที่มีในระนาดเอก เนื่องจากความถี่ของโน้ตเสียงระนาดแต่ละตัวอาจจะมีค่าไม่เท่ากัน ดังนั้นจึงต้องทำการบันทึกเสียงระนาดและนำมาหาค่าเฉลี่ยความถี่ของโน้ตเสียงเพื่อนำมาใช้ในการออกแบบตัวกรองผ่านแถบความถี่

สำหรับคุณสมบัติพื้นฐานของตัวกรองที่ใช้มีค่าดังนี้คือ ชนิดของตัวกรองกำหนดเป็น Elliptic, Passband Attenuation = 1 dB, Stopband Attenuation = 50 dB, ลำดับของตัวกรองเท่ากับ 3 และมีค่าลำดับตัวกรองเท่ากับ 4 สำหรับโน้ตที่ 14, 20, 21 สำหรับตัวกรองผ่านแถบความถี่ที่ใช้นั้นได้กำหนดไว้คุณลักษณะดังแสดงในรูปที่ 2.11 และกำหนดค่าช่วงผ่านแถบความถี่ดังแสดงใน ตารางที่ 3-1

ตารางที่ 3-1การกำหนดค่าของตัวกรองผ่านแถบความถี่

ลำดับของตัวกรอง	ความถี่ (Hz)				
	f_{notes}	f_{pass1}	f_{pass2}	f_{stop1}	f_{stop2}
1	171	166	172	156.3	181.6
2	192	190	194	181.6	201.2
3	208	207	211	199.2	218.8
4	227	224	229	214.8	238.3
5	256	253	258	244.1	267.6
6	280	279	283	271.5	291
7	312	310	315	300.8	324.2
8	344	342	347	334	355.5
9	376	374	379	365.2	388.7
10	424	421	426	412.1	435.5
11	456	454	459	445.3	468.7
12	512	509	514	500	523.4
13	568	566	571	556.6	580.1
14	632	627	636	619.1	642.6
15	688	685	690	675.8	699.2
16	760	757	762	748	771.5
17	848	845	850	835.9	859.4
18	928	925	930	916	939.5
19	1041	1038	1043	1029	1053
20	1161	1157	1166	1150	1174
21	1265	1260	1270	1252	1277

สัญญาณเสียงในแต่ละเฟรมต้องผ่านตัวกรองผ่านแถบความถี่ทั้ง 21 ตัวตามที่ได้ออกแบบไว้ ซึ่งผลลัพธ์ที่ได้จากตัวกรองแต่ละตัว จะมีค่าไม่เท่ากันขึ้นอยู่กับความถี่ที่มีอยู่ในแต่ละเฟรมย่อย ซึ่งถ้าความถี่ของสัญญาณเสียงในเฟรมอยู่ในช่วงผ่านของตัวกรองผ่านแถบความถี่มากจะทำให้ผลลัพธ์ที่ได้มีค่าสูงตามไปด้วย แต่ถ้าหากค่าความถี่ของสัญญาณเสียงอยู่นอกช่วงผ่านของตัวกรองผ่านแถบความถี่ ค่าผลลัพธ์ที่ได้จะน้อยลง ตัวอย่างผลลัพธ์ที่ได้จากการป้อนสัญญาณเสียงของโน้ตตัวที่ 7 เข้าไปยังตัวกรองผ่านแถบความถี่ทั้ง 21 ตัวแสดงไว้ในรูปที่ 3.2



รูปที่ 3.2 ตัวอย่างผลลัพธ์ของ Output Energy ที่ได้จากการป้อนสัญญาณเสียงของโน้ตตัวที่ 7 เข้าไปยังตัวกรองผ่านแถบความถี่ทั้ง 21 ตัว

3.4 การชดเชยค่าความเด่นชัดของตัวโน้ต (Note Compensation)

เนื่องจากตัวโน้ตแต่ละตัวจะมีค่าความเด่นชัดไม่เท่ากัน ซึ่งทั้งนี้เกิดขึ้นมาได้จากหลายปัจจัย เช่น ระยะห่างที่แตกต่างกันระหว่างลูกระนาดกับไมโครโฟนที่ใช้ในการบันทึกเสียง, ความดังของโน้ตแต่ละตัว เป็นต้น ดังนั้นเพื่อให้การอ้างอิงความเด่นชัดของโน้ตแต่ละอยู่ในระดับมาตรฐานเดียวกัน จึงต้องมีการชดเชยค่าความเด่นชัดของโน้ตแต่ละตัว

สำหรับในการออกแบบระบบรู้จำนี้ได้เสนอวิธีการชดเชยค่าความเด่นชัดของโน้ตแต่ละตัว ซึ่งแสดงอยู่ในรูปของค่าสัมประสิทธิ์ชดเชย (μ) โดยค่าสัมประสิทธิ์ชดเชยของโน้ตแต่ละตัวสามารถหาค่าได้จากส่วนกลับของค่าความเด่นชัดเฉลี่ยของโน้ตตัวนั้น ($\overline{E}$) และค่าความเด่นชัดเฉลี่ยของตัวโน้ตแต่ละตัวสามารถหาค่าได้จาก ค่าเฉลี่ยความเด่นชัดของโน้ตตัวนั้นจากไฟล์เสียงแบบ Monophonic ทั้งหมดที่ใช้ในการทดลอง ดังแสดงในสมการที่ (3-1), (3-2) และ (3-3) ตามลำดับ

$$\hat{E}_k = \mu_k E_k \quad ;1 \le k \le 21 \tag{3-1}$$

$$\mu_k = \frac{1}{\overline{E}_k} \tag{3-2}$$

$$\overline{E}_k = \frac{\left(\sum_{m=1}^{M} \sum_{n=1}^{N} E_k(m,n) \right)}{M \times N} \tag{3-3}$$

โดยที่

E_k คือ ค่าความเด่นชัดของตัวโน้ตที่ k ก่อนการชดเชย

$\hat{E}_k$ คือ ค่าความเด่นชัดของตัวโน้ตที่ k หลังการชดเชย

μ_k คือ ค่าสัมประสิทธิ์ชดเชยของตัวโน้ตที่ k

$\overline{E}_k$ คือ ค่าความเด่นชัดเฉลี่ยของตัวโน้ตที่ k

M คือ จำนวนไฟล์เสียงทั้งหมดของตัวโน้ตที่ k

N คือ จำนวนเฟรมเสียงทั้งหมดในไฟล์เสียงลำดับที่ m

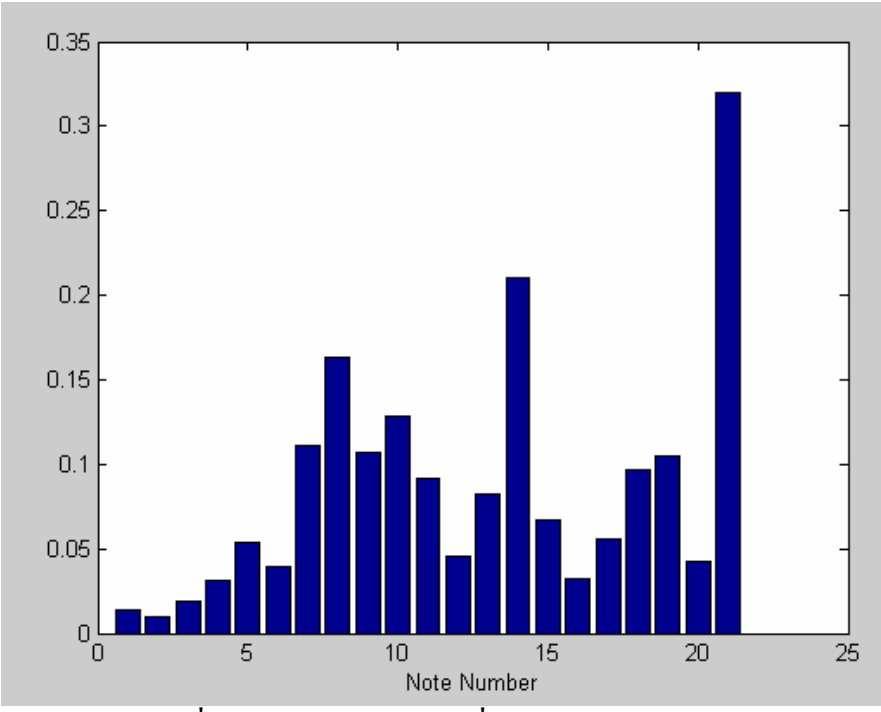
สำหรับผลลัพธ์ของค่าความเด่นชัดเฉลี่ยของแต่ละตัวโน้ตและค่าสัมประสิทธิ์ชดเชยของโน้ตแต่ละตัว ที่ใช้ได้ใช้งานในระบบรู้จำที่ได้นำเสนอนี้มีค่าดังตารางที่ 3-2 และแสดงผลเปรียบเทียบได้ดังรูปที่ 3.3 และรูปที่ 3.4

ตารางที่ 3-2 ค่าความเด่นชัดเฉลี่ยและค่าสัมประสิทธิ์ชดเชยของแต่ละตัวโน้ต

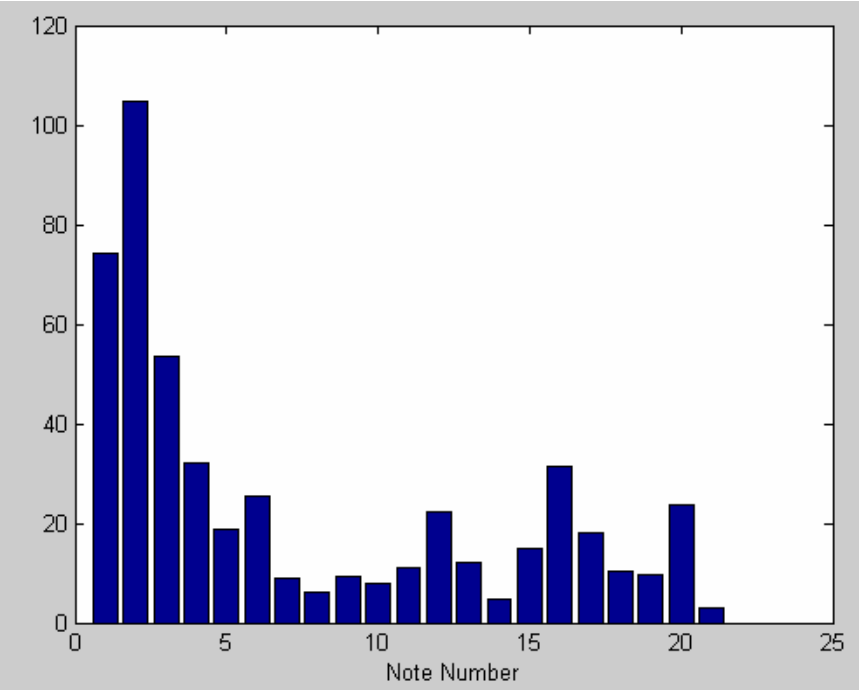
ตัวโน้ตที่	ค่าความเด่นชัดเฉลี่ย	ค่าสัมประสิทธิ์ชดเชย
1	0.0135	74.1797
2	0.0095	104.7386
3	0.0187	53.4791
4	0.0312	32.0096
5	0.0535	18.6936
6	0.0396	25.2746
7	0.1114	8.9787
8	0.1633	6.1231
9	0.1073	9.3212
10	0.1281	7.8088
11	0.0918	10.8905
12	0.0452	22.1270

ตารางที่ 3-2 (ต่อ) ค่าความเด่นชัดเฉลี่ยและค่าสัมประสิทธิ์ชดเชยของแต่ละตัวโน้ต

ตัวโน้ตที่	ค่าความเด่นชัดเฉลี่ย	ค่าสัมประสิทธิ์ชดเชย
13	0.0821	12.1814
14	0.2104	4.7537
15	0.0673	14.8527
16	0.0319	31.3468
17	0.0556	17.9720
18	0.0965	10.3677
19	0.1053	9.4992
20	0.0425	23.5179
21	0.3194	3.1305



รูปที่ 3.3 ค่าความเด่นชัดเฉลี่ยของตัวโน้ตแต่ละตัว



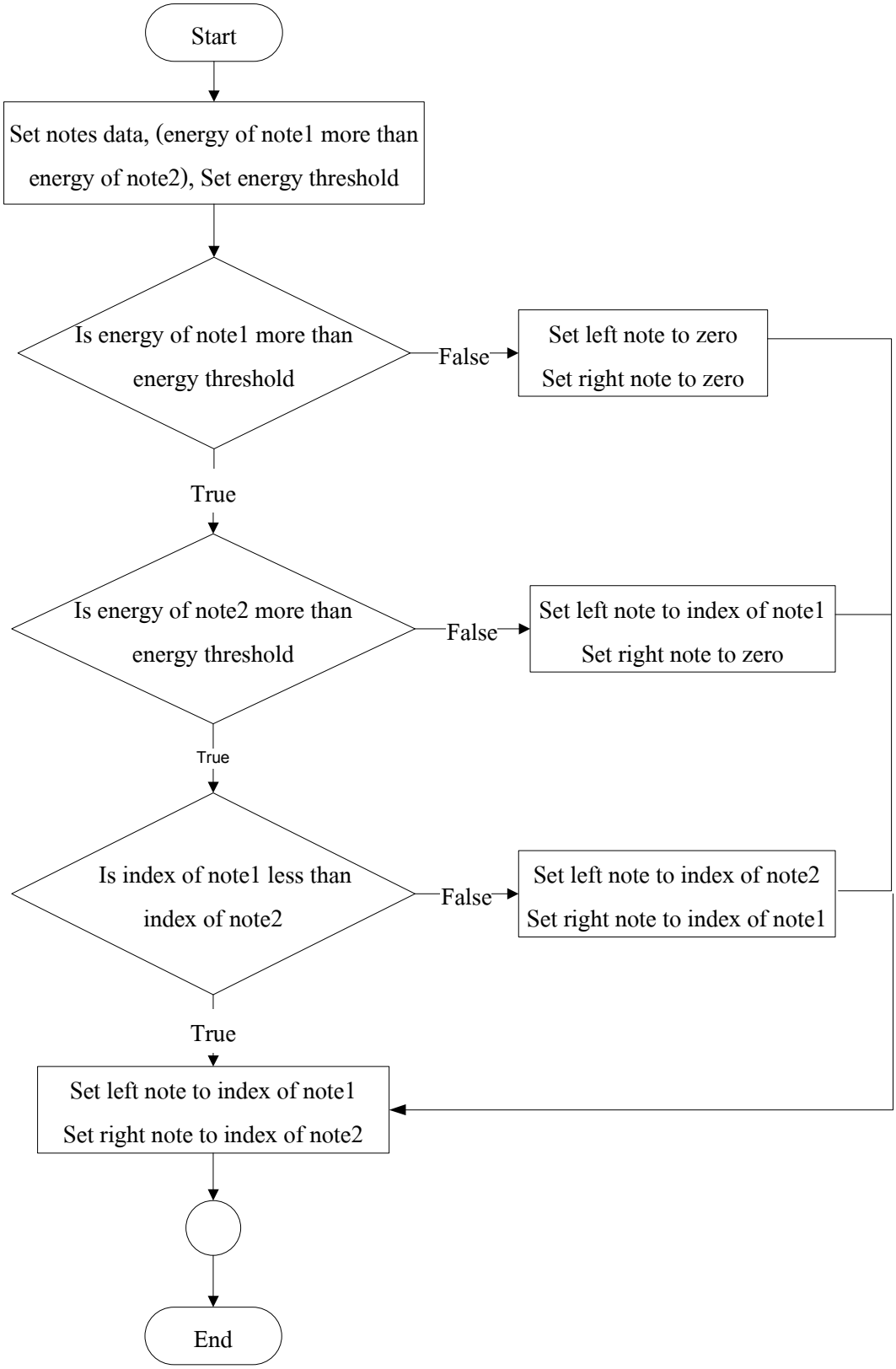
รูปที่ 3.4 ค่าสัมประสิทธิ์ชดเชยของตัวโน้ตแต่ละตัว

3.5 การตัดสินใจระบุตัวโน้ต (Note Selection)

การตัดสินใจระบุโน้ตในแต่ละเฟรม จะเป็นการพิจารณาว่า โน้ตที่อยู่ในเฟรมควรเป็นโน้ตตัวใด ซึ่งด้วยข้อจำกัดของการเล่นระนาดเอก ทำให้โน้ตที่สามารถเป็นไปได้นั้น มีได้ไม่เกิน 2 ตัว โน้ตในแต่ละเฟรมของข้อมูลเสียง สำหรับเทคนิคที่ใช้ในการพิจารณาตัดสินใจเลือกตัวโน้ตนั้นได้นำเสนอด้วยกัน 2 วิธีคือ การใช้ค่าพลังงานอ้างอิง (Energy Threshold) และ การใช้โครงข่ายประสาทเทียมในการตัดสินใจ

3.5.1 การใช้ค่าพลังงานอ้างอิง (Energy Threshold)

การใช้ค่าพลังงานอิงในการตัดสินใจเลือกตัวโน้ตนั้นจะพิจารณาค่าพลังงานหรือความเด่นชัดของโน้ตที่ได้หลังจากผ่านตัวกรองผ่านแถบความถี่และได้รับการชดเชยค่าความเด่นชัดแล้ว เนื่องจากข้อจำกัดของการเล่นระนาดเอกทำให้ตัวโน้ตของระนาดเอกมีได้ไม่เกินสองตัวโน้ต ดังนั้นอัลกอริทึม ในการพิจารณาตัวโน้ตในแต่ละเฟรมจึงมีได้ไม่เกิน 2 ตัวและพิจารณายอมรับและเลือกตัวโน้ตจำนวน 2 ตัวที่มีค่าพลังงานหรือความเด่นชัดสูงสุดและมีค่ามากกว่าค่าพลังงานอ้างอิงที่กำหนดไว้ แต่ถ้าหากค่าพลังงานของตัวโน้ตตัวใดตัวหนึ่งหรือทั้งสองตัวมีค่าน้อยกว่าค่าอ้างอิงที่กำหนดก็จะปฏิเสธการยอมรับโน้ตตัวนั้นไป ดังแสดงในรูปที่ 3.5



รูปที่ 3.5 แผนผังกระบวนการตัดสินใจเลือกโน้ตด้วยการใช้ค่าพลังงานอ้างอิง

3.5.2 การใช้โครงข่ายประสาทเทียม

การใช้โครงข่ายประสาทเทียมในการพิจารณาเลือกตัวโน้ตในแต่ละเฟรมนั้น จะใช้เพื่อพิจารณาตัดสินใจว่าโน้ตเสียงที่อยู่ในเฟรมนั้นเป็นโน้ตเสียงแบบใด คือ โน้ตแบบโมโนโฟนิก หรือโน้ตแบบโพลีโฟนิก จากนั้นจึงทำการเลือกและยอมรับตัวโน้ตที่มีค่าพลังงานสูงสุดตามจำนวนที่พิจารณาได้จากโครงข่ายประสาทเทียม

สำหรับโครงข่ายประสาทเทียมที่ได้ออกแบบเพื่อใช้ในการทดสอบมีคุณสมบัติดังนี้คือ

1. กำหนดให้ใช้ Feed-Forward Neural Network ทีมีการเรียนรู้แบบ Backpropagation
2. กำหนดมีจำนวน 21 neurons ในชั้นอินพุต, 2 neurons ในชั้นเอาท์พุต และ มี 1 hidden layer โดยมีจำนวน neuron ที่ใช้ในการทดสอบ อยู่ในช่วง 5-40 neurons
3. กำหนด transfer function (activate function) ที่ใช้ทดสอบคือ Hyperbolic tangent sigmoid (tansig) และ Logarithmic sigmoid (logsig)
4. การกำหนด target output

   4.1 tansig กำหนดให้ค่า [1,-1] สำหรับโน้ตเสียงแบบ monophonic และ [-1, 1] สำหรับโน้ตเสียงแบบ polyphonic

   4.2 logsig กำหนดให้ค่า [1, 0] สำหรับโน้ตเสียงแบบ monophonic และ [0, 1] สำหรับโน้ตเสียงแบบ polyphonic
5. พิจารณาเลือกผลลัพธ์จากเอาท์พุตที่ได้ว่าเป็นโน้ตแบบ monophonic หรือ โน้ตแบบ polyphonic
6. ข้อมูลที่ใช้ในการสอนระบบโครงข่ายประสาทเทียมได้มาจากข้อมูลจำนวนครึ่งหนึ่งของข้อมูลทั้งหมดในฐานข้อมูลเสียงที่ได้จัดทำไว้ในหัวข้อ 4.1 ซึ่งมีรายละเอียดคือ โน้ตเดี่ยวตั้งแต่โน้ตตัวที่ 1 ถึง โน้ตตัวที่ 21 และโน้ตผสมซึ่งเกิดจากตัวที่ 1 ถึงโน้ตตัวที่ 21 รวมมีชุดตัวโน้ตทั้งสิ้น 231 ชุด โดยในแต่ละชุดจะบันทึกชุดละ 25 ไฟล์เสียง รวมเป็นไฟล์เสียงที่ใช้ทั้งหมด 5,775 ไฟล์

เนื่องจากขนาดของอินพุตที่ป้อนเข้ามายังโครงข่ายประสาทเทียมนั้น ถ้าหากมีขนาดที่แตกต่างกันมากจะทำให้มีผลกระทบที่ทำให้มีความผิดพลาดเกิดขึ้นได้มาก ดังนั้นเพื่อลดผลกระทบที่เกิดขึ้นจากความแตกต่างจากขนาดของตัวอินพุตแต่ละตัว ดังนั้นจึงได้ทำการนอร์มอลไลซ์อินพุต (Input Normailizing) ก่อนที่จำไปใช้งานซึ่งสามารถทำได้ดังสมการที่ (3-4)

$$S_i = \frac{X_i - mean}{std} \qquad (3\text{-}4)$$

โดยที่

X_i คือ ค่าอินพุตก่อนทำนอร์มอลไลซ์

S_i คือ ค่าอินพุตหลังทำนอร์มอลไลซ์

mean คือ ค่าเฉลี่ยของอินพุต
std คือ ส่วนเบี่ยงเบนมาตรฐาน(Standard Deviation)

สำหรับค่าเฉลี่ยของอินพุต และ ส่วนเบี่ยงเบนมาตรฐานที่ใช้นั้นจะได้มาจากค่าข้อมูลโน้ตเสียงในแต่ละเฟรมที่ใช้การสอนระบบโครงข่ายประสาทเทียมซึ่งมีค่าดังแสดงในตารางที่ 3-3

ตารางที่ 3-3 ค่าเฉลี่ยอินพุตและส่วนเบี่ยงเบนมาตรฐานของโน้ตเสียงในแต่ละเฟรมที่สอนระบบโครงข่ายประสาทเทียม

ตัวโน้ตที่	ค่าเฉลี่ยอินพุต (Mean)	ส่วนเบี่ยงเบนมาตรฐาน(SD)
1	0.5546	0.5496
2	0.4004	0.2843
3	0.2928	0.3278
4	0.2730	0.3555
5	0.3163	0.6154
6	0.4062	0.8252
7	0.3108	0.7365
8	0.3679	0.9493
9	0.4693	1.1887
10	0.4142	1.0954
11	0.2817	0.6278
12	0.3487	0.6737
13	0.2708	0.6143
14	0.3329	0.7548
15	0.2629	0.6308
16	0.4277	0.9982
17	0.2870	0.7780
18	0.1673	0.3808
19	0.1695	0.4919
20	0.9411	1.7423
21	0.2210	0.5275

เนื่องจากในสอนโครงข่ายประสาทเทียมในแต่ละครั้งจะมีการกำหนดค่า Weight เริ่มต้นที่แตกต่างกันซึ่งจะส่งผลให้ผลลัพธ์ที่ได้แตกต่างกัน ดังนั้นเพื่อลดความคลาดเคลื่อนที่เกิดขึ้นจึง

ได้ออกแบบและทดสอบหลาย ๆครั้งและเลือกผลลัพธ์ที่ดีที่สุดสำหรับโครงข่ายประสาทเทียมแต่ละอันนำไปใช้ในการพิจารณาต่อไป

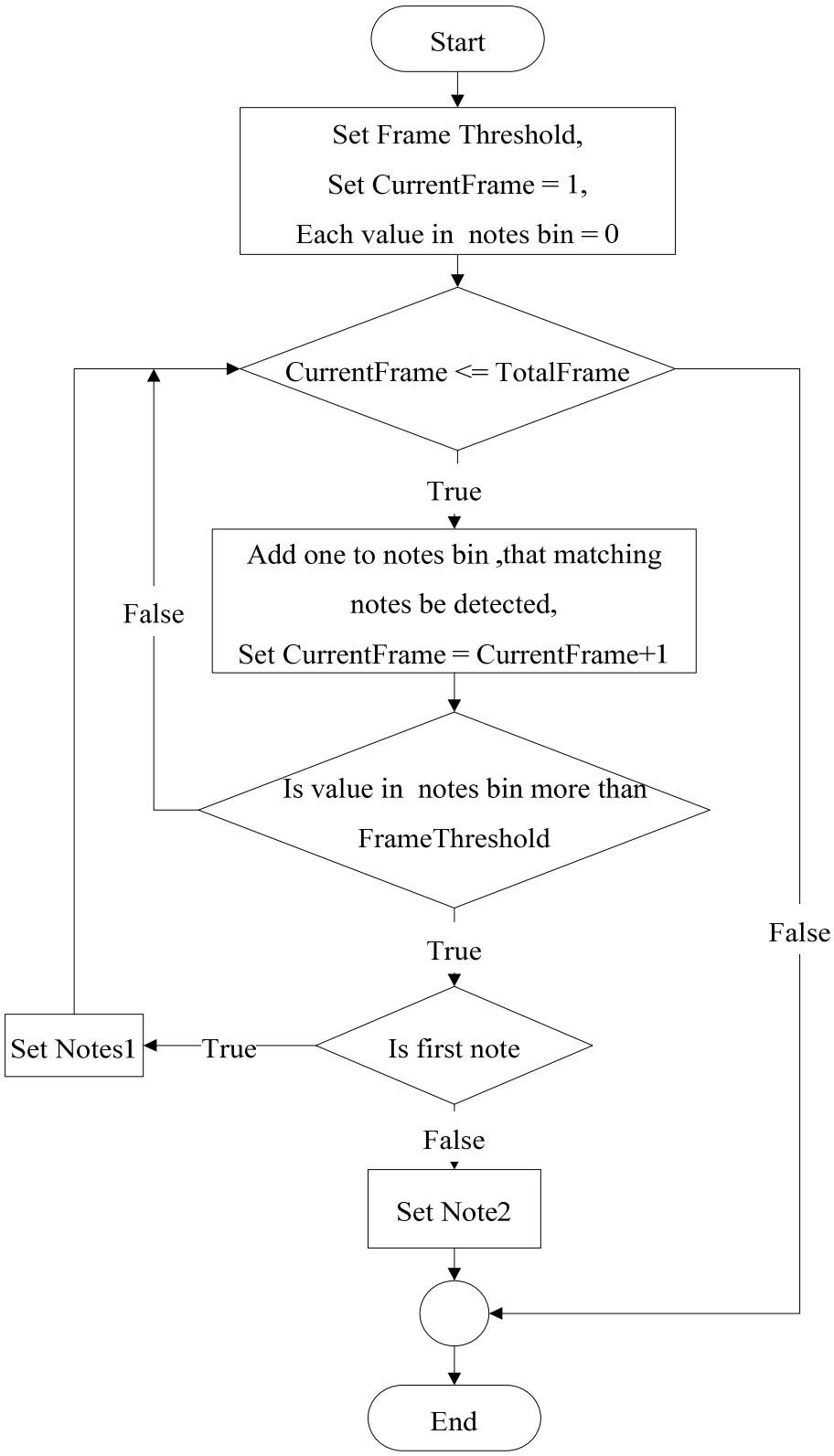
3.6 การทำโพสโปรเซสซิ่ง (Post-Processing)

เนื่องจากผลลัพธ์การระบุตัวโน้ตในแต่ละเฟรมอาจจะมีความผิดพลาดไปบ้างในบางเฟรม ดังนั้นจึงต้องมีการตัดสินใจทั้งหมดอีกครั้ง เรียกการทำงานนี้ว่า กระบวนการโพสโปรเซสซิ่ง ซึ่งเป็นการนำเอาผลลัพธ์ข้อมูลตัวโน้ตในแต่ละเฟรมมารวมกันและพิจารณาตัดสินใจว่าผลลัพธ์โดยรวมทั้งหมดควรเป็นตัวโน้ตตัวใด โดยในการพิจารณาจะใช้หลักการว่าโน้ตที่ระบุได้จะต้องมีช่วงระยะเวลานานเพียงพอ หรือ มีจำนวนเฟรมของโน้ตติดต่อกันมากเพียงพอ ดังแสดงในรูปที่ 3.6

สำหรับในการออกแบบการทดลองที่ได้นำเสนอนี้ได้กำหนดค่าจำนวนเฟรมอ้างอิง (Frame Threshold) เท่ากับ 4 เฟรม ดังนั้นโน้ตที่ระบุได้จะต้องมีความยาวอย่างน้อยเท่ากับ $4 \times 10 = 40$ มิลลิวินาที (จำนวนเฟรมอ้างอิง$\times$อัตราการขยับเฟรม) ซึ่งหากพิจาณาจากข้อกำหนดเบื้องต้นของสัญญาณเสียงเพลงที่มีค่า Tempo เป็นจำนวนมากสุดคือ 120 บีตต่อนาที (Beats/Minute) ซึ่งจะทำให้โน้ตเพลงแต่ละตัวมีช่วงระยะเวลาดังแสดงในตารางที่ 3-4 ซึ่งจะทำให้ระบบสามารถระบุโน้ตได้ถึงระดับตัวโน้ตเขบ็ตสามชั้น

ตารางที่ 3-4 ช่วงระยะเวลาของตัวโน้ตเมื่อกำหนดค่า Tempo เท่ากับ 120 บีตต่อนาที

ชนิดของตัวโน้ต	ระยะเวลา (มิลลิวินาที)
ตัวกลม	2000
ตัวขาว	1000
ตัวดำ	500
เขบ็ตหนึ่งชั้น	250
เขบ็ตสองชั้น	125
เขบ็ตสามชั้น	62.5



รูปที่ 3.6 แผนผังกระบวนการทำโพสโปรเซสซิ่ง (Post-processing)

3.7 การแบ่งช่วงโน้ตเสียงแบบต่อเนื่อง (Boundary Detection)

ส่วนของการแบ่งช่วงเสียงโน้ตต่อเนื่อง จะทำการแบ่งช่วงโน้ตเสียงที่ได้รับเข้ามาออกเป็นส่วน ๆ ซึ่งแต่ละส่วนจะมีค่าข้อมูลที่จะถูกส่งไปพิจารณาตัดสินใจว่าเป็นโน้ตตัวใด ดังนั้นความสำคัญของการแบ่งโน้ตเสียงคือจะทำให้เราทราบได้ว่ามีการเล่นโน้ตเพลงทั้งหมดเป็นจำนวนกี่ครั้ง โดยทั่วไปแล้วการแบ่งช่วงโน้ตเสียงที่มีการใช้งานนั้นสามารถทำได้สองอย่างคือ พิจารณาเชิงความถี่ และ พิจารณาเชิงพลังงาน สำหรับในการทดลองที่ได้นำเสนอนี้จะได้ใช้หลักการพิจารณาเชิงพลังงาน โดยใช้การพิจารณาอัตราการเปลี่ยนแปลงพลังงาน (Difference of Energy) ดังแสดงในสมการที่ (3-5) และ (3-6) ตามลำดับ ทั้งนี้เนื่องจากระนาดเอกเป็นเครื่องดนตรีที่ใช้การเคาะ ดังนั้นจึงมีค่าการเปลี่ยนแปลงพลังงานค่อนข้างเด่นชัด อย่างไรก็ตามเพื่อความเหมาะสมค่าพลังงานอ้างอิงที่ใช้ในการแบ่งช่วงตัวโน้ตอาจจะต้องเปลี่ยนไปตามลักษณะและแรงเคาะระนาดในแต่ละครั้งที่เล่นระนาด

$$E_i - E_j = \Delta E_{ij} \tag{3-5}$$

$$\Delta E_{ij} > E_{SegThreshold} \tag{3-6}$$

โดยที่

i คือตำแหน่งเฟรมที่ i

j คือตำแหน่งของเฟรมที่ j เมื่อ $j = i + 1$

E_i คือ ค่าพลังงานของเฟรมที่ i

E_j คือ ค่าพลังงานของเฟรมที่ j

ΔE_{ij} คือ ค่าความแตกต่างของพลังงานของเฟรมที่ i และเฟรมที่ j

$E_{SegThreshold}$ คือ ค่าพลังงานอ้างอิงที่ใช้ในการแบ่งช่วงตัวโน้ต

3.8 สรุป

ในบทนี้ได้นำเสนอวิธีการรู้จำตัวโน้ตของเสียงระนาดเอกที่เป็นแบบโมโนโฟนิกและโน้ตแบบโพลีโฟนิกที่มีจำนวน 2 ตัวโน้ตตามข้อจำกัดโดยทั่วไปตามวิธีในการเล่นระนาดเอก โดยใช้วิธีการกรองผ่านตัวกรองผ่านแถบความถี่ ซึ่งออกแบบตามความถี่มูลฐานของโน้ตระนาดแต่ละตัว สำหรับวิธีการตัดสินใจระบุตัวโน้ตนั้นได้นำเสนอด้วยกัน 2 แนวทางด้วยกันคือการใช้ค่าพลังงานอ้างอิงและการนำโครงข่ายประสาทเทียมมาช่วยในกันตัดสินใจในการยอมรับหรือปฏิเสธตัวโน้ต รวมถึงการเสนอขั้นตอนในการทำโพสโปรเซสซิ่งและสุดท้ายคือ การนำวิธีการหาค่าอัตราการเปลี่ยนแปลงพลังงานเพื่อนำมาใช้ในแบ่งแยกช่วงตัวโน้ตแต่ละตัวเพื่อใช้สำหรับแบ่งช่วงตัวโน้ต

สำหรับเสียงระนาดที่เป็นแบบโน้ตต่อเนื่อง สำหรับผลการทดสอบของการทำงานของระบบที่ได้ออกแบบไว้จะได้นำเสนอในบทที่ 4 มีการนำเสนอการพัฒนาโปรแกรมตัวอย่างเพื่อนำไปใช้งานในบทที่ 5 และสรุปผลในบทที่ 6 ตามลำดับ