บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของวิทยานิพนธ์

ในปัจจุบันนี้ โลกของการสื่อสารได้พัฒนาขึ้นมาอย่างรวดเร็ว ทำให้ประชาชนแต่ละประเทศสามารถติดต่อสื่อสารและรับทราบข่าวสารกันได้อย่างรวดเร็ว รวมถึงการรับรู้และนำเอาวัฒนธรรมของต่างประเทศมาปฏิบัติกันมากขึ้น ประเทศไทยก็เป็นหนึ่งในนานาประเทศที่ประชาชนได้รับสิทธิเสรีภาพในการติดต่อสื่อสาร ดังนั้นจึงหลีกเลี่ยงไม่ได้ที่จะได้รับรู้เกี่ยวกับข่าวสารและวัฒนธรรมของต่างประเทศ ซึ่งในปัจจุบันนี้วัฒนธรรมจากต่างประเทศได้เผยแพร่เข้ามามีอิทธิพลเป็นอย่างมาก เช่น วัฒนธรรมการกินอาหาร วัฒนธรรมการแต่งกาย วัฒนธรรมการแสดง เป็นต้น การรับเอาวัฒนธรรมจากต่างประเทศเข้ามาใช้ บางอย่างก็เป็นสิ่งที่ดี ที่ช่วยในการพัฒนาคุณภาพและวิถีชีวิต แต่อย่างไรก็ควรที่จะรักษาวัฒนธรรมไทยของเราไว้ด้วยเช่นกัน

ดนตรีไทย เป็นหนึ่งในวัฒนธรรมไทยที่ได้รับผลกระทบจากการเข้ามามีอิทธิพลของวัฒนธรรมดนตรีจากต่างประเทศ ซึ่งจากการสังเกตและผลการสำรวจโดยทั่วไปแล้ว พบว่าประชาชนไทยโดยส่วนใหญ่ในปัจจุบันนี้ ไม่ว่าจะเป็นกลุ่มวัยเรียนหรือวัยทำงาน โดยส่วนใหญ่แล้วให้ความสนใจและรู้จักเครื่องดนตรีจากต่างประเทศเป็นอย่างดี แต่กลับรู้จักและให้ความสนใจเครื่องดนตรีไทยน้อยลง ดังนั้นอาจจะทำให้ดนตรีไทยสูญหายไป ซึ่งเป็นเรื่องที่น่าเสียดายเป็นอย่างมาก

ปัจจุบันนี้มีซอฟต์แวร์ที่ใช้ในการประมวลผลทางด้านสัญญาณเสียงดนตรีอยู่มากมาย ไม่ว่าจะเป็นตั้งแต่ขั้นพื้นฐานกระทั่งถึงที่สลับซับซ้อน แต่โดยมากแล้วจะรองรับการใช้งานกับเครื่องดนตรีที่เป็นอุปกรณ์ไฟฟ้า ที่มีระบบการจัดการเสียงไฟล์มีดี ดังนั้นจึงมีข้อจำกัดกับอุปกรณ์บางประเภทที่ไม่ได้เป็นอุปกรณ์ไฟฟ้า โดยเฉพาะเครื่องดนตรีไทย

วิทยานิพนธ์นี้ได้เสนอการออกแบบและวิธีการรู้จำตัวโน้ตจากเสียงเครื่องดนตรีไทย โดยมุ่งเน้นเพื่อเพิ่มประสิทธิภาพทางด้านการประมวลผลสัญญาณเสียงดนตรีไทยคือ ระนาดเอก (Thai xylophone) ซึ่งเป็นเครื่องดนตรีที่มีทั้งแบบที่เล่นโน้ตตัวเดียว (Monophonic) และแบบเล่นโน้ตสองตัวพร้อมกัน (Polyphonic) และแสดงผลลัพธ์ออกมาในรูปแบบของสัญญาณเสียงแบบไฟล์มีดี (Musical Instrument Digital Interface: MIDI) ทั้งนี้เพื่อเป็นการเปิดโอกาสให้ดนตรีไทย สามารถสื่อออกไปเป็นสากลยิ่งขึ้น และเพื่อเป็นการอนุรักษ์ดนตรีไทยซึ่งเป็นวัฒนธรรมไทยไว้ด้วย

1.2 การตรวจเอกสาร (Literature Review)

ในการตรวจเอกสารของวิทยานิพนธ์นี้ ได้มุ่งเน้นศึกษาประเด็นที่เกี่ยวกับเทคนิคที่ได้มีการนำมาใช้ในการรู้จำตัวโน้ตของเครื่องดนตรีไทยและวิจัยที่เกี่ยวข้องกับการระบุตัวโน้ตจากเสียงเครื่องดนตรีไทย และพบว่ามีผู้นำเสนองานวิจัยไม่มากนัก โดยส่วนมากแล้วจะเป็นงานวิจัยที่เกี่ยวข้องกับเครื่องดนตรีทางตะวันตก ซึ่งจากการศึกษาพบว่าจากความแตกต่างขององค์ประกอบทางดนตรี เช่น การแบ่งช่วงตัวโน้ตของดนตรีไทยและดนตรีสากล รวมถึงคุณสมบัติเฉพาะตัวของเครื่องดนตรีแต่ละชิ้น ทำให้การรู้จำตัวโน้ตซึ่งเป็นงานที่ไม่ยากนักสำหรับนักดนตรีที่มีความชำนาญกลับกลายเป็นงานที่ค่อนข้างยากสำหรับระบบคอมพิวเตอร์

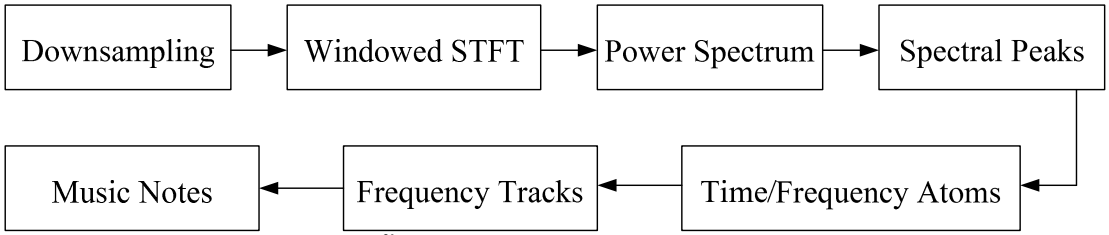
วีระ ทองไพบูลย์และคณะฯ ได้นำเสนอระบบการรู้จำโน้ตจากเสียงเครื่องดนตรีไทยในบทความ [1] โดยได้แสดงถึงอัลกอริทึมในการหาค่าขอบเขตของตัวโน้ตแต่ละตัวโดยการใช้ความแตกต่างของสเปกตรัม (Spectrum) ที่คำนวณได้จากการใช้สมการฟูริเยร์เต็มหน่วยในช่วงระยะเวลาสั้น (Short-Time Fourier Transform: STFT) และหลังจากได้ขอบเขตของโน้ตเสียงแล้วก็จะนำเอาเสียงในแต่ละขอบเขตไปหาค่าสเปกตรัมโดยกำหนดให้สเปกตรัมบนตำแหน่งความถี่ที่มีค่าสูงสุดเป็นความถี่ของโน้ตเสียงจากนั้นจึงนำค่าความถี่ที่ระบุได้ไปเทียบกับช่วงค่าความถี่อ้างอิงต่ำสุดถึงสูงสุดที่รวบรวมมาจากหลายสำนัก [2][3] เพื่อเทียบหาตัวโน้ต

เทคนิคการรู้จำเสียงที่ได้นำเสนอใน [1] มีข้อดีคือไม่จำเป็นต้องทำการสอนระบบในเรียนรู้ตัวโน้ตก่อนใช้งาน แต่อย่างไรก็ตามยังมีข้อจำกัดอยู่ เช่น ในการหาค่าขอบเขตโดยการใช้ความแตกต่างของของสเปกตรัมที่เปลี่ยนแปลงไปนั้น ถ้าหากเป็นเครื่องดนตรีบางชิ้น เช่น ขลุ่ย ถ้าหากเป่าขลุ่ยให้มีเสียงโน้ตตัวเดียวกันสองครั้งแบบต่อเนื่องกันอย่างรวดเร็วอาจจะทำให้ไม่สามารถหาค่าขอบเขตของระหว่างตัวโน้ตทั้งสองตัวได้ นอกจากนี้ การใช้สเปกตรัมบนตำแหน่งความถี่ที่มีค่าสูงสุดเป็นความถี่ของโน้ตเสียงนั้น ไม่สามารถใช้ในกรณีที่ช่วงโน้ตที่ต้องการระบุนั้นมีตัวโน้ตมากกว่าหนึ่งตัวได้

ฤดีรัตน์ ชินเวชกิจวานิชย์และคณะฯ ได้นำเสนอการวิเคราะห์ค่าระดับเสียงของสัญญาณเสียงดนตรีไทยในบทความ [4] โดยมุ่งเน้นไปที่เครื่องดนตรีที่เป็นเครื่องดนตรีสำหรับการเทียบเสียงวงเครื่องสายและมโหรี คือ โน้ตเสียงหลักของขลุ่ยเพียงออและระนาดเอกเหล็ก โดยได้วิเคราะห์สัญญาณเสียงด้วยเทคนิคการแปลงฟูริเยร์เต็มหน่วย (Discrete Fourier Transform: DFT), เทคนิคการแปลงฟูริเยร์ในช่วงเวลาสั้น, เทคนิคการวิเคราะห์สัญญาณข้อมูลด้วยแบบจำลองเออาร์ (Autoregressive model: AR) และเทคนิคการกระจายเชิงโมด (Modal Distribution: MD) [4] ซึ่งผลการทดสอบพบว่าเสียงดนตรีไทยนั้นมีความเป็นเอกลักษณ์ของความเป็นธรรมชาติมากเนื่องจากในการปรับตั้งเสียงของเครื่องดนตรีไทยนั้นใช้ทักษะและความชำนาญของส่วนบุคคลโดยไม่ได้ใช้เครื่องมือวัดละเอียดใด ๆเข้ามาช่วย ดังนั้น ความถี่ของโน้ตแต่ละเสียงจึงไม่สามารถแสดงด้วยตัวเลขจำนวนเต็มได้ลงตัวและยังมีอาการแกว่งของระยะพิตช์

(Pitch) ส่งผลทำให้ระยะพิตช์ของเสียงดนตรีไทยไม่คงที่เหมือนสมมุติฐานดั้งเดิมของมอร์ตัน [5]

S.Dixon ได้นำเสนอการรู้จำตัวโน้ตจากเสียงเปียโนในบทความ [6] โดยกำหนดข้อมูลเสียงเปียโนที่รับเข้ามานั้นเป็นเสียงแบบโพลีโฟนิกและสร้างผลลัพธ์ออกมาเป็นไฟล์มีดี ดังนั้นข้อมูลที่ต้องตรวจสอบเพื่อนำไปใช้คือ ระดับเสียง, ระยะเวลา และ ความดังของตัวโน้ต โดยมีขั้นตอนการทำงานเป็นไปดังรูปที่ 1.1



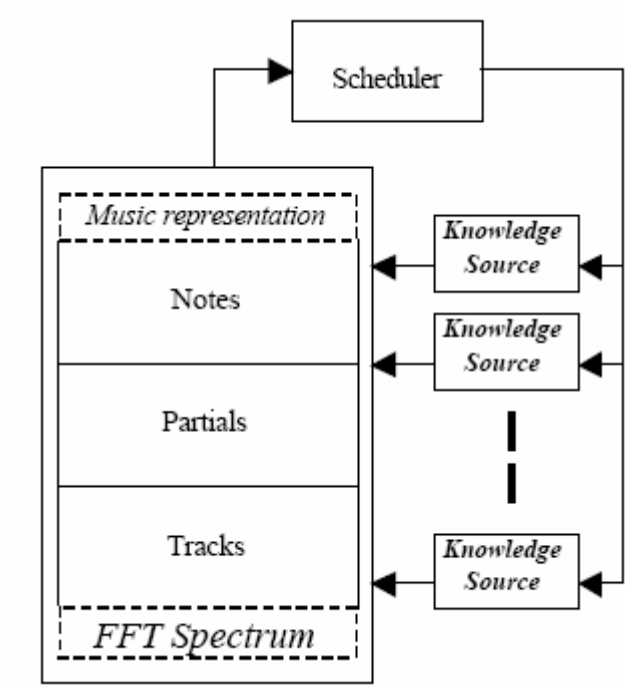
รูปที่ 1.1 ขั้นตอนการทำงานการรู้จำเสียงเปียโน [6]

เทคนิคการหาค่าระดับเสียง (Peak-picking) ที่นำเสนอในบทความ [6] ได้ใช้การเปรียบเทียบกับค่าอ้างอิงโดยค่าสเปกตรัมของระดับเสียงที่หาได้ในช่วงเวลาสั้น ๆนั้นจะต้องมีค่ามากกว่าค่าอ้างอิงต่ำสุดและต้องมีค่าระดับพลังงานอย่างน้อยร้อยละ 1 ของพลังงานทั้งหมดของสัญญาณเสียงในช่วงเวลานั้น ซึ่งมีข้อดีคือ ทำให้สามารถระบุตัวโน้ตได้หลายตัวในช่วงระยะเวลาเดียวกันซึ่งจากรายงานการทดสอบใน [6] ได้แสดงผลค่าคะแนนในการระบุตัวโน้ตถูกต้องได้เป็นจำนวนสูงสุด คือ ร้อยละ 81.8 อย่างไรก็ตามยังมีข้อจำกัดอยู่คือ ถ้าหากนำมาใช้กับเครื่องดนตรีบางชิ้นที่เป็นประเภทเคาะ เช่น ระนาดเอก ช่วงที่เคาะลูกระนาดในช่วงแรกจะมีการสั่นและส่งผลให้ค่าความถี่ของเสียงที่ออกมาในช่วงแรกมีการเปลี่ยนแปลงและไม่คงที่อยู่เป็นระยะเวลาสั้น ๆ ซึ่งจะส่งผลทำให้ระบบรู้จำระบุระดับความถี่เสียงที่ไม่ต้องการว่าเป็นเสียงตัวโน้ตออกมาด้วย ซึ่งทำให้ระบบรู้จำตัวโน้ตมีค่าความผิดพลาดมากยิ่งขึ้น

Juan Pablo Bello, et al ได้นำเสนอและสรุปวิธีการที่มีการใช้ในระบบรู้จำตัวโน้ตสองแบบในบทความ [7] คือ ระบบที่รู้จำเฉพาะเสียงแบบโมโนโฟนิกและระบบที่รู้จำเสียงแบบโพลีโฟนิกโดยในในแต่ละแบบได้มีเทคนิควิธีการแตกต่างกัน

ระบบรู้จำที่เฉพาะเสียงแบบโมโนโฟนิกนั้นใช้เทคนิกการแยกแยะระดับเสียงโดยใช้วิธีการคำนวณหาอัตตสหสัมพันธ์ (Autocorrelation Pitch Tracker) [7][8] ส่วนในการแบ่งตัวโน้ตได้ใช้วิธีการตรวจสอบระดับค่าพลังงานอ้างอิงเพื่อตรวจจุดเริ่มต้นและจุดสิ้นสุดตัวโน้ต ทั้งนี้สัญญาณเสียงของตัวโน้ตที่ยอมรับได้นั้นจะต้องมีช่วงระยะเวลาที่นานพอซึ่งคล้ายกันกับช่วงระยะเวลาที่นานพอของข้อมูลเสียงที่สมองของมนุษย์จะสามารถรับรู้และพิจารณาความหมายได้ [9]

ระบบรู้จำเสียงแบบโพลีโฟนิกได้ใช้วิธีการที่เรียกว่าระบบแบล็คบอร์ด (Blackboard System) ซึ่งได้เคยมีการนำเสนอในบทความ [10][11][12] ซึ่งมีขั้นตอนในการทำงานเป็นดังรูปที่ 1.2



รูปที่ 1.2 การทำงานของระบบแบล็คบอร์ด (Blackboard System) [7]

ระบบแบล็คบอร์ด มีองค์ประกอบหลักในการทำงานอยู่ 3 องค์ประกอบคือคือ ตัวของแบล็คบอร์ด (Blackboard) หรือข้อมูลที่ถูกนำเสนอและจะพัฒนาต่อไป Knowledge Source หรือแหล่งความรู้และกฎเกณฑ์ที่จะใช้ในการตรวจสอบ และองค์ประกอบสุดท้ายคือ Scheduler หรือตัวควบคุมลำดับในการใช้งาน Knowledge Source เพื่อใช้ในการตรวจสอบ ข้อดีของการใช้งานวิธีการระบบแบล็คบอร์ดนั้นจะทำให้เราสามารถแก้ไขปัญหาได้ง่ายขึ้นเนื่องจากระบบสามารถเพิ่มหรือลด Knowledge Source ได้ง่าย แต่ข้อเสียคือ ถ้าหากเป็นระบบที่ซับซ้อนทำให้ต้องมี Knowledge Source เป็นจำนวนมากในการจัดเรียงลำดับ การใช้งาน Knowledge Source ให้เหมาะสมก็เป็นสิ่งที่ยากมากยิ่งขึ้น

Tim Shuttleworth และ Roland Wilson ได้นำเสนอการการใช้งานโครงข่ายประสาทเทียม (Neural Network) ในบทความ [13] สำหรับการแก้ไขปัญหาสองอย่าง ลำดับแรกคือการระบุคอร์ด (Chord) ที่ประกอบไปด้วยกลุ่มโน้ตจำนวนสามตัวโน้ต (Triad) ลำดับที่สองคือการแยะแยะตัวโน้ตจากคอร์ด (Chord) จากผลการทดลองที่ได้พบว่าในการระบุตัวโน้ตที่ได้ยังมีความผิดพลาดเกิดขึ้น เช่น ความผิดพลาดในการแยะแยะโน้ตที่เป็นเสียงโน้ตตัวเดียวกันแต่อยู่คนละออกเทฟ (Octave) นอกจากนี้ยังมีข้อจำกัดเช่น เพื่อให้สามารถระบุโน้ตเสียงได้เป็นจำนวนมาก โครงข่ายประสาทเทียมที่ได้ออกแบบขึ้นใน [13] เพื่อใช้ในระบุคอร์ดจากตัวโน้ตนั้นจึงต้องมี

จำนวนอินพุตโหนด (Input Node) เท่ากับโน้ตที่ต้องการระบุได้ ดังนั้นจึงจำเป็นต้องเตรียมข้อมูลและใช้ระยะเวลาสอนให้ระบบเรียนรู้มากขึ้น และถ้าหากมีการแก้ไขก็จำเป็นต้องมีการสอนระบบใหม่อีกครั้ง

Matija Marol ได้นำเสนอ การเปรียบเทียบการรู้เสียงเปียโนแบบโพลีโฟนิก โดยการใช้งานโครงข่ายประสาทเทียมแบบ Feed Forword Neural Network ในลักษณะต่าง ๆไว้ในบทความ [14] เป็นจำนวน 4 แบบ คือ Mutilayer perceptron (MLP), Time-Dalay Network (TDNNs), RBF Network (RBF) และ Support Vector Machines (SVMs) [15] ซึ่งจากรายงานสรุปผลการทดสอบพบว่าการใช้ Mutilayer perceptron และ Time-Dalay Network ให้ผลการรู้จำใกล้เคียงกันมากและให้ผลลัพธ์ดีกว่า RBF Network และ Support Vector Machine

Peyman Heydarian และ Joshua D. Reiss ได้นำเสนอการประยุกต์ใช้งานของเทคนิคการหาความถี่มูลฐาน (Fundamental Frequency) กับเครื่องดนตรีของทางเปอร์เซีย (Persia Music) ชนิดหนึ่งเรียกว่า ซันตู (Santur) ซึ่งเป็นเครื่องดนตรีที่ที่คล้ายกับขิม ไว้ในบทความ [16] สำหรับเทคนิคในการหาค่าความถี่มูลฐานของเสียงนั้นใช้การคำนวณหาค่า Cross-Correlation ระหว่างตัวอย่างชุดทดสอบกับชุดรูปแบบที่ใช้สำหรับอ้างอิง [17][18]

Jun Yin,Ankur Dhanik.,et al ได้นำเสนอการประยุกต์ใช้งานการวิเคราะห์สัญญาณเสียงโน้ตไวโอลินแล้วนำไปแสดงผลลัพธ์เป็นภาพเคลื่อนไหวแสดงลักษณะการเล่นไวโอลินไว้ใน [19] สำหรับเทคนิคที่ใช้ในการหาค่าระดับเสียงนั้นได้อาศัยการศึกษาและพิจาณาจากลำดับโครงสร้างฮาร์โมนิกของไวโอลิน โดยใช้ค่าของฮาร์โมนิก 5 ลำดับแรกคือฮาร์โมนิกลำดับที่ 12,19,24,28 และ 31 มาหาผลรวมแล้วจึงพิจารณาว่าค่าความถี่มูลฐานค่าใดให้ค่าผลลัพธ์มากที่สุด ก็จะถือว่าเป็นระดับเสียงในช่วงเวลานั้น ข้อดีของการใช้วิธีนี้คือไม่จำเป็นต้องสอนระบบให้มีการเรียนรู้ก่อนการใช้งาน แต่อย่างไรก็ตามยังมีข้อจำกัดคือ ถ้าต้องการใช้งานกับเครื่องดนตรีชนิดใดจะต้องศึกษาค่าฮาร์โมนิกของของเครื่องดนตรีชนิดนั้นว่ามีค่าฮาร์โมนิกลำดับใดบ้างที่เด่นชัดขึ้นมา เพื่อนำมาใช้ในการพิจารณาหาค่าระดับเสียงที่แท้จริง

1.3 วัตถุประสงค์

1) เพื่อศึกษาและออกแบบวิธีการในการระบุตัวโน้ตจากเสียงระนาดเอก
2) เพื่อพัฒนาระบบที่สามารถช่วยอำนวยความสะดวกในการแปลงเสียงระนาดเอกให้เป็นตัวโน้ต

1.4 ขอบเขตของการวิจัย

1) โน้ตดนตรีที่สามารถระบุได้เป็นได้ทั้งแบบที่เล่นโน้ตตัวเดียวและแบบเล่นโน้ตสองตัวพร้อมกัน
2) ผลลัพธ์ของการระบุตัวโน้ตจัดอยู่ในรูปแบบของไฟล์มีดีในรูปแบบที่ 1 (Format Type 1) โดยกำหนดให้มีจำนวน 2 แทรค (Track) และ 1 ช่องสัญญาณเสียง
3) ในการเล่นระนาดมีเทคนิควิธีการเคาะลูกระนาดหลายแบบแต่ในงานวิจัยนี้ได้ใช้การเคาะลูกระนาดแบบปกติ คือ หนึ่งโน้ตต่อการเคาะหนึ่งครั้งเท่านั้น
4) การบันทึกข้อมูลเสียงดนตรีจัดทำในสภาพแวดล้อมที่เงียบ โดยกำหนดค่าความถี่ในการสุ่มตัวอย่าง (Sampling Rate) เท่ากับ 16,000 เฮิรตซ์ (Hz) และค่าความละเอียดของข้อมูลเท่ากับ 16 บิต
5) การทดสอบวิธีการที่นำเสนอประกอบด้วยการทดสอบการรู้จำตัวโน้ตและการหาขอบเขตของของโน้ตเสียงระนาด
6) มีการพัฒนาโปรแกรมตัวอย่างที่มีความสามารถในการระบุโน้ตเสียงของระนาดเอก โดยกำหนดให้ผลลัพธ์จัดอยู่ในรูปแบบของไฟล์มีดีได้ตามที่กำหนดและเล่นสัญญาณเสียงกลับออกมาได้

1.5 ขั้นตอนและวิธีการดำเนินการวิจัย

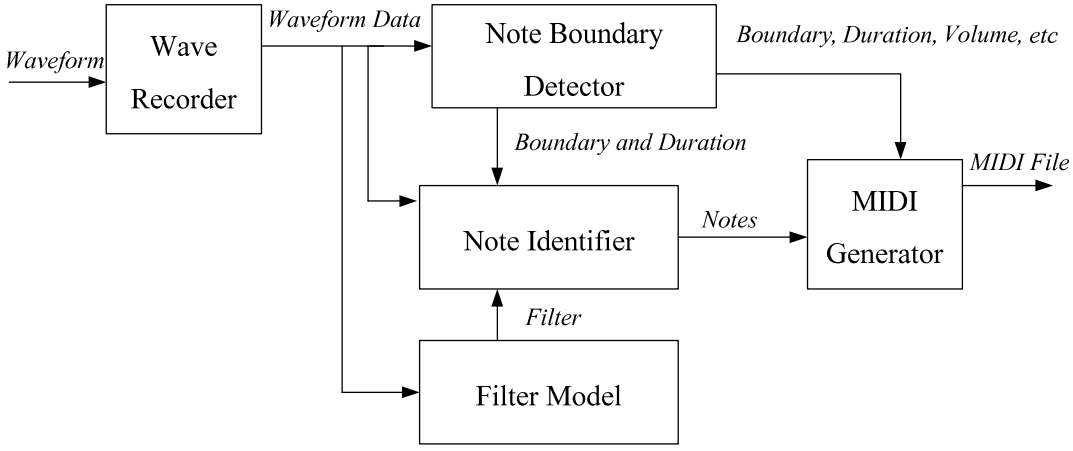
1) ศึกษาหลักการของการประมวลผลสัญญาณดิจิตอล (Digital Signal Processing)
2) ศึกษาข้อมูลและทฤษฎีเบื้องต้นเกี่ยวหลักการทางด้านดนตรี เช่น คุณสมบัติของเครื่องดนตรี การแบ่งตัวโน้ตสากล ความถี่ของตัวโน้ต เป็นต้น
3) ศึกษาวิธีการในการระบุตัวโน้ตจากเสียงเครื่องดนตรีและเอกสารอื่น ๆที่เกี่ยวข้อง ที่ได้มีการตีพิมพ์ลงในบทความทางวิชาการ
4) จัดทำฐานข้อมูลเสียงระนาดเอก
5) ออกแบบวิธีการในการรู้จำตัวโน้ตจากเสียงระนาดเอก
6) ทดสอบและเปรียบเทียบวิธีการในการรู้จำตัวโน้ต
7) ทดสอบวิธีการหาขอบเขตของของโน้ตเสียงระนาด
8) นำวิธีการที่ได้ออกแบบไว้มาพัฒนาเป็นโปรแกรมตัวอย่างเพื่อทดสอบความสามารถในการระบุโน้ตเสียงของระนาดเอก
9) รวบรวมผลการทดลอง สรุปผล และจัดทำวิทยานิพนธ์

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1) ระบบที่สร้างขึ้นสามารถนำไปใช้ในการแปลงเสียงระนาดเอกเป็นตัวโน้ตซึ่งช่วยให้ไม่ต้องเขียนโน้ตและโน้ตที่ได้อยู่ในรูปแบบไฟล์มีดีซึ่งสามารถนำไปประยุกต์ใช้งานได้สะดวก
2) ได้วิธีการที่สามารถนำไปใช้ในการระบุตัวโน้ตจากเสียงเครื่องดนตรีไทยคือระนาดเอก
3) เพื่อเป็นการส่งเสริมความรู้ความสนใจเกี่ยวกับดนตรีไทยโดยเฉพาะระนาดเอกให้เป็นที่รู้จักต่อบุคคลทั่วไปมากยิ่งขึ้น
4) สามารถนำความรู้ที่ได้ไปประยุกต์ใช้กับการรู้จำเครื่องดนตรีชนิดอื่น ๆ

1.7 ภาพรวมของระบบ

ในวิทยานิพนธ์นี้ ภาพรวมของระบบเป็นดังรูปที่ 1.3 ระบบจะอ่านไฟล์ข้อมูลเสียงระนาดเอกเข้ามา โดยข้อมูลเสียงระนาดเอกที่ได้รับนั้นสามารถเป็นเสียงเคาะลูกระนาดแบบต่อเนื่องได้ ดังนั้นระบบจะแบ่งขอบเขตของโน้ตเสียงและระยะเวลาของโน้ตแต่ละตัวเสียก่อน ต่อจากนั้นจึงระบุตัวโน้ตในแต่ละขอบเขตว่าเป็นตัวโน้ตระนาดเอกตัวใด สุดท้ายจึงจัดรูปแบบผลลัพธ์ในการรู้จำตัวโน้ตให้อยู่ในรูปของไฟล์มีดี เพื่อที่จะได้นำไปแสดงผลต่อไป



รูปที่ 1.3 ภาพรวมของระบบ