# บทที่ 2

# ทฤษฎีและวิทยานิพนธ์ที่เกี่ยวข้อง

วิทยานิพนธ์นี้ทำการศึกษาเกี่ยวกับการจำแนกประเภทดนตรีคันทรี ป๊อบ ร๊อค และแจ็ส ใช้เพลงที่แตกต่างกันโดยนำเพลงภาษาไทยและฝรั่งมาใช้ในการทดลอง ทฤษฎีที่เกี่ยวข้องเพื่อนำมาสนับสนุน ในงานวิทยานิพนธ์ชิ้นนี้ ประกอบด้วยทฤษฎีของเสียง และประเภทดนตรี การดึงค่าคุณลักษณะเด่นของงานวิจัยนี้ประกอบด้วยการดึงค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) ประกอบด้วย MFCC,LPC,ZC,SF,SC,SR และ SSB ส่วนในวิทยานิพนธ์ที่เกี่ยวข้องที่นำมาใช้ในการอ้างอิงในเรื่องของการจำแนกประเภทเพลงโดยใช้วิธีการดึงค่าคุณลักษณะที่หลากหลายใช้ตัวจำแนกที่และอัลกอริทึ่มที่แตกต่างกันซึ่งจะแสดงรายละเอียดของทฤษฎี หลักการ และแนวทางการศึกษาเพื่อดำเนินวิทยานิพนธ์ต่อไป

## 2.1 ทฤษฎีที่เกี่ยวข้อง

ในส่วนนี้นำเสนอความหมายของดนตรีและคุณลักษณะต่าง ๆ ของค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) เพื่อทำให้เข้าใจกระบวนการความสามารถรวมถึงที่มาต่าง ๆ ของค่าคุณลักษณะที่นำมาใช้ในการจำแนก

“เครื่องดนตรี” พจนานุกรมฉบับราชบัณฑิตยสถาน (2525) ให้ความหมายไว้ว่าเป็นเครื่องบรรเลงซึ่งมีเสียงดัง ทำให้รู้สึกเพลิดเพลิน หรือเกิดอารมณ์รัก โศกเศร้า และรื่นเริงได้ตามทำนองเพลง

ดนตรี พจนานุกรมฉบับราชบัณฑิตสถาน (2542) คือ เสียงที่ประกอบกันเป็นทำนองเพลง เครื่องบรรเลงซึ่งมีเสียงดังทำให้เกิดความรู้สึกเพลิดเพลินหรือเกิดอารมณ์รัก โศกเศร้า และรื่นเริงเป็นต้น

ดนตรีเป็นสิ่งที่ธรรมชาติให้มาพร้อม ๆ กับชีวิตมนุษย์โดยที่มนุษย์เองไม่รู้ตัว ดนตรีเป็นทั้งศาสตร์และศิลป์อย่างหนึ่งที่ช่วยให้มนุษย์มีความสุข สนุกสนานรื่นเริง ช่วยผ่อนคลายความเคลียดทั้งทางตรงและทางอ้อม ดนตรีเป็นเครื่องกล่อมเกลาจิตใจของมนุษย์ ให้มีความเบิกบานหรรษาให้เกิดความสงบและพักผ่อน (ณณฐ วิโย, 2553)

เพลงคือ สำเนียงที่ขับร้องหรือเสียงทางดนตรีที่ประดิษฐ์ขึ้นโดยมีสัดส่วนและวรรคตอนตามจินตนาการของผู้ประพันธ์ ซึ่งมีลักษณะเดียวกับบทกวีที่แต่งขึ้นโดยใช้ถ้อยคำ เสียง

อักษร สระ และวรรณยุกต์ ตามกฎแห่งฉันทลักษณ์ เพลงหนึ่งจะมีกี่จังหวะหรือกี่ท่อนนั้น ไม่มีปรากฏตายตัว (พจนานุกรมฉบับราชบัณฑิตยสถาน, 2540)

เพลงสามารถจำแนกได้หลายวิธี การจำแนกประเภทเพลงตามหลักวิชาการดนตรี แบ่งได้ 4 ประเภท 1) เพลงที่มีเนื้อและทำนอง 2) เพลงบรรเลงแต่ทำนอง 3) เพลงที่มีแต่ดนตรีประกอบ 4) เพลงบรรเลงด้วยเครื่องดนตรี หากจำแนกตามจังหวะแบ่งได้เป็น เพลงคลาสสิค (Classic Music) เพลงแจ๊ส (Jazz Music) เพลงบูลส์ (Blue Music) และเพลงริธึ่มแอนบูลส์ (Rhythm and Blue) หากแบ่งเพลงตามความหมายที่เกี่ยวข้องกับชีวิตสามารถจำแนกเป็นประเภทได้เป็น เพลงพื้นบ้าน (Folk Music) เพลงเพื่อชีวิต เพลงลูกทุ่ง (Country Music) เพลงลูกกรุงหรือเพลงป๊อบ (Light Music or Popular Music) (ทรัพยากรสารสนเทศด้านดนตรี, 2551)

องค์ประกอบของดนตรีคือ ส่วนสำคัญพื้นฐานที่ทำให้ดนตรีเป็นรูปร่างขึ้นมาได้ ดนตรีของทุกชาติ ทุกภาษา จะมีองค์ประกอบของดนตรีที่เป็นหลักอยู่เหมือน ๆ กันหรือคล้ายคลึงกันแต่จะมีความแตกต่างกันในส่วนรายละเอียดและการปรุงแต่งเพื่อให้เกิดความเหมาะสมตามสภาพวัฒนธรรมของตน (ณณฐ วิโย, 2553) ซึ่งประกอบไปด้วยห้าองค์ประกอบต่าง ๆ ดังนี้

1. **จังหวะ (Rhythm)** หมายถึง ความสม่ำเสมอหรือช่วงเท่า ๆ กันในการแบ่งระยะของทำนองเพลง โดยใช้เครื่องดนตรีประเภทเครื่องกำกับจังหวะเป็นตัวควบคุม จังหวะถือว่าเป็นหัวใจสำคัญของดนตรี เพราะจังหวะเปรียบเสมือนเครื่องกำหนดระเบียบแบบแผนในการบรรเลงและการประพันธ์บทเพลง

2. **ทำนอง (Melody)** หมายถึง การจัดเรียงของเสียงที่มีความแตกต่างระดับเสียงและความยาวของเสียงหรือระดับเสียงสูงต่ำ สั้นบ้างยาวบ้างสลับกันไป บางครั้งก็ดังบางครั้งก็ค่อย ทั้งนี้ขึ้นอยู่กับความประสงค์ของผู้ประพันธ์ โดยทั่วไปดนตรีจะประกอบด้วยทำนองซึ่งเป็นองค์ประกอบที่ง่ายต่อการจดจำมากที่สุด

3. **พื้นที่ผิวและการประสานเสียง** (Texture and Harmony) หมายถึงลักษณะการจัดความสัมพันธ์ของดนตรีระหว่างแนวทำนองแต่ละแนวโดยพิจารณาในแนวนอน ส่วนการประสานเสียง (Harmony) เป็นการจัดความสัมพันธ์ของเสียงในแนวดิ่งหรือแนวตั้ง ในบทเพลงแต่ละบทเพลงจะประกอบทั้งสองอย่างเข้าด้วยกัน พื้นผิวทางดนตรี (Musical Texture) เป็นลักษณะการจัดความสัมพันธ์ระหว่างทำนองกับการประสานเสียงในดนตรี ลักษณะพื้นผิวในดนตรีขึ้นอยู่กับการแปรทำนองและการประสานเสียงพื้นผิวทางดนตรี (Musical Texture) คือองค์ประกอบของเสียงที่เกิดจากแนวทำนองของเสียงหลาย ๆ แนวมาประสานกัน

**4. สีสันหรือคุณภาพของเสียง** (Tone Color of Quality) ในวงดนตรีต่าง ๆ ไม่ว่าจะเป็นของชาติใดก็ตามจะประกอบด้วยเครื่องดนตรีต่างชนิดมากบ้างน้อยบ้างผสมกัน ทั้งนี้เนื่องมาจากเสียงของเครื่องดนตรีแต่ละชนิดจะให้ความรู้สึกในการฟังที่แตกต่างกัน เช่น ซอ กับขลุ่ย ก็มีคุณภาพเสียงไม่เหมือนกันลักษณะนี้ ในทางดนตรีเรียกว่ามี “สี” (Tone color) ต่างกัน สีสันหรีคุณภาพของเสียง (Tone Color of Quality) เป็นลักษณะเฉพาะตัวของเสียงเครื่องดนตรีและเสียงขับร้องซึ่งคุณภาพของเสียงจะไม่เหมือนกันและแตกต่างกัน

**5. คีตลักษณ์หรือรูปแบบ** (Forms) หมายถึง ลักษณะของบทเพลงแต่ละแบบที่มีความแตกต่างกัน โดยขึ้นอยู่กับผู้แต่งว่าจะดำเนินไปในลักษณะใด

### 2.1.1 องค์ประกอบของดนตรี

เสียงเกิดจากการสั่นสะเทือนของวัตถุ เสียงที่ผ่านเข้าไปในหูของคนเรา ไม่ว่าจะเป็นเสียง บนท้องถนน เสียงพูด เสียงร้อง เราใช้เสียงในการติดต่อสื่อสาร เสียงสามารถแสดงออกได้ทั้งการมีความสุข ความพอใจ โดยองค์ประกอบของดนตรีประกอบด้วย ระดับเสียง ระดับความเข้มเสียง คุณภาพเสียง จังหวะ และทำนอง

**2.1.1.1 ระดับเสียง** (Pitch)

เป็นระดับความสูงต่ำของเสียง ระดับเสียงขึ้นอยู่กับความถี่ (Frequency) ของการสั่นสะเทือน การสั่นสะเทือนยิ่งเร็วเสียงจะยิ่งสูง แต่ถ้าการสั่นสะเทือนยิ่งช้า เสียงก็จะยิ่งต่ำ ความถี่ของการสั่นสะเทือนปกติจะวัดเป็นรอบ/วินาที (Cycle/Second) โดยในเปียโนมีความสั่นสะเทือนเสียงที่สูงที่สุดอยู่ที่ 4,186 รอบ/วินาที และความถี่ต่ำที่สุดคือ 27 รอบ/ วินาที

**2.1.1.2 ระดับความเข้มเสียง (Volume หรือ Dynamic)**

ระดับความดังค่อยในทางดนตรีจะเรียกว่า Dynamic เป็นองค์ประกอบสำคัญทางดนตรี มีความสัมพันธ์กับความกว้างของคลื่นเสียง (Amplitude) ในการสั่นสะเทือนที่ทำให้เกิดเสียงนั้นช่วงความกว้างมากเสียงจะดังมาก และช่วงกว้างน้อยเสียงจะเบา หรือวัตถุสั่นน้อยเสียงจะเบา วัตถุสั่นสะเทือนมากเสียงจะดัง ในการบรรเลงดนตรีในวงให้มีความดังมากแล้วค่อยลดลงหรือมีการเปลี่ยนแปลงจำนวนของเครื่องดนตรีที่เล่นให้น้อยลง ผลที่ตามมาก็คือจะมีการเปลี่ยนแปลงความดัง-ค่อย (Dynamic) ซึ่งจะมีผลต่อความรู้สึกของผู้ฟังอาจจะทำให้รู้สึก ตื่นเต้น ยิ่งใหญ่

2.1.1.3 **คุณภาพเสียงหรือสีสันเสียง** (Timbre **หรือ** Tone Color of Quality)

สีสัน (Tone color) คือ เป็นคุณสมบัติเฉพาะของเครื่องดนตรี รวมทั้งเสียงร้องของมนุษย์ ซึ่งมีความแตกต่างโดยธรรมชาติของแหล่งกำเนิดเสียง เครื่องดนตรีแต่ละชนิดจะมีเสียงที่ให้อารมณ์แตกต่างกัน ในทำนองเดียวกัน ถ้าเราเปลี่ยนสีสันของเสียงดนตรีหรือเสียงร้องแล้วทำให้เกิดความรู้ที่ตัดกัน เช่น เมื่อเราให้เครื่องดนตรีชนิดหนึ่งเล่นทำนองหนึ่งแล้วเปลี่ยนให้เครื่องดนตรีอีกชนิดเล่นทำนองเดียวกัน ก็จะผลต่ออารมณ์และความรู้สึก คุณลักษณะนี้ทำให้เกิดความรู้สึกที่แตกต่างได้

2.1.1.4 **จังหวะ** (Rhythm)

ในทางดนตรีจังหวะหมายถึงการเคลื่อนที่ของแนวทำนองหรือเสียงในช่วงเวลาหนึ่ง โดยปกติจังหวะประกอบด้วย จังหวะหนัก จังหวะเบา จังหวะทำนอง (Duration) หมายถึง ความสั้นยาวของเสียงทุกเสียงในทำนองเพลง

2.1.1.5.**ทำนอง** (Melody)

ทำนองคือเสียงที่เปล่งออกมาโดยมีความต่อเนื่องกันเป็นระบบ ทำนองเปรียบเหมือนรูปร่างของบทเพลงที่มีเสียงสูง ต่ำและ สั้น ยาว ประกอบเข้าไว้ด้วยกัน โดยทั่วไปดนตรีประกอบด้วยทำนองซึ่งเป็นองค์ประกอบที่สามารถจำได้ง่าย องค์ประกอบของทำนองประกอบด้วย จังหวะทำนอง (Melodic Rhythm) มิติ (Melodic Dimensions) ความยาว (Length) และช่วงความกว้างของเสียง

ณรุทธ์ สุทธจิตต์ (2540) กล่าวว่า สุนทรียภาพทางดนตรีเกิดขึ้นได้เนื่องจากผู้นั้นได้สัมผัสกับสุนทรีวัตถุโดยกระบวนการทางปัญญาเชิงคุณภาพ เมื่อนำเอาการประเมินคุณค่ามาเกี่ยวข้องด้วยก็จะนำไปสู่ความซาบซึ้งอันเนื่องมาจากสุนทรียประสบการณ์นั้น ดนตรีเป็นศิลปะที่อาศัยเสียงเพื่อถ่ายทอดอารมณ์ไปสู่ผู้ฟัง เป็นศิลปะที่ง่ายต่อการสัมผัส ก่อให้เกิดความสุข ความปิติพอใจแก่มนุษย์ ดังคำกล่าวว่า “ดนตรีเป็นภาษาสากล” เพราะสามารถเป็นสื่อความรู้สึกของชนทุกชาติได้ เสาวณีย์ สังฆโสภณ วารสารเพลงดนตรี (2541) กล่าวว่า การฟังเพลงเป็นกิจกรรมทางดนตรีอย่างหนึ่งที่ได้รับความนิยมจากผู้ฟัง ทุกชาติ ทุกภาษา ทุกเพศ ทุกวัย ทั้งคนปกติ ผู้ป่วย และคนพิการ เพราะการฟังเพลงก่อให้เกิดความสุข ความเบิกบานใจ ผ่อนคลายความเครียดทั้งทางกายและจิตใจ ช่วยทำให้สุขภาพจิตดี ช่วยพัฒนาสมอง ทำให้เกิดความจำ การรับรู้ และการเรียนรู้ได้ดีขึ้น ในงานวิจัยนี้นำเสนอประเภทดนตรีที่เกี่ยวข้อง 4 ประเภท

### 2.1.2 ประเภทดนตรี

ประเภทดนตรีของแต่ละพื้นที่แต่ละชนชาติมีลักษณะที่แตกต่างกันซึ่งมีเอกลักษณ์ของตนเองที่แตกต่างกันไป ในงานวิจัยชุดนี้จะศึกษาประเภทดนตรีที่เป็นที่สากลและเป็นที่นิยมซึ่งประกอบด้วย ร็อค ป็อบ คันทรี และ แจ๊ส ซึ่งในส่วนนี้จะอธิบายที่มาและลักษณะของดนตรีดังต่อไปนี้

#### 2.1.2.1 ดนตรี แจ๊ส (Jazz)

ดนตรีแจ๊ส เป็นแนวเพลงที่มีลักษณะดนตรีแอฟริกาแถบตะวันตกเป็นต้นกำเนิดของดนตรีแจ๊สมีลักษณะการสร้างสรรค์ แบบการด้นสดหรือการอิมโพรไวเซชั่น (Improvisation) เน้นที่จังหวะกลองและจังหวะที่ซับซ้อน นอกจากนี้ลักษณะที่เรียกว่า การโต้ตอบหรือ Call and Respond ซึ่งพบได้ในเพลงแจ๊ส โดยเฉพาะการร้องโต้ตอบของนักร้องเดี่ยวกับกลุ่มนักร้องประสานเสียง มีการใช้จังหวะขัด จังหวะตบที่สม่ำเสมอ และสีสันที่โดดเด่นรวมทั้งลักษณะเฉพาะของการบรรเลงดนตรี ที่มาของการเรียกว่า ดนตรีแจ๊ซ สันนิษฐานว่า เรียกตามวงดนตรีวงแรกที่นำแนวเพลงแจ๊ซมาสู่คนฟังหมู่มาก มีการบันทึกเสียงออกจำหน่ายอย่างเป็นทางการ คือ วง ดิ ออริจินัล ดิกซีแลนด์ แจ๊ซ (The Original Dixiel and Jazz Band หรือ วงโอดีเจบี) โดยเริ่มเรียกใช้กันราวปี ค.ศ.1917 ดนตรีแจ๊ซมีการพัฒนารูปแบบแตกต่างกันออกไป หลายประเภท เช่น แบบนิวออร์ลีน หรือดิกซีแลนด์ สวิง บีบ็อพ คูล ฟรีแจ๊ซ และ แจ๊ซร็อก เป็นต้น องค์ประกอบของดนตรีแจ๊ซ มีเครื่องดนตรีในกลุ่มเครื่องเป่า แซกโซโฟน และทรัมเป็ต เป็นหลักและมีเครื่องดนตรีในกลุ่มเครื่องประกอบจังหวะผสมวงบรรเลงร่วมกัน

#### 2.1.2.2 ดนตรีป็อป (Pop)

ดนตรีป็อบ (Pop music)หรือเพลงป็อป (pop song) เป็นแนวเพลงที่มีลักษณะทำนองเพลงง่าย ๆ โครงสร้างเพลงไม่สลับซับซ้อนจะรวมเพลงหลายๆประเภทเช่น ร็อก ฮิปฮอป เร้กเก้ แด๊นส์ อาร์แอนด์บี ฟังค์

#### 2.1.2.3 ดนตรีร็อค (Rock)

ดนตรีร็อค คือ ดนตรีอเมริกันประเภทหนึ่งซึ่งเริ่มได้รับความนิยมขึ้นมาในช่วงปลายปี ค.ศ. 1950 และต้นปี ค.ศ.1960 สิ่งที่สำคัญของดนตรีประเภทนี้คือ การใช้จังหวะที่หนักหน่วงและแบ่งอัตราส่วนของจังหวะออกเป็นโน๊ตเขบ็จ 1 ชั้น (Eight Note) จังหวะตกและจังหวะยกสลับต่อเนื่องกันไปทั้งเพลง และมักใช้บันไดเสียงประเภทโหมด (Mode) เนื้อหาของเพลงร็อคไม่ได้จำกัดอยู่ในเรื่อง ความรัก แต่อาจจะเป็นการเมือง ปัญหาสังคม หรือต่อต้านสังคม เช่น บ็อบ ดีแลน ดนตรีที่ใช้มักเป็นตัวหลักมักเป็น กีตาร์ไฟฟ้า

2.1.2.4 **ดนตรีคันทรีแอนด์เวสเทิร์น** (Country and Western)

ดนตรีคันทรี คือ ดนตรีพื้นบ้าน หรือ ดนตรีลูกทุ่ง ซึ่งเกิดจากเพลงบัลลาด (Ballad คือ เพลงที่เล่าเรื่องบุคคลสำคัญหรือเหตุการณ์สำคัญ) และการเต้นรำของชาวอังกฤษ โดยใช้เครื่องดนตรีจำพวกซอ (Fiddle) และดัลซิเมอร์ (Dulcimer) ดนตรีประเภทนี้รุ่งเรืองมากในทางตอนใต้ของอเมริกาแถบหุบเขาเวอร์จิเนีย แคโรไลนา เคนตักกี เทนเนสซี จนกลายเป็นรูปแบบของดนตรีอเมริกันที่แท้จริงที่เรียกว่า บทเพลงฮิลล์บิลลี (Hillbilly Music) จากนั้นก็ได้มีการผสมผสานแลกเปลี่ยนกับบทเพลงต่างๆ ในท้องถิ่นอื่น เช่น ดนตรี โพลกา (Polka) ของโบฮีเมียและเพลงบูลส์ของเท็กซัสจนได้ลักษณะที่แตกต่างออกมา เช่น เวสเทรินสวิง (Western Swing) เป็นดนตรีเต้นรำแบบคันทรีที่ประกอบด้วย เปียโน คลาริเน็ต แซ็กโซโน เครื่องเป่า กีตาร์ กลอง ตลอดจนไวโอลินเป็นตัวยืน อีกประเภทหนึ่งคือ บูลกราสศ์ ( Bluegrass) เกิดจากการผสมระหว่างรูปแบบบทเพลงฮิลล์บิลลีและเครื่องสายจากท้องถิ่นอื่น เช่น กีตาร์ แบนโจ และ แมนโดลิน

2.1.3 **การดึงค่าคุณลักษณะของเสียง** (Feature Extraction)

การดึงค่าคุณลักษณะ (Feature Extraction) เป็นขั้นตอนที่แปลงสัญญาณให้มีค่าคุณลักษณะที่สามารถนำมาจำแนก เพื่อการเตรียมพร้อมสำหรับขั้นตอนในการจำแนก การดึงค่าคุณลักษณะเป็นขั้นตอนที่สำคัญ เพราะเป็นขั้นตอนที่บ่งบอกถึงความแตกต่างของสัญญาณในขณะที่การใช้ค่าที่มีการดึงลักษณะเด่นได้น้อยจะทำให้ไม่สามารถจำแนกค่า เพื่อใช้ในการเรียนรู้หรือทำนายได้อย่างลำบาก ค่าที่ได้จากเสียงนั้นเป็นตัวแทนที่นำมาแทนข้อมูลเสียง เพื่อนำไปใช้สร้างแบบจำลอง และนำไปใช้ในการจำแนกหมวดหมู่ต่อไป ในวิทยานิพนธ์นี้ใช้ค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Features) ซึ่งเป็นค่าคุณลักษณะที่ใช้บ่งบอกความเป็นเอกลักษณ์ของเสียง

Timbral เป็นคุณภาพของเนื้อเสียง ลักษณะเสียงของตัวเอง ซึ่งจะขึ้นกับค่าระดับเสียงความดัง โดยเนื้อเสียงจะสามารถกล่าวถึงความสัมพันธ์ระหว่าง Timbre Quality และ Timbre Identity โดย Quality จะพูดในเรื่องความสามารถที่มนุษย์สามารถแยกแยะเสียงที่แตกต่างออกจากกันแม้ว่าจะมีแหล่งกำเนิดเดียวกัน ส่วน Timbre Identity เป็นเรื่องของความสามารถมนุษย์ที่แยกแยะความแตกต่างของเสียง 2 เสียงจากแหล่งกำเนิด 2 แหล่ง (Lerch, 2008)

Timbral เป็นคุณภาพของเสียงที่บ่งบอกถึงความแตกต่างระหว่างเสียง และความดังที่อยู่ในระดับเดียวกันโดยใช้เสียงจากเครื่องดนตรีที่แตกต่างกัน จากลักษณะทางกายภาพ timbre จะขึ้นอยู่กับพลังงานที่ถูกกระตุ้น รูปร่างของคลื่น ความดังของเสียง ที่ตั้งความถี่ของ Spectrum และลักษณะชั่วคราวของคลื่นที่ถูกกระตุ้น (Zhang and Kuo, 1998)

Timbral feature เป็นค่าคุณลักษณะที่มาตรฐานที่ใช้เสียงในเรื่อง Music Speech discrimination และเรื่อง Speech recognition จะใช้ Short Time Fourier Transform (STFT) ในทุกเฟรมของเสียง และใช้ STFT กับสัมประสิทธิ์เซปตรัมบนสเกลเมลซึ่งใช้ในการแยกเสียงเพลงกับเสียงมนุษย์ (Tzanetakis and Cook, 2002) นำเสนอ Timbral Feature ที่มีค่าคุณลักษณะดังนี้ Mel Frequency Cepstral Coefficients (MFCC), Zero Crossing Rate (ZC), Linear Predictive Coefficients (LPC), Spectral Centroid (SC), Spectral Flux (SF) และ Spectral Roll- Off (SR)

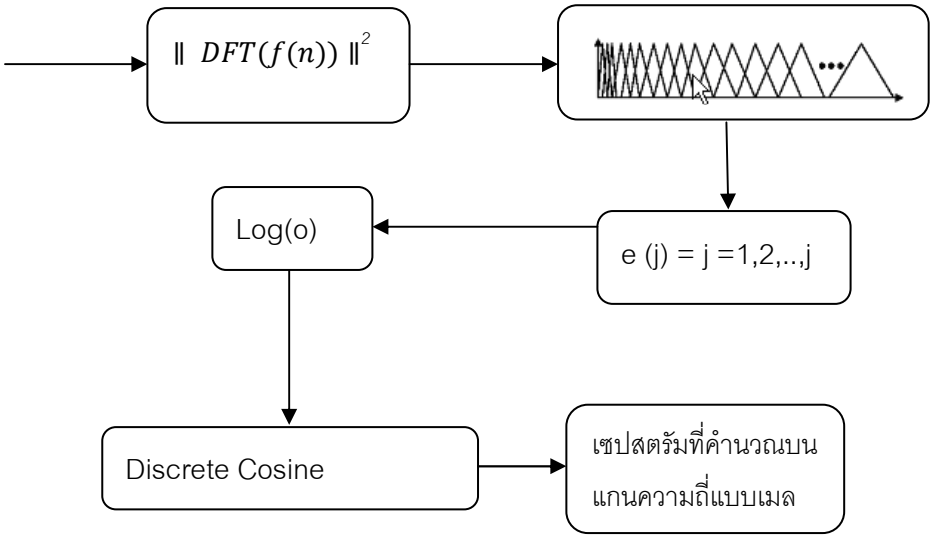
2.1.3.1 **สัมประสิทธิ์เซปตรัมบนสเกลเมล (Mel- Frequency Cestrum Coefficient, MFCC)**

สัมประสิทธิ์เซปตรัมบนสเกลเมล เป็นเทคนิคที่ปรับปรุงจากเซปสตรัมด้วยการปรับสเกลของสเปคตรัมให้อยู่บนสเกลที่เหมาะสมสำหรับการรับฟังของมนุษย์ เทคนิคสัมประสิทธิ์เซปสตรัมบนแกนความถี่เมล Karn (2543) เป็นลักษณะเด่นที่ปรับปรุงมาจากสัมประสิทธิ์เซปตรัมปกติด้วยการปรับสเกลของสเปกตรัมให้อยู่บนสเกลที่เหมาะสมสำหรับการรับฟังของมนุษย์ จึงมีการออกแบบสเกลของสเปกตรัมให้สามารถเก็บรายละเอียดของสัญญาณเสียงช่วงความถี่ต่ำได้มากกว่า เรียกว่า สเกลเมล (Mel scale) สูตรในการแปลงหน่วยจากเฮิรตซ์ไปเป็นหน่วยของสเกลเมล โดยสังเกตจากลักษณะของสัญญาณเสียงพูด เนื่องจากสัญญาณเสียงพูดในช่วงความถี่ต่ำจะมีความสำคัญมากกว่าช่วงความถี่สูง มี 3 ขั้นตอนดังนี้ ในขั้นแรกเป็นการสัญญาณเสียงเข้ามาจากภายนอก แล้วมาทำการแปลง Discrete Fourier Transform จากนั้นนำแอมพลิจูดที่ได้ไปยกกำลังสอง เพื่อให้กลายเป็น Power Spectrum ซึ่งจะได้สัญญาณในแกนความถี่ ขั้นต่อมานำไปคูณกับ Mel-Scale Filter Bank ซึ่งเป็นตัวกรองความถี่เพื่อทำการกรองสัญญาณที่มนุษย์ไม่ได้ยินออกไป หลังจากนั้นจะได้ค่าพลังงานจากตัวกรองแต่ละตัว (สามเหลี่ยมแต่ละลูก) ตั้งแต่ตัว ที่ 1 ถึง J แล้วนำค่าพลังงานที่ได้แต่ละตัวไปทำการผ่านฟังก์ชันลอการิทึม ซึ่งจะได้เป็นค่าเดซิเบลของแต่ละตัวออกมา จากนั้นนำไปผ่านฟังก์ชันการแปลงโคไซน์แบบไม่ต่อเนื่อง (Discrete Cosine Transform) ดังสมการด้านล่าง

$$MFCC[i][j] = \sqrt{\frac{2}{p}\sum_{j=1}^{r}\{(\log e[j][t])\cos(\frac{\pi i}{p}(j-0.5))\}} \qquad (2.1)$$

ภาพที่ 2.1

แสดงการทำงานของ MFCC




ที่มา: “Music Search Engine” โดย Songrit et al. (สืบค้น 5 ต.ค2553) จาก http://cpe.kmutt.ac.th/previousproject/2005/26/theory.html

**2.1.3.2 การประมาณเชิงเส้นหรือกระบวนการแอลพีซี (Linear Predictive Coefficient, LPC)**

บุญธิดา โฆษิตทรัพย์ (2547) กล่าวว่าได้มีประยุกต์ใช้การประมาณเชิงเส้นหรือกระบวนการแอลพีซี ในการวิเคราะห์หาพารามิเตอร์ต่างๆ ของสัญญาณสียง เช่น คาบเวลาพิทช์ ความถี่ฟอร์แมนท์ สเปกตรัม ฟังก์ชันของพื้นที่โวคัลแทร็ค (Vocal Tract Area Function) และการส่งสัญญาณเสียงด้วยอัตราความเร็วบิตต่ำ (Low Bit Rate Voice) แนวคิดพื้นฐานของกระบวนการ แอลพีซี ก็คือการสุ่มตัวอย่างของสัญญาณเสียงหลายๆตัวอย่างที่ผ่านมาในอดีต โดยการทำให้ค่าผลรวมกำลังของความแตกต่าง หรือค่าผิดพลาดระหว่างค่าตัวอย่างจริงและค่าที่ทำนายได้ ซึ่งก็คือค่าสัมประสิทธ์ให้น้ำหนัก ( Weighting Coefficient) ที่ใช้ในผลรวมเชิงเส้น

รูปแบบและวิธีการที่ใช้ในกระบวนการแอลพีซีนั้นมีอยู่หลายแบบด้วยกัน ความแตกต่างของแต่ละรูปแบบอยู่ที่การมองปัญหา หรืออีกนัยหนึ่งคือ วิธีการคำนวณเพื่อที่จะให้ได้มาซึ่งสัมประสิทธิ์ของตัวทำนายรูปแบบต่างๆดังนี้

1) วิธีโควาเรียนซ์ (Covariance Method)

2) รูปแบบอัตสหสัมพันธ์ (Autocorrelation Formulation)

3) วิธีการแลตติช (Lattice Method)

4) รูปแบบตัวกรองความถี่กลับ (Inverse Filter Formulation)

5) รูปแบบการประมาณเชิงสเปกตรัม (Spectral Estimation Formulation)

6) รูปแบบความน่าจะเป็นสูงสุด (Maximum Estimation Formulation)

7) รูปแบบผลคูณภายใน (Inner Product Formulation)

รูปแบบของกระบวนการแอลพีซีซึ่งเป็นกระบวนการที่นิยมใช้ คือ วิธีโควาเรียนซ์ (Covariance Method) และ รูปแบบอัตสหสัมพันธ์ (Autocorrelation Formulation) รูปแบบอัตสหสัมพันธ์ (Autocorrelation Formulation) ในการวิเคราะห์การทำงานแบบเชิงเส้น ช่วงสั้น จากการหาการวิเคราะห์แบบหน้าต่างสัญญาณเสียงและตัวอย่างสัญญาณภายนอกหน้าต่างนี้มีค่าเท่ากับศูนย์ แล้วจึงทำให้ค่าผิดพลาด (อานนท์ นามสนิท, 2549 ) ตามสมการที่ 2.1 ค่าต่ำสุดซึ่งสมการที่ 2 .1 เมื่อนำมาขยายผลจะได้ผลดังนี้

$$E = \sum_{n=-\infty}^{\infty} e_n^2 \qquad 2.2$$

$$E = \sum_{n=-\infty}^{\infty} s_n^2 - 2\sum_{k=-\infty}^{P} a_k \sum_{n=-\infty}^{\infty} S_n S_{n-1} + [\sum_{k=1}^{p} a_k s_{n-k}]^2 \qquad 2.3$$

การหาค่า $a_k$ ที่ทำให้ E มีค่าต่ำสุดสามารถกระทำได้โดยการหาในสมการที่ 2.3 ของแต่ละค่า i = 1,2,3,…,p เมื่อ k = 1,2,3,…,p ได้เป็นสมการเชิงเส้น p สมการแต่ละสมการมี k ค่า ซึ่งก็คือ $a_k$ ที่ไม่ทราบค่า *p* ตัวดังนี้

$$\frac{\partial E}{\partial a_i} = 0 = -2\sum_{n=-\infty}^{\infty} S_n S_{n-1} + 2\sum_{n=-\infty}^{\infty} [\sum_{k=1}^{p} a_k s_{n-k}] S_{n-1} \qquad 2.4$$

$$\sum_{n=-\infty}^{\infty} S_{n-1} S_n = \sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} S_{n-i} S_{n-k} \qquad 2.5$$

เนื่องจากพจน์แรกเป็นอัตสหสัมพันธ์ $r_i$ ของ $S_n$ ที่มีความยาวจำกัดสามารถเขียนสมการที่ 2.5 ใหม่ได้เป็นสมการที่ 2.6

$$\sum_{k=1}^{p} r_{|r-k|} = -r_j \quad \text{เมื่อ} \quad 1 \le i \le p \qquad 2.6$$

เมื่อแก้สมการที่ 2.4 ซึ่งประกอบด้วยสมการเชิงเส้น *P* สมการได้เป็นค่าสัมประสิทธิ์ $a_k$ ที่ต้องการออกมาโดย เขียนสมการให้อยู่ในรูปเมตริกซ์ $R_a = -r$

$$R = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ \vdots & & & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix} \qquad 2.7$$

$$R = [a_1, a_2, a_3, \ldots a_p] \qquad 2.8$$

$$R = [r_1, r_2, r_3, \ldots r_p]^t \qquad 2.9$$

จากสมการที่ *2.7* เมตริกซ์ *R* เรียกว่า เมตริกซ์อัตสหสัมพันธ์ (Autocorrelation Method) ซึ่งมีโครงสร้างแบบโทพสิทธ์ (Toeplitz) เป็นเมตริกซ์แบบสมมาตร คือมีค่าในแนวทแยงมุมเท่ากันทั้งหมดทุกแถวสำหรับวิธีที่ช่วยหาผลเฉลยของสมการยูล – วอล์กเกอร์ ในสมการที่ *2.2* และหาสัมประสิทธิ์การทำนายแบบเชิงเส้นในสมการที่ *2.3* มีหลายวิธีเช่นวิธีเลวินสัน – เดอร์บิน (Levinson – Durbin Algorithm) และ วิธีเชอร์ (Schur Algorithm)

**2.1.3.3 อัตราค่าตัดศูนย์** (Zero Crossing Rate,ZC)

จำนวนครั้งของสัญญาณเสียงที่ตัดแกนศูนย์ในช่วงเวลาใดๆ ภายในหน้าต่างที่กำหนด ปกติสัญญาณเสียงที่มีค่าจุดตัดสูงจะเป็นเสียงไม่ก้องและสัญญาณเสียงที่มีค่าจุดตัดต่ำจะเป็นเสียงก้อง โดยทั่วไปแล้วจะใช้ในการนับจำนวนครั้งที่สัญญาณมีการเปลี่ยนเครื่องหมายจากบวกเป็นลบ หรือจากลบเป็นบวก นอกจากนั้นเพื่อลบผลของสัญญาณลบกวน การนับจำนวนครั้งจะมีการพิจารณา การเปลี่ยนแปลงขนาดสัญญาณดนตรีควบคู่ไปด้วย ซึ่งจำนวนครั้งของการตัดผ่าน (ชัย วุฒิวิวัฒน์ชัย, 2540) ซึ่งจะสามารถคำนวณได้จากสมการ

$$Z(i) = \frac{1}{2N}\sum_{n=0}^{N-1}|\,\mathrm{sgn}[x_i(n)] - \mathrm{sgn}[x_i(n-1)]\,| \qquad 2.10$$

โดย

$$\mathrm{sgn}[x_i(n) = \begin{cases} 1, x_i(n) \geq 0 \\ 0, x_i(n) < 0 \end{cases}$$

$x_i\ (n)$ เป็นสัญญาณที่ i ใน โดเมนเวลา, $sgn$ เป็น 1 สำหรับ อะกิวเม็นท์ที่เป็นบวก และเป็น 0 สำหรับอะกิวเม็นท์ที่เป็นลบ

Zhang and Kuo (1998) ได้อธิบายว่า รูปร่างของคลื่น ความดังของเสียง ที่ตั้งความถี่ของ Spectrum และลักษณะชั่วคราวของคลื่นที่ถูกกระตุ้น (Spectral Shape Feature) ซึ่งแบ่งออกได้อีกสี่คุณลักษณะคือ ค่าคุณลักษณะที่แสดงถึงค่ากึ่งกลางของ Spectrum (Spectral Centroid) *ค่าภายใต้* 85 %ของการกระจายค่าความถี่ใช้วัดค่ารูปร่างที่เปลียนแปลง (Spectral Roll off) และค่าความแตกต่างความกว้างคลื่นเสียงของแต่ละเฟรม (Spectral Flux) คุณลักษณะเหล่านี้นำมาใช้ในการจำแนกลักษณะสำคัญของสัญญาณเสียงดนตรี

**2.1.3.4 สเปกทรอลเซนทรอยด์** (Spectral Centriod,SC)

คือค่าคุณลักษณะที่แสดงถึงค่ากึ่งกลางของ *Spectrum* โดยหาได้จากสมการ 2.11

$$Centriod = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \qquad 2.11$$

เมื่อ $x(n)$ คือขนาดของ *bin* จำนวน *n*

และ $f(n)$ คือค่ากึ่งของความถี่ของ *bin*

**2.1.3.5 สเปกทรอล ฟลักซ์** (Spectral Flux,SF)

เป็นค่าที่คำนวณความแตกต่างระหว่างค่าปัจจุบันของแต่ละ Spectrum กับค่าที่สอดคล้องกับ Spectrum ที่สอดคล้องกับค่าก่อนโดยแต่ละค่าที่แตกต่างจะถอดด้วยรากที่ของผลรวมจากการยกกำลังสองดังแสดงในสมการที่ 2.12

$$F_t = \sum_{i=1}^{N} (N_t[n] - N_{t-1}[n])^2 \qquad 2.12$$

$N_t[n] - N_{t-1}[n]$ เป็นค่าที่ถูกทำให้เป็น Normalize ณ Fram t และ $N_{t-1}$ เป็นค่า ณ Fram t ก่อนหน้า

**2.1.3.6 สเปกทรอล โรล ออฟ** (Spectral Roll off,SR)

สเปกทรอล โรล ออฟเป็นค่าของความถี่ $\mathbf{R_t}$ ภายใต้ 85 %ของการกระจายค่าความถี่ใช้วัดค่ารูปร่างที่เปลียนแปลง

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^{N} M_t[n] \qquad 2.13$$

$M_t$ เป็นค่าพลังงานเสียงที่ได้การผ่าน Fourier Transform ณ เฟรม t

$n$ เป็นค่าที่ได้จากการทำ normalized magnitude

**2.1.3.7 สเตรท ออฟ สตรอง บีท** (Strength of Strongest Beat,SSB)

เป็นค่าที่ได้ค่า Beat Histogram แสดงกำลังคาบจังหวะที่แตกต่างของสัญญาณ โดยคำนวณจากการใช้ RMS ของ 256 ค่าจากนั้นใช้ Fast Fourier Transform ซึ่งได้เป็นค่า Beat Histogram ส่วน Strength of Strongest Beat นำค่าสัญญาณ จาก Beat Histogram โดยใช้ค่าที่สอดคล้องกับ Strongest Beat ค่าของผลรวมที่ถูกแบ่งจากสัญญาณ Beat Histogram ที่ผ่านเข้ามา

$$RMS = \sqrt{\frac{\sum_{i=1}^{i=32} (M[i]^2)}{32}} \qquad 2.14$$

$M$ เป็นค่าพลังงานเสียงที่ได้การผ่าน Fourier Transform

## 2.1.4 ทฤษฎีในศาสตร์ปัญญาประดิษฐ์

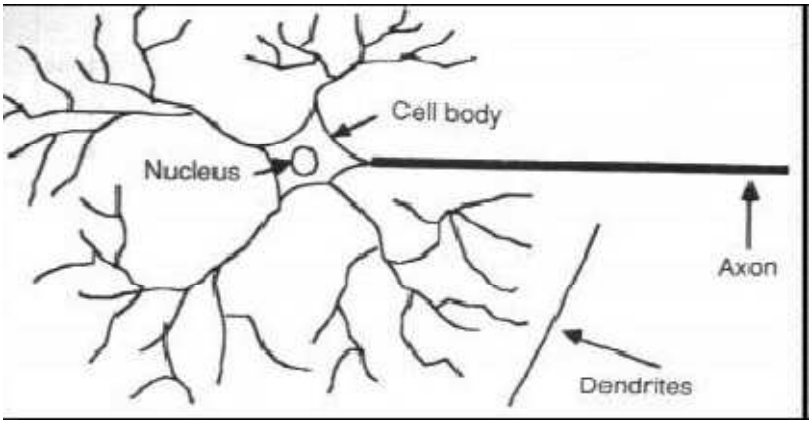
ปัญญาประดิษฐ์เป็นวิชาที่ว่าด้วยการศึกษาเพื่อให้เข้าใจถึงความฉลาดและสร้างระบบคอมพิวเตอร์ที่ชาญฉลาด และนำมาทำงานแทนหรือช่วยมนุษย์ทำงานที่ต้องใช้ความฉลาดนั้นๆ" (บุญเสริม กิจศิริกุล, 2546) ในส่วนของโครงข่ายประสาทเทียม (Artificial Neural Networks) มีพื้นฐานมาจากการจำลองการทำงานของสมองมนุษย์ ด้วยโปรแกรมคอมพิวเตอร์ จุดมุ่งหมายของโครงข่ายประสาทเทียมคือต้องการให้คอมพิวเตอร์มีความชาญฉลาดในการเรียนรู้เหมือนที่ มนุษย์มีการเรียนรู้ สามารถฝึกฝนได้ และสามารถนำความรู้และทักษะ รวมทั้งสามารถ

นำไปประยุกต์ใช้ได้ดีกับปัญหา Classification, Regression และ Clustering เทคนิคนี้ มักถูกเรียกว่า “Black Box” (ภรัณยา อำมฤครัตน์ และคณะ, 2009)

**2.1.4.1 โครงข่ายประสาทเทียม (Artificial Neural Network,ANN)**

โครงข่ายประสาทเทียม คือ โมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณ เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้ การจดจำแบบรูป (Pattern Recognition) และการอุปมานความรู้ (Knowledge Deduction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (Bioelectric Network) ในสมองซึ่งประกอบด้วย เซลล์ประสาท หรือนิวรอนซึ่งประกอบด้วย เซลล์ประสาท หรือ นิวรอน (Neurons) และ จุดประสานประสาท (Synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า "เดนไดรท์" (Dendrite) ซึ่งเป็น Input และปลายในการส่งกระแสประสาทเรียกว่า "แอคซอน" (Axon) ซึ่งเป็นเหมือน Output ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทางแอคซอนของมัน

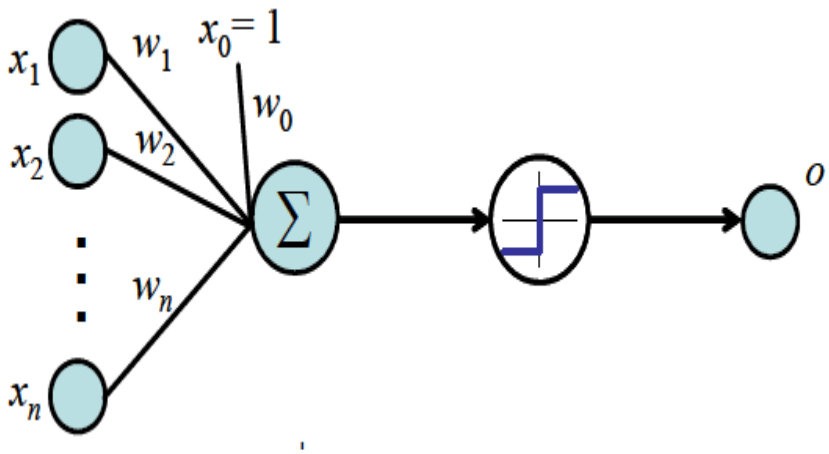
ภาพที่ 2.2
แสดงส่วนประกอบของเซลล์ประสาท




ที่มา: Christos Stergiou , Dimitrios Siganos. Neural Network. Available (สืบค้น 23 ตุลาคม 2553) จาก http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

Perceptron เป็นโครงข่าย อย่างง่ายเพียงหน่วยเดียวที่จำลองแบบเซลล์ประสาท ซึ่ง Input จะเป็น Vector เป็นจำนวนจริง แล้วคำนวณผลรวมเชิงเส้นแบบถ่วงน้ำหนัก ถ้าผลรวมที่ได้มีค่ามากกว่า (θ) และ เป็น -1 ค่าจะไม่เกินขอบเขต Function กระตุ้น Activation Function ซึ่งจะแสดง Output เป็น -1 กับ 1 ส่วน Function กระตุ้นอาจจะแสดงในรูป 0 และ 1 (บุญเสริม กิจศิริกุล, 2546)

ภาพที่ 2.3
เป็นโครงข่าย อย่างง่ายเพียงหน่วยเดียว



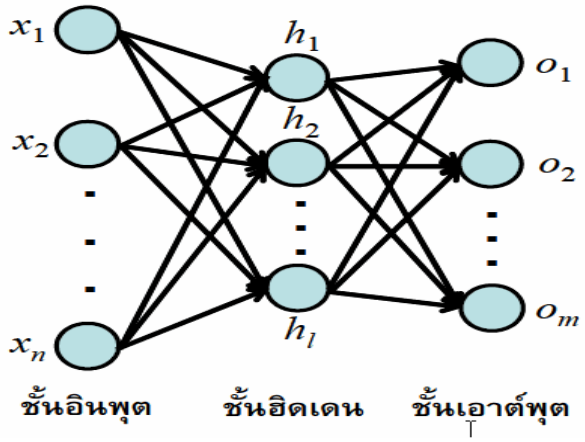

ที่มา: “ปัญญาประดิษฐ์,”โดย บุญเสริม กิจศิริกุล, 2548, วิศวกรรมคอมพิวเตอร์, จุฬาลงกรณ์มหาวิทยาลัย

ซึ่งสามารถสรุปได้ว่ากลุ่มตัวอย่างบางตัวอย่าง สามารถแยกได้เป็นตัวอย่างบวกและลบ จากระนาบที่แตกต่างกัน ซึ่งสามารถเรียกตัวอย่างที่เกิดขึ้นแบบเชิงเส้น(Linearly Separable Sets of Examples) แต่ด้วยลักษณะปัญหาบางปัญหาซึ่งไม่สามารถแก้ปัญหาด้วย Perceptron เดี่ยวได้การนำ Perceptron มาเชื่อมกันหลายชั้นเพื่อจำแนกค่าเราเรียกโครงข่ายประสาทเทียมหลายชั้น (Multilayer Neural Network ) สถาปัตยกรรมของโครงข่ายประสาทเทียมคือ การเชื่อมต่อกับโครงข่ายประสาทเทียมหนึ่งเชื่อมกับอีกโครงข่ายประสาทเทียมอีกกลุ่มหนึ่งคล้ายกับการเชื่อมกันของเซลล์ประสาทโดยลักษณะการเชื่อมมีหลายรูปแบบและมีลักษณะเพื่อที่จะเหมาะกับการใช้งานโดยลักษณะทั่วไปก็คือ Input Layer Hidden Layer และ Output

Layer จากโครงข่ายป้อนไปยังข้างหน้าหลายชั้น (Multilayer Feed Forward Network) ซึ่งประกอบด้วย ซึ่งประกอบด้วย 3 ชั้น Input ชั้น Hidden ชั้น Output ซึ่งแต่ละชั้นจะเชื่อมกันหมด (Fully Connection) ซึ่งจะเชื่อมต่อกันไปโดยชั้น Input จะเชื่อมกับชั้น Hidden และชั้น Hidden จะเชื่อมกับชั้น Output โดยจะเชื่อมแบบไปข้างหน้าอย่างเดียวไม่มีเส้นเชื่อมจากชั้น Output กลับมายังชั้น Hidden และ Input

ข่ายงานหลายชั้นและการแพร่กระจายย้อนกลับ (Back Propagation Algorithm) เป็นโครงข่ายประสาทเทียมที่อาศัยค่าการเรียนรู้เวกเตอร์น้ำหนักโดยป้อนข้อมูลหลายชั้น (Multilayer Feed Forward Network) โดยที่เคลื่อนที่ตามความชันเพื่อหาค่าต่ำสุดระหว่าง Output ของข่ายงานและ Output เป้าหมายที่ผิดพลาดของทั้งสองค่า ลักษณะการทำงานของโครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับ จะมีความแตกต่างจาก Multilayer Feed Forward Network คือสามารถทำงานกับ Function ที่มีรูปแบบไม่เป็นเชิงเส้นได้ (Nonlinear Function) โดยใช้หน่วยของ ซิกมอยด์ (Sigmoid Unit) โดยลักษณะการทำงานของซิกมอยด์จะใช้ผลคูณของเวกเตอร์อินพุตกับค่าน้ำหนักเส้นเชื่อม Input Cell ชั้นถัดไป หลังจากผ่านจาก (Threshold) ค่าของหน่วย Sigmoid ออกมาเป็นฟันก์ชั่นต่อเนื่องที่มีค่าใกล้กับ 0,1 ซึ่งในส่วนของ Perceptron จะให้ค่าออกมาไม่ต่อเนื่อง

ภาพที่ 2.4
ข่ายงานหลายชั้นและการแพร่กระจายย้อนกลับ




ที่มา: “ปัญญาประดิษฐ์,”โดย บุญเสริม กิจศิริกุล, 2548, วิศวกรรมคอมพิวเตอร์, จุฬาลงกรณ์มหาวิทยาลัย

โครงข่ายประสาทเทียมมีหลากหลายแบบทั้งเป็นการฝึกสอนแบบควบคุม (Supervised Learning) และการฝึกสอนแบบอิสระ (Unsupervised Learning) ในมุ่งเน้นการรู้จำแบบ การฝึกสอนแบบควบคุม โครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron :MLP) โดยใช้วิธีการเรียนรู้แบบแพร่กระจายกลับ (Back-Propagation Learning Algorithm) "วิธีการนี้มีข้อดีคือเป็นวิธีพื้นฐานมีการคำนวณที่ไม่ซับซ้อน และสามารถรู้จำได้ดีไม่ด้อยกว่าวิธีอื่นๆ"

**2.1.4.2 ขั้นตอนการฝึกฝน (Training หรือ Learning)**

โครงข่ายประสาทเทียมแบบหลายชั้นโดยอาศัยวิธีการฝึกฝนแบบการเรียนรู้แบบแพร่กระจายกลับ จะเรียนรู้จากแต่ละตัวอย่างเวกเตอร์ข้อมูลเข้า (i) และเวกเตอร์ข้อมูลออกที่ต้องการ (t) ในชุดฝึกฝน (Training Set) มีขั้นตอนในการฝึกฝนดังนี้

ป้อนค่าเวกเตอร์ข้อมูลเข้า ให้กับระดับชั้นข้อมูลเข้า

แพร่กระจายไปข้างหน้า (Feed Forward) ค่าข้อมูลเข้าไปยังแต่ละระดับชั้นจนกระทั่งได้ค่าเวกเตอร์ข้อมูลออก ภาพที่ 2.5 แสดงรายละเอียดของโหนด j ซึ่งอยู่ในระดับชั้นใดๆ (ระดับชั้นที่ h) ในโครงข่ายประสาทเทียมแบบ MLP ค่าข้อมูลออกของโหนด j($O_j$) สามารถคำนวณได้จากสมการ

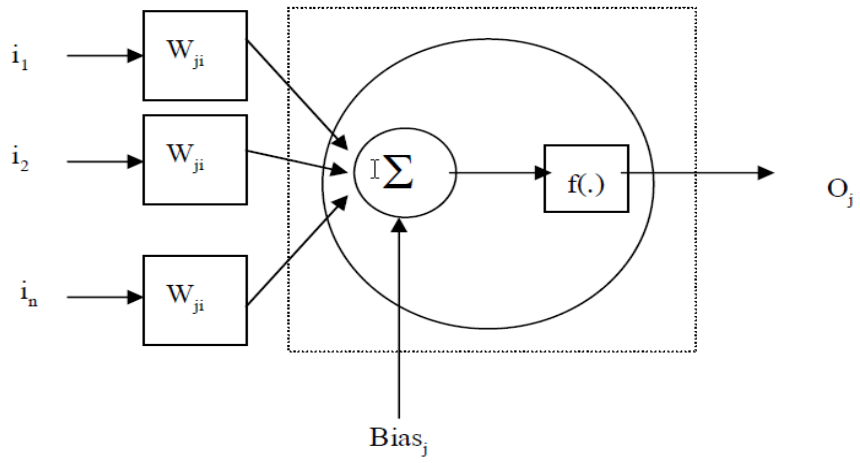
$$O_j = f(net_j) \qquad 2.14$$

โดยที่ $net_j = \sum W_{ji} O_i + bias$ 2.15

เมื่อ $o_i$ เป็นค่าของข้อมูลออกของโหนด i ซึ่งโหนดใดๆ ที่อยู่ในระดับชั้นก่อนหน้าโหนด *j* นี้ (ระดับชั้นที่ h-1) และจะเป็นค่าของข้อมูลเข้า $i_i$ ถ้าโหนดที่ j เป็นโหนดในระดับชั้นข้อมูลเข้า $W_{ij}$ เป็นค่าน้ำหนักการเชื่อมต่อจากโหนด i ไปยังโหนด j $bias_j$ เป็นค่าที่ใช้ปรับ $net_j$ มีค่าไม่เป็นศูนย์ แม้ว่าค่า $O_j$ จะมีค่าเป็นศูนย์หมด $f(net_j)$เป็นค่าของฟังก์ชันกระตุ้นแบบไม่เป็นเชิงเส้น (Nonlinear Activation Function) ซึ่งต้องเป็นฟังก์ชันเพิ่ม และสามารถหาอนุพันธ์ได้ ในวิทยานิพนธ์นี้เลือกใช้ฟังก์ชัน Tansig ดังภาพที่ 2.5 แสดงรูปแบบ Neural Network อย่างง่ายของ Herve Abdei

ภาพที่ 2.5

แสดง รูปแบบอย่างง่าย Herve Abdei 1994




ที่มา” Neural Network, “ The University of Texas,Dallas,1994

ปรับน้ำหนักการเชื่อมต่อระหว่างโหนดใดๆ ด้วยวิธีการเรียนรู้แบบแพร่กระจายกลับ โดยบวกเพิ่มค่า $w_{ji}$ ด้วยค่า $\Delta w_{ji}$ ซึ่งคำนวณได้จากสมการ

$$\Delta wji(n) = -\varepsilon(\partial \in) / \partial \omega_{ij} + \alpha \Delta wji(n-1) \qquad 2.16$$

โดยที่ $\varepsilon$ คือค่า Learning Rate $\alpha$ คือค่า Momentum Rate ซึ่งเป็นค่าคงที่ทั้งคู่ พจน์ที่สองทางขวาของสมการที่ (2.16) เป็นพจน์ของโมเมนตัมซึ่งช่วยให้การปรับค่า $_{ji}w$ ไม่เกิดการแกว่ง(oscillation) มากเกินไป n เป็นตัวเลขแสดงลำดับครั้งของการปรับ โดยขณะเริ่มต้นจะตั้งค่า $\Delta w_{ij}$เป็น $E$ เป็นค่าความผิดพลาดของค่าข้อมูลออก เมื่อเทียบกับค่าข้อมูลออกที่ต้องการ ดังนี้

$$E = \frac{1}{2}\sum{}_j (t_j - o_j)^2 \qquad 2.17$$

โดย $t_j$ และ $o_j$ เป็นค่าข้อมูลออก และค่าข้อมูลออกที่ต้องการของโหนดในระดับชั้นข้อมูลออก

j=1,2,…,c โดยที่ c เป็นจำนวนโหนดในระดับชั้นข้อมูลออก ดังนั้นจะได้ว่า

$$\frac{\partial x}{\partial w_{ij}} = \frac{\partial x}{\partial o_{ij}} - \frac{d_{o_j}}{dnet_j} - \frac{\partial_{net_j}}{\partial w_{ji}} \qquad 2.18$$

ถ้าโหนด j อยู่ในระดับชั้นข้อมูลออก จากสมการที่ (2-39) จะได้

$$\frac{\partial E}{\partial o_j} = \sum_j (t_j - o_j) \qquad 2.19$$

สำหรับโหนด *j* ที่อยู่ในระดับชั้นใดๆ (ระดับชั้นที่ h) นอกเหนือจากระดับชั้นข้อมูลเข้า

$$\frac{\partial E}{\partial o_j} = \sum_i \frac{\partial E}{\partial o_j} - \frac{d_{o_i}}{dnet_i} - \frac{\partial_{net_j}}{\partial_{o_i}} = \sum_{\tilde{A}} \frac{\partial E}{\partial o_j} \frac{do_i}{\partial o_i} \frac{d_{o_i}}{d_{net_i}} w_i \qquad 2.20$$

ส่วน $\frac{d_{o_j}}{d_{net_j}}$ และ $\frac{\partial_{net_j}}{\partial_{\omega_{ji}}}$ ในสมการที่ (2.16) 2.21

$$\frac{d_{o_j}}{d_{net_j}} = f(net_j) = o_j(1 - o_j) \qquad 2.22$$

$$\frac{\partial_{net_j}}{\partial_{\omega_{ji}}} = o_j \qquad 2.23$$

โดยที่โหนด *i* อยู่ในระดับชั้นที่ถัดจากระดับชั้นของโหนด *j* ไปอีกหนึ่งชั้น (ระดับชั้นที่ h+1)จะเห็นได้ว่า การปรับน้ำหนักการเชื่อมต่อ $_{ij}w$ อาศัยวิธีการเรียนรู้แบบแพร่กระจายกลับของค่าความผิดพลาดของค่าข้อมูลออกของแต่ละโหนดในโครงข่ายประสาทเทียม

เมื่อปรับน้ำหนักการเชื่อมต่อหมดทุกค่าแล้ว จะทำการแพร่กระจายค่าเวกเตอร์ข้อมูลเข้าอีกครั้งโดยใช้ค่าน้ำหนักการเชื่อมต่อค่าใหม่ที่ได้จากการปรับ กระทำเช่นนี้ต่อไปเรื่อยๆ จนกระทั่งค่าความผิดพลาด (*E*) น้อยกว่าค่าความผิดพลาดที่ยอมรับได้ (Error Threshold) เป็นอันว่าสิ้นสุดการฝึกฝนโครงข่ายประสาทเทียม และจะเก็บค่าน้ำหนักการเชื่อมต่อสุดท้ายไว้สำหรับการทดสอบการรู้จำ สรุปขั้นตอนวิธีการเรียนรู้แบบแพร่กระจายกลับเป็นโครงสร้าง

**2.1.4.3 ขั้นตอนทดสอบการรู้จำ** (Recognition Test)

1. ป้อนค่าเวกเตอร์ข้อมูลของชุดทดสอบ เข้าไปยังระดับชั้นข้อมูลเข้า

2. แพร่กระจายค่าข้อมูลเข้า เข้าไปในโครงข่ายประสาทเทียมทำนองเดียวกับการฝึกฝนโดยอาศัยค่าน้ำหนักการเชื่อมต่อที่ได้จากการฝึกฝน จนกระทั่งได้ค่าเวกเตอร์ข้อมูลออกที่ระดับข้อมูลออก

3. ทำการตัดสินใจ โดยใช้กฎเกณฑ์การตัดสินใจเนื่องจากโครงข่ายประสาทเทียมแบบหลายชั้น ที่ผ่านการฝึกฝนมาแล้ว จะให้ค่าข้อมูลออกที่คล้ายคลึงกับค่าข้อมูลออกที่ต้องการซึ่งคู่กับตัวอย่างข้อมูลเข้าในชุดฝึกฝน และโดยปกติจะกำหนดค่าข้อมูลออกที่ต้องการให้มีค่าสูงสุดที่โหนดที่สอดคล้องกับเวกเตอร์ข้อมูลเข้าที่ฝึกฝนในขณะนั้น ดังนั้นเกณฑ์การตัดสินใจที่เลือกใช้คือการเลือกรูปแบบ (Pattern) ที่ตรงกับโหนดในระดับชั้นข้อมูลออก ที่ให้ค่าข้อมูลออกมากที่สุด กฎเกณฑ์การตัดสินใจแบบนี้ เรียกว่าวิธี Maximum Likelihood (ML) ซึ่งเขียนเป็นสมการได้ดังนี้

$$Unknown\ Pattern = O_i \qquad 2.24$$

เมื่อ$O_i=\max(O_1,O_2,\ldots,O_c)$

โดยที่ c คือจำนวนโหนดที่ระดับชั้นข้อมูลออก ซึ่งเท่ากับจำนวนคำศัพท์ที่รู้จำ

## 2.2 วิทยานิพนธ์ที่เกี่ยวข้อง

ในส่วนของวิทยานิพนธ์และบทความที่เกี่ยวข้องกับวิทยานิพนธ์ฉบับนี้ ประกอบไปด้วยวิทยานิพนธ์และบทความการจำแนกหมวดหมู่ดนตรีด้วยวิธีใหม่และการเลือกดึงคุณลักษณะ โดยจะทำการรวบรวมและศึกษาวิทยานิพนธ์เพื่อรวบรวมและนำเอาเทคนิคในการจำแนกเสียงมาประยุกต์หรืออ้างอิงต่อไป

ในส่วนแรกเป็นส่วนจำแนกหมวดหมู่ดนตรีด้วยวิธีใหม่จากบทความต่างประเทศซึ่งบทความแรกได้จำแนกดนตรีพื้นบ้านจีน โดย Liu, Yang ,and Chen (2008) นำเสนอระบบจำแนกดนตรีพื้นบ้านจีนอัติโนมัติ Continuous Hidden Markov Model (CHMM) ที่ใช้อยู่ทั่วไปเพื่อใช้ในเป็นแบบจำลองการประเมิน สัญญาณดนตรี ในทางตรงกันข้าม ข้อดีข้อเสียที่เห็นได้ชัดเวลาที่ใช้ ในละ State ยังไม่ดีเท่าที่ควรใน Paper นี้นำเสนอ 2 วิธี Hidden Semi Markov Model (HSMM) และ Segmentation Duration Based HMM(SDBHMM) ทำการเปรียบเทียบกับ Continuous Hidden Markov Model (CHMM) ผลที่ได้คือ SDBHMM ให้ความถูกต้อง 92.49% และ HSMM 90.02 % ให้ผลดีกว่า CHMM โดย Feature ที่ใช้ Timbre Feature ( MFCC,ZC,SR

LPC,SF และ SC) ร่วมกับ Rhythmic Content Feature ซึ่งใช้ในการตรวจจับ Salient Periodicties ของสัญญาณ Rhythmic Content Feature จะใช้ในการคำนวณ Beat Histogram ซึ่งจะใช้ค่า Mean Standard Deviation Mean of the Derivation และ Standard Deviation of The Derivation ในการคำนวณทุก frame

ต่อมา Wang , Huang ,Wang ,Liang และ Xu (2008) ได้อธิบายถึงความต้องการความถูกต้องของระบบยังไม่เป็นที่น่าพอใจ จึงได้เสนอวิธีการลดความผิดพลาดโดยใช้วิธี Multiple Classifier Fusion โดยอย่างแรกจะใช้ MFCC ร่วมกับ 4 Feature จาก Mpeg- 7 Audio Descriptor ซึ่งจะถูกตัดจากทุกๆเฟรมหลังจากนั้นจะรวมเฟรมเข้าด้วยกัน และใช้ค่า Mean และค่า Variance ของ Short Time Frame มาคำนวณ หลังจากนั้นนำแต่ละส่วนมา Traning และทดสอบโดยใช้ Random Forest (RF) และ โครงข่ายประสาทเทียมแบบหลายชั้นจำแนกอย่างอิสระและน้ำหนักตามแบบ Fusion และผลจากการจำแนกในแต่ละแบบ จะนำมาเป็นการตัดสินใจในการเลือกจากความถี่สูงสุดจากแต่ละส่วน ผลจากการทดลองพบว่า Fusion จะทำให้ผลจากความผิดพลาดลดลง 12.4% จากการเปรียบเทียบจากระบบ

Xu , Wang และ Yan (2008) ลักษณะเฉพาะพื้นฐานของเพลงพื้นบ้านจีนก็คือ เพลงพื้นบ้าน เนื่องจากมีลักษณะที่โดดเด่น การจำแนกเพลงพื้นบ้านของจีน จะตามลักษณะ ภูมิศาสตร์เป็นส่วนสำคัญ ซึ่งจะเห็นได้ว่ามีคนจำนวนมากที่ชื่นชอบเพลงพื้นบ้านจีน การค้นหาบทเพลงจึงมีส่วนสำคัญทั้งในด้านวิชาการและอุตสาหกรรม ในการทดลองนี้จะนำเสนอ การจำแนกเพลงพื้นบ้านจีนโดยอัติโนมัติโดยใช้วิธี Contribution Ratio Based Selection ในการจำแนกและใช้ RBF Neural Network . จากการทดลอง โดยใช้ 74 ค่าคุณลักษณะ จาก 517 เพลง 10 ประเภท ผลแสดงให้เห็นว่า ค่าเฉลี่ยกำลังสองของ SF และ LPC มีประสิทธิภาพมากที่สุดในการจำแนกเพลงพื้นบ้านจีน 10 ประเภทและจะดีขึ้นเมื่อลดจำนวนลักษณะลงเป็น 30 ค่าคุณลักษณะ

การเลือกดึงคุณลักษณะเพลงในส่วนนี้มีทั้งวิทยานิพนธ์ภายในประเทศและนอกประเทศซึ่งในส่วนนี้จะได้รับความสนใจจากผู้วิจัยเป็นอย่างมาก Doraisamy และคณะ (2008) ได้ทำการศึกษาเพื่อเลือกค่าคุณลักษณะและเทคนิคในการจำแนกหมวดหมู่เพลงพื้นบ้านประเทศมาเลเซียซึ่งแบ่งออกเป็น 10 หมวดหมู่คือ Dikir,Bart,Etnik,Sabah,Gamelan,Inang,Joget,Keroncong,Tumbuk kalang,Wayang,Kulit และ Zupin โดยเลือกใช้ค่าคุณลักษณะแบบ STFT MFCC และ Beat ร่วมกัน ในส่วนของการจำแนกหมวดหมู่เพลงพื้นบ้านมาเลเซีย ได้ทำการทดลองเปรียบเทียบวิธีการ 18 วิธีการ ที่มีอยู่ใน Weka ซึ่งเป็นโปรแกรมสำหรับ Mechine Learning และ Data Mining ดังนี้ AIRS (Artificial

Immune System : AIS), Bagging Bayesian Network,Cart Conjunctive Rule Learner (Conj-Rules),Decision Stump Decision Table IB1 J48(C4.5), Kstar Logistic LogitBoost,bn Muti-layer Neural Network with Back Propagation (MLP),Naive Bayesian NBtree PART RBF Network และ Support Vector Machine โดยผลที่จำแนกความถูกต้องดีที่สุดคือ โครงข่ายประสาทเทียมแบบหลายชั้นโดยให้ค่า 88.6 %

อานนท์ นามสนิท (2548) ได้ทำการศึกษาวิจัยการจำแนกกลุ่มเพลงไทยโดยใช้ Support Vector Machine ไว้ดังนี้ เพลงไทยลูกทุ่ง เพลงไทยลูกกรุง และเพลงไทยเดิม โดยเลือกมาเฉพาะแทร็คดนตรีประเภทละ 25 เพลง โดยในแต่ละเพลงนั้นถูกแบ่งเป็นชุดข้อมูลในการเรียนรู้ 15 เพลง และชุดข้อมูลสำหรับการทดสอบ 10 เพลง โดยเพลงที่ใช้ในการทดสอบได้ทำการตัดออกเป็นเพลงละ 2000 เฟรม และใช้ค่าคุณลักษณะแบบ Beat Spectrum, LPC, ZC, MFCC Spectrum Power และ Line Spectral Frequency จากนั้นนำมาทดสอบด้วย Support Vector Machine โดยใช้ เคอร์เนลเรเดียลเบลิสฟังก์ชั่น (Radial Basis Function) ผลที่ได้จำแนกได้ถูกต้อง 83 %

โอภาส แก้วต่าย (2552) ได้นำเสนอการศึกษาการเปรียบเทียบเพลงไทยเดิม 3 หมวดหมู่ คือ มโหรีย์ เครื่องสาย และปี่พาทย์ โดยใช้ค่าคุณลักษณะ Timbrel Feature คือ LPC, SC, ZC, MFCC ร่วมกับ Power Spectrum มาจำแนกหมวดหมู่เพลงไทยเดิมโดยใช้ K- Nearest Neighbor และ Support Vector Machine และ Artificial Neural Network พบว่า Support Vector Machine ให้ความถูกต้องในการจำแนกสูงสุดคือ 86.97 %

ตารางที่ 2.1
แสดงค่าคุณลักษณะที่ได้จากการทบทวนวรรณกรรม

| ชื่อผู้วิจัย | ค่าคุณลักษณะที่ใช้ | วิธีที่ใช้ในการจำแนก | ผลการใช้ค่าคุณลักษณะ |
|---|---|---|---|
| Xu, Maddage and Shao (2007) | LPCC (Linear Prediction Cepstral Coefficients) LPC ZC | SVM | LPC และ LPCC เหมาะสำหรับใช้วิเคราะห์เสียงร้องที่ไม่มีดนตรี ZC เหมาะสำหรับใช้วิเคราะห์สัญญาณดนตรี |
| Wang , Huang ,Wang ,Liang and Xu (2008) | MFCC ZC SR SF SC Mpeg – 7 audio | Random Forest ANN Fusion Result | พบว่า MFCC และ Mpeg – 7 audio จะมีประสิทธิภาพในการจำแนกได้ดี |
| Chen Huang and Chen Hao (2009) | MFCC Log energy | SVM | จากการทดลองพบว่าใช้ค่า MFCC ร่วมกับ Log energy เฉลี่ยจากการจำแนกได้ค่าเปอร์เซ็นความถูกต้อง 82 % |
| XU,Wang and Yan (2008) | MFCC LPC SR SF SC ZC SSB Beat Sum Fraction of Low Energy | RBF neural network with feature selection algorithm Fisher Criterion method | พบว่าใช้ค่าคุณลักษณะ 25 ค่าสามารถจำแนกค่าสูงสุดจากทั้งหมด 74 ค่าโดยค่ากระจาย Standard Deviation จากค่า MFCC ดีกว่าค่าอื่น |
| Sho ,Xu and Kanhalli (2004) | MFCC LPC | HMM | จากการทดลองโดยใช้ระยะเพลง 30 วินาทีโดยใช้ค่า MFCCและLPC โดยใช้ HMM ที่ 5 State ให้ค่าจำแนกสูงสุดที่ 92 % จากเพลง |

| | | | |
|---|---|---|---|
| | | | country |
| Liu, Yang and Chen (2008) | MFCC LPC ZC SR SF SC Beat Histogram | HMM | HMM ให้ความถูกต้อง ในการจำแนก 92.49% |
| Doraisamy et.al.(2008) | MFCC LPC Beat Feature | ANN | ผลที่จำแนกความถูกต้องดีที่สุดคือโดยใช้ค่าMFCC LPC Beat ใช้แบบโครงข่ายประสาทเทียมแบบหลายชั้นโดยให้ค่าเปอร์เซ็นต์ในการจำแนก 88.6 % |
| อานนท์ นามสนิท (2548) | MFCC LPC ZC Line Spectrum Frequency Power Spectrum | SVM | พบว่าค่า LPC และ ZC ให้ค่าสูงสุดในการจำแนกเพลงไทยเดิมและค่า LPC ZC และ Line Spectral frequency ให้ค่าในการจำแนกเพลงไทยลูกทุ่งและลูกกรุง |
| โอภาส แก้วต่าย (2552) | MFCC LPC ZC SC Power Spectrum | ANN<br>KNN<br>SVM | พบว่า MFCC และ Power Spectrum ให้ค่าจำแนกสูงสุดที่ 86.16 โดยวิธี SVM |