บทที่ 2

ทฤษฎี

2.1 การสังเคราะห์เสียง

ในปัจจุบันเราสามารถแยกการสังเคราะห์เสียงออกเป็น 3 กลุ่มใหญ่ๆ (Luksaneeyanawin, 1989)ได้คือ

1. **การสร้างกลไกให้เปล่งเสียง** กล่าวคือจะสร้างเครื่องจักรที่สามารถเปล่งเสียงได้ โดยการควบคุมการเคลื่อนที่ของลมผ่านช่องหรือท่อต่างๆ เพื่อให้เกิดเสียง (Docherty และ Shockey,1988 อ้างถึงใน Luksaneeyanawin,1989)

2. **การใช้เครื่องมืออิเล็กทรอนิกส์สร้างเสียง** ด้วยวิธีการนี้จะอาศัยการเปลี่ยนแปลงของปัจจัยต่างๆที่ทำให้เกิดเสียง 8 - 32 ปัจจัยในการสังเคราะห์ลำดับของหน่วยเสียงเรียงก่อนที่จะเพิ่มหรือลดค่าต่างๆในกระบวนการเติมจังหวะลงบนการสังเคราะห์เสียงพูดทำให้เกิดเสียงพูดตามต้องการได้(Docherty และ Shockey,1988 อ้างถึงใน Luksaneeyanawin,1989) ด้วยเทคโนโลยีในปัจจุบันสามารถสร้างอุปกรณ์ทางอิเล็กทรอนิกส์ที่สามารถสร้างเสียงสังเคราะห์นี้โดยใช้เพียงชิปเดียว ซึ่งทำให้ระบบนี้เป็นระบบที่ฉลาดและราคาไม่แพง แต่มีข้อเสียคือเสียงที่ได้จะเป็นเสียงที่ไม่เป็นธรรมชาติ กล่าวคือไม่เหมือนกับเสียงคนพูดปกติ

3. **การสังเคราะห์เสียงโดยการต่อคลื่นเสียง** วิธีนี้จะอาศัยการบันทึกเสียงพูดจริงๆของมนุษย์เอาไว้ เมื่อต้องการจะพูดคำใดก็จะนำส่วนของเสียงที่บันทึกไว้มาพูดต่อกันเป็นคำพูดนั้นได้ ข้อดีที่เห็นชัดของวิธีนี้คือ คุณภาพของเสียงที่ได้จะคล้ายคลึงเสียงของมนุษย์มากที่สุด ข้อเสียของวิธีนี้คือจะต้องมีเนื้อที่ของหน่วยความจำขนาดใหญ่ในการเก็บส่วนของเสียงต่างๆที่ได้ทำการบันทึกเอาไว้ ซึ่งขึ้นกับขนาดของการสุ่มเก็บข้อมูลเสียงในการบันทึก

เนื่องจากงานวิทยานิพนธ์นี้เป็นงานทางด้านสิ่งประดิษฐ์ประยุกต์ ประกอบกับการที่วิทยานิพนธ์ นี้ได้ใช้วิธีการสังเคราะห์เสียงพูดโดยการต่อคลื่นเสียง ดังนั้นจึงขอกล่าวในรายละเอียดเกี่ยวกับการสังเคราะห์เสียงพูดโดยการต่อคลื่นเสียงเท่านั้นดังต่อไปนี้

2.2 การสังเคราะห์เสียงโดยการต่อคลื่นเสียง

2.2.1 การสังเคราะห์เสียงพูดแบบพยางค์ (Syllable)

การสังเคราะห์เสียงพูดแบบพยางค์นี้จะบันทึกเสียงพูดที่ละพยางค์ที่ต้องการ จากนั้นเมื่อระบบต้องการที่จะพูดคำใดก็จะทำการค้นเสียงที่ต้องการจากฐานข้อมูลที่เก็บไว้เมื่อมีก็จะทำการออกเสียงพยางค์นั้นออกมา ข้อดีคือจะได้เสียงที่เหมือนจริงที่สุด แต่ข้อเสียของระบบนี้ก็คือ ระบบนี้เหมาะสำหรับภาษาที่มีจำนวนพยางค์ไม่มาก และจะออกเสียงได้จำกัดเท่ากับจำนวนพยางค์ที่บันทึกไว้ สำหรับของภาษาไทยนั้น มีงานวิจัยเรื่องระบบการสังเคราะห์เสียง (Luksaneeyanawin, 1989) ที่พบว่ามีพยางค์ในภาษาไทยทั้งหมด 29,928 พยางค์ และมีที่ใช้จริงเพียง 5,912 พยางค์ ซึ่งถ้าจะทำการบันทึกทั้งหมดทุกพยางค์ จะต้องการขนาดหน่วยความจำในการเก็บประมาณ 300 เมกกะไบต์ (โดยประมาณว่าพยางค์หนึ่งพยางค์เฉลี่ยแล้วใช้เนื้อที่ประมาณ 10,000 ไบต์) ซึ่งจะเห็นได้ว่าเป็นขนาดที่ใหญ่มาก

2.2.2 การสังเคราะห์เสียงพูดโดยใช้อัฒพยางค์ (Demisyllable)

การสังเคราะห์เสียงพูดโดยอัฒพยางค์นี้จะใช้หลักที่ว่าพยางค์ใดสามารถจะแบ่งพยางค์ออกเป็นสองหน่วยอัฒพยางค์(Stephen, Tybring และ James, 1989) โดยที่หน่วยอัฒพยางค์ที่เก็บไว้นี้สามารถนำไปใช้กับหน่วยอัฒพยางค์อื่นเพื่อสังเคราะห์เป็นคำใหม่ได้ดังเช่นตารางที่ 2.1 นี้

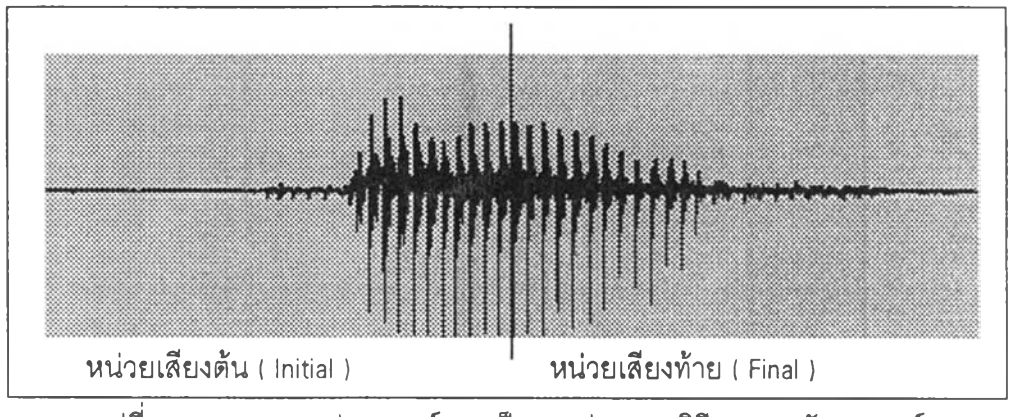
พยางค์	หน่วยเสียงเริ่ม		หน่วยเสียงท้าย
BET	BE	+	ET
SET	SE	+	ET

ตารางที่ 2.1 แสดงตัวอย่างหน่วยเสียงของการสังเคราะห์เสียงพูดแบบอัฒพยางค์

จากตารางที่ 2.1 เป็นตัวอย่างที่ทำการออกเสียงคำว่า BET และ SET ถ้าเป็นในระบบการสังเคราะห์เสียงพูดแบบพยางค์ ก็จะต้องบันทึกทั้ง 2 คำนี้ไว้ ซึ่งทั้งสองเสียงมีเสียงสระและตัวสะกดเดียวกันทำให้ซ้ำซ้อนกัน ดังนั้นด้วยวิธีสังเคราะห์เสียงพูดโดยใช้อัฒพยางค์สามารถประหยัดหน่วยอัฒพยางค์ได้ 1 หน่วยคือ ET จากทั้งหมด 4 หน่วย ดังนั้นถ้ามีพยัญชนะต้นที่ใช้ประกอบกับเสียงสระ ET มากก็จะยิ่งประหยัดมากหน่วยเสียงขึ้นไปอีก สำหรับหน่วยอัฒพยางค์ทั้งสองหน่วยอาจจำแนกได้เป็น 2 ประเภทดังนี้

1. *หน่วยเสียงเริ่ม (Initial)* หน่วยเสียงนี้จะเก็บเสียงของพยัญชนะต้นจนถึงช่วงของเสียงสระที่คงที่ตรงกลางของพยางค์

2. *หน่วยเสียงท้าย (Final)* หน่วยเสียงนี้จะเก็บเสียงของสระและเสียงอื่นๆที่เหลือจนหมด โดยจะเริ่มที่ช่วงของเสียงสระที่คงที่เป็นต้นไป ดังรูปที่ 2.1



รูปที่ 2.1 แสดงการแบ่งพยางค์ออกเป็นสองส่วนตามวิธีการของอัฒพยางค์

ผลที่ได้เมื่อทำการสังเคราะห์ในภาษาอังกฤษคือ มีหน่วยเสียงทั้งหมดเพียง 950 หน่วย และใช้เนื้อที่หน่วยความจำเพียง 250 กิโลไบต์

ข้อเสียก็คือการแบ่งตรงอัฒพยางค์ในช่วงของเสียงสระที่คงที่นั้นขนาดของสัญญาณจะมีขนาดใหญ่ทำให้เกิดความยุ่งยากในการบันทึกเสียงและการแบ่งเก็บที่จะต้องควบคุมความหนักเบาของเสียงที่บันทึกให้ใกล้เคียงกัน เพราะไม่เช่นนั้นแล้วเวลาทำการสังเคราะห์เสียง จะเกิดการสะดุดของสัญญาณได้

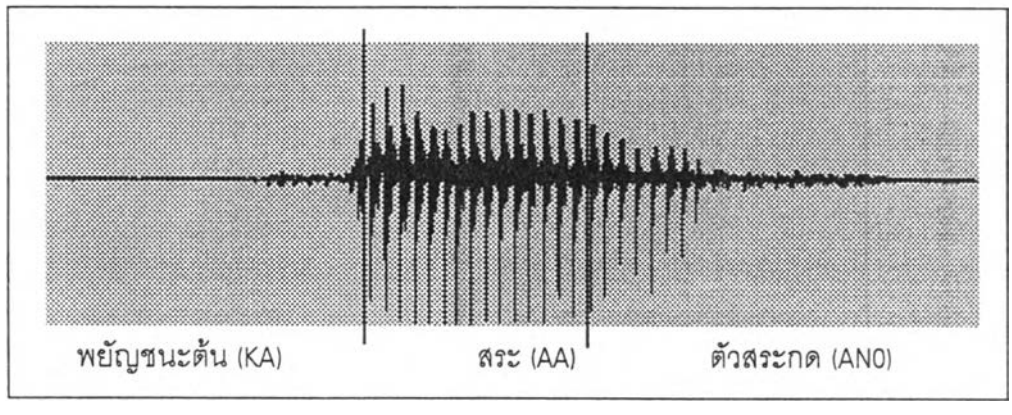
2.2.3 **การสังเคราะห์เสียงโดยใช้หน่วยคู่เสียง** (Diphone)

การออกเสียงในภาษาต่างๆนั้น จะมีลักษณะของเสียงที่ผสมกันของหลายหน่วยเสียง ทำให้หน่วยเสียงข้างเคียงจะมีผลต่อการออกเสียงด้วย ดังนั้นการสังเคราะห์เสียงโดยใช้หน่วยคู่เสียงนี้จะทำการเก็บหน่วยเสียงทุกหน่วยก่อนที่หน่วยเสียงนั้นจะมีการผันแปรตามหน่วยเสียงอื่นที่จะตามมา สำหรับการบันทึกหน่วยคู่เสียงจะเริ่มบันทึกตั้งแต่หน่วยเสียงพยัญชนะต้นจนถึงช่วงผลตอบชั่วครู่ของหน่วยเสียงถัดไป(Luksaneeyanawin,1989) ซึ่งอาจจะเป็นหน่วยเสียงของพยัญชนะต้นอื่น หรือหน่วยเสียงสระหรือหน่วยเสียงตัวสะกดก็ได้ ซึ่งจะต่างกับการสังเคราะห์เสียงแบบอัฒพยางค์ตรงที่การสังเคราะห์เสียงโดยใช้หน่วยคู่เสียงนี้อาจจะต้องการใช้

หน่วยสังเคราะห์หลายหน่วยในการประกอบเป็นหนึ่งพยางค์ก็ได้ ซึ่งขึ้นกับการผันแปรในการออกเสียงของพยางค์นั้น ดังแสดงเป็นตัวอย่างตามตารางที่ 2.2 และรูปที่ 2.2

พยางค์	หน่วยคู่เสียง	หน่วยคู่เสียง	หน่วยคู่เสียง
KAAN0 (การ)	KA	AA	AN0
DAAN0 (ดาน)	DA	AA	AN0
KAN0 (กัน)	KA	A	AN0

ตารางที่ 2.2 ตัวอย่างการแยกหน่วยคู่เสียงของพยางค์ต่างๆในรูปแบบของสัทอักษรที่ดัดแปลงมาจากสัทอักษรสากล



รูปที่ 2.2 ตัวอย่างรูปคลื่นของหน่วยคู่เสียงที่ประกอบกันขึ้นเป็นพยางค์ 'การ'

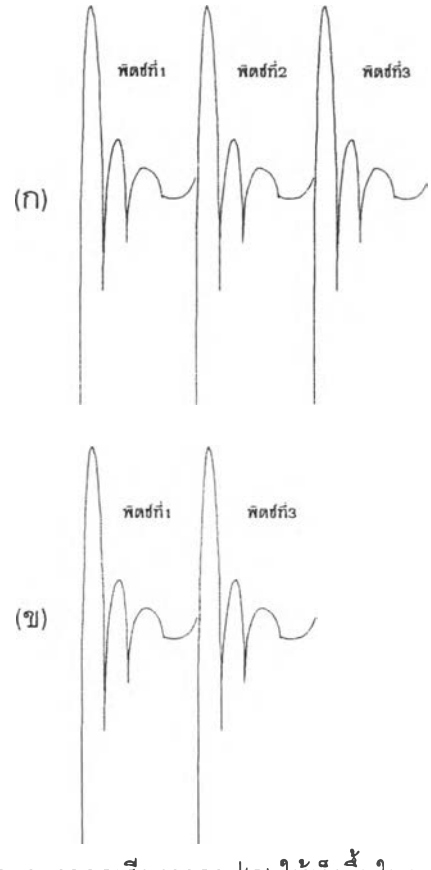
ข้อดีของการสังเคราะห์แบบนี้ก็คือประหยัดหน่วยความจำในการเก็บได้มากกว่าแบบการสังเคราะห์แบบอัฒพยางค์เนื่องจากได้มีการแยกพยางค์ออกเป็นหลายหน่วยย่อยในการประกอบกันทำให้ลดการซ้ำซ้อนกันของหน่วยเสียงที่อาจจะไปปรากฏในหน่วยเสียงอื่น แต่ทั้งนี้ต้องขึ้นกับว่าหน่วยย่อยดังกล่าวเป็นหน่วยย่อยที่สามารถเกิดร่วมกับหน่วยอื่นได้มากน้อยแค่ไหน ด้วยวิธีการสังเคราะห์นี้จะไม่ค่อยมีปัญหาความราบรื่นของการนำหน่วยเสียงมาต่อกัน เพราะช่วงที่นำมาต่อกันนั้นเป็นช่วงของภาวะชั่วครู่ซึ่งมีขนาดของสัญญาณไม่มาก ทำให้เวลาการนำมาประกอบกันค่อนข้างราบรื่นกว่า

ข้อเสียของวิธีนี้ก็คือถ้ามีการแบ่งพยางค์ออกเป็นหน่วยคู่เสียงหลายหน่วย ก็จะทำให้คุณภาพของเสียงไม่ดีนัก กล่าวคือจะเป็นเสียงที่ไม่เป็นธรรมชาตินัก ฟังดูจะค่อนข้างแปลกหูบ้างสำหรับคนที่ไม่คุ้นเคย

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

2.3 การปรับระดับความเร็วของการพูด

การเปลี่ยนความเร็วของเสียงพูดนั้นทำได้ทั้งในระบบแอนะล็อกและดิจิตอล ในระบบแอนะล็อกนั้นใช้วิธีบันทึกลงบนเทปแม่เหล็กหรือแผ่นเสียง จากนั้นก็นำมาเล่นโดยปรับความเร็วของมอเตอร์ที่ใช้ขับเครื่องเล่นให้เร็วหรือช้าตามต้องการ ในระบบดิจิตอลนั้นทำได้โดยการสุ่มสัญญาณด้วยความถี่หนึ่งแล้วเล่นกลับด้วยความถี่ที่แตกต่างกันออกไป วิธีที่กล่าวข้างต้นนี้สามารถปรับความเร็วของเสียงตามต้องการได้ แต่ก็มีข้อเสียว่าความถี่ของเสียงหรือพิตช์(pitch) จะเปลี่ยนไปด้วย ถ้าเล่นช้าลงความถี่ของเสียงก็จะต่ำลงจนเป็นที่น่ารำคาญหรืออาจจะฟังไม่รู้เรื่อง ถ้าเล่นเร็วขึ้นความถี่ของเสียงก็จะสูงขึ้นทำให้กลายเป็นเสียงเล็กแหลม แต่วิธีที่จะนำมาใช้ในวิทยานิพนธ์จะสามารถคงความถี่ของเสียงไว้เท่าเดิมโดยใช้การลดจำนวนพิตช์ (อภิชาติ ตั้งทางธรรม, 2537) ดังแสดงเป็นตัวอย่างในรูปที่ 2.3



รูปที่ 2.3 การปรับความเร็วของการออกเสียงจากรูป(ก) ให้เร็วขึ้นในรูป(ข) โดยการตัดพิตช์ที่ 2 ออก

2.4 การผันเสียงวรรณยุกต์ในพยางค์ไทยคำเดียว

ปัญหาใหญ่ในการออกเสียงพยางค์ต่างๆในภาษาไทยอยู่ตรงที่ส่วนของวรรณยุกต์ ซึ่งมีรูปและเสียงไม่จำเป็นต้องตรงกันเสมอไป ในทางปฏิบัติจะยึดหลักไตรยางค์ และเรื่องคำเป็นคำตายตามหลักไวยากรณ์ไทยดั้งเดิม (สมชาย ลำดวน, 2526) ดังนี้

2.4.1 คำเป็น-คำตาย

คำเป็น คือพยางค์ที่ประสมสระเสียงยาวในแม่ก.กา เช่น กา กี กือ กู และพยางค์ที่ประสมสระเสียงยาวหรือสั้นในแม่ กง กน กม เกย และ เกอว

คำตาย คือพยางค์ที่ประสมสระเสียงสั้นในแม่ ก.กา เช่น กะ กิ กึ กุ (เว้นสระ อำ ไอ ใอ เอา) และพยางค์ที่ประสมสระเสียงยาวหรือสั้นแม่ กก กด กบ

2.4.2 ไตรยางค์

ไตรยางค์ คือการแบ่งพยัญชนะออกเป็น 3 พวก เพื่อให้สอดคล้องกับวิธีการผันเสียงวรรณยุกต์ โดยกำหนดให้อักษรกลุ่มหนึ่งมีจำนวน 9 ตัว ผันเสียงวรรณยุกต์ได้ครบ 5 เสียง เรียกว่าอักษรกลาง ได้แก่ ก จ ฎ ฏ ด ต บ ป และ อ เช่น กา ก่า ก้า ก๊า ก๋า

อักษรอีกกลุ่มหนึ่งจำนวน 11 ตัว ได้ ข ฃ ฉ ฐ ถ ผ ฝ ส ศ ษ ห มีพื้นเสียงเป็นเสียงสูง (เสียงจัตวา) จึงเรียกว่าอักษรสูง อักษรประเภทนี้ผันเสียงวรรณยุกต์ได้เพียง 3 เสียงคือ จัตวา เอก โท (ขา ข่า ข้า)

อักษรอีก 24 ตัวที่เหลือ เรียกว่าอักษรต่ำ จะแบ่งได้เป็น 2 กลุ่มย่อยคือ อักษรต่ำที่มีคู่กับอักษรสูงมีอยู่ 14 ตัวคือ ค ฅ ฆ ช ฌ ซ ฑ ฒ ธ ท พ ภ ฟ และ ฮ และอักษรต่ำเดี่ยวมีอยู่มีทั้งหมด 10 ตัวคือ ง ญ ณ น ม ย ร ล ว ฬ

การผันเสียงของแต่ละกลุ่มอักษรเป็นดังตารางที่ 2.3 - 2.5

	เสียง	สามัญ	เอก	โท	ตรี	จัตวา
คำเป็น		กา	ก่า	ก้า	ก๊า	ก๋า
คำตาย	รัสสะ	-	กะ	ก้ะ	ก๊ะ	-
	ทีฆะ	-	กาก	ก้าก	ก๊าก	-

ตารางที่ 2.3 แสดงการผันเสียงวรรณยุกต์ของอักษรกลาง

	เสียง	สามัญ	เอก	โท	ตรี	จัตวา
คำเป็น		คา	ข่า	ข้า	ค้า	ขา
คำตาย	รัสสะ	-	ขะ	ข้ะ,ค่ะ	คะ	-
	ทีฆะ	-	ขาก	ข้าก,คาก	ค้าก	-

ตารางที่ 2.4 แสดงการผันเสียงวรรณยุกต์ของอักษรต่ำคู่กับอักษรสูง

		เสียง	สามัญ	เอก	โท	ตรี	จัตวา
คำเป็น		กลางนำ	ยา	อย่า	อย้า,ย่า	ย้า	อย๋า
		สูงนำ	ยา	หย่า	หย้า,ย่า	ย้า	หยา
คำตาย	รัสสะ	กลางนำ	-	อยะ	อย้ะ	อย๊ะ,ยะ	-
		สูงนำ	-	หยะ	หย้ะ,ย่ะ	ยะ	-
	ทีฆะ	กลางนำ	-	อยาก	อย้าก, ยาก	อย๊าก	-
		สูงนำ	-	หยอก	หย้อก, ยอก	ย้อก	-

ตารางที่ 2.5 แสดงการผันเสียงวรรณยุกต์ของอักษรต่ำเดี่ยวกับอักษรนำ

2.5 การบีบอัดข้อมูล (Data Compression)

จำแนกการบีบอัดข้อมูลหรือสัญญาณต่างๆตามความถูกต้องได้เป็น 2 แบบ แบบแรกเป็นแบบไม่สูญเสียข้อมูล (Lossless) คือข้อมูลที่ผ่านการบีบอัด เมื่อได้รับการคลายข้อมูลแล้วจะเหมือนกับข้อมูลต้นฉบับทุกประการ โดยทั่วไปวิธีการนี้ นิยมใช้ในการบีบอัดข้อมูลที่เกี่ยวกับ ฐานข้อมูล, สเปรดชีต หรือโปรแกรมการประมวลผลคำต่างๆ แบบที่สองเป็นแบบยอมสูญเสียบางส่วน (Lossy) เพื่อให้ได้อัตราส่วนของการบีบอัดดีที่ขึ้นกล่าวคือวิธีการในแบบนี้จะมีการจัดการปรับแต่งข้อมูลที่ต้องการบีบอัดก่อนที่จะทำการบีบอัดข้อมูล หลังจากนั้นก็จะใช้วิธีการบีบอัดข้อมูลในแบบแรกในการบีบอัดข้อมูล วิธีการในแบบหลังนี้เหมาะสำหรับข้อมูลแบบภาพ และข้อมูลเสียง

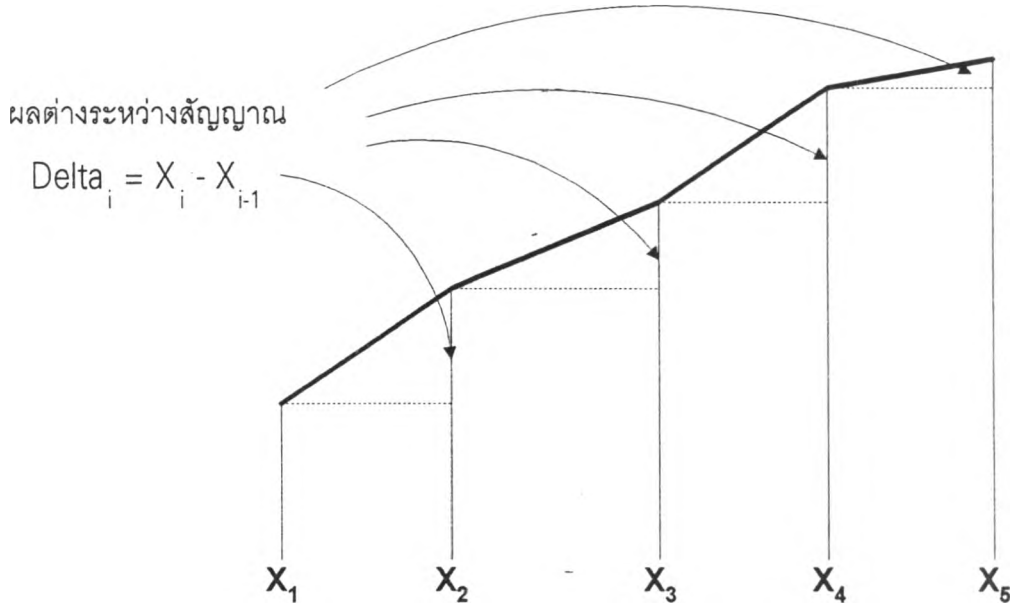
2.5.1 หลักเกณฑ์ทั่วไปในการบีบอัดข้อมูล

หลักในการบีบอัดข้อมูลโดยทั่วไปจะประกอบด้วย 2 ส่วนที่สำคัญ คือ การสร้างแบบจำลอง (Modeling) และ การเข้ารหัส (Coding) แบบจำลองมีอยู่ 2 ชนิดด้วยกันคือ แบบจำลองทางสถิติ (Statistical Model) และแบบจำลองทางพจนานุกรม (Dictionary Based) สำหรับการเข้ารหัสที่รู้จักกันดีคือ การเข้ารหัสแบบแชนนอน ฟาโน (Shannon-Fano) และการเข้ารหัสแบบฮาฟฟ์แมน (Huffman Coding) ซึ่งจะคล้ายกับวิธีการเข้ารหัสของ แชนนอน ฟาโน กล่าวคือจะมีใช้คุณลักษณะส่วนใหญ่ของแชนนอน ฟาโน แต่ในส่วนของที่แตกต่างกันก็คือ ฮาฟฟ์แมนจะสร้างต้นไม้แบบแตกสองแบบล่างขึ้นบน ในขณะที่ของแชนนอน ฟาโนจะสร้างจากบนลงล่าง ปัจจุบันมีการนำวิธีเข้ารหัสของฮาฟฟ์แมนไปใช้อย่างแพร่หลาย อาทิเช่นในเครื่องโทรสาร ,ใช้เป็นมาตรฐานของเจเปก (JPEG) และอื่นๆ

สำหรับการบีบอัดข้อมูลที่เป็นสัญญาณเสียงที่นิยมใช้กันก็คือ วิธีเอดีพีซีเอ็ม (ADPCM ,Adaptive Differential Pulse Code Modulation)(Nejat,1992) จะทำงานแบบเวลาจริง(real time) และสามารถลดขนาดของข้อมูลได้ 50 เปอร์เซ็นต์ โดยจะยอมให้เกิดโอเวอร์โหลดทางความชันได้ จึงจัดได้ว่าเป็นวิธีการบีบอัดข้อมูลที่ยอมสูญเสียข้อมูลบางส่วนในกรณีที่เกิดโอเวอร์โหลดทางความชัน แต่ในวิทยานิพนธ์นี้จะไม่ยอมให้เกิดการผิดเพี้ยนของข้อมูลเสียงที่ผ่านการบีบอัดสัญญาณ จึงได้นำวิธีการของเอดีพีซีเอ็มมาปรับปรุงเสียใหม่ดังนี้

2.5.2 วิธีการบีบอัดสัญญาณแบบแอลเอดีพีซีเอ็ม (Lossless Adaptive Differential Pulse Code Modulation)

วิธีการบีบอัดข้อมูลแบบแอลเอดีพีซีเอ็ม จะไม่ทำงานแบบเวลาจริง เพราะข้อมูลเสียงที่จะนำมาบีบอัดนั้นถูกบันทึกเก็บเอาไว้แล้วด้วยความถี่สุ่มเท่ากับ 16 กิโลเฮิรตซ์ขนาดข้อมูล 8 บิต โดยที่การบีบอัดสัญญาณจะใช้การเก็บข้อมูลที่เป็นผลต่างระหว่างสัญญาณที่ติดกันดังแสดงในรูปที่ 2.4 ซึ่งผลต่างเหล่านี้จะมีค่าน้อยทำให้สามารถเก็บเป็นข้อมูลที่มีขนาดของผลต่าง(บิต) ที่เล็กกว่าหรือเท่ากับ 8 บิตโดยอยู่ในรูปของกระแสของบิต(bit stream) และถ้าขนาดของผลต่างของสัญญาณมีขนาดน้อยกว่า 8 บิตจำนวนมากๆ ก็จะทำให้อัตราส่วนของการบีบอัดสัญญาณได้มาก ดังนั้นวิธีนี้จะแทรกคำสั่งพิเศษดังแสดงในตารางที่ 2.6 เพื่อระบุจำนวนบิตที่ใช้ในการเก็บขนาดผลต่างดังนี้



รูปที่ 2.4 แสดงการหาผลต่างระหว่างสัญญาณเสียงเพื่อประหยัดจำนวนบิตที่จะต้องใช้เก็บ

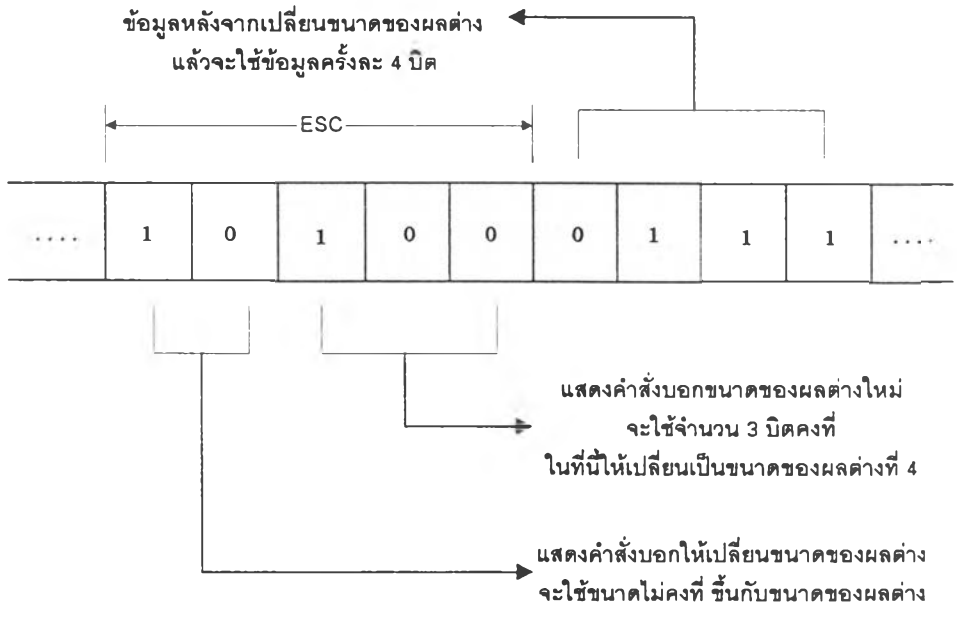
1. คำสั่งที่บอกว่าจะต้องเปลี่ยนขนาดของผลต่าง ซึ่งจะใช้ค่าที่เป็นลบที่มากที่สุดในช่วงขนาดของผลต่างในขณะนั้นเป็นตัวบอก ซึ่งจำนวนบิตที่ใช้ไม่แน่นอนขึ้นกับขนาดของผลต่างในขณะนั้น

2. คำสั่งที่บอกขนาดของผลต่างใหม่ มีขนาดของคำสั่งที่คงที่คือ 3 บิตและจะต้องอยู่ติดกับคำสั่งแรกเสมอ

เนื่องจากคำสั่งทั้งสองนี้จะอยู่ติดกันเสมอจึงขอใช้สัญลักษณ์ ESC แทนคำสั่งทั้งสองนี้รวมกัน ดังแสดงเป็นดังตัวอย่างในรูปที่ 2.5

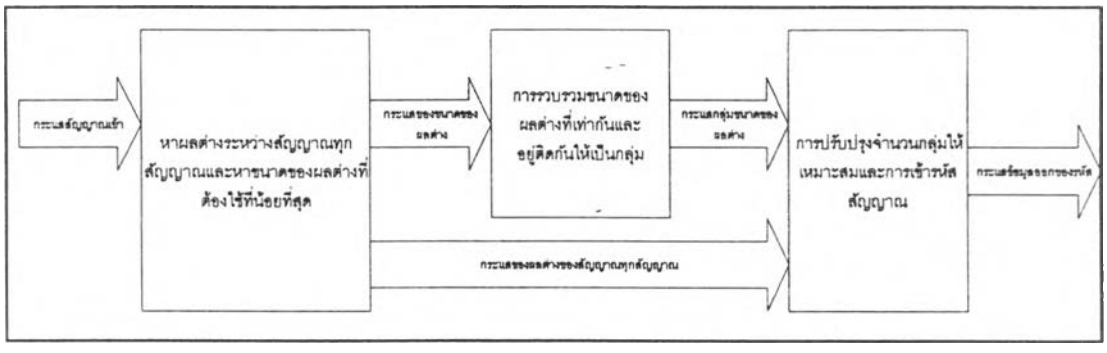
คำสั่งบอกขนาดของผลต่างใหม่ (ใช้ 3 บิต)	ขนาดของผลต่าง (บิต)	ช่วงของข้อมูลในแต่ละขนาดของผลต่าง	คำสั่งที่บอกให้ต้องเปลี่ยนขนาดของผลต่าง
000	8	-128 ถึง +127	-128
001	ข้อมูลสุดท้ายของสัญญาณ	-	-
010	2	-2 ถึง +1	-2
011	3	-4 ถึง +3	-4
100	4	-8 ถึง +7	-8
101	5	-16 ถึง +15	-16
110	6	-32 ถึง +31	-32
111	7	-64 ถึง +64	-64

ตารางที่ 2.6 แสดงคำสั่งการเปลี่ยนขนาดของผลต่าง



รูปที่ 2.5 แสดงตัวอย่างการใช้คำสั่งเปลี่ยนขนาดของผลต่าง

สำหรับรูปแบบการบีบอัดสัญญาณจะเป็นดังรูปที่ 2.6 ซึ่งมีอยู่ 3 ส่วนที่สำคัญคือ



รูปที่ 2.6 แสดงการบีบอัดสัญญาณแบบแอลเอดีพีซีเอ็ม

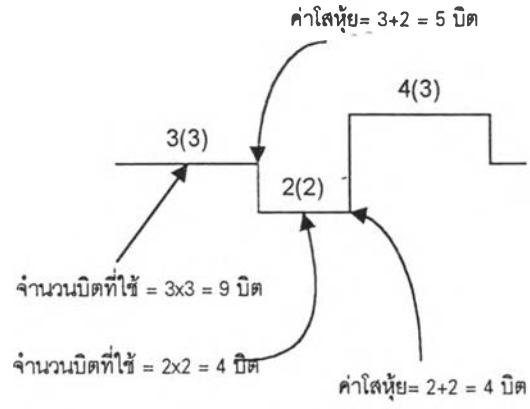
2.5.2.1 ส่วนของการหาผลต่างของสัญญาณและขนาดของผลต่างที่ต้องใช้ที่น้อยที่สุด ส่วนนี้จะให้ข้อมูลออกเป็น 2 ชุดคือ ข้อมูลชุดแรกจะเป็นการจัดการหาผลต่างของสัญญาณทั้งหมดให้กับส่วนของการเข้ารหัสสัญญาณ และข้อมูลชุดที่สองจะเป็นข้อมูลที่ได้จากการแปลงผลต่างเหล่านั้นให้อยู่ในรูปของขนาดของผลต่างที่น้อยที่สุดให้กับส่วนของการรวบรวมขนาดของผลต่างที่เท่ากันและอยู่ติดกันเข้าเป็นกลุ่มเดียวกันต่อไป ตัวอย่างของการแปลงผลต่างให้เป็นขนาดของผลต่างเป็นดังนี้ เช่นผลต่างของสัญญาณคู่หนึ่งเป็น -3 จากตารางที่ 2.5 จะได้ว่าขนาดของผลต่างที่น้อยที่สุดที่สามารถครอบคลุมได้ก็คือ 3

2.5.2.2 ส่วนของการรวบรวมขนาดของผลต่างที่เท่ากันและอยู่ติดกันเข้าไว้เป็นกลุ่มเดียวกัน ส่วนนี้จะทำการแปลงข้อมูลที่เป็นกระแสของขนาดของผลต่างให้เป็นกลุ่มขนาดของผลต่าง เพื่อให้ส่วนของการปรับปรุงจำนวนกลุ่มให้เหมาะสมและการเข้ารหัสสัญญาณนั้นวิเคราะห์ข้อมูลได้สะดวก ตัวอย่างเช่น

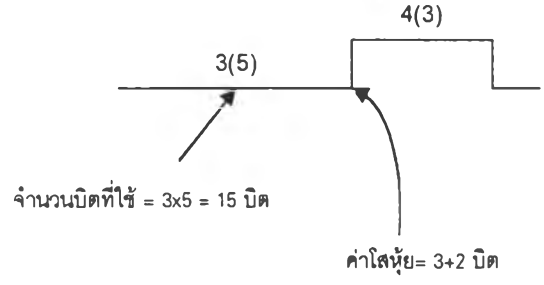
ข้อมูลเข้า = 2,2,2,2,2,2,3,3,3,3,4,4,4,3,2,2,2,2,5,5,5,6,6

ข้อมูลออก = 2(6),3(4),4(3),3(1),2(4),5(3),6(2)

2.5.2.3 ส่วนของการปรับปรุงกลุ่มให้เหมาะสมและการเข้ารหัส ส่วนนี้จะทำการวิเคราะห์และปรับปรุงข้อมูลเข้าที่เป็นกระแสกลุ่มของขนาดของผลต่างให้มีความเหมาะสม เนื่องจากการเปลี่ยนขนาดของผลต่างแต่ละครั้งจะต้องใช้คำสั่ง ESC ซึ่งถือว่าส่วนนี้เป็นค่าโสหุ้ย ดังนั้นถ้ามีการเปลี่ยนขนาดของผลต่างถี่ๆจะส่งผลให้ผลของการบีบอัดสัญญาณไม่ดี ทำให้ต้องมีการปรับปรุงกลุ่มโดยการยุบกลุ่มบางกลุ่มให้รวมกับกลุ่มอื่นที่ติดกันเพื่อให้ได้ผลของการบีบอัดที่ดีเช่นตัวอย่างในรูปที่ 2.7



(ก) รวมจำนวนบิตที่ต้องใช้ = 9+5+4+4 = 22 **บิต**



(ข) รวมจำนวนบิตที่ต้องใช้ = 15+5 = 20 **บิต**

รูปที่ 2.7 แสดงการยุบกลุ่มของขั้นขนาดของผลต่างเพื่อให้ผลการบีบอัดสัญญาณที่ดี โดยที่ระดับความสูงของแต่ละกลุ่มหมายถึงขนาดของผลต่าง และความกว้างของแต่ละกลุ่มคือจำนวนข้อมูลในกลุ่ม

จากรูปที่ 2.7 (ก) จะเป็นการพิจารณาข้อมูลของ 2 กลุ่มแรกที่เข้ามาโดยยังไม่มีการรวมกลุ่ม ซึ่งจะใช้จำนวนบิตในการเข้ารหัสเท่ากับ 22 บิต โดยมีการเปลี่ยนขนาดของผลต่าง 2 ครั้งทำให้ต้องมีบิตคำสั่งเท่ากับ 9 บิต และบิตที่เป็นข้อมูลอีก 13 บิต แต่ในรูป (ข) จะมีการรวมกลุ่มที่สองเข้ารวมกับกลุ่มแรกทำให้ต้องใช้บิตที่เป็นข้อมูลเพิ่มขึ้นเป็น 15 บิต แต่ในขณะที่บิตที่เป็นคำสั่งลดเหลือ 5 บิตทำให้ผลรวมของจำนวนบิตเท่ากับ 20 บิต ซึ่งจะเห็นได้ว่าประหยัดกว่ารูป (ก) อยู่ 2 บิต ดังนั้นรวมกลุ่มจะดีกว่า

2.5.2.3.1 หลักของการปรับปรุงกลุ่ม

สำหรับหลักในการปรับปรุงกลุ่มที่ใช้ในวิทยานิพนธ์นี้ไม่ใช้หลักที่ให้ผลการบีบอัดสัญญาณที่ดีที่สุด แต่เป็นหลักการปรับปรุงกลุ่มที่สามารถพัฒนาได้

ง่ายและให้ผลของการบีบอัดสัญญาณโดยเฉลี่ยได้ถึง 50 เปอร์เซ็นต์ หลักในการปรับปรุงนี้จะเริ่มจากการพิจารณาเสถียรภาพของกลุ่มทุกกลุ่มที่มีขนาดของผลต่างที่เท่ากันโดยจะเริ่มจากขนาดผลต่างที่เท่ากับ 2 บิตก่อนจากนั้นเมื่อหมดกลุ่มที่มีขนาดของผลต่างเป็น 2 แล้วก็จะทำการพิจารณาเสถียรภาพของกลุ่มที่มีขนาดของผลต่างเพิ่มขึ้นอีก 1 บิตเป็น 3 บิต ทำเช่นนี้ต่อไปจนกระทั้งหมดกลุ่มของขนาดของผลต่างที่ 7 บิต ถ้าเกิดว่ามีกลุ่มที่ไม่เสถียรภาพก็จะต้องปรับขนาดของผลต่างเพิ่มขึ้นเพื่อให้เข้าสู่กลุ่มที่เสถียรภาพให้ได้ โดยจะลองเพิ่มขนาดของผลต่างให้กับกลุ่มนั้นที่ละ 1 บิตเช่นถ้าขนาดของผลต่างเป็น 3 บิตก็จะเพิ่มให้เป็น 4 บิต ทำให้เมื่อเสร็จจากการพิจารณาเสถียรภาพในขนาดของผลต่างใดแล้ว ก่อนที่จะเริ่มพิจารณาในขนาดของผลต่างต่อไปจะต้องทำการจับกลุ่มขนาดของผลต่างใหม่ให้ทันกาล แต่ถ้าเกิดกลุ่มนั้นเสถียรภาพก็ไม่ต้องทำอะไรกับขนาดของผลต่างของกลุ่มนั้น

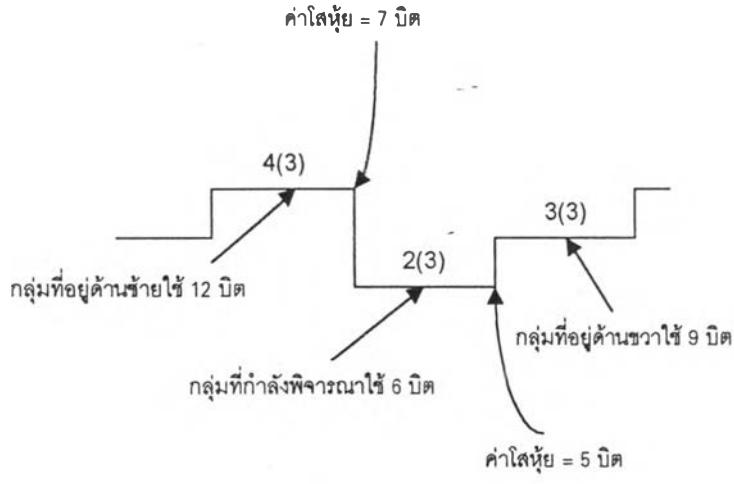
การพิจารณาเสถียรภาพของกลุ่มจะมีการใช้ข้อมูลของกลุ่มข้างเคียงช่วยในพิจารณา กลุ่มข้างเคียงที่สำคัญคือ *กลุ่มทางซ้าย*ของกลุ่มที่กำลังพิจารณาเสถียรภาพ กลุ่มที่สองได้แก่ *กลุ่มที่อยู่ทางขวา*ของกลุ่มที่กำลังพิจารณา โดยจะพิจารณาเสถียรภาพดังนี้

1. คำนวณหาจำนวนบิตที่ต้องใช้ทั้งหมดรวม 3กลุ่มโดยไม่มีการรวมกลุ่ม ดังตัวอย่างที่แสดงในรูปที่ 2.8(ก)

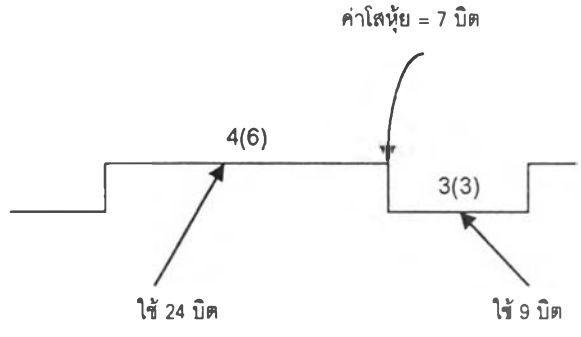
2. ให้ยุบกลุ่มที่กำลังพิจารณารวมกับกลุ่มทางซ้ายมือและคำนวณหาจำนวนบิตที่ต้องใช้ทั้งหมด ซึ่งจะเหลือเพียง 2 กลุ่มคือ กลุ่มทางซ้ายที่รวมกับกลุ่มที่กำลังพิจารณา และกลุ่มที่อยู่ทางขวา ดังตัวอย่างที่แสดงในรูปที่ 2.8(ข)

3. ให้ยุบกลุ่มที่กำลังพิจารณารวมกับกลุ่มทางขวามือและคำนวณหาจำนวนบิตที่ต้องใช้ทั้งหมด ซึ่งจะเหลือเพียง 2 กลุ่มคือ กลุ่มทางซ้ายและกลุ่มทางขวาที่รวมกับกลุ่มที่กำลังพิจารณาดังตัวอย่างที่แสดงในรูปที่ 2.8(ค)

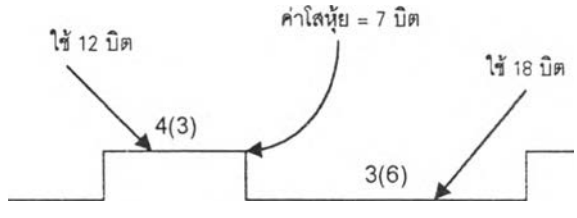
4. จากนั้นให้ดูว่ากลุ่มที่กำลังพิจารณานั้นเสถียรภาพไหม โดยดูจากจำนวนบิตที่ใช้ในข้อ 1 นั้นจะต้องน้อยกว่าจำนวนบิตในข้อ 2 และข้อ 3 ถ้าน้อยกว่าทั้งคู่จริงก็แสดงว่ากลุ่มที่กำลังพิจารณาอยู่นี้เสถียรภาพ แต่ถ้าเกิดจำนวนบิตในข้อ 1 มากกว่าจำนวนบิตในข้อ 2 หรือ 3 แสดงว่ากลุ่มนี้ไม่เสถียรภาพ



(ก) **รวมจำนวนบิตที่ต้องใช้ในกรณีที่ไม่รวมกลุ่ม = 39 บิต**



(ข) รวมจำนวนบิตที่ต้องใช้ในกรณีที่รวมกลุ่มทางซ้าย = 40 บิต



(ค) รวมจำนวนบิตที่ต้องใช้ในกรณีที่รวมกลุ่มทางขวา = 37 บิต

รูปที่ 2.8 แสดงการหาเสถียรภาพของกลุ่มโดยการหาจำนวนบิตที่ต้องใช้ตามกรณีต่างๆจะเห็นได้ว่าในกรณีที่รวมกลุ่มทางขวามือ(ค) จะทำให้ประหยัดจำนวนบิตที่ต้องใช้มากที่สุด แสดงว่ากลุ่มที่กำลังพิจารณาอยู่นั้นไม่เสถียรภาพ ดังนั้นให้ทำการเพิ่มขนาดของผลต่างให้กับกลุ่มที่กำลังพิจารณานี้ 1 บิตทำให้มีขนาดของผลต่างเป็น 3 บิต

2.5.2.3.2 การเข้ารหัส

การเข้ารหัสในที่นี้จะนำเอาผลต่างของสัญญาณที่หาได้ใน 2.5.2.1 ซึ่งอยู่ในขนาดของข้อมูล 8 บิตมาที่ละตัวมาเข้ารหัสให้เหลือจำนวนบิตเท่ากับ

ขนาดของผลต่างที่มีลำดับเดียวกันในกลุ่มของขนาดผลต่างที่ได้ปรับปรุงกลุ่มแล้ว ตัวอย่างเช่นมีกลุ่มของขนาดผลต่างที่ปรับปรุงแล้วเป็นดังนี้ 2(2),3(2),2(1) และมีผลต่างของสัญญาณเป็น -1,0,-2,-2,1 ผลของการเข้ารหัสจะเป็นดังรูปที่ 2.9

1	1	0	0	1	0	0	1	1	1	0	1	1	0	1	1	0	0	0	1	0	0	1
-1		0		ESC					-2			-2			ESC						+1	

รูปที่ 2.9 แสดงผลที่ได้จากการเข้ารหัส

สำหรับโครงสร้างของข้อมูลแต่ละชุดที่เกิดจากการเข้ารหัสในวิทยานิพนธ์นี้จะเป็นดังนี้ กำหนดให้ขนาดของผลต่างเป็น 2 บิตเป็นค่าโดยปริยายในตอนเริ่มต้น ข้อมูลไบต์แรกจะเป็นสัญญาณเสียงจุดแรกเพื่อให้เป็นค่าเริ่มต้นในการคลายสัญญาณ ข้อมูลในไบต์ถัดมาจึงจะเป็นข้อมูลที่ถูกเข้ารหัสไว้ และจะปิดท้ายข้อมูลด้วยขนาดของผลต่าง 1 บิต