บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 การเข้ารหัสเสียงแบบต่างๆ [5],[6]

ที่ผ่านมาได้มีการพัฒนาวิธีการเข้ารหัสหรือบีบอัดสัญญาณเสียงด้วยกันหลายๆ วิธีซึ่งแต่ละวิธีก็พยายามที่จะให้มีอัตราการเข้ารหัสที่ต่ำ โดยที่ยังคงคุณภาพของเสียงไว้ได้ในระดับที่ดี ตารางที่ 2.1 แสดงการเปรียบเทียบคุณสมบัติการเข้ารหัสเสียงพูดแบบต่างๆ การเข้ารหัสเริ่มตั้งแต่การเข้ารหัสด้วยวิธี PCM (Pulse Code Modulation) มีอัตราการเข้ารหัส 64 kbits/s เป็นการเข้ารหัสเสียงที่ใช้อ้างอิงเพื่อเปรียบเทียบกับการเข้ารหัสเสียงแบบอื่นๆ และเป็นมาตรฐานการเข้ารหัสเสียง ITU-T G.711 เมื่อปี ค.ศ. 1960 เป็นมาตรฐานนานาชาติที่ตั้งขึ้นโดย ITU-T (International Telecommunication Union) ชื่อเดิมเรียก CCITT (Consultative Committee of International Telegraph and Telephone) ต่อมามีการพัฒนาการเข้ารหัสเสียงแบบ ADPCM (Adaptive Differential Pulse Code Modulation) หรือ ITU-T G.721 มีอัตราการเข้ารหัส 32 kbits/s (เป็นมาตรฐานที่ได้รับความนิยมและมีการใช้งานอย่างกว้างขวางในเครื่องบันทึกเสียงพูดดิจิตอลและใช้ในการสื่อสารในปัจจุบัน) แต่มีข้อเสียของการเข้ารหัสแบบ ADPCM ก็คือยังมีอัตราการเข้ารหัสที่สูง ซึ่งทำให้ยังต้องใช้หน่วยความจำในการเก็บบันทึกค่อนข้างสูง ต่อมามีการพัฒนาการเข้ารหัสเสียงพูดหรือบีบอัดเสียงพูดหลายๆ วิธีเพื่อต้องการให้คุณภาพของเสียงที่สูง ในขณะที่ใช้อัตราการเข้ารหัสต่ำลง เช่น การเข้ารหัสแบบ LD-CELP (Low-Delay Code Excited Linear Prediction) หรือ ITU-T G.728 มีอัตราการเข้ารหัสคือ 16 kbits/s และต่อมาทาง ITU-T ได้มีการพัฒนาการเข้ารหัสเสียงพูดแบบ CS-ACELP (Conjugate Structure and Algebraic Code Excited Linear Prediction) หรือ ITU-T G.729 ในเดือนกุมภาพันธ์ ปี ค.ศ. 1995 นับเป็นมาตรฐานการเข้ารหัสเสียงพูดล่าสุดของ ITU-T มีอัตราการเข้ารหัสคือ 8 kbits/s ซึ่งถือว่าเป็นอัตราการเข้ารหัสที่ต่ำแต่ยังคงสามารถรักษาคุณภาพของเสียงได้ในระดับที่ดี

2.2 การวัดสมรรถนะของการเข้ารหัสเสียงพูด

สมรรถนะในการเข้ารหัสเสียงพูดจะพิจารณาจากคุณสมบัติต่างๆ เช่น อัตราการเข้ารหัส (Bit rate) คุณภาพเสียงที่ผ่านการเข้ารหัส (Speech Quality) ความซับซ้อนของการเข้ารหัส (Complexity) มีหน่วยเป็น MIPS (Million Instructions Per Second) ค่าประวิงเวลา (Delay

time) ความทนทานต่อความผิดพลาดที่เกิดภายในของช่องสัญญาณ(Robustness) หรือการแทรกสอดที่เกิดจากเสียงอื่นๆ (Acoustic Interferences) เช่น เสียงรบกวน

การสื่อสารข้อมูลแบบดิจิตอล คุณภาพของเสียงจะถูกแบ่งออกเป็น 4 ระดับได้แก่

1. ระดับกระจายเสียง (Broadcast) เสียงพูดในระดับนี้จะอ้างถึงเสียงพูดบรรยายที่มีคุณภาพสูง โดยปรกติจะมีอัตราการเข้ารหัสตั้งแต่ 64 kbits/s ขึ้นไป
2. ระดับเครือข่าย (Toll or Network) คุณภาพเสียงจะสามารถเทียบได้กับเสียงพูดในระบบแอนาลอกในช่วงความถี่ 200-3200 Hz โดยปรกติจะมีอัตราการเข้ารหัสตั้งแต่ 16 kbits/s ขึ้นไป
3. ระดับสื่อสาร (Communication) ยอมให้คุณภาพของเสียงลดลงได้บ้างแต่ยังคงความเป็นธรรมชาติของเสียงอยู่มีคุณภาพเพียงพอที่จะใช้ในการสื่อสารสามารถสร้างได้โดยอัตราการเข้ารหัสตั้งแต่ 4.8 kbits/s ขึ้นไป แต่ปัจจุบันพยายามปรับอัตราการเข้ารหัสลงต่ำประมาณ 4.0 kbits/s
4. ระดับเสียงสังเคราะห์ (Synthetic) สามารถรับฟังได้เข้าใจแต่ไม่เป็นธรรมชาติและสูญเสียคุณสมบัติในการรู้จำของเจ้าของเสียงพูด มีอัตราการเข้ารหัสต่ำกว่า 4.0 kbits/s

ตารางที่ 2.1 การเปรียบเทียบคุณสมบัติการเข้ารหัสเสียงพูดแบบต่างๆ

การเข้ารหัส	อัตราการเข้ารหัส (Kbits/s)	คุณภาพเสียง (MOS)	ความซับซ้อน (MIPS)	ขนาดเฟรม (ms)
PCM G.711	64	4.3	0.01	0
ADPCM G.721	32	4.1	2	0.125
ADPCM G.726	16 24 32 40	Toll	2	0.125
LD-CELP G.728	16	4.0	30	0.625
CS-ACELP G.729	8	4.0	20	10
CS-ACELP G.729 (Annex A)	8	3.8	11	10
GSM HR VSELP	6.3	3.4	14	20
IS-54 VSELP	8	3.5	14	20
MELP	2.4	3.2	40	22.5
FS 1015 LPC 10E	2.4	2.3	7	
FS 1016-CELP	4.8	3.2	16	30

2.2.1 **การวัดคุณภาพเสียงโดยใช้ค่าอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวน** (Power Signal-to-Noise Ratio : PSNR)

เป็นการวัดค่าเชิงวัตถุ (Objective Measurement) ใช้สำหรับวัดคุณสมบัติของอัลกอริทึมในการบีบอัดข้อมูล โดยมีสมการการคำนวณดังนี้

$$PSNR = 10\log_{10}\left\{\frac{\sum_{n=0}^{M-1} s^2(n)}{\sum_{n=0}^{M-1}(s(n)-s'(n))^2}\right\} \quad (2.1)$$

โดยที่ $s(n)$ คือสัญญาณเสียงดั้งเดิม

$s'(n)$ คือสัญญาณเสียงที่ผ่านการเข้ารหัสแล้ว

ค่า PSNR ถือได้ว่าเป็นการวัดแบบช่วงยาว (long-term) สำหรับวัดหาความถูกต้องของการสร้างสัญญาณเสียงกลับขึ้นมาใหม่

การเปลี่ยนแปลงอย่างฉับพลันสามารถตรวจจับและประเมินได้ โดยการใช้ PSNR ในช่วงสั้นคือการคำนวณ PSNR สำหรับแต่ละส่วนของเสียงพูดที่มีอยู่ N จุด เรียกการวัดเช่นนี้ว่า การหาอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนเป็นส่วน (SegPSNR : Segmental Power Signal to Noise Ratio) มีสมการดังนี้

$$SegPSNR = \frac{10}{L}\sum_{i=0}^{L-1}\log 10\left\{\frac{\sum_{n=0}^{N-1} s^2(iN+n)}{\sum_{n=0}^{N-1}(s(iN+n)-s'(iN+n))^2}\right\} \quad (2.2)$$

2.2.2 **การวัดคุณภาพเสียงโดยใช้วิธีการให้คะแนนความเห็นเฉลี่ย** (Mean Opinion Score)

วิธีการวัดคุณภาพเสียงอีกประเภทหนึ่งคือใช้การรับรู้และความรู้สึกของมนุษย์เป็นเกณฑ์ในการตัดสิน คือ การวัดในเชิงของผู้ฟัง (Subjective Measurement) ซึ่งมีหลายวิธีแต่วิธีที่ใช้กันอย่างแพร่หลายมากที่สุด คือ Mean Opinion Score (MOS) วิธี MOS จะใช้ผู้ฟังประมาณ 12-24 คน (อาจเปลี่ยนไปแล้วแต่มาตรฐานที่จะเป็นตัวกำหนด) ซึ่งถ้าทำการทดสอบตามมาตรฐาน ITU นั้นจะมีรายละเอียดขั้นตอนในการทดสอบดังนี้

1. ใช้คนกลุ่มเล็กๆ ทำการอ่านประโยคทดสอบและทำการอัดสัญญาณเสียงเหล่านี้ไว้
2. ทำการเข้ารหัสสัญญาณเสียงเหล่านี้ด้วยวิธีที่ต้องการจะทดสอบ

3. ทดสอบคุณภาพของเสียงกับกลุ่มคนประมาณ 12-24 คน โดยที่แต่ละคนจะให้คะแนนที่มีค่าอยู่ระหว่าง 1-5 ตามคุณภาพของสัญญาณที่ตัวเองรู้สึก รายละเอียดของคะแนนแต่ละขั้นแสดงไว้ในตารางที่ 2.2
4. นำค่าเฉลี่ยที่ได้ไปใช้ ซึ่งมีชื่อเรียกว่าค่า MOS

ตารางที่ 2.2 ระดับคะแนนในการวัด MOS

คะแนน	คุณภาพของเสียง	
5	ดีมาก	คุณภาพเสียงชัดเจนและเข้าใจง่าย
4	ดี	คุณภาพเสียงดีและเข้าใจง่าย แต่อาจได้ยินเสียงรบกวนบ้าง
3	พอใช้	คุณภาพเสียงเข้าใจได้แต่อาจต้องอาศัยความตั้งใจหรือบางทีต้องขอให้พูดซ้ำ
2	เลว	คุณภาพเสียงจะเข้าใจได้ก็ต่อเมื่อมีความตั้งใจมากๆ และบ่อยครั้งที่ต้องขอให้พูดซ้ำ
1	เลวมาก	คุณภาพเสียงฟังไม่รู้เรื่องเลย

ระบบการใช้งานแต่ละระบบมีความต้องการคุณภาพเสียงหรือค่า MOS ที่แตกต่างกันออกไป สามารถสรุปค่า MOS ที่เหมาะสมกับการใช้งานต่างๆ ดังตารางที่ 2.3

ตารางที่ 2.3 ค่า MOS ของคุณภาพเสียงในระดับต่างๆ

MOS	การใช้งาน
4.5-5.0	ระดับกระจายเสียง
4.0-4.5	ระดับเครือข่าย
3.5-4.0	ระดับสื่อสาร
2.5-3.5	ระดับเสียงสังเคราะห์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย