

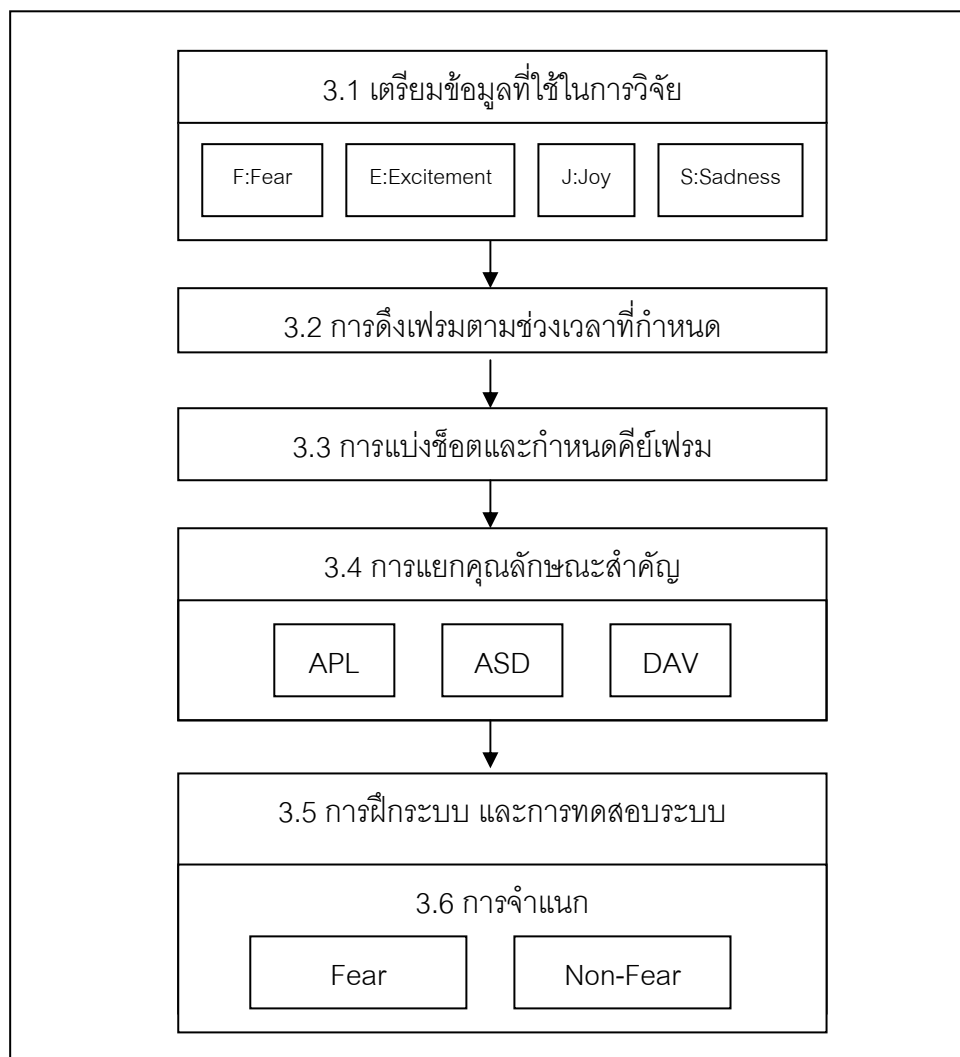
บทที่ 3

วิธีดำเนินงานวิจัย

เนื้อหาภายในบทนี้กล่าวถึงวิธีการดำเนินงานวิจัยเพื่อให้บรรลุตามวัตถุประสงค์ดังที่กำหนดไว้แบ่งได้หกขั้นตอนดังภาพที่ 3.1 ได้แก่ การเตรียมข้อมูลที่ใช้ในการวิจัย การดึงเฟรมตามช่วงเวลาที่กำหนด การแบ่งคลิปวิดีโอออกเป็นข้อตและการกำหนดคีย์เฟรม (Key Frame) การแยกคุณลักษณะสำคัญ (Feature Extraction) การฝึกระบบและการทดสอบระบบ และสุดท้ายเป็นขั้นตอนการจำแนก (Classification)

ภาพที่ 3.1

ภาพรวมการดำเนินงานวิจัย



3.1 ขั้นตอนการเตรียมข้อมูลวิดีโอที่ใช้ในงานวิจัย

3.1.1 ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในการวิจัยเป็นคลิปวิดีโอที่มีความยาวประมาณ 25 ถึง 60 วินาทีรวมทั้งหมด 80 คลิป โดยคลิปทั้งหมดถูกตัดมาจากฉากในภาพยนตร์ฮอลลีวู้ด ช่วงปี ค.ศ.2000 ถึง ค.ศ.2008 จำนวน 17 เรื่องซึ่งบันทึกในระบบ PAL (Phase Alternate Line) รูปแบบของคลิปวิดีโออยู่ในรูปแบบไฟล์ MPEG1 มีอัตราการแสดงผลภาพ (Frame Rate) เท่ากับ 25 เฟรมต่อวินาที (fps) โดยกำหนดขนาดความกว้างคูณความยาวเท่ากับ 352 x 288 จุดภาพ (pixels) คลิปวิดีโอทั้งหมดจะแบ่งตามอารมณ์หลักที่ปรากฏอยู่ในคลิป ดังแสดงในตารางที่ 3.1 คือ กลุ่มอารมณ์กลัว 35 คลิป กลุ่มอารมณ์อื่นๆ ประกอบด้วยกลุ่มอารมณ์ตื่นเต้น อารมณ์สนุกสนาน และอารมณ์เศร้ากลุ่มละ 15 คลิป

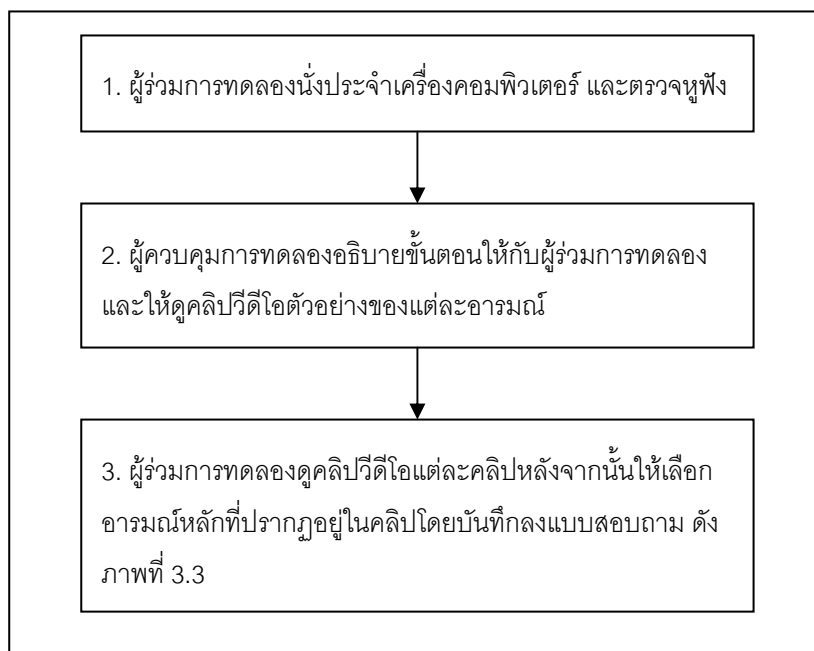
ตารางที่ 3.1
คลิปวิดีโอที่ใช้ในงานวิจัย

ลักษณะของอารมณ์			
อารมณ์กลัว	กลุ่มอารมณ์อื่นๆ		
	อารมณ์ตื่นเต้น	อารมณ์สนุกสนาน	อารมณ์เศร้า
35 คลิป	15 คลิป	15 คลิป	15 คลิป

3.1.2 การจำแนกกลุ่มอารมณ์หลักในคลิปวิดีโอ

การจำแนกกลุ่มอารมณ์หลักในคลิปวิดีโอเพื่อใช้เป็นหลักในการวิจัย (Ground Truth) โดยให้ผู้ร่วมการทดลองเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่หกจากโรงเรียนอัสสัมชัญธนบุรี จำนวนทั้งสิ้น 32 คน ผู้ร่วมการทดลองแต่ละคนจะดูคลิปวิดีโอทั้งหมด 80 คลิป โดยใช้เวลาประมาณสองชั่วโมงตามขั้นตอนในภาพที่ 3.2

ภาพที่ 3.2
สรุปขั้นตอนของผู้ร่วมการทดลอง



ภาพที่ 3.3
ตัวอย่างแบบสอบถามการแสดงข้อคิดเห็นของผู้ร่วมการทดลอง

ชุดที่1-32....

เพศ ☐ ชาย ☐ หญิง อายุ ปี ระดับการศึกษา

Clip#	อาการหลักที่ท่านได้รับ				คลิปวิดีโอที่ท่านเคยดูหรือไม่		
	กลัว	ตื่นเต้น	สับสน	เศร้า	เคย	ไม่เคย	ไม่แน่ใจ
Clip_01							
Clip_02							
Clip_03							
...							
Clip_80							

หลังจากผู้ร่วมการทดลองทุกคนทำการทดลองเสร็จสิ้นจึงนำผลการทดลองที่ได้ทั้งหมดมาบันทึกลงตารางดังแสดงในตารางที่ 3.2 เพื่อเลือกคลิปวิดีโอที่มีกลุ่มอารมณ์ในคลิปวิดีโอเดียวกันตรงกันมากกว่า 75 เปอร์เซนต์เพื่อนำมาใช้ในการฝึกระบบและทดสอบระบบต่อไป

ตารางที่ 3.2

การเลือกคลิปวิดีโอเพื่อใช้ในการวิจัย

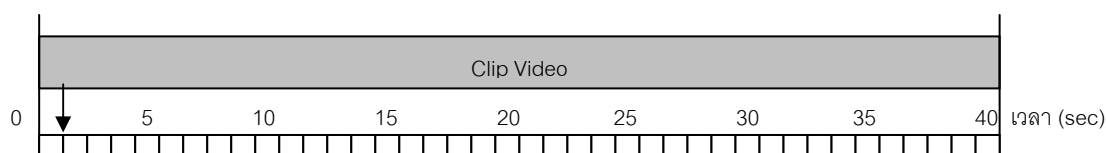
Clip#	Movie_Scene#	Emotion				Fear / Non-Fear	ชุดที่ 1-32			Classification(%)			
		F	E	J	S		1	...	32	F	E	J	S
Clip_01	SLITHER_ 1	✓				Fear	F	...	E				
Clip_80	SLITHER: 2		✓			Non-Fear	F	...	S				

3.2 ขั้นตอนการดึงเฟรมตามช่วงเวลาที่กำหนด

ในขั้นตอนนี้จะนำคลิปวิดีโอทั้งหมดที่ผ่านเกณฑ์จากขั้นตอนที่ผ่านมาทำการดึงเฟรมทุกๆหนึ่งวินาที (second) เพื่อเตรียมการแบ่งข้อในขั้นตอนต่อไปดังตัวอย่างภาพที่ 3.4 เป็นการดึงเฟรมทั้งหมด 40 เฟรม

ภาพที่ 3.4

การดึงเฟรมทุกๆ หนึ่งวินาที



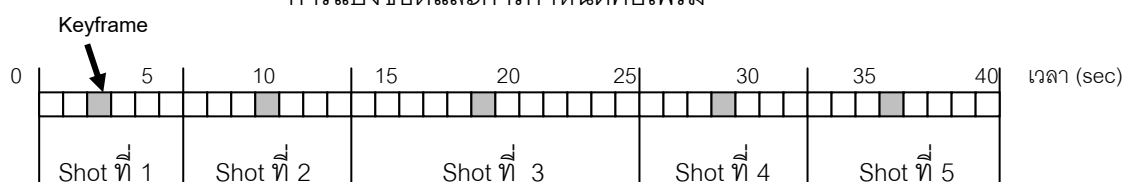
3.3 ขั้นตอนการแบ่งช็อตและการกำหนดคีย์เฟรม (Key Frame)

หลังจากนั้นจะนำเฟรมที่ดึงมาได้ทั้งหมดในแต่ละคลิปเปรียบเทียบความแตกต่างระหว่างเฟรมที่ติดกันเพื่อทำการแยกช็อตโดยเฟรมที่อยู่ในช็อตเดียวกันจะมีความคล้ายคลึงกัน และกำหนดให้เฟรมที่อยู่ตรงกลางของช็อตเป็นคีย์เฟรมเพื่อเป็นตัวแทนของช็อตนั้นดังภาพที่

3.5 กรณีช็อตมีจำนวนเฟรมเป็นเลขคู่คีย์เฟรมที่อยู่ตรงกลางก็คือเฟรมใดก็ได้ที่อยู่กลางช็อต ตัวอย่างการแบ่งช็อตและการกำหนดคีย์เฟรมของคลิปวิดีโอกลุ่มอารมณ์กลัว ดังภาพที่ 3.6

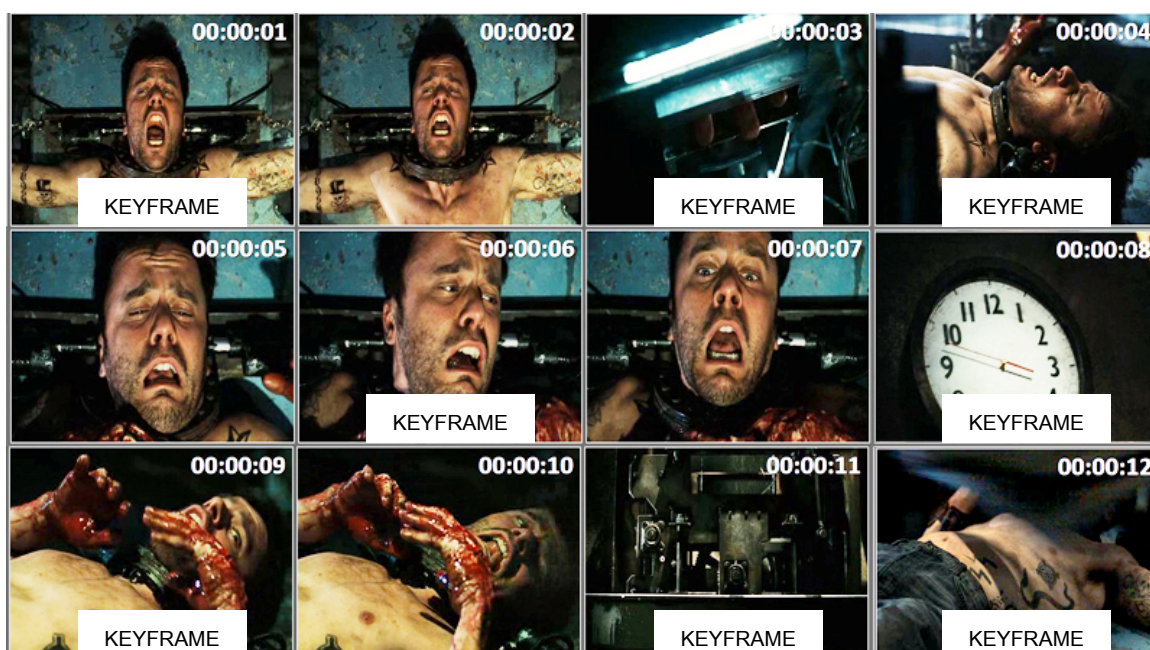
ภาพที่ 3.5

การแบ่งช็อตและการกำหนดคีย์เฟรม



ภาพที่ 3.6

ตัวอย่างการแบ่งช็อตและการกำหนดคีย์เฟรม



3.4.1 ค่าเปอร์เซ็นต์เฉลี่ยแสงในคลิปวิดีโอ (Average percent of light in the clip : APL)

1) นำคีย์เฟรมทั้งหมดแปลงค่าสีจาก RGB เป็น HSV (Hue , Saturation and Value) ในงานวิจัยใช้ค่าความสว่างของสี (Value) เพื่อหาคำนวนค่าเปอร์เซ็นต์เฉลี่ยของแสงในคลิป จากสมการที่ 3.1 (Saowaluk, 2008)

โดย F = จำนวนคีย์เฟรมทั้งหมด
 P = จำนวนจุดภาพทั้งหมดของคีย์เฟรม สามารถคำนวณได้จากสมการที่ 3.2
 V = ค่า Value ซึ่งอยู่ระหว่าง 0 ถึง 1

โดย m = ขนาดความสูงของคีย์เฟรม
 n = ขนาดความกว้างของคีย์เฟรม

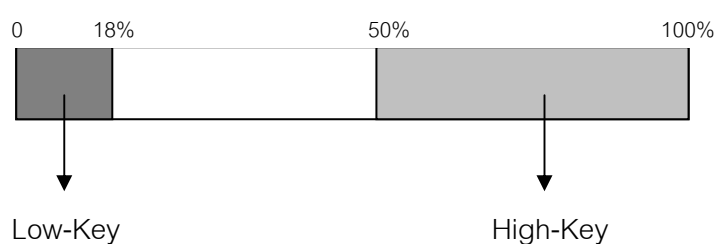
2) นำค่าเปอร์เซ็นต์เฉลี่ยของแสงในคลิปมาจัดเกณฑ์ดังภาพที่ 3.7

APL ≤ 18 แสดงว่า Low-Key (Wang & Cheong, 2006)

APL ≥ 50 แสดงว่า High-Key (Saowaluk, 2008)

ภาพที่ 3.7

เกณฑ์การจำแนกระหว่าง Low-Key และ High-Key



Low-Key หมายถึงภาพโทนมืดมีงานวิจัยที่กล่าวถึงกลุ่มอารมณ์ที่ใช้เกณฑ์นี้สองกลุ่มอารมณ์คือ กลุ่มอารมณ์กลัว (Huang, 2007 ; Rasheed, 2005 ; Wang & Cheong, 2006) และอารมณ์เศร้า (Dunker, 2008 ; Saowaluk, 2008 ; Soleymani, 2008)

High-Key หมายถึงภาพโทนสว่างมีงานวิจัยที่กล่าวถึงกลุ่มอารมณ์ที่ใช้เกณฑ์นี้สองกลุ่มอารมณ์คือกลุ่มอารมณ์ดีใจ (Huang, 2007 ; Rasheed, 2005) และกลุ่มอารมณ์สนุกสนาน (Dunker, 2008 ; Saowaluk, 2008 ; Soleymani, 2008)

3.4.2 ค่าเฉลี่ยความยาวในแต่ละช็อต (Average shot duration : ASD)

การคำนวณค่าเฉลี่ยความยาวในแต่ละช็อตเริ่มจากการนำจำนวนเฟรมทั้งหมดของแต่ละคลิปหารด้วยจำนวนของช็อตทั้งหมดในคลิปนั้น (Rasheed, 2005 ; Saowaluk, 2008) ตามสมการที่ 3.3

$$ASD = \frac{\text{Total}_{\text{Frame}}}{\text{Total}_{\text{Shot}}} \quad (3.3)$$

โดย $\text{Total}_{\text{Frame}}$ คือ จำนวนเฟรมทั้งหมดในคลิป
 $\text{Total}_{\text{Shot}}$ คือ จำนวนของช็อตทั้งหมดในคลิป

การหาจำนวนเฟรมทั้งหมดในคลิปของงานวิจัยนี้สามารถคำนวณจากเวลาทั้งหมดของแต่ละคลิปแทนได้เนื่องจากกำหนดให้มีการดิงเฟรมทุกๆหนึ่งวินาทีในขั้นตอนที่ 3.2 และจำนวนของข้อต่อทั้งหมดในคลิปหมายถึงจำนวนทั้งหมดของคีย์เฟรมในแต่ละคลิبنั้นคำนวณได้จากขั้นตอนที่ 3.3

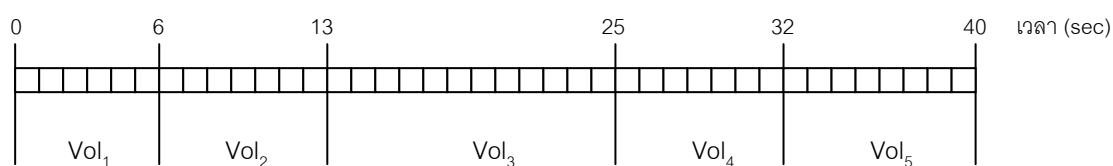
มีงานวิจัยของ (Wang & Cheong ,2006 ; Saowaluk, 2008 ; Soleymani, 2008) อธิบายค่าเฉลี่ยความยาวในแต่ละข้อต่อในทิศทางเดียวกันโดยเรียงจากค่ามากที่สุดไปน้อยที่สุด ดังนี้อารมณ์เศร้า อารมณ์สนุกสนาน และอารมณ์ตื่นเต้น

3.4.3 ค่าความแตกต่างระหว่างค่าเฉลี่ยความดังเสียงที่ดังมากที่สุดของข้อต่อกับค่าเฉลี่ยความดังเสียงของข้อต่อก่อนหน้า (Difference average of maximum volume shot and its predecessor : DAV)

เริ่มจากการนำคลิปเสียงที่ถูกแยกออกจากไฟล์วิดีโอซึ่งอยู่ในรูปแบบไฟล์ wav อัตราการสุ่มตัวอย่าง (Sampling Rate) เท่ากับ 44,100 เฮิร์ตซ์ (Hz) ระดับความลึกของเสียง (Bit Depth) 16 บิต ช่องของเสียง (Channel) เป็นแบบหนึ่งช่องทาง (mono) การคำนวณมีขั้นตอนดังนี้

ภาพที่ 3.8

การหาค่าเฉลี่ยความดังเสียงของแต่ละข้อต่อ



- 1) คำนวณหาค่าเฉลี่ยความดังเสียงของแต่ละข้อต่อดังภาพที่ 3.8 จากสมการที่ 3.4

$$Vol_n = \frac{\sum_{i=T1}^{T2} S_n(i)}{I} \quad (3.4)$$

โดย Vol_n คือ ค่าเฉลี่ยความดังเสียงของข้อต่อที่ n

$T1$ คือ เวลาเริ่มต้นของข้อตุดูณด้วยอัตราการสุ่มตัวอย่าง

T2 คือ เวลาสิ้นสุดของข้อต่อนด้วยอัตราการสุ่มตัวอย่าง

I คือ จำนวนการสุ่มตัวอย่าง (Sample) คำนวณจาก $T2 - T1$

$S_n(i)$ คือ ค่าความดังเสียงของการสุ่มตัวอย่างที่ i ของข้อที่ n

2) จากนั้นนำค่าเฉลี่ยความดังเสียงที่ดังมากที่สุดของข้อต้นนำมาหาความแตกต่างจากค่าเฉลี่ยความดังเสียงของข้อต่อก่อนหน้าจากสมการที่ 3.5

$$DAV = \text{Max}_{Voln} - \text{Max}_{Voln-1} \quad (3.5)$$

โดย Max_{Voln} คือ ค่าเฉลี่ยความดังเสียงที่มากที่สุดของข้อที่ n

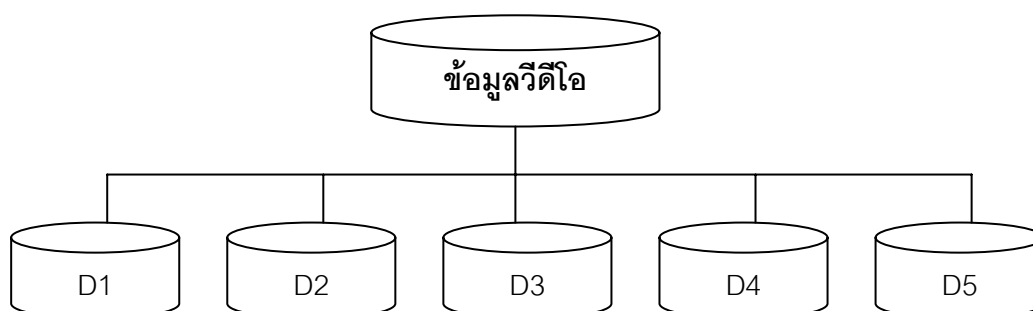
Max_{Voln-1} คือ ค่าเฉลี่ยความดังเสียงก่อนหน้า Max_{Voln}

3.5 การฝึกระบบ และการทดสอบระบบ

งานวิจัยนี้จะแบ่งข้อมูลออกเป็นสองชุดได้แก่ข้อมูลสำหรับฝึกระบบ (Training Set) และข้อมูลสำหรับการทดสอบ (Test Set) เพื่อไม่ให้เกิดความโน้มเอียงในการแบ่งชุดฝึกระบบกับชุดทดสอบระบบจึงใช้วิธีการทดสอบความถูกต้องแบบข้าม (cross validation) ดังภาพที่ 3.9 โดยแบ่งข้อมูลออกเป็น 5 ชุดข้อมูลย่อย (fold) คือ D1, D2, D3, D4 และ D5 ในแต่ละชุดข้อมูลย่อยจะประกอบด้วยข้อมูลที่กระจายกันจากนั้นทำการทดลอง 5 ครั้งดังภาพที่ 3.10 โดยครั้งที่หนึ่งจะมีชุดข้อมูล D2, D3, D4 และ D5 เป็นชุดข้อมูลสำหรับการฝึกระบบและชุดข้อมูล D1 เป็นชุดข้อมูลสำหรับการทดสอบความถูกต้อง การทดลองครั้งที่สองชุดข้อมูล D1, D3, D4 และ D5 เป็นชุดข้อมูลสำหรับการฝึกระบบและชุดข้อมูล D2 เป็นชุดข้อมูลสำหรับการทดสอบความถูกต้องทำแบบนี้ตามลำดับไปจนถึงการทดลองครั้งที่ห้า

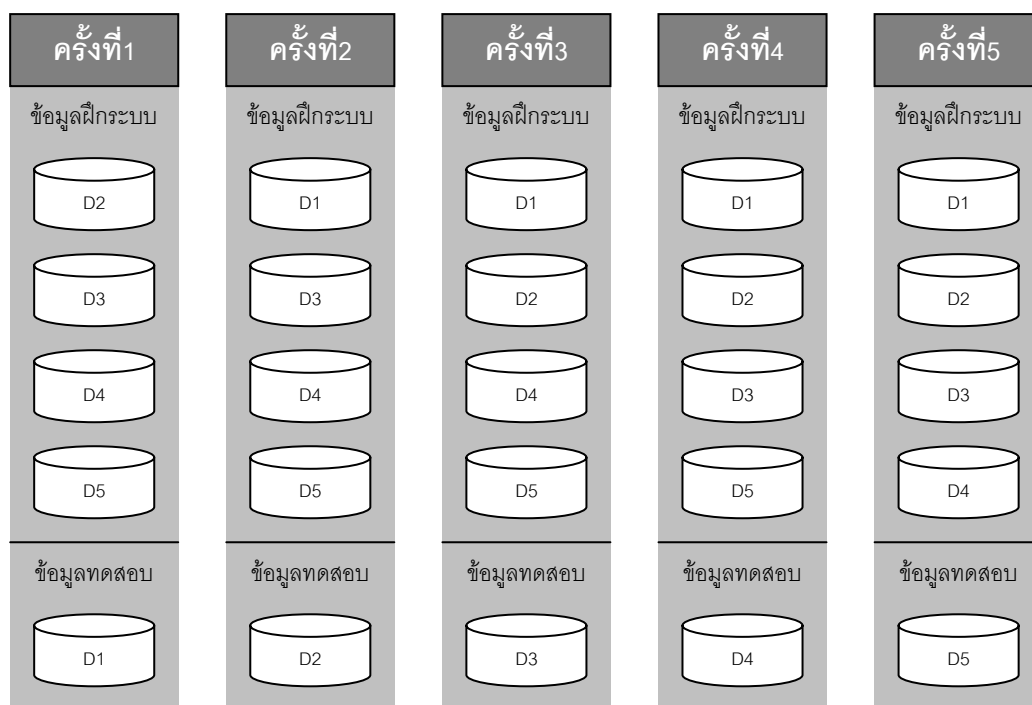
ภาพที่ 3.9

การแบ่งข้อมูลออกเป็น 5 ชุดข้อมูลย่อย (fold)



ภาพที่ 3.10

วิธีการทดสอบความถูกต้องแบบข้าม (cross validation)

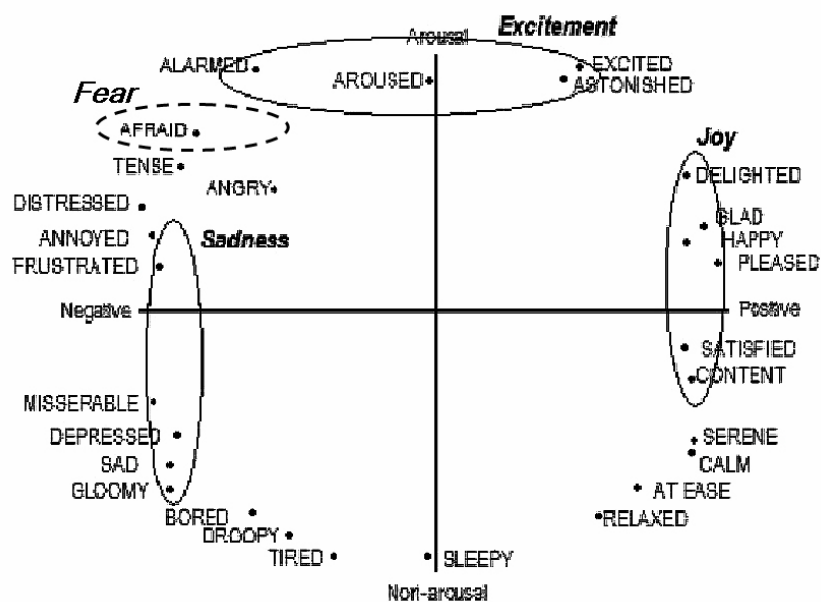


3.6 การจำแนก (Classification)

งานวิจัยได้จำแนกคลิปวิดีโอตามอารมณ์หลักของผู้ชมออกเป็นสองกลุ่มได้แก่กลุ่มอารมณ์กลัว (Fear) และกลุ่มอารมณ์อื่นๆ ได้แก่อารมณ์ตื่นเต้น (Excitement) อารมณ์สนุกสนาน (Joy) และอารมณ์เศร้า (Sadness) โดยกลุ่มอารมณ์กลัวเป็นอารมณ์ที่อยู่ระหว่างกลุ่มอารมณ์ที่ถูกกระตุ้นกับกลุ่มอารมณ์เชิงลบ (Russell, 1980 ; Wang & Cheong, 2006 ; Xu, Jin, Luo, & Duan, 2008) ดังภาพที่ 3.11 โดยงานวิจัยเลือกเทคนิคการจำแนกแบบโครงข่ายประสาทเทียมโดยใช้การเรียนรู้แบบแพร่กระจายย้อนกลับ (Backpropagation Learning) กำหนดให้สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron: MLP) ดังภาพที่ 3.12 มีโครงสร้างดังนี้ชั้นข้อมูลเข้า (Input Layer) มีจำนวนนิวรอนเท่ากับ 3 คือ APL, ASD, DAV ชั้นซ่อน (Hidden Layer) มีจำนวนนิวรอนเท่ากับ 2 และชั้นผลลัพธ์ (Output) มีจำนวนนิวรอนเท่ากับ 2 คือ Fear และ Non-Fear

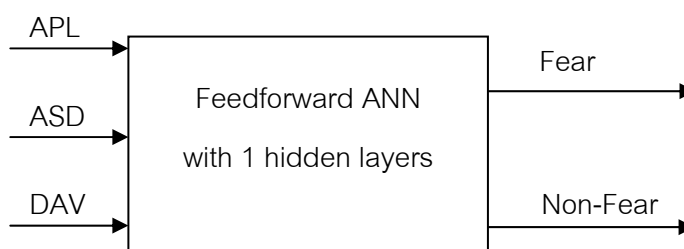
ภาพที่ 3.11

โมเดลกลุ่มอารมณ์สำหรับงานวิจัย



ภาพที่ 3.12

โมเดลสำหรับการจำแนก



APL Average Percent of Light in the clip

ASD Average Shot Duration

DAV Difference Average of maximum Volume shot and its predecessor

3.7 เครื่องมือและซอฟต์แวร์ที่ใช้ในการวิจัย

เครื่องมือและซอฟต์แวร์ที่ใช้ในการวิจัยแสดงในตารางที่ 3.3 ขั้นตอนเริ่มจากการเตรียมข้อมูลเกี่ยวกับวีดีโอและเสียงจากนั้นนำมาหาคุณลักษณะสำคัญโดยใช้โปรแกรม MATLAB และการจำแนกคลิปภาพยนตร์ระหว่างกลุ่มอารมณ์กลัวและกลุ่มอารมณ์อื่นๆใช้โปรแกรม WEKA ซึ่งทำงานบนระบบปฏิบัติการ Windows XP

ตารางที่ 3.3 เครื่องมือและซอฟต์แวร์ที่ใช้ในการวิจัย

ฮาร์ดแวร์ (Hardware)	CPU Intel Core 2 Duo E4600 2.4GHz RAM 2 GB
ระบบปฏิบัติการ (Operating System)	Microsoft Windows XP Professional Service Pack 2
ซอฟต์แวร์ (Software)	MATLAB Version R2008b Adobe Premiere Pro CS3 AVI MPEG RM WMV Splitter Video Thumbnails Maker PhotoScape WEKA 3.6