

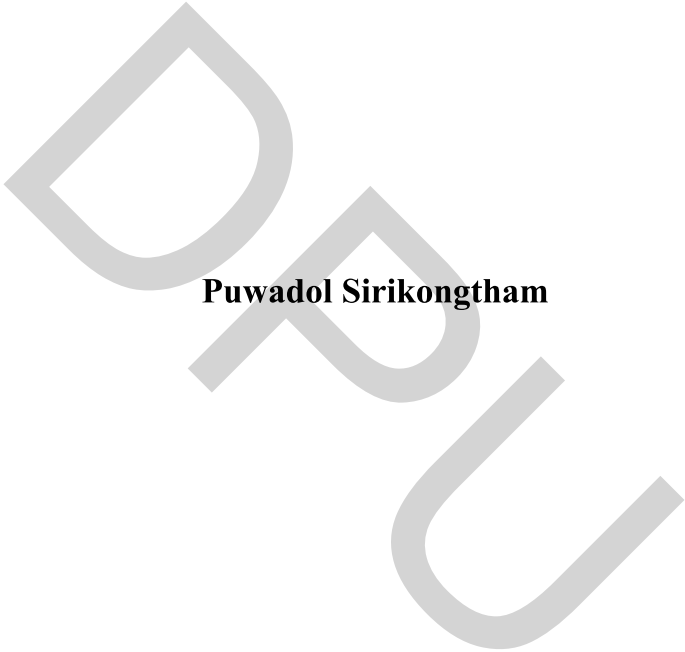
เทคนิคสำหรับการควบคุมอินเทอร์เน็ตที่วีโดยกีเนค
ด้วยการผสมผสานระหว่างความรู้จำเสี่ยงและการตรวจสอบการเคลื่อนไหว

ภูวดล ศิริกองธรรม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2556

**A Technique for Internet Television Control using Kinect
Combination of Voice Recognition and Motion Detection Approach**



Puwadol Sirikongtham

**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Web Engineering
Faculty of Information Technology, Dhurakij Pundit University**

2013

หัวข้อวิทยานิพนธ์	เทคนิคสำหรับการควบคุมอินเทอร์เนตทีวีโดยคีนค ด้วยการผสมผสานระหว่างการรู้จำเสียงและการ ตรวจสอบการเคลื่อนไหว
ชื่อผู้เขียน	ภูวดล ศิริกองธรรม
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.วรสิทธิ์ ชูชัยวัฒนา
สาขาวิชา	วิศวกรรมเว็บ
ปีการศึกษา	2555

บทคัดย่อ

การรู้จำเสียงหรือการตรวจสอบการเคลื่อนไหวได้ถูกนำมาช่วยทำให้การควบคุมอินเทอร์เนตทีวี มีประสิทธิภาพและสะดวกสบายมากขึ้น แต่เทคโนโลยีดังกล่าวที่กล่าวมานั้น ต่างมีจุดเด่นและข้อด้อยในตัว ผู้วิจัยจึงเสนอเทคนิคที่ใช้ควบคุมอินเทอร์เนตทีวีแบบใหม่ด้วยการผสมผสานจุดเด่นระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว เพื่อให้การควบคุมอินเทอร์เนตทีวีมีประสิทธิภาพมากยิ่งขึ้น ในการทำการทดลองเพื่อวัดประสิทธิภาพ ผู้วิจัยทำการเปรียบเทียบเทคนิคที่นำเสนอ ซึ่งได้แก่การผสมผสานระหว่างการรู้จำเสียงและท่าทางโดยทำการทดลอง 2 ครั้ง โดยครั้งแรกจะทำการทดลองแบบเบื่องต้น ในสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวน ซึ่งค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 เดซิเบลและสภาพแวดล้อมที่มีเสียงรบกวน โดยกำหนดให้เสียงรบกวนอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยิน รับรู้เสียงที่ออกมาจากอินเทอร์เนตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่ เฉลี่ย 70 เดซิเบล และการทดลองครั้งที่สอง จะใช้โจทย์ทดสอบให้มีความใกล้เคียงกับการควบคุมทีวีในปัจจุบันมากที่สุดและมีเสียงที่ดังออกมาจากตัวทีวี ความดังของเสียงอยู่ที่ 70 dB จากการทดลอง พบว่าขั้นตอนและวิธีการควบคุมอินเทอร์เนตทีวีด้วยการผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว ให้ผลลัพธ์ที่ดีกว่าในด้านของความเร็วและความถูกต้อง กับใช้เสียงสั่งงานเพียงอย่างเดียว ในสภาพแวดล้อมที่มีเสียงรบกวน แม้ว่าวิธีการตรวจสอบการเคลื่อนไหวเป็นวิธีที่ใช้เวลาน้อยที่สุด และมีความถูกต้องสูงที่สุดก็ตาม วิธีดังกล่าวอาจจะไม่เหมาะสมสำหรับผู้ที่มีปัญหาเกี่ยวกับการเคลื่อนไหว

Thesis Title	A Technique for Internet Television Control using Kinect Combination of Voice Recognition and Motion Detection Approach
Author	Puwadol Sirikongtham
Thesis Advisor	Asst.Prof.Dr. Worasit Choochaiwattana
Academic Program	Web Engineering
Academic Year	2012

ABSTRACT

Voice recognition and motion detection approach were applied to Internet television controlling and made the task more efficiency and more convenient. These two approaches have their own distinctive points and drawbacks. This research aims at proposing a new technique for Internet television controlling by combing distinctive points of voice recognition and motion detection approach to make the proposed technique more efficient. Two experiments have been conducted to evaluation the proposed technique by comparing with the voice recognition and motion detection. Participants in both experiments were given tasks of controlling the television. In the first experiment, the proposed technique was evaluated in two environmental settings, which were quiet environment (46 dB) and noisy environment (70 dB). In the second experiment, the longer and more realistic tasks of controlling the television were assigned to the participants under the noisy environment. The results showed that the combination of the voice recognition and motion detection approach outperformed the voice recognition only approach in term of speed and accuracy under noisy environment. Although, the motion detection approach consumes shortest time and provides more accuracy of the television controlling task, this approach is not appropriate for people who have a motion disability.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยความอนุเคราะห์จาก ผศ.ดร.วรสิทธิ์ ชูชัยวัฒนา ผู้จัดทำวิทยานิพนธ์ใคร่ขอกราบขอบพระคุณที่ท่านอาจารย์กรุณาให้คำปรึกษา ให้ข้อคิดเห็น และสละเวลาให้กับผู้จัดทำวิทยานิพนธ์ตลอดมา ตลอดจนช่วยตรวจสอบต้นฉบับและแก้ไขข้อบกพร่องของงานวิจัยเพื่อให้วิทยานิพนธ์ฉบับนี้เสร็จสิ้นไปได้ด้วยดี

ขอขอบคุณนักศึกษา สถาบันไทย-ญี่ปุ่นที่ช่วยสละเวลาเข้าร่วมเป็นผู้ทดสอบ ในงานวิจัย ให้ผู้จัดทำวิทยานิพนธ์เก็บข้อมูลที่สถาบันอย่างไม่รู้จักเหน็ดเหนื่อย

ผู้วิจัยขอกราบขอบพระคุณคณาจารย์ คณะเทคโนโลยีสารสนเทศ สาขาวิศวกรรมเว็บ มหาวิทยาลัยธุรกิจบัณฑิตย์ ด้วยความเคารพอย่างสูง ผู้วิจัยรู้สึกซาบซึ้งยิ่งนัก จึงขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ภูวดล ศิริกองธรรม

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ผ
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ณ
บทที่	
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของงาน.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	2
1.3 ประโยชน์และผลที่คาดว่าจะได้รับ.....	2
1.4 ขอบเขตการศึกษา.....	2
1.5 นิยามคำศัพท์.....	3
2. วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 คิเนค (Kinect).....	4
2.2 Extensible Application Markup Language XAML.....	9
2.3 การตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect และ Kinect SDK Beta 1.....	9
2.4 การรู้จำเสียง (Speech Recognition).....	12
2.5 Microsoft Speech Application Programming Interface (SAPI)	14
2.6 งานวิจัยที่เกี่ยวข้อง.....	17
3. วิธีการดำเนินการและเครื่องมือ.....	26
3.1 การวางแผนการดำเนินงาน.....	26
3.2 การศึกษาค้นคว้าหาข้อมูล.....	28
3.3 การออกแบบและพัฒนา Interface ส่วนควบคุมอินเทอร์เนตที่วีจ่าลอง.....	29
3.4 การพัฒนาระบบการรู้จำเสียง.....	30
3.5 การพัฒนาระบบการตรวจจับการเคลื่อนไหว.....	33
3.6 การพัฒนาระบบเทคนิคการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว.....	35

สารบัญ (ต่อ)

บทที่	หน้า
3.7 การกำหนดแบบแผนการทดลองและกลุ่มผู้ทดสอบ.....	37
3.8 การวัดผลทั้งด้านความเร็วและความถูกต้องในการควบคุมอินเทอร์เนตทีวี จำลอง.....	40
3.9 เครื่องมือที่ใช้จัดทำระบบควบคุมอินเทอร์เนตทีวีจำลอง.....	41
4. ผลการดำเนินงาน.....	43
4.1 ผลการวิจัยการทดสอบเบื้องต้นในสภาพแวดล้อมที่มีเสียงรบกวนและไม่มีเสียง รบกวน.....	43
4.2 ผลการวิจัยการทดสอบครั้งที่ 2 ปรับปรุงโจทย์ทดสอบให้มีคำสั่งมากขึ้นและมี เสียงรบกวนที่ดังออกจากทีวี.....	54
5. สรุปอภิปรายผลการศึกษาและข้อเสนอแนะ.....	60
5.1 สรุปผลและวิจารณ์.....	60
5.2 ปัญหาที่พบในงานวิจัย.....	61
5.3 ข้อเสนอแนะในการศึกษาขั้นต่อไป.....	62
บรรณานุกรม.....	63
ภาคผนวก	
ก วิธีติดตั้ง Kinect for Windows SDK.....	67
ข ตารางบันทึกผลการทดลองครั้งที่ 1.....	73
ค ตารางบันทึกผลการทดลองครั้งที่ 2.....	85
ง ผลงานวิจัยในการประชุมวิชาการ การบริหารและการจัดการ ครั้งที่ 8 “การบริหารจัดการแห่งโลกพลวัต” โดยมหาวิทยาลัยธุรกิจบัณฑิตฯ ร่วมกับ เครือข่ายวิจัยประชาชน.....	93
จ หนังสือรับรองผลงานวิจัยในการประชุมวิชาการ การบริหารและการจัดการครั้งที่ 8 “การบริหารจัดการแห่งโลกพลวัต” โดยมหาวิทยาลัยธุรกิจบัณฑิตฯ ร่วมกับ เครือข่ายวิจัยประชาชน.....	106
ประวัติผู้เขียน	109

สารบัญตาราง

ตารางที่	หน้า
2.1 แสดงความสามารถพื้นฐานของกล้อง Kinect.....	7
3.1 ระยะเวลาในการดำเนินการพัฒนาระบบ.....	27
3.2 แสดงถึงงานที่ผู้ทดสอบครั้งที่ 1 ต้องทำในแต่ละชุดการทดสอบในทำยีน.....	37
3.3 แสดงถึงงานที่ผู้ทดสอบครั้งที่ 2 ต้องทำในแต่ละชุดการทดสอบในทำยีน.....	39
3.4 ตารางแสดงลำดับการควบคุมในแต่ละเทคนิคการควบคุมอินเทอร์เนตทีวี.....	39
4.1 แสดงถึงงานที่ผู้ทดสอบเบื้องต้นในสภาพแวดล้อมที่มีเสียงรบกวนและไม่มีเสียงรบกวน ต้องทำในแต่ละชุดการทดสอบในทำยีน.....	44
4.2 แสดงถึงผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยีน.....	47
4.3 แสดงถึงค่า Extreme จากการทดสอบเบื้องต้นแบบไม่มีเสียงรบกวน (46 dB) ในทำยีน.....	48
4.4 ตารางสรุปค่าเฉลี่ยเวลาที่ใช้ และจำนวนครั้งที่ผิดพลาดในแต่ละเทคนิคผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยีน.....	49
4.5 แสดงถึงผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำยีน.....	50
4.6 แสดงถึงค่า Extreme จากการทดสอบเบื้องต้นแบบมีเสียงรบกวน (70 dB) ในทำยีน.....	51
4.7 ตารางสรุปถึงค่าเฉลี่ยเวลาที่ใช้ และจำนวนครั้งที่ผิดพลาดในแต่ละเทคนิคผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำยีน.....	52
4.8 ตารางแสดงลำดับการควบคุมในแต่ละเทคนิคการควบคุมอินเทอร์เนตทีวี.....	55
4.9 แสดงถึงผลการทดสอบครั้งที่ 2 แบบมีเสียงรบกวน (70 dB) ในทำยีน.....	56
4.10 แสดงค่า Extreme Value จากการทดสอบครั้งที่ 2 แบบมีเสียงรบกวน (70 dB) ในทำยีน.....	57
4.11 ตารางสรุปถึงค่าเฉลี่ยเวลาที่ใช้ และจำนวนครั้งที่ผิดพลาดในแต่ละเทคนิคผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำยีน.....	58

สารบัญภาพ

ภาพที่	หน้า
2.1 สถาปัตยกรรมบนกล้อง Kinect.....	4
2.2 แสดงค่าสีที่ได้จากกล้อง Kinect.....	5
2.3 แสดงอินฟราเรดที่ใช้หาความลึกของภาพ.....	5
2.4 ไมโครโฟนบนตัว Kinect.....	6
2.5 แสดงการแยกรูปร่างของวัตถุโดยใช้คุณสมบัติของความลึก.....	6
2.6 แสดงสถาปัตยกรรม Kinect SDK (1).....	8
2.7 แสดงสถาปัตยกรรม Kinect SDK (2).....	8
2.8 แกน X, Y และ Z ของกล้อง Kinect เมื่อใช้ Kinect SDK Beta.....	9
2.9 แสดงแกนทางด้าน Positive และ Negative.....	10
2.10 ตำแหน่งของข้อต่อที่สัมพันธ์กับร่างกายมนุษย์.....	10
2.11 รวมการทำงานของ Speech Recognition.....	13
2.12 แสดงมุมมองและทิศทางในการรับเสียงจาก Kinect บน SAPI 5.1.....	16
2.13 การตรวจจับการเคลื่อนไหวของคนและโมเดล.....	18
2.14 ระบบการลองเสื้อผ้าเสมือนจริง.....	19
2.15 ระบบการบ่งบอกคุณลักษณะภายนอกของบุคคลโดยใช้กล้อง Kinect.....	19
2.16 ตัวอย่างพฤติกรรมของเด็กที่บันทึกโดยใช้กล้อง Kinect.....	21
2.17 ระบบการควบคุมการแสดงรูปภาพทางการแพทย์ด้วยท่าทางของผู้ใช้.....	22
2.18 ระบบแสดงผลที่ควบคุมโดยการแสดงท่าทางและการแสดงออกทางหน้าตาของผู้ใช้.....	23
2.19 ระบบแสดงผลที่ควบคุมโปรแกรมกูเกิ้ลเอิร์ธโดยใช้ท่าทางและเสียง.....	23
2.20 โครงร่างมนุษย์ที่กล้องแต่ละตัวตรวจจับได้ และโครงร่างมนุษย์ประกอบรวม....	24
2.21 ตัวอย่าง Smart TV รุ่น 55ES8000.....	25
3.1 แสดงอินเตอร์เฟซของการระบบควบคุมอินเทอร์เน็ตทีวีจำลอง.....	29
3.2 แสดงวิธีการเปิดปิดระบบการสั่งการด้วยเสียง.....	30
3.3 แสดง Interface ระบบการสั่งงานด้วยเสียง.....	31
3.4 Flow Chart แสดงการทำงานระบบรู้จำเสียง.....	32

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
3.5 แสดงจุดข้อต่อที่ API ส่งค่ามา.....	33
3.6 Flow Chart แสดงการทำงานระบบการตรวจจับการเคลื่อนไหว.....	34
3.7 แสดงรูปแบบการทำงานของวิธีการผสมผสานเทคนิคการรู้จำเสียงและการ ตรวจสอบการเคลื่อนไหว.....	35
3.8 Flow Chart แสดงการทำงานระบบการผสมผสานรู้จำเสียงและการตรวจจับการ เคลื่อนไหว.....	36
4.1 กราฟแสดงค่า Outlier ของข้อมูลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยีน แบบ Voice Only.....	48
4.2 กราฟสรุปผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยีน.....	49
4.3 กราฟแสดงค่า Outlier ของข้อมูลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำยีน แบบ Voice Only.....	51
4.4 กราฟสรุปผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (70dB) ในทำยีน.....	52
4.5 กราฟเปรียบเทียบผลการทดสอบเบื้องต้นในประสิทธิภาพด้านความเร็ว แบบมี เสียงรบกวน (70 dB) และไม่มีเสียงรบกวน (46dB) ในทำยีน.....	53
4.6 กราฟเปรียบเทียบผลการทดสอบเบื้องต้น ประสิทธิภาพด้านความถูกต้องในการ สั่งงาน แบบมีเสียงรบกวน (70 dB) และไม่มีเสียงรบกวน (46dB) ในทำยีน.....	53
4.7 กราฟแสดงค่า Outlier ของข้อมูลทดสอบครั้งที่2 แบบมีเสียงรบกวน (70 dB) ใน ทำยีน แบบ Motion Detection.....	57
4.8 กราฟเปรียบเทียบผลการทดสอบครั้งที่2 ประสิทธิภาพด้านความเร็ว แบบมีเสียง รบกวน (70 dB) ในทำยีน.....	58
4.9 กราฟเปรียบเทียบผลการทดสอบครั้งที่ 2 ประสิทธิภาพด้านความถูกต้องในการ สั่งงาน แบบมีเสียงรบกวน (70 dB) ในทำยีน.....	58

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของงาน

เทคโนโลยีเป็นส่วนสำคัญอย่างมากที่จะทำให้การดำรงชีวิตอยู่ของมนุษย์ สะดวกสบายมากขึ้น โทรทัศน์ก็เป็นอีกหนึ่งของเทคโนโลยีที่ช่วยเพิ่มความสามารถในการรับรู้ข่าวสารและให้ความบันเทิง กับผู้ใช้ตลอดระยะเวลาที่ผ่านมาโทรทัศน์ได้ถูกพัฒนาอย่างต่อเนื่องในเรื่องของการแสดงผลภาพ แต่ก็ยังใช้วิธีการควบคุมโดย เครื่องควบคุมระยะไกล (Remote Control) ในการควบคุมโทรทัศน์มาอย่างยาวนาน ซึ่งเป็นปัญหาในเรื่องของการควบคุมตัวโทรทัศน์ เนื่องจากรีโมทที่ใช้ควบคุมโทรทัศน์เป็นอีกหนึ่งอุปกรณ์ที่แยกจากตัวโทรทัศน์เองและจำเป็นต้องใช้แหล่งพลังงานเสริม และตัวรีโมทที่แยกออกมานั้นอาจมีการสูญหายทำให้การควบคุมโทรทัศน์เสียไป

อย่างไรก็ตามในปัจจุบันเริ่มมีการนำเทคโนโลยีการรู้จำเสียงและท่าทางมาใช้แทนที่สำหรับเครื่องควบคุมระยะไกล (Remote Control) แต่ก็ยังไม่เป็นที่นิยมเท่าที่ควรทั้งในเรื่องของฮาร์ดแวร์ที่มีราคาสูงและ เทคโนโลยีทั้งสองมีข้อด้อยกล่าวคือ เทคโนโลยีการรู้จำเสียงนั้นหากมีเสียงเข้ามารบกวนจะทำให้การทำงานเกิดความผิดพลาด เทคโนโลยีตรวจสอบการเคลื่อนไหวนั้นต้องใช้เวลาค่อนข้างมากในการสั่งงาน แต่ในปัจจุบันเทคโนโลยีทางด้รับรู้ภาพและเสียงได้ถูกพัฒนาไปอีกระดับหนึ่ง โดยทางบริษัท ไมโครซอฟท์สามารถพัฒนาอุปกรณ์คิเนค (Kinect) ซึ่งเป็นอุปกรณ์ที่สามารถรับรู้ข้อมูลมิติความลึกของภาพและเสียงได้ในเวลาเดียวกัน ทำให้เพิ่มศักยภาพในการใช้ควบคุมโทรทัศน์ด้วยการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว

ดังนั้นงานวิจัยนี้ จึงเสนอเทคนิคในการควบคุมอินเทอร์เนตทีวี โดยใช้การผสมผสานระหว่างเทคนิคการรู้จำเสียงและเทคนิคการตรวจสอบการเคลื่อนไหว ผ่านอุปกรณ์คิเนค(Kinect) เพื่อทดสอบถึงความถูกต้องและความเร็วในการควบคุมอินเทอร์เนตทีวีจำลอง ภายใต้สภาพแวดล้อมที่มีเสียงรบกวน

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาตัวแบบสำหรับการควบคุมอินเทอร์เน็ตทีวีจำลองแบบใหม่ด้วยการผสมผสานเทคนิคการรู้จำเสียงและเทคนิคการตรวจจับการเคลื่อนไหว
2. เพื่อทดสอบประสิทธิภาพของวิธีการควบคุมอินเทอร์เน็ตทีวีจำลอง ทั้งสามวิธี คือการรู้จำเสียง การตรวจจับการเคลื่อนไหว และการผสมผสานการรู้จำเสียงและการเคลื่อนไหว

1.3 ประโยชน์และผลที่คาดว่าจะได้รับ

1. ได้ตัวต้นแบบระบบการตรวจที่ใช้ควบคุมอินเทอร์เน็ตทีวีที่รองรับการใช้กล้อง Kinect
2. ได้วิธีการผสมผสานเทคนิคการรู้จำเสียงและการตรวจจับการเคลื่อนไหว โดยใช้กล้อง Kinect ที่มีผลลัพธ์ที่ขอปรับได้ สามารถนำไปใช้ในการควบคุมอินเทอร์เน็ตทีวี ได้ในอนาคตต่อไป

1.4 ขอบเขตการศึกษา

1.4.1 ขอบเขตของระบบ

1. ระบบที่จำลองขึ้นสามารถควบคุมได้ทั้งเทคนิคการรู้จำเสียง การตรวจจับการเคลื่อนไหว และเทคนิคการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว โดยแบ่งระบบออกเป็นสามแอปพลิเคชันตามแต่ละเทคนิคการทดสอบ
2. ระบบที่จำลองขึ้นสามารถรองรับการสั่งการได้ดังนี้ เปลี่ยนช่อง 3,5,7,9,11 และ Thai PBS เพิ่มลดเสียงได้ 5 ระดับ
3. ระบบที่จำลองขึ้นในช่องต่างๆที่จะแสดงผลให้กับผู้ทดสอบเป็นการจำลองโดยใช้วิดีโอในการแสดงผลภาพ ไม่ได้ดึงจากสัญญาณอินเทอร์เน็ต เนื่องจากการทดสอบบางครั้งระดับสัญญาณอินเทอร์เน็ตไม่คงที่อาจทำให้ผลการทดสอบเปลี่ยนไปได้
4. สภาพแวดล้อมที่ใช้ทดสอบจะแบ่งเป็น 2 สภาพแวดล้อม ที่ไม่มีเสียงรบกวนเฉลี่ยอยู่ที่ 46 เดซิเบล และมีเสียงรบกวนเฉลี่ยอยู่ที่ 70 เดซิเบล ทำการวัดหาค่าเฉลี่ยโดยใช้ แอปพลิเคชันบนแอนดรอยด์สมาร์ตโฟนชื่อ “Sound Meter” ในการตรวจสอบค่าเฉลี่ยของระดับเดซิเบล
5. การวัดประสิทธิภาพด้านความเร็วและความถูกต้องในการควบคุมอินเทอร์เน็ตทีวีจำลองนั้น จะเป็นแบบจดบันทึกข้อมูลที่ได้นบนกระดาษ แล้วจึงนำค่าที่ได้มาคำนวณวัดผลการทดสอบ

1.4.2 ประชากรกลุ่มตัวอย่าง

ประชากรกลุ่มตัวอย่างที่ใช้ในงานวิจัยชิ้นนี้เป็นนักศึกษาชั้นปีที่ 1-3 ช่วงอายุประมาณ 19-21 ปี สถาบันเทคโนโลยีไทย-ญี่ปุ่น จำนวน 60 คน เพื่อใช้ทดสอบประสิทธิภาพของวิธีการควบคุมอินเทอร์เน็ตทีวีจำลอง ทั้งสามวิธี คือการรู้จำเสียง การตรวจจับการเคลื่อนไหว และการผสมผสานการรู้จำเสียงและการเคลื่อนไหว โดยจะแบ่งการทดสอบเป็น 2 กลุ่ม กลุ่มละ 30 คน โดยให้กลุ่มแรกทำการทดสอบในสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 เดซิเบล และกลุ่มที่สองอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เน็ตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 เดซิเบล แล้วประเมินผลในเรื่องของความเร็วและความถูกต้องในการทำงาน

1.5 นิยามศัพท์

ในวิทยานิพนธ์ฉบับนี้ได้ใช้นิยามศัพท์ที่ใช้ในการวิจัยไว้ ดังนี้

การรู้จำเสียง (Speech Recognition) หมายถึง การทำให้คอมพิวเตอร์สามารถที่จะฟังคำพูดและตัดสินใจได้ว่าคำพูดนั้นเป็นคำว่าอะไรหรืออีกความหมายหนึ่งคือการนำ File Audio ที่บรรจุเสียงพูดนำมาแปลงเป็น Text ได้

การตรวจจับการเคลื่อนไหว (Motion Detection) หมายถึง ระบบที่มีไว้สำหรับตรวจสอบการเคลื่อนไหวของวัตถุเพื่อใช้ใน ระบบการตรวจตรา จับการเคลื่อนไหว

ไมโครซอฟท์ คิเนค (Microsoft Kinect) หมายถึง อุปกรณ์รับรู้การเคลื่อนไหวทั้งภาพและเสียงที่บริษัทไมโครซอฟท์พัฒนาขึ้น ใช้การทำงานผสมผสานกันระหว่างฮาร์ดแวร์และซอฟต์แวร์

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

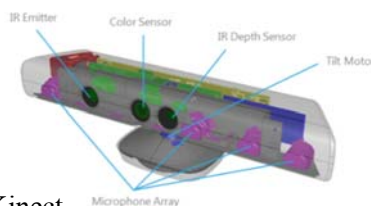
งานวิจัยเรื่อง “เทคนิคสำหรับการควบคุมอินเทอร์เน็ตทีวีโดยคีนคด้วยการผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว” ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดดังนี้

- 2.1 คีนค (Kinect)
- 2.2 Extensible Application Markup Language (XAML)
- 2.3 การตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect และ Kinect SDK Beta 1
- 2.4 การรู้จำเสียง (SPEECH RECOGNITION)
- 2.5 งานวิจัยที่เกี่ยวข้อง

2.1 คีนค (Kinect)

Michal Czerwinka(2010)ได้ให้ความหมายของคีนคว่าเป็นอุปกรณ์รับรู้การเคลื่อนไหวที่ใช้การทำงานผสมผสานกันระหว่างฮาร์ดแวร์และซอฟต์แวร์ ฟังก์ชันหลักของ Kinect มีสองฟังก์ชันคือ สร้างภาพเคลื่อนไหวสามมิติของวัตถุในมุมมองที่กำหนด และแยกแยะมนุษย์ออกจากวัตถุเหล่านั้นได้ ในปัจจุบัน Kinect มีความละเอียดของภาพอยู่ที่ 640x480 pixel และสามารถทำงานได้ที่ 30 frames ต่อวินาทีสำหรับ Hardware ของ Kinect สามารถแบ่งออกเป็น 3 ส่วนประกอบสำคัญ ได้แก่

- 2.1.1 Color VGA video
- 2.1.2 Depth sensor
- 2.1.3 Multi-array microphone



ภาพที่ 2.1 สถาปัตยกรรมบนกล้อง Kinect

ที่มา: Dan Fernandez (2010)

2.1.1 Color VGA video camera

เป็นกล้องถ่ายภาพเคลื่อนไหวซึ่งช่วยในการจดจำใบหน้า (Face Recognition) และการตรวจจับลักษณะเด่นอื่นๆ โดยใช้ความสามารถในการตรวจจับองค์ประกอบของสีทั้งสามสีอันได้แก่ สีแดง สีเขียว และสีน้ำเงิน



8-bit VGA RGB
640 x 480



11-bit monochrome
320 x 240

ภาพที่ 2.2 แสดงค่าสีที่ได้จากกล้องKinect

ที่มา: Dan Fernandez (2010)

2.1.2 Depth sensor

ประกอบด้วยการทำงานร่วมกันของต้นกำเนิดแสงอินฟราเรด (Infrared Projector) และตัวรับรู้แบบ monochrome CMOS (Complimentary Metal-Oxide Semiconductor) ซึ่งทำหน้าที่รับแสงอินฟราเรดที่ถูกสะท้อนกลับมาจากวัตถุ จากนั้นทำการวัดเวลาในการเดินทาง (Time of Flight) แสงอินฟราเรดนี้ใช้หลักการทำงานเช่นเดียวกับคลื่นโซนาร์คือ หากรู้ระยะเวลาที่แสงอินฟราเรดใช้ในการเดินทางไปกลับก็จะสามารถคำนวณระยะห่างระหว่างตัวรับรู้ความลึกกับวัตถุได้ โดยการทำงานด้วยความเร็วแสงหลายๆรอบทำให้สามารถระบุระยะห่างที่แน่นอนได้โดยไม่ต้องคำนึงถึงสภาพแสงสว่าง

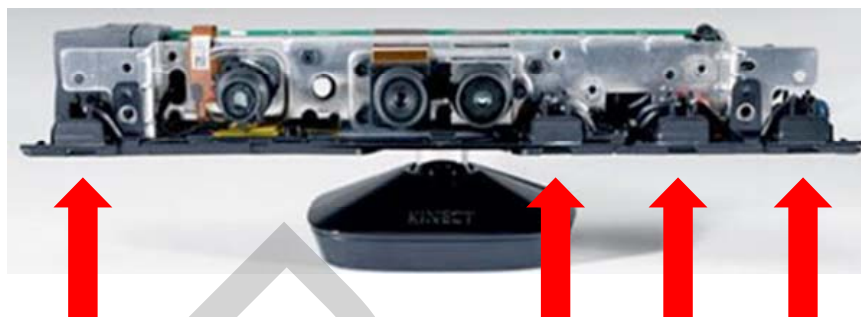


ภาพที่ 2.3 แสงอินฟราเรดที่ใช้หาความลึกของภาพ

ที่มา: Dan Fernandez (2010)

2.1.3 Multi-array microphone

เป็นอาร์เรย์ของไมโครโฟน 4 ตัวซึ่งสามารถแยกแยะเสียงของผู้ใช้ออกจากเสียงรบกวนภายในห้องได้



ภาพที่ 2.4 ไมโครโฟนบนตัวKinect

ที่มา: Dan Fernandez (2010)

หลักการในการตรวจจับการเคลื่อนไหวของKinect ทำให้การประมวลผลภาพทำได้ง่ายยิ่งขึ้นโดยอาศัยความลึกมาเป็นอีกหนึ่งปัจจัยในการแยกแยะวัตถุ จากรูปจะพบว่าการแยกตัวคนออกจากพื้นหลังโดยใช้คุณสมบัติความแตกต่างของสีและพื้นผิวจะทำได้ยาก เพราะสีของพื้นหลังและสีของเสื้อมีความคล้ายคลึงกันมาก แต่จะเป็นเรื่องที่ยากมากถ้าใช้คุณสมบัติของความลึกเป็นเกณฑ์ในการแบ่ง



ภาพที่ 2.5 แสดงการแยกรูปร่างของวัตถุโดยใช้คุณสมบัติของความลึก

ที่มา: Dan Fernandez (2010)

โดยทางบริษัท Microsoft ได้เผยแพร่ชุดพัฒนาซอฟต์แวร์บน Windows Platform สำหรับ Kinect (The Kinect for Windows SDK beta) ซึ่งเป็นชุดเครื่องมือโปรแกรมมิ่งสำหรับนักพัฒนาแอปพลิเคชัน ทำให้ผู้ที่สนใจในการพัฒนาสามารถเข้าถึงการใช้งานอุปกรณ์ Microsoft Kinect ได้อย่างง่ายดายด้วยการใช้งานเชื่อมต่อผ่านระบบปฏิบัติการ Windows 7 โดยชุดพัฒนาซอฟต์แวร์นี้มีลักษณะเด่นดังนี้

Raw sensor streams ทำให้สามารถเข้าถึงข้อมูลดิบจากตัวรับรู้ความลึก ตัวรับรู้เสียง กล้อง และ Four-element Microphone Array

Skeletal tracking ทำให้สามารถติดตามโครงร่างกระดูกของมนุษย์หนึ่งหรือสองคนที่กำลังเคลื่อนที่ได้ ทำให้สามารถสร้างแอปพลิเคชันที่บังคับด้วยท่าทางได้

Advanced audio capabilities ทำให้สามารถประมวลผลเสียง กำจัดเสียงรบกวนที่ซับซ้อน กำจัดเสียงสะท้อน ระบุแหล่งที่มาของเสียง และสามารถบูรณาการร่วมกับ Windows speech recognition API ได้

Sample code and documentation ประกอบด้วยเอกสารเชิงเทคนิคมากกว่า 100 หน้า เอกสารตัวอย่างต่างๆ และ Built-in help files

Easy installation สามารถติดตั้งได้อย่างรวดเร็ว ไม่มีการตั้งค่าที่ซับซ้อน และขนาดของตัวติดตั้งน้อยกว่า 100MB

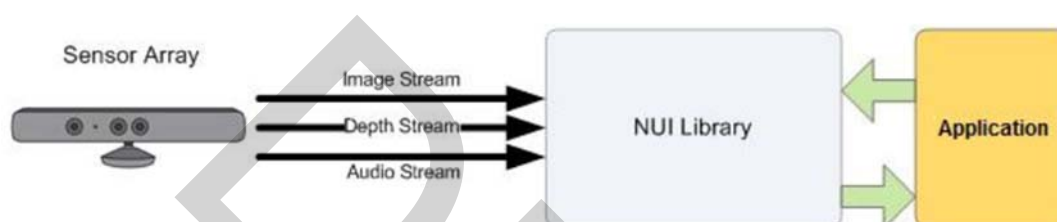
ตารางที่ 2.1 แสดงความสามารถพื้นฐานของกล้อง Kinect

Playable Ranges for the Kinect for Windows Sensor	
Sensor item	Playable range
Color and depth stream	4 to 11.5 feet (1.2 to 3.5 meters)
Skeletal tracking	4 to 11.5 feet (1.2 to 3.5 meters)
Viewing angle	43° vertical by 57° horizontal field of view
Mechanized tilt range (vertical)	±28°
Frame rate (depth and color stream)	30 frames per second (FPS)
Resolution, depth stream	QVGA (320 × 240)
Resolution, color stream	VGA (640 × 480)
Audio format (PCM)	16-kHz, 16-bit mono pulse code modulation

Audio input characteristics	A four-microphone array with 24-bit analog-to-digital converter (ADC) and Kinect-resident signal processing such as acoustic echo cancellation and noise suppression
-----------------------------	--

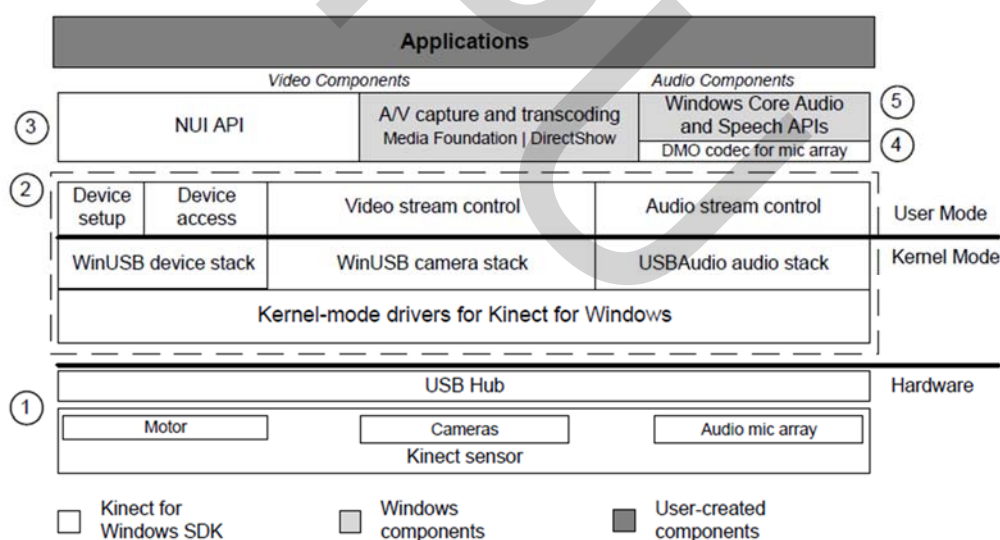
ที่มา: (<http://msdn.microsoft.com/en-us/library/jj131023.aspx>,2012)

Kinect for Windows Architecture



ภาพที่ 2.6 แสดงสถาปัตยกรรมKinect SDK(1)

ที่มา: (<http://msdn.microsoft.com/en-us/library/jj131023.aspx>,2012)



1. Kinect hardware 2. Microsoft Kinect drivers 3. NUI API 4. Kinect Audio DMO 5. Windows 7 standard APIs

ภาพที่ 2.7 แสดงสถาปัตยกรรมKinect SDK(2)

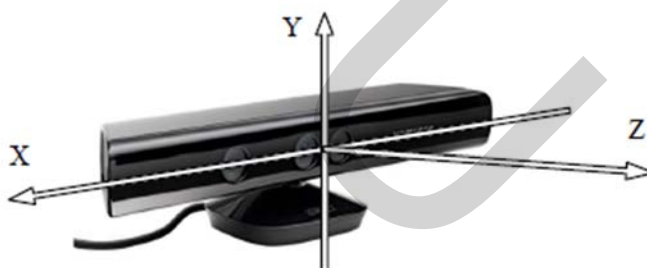
ที่มา: (<http://msdn.microsoft.com/en-us/library/jj131023.aspx>,2012)

2.2 Extensible Application Markup Language (XAML)

เป็นภาษามาร์กอัปสำหรับกำหนดส่วนติดต่อผู้ใช้หรือ User Interface ในการพัฒนาโปรแกรม ใช้สำหรับกำหนดวัตถุ คุณลักษณะ ความสัมพันธ์ และการโต้ตอบของวัตถุ XAML นั้นถูกใช้เป็นหัวใจสำคัญในการสร้างส่วนติดต่อผู้ใช้งานด้วย Windows Presentation Foundation (WPF) (หรือ Avalon) ซึ่งเป็นไลบรารีส่วนสำหรับการจัดการส่วนติดต่อผู้ใช้งานแบบใหม่ใน Microsoft .NET Framework 3.0 XAML เป็นภาษาสำหรับกำหนดส่วนติดต่อผู้ใช้ที่พัฒนามาจาก XML เช่นเดียวกับภาษา XML User Interface Language ออกแบบส่วนติดต่อผู้ใช้งานผ่านเครื่องมือพัฒนาอย่างเช่น Visual Studio หรือ XAMLPad

2.3 การตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect และ Kinect SDK Beta 1

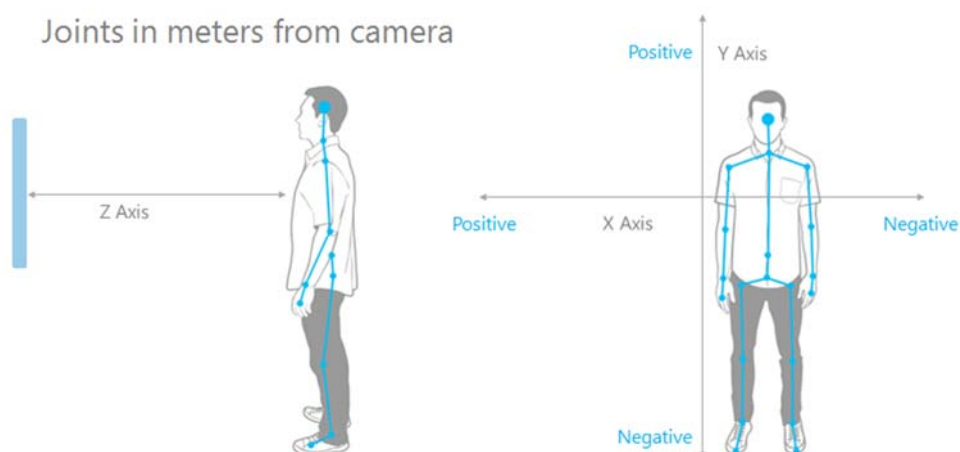
การตรวจจับการเคลื่อนไหวของร่างกายมนุษย์โดยใช้กล้อง Kinect และ Kinect SDK Beta 1 ข้อมูลตำแหน่งของข้อต่อที่ได้รับจาก Kinect SDK Beta 1 [4] จะอยู่ในรูปแบบจุดพิกัดสามมิติ (X, Y และ Z) โดยตำแหน่งของจุดกำเนิด ($X = Y = Z = 0$) จะเป็นตำแหน่งของกล้องที่ใช้ในการตรวจจับการเคลื่อนไหว และมีทิศทางของแกน X, Y และ Z ตามที่แสดงในภาพที่ 2.8 ซึ่งแกน Z จะเป็นทิศทางที่กล้องตรวจจับ



ภาพที่ 2.8 แกน X, Y และ Z ของกล้อง Kinect เมื่อใช้ Kinect SDK Beta

ที่มา: นราวุฒิ พัฒโนทัย (2554: 252)

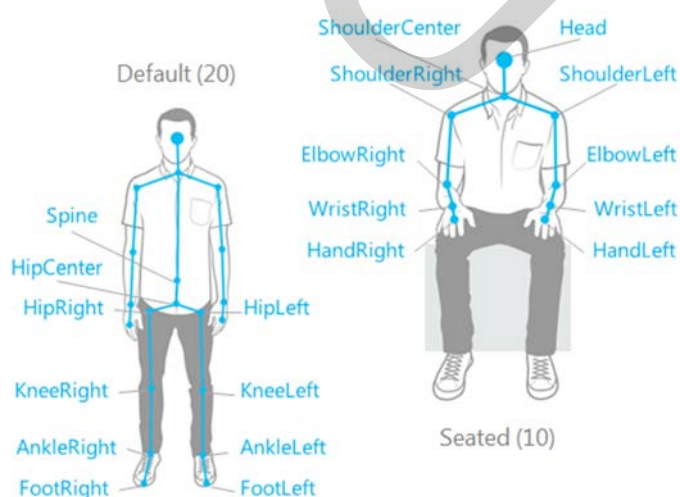
ค่า X จะเป็นค่าแสดงระยะทางที่ห่างออกไปจากจุดกำเนิดตามแนวนอนสัมพันธ์กับกล้องที่ตรวจจับ จะมีค่าเป็นบวกเมื่อตำแหน่งที่ถูกตรวจจับได้อยู่ทางด้านขวาของกล้อง ค่า Y จะเป็นค่าแสดงระยะทางที่ห่างออกไปจากจุดกำเนิดตามแนวตั้งที่กล้องตรวจจับ จะมีค่าเป็นบวกเมื่อตำแหน่งที่ถูกตรวจจับได้อยู่สูงกว่าตำแหน่งของกล้อง และค่า Z จะเป็นระยะทางที่ห่างออกไปจากกล้องโดยจะมีค่า เป็นบวกเสมอ



ภาพที่ 2.9 แสดงแกนทางด้าน Positive และ Negative

ที่มา: (<http://video.ch9.ms/teched/2012/eu/DEV330.pptx,2012>)

ตำแหน่งของข้อต่อที่ได้รับจาก API จะสัมพันธ์กับตำแหน่งของร่างกายมนุษย์ 20 ตำแหน่งในท่ายืน และ 10 ตำแหน่งในท่านั่ง ตามที่แสดงในภาพที่ 2.10 เมื่อผู้ใช้นั้นหันหน้าเข้าหากล้อง ตำแหน่งของข้อต่อข้างซ้ายและขวาจะสลับข้างกัน แตกต่างจากร่างกายจริง แต่ละตำแหน่งของข้อต่อจะมีค่า X, Y และ Z ในหน่วยมิลลิเมตร และมีฟิลด์ Confidence แสดงสถานะการตรวจจับตำแหน่งข้อต่อนั้นๆ โดยจะมีค่าระหว่าง 0 – 1 ซึ่ง 1 หมายถึง สามารถตรวจจับได้ และหากมีค่าน้อยกว่า 1 (เป็นตัวเลขทศนิยม) หมายถึง ไม่สามารถตรวจจับได้ ข้อมูลตำแหน่งของข้อต่อในส่วนนั้นจะเป็นค่าประมาณ ในบางครั้งอาจเป็นค่าที่ตรวจจับได้ล่าสุดก่อนที่จะตรวจจับไม่ได้ในเวลาต่อมา หรืออาจเป็นค่าที่ไม่มีความหมายก็ได้



ภาพที่ 2.10 ตำแหน่งของข้อต่อที่สัมพันธ์กับร่างกายมนุษย์

ที่มา: (<http://video.ch9.ms/teched/2012/eu/DEV330.pptx,2012>)

ในการตรวจจับการเคลื่อนไหว หากผู้ใช้ภายในมวกกล้อง ห่างจากกล้องประมาณ 1.2 – 3.5 เมตร (ระยะห่างอุดมคติตามที่ Microsoft แนะนำคือ ประมาณ 2.5 เมตร) ค่าตำแหน่งของข้อต่อที่ ถูกตรวจจับได้ จะมีค่าใกล้เคียงกับการวัดจริง เช่น ความยาวของข้อศอกไปจนถึงกลางฝ่ามือ เป็นต้น และฟิวด์ Confidence ของข้อต่อจะมีค่าเป็น 1 แต่หากมีบางข้อต่อที่กล้องไม่สามารถตรวจจับได้ หรือถูกบดบังไป เช่น ยืนมือโพล่หลัง เป็นต้น ฟิวด์ Confidence ของข้อต่อนั้นจะมีค่าน้อยกว่า 1

คำแนะนำเพิ่มเติมจาก Microsoft ในการตรวจจับการเคลื่อนไหวด้วย Kinect SDK

1. ผู้ใช้ไม่ควรสวมชุดที่หลวมโคร่งในขณะที่ทำการตรวจจับการเคลื่อนไหว เพราะจะทำให้ ไม่สามารถตรวจจับตำแหน่งของข้อต่อได้
2. การตรวจจับข้อต่อที่แขนจะมีความเสถียรน้อย หากแขนนั้นอยู่ใกล้ชิดกับส่วนของร่างกาย โดยเฉพาะอยู่ชิดกับลำตัว ถ้าแขนทั้งสองอยู่ชิดกับลำตัวหรืออยู่ใกล้กับส่วนอื่น อาจรวมแขนนั้นเป็นส่วนหนึ่งของร่างกายส่วนนั้นได้ และจะทำให้ไม่สามารถตรวจจับข้อต่อ ที่แขนได้
3. การตรวจจับข้อต่อที่ขาจะมีการตรวจจับที่ดีขึ้น หากผู้ใช้ไม่ยืนชิดขา
4. การเคลื่อนไหวอย่างรวดเร็วหรือการแสดงท่าทางที่ซับซ้อน อาจทำให้ไม่สามารถตรวจจับการเคลื่อนไหวได้ เช่น การทำท่าเตะฟุตบอลอย่างรวดเร็ว เป็นต้น

จากการใช้งานกล้อง Kinect และ Kinect SDK ในการตรวจจับการเคลื่อนไหวของร่างกายมนุษย์ ในเบื้องต้น ผู้วิจัยพบว่า ตำแหน่งที่ดีที่สุดในการตั้งกล้องตรวจจับการเคลื่อนไหวคือการตั้งกล้อง ให้เผชิญหน้ากับผู้ใช้ และไม่มีสิ่งทีบดบังการตรวจจับ เพราะส่วนของร่างกายเกือบทุกๆ ส่วน จะสามารถตรวจจับได้ตลอดเวลา และหากมีสิ่งของที่อยู่ใกล้กับผู้ใช้ เช่น ผู้ใช้นั่งอยู่บนเก้าอี้ เป็นต้น Kinect SDK อาจรวมสิ่งของที่อยู่ใกล้นั้นเป็นส่วนหนึ่งของร่างกายผู้ใช้ และระบุตำแหน่งของข้อต่อ ที่ผิดพลาดได้ การระบุส่วนของร่างกายด้านซ้ายและขวาอาจสามารถสลับกันได้ เนื่องจาก Kinect SDK ไม่สามารถวิเคราะห์ภาพโครงร่างมนุษย์ที่กล้องตรวจจับได้ว่าเป็นด้านหน้าหรือด้านหลังของร่างกาย หากผู้ใช้หันหลังให้กับกล้องตั้งแต่เริ่มการตรวจจับการเคลื่อนไหว ตำแหน่งของข้อต่อที่ควรจะเป็นด้านซ้ายก็จะสลับไปเป็นด้านขวาได้ นอกจากนี้ หากมีสิ่งของที่มีรูปร่างคล้ายกับโครงร่างมนุษย์ เช่น พัดลมตั้งพื้น เป็นต้น Kinect SDK อาจทำการตรวจจับและให้ค่าตำแหน่งได้

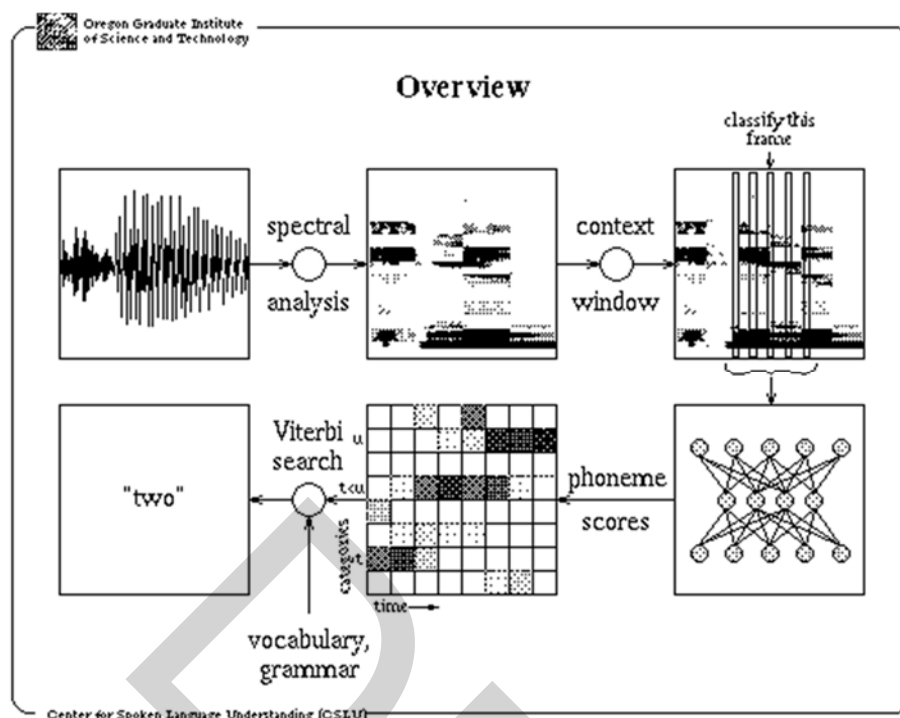
2.4 การรู้จำเสียง (Speech Recognition)

Parinya (2001) กล่าวว่า Speech recognition หมายถึง การทำให้คอมพิวเตอร์สามารถที่จะฟังคำพูดและตัดสินใจได้ว่าคำพูดนั้นเป็นคำอะไรหรืออีกความหมายหนึ่งคือการนำ File Audio ที่บรรจุเสียงพูดนำมาแปลงเป็น Text ได้

Automatic Speech Recognition (ASR) เป็นเทคโนโลยีที่เกิดขึ้นเพื่อที่จะให้คอมพิวเตอร์สามารถแยกแยะคำพูดต่างๆที่มนุษย์สามารถพูดใส่อุปกรณ์ไมโครโฟนหรือเครื่องโทรศัพท์หรืออื่นๆเท่าที่จะเป็นไปได้ “ Holy grail ” of ASR Search เป็นโปรแกรมที่อนุญาตให้คอมพิวเตอร์เข้าใจคำศัพท์ทุกคำอย่างถูกต้อง 100% ซึ่งมีความสามารถเข้าใจถึงคำพูดได้ดีไม่ว่าจะเป็นคำพูดของใครก็ตาม เป็นอิสระจากขนาดของกลุ่มคำศัพท์ , ความดัง , ลักษณะของผู้พูด และการออกเสียง หรือเงื่อนไขของช่องทางต่างๆที่เป็นไปได้

เทคโนโลยีที่เป็นส่วนสำคัญที่ใช้ทำ ASR ถูกเรียกว่า “ Hidden Markov Model ” หรือ “HMM” เทคโนโลยีชนิดนี้สามารถที่จะเข้าใจถึงคำพูดโดยการประมาณการถึงความเป็นไปได้ของแต่ละหน่วยที่เป็นพื้นฐานของเสียงที่อยู่ติดๆกัน ซึ่งแต่ละเสียงจะมีขอบเขตของตัวสัญญาณ คำแต่ละคำในกลุ่มของคำศัพท์เหล่านั้น ต่างก็มีลักษณะเฉพาะที่มีความแตกต่างกัน โดยสังเกตจากส่วนประกอบของหน่วยที่เป็นพื้นฐานของเสียง

Procedure ที่เกี่ยวกับการค้นหาที่มีชื่อเรียกว่า “ Viterbi Search ” ถูกใช้เพื่อกำหนดถึงความต่อเนื่องของหน่วยพื้นฐานของเสียงด้วยความเป็นไปได้สูงสุด เทคนิคชนิดนี้ถูกจำกัดเฉพาะมองหาคำจากความต่อเนื่องของหน่วยพื้นฐานของเสียง ซึ่งตรงกันกับความคำต่างๆในกลุ่มของคำศัพท์ที่มี และความต่อเนื่องของหน่วยพื้นฐานของเสียงเหล่านั้นด้วยความเป็นไปได้สูงสุดทั้งหมดถูกระบุด้วยคำศัพท์ที่ถูกพูดออกมา ในมาตรฐานของ HMMs ความเป็นไปได้อาจจะถูกคำนวณโดยใช้ “ a Gaussian Mixture Model ” : ซึ่งอยู่ในขอบเขต HMM/ANN ซึ่งค่าเหล่านี้จะถูกคำนวณโดย Artificial neural network (ANN)



ภาพที่ 2.11 รวมการทำงานของ Speech Recognition

ที่มา: (http://vclass.mgt.psu.ac.th/~parinya/project2001/speech_recognition.html, 2001)

แบบจำลองฮิดเดินมาร์คอฟ สามารถโมเดลคำหรือหน่วยเสียงก็ได้ แต่เหตุผลของการที่จะโมเดลให้เป็นคำหรือหน่วยเสียง จะขึ้นอยู่กับการใช้งานของระบบการรู้จำเสียงเช่น ถ้าระบบรู้จำเสียงมีข้อจำกัดคือให้รู้จำคำศัพท์ที่น้อยและรู้จำคำพูดที่ไม่ติดต่อกัน การโมเดลเสียงให้เป็นคำเสียงจะดีกว่าการโมเดลหน่วยเสียงเนื่องจากการโมเดลเป็นคำจะให้ความแม่นยำมากกว่าและวิธีการของระบบรู้จำคำเสียงจะง่ายกว่าการโมเดลหน่วยเสียง การโมเดลหน่วยเสียงต้องการโมเดลแต่ละหน่วยเสียงและจะต้องรู้จำหน่วยเสียงนั้นๆก่อน เมื่อรู้จำหน่วยเสียงได้แล้วก็ต้องมีการสร้างคำจากหน่วยเสียงที่ได้รู้จำมา ถ้ามีการรู้จำหน่วยเสียงหนึ่งผิดไปก็จะทำให้คำคำนั้นรู้จำผิดได้ แต่สำหรับการโมเดลคำเสียง ความแม่นยำในการรู้จำจะขึ้นโดยตรงกับโมเดลคำเสียงที่ได้ฝึกฝนมาและไม่ต้องการวิธีสร้างคำจากหน่วยเสียง แต่สำหรับระบบที่มีคำศัพท์มาก (มากกว่า 1000 คำ) และรู้จำเสียงแบบพูดติดต่อกันระบบรู้จำที่มีโมเดลหน่วยเสียงจะมีประสิทธิภาพมากกว่าเพราะจำนวนโมเดลของหน่วยเสียงที่ต้องการจะมีน้อยกว่าโมเดลของคำเสียงและทำให้ระบบรู้จำเสียงใช้เวลาในการค้นหาคำเสียงที่มีความน่าจะเป็นมากที่สุดมากกว่า อีกประการหนึ่งที่สำคัญก็คือการรู้จำคำพูดแบบติดต่อกัน มีความเป็นไปได้สูงที่คำหนึ่งคำจะเปลี่ยนไปเมื่อใช้กับคำอื่น ถ้าระบบมีการฝึกฝน

ที่ไม่ดีก็จะทำให้ระบบรู้จำมีความแม่นยำต่ำ แต่สำหรับหน่วยเสียงจะมีการเปลี่ยนแปลงน้อยกว่าเมื่อใช้กับคำคำอื่นจึงทำให้มีความแม่นยำมากกว่า

ตามทฤษฎีแบบจำลองฮิดเดินมาร์คอฟ คือ การเก็บรวบรวมของสถานะหลายๆ สถานะที่ถูกเชื่อมโยงโดยการเปลี่ยนสถานะในที่นี้คือ ลักษณะของสัญญาณเสียงหรือเครื่องหมายที่หมายถึงสัญญาณเสียงนั้น ณ เวลาหนึ่งๆ ส่วนการเปลี่ยนสถานะคือ การเลื่อนหรือเปลี่ยนแปลงของสัญญาณเสียงจากเวลาหนึ่งไปหาอีกเวลาหนึ่ง แบบจำลองฮิดเดินมาร์คอฟ มีความน่าจะเป็นอยู่สองชนิดคือ ค่าความน่าจะเป็นที่การเปลี่ยนสถานะหนึ่งที่จะเกิดขึ้น (transition probability) และ ค่าความน่าจะเป็นของผลลัพธ์ (output symbol) เมื่อมีการเปลี่ยนสถานะหนึ่งเกิดขึ้นผลลัพธ์ที่ว่าจะนำไปใช้ในการตัดสินใจว่าเสียงที่ได้ยินนั้นเป็นเสียงอะไร (output probability)

ปัญหาของแบบจำลองฮิดเดินมาร์คอฟมีอยู่สามข้อด้วยกันคือ

1. Evaluation Problem: แบบจำลองสามารถสร้างหรือเข้าใจเสียงที่ได้ยินดีแค่ไหน
2. Decoding Problem: ลำดับของสถานะในแบบจำลองสร้างเสียงที่ได้ยินดีแค่ไหน
3. Learning Problem: ตัวแปร (parameter) ของแบบจำลองควรเป็นอย่างไรเพื่อที่จะทำให้มีความผิดพลาดในการสร้างเสียงหรือเข้าใจเสียงที่ได้ยินน้อยที่สุด

ถ้าปัญหา Evaluation Problem ถูกแก้ไข ก็จะมีทางที่ทำให้แบบจำลองมีลักษณะเดียวกันกับเสียงที่ได้ยิน ซึ่งสามารถใช้ในการรู้จำเสียงพูดที่ไม่ติดต่อกันได้ ถ้าปัญหา Decoding Problem ถูกแก้ไข ก็จะสามารถหาลำดับของสถานะที่ดีที่สุดในการทำความเข้าใจหรือเปรียบเทียบกับเสียงที่ได้ยิน เพื่อที่จะนำมาใช้ในการรู้จำเสียงที่ติดต่อกันได้ และที่สำคัญที่สุดคือถ้าปัญหา Learning Problem ถูกแก้ไข ก็จะทำให้มีระบบที่สามารถเรียนรู้ได้แบบอัตโนมัติจากค่าตัวแปร (Parameter) ต่างๆ โดยการฝึกฝนจากชุดข้อมูลที่ป้อนเข้าระบบ

2.5 Microsoft Speech Application Programming Interface (SAPI)

โปรแกรมสำหรับออกเสียงสนทนาหรือ SAPI คือ API ซึ่งพัฒนาโดย Microsoft รองรับการรู้จำเสียงสนทนา และการสังเคราะห์เสียงในโปรแกรมสำหรับ Windows เวอร์ชันปัจจุบันที่ถูกปล่อยออกมาซึ่งถูกนำไปประกอบกับส่วนต่างๆ ของ Speech SDK หรือแต่ละส่วนของระบบปฏิบัติการ Windows เอง ประกอบด้วย Microsoft Office, Microsoft Agent และ Microsoft Speech Server ในเวอร์ชันทั่วไปของ API ถูกออกแบบมาเพื่อให้นักพัฒนาสามารถเขียนโปรแกรมประยุกต์เพื่อรู้จำเสียงสนทนาและสังเคราะห์เสียงโดยใช้มาตรฐานเดียวกัน สามารถใช้ภาษาในการพัฒนาได้หลากหลาย นอกจากนี้ยังมีความเป็นไปได้สำหรับบริษัทอื่นๆ ที่จะพัฒนา Text-to-Speech หรือ Speech-to-Text ของตัวเองโดยพัฒนาจาก SAPI โดยหลักแล้วสามารถใช้งานกว่าจะพัฒนาได้ดี

ยิ่งกว่าหรือใช้แทนของ Microsoft ได้ โดยทั่วไป Speech API สามารถเผยแพร่ได้อย่างอิสระซึ่งสามารถส่งไปพร้อมกับโปรแกรมประยุกต์ของ Windows ที่ใช้เทคโนโลยีการพูด แต่ไม่ใช่ทุกเวอร์ชันของการรู้จำเสียงสนทนาและการสังเคราะห์เสียงจะไม่เสียค่าใช้จ่าย มีอยู่ 2 ตระกูลหลักด้วยกันของ Microsoft Speech API โดย SAPI 1-4 จะคล้ายกับตัวอื่นๆแต่จะมีเพิ่มคุณลักษณะพิเศษเข้ามาในเวอร์ชันที่ใหม่กว่า SAPI เวอร์ชัน 5 ถูกสร้างในปี 2000 และมีเวอร์ชันย่อยออกมาหลังจากนั้นเป็นระยะๆ

2.5.1 SAPI (Speech Application Programming Interface)

แบ่งได้ 2 ประเภท คือ Text-to-Speech และ Speech-to-Text

Text-to-Speech เป็นส่วนที่สร้างเสียงพูดของมนุษย์ ด้วยการผสมเสียงจากตัวอักษรที่ประกอบเป็นคำหรือประโยค เพื่อให้เครื่องคอมพิวเตอร์สามารถสร้างเสียงเลียนแบบเสียงมนุษย์และความหมายที่มนุษย์สามารถเข้าใจ

Speech-to-Text เป็นที่ใช้ในการวิเคราะห์จำนวนเสียง และควบคุมการทำงานคอมพิวเตอร์ด้วยเสียง โดยจะพยายามทำความเข้าใจในข้อมูลเสียงที่ได้รับ โดยจะทำการตรวจสอบโครงสร้างและรูปแบบของเสียงว่าถูกต้องตามที่กำหนดหรือไม่ และดำเนินการตามเงื่อนไขของโปรแกรมที่กำหนดไว้

2.5.2 SAPI 5.1

ในเวอร์ชันนี้ถูกปล่อยออกมาหลังจากปี 2001 เป็นส่วนหนึ่งของ Speech SDK 5.1 ซึ่งได้เพิ่ม Automation-compliant interfaces เข้ามารองรับการใช้งาน โดย Visual Basic ภาษาสคริปต์ได้แก่ JScript และ managed code เวอร์ชันนี้ถูกปล่อยมาพร้อมกับ Windows XP ได้พัฒนาเวอร์ชัน 6 ออกมาใน Office 2003 และ Windows XP Tablet PC Edition

2.5.3 สถาปัตยกรรมพื้นฐาน (SAPI)

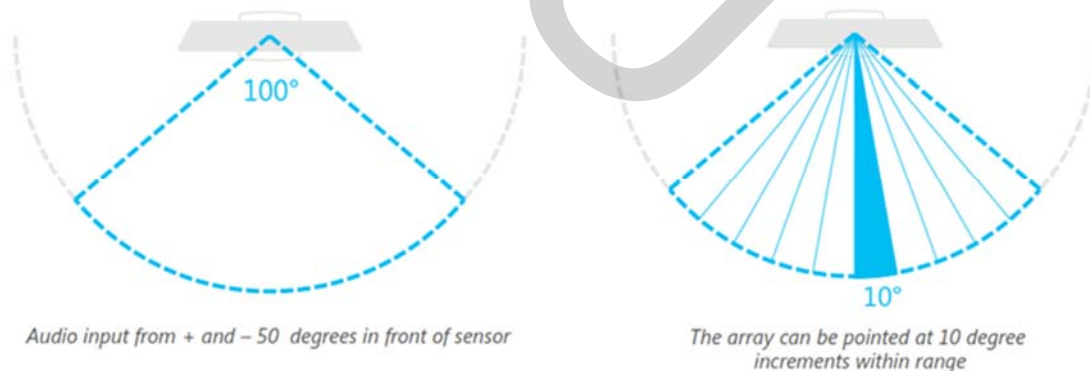
Speech API เป็นที่รู้จักอย่างกว้างขวางในลักษณะของอินเทอร์เฟซ (Interface) ซึ่งถูกจัดให้อยู่ระหว่างแอปพลิเคชัน (Application) และ Speech engine ใน SAPI 1 - 4, แอปพลิเคชันสามารถสื่อสารกับเอนจินได้โดยตรง มี interface definition ที่แอปพลิเคชันและเอนจินได้ทำการปรับเปลี่ยนให้สามารถทำงานด้วยได้ และแอปพลิเคชันก็ยังสามารถใช้ Object ระดับสูงได้โดยง่าย โดยที่ไม่ต้องเรียกใช้เมธอด (Method) จากเอนจิน

แต่ใน API เวอร์ชัน 5 นั้น แอปพลิเคชันและเอนจินไม่ต้องสื่อสารกันโดยตรงอีกต่อไป แต่ติดต่อกับรันไทม์ (sapi.dll) แทน มี API Implement ที่กระทำโดยตัวรันไทม์นี้และอีกตัวถูกตั้งขึ้นเพื่อเป็น interface ของเอนจินในเวอร์ชัน 5 คำสั่งของแอปพลิเคชันเรียกผ่าน API (ตัวอย่างเช่น การจดจำแกรมมาหรือการเตรียมข้อความเพื่อนำไปสังเคราะห์) sapi.dll รันไทม์ ตัวนี้จะทำการแปลง

คำสั่งและประมวลผล เมื่อมีการเรียกใช้เอนจินเกิดขึ้นผ่านเอนจิน อินเตอร์เฟส (เช่น การใช้แกรมมาจากไฟล์ที่ได้กระทำเสร็จแล้ว แต่ข้อมูลแกรมมาถูกส่งไปยัง recognition engine เพื่อใช้ในส่วนของการ recognition) recognition และ synthesis engine ทำการสร้าง event ในขณะที่ประมวลผล(เช่น การออกเสียงถูกนำไปต่อกันเพื่อสร้างเสียงสังเคราะห์ขึ้นมา) ทั้งหมดนี้เกิดขึ้นในทิศทางที่ตรงกันข้าม จากเอนจิน ผ่านรันไทม์ dll และไปสู่ event sink ที่อยู่ในแอปพลิเคชัน

จากคำจำกัดความของ API และ dll รันไทม์แล้ว ส่วนประกอบอื่นๆถูกเพิ่มเข้าไปเพื่อให้ toolkit นี้สมบูรณ์ ส่วนประกอบต่อไปนี้มีอยู่ในเกือบทุกเวอร์ชันของ Speech SDK

1. API definition files – อยู่ใน MIDL พัฒนาโดยภาษา C หรือ C++
2. Runtime components – เช่น sapi.dll.
3. Control Panel applet – มีไว้สำหรับเลือก default speech recognizer และ synthesizer.
4. Text-To-Speech engines - รองรับหลายภาษา
5. Speech Recognition engines - รองรับหลายภาษา
6. Redistributable components อนุญาตให้ผู้พัฒนานำไปใช้ในโค้ดและแอปพลิเคชันที่สร้างขึ้นได้
7. Sample application code.
8. Sample engines – คำสั่งที่จำเป็นของเอนจิน แต่ไม่ใช่การประมวลผลเสียงอ่านที่แท้จริง



ภาพที่ 2.12 แสดงมุมมองและทิศทางในการรับเสียงจาก Kinect บน SAPI 5.1

ที่มา: (<http://video.ch9.ms/teched/2012/eu/DEV330.pptx>, 2012)

2.6 งานวิจัยที่เกี่ยวข้อง

การตรวจจับการเคลื่อนไหวของร่างกายมนุษย์ (Human Motion Capture) เป็นกระบวนการการบันทึกการเคลื่อนไหวของร่างกายมนุษย์ให้อยู่ในรูปแบบดิจิทัล การตรวจจับการเคลื่อนไหวของร่างกายสามารถแบ่งออกได้เป็น 2 รูปแบบหลัก ได้แก่ การตรวจจับการเคลื่อนไหวโดยการทำเครื่องหมายตามตำแหน่งต่างๆ บนร่างกาย (Marker Motion Capture) และการตรวจจับการเคลื่อนไหวโดยปราศจากการทำเครื่องหมายตามตำแหน่งต่างๆ บนร่างกาย (Markerless Motion Capture)

การตรวจจับการเคลื่อนไหวโดยการทำเครื่องหมายตามตำแหน่งต่างๆ บนร่างกาย จะใช้วิธีการติดเครื่องหมายตามข้อต่อต่างๆ บนร่างกาย หรือติดเครื่องหมายลงบนชุดที่สวมใส่ แล้วใช้กล้องตรวจจับเครื่องหมายที่ติดไว้ จากนั้นจะใช้ซอฟต์แวร์ทำการวิเคราะห์ตำแหน่งของเครื่องหมายเพื่อสร้างตำแหน่งของข้อต่อของร่างกายในรูปแบบสามมิติ การทำเครื่องหมายตามตำแหน่งต่างๆ บนร่างกายอีกวิธีการหนึ่ง จะใช้วิธีการติดอุปกรณ์บอกตำแหน่งไว้ที่ร่างกายโดยตรง วิธีการนี้จึงไม่ต้องใช้กล้องในการตรวจจับ และทำให้ได้ตำแหน่งของข้อต่อในสามมิติโดยตรง

การตรวจจับการเคลื่อนไหวโดยปราศจากการทำเครื่องหมายตามตำแหน่งต่างๆ บนร่างกาย จะใช้กล้องตรวจจับภาพการเคลื่อนไหว แล้วใช้ซอฟต์แวร์ทำการวิเคราะห์ภาพที่กล้องตรวจจับได้ เพื่อแยกภาพร่างกายออกจากภาพพื้นหลัง จากนั้นจะทำการวิเคราะห์ภาพร่างกายเพื่อสร้างตำแหน่งของข้อต่อของร่างกายในรูปแบบสามมิติ

2.6.1 Evaluating a Dancer's Performance using Kinect-based Skeleton Tracking

งานวิจัยของ D. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, และ M. B. Moussa (2011) ได้ทำการศึกษาเกี่ยวกับการวัดประสิทธิภาพของนักเต้น โดยใช้กล้อง Kinect ในการตรวจจับการเคลื่อนไหว ในการให้ข้อมูลการตรวจจับงานวิจัยนี้ใช้การเปรียบเทียบตำแหน่งของข้อต่อของนักเต้นสองคนที่ได้จากการตรวจจับตามเวลาจริงมาเป็นส่วนหนึ่งในการให้คะแนนการเต้น ผลการศึกษางานวิจัยนี้แสดงให้เห็นว่าสามารถนำข้อมูลตำแหน่งของข้อต่อและมือ ไปใช้งานในด้านการตรวจจับการเคลื่อนไหวได้ภาพที่ 2.13



ภาพที่ 2.13 การตรวจจับการเคลื่อนไหวของคนและโมเดล

ที่มา: (<http://doras.dcu.ie/16574/>)

2.6.2 Free Viewpoint Virtual Try-On With Commodity Depth Cameras

งานวิจัยของ S. Hauswiesner, M. Straka และ G. Reitmayr (2011) ได้เสนอระบบการลองเสื้อผ้าจากที่บ้านผ่านระบบเสมือนจริง โดยใช้กล้อง Kinect ตรวจจับผู้ใช้ และตรวจจับเสื้อผ้าที่ต้องการ จากนั้นระบบจะทำการสร้างแบบจำลองสามมิติของผู้ใช้และเสื้อผ้า แล้วนำมาแสดงประกอบกันเป็นภาพผู้ใช้สวมใส่เสื้อผ้าตามที่ได้เลือกไว้ โดยแสดงท่าทางเดียวกันกับผู้ใช้ตามเวลาจริง ตามที่แสดงในภาพที่ 2.14 งานวิจัยนี้ทดสอบระบบที่เสนอโดยใช้แบบสอบถามสอบถามผู้ทดลอง ซึ่งปรากฏผลตอบรับที่ดีจากผู้ทดลอง

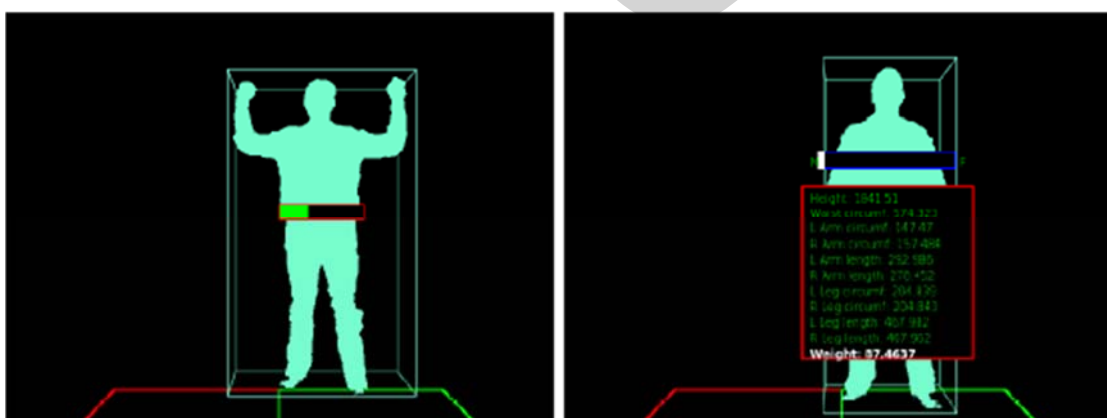


ภาพที่ 2.14 ระบบการลงเสื้อผ้าเสมือนจริง

ที่มา: (<http://dl.acm.org/citation.cfm?id=2087759>)

2.6.3 Real Time Extraction of Body Soft Biometric from 3D Videos

งานวิจัยของ C. Velardo และ J. Dugelay (2011) ได้เสนอระบบการบ่งบอกคุณลักษณะภายนอกของบุคคล (Body Soft Biometric) โดยใช้กล้อง Kinect ตรวจจับร่างกายของผู้ใช้ จากนั้นจะนำข้อมูลการตรวจจับที่ได้ไปคำนวณหาส่วนสูง น้ำหนัก และเพศของผู้ใช้คนนั้นโดยประมาณตามที่แสดงใน ภาพที่ 2.15 งานวิจัยนี้แสดงให้เห็นถึงแนวทางในการระบุตัวตนของบุคคลโดยใช้กล้อง Kinect



ภาพที่ 2.15 ระบบการบ่งบอกคุณลักษณะภายนอกของบุคคลโดยใช้กล้อง Kinect

ที่มา (<http://dl.acm.org/citation.cfm?id=2072298.2072454>)

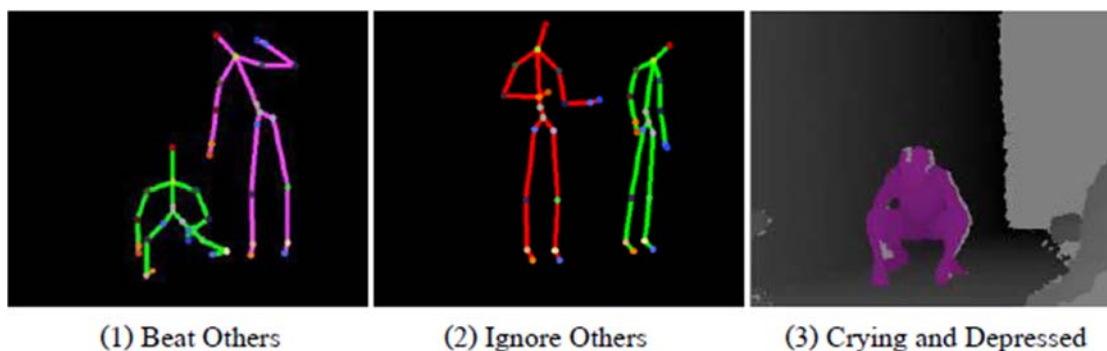
2.6.4 Workshop on Digital Media and Digital Content Management

งานวิจัยของ M. K. M. bin Sidik, M. S. bin Sunar, I. bin Ismail, M. K. bin Mokhtar และ N. binti M. Jusoh (2011) ได้ทำการศึกษาเกี่ยวกับการปฏิสัมพันธ์ระหว่างผู้ใช้กับคอมพิวเตอร์ โดยการตรวจจับการเคลื่อนไหวด้วยการใช้ภาพความลึก (Depth Image) การศึกษานี้ได้แสดงให้เห็นถึงโครงสร้างของระบบการวิเคราะห์การเคลื่อนไหวของร่างกายมนุษย์ ซึ่งประกอบด้วย 4 ขั้นตอนดังนี้

1. Initialization เป็นขั้นตอนในการเตรียมความพร้อมก่อนที่จะเริ่มทำการตรวจจับการเคลื่อนไหว
 2. Tracking เป็นขั้นตอนในการแบ่งแยกระหว่างผู้ถูกตรวจจับกับฉากพื้นหลัง
 3. Pose Estimation เป็นขั้นตอนในการประมาณการแสดงท่าทางของผู้ถูกตรวจจับ
 4. Recognition เป็นขั้นตอนในการวิเคราะห์และจำแนกการกระทำของผู้ถูกตรวจจับ
- ผลการศึกษางานวิจัยนี้แสดงให้เห็นถึงประโยชน์ของการใช้ภาพความลึกในการตรวจจับการเคลื่อนไหวของร่างกายมนุษย์ ซึ่งการใช้ภาพความลึกจะช่วยลดขั้นตอนและเวลาในการแบ่งแยกระหว่างผู้ถูกตรวจจับกับฉากพื้นหลัง

2.6.5 Third Chinese Conference on Intelligent Visual Surveillance

งานวิจัยของ X. Yu, L. Wu, Q. Liu และ H. Zhou (2011) ได้เสนอวิธีการประเมินการแสดงอารมณ์ด้านลบของเด็กด้วยข้อมูลที่ได้รับการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect ตามที่แสดงในภาพที่ 2.16 งานวิจัยนี้นำข้อมูลการตรวจจับการเคลื่อนไหวที่ได้ไปวิเคราะห์ด้วยอัลกอริทึมในการวิเคราะห์พฤติกรรมแบบ Stochastic Grammar งานวิจัยนี้ทำการทดสอบระบบต้นแบบเปรียบเทียบกับระบบบันทึกภาพเคลื่อนไหวและการจดบันทึกพฤติกรรมที่เด็กแสดงออก ผลของงานวิจัยนี้แสดงให้เห็นถึงระบบต้นแบบที่ช่วยในการวิเคราะห์พฤติกรรมและการแสดงอารมณ์ด้านลบของเด็ก และช่วยในการแก้ไขพฤติกรรมของเด็ก



ภาพที่ 2.16 ตัวอย่างพฤติกรรมของเด็กที่บันทึกโดยใช้กล้อง Kinect

ที่มา: (<http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6153167>)

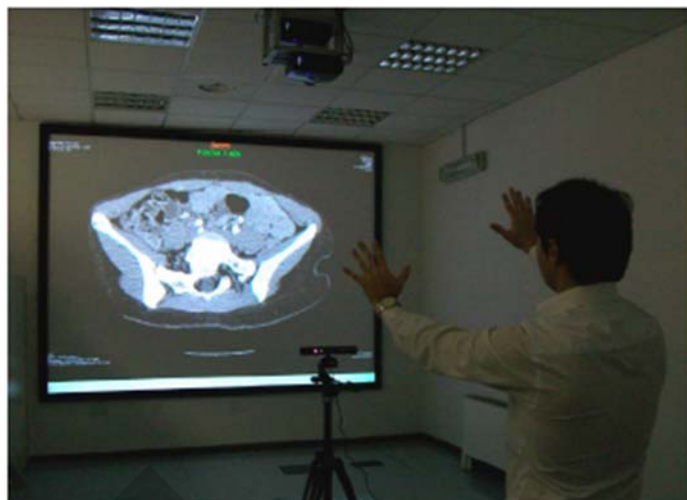
2.6.6 Development and Evaluation of Low Cost Game-Based Balance Rehabilitation Tool

Using the Microsoft Kinect Sensor

งานวิจัยของ B. Lange, C. Chang, E. Suma, B. Newman, A. S. Rizzo และ M. Bolas (2011) ได้เสนอเกมสำหรับฝึกการทรงตัวของผู้ป่วยทางระบบประสาท โดยใช้กล้อง Kinect ในการตรวจจับการเคลื่อนไหวของผู้ป่วย ในการให้ข้อมูลตำแหน่งของข้อต่อ แล้วนำข้อมูลการตรวจจับการเคลื่อนไหวที่ได้ไปควบคุมตัวการ์ตูนในเกม งานวิจัยนี้ทำการทดสอบเกมที่นำเสนอกับผู้ป่วยทางระบบประสาท ผลของงานวิจัยนี้แสดงให้เห็นถึงเกมที่ช่วยในการพัฒนาการทรงตัวของผู้ป่วยที่มีความผิดปกติ และผู้ป่วยมีความพึงพอใจมาก

2.6.7 Controller-free exploration of medical image data: experiencing the Kinect

งานวิจัยของ L. Gallo, A. P. Placitelli และ M. Ciampi (2011) ได้เสนอระบบการควบคุมการแสดงรูปภาพทางการแพทย์ด้วยท่าทางของผู้ใช้ ตามที่แสดงในภาพที่ 2.17 งานวิจัยนี้ใช้กล้อง Kinect ในการตรวจจับท่าทางของผู้ใช้ ในการให้ข้อมูลตำแหน่งของข้อต่อ แล้วนำข้อมูลตำแหน่งของข้อต่อที่มือและแขนไปวิเคราะห์เป็นคำสั่งในการควบคุมการแสดงรูปภาพ ผลการศึกษาของงานวิจัยนี้แสดงให้เห็นถึงระบบการควบคุมการแสดงรูปภาพทางการแพทย์ที่ปราศจากการใช้อุปกรณ์ควบคุม มีค่าใช้จ่ายน้อย และสามารถหาซื้ออุปกรณ์ได้ตามร้านค้าทั่วไป



ภาพที่ 2.17 ระบบการควบคุมการแสดงผลภาพทางการแพทย์ด้วยท่าทางของผู้ใช้

ที่มา: ([www.researchgate.net/.../224255259_Controller-](http://www.researchgate.net/.../224255259_Controller-free_exploration_of_medical_image_data_Experiencing_the_Kinect)

[free_exploration_of_medical_image_data_Experiencing_the_Kinect](http://www.researchgate.net/.../224255259_Controller-free_exploration_of_medical_image_data_Experiencing_the_Kinect))

2.6.8 Interactive Display Using Depth and RGB Sensors for Face and Gesture Control

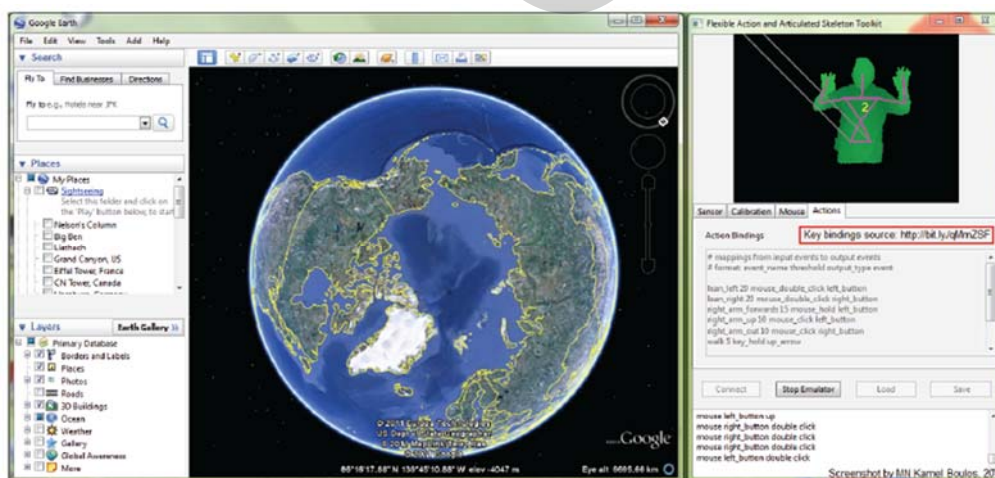
งานวิจัยของ C. Bellmore, R. Ptucha, และ A. Savakis (2011) ได้เสนอระบบแสดงผลแบบโต้ตอบ (Interactive Display System) ที่ควบคุมโดยการแสดงท่าทางของผู้ใช้ และการแสดงออกทางหน้าตาของผู้ใช้ ตามที่แสดงในภาพที่ 2.18 งานวิจัยนี้ใช้กล้อง Kinect ในการตรวจจับการเคลื่อนไหวของผู้ใช้ และใช้ภาพสีที่กล้องตรวจจับได้นำไปวิเคราะห์คำสั่งที่มาจากการแสดงออกทางหน้าตา สำหรับการควบคุมโดยใช้การแสดงท่าทาง จะนำค่าตำแหน่งของข้อต่อที่ได้รับจาก โปรแกรมที่อยู่ในรูปแบบ จุดพิกัดสามมิติไปเปลี่ยนให้เป็นจุดพิกัดสองมิติของรูปภาพก่อนนำค่าตำแหน่งของข้อต่อที่อยู่ในรูปแบบสองมิติไปวิเคราะห์คำสั่งของผู้ใช้ งานวิจัยนี้ทำการทดสอบวิธีการควบคุมระบบแสดงผล ที่นำเสนอด้วยระบบร้านค้ากับผู้ใช้งานมากกว่า 100 คน ผลการศึกษาของงานวิจัยนี้แสดงให้เห็นถึงแนวทางที่เป็นไปได้ในการปฏิสัมพันธ์กับระบบด้วยร่างกายมนุษย์



ภาพที่ 2.18 ระบบแสดงผลที่ควบคุมโดยการแสดงท่าทางและการแสดงออกทางหน้าตาของผู้ใช้
ที่มา: (<http://www.deepdyve.com/lp/institute-of-electrical-and-electronics-engineers/interactive-display-using-depth-and-rgb-sensors-for-face-and-gesture-7fV0UGTQ37>)

2.6.9 Web GIS in practice X a Microsoft Kinect natural

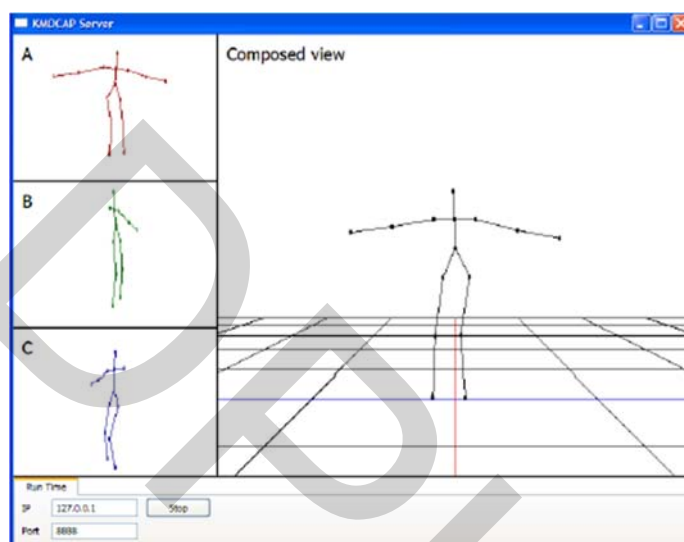
งานวิจัยของ Maged N Kamel Boulos, Bryan J Blanchard, Cory Walker, Julio Montero, Aalap Tripathy และ Ricardo และ Gutierrez-Osuna (2011) ได้ทำการศึกษาการใช้ Kinect ในการใช้ท่าทางและการรู้จำเสียงในการควบคุมส่วนติดต่อกับผู้ใช้บนโปรแกรมกูเกิ้ลเอิร์ธ โดยผลการศึกษสามารถควบคุมได้ทั้งเสียงและท่าทางผลการศึกษางานวิจัยนี้แสดงให้เห็นว่าสามารถนำวิธีการรู้จำเสียงและท่าทางมาใช้ในการใช้เสียงตรวจจับคำพูดได้



ภาพที่ 2.19 ระบบแสดงผลที่ควบคุมโปรแกรมกูเกิ้ลเอิร์ธโดยใช้ท่าทางและเสียง
ที่มา: (<http://www.ij-healthgeographics.com/content/10/1/45>)

2.6.10 การประกอบรวมโครงร่างมนุษย์จากการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว

งานวิจัยของ นราวุฒิ พัฒโนทัย, พรชัย มงคลนาม และ บัณฑิต วรรณภา(2012)ได้ ทำการศึกษาการใช้ Kinect สามตัวต่อเข้าด้วยกันในระบบ Client – Server โดยวางมุมกล้องทำมุม 120 องศาต่อกัน การศึกษาของงานวิจัยนี้แสดงให้เห็นว่าสามารถเพื่อช่วยเพิ่มโอกาสในการมองเห็น จุด Skeleton ของคนในการจับการเคลื่อนไหวได้ดีขึ้น



ภาพที่ 2.20 โครงร่างมนุษย์ที่กล้องแต่ละตัวตรวจจับได้ และโครงร่างมนุษย์ประกอบรวม
ที่มา: (www2.sit.kmutt.ac.th/.../1%20Narawut_NCIT2012%20motionMovement.pdf)

2.6.11 รีวิว Samsung Smart TV 55ES8000

ในปี 2012 บริษัท Samsung Electronics (2012)ได้นำเทคโนโลยีการรู้จำเสียงและการตรวจสอบการเคลื่อนไหวใส่กับอินเทอร์เน็ตทีวี หลักการก็คล้ายๆกับ Kinect ที่กล้องจะจับท่าทาง และตัวทีวีก็คอยจะตอบสนองคำสั่งตามท่าทางของเรา ตัวอย่างเช่น "การแบมือ" ทีวีหน้ากล้องก็จะเริ่มต้นคำสั่ง หลังจากนั้นตัว Cursor ลูกศรก็จะปรากฏขึ้นมาบนหน้าจอ เลื่อนมือไปในทิศทางต่างๆ เพื่อเลื่อนตัวลูกศรที่เป็น Cursor ไปในทิศทางที่เราต้องการ หากต้องการกดปุ่มใดๆบนหน้าจอ ก็กำมือ ซึ่งเปรียบเสมือนการคลิกเมาส์ การสั่งงานด้วยเสียง เพียงแค่พูดว่า "HI TV" เพื่อเปิดหรือปิดการใช้งาน Voice Control หลังจากนั้นจะมี "แถบชุดเมนูคำสั่ง" ขึ้นมาด้านล่างชุดคำสั่ง Voice Control

1. "TV Power Off": ปิดทีวี
2. "Source": เปลี่ยนแหล่งสัญญาณ

3. "Channel Number": เลือกช่องทีวีตามที่ต้องการ
4. "Channel Up หรือ Down": เลื่อนช่องทีวีขึ้นหรือลงทีละช่อง
5. "Volume Up/Down": เพิ่มระดับเสียงดัง/เบา
6. "Mute": ปิดเสียง



ภาพที่ 2.21 ตัวอย่าง Smart TV รุ่น 55ES8000

ที่มา: (http://www.lcdtvthailand.com/review/detail.asp?desc=1¶m_id=1277,2012)

ดังนั้นงานวิจัยนี้จึงได้เสนอและออกแบบและพัฒนาเทคนิคการผสมผสานเทคนิค ระหว่างการรู้จำเสียงและเทคนิคการตรวจสอบการเคลื่อนไหวเข้าด้วยกันเพื่อใช้ในการควบคุม อินเทอร์เน็ตทีวีโดยออกแบบการควบคุมไว้ดังนี้ เปลี่ยนช่อง 3,5,7,9,11,thai PBS เพิ่มลดเสียงได้ 5 ระดับและการเปิดปิด TV

บทที่ 3

วิธีการดำเนินการและเครื่องมือ

บทนี้จะกล่าวถึงวิธีการผสมผสานเทคนิคการรู้จำเสียงและการตรวจจับการเคลื่อนไหว โดยใช้กล้องคีนะ วิถีการทดสอบ โดยสร้างสถานการณ์การใช้เทคนิคควบคุมอินเทอร์เน็ตทีวีจำลองสามเทคนิค ได้แก่ การรู้จำเสียง, การตรวจจับการเคลื่อนไหว และเทคนิคผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว และพัฒนาโปรแกรมส่วนควบคุมอินเทอร์เน็ตทีวี ดังต่อไปนี้

- 3.1 การวางแผนการดำเนินงาน
- 3.2 การศึกษาค้นคว้าหาข้อมูล
- 3.3 การออกแบบและพัฒนา Interface ส่วนควบคุมอินเทอร์เน็ตทีวีจำลอง
- 3.4 การพัฒนาระบบการรู้จำเสียง
- 3.5 การพัฒนาระบบการตรวจจับการเคลื่อนไหว
- 3.6 การพัฒนาระบบเทคนิคการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว
- 3.7 การกำหนดแบบแผนการทดลองและกลุ่มผู้ทดสอบ
- 3.8 การวัดผลทั้งด้านความเร็วและความถูกต้องในการควบคุมอินเทอร์เน็ตทีวีจำลอง
- 3.9 เครื่องมือที่ใช้จัดทำระบบควบคุมอินเทอร์เน็ตทีวีจำลอง

3.1 การวางแผนการดำเนินงาน

การวางแผนการดำเนินงานเป็นขั้นตอนสำคัญ และเป็นขั้นตอนแรก ก่อนที่ผู้วิจัยจะเริ่มทำงานวิจัย เพราะในการวางแผนการดำเนินงานนั้นจะบอกถึงรายละเอียดงานที่ผู้วิจัยควรทำก่อนหลัง รวมไปถึงระยะเวลาในการทำงานในแต่ละขั้นตอนของงาน ทำให้ผู้วิจัยสามารถดำเนินงานได้เสร็จตรงตามเวลาที่กำหนดไว้ได้ โดยได้ผลของการวางแผนดำเนินงานตามภาพที่ 3.1

3.2 การศึกษาค้นคว้าหาข้อมูล

ผู้วิจัยได้ศึกษาและค้นคว้าหาข้อมูลจากแหล่งต่างๆ ที่จะสร้างการควบคุมอินเทอร์เน็ตที่จำลองด้วย ผสมผสานเทคนิคการรู้จำเสียงและการตรวจจับการเคลื่อนไหว ในหัวข้อต่อไปนี้

3.2.1 ศึกษาการใช้งานกล้องคิเนค

โดยผู้วิจัยได้ศึกษาหาข้อมูลเกี่ยวกับอุปกรณ์ฮาร์ดแวร์ต่างๆ ที่มีคุณสมบัติที่มีความสามารถในการรับรู้ภาพและเสียงได้ในเวลาเดียวกัน ซึ่งอุปกรณ์ Kinect ของไมโครซอฟท์ นั้นมีความสามารถในการรับรู้ภาพและเสียง โดยมีคุณสมบัติเด่นๆดังนี้

3.2.1.1 การรับรู้ภาพสี RGB มีความสามารถถ่ายภาพหรือบันทึกภาพเป็นวิดีโอได้

3.2.1.2 Depth sensor มีความสามารถในการแยกแยะภาพทางลึกของภาพได้

3.2.1.3 Multi-array microphone มีความสามารถในการรับรู้เสียง บันทึกเสียงได้ และสามารถแยกแยะทิศทางเสียงที่ได้เข้าสู่ตัวไมโครโฟน มีระบบตัดเสียงรบกวนผู้วิจัยจึงเลือกใช้กล้องคิเนคในงานวิจัย

3.2.2 ศึกษาวิธีสร้างการรู้จำเสียง

โดยผู้วิจัยได้ศึกษาตัวแบบของการรู้จำเสียง จากบทความที่ตีพิมพ์ในนิตยสารวิชาการต่างประเทศและในประเทศ รวมถึงบทความจากเว็บไซต์ต่างๆ หลากหลายตัวแบบผู้วิจัยจึงเลือกใช้โมดูลของไมโครซอฟท์ Speech SDK 5.1 ซึ่งมีคุณสมบัติรองรับภาษาต่างๆมากมาย ทั้งภาษาอังกฤษ จีน ญี่ปุ่น และสามารถทำงานร่วมกับอุปกรณ์กล้องคิเนคที่ผู้วิจัยใช้ได้เป็นอย่างดี โดยผู้วิจัยได้เลือกใช้วิธี Speech To Text (STT) ในการสร้างกระบวนการรู้จำเสียง

3.2.3 ศึกษาวิธีสร้างการตรวจจับการเคลื่อนไหว

ผู้วิจัยได้เลือกใช้ไลบรารี Kinect SDK Beta 1 มาเป็นเครื่องมือในการสร้างการตรวจจับการเคลื่อนไหว โดยใช้ภาษา C# โดยศึกษาวิธีการใช้งานจาก ตำราภาษาต่างประเทศ และบทความต่างๆทั้งในวารสารหรือบทความที่อยู่ในเว็บไซต์ในอินเทอร์เน็ต เพื่อนำความรู้ที่ได้มาสร้างกระบวนการตรวจจับการเคลื่อนไหว

3.2.4 ศึกษาวิธีสร้าง User Interface สำหรับติดต่อผู้ใช้

ผู้วิจัยได้เลือกใช้ Windows Presentation Foundation (WPF) ในการเขียนโปรแกรมเพื่อสร้าง User Interface สำหรับอินเทอร์เน็ตที่จำลอง โดยศึกษาวิธีการใช้งานจาก ตำราภาษาต่างประเทศ และบทความต่างๆทั้งในวารสารหรือบทความที่อยู่ในเว็บไซต์ในอินเทอร์เน็ต เพื่อนำความรู้ที่ได้มาสร้าง User Interface สำหรับติดต่อผู้ใช้

3.3 การออกแบบและพัฒนา Interface ส่วนควบคุมอินเทอร์เน็ตทีวีจำลอง

ในการออกแบบหน้าต่างของอินเทอร์เน็ตเฟส จะใช้ Windows Presentation Foundation (WPF) และใช้ XAML ในการออกแบบ Interface โดยจะมีรูปภาพบอกถึงสถานการณ์เข้าในงาน โหมด การรู้จำเสียง และในขณะที่เปิดโหมดการทำงานนี้เมื่อผู้ใช้งานสั่งงานอินเทอร์เน็ตทีวีใดๆ ก็จะถูกแสดงผลในรูปแบบของ ข้อความที่หน้าจอ เพื่อบอกว่าผู้ใช้ได้ออกเสียงเพื่อสั่งงานอะไรไป หรือ Curser มือที่จะใช้สำหรับการการตรวจจับการเคลื่อนไหว



ภาพที่ 3.1 แสดงอินเทอร์เน็ตเฟสของระบบควบคุมอินเทอร์เน็ตทีวีจำลอง

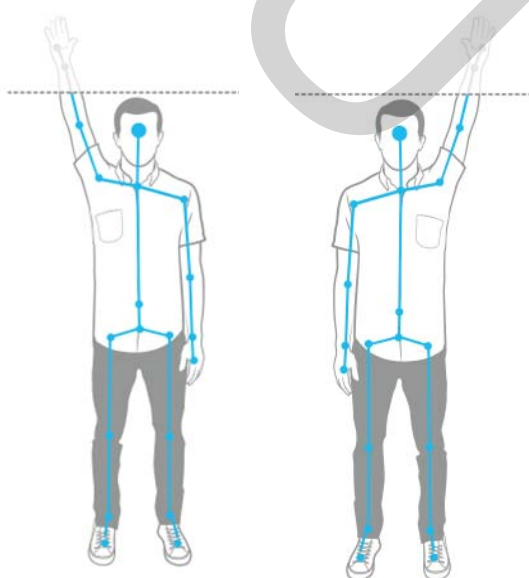
1. แสดงผลอินเทอร์เน็ตทีวีช่องที่กำลังใช้งาน
2. แสดงสถานการณ์เปิดหรือปิดระบบการรู้จำเสียง
3. แสดงข้อความที่ผู้ใช้ออกเสียงในการเปลี่ยนช่องอินเทอร์เน็ตทีวี
4. ระดับเสียงของอินเทอร์เน็ตทีวี
5. แสดงผังรายการให้ผู้ใช้
6. แสดงภาพของผู้ใช้ในกรณีที่ใช้โหมดการตรวจสอบการเคลื่อนไหว
7. แสดงรูป Curser มือที่ใช้ในกรณีการตรวจจับการเคลื่อนไหวของมือขวาของผู้ใช้

3.4 การพัฒนาระบบการรู้จำเสียง

ในการตรวจสอบการรู้จำเสียงอินพุตที่เข้ามาผ่านไมโครโฟนบนตัวของกล้อง Kinect ได้นำไลบรารีของ Microsoft Speech Platform Runtime 11 โดยจะรองรับภาษา En-Us เขียนโปรแกรมผ่าน Visual Studio 2010 เพื่อนำอินพุตเสียงที่ได้แปลงเป็นตัวอักษรโดยใช้เทคนิค Speech to Text (STT) จากนั้นนำคำที่ได้ไปเปรียบเทียบกับคำที่ระบุไว้หากไม่ตรงตามที่ระบุไว้ก็จะ Reject ทิ้ง หากตรงตามเงื่อนไขก็จะ Accepted

รูปแบบของการ Capture Audio โดยจะให้คุณภาพของเสียงอยู่ที่ อัตราการสุ่มเสียง 16000 Sampling Rate ค่าความลึกของเสียง 16 bit/sec ช่องสัญญาณเสียง 1 Channel อัตราการ Stream ข้อมูลเสียง 32000 byte/sec

โดยการสั่งงานด้วยเสียงผู้วิจัยจะออกแบบให้มีกระบวนการเปิดและปิดระบบการรู้จำเสียงดังภาพที่ 3.4 เพื่อป้องกันเสียงรบกวนในขณะที่ผู้ใช้งานยังไม่ได้มีความต้องการในการใช้คำสั่ง จะกำหนดให้ผู้ใช้นิ้วชี้ขึ้นเหนือศีรษะจะเป็นการเปิดระบบการรู้จำเสียงและหากต้องการปิดให้ผู้ใช้นิ้วชี้ขึ้นเหนือศีรษะจะเป็นการปิดระบบเสียง โดยจะอาศัยตำแหน่งของข้อต่อที่ได้รับจาก API และแต่ละตำแหน่งของข้อต่อจะมีค่า X,Y,Z ในหน่วยมิลลิเมตร จึงได้นำตำแหน่ง Joint Hand Right และ Joint Hand Left ทำการเปรียบเทียบตำแหน่ง Position ด้านแกน Y กับ Joint Head ดังภาพที่ 3.3



ภาพที่ 3.2 แสดงวิธีการเปิดปิดระบบการสั่งการด้วยเสียง

ที่มา: (<http://video.ch9.ms/teched/2012/eu/DEV330.pptx>,2012)

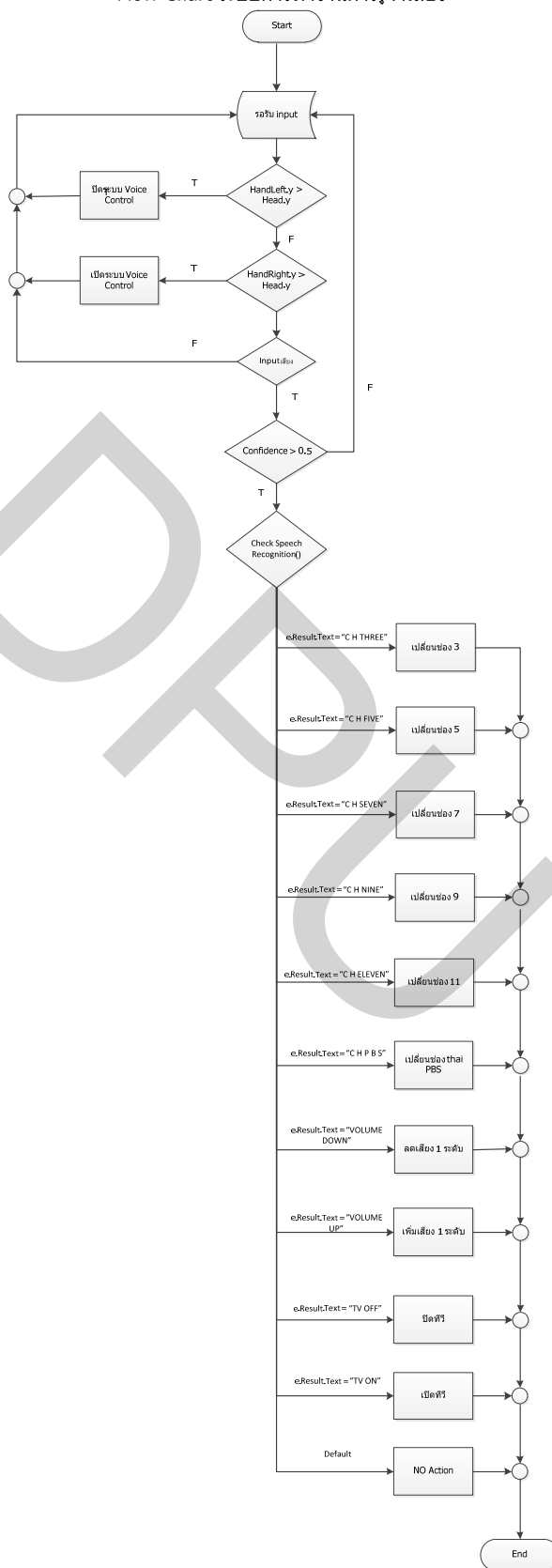


ภาพที่ 3.3 แสดง Interface ระบบการสั่งงานด้วยเสียง

ที่มา: (<http://video.ch9.ms/teched/2012/eu/DEV330.pptx>, 2012)

ในระบบการรู้จำเสียงของ Input ที่เข้ามา โดยจะทำการเปลี่ยนสัญญาณเสียงที่เข้าเพื่อนำไปเปรียบเทียบกับคำที่ระบุไว้จะมีฟิลด์ Confidence แสดงสถานะ การตรวจจับของเสียงนั้นๆ โดยจะมีค่าระหว่าง 0 – 1 ซึ่ง 1 หมายถึง สามารถตรวจจับได้ และตรงกับเงื่อนไขของคำที่ระบุไว้ โดยจะระบุ ถ้ามากกว่า 0.5 ก็จะเป็นการ Accepted คำนั้นๆเพื่อนำไปตรวจสอบ หากน้อยกว่า 0.5 ก็จะ Reject คำนั้นๆทิ้งไป

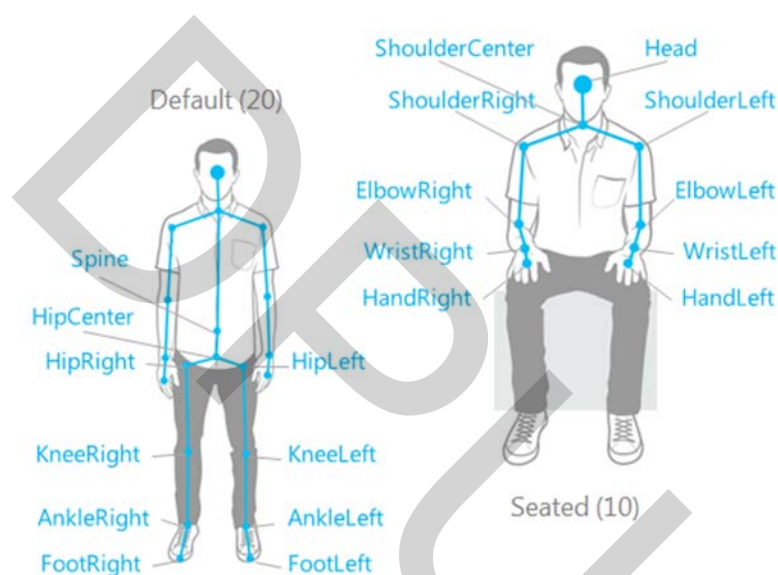
Flow Chart ระบบการทำงานการรู้จำเสียง



ภาพที่ 3.4 Flow Chart แสดงการทำงานของระบบรู้จำเสียง

3.5 การพัฒนาระบบการตรวจจับการเคลื่อนไหว

ในการตรวจสอบการเคลื่อนไหวภาพผ่านหน้ากล้อง Kinect ได้นำไลบรารีของ Kinect SDK โดยจะเขียนโปรแกรมผ่านโปรแกรม Visual Studio 2010 โดยทำการตรวจจับรูปร่างของคน แล้วจะนำ Skeleton ไปซ้อนทับบนภาพ โดยทำขึ้นจะกำหนดเป็น 20 Joint และทำนั้นจะกำหนดเป็น 10 Joint จากนั้นก็ทำการเขียนโปรแกรมเก็บค่า เฉพาะ Joint Hand Right และ Joint Hand Left เพื่อนำค่า Position ของตำแหน่งมือไปเปรียบเทียบกับตำแหน่ง Cursor มือนบนจอภาพ เพื่อให้ได้ตำแหน่งและพิกัด สำหรับการควบคุมอินเทอร์เนตทีวี โดยจะแบ่งขั้นตอนออกเป็นดังนี้

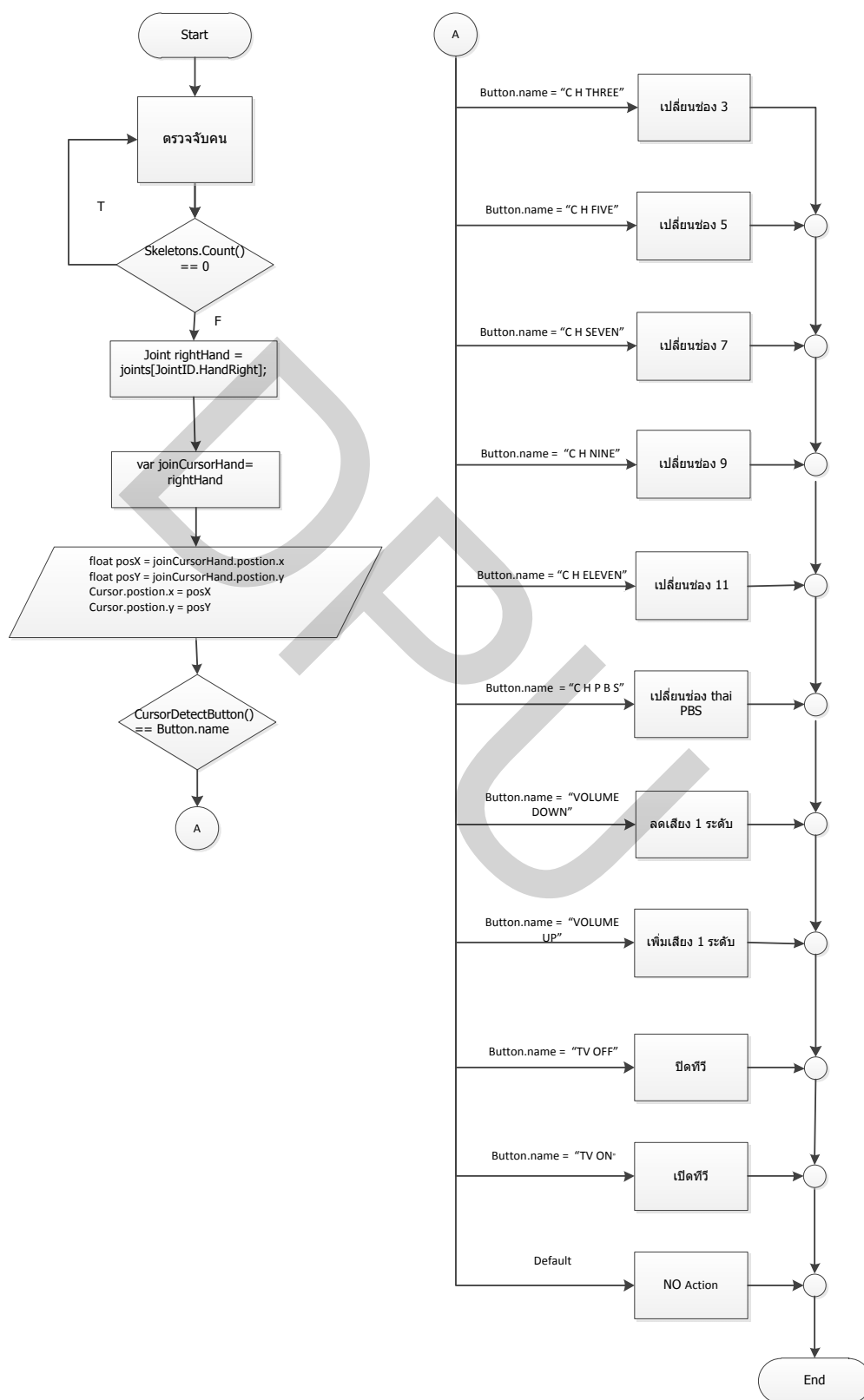


ภาพที่ 3.5 แสดงจุดข้อต่อที่ API ส่งค่ามา

ที่มา: (<http://video.ch9.ms/teched/2012/eu/DEV330.pptx>,2012)

Hand Tracking โดยจะเป็นขั้นตอนของการเชื่อมโยงข้อมูลที่ได้รับจากข้อต่อ API ในส่วนของ Hand Right กับรูปภาพ Curser มือนบนจอภาพ โดยอันดับแรกให้โปรแกรมทำการเรียกข้อมูลข้อต่อสำหรับ คน 1 คนขึ้นมา จากนั้นเมื่อได้ข้อมูลชุดข้อต่อจากที่สามารถเรียกได้ผ่านตัวแปร first Person แล้วก็ทำการนำข้อมูลข้อต่อของโหนด Hand Right หาค่า Position X, Y ออกมา และนำไปเปรียบเทียบกับตำแหน่งของ Cursor Hand บนจอภาพ

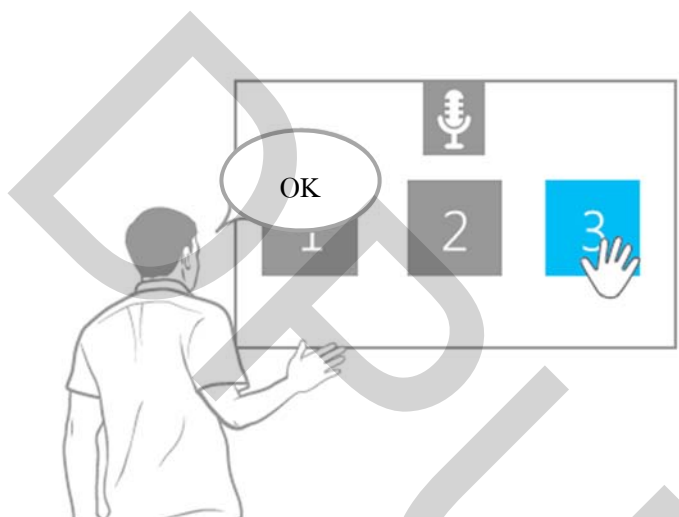
Flow Chart ระบบการทำงานการตรวจจับการเคลื่อนไหว



ภาพที่ 3.6 Flow Chart แสดงการทำงานของระบบการตรวจจับการเคลื่อนไหว

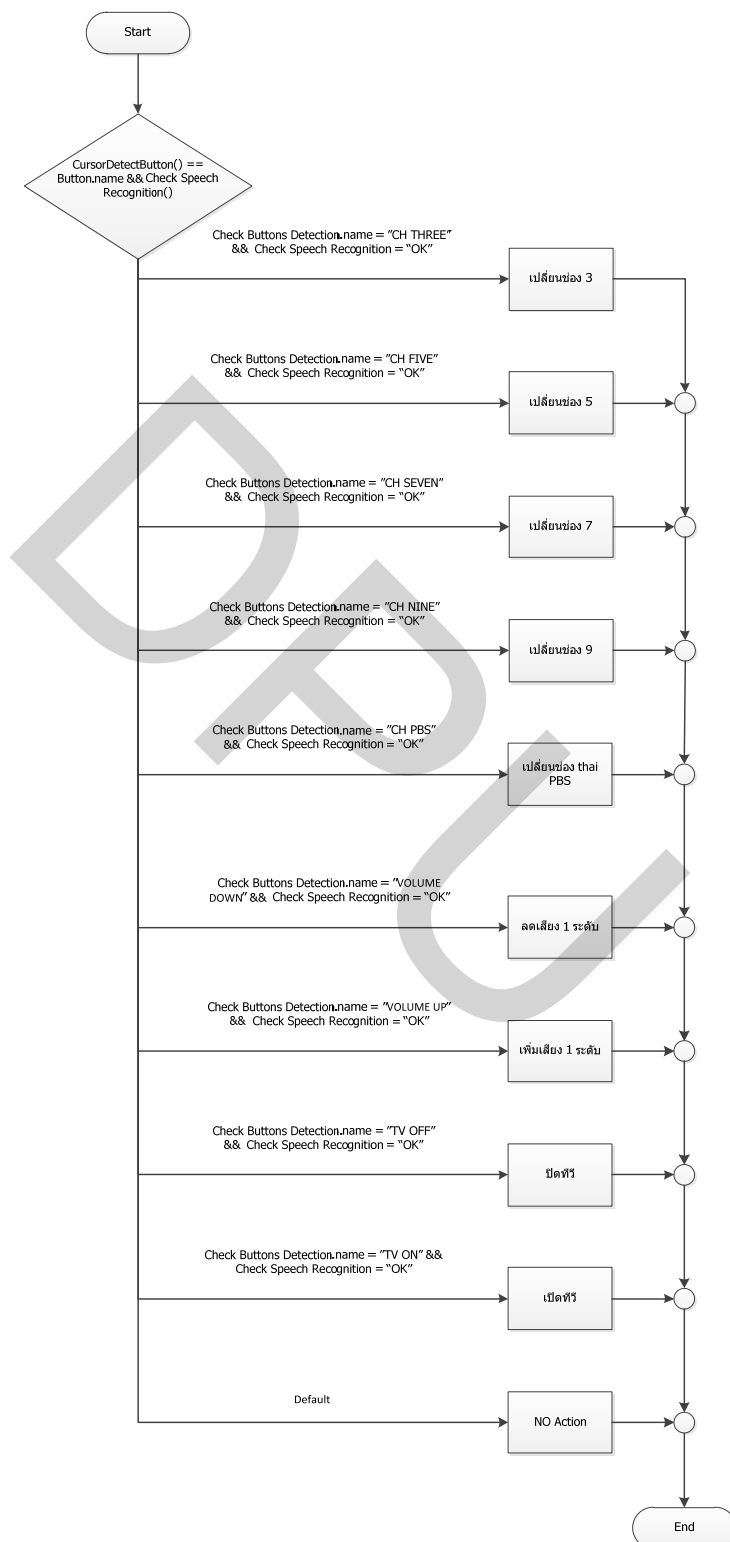
3.6 การพัฒนาระบบเทคนิคการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว

เมื่อได้วิธีการรู้จำเสียงและการตรวจจับการเคลื่อนไหวแล้ว จึงได้ออกแบบวิธีการผสมผสานเทคนิคการรู้จำเสียงและการตรวจสอบการเคลื่อนไหวโดยกำหนดให้ เมื่อ Position ของ Cursor บนหน้าจออินเทอร์เน็ตทีวี ไป Detect กับปุ่มใดๆบนอินเตอร์เฟซบนหน้าจอพร้อมทั้งออกเสียง “OK” ก็จะเป็นการสั่งงานกับเมนูนั้นๆ ดังภาพที่ 3.8 แต่ถ้าหาก Cursor ไป Detect กับปุ่มใดๆบนอินเตอร์เฟซบนหน้าจอแล้วมีการออกเสียงอื่นโปรแกรมก็จะไม่มีการทำงาน หรือถ้าออกเสียงอย่างเดียวแล้วไม่ได้มีการนำ Cursor ไป Detect กับปุ่มใดๆ โปรแกรมก็จะไม่มีการทำงานอีกเช่นกัน



ภาพที่ 3.7 แสดงรูปแบบการทำงานของวิธีการผสมผสานเทคนิคการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว

Flow Chart ระบบการทำงานการผสมผสานการรู้จำเสียง และการตรวจจับการเคลื่อนไหว



ภาพที่ 3.8 Flow Chart แสดงการทำงานของระบบการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว

3.7 การกำหนดแบบแผนการทดลองและกลุ่มผู้ทดสอบ

งานวิจัยนี้ทำการทดสอบประสิทธิภาพโดยการวัดความเร็วและความถูกต้องของวิธีการที่นำเสนอและได้ทำการทดลองทั้งหมด 2 ครั้ง โดยครั้งที่ 1 จะเป็นการทดสอบเบื้องต้น(ตารางที่ 3.2) และการทดสอบครั้งที่ 2 (ตารางที่ 3.3) จะมีการปรับโจทย์การทดสอบให้เหมือนกับการใช้คำสั่งในชีวิตประจำวันมากขึ้น โดยการพัฒนาโปรแกรมทั้งสามระบบ คือ 1.ระบบการรู้จำเสียงสำหรับการควบคุมอินเทอร์เนตทีวี 2.ระบบการตรวจสอบการเคลื่อนไหวสำหรับการควบคุมอินเทอร์เนตทีวี 3.ระบบผสมผสานระหว่างการรู้จำเสียงและการตรวจจับการเคลื่อนไหว ขึ้นมาด้วยภาษา C# ใช้ Kinect SDK เวอร์ชัน Beta 1 และสร้างสถานการณ์สภาพแวดล้อมในห้องขึ้นมาโดยใช้ห้องประชุมขนาดประมาณ 4x8 เมตร โดยให้สภาพแวดล้อมแรกเป็นสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวนโดยที่ค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB สภาพแวดล้อมที่สองคือสภาพแวดล้อมที่มีเสียงรบกวนโดยกำหนดให้เสียงรบกวนอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เนตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB ในการตรวจจับตั้งกล้องอยู่บนระนาบที่ขนานกับระนาบพื้นที่ทำกรตรวจจับสูงจากระนาบพื้น 0.6 เมตร ห้างออกไป 1.5 เมตร โดยกำหนดโจทย์ที่ใช้ทดสอบครั้งที่หนึ่ง มีรายละเอียดดังต่อไปนี้ ปิด TV, เลือกช่อง (เปิด TV), เปลี่ยนช่อง, เพิ่มเสียง, ลดเสียง

ตารางที่ 3.2 แสดงถึงงานที่ผู้ทดสอบครั้งที่ 1 ต้องทำในแต่ละชุดการทดสอบในทำขึ้น

โจทย์สำหรับใช้ทดสอบ		
ชุดทดสอบ A	ชุดทดสอบ B	ชุดทดสอบ C
เลือกช่อง 3	เลือกช่อง 5	เลือกช่อง 9
เปลี่ยนช่อง 7	เพิ่มเสียงขึ้น 1 ระดับ	เพิ่มเสียงขึ้น 1 ระดับ
ลดเสียงลง 1 ระดับ	เปลี่ยนช่อง 9	ปิด TV
เพิ่มเสียงขึ้น 1 ระดับ	ลดเสียงลง 1 ระดับ	เปลี่ยนช่อง 11
ปิด TV	ปิด TV	ลดเสียงลง 1 ระดับ

โดยจะใช้ผู้ทดสอบจำนวน 60 คนนักศึกษาชั้นปี 1 – 3 ช่วงอายุประมาณ 19 – 21 ปี สถาบันเทคโนโลยีไทย-ญี่ปุ่น ซึ่งไม่มีประสบการณ์การใช้ระบบเหล่านี้มาก่อน โดยจะแบ่งการทดสอบเป็น 2 กลุ่ม กลุ่มละ 30 คนโดยให้กลุ่มแรกทำการทดสอบในสภาพแวดล้อมที่ไม่มีเสียงเข้า

มารบกวณค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB และกลุ่มที่สองอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เนตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB โดยให้แต่ละกลุ่มทำการทดสอบดังนี้

1.ให้ผู้ทดสอบ สุ่มเลือก เทคนิคสำหรับควบคุมอินเทอร์เนตทีวี จาก 3 วิธี มา 1 วิธี พร้อมทั้งนำหัวข้อเทคนิคที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก

2.จากนั้นให้ผู้ทดสอบ สุ่มเลือก ชุดทดสอบ จาก 3 ชุดทดสอบ มา 1 ชุดทดสอบ พร้อมทั้งนำหัวข้อชุดทดสอบ ที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก

3.ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมินจะเริ่มทำการจับเวลาในการทดสอบ

4.เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป

5.ให้ผู้ทดสอบทำการ สุ่มเลือก เทคนิคสำหรับควบคุมอินเทอร์เนตทีวี จาก 2 วิธีที่เหลือ มา 1 วิธี พร้อมทั้งนำหัวข้อเทคนิคที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก

6.จากนั้นให้ผู้ทดสอบ สุ่มเลือก ชุดทดสอบ จาก 2 ชุดทดสอบ มา 1 ชุดทดสอบ พร้อมทั้งนำหัวข้อชุดทดสอบ ที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก

7.ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมินจะเริ่มทำการจับเวลาในการทดสอบ

8.เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป

9.จากนั้นให้ผู้ทดสอบทำการทดสอบ เทคนิคสำหรับควบคุมอินเทอร์เนตทีวีและชุดทดสอบ ที่เหลืออยู่

10.ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมินจะเริ่มทำการจับเวลาในการทดสอบ

11.เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป

12.นำข้อมูลที่ได้จากทั้ง 3 กลุ่ม มาวิเคราะห์เปรียบเทียบความแตกต่างในเรื่องของความเร็วและความถูกต้องของการทำงาน

โจทย์ที่ใช้ทดสอบครั้งที่ 2 มีรายละเอียดดังต่อไปนี้ เปิด TV, ปิด TV, เลือกช่อง, เปลี่ยนช่อง, เพิ่มเสียง, ลดเสียง

ตารางที่ 3.3 แสดงถึงงานที่ผู้ทดสอบครั้งที่ 2 ต้องทำในชุดการทดสอบในทำขึ้น

ลำดับที่	โจทย์สำหรับใช้ทดสอบ
1	เปิด TV (TV ON)
2	เปลี่ยนช่อง 5 (CH FIVE)
3	เปลี่ยนช่อง 9 (CH NINE)
4	เพิ่มเสียง 1 ระดับ (Volume UP)
5	เปลี่ยนช่อง 3 (CH THREE)
6	เปลี่ยนช่อง 7 (CH SEVEN)
7	ลดเสียง 1 ระดับ (Volume DOWN)
8	เพิ่มเสียง 1 ระดับ (Volume Up)
9	ปิด TV (TV OFF)

โดยจะใช้ผู้ทดสอบจำนวน 30 คน นักศึกษาชั้นปี 1 – 3 ช่วงอายุประมาณ 19 – 21 ปี สถาบันเทคโนโลยีไทย-ญี่ปุ่น ซึ่งไม่มีประสบการณ์การใช้ระบบเหล่านี้มาก่อน โดยจะแบ่งการโดยให้ทำการทดสอบในสภาพแวดล้อมที่มีเสียงเข้ามารบกวนอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เน็ตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB โดยให้แต่ละกลุ่มทำการทดสอบดังนี้

1. ใส่ลำดับเทคนิคสำหรับควบคุมอินเทอร์เน็ตทีวี ทั้ง 6 ลำดับ ลำดับละ 5 ชุด

ตารางที่ 3.4 ตารางแสดงลำดับการควบคุมในแต่ละเทคนิคการควบคุมอินเทอร์เน็ตทีวี

ลำดับการควบคุม	1 : Voice Only	2 : Motion Detection Only	3 : A Combination
A	1	2	3
B	1	3	2

C	2	1	3
D	2	3	1
E	3	1	2
F	3	2	1

2. ผู้เข้าทดสอบจะสุ่มหยิบลำดับการควบคุม 1 ลำดับการควบคุมจากกองสุ่มเลือก (1 คน ต้องทำทั้ง 3 เทคนิค)

3. ผู้เข้าทดสอบหยิบโจทย์ที่ใช้ในการควบคุม (Task)

4. ผู้เข้าทดสอบ เข้ามายืนในจุด เริ่มการทดสอบ ผู้ควบคุมเวลาจะเริ่มนับจับเวลา

5. ระหว่างที่ผู้ทดสอบทำโจทย์ทดสอบ ผู้ควบคุมเวลาจะคอยบันทึก ข้อผิดพลาดที่เกิดขึ้นในการควบคุม TV

6. เมื่อผู้ทดสอบทำ โจทย์ จนครบ ผู้ควบคุมเวลาจะบันทึกเวลาที่ใช้ในแต่ละวิธีในการควบคุม TV

7. ผู้ทดสอบคนต่อไปเข้ามาสุ่มหยิบลำดับและทำกระบวนการนี้ ซ้ำๆ จนครบทั้ง 30 คน

3.8 การวัดผลทั้งด้านความเร็วและความถูกต้องในการควบคุมอินเทอร์เน็ตทีวีจำลอง

ในส่วน of เครื่องมือในการวิจัยนั้น นอกจากการออกแบบเทคนิคการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหวในการควบคุมอินเทอร์เน็ตทีวีแล้ว ยังมีการวัดสถานะของระบบ ในเรื่องของความเร็วในการควบคุมอินเทอร์เน็ตทีวี และความถูกต้องในการสั่งการอินเทอร์เน็ตทีวี ทั้ง 3 วิธี คือ 1.ระบบการรู้จำเสียงสำหรับการควบคุมอินเทอร์เน็ตทีวี 2.ระบบการตรวจสอบการเคลื่อนไหวสำหรับการควบคุมอินเทอร์เน็ตทีวี 3.ระบบผสมผสานระหว่างการรู้จำเสียงและการตรวจจับการเคลื่อนไหว

3.8.1 การวัดผลด้านความเร็วในการควบคุมอินเทอร์เน็ตทีวีจำลอง

ในการวัดผลด้านความเร็วในการควบคุมอินเทอร์เน็ตทีวีจำลองนั้น จะให้ผู้ทดสอบทำโจทย์การทดสอบ จนกว่าจะเสร็จสิ้นกระบวนการสั่งการอินเทอร์เน็ตทีวีจำลอง ดังตารางที่ 3.2 จากนั้นจะนำค่าเวลาที่ได้ ของแต่ละคนทั้ง 30 คนมาหาค่าเฉลี่ยความเร็วในการควบคุมอินเทอร์เน็ตทีวีจำลอง ทั้ง 3 วิธี และ 2 สถานการณ์ คือ ในสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวน ค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB และกลุ่มที่สองอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินและรับรู้เสียงที่ออกมาจากอินเทอร์เน็ตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB ผู้วิจัยใช้สูตรดังนี้

$$\text{Average}_{(v,z)} = \frac{\sum N}{Y}$$

เมื่อ

Average = ค่าเฉลี่ยความเร็วในการควบคุมอินเทอร์เนตทีวีจำลองของแต่ละวิธี

V = รูปแบบที่ใช้ในการควบคุมอินเทอร์เนตทีวีจำลอง

Z = สภาพแวดล้อมในการทดสอบ

N = เวลาที่ได้จากการทำแบบทดสอบของแต่ละวิธี (วินาที)

Y = จำนวนผู้ทดลองทั้งหมดในแต่ละวิธี

3.8.2 การวัดผลด้านความถูกต้องในการควบคุมอินเทอร์เนตทีวีจำลอง

ในการวัดผลด้านความถูกต้องในการควบคุมอินเทอร์เนตทีวีจำลองนั้น จะให้ผู้ทดสอบทำโจทย์การทดสอบ จนกว่าจะเสร็จสิ้นกระบวนการการสั่งการอินเทอร์เนตทีวีจำลอง ดังตารางที่ 3.2 และ 3.3 แล้วนับจำนวนครั้งที่ผู้ทดสอบออกคำสั่งผิดพลาด หรือระบบทำงานผิดพลาด และ 2 สถานการณ์ คือ ในสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB และกลุ่มที่สองอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เนตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB

3.9 เครื่องมือที่ใช้จัดทำระบบควบคุมอินเทอร์เนตทีวีจำลอง

ในการจัดทำระบบควบคุมอินเทอร์เนตทีวีจำลองนั้นใช้ เครื่องมือประกอบด้วยส่วนของฮาร์ดแวร์ (Hardware) และส่วนของซอฟต์แวร์ (Software) สรุปได้ดังนี้

3.9.1 ฮาร์ดแวร์ (Hardware)

ในส่วนฮาร์ดแวร์นั้นผู้วิจัยได้ใช้เครื่องคอมพิวเตอร์แบบพกพา (Notebook Computer) ในการจัดทำระบบควบคุมอินเทอร์เนตทีวีจำลองและในการวัดประสิทธิภาพของระบบ ซึ่งคอมพิวเตอร์ที่ใช้นั้นมีรายละเอียดดังนี้

Brand: MSI

Model: GX640

CPU: Inter Core I5-420M

RAM: Samsung DD3 Ram 8 GB

VGA: ATI HD5850

Hard disk: Hitachi 500 GB SATA II 5200 rpm

Monitor: LED Display 15" resolution 1280 x 800

Microsoft Kinect Xbox360

3.8.2 ซอฟต์แวร์ (Software)

1. Microsoft Windows 7 Ultimate x64
2. Microsoft Kinect for Windows SDK BETA
3. Microsoft Visual Studio 2010
4. Microsoft Speech SDK 5.1
5. Adobe Photoshop CS6

บทที่ 4

ผลการดำเนินงาน

การวิจัยครั้งนี้ มีวัตถุประสงค์ในการพัฒนาตัวแบบสำหรับการควบคุมอินเทอร์เนตทีวีจำลองแบบใหม่ด้วยการผสมผสานเทคนิคการรู้จำเสียงและเทคนิคการตรวจจับการเคลื่อนไหวและทดสอบประสิทธิภาพความเร็วในการควบคุมและความถูกต้องแม่นยำในการสั่งการของวิธีการควบคุมอินเทอร์เนตทีวีจำลอง ทั้งสามวิธี คือการรู้จำเสียง การตรวจจับการเคลื่อนไหว และการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว ทดสอบนักศึกษาชั้นปีที่ 1-3 จำนวนทั้งสิ้น 90 คน ซึ่งได้ผลการวิจัยจำแนกได้เป็นหัวข้อ ดังนี้

4.1 ผลการวิจัยการทดสอบเบื้องต้นในสภาพแวดล้อมที่มีเสียงรบกวนและไม่มีเสียงรบกวน

4.2 ผลการวิจัยการทดสอบครั้งที่ 2 ปรับปรุงโจทย์ทดสอบให้มีคำสั่งมากขึ้นและมีเสียงรบกวนที่ดังออกจากทีวี

4.1 ผลการวิจัยการทดสอบเบื้องต้นในสภาพแวดล้อมที่มีเสียงรบกวนและไม่มีเสียงรบกวน

ในผลการทดสอบเบื้องต้นจะใช้ผู้ทดสอบจำนวน 60 คน นักศึกษาชั้นปี 1 – 3 ช่วงอายุประมาณ 19 – 21 ปี สถาบันเทคโนโลยีไทย-ญี่ปุ่น ซึ่งไม่มีประสบการณ์การใช้ระบบเหล่านี้มาก่อน โดยจะแบ่งการทดสอบเป็น 2 กลุ่ม กลุ่มละ 30 คน โดยให้กลุ่มแรกทำการทดสอบในสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB และกลุ่มที่สองอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เนตทีวี ค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB โดยใช้ห้องประชุมขนาดประมาณ 4x8 เมตร ตั้งกล้องอยู่บนระนาบที่ขนานกับระนาบพื้นที่ทำการตรวจจับสูงจากระนาบพื้น 0.6 เมตร ห่างออกไป 1.5 เมตร มีรายละเอียดดังนี้

4.1.1 เทคนิคที่ทดสอบ

1. Voice Only
2. Motion Detection Only

3. A Combination (Voice and Motion Detection)

4.1.2 Environment

1. สภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB ห้องประชุมขนาดประมาณ 4x8 เมตร

2. สภาพแวดล้อมที่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 70dB ห้องประชุมขนาดประมาณ 4x8 เมตร

4.1.3 Task

- 1.เปิด TV
- 2.เลือกช่อง (เปิด TV)
- 3.เปลี่ยนช่อง
- 4.เพิ่มเสียง
5. ลดเสียง

4.1.4 โจทย์สำหรับใช้ทดสอบ

ตารางที่4.1 แสดงถึงงานที่ผู้ทดสอบเบื้องต้นในสภาพแวดล้อมที่มีเสียงรบกวนและไม่มีเสียงรบกวนต้องทำในแต่ละชุดการทดสอบในทำขึ้น

โจทย์สำหรับใช้ทดสอบ		
ชุดทดสอบ A	ชุดทดสอบ B	ชุดทดสอบ C
เลือกช่อง 3	เลือกช่อง 5	เลือกช่อง 9
เปลี่ยนช่อง 7	เพิ่มเสียงขึ้น 1 ระดับ	เพิ่มเสียงขึ้น 1 ระดับ
ลดเสียงลง 1 ระดับ	เปลี่ยนช่อง 9	ปิด TV
เพิ่มเสียงขึ้น 1 ระดับ	ลดเสียงลง 1 ระดับ	เปลี่ยนช่อง 11
ปิด TV	ปิด TV	ลดเสียงลง 1 ระดับ

4.1.5 วิธีการดำเนินการทดสอบเบื้องต้นในสภาพแวดล้อมที่มีเสียงรบกวนและไม่มีเสียง

รบกวน

1. ให้ผู้ทดสอบ สุ่มเลือก เทคนิคสำหรับควบคุมอินเทอร์เนตทีวี จาก 3 วิธี มา 1 วิธี พร้อมทั้งนำหัวข้อเทคนิคที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
2. จากนั้นให้ผู้ทดสอบ สุ่มเลือก ชุดทดสอบ จาก 3 ชุดทดสอบ มา 1 ชุดทดสอบ พร้อมทั้งนำหัวข้อชุดทดสอบ ที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
3. ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมิน จะเริ่มทำการจับเวลาในการทดสอบ
4. เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป
5. ให้ผู้ทดสอบทำการ สุ่มเลือก เทคนิคสำหรับควบคุมอินเทอร์เนตทีวี จาก 2 วิธีที่เหลือ มา 1 วิธี พร้อมทั้งนำหัวข้อเทคนิคที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
6. จากนั้นให้ผู้ทดสอบ สุ่มเลือก ชุดทดสอบ จาก 2 ชุดทดสอบ มา 1 ชุดทดสอบ พร้อมทั้งนำหัวข้อชุดทดสอบ ที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
7. ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมิน จะเริ่มทำการจับเวลาในการทดสอบ
8. เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป
9. จากนั้นให้ผู้ทดสอบทำการทดสอบ เทคนิคสำหรับควบคุมอินเทอร์เนตทีวีและชุดทดสอบ ที่เหลืออยู่
10. ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมิน จะเริ่มทำการจับเวลาในการทดสอบ
11. เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป
12. นำข้อมูลที่ได้จากทั้ง 3 กลุ่ม มาวิเคราะห์เปรียบเทียบความแตกต่างในเรื่องของ ความเร็วและความถูกต้องของการทำงาน

จำนวนครั้งที่ผิดพลาด (Error) หมายถึง การที่ผู้ทดสอบแสดงออกถึงการสั่งงานกับอินเทอร์เน็ตทีวี ไม่ว่าจะเป็น การพูดการใช้ท่าทาง เพื่อควบคุมอินเทอร์เน็ตทีวี แล้วมีการประมวลผลข้อมูลที่ผิดพลาด หรือไม่มีที่การทำงานของระบบ หรือไม่ได้ผลการบังคับตามที่ผู้ทดสอบสั่งการ

4.1.6 ตารางบันทึกการทดสอบเบื้องต้นในสภาพแวดล้อมที่ไม่มีเสียงรบกวน(46 dB)ในด้านความเร็วและความถูกต้องในการสั่งงานอินเทอร์เน็ตทีวี

ตารางที่ 4.2 แสดงถึงผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยื่น

ผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยื่น						
ผู้ทดสอบ	Voice		Motion Detection		A Combination	
	เวลาวินาที	Error(ครั้ง)	เวลาวินาที	Error(ครั้ง)	เวลาวินาที	Error(ครั้ง)
คนที่ 1	21.44	2	20.15	0	25.8	0
คนที่ 2	19	2	23.7	0	32.6	0
คนที่ 3	17.9	0	18.11	0	30.2	0
คนที่ 4	14.7	1	13.78	0	24.1	0
คนที่ 5	20.3	0	20.11	0	30.2	0
คนที่ 6	12.05	0	20.72	0	27.3	0
คนที่ 7	11.25	0	25.1	1	29.2	0
คนที่ 8	8.7	0	22.89	0	22.3	0
คนที่ 9	57.2	10	19.14	0	35.6	0
คนที่ 10	11.8	0	19.5	0	20.44	0
คนที่ 11	18.7	3	34.41	2	20.7	0
คนที่ 12	9.8	0	30.57	1	35.5	1
คนที่ 13	13.6	1	20.02	0	26.7	0
คนที่ 14	16.8	3	38.14	0	18.9	0
คนที่ 15	11.19	0	20.02	0	20	0
คนที่ 16	15.4	1	25	0	23.7	0
คนที่ 17	17.57	1	22.44	0	19.23	0
คนที่ 18	22.85	2	33.43	1	21.17	0
คนที่ 19	29.03	4	12.85	0	15.08	0
คนที่ 20	13.29	0	15.5	0	22.17	0
คนที่ 21	16.87	1	18.98	0	17.68	0
คนที่ 22	21.16	1	21.94	0	25.23	0
คนที่ 23	41.03	4	15.77	0	29.26	1
คนที่ 24	16.76	0	19.92	0	27.68	1
คนที่ 25	22.49	0	25.94	1	25.01	0
คนที่ 26	25.39	2	29.34	1	16.36	0
คนที่ 27	13.55	2	14.81	0	15.24	0
คนที่ 28	29.16	3	15.38	0	19.16	0
คนที่ 29	11.86	0	22.19	1	21.63	0
คนที่ 30	18.78	0	20.29	0	20.25	0

เนื่องจากการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำขึ้น แบบ Voice Only มีค่าที่เป็น Extreme Value อยู่ คือ Case ที่ 9 จึงไม่เอาค่าดังกล่าวมาคิดคำนวณ

ตารางที่ 4.3 แสดงถึงค่า Extreme จากการทดสอบเบื้องต้นแบบไม่มีเสียงรบกวน (46 dB) ในทำขึ้น

			Case Number	Value
Voice Only	Highest	1	9	57.20
		2	23	41.03
		3	28	29.16
		4	19	29.03
		5	26	25.39
	Lowest	1	8	8.70
		2	12	9.80
		3	15	11.19
		4	7	11.25
		5	10	11.80

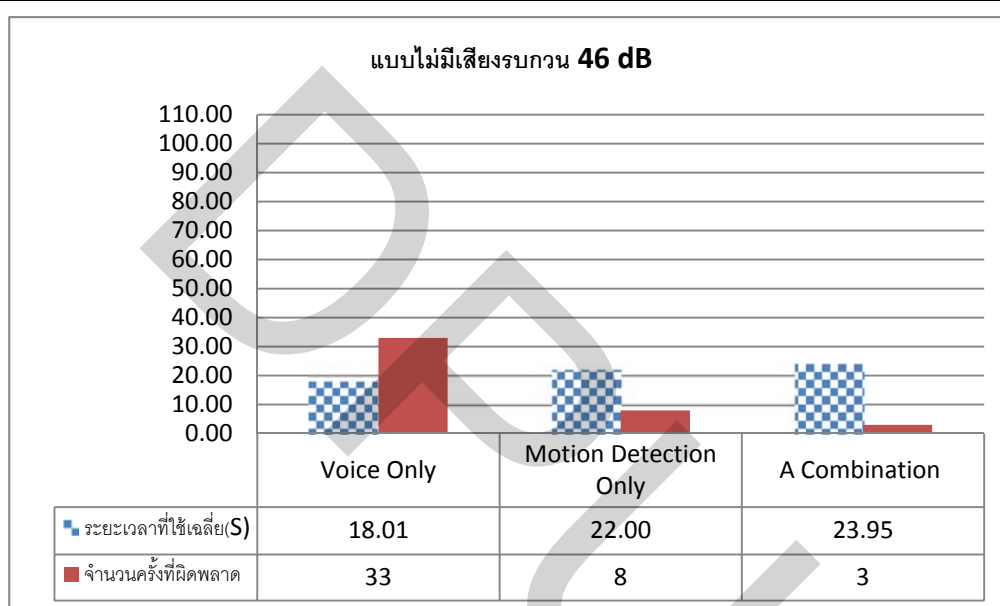
			Case Number	Value
Voice Only	Highest	1	9	10.00
		2	19	4.00
		3	23	4.00
		4	11	3.00
		5	14	3.00 ^a
	Lowest	1	30	0.00
		2	29	0.00
		3	25	0.00
		4	24	0.00
		5	20	.00 ^a



ภาพที่ 4.1 กราฟแสดงค่า Outlier ของข้อมูลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำขึ้น แบบ Voice Only

ตารางที่ 4.4 ตารางสรุปค่าเฉลี่ยเวลาที่ใช้ และจำนวนครั้งที่ผิดพลาดในแต่ละเทคนิคผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยื่น

เทคนิคที่ทดสอบ	ระยะเวลาที่ใช้เฉลี่ยวินาที	จำนวนครั้งที่ผิดพลาด
Voice Only	18.01	33
Motion Detection Only	22.00	8
A Combination	23.95	3



ภาพที่ 4.2 กราฟสรุปผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (46 dB) ในทำยื่น

จากรูปภาพที่ 4.1 กราฟแสดงผลการทดสอบเบื้องต้น กราฟแสดงผลการทดสอบเบื้องต้นแบบไม่มีเสียงรบกวน (46 dB) ในทำยื่นพบว่าเทคนิคการสั่งงานอินเทอร์เน็ตทีวีโดยใช้เสียงในการสั่งงานจะทำได้อย่างรวดเร็วพบว่า ค่าเฉลี่ยของผู้ทดสอบทั้ง 30 คน คำสั่ง 5 คำสั่ง(ตามตารางที่ 4.1) ทำงานได้เสร็จสิ้นเร็วที่สุดคือ 18.02 วินาที แต่ก็มีข้อผิดพลาดในการสั่งการมากที่สุดเช่นกัน คือ 33 ครั้ง จาก 150 คำสั่ง ผู้เข้าทดสอบ ทั้ง 30 คน และเทคนิคผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหวพบว่ามีข้อผิดพลาดในการสั่งงานน้อยที่สุดคือ 3 ครั้ง จาก 150 คำสั่งของผู้เข้าทดสอบ ทั้ง 30 คน และใช้เวลาสั่งการเฉลี่ยประมาณ 23.95 วินาที ต่อ 1 Task

4.1.7 ตารางบันทึกการทดสอบเบื้องต้นในสภาพแวดล้อมที่มีเสียงรบกวน(70 dB)ในด้านความเร็วและความถูกต้องในการสั่งงานอินเทอร์เน็ตทีวี

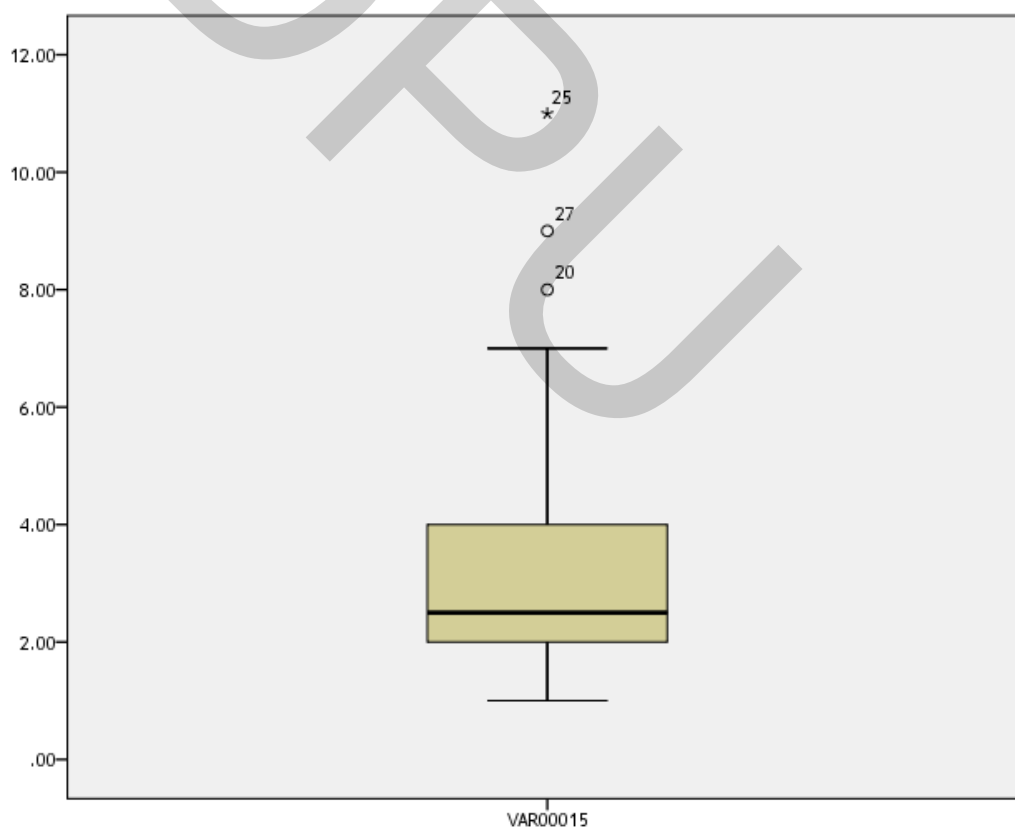
ตารางที่ 4.5 แสดงถึงผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำขึ้น

ผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำขึ้น						
ผู้ทดสอบ	Voice		Motion Detection		A Combination	
	เวลาวินาที	Error(ครั้ง)	เวลาวินาที	Error(ครั้ง)	เวลาวินาที	Error(ครั้ง)
คนที่ 1	30.35	3	20.15	0	38.73	2
คนที่ 2	20.14	2	23.7	0	42.53	3
คนที่ 3	21.03	2	18.11	0	40.2	0
คนที่ 4	47.14	5	13.78	0	38.56	3
คนที่ 5	37.56	4	20.11	0	26.45	2
คนที่ 6	19.55	1	20.72	0	38.13	5
คนที่ 7	21.42	2	25.1	1	35.69	2
คนที่ 8	32.89	2	22.89	0	28.29	2
คนที่ 9	24.9	2	19.14	0	31.82	2
คนที่ 10	19.45	1	19.5	0	40.71	2
คนที่ 11	36.63	5	34.41	2	32.14	3
คนที่ 12	26.61	3	30.57	1	41.75	3
คนที่ 13	21.61	3	20.02	0	28.19	1
คนที่ 14	19.66	2	38.14	0	41.84	3
คนที่ 15	26.63	2	20.02	0	23.68	1
คนที่ 16	21.91	2	25	0	21.89	1
คนที่ 17	52.38	7	22.44	0	29.6	3
คนที่ 18	36.73	6	33.43	1	33.99	1
คนที่ 19	26.8	3	12.85	0	27.75	1
คนที่ 20	42.2	8	15.5	0	33.36	1
คนที่ 21	22.56	3	18.98	0	24.23	0
คนที่ 22	19.58	2	21.94	0	26.54	0
คนที่ 23	32.06	4	15.77	0	45.43	1
คนที่ 24	21.72	2	19.92	0	40.81	3
คนที่ 25	71.05	11	25.94	1	22.26	1
คนที่ 26	16.69	1	29.34	1	25.26	0
คนที่ 27	62.05	9	14.81	0	32.17	2
คนที่ 28	19.81	2	15.38	0	27.94	1
คนที่ 29	20.07	2	22.19	1	28.39	2
คนที่ 30	37.81	4	20.29	0	34.51	2

เนื่องจากการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำขึ้น แบบ Voice Only มีค่าที่เป็น Extreme Value อยู่ คือ Case ที่ 25 จึงไม่เอาค่าดังกล่าวมาคิดคำนวณ

ตารางที่ 4.6 แสดงถึงค่า Extreme จากการทดสอบเบื้องต้นแบบมีเสียงรบกวน (70 dB) ในทำขึ้น

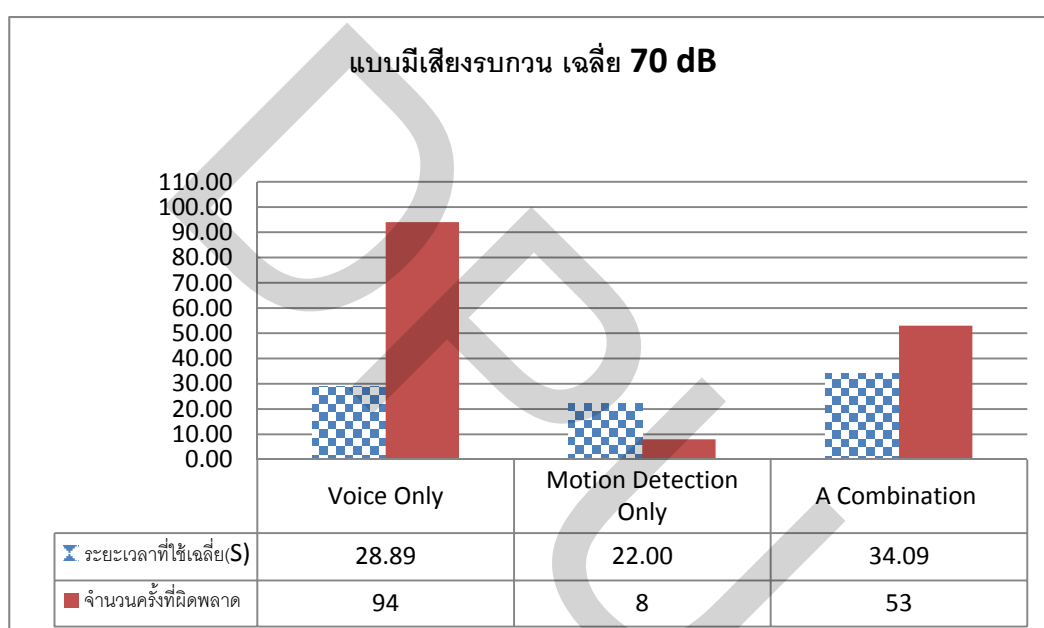
Voice Only Extreme Values 70dB มีเสียงรบกวน (ระยะเวลาที่ใช้)					Voice Only Extreme Values 70dB มีเสียงรบกวน (จำนวนครั้งที่ผิดพลาด)				
			Case Number	Value				Case Number	Value
Voice Only	Highest	1	25	71.05	VAR00015	Highest	1	25	11.00
		2	27	62.05			2	27	9.00
		3	17	52.38			3	20	8.00
		4	4	47.14			4	17	7.00
		5	20	42.20			5	18	6.00
	Lowest	1	26	16.69		Lowest	1	26	1.00
		2	10	19.45			2	10	1.00
		3	6	19.55			3	6	1.00
		4	22	19.58			4	29	2.00
		5	14	19.66			5	28	2.00 ^a



ภาพที่ 4.3 กราฟแสดงค่า Outlier ของข้อมูลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำขึ้น แบบ Voice Only

ตารางที่ 4.7 ตารางสรุปถึงค่าเฉลี่ยเวลาที่ใช้ และจำนวนครั้งที่ผิดพลาดในแต่ละเทคนิคผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำยื่น

เทคนิคที่ทดสอบ	ระยะเวลาที่ใช้เฉลี่ยวินาที	จำนวนครั้งที่ผิดพลาด
Voice Only	28.89	94
Motion Detection Only	22	8.00
A Combination	32.76	53

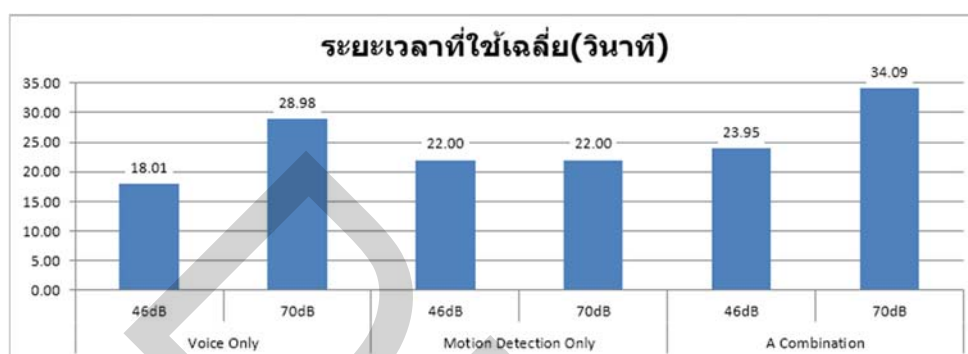


ภาพที่ 4.4 กราฟสรุปผลการทดสอบเบื้องต้น แบบไม่มีเสียงรบกวน (70dB) ในทำยื่น

จากภาพที่ 4.2 กราฟแสดงผลการทดสอบเบื้องต้น กราฟแสดงผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70dB) ในทำยื่นในการทดสอบนี้จะเห็นได้ชัดว่าการใช้เทคนิคการสั่งงานด้วยเสียงในสภาพแวดล้อมที่มีเสียงรบกวนภายในห้องอยู่ที่ 70dB นั้น จะทำให้มีการสั่งงานที่ผิดพลาดในอัตราสูงที่สุด คือ 94 ครั้ง จาก 150 คำสั่ง ผู้เข้าทดสอบ 30 คน และเนื่องจากมีข้อผิดพลาดในการสั่งงานก็จะทำให้ระยะเวลาที่ใช้เฉลี่ยอยู่ที่ 28.89 วินาที ในการสั่งงาน 1Task ส่วนวิธีการควบคุมด้วยท่าทาง จะใช้ค่าทดสอบเดิมเนื่องจากเสียงที่ดังออกจากทีวีจะไม่มีผลกระทบกับการสั่งงานด้วยท่าทาง และเทคนิคการผสมเทคนิคผสมผสานการรู้จำเสียงและการตรวจจับการ

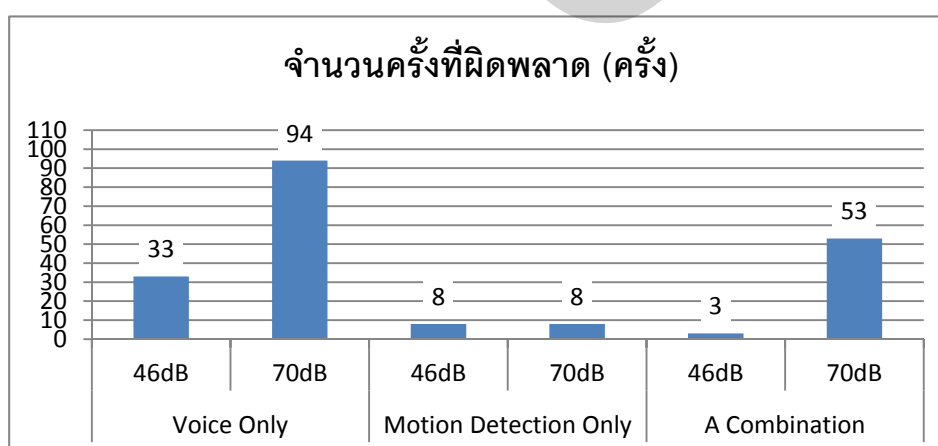
เคลื่อนไหวพบว่ามียัตราการผิดพลาดในการสั่งงานน้อยกว่าการสั่งงานด้วยเสียงอย่างเดียวโดยอยู่ที่ 53 ครั้ง จาก 150 คำสั่ง ผู้เข้าทดสอบ 30 คน และใช้เวลาในการทำงานเสร็จ 1 Task เฉลี่ยอยู่ที่ 53 วินาที

4.1.8 รูปภาพเปรียบเทียบประสิทธิภาพด้านความเร็วในการสั่งงานอินเทอร์เนตทีวี ทั้ง 3 เทคนิคการควบคุม ในสภาพแวดล้อมทั้งมีเสียงรบกวนและไม่มีเสียงรบกวน



ภาพที่ 4.5 กราฟเปรียบเทียบผลการทดสอบเบื้องต้นในประสิทธิภาพด้านความเร็ว แบบมีเสียงรบกวน(70 dB)และไม่มีเสียงรบกวน (46dB) ในทำขึ้น

4.1.9 รูปภาพเปรียบเทียบประสิทธิภาพด้านความถูกต้องในการสั่งงานอินเทอร์เนตทีวี ทั้ง 3 เทคนิคการควบคุม ในสภาพแวดล้อมทั้งมีเสียงรบกวนและไม่มีเสียงรบกวน



ภาพที่ 4.6 กราฟเปรียบเทียบผลการทดสอบเบื้องต้น ประสิทธิภาพด้านความถูกต้องในการสั่งงานแบบมีเสียงรบกวน(70 dB) และไม่มีเสียงรบกวน (46dB) ในทำขึ้น

4.2 ผลการวิจัยการทดสอบครั้งที่ 2 ปรับปรุงโจทย์ทดสอบให้มีค่าสูงมากขึ้นและมีเสียงรบกวนที่ดังออกจากทีวี

ในผลการทดสอบครั้งที่ 2 จะใช้ผู้ทดสอบจำนวน 30 คน นักศึกษาชั้นปี 1 – 3 ช่วงอายุประมาณ 19 – 21 ปี สถาบันเทคโนโลยีไทย-ญี่ปุ่น ซึ่งไม่มีประสบการณ์การใช้ระบบเหล่านี้มาก่อน โดยมีการปรับปรุงโจทย์ทดสอบให้มีความใกล้เคียงกับการใช้งานทีวีในปัจจุบัน และในระดับความดังของเสียงที่ได้ยินออกมาจากอินเทอร์เน็ตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB โดยใช้ห้องประชุมขนาดประมาณ 4x8 เมตร ตั้งกล้องอยู่บนระนาบที่ขนานกับระนาบพื้นที่ทำการตรวจจับสูงจากระนาบพื้น 0.6 เมตร ห่างออกไป 1.5 เมตร มีรายละเอียดดังนี้

4.2.1 เทคนิคที่ทดสอบ

1. Voice Only
2. Motion Detection Only
3. A Combination (Voice and Motion Detection)

4.2.2 Environment

1. สภาพแวดล้อมที่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 70 dB ห้องประชุมขนาดประมาณ 4x8 เมตร

4.2.3 โจทย์สำหรับใช้ทดสอบ (Task)

1. เปิด TV (TV ON)
2. เปลี่ยนช่อง 5 (CH FIVE)
3. เปลี่ยนช่อง 9 (CH NINE)
4. เพิ่มเสียง 1 ระดับ (Volume UP)
5. เปลี่ยนช่อง 3 (CH THREE)
6. เปลี่ยนช่อง 7 (CH SEVEN)
7. ลดเสียง 1 ระดับ (Volume DOWN)
8. เพิ่มเสียง 1 ระดับ (Volume Up)
9. ปิด TV (TV OFF)

4.2.4 วิธีการดำเนินการทดสอบครั้งที่ 2 ในสภาพแวดล้อมที่มีเสียงรบกวนที่ออกมาจากทีวี

1. ไล่ลำดับเทคนิคสำหรับควบคุมอินเทอร์เน็ตทีวี ทั้ง 6 ลำดับ ลำดับละ 5 ชุด ลงบนกองสุ่มเลือกรายละเอียดดังนี้

ตารางที่ 4.8 ตารางแสดงลำดับการควบคุมในแต่ละเทคนิคการควบคุมอินเทอร์เน็ตทีวี

ลำดับการควบคุม	1 : Voice Only	2 : Motion Detection Only	3 : A Combination
A	1	2	3
B	1	3	2
C	2	1	3
D	2	3	1
E	3	1	2
F	3	2	1

1. ผู้เข้าทดสอบจะสุ่มหยิบลำดับการควบคุม 1 ลำดับการควบคุมการกองสุ่มเลือก
 2. ผู้เข้าทดสอบ หยิบ โจทย์ที่ใช้ในการควบคุม (Task)
 3. ผู้เข้าทดสอบ เข้ามายืนในจุด เริ่มการทดสอบ ผู้ควบคุมเวลาจะเริ่มนับจับเวลา
 4. ระหว่างที่ผู้ทดสอบทำโจทย์ทดสอบ ผู้ควบคุมเวลาจะคอยบันทึก ข้อผิดพลาดที่เกิดขึ้นในการควบคุม TV
 5. เมื่อผู้ทดสอบทำ โจทย์ จนครบ ผู้ควบคุมเวลาจะบันทึกเวลาที่ได้ในแต่ละวิธีในการควบคุม TV
 6. ผู้ทดสอบคนต่อไปเข้ามาสุ่มหยิบลำดับแต่ทำกระบวนการนี้ ซ้ำๆ จนครบทั้ง 30 คน
- 4.1.5 ตารางบันทึกการทดสอบครั้งที่ 2 ในสภาพแวดล้อมมีเสียงรบกวน(70dB)ในด้านความเร็วและความถูกต้องในการสั่งงานอินเทอร์เน็ตทีวี
- จำนวนครั้งที่ผิดพลาด (Error) หมายถึง การที่ผู้ทดสอบแสดงออกถึงการสั่งงานกับอินเทอร์เน็ตทีวี ไม่ว่าจะเป็น การพูดการใช้ท่าทาง เพื่อควบคุมอินเทอร์เน็ตทีวี แล้วมีการประมวลผลข้อมูลที่ผิดพลาด หรือไม่มีที่การทำงานของระบบ หรือไม่ได้ผลการบังคับตามที่ผู้ทดสอบสั่งการ

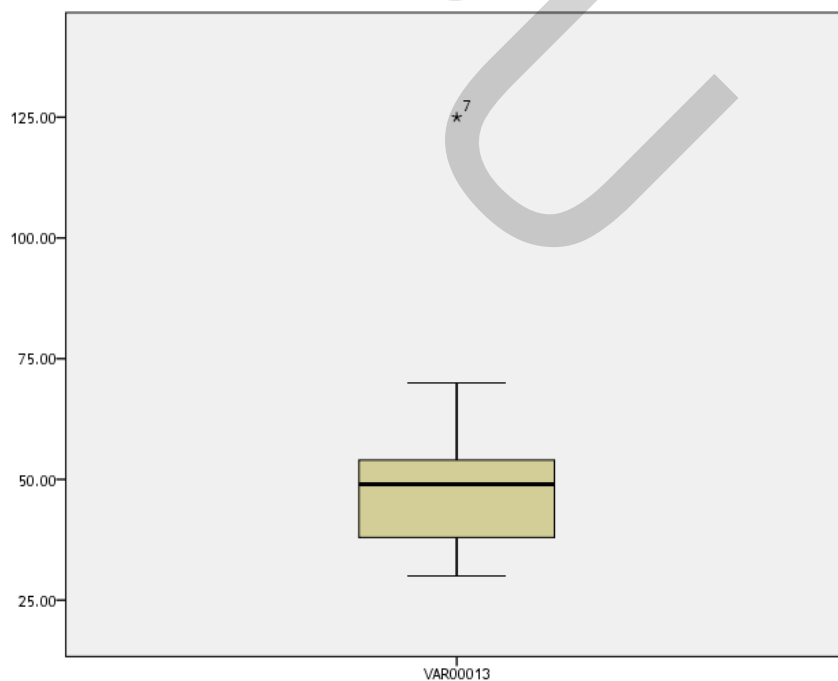
ตารางที่ 4.9 แสดงถึงผลการทดสอบครั้งที่ 2 แบบมีเสียงรบกวน (70 dB) ในทำขึ้น

ผลการทดสอบครั้งที่ 2 แบบมีเสียงรบกวน (70 dB) ในทำขึ้น						
ผู้ทดสอบ	Voice		Motion Detection		A Combination	
	เวลาวินาที	Error(ครั้ง)	เวลาวินาที	Error(ครั้ง)	เวลาวินาที	Error(ครั้ง)
คนที่ 1	87	10	44	1	50	1
คนที่ 2	56	10	44	0	129	5
คนที่ 3	47	5	38	0	47	1
คนที่ 4	94	15	63	1	101	7
คนที่ 5	114	11	50	0	117	5
คนที่ 6	67	6	49	0	131	7
คนที่ 7	93	7	125	1	96	5
คนที่ 8	58	5	70	0	78	7
คนที่ 9	83	10	35	1	53	6
คนที่ 10	57	3	30	0	65	4
คนที่ 11	133	10	40	0	60	1
คนที่ 12	83	11	32	0	67	2
คนที่ 13	49	2	36	1	69	2
คนที่ 14	110	7	32	0	96	4
คนที่ 15	195	10	40	0	80	5
คนที่ 16	58	2	31	1	45	2
คนที่ 17	131	8	35	0	48	0
คนที่ 18	79	4	56	0	84	0
คนที่ 19	3.38	15	54	1	83	0
คนที่ 20	1.11	9	51	0	64	0
คนที่ 21	223	15	49	0	92	3
คนที่ 22	135	8	49	1	70	2
คนที่ 23	175	12	49	0	88	4
คนที่ 24	240	16	57	2	59	5
คนที่ 25	195	13	49	0	55	0
คนที่ 26	240	16	45	1	53	3
คนที่ 27	52	0	50	0	59	2
คนที่ 28	165	9	45	0	75	1
คนที่ 29	199	16	58	1	137	4
คนที่ 30	185	12	54	0	86	3

เนื่องจากการทดสอบครั้งที่ 2 แบบมีเสียงรบกวน (70 dB) ในทำยื่น แบบ Motion Detection มีค่าที่เป็น Extreme Value อยู่ คือ Case ที่ 7 จึงไม่เอาค่าดังกล่าวมาคิดคำนวณ

ตารางที่ 4.10 แสดงค่า Extreme Value จากการทดสอบครั้งที่ 2 แบบมีเสียงรบกวน (70 dB) ในทำยื่น

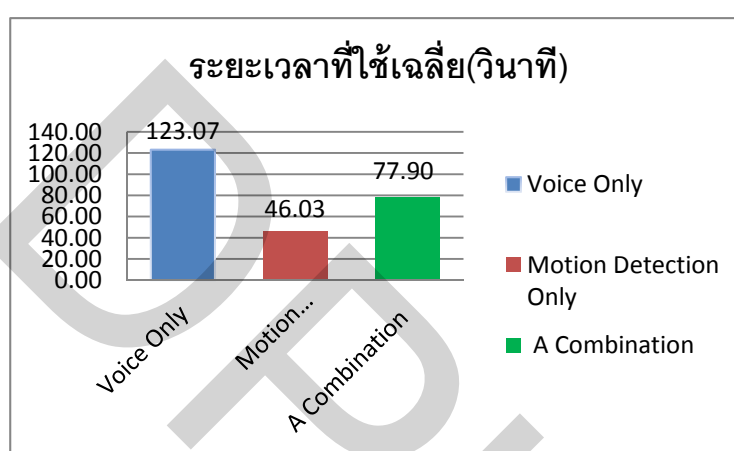
Motion Detection ครั้งที่ 2 Extreme Values 70dB มีเสียงรบกวน (ระยะเวลาที่ใช้)				
			Case Number	Value
Motion Detection	Highest	1	7	125.00
		2	8	70.00
		3	4	63.00
		4	29	58.00
		5	24	57.00
	Lowest	1	10	30.00
		2	16	31.00
		3	14	32.00
		4	12	32.00
		5	17	35.00 ^a



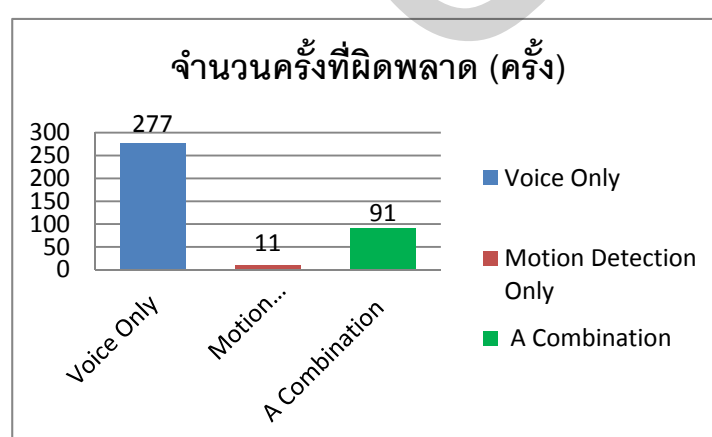
ภาพที่ 4.7 กราฟแสดงค่า Outlier ของข้อมูลทดสอบครั้งที่ 2 แบบมีเสียงรบกวน (70 dB) ในทำยื่น
แบบ Motion Detection

ตารางที่ 4.11 ตารางสรุปถึงค่าเฉลี่ยเวลาที่ใช้ และจำนวนครั้งที่ผิดพลาดในแต่ละเทคนิคผลการทดสอบเบื้องต้น แบบมีเสียงรบกวน (70 dB) ในทำขึ้น

เทคนิคที่ทดสอบ	ระยะเวลาที่ใช้เฉลี่ยวินาที	จำนวนครั้งที่ผิดพลาด
Voice Only	113.58	277
Motion Detection Only	46.03	11
A Combination	77.90	91



ภาพที่ 4.8 กราฟเปรียบเทียบผลการทดสอบครั้งที่ 2 ประสิทธิภาพด้านความเร็ว แบบมีเสียงรบกวน (70 dB) ในทำขึ้น



ภาพที่ 4.9 กราฟเปรียบเทียบผลการทดสอบครั้งที่ 2 ประสิทธิภาพด้านความถูกต้องในการสั่งงานแบบมีเสียงรบกวน (70 dB) ในทำขึ้น

จากรูปภาพที่ 4.5 และ 4.6 กราฟเปรียบเทียบผลการทดสอบครั้งที่ 2 ประสิทธิภาพด้านความเร็ว และความถูกต้อง แบบมีเสียงรบกวน (70 dB) ในทำขึ้นในการทดสอบนี้จะเห็นได้ชัดว่า การใช้เทคนิคการสั่งงานด้วยเสียงในสภาพแวดล้อมที่มีเสียงรบกวนภายในห้องอยู่ที่ 70dB นั้น ใช้เวลาเฉลี่ย ประมาณ 113.58 วินาที ใน 9 คำสั่ง ซึ่งใช้เวลามากที่สุด และมีอัตราการผิดพลาดในการสั่งงานอินเทอร์เน็ตทีวีสูงถึง 277 ครั้ง ใน 270 คำสั่ง จากผู้ทดสอบทั้ง 30 คน ในส่วนของการควบคุมด้วยการตรวจจับการเคลื่อนไหวนั้นมีความเร็วในการสั่งงานอินเทอร์เน็ตทีวีอยู่ที่ 46.03 วินาที ใน 9 คำสั่ง และมีอัตราความผิดพลาด อยู่ที่ 11 ครั้ง ใน 270 คำสั่ง จากผู้ทดสอบทั้ง 30 คน ส่วนวิธีเทคนิคการผสมผสานการรู้จำเสียงและการตรวจจับการเคลื่อนไหว มีความเร็วในการสั่งงานอินเทอร์เน็ตทีวีอยู่ที่ 91 นาที ใน 9 คำสั่ง และอัตราความผิดพลาด อยู่ที่ 91 ครั้ง ใน 270 คำสั่ง

บทที่ 5

สรุปอภิปรายผลการศึกษาและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงข้อสรุปจากการดำเนินงานวิจัยปัญหาและอุปสรรคระหว่างการพัฒนา รวมทั้งข้อเสนอแนะต่างๆ ในการศึกษาต่อไป โดยมีรายละเอียดดังต่อไปนี้

5.1 สรุปผลและวิจารณ์

จากผลการทดลองส่วนวิเคราะห์การใช้วิธีรู้จำเสียงในการสั่งงานอินเทอร์เน็ตทีวี พบว่าสามารถสั่งงานได้รวดเร็วแต่จะมีอัตราความผิดพลาดมากทั้งสองสถานการณ์ ทั้งในสภาพแวดล้อมที่มีเสียงรบกวน (70dB) และไม่มีเสียงรบกวน (46dB) เนื่องจากคำศัพท์ที่ใช้ในการควบคุมมีถึง 10 คำสั่ง ก็จะทำให้ความเป็นไปได้ของเสียงที่จะออกถูกต้องมีน้อยลง อีกทั้งเวลาควบคุมมีเสียงจากทีวีเข้าสู่ตัวรับเสียงก็ยิ่งทำให้เกิดอัตราการผิดพลาดของคำสั่งมากขึ้นไปอีกด้วย อาจแก้ไขด้วยวิธีการย้ายอุปกรณ์รับเสียงให้อยู่ใกล้ตัวผู้สั่งการมากที่สุด เพื่อที่จะรับเสียงได้อย่างถูกต้องชัดเจนและเสียงรบกวนในห้องน้อยที่สุดเพื่อเพิ่มความถูกต้องของการทำงาน

ส่วนวิเคราะห์การตรวจสอบการเคลื่อนไหวในการสั่งงานอินเทอร์เน็ตทีวี นั้นสามารถทำได้ดีในทั้งสองสถานการณ์ทั้งสองเนื่องจากสภาพแวดล้อมที่มีเสียงรบกวนจะไม่มีผลกระทบต่อเทคนิคนี้ และทำงานได้อย่างถูกต้อง แต่จะทำงานได้ค่อนข้างช้า เนื่องจากผู้สั่งการต้องอยู่ในระยะที่กล้องสามารถตรวจจับได้คือระยะตั้งแต่ 1.2 เมตรถึง 3.5 เมตร ซึ่งเป็นระยะที่กล้องก็เนคสามารถมองเห็นได้ แต่ที่ผู้วิจัยทดลองคือระยะ 1.5 เมตร และแสงสว่างภายในห้องต้องเพียงพอต่อการมองเห็นของกล้อง ก็จะทำให้การทำงานถูกต้องมากยิ่งขึ้น

ส่วนวิเคราะห์การผสมผสานระหว่างการใช้รู้จำเสียงและการตรวจสอบการเคลื่อนไหว จะพบว่าในสถานการณ์ที่หนึ่งคือเสียงในห้องอยู่ในระดับ 46 dB การใช้วิธีนี้จะได้ผลการทำงานที่ถูกต้องมากที่สุดเนื่องจากตัวโปรแกรมต้องยืนยันการทำงานถึงสองกระบวนการคือตรวจสอบจากการเคลื่อนไหวของมือเมื่อการเคลื่อนไหวของมือถูกต้องจึงมาตรวจสอบด้วยเสียงอีกครั้งหนึ่ง ข้อดีของวิธีนี้คือจะมีคำที่ใช้ยืนยันแค่คำเดียวคือ OK เพื่อลดความผิดพลาดในเรื่องของคำหรือประโยคที่คล้ายๆกันในการสั่งการอินเทอร์เน็ตทีวี ทำให้การทำงานออกมาค่อนข้างถูกต้องสมบูรณ์ มากที่สุดส่วนในกรณีที่สองคือเสียงภายในห้องเฉลี่ยอยู่ที่ 70 dB การทำงานจะมีความผิดพลาดเนื่องจากมีเสียงรบกวนจากภายนอกทำให้ตัวโปรแกรมไม่สามารถตรวจจับคำได้ แต่อย่างไรก็ตามวิธีนี้ให้ผล

ลัพท์ที่ดีกว่าในด้านของความเร็วและความถูกต้อง กับใช้เสียงสังงานเพียงอย่างเดียว ในสภาพแวดล้อมที่มีเสียงรบกวน

ส่วนวิเคราะห์งานวิจัย เนื่องจากผู้วิจัยได้ดำเนินการวิจัย 2 ครั้ง เพื่อให้ได้ผลการวิจัยที่มีถูกต้องมากที่สุด โดยผู้วิจัยพบว่า ผลการทดลองครั้งที่ 1 และครั้งที่ 2 ให้ผลตอบรับไปในทิศทางเดียวกันกล่าวคือเทคโนโลยีการรู้จำเสียงไม่เหมาะสมสำหรับการควบคุมอุปกรณ์อิเล็กทรอนิกส์ ที่สามารถส่งเสียงออกมาจากตัวมันเองได้ และอุปกรณ์ที่ใช้ในการรับเสียงเข้าไปไม่ควรอยู่ใกล้ อุปกรณ์อิเล็กทรอนิกส์ที่ เพราะอาจจะทำให้การควบคุมผิดพลาดได้อย่างมาก หรือหากจำเป็นต้องใช้เทคโนโลยีดังกล่าว อาจนำเทคนิคที่ผู้วิจัยเสนอคือเทคนิควิธีการผสมผสานเทคนิคการรู้จำเสียง และการตรวจจับการเคลื่อนไหว เข้าไปช่วยลดข้อจำกัดในเรื่องของเทคโนโลยีเสียง

ส่วนวิเคราะห์การนำระบบไปพัฒนาในอนาคตเนื่องจากเทคโนโลยี การรู้จำเสียง, การตรวจสอบการเคลื่อนไหวและเทคนิคการผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว นั้นต่างก็มีข้อดีข้อเสียต่างๆ บางเทคนิคก็เหมาะสมกับสภาพแวดล้อมที่ไม่เหมือนกัน และสภาพผู้ใช้งานแต่ละวัย ดังนั้นหากมีการนำเทคนิคทั้งสามชนิด มาปรับปรุงโดยให้สามารถปรับเปลี่ยนเทคนิคการทำงานต่างสภาพแวดล้อมและผู้ใช้งาน ก็จะทำให้การควบคุมมีประสิทธิภาพมากยิ่งขึ้น

ส่วนวิเคราะห์การสั่งการอินเทอร์เนตทีวี ของผู้เข้าทดสอบทั้งสามเทคนิค เทคนิคเสียง เนื่องจากคำสั่งที่ใช้เป็นคำศัพท์ภาษาอังกฤษ ผู้เข้าทดสอบบางคนอาจออกเสียงได้ไม่ชัดเจน หรือเสียงที่เปล่งออกมาเบาเกินกว่าเสียงสภาพแวดล้อม ทำให้ไมโครโฟนที่ติดตั้งอยู่ในกล่องกินเนสส์ไม่สามารถแยกออกได้ว่าเป็นเสียงคำพูดหรือเสียงรบกวน เทคนิคการตรวจสอบการเคลื่อนไหว เนื่องจากการจากควบคุมเทคนิคนี้ผู้เข้าทดสอบจำเป็นต้องใช้มือแตะค้างไว้ที่ไอคอนที่ต้องการควบคุมเป็นเวลา 1.5 วินาทีที่ผู้เข้าทดสอบบางคนขยับมือไปมาไม่อยู่ในไอคอนในเวลาที่กำหนดไว้ทำให้เกิดความผิดพลาด และเทคนิคการผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว นั้นความผิดพลาดส่วนใหญ่จะเกิดจาก ความสับสนระหว่างการใช้เทคนิคต่างๆ เข้าด้วยกัน โดยบางคนใช้มือแตะที่ไอคอนที่กำหนดโดยที่ยังไม่ได้ออกเสียงคำสั่ง ก็ปล่อยมือออกจากไอคอนที่ควบคุม จึงให้โปรแกรมไม่สามารถทำตามที่ใช้ระบบต้องการได้

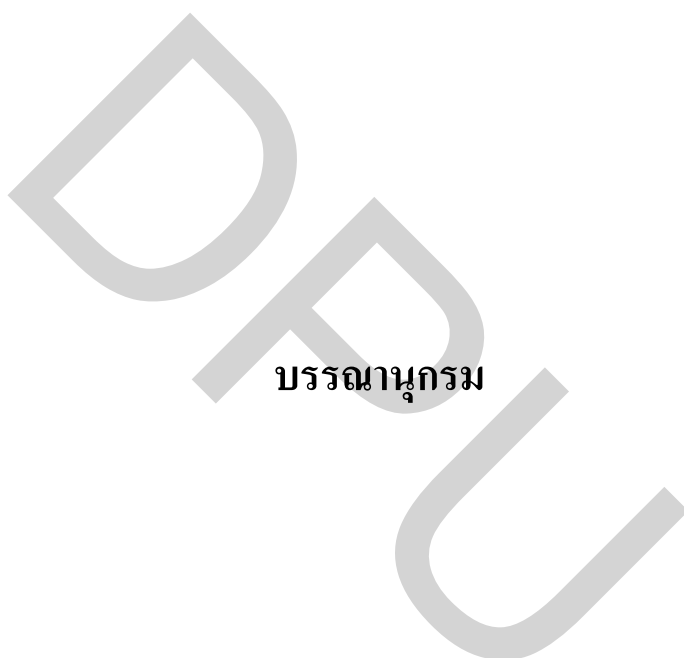
5.2 ปัญหาที่พบในงานวิจัย

5.2.1 เนื่องจากเป็นเทคโนโลยีใหม่ ผู้เข้าทดสอบบางคนไม่เคยใช้เทคโนโลยีเหล่านี้ อาจทำให้เกิดปัญหาในระหว่างการทำทดสอบ ซึ่งสามารถแก้ไขได้โดยอธิบายถึงหลักการทำงานของเทคโนโลยีต่างๆ ก่อนที่ผู้เข้าทดสอบจะเริ่มทำการทดสอบ

5.2.2 ผู้วิจัยได้ทำการพัฒนาระบบ แยกก่อน จากนั้นจึงนำระบบทั้ง 2 คือ การรู้จำเสียง และการตรวจจับการเคลื่อนไหวเข้าด้วยกัน ต้องอาศัยการศึกษาข้อมูลอย่างมาก เนื่องจากในปัจจุบันยังไม่ค่อยมีการทำระบบทั้งสองเข้ามารวมกันทำงาน เพื่อผลลัพธ์ใดผลลัพธ์หนึ่ง

5.3 ข้อเสนอแนะในการศึกษาขั้นต่อไป

งานวิจัยชิ้นนี้ได้นำเสนอเทคนิคสำหรับการควบคุมอินเทอร์เน็ททีวีโดยทีเเนคด้วยการผสมผสานระหว่างการใช้รู้จำเสียงและการตรวจสอบการเคลื่อนไหวจากผลของวิธีการที่นำเสนอแสดงให้เห็นว่า ผลลัพธ์ที่ได้ในเรื่องของประสิทธิภาพของความถูกต้องในการควบคุมอินเทอร์เน็ททีวีมีความถูกต้องมากที่สุดในสถานการณ์ที่มีเสียงรบกวนน้อย คืออยู่ที่ 46 dB เนื่องจากการตรวจสอบถึงสองขั้นตอนวิธี ทั้งการใช้รู้จำเสียงและการตรวจสอบการเคลื่อนไหว แต่ก็มีข้อเสียในเรื่องของความเร็ว อาจเป็นเพราะผู้ทดสอบเพิ่งเคยได้ใช้เทคโนโลยีที่ไม่เคยได้ถูกฝึกมาก่อนแต่เพื่อประโยชน์อย่างแท้จริงในอนาคตควรที่จะมีการทดลองในรูปแบบต่างๆมากขึ้นเช่น กลุ่มทดลองกลุ่มที่หนึ่งที่ได้รับการฝึกฝนการใช้กับอีกกลุ่มที่ไม่ได้รับการฝึกฝนการใช้งานเพื่อวัดผลทางด้านประสิทธิภาพและประสิทธิผลของแต่ละวิธี



บรรณานุกรม

บรรณานุกรม

ภาษาไทย

นราวุฒิ พัฒโนทัย. พรชัย มงคลนาม. บัณฑิต วรรณภา. 2550 การประกอบรวมโครงร่างมนุษย์จาก
การ ตรวจจับการเคลื่อนไหว โดยใช้กล้อง *Kinect* หลายตัว. National Conference on
Information Technology 2012. 370-374

Microsoft Corporation, *Kinect for Xbox 360 - Xbox.com*. สืบค้นเมื่อสิงหาคม 2555, จาก
<http://www.xbox.com/en-GB/kinect>

Microsoft Corporation, *Speech recognition – Sapi5.1*. สืบค้นเมื่อสิงหาคม 2555, จาก
<http://www.microsoft.com/en-us/download/details.asp?id=10121>

Roman and Boomerang, รีวิว *Samsung Smart TV 55ES8000*. สืบค้นเมื่อสิงหาคม 2555, จาก
http://www.lcdtvthailand.com/review/detail.asp?desc=1¶m_id=1277

Wikipedia, *Motion capture*. สืบค้นเมื่อสิงหาคม 2555, จาก
http://en.wikipedia.org/wiki/Motion_capture

Microsoft Corporation, *Kinect for Windows SDK from Microsoft Research*. สืบค้นเมื่อสิงหาคม
2555, จาก
<http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/>

Microsoft Corporation, *Human Interface Guidelines_v1.5.0*. สืบค้นเมื่อสิงหาคม 2555, จาก
http://www.download.microsoft.com/.../Human_Interface

ภาษาต่างประเทศ

BOOKS

- J D. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, and M. B. Moussa, *Evaluating a Dancer's Performance using Kinect-based Skeleton Tracking* : Proceedings of the 19th ACM international conference on Multimedia. pp. 659-662:Nov. 2011.
- Velardo, C. and Dugelay, *Real Time Extraction of Body Soft Biometric from 3D Videos* : Proceedings of the 19th ACM international conference on Multimedia, 28 November-1 December 2011, Arizona, USA, pp. 781-782.
- K. Berger, K. Ruhl, Y. Schroeder, C. Bruemmer, Al. Scholz, and M. Magnor, *Markerless Motion Capture using multiple Color-Depth Sensors* : Proc. Vision, Modeling and Visualization (VMV) 2011, pp. 317-324, Oct. 2011.
- J. Vince, *Mathematics for Computer Graphics, 3rd ed.* London:Springer, 2010.
- B. Siciliano and O. Khatib, *Springer Handbook of Robotics*, Springer, 2008.

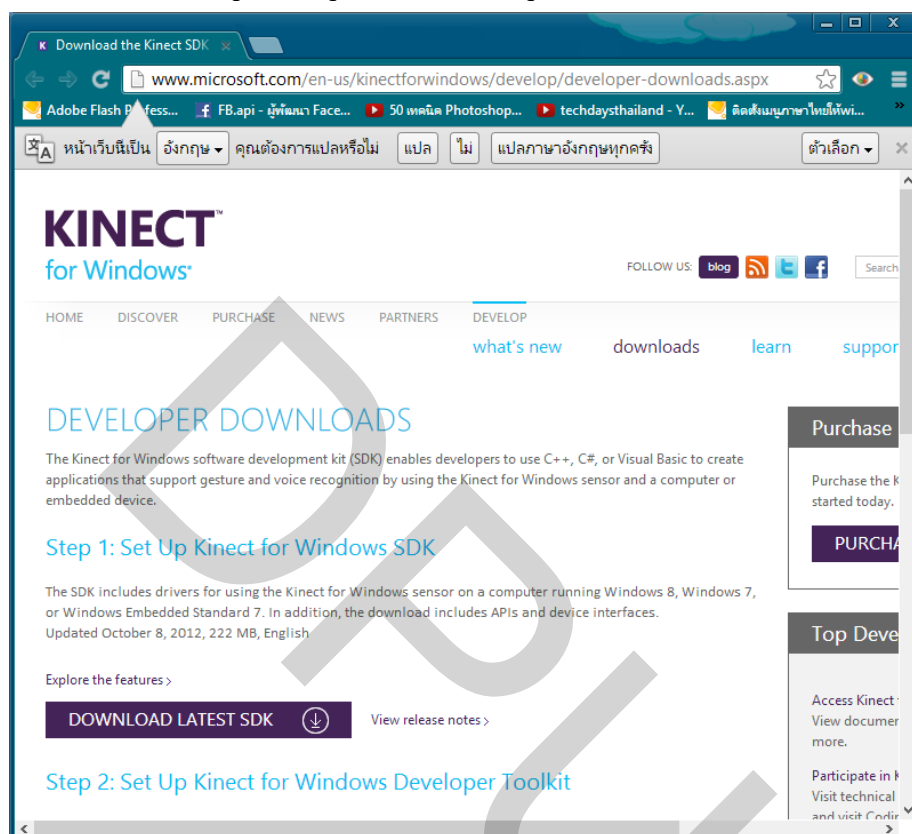
DRU

ภาคผนวก

ภาคผนวก ก
วิธีติดตั้ง Kinect for Windows SDK

คู่มือการติดตั้ง Kinect for Windows sdk

1. ก่อนอื่นเข้าไป download ที่ลิงค์ <http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx>



ภาพที่ 1 แสดงหน้าเว็บ Download

เลื่อนหา ข้อความดังภาพ แล้วคลิกเพื่อดาวน์โหลด แล้วจะได้ไฟล์ติดตั้ง .exe มา

Related Downloads

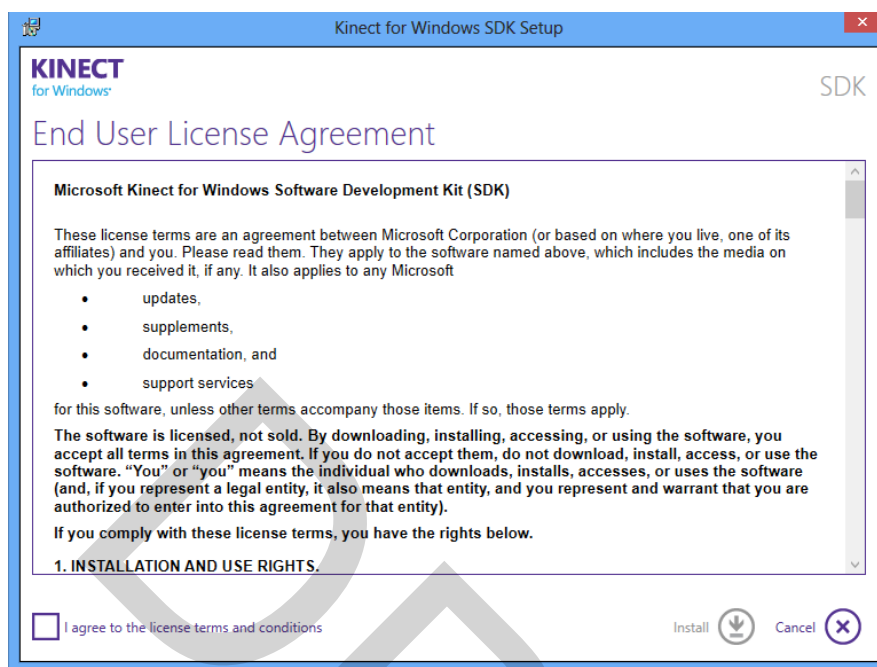
- Kinect for Windows Runtime v1.6
- .NET Framework 4.5
- .NET Framework 4.0
- XNA Framework Redistributable 4.0

Previous Versions

- Kinect for Windows SDK 1.5
- Kinect for Windows SDK 1.0
- Kinect for Windows SDK beta
- Kinect for Windows Developer Toolkit v1.5.2
- Kinect for Windows Runtime v1.5

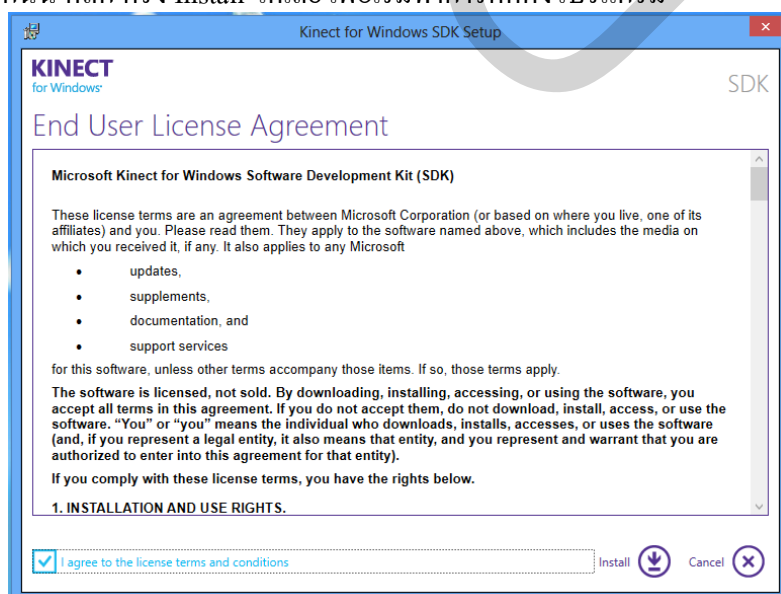
ภาพที่ 2 แสดงหน้า URL Download

2. เปิดไฟล์ kinectSDK-1.5.exe ที่ดาวน์โหลดมาไว้บนเครื่อง ก็จะได้นหน้าต่างดังภาพ



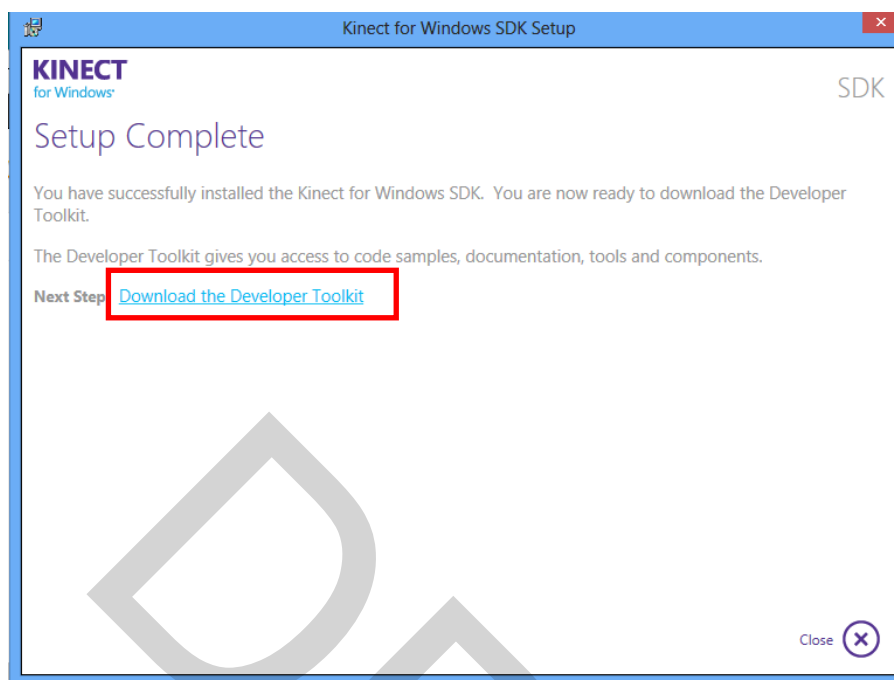
ภาพที่ 3 แสดงขั้นตอนการติดตั้ง

- เลือกเครื่องหมายถูก ในช่อง ตามภาพ
- จากนั้น คลิก ตรง Install ได้เลย เพื่อเริ่มทำการติดตั้งโปรแกรม



ภาพที่ 4 แสดงขั้นตอนการติดตั้ง

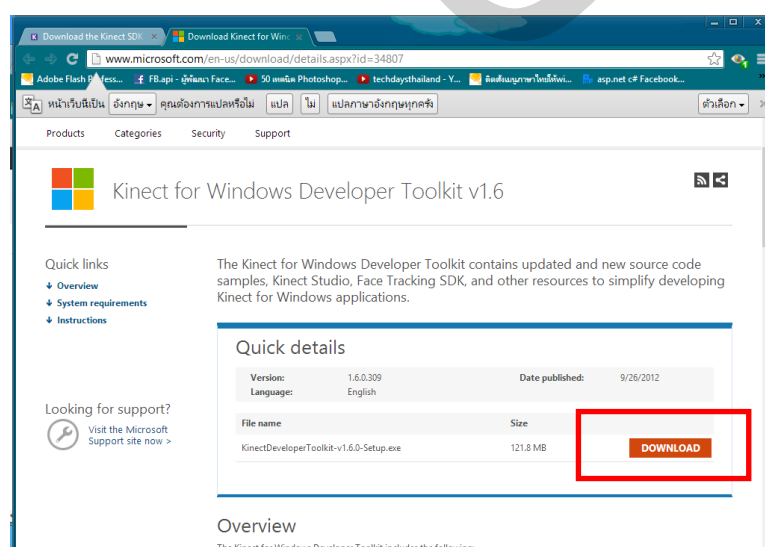
3. เมื่อเลือกที่ Install แล้ว ก็จะแสดงหน้าต่างดังภาพ



ภาพที่ 5 แสดงขั้นตอนการติดตั้ง

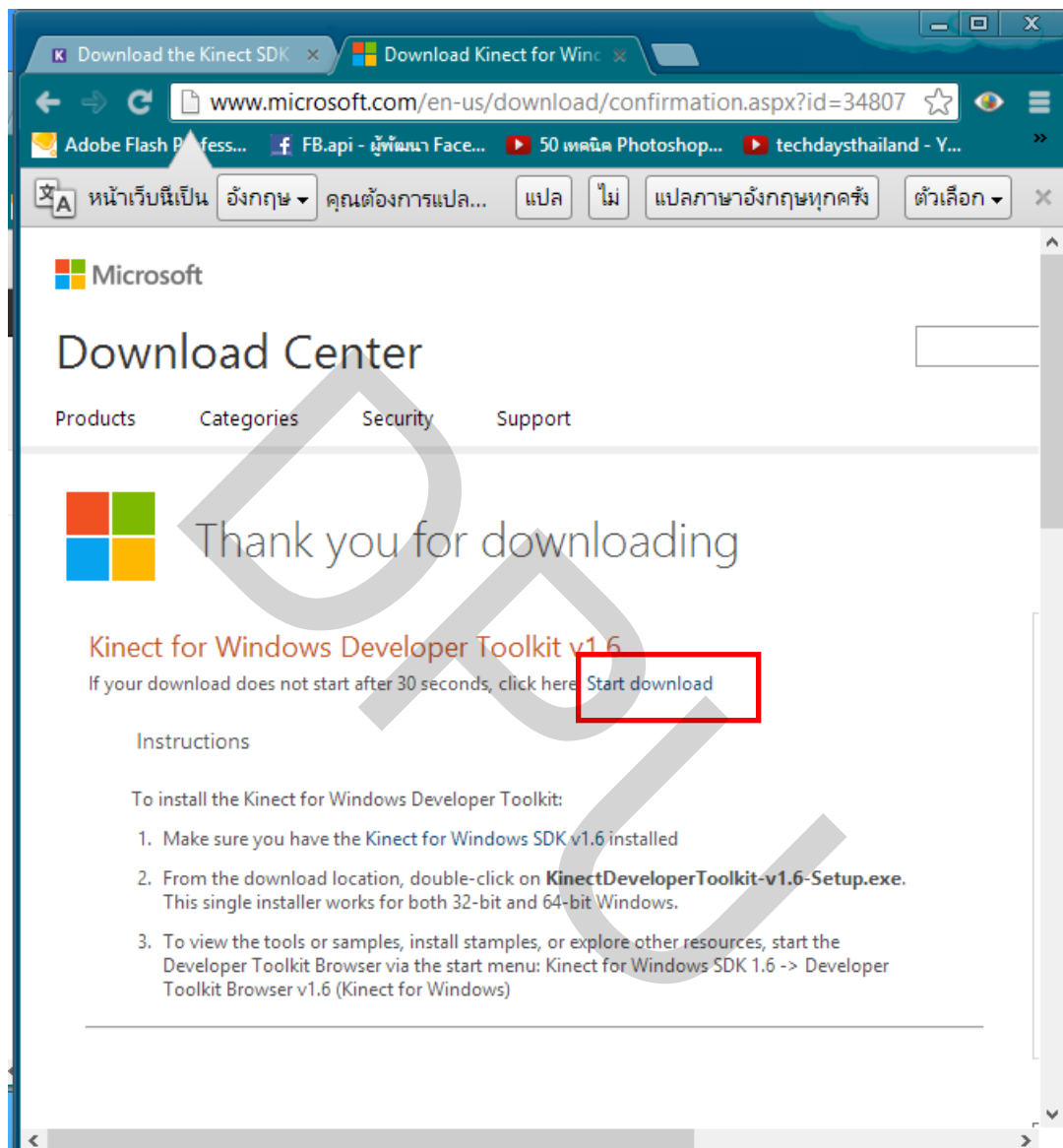
- เมื่อได้หน้าต่างตามนี้แล้วให้ คลิก Download the Developer Toolkit ตามภาพ เพื่อที่จะทำการ Download Kinect for Windows Developer Toolkit

4 เมื่อทำตามข้อ 3 แล้ว ก็จะมีหน้าต่างบน web browser ขึ้นมาเพื่อจะทำการ Download ต่อไป



ภาพที่ 6 แสดงขั้นตอนการติดตั้ง

5 คลิกที่ Start download เพื่อทำการ Download เป็นขั้นตอนสุดท้ายของการ Download Kinect for Windows Developer Toolkit



ภาพที่ 7 แสดงขั้นตอนการติดตั้ง

download เสร็จแล้วก็จะได้ file ติดตั้ง .exe

6 เมื่อ Download file (.exe) มาเก็บไว้ที่เครื่องแล้วก็ทำการเปิดไฟล์นั้นขึ้นมาเลย เพื่อที่จะทำการติดตั้ง Kinect for Windows Developer Toolkit เมื่อเปิดขึ้นมาแล้วก็จะได้เห็นหน้าต่างในภาพ



ภาพที่ 8 แสดงขั้นตอนการติดตั้ง

- เลือก เครื่องหมายถูก ในช่อง ตามภาพ
- คลิกที่ Install เพื่อเริ่มทำการติดตั้ง

ภาคผนวก ข
ตารางบันทึกผลการทดลองครั้งที่ 1

การทดลองครั้งที่ 1 จำนวน 30 คน ไม่มีเสียงรบกวน

ตารางที่ 1 ตารางบันทึกผลการทดลองกลุ่มที่ 1 Voice Only (ไม่มีเสียงรบกวน)

ลำดับ ที่	ชุด ทดสอบ ที่	ระยะเวลา ที่ใช้ ทดลอง(s)	จำนวน ครั้งที่ ผิดพลาด	รายละเอียดของข้อผิดพลาด
1	C	21.44	2	ออกเสียง CH 11, Tv Off แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
2	B	19	2	ออกเสียง Volume Up แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
3	C	17.9	0	
4	C	14.7	1	ออกเสียง CH 9 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
5	A	20.3	0	-
6	B	12.05	0	-
7	A	11.25	0	-
8	B	8.7	0	-
9	A	57.2	10	ออกเสียง CH 7, CH 3, Tv Off แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
10	B	11.8	0	-
11	C	18.7	3	ออกเสียง CH 9, CH 11 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
12	B	9.8	0	-
13	C	13.6	1	ออกเสียง CH 9 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
14	A	16.8	3	ออกเสียง CH 3, CH 7 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
15	B	11.19	0	-

16	A	15.4	1	ออกเสียง Ch 7 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
17	C	17.57	1	ออกเสียง CH 9 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
18	A	22.85	2	ออกเสียง CH 3 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
19	C	29.03	4	ออกเสียง Ch9, Volum Down, Tv Off แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
20	C	13.29	0	-
21	A	16.87	1	ออกเสียง Volum Down แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
22	B	21.16	1	ออกเสียง CH 11 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
23	C	41.03	4	ออกเสียง CH 11, Tv Off แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
24	A	16.76	0	-
25	B	22.49	0	-
26	C	25.39	2	ออกเสียง Volume Up, Tv Off แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
27	A	13.55	2	ออกเสียง Ch7 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
28	C	29.16	3	ออกเสียง CH 11, Volume Up แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
29	A	11.86	0	-
30	A	18.78	0	-

ตารางที่ 2 ตารางบันทึกการทดลองกลุ่มที่ 2 Motion Detection Only (ไม่มีเสียงรบกวน)

ลำดับ ที่	ชุด ทดสอบ ที่	ระยะเวลา ที่ใช้ ทดลอง(s)	จำนวน ครั้งที่ ผิดพลาด	รายละเอียดของข้อผิดพลาด
1	A	20.15	0	-
2	A	23.70	0	-
3	B	18.11	0	-
4	A	13.78	0	-
5	B	20.11	0	-
6	A	20.72	0	-
7	B	25.10	1	เลื่อนไป Detection Volume Up แล้วแต่ตัว Pointer อยู่ไม่ครบเวลา
8	A	22.89	0	-
9	C	19.14	0	-
10	A	19.50	0	-
11	B	34.41	2	เลื่อนไป Detection Volume Down แล้วแต่ตัว Pointer อยู่ไม่ครบเวลา
12	A	30.57	1	เลื่อนไป Detection CH 3 แล้วแต่ตัว Pointer อยู่ไม่ ครบเวลา
13	B	20.02	0	-
14	C	38.14	0	-
15	A	20.02	0	-
16	B	25	0	-
17	A	22.44	0	-
18	B	33.43	1	เลื่อนไป Detection CH 7 แล้วแต่ตัว Pointer อยู่ไม่

				ครบเวลา
19	A	12.85	0	-
20	A	15.50	0	-
21	B	18.98	0	-
22	A	21.94	0	-
23	A	15.77	0	-
24	B	19.92	0	-
25	B	25.94	1	เลื่อนไป Detection CH 7แล้วแต่ตัว Pointer อยู่ไม่ ครบเวลา
26	B	29.34	1	เลื่อนไป Detection CH 7แล้วแต่ตัว Pointer อยู่ไม่ ครบเวลา
27	B	14.81	0	-
28	B	15.38	0	-
29	C	22.19	1	เลื่อนไป Detection CH 9แล้วแต่ตัว Pointer อยู่ไม่ ครบเวลา
30	B	20.29	0	-

ตารางที่ 3 ตารางบันทึกการทดลองกลุ่มที่ 3 A Combination (ไม่มีเสียงรบกวน)

ลำดับ ที่	ชุด ทดสอบ ที่	ระยะเวลา ที่ใช้ ทดลอง(s)	จำนวน ครั้งที่ ผิดพลาด	รายละเอียดของข้อผิดพลาด
1	B	25.8	0	-
2	C	32.6	0	-
3	A	30.2	0	-
4	B	24.1	0	-
5	C	30.2	0	-
6	C	27.3	0	-
7	C	29.2	0	-
8	C	22.3	0	-
9	B	35.6	0	-
10	C	20.44	0	-
11	A	20.7	0	-
12	C	35.5	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
13	A	26.7	0	-
14	B	18.9	0	-
15	C	20	0	-
16	C	23.7	0	-
17	B	19.23	0	-
18	C	21.17	0	-
19	B	15.08	0	-

20	B	22.17	0	
21	C	17.68	0	-
22	B	25.23	0	-
23	B	29.26	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
24	C	27.68	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
25	C	25.01	0	-
26	A	16.36	0	-
27	C	15.24	0	-
28	A	19.16	0	-
29	B	21.63	0	-
30	C	20.25	0	-

การทดลองครั้งที่ 1 จำนวน 30 คน มีเสียงรบกวน

ตารางที่ 4 ตารางบันทึกผลการทดลองกลุ่มที่ 1 Voice Only (มีเสียงรบกวน)

ลำดับ ที่	ชุด ทดสอบ ที่	ระยะเวลา ที่ใช้ ทดลอง(s)	จำนวน ครั้งที่ ผิดพลาด	รายละเอียดของข้อผิดพลาด
1	A	30.35	3	ออกเสียง Ch3, Volume Down, Volume Up แล้วไม่มีการตอบสนองหรือตอบสนองผิด
2	C	20.14	2	ออกเสียง Ch9 แล้วไม่มีการตอบสนองหรือตอบสนองผิด
3	B	19.53	1	ออกเสียง Ch5 แล้วไม่มีการตอบสนองหรือตอบสนองผิด
4	C	47.14	5	ออกเสียง Ch9, Ch11, Volume Up, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
5	B	17.56	1	ออกเสียง Volume Down แล้วไม่มีการตอบสนองหรือตอบสนองผิด
6	C	18.55	1	ออกเสียง Ch 11 แล้วไม่มีการตอบสนองหรือตอบสนอง
7	B	18.53	1	ออกเสียง Ch 5 แล้วไม่มีการตอบสนองหรือตอบสนองผิด
8	C	32.89	2	ออกเสียง Volume Down แล้วไม่มีการตอบสนองหรือตอบสนองผิด
9	A	24.90	2	ออกเสียง Ch7 แล้วไม่มีการตอบสนองหรือตอบสนองผิด
10	B	19.45	1	ออกเสียง Ch 9 แล้วไม่มีการตอบสนองหรือตอบสนองผิด
11	A	36.63	5	ออกเสียง Ch3, Ch7, Volume Up แล้วไม่มีการตอบสนองหรือตอบสนองผิด
12	C	26.61	3	ออกเสียง Ch11, Volume Donw แล้วไม่มีการตอบสนองหรือตอบสนองผิด

13	B	21.61	3	ออกเสียง Ch 5, Volume Up แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
14	C	19.66	1	ออกเสียง Ch9 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
15	A	26.63	2	ออกเสียง Ch3 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
16	B	21.91	1	ออกเสียง Volume Down แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
17	C	52.38	7	ออกเสียง Ch9, Volume Down, Ch11, Volume Up แล้วไม่มีการตอบสนองหรือตอบสนองผิด
18	A	36.73	6	ออกเสียง Ch3, Ch7, Volume Down แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
19	B	26.80	3	ออกเสียง Ch5, Volume Up แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
20	A	42.20	8	ออกเสียง Ch3, Ch7, Volume Down, Volume Up, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
21	C	22.56	1	ออกเสียง Volume Down แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
22	B	19.58	1	ออกเสียง Ch9, Volume Down Up แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
23	C	32.06	4	ออกเสียง Ch9, Volume Up แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
24	A	21.72	2	ออกเสียง Ch7 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
25	C	71.05	11	ออกเสียง Ch9, Ch11, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด

26	B	16.69	1	ออกเสียง Volume Up Up แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
27	A	62.25	9	ออกเสียง Ch3, Ch7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
28	C	19.81	1	ออกเสียง Ch11 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
29	B	20.07	1	ออกเสียง Ch5 แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
30	A	37.81	4	ออกเสียง Ch7 Volume Up, Volume Down

ตารางที่ 5 ตารางบันทึกผลการทดลองกลุ่มที่ 3 A Combination (มีเสียงรบกวน)

ลำดับ ที่	ชุด ทดสอบ ที่	ระยะเวลา ที่ใช้ ทดลอง(s)	จำนวน ครั้งที่ ผิดพลาด	รายละเอียดของข้อผิดพลาด
1	B	38.73	5	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
2	A	56.21	8	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
3	C	40.20	0	-
4	A	38.56	3	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
5	C	26.45	2	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
6	B	38.13	5	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
7	C	35.69	3	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
8	A	28.29	2	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
9	C	31.82	2	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
10	B	40.71	3	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
11	A	58.45	5	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
12	B	41.75	4	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
13	C	28.19	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
14	B	41.84	3	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
15	A	23.68	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง

16	C	21.89	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
17	A	29.60	3	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
18	B	33.99	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
19	C	27.75	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
20	A	33.36	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
21	C	24.23	0	-
22	B	26.54	0	-
23	A	45.43	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
24	B	40.81	6	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
25	A	22.26	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
26	C	25.26	0	-
27	B	32.17	2	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
28	C	27.94	1	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
29	A	28.39	2	ออกเสียง OK แล้ว Pointer หลุดการแถบเปลี่ยนช่อง
30	B	34.51	4	ออกเสียง OK แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด

ภาคผนวก ค

ตารางบันทึกผลการทดลองครั้งที่ 2

การทดลองครั้งที่ 2 จำนวน 30 คน มีเสียงรบกวน

ตารางที่ 6 ตารางบันทึกผลการทดลองครั้งที่ 2 มีเสียงรบกวน (70 dB)

ลำดับ ที่	ลำดับ การ ควบคุม	เทคนิคที่ใช้ ควบคุม	ระยะเวลาที่ ใช้ทดลอง (นาที)	จำนวน ครั้งที่ ผิดพลาด	รายละเอียดของข้อผิดพลาด
1	D	Voice	1.27	10	ออกเสียง Ch3, Ch7 แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
		Motion	0.44	2	เลื่อนไป Ch3, Ch5 แล้วแต่ตัว Pointer อยู่ไม่ครบเวลา
		Combination	0.50	1	ออกเสียง OK แล้ว Pointer หลุดการ แถบเปลี่ยนช่อง
2	C	Voice	0.56	10	ออกเสียง Ch3, Vol up, Ch 7, TV Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
		Motion	0.44	-	-
		Combination	1.09	5	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
3	B	Combination	0.47	1	ออกเสียง OK แล้ว Pointer หลุดการ แถบเปลี่ยนช่อง
		Motion	0.38	-	-
		Voice	0.47	5	ออกเสียง TV ON, Ch5, Vol Up แล้ว ไม่มีการตอบสนองหรือตอบสนองผิด
4	A	Motion	1.03	-	-
		Combination	1.41	7	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Voice	1.34	15	ออกเสียง TV ON, Vol Up, Ch3, Tv On, Tv Off แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
5	E	Voice	1.53	11	ออกเสียง TV ON, Ch3, Ch5 แล้วไม่มี

					การตอบสนองหรือตอบสนองผิด
		Combination	1.57	5	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Motion	0.50	-	-
6	F	Motion	0.49	-	-
		Voice	1.07	6	ออกเสียง Ch3, Vol Down แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Combination	2.11	7	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
7	B	Combination	1.36	5	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Motion	2.05	-	-
		Voice	1.33	7	ออกเสียง Vol Up, Ch3, Vol Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
8	E	Voice	0.58	5	ออกเสียง Ch3, Ch9, TVOn แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Combination	1.18	7	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Motion	1.10	-	-
9	D	Combination	0.53	6	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Voice	1.23	10	ออกเสียง Ch5, Ch9, Ch3, Ch 11, Volume Up, TV On, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Motion	0.35	-	-

10	F	Motion	0.30	-	-
		Voice	0.57	3	ออกเสียง Volume Down, Volume Up แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
		Combination	1.05	4	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
11	E	Voice	2.13	10	ออกเสียง Tv On, Ch 5, Ch9, Ch7 Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
		Combination	1.00	1	ออกเสียง OK แล้ว Pointer หลุดการ แถบเปลี่ยนช่อง
		Motion	0.40	-	-
12	E	Voice	1.23	11	ออกเสียง Tv On, Ch 5, Ch9, Ch7, Ch3 Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
		Combination	1.07	2	ออกเสียง OK แล้ว Pointer หลุดการ แถบเปลี่ยนช่อง
		Motion	0.32	-	-
13	D	Combination	1.09	2	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Voice	0.49	2	ออกเสียง Ch3 แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Motion	0.36	-	-
14	D	Combination	1.36	4	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Voice	1.50	7	ออกเสียง Tv On, Ch5, Ch9, Ch7, Volume Down แล้วไม่มีการ

					ตอบสนองหรือตอบสนองผิด
		Motion	0.32	-	-
15	B	Combination	1.20	5	Ch5, Ch9, Ch7, Volume Up
		Motion	0.40	-	-
		Voice	3.15	10	ออกเสียง Tv On, Ch5, Ch3, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
16	E	Voice	0.58	2	ออกเสียง Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Combination	0.45	2	ออกเสียง OK แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Motion	0.31	-	-
17	F	Motion	0.35	-	-
		Voice	2.11	8	ออกเสียง Tv On, Ch3, Ch7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Combination	0.48	-	-
18	C	Voice	1.19	4	ออกเสียง Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Motion	0.56	-	-
		Combination	1.24	-	-
19	F	Motion	0.54	-	-
		Voice	3.38	15	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Combination	1.23	-	-
20	A	Motion	0.51	-	-
		Combination	1.04	-	-

		Voice	1.11	9	ออกเสียง Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
21	D	Combination	1.32	3	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Voice	3.43	15	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
		Motion	0.49	-	-
22	A	Motion	0.49	-	-
		Combination	1.10	2	ออกเสียง OK แล้ว Pointer หลุดการ แถบเปลี่ยนช่อง
		Voice	2.15	8	ออกเสียง Ch9, Ch3, Ch 7, Volume Up, Volume Down แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
23	F	Motion	0.49	-	-
		Voice	2.55	12	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
		Combination	1.28	4	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
24	A	Motion	0.57	-	-
		Combination	0.59	5	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Voice	4.00	16	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv

					Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
25	C	Voice	3.15	13	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
		Motion	0.49	-	-
		Combination	0.55	-	-
26	B	Combination	0.53	3	ออกเสียง OK แล้วไม่มีการตอบสนอง หรือตอบสนองผิด
		Motion	0.45	-	-
		Voice	4.00	16	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือ ตอบสนองผิด
27	A	Motion	0.50	-	-
		Combination	0.59	2	ออกเสียง OK แล้ว Pointer หลุดการ แถบเปลี่ยนช่อง
		Voice	0.52	-	-
28	C	Voice	2.45	9	ออกเสียง Volume Up, Volume Down, Tv Off แล้วไม่มีการ ตอบสนองหรือตอบสนองผิด
		Motion	0.45	-	-
		Combination	1.15	1	ออกเสียง OK แล้ว Pointer หลุดการ แถบเปลี่ยนช่อง
29	B	Combination	2.17	4	Tv On, Volume Down, Tv Off
		Motion	0.58	-	-
		Voice	3.19	16	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch

					7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
30	C	Voice	3.05	12	ออกเสียง Tv On, Ch5, Ch9, Ch3, Ch 7, Volume Up, Volume Down, Tv Off แล้วไม่มีการตอบสนองหรือตอบสนองผิด
		Motion	0.54	-	-
		Combination	1.26	3	ออกเสียง OK แล้วไม่มีการตอบสนองหรือตอบสนองผิด

ภาคผนวก ง

ผลงานวิจัยในการประชุมวิชาการ การบริหารและการจัดการครั้งที่ 8
“การบริหารจัดการแห่งโลกพลวัต” โดยมหาวิทยาลัยธุรกิจบัณฑิตย์
ร่วมกับเครือข่ายวิจัยประชาชน

เทคนิคสำหรับการควบคุมอินเทอร์เน็ตทีวีโดยคิเนค ด้วยการผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว

A Technique for Internet Television Control using Kinect Combination of Voice Recognition and Motion Detection Approach

ภูวดล ศิริกองธรรม¹ และ วรสิทธิ์ ชูชัยวัฒนา²

บทคัดย่อ

การรู้จำเสียงหรือการตรวจสอบการเคลื่อนไหว ได้ถูกนำมาเข้ามาช่วยทำให้การควบคุมอินเทอร์เน็ตทีวี มีประสิทธิภาพและสะดวกสบายมากขึ้น แต่เทคโนโลยีดังกล่าวที่กล่าวมานั้น ต่างมีจุดเด่นและข้อด้อยในตัว ผู้วิจัยจึงเสนอเทคนิคที่ใช้ควบคุมอินเทอร์เน็ตทีวีแบบใหม่ด้วยการผสมผสานจุดเด่นระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว เพื่อให้การควบคุมอินเทอร์เน็ตทีวีมีประสิทธิภาพมากยิ่งขึ้น ในการทำการทดลองเพื่อวัดประสิทธิภาพ ผู้วิจัยทำการเปรียบเทียบเทคนิคที่น่าเสนอ ซึ่งได้แก่การผสมผสานระหว่างการรู้จำเสียงและท่าทาง ในสภาพแวดล้อมที่ไม่มีเสียงรบกวน ซึ่งค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 เดซิเบล และสภาพแวดล้อมที่มีเสียงรบกวน โดยกำหนดให้เสียงรบกวนอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยิน รับรู้เสียงที่ออกมาจากอินเทอร์เน็ตทีวีค่าเฉลี่ยของเสียงในห้องอยู่ที่ เฉลี่ย 70 เดซิเบล จากการทดลอง พบว่าขั้นตอนและวิธีการควบคุมอินเทอร์เน็ตทีวีด้วยการผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว สามารถทำงานได้อย่างถูกต้องทั้งในสภาพแวดล้อมทั้งสองแบบ

คำสำคัญ-- การรู้จำเสียง; ตรวจสอบการเคลื่อนไหว; อินเทอร์เน็ตทีวี;

Abstract

Voice recognition techniques and motion detection techniques were applied to improve an effectiveness of an internet television controlling task and make the task more convenient.

¹นักศึกษาระดับบัณฑิตศึกษาศาสาวิชาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ

มหาวิทยาลัยธุรกิจบัณฑิต

²อาจารย์ประจำสาขาวิชาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิต

These techniques, however, provide some benefit and drawback to the task. A combination of voice recognition and motion detection techniques was proposed to further improve an effectiveness of the controlling task. To evaluate the proposed technique, a series of experiments were conducted. The proposed technique was compared with a voice recognition technique only and motion detection techniques only. The experiments were conducted in two different environments, which are a room with an average of 46 dB sound level and a room with an average of 70 dB sound level. The result of experiment showed that the combination of voice recognition and motion detection technique provide a better performance in both environments.

Keyword-- Speech recognition; Motion Detection; Internet TV;

1. บทนำ

1.1 ที่มาและความสำคัญของปัญหา

เทคโนโลยีเป็นส่วนสำคัญอย่างมากที่จะทำให้การดำรงชีวิตอยู่ของมนุษย์ สะดวกสบายมากขึ้น โทรทัศน์ก็เป็นอีกหนึ่งของเทคโนโลยีที่ช่วยเพิ่มความสามารถในการรับรู้ข่าวสาร และให้ความบันเทิง กับผู้ใช้ตลอดระยะเวลาที่ผ่านมาโทรทัศน์ได้ถูกพัฒนาอย่างต่อเนื่องในเรื่องของการแสดงผลภาพ แต่ก็ยังใช้วิธีการควบคุมโดย เครื่องควบคุมระยะไกล (Remote Control) ในการควบคุมโทรทัศน์มาอย่างยาวนาน ซึ่งเป็นปัญหาในเรื่องของการควบคุมตัวโทรทัศน์ เนื่องจากรีโมทที่ใช้ควบคุมโทรทัศน์เป็นอีกหนึ่งอุปกรณ์ที่แยกจากตัวโทรทัศน์เองและจำเป็นต้องใช้แหล่งพลังงานเสริม และตัวรีโมทที่แยกออกมานั้นอาจมีการสูญหายทำให้การควบคุมโทรทัศน์เสียไป

อย่างไรก็ตามในปัจจุบันเริ่มมีการนำเทคโนโลยีการรู้จำเสียงและท่าทางมาใช้แทนที่สำหรับเครื่องควบคุมระยะไกล (Remote Control) แต่ก็ยังไม่เป็นที่นิยมเท่าที่ควรทั้งในเรื่องของฮาร์ดแวร์ที่มีราคาสูงและ เทคโนโลยีทั้งสองมีข้อด้อยกล่าวคือ เทคโนโลยีการรู้จำเสียงนั้นหากมีเสียงเข้ามารบกวนจะทำให้การทำงานเกิดความผิดพลาด เทคโนโลยีตรวจสอบการเคลื่อนไหวนั้นต้องใช้เวลาค่อนข้างมากในการสั่งงาน แต่ในปัจจุบันเทคโนโลยีทางตัวรับรู้ภาพและเสียงได้ถูกพัฒนาไปอีกระดับหนึ่ง โดยทางบริษัท ไมโครซอฟท์สามารถพัฒนาอุปกรณ์คิเนค (Kinect) ซึ่งเป็นอุปกรณ์ที่สามารถรับรู้ข้อมูลมิติความลึกของภาพและเสียงได้ในเวลาเดียวกัน ทำให้เพิ่มศักยภาพในการใช้ควบคุมโทรทัศน์ด้วยการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว

ดังนั้นงานวิจัยนี้ จึงเสนอเทคนิคในการควบคุมอินเทอร์เน็ตทีวี โดยใช้การผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหว ผ่านอุปกรณ์คิเนค เพื่อทดสอบถึงความ

ถูกต้องและความเร็วในการควบคุมอินเทอร์เนตที่วิทย์ได้สภาพแวดล้อมที่มีเสียงรบกวนและไม่รบกวน

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาการใช้งานกล้อง Kinect ในการตรวจจับการเคลื่อนไหว และการรู้จำเสียงของมนุษย์
2. เพื่อเสนอเทคนิควิธีที่ใช้ควบคุมอินเทอร์เนตที่วิทย์แบบใหม่ด้วยการผสมผสานระหว่างการรู้จำเสียงและการตรวจจับการเคลื่อนไหว
3. เพื่อทดสอบประสิทธิภาพของวิธีการควบคุมอินเทอร์เนตที่วิทย์ทั้งสามวิธี คือการรู้จำเสียง การตรวจจับการเคลื่อนไหว และการผสมผสานการรู้จำเสียงและการเคลื่อนไหว
4. เพื่อเสนอโปรแกรมแสดงผลการควบคุมอินเทอร์เนตที่วิทย์

1.3 ขอบเขตของการศึกษา

วิทยานิพนธ์นี้ประกอบไปด้วย สองส่วนหลักๆ คือการออกแบบและกระประเมินผล ในส่วนแรกคือการออกแบบเทคนิคสำหรับการควบคุมอินเทอร์เนตที่วิทย์โดย คีเนคด้วยการผสมผสานระหว่างการรู้จำเสียงและการตรวจสอบการเคลื่อนไหวมาใช้กับระบบอินเทอร์เนตที่วิทย์จำลองขึ้น ในส่วนที่สองเป็นการทดลองประเมินผลระบบด้วยเทคนิคที่นำเสนอและออกแบบไว้ โดยกลุ่มทดสอบจำนวน 60 คนนักศึกษาชั้นปี 1 – 3 ช่วงอายุประมาณ 19 – 21 ปี โดยจะแบ่งการทดสอบเป็น 2 กลุ่ม กลุ่มละ 30 คนโดยให้กลุ่มแรกทำการทดสอบในสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB และกลุ่มที่สองอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เนตที่วิทย์ค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB แล้วประเมินผลในเรื่องของความเร็วและความถูกต้องในการทำงาน

1.4 ความรู้เบื้องต้นและงานวิจัยที่เกี่ยวข้อง

1.4.1 คีเนค (Kinect) เป็นอุปกรณ์รับรู้การเคลื่อนไหวที่ใช้การทำงานผสมผสานกันระหว่างฮาร์ดแวร์และซอฟต์แวร์ ฟังก์ชันหลักของ คีเนค[1] มีสองฟังก์ชันคือ สร้างภาพเคลื่อนไหวสามมิติของวัตถุในมุมมองที่กำหนด และแยกแยะมนุษย์ออกจากวัตถุเหล่านั้นได้ ในปัจจุบัน คีเนคมีความละเอียดของภาพอยู่ที่ 640x480 pixel และสามารถทำงานได้ที่ 30 frames ต่อวินาที สำหรับ Hardware ของ คีเนคสามารถแบ่งออกเป็น ส่วนประกอบสำคัญ ได้แก่



ภาพที่ 1 สถาปัตยกรรมคีเนค

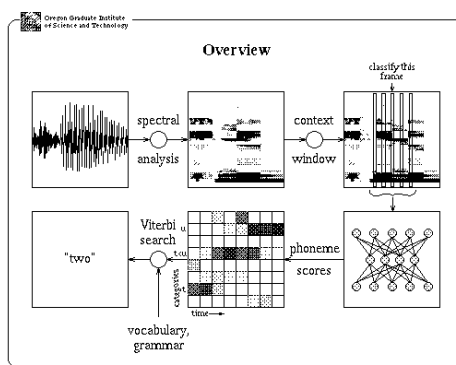
Color VGA video camera เป็นกล้องถ่ายภาพเคลื่อนไหวซึ่งช่วยในการจดจำใบหน้า (Face Recognition) และการตรวจจับลักษณะเด่นอื่นๆ โดยใช้ความสามารถในการตรวจจับองค์ประกอบของสีทั้งสามสีอื่น ได้แก่ สีแดง สีเขียว และสีน้ำเงิน

Depth sensor ประกอบด้วยการทำงานร่วมกันของต้นกำเนิดแสงอินฟราเรด (Infrared Projector) และตัวรับรู้แบบ monochrome CMOS (Complimentary Metal-Oxide Semiconductor) ซึ่งทำหน้าที่รับแสงอินฟราเรดที่ถูกสะท้อนกลับมาจากวัตถุ จากนั้นทำการวัดเวลาในการเดินทาง (Time of Flight) แสงอินฟราเรดนี้ใช้หลักการการทำงานเช่นเดียวกับโซนาร์คือ หากรู้ระยะเวลาที่แสงอินฟราเรดใช้ในการเดินทางไปกลับก็จะสามารถคำนวณระยะห่างระหว่างตัวรับรู้ความลึกกับวัตถุได้ โดยการทำงานด้วยความเร็วแสงหลายๆรอบทำให้สามารถระบุระยะห่างที่แน่นอนได้โดยไม่ต้องคำนึงถึงสภาพแสงสว่าง

Multi-array microphone เป็นอาร์เรย์ของไมโครโฟน 4 ตัวซึ่งสามารถแยกแยะเสียงของผู้ใช้ออกจากเสียงรบกวนภายในห้องได้

1.4.2 ระบบการรู้จำเสียง (Speech recognition) [2] หมายถึง การทำให้คอมพิวเตอร์สามารถที่จะฟังคำพูดและตัดสินใจได้ว่าคำพูดนั้นเป็นคำว่าอะไรหรืออีกความหมายหนึ่งคือการนำ File Audio ที่บรรจุเสียงพูดนำมาแปลงเป็น Text ได้

Automatic Speech Recognition (ASR) เป็นเทคโนโลยีที่เกิดขึ้นเพื่อที่จะให้คอมพิวเตอร์สามารถแยกแยะคำพูดต่างๆที่มนุษย์สามารถพูดใส่อุปกรณ์ไมโครโฟนหรือเครื่องโทรศัพท์หรืออื่นๆเท่าที่จะเป็นไปได้ “Holy grail” of ASR Search เป็นโปรแกรมที่อนุญาตให้คอมพิวเตอร์เข้าใจคำศัพท์ทุกคำอย่างถูกต้อง 100% ซึ่งมีความสามารถเข้าใจถึงคำพูดได้ดีไม่ว่าจะเป็นคำพูดของใครก็ตาม เป็นอิสระจากขนาดของกลุ่มคำศัพท์ , ความดัง, ลักษณะของผู้พูด และการออกเสียง หรือเงื่อนไขของช่องทางต่างๆที่เป็นไปได้



ภาพที่ 2. ภาพรวมการทำงานของ Speech Recognition

1.4.3 การตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect และ Kinect SDK Beta

ตำแหน่งของข้อต่อที่ได้รับจาก API จะสัมพันธ์กับตำแหน่งของร่างกายมนุษย์ 20 ตำแหน่งในท่ายืน และ 10 ตำแหน่งในท่านั่ง ตามที่แสดงในภาพที่ 2.10 เมื่อผู้ใช้นั่งหันหน้าเข้าหากล้อง ตำแหน่งของข้อต่อข้างซ้ายและขวาจะสลับข้างกัน แตกต่างจากร่างกายจริง แต่ละตำแหน่งของข้อต่อจะมีค่า X, Y และ Z ในหน่วยมิลลิเมตร และมีฟิลด์ Confidence แสดงสถานะ การตรวจจับตำแหน่งข้อต่อนั้นๆ โดยจะมีค่าระหว่าง 0 – 1 ซึ่ง 1 หมายถึง สามารถตรวจจับได้ และหากมีค่าน้อยกว่า 1 (เป็นตัวเลขทศนิยม) หมายถึง ไม่สามารถตรวจจับได้ ข้อมูลตำแหน่งของข้อต่อในส่วนนั้นจะเป็นค่าประมาณ ในบางครั้งอาจเป็นค่าที่ตรวจจับได้ล่าสุดก่อนที่จะตรวจจับไม่ได้ในเวลาต่อมา หรืออาจเป็นค่าที่ไม่มีความหมายก็ได้ ในการตรวจจับการเคลื่อนไหว หากผู้ใช้ยืนอยู่ในมุมกล้อง ห่างจากกล้อง ประมาณ 1.2 – 3.5 เมตร (ระยะห่างอุดมคติตามที่ Microsoft แนะนำคือ ประมาณ 2.5 เมตร) ค่าตำแหน่งของข้อต่อที่ ถูกตรวจจับได้ จะมีค่าใกล้เคียงกับการวัดจริง เช่น ความยาวของข้อศอกไปจนถึงกลางฝ่ามือ เป็นต้น และฟิลด์ Confidence ของข้อต่อจะมีค่าเป็น 1 แต่หากมีบางข้อต่อที่กล้องไม่สามารถตรวจจับได้ หรือถูกบดบังไป เช่น ยืนมือไขว่หลัง เป็นต้น ฟิลด์ Confidence ของข้อต่อนั้นจะมีค่าน้อยกว่า 1

1.4.4 งานวิจัยที่เกี่ยวข้อง

งานวิจัยเรื่อง “Evaluating a Dancer's Performance using Kinect-based

Skeleton Tracking” [3] (D. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, and M. B. Moussa, 2011 : 659 - 662) ได้ทำการศึกษาเกี่ยวกับการวัดประสิทธิภาพของนักเต้นโดยใช้กล้องคิเนคในการตรวจจับการเคลื่อนไหว และใช้ OpenNI เป็น API ในการให้ข้อมูลการตรวจจับ งานวิจัยนี้ใช้การเปรียบเทียบตำแหน่งของข้อต่อของนักเต้นสองคนที่ได้จากการตรวจจับตามเวลาจริงมาเป็นส่วนหนึ่งในการให้คะแนนการเต้น ผลการศึกษาของงานวิจัยนี้แสดงให้เห็นว่าสามารถนำข้อมูลตำแหน่งของข้อต่อและมือ ไปใช้งานในด้านการตรวจจับการเคลื่อนไหวได้

งานวิจัยเรื่อง “Real Time Extraction of Body Soft Biometric from 3D Videos” [4]

(C. Velardo and J. Dugelay, 2011 : 781-782) ได้ทำการศึกษาการใช้ คิเนคในการใช้ท่าทางและการรู้จำเสียงในการควบคุมส่วนติดต่อกับผู้ใช้นโปรแกรมมูเกิลเอิร์ธ โดยผลการศึกษาสามารถควบคุมได้ทั้งเสียงและท่าทางผลการศึกษาของงานวิจัยนี้แสดงให้เห็นว่าสามารถนำวิธีการรู้จำเสียงและท่าทางมาใช้ในการใช้เสียงตรวจจับคำพูดได้

งานวิจัยเรื่อง “การประกอบรวมโครงร่างมนุษย์จากการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว” [5] (นราวุฒิ พัฒโนทัย, พรชัย มงคลนาม และ บัณฑิต วรรณภา, 2012 : 370-374) ได้ทำการศึกษาการใช้ คิเนคสามตัวต่อเข้าด้วยกันในระบบ Client – Server โดยวางมุม

กล้องทำมุม 120 องศาต่อกัน การศึกษาของงานวิจัยนี้แสดงให้เห็นว่าสามารถเพื่อช่วยเพิ่มโอกาสในการมองเห็นจุด Skeleton ของคนในการจับการเคลื่อนไหวได้ดีขึ้น

ในปี 2012 บริษัท Samsung Electronic[6] ได้นำเทคโนโลยีการรู้จำเสียงและการตรวจสอบการเคลื่อนไหวใส่กับอินเทอร์เนตทีวี หลักการก็คล้ายๆกับ คีเนคที่กล้องจะจับท่าทางและตัวทีวีก็คอยจะตอบสนองคำสั่งตามท่าทางของเรา ตัวอย่างเช่น "การแบมือ" โชว์หน้ากล้องก็จะเริ่มต้นคำสั่ง หลังจากนั้นตัว Cursor ลูกศรก็จะปรากฏขึ้นมาบนหน้าจอ เลื่อนมือไปในทิศทางต่างๆ เพื่อเลื่อนตัวลูกศรที่เป็น Cursor ไปในทิศทางที่เราต้องการ หากต้องการกดปุ่มใดๆบนหน้าจอ ก็กำมือ ซึ่งเปรียบเสมือนการคลิกเมาส์ การสั่งงานด้วยเสียง เพียงแค่พูดว่า "HI TV" เพื่อเปิดหรือปิดการใช้งาน Voice Control หลังจากนั้นจะมี "แถบชุดเมนูคำสั่ง" ขึ้นมาด้านล่าง

ชุดคำสั่ง Voice Control

1. "TV Power Off" :ปิดทีวี
2. "Source" :เปลี่ยนแหล่งสัญญาณ
3. "Channel Number" :เลือกช่องทีวีตามที่ต้องการ
4. "Channel Up หรือ Down" :เลื่อนช่องทีวีขึ้นหรือลงทีละช่อง
5. "Volume Up/Down" :เพิ่มระดับเสียงดัง/เบา
6. "Mute" :ปิดเสียง

ดังนั้นงานวิจัยนี้จึงได้เสนอและออกแบบและพัฒนาเทคนิคการผสมผสานเทคนิคระหว่างการรู้จำเสียงและเทคนิคการตรวจสอบการเคลื่อนไหวเข้าด้วยกันเพื่อใช้ในการควบคุมอินเทอร์เนตทีวีโดยออกแบบการควบคุมไว้ดังนี้ เปลี่ยนช่อง 3,5,7,9,11,thai PBS เพิ่มลดเสียงได้ 5 ระดับและการเปิดปิดอินเทอร์เนตทีวี

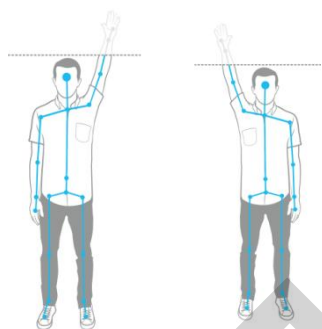
2. วิธีดำเนินการวิจัย

วิธีการผสมผสานระหว่างเทคนิคการรู้จำเสียงและการตรวจสอบการเคลื่อนไหวโดยใช้กล้องคีเนคประกอบไปด้วยขั้นตอนที่มีรายละเอียดดังต่อไปนี้

2.1 การตรวจสอบการรู้จำเสียง

ในการตรวจสอบการรู้จำเสียงอินพุตที่เข้ามาผ่านไมโครโฟนบนตัวของกล้องคีเนคได้นำไลบรารีของ Microsoft Speech Application Programming Interface 5.1 (SAPI 5.1) โดยจะเขียนโปรแกรมผ่าน Visual Studio 2010 เพื่อนำอินพุตเสียงที่ได้แปลงเป็นตัวอักษรโดยใช้เทคนิค Speech to Text (STT) จากนั้นนำคำที่ได้ไปเปรียบเทียบกับคำที่ระบุไว้หากไม่ตรงตามที่ระบุไว้ก็จะ Reject ทิ้ง หากตรงตามเงื่อนไขก็จะ Accepted โดยจะมีกระบวนการเปิดและปิดระบบการรู้จำเสียง จะ

กำหนดให้ผู้ใช้นิ้วชี้ขึ้นเหนือศีรษะจะเป็นการเปิดระบบการรู้จำเสียงและหากต้องการปิดให้ผู้ใช้นิ้วชี้ขึ้นเหนือศีรษะจะเป็นการปิดระบบเสียง โดยจะอาศัย Joint Hand Right และ Joint Hand Left ทำการเปรียบเทียบตำแหน่ง Position ด้านแกน Y กับ Joint Head ดังภาพที่ 3.



ภาพที่ 3. วิธีการเปิดปิดระบบการสั่งการด้วย

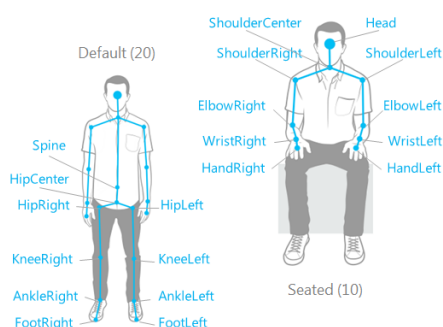


ภาพที่ 4. แสดงรูปภาพระบบการสั่งงานด้วยเสียง

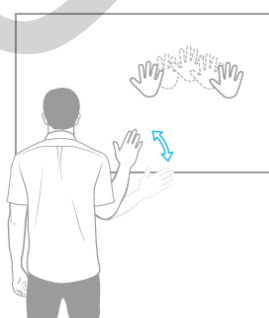
2.2 วิธีการตรวจสอบการเคลื่อนไหว

ในการตรวจสอบการเคลื่อนไหวภาพผ่านกล้อง Kinect ได้นำไลบรารีของ Kinect SDK โดยจะเขียนโปรแกรมผ่านโปรแกรม Visual Studio 2010

โดยทำการตรวจจับรูปร่างของคนแล้วจะนำ Skeleton ไปซ้อนทับบนภาพ โดยทำเช่นนี้จะกำหนดเป็น 20 Joint และทำเช่นนี้จะกำหนดเป็น 10 Joint จากนั้นก็ทำการเขียนโปรแกรมเก็บค่าเฉพาะ Joint Hand Right และ Joint Hand Left เพื่อนำค่า Position ของตำแหน่งมือไปเปรียบเทียบกับตำแหน่ง Cursor ลูกศรบนจอภาพ เพื่อให้ได้ตำแหน่งและพิกัด สำหรับการควบคุมอินเทอร์เนตทีวี



ภาพที่ 5. แสดงชื่อ Skeleton Joints



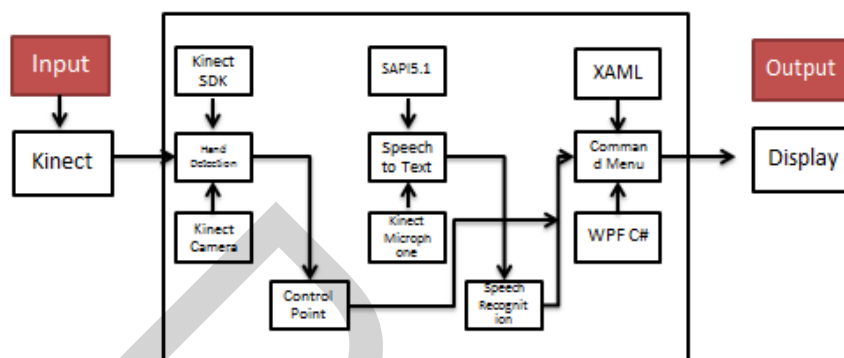
ภาพที่ 6. Cursor ที่ใช้เปรียบเทียบกับ Position

2.3 การผสมผสานวิธีการรู้จำเสียงและการตรวจจับการเคลื่อนไหว

เมื่อได้วิธีการรู้จำเสียงและการตรวจจับการเคลื่อนไหวแล้ว จึงได้ออกแบบวิธีการผสมผสานเทคนิคการรู้จำเสียงและการตรวจสอบการเคลื่อนไหวโดยกำหนดให้ เมื่อ Position ของ Cursor บนหน้าจออินเทอร์เนตทีวี ไป Detect กับปุ่มใดๆบนอินเทอร์เฟซบนหน้าจอพร้อมทั้งออกเสียง

“Yes” ก็จะเป็นการสั่งงานกับเมนูอื่นๆ ดังรูปที่ 1 แต่ถ้าหาก Cursor ไป Detect กับปุ่มใดๆบน อินเทอร์เน็ตหน้าจอแล้วมีการออกเสียงอื่น โปรแกรมก็จะไม่มี

การทำงาน หรือถ้าออกเสียงอย่างเดียวแล้วไม่ได้มีการนำ Cursor ไป Detect กับปุ่มใดๆ โปรแกรมก็จะไม่มีการทำงานอีกเช่นกัน



ภาพที่ 7. แสดงรูปแบบการทำงานของวิธีการ
ผสมผสานเทคนิคการรู้จำเสียงและการตรวจสอบการ

3. ผลการวิจัย

งานวิจัยนี้ทำการทดสอบประสิทธิภาพโดยการวัดความเร็วและความถูกต้องของวิธีการที่นำเสนอ โดยการพัฒนาโปรแกรมทั้งสามระบบ คือ 1.ระบบการรู้จำเสียงสำหรับการควบคุมอินเทอร์เน็ตทีวี 2.ระบบการตรวจสอบการเคลื่อนไหวสำหรับการควบคุมอินเทอร์เน็ตทีวี 3.ระบบผสมผสานระหว่างการรู้จำเสียงและการตรวจจับการเคลื่อนไหว ขึ้นมาด้วยภาษา C# ใช้ Kinect SDK เวอร์ชัน Beta 1 และสร้างสถานการณ์สภาพแวดล้อมในห้องขึ้นมา โดยให้สภาพแวดล้อมแรกเป็นสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวน โดยที่ค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB สภาพแวดล้อมที่สองคือสภาพแวดล้อมที่มีเสียงรบกวนโดยกำหนดให้เสียงรบกวนอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เน็ตทีวีที่ค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB ในการตรวจจับตั้งกล้องอยู่บนระนาบที่ขนานกับระนาบพื้นที่ทำการตรวจจับสูงจากระนาบพื้น 0.6 เมตร ห่างออกไป 1.5 เมตร โดยกำหนดโจทย์ที่ใช้ทดสอบดังต่อไปนี้ ปิด TV, เลือกร่อง (เปิด TV), เปลี่ยนช่อง, เพิ่มเสียง, ลดเสียง

ตารางที่ 1. แสดงถึงงานที่ผู้ทดสอบต้องทำในแต่ละชุดการทดสอบ

โจทย์สำหรับใช้ทดสอบ		
ชุดทดสอบ A	ชุดทดสอบ B	ชุดทดสอบ C
เลือกช่อง 3	เลือกช่อง 5	เลือกช่อง 9
เปลี่ยนช่อง 7	เพิ่มเสียงขึ้น 1 ระดับ	เพิ่มเสียงขึ้น 1 ระดับ
ลดเสียงลง 1 ระดับ	เปลี่ยนช่อง 9	ปิด TV
เพิ่มเสียงขึ้น 1 ระดับ	ลดเสียงลง 1 ระดับ	เปลี่ยนช่อง 11
ปิด TV	ปิด TV	ลดเสียงลง 1 ระดับ

โดยจะใช้ผู้ทดสอบจำนวน 60 คนนักศึกษาชั้นปี 1 – 3 ช่วงอายุประมาณ 19 – 21 ปี โดยจะแบ่งการทดสอบเป็น 2 กลุ่ม กลุ่มละ 30 คน โดยให้กลุ่มแรกทำการทดสอบในสภาพแวดล้อมที่ไม่มีเสียงเข้ามารบกวนค่าเฉลี่ยเสียงภายในห้องที่ทดสอบอยู่ที่เฉลี่ย 46 dB และกลุ่มที่สองอยู่ในระดับความดังที่ผู้ทดสอบจะสามารถได้ยินรับรู้เสียงที่ออกมาจากอินเทอร์เน็ตทีวีที่ค่าเฉลี่ยของเสียงในห้องอยู่ที่เฉลี่ย 70 dB โดยให้แต่ละกลุ่มทำการทดสอบดังนี้

- 1.ให้ผู้ทดสอบ สุ่มเลือก เทคนิคสำหรับควบคุมอินเทอร์เน็ตทีวี จาก 3 วิธี มา 1 วิธี พร้อมทั้งนำหัวข้อเทคนิคที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
- 2.จากนั้นให้ผู้ทดสอบ สุ่มเลือก ชุดทดสอบ จาก 3 ชุดทดสอบ มา 1 ชุดทดสอบ พร้อมทั้งนำหัวข้อชุดทดสอบ ที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
- 3.ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมินจะเริ่มทำการจับเวลาในการทดสอบ
- 4.เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป
- 5.ให้ผู้ทดสอบทำการ สุ่มเลือก เทคนิคสำหรับควบคุมอินเทอร์เน็ตทีวี จาก 2 วิธีที่เหลือ มา 1 วิธี พร้อมทั้งนำหัวข้อเทคนิคที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
- 6.จากนั้นให้ผู้ทดสอบ สุ่มเลือก ชุดทดสอบ จาก 2 ชุดทดสอบ มา 1 ชุดทดสอบ พร้อมทั้งนำหัวข้อชุดทดสอบ ที่ผู้ใช้เลือกได้ออกจากกองสุ่มเลือก
- 7.ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมินจะเริ่มทำการจับเวลาในการทดสอบ

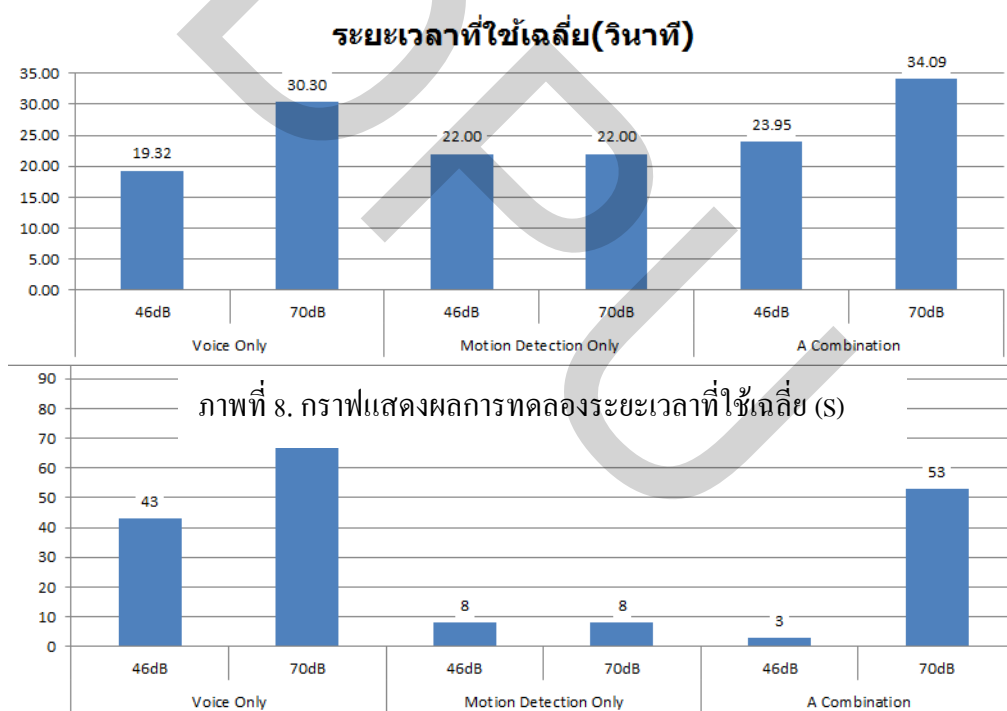
8.เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป

9.จากนั้นให้ผู้ทดสอบทำการทดสอบ เทคนิคสำหรับควบคุมอินเทอร์เนตทีวีและชุดทดสอบ ที่เหลืออยู่

10.ให้ผู้ทดสอบ เข้าอยู่ในบริเวณที่จะทดสอบ หากผู้ทดสอบเตรียมตัวพร้อมผู้ประเมิน จะเริ่มทำการจับเวลาในการทดสอบ

11.เมื่อผู้ทดสอบ ทำตามเงื่อนไขจากชุดทดสอบจนเสร็จ ผู้ประเมินจะทำการหยุดเวลา พร้อมจดบันทึกเวลาที่ผู้ทดสอบได้ใช้ไป

12.นำข้อมูลที่ได้จากทั้ง 3 กลุ่ม มาวิเคราะห์เปรียบเทียบความแตกต่างในเรื่องของ ความเร็วและความถูกต้องของการทำงาน



ภาพที่ 9. กราฟแสดงผลการทดลองจำนวนครั้งที่ผิดพลาดในการ

จากผลงานทดลองส่วนวิเคราะห์การใช้วิธีรู้จำเสียงในการสั่งงานอินเทอร์เนตทีวีพบว่าสามารถสั่งงานได้รวดเร็วแต่จะมีความผิดพลาดค่อนข้างมากทั้งสองสถานการณ์เนื่องจากคำสั่งที่ใช้ในการควบคุมมีปริมาณของคำสั่งมากทำให้ความเป็นไปได้ของเสียงที่จะออกถูกต้องมีน้อยลง อีกทั้งหากเวลาควบคุมมีเสียงจากทีวีเข้าสู่ตัวรับเสียงก็ยิ่งทำให้เกิดอัตราการผิดพลาดของคำสั่งมากขึ้นไปอีกด้วย อาจแก้ไขด้วยวิธีการย้ายอุปกรณ์รับเสียงให้อยู่ใกล้ตัวผู้สั่งการมากที่สุด

เพื่อที่จะรับเสียงได้อย่างถูกต้องชัดเจนและเสียงรบกวนในห้องน้อยที่สุดเพื่อเพิ่มความถูกต้องของการทำงาน

ส่วนวิเคราะห์การตรวจสอบการเคลื่อนไหว นั้นสามารถทำได้ดีในทั้งสถานการณ์ทั้งสอง เนื่องจากเสียงรบกวนไม่มีผลกระทบต่อเทคนิคนี้ และทำงานได้อย่างถูกต้อง แต่จะทำงานได้ค่อนข้างช้า เนื่องจากผู้สั่งการต้องอยู่ในระยะที่กล้องสามารถตรวจจับได้และแสงสว่างภายในห้องต้องเพียงพอต่อการมองเห็นของกล้อง ก็จะทำให้การทำงานถูกต้องมากยิ่งขึ้น

ส่วนวิเคราะห์การผสมผสานระหว่างการใช้เสียงและการตรวจสอบการเคลื่อนไหว จะพบว่าในสถานการณ์ที่หนึ่งคือเสียงในห้องอยู่ในระดับ 46 dB การใช้วิธีนี้ได้ผลการทำงานที่ถูกต้องมากที่สุดเนื่องจากตัวโปรแกรมต้องยืนยันการทำงานถึงสองกระบวนการคือตรวจสอบจากเคลื่อนไหวของมือเมื่อการเคลื่อนไหวของมือถูกต้องจึงมาตรวจสอบด้วยเสียงอีกครั้งหนึ่งข้อดีของวิธีนี้เราจะมีค่าที่ใช้ยืนยันแค่ค่าเดียวคือ Yes เพื่อลดความผิดพลาดในเรื่องของคำหรือประโยคที่คล้ายๆกันในการสั่งการอินเทอร์เน็ตทีวี ทำให้การทำงานออกมาค่อนข้างถูกต้องสมบูรณ์ มากที่สุดในกรณีที่หนึ่ง ส่วนในกรณีที่สองคือเสียงภายในห้องเปลี่ยนอยู่ที่ 70 dB การทำงานเริ่มจะมีความผิดพลาดเนื่องจากมีเสียงรบกวนจากภายนอกทำให้ตัวโปรแกรมไม่สามารถตรวจจับคำได้ แต่อย่างไรก็ดีวิธีนี้โปรแกรมก็ยังสามารถตรวจจับคำได้ดีกว่าการใช้วิธีการรู้จำเสียงอย่างเดียวเพราะโปรแกรมจะสามารถจับคำหรือประโยคสั้นๆได้ดีกว่าประโยคยาวๆที่มีหลากหลายของคำตั้ง

4. อภิปรายผล

งานวิจัยชิ้นนี้ได้นำเสนอเทคนิคสำหรับการควบคุมอินเทอร์เน็ตทีวีโดยทีเนคด้วยการผสมผสานระหว่างการใช้เสียงและการตรวจสอบการเคลื่อนไหวจากผลของวิธีการที่นำเสนอแสดงให้เห็นว่า ผลลัพธ์ที่ได้ในเรื่องของประสิทธิภาพของความถูกต้องในการควบคุมอินเทอร์เน็ตทีวีมีความถูกต้องมากที่สุดในสถานการณ์ที่มีเสียงรบกวนน้อย คืออยู่ที่ 46 dB เนื่องจากเป็นการตรวจสอบถึงสองขั้นตอนวิธี ทั้งการใช้เสียงและการตรวจสอบการเคลื่อนไหว แต่ก็มีข้อเสียในเรื่องของความเร็ว อาจเป็นเพราะผู้ทดสอบเพิ่งเคยได้ใช้เทคโนโลยีที่ไม่เคยได้ถูกฝึกมาก่อน แต่เพื่อประโยชน์อย่างแท้จริงในอนาคตควรที่จะมีการทดลองในรูปแบบต่างๆมากขึ้นเช่น กลุ่มทดลองกลุ่มที่หนึ่งที่ได้รับการฝึกฝนการใช้กับอีกกลุ่มที่ไม่ได้รับการฝึกฝนการใช้งานเพื่อวัดผลทางด้านประสิทธิภาพและประสิทธิผลของแต่ละวิธี

บรรณานุกรม

- [1] Microsoft. **Kinect for Xbox 360 - Xbox.com**. [Online]. Available: www.xbox.com/en-GB/kinect, 2012.
- [2] Microsoft. **Speech recognition – Sapi5.1**. [Online]. Available: www.microsoft.com/en-us/download/details.aspx?id=10121
- [3] D. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, and M. B. Moussa, "Evaluating a Dancer's Performance using Kinect-based Skeleton Tracking" **Proceedings of the 19th ACM international conference on Multimedia**. pp. 659-662:Nov. 2011.
- [4] Velardo, C. and Dugelay, J., 2011, "Real Time Extraction of Body Soft Biometric from 3D Videos", **Proceedings of the 19th ACM international conference on Multimedia**, 28 November-1 December, Arizona, USA, pp. 781-782.
- [5] นราวุฒิ พัฒโนทัย. พรชัย มงคลนาม. บัณฑิต วรรณภา. "การประกอบรวมโครงร่างมนุษย์จากการตรวจจับการเคลื่อนไหวโดยใช้กล้อง Kinect หลายตัว" **National Conference on Information Technology 2012**. 370-374 : April. 2012.
- [6] Roman and Boomerang. **รีวิว Samsung Smart TV 55ES8000**. [Online]. Available: http://www.lcdtvthailand.com/review/detail.asp?desc=1¶m_id=1277

ภาคผนวก จ

หนังสือรับรองผลงานวิจัยในการประชุมวิชาการ การบริหารและการ
จัดการครั้งที่ 8 “การบริหารจัดการแห่งโลกพลวัต” โดยมหาวิทยาลัย
ธุรกิจบัณฑิตร่วมกับเครือข่ายวิจัยประชาชน

ภาพที่ 9 ประกาศนียบัตรรับรองในการนำเสนอผลงานวิชาการ



รายชื่อผู้ทรงคุณวุฒิพิจารณาผลงานวิชาการ

ศาสตราจารย์ ดร.ไพฑูรย์ สินลารัตน์
 รองศาสตราจารย์ ดร.เชาว์ โรจน์แสง
 รองศาสตราจารย์ ดร.นรา สมประสงค์
 รองศาสตราจารย์ ดร.ประคินันท์ อุปรมัย
 รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร
 รองศาสตราจารย์ ดร.สุชาติ ณ หนองคาย
 รองศาสตราจารย์ ดร.อุษา บิ๊กกินส์
 รองศาสตราจารย์ พินิจ ทิพย์มณี
 รองศาสตราจารย์ ยุทธนา ธรรมเจริญ
 รองศาสตราจารย์ สมประสงค์ วิทยเกียรติ
 รองศาสตราจารย์ อดิศักดิ์ ปัทมานิช
 ผู้ช่วยศาสตราจารย์ ดร.เพียรศิริ เอกนิคม
 ผู้ช่วยศาสตราจารย์ ดร.จิตติมา สิทธิพงษ์พานิช
 ผู้ช่วยศาสตราจารย์ ดร.นภาพร ธีระชาญ
 ผู้ช่วยศาสตราจารย์ ดร.ลีลา เดียงสูงเนิน
 ผู้ช่วยศาสตราจารย์ ดร.วรสิทธิ์ ชูชัยวัฒนา
 ผู้ช่วยศาสตราจารย์ ดร.วราพร จิระพันธุ์ทอง
 ผู้ช่วยศาสตราจารย์ ดร.วันวร จะนู
 ผู้ช่วยศาสตราจารย์ ดร.อดิศักดิ์ พงษ์กล้า
 ผู้ช่วยศาสตราจารย์ ปิยวรรณ แสงสว่าง
 ผู้ช่วยศาสตราจารย์ ผ่องใส เพ็ชรรักษ์
 ผู้ช่วยศาสตราจารย์ สุเทพ พันประสิทธิ์
 ผู้ช่วยศาสตราจารย์ เฉลิมพร อภิชนาพงศ์
 ผู้ช่วยศาสตราจารย์ เชียง เกาซิด
 ผู้ช่วยศาสตราจารย์ เลิศพร ภาระสกุล
 ผู้ช่วยศาสตราจารย์ จุลศักดิ์ ชาญณรงค์
 ดร.เกียรติอนันต์ ล้วนแก้ว
 ดร.เขวาลักษณ์ ราชแพทยาคม
 ดร.เอกพงศ์ กิตติสาร
 ดร.ธัญรัช วิกิตภูมิประเทศ
 ดร.ปริญ เพ็ญวุฒิ
 ดร.รชฏ ขำบุญ
 ดร.สาวิตรี สมุทรจักร
 ดร.อรพรรณ อัมสมบัติ
 อาจารย์วิญญู นิรนาทล้ำพงศ์

ประวัติผู้เขียน

ชื่อ-นามสกุล

ภูวดล ศิริกองธรรม

ประวัติการศึกษา

สำเร็จการศึกษาระดับปริญญาตรีสาขาวิชา
วิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิต
ปีการศึกษา 2553

ตำแหน่งและสถานที่ทำงานปัจจุบัน

อาจารย์ประจำ สาขาวิชาเทคโนโลยีมัลติมีเดีย
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น