NUEJ
NARESUAN UNIVERSITY
ENGINEERING JOURNAL

# การพัฒนาระบบวิเคราะห์เสียงสำหรับการควบคุมห้องอัจฉริยะ

# Development of a voice analysis system for controlling smart rooms

สุเมธ เหมะวัฒนะชัย[1,3*], มงคลชัย รุ่งเรือง[1,3], สะการะ ตันโสภณ[2]

Sumet Heamawatanachai[1,3*] , Mongkolchai Rungrueang[1,3] , Sakara Tunsophon[2]

[1]ภาควิชาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

[2]ภาควิชาสรีรวิทยา คณะวิทยาศาสตร์การแพทย์ มหาวิทยาลัยนเรศวร

[3]หน่วยวิจัยเทคโนโลยีด้านวิศวกรรมความเที่ยงตรงและการแพทย์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร

*Corresponding author: sumeth@nu.ac.th, sumet_h@yahoo.com

**บทคัดย่อ**

งานวิจัยนี้เป็นการศึกษาและพัฒนาระบบวิเคราะห์เสียงเพื่อนำไปประยุกต์ใช้กับระบบควบคุมห้องอัจฉริยะ เพื่อช่วยอำนวยความสะดวกสำหรับผู้ป่วยติดเตียงในการควบคุมเครื่องใช้ไฟฟ้าต่าง ๆ โดยทั่วไปพบว่ากลุ่มผู้ป่วยติดเตียง มีปัญหาหลักคือการเคลื่อนไหวและช่วยเหลือตนเองได้ยากลำบาก จึงจำเป็นต้องรอการช่วยเหลือจากผู้ดูแล ซึ่งส่วนใหญ่ผู้ป่วยเหล่านี้ยังสามารถออกเสียงพูดได้ ดังนั้นงานวิจัยนี้จึงเป็นการพัฒนาโปรแกรมเพื่อตรวจจับและวิเคราะห์สัญญาณเสียงจากผู้ป่วยหรือผู้ใช้งาน เพื่อนำไปประยุกต์ใช้ในการควบคุมเครื่องใช้ไฟฟ้า โดยโปรแกรมนี้ถูกพัฒนาด้วยภาษา LabVIEW ทำงานโดยรับสัญญาณเสียงจากไมโครโฟนแล้วนำมาประมวลผลด้วยเทคนิค Fast Fourier Transform (FFT) เพื่อให้ได้ข้อมูลเชิงความถี่ซึ่งจะถูกแบ่งกลุ่มแล้วเปรียบเทียบกับฐานข้อมูลที่ได้ทำการบันทึกไว้แล้วด้วยวิธีการ K-Nearest Neighbor Classification (KNN) เพื่อจำแนกเสียงพูดออกเป็น 5 เสียงพื้นฐานได้แก่ เอ อี โอ อา และ อู โดยการทดสอบระบบวิเคราะห์สัญญาณเสียงโดยใช้เสียงทั้ง 5 พบว่าระบบให้ความถูกต้องเฉลี่ยมากกว่าร้อยละ 90 ในกรณีที่เลือกเสียงกลุ่มละสองเสียงที่เหมาะสมจากฐานข้อมูลมาใช้ในการสร้างรหัสควบคุม เช่น “โอ อา” “อา อู” และ “อี อา” จะสามารถให้ผลที่ถูกต้องถึงร้อยละ 99 ในกรณีที่เลือกกลุ่มละสามเสียงที่เหมาะสมเช่น “อี โอ อา” มาใช้ในการสร้างรหัสควบคุมจะให้ผลการตรวจจับที่ถูกต้องถึงร้อยละ 99 โดยสามารถปรับเลือกใช้กลุ่มเสียงให้เหมาะกับเสียงจากผู้ใช้แต่ละคนได้ โดยสรุปแล้ว ระบบที่พัฒนาขึ้นมีความถูกต้องสูงและมีศักยภาพในการนำไปช่วยผู้ป่วยติดเตียง ในการควบคุมเครื่องใช้ไฟฟ้าด้วยเสียงได้

**คำสำคัญ:** ห้องอัจฉริยะ, ผู้ป่วยติดเตียง, สั่งการด้วยเสียง, KNN

**Abstract**

This research presents the development of a voice activated system to assist bedridden patient in controlling of room appliance. Generally, it is hard for bedridden patients to move or to control electrical devices by themselves, thus they have to wait for caregiver to assist. However, many of them are able to speak or generate voices. Therefore, the objective of this study was to develop the software to detect and analyze the patient’s voices which can be applied for controlling room appliances. The software in this research was developed using LabVIEW. The voice signal was sent to computer via microphone, then the Fast Fourier Transform technique was applied to get frequency domain data. The k-Nearest Neighbors (KNN) classification was used to compare the frequency data with the recorded database to recognize the voice. There were five defined voices in this research (a_e, ee, o_e, aa, oo) to be used for controlling. From the experiment results, using 5 voices, the developed voice recognition software achieved accuracy greater than 90%. In case of using 2 appropriated voices from 5 voices to construct control code such as “o_e, aa”, “aa, oo” and “ee, aa” the system achieve 99% accuracy. In case of using 3 suitable voices to construct control code such as “ee, o_e, aa”, the system also achieve high accuracy at

99%. It is also possible to select suitable voices to match with individual user. In conclusion, the developed system has high accuracy and has high potential to assist bedridden patients to control of room appliances using their voices.

**Keywords:** Smart room, Bedridden patient, Voice activated system, KNN classification.

## 1. บทนำ

กลุ่มผู้ป่วยติดเตียงหรือกลุ่มผู้ที่ต้องอาศัยอยู่บนเตียงเป็นระยะเวลานาน ซึ่งอาจเคลื่อนไหวและช่วยเหลือตัวเองได้ยากลำบาก โดยกลุ่มผู้ป่วยติดเตียงมีทั้งที่เป็นผู้ได้รับบาดเจ็บ ผู้ป่วยและผู้สูงอายุ ปัญหาผู้ป่วยติดเตียงมีมานานแล้ว ซึ่งกลุ่มผู้ป่วยเหล่านี้อาจมาจากผู้ที่ประสบอุบัติเหตุ ผู้เข้ารับการรักษาพยาบาล เช่น การผ่าตัด หรือการพักฟื้น กลุ่มผู้ป่วยที่เป็นโรคบางอย่างที่ทำให้ไม่สามารถเคลื่อนไหวได้ เช่น โรคอัมพฤกษ์ อัมพาต เป็นต้น และกลุ่มผู้พิการ โดยในบางกรณีผู้ป่วยติดเตียงอาจจะถึงขั้นที่ไม่สามารถพูดได้อย่างชัดเจน แต่ยังสามารถเปล่งเสียงได้บ้าง เช่น อา-อู เป็นต้น ส่วนผู้ติดเตียงที่เป็นผู้สูงอายุนั้น พบว่าปัจจุบันมีแนวโน้มมากขึ้น จากผลสำรวจของมูลนิธิสถาบันวิจัยและพัฒนาผู้สูงอายุไทยได้เผยว่า สังคมไทยมีแนวโน้มประชากรผู้สูงอายุที่มากขึ้น [1] ดังนั้นจำนวนผู้สูงอายุที่ต้องการการดูแลอย่างใกล้ชิดและผู้สูงอายุที่นอนติดเตียงจึงมีมากขึ้นด้วย

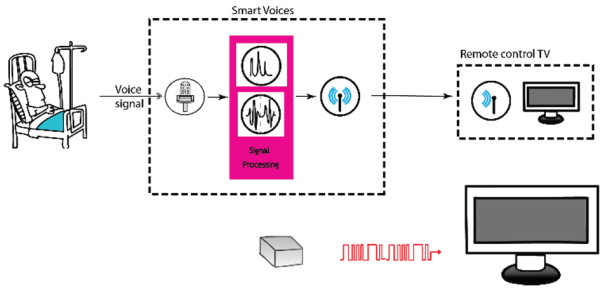
ปัญหาหลักที่พบได้ในกลุ่มผู้ติดเตียง ไม่ว่าจะเป็นผู้ป่วยติดเตียงที่ประสบอุบัติเหตุ เป็นโรคบางชนิด ผู้พิการ หรือผู้สูงอายุก็คือ การขยับเคลื่อนที่หรือการช่วยเหลือตัวเองได้ยากลำบาก ทำให้จำเป็นต้องรอการช่วยเหลือหรือมีผู้ดูแลอย่างใกล้ชิดจากเจ้าหน้าที่แพทย์ พยาบาล เพื่อน หรือญาติพี่น้อง อาทิเช่น การขอความช่วยเหลือในยามฉุกเฉิน การเปิด-ปิดอุปกรณ์ภายในห้องพัก เช่น เครื่องปรับอากาศ โทรทัศน์ หรือหลอดไฟ เป็นต้น จากการศึกษาระดับคุณภาพชีวิตในผู้ป่วยติดเตียงในประเทศอินเดียและญี่ปุ่น พบว่า ผู้ป่วยเหล่านี้มีระดับคุณภาพชีวิตสัมพันธ์กับระดับของการช่วยเหลือตัวเองในการทำกิจวัตรประจำวัน นอกจากนี้ พบว่าผู้ป่วยติดเตียงมีความเครียดเพิ่มมากขึ้น [2,3] ซึ่งเกิดจากการที่ผู้ป่วยรู้สึกว่าตนเองเป็นภาระแก่คนรอบข้าง แต่ถ้าผู้ป่วยสามารถทำอะไรได้ด้วยตนเอง ก็จะทำให้เขารู้สึกว่าตนเองมีคุณค่ามากขึ้น

งานวิจัยที่พัฒนาขึ้นเพื่อช่วยผู้ป่วยหรือผู้พิการมีอย่างมากมาย ตัวอย่างเช่น การพัฒนาระบบควบคุมรถเข็นผู้ป่วยด้วยการสั่งการด้วยเสียง [4, 5, 6] การพัฒนาวิธีการประมวลผลสัญญาณเสียงเพื่อนำมาประยุกต์ใช้กับผู้ป่วย [7, 8, 9] การออกแบบสภาพแวดล้อมที่เหมาะสมกับผู้ป่วยและผู้ดูแล [10] ฯลฯ ทั้งนี้ระบบที่มีอยู่เหมาะกับผู้ใช้ที่สามารถออกเสียงเป็นประโยคได้ชัดเจน

คณะผู้วิจัยได้เห็นถึงความสำคัญในการช่วยเหลือผู้ป่วยติดเตียง จึงได้พัฒนาระบบควบคุมห้องอัจฉริยะสั่งการด้วยเสียง โดยภายในห้องอัจฉริยะนี้ผู้ป่วยติดเตียงจะสามารถควบคุมอุปกรณ์ต่าง ๆ ภายในห้องได้โดยไม่ต้องเคลื่อนที่ออกจากเตียงหรือขยับมือสัมผัสปุ่มกดเพื่อสั่งการควบคุม แต่ผู้ป่วยติดเตียงจะใช้แค่การออกเสียงเพื่อสั่งการ ซึ่งจะช่วยเพิ่มความสะดวกในการควบคุมอุปกรณ์ต่าง ๆ และเป็นทางออกที่ดีสำหรับผู้ป่วยติดเตียงที่เคลื่อนไหวลำบาก โดยระบบจะสามารถรับรู้คำพูดแบบเสียงสระง่าย ๆ ได้แก่ เอ อี อา โอ อู ซึ่งจะวิเคราะห์จากลักษณะของเสียงคำพูด แล้วจะสั่งการเพื่อควบคุมอุปกรณ์ไฟฟ้า ระบบดังกล่าวนี้ผู้วิจัยได้ใช้โปรแกรม LabVIEW ในการวิเคราะห์ประมวลผล และส่งคำสั่งไปยังแผงวงจรไมโครคอนโทรลเลอร์เพื่อควบคุมตัวส่งสัญญาณอินฟราเรดเพื่อสั่งการอุปกรณ์ ได้แก่ โทรทัศน์ หรืออุปกรณ์อื่นๆที่รองรับสัญญาณอินฟราเรด โดยผลที่ได้จะทำให้ผู้ป่วยติดเตียงสามารถช่วยเหลือตัวเองได้ เป็นการลดภาวะเครียด รวมทั้งทำให้คุณภาพชีวิตผู้ป่วยติดเตียงดีขึ้น โดยในบทความนี้จะเน้นเฉพาะในส่วนของการพัฒนาระบบวิเคราะห์เสียงที่เหมาะสมสำหรับแยกแยะเสียงสระอย่างง่าย และหลักการสร้างรหัสควบคุม

## 2. การออกแบบการทำงานของระบบควบคุม

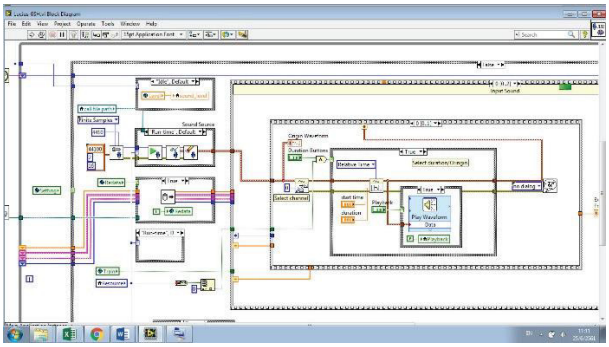
หลักการของระบบควบคุมห้องอัจฉริยะสั่งการด้วยเสียง โดยตัวอย่างสำหรับการควบคุมโทรทัศน์ แสดงดังรูปที่ 1 ซึ่งประกอบไปด้วย 3 ส่วนสำคัญ คือ ส่วนรับสัญญาณเสียงจากผู้ป่วยผ่านไมโครโฟน ส่วนของโปรแกรมการวิเคราะห์เสียง และส่วนของการส่งสัญญาณอินฟราเรดเพื่อควบคุมโทรทัศน์ ซึ่งหากผู้ใช้ส่งสัญญาณเสียงตรงกับรหัสฐานข้อมูลที่ตั้งไว้ ระบบจะส่งสัญญาณอินฟราเรดไปควบคุมเครื่องใช้ไฟฟ้าเช่นโทรทัศน์ หรือเครื่องปรับอากาศได้ โดยในการเข้ารหัสเสียงจะใช้เสียงสระพื้นฐาน ได้แก่ ‘เอ’ ‘อี’ ‘โอ’ ‘อา’ และ ‘อู’ เพื่อให้สามารถใช้งานได้ง่าย โดยผู้ป่วยไม่จำเป็นต้องพูดชัดเจนเป็นประโยคเช่น “เปิด ที วี” หรือ “เพิ่ม เสียง” แต่จะใช้รหัสจากเสียงสระ เช่น “อา โอ” หรือ “อู อี เอ” แทน




**รูปที่ 1** กระบวนการทำงานของระบบห้องอัจฉริยะสั่งการด้วยเสียงสำหรับควบคุมโทรทัศน์

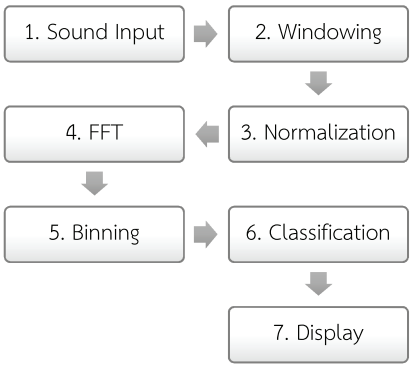
*2.1.1 หลักการทำงานของโปรแกรมวิเคราะห์สัญญาณเสียง*

โปรแกรมวิเคราะห์สัญญาณเสียงในงานวิจัยนี้ถูกพัฒนาบนโปรแกรม NI LabVIEW ซึ่งเป็นภาษาโปรแกรมแบบไดอะแกรมรูปภาพ ซึ่งเหมาะต่อการพัฒนาโปรแกรมที่มีการเชื่อมต่อกับอุปกรณ์และมีการคำนวณหลาย ด้านพร้อม ๆ กัน โดยการเขียนโค้ดจะเขียนด้วยรูปสัญลักษณ์ภาพแทนการเขียนด้วยตัวอักษร ดังรูปที่ 2



**รูปที่ 2** ลักษณะการเขียนโค้ดโปรแกรม LabVIEW

กระบวนการทำงานของโปรแกรมประกอบด้วย การรับสัญญาณเสียง การแบ่งหน้าต่าง การปรับค่ามาตรฐาน การแปลงฟูริเย การแบ่งข้อมูล การจำแนกข้อมูล และการแสดงผล ดังแสดงในรูปที่ 3.



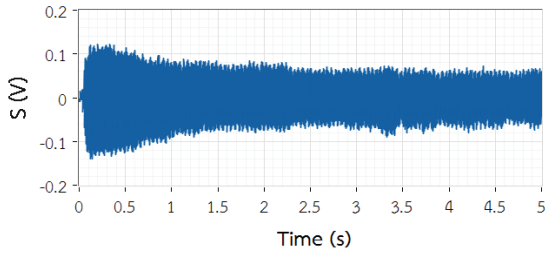

**รูปที่ 3** กระบวนการวิเคราะห์สัญญาณเสียง

*2.1.2 การรับข้อมูลของสัญญาณเสียง*

ส่วนแรกของการทำงานคือการรับข้อมูลของสัญญาณเสียงจากไมโครโฟน ข้อมูลที่ได้รับจะอยู่ในลักษณะของคลื่น (Waveform) ซึ่งมีเป็นค่าขนาดในเชิงเวลา (Time Domain) ตามสมการที่ (1) โดยในรูปที่ 4 ได้แสดงตัวอย่างสัญญาณคลื่นเสียงที่บันทึกในช่วงเวลา 5 วินาที

$$S = \{ (t, M) \mid t \geq 0, M \in R \} \quad (1)$$

โดยที่ $S$ คือ เซตของขนาดของสัญญาณ ณ เวลาต่าง ๆ



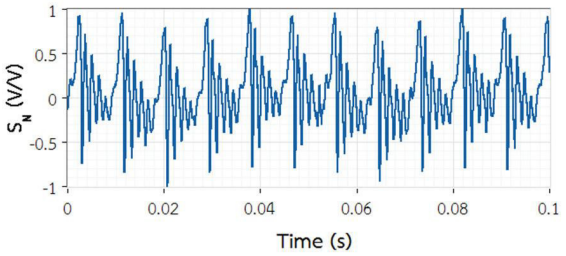

**รูปที่ 4** สัญญาณคลื่นเสียงที่บันทึกในช่วงเวลา 5 วินาที

*2.1.3 การแบ่งหน้าต่าง (Windowing)*

เนื่องจากกระบวนการรับสัญญาณเสียงจะส่งผลกระทบต่อการวิเคราะห์อย่างมาก เพราะจำนวนข้อมูลเสียงที่รับมาจะมีผลต่อระยะเวลาในการประมวลผลของโปรแกรม ซึ่งขณะที่กำลังประมวลผลอยู่นั้น โปรแกรมจะไม่สามารถรับข้อมูลเสียงได้จนกว่าจะประมวลผลเสร็จ ดังนั้นทำให้ข้อมูลจากสัญญาณเสียงขาดหายไปบางช่วง หากยิ่งรับข้อมูลจำนวนมากขึ้นก็จะยิ่งมีข้อมูลที่ขาดหายไปมากขึ้นเช่นกัน การกำหนดช่วงเวลาในการรับและวิเคราะห์ข้อมูลจึงเป็นส่วนสำคัญ โดยหากใช้ข้อมูลสัญญาณเสียงด้วยช่วงเวลาน้อย ๆ (ระหว่าง 5-100 มิลลิวินาที) [11] จะช่วยให้การวิเคราะห์ความถี่มีผลที่ชัดเจนและลดปัจจัยจากความถี่จากช่วงเวลาอื่นที่ปนเข้ามาได้ เมื่อเทียบกับการรับข้อมูลในช่วงเวลามาก ๆ ดังนั้นการแบ่งหน้าต่าง (Windowing) เป็นการเลือกข้อมูลเฉพาะในระยะเวลาหรือขอบเขตที่ต้องการ เพื่อทำให้จำนวนการรับข้อมูลในแต่ละครั้งจะมีปริมาณข้อมูลที่เหมาะแก่การนำไปประมวลผล ซึ่งสำหรับโปรแกรมที่พัฒนาขึ้นนี้สามารถปรับการรับข้อมูลให้มีระยะเวลาน้อยที่สุดที่ 100 มิลลิวินาที ดังนั้นโปรแกรมในงานวิจัยนี้จึงรับข้อมูลเสียงจำนวนครั้งละ 4,410 sample ที่ อัตราการสุ่มสัญญาณ 44,100 sample/s ดังนั้นหนึ่งรอบของการอ่านข้อมูล 4,410 sample จะใช้เวลา 100 มิลลิวินาที ตามสมการที่ (2)

$$Sampling\ rate = \frac{n(S)}{t} \quad (2)$$

โดยที่ *Sampling Rate* คือ อัตราการสุ่มข้อมูล โดย $n(S)$ คือ จำนวนข้อมูลในเซต $S$ และ $t$ คือ ระยะเวลาในการรับข้อมูล โดยรูปที่ 5 แสดงตัวอย่างสัญญาณเสียง “อา” ที่ถูกเลือกมาในช่วงเวลา 100 มิลลิวินาที




**รูปที่ 5** สัญญาณเสียง ‘อา’ จากข้อมูลที่วินาที 4.9 – 5.0

*2.1.4 การปรับค่ามาตรฐาน (Normalization)*

ก่อนการวิเคราะห์ข้อมูลเพื่อหาความถี่ ข้อมูลจะถูกปรับค่าโดยวิธี Peak Normalization ตามสมการที่ (3) ซึ่งจะทำให้ข้อมูลมีช่วงสเกลที่เท่ากันคืออยู่ระหว่าง -1 ถึง 1 ก่อนนำมาใช้ในการคำนวณ (ดังแสดงในรูปที่ 5)

$$S_N = \frac{S}{max(S)} \tag{3}$$

โดยที่ $S_N$ คือ เซตของข้อมูลสัญญาณเสียงที่ทำการปรับค่าโดยวิธี Peak Normalization และ $max(S)$ คือ ค่าที่มากที่สุดของข้อมูลสัญญาณเสียง

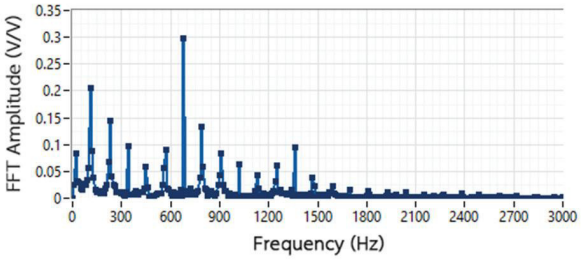
*2.1.5 Fast Fourier Transform*

ข้อมูลของสัญญาณที่ถูกปรับค่าแล้วจะถูกวิเคราะห์ด้วยเทคนิค Fast Fourier Transform (FFT) เพื่อเปลี่ยนข้อมูลเชิงเวลาให้เป็นข้อมูลเชิงความถี่โดย *FFT* เป็นฟังก์ชันของ $S_N$ ตามสมการที่ (4) และ *FFT* เป็นเซตคู่อันดับความถี่และแอมพลิจูด ซึ่งแสดงได้ตามสมการที่ (5) หลังจากนั้นในงานวิจัยนี้จะเลือกข้อมูลความถี่ที่มีแอมพลิจูดสูงที่สุดจำนวน 20 อันดับแรกมาใช้เพื่อลดระยะเวลาการทำงานของโปรแกรมให้น้อยลง ข้อมูลที่ได้จะแสดงดังรูปที่ 6

$$FFT = f(S_N) \tag{4}$$

$$FFT = \{ (f, A) \mid f \geq 0, A \geq 0 \} \tag{5}$$

โดยที่ *FFT* คือ เซตของคู่อันดับความถี่ $f$ กับขนาดของสัญญาณ $A$ ที่ความถี่นั้น ซึ่ง $f$ และ $A$ ต้องมีค่าไม่น้อยกว่า 0, โดยเซตของ *FFT* ในกรณีนี้จะมีสมาชิก 20 คู่อันดับ




**รูปที่ 6** กราฟแสดงข้อมูลเชิงความถี่ของเสียง ‘อา’ ที่วินาทีที่ 4.9 – 5.0

*2.1.6 การแบ่งข้อมูลตามช่วงความถี่ (Binning in frequency domain)*

เสียงของมนุษย์ส่วนใหญ่จะมีความถี่หลักอยู่ในช่วงความถี่ประมาณ 60 – 1,500 Hz และความถี่ที่ไวต่อการได้ยินจะอยู่ในช่วงความถี่ประมาณ 1,000 – 4,000 Hz ดังนั้นโปรแกรมที่พัฒนาขึ้นจึงกำหนดให้ใช้ข้อมูลเชิงความถี่ เฉพาะที่อยู่ในช่วง 50 - 3,000 Hz [12] สำหรับการคำนวณ (การใช้ข้อมูลความถี่ตั้งแต่ 50 Hz ขึ้นไป เป็นการกรองสัญญาณเบื้องต้นเพื่อกรองสัญญาณรบกวนจากระบบไฟฟ้า และสัญญาณกระแสตรงออก) ข้อมูลเชิงความถี่นี้จะถูกนำมาแบ่งเป็นกลุ่มย่อย (Binning) สมการที่ (5 – 14) เพื่อจะกำหนดเป็นลักษณะเฉพาะใช้สำหรับการจำแนกข้อมูล โดยการแบ่งข้อมูลจะแบ่งออกเป็น 10 ส่วน แต่ละส่วนจะมีความกว้าง 300 Hz (ยกเว้นส่วนแรกซึ่งมีช่วงกว้าง 250 Hz เนื่องจากตัดผลของกระแสตรงและความถี่ต่ำกว่า 50 Hz ออก) โดยรูปที่ 7 ได้แสดงภาพตัวอย่างการแบ่งข้อมูลเชิงความถี่เป็นช่วงต่าง ๆ ทั้ง 10 ส่วน

$$FFT_{50\text{ - }300} = \{ A \mid (f, A) \in FFT, 50 < f \leq 300 \} \tag{5}$$

$$FFT_{300\text{ - }600} = \{ A \mid (f, A) \in FFT, 300 < f \leq 600 \} \tag{6}$$

$$FFT_{600\text{ - }900} = \{ A \mid (f, A) \in FFT, 600 < f \leq 900 \} \tag{7}$$

$$FFT_{900\text{ - }1200} = \{ A \mid (f, A) \in FFT, 900 < f \leq 1200 \} \tag{8}$$

$$FFT_{1200\text{-}1500} = \{ A \mid (f,A) \in FFT, 1200 < f \leq 1500 \} \tag{9}$$

$$FFT_{1500\text{-}1800} = \{ A \mid (f,A) \in FFT, 1500 < f \leq 1800 \} \tag{10}$$

$$FFT_{1800\text{-}2100} = \{ A \mid (f,A) \in FFT, 1800 < f \leq 2100 \} \tag{11}$$

$$FFT_{2100\text{-}2400} = \{ A \mid (f,A) \in FFT, 2100 < f \leq 2400 \} \tag{12}$$

$$FFT_{2400\text{-}2700} = \{ A \mid (f,A) \in FFT, 2400 < f \leq 2700 \} \tag{13}$$

$$FFT_{2700\text{-}3000} = \{ A \mid (f,A) \in FFT, 2700 < f \leq 3000 \} \tag{14}$$

ข้อมูลแต่ละส่วนที่ถูกแบ่งออกมาจะถูกทำให้อยู่ในรูปของผลรวมของสมาชิกในเซต ตามสมการที่ (15 – 24)

$$bin_1 = \bigcup FFT_{50-300} \quad (15)$$

$$bin_2 = \bigcup FFT_{300-600} \quad (16)$$

$$bin_3 = \bigcup FFT_{600-900} \quad (17)$$

$$bin_4 = \bigcup FFT_{900-1200} \quad (18)$$

$$bin_5 = \bigcup FFT_{1200-1500} \quad (19)$$

$$bin_6 = \bigcup FFT_{1500-1800} \quad (20)$$

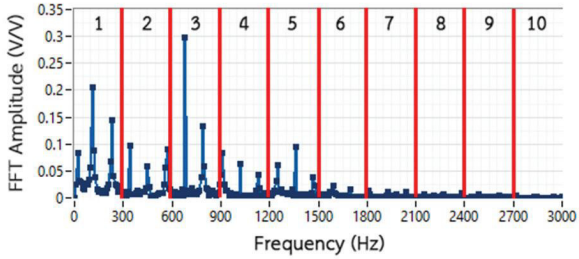
$$bin_7 = \bigcup FFT_{1800-2100} \quad (21)$$

$$bin_8 = \bigcup FFT_{2100-2400} \quad (22)$$

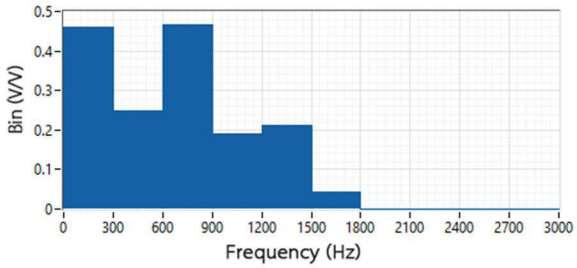
$$bin_9 = \bigcup FFT_{2400-2700} \quad (23)$$

$$bin_{10} = \bigcup FFT_{2700-3000} \quad (24)$$

สมการที่ (15) แสดงถึงผลรวมของสมาชิกในเซต $FFT_{50-300}$ ซึ่งเป็นการรวมกันของค่าแอมพลิจูดในช่วงความถี่ตั้งแต่ 50 ถึง 300 Hz โดยผลรวมของแต่ละเซตจะสามารถนำมาแสดงในลักษณะของแผนภูมิแท่งได้ดังรูปที่ 8 และเมื่อได้ค่าผลรวมทั้ง 10 ค่า ($bin_1$ - $bin_{10}$) แล้ว ค่าเหล่านี้จะถูกนำมากำหนดเป็นคุณลักษณะเฉพาะ ที่จะใช้สำหรับเปรียบเทียบข้อมูลเสียงกับฐานข้อมูลที่ได้มีการบันทึกไว้ เพื่อจะสามารถระบุได้ว่าข้อมูลเสียงที่ได้รับจากไมโครโฟนนั้นมีคุณสมบัติใกล้เคียงกับฐานข้อมูลเสียงกลุ่มใด




**รูปที่ 7** การแบ่งข้อมูลเชิงความถี่ออกเป็น 10 ส่วน




**รูปที่ 8** กราฟแท่งแสดงผลรวมของสมาชิกในแต่ละเซต *FFT*

*2.1.7 การจำแนกกลุ่ม (Classification)*

ในเสียงแต่ละเสียงจะมีค่า $bin_1$ - $bin_{10}$ ที่มีความแตกต่างกัน ซึ่งบางเสียงอาจจะสามารถสังเกตเห็นความแตกต่างจากกราฟสัญญาณในโดเมนความถี่ได้ด้วยตาเปล่า แต่บางเสียงจะมีลักษณะที่ใกล้เคียงกันมาก เช่น เสียง ‘โอ’ กับ ‘อู’ และแต่ละคนก็มีเสียงที่แตกต่างกันแม้จะออกเสียงเดียวกัน ดังนั้นการจะระบุข้อมูลเสียงที่รับมาว่าจัดอยู่ในกลุ่มเสียงใดอย่างชัดเจนนั้นจำเป็นต้องใช้กระบวนการจำแนกข้อมูลซึ่งมีอยู่หลายวิธีด้วยกัน เช่น วิธี K-Nearest Neighbor (KNN) และ Support Vector Machine (SVM) เป็นต้น สำหรับงานวิจัยนี้เลือกใช้วิธีแบบ KNN ซึ่งมีข้อดีคือพัฒนาได้ง่าย และสามารถนำมาใช้สำหรับการจำแนกกลุ่มข้อมูลของเสียงพูดได้ [13]

การจำแนกกลุ่มด้วยวิธี KNN จะเป็นการนำคุณสมบัติของข้อมูลที่ต้องการจัดกลุ่มและฐานข้อมูลมาเทียบกัน เพื่อหากลุ่มข้อมูลใดจากฐานข้อมูลที่มีคุณสมบัติใกล้เคียงกับข้อมูลที่ทดสอบมากที่สุด ซึ่งผลจากการเปรียบเทียบข้อมูลแต่ละตัวนั้นจะถูกนำมาเรียงลำดับความใกล้เคียงจากมากไปน้อย โดยจะเลือกความใกล้เคียงที่มากที่สุดจำนวน K ค่าแรก เพื่อจะระบุว่าข้อมูลที่ต้องการจัดกลุ่มนั้นมีคุณสมบัติ ใกล้กลุ่มข้อมูลใดที่สุด

เทคนิค KNN ประกอบด้วยกระบวนการหลัก 3 ขั้นตอน ได้แก่ การกำหนดเงื่อนไขหรือลักษณะเฉพาะที่ต้องการจัดกลุ่มหรือเปรียบเทียบ การคำนวณหาค่าระยะห่าง (เพื่อหาว่าข้อมูลมีความใกล้เคียงกับฐานข้อมูลกลุ่มใดมากที่สุด) แล้วระบุกลุ่มข้อมูลจากผลจากเปรียบเทียบระยะห่างจำนวน K ค่า

การกำหนดเงื่อนไขหรือลักษณะเฉพาะที่ต้องการจัดกลุ่มในงานวิจัยนี้มีการกำหนดเป็น 7 ค่า (7 Attribute) ลักษณะเฉพาะ 5 ค่าแรกจะกำหนดให้มีค่าเท่ากับ bin ที่ 1, 2, 3, 4 และ 5 ตามลำดับ ซึ่งแสดงในสมการที่ (25)

$$Attribute_i = bin_i \quad ; i = 1, 2, 3, 4, 5 \quad (25)$$

และเนื่องจาก $bin_6$ ถึง $bin_{10}$ มีค่าน้อยเพราะเป็นย่านความถี่สูง ลักษณะเฉพาะที่ 6 ($Attribute_6$) จึงถูกกำหนดให้เป็นผลรวมของ $bin_6$ และ $bin_7$ ส่วนในลักษณะเฉพาะที่ 7 จะเป็นผลรวมของ $bin_8$, $bin_9$, และ $bin_{10}$ ซึ่งแสดงตามสมการที่ (26–27) ตามลำดับ

$$Attribute_6 = bin_6 + bin_7 \quad (26)$$

$$Attribute_7 = bin_8 + bin_9 + bin_{10} \quad (27)$$

**ตารางที่ 1** ตัวอย่างค่าลักษณะเฉพาะของผู้ทดสอบท่านหนึ่ง

| Voice | A (เอ) | E (อี) | O (โอ) | R (อา) | U (อู) |
|---|---|---|---|---|---|
| *Att1* | 0.720 | 0.620 | 0.390 | 0.460 | 0.585 |
| *Att2* | 0.500 | 0.471 | 0.565 | 0.249 | 0.416 |
| *Att3* | 0.132 | 0.017 | 0.389 | 0.465 | 0.239 |
| *Att4* | 0.023 | 0.000 | 0.071 | 0.191 | 0.022 |
| *Att5* | 0.000 | 0.000 | 0.013 | 0.211 | 0.000 |
| *Att6* | 0.144 | 0.000 | 0.000 | 0.042 | 0.000 |
| *Att7* | 0.080 | 0.167 | 0.007 | 0.000 | 0.000 |

ตารางที่ 1 ได้แสดงตัวอย่างการคำนวณลักษณะเฉพาะของเสียง ‘เอ’ ‘อี’ ‘โอ’ ‘อา’ และ ‘อู’ ชุดหนึ่งจากฐานข้อมูลจำนวน 200 ชุด ของผู้ทดสอบท่านหนึ่ง จะสังเกตได้ว่าลักษณะเฉพาะของแต่ละเสียงนั้นบางเสียง ค่อนข้างมีค่าที่แตกต่างกันอย่างชัดเจน เช่นหากผู้ทดสอบนี้ออกเสียง ‘อา’ เมื่อนำสัญญาณ มาคำนวณโดยส่วนใหญ่จะมีค่า Att1-Att6 มากกว่าศูนย์ทุกตัว เมื่อเทียบกับเสียง ‘อู’ ซึ่งมักจะมีค่า Att5 และ Att6 เป็นศูนย์ ดังนั้นจึงพอจะแยกระหว่างเสียง ‘อู’ และ ‘อา’ จากค่า Att1-Att7 ได้ง่าย ทั้งนี้เมื่อพิจารณาจากตารางจะพบว่าการแยกระหว่างเสียง ‘เอ’ และ ‘อี’ จะทำได้ยากกว่าเนื่องจากมีค่า Att1-Att7 ค่อนข้างใกล้เคียงกัน นอกจากนี้ข้อมูลลักษณะเฉพาะของเสียงในชุดฐานข้อมูลที่บันทึกไว้ต่าง ๆ ก็ยังมีค่าแปรเปลี่ยนไปบ้างทุกค่า ซึ่งเป็นความไม่แน่นอนที่เกิดขึ้นได้ทั้งจากตัวผู้พูดเองและสัญญาณรบกวนจากสิ่งแวดล้อม ดังนั้นในการแยกแยะว่าสัญญาณเสียงที่ได้รับมาเป็นเสียงใดจะต้องคำนวณค่าลักษณะเฉพาะ แล้วนำไปเทียบกับฐานข้อมูลทั้งหมด ซึ่งต้องอาศัยเทคนิคการจำแนกเพื่อช่วยให้การจำแนกกลุ่มมีความถูกต้องและแม่นยำมากขึ้น

วิธีการจำแนกด้วยเทคนิค KNN นั้นจะใช้การเปรียบเทียบลักษณะเฉพาะกับข้อมูลหลาย ๆ กลุ่ม เพื่อหาความใกล้เคียงกันว่าข้อมูลที่ต้องการทดสอบนั้นอยู่ใกล้กับฐานข้อมูลกลุ่มใดที่สุด โดยความใกล้เคียงระหว่างข้อมูลจะสามารถหาได้โดยการหาความแตกต่างระหว่างลักษณะเฉพาะประเภทเดียวกันของข้อมูลทดสอบกับฐานข้อมูลที่มีอยู่ ค่าความแตกต่างของคุณสมบัติหรือลักษณะเฉพาะจะถูกเรียกว่า *“ระยะห่าง (Distance)”* โดยคุณสมบัติของระยะห่างนั้นจะประกอบด้วย

1. เมื่อข้อมูลทดสอบกับข้อมูลที่มีอยู่นั้นเป็นข้อมูลเดียวกัน ระยะห่างที่ได้ต้องเท่ากับศูนย์
2. ค่าระยะห่างจะต้องไม่ติดลบ

ในการหาระยะห่างนั้นสามารถหาได้หลายวิธี แต่สำหรับโปรแกรมวิเคราะห์สัญญาณเสียงนี้ได้เลือกใช้ การหาค่าสมบูรณ์ของผลต่างต่อค่าเฉลี่ย ซึ่งแสดงในสมการที่ (28)

$$D_i = \left| \frac{F_t - F_{db,i}}{\bar{F}_{db}} \right| \quad (28)$$

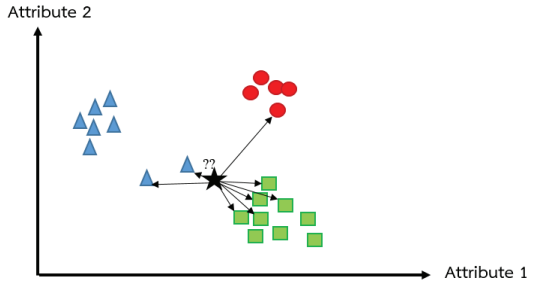
โดยที่ $F_t$ คือ ข้อมูลทดสอบที่ต้องการจัดกลุ่ม, $F_{db,i}$ คือ ข้อมูลที่ $i$ ใด ๆ ซึ่งเป็นฐานข้อมูลที่บันทึกไว้ทั้งหมดทุกเสียง, $\bar{F}_{db}$ คือ ค่าเฉลี่ยของฐานข้อมูลทั้งหมด และ $D_i$ คือ ระยะห่างของข้อมูลทดสอบกับข้อมูลตัวที่ $i$ ของฐานข้อมูลที่บันทึกไว้

เนื่องจากคุณสมบัติที่นำมาใช้ในการเปรียบเทียบมีอยู่หลายค่า จึงต้องรวมระยะห่างของแต่ละคุณสมบัติเข้าด้วยกันด้วยวิธี Euclidean distance ซึ่งเป็นการหารากที่สอง (Square Root) ของผลรวมระยะห่างในแต่ละคุณสมบัติยกกำลังสองตามสมการที่ (29) ซึ่งวิธีนี้จะมีลักษณะคล้ายกับการหาขนาดเวกเตอร์รวมจากเวกเตอร์ในแต่ละมิตินั่นเอง

$$D_S = \sqrt{\sum_{i=1}^{n} D_i^2} \quad (29)$$

โดยที่ $D_S$ คือ ผลรวมของระยะห่าง, $D_i$ คือ ระยะห่างของคุณสมบัติที่ $i$ ใด ๆ และ $n$ คือ จำนวนของมิติ

ส่วนสุดท้ายของการจำแนกข้อมูลคือการระบุข้อมูลที่มีความใกล้เคียงกันมากที่สุด โดยการนำค่าระยะห่างที่ได้จากการเปรียบเทียบมาเรียงลำดับจากน้อยไปมาก แล้วเลือกค่าระยะห่างจำนวน 15 ค่าแรก (กำหนดค่า K เป็น 15) มานับจำนวนประเภทข้อมูลซึ่งเป็นประเภทเดียวกัน ฐานข้อมูลที่มีจำนวนค่าระยะห่างในจำนวนค่าระยะห่างที่เลือกมามากที่สุด ก็จะสามารถระบุได้ว่าข้อมูลทดสอบนั้นมีความใกล้เคียงกับฐานข้อมูลนั้นมากที่สุด




**รูปที่ 9** การหาระยะห่างของข้อมูลทดสอบกับข้อมูลที่มีอยู่โดยใช้ลักษณะเฉพาะที่ 1 และ 2

รูปที่ 9 แสดงตัวอย่างการเปรียบเทียบข้อมูลทดสอบกับกลุ่มข้อมูลที่มีอยู่ โดยใช้ลักษณะเฉพาะสองด้าน จากรูปจะสังเกตได้ว่าข้อมูลทดสอบ (รูปดาว) มีความใกล้เคียงกับข้อมูลสามเหลี่ยมมากกว่าข้อมูลสี่เหลี่ยม แต่เมื่อเลือกข้อมูลที่ใกล้ที่สุดจำนวน 8 ค่า จะพบว่าข้อมูลสี่เหลี่ยมมีจำนวนมากที่สุด จึงระบุได้ว่าข้อมูลทดสอบนั้นควรจัดอยู่ในกลุ่มข้อมูลสี่เหลี่ยมเนื่องด้วยลักษณะเฉพาะที่ 1 ($Attribute_1$) และที่ 2 ($Attribute_2$) ที่ค่า K = 8

## 3. ผลการทดลอง

การทดสอบโปรแกรมวิเคราะห์สัญญาณเสียงนี้เป็นการศึกษานำร่องเพื่อประเมินความถูกต้องของระบบ โดยการนำเสียงที่บันทึกจากผู้ทดสอบจำนวน 9 ท่าน มาใช้ในการทดสอบ โดยแต่ละคนจะบันทึก 5 เสียง ได้แก่ เสียง ‘เอ’ ‘อี’ ‘โอ’ ‘อา’ และ ‘อู’ แล้วข้อมูลเสียงเหล่านี้จะถูกนำมาคำนวณเป็นข้อมูลลักษณะเฉพาะ เนื่องจากข้อมูลเสียงระยะเวลา 0.1 วินาที จะสามารถคำนวณได้ลักษณะเฉพาะหนึ่งชุด (ซึ่งประกอบไปด้วยลักษณะเฉพาะ 7 ค่า) ดังนั้นผู้ทดสอบจะออกเสียงเป็นเวลา 5 วินาที ทั้งหมด 5 ครั้ง เพื่อจะให้ได้ข้อมูลลักษณะเฉพาะ 250 ชุด โดยข้อมูลตั้งแต่ชุดที่ 1 - 200 (จำนวน 200 ชุดข้อมูล) จะใช้เป็นฐานข้อมูลสำหรับการวิเคราะห์ และข้อมูลตั้งแต่ชุดที่ 201 – 250 (จำนวน 50 ชุดข้อมูล) จะใช้เป็นข้อมูลสำหรับทดสอบกับฐานข้อมูลเพื่อหาความถูกต้องของการวิเคราะห์ของโปรแกรมวิเคราะห์สัญญาณเสียง

การทดสอบความถูกต้องในงานวิจัยนี้แบ่งออกเป็น 4 การทดสอบ โดยแต่ละการทดสอบจะเปลี่ยนจำนวนเสียงของฐานข้อมูลเพื่อหาเสียงและจำนวนฐานข้อมูลที่มีความเหมาะสมที่สุด โดยเสียงที่ใช้เป็นฐานข้อมูลในแต่ละการทดสอบจะแสดงในตารางที่ 2

การทดสอบที่ 1 จะใช้จำนวนฐานข้อมูล 2 เสียง จาก 5 เสียง ซึ่งจะได้ทั้งหมด 10 รูปแบบ

การทดสอบที่ 2 จะใช้จำนวนฐานข้อมูล 3 เสียง จาก 5 เสียง ซึ่งจะได้ทั้งหมด 10 รูปแบบ

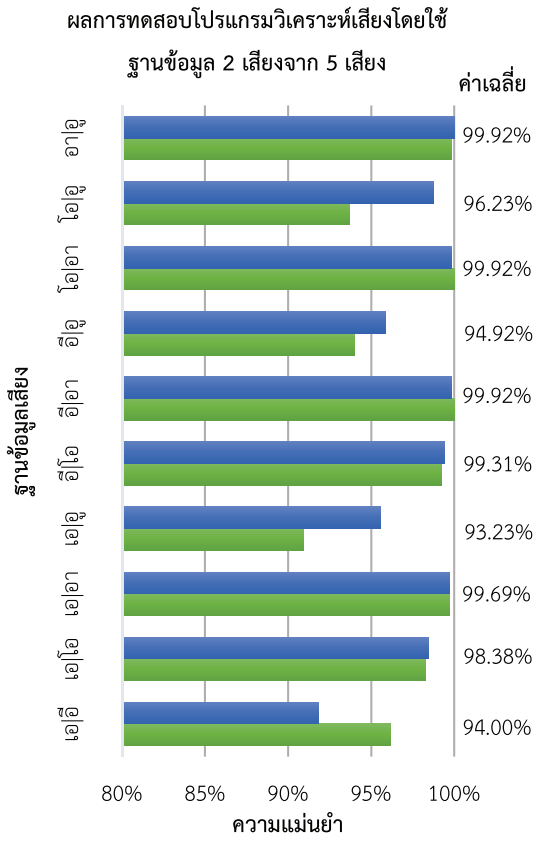
การทดสอบที่ 3 จะใช้จำนวนฐานข้อมูล 4 เสียง จาก 5 เสียง ซึ่งจะได้ทั้งหมด 5 รูปแบบ

การทดสอบที่ 4 จะใช้จำนวนฐานข้อมูลทั้ง 5 เสียง ในการวิเคราะห์

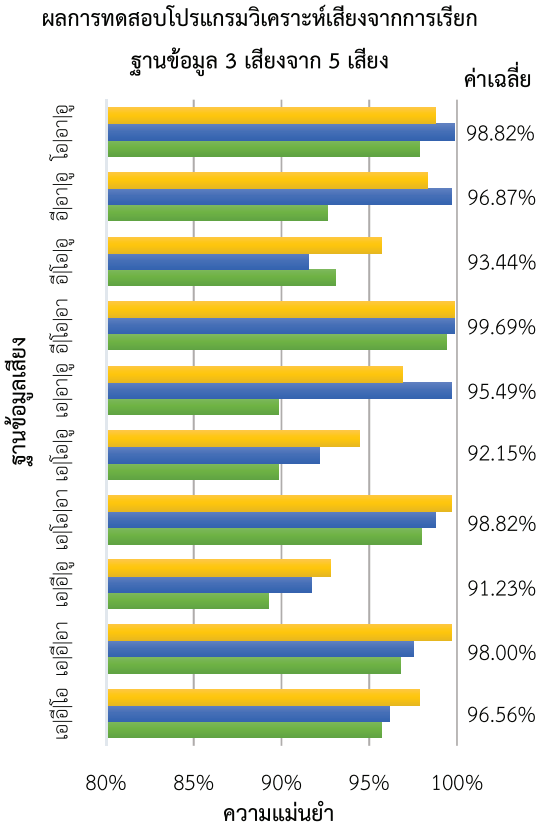
จากการทดสอบที่ 1 ดังแสดงในรูปที่ 10 พบว่าโปรแกรมสามารถวิเคราะห์เสียงด้วยการใช้ฐานข้อมูลจำนวน 2 เสียงได้อย่างถูกต้อง โดยการวิเคราะห์ข้อมูลเสียงของทั้ง 9 ท่าน มีค่าความแม่นยำเฉลี่ยอยู่ในช่วงร้อยละ 91 ถึง 100 จะเห็นได้ว่าความแม่นยำที่มีค่าเฉลี่ยน้อยที่สุดเมื่อใช้ฐานข้อมูลเสียง “เอ อู” แต่ถ้าเลือกใช้เสียง “อา อู”, “โอ อา” และ “อี อา” จะได้ความถูกต้องถึงร้อยละ 99 โดยหลักการสร้างรหัสควบคุมจากการเลือกเสียงสองเสียงเช่น “อา อู” ไปใช้ตัวอย่างเช่น หากใช้รหัสสองคำ จะได้รหัสควบคุมที่แตกต่างกัน 4 แบบ คือ อาอา, อาอู, อูอา และ อูอู และหากใช้รหัสสามคำ จะได้รหัสควบคุมที่แตกต่างกัน 8 แบบ (คำนวณจาก 2x2x2)

**ตารางที่ 2** รูปแบบการเลือกฐานข้อมูลเสียงสำหรับการทดสอบ

| แบบที่ | การทดสอบที่ 1 | การทดสอบที่ 2 | การทดสอบที่ 3 | การทดสอบที่ 4 |
|---|---|---|---|---|
| *1* | ‘เอ’ ‘อี’ | ‘เอ’ ‘อี’ ‘โอ’ | ‘เอ’ ‘อี’ ‘โอ’ ‘อา’ | ‘เอ’ ‘อี’ ‘โอ’ ‘อา’ ‘อู’ |
| *2* | ‘เอ’ ‘โอ’ | ‘เอ’ ‘อี’ ‘อา’ | ‘เอ’ ‘อี’ ‘โอ’ ‘อู’ | |
| *3* | ‘เอ’ ‘อา’ | ‘เอ’ ‘อี’ ‘อู’ | ‘เอ’ ‘อี’ ‘อา’ ‘อู’ | |
| *4* | ‘เอ’ ‘อู’ | ‘เอ’ ‘โอ’ ‘อา’ | ‘เอ’ ‘โอ’ ‘อา’ ‘อู’ | |
| *5* | ‘อี’ ‘อา’ | ‘เอ’ ‘โอ’ ‘อู’ | ‘อี’ ‘โอ’ ‘อา’ ‘อู’ | |
| *6* | ‘อี’ ‘โอ’ | ‘เอ’ ‘อา’ ‘อู’ | | |
| *7* | ‘อี’ ‘อู’ | ‘อี’ ‘โอ’ ‘อา’ | | |
| *8* | ‘โอ’ ‘อา’ | ‘อี’ ‘โอ’ ‘อู’ | | |
| *9* | ‘โอ’ ‘อู’ | ‘อี’ ‘อา’ ‘อู’ | | |
| *10* | ‘อา’ ‘อู’ | ‘โอ’ ‘อา’ ‘อู’ | | |



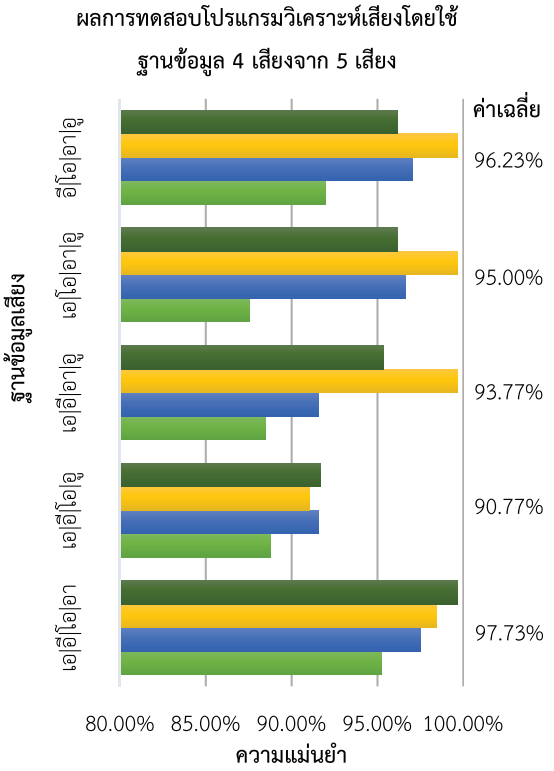

**รูปที่ 10** ผลการทดสอบความแม่นยำของโปรแกรม
วิเคราะห์เสียงด้วยการใช้ฐานข้อมูล 2 เสียง (การทดสอบที่ 1)




**รูปที่ 11** ผลการทดสอบความแม่นยำของโปรแกรม
วิเคราะห์เสียงด้วยการใช้ฐานข้อมูล 3 เสียง (การทดสอบที่ 2)

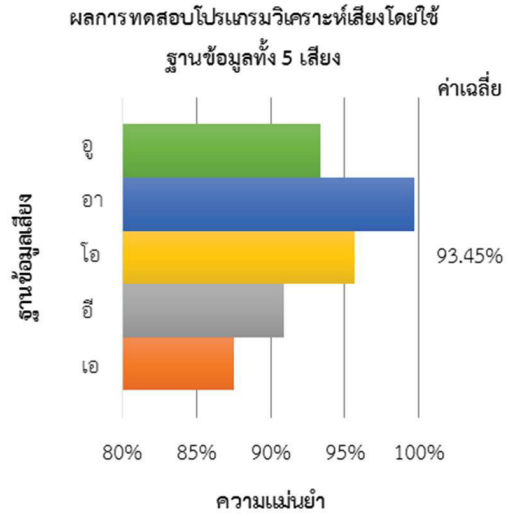
การทดสอบที่ 2 เป็นการทดสอบโปรแกรมวิเคราะห์เสียงด้วยการใช้ฐานข้อมูลจำนวน 3 เสียง ซึ่งพบว่าความแม่นยำเฉลี่ยของการวิเคราะห์เสียงของผู้ทดสอบ 9 ท่าน มีค่าอยู่ในช่วง ร้อยละ 89.23 ถึง 99.85 โดยมีความเฉลี่ยน้อยที่สุดเมื่อใช้ฐานข้อมูลเสียง “เอ อี อู” (แสดงดังรูปที่ 11) ซึ่งน้อยกว่าการทดสอบโดยใช้ฐานข้อมูลเสียงเพียง 2 เสียงเพียงเล็กน้อย และหากเลือกใช้ฐานข้อมูล “อี โอ อา” จะได้ความถูกต้องมากกว่าร้อยละ 99 โดยหลักการสร้างรหัสควบคุมจากการเลือกเสียงสามเสียงเช่น “อี โอ อา” ไปใช้ตัวอย่างเช่น หากใช้รหัสสองคำจะได้รหัสควบคุมที่แตกต่างกัน 9 แบบ (คำนวณจาก 3x3) และหากใช้รหัสสามคำ จะได้รหัสควบคุมที่แตกต่างกัน 27 แบบ (คำนวณจาก 3x3x3)

การทดสอบที่ 3 เป็นการทดสอบโปรแกรมวิเคราะห์เสียงซึ่งใช้ฐานข้อมูลจำนวน 4 เสียง ดังแสดงในรูปที่ 12 พบว่ามีค่าความแม่นยำเฉลี่ยของผู้ทดสอบ 9 ท่าน มีค่าอยู่ในช่วง ร้อยละ 87.54 ถึง 99.69 ซึ่งน้อยกว่าการทดสอบโดยใช้ฐานข้อมูลจำนวน 2 เสียงและจำนวน 3 เสียงเพียงเล็กน้อย โดยกลุ่มฐานข้อมูลเสียงที่มีความแม่นยำเฉลี่ยต่ำที่สุดคือ “เอ อี โอ อู” ซึ่งมีค่าร้อยละ 90.77 จากรูปแสดงได้อีกว่าเสียง ‘เอ’ เป็นเสียงที่เมื่อนำมาทดสอบแล้วจะมีความแม่นยำน้อยที่สุดในทุกกลุ่มฐานข้อมูลเสียง สำหรับการเลือกกลุ่มฐานข้อมูลที่เหมาะสมกับการนำไปใช้นั้น ควรเลือกใช้ฐานข้อมูลเสียง “เอ อี โอ อา” เพราะจะให้ผลการวิเคราะห์ที่ถูกต้องมากกว่าร้อยละ 95 ในทุกเสียง และมีความแม่นยำเฉลี่ยมากที่สุด




**รูปที่ 12** ผลการทดสอบความแม่นยำของโปรแกรมวิเคราะห์เสียงด้วยการใช้ฐานข้อมูล 4 เสียง (การทดสอบที่ 3)

ในการทดสอบที่ 4 โปรแกรมวิเคราะห์เสียงซึ่งใช้ฐานข้อมูลจำนวน 5 เสียง (รูปที่ 13) พบว่าโปรแกรมสามารถวิเคราะห์เสียงได้อย่างถูกต้องโดยมีค่าความแม่นยำเฉลี่ยอยู่ในช่วงร้อยละ 87.54 ถึง 99.69 โดยเสียง ‘เอ’ มีค่าความแม่นยำเฉลี่ยต่ำที่สุดที่ ร้อยละ 87.54 และเสียง ‘อา’ มีความแม่นยำเฉลี่ยสูงที่สุดที่ร้อยละ 99.69




**รูปที่ 13** ผลการทดสอบความแม่นยำของโปรแกรมวิเคราะห์เสียงด้วยการใช้ฐานข้อมูล 5 เสียง (การทดสอบที่ 4)

## 4. สรุปผลการทดลอง

จากผลการทดสอบโปรแกรมวิเคราะห์สัญญาณเสียง ทำให้เห็นว่าความแม่นยำการวิเคราะห์สัญญาณเสียงของโปรแกรมมีแนวโน้มที่จะลดลงเล็กน้อย เมื่อฐานข้อมูลจำนวนเสียงที่นำมาใช้ มีจำนวนมากขึ้นและยังพบอีกว่าเสียงที่มีความแม่นยำมากที่สุดคือเสียง ‘อา’ และเสียงส่วนใหญ่ที่มีความแม่นยำน้อยคือเสียง ‘เอ’ แต่ทั้งนี้ในทุกการทดสอบ ก็ยังมีค่าความแม่นยำของเสียง ‘เอ’ มากกว่าร้อยละ 85 ผลการทดสอบนี้จึงแสดงให้เห็นว่าโปรแกรมสามารถวิเคราะห์สัญญาณเสียงได้อย่างแม่นยำและมีความถูกต้องสูงในทุกฐานข้อมูล โดยความผิดพลาดที่เกิดขึ้นอาจจะเกิดจากที่เสียงของผู้ทดสอบบางเสียง เช่น ‘เอ’ และ ‘อี’ ซึ่งอาจมีความคล้ายคลึงกันมาก ทำให้โปรแกรมคำนวณหาลักษณะเฉพาะได้ค่าที่ใกล้เคียงกันซึ่งทำให้การวิเคราะห์เกิดความคลาดเคลื่อนได้ ทั้งนี้หากเลือกเสียงที่เหมาะสมกับผู้ใช้นั้น ๆ จะสามารถทำให้ระบบมีความถูกต้องสูงขึ้น สำหรับผลการทดลองโดยรวมนั้น ในกรณีที่เลือกฐานข้อมูลสองเสียงที่เหมาะสม เช่น “อา โอ” “อา อู” และ “อี อา” จะสามารถให้ผลที่ถูกต้องถึงร้อยละ 99 ในกรณีที่เลือกกลุ่มละสามเสียงที่เหมาะสมเช่น “อี โอ อา” มาใช้ในการสร้างรหัสควบคุมจะให้ผลการตรวจจับที่ถูกต้องถึงร้อยละ 99 โดยผลจากการทดสอบโปรแกรมวิเคราะห์สัญญาณเสียงในกลุ่มตัวอย่างขนาดเล็กนี้ ให้ข้อสรุปได้ว่า โปรแกรมที่พัฒนาขึ้นมีประสิทธิภาพสูงและมีศักยภาพในการนำไปใช้งานจริงในบุคคลทั่วไปรวมทั้งผู้ป่วยติดเตียง ทั้งนี้ควรศึกษาและทดลองเพิ่มเติมในกลุ่มตัวอย่างขนาดใหญ่ขึ้นเพื่อให้สามารถประเมินความถูกต้องและความไม่แน่นอนของระบบในการนำไปใช้ในกลุ่มผู้ใช้ต่าง ๆ ได้

## 5. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากสำนักงานคณะกรรมการอุดมศึกษา (สกอ) ประจำปีงบประมาณ 2559

## 6. เอกสารอ้างอิง

[1] Foundation of Thai Gerontology Research and Development Institute. (2014). The Situation of the Elderly in Thailand in 2014. Retrieved from http://thaitgri.org/?p=37626

[2] Kumar, V., Singh, A., Tewari, M. K., & Kaur, S. (2016). Comprehension and compliance with the discharge advice and quality of life at home among the postoperative neurosurgery patients discharged from PGIMER,

Chandigarh, India. *Asian Journal of Neurosurgery*, *11*(4), 372–377. https://doi.org/10.4103/1793-5482.144190

[3] Takemasa, S., Nakagoshi, R., Murakami, M., Uesugi, M., Inoue, Y., Gotou, M., ... Naruse, S. (2014). Factors Affecting Quality of Life of the Homebound Elderly Hemiparetic Stroke Patients. *Journal of Physical Therapy Science*, *26*(2), 301–303. https://doi.org/10.1589/jpts.26.301

[4] Ruíz-Serrano, A., Posada-Gómez, R., Sibaja, A. M., Rodríguez, G. A., Gonzalez-Sanchez, B. E., & Sandoval-Gonzalez, O. O. (2013). Development of a Dual Control System Applied to a Smart Wheelchair, using Magnetic and Speech Control. *Procedia Technology*, *7*, 158–165. https://doi.org/10.1016/j.protcy.2013.04.020

[5] Sasou, A., & Kojima, H. (2009). Noise robust speech recognition applied to voice-driven wheelchair. *Eurasip Journal on Advances in Signal Processing*, *2009*. https://doi.org/10.1155/2009/512314

[6] Falk, T. H., Andrews, A., Hotzé, F., Wan, E., & Chau, T. (2012). Evaluation of an ambient noise insensitive hum-based powered wheelchair controller. *Disability and Rehabilitation: Assistive Technology*, *7*(3), 242–248. https://doi.org/10.3109/17483107.2011.629330

[7] Šalna, B., & Kamarauskas, J. (2010). Evaluation of effectiveness of different methods in speaker recognition. *Elektronika Ir Elektrotechnika*, (2), 67–70. https://doi.org/10.5755/j01.eee.98.2.9928

[8] AnanthaNatarajan, V., & Jothilakshmi, S. (2012). Segmentation of Continuous Speech into Consonant and Vowel Units using Formant Frequencies. *International Journal of Computer Applications*, *56*(15), 24–27. https://doi.org/10.5120/8968-3185

[9] Kuwabara, H., & Ohgushis, K. (1987). Role of Formant Frequencies and Bandwidths in Speaker Perception, *Elecronics and Communication in Japan (Part I: Communications)*, *70(9)*, 11-21

[10] Kartakis, S., Sakkalis, V., Tourlakis, P., Zacharioudakis, G., & Stephanidis, C. (2012). Enhancing Health Care Delivery Through Ambient Intelligence Applications, *Sensors*, *12(9)*, 11435-11450

[11] Rabiner, L., & Juang, B. H. (2005). *Fundamental of Speech Recognition.* Delhi, India: Pearson Education.

[12] Wolfe, J., Garnier, M., & Smith, J. (2009). Voice acoustics: an introduction, *UNSW*. Retrieved from http://www.animations.physics.unsw.edu.au/jw/voice.html

[13] Lakshmi Priya, T., Raajan, N. R., Raju, N., Preethi, P., & Mathini, S. (2012). Speech and non-speech identification and classification using KNN algorithm. In *Procedia Engineering* (Vol. 38, pp. 952–958). Elsevier Ltd. https://doi.org/10.1016/j.proeng.2012.06.120