โปรแกรมควบคุมการเคลื่อนไหวหุ่นยนต์โดยใช้การรู้จำเสียงพูด

Program Control the Robot Motion Using Speech Recognition

ธนะศักดิ์ พันธ์ประสิทธิ์

ภาควิชาวิศวกรรมอิเล็กทรอนิกส์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยกรุงเทพ วิทยาเขตรังสิต ปทุมธานี

บทคัดย่อ

ในการควบคุมการเคลื่อนไหวของหุ่นยนต์เพื่อให้ได้ประสิทธิภาพสูงสุดต้องคำนึงถึงการตอบสนองต่อหน้าที่ และให้ความปลอดภัยสูงสุดตามที่ผู้ออกแบบได้กำหนดไว้ ในบทความนี้นำเสนอการออกแบบโปรแกรมควบคุมการเคลื่อนไหวหุ่นยนต์โดยการใช้คำสั่งเสียงพูดภาษาไทยแบบกำหนดเสียงพูด คือคำว่า ซ้าย ขวา หน้า หลังและหยุด โดยไม่ขึ้นกับผู้พูด เพื่อความสะดวกในการใช้งานทางปฏิบัติ หลักการทำงานของระบบประกอบด้วย 1) การประมวลผลสัญญาณเสียงพูดเบื้องต้น โดยใช้วิธีการหาค่าอัตราส่วนการตัดค่าศูนย์ 2) ใช้การแปลงเสียงพูดในโดเมนความถี่ ด้วยวิธีการแปลงฟูเรียร์อย่างเร็ว (Fast Fourier Transform, FFT) 3) รู้จำเสียงพูดโดยใช้โครงข่ายประสาทเทียม (Artificial Neural Network, ANN) ผลการทดสอบจากวิธีที่นำเสนอโดยใช้ผู้พูดเป็นผู้หญิงและผู้ชาย เพศละ 5 คน แต่ละคนใช้คำสั่งคำละ 5 ครั้ง มีค่าเปอร์เซ็นต์ความถูกต้องโดยเฉลี่ยของเพศหญิงและเพศชายเท่ากัน มีค่าเท่ากับ 68.00% ตามลำดับ

คำสำคัญ : การแปลงฟูเรียร์อย่างเร็ว โครงข่ายประสาทเทียม การรู้จำเสียงพูด

Abstract

In order to control the robot motion for high achievement performance of command functions and gives the highest safety belongs to it designed. This paper present the robot motion controlled by speech recognition programmed that command in Thai speech language such as “Sai”, “Khwa”, “Na”, “Lung” and “Yoot” that representation for turn left, turn right, forward, backward and stop respectively. The speech commands are independent from any users, so it can be gain benefit for generally use in practical. The three mainly parts of this paper are consists of preprocessing part, Fast Fourier Transform (FFT) and Back Propagation Neural Network (BPN). The preprocessing technique is based on zero crossing calculation for finding the start and stop of speech frequency range and after that the FFT is used to transform the speech signal in time domain to frequency domain. The BPN is used in supervise training in order to recognize the five speech commands based on Radial Basis Function (RBF). The experimental results are reviewed that the average accuracy percentage is equal to 68.00% both for men and women

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

commands. This result is strongly confirmed that this speech recognition technique is independent from any user and quite enough percentage of its accuracy command.

Keywords : FFT, ANN, Speech recognition

1. บทนำ

ในปัจจุบัน ระบบคอมพิวเตอร์และหุ่นยนต์ ได้เข้ามามีบทบาทในชีวิตประจำวันของมนุษย์มากขึ้น โดย เฉพาะอย่างยิ่งในสถานที่ที่มีสภาพของสิ่งแวดล้อมที่เป็นอันตราย เช่น การเก็บกู้ระเบิด การเข้าไปหาข้อมูลในสนามรบทางทหาร ในอาคารสถานที่เกิดเพลิงไหม้ เป็นต้น ถ้าเราสามารถควบคุมหุ่นยนต์ให้มีการเคลื่อนไหวไปตามคำสั่งสัญญาณเสียงพูดผ่านไมโครโฟนได้ ผลทำให้ลดความเสี่ยงและอันตรายต่อผู้ปฏิบัติงาน โดยระบบควบคุมใช้การติดต่อสื่อสารผ่านระบบไร้สาย (Wireless System) ระบบควบคุมประกอบด้วยระบบฮาร์ดแวร์ (Hardware System) และระบบโปรแกรม (Software System) ดังจะเห็นได้จากมีนักวิจัยจำนวนมากได้นำวิธีการรู้จำเสียงพูดไปควบคุมการเคลื่อนไหวหุ่นยนต์ เช่น [1] งานวิจัยนี้นำเสนอการควบคุมรถเข็นคนพิการด้วยเสียงร่วมกับการใช้เซ็นเซอร์ในการระบุและหลีกเลี่ยงอุปสรรคในเส้นทาง ผลการทดลองทำการเปรียบเทียบในกรณีทั้งที่มีและไม่มีการนำความช่วยเหลือมาช่วยผู้พิการ [2] นักวิจัยกลุ่มนี้ประยุกต์ใช้ เทคนิค Ella Voice ซึ่งก็คือเทคนิคที่ขึ้นอยู่กับเสียงของผู้พูดร่วมกับเทคนิค การข้ามคำ (cross-words) และการสร้างแม่แบบการอ้างอิงหรือการตรวจหาจุดสิ้นสุดของคำ รวมถึงโปรแกรมเขียนด้วย MATLAB [3] นักวิจัยกลุ่มนี้นำเสนอการรู้จำเสียงแบบ hands-free (human-robot) อัตราการรู้จำสำหรับคำของผู้ใช้เป็น 89.93% และ [4] ในงานวิจัยนี้นำเสนอเทคนิค Multiregdilet transform ใช้สำหรับการรู้จำเสียงพูดรวมถึงใช้โครงข่ายประสาทเทียมร่วมกับอินเตอร์เฟซฮาร์ดแวร์ชนิดพิเศษ ผลได้รับอัตราการรู้จำเสียงพูดเท่ากับ 98% จากงานวิจัยก่อนหน้านี้พบว่าการรู้จำรู้จำเสียงพูดสำหรับหุ่นยนต์ต้องใช้ระบบนำทางที่ขึ้นอยู่กับเซ็นเซอร์เพิ่มเติมเช่น อินฟราเรด กล้อง คลื่นเสียงความถี่สูง ฯลฯ

ในงานวิจัยนี้จึงได้นำเสนอระบบโปรแกรมบังคับการเคลื่อนไหวหุ่นยนต์โดยใช้การรู้จำเสียงพูดแบบไม่ขึ้นกับผู้พูด โดยหัวข้องานวิจัยประกอบด้วย 1. บทนำ 2. ทฤษฎีและหลักการ 3. ขั้นตอนการทำงานของระบบ 4. โครงข่ายประสาทเทียม (ANN) 5. การทดลอง และ 6. สรุปผล

2. ทฤษฎีและหลักการ

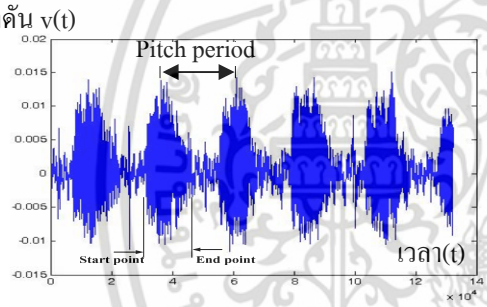
การบังคับการเคลื่อนไหวหุ่นยนต์โดยใช้การรู้จำเสียงพูด ในบทความฉบับนี้ใช้การวิเคราะห์สัญญาณเสียงพูดในโดเมนของความถี่ และใช้การคำนวณทางคณิตศาสตร์ฟูเรียร์อย่างเร็ว (FFT) เพื่อคำนวณหาค่าขนาด (Magnitude) และทิศทาง (Phase) ในลำดับต่อไปข้อมูลทิศทางจะไม่ถูกนำมาใช้ในการคำนวณ ทำการคำนวณหาความแตกต่างระหว่างค่าขนาดเสียงพูดของแต่ละคำด้วยคณิตศาสตร์โครงข่ายประสาทเทียม (ANN) ค่าความแตกต่างของแต่ละคำจะถูกแทนด้วยค่าน้ำหนัก (weight) และค่าไบแอส (Bias term) ค่าทั้งสองจะถูกนำมาใช้เป็นฐานข้อมูล (Data Base) ของเสียงพูดแต่ละคำ และในลำดับต่อไปจะกล่าวถึงการกำหนดตำแหน่งเริ่มต้นและสิ้นสุดของเสียงพูด

2.1 การกำหนดตำแหน่งเริ่มต้นและสิ้นสุดของเสียงพูด

ตัวอย่างของสัญญาณเสียงพูดคำว่า “ขวา” มีลักษณะองค์ประกอบดังแสดงในรูปที่ 1 ตำแหน่งที่มีค่ายอดสูงสุดเรียกว่า ตำแหน่งค่าพิตช์สูงสุด และระยะเวลาระหว่างตำแหน่งค่าพิตช์สูงสุดถึงตำแหน่งค่าพิตช์สูงสุดในลูกถัดไปจะเรียกว่า คาบเวลาพิตช์ (Pitch period) ในการตรวจจับตำแหน่งค่าพิตช์สูงสุดถึงตำแหน่งค่าพิตช์สูงสุดโดยการใช้หน้าต่างชนิดสีเหลี่ยม (Rectangular Window)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผ่านสัญญาณเสียงพูด พิจารณาให้ตำแหน่งค่าพิตช์สูงสุดถึงตำแหน่งค่าพิตช์สูงสุดในลูกถัดไปอยู่ที่ตำแหน่งกึ่งกลางของหน้าต่าง และจะต้องอยู่บนเงื่อนไขที่ว่า ความกว้างของหน้าต่างจะต้องมีค่ามากกว่าค่าคาบเวลาพิตช์และมีค่าน้อยกว่าสองเท่าของค่าคาบเวลาพิตช์ และเงื่อนไขในการบันทึกเสียงพูดใช้รูปแบบการบันทึกข้อมูลขนาด 8 บิต แม้ว่าค่าความถี่เสียงจะมีค่าอยู่ระหว่าง 20 Hz – 20,000 Hz ในงานวิจัยนี้ใช้ค่าอัตราการสุ่มของสัญญาณเท่ากับ 8,000 Hz เพราะว่าสัญญาณเสียงพูดที่มนุษย์เราสามารถรับฟังได้เข้าใจจะมีค่าอยู่ระหว่าง 20 Hz – 4,000 Hz เท่านั้น จากนั้นเสียงพูดจะถูกนำมาทำการปรับแต่งโดยการใช้การตัดจุดเริ่มต้น (Start point) – จุดสิ้นสุด (End point) จากนั้นนำเสียงพูดไปผ่านขั้นตอนของการปรับค่าบรรทัดฐานทางเวลาเพื่อที่จะทำให้เสียงพูดทุกคำมีช่วงคาบความยาวที่เท่ากัน



รูปที่ 1. ตัวอย่างเสียงพูดคำว่า “ขวา”

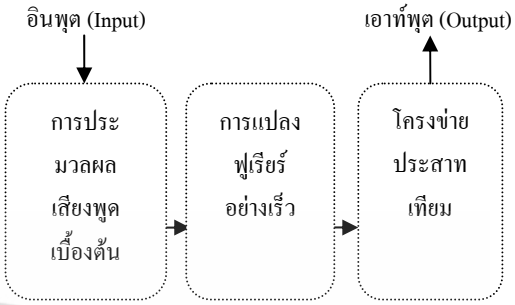
3. ขั้นตอนการทำงานของระบบ

ในการออกแบบและการเขียนโปรแกรมมีโครงสร้างของโปรแกรมบังคับการเคลื่อนไหวหุ่นยนต์โดยใช้การรู้จำเสียงพูด ประกอบด้วยส่วนหลักๆ ดังนี้ คือ 1) หน่วยการประมวลผลเสียงพูดเบื้องต้น 2) หน่วยการแปลงฟูเรียร์อย่างเร็ว (FFT) 3) หน่วยโครงข่ายประสาทเทียม (ANN) สำหรับประมวลผล และมีบล็อกไดอะแกรมของโปรแกรมดังแสดงในภาพที่ 2

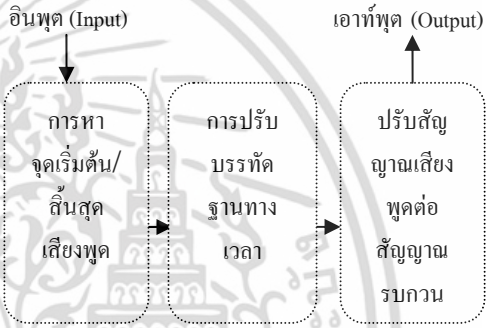
3.1 การประมวลผลเสียงพูดเบื้องต้น

ในขั้นตอนของการประมวลผลสัญญาณเสียงพูดเบื้องต้นมีขั้นตอนดังแสดงในรูปที่ 3.

การหาจุดเริ่มต้น/สิ้นสุดเสียงพูด ในงานวิจัยนี้ใช้วิธีการหาค่าอัตราส่วนการตัดค่าผ่านศูนย์ (Zero Cross) และค่าพลังงาน วิธีนี้เป็นวิธีที่สามารถหาจุดเริ่มต้น/สิ้นสุดเสียงพูดได้ถูกต้องมากกว่าวิธีอื่นๆ และในการจำลองการทำงานเขียนโปรแกรมด้วยภาษาแมทแลป (MATLAB)



รูปที่ 2. บล็อกไดอะแกรมการทำงานของระบบ



รูปที่ 3. การประมวลผลเสียงพูดเบื้องต้น

การปรับบรรทัดฐานทางเวลา เนื่องจากเสียงพูดคำว่า ซ้าย ขวา หน้า หลัง หยุด มีค่าความยาวไม่เท่ากัน ดังนั้นจึงต้องปรับบรรทัดฐานทางเวลาเสียงพูดของแต่ละคำให้มีค่าช่วงเวลาที่เท่ากัน และในงานวิจัยนี้ใช้การประมาณค่าในรูปแบบเชิงเส้น เพราะว่าจะทำให้ใช้เวลาในการประมวลผลน้อย และใช้หน่วยความจำจำกัด จากวิธีที่นำ เสนอแล้วนั้น ในงานวิจัยนี้ใช้สัญญาณเสียงพูดยาวเท่ากับ 8,000 จุดข้อมูล

การปรับสัญญาณเสียงพูดต่อสัญญาณรบกวน กรรมวิธีนี้จะมีผลต่อค่าประสิทธิภาพของสัญญาณเสียง พูด ผลทำให้ค่าสัญญาณเสียงพูดต่อสัญญาณรบกวนมีค่าสูงขึ้น ในงานวิจัยที่นำเสนอนี้ใช้วิธีการกรองแบบดิจิตอลอันดับหนึ่ง (First Order Digital Filter) ดังสมการ

$$H(z) = 1 - az^{-1} \qquad (1)$$

โดยที่ a คือค่าสัมประสิทธิ์ของตัวกรอง โดยทั่วไปแล้วจะมีค่าอยู่ระหว่าง 0.92 – 0.97 จากผลการทดลองเลือกค่า a = 0.94 เป็นค่าที่ให้ผลการทดลองดีที่สุด และในลำดับต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะกล่าวถึงขั้นตอนการแปลงสัญญาณเสียงพูด ให้เข้าสู่โดเมนของความถี่โดยใช้การแปลงฟูเรียร์อย่างเร็วในหัวข้อถัดไป

3.2 การแปลงฟูเรียอย่างเร็ว

จากบล็อกไดอะแกรมการทำงานของระบบในรูปที่ 2. สัญญาณเสียงพูดจะถูกแปลงให้เข้าสู่โดเมนของความถี่โดยใช้การแปลงฟูเรียร์อย่างเร็ว วิธีนี้จะคำนวณหาการแปลงฟูเรียร์แบบเป็นช่วง (Discrete Fourier Trans - form, DFT) ดังสมการ

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-jk(2\pi/N)n} \quad (2)$$

$$= \sum_{n=0}^{N-1} x(n) W_N^{nK} \quad (3)$$

โดยการใช้ FFT กรณีที่ *X(n)* มีค่าเป็นเลขจำนวนจริง (Real number) และมีค่า DFT ของ *X(K)* ดังสมการ

$$X(k) = X_r(k) + jX_i(k) \quad (4)$$

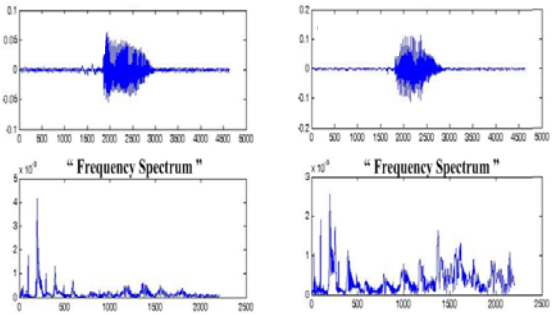
โดยที่ $X_r(k)$ คือส่วนที่เป็นเลขจำนวนจริง: $X_i(k)$ คือส่วนที่เป็นเลขจำนวนจินตภาพ: $x(n)$ คือค่าสัญญาณเสียงพูดอินพุตที่สุ่มมา: W คือค่ารากของหนึ่งที่มีอันดับ N: k มีค่าเท่ากับ 0,1,…,N-1: n มีค่าเท่ากับ 0,1,…,N-1 จากนั้นนำค่า $X(k)$ ที่ได้ไปคำนวณหาค่าของขนาด (Magnitude value) และมุมเฟส (Phase angle) ดังสมการ

ค่าของขนาด $$|X(k)| = \sqrt{X_r(k)^2 + X_i(k)^2} \quad (5)$$

ค่าของมุมเฟส $$\phi = \tan^{-1} \frac{X_i(k)}{X_r(k)} \quad (6)$$

การคำนวณหาค่าการแปลงฟูเรียร์อย่างเร็ว *X(k)* จะคำนวณเฉพาะค่าสัญญาณเสียงพูดอินพุตที่สุ่มมาเป็นเลขจำนวนจริงเท่านั้น ค่าส่วนที่เป็นเลขจำนวนจินตภาพจะไม่ถูกนำมาใช้ และค่าผลลัพธ์ที่ได้ *X(k)* เฉพาะค่าของขนาดดังสมการที่ (5) เท่านั้นที่จะนำมาใช้กับการคำนวณด้วยคณิตศาสตร์โครงข่ายประสาทเทียม (ANN) ส่วนค่ามุมเฟสดังสมการที่ (6) จะไม่ถูกนำมาใช้กับการคำนวณ

ในรูปที่ 4. แสดงตัวอย่างคลื่นสัญญาณเสียงพูด และค่าสเปคตรัมที่ใช้เป็นข้อมูลในการทดสอบ



ก) คำว่า "ซ้าย" ข) คำว่า "ขวา"

รูปที่ 4. ตัวอย่างคลื่นสัญญาณเสียงพูด และค่าสเปคตรัม (Spectrum signal)

ในแต่ละตัวอย่างคลื่นสัญญาณเสียงพูดคือค่าสัญญาณเสียงพูดอินพุตที่สุ่มมา $x(n)$ และแต่ละตัวอย่างคลื่นสัญญาณเสียงพูดใช้จำนวนจุดข้อมูลเท่ากับ 8,000 จุด ค่าสัญญาณเสียงพูดอินพุตที่สุ่มมาจะถูกนำมาจัดเรียงใหม่เป็นข้อมูลขนาด 2 มิติ (2x4,000) และข้อมูลขนาด 2 มิติจะถูกนำมาใช้กับคณิตศาสตร์โครงข่ายประสาทเทียม (ANN)

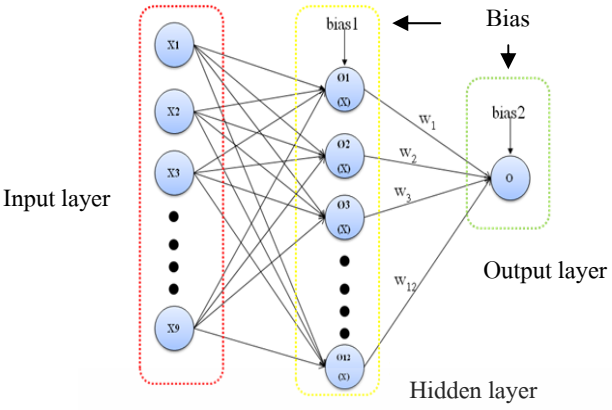
4. โครงข่ายประสาทเทียม

ส่วนนี้เป็นส่วนที่ทำการรู้จำข้อมูลเสียงพูดโดยเลือกใช้ทฤษฎีโครงข่ายประสาทเทียม (ANN) มาทำหน้าที่ในการรู้จำเสียงพูด โครงข่ายประสาทเทียม คือคณิตศาสตร์รูปแบบหนึ่ง และมีหลายรูปแบบ เช่นโครงข่าย Radial Basis Function (RBF) โครงข่าย Multilayer Perceptron (MP) ในงานวิจัยหลายบทความได้มีการรายงานประสิทธิภาพการทำงานของโครงข่าย RBF ดีกว่าโครงข่าย MP แต่ก็มีบางงานวิจัยรายงานประสิทธิภาพการทำงานในทางตรงกันข้าม [5] อย่างไรก็ตามในงานวิจัยนี้เลือกใช้โครงข่าย RBF สำหรับการรู้จำเสียงพูด และโครงข่าย RBF มีโครงสร้างดังรูปที่ 5 โครงข่าย RBF ประกอบด้วย 3 ชั้น (layer) ดังนี้ ชั้นที่หนึ่ง (First layer) รับค่าอินพุตมาจากเอาท์พุตของการแปลงฟูเรียร์ อย่างเร็ว ชั้นที่สอง (Second layer) เรียกว่า ชั้นซ่อน(Hidden layer) มีการคำนวณดังสมการ

$$Z_j = \frac{1}{M} \sum_{m=1}^{M} W_{mj} \phi_m + b_j \quad (7)$$

โดยที่ Z_j คือการขอใช้งานเอาท์พุตในลำดับที่ jth , m คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5. โครงขาย Radial Basis Function

จำนวน neurons, w_{mj} คือค่าน้ำหนัก (weight) ระหว่าง m^{th} ในชั้นที่สองและเอาท์พุต, b_j คือค่าไบแอส (Bias term) ชั้นที่สาม (Third layer) เรียกว่า ชั้นเอาท์พุต ซึ่งจะทำการแปลงค่าจากชั้นที่สองและค่าไบแอสโดยการใช้ฟังก์ชั่นเกาส์เซียน (Gaussian function) โดยที่อินพุตของฟังก์ชั่นเกาส์เซียนจะรับค่ามาจากเอาท์พุตของชั้นที่สองหรือชั้นซ่อนเพื่อคำนวณหาระยะ ทางระหว่างเสียงพูด ตัวอย่างฟังก์ชั่นเกาส์เซียนดังแสดงในสมการ

$$G(\|x-t_i\|) = exp(-\frac{m_1}{d^2}\|x_j - t_i\|^2) \qquad (8)$$

โดยที่ : $G_1(x), G_2(x), G_3(x), ..., G_k(x)$ คือ ฟังก์ชั่นการแปลงแบบเกาส์เซียน, $X_1, X_2, X_3, ..., X_n$ คืออินพุตรับค่ามาจากเอาท์พุตของชั้นที่สองหรือชั้นซ่อน, m_1 คือค่าการเลือกที่ตำแหน่งส่วนกลาง (Central location), t_i คือศูนย์กลาง (Centre of gravity), d คือระยะทางระหว่างศูนย์กลาง (distance between the centre of each data)

5. การทดลอง

ในบทความวิจัยนี้จุดมุ่งหมายหลักคือต้องการศึกษาการรู้จำสัญญาณเสียงพูดแบบไม่ขึ้นกับผู้พูด ในการทดลองจะเริ่มทำการศึกษาจากสัญญาณเสียงพูดแบบไม่ขึ้นกับผู้พูดที่เป็นเพศหญิง และเพศชาย เพศละ 5 คน คำละ 5 ครั้ง ใช้คลื่นสัญญาณเสียงพูด ข) ภาพด้านหน้า ซ้าย = ซ ") (" ขวา = ข ") (" หน้า = น ") (" หลัง = ล ") และคำว่า (" หยุด = ย ") โปรแกรมเขียนด้วยภาษาซี ผลการทดลองการรู้จำเสียงพูดแบบไม่ขึ้นกับผู้พูดดังแสดงในตารางที่ 2-6 และทดสอบกับหุ่นยนต์ต้นแบบ รูปที่ 6

ก) ภาพด้านบน ข) ภาพด้านหน้า

ค) ภาพด้านข้าง ค) ภาพสามมิติ

รูปที่ 6. หุ่นยนต์ต้นแบบ

ตารางที่ 2. เสียงบุคคลที่มีในฐานข้อมูล (เพศหญิง)

คำ	ลำดับที่ของเสียงผู้พูด					รวม
	1	2	3	4	5	เฉลี่ย
	20%	20%	20%	20%	20%	100%
ซ	ซ	ข	ข	ซ	ซ	60.00
ข	ข	ข	ข	ข	ซ	80.00
น	น	น	ซ	ซ	น	60.00
ล	ล	ล	ล	น	ล	80.00
ย	ย	ย	ย	ล	ย	80.00
เฉลี่ย	99.99	80.00	60.00	40.00	80.00	72.00

จากผลการทดลองในตารางที่ 2 พบว่าในกรณีใช้เสียงบุคคลที่มีในฐานข้อมูล (เพศหญิง) มีค่าความถูกต้องเฉลี่ยค่าเท่ากับ 72.00 % และลำดับที่ของเสียงผู้พูด ที่มีค่าความถูกต้องเฉลี่ยสูงสุดคือครั้งที่ 1 มีค่าเท่ากับ 99.99 % และเฉลี่ยต่ำสุดคือครั้งที่ 4 มีค่าเท่ากับ 40.00 %

จากผลการทดลองในตารางที่ 3 พบว่าในกรณีใช้เสียงบุคคลที่ไม่มีในฐานข้อมูล (เพศหญิง) มีค่าความถูกต้องเฉลี่ยค่าเท่ากับ 68.00 % และลำดับที่ของเสียงผู้พูดเฉลี่ยสูงสุดคือครั้งที่ 4 มีค่าเท่ากับ 99.99 % และเฉลี่ยต่ำสุดคือครั้งที่ 5 มีค่าเท่ากับ 20.00 %

จากผลการทดลองในตารางที่ 4-5 พบว่า % เฉลี่ยความถูกต้องเท่ากับ 80% และ 68% ในกรณีนี้ขึ้นและไม่ขึ้นอยู่กับเสียงผู้พูด ไม่ใช้วงจรฮาร์ดแวร์ (ไม่ใช้ต้นทุนกับฮาร์ดแวร์) และเมื่อเปรียบเทียบกับ [2] % เฉลี่ยความถูกต้องจะขึ้นอยู่กับเสียงผู้พูด และเมื่อเปรียบเทียบกับ [3]–[4] ใช้ร่วมกับอินเตอร์เฟซฮาร์ดแวร์ชนิดพิเศษ % เฉลี่ยความถูกต้อง 89% และ 98% (ใช้ต้นทุนกับฮาร์ดแวร์ที่สูง)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

% เฉลี่ยความถูกต้องในตารางที่ 4-5 ดูเหมือนจะน้อย ความผิดพลาดส่วนหนึ่งเกิดจาก สัญญาณรบกวนและสถานที่ทดลองเป็นบริเวณที่เปิด และใช้ไมโครโฟนที่มีราคาถูก คุณภาพต่ำ และถ้าปรับปรุงข้อผิดพลาดนี้ ค่า% เฉลี่ยความถูกต้องน่าจะสูงขึ้น

ตารางที่ 3. เสียงบุคคลที่ไม่มีในฐานข้อมูล (เพศหญิง)

คำ	ลำดับที่ของเสียงผู้พูด					รวม
	1	2	3	4	5	เฉลี่ย
	20%	20%	20%	20%	20%	100%
ซ	ซ	ข	ซ	ซ	ล	60.00
ข	น	ข	ข	ข	ข	80.00
น	น	น	ล	น	ล	60.00
ล	ข	ล	ล	ล	ซ	60.00
ย	ย	ย	ย	ย	ล	80.00
เฉลี่ย	60.00	80.00	80.00	99.99	20.00	68.00

ตารางที่ 4. เสียงบุคคลที่มีในฐานข้อมูล (เพศชาย)

คำ	ลำดับที่ของเสียงผู้พูด					รวม
	1	2	3	4	5	เฉลี่ย
	20%	20%	20%	20%	20%	100%
ซ	ซ	ข	ข	ซ	ซ	60.00
ข	ข	ข	ข	ข	ข	99.99
น	น	น	ซ	ซ	น	60.00
ล	ล	ล	ล	ล	ล	99.99
ย	ย	ย	ย	ล	ย	80.00
เฉลี่ย	99.99	80.00	60.00	60.00	99.99	80.00

ตารางที่ 5. เสียงบุคคลที่ไม่มีในฐานข้อมูล (เพศชาย)

คำ	ลำดับที่ของเสียงผู้พูด					รวม
	1	2	3	4	5	เฉลี่ย
	20%	20%	20%	20%	20%	100%
ซ	ซ	ข	ซ	ซ	ล	60.00
ข	ข	ข	ล	ข	ข	80.00
น	น	น	ล	น	น	80.00
ล	ข	ล	ล	ข	ล	60.00
ย	ย	ล	ย	ข	ย	60.00
เฉลี่ย	80.00	60.00	60.00	60.00	80.00	68.00

6. สรุปผล

บทความนี้นำเสนอการออกแบบโปรแกรมควบคุมการเคลื่อนไหวหุ่นยนต์โดยการใช้คำสั่งเสียงพูดภาษาไทยแบบกำหนดเสียงพูด คือคำว่า ซ้าย ขวา หน้า หลังและหยุด ไม่ขึ้นกับผู้พูด ผลการทดสอบจากวิธีที่นำเสนอสำหรับเสียงบุคคลที่มีในฐานข้อมูลมีค่าเปอร์เซ็นต์ความถูกต้องเฉลี่ยสำหรับเพศหญิงเท่ากับ 72.00% เพศชายเท่ากับ 80% และในกรณีเสียงบุคคลที่ไม่มีในฐานข้อมูล สำหรับเพศหญิงเท่ากับ 68.00% เพศชายเท่ากับ 68% นั่นหมายความว่าในกรณีเสียงบุคคลที่ไม่มีในฐานข้อมูลจะไม่ขึ้นอยู่กับเพศของผู้พูด ในทางตรงกันข้าม (เสียงบุคคลที่มีในฐานข้อมูล) เพศชายมีค่าความถูกต้องเฉลี่ยสูงกว่าเพศหญิง 80%

7. เอกสารอ้างอิง

[1] R. Simpson and Levine S., "Voice Control of Powered Wheelchair", IEEE Transaction on Neural System and Rehabilition Engineering, vol. 10, no. 2, pp. 122-125, June 2002.

[2] M. Hrncar, 2007, "Voice Command Control for Mobile Robots", Department of Control and Information Systems Faculty of Electrical Engineering, University of Zilina, Zilina 2007.

[3] T. Takiguchi , Yamagata., Sako., Miyake N., Revaud J. and Ariki Y., "Human-Robot Interface Using System Request Utterance Detection Base on Acoustic Features", International Journal of Hybrid Information Technology, vol. 1, no. 3, July, 2008.

[4] Q. Ahmed and AL-Thahab, "Controled of Mobile Robots by Using Speech Recognition" , Journal of Babylon University/Pure and Applied Sciences, no. 3, vol. 19, 2011.

[5] S. Hyontai, "Performance Comparison of RBF networks and MLPs for Classification," Proceedings of the 9th WSEAS International Conference on Applied Information and Communications (AIC'09), pp 450-455, 2009.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้