ว.วิทย. มข. 42(3) 658-669 (2557) KKU Sci. J. 42(3) 658-669 (2014)

การลดความซับซ้อนของการประมาณค่าทิศทางเสียงแบบ 2 มิติในแนวนอน โดยใช้วิธีสหสัมพันธ์แบบไขว้ สำหรับการรับรู้ทิศทางเสียงของหุ่นยนต์

Reduce the Complexity of Estimation a 2D Horizontal Directional Sound Using Cross-correlation Method for the Recognition of Robot Voice Directions

อรรถชัย คนยัง[1] *สายยัญ สายยศ*[1*] *และ พิเชษ วะยะลุน*[1]

บทคัดย่อ

การพัฒนาหุ่นยนต์ให้มีความสามารถคล้ายคลึงกับมนุษย์ให้ได้มากที่สุด เพื่ออนาคตจะสามารถนำไปใช้ในงานที่มีความเสี่ยงต่ออันตรายแทนมนุษย์ได้ การรับรู้ทิศทางต้นกำเนิดเสียงจึงมีความสำคัญที่จะช่วยให้หุ่นยนต์สามารถเคลื่อนที่ไปในทิศทางที่กำเนิดเสียงได้อย่างถูกต้อง และยังสามารถนำไปประยุกต์ใช้กับระบบนำทางแก่ผู้พิการทางหูได้ ดังนั้นงานวิจัยนี้ได้นำเสนอวิธีการประมาณค่าทิศทางเสียงด้วยการคำนวณสหสัมพันธ์แบบไขว้ (cross-correlation) จากไมโครโฟน 2 ตัว ที่เปรียบเสมือนหูซ้ายและหูขวาของมนุษย์ด้วยคุณสมบัติของฟังก์ชันการถ่ายโอนที่เกี่ยวข้องกับศีรษะ (head-related transfer function: HRTF) และลดความซับซ้อนของการประมาณค่าทิศทางเสียงด้วยสมการถดถอยแบบพหุนาม (polynomial regressions) สำหรับประมาณค่าจากแนวโน้มของค่าสหสัมพันธ์ อีกทั้งยังสามารถประมาณค่าทิศทางเสียงแบบเวลาจริง (real-time) ให้มีความรวดเร็วยิ่งขึ้น ซึ่งวิธีที่นำเสนอมีค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (root mean square error) เพียง 0.96 องศาเท่านั้น

[1]ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น 40002

*Corresponding Author, E-mail: saiyan@kku.ac.th

ABSTRACT

Developing robots ability to be able to act similar to human being most for being substituted from human being in the risky and dangerous work in the future. Perceiving sound source origin is important to help robots move toward to the origin of sound. In addition, this idea can be applied as the leading system for the hearing disability people. Therefore, this research proposes Cross-correlation method to estimate the direction of sound from two microphones to be presented as human begins left ear and right ear with the qualifications of the transfer function which is related to the head (Head-Related Transfer Function: HRTF) and reduces the complexity of the sound direction estimation with the application of polynomial regressions. To estimated the tendency of correlation of sound direction in real time is faster. The method that has presented shown the approximation error as root mean square only 0.96 degrees.

คำสำคัญ: ประมาณค่าทิศทางเสียง หูหุ่นยนต์ สหสัมพันธ์แบบไขว้ สมการถดถอย พหุนาม

Keywords: Sound direction estimation, Robot ears, Cross-correlation, Regressions equation, Polynomial

บทนำ

ปัจจุบันนี้เทคโนโลยีทางด้านหุ่นยนต์เข้ามามีบทบาทในการดำเนินชีวิตของมนุษย์เป็นอย่างมาก การเคลื่อนที่ของหุ่นยนต์ตามทิศทางที่กำเนิดเสียงเป็นอีกหนึ่งเทคโนโลยีที่ก้าวหน้าของหุ่นยนต์ในปัจจุบัน ในด้านการรับรู้ทิศทางต้นกำเนิดเสียงจึงมีความจำเป็นที่จะช่วยให้หุ่นยนต์สามารถเคลื่อนที่ไปในทิศทางที่กำเนิดเสียงได้อย่างถูกต้อง ทฤษฎีฟังก์ชันการถ่ายโอนที่เกี่ยวข้องกับศีรษะจึงเป็นหนึ่งในวิธีการที่จะทำให้แยกแยะทิศทางได้ และมีงานวิจัยที่เกี่ยวข้องที่กล่าวถึงการประมาณค่าทิศทางเสียงโดยใช้หูเทียมสำหรับหุ่นยนต์ (Hwang et al., 2011) ได้เสนอวิธีการใช้ไมโครโฟน 4 ตัวต่อหู 1 ข้างในการประมาณค่าทิศทางเสียงแบบ 3 มิติ ซึ่งการใช้ไมโครโฟนเป็นจำนวนมากทำให้การคำนวณสหสัมพันธ์ (correlation) มีความซับซ้อนมากยิ่งขึ้น อีกทั้งยังใช้ตัวกรองแบบเกาส์เซียน (gaussian filter) ในการกรองคลื่นเสียงให้มีความราบเรียบและขจัดเสียงรบกวนที่ไม่ต้องการ ส่งผลให้รายละเอียดที่สำคัญบางส่วนหายไป นอกจากนี้ยังมีงานวิจัยที่เกี่ยวข้องที่กล่าวถึงการประมาณค่าทิศทางเสียงสำหรับการมีปฏิสัมพันธ์ของหุ่นยนต์ปลาโลมาเพื่อความบันเทิง (Daejung et al., 2008) โดยได้เสนอการใช้วงจรไมโครโฟนสเตอริโอในการรับเสียงและใช้ไมโครคอนโทรลเลอร์ในการประมาณค่าทิศทางเสียง แต่ยังมีข้อจำกัดในเรื่องของหน่วยความจำ รวมทั้งการประมาณค่าทิศทางเสียงโดยการหาแนวโน้มของค่าที่ได้จากฟังก์ชันขนาดที่แตกต่างเฉลี่ย (average magnitude difference function: AMDF) ซึ่งมีความแม่นยำที่ยังไม่ดีเท่าที่ควร

จากปัญหาที่ได้กล่าว งานวิจัยนี้จึงนำเสนอวิธีการประมาณค่าทิศทางเสียงในแนวนอนหรือแนวราบ โดยใช้ไมโครโฟนเพียง 2 ตัวในการประมาณค่าทิศทางเสียง ด้วยการคำนวณสหสัมพันธ์จากคุณสมบัติของฟังก์ชันการถ่าย

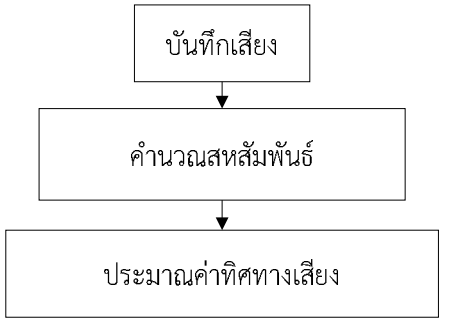
โอนที่เกี่ยวข้องกับศีรษะ และลดความซับซ้อนในการประมาณค่าทิศทางเสียงด้วยสมการถดถอยแบบพหุนาม ทำให้เกิดเป็นสมการใหม่สามารถนำไปใช้งานและประมาณค่าทิศทางเสียงแบบเวลาจริงได้ทันที

วิธีการดำเนินการวิจัย

วิธีการประมาณค่าทิศทางเสียงที่นำเสนอ

วิธีการประมาณค่าทิศทางเสียงที่นำเสนอ เป็นหนึ่งในส่วนของระบบการรับรู้เสียงของหุ่นยนต์ ซึ่งระบบจะประกอบด้วยการบันทึกเสียง การคำนวณสหสัมพันธ์และการประมาณค่าทิศทางเสียง

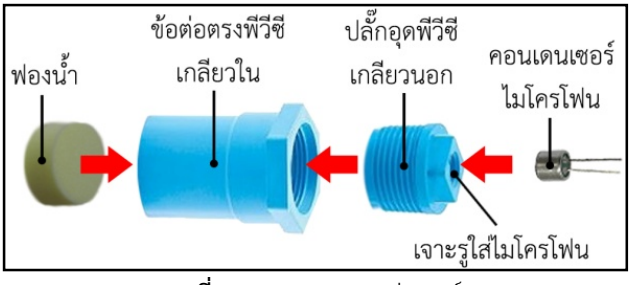
ขั้นตอนนี้เป็นขั้นตอนการดำเนินงานโดยทั่วไป ซึ่งส่วนสำคัญคือขั้นตอนคำนวณสหสัมพันธ์แล้วประมาณค่าทิศทางเสียง โดยสามารถดำเนินงานขั้นตอนต่าง ๆ ได้ดังรูปที่ 1



รูปที่ 1 ขั้นตอนการประมาณค่าทิศทางเสียงที่นำเสนอ

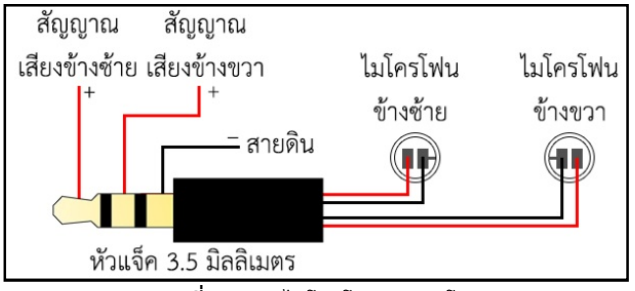
การประดิษฐ์หูหุ่นยนต์

สำหรับการประดิษฐ์หูหุ่นยนต์ให้สามารถรับเสียงได้ ซึ่งงานวิจัยนี้ได้ประยุกต์ใช้คอนเดนเซอร์ไมโครโฟน 2 ตัว เป็นหูด้านซ้ายและด้านขวา โดยจะใช้ข้อต่อตรงพีวีซีเป็นหูของหุ่นยนต์ จากนั้นนำปลั๊กอุดพีวีซีมาเจาะรูตรงกลางขนาดเท่ากับคอนเดนเซอร์ไมโครโฟน ประมาณ 8 มิลลิเมตร แล้วนำเอาคอนเดนเซอร์ไมโครโฟนมาสอดใส่รูที่เจาะเอาไว้ โดยให้หัวไมโครโฟนหันเข้าจากทางด้านท้ายของปลั๊กอุดพีวีซี แล้วนำปลั๊กอุดพีวีซีมาหมุนเข้ากับข้อต่อตรงพีวีซีให้แน่นและนำฟองน้ำที่ไม่ทึบจนเกินไป เสียงสามารถผ่านได้มาตัดเป็นวงกลมขนาดเท่ากับความกว้างภายในข้อต่อตรงพีวีซี ความหนาของฟองน้ำไม่เกิน 1 เซนติเมตร วิธีนี้จะช่วยดูดซับเสียงที่เกิดการสั่นพ้องของเสียงน้อยลง รวมทั้งยังช่วยลดเสียงลมที่ผ่านปากท่อเข้าไปภายในท่อได้ ดังวิธีการในรูปที่ 2



รูปที่ 2 การออกแบบหูหุ่นยนต์

หลังจากได้ประกอบหูหุ่นยนต์เรียบร้อยแล้วให้ทำหูหุ่นยนต์อีกหนึ่งข้างสำหรับเป็นหูด้านซ้ายและด้านขวา จากนั้นให้ต่อวงจรไมโครโฟนสเตอริโอโดยใช้สายสเตอริโอที่มีหัวแจ็คขนาด 3.5 มิลลิเมตร ต่อเข้ากับคอนเดนเซอร์ไมโครโฟนที่ประกอบเป็นหูหุ่นยนต์ทั้งด้านซ้ายและด้านขวา ดังรูปที่ 3

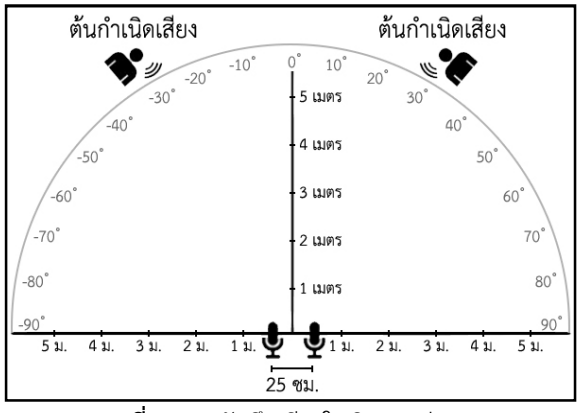


รูปที่ 3 วงจรไมโครโฟนสเตอริโอ

โดยปกติแล้วตำแหน่งของหูมนุษย์จะมีรูหูรวมทั้งแก้วหูอยู่ภายในหัว แต่จะมีส่วนของใบหูเท่านั้นที่ยื่นออกมาจากหัวเพื่อรวบรวมคลื่นเสียงที่มาจากทิศทางต่าง ๆ แล้วส่งเข้าสู่รูหู แต่สำหรับหุ่นยนต์แล้วเป็นเรื่องยากในการออกแบบให้มีใบหูที่สามารถรวบรวมคลื่นเสียงแล้วส่งเข้าสู่รูหูที่อยู่ภายในหัวเหมือนอย่างมนุษย์ได้ ดังนั้นหูหุ่นยนต์ที่ออกแบบมานี้จะวางอยู่ในตำแหน่งด้านข้างของหัว โดยระยะห่างระหว่างหูด้านซ้ายถึงหูด้านขววัดจากจุดกึ่งกลางของหูหุ่นยนต์กว้าง 25 เซนติเมตร ในการทดลองจริงจะต้องมีส่วนที่เป็นหัวหุ่นยนต์สำหรับกั้นเสียงระหว่างหูด้านซ้ายและด้านขวาด้วย เพื่อให้ระยะเวลาที่เสียงเดินทางถึงหูทั้งสองข้างมีความแตกต่างกันมากขึ้น

การบันทึกเสียง

เสียงพูดของมนุษย์ที่ใช้ในระบบจะเก็บเป็นไฟล์ .WAV โดยเงื่อนไขที่สำคัญหลัก ๆ คือเสียงที่ใช้เป็นเสียงพูดปกติของมนุษย์ โดยจะบันทึกจากไมโครโฟน 2 ตัวพร้อมกันในรูปแบบเสียงสเตอริโอ ซึ่งระดับความดังของการพูดอยู่ที่ 65-75 เดซิเบล (decibel) ลักษณะทิศทางในการบันทึกเสียงเป็นแบบทิศทางด้านหน้าในแนวนอนหรือแนวราบ จากทางด้านซ้ายมือ กำหนดให้เป็นมุม -90 องศา ถึงด้านขวาขวามือ กำหนดให้เป็นมุม 90 องศา และตรงกลางกำหนดให้เป็นมุม 0 องศา โดยเลื่อนตำแหน่งทีละ 10 องศา องศาละ 5 ระยะ คือที่ระยะ 1 เมตร 2 เมตร 3 เมตร 4 เมตร และ 5 เมตร ดังตัวอย่างรูปที่ 4



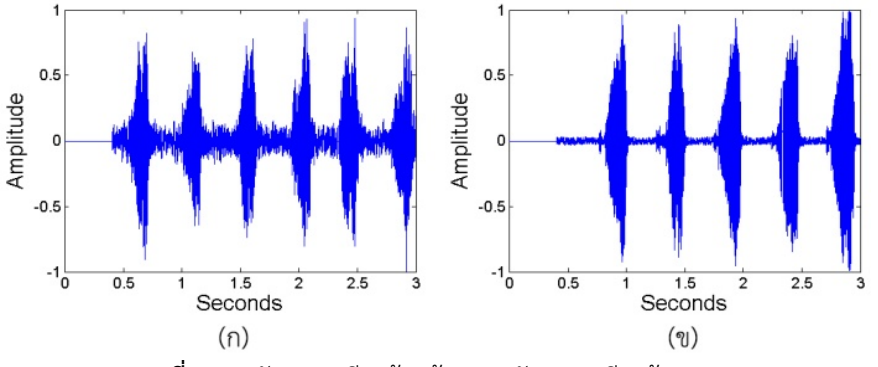
รูปที่ 4 การบันทึกเสียงในทิศทางต่าง ๆ

การบันทึกเสียงจะมีเสียงจากไมโครโฟนด้านซ้ายและด้านขวาอยู่ในไฟล์เดียวกัน เรียกว่าเสียงสเตอริโอซึ่งเป็นเสียงที่ประกอบด้วย 2 สัญญาณเสียง (ด้านซ้ายและด้านขวา) และเนื่องจากการรับเสียงจากไมโครโฟนเป็นแบบเวลาจริง คือการตอบสนองทันทีโดยที่ไมโครโฟนจะรอรับเสียงแบบต่อเนื่องทำให้เสียงผ่านไมโครโฟนเข้ามาเรื่อยๆ ดังนั้นจึงต้องมีการกำหนดขนาดชุดข้อมูลในแต่ละชุดที่จะบันทึก โดยกำหนดให้ 1 ขนาดชุดข้อมูลเท่ากับ 3 วินาที สำหรับการทดลองในวิจัยนี้จะใช้อัตราการสุ่มตัวอย่างที่ช่วงความถี่ 8 kHz (กิโลเฮิรตซ์) 11 kHz 22 kHz 44 kHz 48 kHz และ 96 kHz ในการบันทึกเสียง อีกทั้งยังได้ทดลองความแตกต่างระหว่างไฟล์เสียงที่บันทึกเสียงด้วยอัตราบิต (bit rate) ขนาด 8 bit/s (บิตต่อวินาที) และ 16 bit/s ด้วย ดังนั้นการบันทึกเสียงเริ่มต้นจะกำหนดอัตราการสุ่มตัวอย่างสูงสุดในการทดลองก่อนคือที่ช่วงความถี่ 96 kHz และอัตราบิตขนาด 16 bit/s เมื่อบันทึกเสียงครบทุกมุมองศาและทุกระยะที่กำหนดไว้แล้ว จะเพิ่มผลการทดลองของอัตราการสุ่มตัวอย่างอื่นๆ คือที่ช่วงความถี่ 8 kHz 11 kHz 22 kHz 44 kHz และ 48 kHz รวมทั้งที่อัตราบิตขนาด 8 bit/s ด้วยวิธีการเปลี่ยนแปลงอัตราการสุ่มตัวอย่าง (Resampling) (Andrew, 2002)

เพื่อให้ผลการวิจัยมีความน่าเชื่อถือและเพื่อทดสอบประสิทธิภาพของระบบ จะบันทึกเสียงผู้ชายและผู้หญิงซ้ำทั้งหมด 3 รอบ โดยเสียงของผู้บันทึกจะมีช่วงอายุที่แตกต่างกัน บันทึกเสียงในสถานที่เดียวกัน ช่วงเวลาเดียวกัน และอุณหภูมิที่ใกล้เคียงกัน โดยวัดอุณหภูมิที่จุดวางไมโครโฟนประมาณ 32 องศาเซลเซียส

การคำนวณสหสัมพันธ์ (correlation)

ในการคำนวณสหสัมพันธ์จะใช้วิธีการคำนวณสหสัมพันธ์แบบไขว้ (cross-correlation) เนื่องจากเป็นการหาความสัมพันธ์ระหว่างสองสัญญาณ (Bracewell, 1965) คือสัญญาณเสียงด้านซ้ายและด้านขวา ดังนั้นจึงต้องแยกสัญญาณเสียงด้านซ้ายและด้านขวาที่อยู่ในไฟล์เสียงแบบสเตอริโอออกจากกันก่อน ดังตัวอย่างรูปที่ 5 แสดงสัญญาณเสียงด้านซ้ายและด้านขวาของไฟล์เสียงที่มุม 90 องศา



รูปที่ 5 (ก) สัญญาณเสียงด้านซ้าย (ข) สัญญาณเสียงด้านขวา

จากรูปที่ 5 จะเห็นว่าความสูงของคลื่นเสียงที่สัญญาณเสียงด้านขวาจะสูงกว่าสัญญาณเสียงด้านซ้าย แสดงว่าเสียงมาจากทิศทางด้านขวาและเสียงด้านขวามีระดับความดังมากกว่าเสียงด้านซ้าย พิสูจน์ได้โดยการนำสัญญาณเสียงทั้งสองข้างมาคำนวณสหสัมพันธ์แบบไขว้ด้วยสมการดังนี้

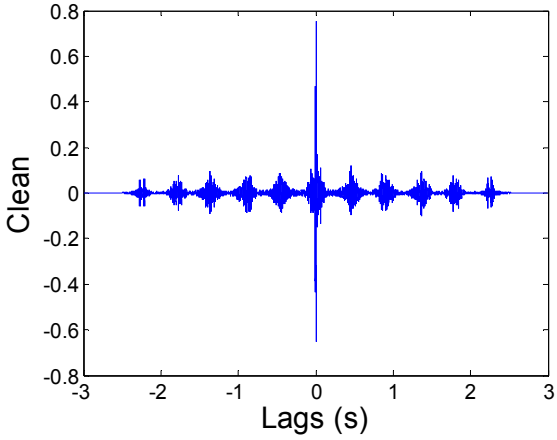
$$(f * g)(\tau) \overset{\text{def}}{=} \int_{-\infty}^{\infty} f^*(t) g(t+\tau) dt \qquad (1)$$

เมื่อ $f*$ หมายถึงการร่วมกันที่ซับซ้อนของ f

τ คือ เวลาความล่าช้า

ในทางทฤษฎีความน่าจะเป็นสหสัมพันธ์จะถูกนำมาใช้ในการรวมปัจจัยมาตรฐานในลักษณะที่มีความสัมพันธ์ระหว่างค่า -1 ถึง 1 และสหสัมพันธ์แบบไขว้หมายถึงความสัมพันธ์ $corr(x, y)$ ของตัวแปรสุ่ม x และ y ในขณะที่ความสัมพันธ์ของเวกเตอร์แบบสุ่ม x จะถือว่าเป็นเวกเตอร์ของสหสัมพันธ์ระหว่างองค์ประกอบของตัวแปร x

เมื่อคำนวณสหสัมพันธ์สัญญาณเสียงด้านซ้ายและด้านขวาโดยใช้สมการสหสัมพันธ์แบบไขว้แล้ว สามารถแสดงผลลัพธ์ของการคำนวณสหสัมพันธ์เป็นกราฟได้ดังรูปที่ 6



รูปที่ 6 กราฟสหสัมพันธ์

จากกราฟสหสัมพันธ์ในรูปที่ 6 จะเห็นว่าเมื่อนำสองสัญญาณเสียงมาคำนวณสหสัมพันธ์แบบไขว้จะมีค่าที่ได้จากการหาความสัมพันธ์ระหว่างเสียงด้านซ้ายและด้านขวาอยู่หลายค่า แต่มีค่าที่สนใจและจะนำมาใช้ประมาณค่าทิศทางเสียงคือค่าตำแหน่งสูงสุด (peaks position value) โดยค่าตำแหน่งสูงสุดจะขึ้นอยู่กับความสัมพันธ์ระหว่างสองสัญญาณเสียง หากทั้งสองสัญญาณเสียงเหมือนกัน หมายความว่าเสียงมาจากตรงกลางค่าตำแหน่งสูงสุดจะอยู่ที่จุดกึ่งกลางของกราฟ สามารถคำนวณหาค่าตำแหน่งสูงสุดได้จาก $max_r = max(R)$ เมื่อ R คือค่าที่ได้จากการคำนวณสหสัมพันธ์แบบไขว้ระหว่างสองสัญญาณเสียง และ max_r คือค่าตำแหน่งสูงสุด เมื่อได้ค่าตำแหน่งสูงสุดแล้วให้คำนวณหาค่าความยาวของกราฟสหสัมพันธ์ดังสมการที่ 2

$$length_r = (Fs \times s \times 2) - 1 \tag{2}$$

เมื่อ $length_r$ คือความยาวของกราฟสหสัมพันธ์

Fs คือ อัตราการสุ่มตัวอย่างมีหน่วยเป็น Hz (เฮิรตซ์)

s คือ ขนาดชุดข้อมูลมีหน่วยเป็นวินาที

ตัวอย่างในการคำนวณหาค่าความยาวของกราฟสหสัมพันธ์ เช่น อัตราการสุ่มตัวอย่างที่บันทึกเสียงเท่ากับ 44, 100 Hz มีขนาดชุดข้อมูลเท่ากับ 3 วินาที ทดลองแทนค่าในสมการที่ 2 จะได้ค่าความยาวของกราฟ

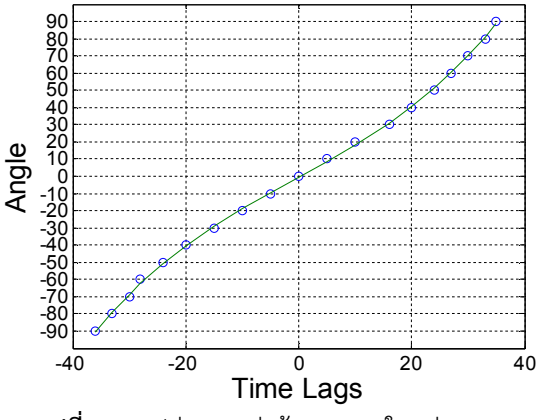
เท่ากับ 264,599 และเมื่อได้ค่าความยาวของกราฟแล้วให้คำนวณหาค่าตำแหน่งกึ่งกลางของกราฟสหสัมพันธ์ สามารถคำนวณได้จากสมการที่ 3

$$center_r = (length_r + 1)/2 \tag{3}$$

เมื่อ $center_r$ คือค่าตำแหน่งกึ่งกลางของกราฟสหสัมพันธ์ หลังจากได้ค่าต่าง ๆ ตามสมการข้างต้นแล้ว ทดลองนำค่าตำแหน่งสูงสุดลบด้วยค่าตำแหน่งกึ่งกลาง ผลลัพธ์ที่ได้จะเป็นเลขจำนวนเต็มและเรียกค่านี้ว่าค่าความล่าช้าของเวลา (time lags) โดยค่านี้สามารถบอกได้ว่าเสียงมากจากด้านซ้ายหรือด้านขวา คือเมื่อค่าความล่าช้าของเวลามีค่าติดลบแสดงว่าเสียงมาจากด้านซ้าย และเมื่อค่าความล่าช้าของเวลามีค่าเป็นบวกแสดงว่าเสียงมาจากด้านขวา หากค่าความล่าช้าของเวลามีค่าเท่ากับศูนย์แสดงว่าเสียงมาจากตรงกลางคือที่ 0 องศา ตัวอย่างเช่น ค่าตำแหน่งสูงสุดที่มุม 10 องศา เท่ากับ 132,305 และค่าตำแหน่งกึ่งกลางของกราฟสหสัมพันธ์เท่ากับ 132,300 เมื่อนำสองค่านี้มาลบกันจะได้ผลลัพธ์เท่ากับ 5 ซึ่งมีค่าเป็นบวก แสดงว่าเสียงมาจากด้านขวาของหุ่นยนต์

การประมาณค่าทิศทางเสียง

เมื่อคำนวณสหสัมพันธ์และได้ค่าความล่าช้าของเวลาที่มุม -90 องศา ถึง 90 องศา ในระยะ 1 เมตร โดยอัตราการสุ่มตัวอย่างที่บันทึกเสียงอยู่ในช่วงความถี่ 44 kHz และมีอัตราบิตขนาด 16 bit/s และเมื่อนำมาแสดงเป็นกราฟ โดยกำหนดให้มุมองศาเป็นแกน y และค่าความล่าช้าของเวลาเป็นแกน x จะเห็นแนวโน้มค่าความล่าช้าของเวลาที่มุมต่าง ๆ ดังรูปที่ 7



รูปที่ 7 กราฟค่าความล่าช้าของเวลาในแต่ละมุม

จากรูปที่ 7 จะเห็นว่ากราฟค่าความล่าช้าของเวลามีลักษณะถดถอยที่ไม่เป็นเชิงเส้นตรง ดังนั้นเราสามารถประมาณค่าทิศทางเสียงได้จากสมการถดถอยแบบพหุนาม ซึ่งสามารถคำนวณหาแนวโน้มที่มีความซับซ้อนได้ (Funkhouser, 1930) ดังสมการที่ 4

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \tag{4}$$

สมการที่ 4 คือสมการพหุนามทั่วไปใด ๆ ของดีกรีระดับ n เป็นสัมประสิทธิ์ของจำนวนจริงหรือตัวเลขที่ซับซ้อนและไม่เท่ากับศูนย์ เป็นที่รู้จักกันในทฤษฎีมูลฐาน (fundamental theorem) ของพีชคณิตจะมี n เป็นรากที่ซับซ้อนของ $x_1, x_2, \ldots, x_n$ จำนวนของค่าสัมประสิทธิ์ขึ้นอยู่กับระดับดีกรีของพนุนาม ในบางกรณีพหุนามดีกรี

ระดับ 3 ก็เพียงพอที่จะทำให้การพยากรณ์มีความใกล้เคียง เพราะยิ่งใช้พหุนามดีกรีระดับสูงมากขึ้นทำให้ความซับซ้อนของการคำนวณก็จะยิ่งมากขึ้นด้วย ซึ่งอาจทำให้ความแม่นยำเพิ่มขึ้นไม่มากนัก สมการที่เกี่ยวข้องกับค่าสัมประสิทธิ์ของพหุนาม $\{a_k\}$ สำหรับหาผลรวมของราก $\{x_i\}$ หาได้จากสมการที่ 5

$$\begin{cases} x_1 + x_2 + \ldots + x_{n-1} + x_n = \dfrac{a_{n-1}}{a_n} \\ (x_1x_2 + x_1x_3 + \ldots + x_1x_n) + (x_2x_3 + x_2x_4 + \ldots + x_2x_n) + \ldots + x_{n-1}x_n = \dfrac{a_{n-2}}{a_n} \\ \quad \vdots \\ x_1x_2 \ldots x_n = (-1)^n \dfrac{a0}{a_n}. \end{cases} \tag{5}$$

เมื่อนำค่าจากรูปที่ 7 มาคำนวณด้วยสมการถดถอยแบบพหุนาม โดยกำหนดให้แกน x คือค่าความล่าช้าของเวลา แกน y คือมุมองศา และกำหนดให้ดีกรีของพหุนามระดับ 3 ซึ่งค่าสัมประสิทธิ์ที่คำนวณได้ ได้แก่ 0.0005891585 0.0012155033 1.7807293888 และ -0.4252343901 จากนั้นเมื่อนำค่าสัมประสิทธิ์ที่คำนวณได้มาแทนค่าในสมการพหุนามจะได้สมการใหม่ดังสมการที่ 6

$$angle^o = 0.0005891585 \times x^3 + 0.0012155033 \times x^2 + 1.7807293888 \times x + (-0.4252343901) \tag{6}$$

เมื่อ x คือค่าความล่าช้าของเวลาที่ได้จากการคำนวณสหสัมพันธ์ ทดลองแทนค่าในสมการจะให้ผลลัพธ์เป็นมุมองศาที่พยากรณ์ได้ สามารถนำสมการข้างต้นไปประมาณค่าทิศทางเสียงแบบเวลาจริงได้ทันทีโดยไม่ต้องคำนวณหาค่าสัมประสิทธิ์ใหม่ วิธีนี้จะช่วยลดระยะเวลาในการประมวลผลและลดความซับซ้อนของการประมาณค่าได้มาก ค่าสัมประสิทธิ์จะสามารถเปลี่ยนแปลงได้ขึ้นอยู่กับองค์ประกอบอื่น ๆ เช่น เปลี่ยนแปลงอัตราการสุ่มตัวอย่าง เปลี่ยนแปลงระดับดีกรีของพหุนาม เป็นต้น และหากมีการเปลี่ยนแปลงอัตราการสุ่มตัวอย่างจะมีผลทำให้ต้องคำนวณหาค่าความล่าช้าของเวลาทุกมุมองศาใหม่ทั้งหมด

เพื่อพิสูจน์ว่าสมการใหม่ที่ได้สามารถประมาณค่าทิศทางเสียงได้จริง จะกำหนดให้ x เป็นค่าความล่าช้าของเวลาที่มุม -60 องศา ซึ่งมีค่าเท่ากับ -28 เมื่อแทนค่าในสมการที่ 6 จะให้ค่าพยากรณ์เท่ากับ -62.2659 องศา จากผลลัพธ์พบว่าค่าพยากรณ์มีความใกล้เคียงกับมุมที่ได้ทำการทดลองมาก โดยมีค่า RMSE (ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย) (ชมพูนุช, 2549) เพียง 2.2659 องศา

ผลการวิจัย

จากการทดลองที่กล่าวมาแล้วในข้างต้นจะทดลองใช้ดีกรีของพหุนามระดับที่สูงขึ้นคือ ดีกรีของพหุนามระดับ 5 ซึ่งค่าสัมประสิทธิ์ที่คำนวณได้ ได้แก่ 0.0000001843 -0.0000000307 0.0002156368 0.0000342276 1.9248140251 และ -0.0057774617 จากนั้นเมื่อนำค่าสัมประสิทธิ์ที่คำนวณได้มาแทนค่าในสมการพหุนามจะได้สมการใหม่ดังสมการที่ 7

$$angle^{o} = 0.0000001843 \times x^{5} + (-0.0000000307) \times x^{4} + 0.0002156368 \times x^{3} + 0.0000342276 \times x^{2} + 1.9248140251 \times x + (-0.0057774617) \quad (7)$$

เพื่อแสดงให้เห็นว่าดีกรีของพหุนามระดับสูงขึ้นมีความแม่นยำมากขึ้น จะกำหนดให้ x เป็นค่าความล่าช้าของเวลาที่มุม -60 องศา ซึ่งมีค่าเท่ากับ -28 เมื่อแทนค่าในสมการที่ 7 จะให้ค่าพยากรณ์เท่ากับ -61.7982 ซึ่งมีความแม่นยำมากกว่าดีกรีของพหุนามระดับ 3 โดยมีค่า RMSE เพียง 1.7982 องศาเท่านั้น

จากการทดลองจึงเลือกใช้ดีกรีของพหุนามระดับ 5 ในการทดลอง ซึ่งเป็นระดับที่เหมาะสม มีความซับซ้อนไม่มากนัก และให้ผลลัพธ์ที่เพียงพอเมื่อเปรียบเทียบกับดีกรีของพหุนามระดับที่มากกว่า ทดลองคำนวณหาค่าความล่าช้าของเวลาที่อัตราบิตขนาด 8 bit/s กับ 16 bit/s และนำมาเปรียบทียบกัน แสดงในตารางที่ 1

จากผลการทดลองจะเห็นว่าแต่ละมุมองศาที่อัตราบิตขนาด 8 bit/s กับ 16 bit/s มีค่าความล่าช้าของเวลาที่เท่ากัน ดังนั้นในการคำนวณหาค่าความล่าช้าของเวลาสามารถใช้อัตราบิตขนาด 8 bit/s แทนได้ เพื่อให้ไฟล์เสียงมีขนาดเล็กลงและยังประมวลผลได้อย่างรวดเร็วยิ่งขึ้น

ตารางที่ 1 ค่าความล่าช้าของเวลาที่อัตราบิตขนาด 8 bit/s กับ 16 bit/s

มุม (°)	ค่าความล่าช้าของเวลา		
	8 bit/s	16 bit/s	ค่าเฉลี่ย
-90°	-36	-36	-36
-60°	-28	-28	-28
-30°	-15	-15	-15
0°	0	0	0
30°	16	16	16
60°	27	27	27
90°	35	35	35

ในงานวิจัยนี้ได้ทดลองระยะห่างจากหูหุ่นยนต์เป็น 5 ระยะ โดยที่แต่ละระยะมีความดังของต้นกำเนิดเสียงอยู่ระหว่าง 65-75 เดซิเบล ซึ่งอยู่ในช่วงเสียงพูดปกติของมนุษย์ เมื่อทดลองนำค่าความล่าช้าของเวลาที่ได้จากการคำนวณสหสัมพันธ์ของไฟล์เสียงที่ระยะ 1 2 3 4 และ 5 เมตร มาเปรียบเทียบกันพบว่าค่าความล่าช้าของเวลาแต่ละมุมที่คำนวณได้ไม่แตกต่างกันมากนัก จะมีค่าที่แตกต่างกันเพียงบางมุมเท่านั้น แสดงในตารางที่ 2

ตารางที่ 2 ค่าความล่าช้าของเวลาที่ระยะ 1 เมตร 2 เมตร 3 เมตร 4 เมตร และ 5 เมตร

มุม (°)	ค่าความล่าช้าของเวลา					
	1 เมตร	2 เมตร	3 เมตร	4 เมตร	5 เมตร	ค่าเฉลี่ย
-90°	-36	-36	-36	-35	-36	-35.8
-60°	-28	-28	-27	-27	-27	-27.4
-30°	-15	-15	-15	-15	-15	-15
0°	0	0	0	0	0	0
30°	16	16	15	15	15	15.4
60°	27	27	27	26	27	26.8
90°	35	35	36	35	35	35.2

จากตารางที่ 2 ค่าความล่าช้าของเวลาที่แตกต่างกันเนื่องจากเมื่อเลื่อนตำแหน่งไกลออกไปทำให้ต้นกำเนิดเสียงไม่ได้อยู่ในมุมเดิม กล่าวคือหากแหล่งกำเนิดเสียงเอียงไปทางซ้ายหรือทางขวา ไม่ตรงกับมุมเดิมในระยะก่อนหน้า ก็จะทำให้ค่าความล่าช้าของเวลาที่คำนวณได้มีความคลาดเคลื่อนตามไปด้วย สามารถประมาณค่าทิศทางเสียงและคำนวณค่า RMSE ได้ดังตารางที่ 3 สังเกตว่าค่า RMSE มีความแตกต่างกันม่ถึง 1 องศา

ตารางที่ 3 ค่า RMSE ที่ระยะ 1 เมตร 2 เมตร 3 เมตร 4 เมตร และ 5 เมตร

ระยะทาง	1 เมตร	2 เมตร	3 เมตร	4 เมตร	5 เมตร
ค่า RMSE	0.9035	1.0474	0.7959	1.2695	0.6785

เพื่อให้ระบบสามารถประมาณค่าทิศทางเสียงได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้ได้ทดลองเกี่ยวกับความถี่เสียงผู้ชายกับเสียงผู้หญิงที่มีความแตกต่างกัน เพื่อทดสอบว่าความถี่เสียงมีผลกระทบต่อการคำนวณหาค่าความล่าช้าของเวลามากน้อยเพียงใด เมื่อทดลองคำนวณหาค่า RMSE โดยกำหนดอัตราการสุ่มตัวอย่างที่ความถี่ 96 kHz ที่อัตราบิตขนาด 8 bit/s ระยะทาง 1 เมตร แสดงในตารางที่ 4

ตารางที่ 4 ค่า RMSE ของความถี่เสียงผู้ชายและความถี่เสียงผู้หญิง

ความถี่เสียง	ผู้ชาย	ผู้หญิง
ค่า RMSE	0.9612	1.0165

จากตารางที่ 4 จะเห็นว่าค่าความล่าช้าของเวลาของความถี่เสียงผู้ชายและเสียงผู้หญิงมีค่าแตกต่างกันเล็กน้อยเท่านั้น โดยในแต่ละมุมมีค่าที่ใกล้เคียงกันมาก เนื่องจากเป็นการคำนวณทางด้านเวลาด้านความถี่เสียงจึงไม่มีผลกระทบกับการประมาณค่าทิศทางเสียง

ในการกำหนดอัตราการสุ่มตัวอย่างที่ความถี่แตกต่างกันจะทำให้ค่าความล่าช้าของเวลาที่คำนวณได้มีความแตกต่างกันไปด้วย เนื่องจากอัตราการสุ่มตัวอย่างมีผลกระทบโดยตรงต่อความละเอียดของคลื่นเสียง ยิ่งมีอัตราการสุ่มตัวอย่างสูงความละเอียดของคลื่นเสียงยิ่งมาก ดังในตารางที่ 5

ตารางที่ 5 ค่าความล่าช้าของเวลาของอัตราการสุ่มตัวอย่างที่ความถี่ 8 kHz 11k Hz 22 kHz 44 kHz 48 kHz และ 96 kHz

มุม (°)	**ค่าความล่าช้าของเวลา**					
	8 kHz	**11 kHz**	**22 kHz**	**44 kHz**	**48 kHz**	**96 kHz**
-90°	-2	-9	-18	-36	-39	-78
-60°	-5	-7	-14	-28	-30	-60
-30°	-3	-4	-8	-15	-16	-33
0°	0	0	0	0	0	-1
30°	3	4	8	16	17	35
60°	5	7	14	27	30	59
90°	6	9	18	35	38	77

ตามที่ได้กล่าวมาแล้วข้างต้น ในกรณีเดียวกันหากมีอัตราการสุ่มตัวอย่างที่ต่ำความละเอียดของคลื่นเสียงก็จะน้อยตามไปด้วยทำให้มีความละเอียดไม่เพียงพอต่อการนำมาใช้งาน เมื่อนำแต่ละอัตราการสุ่มตัวอย่างมาประมาณค่าทิศทางเสียงและคำนวณหาค่า RMSE ดังตารางที่ 6

ตารางที่ 6 ค่า RMSE ของอัตราการสุ่มตัวอย่างที่ 8 kHz 11 kHz 22 kHz 44 kHz 48k Hz และ 96 kHz

อัตราการสุ่มตัวอย่าง	8 kHz	11 kHz	22 kHz	44 kHz	48 kHz	96 kHz
ค่า RMSE	22.3883	2.6174	1.9700	1.0852	1.3631	0.9612

จากตารางที่ 6 จะเห็นว่าอัตราการสุ่มตัวอย่างที่สูงที่สุดคืออัตราการสุ่มตัวอย่างที่ความถี่ 96 kHz มีค่า RMSE น้อยที่สุด และอัตราการสุ่มตัวอย่างที่ต่ำที่สุดคืออัตราการสุ่มตัวอย่างที่ความถี่ 8 kHz มีค่า RMSE มากที่สุด

สรุปผลการวิจัยและวิจารณ์ผล

จากผลการทดลองการประมาณค่าทิศทางเสียงด้วยการคำนวณสหสัมพันธ์แบบไขว้และลดความซับซ้อนของการประมาณค่าทิศทางเสียงด้วยสมการถดถอยแบบพหุนามในงานวิจัยนี้ สามารถแบ่งผลการทดลองออกเป็น 4 ด้าน ได้แก่ ด้านที่หนึ่งคือด้านอัตราบิต พบว่าอัตราบิตไม่มีผลต่อการคำนวณหาค่าความล้าช้าของเวลา โดยค่าความล้าช้าของเวลาที่มุมเดียวกันที่อัตราบิตขนาด 8 bit/s และ 16 bit/s มีค่าเท่ากัน ด้านที่สองคือด้านระยะทาง ในด้านนี้มีผลต่อการคำนวณค่าความล่าช้าของเวลาเพียงเล็กน้อยเท่านั้น หากต้นกำเนิดเสียงที่ไกลออกไปอยู่ในมุมเดิมพอดี ก็จะทำให้ค่าความล่าช้าของเวลาแต่ละระยะเท่ากัน ส่วนด้านที่สามคือด้านความถี่เสียง จะเห็นว่าค่าความล่าช้าของเวลาของความถี่เสียงผู้ชายและความถี่เสียงผู้หญิงมีผลกระทบต่อการคำนวณค่าความล่าช้าของเวลาน้อยมาก ในแต่ละมุมมีค่าที่ใกล้เคียงกันมาก เนื่องจากการประมาณค่าทิศทางเสียงจะคำนวณสหสัมพันธ์ของเวลาที่เสียงมาถึงหูทั้งสองข้างเท่านั้น ซึ่งเป็นการคำนวณด้านเวลา ดังนั้นด้านความถี่เสียงจึงไม่มีผลกระทบต่อการคำนวณในด้านเวลาและการประมาณค่าทิศทางเสียง ในด้านที่สี่คือด้านอัตราการสุ่มตัวอย่าง ด้านนี้จะมีผลกระทบต่อการคำนวณหาค่าความล่าช้าของเวลาและการประมาณค่าทิศทางเสียงโดยตรง หากกำหนดอัตราการสุ่มตัวอย่างที่ต่ำความละเอียดของสัญญาณเสียงก็จะมีน้อย ทำให้การประมาณค่าทิศทางเสียงไม่มีความแม่นยำพอ ในขณะเดียวกันหากกำหนดอัตราการสุ่มตัวอย่างที่สูง ความละเอียดของสัญญาณเสียงก็มากตามไปด้วย และยังส่งผลให้ค่าที่พยากรณ์ได้มีความแม่นยำมากขึ้น โดยอัตราการสุ่มตัวอย่างที่ความถี่ 96 kHz มีค่า RMSE เพียง 0.96 องศาเท่านั้น

ดั้งนั้นการประมาณค่าทิศทางเสียงแบบ 2 มิติในแนวนอน โดยใช้ฟังก์ชัน เฮด-รีเลท ทรานส์เฟอร์ ในงานวิจัยนี้มีความเหมาะสมแม่นยำ สามารถนำไปใช้งานได้จริง ซึ่งใช้อัลกอริทึมที่ไม่ยากมากเพื่อลดความซับซ้อนของการพัฒนา สามารถนำไปประยุกต์ใช้งานกับหุ่นยนต์โดยให้รับเสียงแบบเวลาจริงได้ เพียงนำค่าสัมประสิทธิ์จากในงานวิจัยนี้กับค่าสหสัมพันธ์ที่ได้มาคำนวณในสมการพหุนามระดับดีกรี และในการบันทึกเสียงควรอยู่ในสภาพแวดล้อมเดียวกันกับในงานวิจัยนี้ ผลลัพธ์ของการคำนวณจะประมาณค่าทิศทางเสียงเป็นมุมองศา สามารถนำไปใช้งานได้ทันที ช่วยให้ใช้เวลาในการประมวลผลมีความรวดเร็วโดยไม่ต้องคำนวณหาค่าสัมประสิทธิ์ใหม่ทุกครั้งที่ประมาณค่าทิศทางเสียง นอกจากนี้ยังสามารถนำไปประยุกต์ใช้กับระบบรู้จำเสียงพูดเพื่อให้หุ่นยนต์เคลื่อนไปยังทิศทางต้นกำเนิดเสียงผู้ที่เป็นเจ้าของได้

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณทางภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่นที่สนับสนุนอุปกรณ์สำหรับการวิจัยในครั้งนี้

เอกสารอ้างอิง

ชมพูนุช วีระวัธนชัย. (2549). การทำนายระดับยาคาร์บาร์มาซีปีนในซีรั่มที่สภาวะคงที่ โดยแบบจำลองประชากรไทย. วิทยานิพนธ์เภสัชศาสตรมหาบัณฑิต, มหาวิทยาลัยศิลปากร. นครปฐม: 126 หน้า

Andrew, I.R. (2002). Regular and Irregular Signal Resampling. *In* partial fulfillment of the requirements for the degree of Doctor of Philosophy. Massachusetts: Massachusetts Institute of Technology. pp. 69-98

Daejung, S., Seung, N., Jin, K., Seong, B. and So, M. (2008). Estimation of sound direction for improved interaction of entertainment dolphin robots. In: The international conference on Software Engineering proceeding of IEEE International Symposium 10-18 May 2008. IEEE, Germany. 1-4.

Bracewell, R. (1965). Pentagram Notation for Cross Correlation. The Fourier Transform and Its Applications. New York: McGraw-Hill. pp. 46 and 243.

Funkhouser, H. (1930). A Short Account of the History of Symmetric Functions of Roots of Equations. American Mathematical Monthly 37(7): 357-365.

Hwang, S., Park, Y. and Park, Y.-s. (2011). Sound direction estimation using an artificial ear for robots. Robotics and Autonomous Systems 59(2011): 208-217.

❑❑❑❑❑