การวิจัยเชิงสำรวจเรื่องการประมาณค่าทิศทางเสียงในแนวดิ่ง
A Survey on Vertical Sound Direction Determination

ปิยณัฐ ศิริสวัสดิ์ (Piyanat Sirisawat)[1]
[1]ภาควิชาวิทยาการคอมพิวเตอร์คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
piyanat_siri@kkumail.com

บทคัดย่อ

ประสาทการรับรู้เสียงของมนุษย์ สามารถทำให้มนุษย์รับรู้เสียงรอบตัวได้ นอกจากนี้ยังสามารถบอกได้ว่า เสียงนั้นมาจากทิศทางไหน โดยไม่จำเป็นต้องพึ่งประสาทการรับรู้อื่น ๆ ช่วย โดยมีทฤษฎีที่เรียกว่า Duplex Theory ซึ่งใช้หลักการของความแตกต่างของเวลาที่มาถึง (ITD) และความแตกต่างของความดังเสียงที่มาถึง (IID) ของเสียงที่เข้ามายังหูทั้งสองข้างในการบอกทิศทาง อย่างไรก็ตาม ทฤษฎีนี้ยังไม่สามารถใช้บอกเสียงที่มาจากทิศทางในแนวดิ่งได้ จึงอาจจะมีปัจจัยอื่น ๆ ที่มีความสำคัญต่อการรับรู้ทิศทางเสียงในแนวดิ่งของมนุษย์ โดยมีการศึกษาถึงการรับรู้ทิศทางเสียงในแนวดิ่ง ทั้งจากการใช้งานไมโครโฟนหลายตัว ไปจนถึงการใช้งานใบหูมนุษย์เทียมขึ้นมา โดยพยายามใช้ทฤษฎี Pinna Filtering Effect ในการช่วยประมาณค่าทิศทางเสียงในแนวดิ่ง

คำสำคัญ: การประมาณค่าทิศทางเสียง, ใบหูเทียม, TDOA, HRTF, ทิศทางเสียงในแนวดิ่ง

Abstract

Human sense of hearing can also be used to determine the direction of the sound even without the aid of visual perception. There are duplex theory that human use the Interaural Time Difference (ITD) and Interaural Intensity Difference (IID) to tell the direction of the sound. However, the duplex theory cannot be used to tell the sound direction in median plane or any vertical direction. There must be other factors played an important role in median plane sound localization. There are studies explore the possibility of human pinna that have an impact on vertical sound direction estimation. They call it Pinna Filtering Effect Theory. There also another approach that using microphones array. We will explore the model or improvement from old theory or possibility for the existing model to be use in median plane sound localization.

Keyword: Sound Direction Determination, Artificial Ear, TDOA, HRTF, Vertical sound localization.

1. บทนำ

ประสาทการรับรู้เสียงของมนุษย์นั้น นอกจากจะสามารถทำให้มนุษย์เรารับรู้ถึงเสียงและทำความเข้าใจถึงเสียงต่าง ๆ รอบข้างได้แล้ว ยังสามารถทำให้มนุษย์เราบอกถึงทิศทางเสียงที่ได้ยินได้อีกด้วย โดยใช้ทฤษฎีที่เรียกว่า Duplex Theory ซึ่งมี 2 หลักการ คือ ความแตกต่างของเวลาที่มาถึง (Interaural Time Difference : ITD) และความแตกต่างของความดังเสียงที่มาถึง (Interaural Intensity Difference : IID) ซึ่งสามารถทำให้หุ่นยนต์สามารถบอกทิศทางเสียงในแนวระนาบได้โดยใช้ไมโครโฟนเพียง 2 ตัวเท่านั้น โดยใช้ความแตกต่างของเวลาที่มาถึง (Time Difference of Arrival : TDOA) ถึงอย่างไรก็ตาม ยังคงมีปัญหาเรื่องความสับสนระหว่างหน้า-หลัง (Front-Back Confusion) ของเสียงที่เข้ามาในระบบ และไม่สามารถบอกทิศทางเสียงในแนวดิ่งได้

จึงมีการนำเสนอทฤษฎีอื่น ๆ โดยบอกว่า ใบหูของมนุษย์มีผลต่อเสียงที่ได้ยิน และเป็นตัวหลักที่ทำให้มนุษย์สามารถรับรู้ทิศทางเสียงได้ โดยเรียกว่า Pinna Filtering Effect ซึ่งบอกว่า การที่เสียงกระทบใยหูมนุษย์ ทำให้

ลักษณะของเสียงเปลี่ยนไป และเนื่องจากเสียงที่มาจากต่างทิศทางกันจะกระทบใบหูไม่เหมือนกัน จึงทำให้มนุษย์สามารถบอกทิศทางของเสียงได้จากความแตกต่างนั้น [1] และพบว่าเสียงที่กระทบใบหูนี้ จะมีลักษณะความถี่ที่เปลี่ยนแปลงไป และเรียกการเปลี่ยนแปลงของความถี่เหล่านั้นว่า Head-Related Transfer Function (HRTF) โดยลักษณะการเปลี่ยนแปลงที่สำคัญต่อการประมาณค่าทิศทางในแนวดิ่งคือ การลดลงของความถี่บางช่วงหรือ Spectral Notch

ถึงกระนั้น HRTF ยังไม่สามารถนำมาใช้งานในการบอกทิศทางของเสียงได้โดยตรง เนื่องจากคุณสมบัติของเสียงที่แต่ละคนมีความถี่ไม่เท่ากัน หรือคนเดียวกันพูดต่างคำกันก็มีความถี่ที่แตกต่างกัน และค่า HRTF ก็จะเปลี่ยนแปลงไปตามแต่ละคำพูดและแต่ละทิศทาง จึงเป็นความยากที่นำมาใช้งานโดยตรง ดังนั้น Survey นี้จึงจะทำการค้นหาถึงการศึกษาต่าง ๆ เกี่ยวกับการรับรู้ทิศทางเสียงในแนวดิ่ง และความเป็นไปได้ที่จะพัฒนาวิธีการรับรู้ทิศทางเสียงในแนวดิ่งโดยการใช้งานวิธีการต่าง ๆ

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 Duplex Theory

ทฤษฎีนี้ กล่าวถึงความแตกต่างของเสียงทั้ง 2 ข้าง โดยจะเน้นในเรื่องของเวลาที่แตกต่างกัน (ITD) และความดังของเสียงที่แตกต่างกัน (IID)

เวลาที่แตกต่างกันนั้น เกิดขึ้นเนื่องจากเสียงมีความเร็วในการเดินทางในอากาศด้วยความเร็ว 343 เมตรต่อวินาที และเนื่องจากเสียงต้องใช้เวลาในการเดินทาง ทำให้เวลาที่เข้าถึงหูด้านซ้ายและด้านขวาไม่เท่ากัน และความแตกต่างของเวลานี้จะขึ้นอยู่กับทิศทางเสียงที่เข้ามายังใบหูทั้งสองข้าง เช่น หากเสียงมาจากด้านขวา เสียงจะเข้าสู่หูขวาก่อนหูซ้าย และกลับกันหากเสียงมาจากด้านซ้าย เสียงก็จะเข้าสู่หูซ้ายก่อนหูขวา แต่หากเสียงมาจากด้านหน้า เสียงจะเข้าสู่หูทั้ง 2 ข้างในเวลาใกล้เคียงกัน

ความดังเสียงที่แตกต่างกันนั้นก็มีความใกล้เคียงกับเวลาที่แตกต่างกัน กล่าวคือ เสียงที่เดินทางในอากาศนั้นจะสูญเสียระดับความดังของเสียงไปตามระยะทางที่เสียงต้องเดินทาง ทำให้การได้ยินเสียงนั้นจะได้ยินเพียงระยะทางหนึ่งเท่านั้น หากเกินระยะทางนั้นแล้วจะไม่ได้ยิน หรือได้ยินเสียงเบามาก ดังนั้น ใบหูด้านที่ใกล้กับแหล่งกำเนิดเสียง ก็จะได้ยินเสียงในระดับที่ดังมากกว่าใบหูที่ห่างออกไปจากแหล่งกำเนิดเสียง และเสียงที่มาจากด้านหน้าก็จะมีระดับเสียงเท่า ๆ กันในใบหูทั้ง 2 ข้าง เนื่องจากระยะทางใกล้เคียงกัน

ถึงแม้ว่าทั้งเวลาที่แตกต่างกันและความดังของเสียงที่แตกต่างกัน จะมีความสามารถในการบอกทิศทางเสียงที่ดีและสามารถบอกทิศทางเสียงได้โดยไม่สนใจปัจจัยอื่น ๆ แต่ทฤษฎีนี้สามารถบอกได้เพียงเสียงในแนวระนาบ และไม่สามารถบอกได้ว่าเสียงนั้นมาจากด้านหน้าหรือด้านหลัง เนื่องจากเสียงที่มาจากด้านหน้าและด้านหลังนั้นจะมีค่าเวลาที่แตกต่างกันและความดังของเสียงที่แตกต่างกันใกล้เคียงกันมาก จึงทำให้ระบบเกิดความสับสนของเสียงที่เข้ามาจากด้านหน้าและด้านหลัง

2.2 Pinna Filtering Effect

ทฤษฎีนี้ได้กล่าวถึงลักษณะของใบหูมนุษย์ที่มีผลต่อลักษณะของเสียงที่เข้ามายังใบหูมนุษย์ โดยกล่าวว่าใบหูของมนุษย์นั้นนอกจากจะรับเสียงรอบข้างเข้ามายังหูส่วนกลางแล้ว ลักษณะการโค้งเว้าที่แตกต่างกันในแต่ละส่วนของใบหู ยังทำให้ลักษณะของเสียงที่เข้ามากระทบใบหูแตกต่างกันออกไปตามทิศทางของเสียงที่เข้ามากระทบอีกด้วย ถึงอย่างไรก็ตาม การที่จะศึกษาถึงลักษณะของใบหูมนุษย์และลักษณะของเสียงที่แตกต่างกันนั้นทำได้ยาก เนื่องจากมนุษย์มีลักษณะของใบหูที่แตกต่างกันออกไปในแต่ละบุคคล ถึงอย่างไรก็ตาม การศึกษาของเปลี่ยนลักษณะโครงสร้างของใบหูมนุษย์ [2] พบว่า ในวันแรกที่มีการเปลี่ยนโครงสร้างใบหูมนุษย์ ผู้ทดสอบสูญเสียความสามารถในการบอกทิศทางเสียงในแนวดิ่งไปโดยสิ้นเชิง แต่ยังคงความสามารถในการบอกทิศทางเสียงในแนวระนาบ และเมื่อเวลาผ่านไป ผู้ทดสอบสามารถกลับมาบอกทิศทางเสียงในแนวดิ่งได้ เหมือนกับการปรับตัวและเรียนรู้ลักษณะของใบหูใหม่ที่เปลี่ยนแปลงไปและสามารถใช้งานได้เกือบเหมือนปกติ

2.3 Head-Related Transfer Function (HRTF)

การศึกษาถึงลักษณะของใบหูที่มีผลต่อลักษณะของเสียงที่รับเข้ามาในใบหูชั้นกลาง ได้มีการนำเอาการวิเคราะห์ความถี่เข้ามาใช้งานร่วมด้วย เนื่องจากเป็นลักษณะการวิเคราะห์ลักษณะของหูแต่ละข้าง จึงไม่สามารถนำเอา Duplex Theory มาใช้งานได้ และการวิเคราะห์ลักษณะของความถี่ก็แสดงให้เห็นถึงลักษณะความถี่ที่แตกต่างกัน ของเสียงเดียวกันที่มาจากความดิ่งที่ต่างทิศทางกัน โดยแสดงออกในลักษณะของความถี่บางช่วงที่มีการเพิ่มเติม (Spectral Peak) หรือความถี่บางช่วงที่มีการลดลง (Spectral Notch) แม้ว่าจะเป็นเสียงลักษณะเดียวกัน แต่มาจากต่างทิศทางกัน

2.4 การศึกษาแม่แบบใบหูมนุษย์

การวิเคราะห์หรือการที่จะนำเอาใบหูมนุษย์มาใช้งานนั้นจำเป็นต้องมีการศึกษาถึงลักษณะของใบหูเทียมและความแตกต่างระหว่างใบหูมนุษย์จริงและใบหูมนุษย์เทียม โดยพบว่า แม้จะใช้งานใบหูเทียม ลักษณะของ Spectral Notch ที่เป็นลักษณะสำคัญในการค้นหาทิศทางเสียงในแนวดิ่งนั้น ยังคงสามารถพบได้ และสามารถนำมาใช้งานเพื่อค้นหาทิศทางเสียงในแนวดิ่งได้ [3]

นอกจากนี้ยังมีความพยายามที่จะทำการศึกษาและสร้างแม่แบบกลางสำหรับทำการศึกษาโครงสร้าง HRTF โดยรวมสำหรับการใช้งานกับลักษณะใบหูในทุกรูปแบบ [4] แต่ผลลัพธ์นั้นพบว่า เนื่องจากความแตกต่างของลักษณะใบหูที่มากในแต่ละบุคคล ทำให้โครงสร้างลักษณะ HRTF แตกต่างกัน และแม่แบบกลางนั้นไม่สามารถนำไปใช้แทนที่ลักษณะ HRTF ของมนุษย์ได้

เนื่องจากลักษณะ HRTF ของมนุษย์แต่ละคนที่แตกต่างกัน นำไปสู่การศึกษาถึงการแยกเอา Spectral Notch ออกมาจากรูปภาพลักษณะของใบหูร่วมกับลักษณะของ HRTF ของเสียงที่เข้ามายังใบหู [5] ซึ่งนำไปสู่การแยก Spectral Notch ที่สำคัญต่อการประมาณค่าทิศทางเสียงในแนวดิ่งที่รวดเร็ว และแม่นยำมากขึ้น

นอกจากนี้ยังมีการศึกษาถึงความสัมพันธ์ของ Spectral Notch 3 ตัวแรกกับการค้นหาทิศทางในแนวดิ่งด้านหน้า [6] โดยพบว่า Spectral Notch ตัวที่ 1 มีความสัมพันธ์ในแบบเส้นตรงและมีความเป็นไปได้ที่จะทำการสร้างความสัมพันธ์นี้จากภาพถ่ายของใบหู และเป็นไปได้ที่จะใช้ Spectral Notch เหล่านี้ค้นหาทิศทางในแนวดิ่ง ถึงอย่างไรก็ตาม เนื่องจากความแตกต่างของ HRTF ในแต่ละบุคคล จึงต้องมีการศึกษากับผู้ทดสอบจำนวนที่มากขึ้น

2.5 การประมาณค่าทิศทางเสียง

การประมาณค่าทิศทางเสียงในแบบปกตินั้น จะใช้ไมโครโฟนอย่างน้อย 2 ตัวขึ้นไปในการค้นหาทิศทางเสียงเพื่อที่จะสามารถใช้การหา ITD และ IID ได้ และประมาณค่าทิศทางเสียงในแนวระนาบได้ แต่เนื่องจากการมีความสับสนระหว่างหน้า-หลัง การใช้งานไมโครโฟนเพียง 2 ตัว จึงไม่สามารถใช้งานในแนวระนาบทั้งหมดได้ และได้เพียงด้านใดด้านหนึ่งเท่านั้น ถึงอย่างนั้นยังมีการศึกษาเพื่อลดความซับซ้อนในการคำนวณโดยการใช้งาน CORDIC [7] บนไมโครโฟนจำนวน 5 ตัว โดยระบบสามารถลดการคำนวณลงมากได้ถึง 65 เท่าและยังคงความแม่นยำเอาไว้

นอกจากการใช้งาน ITD และ IID แล้ว การศึกษาในปัจจุบันยังมีการนำเอา Steered response power phase transform (SRP-PHAT) ซึ่งเป็นวิธีการที่มีความแม่นยำสูงในการประมาณค่าทิศทางเสียง โดยการใช้ไมโครโฟนหลายตัวเรียงกันเป็นรูปแบบวงกลมในแนวระนาบและค้นหาว่าเสียงกระทบไมโครโฟนใดก่อนและมีระดับเสียงแตกต่างกันเท่าใด แต่เนื่องจากการค้นหาทิศทางเสียงที่ต้องเทียบไมโครโฟนหลายตัว จึงมีการศึกษาเพื่อลดความซับซ้อนในการคำนวณลงโดยใช้งาน two level search space clustering (TL-SSC) [8] ทำให้ระบบสามารถค้นหาทิศทางเสียงได้อย่างรวดเร็วบนไมโครโฟนหลายตัวและมีความแม่นยำมากยิ่งขึ้น ถึงอย่างไรก็ตาม SRP-PHAT นั้นสามารถใช้งานในระนาบใดระนาบหนึ่งเท่านั้น จึงมีความเป็นไปได้ที่จะใช้งาน SRP-PHAT ในแนวดิ่งได้เช่นกัน

2.6 การประมาณค่าทิศทางเสียงโดยใช้ใบหูมนุษย์เทียม

การศึกษาโดยใช้ใบหูมนุษย์เทียมนั้น จะเน้นไปที่การค้นหาทิศทางเสียงโดยใช้ใบหูข้างเดียวหรือการค้นหาทิศทางเสียงในแนวดิ่ง โดยการศึกษาการประมาณค่าทิศทางเสียงด้วยหูข้างเดียวโดยใช้งาน monaural modulation spectrum (MMS) [9] โดยพบว่า ลักษณะของ

MMS มีความสัมพันธ์กับทิศทางเสียงที่เข้ามายังใบหูและทิศทางเสียงในแนวดิ่งก็ทำให้มีความแตกต่างในค่า MMS เช่นกัน จึงความเป็นไปได้ที่จะสามารถใช้งานวิธีการนี้ในการค้นหาทิศทางเสียงในแนวดิ่งได้ หรือแม้แต่การค้นหาทิศทางเสียงในแนวระนาบโดยใช้หูเพียงข้างเดียว

แม้ว่าการศึกษาโดยใช้ใบหูมนุษย์จะเน้นไปที่การใช้งานไมโครโฟนไม่เกิน 2 ตัว การศึกษาของ Keyrouz [10] มีการใช้งานไมโครโฟนเพิ่มอีก 1 คู่รวมเป็น 4 ตัว โดยจะมีไมโครโฟนด้านในใบหู 2 ตัวและด้านนอกใบหู 2 ตัว และใช้งานเสียงที่เข้ามายังไมโครโฟนเหล่านั้นในการสร้าง Database HRTF สำหรับการค้นหาทิศทางเสียง โดยตัวระบบสามารถค้นหาทิศทางเสียงได้ทั้งในแนวระนาบและแนวดิ่ง ขึ้นอยู่กับเสียงและ Database HRTF ที่ใช้งาน

ถึงอย่างนั้น ก็ยังมีการศึกษาถึงการค้นหาทิศทางเสียงโดยใช้ไมโครโฟนเพียง 2 ตัว และมีความแม่นยำที่สูง [11] ซึ่งใช้งาน ITD สำหรับทิศทางในแนวระนาบและ IID ร่วมกับ Spectral Notch ในการค้นหาเสียงในแนวดิ่ง และเมื่อเทียบกับวิธีการอื่น ๆ แล้ว มีความผิดพลาดที่ต่ำลงกว่ามาก ถึงอย่างไรก็ตาม การใช้งาน Spectral Notch นั้นขึ้นกับ HRTF และขึ้นกับ Database ที่ใช้งาน จึงไม่สามารถนำไปใช้งานกับเสียงทั่วไปได้

ตารางที่ 1: เปรียบเทียบผลการศึกษาทิศทางเสียงในแนวดิ่ง

อ้างอิง	Database	วิธีการ	ผลลัพธ์
[4]	CIPIC	HRTF Modeling	ความผิดพลาด 3 - 15°
[5]	CIPIC	Spectral Notch Modeling	RMSE = 0.59 – 1.27 kHz
[8]	TIMIT	TL-SSC	ความผิดพลาดน้อยกว่า 1°
[10]	HRTF	GCC และ Bayesian network	ความแม่นยำ 84%
[11]	IRCAM และ RIEC	Weighted interaural และ spectral cues	ความผิดพลาดลดลง

จากตารางที่ 1 จะเห็นว่าในแต่ละงานวิจัยที่ยกมานั้น มีการรายงานความแม่นยำหรือความผิดพลาดที่แตกต่างกัน นอกจากนี้งานวิจัยที่ทำในแนวดิ่งนั้นส่วนมากมีน้อย หรือเน้นการใช้งานไมโครโฟนหลายตัว ซึ่งโดยหลักการแล้วไม่แตกต่างไปจากการใช้ ITD และ IID การ Survey นี้จึงต้องการเน้นที่การใช้งานวิธีการที่แตกต่างออกไปจากการใช้งาน ITD และ IID ปกติ

นอกจากที่แสดงในตารางที่ 1 แล้ว การวิเคราะห์ลักษณะของใบหูยังมีความสำคัญในการแยกค่า Spectral Notch ที่มีความสำคัญในการประมาณค่าทิศทางเสียงในแนวดิ่ง และสามารถนำไปสร้าง Database สำหรับการประมาณค่าทิศทางเสียงได้เช่นกัน

3. การอภิปรายผล

การศึกษาการประมาณค่าทิศทางเสียงในแนวดิ่งนั้น มีการศึกษาที่หลากหลาย ทั้งการแยกค่าที่มีความสำคัญค่อการประมาณค่าทิศทางเสียงในแนวดิ่ง (HRTF, Spectral Notch) รวมถึงการลดการคำนวณหรือลดขั้นตอนการคำนวณ ทำให้การคำนวณมีความรวดเร็วมากขึ้น และมีความแม่นยำมากขึ้นในบางวิธีการ ทำให้การใช้งานไมโครโฟนหลายตัวที่มีข้อเสียที่จำเป็นต้องมีการคำนวณสูงนั้นสามารถใช้งานได้ง่ายและรวดเร็วขึ้น

ขณะที่การศึกษาการใช้งานใบหูมนุษย์ในการประมาณค่าทิศทางเสียงนั้นยังคงพบปัญหาเดิม ๆ คือ Database ที่ใช้งานนั้นสามารถใช้งานกับใบหู หรือ ใช้งานกับเสียงที่ใช้ในการศึกษานั้น ๆ เท่านั้น แม้ว่าจะมีความพยายามในการเพิ่มความยืดหยุ่นและสามารถนำไปใช้กับเสียงอื่น ๆ ได้บางส่วนก็ตาม ก็ยังคงมีความจำกัดในการใช้งานเพื่อความแม่นยำที่สูงขึ้น

ดังนั้น การศึกษาการประมาณค่าทิศทางเสียงในแนวดิ่งจึงมีแนวทาง 2 แนวทางหลักคือ การใช้งานไมโครโฟนหลายตัวในระบบ โดยจัดเรียงให้ระบบมีความสามารถในการประมาณค่าทิศทางเสียงในแนวดิ่งร่วมด้วย หรือการใช้งานใบหูมนุษย์เพื่อลดการใช้งานไมโครโฟนและการแยกค่า HRTF สำหรับการค้นหาทิศทางเสียงที่อยู่นอกเสียงจาก Database ได้

เอกสารอ้างอิง

[1] D. W. Batteau and H. E. Huxley, “The role of the pinna in human localization,” *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 168, no. 1011, pp. 158–180, Aug. 1967.

[2] P. M. Hofman, J. G. A. V. Riswick, and A. J. V. Opstal, “Relearning sound localization with new ears,” *Nature Neuroscience*, vol. 1, no. 5, pp. 417–421, Sep. 1998.

[3] S. Hwang, K. H. Shin, and Y. Park, “Artificial Ear for Robots,” in *2006 5th IEEE Conference on Sensors*, 2006, pp. 1460–1463.

[4] M. Geronazzo, A. Carraro, and F. Avanzini, “Evaluating vertical localization performance of 3D sound rendering models with a perceptual metric,” in *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, 2015, pp. 1–5.

[5] A. Sohni, C. Ahuja, and R. M. Hegde, “Extraction of pinna spectral notches in the median plane of a virtual spherical microphone array,” in *2014 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2014, pp. 142–146.

[6] S. Spagnol and F. Avanzini, “Frequency estimation of the first pinna notch in Head-Related Transfer Functions with a linear anthropometric model,” 2015.

[7] A. N. Dwiputra, R. H. G. Nainggolan, and M. A. M. Nasution, “3 dimension sound source localization with cross-correlation and CORDIC algorithm on FPGA,” in *2016 International Symposium on Electronics and Smart Devices (ISESD)*, 2016, pp. 365–370.

[8] D. Yook, T. Lee, and Y. Cho, “Fast Sound Source Localization Using Two-Level Search Space Clustering,” *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 20–26, Jan. 2016.

[9] D. Morikawa, M. Ando, and M. Unoki, “Feasibility of Estimating Direction of Arrival Based on Monaural Modulation Spectrum,” in *2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 2015, pp. 384–387.

[10] F. Keyrouz, “Advanced Binaural Sound Localization in 3-D for Humanoid Robots,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 9, pp. 2098–2107, Sep. 2014.

[11] B. R. Hammond and P. J. B. Jackson, “Robust Full-sphere Binaural Sound Source Localization Using Interaural and Spectral Cues,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 421–425.