บทที่ 2

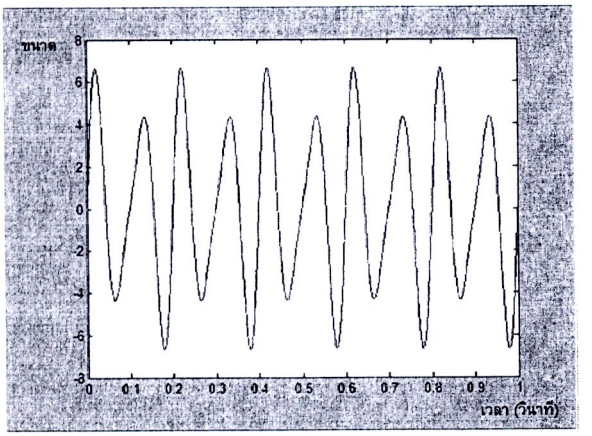
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

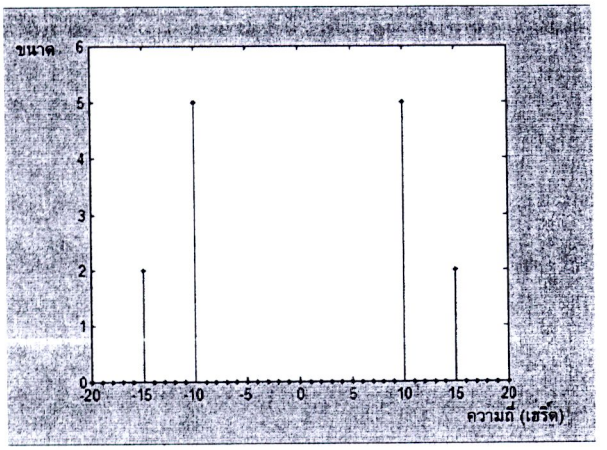
ในส่วนนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับสัญญาณเสียง ซึ่งสามารถแบ่งได้เป็นทฤษฎีเกี่ยวกับการวิเคราะห์สัญญาณเสียงพูดโดยใช้หลักการของการแปลงฟูริเยร์แบบวิยุต (Discrete Fourier Transform) สเปกโตรแกรม (Spectrogram) อัตสหสัมพันธ์ (Autocorrelation) และทฤษฎีที่เกี่ยวข้องกับการเปล่งเสียงชนิดต่าง ๆ

2.1.1 การแปลงฟูริเยร์แบบวิยุต [6]

การแปลงฟูริเยร์ใช้ในการเปลี่ยนแปลงสัญญาณที่อยู่ในโดเมนเวลา ให้กลายเป็นโดเมนความถี่ โดยการวิเคราะห์สัญญาณในโดเมนความถี่จะเหมาะสมกับสัญญาณที่มีลักษณะเป็นรายคาบ เพราะสามารถแยกองค์ประกอบความถี่ต่าง ๆ ออกมาอย่างเห็นได้ชัด อาทิเช่น ในรูปที่ 2.1 เป็นสัญญาณในโดเมนเวลาที่ประกอบด้วยสัญญาณของฟังก์ชัน sin ที่ความถี่ 10 เฮิรตซ์ และมีขนาดเป็น 5 และสัญญาณของฟังก์ชัน sin ที่มีความถี่ 15 เฮิรตซ์ และมีขนาดเป็น 2 เมื่อนำสัญญาณในรูปที่ 2.1 ไปผ่านการแปลงฟูริเยร์แบบต่อเนื่อง (Continuous Fourier Transform) ส่งผลให้ได้ผลลัพธ์ดังรูปที่ 2.2 ซึ่งอยู่ในโดเมนความถี่ และสามารถแสดงให้เห็นถึงองค์ประกอบในความถี่ต่าง ๆ ได้



รูปที่ 2.1 สัญญาณในโดเมนเวลา



รูปที่ 2.2 สัญญาณในโดเมนความถี่

สมการสำหรับการแปลงฟูริเยร์แบบต่อเนื่อง เป็นดังสมการที่ 2.1 โดย $f(t)$ แทนถึงสัญญาณแบบต่อเนื่องที่รับเข้าในการแปลงฟูริเยร์แบบต่อเนื่อง และ $f(v)$ เป็นผลลัพธ์จากการแปลงฟูริเยร์แบบต่อเนื่อง

$$f(v) = \mathcal{F}_t[f(t)](v) = \int_{-\infty}^{\infty} f(t)e^{-2\mu ivt}dt \qquad (2.1)$$

จากสมการที่ 2.1 แสดงให้เห็นว่าการแปลงฟูริเยร์แบบต่อเนื่องไม่สามารถนำมาใช้ได้จริงในการคำนวณทางคอมพิวเตอร์ เนื่องจากสัญญาณที่รับเข้าต้องเป็นสัญญาณแบบต่อเนื่อง และสัญญาณต้องมีความยาวอนันต์ จึงได้มีการนำการแปลงฟูริเยร์แบบต่อเนื่องมาปรับปรุงให้สามารถใช้ในการคำนวณทางคอมพิวเตอร์ได้ โดยเรียกว่าการแปลงฟูริเยร์แบบวิยุต ซึ่งมีสมการเป็นดังสมการที่ 2.2 โดย f_k แทนถึงลำดับของสัญญาณรับเข้า และ f_n แทนถึงผลลัพธ์จากการแปลงฟูริ-เยร์แบบวิยุต

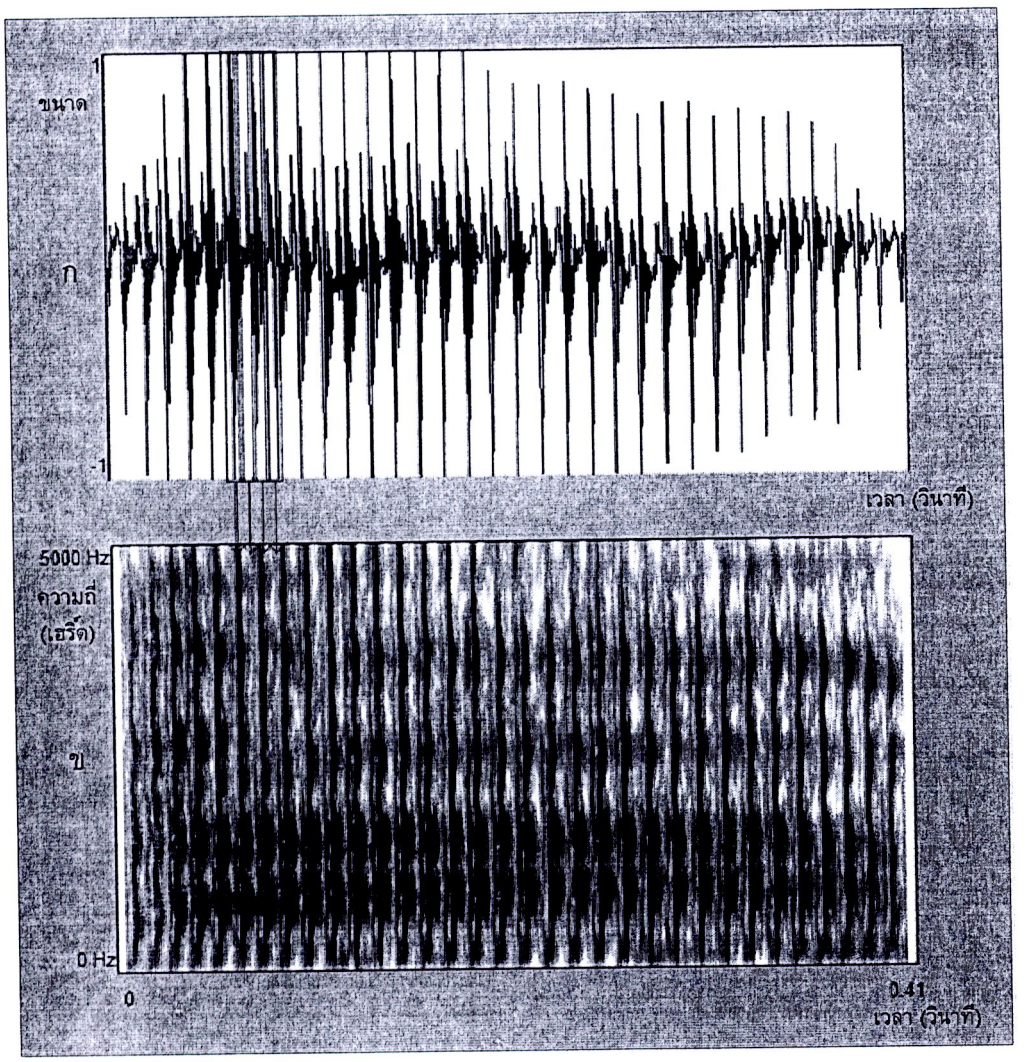
$$Fn \equiv \sum_{k=0}^{n-1} F_k e^{-2\pi ink/n} \qquad (2.2)$$

2.1.2 สเปกโตรแกรม [6]

สเปกโตรแกรมคือเครื่องมือที่ช่วยในการวิเคราะห์สัญญาณเสียง เพราะสามารถแสดงให้เห็นถึงการเปลี่ยนแปลงความเข้มของสัญญาณในช่วงความถี่ต่าง ๆ เทียบกับเวลา และแสดงภาพรวมของสัญญาณในโดเมนความถี่ จึงทำให้ง่ายต่อการวิเคราะห์

สเปกโตรแกรมมีลักษณะเป็นกราฟโดยแกนนอนแทนถึงเวลา และแกนตั้งแทนถึงความถี่ต่าง ๆ ดังแสดงในรูปที่ 2.3 วิธีการอ่านค่าของสเปกโตรแกรมในเวลาใด ๆ สามารถทำได้โดยเลือกเวลาที่ต้องการจากแกนนอน โดยค่าตามแกนตั้ง แสดงถึงความเข้มของสัญญาณในช่วงความถี่ต่าง ๆ ตามที่ระบุในแกนตั้ง โดยใช้สีต่าง ๆ เพื่อแสดงถึงความเข้มของสัญญาณ โดยในที่นี้ได้กำหนดให้สีขาวแทนถึงความเข้มของสัญญาณน้อย และสีดำแทนถึงความเข้มของสัญญาณมาก

การได้มาของสเปกโตรแกรมในแต่ละช่วงเวลานั้น เกิดจากการนำสัญญาณในโดเมนเวลามาแบ่งเป็นกรอบของสัญญาณ (Frame) ที่มีความกว้างเท่ากัน โดยความกว้างของหน้าต่างนั้นสามารถกำหนดได้โดยผู้ใช้ จากนั้นนำหน้าต่างเหล่านั้นคำนวณฟูริเยร์ในช่วงเวลาสั้น เพื่อให้ได้เป็นสัญญาณในโดเมนความถี่ของหน้าต่างนั้น แล้วนำสัญญาณในโดเมนความถี่นั้นไปวาดลงในสเปกโตรแกรม



รูปที่ 2.3 ความสัมพันธ์ระหว่างสัญญาณทางเวลา และสเปกโตรแกรม

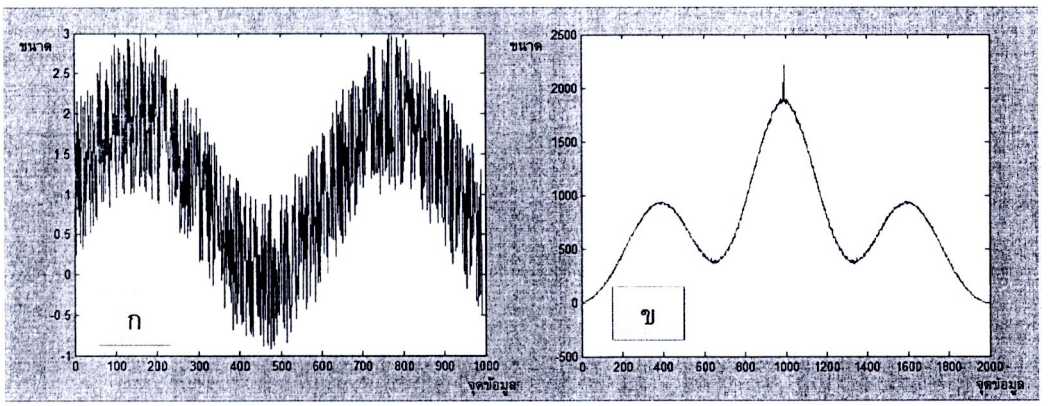
2.1.3 อัตสหสัมพันธ์ [6]

ผลลัพธ์ของการหาค่าอัตสหสัมพันธ์ สามารถบ่งบอกได้ถึงรูปแบบที่ซ้ำกันที่เกิดขึ้นในฟังก์ชัน หรือเพื่อใช้ในการหาความถี่มูลฐานของฟังก์ชัน การหาค่าอัตสหสัมพันธ์ระหว่างฟังก์ชั่นที่ไม่ต่อเนื่อง (Discrete Function) $f[n]$ และ $g[n]$ เป็นไปตามสมการที่ 2.3

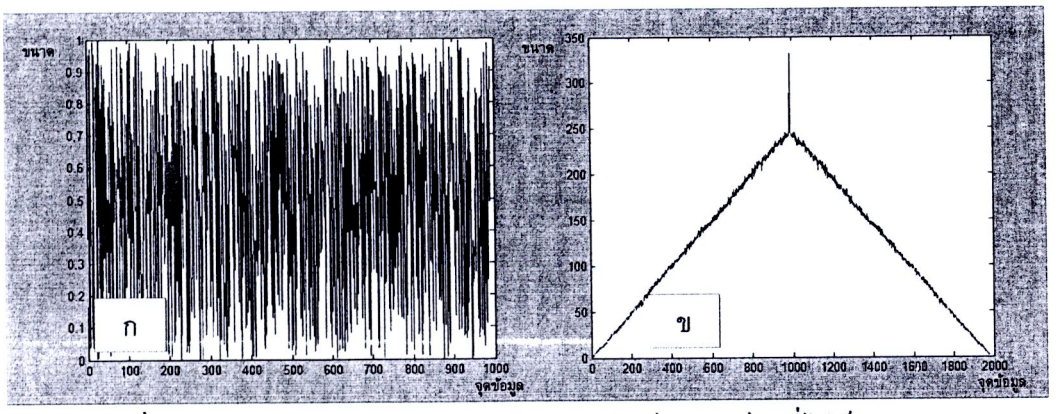
$$f[n] * g[n] = \sum_{m=-\infty}^{\infty} f[m]g[n+m] \qquad (2.3)$$

จากสมการ 2.3 แสดงให้เห็นว่าวิธีการหาค่าอัตสหสัมพันธ์ มีวิธีการคำนวณเหมือนกับการหาค่าสังวัตนาการ (Convolution) แต่ต่างในส่วนของการกลับค่าของโดเมนของฟังก์ชัน เพราะการหาค่าสังวัตนาการ ต้องทำการกลับค่าของโดเมนของฟังก์ชันก่อนแล้วจึงนำมาบวกรวมกัน ต่างจากการหาค่าอัตสหสัมพันธ์ ที่ไม่ต้องทำการกลับค่าของโดเมนของฟังก์ชัน

ผลลัพธ์ของการคำนวณค่าอัตสหสัมพันธ์ของฟังก์ชั่น จะแสดงให้เห็นถึงรูปแบบที่ซ้ำกัน ดังเช่นในรูปที่ 2.4ก แสดงให้เห็นถึงฟังก์ชันที่ประกอบด้วยฟังก์ชัน sin และสัญญาณรบกวน และรูปที่ 2.4ข คือการนำฟังก์ชันดังกล่าวไปผ่านการคำนวณอัตสหสัมพันธ์ จะพบว่าฟังก์ชันที่ผ่านการทำอัตสหสัมพันธ์ มีรูปแบบที่ซ้ำกันโดยระยะห่างระหว่างจุดสูงสุดสัมพัทธ์สามารถนำมาคำนวณหาความถี่มูลฐานได้ ต่างจากรูปที่ 2.5ก ที่มีแค่เพียงสัญญาณรบกวน และเมื่อผ่านการคำนวณค่าอัต-สหสัมพันธ์ จะพบว่าเป็นเพียงสัญญาณที่ไม่ใช่รายคาบ ดังแสดงในรูปที่ 2.5ข



รูปที่ 2.4 ผลลัพธ์จากการคำนวณหาค่าอัตสหสัมพันธ์ของฟังก์ชันรายคาบ



รูปที่ 2.5 ผลลัพธ์จากการคำนวณหาค่าอัตสหสัมพันธ์ของฟังก์ชันที่ไม่เป็นรายคาบ

สำหรับวิธีการหาความถี่มูลฐาน (Fundamental Frequency, f_0) ของฟังก์ชัน สามารถพิจารณาได้จากฟังก์ชันที่ผ่านการคำนวณค่าอัตสหสัมพันธ์มาแล้ว โดยระยะห่างระหว่างจุดสูงสุดสัมพัทธ์ที่อยู่ติดกันซึ่งหมายถึงคาบของฟังก์ชัน (t) โดยค่าที่วัดได้มีหน่วยเป็นจุดตัวอย่าง ซึ่งสามารถเปลี่ยนเป็นค่าความถี่มูลฐานได้จากสมการที่ 2.4 โดย f_s แทนถึงอัตราการชักตัวอย่าง (Sampling rate)

$$f_0 = \frac{f_s}{t} \tag{2.4}$$

2.1.4 เสียงชนิดต่าง ๆ

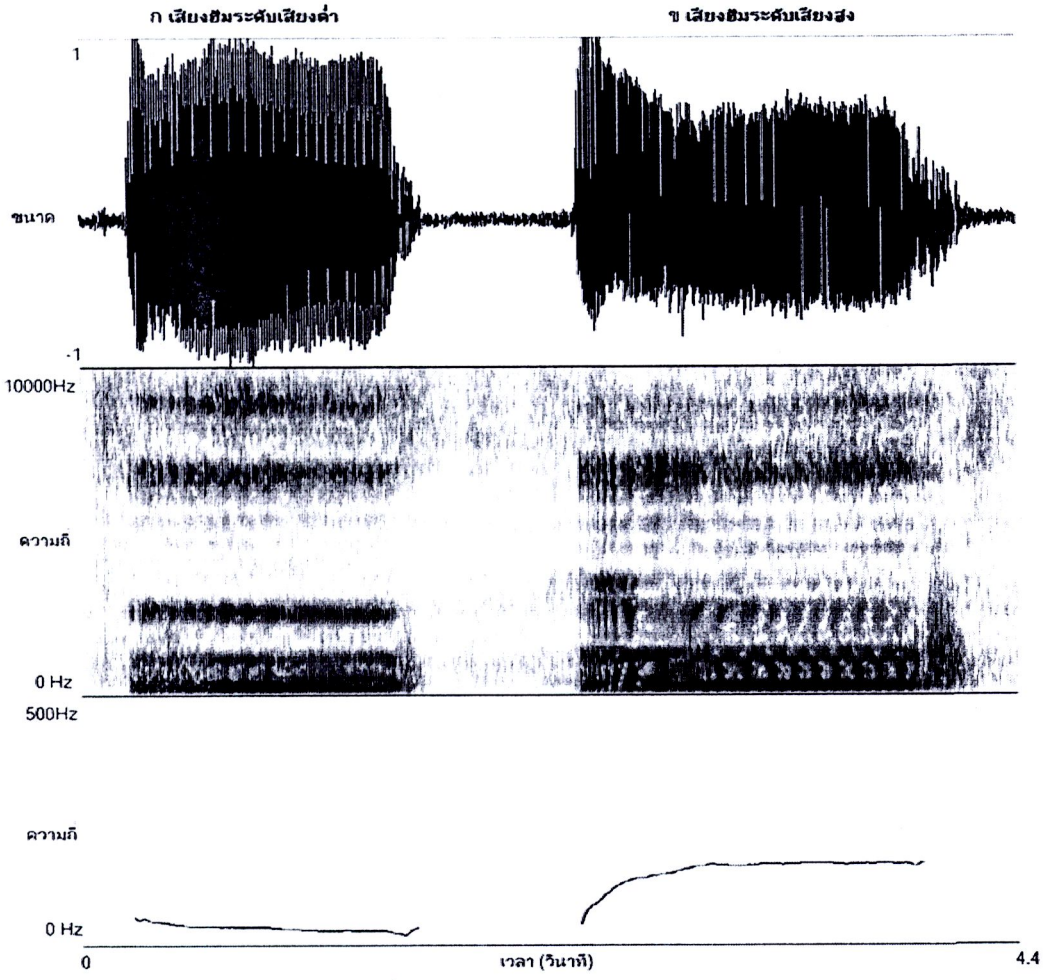
ในส่วนนี้จะบรรยายวิธีการออกเสียง และลักษณะของสัญญาณเสียงที่เกิดขึ้นของเสียงชนิดต่าง ๆ เฉพาะเสียงที่เกี่ยวข้องกับงานวิจัยนี้ ซึ่งประกอบด้วยเสียงฮัม และเสียงเสียดแทรก

1) เสียงฮัม

เสียงฮัมมักถูกใช้เพื่อความบันเทิง เช่นการฮัมตามจังหวะเสียงดนตรีต่าง ๆ โดยเกิดจากการเปล่งเสียงที่มีลักษณะเป็นกึ่งรายคาบ โดยที่ไม่มีการเปิดปาก โดยลมจะถูกปล่อยออกมาจากจมูก ส่งผลให้ความแตกต่างของเสียงฮัมขึ้นอยู่กับการสั่นของเส้นเสียงเท่านั้น โดยสามารถละทิ้งการวางตัวของอวัยวะภายในช่องปาก เพราะไม่มีอากาศไหลผ่านไปยังอวัยวะภายในช่องปาก

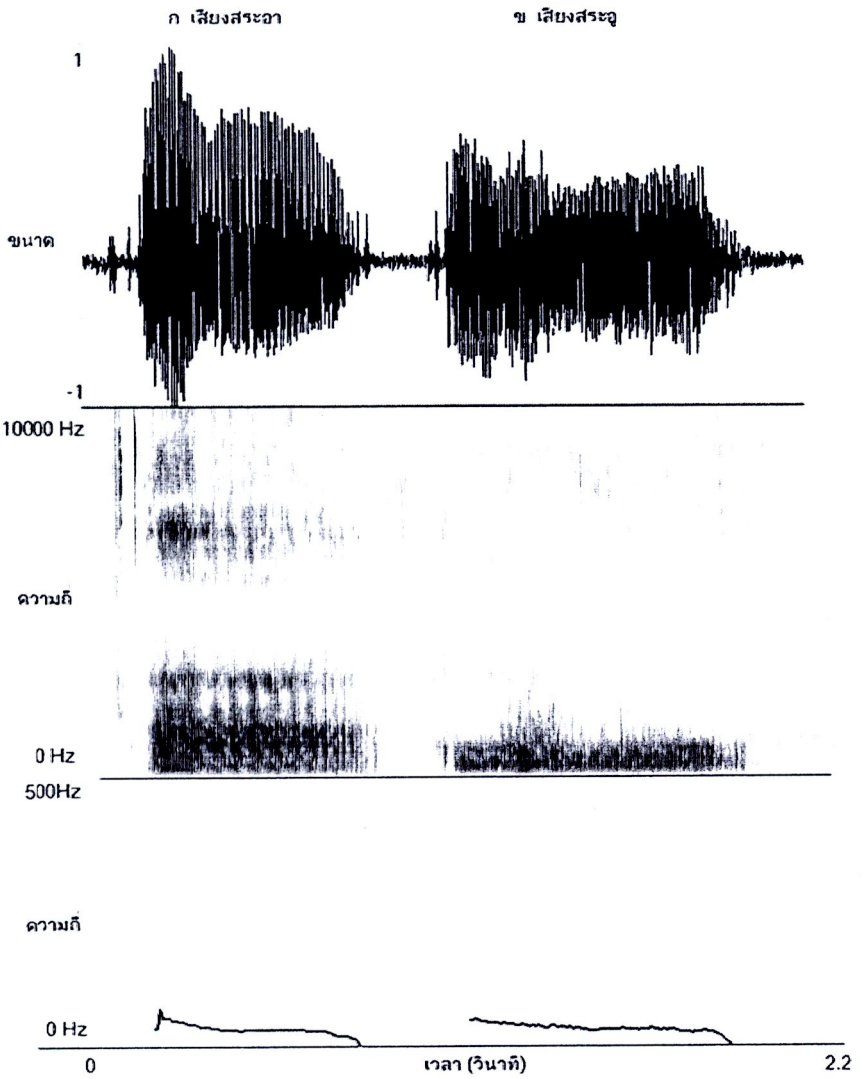
ความถี่ในการสั่นของเส้นเสียงทำให้เสียงมีความความทุ้ม และแหลมที่ไม่เท่ากัน ซึ่งความถี่ในการสั่นของเส้นเสียงสามารถวัดได้จากคุณลักษณะความถี่มูลฐาน ดังนั้นจึงใช้ค่าของความถี่มูลฐานที่แตกต่างกันในการแบ่งระดับของเสียงฮัม โดยสามารถหาค่าดังกล่าวได้โดยใช้หลักการของการคำนวณหาค่าอัตสหสัมพันธ์ตามหัวข้อที่ 2.1.3

ความแตกต่างของเสียงฮัมขึ้นอยู่กับความถี่มูลฐาน ดังแสดงในรูปที่ 2.6ก ที่แสดงถึงเสียงฮัมที่มีระดับเสียงสูง และรูป 2.6ข ที่แสดงถึงเสียงฮัมที่มีระดับเสียงต่ำ โดยสังเกตได้ว่าความแตกต่างของเสียงฮัมทั้งสองระดับเสียงไม่แสดงอย่างเด่นชัดในผลลัพธ์จากการแปลงฟูริเยร์แบบวิยุต เพราะช่วงความถี่ที่มีระดับพลังงานสูงอยู่ในช่วงเดียวกัน



รูปที่ 2.6 ความแตกต่างของเสียงฮัม

เมื่อเปรียบเทียบเสียงฮัมกับเสียงสระ พบว่าเสียงสระสามารถหาค่าของความถี่มูลฐานได้เช่นกัน แต่ค่าของความถี่มูลฐานของเสียงสระ ไม่สามารถใช้ในการแบ่งแยกเสียงได้ โดยในกรณีของเสียงสระจำเป็นต้องความถี่สั่นพ้องลำดับที่ 1 และ 2 ซึ่งสังเกตได้จากผลลัพธ์ของการแปลงฟูริ-เยร์แบบวิยุต ดังแสดงในรูปที่ 2.7ก คือตัวอย่างเสียงสระอู และในรูปที่ 2.7ข คือตัวอย่างเสียงสระอา ที่ช่วงของความถี่ที่มีพลังงานสูงอยู่ต่างช่วงกัน

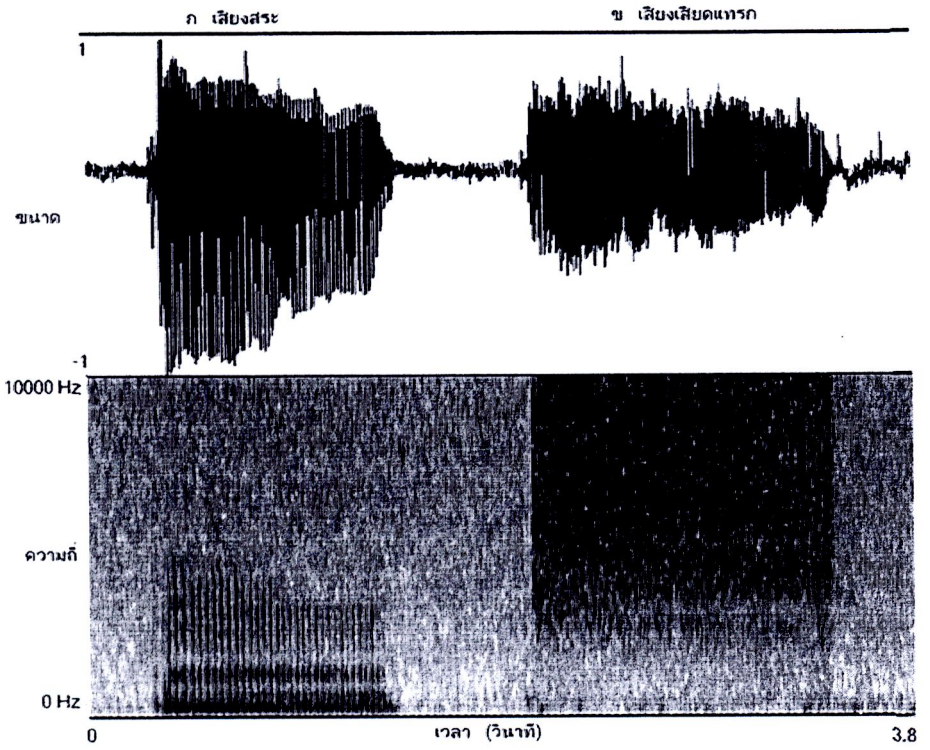


รูปที่ 2.7 ความแตกต่างของเสียงสระ

2) เสียงเสียดแทรก

เสียงเสียดแทรกคือเสียงที่มีเกิดจากการสร้างช่องแคบบริเวณเพดานปาก แล้วให้อากาศไหลผ่านช่องแคบดังกล่าว อาทิเช่นการเปล่งเสียงตัวสะกด ของคำว่า ทราย ซึ่งเปรียบเสมือนกับการพ่นลมผ่านเข้าไปยังท่อ ซึ่งทำให้เกิดการสั่นพ้องที่ความถี่ใดความถี่หนึ่ง

เสียงเสียดแทรกมีลักษณะของสัญญาณเสียงมีการกระจายตัวอยู่ในช่วงความถี่ที่สูงกว่าเสียงพูด ดังแสดงในรูปที่ 2.8 ที่แสดงให้เห็นว่าเสียงเสียดแทรก ดังรูปที่ 2.8ข มีกลุ่มของพลังงานในช่วงความถี่สูง ต่างจากเสียงพูดทั่วไป ดังรูป 2.8ก ซึ่งสามารถใช้หลักการดังกล่าวในการตรวจจับว่าเสียงที่เข้ามาในระบบเป็นเสียงเสียดแทรกหรือไม่

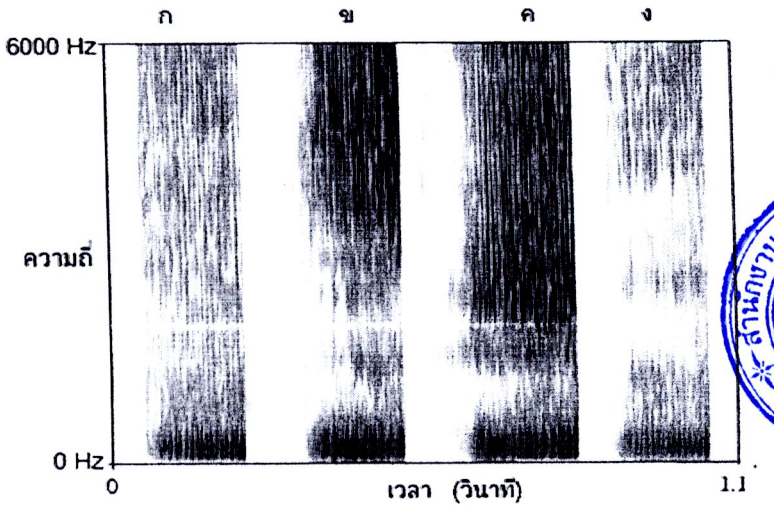


รูปที่ 2.8 แสดงความแตกต่างของเสียงสระ (ก) และเสียงเสียดแทรก (ข)

เสียงเสียดแทรกสามารถจำแนกได้ 3 ประเภท ตามฐานกรณ์ของการเกิดเสียง ได้แก่

- เสียดเสียดแทรกมีฐานกรณ์บริเวณริมฝีปาก ตัวอย่างเช่นเสียงตัวสะกดของคำว่า ฟัน
- เสียงเสียดแทรกที่มีฐานกรณ์บริเวณปุ่มเหงือก ตัวอย่างเช่นเสียงตัวสะกดของคำว่า สาน
- เสียงเสียดแทรกที่มีฐานกรณ์บริเวณเพดานปาก ตัวอย่างเช่นเสียงตัวสะกดของคำว่า Shy (ออกเสียงตามหน่วยเสียงของภาษาอังกฤษ)
- เสียงเสียดแทรกที่มีฐานกรณ์บริเวณฟัน ตัวอย่างเช่นเสียงตัวสะกดของคำว่า Thief (ออกเสียงตามหน่วยเสียงของภาษาอังกฤษ)

ข้อแตกต่างระหว่างเสียงเสียดแทรกที่มีฐานกรณ์แตกต่างกันคือ ช่วงความถี่ของบริเวณที่มีพลังงานสูงจะแตกต่างกัน ดังแสดงในรูปที่ 2.9



สำนักงานคณะกรรมการวิจัยแห่งชาติ ห้องสมุดงานวิจัย

รูปที่ 2.9 ความแตกต่างของเสียงเสียดแทรก ที่มีฐานกรณ์บริเวณริมฝีปาก

2.2. ทบทวนวรรณกรรม

ในส่วนของการทบทวนวรรณกรรมได้แบ่งรูปแบบการควบคุมคอมพิวเตอร์โดยใช้เสียงเป็นส่วนรับเข้าออกเป็น 2 แบบคือ แบบที่นำเสียงที่รับเข้ามาเพื่อตรวจจับหาคำ หรือกลุ่มของคำที่ต่อเนื่องกันเพื่อใช้ในการควบคุมต่าง ๆ และแบบที่นำเสียงที่รับเข้ามาเพื่อตรวจจับหาคุณลักษณะต่าง ๆ เพื่อใช้ในการควบคุมคอมพิวเตอร์ ซึ่งส่งผลให้เสียงที่รับเข้าในแบบที่ 2 ไม่จำเป็นต้องเป็นเสียงที่มีความหมาย

การพัฒนาส่วนรับเข้าที่ใช้เสียง เริ่มจากการใช้เทคโนโลยีของการรู้จำเสียงเพื่อใช้ในตรวจจับหาความหมายจากเสียงที่รับเข้ามา และความหมายที่ได้จากเสียงเหล่านั้นเชื่อมโยงเข้ากับคำสั่งต่าง ๆ ที่กำหนดไว้ ลักษณะของคำสั่งสามารถแยกได้เป็น 2 ประเภทคือ ประเภทที่คำสั่งแทนถึงคำสั่งย่อยหลายคำสั่งเพื่อทำงานที่เฉพาะเจาะจงบางอย่าง เช่นคำสั่งในการเปิดโปรแกรมใดโปรแกรมหนึ่ง และประเภทที่คำสั่งแทนถึงการทำงานของอุปกรณ์รับเข้ามาตรฐาน เช่นเมาส์ และแป้นพิมพ์ เช่นการสั่งการให้ตัวชี้ตำแหน่งเคลื่อนที่ไปยังทิศทาง หรือตำแหน่งที่ต้องการ โดยใช้คำสั่ง “Move Left” หรือ “Go to 100 100” ดังเช่นในงานวิจัยของ [9]

ประสิทธิภาพของการใช้เทคโนโลยีการรู้จำเสียงเพื่อควบคุมคอมพิวเตอร์ขึ้นอยู่กับความถูกต้องในการรู้จำเสียงดังแสดงในงานวิจัยของ [10] ปัจจัยหนึ่งที่ส่งผลต่อความถูกต้องของระบบรู้จำเสียงคือจำนวนคำศัพท์ และความซับซ้อนของไวยากรณ์ ซึ่งหมายถึงจำนวนของคำศัพท์ที่นำมาต่อกัน ซึ่งสามารถสรุปได้ว่ายิ่งมีจำนวนคำสั่งในระบบมากขึ้นเท่าใด ก็จะส่งผลให้ความถูกต้องของการรู้จำเสียงมีค่าน้อยลง ซึ่งเรื่องดังกล่าวส่งผลกระทบต่อจำนวนคำสั่งที่สามารถใช้ในการควบคุม และทำให้เห็นว่าการใช้คำสั่งเพื่อแทนงานที่เฉพาะเจาะจงนั้น ไม่เหมาะสมในการนำมาใช้ควบคุม

สำนักงานคณะกรรมการวิจัยแห่งชาติ
ห้องสมุดงานวิจัย
วันที่ 23 ก.ค. 2555
เลขทะเบียน 247146
เลขเรียกหนังสือ

คอมพิวเตอร์ เพราะต้องใช้จำนวนคำศัพท์จำนวนมากเพื่อให้ครอบคลุมการทำงานที่เป็นไปได้ทั้งหมด

ในงานวิจัย [12] แสดงให้เห็นว่าการหน่วงเวลาของการใช้เทคโนโลยีการรู้จำเสียง ส่งผลโดยตรงในการหยุดตัวชี้ตำแหน่งให้ตรงกับเป้าหมายที่ต้องการ อันเนื่องมาจากการหน่วงเวลาที่เกิดขึ้นจากการสั่งคำสั่งหยุด ทำให้ไม่สามารถใช้ตัวชี้ตำแหน่งได้อย่างมีประสิทธิภาพ

ในงานวิจัย [12] ได้เสนอแนวทางเพื่อแก้ปัญหาตัวชี้ตำแหน่งหยุดไม่ตรงกับเป้าหมาย โดยการสร้างตัวชี้ตำแหน่งเสมือน หรือเรียกว่าตัวชี้ตำแหน่งพรีดิกทีฟ เคลื่อนที่ตามตัวชี้ตำแหน่งจริง โดยระยะห่างระหว่างตัวชี้ตำแหน่งเสมือน และตัวชี้ตำแหน่งจริงคำนวณจากค่าเฉลี่ยของระยะทางที่ตัวชี้ตำแหน่งเคลื่อนที่เลยเป้าหมาย โดยวิธีการของงานวิจัยนี้คือให้ผู้ใช้ทำการเปล่งเสียงหยุดเมื่อตัวชี้ตำแหน่งจริงเคลื่อนที่ถึงเป้าหมาย ซึ่งส่งผลให้ตัวชี้ตำแหน่งจริงเคลื่อนที่เลยเป้าหมายไป แต่ตัวชี้ตำแหน่งเสมือนที่เคลื่อนที่ตามมาด้านหลังจะหยุดตรงกับเป้าหมาย และสั่งให้ตัวชี้ตำแหน่งจริงเคลื่อนที่กลับมาที่ตำแหน่งเดียวกับตัวชี้ตำแหน่งเสมือน แต่อย่างไรก็ตามวิธีดังกล่าวไม่สามารถแก้ไขปัญหาตัวชี้ตำแหน่งหยุดเลยจากเป้าหมายได้อย่างมีประสิทธิภาพ เพราะในการสั่งการแต่ละครั้ง จะมีค่าการหน่วงเวลาที่ไม่เท่ากัน ซึ่งขึ้นกับช่วงเวลาที่ใช้เปล่งคำสั่งหยุด ส่งผลให้ตัวชี้ตำแหน่งทำนายหยุดไม่ตรงกับเป้าหมายเช่นเดียวกัน

ในงานวิจัย [10] ได้พัฒนารูปแบบของการเข้าถึงเป้าหมายที่เรียกว่าตัวชี้ตำแหน่งแบบตารางที่มีจุดประสงค์ในการลดจำนวนคำสั่งที่ต้องใช้ในการสั่งการ โดยการแบ่งหน้าจอออกเป็น 9 ช่อง (3 ช่องในแนวตั้ง 3 ช่องในแนวนอน) และกำหนดตัวเลข 1 - 9 แทนช่องต่าง ๆ ผู้ใช้เข้าถึงช่องต่าง ๆ โดยการเปล่งเสียงตัวเลขที่แทนช่องเหล่านั้น และทำการแบ่งช่องที่ถูกเลือก ออกเป็น 9 ช่องอีกครั้ง และวนรอบจนกระทั่งถึงเป้าหมายที่ต้องการ ซึ่งแสดงให้เห็นว่าผู้ใช้สามารถเข้าถึงเป้าหมายต่าง ๆ โดยใช้คำสั่งเพียง 9 คำสั่ง และการควบคุมตัวชี้ตำแหน่งแบบตารางไม่มีข้อจำกัดทางด้านเวลา ส่งผลให้ละทิ้งปัญหาทางด้านการหน่วงเวลาไปได้

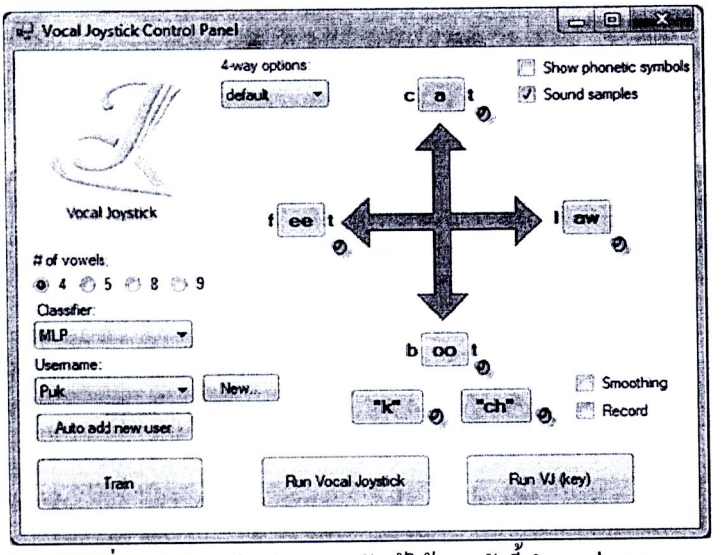
แต่อย่างไรก็ตามการใช้เทคโนโลยีการรู้จำเสียง มีข้อเสียที่มีการยึดติดกับภาษา เพราะต้องเปล่งเสียงเป็นคำที่มีความหมาย ที่แตกต่างกันไปตามภาษาที่ใช้ จึงทำให้มีการใช้แนวคิดการใช้เสียงที่ไม่ต้องการความหมายในการสั่งการ ดังแสดงในงานวิจัยของ [11] โดยใช้ค่าคุณลักษณะเฉพาะของเสียง เช่นค่าความถี่มูลฐานของเสียงในการควบคุม แทนที่จะใช้ความหมายของเสียงที่รับเข้าไป ส่งผลให้ระบบไม่ยึดติดกับภาษา เพราะไม่จำเป็นต้องใช้ความหมายจากเสียงที่รับเข้าไป และทำให้ระบบตอบสนองต่อผู้ใช้ได้รวดเร็วกว่า เพราะการใช้เสียงที่ไม่ต้องการความหมายมีการประมวลผลและคืนผลลัพธ์ไปยังระบบทุก ๆ กรอบของเสียงทันทีที่ผู้ใช้ทำการเปล่งเสียง ต่างจากการใช้เทคโนโลยีการรู้จำเสียง ที่จะมีการคืนผลลัพธ์ไปยังระบบหลังจากที่ผู้ใช้เปล่งเสียงเสร็จ ซึ่งส่งผลให้การใช้เสียงที่ไม่ต้องการความหมาย สามารถตอบสนองได้รวดเร็วกว่าการใช้เสียงพูด และการประมวลผลในแต่ละกรอบของเสียงใช้เวลาเพียง 2-3 มิลลิวินาที

ด้วยความเร็วในการประมวลผลที่รวดเร็ว และการตอบสนองต่อผู้ใช้ จึงทำให้การใช้เสียงที่ไม่ต้องการความหมายเหมาะสมในการนำไปใช้ควบคุมคำสั่งที่ต้องการการตอบสนองอย่างรวดเร็วและต่อเนื่อง เช่นใช้ในการควบคุมวัตถุ หรือสิ่งระบุตำแหน่งที่มีการเคลื่อนที่ ดังแสดงในงานวิจัย [19] ที่ใช้เสียงเป็นส่วนรับเข้าเพื่อใช้การเล่นเกมต่อกล่อง หรือที่มีชื่อเฉพาะว่าเททิส โดยเกมต่อกล่องเป็นเกมที่มีข้อจำกัดทางด้านเวลาในการเล่น โดยเสียงที่ใช้ในการเปรียบเทียบประกอบด้วยเสียงที่เป็นเสียงพูดที่เป็นคำสั่ง และเสียงที่ไม่ต้องการความหมายโดยเลือกใช้เสียงฮัม โดยคำสั่งต่าง ๆ ใช้ในการหมุนกล่องที่กำลังเลื่อนลงมา และสั่งให้กล่องเคลื่อนที่ลงมา โดยผลการทดลองแสดงให้เห็นว่าการใช้เสียงที่ไม่ต้องการความหมาย สามารถเล่นเกมเททิสได้อย่างมีประสิทธิภาพมากกว่า

งานวิจัย [20] ได้เสนอรูปแบบการผสมผสานการใช้เสียงที่มีความหมาย และเสียงที่ไม่ต้องการความหมาย ควบคุมโดยการสร้างตัวชี้ตำแหน่งเสมือนพร้อมตัวเลขกำกับในทิศทางที่ผู้ใช้ต้องการเคลื่อนที่ไป เพื่อให้ผู้ใช้เปล่งเสียงตัวเลขของตัวชี้ตำแหน่งเสมือนที่อยู่ใกล้เป้าหมาย ในกรณีที่ตัวชี้ตำแหน่งเสมือน ไม่ตรงกับเป้าหมายที่ต้องการ ผู้ใช้สามารถสั่งโดยการใช้เสียงที่ไม่ต้องการความหมาย เพื่อให้ตัวชี้ตำแหน่งเคลื่อนที่อย่างต่อเนื่องไปยังเป้าหมายที่ต้องการ แต่อย่างใดก็ตาม การออกแบบตามงานวิจัย [20] ไม่ได้เป็นที่นิยม

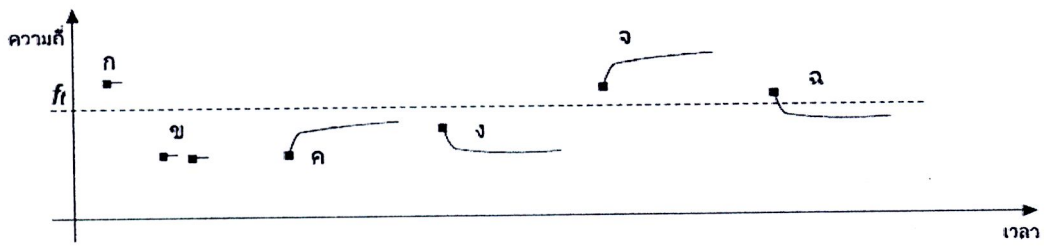
ในงานวิจัยที่พัฒนาตัวชี้ตำแหน่งสั่งการด้วยเสียงของ [13] เรียกว่าโวคอลจอยสติก เรียกโดยย่อว่า VJ และ [14] เรียกว่าตัวชี้ตำแหน่งยูสามไอ เรียกโดยย่อว่า U3I ได้นำแนวคิดของการใช้เสียงที่ไม่ต้องการความหมายมาใช้ในการควบคุมตัวชี้ตำแหน่งแบบ โดยให้ตัวชี้ตำแหน่งเคลื่อนที่ไปตามทิศทางที่กำหนดไว้ 4 ทิศทางตามทิศหลัก โดยในงาน [13] เลือกใช้เสียงสระ และในงาน [14] เลือกใช้การเปลี่ยนแปลงของเสียงฮัม และทั้งสองงานวิจัยกำหนดให้ตัวชี้ตำแหน่งเคลื่อนที่เมื่อมีการเปล่งเสียง และหยุดการเคลื่อนที่เมื่อไม่มีการเปล่งเสียง โดยความเร็วในการเคลื่อนที่ของตัวชี้ตำแหน่งในงานวิจัย [13] ขึ้นอยู่กับระดับความดังของเสียง ต่างจากในงานวิจัย [14] ที่ใช้ระดับความแตกต่างของระดับเสียงเมื่อเริ่มเปล่งเสียงกับระดับเสียงในขณะนั้น

เสียงที่ใช้ในการควบคุมตัวชี้ตำแหน่ง VJ ได้แก่เสียงสระที่อยู่ในคำว่า boot cat feet และ law เพื่อใช้ในการควบคุมทิศทางทั้งสี่ และใช้เสียงของหน่วยเสียง k เพื่อใช้แทนการคลิก และใช้เสียงของหน่วยเสียง ch เพื่อใช้แทนการลาก ดังแสดงในรูปที่ 2.10



รูปที่ 2.10 ส่วนต่อประสานกับผู้ใช้ของตัวชี้ตำแหน่ง VJ

ในการควบคุมตัวชี้ตำแหน่ง U3I ผู้ใช้ต้องทำการฝึกฝนการเปล่งเสียงฮัมที่มีเสียงในระดับกลาง เพื่อใช้เป็นระดับเสียงขีดแบ่ง (f_t) ในการแบ่งระดับเสียง โดยการเปลี่ยนแปลงของระดับเสียงฮัมที่นำมาใช้ได้แก่ การฮัมให้มีระดับเสียงสูงกว่าระดับเสียงที่กำหนดไว้ แล้วฮัมให้มีระดับเสียงสูงขึ้น หรือต่ำลง และการฮัมให้มีระดับเสียงต่ำกว่าระดับเสียงที่กำหนดไว้ แล้วเปลี่ยนแปลงระดับเสียงฮัมให้สูงขึ้นหรือต่ำลง เพื่อใช้แทนการเคลื่อนที่ทั้ง 4 ทิศทาง และใช้เสียงฮัมที่มีระยะเวลาสั้นเพื่อแทนการคลิก ดังแสดงในรูปที่ 2.11 ที่แกนนอนแสดงถึงเวลา แกนตั้งแสดงถึงความถี่ และเส้นที่ปรากฏอยู่ในกราฟแสดงถึงความถี่มูลฐานของเสียงที่รับเข้ามา โดยรูปที่ 2.11ก แทนถึงการคลิก รูปที่ 2.11ข แทนถึงการคลิกสองครั้ง รูปที่ 2.11ค แทนถึงการเคลื่อนที่ไปทางขวา รูปที่ 2.11ง แทนถึงการเคลื่อนที่ไปทางซ้าย 2.15จ แทนถึงการเคลื่อนที่ไปทางด้านบน และรูปที่ 2.11ฉ แทนถึงการเคลื่อนที่ไปทางด้านล่าง



รูปที่ 2.11 การฮัมในลักษณะต่างๆ เพื่อใช้ควบคุมตัวชี้ตำแหน่งยูไอ

ในงานวิจัยที่ [15] ได้เปรียบเทียบตัวชี้ตำแหน่งวีเจ และตัวชี้ตำแหน่งยูไอ โดยพบว่าผู้ใช้สามารถเรียนรู้การเปล่งเสียงสระได้เร็วกว่าการเรียนรู้การเปล่งเสียงฮัม และผู้ใช้บางคนไม่สามารถ

จดจำระดับของเสียงฮัมที่ถูกต้องได้ การกำหนดให้ความเร็วของตัวชี้ตำแหน่งเปลี่ยนตามระดับของเสียงฮัมในตัวชี้ตำแหน่ง U3I ส่งผลให้ผู้ใช้ไม่สามารถควบคุมความเร็วที่ต้องการได้ และเคลื่อนที่ออกนอกเป้าหมายมากกว่าตัวชี้ตำแหน่ง VJ แต่การใช้ระดับเสียงของเสียงฮัมในการควบคุมตัวชี้ตำแหน่งส่งผลให้ผู้ใช้รู้สึกเหนื่อยน้อยกว่า

ในงานวิจัยของ [16] ได้นำแนวคิดของการใช้เสียงที่ไม่ต้องการความหมายมาใช้ในการควบคุมตัวชี้ตำแหน่งแบบตาราง แต่แบ่งหน้าจอออกเป็น 4 ส่วน และใช้เสียงฮัมเสียงสูงและเสียงต่ำพร้อมกับตัวแปรด้านเวลา เพื่อสร้างเป็น 4 คำสั่งในการเข้าถึงทั้ง 4 ส่วน โดยผลการทดลองระบุว่าการใช้เสียงที่ไม่ต้องการความหมายมีความถูกต้องในการสั่งการไม่แตกต่างจากการสั่งการด้วยเสียงที่ต้องการความหมาย แต่ได้เปรียบในด้านของความเร็วในการประมวลผล

จากงานวิจัย [15] และ [16] ได้แสดงให้เห็นว่าการเลือกใช้เสียงที่แตกต่างกันเพื่อใช้เป็นส่วนรับเข้า ส่งผลให้ผลลัพธ์ของโปรแกรมที่แตกต่างกัน ทั้งด้านประสิทธิภาพของโปรแกรม และความพอใจของผู้ใช้ โดยปัจจัยที่ส่งผลต่อเรื่องดังกล่าวได้แก่ความยากในการเปล่งเสียง ความถูกต้องในการรู้จำเสียง ความเร็วในการตอบสนองต่อผู้ใช้ และความพยายามในการเปล่งเสียงขณะใช้งาน

ในงานวิจัย [17] ได้เปรียบเทียบประสิทธิภาพของการใช้รูปแบบการเคลื่อนที่ของตัวชี้ตำแหน่งแบบตาราง เทียบกับตัวชี้ตำแหน่ง VJ พบว่ารูปแบบการเคลื่อนที่ของตัวชี้ตำแหน่งแบบตารางเหมาะสมกับการเข้าถึงเป้าหมายที่อยู่ไกล เพราะตัวชี้ตำแหน่งแบบตารางสามารถกระโดดไปยังเป้าหมายที่ต้องการได้ แต่ตัวชี้ตำแหน่ง VJ ที่มีรูปแบบการเคลื่อนที่แบบต่อเนื่อง เหมาะสมกับการเข้าถึงเป้าหมายที่อยู่ใกล้ เช่นการเข้าถึงในลักษณะของรายการเลือก ซึ่งแสดงให้เห็นว่าการออกแบบการเคลื่อนที่ของตัวชี้ตำแหน่งที่แตกต่างกันส่งผลต่อความเหมาะสมของลักษณะงาน

จากงานวิจัยที่เกี่ยวกับการควบคุมคอมพิวเตอร์ด้วยการใช้เสียงเป็นส่วนรับเข้า พบว่าการใช้เสียงในการควบคุมตัวชี้ตำแหน่งนั้นสามารถใช้ในการควบคุมคอมพิวเตอร์ได้ แต่เนื่องจากข้อจำกัดในรูปแบบการควบคุม จึงทำให้การใช้เสียงในการควบคุมไม่สามารถควบคุมตัวชี้ตำแหน่งได้อย่างมีประสิทธิภาพได้เทียบเท่ากับการใช้เมาส์ในการควบคุม และรูปแบบการเคลื่อนที่ของตัวชี้ตำแหน่งที่มีการนำเสนอมา พบว่ามีข้อดีข้อเสียที่แตกต่างกันไป จากเหตุผลดังกล่าว ทำให้งานวิจัยนี้ไม่มุ่งเน้นในการพัฒนาวิธีการควบคุม หรือรูปแบบของเสียงที่ใช้ในการสั่งการตัวชี้ตำแหน่งเพียงอย่างเดียว แต่เน้นในการสร้างระบบที่ทำหน้าที่เป็นส่วนรับเข้าสั่งการด้วยเสียง ซึ่งมีตัวชี้ตำแหน่งเป็นส่วนประกอบ พร้อมกับพัฒนารูปแบบการควบคุมในลักษณะอื่นเพิ่มเติมขึ้นมา ที่ทำให้สามารถใช้คอมพิวเตอร์ได้อย่างมีประสิทธิภาพมากขึ้น ภายใต้ชุดคำสั่งเสียง และรูปแบบการควบคุมที่เหมือนกัน และมีอยู่อย่างจำกัด