

บทที่ 2

หลักการและทฤษฎี

2.1 ความรู้ทั่วไปเกี่ยวกับเสียง

2.1.1 ความหมายของเสียง

เสียง (Sound) ทางกายภาพ หมายถึง ความสั่นสะเทือนของตัวกลาง หรืออากาศ ที่เกิดจากการเปลี่ยนแปลงความดันบรรยากาศ จากแหล่งกำเนิดที่ทำให้เกิดการสั่นนั้น

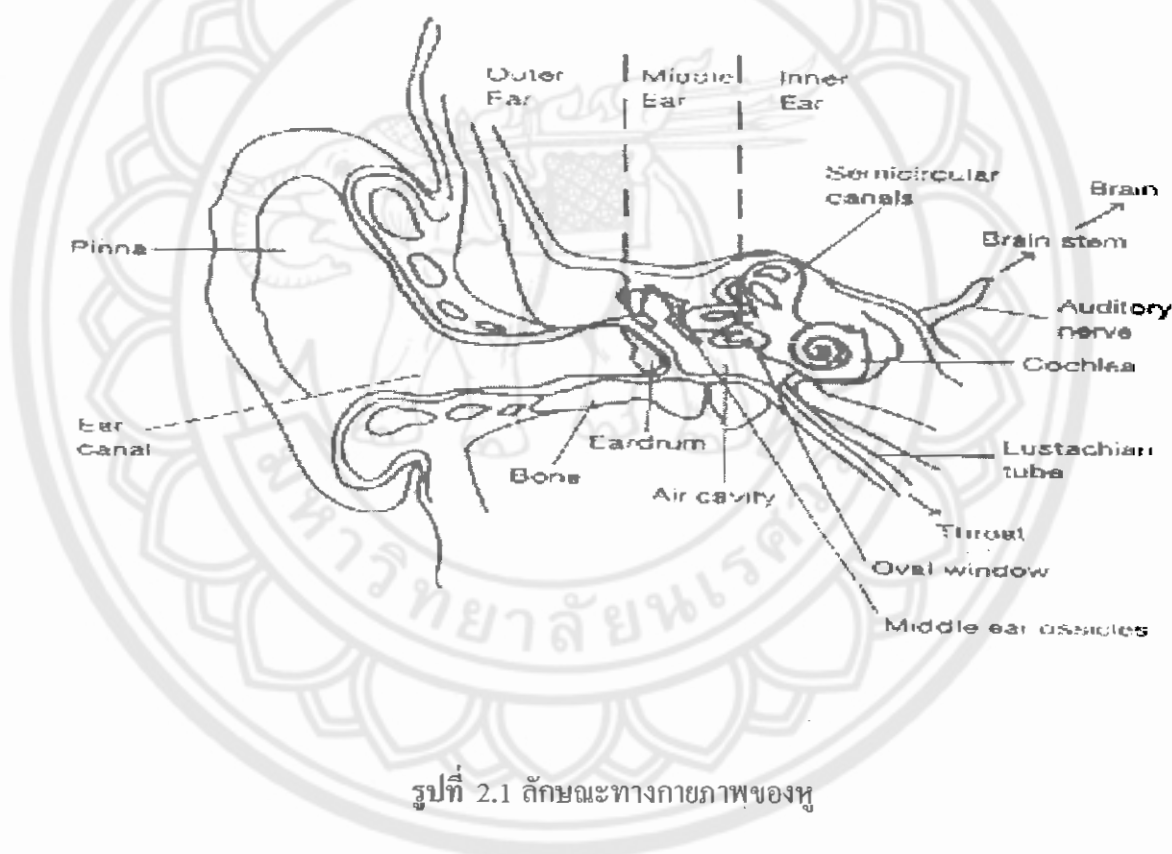
2.1.2 ลักษณะทางกายภาพของหู กับการได้ยิน

หู (Ears) คือ อวัยวะรับเสียงของมนุษย์ ซึ่งเป็นระบบเปิดที่สามารถรับรู้โดยการได้ยินเสียงจากการเปลี่ยนแปลงความดันบรรยากาศ มีช่วงความถี่ที่ได้ยินประมาณ 20-20,000 Hz และระดับความดังเสียง (Sound pressure level, SPL in dB) ประมาณ 0-130 dB หูประกอบด้วยส่วนหลัก 3 ส่วน ได้แก่ หูชั้นนอก หูชั้นกลาง และหูชั้นใน (รูปที่ 2.1)

หูชั้นนอก (Outer ear) ประกอบด้วย ใบหู (Pinna หรือ Auricle) เป็นส่วนที่สามารถมองเห็นได้จากภายนอกต่อกับอวัยวะลักษณะเป็นท่อ (Ear canal) มีความยาวประมาณ 27 มิลลิเมตรเส้นผ่านศูนย์กลางประมาณ 5-7 มิลลิเมตร ซึ่งมีการสั่นพ้องพื้นฐานโดยประมาณที่ 2,700 และ 3,500 Hz ปลายท่อด้านในต่อกับเยื่อแก้วหู (Ear drum or Tympanic membrane) ซึ่งเป็นแผ่นบางๆ ที่มีการเคลื่อนที่เมื่อมีเสียงมากระทบ

หูชั้นกลาง (Middle ear) เป็นส่วนที่ต่อจากเยื่อแก้วหูที่มีลักษณะเป็นโพรงอากาศที่มีปริมาตรประมาณ 2 ลูกบาศก์เซนติเมตร ประกอบด้วยกระดูกอ่อน 3 ชิ้น คือ กระดูกน้อน (Malleus) กระดูกทั่ง (Incus) และกระดูกโกลน (Stapes) เป็นระบบอวัยวะที่ทำหน้าที่ส่งผ่านความเปลี่ยนแปลงของความสั่นสะเทือนของอากาศที่ส่งผ่านมาจากการสั่นของเยื่อแก้วหู พร้อมทั้งทำหน้าที่ป้องกันไม่ให้หูได้รับอันตรายจากการได้รับเสียงดัง หรือการเปลี่ยนแปลงความดันบรรยากาศมากๆ ด้วย โดยจะมีกล้ามเนื้อ 2 ชิ้น ที่ทำหน้าที่พยุงระบบกระดูกอ่อนไว้ คือ Tensor tympani muscular และ Stapedius muscular ด้านล่างสุดของหูชั้นกลางจะติดต่อกับด้านหลังของจมูกด้วยท่อ Eustachian ที่ทำหน้าที่ปรับความดันบรรยากาศภายในหูและภายนอกให้เท่ากัน

หูชั้นใน (Inner ear) เป็นส่วนที่มีความสลับซับซ้อนอยู่ติดกับกระดูกโกลนของหูชั้นกลาง และอีกด้านหนึ่งติดกับ Oval window ซึ่งประกอบด้วย โพรงที่บรรจุด้วยของเหลวที่อยู่ในกระดูก Temporal ต่อกับปลายประสาทสมองส่วนรับรู้การได้ยินและการทรงตัวปลายประสาทการได้ยิน นี้อยู่ภายในกระดูกรูปก้นหอย (Cochlea) ที่มีความยาวเมื่อขี้ออกประมาณ 35 มิลลิเมตร มีเซลล์ขน (Hair cells) หรือ Organ of Corti ที่ประกอบด้วย Basilar membrane ที่มีความอ่อนไหวหรือตอบสนองต่อความถี่ของเสียงที่ไม่เท่ากัน ความถี่สูงจะเกิดการสั่นของต่ำจะเกิดการสั่นของ Basilar membrane บริเวณปลายอีกข้างหนึ่ง ซึ่งการรับรู้เสียงหรือการได้ยินเสียงของมนุษย์ จะสอดคล้องกับการสั่นของเซลล์ขน



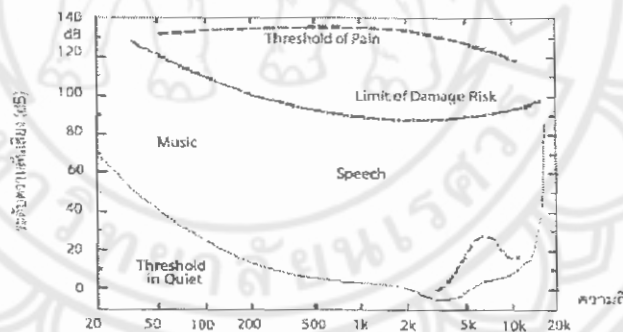
รูปที่ 2.1 ลักษณะทางกายภาพของหู

2.1.3 ความสามารถในการได้ยินของมนุษย์ (Auditory field and characteristic of human hearing)

มนุษย์ได้ยินเสียงโดยการนำคลื่นเสียงทางอากาศมาสู่กระดูก แล้วถูกเปลี่ยนเป็นพลังงานไฟฟ้าและถูกส่งต่อไปยังปลายประสาทสมองรับรู้การได้ยินจนเกิดการแปลความหมายทำให้มนุษย์เข้าใจความหมายของเสียงที่ได้ยิน การสั่นของเซลล์ขนเร็วหรือช้าสัมพันธ์กับระดับความดันเสียงที่เปลี่ยนแปลงในขณะที่เซลล์ขนแต่ละเซลล์ที่ถูกทำให้สั่นจะสัมพันธ์กับความถี่ของเสียง

จะเห็นว่า ความดังไม่ใช่องค์ประกอบอย่างเดียวในการอธิบาย หรือเข้าใจในเรื่องของเสียงและผลกระทบของเสียง ซึ่งเรียกกระบวนการนี้ว่าการได้ยินเสียงโดยการนำเสียงทางอากาศ (Airborne conduction) การได้ยินเสียงอีกอย่างหนึ่งของมนุษย์ คือ การนำเสียงทางกระดูก (Bone conduction) ซึ่งจะเกิดในกรณีที่เสียงมีความถี่ต่ำกว่า 1,000 Hz ลงไปและกรณีที่มีการสั่นสะเทือนของกระดูกในกะโหลกศีรษะ ที่ในสภาวะปกติเราอาจไม่ได้ยินหรือได้ยินเบามากแต่ถ้าเราอุดหู หรือใส่อุปกรณ์ป้องกันเสียงจะได้ยินเสียงนี้ชัดเจน

มนุษย์สามารถได้ยินเสียงขึ้นอยู่กับระดับความดัง (0-130 dB) ในช่วงความถี่ (20-20,000 Hz) ดังรูปที่ 2.2 ซึ่งสัมพันธ์กับการสั่นของเซลล์ขนภายในหูชั้นใน และมีความสัมพันธ์แบบไม่เป็นเส้นตรง (Non-linear relationship) กล่าวคือที่ระดับเสียงเท่ากัน มนุษย์จะได้ยินความดังที่ไม่เท่ากันทุกความถี่ มีนักวิจัยหลายท่านพยายามอธิบายลักษณะการได้ยินของมนุษย์ออกมาในรูปกราฟความสัมพันธ์เพื่อให้เข้าใจถึงลักษณะการได้ยินของหูที่ระดับเสียงและความถี่ต่างๆ กัน จะเห็นว่าที่ความถี่ 20 Hz ระดับเสียงต้องดัง 80 dB มนุษย์จึงจะเริ่มได้ยินเสียงนั้นและที่ความถี่ 1,000 Hz แค่เพียง 5 dB ก็สามารถได้ยินเสียงนั้นแล้ว ที่ความถี่ 4,000 Hz มนุษย์จะมีความไวต่อการได้ยินสูง



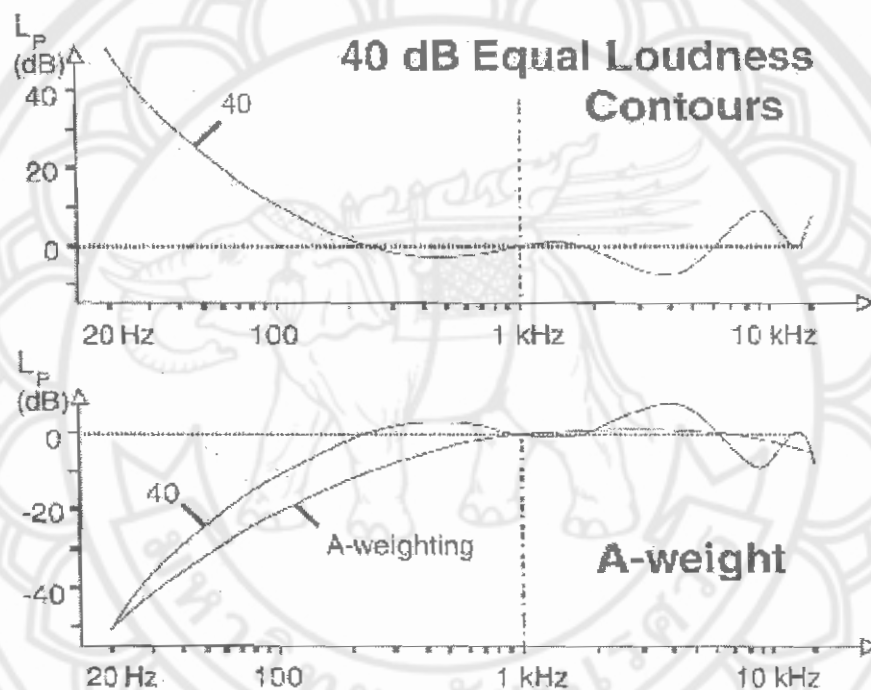
อ้างอิง : Equal loudness contours for pure tone, Kryter & Zwicker

รูปที่ 2.2 ขอบเขตการได้ยิน (Auditory field)

ขอบเขตการได้ยินของมนุษย์ (Auditory field) อยู่ในช่วงความถี่ 20-20,000 Hz และ 0-130 dB แต่ละคนได้ยินเสียงไม่เท่ากัน บางคนอาจได้ยินเสียงที่มีความถี่ต่ำมากๆ หรือ ระดับเสียงต่ำมากๆเมื่อเทียบกับคนอื่น ในขณะที่บางคนอาจไม่ได้ยินเสียงที่ต่ำกว่า 25 dB ที่ความถี่ 1,000 Hz เลย ตามปกติขีดเริ่มการได้ยิน (Threshold of hearing) ของมนุษย์ในระดับความดังที่ต่ำที่สุดคือ 20 dB เมื่อเปรียบเทียบกับขอบเขตของเสียงดนตรี (Music field) ซึ่งมีระดับเสียงประมาณ 20-90 dB ช่วงความถี่ประมาณ 50-10,000 Hz ขอบเขตของเสียงสนทนาของมนุษย์ (Human speech) จะมีระดับเสียงประมาณ 25-80 dB ในช่วงความถี่ประมาณ 100-4,000 Hz และขีดเริ่มของความเจ็บปวด

(Threshold of pain) ของมนุษย์อยู่ที่ประมาณ 130 dB ลักษณะการได้ยินของหูที่มีความสัมพันธ์แบบไม่เป็นเส้นตรงนี้ ทำให้เครื่องมือที่จะใช้ในการตรวจวัดระดับเสียง

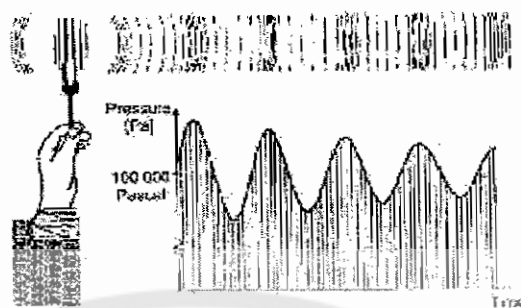
เพื่อศึกษาผลกระทบที่เกิดกับมนุษย์จำเป็นต้องมีวงจรวัดน้ำหนักความถี่ให้เครื่องมือวัดเสียงได้เหมือนลักษณะการได้ยินของมนุษย์มากที่สุด ที่นิยมใช้ในปัจจุบันได้แก่ วงจรวัดน้ำหนักแบบ A ซึ่งจะมีค่าใกล้เคียงกับเส้นผกผันของเส้นเท่าระดับเสียงต่ำสุด (40 dB Equal loudness contours) หรือขีดเริ่มการได้ยิน (รูปที่ 2.3)



รูปที่ 2.3 เส้นเท่าระดับเสียง และเส้นถ่วงน้ำหนักแบบ A

2.1.4 คลื่นเสียง (Sound wave)

คลื่นเสียงเป็นคลื่นตามยาว (Longitudinal wave) ดังรูปที่ 4 เมื่อเคาะส้อมเสียงจะมีการสั่นของโมเลกุลของส้อมเสียงทำให้เกิดการเปลี่ยนแปลงความดันบรรยากาศรอบๆส้อมเสียง เกิดคลื่นออกมาโดยรอบส้อมเสียง คล้ายกับคลื่นน้ำเมื่อเราโยนก้อนหินลงไปในน้ำ เรียกว่าการแผ่กระจายของคลื่นเสียง (Sound propagation)



รูปที่ 2.4 คลื่นเสียง (Sound wave)

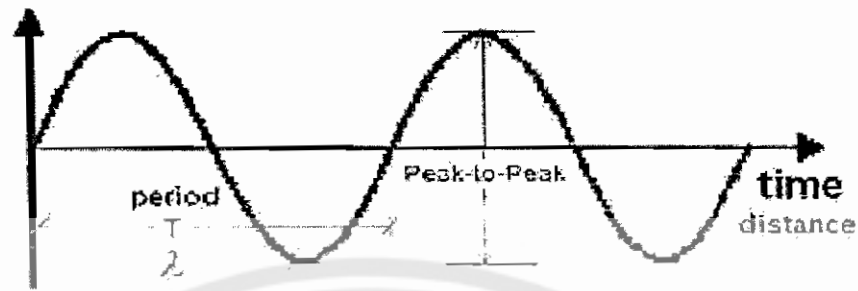
2.1.4.1 คุณสมบัติทางฟิสิกส์ของคลื่นเสียง (Physics properties of sound wave)

คลื่นเสียงที่เราได้ยินนั้นไม่ว่าจะเป็นเสียงจากแหล่งกำเนิดเสียงใด ส่วนใหญ่จะเป็นเสียงที่ซับซ้อนเกิดจากการรวมกันของคลื่นเสียงที่มีความถี่และระดับความดันเสียงต่างๆ เข้าด้วยกัน ในการศึกษาเรื่องฟิสิกส์ของคลื่นเสียง เรามักอธิบายเป็นสมการคณิตศาสตร์ให้เข้าใจได้ง่ายขึ้นในรูปแบบของ Sine wave เช่น การเคาะส้อมเสียง เกิดการเคลื่อนที่หรือความสั่นสะเทือนของโมเลกุลอากาศที่มาจากการเปลี่ยนแปลงความดันบรรยากาศโดยรอบ เป็นส่วนอัดส่วนขยาย เรียกว่า ความดันเสียงเคลื่อนที่ส่งผ่าน โมเลกุลของตัวกลางไปยังผู้รับเสียง (รูปที่ 2.4)

พารามิเตอร์ทางฟิสิกส์ที่สำคัญที่จะอธิบายธรรมชาติของการสั่นของโมเลกุลอากาศ จากการเปลี่ยนแปลงความดันอากาศหรือการเคลื่อนที่ของคลื่นเสียงในตัวกลาง ในที่นี้คือ อากาศผ่านกระบวนการได้ยินของมนุษย์จนรับรู้ความรู้สึกของเสียงนั้น มีอยู่ 2 พารามิเตอร์สำคัญคือ ความถี่ของเสียง และความยาวคลื่นเสียง ซึ่งมีความเกี่ยวข้องกับอัตราเร็วของเสียงในอากาศ ทิศทางการแผ่กระจายของเสียง และเวลาที่เสียงเดินทางจากแหล่งกำเนิดไปยังผู้รับเสียง

2.1.4.2 ความถี่ (Frequency, f in Hz)

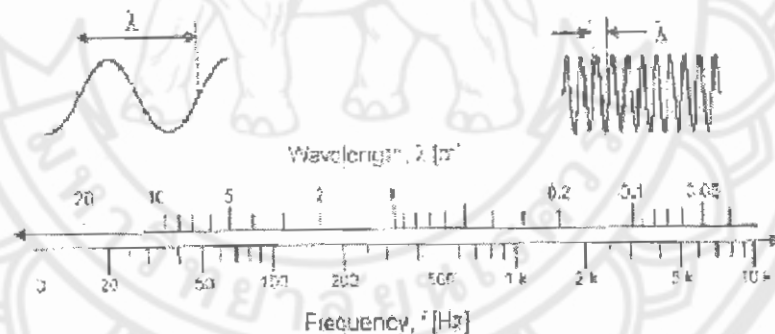
การเคลื่อนที่ของความดันอากาศในตัวอย่างการเคาะส้อมเสียงหรือการเคลื่อนที่ของลูกสูบ ในท่อเกิดปรากฏการณ์ส่วนอัดส่วนขยายเดินไปตามท่อ ซึ่งนำมาเขียนกราฟระหว่างแอมพลิจูดกับเวลา หรือแอมพลิจูดกับระยะทาง จะได้กราฟรูป Sine wave (รูปที่ 2.5) จำนวนรอบการเกิดส่วนอัดส่วนขยาย เราเรียกว่า ความถี่ของคลื่น มีหน่วยรอบต่อวินาที



รูปที่ 2.5 Sine wave

2.1.4.3 ความยาวคลื่น (Wavelength, λ in meters)

จากกราฟของการเคลื่อนที่ของคลื่นเสียงระหว่างแอมพลิจูดกับระยะทาง ความยาวคลื่นเสียง คือระยะทางจากยอดคลื่นหนึ่งหรือระยะทางของการเกิดคลื่น 1 คลื่น (รูปที่ 2.6) มีหน่วยเป็น เมตร และความยาวคลื่นนั้นมีความสำคัญอย่างมากในการศึกษาการควบคุมเสียง



รูปที่ 2.6 ความยาวคลื่นกับความถี่

2.1.5 ความถี่โชนาธรรมชาติทางธรรมชาติ (Natural resonant frequency)

จากที่ทราบกันแล้วว่าเสียงเกิดขึ้นจากการสั่นสะเทือนของวัตถุ วัตถุที่สั่นสะเทือนสามารถเกิดได้หลายแบบ เช่น การสั่นของสายกีตาร์ เสียงสัญญาณไฟไซเรน หรือเสียงจากเครื่องดนตรี

ตัวอย่างที่ดีของการสั่นแล้วเกิดเสียง ก็ได้แก่เสาอากาศวิทยุของรถยนต์ ถ้าเราออกแรงค้ำมัน เราจะสังเกตเห็นว่ามันจะเกิดการสั่นขึ้นทันทีที่ความถี่เฉพาะตัววัตถุหลายอย่าง เมื่อเราทุบ ตี ดึง หรือ กระแทกจะเกิดการสั่นขึ้น เมื่อเราโยนเหรียญลงบนพื้นที่มีความแข็ง มันก็จะเริ่มสั่นเช่นกัน เมื่อเราดึงสายกีตาร์ก็จะเกิดการสั่น เมื่อแต่ละวัตถุเกิดการสั่น ก็มีแนวโน้มที่จะมีความถี่เฉพาะตัวซึ่งก็คือความถี่ทางธรรมชาติ (Natural frequency) หากความถี่ที่เกิดขึ้นนั้นอยู่ในช่วงที่มนุษย์สามารถ

ได้ยิน และมีความดังมากพอ เราจะสามารถได้ยินเสียงที่เกิดขึ้นจากการสั่นนั้นได้ วัตถุทุกชนิดมีความถี่ทางธรรมชาติหรือมีกลุ่มความถี่ที่พวกมันมักจะสั่น วัตถุบางอย่างสั่นที่ความถี่ทางธรรมชาติเดียว (Single natural frequency) ในขณะที่วัตถุอื่นๆอาจมีหลาย โหมด (mode) ของการสั่น และสามารถสั่นได้ในทุกโหมดของการสั่น

ขลุ่ยหรือเสาอากาศวิทยุมีแนวโน้มที่จะสั่นที่ความถี่ทางธรรมชาติเดียวเท่านั้น ส่วนดินสอจะสั่นที่หลายๆ โหมดทำให้สามารถจำแนกเสียงของดินสอได้ วัตถุบางอย่างเช่น สายกีตาร์มีจำนวนโหมดของการสั่นซึ่งเกี่ยวข้องกับทางคณิตศาสตร์ สายกีตาร์ที่มีโหมดของการสั่นที่สัมพันธ์กันจะให้เสียงที่ไพเราะมากกว่าแผ่นโลหะซึ่งไม่มีโหมดการสั่นที่เกี่ยวข้องกัน ดังนั้นเสียงที่ได้จะไม่ไพเราะ

มีคำถามเกิดขึ้นว่า วัตถุจะสั่นที่ความถี่ใด คำตอบที่ได้คือ มันขึ้นอยู่กับองค์ประกอบทางกายภาพหลายๆ อย่าง เช่น มวลและความยืดหยุ่นของวัตถุ ยกตัวอย่างเช่น ความถี่ทางธรรมชาติของสายกีตาร์ง่ายต่อการเปลี่ยนความถี่ทางธรรมชาติของการสั่นสะเทือน (Natural vibration frequency) ของตัวมันเอง โดยการลดระยะความยาวของสายขณะที่กดคอร์ด หรือการเปลี่ยนความตึงของสายกีตาร์ ทำให้สามารถไล่ระดับและผสมผสานเสียงให้เกิดเป็นเสียงดนตรีได้

หลักการเดียวกันนี้นำมาประยุกต์ใช้กับเครื่องดนตรีที่ใช้แรงลม โดยที่ความถี่ทางธรรมชาติถูกควบคุม โดยการควบคุมลมที่จะเข้าไปในช่องรับลมยกตัวอย่างเช่น Trombone ความยาวของช่องรับลมถูกควบคุมโดยการเลื่อนเข้าเลื่อนออกซึ่งจะทำให้ความยาวของช่องรับลมเปลี่ยนไป มีผลทำให้รัโซแนนซ์ทางธรรมชาติ (Natural resonance) ของเครื่องดนตรีเปลี่ยนไปโดยสรุปแล้ว วัตถุจะมีความถี่ทางธรรมชาติที่ความถี่เดียว หรือมากกว่า 1 ความถี่ ซึ่งจะเกิดการสั่น เมื่อเราออกแรงตี หรือทำการรบกวนด้วยวิธีอื่นๆ อันที่จริงแล้วความถี่ทางธรรมชาติจะขึ้นกับวัตถุและองค์ประกอบทางกายภาพอื่นๆ

2.1.6 การอธิบายเสียงที่ได้ยินในเชิงสมการคณิตศาสตร์

เสียงเดินทางผ่านอากาศมาที่อวัยวะรับการได้ยินของมนุษย์ คือ หู ซึ่งเป็นระบบเปิด ที่มีความไวต่อความดันอากาศมาก มนุษย์จะรู้สึกได้ยินก็ต่อเมื่อ มีการนำสัญญาณนั้นผ่านระบบการได้ยินในหูชั้นต่างๆ แปลงเป็น สัญญาณประสาทโดยอวัยวะที่เรียกว่า เซลล์ขน สัญญาณประสาทส่งไปแปลผลที่สมอง มนุษย์จึงรับรู้ได้ว่าเสียงที่ได้ยินนั้นเป็นอย่างไร

เพื่อจะอธิบายเสียงที่ได้ยินในเชิงสมการคณิตศาสตร์สามารถทำได้ในสองลักษณะคือ ฟังก์ชันของเวลา (Time domain) กับฟังก์ชันของความถี่ (Frequency domain) ฟังก์ชันของเวลา สามารถอธิบายในลักษณะระดับความดังของเสียง ในช่วงเวลาที่เปลี่ยนแปลงไปเท่านั้น แต่ไม่มีคำอธิบาย

ใดๆ ในเชิงของความถี่นั้นว่ามีความถี่ต่ำหรือสูง ในขณะที่ฟังก์ชันของความถี่สามารถอธิบายลักษณะของเสียงว่ามีความถี่เป็นอย่างไรสูงหรือต่ำ ซึ่งเป็นข้อมูลที่สำคัญ ในการเข้าถึงธรรมชาติของเสียงนั้นๆ การอธิบายจะต้องทำการวิเคราะห์ความถี่ของเสียง ซึ่งมีอยู่ด้วยกันหลายวิธีที่นิยมใช้ได้แก่ การวิเคราะห์ความถี่แบบ ออกเทฟ (Octave analysis) และ Fast Fourier Transform (FFT) ซึ่งจะให้ข้อมูลองค์ประกอบความถี่เด่นของความถี่ในระดับเสียงที่เราสนใจ และสามารถหาแหล่งกำเนิดที่แท้จริงของเสียงได้ หากนำฟังก์ชันทั้งสองมาอธิบายลักษณะของเสียงในเวลาเดียวกัน จะทำให้ทราบข้อมูลทั้งระดับความถี่เสียงที่เปลี่ยนไปในช่วงเวลาและองค์ประกอบความถี่เด่นต่างๆ ของเสียงนั้นๆ พร้อมกับทำให้ สามารถเข้าใจถึงลักษณะธรรมชาติของเสียงนั้นได้อย่างแท้จริง และสามารถหาแหล่งต้นกำเนิดที่แท้จริงของเสียงนั้นจากองค์ประกอบความถี่เด่น ตัวอย่างเช่น เสียงจากมอเตอร์ ซึ่งเป็นแหล่งกำเนิดเสียงที่มีองค์ประกอบความถี่เด่นชัดเจนจนเป็นลักษณะเสียง Pure tone หรือการวัดระดับเสียงในโรงงานอุตสาหกรรม เมื่อทำการตรวจวัดในฟังก์ชันของเวลาและความถี่ จะพบข้อมูลที่แสดงให้เห็นถึงระดับเสียงที่ความถี่เด่น (Peak frequency) ของเสียงในโรงงานนั้น ทำให้ทราบสาเหตุที่แท้จริงของปัญหา (Source identification) เรานำสมการคณิตศาสตร์มาอธิบายการเปลี่ยนฟังก์ชันของเวลาให้เป็นฟังก์ชันของความถี่ทำได้โดยใช้สมการที่เรียกว่า “Fourier transform”

2.1.6.1 Fourier Transform

วิธีของ Fourier ในการวิเคราะห์สัญญาณมีอยู่ 4 วิธีด้วยกันคือ Fourier series (FS), Fourier transform (FT), Discrete Fourier series (DFS), Discrete Fourier transform (DFT) ซึ่งแต่ละวิธีนั้นใช้วิเคราะห์สัญญาณต่างชนิดกัน โดยที่ Fourier series จะใช้วิเคราะห์สัญญาณคาบ (Periodic signal) ที่เป็นสัญญาณต่อเนื่อง (Continuous-time signal) ส่วนการวิเคราะห์สัญญาณที่เป็นแบบไม่ต่อเนื่อง (Discrete periodic signal) จะใช้ DFS และสำหรับสัญญาณสัญญาณต่อเนื่องที่ไม่เป็นสัญญาณคาบ (Continuous nonperiodic signal) FT ในการวิเคราะห์ แต่สำหรับสัญญาณไม่ต่อเนื่องที่ไม่เป็นสัญญาณสัญญาณคาบ (Discrete-time nonperiodic signal) จะใช้ DFT ในการวิเคราะห์

2.1.6.2 Discrete Fourier Transform (DFT)

Discrete-time Fourier Transform (DTFT) จะสามารถ Transform สัญญาณจาก Time domain ไปเป็น Frequency domain ถ้าหากเป็นสัญญาณที่ไม่จำกัดและเป็นสัญญาณต่อเนื่องใน Frequency domain ซึ่งเป็นสัญญาณที่ไม่สามารถวิเคราะห์ได้ด้วยคอมพิวเตอร์ ด้วยเหตุนี้จึงนำเอาหลักการของ Discrete Fourier Transform (DFT) เข้ามาช่วยการทำ Discrete Fourier Transform

(DFT) เป็นการ Sampling สัญญาณใน Frequency domain เพื่อให้ได้สัญญาณไม่ต่อเนื่อง และสามารถใช้อุปกรณ์วิเคราะห์และคำนวณได้ ถ้าหากเรา sampling ด้วย Sampling frequency ที่สูงๆ จะทำให้ได้สัญญาณที่ดี เมื่อเราทำการ IDFT จะทำให้ได้สัญญาณคล้ายกับสัญญาณเดิมมาก ซึ่งจะขึ้นอยู่กับ Sampling frequency โดยแบ่งสมการดังนี้

$$X(k) = \sum x(n)WN \quad \dots(8.2-1)$$

และสมการการแปลงกลับเป็นดังนี้

$$x(n) = (1/N) \sum X(k) WN \quad \dots(8.2-2)$$

โดยที่

$$WN = e^{-j(2\pi / N)} \quad \dots(8.2-3)$$

2.1.6.3 Fast Fourier Transform (FFT)

วิธีการ FFT คือการทำ DFT ซ้ำกันหลายๆครั้งโดยใช้หลักการของ Recursion ซึ่ง $O(n)$ ของ DFT มีค่าเท่ากับ N^2 แต่ $O(n)$ ของการทำ FFT อาจมีค่าเป็น $(N/2)\log_2 N$ เมื่อ N เป็นจำนวนยกกำลังของ 2

2.2 ความหมายของจินตคณิตอัลกอริทึม (Genetic Algorithms: GAs)

จินตคณิตอัลกอริทึม (Genetic Algorithms :GAs) ถูกคิดค้นโดยจอห์น ฮอลแลนด์ (John Holland) ในปี 1975 โดยที่ GAs จะเป็นเทคนิคการหาคำที่เหมาะสมที่สุด ซึ่งมีรากฐานแนวคิดมาจากทฤษฎีวิวัฒนาการของชาร์ล ดาร์วิน (Charles Darwin) โดยอิงจากแนวความคิดการอยู่รอดของผู้ที่แข็งแกร่งที่สุด (Survival of the fittest) ซึ่งการทำงานของจินตคณิตอัลกอริทึมนี้จะไปในลักษณะของการหาคำตอบแบบคู่ขนาน (Parallel Search) โดยคำตอบที่ได้จากการหาคำตอบในหนึ่งรุ่น (Generation) จะผ่านการแปลง (Transformation) เพื่อที่จะนำไปสู่การค้นหาคำตอบที่ดีขึ้นในรุ่นถัดไป การเปลี่ยนแปลงที่เกิดขึ้นกับคำตอบ (Solution) หรือสมาชิกของประชากร (Individual) ภายในประชากร (Population) หนึ่งนั้นจะเป็นไปเพื่อการสำรวจพื้นที่ในการค้นหา (Search Space) และส่งเสริมให้มีการถ่ายทอดคุณลักษณะที่ดี (Fit Characteristics) ของคำตอบที่ดี พบในปัจจุบันไปยังรุ่นถัดไป สมาชิกของประชากรที่มี คุณลักษณะที่ดี (Fit Individual) หรือ

คำตอบที่มี คุณลักษณะดี จะมีอยู่หลายคำตอบด้วยกัน ซึ่งจะนำไปสู่คำตอบที่มีค่าเหมาะสมที่สุด (Optimum Solution) นั่นคือสมาชิกของประชากรที่มีลักษณะดีที่สุด (Fittest Individual)

ปัจจุบันได้มีการประยุกต์ใช้จินตคติอัลกอริทึม ในงานต่าง ๆ อย่างแพร่หลาย เช่น การใช้จินตคติอัลกอริทึมในการแก้ปัญหาทางคณิตศาสตร์ การแก้ปัญหาการหาเส้นทางที่มีระยะทางสั้นที่สุด (Traveling Salesman Problem: TSP) การจัดตารางสอน (Timetable Scheduling Problem) การควบคุมหุ่นยนต์ (Robot Control) รวมทั้งการค้นคว้าวิจัยด้านอื่น ๆ อีกด้วย

2.3 องค์ประกอบของจินตคติอัลกอริทึม

จินตคติอัลกอริทึมมีองค์ประกอบที่สำคัญ 5 องค์ประกอบ คือ

- 1) รูปแบบโครโมโซมที่ใช้ในการนำเสนอทางเลือกที่สามารถจะเป็นได้ของแต่ละปัญหา
- 2) วิธีสร้างประชากรต้นกำเนิด (initial population) ของทางเลือกที่สามารถจะเป็นไปได้
- 3) ฟังก์ชันสำหรับประเมินค่าความเหมาะสม (fitness) เพื่อให้คะแนนแต่ละทางเลือก
- 4) จินตคติโอเปอเรเตอร์ (Genetic Operator) ซึ่งใช้ในการปรับเปลี่ยนองค์ประกอบของข้อมูลตลอดกระบวนการ ได้แก่ การคัดเลือก การจับคู่ การมีเพศสัมพันธ์และการคอนเวอร์เจนซ์
- 5) ค่าพารามิเตอร์ต่างๆ ซึ่งต้องใช้สำหรับ จินตคติอัลกอริทึม เช่น ขนาดของประชากร, ความน่าจะเป็นของการใช้จินตคติโอเปอเรเตอร์ และจำนวนรุ่น เป็นต้น

2.4 การทำงานของจินตคติอัลกอริทึม

Genetic Algorithms (GAs) จะเป็นการใช้เทคนิคในการหาคำตอบโดยใช้หลักการทางพันธุกรรมและการคัดเลือกโดยธรรมชาติเป็นพื้นฐาน โดยที่ GAs จะทำให้ข้อมูลจำนวนมากที่ประกอบไปด้วยลักษณะเฉพาะส่วนตัวค่อยๆ ปรากฏขึ้นภายใต้กฎการเลือกซึ่งจะนำไปสู่คำตอบที่ถูกต้องและเหมาะสม ซึ่งในทางทฤษฎีขั้นตอนจะมีด้วยกันหลายแบบแต่ในทางปฏิบัติขั้นตอนของ จินตคติอัลกอริทึมจะคล้ายๆ กันซึ่งเมื่อเรานำมาประยุกต์ใช้กับการจำแนกเสียงจะมีขั้นตอนหลักๆ ดังนี้

2.4.1 การเลือกตัวแปรและการหาค่าของตัวแปร (Cost Function)

ให้คำจำกัดความการหาค่าของตัวแปรและตัวแปรที่เกี่ยวข้องทั้งหมด โดยทั่วไปการหาค่าของตัวแปรจะเป็นการทดลองหาค่าที่เราต้องการ จะได้มาจากการกำหนดค่าตัวแปรเริ่มต้น เช่น การ

วัดค่าที่ปล่อยออกมาที่มากที่สุดของรถยนต์ในการเดินทาง โดยการหาค่าของตัวแปรจะต้องเลือกตัวแปรให้เหมาะสมเพื่อให้ค่าการหาค่าของตัวแปรที่ออกมามีความเป็นไปตามที่ต้องการมากที่สุด

2.4.2 การแปลงรหัสและการถอดรหัสของตัวแปร (Variable Encoding and Decoding)

การที่จะทำงานในโปรแกรม MATLAB นั้นเราจะต้องทำการแปลงรหัสและถอดรหัสของตัวแปรก่อนเช่น ในการทดสอบนี้ คณิตศาสตร์ของเสียงจะถูกวัดออกมาแล้วนำเข้าเครื่องคอมพิวเตอร์ แต่ค่าที่ถูกวัดออกมานี้จะยังไม่สามารถนำไปคำนวณในโปรแกรมได้ เรายังคงต้องทำการถอดรหัสของตัวแปรก่อนเพื่อให้เครื่องสามารถที่จะรับค่าไปคำนวณต่อไปได้

2.4.3 การสร้างกลุ่มตัวอย่างเริ่มต้น (Generate initial population)

กลุ่มตัวอย่างเริ่มต้นจะเป็นจำนวนกลุ่มตัวอย่างเสียงที่ถูกสุ่มขึ้นมาเพื่อใช้ในการคัดเลือก โดยแต่ละเสียงละเสียงจะมีรหัส (รหัสของความถี่) ที่แตกต่างกัน

2.4.4 การหาค่าของแต่ละเสียง (Find cost for each voice)

ในแต่ละระดับเสียงจะมีค่าความถี่ที่แตกต่างกัน ซึ่งเมื่อวัดเสียงมาแล้วเราจะต้องนำเสียงแต่ละเสียงที่ได้นำไปหาความถี่ทั้งหมด

2.4.5 การจับคู่ (Selection mate) วิธีการจับคู่จะมีด้วยกัน 2 วิธี ดังนี้

2.4.5.1 การคัดสรรโดยธรรมชาติ (Natural Selection) จะเป็นการเลือกค่าที่เหมาะสมที่สุดเพื่อดำเนินการต่อไป ส่วนค่าที่ไม่เหมาะสม จะถูกตัดทิ้งโดยข้อดีของการคัดสรรโดยธรรมชาติ ก็คือ จะไม่ต้องทำการคัดเลือกกลุ่มตัวอย่าง (เป็นเพียงค่าการนำค่าที่มีอยู่มาเลือกค่าที่เหมาะสม)

2.4.5.2 การคัดสรรแบบทั่วไป

จะเป็นการเลือกคู่ของเสียงโดยการใช้วิธีการคัดเลือกเสียง (ความถี่) ที่จะมาเข้าคู่กัน มีด้วยกัน 4 วิธี คือ

- การจับคู่จากบนลงล่าง (Pairing form top to bottom) จะเป็นการจับคู่กันของเสียงแต่ละแถว เช่น การจับคู่ระหว่างแถว 1,3,5,...กับแถว 2,4,6,...เป็นต้น ซึ่งวิธีนี้จะ เป็นรูปแบบการคัดเลือกที่ไม่ดีนักแต่จะง่ายต่อการทำโปรแกรม
- การจับคู่แบบสุ่ม (Random pairing) จะใช้การสุ่มในการเลือกคู่ของเสียงแต่ละเสียงว่าเสียงคู่ใดจะมีความใกล้เคียงของเสียงมากที่สุด
- การจับคู่แบบสุ่มถ่วงน้ำหนักอันดับ (Rank weighted random pairing) จะใช้เรื่องความน่าจะเป็นในการออกแบบการจับคู่ของเสียง โดยจะเป็นการคัดเลือกที่อิสระคือเราสามารถเลือกค่าความน่าจะเป็นในการคัดเลือกได้จาก

$$P = \frac{N_{keep} - n + 1}{\sum_{n=1}^{N_{keep}} n}$$

P คือ ค่าความน่าจะเป็น

N_{keep} คือ จำนวนตัวอย่างเสียงทั้งหมดใน 1 กลุ่ม

n คือ ลำดับเสียง

ตัวอย่างการสุ่มเลือกค่าความน่าจะเป็นแสดงในตารางที่ 2.4

ตารางที่ 2.4 แสดงการจับคู่แบบสุ่มถ่วงน้ำหนักอันดับ

N	voice code	P_n	$\sum P_n$
1	1010101	0.4	0.4
2	1001000	0.3	0.7
3	1011001	0.2	0.9
4	1011000	0.1	1.0

ซึ่งถ้าเราเลือกค่าความน่าจะเป็น $P_n = 0.577$ แล้วก็จะตกอยู่ในช่วง $0.4 < P_n < 0.7$ ซึ่งเราจะทำการทดสอบเพียง $n=1$ และ 2 เท่านั้นถ้าเป็นเสียงมีความถี่ตรงกันก็จะหยุดการคัดเลือก แต่ถ้าเป็นเสียงต่างกันเราก็จะทำการเลือกค่าความน่าจะเป็นใหม่แล้วทำการทดสอบซ้ำเดิม จนกว่าจะได้เสียงที่มีความถี่ตรงกัน

ข้อดี คือ ถ้าเราทำการทดสอบครั้งแรกแล้วผ่านเลยจะทำให้ประหยัดเวลาและค่าใช้จ่ายได้มาก

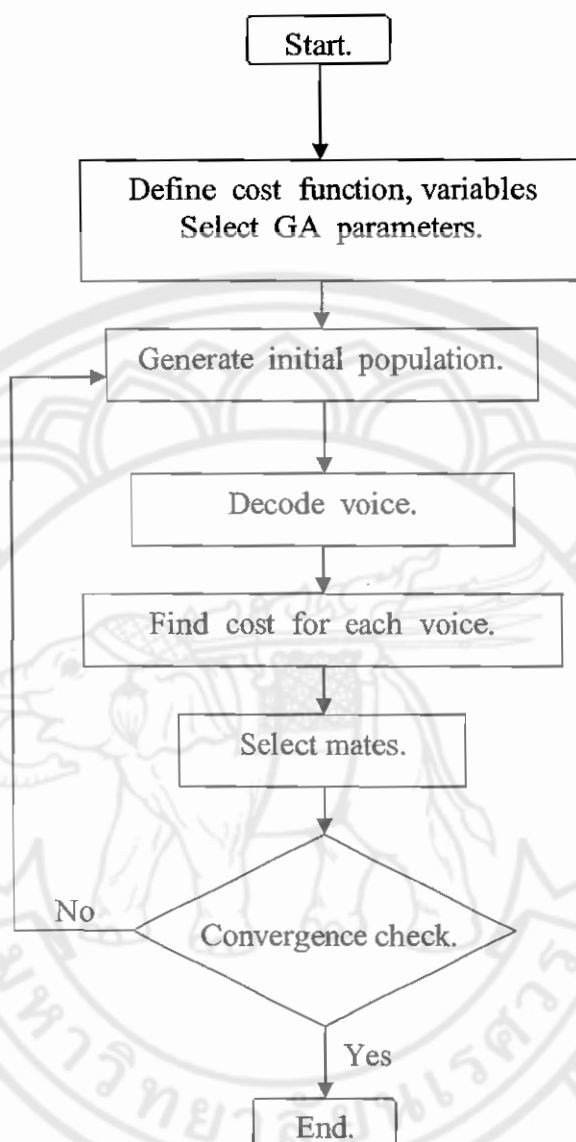
ข้อเสีย คือ ถ้าเราทำการทดสอบครั้งแรกไม่ผ่านแล้วต้องมาทำการทดสอบใหม่ ซึ่งจะทำให้เสียเวลาและค่าใช้จ่ายเพิ่มขึ้น

- การคัดสรรแบบแข่งขัน (Tournament Selection) จะเป็นการแบ่งกลุ่มจำนวนเสียงออกเป็นกลุ่มๆ ละ 2-3 เสียงแล้วทำการเปรียบเทียบเพื่อหาค่าที่ใกล้เคียงกับที่เสียงทดสอบมากที่สุดแล้วแยกออกมาแล้วนำไปรวมกับค่าที่ใกล้เคียงของกลุ่มย่อยอื่นๆ ทำเช่นนี้ไปเรื่อยๆ จนกว่าจะได้เสียงที่มีความถี่ใกล้เคียงที่สุด ซึ่งวิธีนี้จะเหมาะกับ population ที่มีขนาดใหญ่เพราะไม่ต้องทำการคัดเลือกทำให้ประหยัดเวลาในการคัดเลือกได้

สรุป วิธีการคัดสรรแบบแข่งขันและการจับคู่แบบสุ่มถ่วงน้ำหนักอันดับจะเป็นที่นิยมใช้ในมาก ที่สุด เพราะเหมาะสมสำหรับการหาค่าที่เหมาะสมจากกลุ่มตัวอย่างที่มีขนาดใหญ่

2.4.6 Convergence check

เป็นการเปรียบเทียบค่าความถี่เสียงว่ามีความใกล้เคียงกับเสียงตัวอย่างหรือไม่ ถ้ามีความถี่ตรงกันก็จะทำการหยุดการทดสอบแต่ถ้าไม่มีเสียงใดๆ ในกลุ่มตัวอย่างที่มีความถี่ตรงกับเสียงตัวอย่างเลย เราก็จะทำการสุ่มกลุ่มตัวอย่างใหม่แล้วทำการทดสอบซ้ำอีกครั้ง โดยที่ขั้นตอนการทำงานของจินตนิมิตอัลกอริทึมสามารถสรุปเป็นรูปแบบการทำงาน (Flow chart) ได้ดังนี้



รูปที่ 2.7 แผนภาพการทำงานของ GENETIC ALGORITHMS