

การออกแบบระบบการให้คะแนนความสามารถในการร้องเพลงถูกต้องตามโน้ตดนตรีอย่างยุติธรรม

Design of a Fair Scoring System for Singing Accuracy Based on Musical Notes

กฤษฎณ์ ธนพิพัฒนศิริ

มหาวิทยาลัยเกษตรศาสตร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

เลขที่ 50 ถนนงามวงศ์วาน แขวงลาดยาว เขตจตุจักร กรุงเทพมหานคร 10900

บทคัดย่อ

โครงการนี้ออกแบบระบบการให้คะแนนความสามารถในการร้องเพลงถูกต้องตามโน้ตดนตรีอย่างยุติธรรม โดยการวัดความแม่นยำของการร้องเพลงในแต่ละมิติเพื่อนำมาประเมินค่าความแม่นยำผลลัพธ์ในที่สุด มีวัตถุประสงค์ในการออกแบบระบบการให้คะแนนอย่างยุติธรรมเชิงทฤษฎีดนตรีและสะดวกต่อการใช้งาน เริ่มต้นการศึกษาโดยการเลือกรูปแบบไฟล์บันทึกทำนองของเพลงหนึ่ง และรูปแบบไฟล์บันทึกเสียงร้องเพลง แล้วพัฒนาขั้นตอนวิธีในการเปรียบเทียบทำนองเพลงและเสียงร้องบันทึกในแต่ละมิติ พร้อมพัฒนาแนวทางที่ยุติธรรมในการคำนวณค่าความแม่นยำในแต่ละมิติและโดยรวม ซึ่งได้ผลการศึกษาคือรูปแบบไฟล์ที่เหมาะสมคือ MIDI ในการบันทึกทำนองเพลงเนื่องจากความแพร่หลายในการใช้งานและ WAV ในการบันทึกเสียงเนื่องจากความไม่สูญเสียในการเข้ารหัส ขั้นตอนวิธีที่ออกแบบหาความถี่มูลฐานของเสียงร้องจากขั้นตอนวิธี YIN เนื่องจากความแม่นยำในการประมวลผลเสียงมนุษย์ และนำมาเปรียบเทียบกับหมายเลขโน้ตของทำนองในแต่ละกรอบเวลาหลังผ่านการจัดตำแหน่งเชิงเวลาและเชิงระดับเสียง แล้วคำนวณค่าการลงโทษออกจากคะแนนเต็มในมิติความตรงต่อเวลา ระดับเสียง ฤกษ์เสียง และความครอบคลุมทำนองนำมาคำนวณคะแนนรวมในที่สุด โดยให้น้ำหนักความตรงต่อเวลา ระดับเสียง และความครอบคลุมทำนองอย่างมาก สำหรับความตรงต่อระดับเสียงและฤกษ์เสียง ได้พิจารณาแนวคิดทฤษฎีดนตรีเรื่องขั้นคู่เสียงเพื่อความยุติธรรม ผลสรุปคือระบบที่พัฒนาขึ้นมาสามารถให้คะแนนความแม่นยำในการร้องเพลงได้อย่างถูกต้องและยุติธรรมตามจุดหมาย พร้อมได้พัฒนาอินเทอร์เฟซผู้ใช้เพื่อความสะดวกในการใช้งาน หากมีความประสงค์ทดลองใช้งานระบบสามารถเข้าถึงโค้ดแหล่งที่มาได้ที่ <https://github.com/krtchnt/cantometria>

คำสำคัญ: ทฤษฎีดนตรี ขั้นตอนวิธี YIN คาราโอเกะ MIDI FFT

Abstract

This project designs a fair scoring system for singing accuracy based on musical notes by evaluating the accuracy in various aspects to be computed as a resulting accuracy, with the purpose of basing fairness on music theory while keeping accessibility. The study begins with the choice of file types to store melody and audio recording information, then designing a fair algorithm to evaluate the accuracy of the recording in various aspects as well as the resulting accuracy. The study result includes the file type choice being MIDI due to ubiquity and WAV due to lossless encoding. The algorithm employs the YIN algorithm for its accuracy with human voices to compute the fundamental frequency comparing it to the note number of the reference melody in each time frame after time alignment and amplitude alignment. The accuracy will be evaluated from: time alignment, key alignment, pitch alignment and melodic coverage. The resulting accuracy will be computed from these aspects, with an emphasis on time alignment, pitch alignment and melodic coverage. In key and pitch alignment, the concept of interval in music

theory is considered for fairness. In conclusion, the developed system is able to score singing accuracy fairly, and a user interface was also developed for ease of use. Should anyone be interested in testing the developed system, the source code is available at <https://github.com/krtchnt/cantometria>.

Keywords: Music Theory, YIN Algorithm, Karaoke, MIDI, FFT

1. บทนำ

1.1 ที่มาและความสำคัญของปัญหา

โครงการนี้ออกแบบระบบการให้คะแนนความสามารถในการร้องเพลงถูกต้องตามโน้ตของมนุษย์ รวมทั้งคนที่ผ่านการฝึกฝนร้องเพลงมาเป็นมืออาชีพ และคนทั่วไปที่ไม่ผ่านการฝึกฝนมา และเปรียบเทียบคะแนนว่าตอบสนองกับความชำนาญของแต่ละคนหรือไม่ ระบบจะคำนึงถึงแนวคิดเชิงทฤษฎีดนตรี โดยเฉพาะเรื่องขั้นคู่เสียง ควบคู่กับความถี่ของเสียงร้องโดยตรงเพื่อให้มีความยุติธรรมมากขึ้นสำหรับการร้องเพลงที่ไม่ถูกต้องตามโน้ตทุกประการแต่สำหรับคนทั่วไปยังคงฟังดูมีความสุนทรีย์

ทางผู้ดำเนินการโครงการเห็นว่าเรื่องเกี่ยวกับทฤษฎีดนตรี การร้องเพลง การแต่งเพลง ฯลฯ เป็นเรื่องที่สังคมเรารู้สึกว่ากันมานาน ถึงแม้ว่ามีแอปพลิเคชันและศูนย์คาราโอเกะมากมายที่มีระบบให้คะแนนความสามารถในการร้องเพลงถูกต้องตามโน้ตระบบส่วนมากค่อนข้างจะเคร่งครัดเกินไป ไม่อนุญาตให้เกิดข้อผิดพลาดเชิงมนุษย์โดยที่ไม่เสียคะแนนได้เลย หรือให้คะแนนไม่ตรงกับความคิดเห็นของคนอื่นที่ฟังเสียงร้องมาแล้ว จึงพัฒนามิตินี้ของระบบด้วยวิธีใหม่

1.2 วัตถุประสงค์หลักของโครงการ

พัฒนาขั้นตอนวิธีใหม่สำหรับการประมวลผลสัญญาณเสียงร้องมนุษย์และสร้างเกณฑ์ที่เหมาะสมสำหรับเพลงแต่ละเพลง และให้คะแนนการร้องเพลงของคนนั้นให้มีความยุติธรรมไม่เคร่งครัดเกินไป รองรับรอบความผิดพลาดของมนุษย์ รวมถึงข้อจำกัดในการร้องเพลงของมนุษย์ที่ระบบที่ไร้เดียงสาไม่สามารถจับความสุนทรีย์ของดนตรีได้และถือว่าเป็นข้อผิดพลาด นอกจากนี้แล้วได้ใช้ขั้นตอนวิธีที่มีอยู่แล้วมาประยุกต์ใช้ความรู้ด้านการประมวลผลสัญญาณในการแก้ปัญหาดังกล่าวได้อย่างสำเร็จผล เช่นขั้นตอนวิธี FFT หรือขั้นตอนวิธี YIN

1.3 ขอบเขตของโครงการ

การทดลองเป็นในลักษณะการออกแบบโครงสร้างของแนวทางและพัฒนาขั้นตอนวิธีใหม่ในที่สุดโดยเริ่มต้นโดยให้คำจำกัดความของคำว่าความสุนทรีย์ภาพเชิงดนตรีแล้วการออกแบบมาตรวัดภายใต้ขอบเขตของคำจำกัดความดังกล่าวโดยจะบูรณาการแนวคิดเชิงทฤษฎีดนตรีมาช่วยให้คำจำกัดความและเน้นความเป็นยุติธรรมของระบบ โดยได้ออกแบบระบบที่ไม่ยอมรับความผิดพลาดใด ๆ ทั้งสิ้นแล้วจึงเพิ่มการยอมรับข้อผิดพลาดของมนุษย์ในระบบในมาตรที่ถูกตัดสินใจเห็นว่าเหมาะสม

สัญญาณที่นำมาประมวลผลคือสัญญาณเสียงโดยเฉพาะสำหรับเสียงการร้องเพลงของมนุษย์ รวมทั้งเสียงเชิงสูงของผู้หญิงส่วนมากและเสียงเชิงต่ำของผู้ชายส่วนมาก ในกรณีที่เสียงของมนุษย์ที่เปล่งออกมาภายใต้บริบทเชิงดนตรีโครง เช่น เสียงฮัม งานนี้ถือว่าเข้าข่ายขอบเขตเช่นเดียวกัน ถึงแม้ไม่ใช่สิ่งที่ศึกษาโดยตรง โครงการนี้ไม่รับรองเสียงอื่น ๆ ของมนุษย์ หรือเสียงร้องจากสัตว์อื่น ๆ ที่ไม่ใช่มนุษย์และโครงการนี้จะไม่ทำการแยกเสียงมนุษย์ออกจากเสียงเครื่องดนตรี ในบริบทของเสียงร้องเพลงของมนุษย์ที่มีเสียงเครื่องดนตรีของเพลงอยู่ในเบื้องหน้าและเบื้องหลังและจะศึกษาเพียงแค่เสียงของมนุษย์บริสุทธิ์เท่านั้น นอกจากนี้เสียงที่บันทึกมาต้องเริ่มต้นพร้อมกับจังหวะแรกของทำนองอ้างอิง เพราะระบบถือว่าความไม่ตรงเวลาของทำนองเป็นข้อผิดพลาด และถูกนำมาคิดความแม่นยำเช่นกัน

ข้อจำกัดของโครงการคือ ไม่มีคำจำกัดความของคำว่าความสุนทรีย์ภาพเชิงดนตรี ที่ทุกคนสามารถตกลงกันได้ โครงการนี้จึงได้ออกแบบระบบที่ไม่เชิงปฏิบัติจนเกินไปและใช้เส้นพื้นฐานที่คนทั่วไปส่วนมากจะตกลงกัน อย่างไรก็ตาม เนื่องจากโครงการได้ออกแบบแบบจำลองที่ควบคุมโดยพารามิเตอร์ คนที่มีความสนใจอยากต่อยอดโครงการสามารถปรับแก้ค่าของพารามิเตอร์ให้เป็นค่าใหม่ ซึ่งอาจจะเพิ่มหรือลดความยุติธรรมในความคิดเห็น

ส่วนตัวของแต่ละคนได้เสมอ เพราะว่าโค้ดแหล่งที่มาเปิดเผยสู่โลกสาธารณะ

2. วิธีที่นำเสนอ

2.1 แนวทางและทฤษฎีต่าง ๆ ที่ใช้

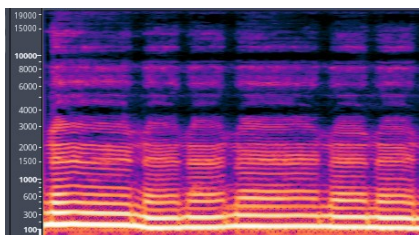
2.1.1 ทฤษฎีและแนวทางจากโสตศาสตร์

ความถี่มูลฐาน กล่าวถึงความถี่ที่ต่ำที่สุดของรูปคลื่นเป็นระยะ นิยมเขียนแทนด้วย f_0 และส่วนกลับคือ T เรียกว่าคาบมูลฐาน ซึ่งนิยามว่าเป็นจำนวนจริงบวกที่เล็กที่สุดที่ทำให้สมการดังกล่าวเป็นจริง สำหรับรูปคลื่น $x(t)$

$$\forall t \in \mathbb{R} \ x(t) = x(t+T) \quad (1)$$

พหุคูณจำนวนเต็มบวกที่มากกว่า 1 ของความถี่มูลฐานเรียกว่าเป็นความถี่โอเวอร์โทน และเรียกพหุคูณรวมทั้งหมดว่าเป็นสมาชิกของอนุกรมฮาร์โมนิกเดียวกัน

เสียงของมนุษย์ มีลักษณะประกอบด้วยความถี่มูลฐาน ซึ่งเป็นระดับเสียงพื้นฐานที่เราได้ยิน และความถี่โอเวอร์โทนในหลาย ๆ ระดับสูงขึ้นไป ซึ่งสามารถเห็นได้อย่างชัดเจนจากรูปภาพดังต่อไปนี้



รูปที่ 1 สเปกโตรแกรมของเสียงมนุษย์ร้องเพลง เส้นสีขาวแทนความถี่ที่โดดเด่นในการบันทึก ณ เวลานั้น

ขั้นตอนวิธีของ YIN สามารถประมาณค่าความถี่มูลฐานที่ถูกพัฒนามาให้เหมาะสมกับเสียงพูดมนุษย์ มีข้อดีคือสามารถลดผลกระทบของฮาร์โมนิกส์และเสียงรบกวนได้ดี โดยมีหลักการการทำงานอย่างสังเขปคือ คำนวณฟังก์ชันความแตกต่าง แทนฟังก์ชันอัตราสหสัมพันธ์มาตรฐานเพื่อลดผลกระทบจากฮาร์โมนิก เมื่อมีสัญญาณเสียง $x(t)$ และกรอบเวลายาว N

$$d(\tau) = \sum_{t=1}^{N-\tau} (x(t) - x(t+\tau))^2 \quad (2)$$

แล้วคำนวณฟังก์ชันผลต่างปกติเฉลี่ยสะสม ลดผลกระทบจากฮาร์โมนิก

$$\tilde{d}(\tau) = \frac{d(\tau)}{\frac{1}{\tau} \sum_{j=1}^{\tau} d(j)} \quad (3)$$

หาค่าแห่งเวลา τ_0 คือค่า τ ต่ำที่สุดที่ทำให้ $\tilde{d}(\tau)$ มีค่าน้อยกว่าค่าเกณฑ์ที่กำหนด แล้วหลังจากการสอดคล้องทั้งสองรอบ ๆ ค่าที่พบเพื่อเพิ่มความแม่นยำ สุดท้ายแล้ว เมื่อ f_s เป็นอัตราการซีกตัวอย่าง จะได้ความถี่มูลฐานดังนี้

$$f_0 = \frac{f_s}{\tau_0} \quad (4)$$

2.1.2 ทฤษฎีและแนวทางจากดนตรี

ขั้นคู่เสียง กล่าวถึงความแตกต่างของระดับเสียงระหว่างเสียงสองเสียง ในดนตรีสากล หน่วยขั้นคู่เสียงที่เล็กที่สุดคือครึ่ง 1 ครั้งเสียง และแต่ละขั้นคู่เสียงก็มีชื่อเรียกเฉพาะ ซึ่งสอดคล้องกับพหุคูณของความถี่มูลฐานดังตารางต่อไปนี้

ตารางที่ 1 ความสัมพันธ์ระหว่างชื่อของขั้นคู่เสียงที่พบบ่อยและพหุคูณของความถี่มูลฐาน

ชื่อเรียกของขั้นคู่เสียง	พหุคูณของความถี่มูลฐาน
คู่แปดสมบูรณ์	2:1
คู่ห้าสมบูรณ์	3:2
คู่สี่สมบูรณ์	4:3
คู่สามเมเจอร์	5:4
คู่สามไมเนอร์	6:5
คู่สองเมเจอร์ / เต็มเสียง	9:8
คู่สองไมเนอร์ / ครึ่งเสียง	16:15

ส่วนมากแล้วความแตกต่างระดับเสียงจะวัดเป็นพหุคูณของครึ่งเสียงคูณ 100 เรียกว่าหน่วยเซนต์ (cent) และจะวัดระดับเสียงด้วยตัวเลขโน้ต เมื่อเลข 60 คือโน้ตโดกลาง และความแตกต่างตัวเลขโน้ต ทุก ๆ 1 เท่ากับความแตกต่าง 1 ครั้งเสียง

เมื่อเราับฟังเสียงระดับเสียงแตกต่างกันเป็นคู่แปดสมบูรณ์ เราจะถือว่าเป็นโน้ตเดียวกันและในขั้นคู่เสียงอื่นที่เป็น

พหุคูณของความถี่มูลฐานอย่างง่าย เช่น โน้ตหนึ่ง และคู่ห้า สมบูรณ์ของโน้ตนั้น เราจะรับฟังเสียงประสานเหล่านี้กลมกลืนกัน และมีความสุนทรีทางดนตรี เนื่องจากมนุษย์ถูกวิวัฒนาการมา ให้รับฟังผลต่างของระดับเสียงเป็นแบบลอการิทึม กล่าวคือ สามารถเขียนรูปความสัมพันธ์มาตรฐานระหว่างหมายเลขโน้ต และความถี่ในหน่วยเป็นเฮิรตซ์ได้ดังนี้ เมื่อ n แทนตัวเลขโน้ตและ f แทนความถี่

$$n = 69 + 12 \cdot \log_2 \frac{f}{440} \quad (5)$$

เทมโป เป็นความเร็วในของงานประพันธ์ดนตรีหนึ่ง ซึ่งส่วนมากจะวัดเป็น จังหวะต่อนาที แต่บางครั้งก็วัดเป็น มิลิวินาที ต่อโน้ตตัวดำ

2.1.3 รายละเอียดของไฟล์ MIDI

MIDI เป็นนโยบายมาตรฐานสำหรับระบบติดต่อการสื่อสารทางดนตรีของอุปกรณ์อิเล็กทรอนิกส์ทางดนตรี โดยไม่ได้มีการเก็บเสียงดนตรีใด ๆ ไว้เหมือนเสียงบันทึก ข้อมูลทั้งหมดจะอยู่ในรูปของคำสั่งที่จะสั่งให้เครื่องดนตรีทำงาน เช่น กดหรือปล่อย โน้ตหลายเลขใด ณ เวลาไหน และความดังเท่าไร จึงทำให้เป็นไฟล์ มีขนาดเล็กมาก และสะดวกในการส่งต่อกันบนโลกอินเทอร์เน็ต และพัฒนาระบบเกี่ยวกับดนตรีเนื่องจากความเรียบง่ายของโครงสร้างไฟล์ ตามมาตรฐานแล้วไฟล์จะไม่ได้เก็บเวลาคำสั่งเป็นหน่วยเวลา แต่เก็บเป็นหน่วยตึก (Tick) ซึ่งสามารถแปลงกลับเป็น เวลาได้โดยเมตาตาต้าในไฟล์ส่วนการแบ่งหน่วยเวลา ซึ่งหากอยู่ในรูปแบบ ตึกต่อโน้ตตัวดำ สามารถเปลี่ยนเป็นเวลาหน่วยวินาที ได้ดังต่อไปนี้ เมื่อ τ แทนจำนวนตึก t แทนเวลาหน่วยวินาที r_{bpm} แทนเทมโปหน่วยจังหวะต่อนาที และ r_{tpqn} แทนเทมโปหน่วยตึกต่อโน้ตตัวดำ

$$t = \frac{\tau \cdot r_{bpm}}{r_{tpqn} \cdot 10^6} \quad (6)$$

2.1.4 รายละเอียดของไฟล์ WAV

WAV เป็นมาตรฐานรูปแบบไฟล์เสียงในการเก็บบิตสตรีมของเสียงบนระบบคอมพิวเตอร์ส่วนตัว ส่วนมากแล้วจะเก็บข้อมูลเสียงแบบไม่ถูกบีบอัดและเข้ารหัสแบบไม่สูญเสีย ตามมาตรฐาน แล้วตัวไฟล์เองจะมีเมตาตาต้าเก็บข้อมูลอัตราการซิกตัวอย่าง

จำนวนของช่องเสียง และข้อมูลอื่น ๆ ไว้ เมื่อทำการสตรีมแต่ละบิตในไฟล์ จะได้ตัวอย่างเสียงในแต่ละช่องเสียงไล่ตามลำดับและไล่วนลำดับเดิมสำหรับตัวอย่างถัดไปจนกว่าจะหมดตัวอย่าง เช่น หากมีช่องเสียง 2 ช่อง คือซ้ายและขวา จะได้ลำดับของตัวอย่างดังนี้: $l_1, r_1, l_2, r_2, l_3, r_3, \dots$

หากเราทำการแบ่งตัวอย่างเป็นกลุ่มติดกันขนาดเท่ากัน ซึ่งอาจจะเรียกว่าหน้าต่าง สามารถคำนวณขนาดของแต่ละกลุ่ม Δt ได้ดังนี้ เมื่อ n_c แทนจำนวนช่องเสียง และ N_w แทนจำนวนตัวอย่างในแต่ละกลุ่ม และ f_s แทนอัตราการซิกตัวอย่าง

$$\Delta t = \frac{N_w}{n_c \cdot f_s} \quad (7)$$

2.1.5 การถือค่าอันดับที่ศูนย์

การถือค่าอันดับที่ศูนย์ (Zero-order hold) เป็นแบบจำลองคณิตศาสตร์ในการสร้างสัญญาณใหม่ในเชิงปฏิบัติเช่นเดียวกับเครื่องแปลงสัญญาณดิจิทัลเป็นอนาล็อกธรรมดา นั่นคือการแปลงสัญญาณเวลาไม่ต่อเนื่องเป็นสัญญาณเวลาต่อเนื่องโดยการถือค่าตัวอย่างในแต่ละค่าไว้ตลอดหนึ่งช่วงซิกตัวอย่าง หรือกล่าวอีกนัยที่ทั่วไปกว่า คือตลอดช่วงเวลา $\tau[n]$ แทนระยะเวลาจนกว่าตัวอย่างต่อไปสำหรับสัญญาณไม่สม่ำเสมอ สามารถเขียนความสัมพันธ์ได้ดังนี้ เมื่อ $x[n]$ แทนสัญญาณเวลาไม่ต่อเนื่อง $y(t)$ แทนสัญญาณเวลาต่อเนื่อง และ $\Pi(t)$ แทนฟังก์ชันสี่เหลี่ยมหน่วย

$$y(t) = \sum_{n=-\infty}^{\infty} x[n] \cdot \Pi\left(\frac{t - \frac{\tau[n]}{2} - n\tau[n]}{\tau[n]}\right) \quad (8)$$

2.1.6 การเติมศูนย์

การเติมศูนย์เป็นการเติมค่า 0 ตรงปลายของสัญญาณโดเมนเวลาเพื่อเพิ่มขนาด ส่วนมากแล้วมีจุดประสงค์เพื่อเพิ่มประสิทธิภาพของขั้นตอนวิธี FFT เช่น เติม 0 ให้ขนาดของจำนวนของตัวอย่างเป็นกำลังของสอง ทำให้ใช้ขั้นตอนวิธี FFT ที่รากเลข 2 ได้อย่างมีประสิทธิภาพ อีกจุดประสงค์คือสำหรับการเปรียบเทียบสัญญาณสองสัญญาณที่มีขนาดใกล้เคียงกัน การเติมศูนย์สามารถทำให้เปรียบเทียบสัญญาณได้แบบหนึ่งต่อหนึ่งได้ไปตลอดทั้งสองสัญญาณ ครอบคลุมทุกตัวอย่าง

2.1.7 ขั้นตอนวิธี FFT และการจัดตำแหน่งเวลาของสัญญาณ

ขั้นตอนวิธี FFT หรือการแปลงฟูเรียร์แบบเร็ว (Fast Fourier Transform) เป็นการหาผลแปลงฟูเรียร์ไม่ต่อเนื่อง หรือผลแปลงย้อนกลับของลำดับหนึ่งได้อย่างรวดเร็วโดยการแยกตัวประกอบเมทริกซ์ผลแปลงฟูเรียร์ไม่ต่อเนื่องให้เป็นผลคูณที่มากเลขศูนย์ ซึ่งมีผลทำให้ลดความซับซ้อนเวลาการทำงานลงจาก $O(n^2)$ เป็น $O(n \log n)$ เมื่อ n เป็นขนาดของข้อมูล

เมื่อก้าวในนัยเชิงคณิตศาสตร์แล้ว สำหรับการจัดตำแหน่งเวลาของสัญญาณสองสัญญาณให้ได้เวลาเลื่อนที่เหมาะสมที่สุด τ_a สามารถคำนวณได้จากการหาสหสัมพันธ์ไขว้ เมื่อ $T[n]$ แทนสัญญาณเป้าหมายและ $I[n]$ แทนสัญญาณเข้า

$$T \star I = \sum_i T[i] I[i + \tau] \quad (9)$$

จากทฤษฎีบทสังวัตนาการ

$$\mathcal{F}(T \star I) = \mathcal{F}(T) \cdot \overline{\mathcal{F}(I)} \quad (10)$$

หลังจากนำผลการแปลงฟูเรียร์มาหาผลแปลงย้อนกลับ ก็จะได้ค่าสหสัมพันธ์ไขว้มา และสุดท้ายแล้ว สามารถคำนวณหาเวลาเลื่อนที่เหมาะสมที่สุด τ_a จาก

$$\tau_a = \arg \max_{\tau \in 0, \dots, n-1} \Re((T \star I)[\tau]) \quad (11)$$

ข้อสังเกตคือ หาก $\tau_a > \frac{n}{2}$ นั้นแปลว่าเวลาเลื่อนที่เหมาะสมที่สุดต้องเป็นลบ เนื่องจากขั้นตอนวิธี FFT เปรียบเสมือน input ว่าเป็นระยะ สุดท้ายแล้ว หมุนอาร์เรย์สัญญาณเข้าเสมือนวงกลมไป τ_a ตำแหน่งเลื่อนก็จะได้สัญญาณเข้าใหม่ที่ตรงกับสัญญาณเป้าหมายทางเวลามากที่สุด

2.1.8 การจัดตำแหน่งช่วงกว้างของสัญญาณ

การจัดตำแหน่งความกว้างของสัญญาณนั้น หนึ่งในวิธีที่สามารถคำนวณได้อย่างเรียบง่ายคือการหาค่าเฉลี่ยเลขคณิตของผลต่างเชิงคู่ ซึ่งต้องคำนวณเซตของดัชนีที่สัญญาณทั้งสองไม่เป็นศูนย์ S นิยามดังนี้ เมื่อ $T[n]$ แทนสัญญาณเป้าหมายและ $I[n]$ แทนสัญญาณเข้า ทั้งคู่จำนวนตัวอย่าง N

$$S = \{j \in [1, N] | (T[j], I[j]) \neq (0, 0)\} \quad (12)$$

ค่าเฉลี่ยเลขคณิตของผลต่างเชิงคู่ Δ หาได้จาก

$$\Delta = \frac{1}{|S|} \sum_{j \in S} (I[j] - T[j]) \quad (13)$$

สุดท้ายแล้ว นำ Δ มาบวกแต่ละค่าของตัวอย่างสัญญาณเข้าที่ไม่ใช่ศูนย์ทุกค่า

2.2 วิธีการแก้ปัญหา

2.2.1 เปิดไฟล์ MIDI ที่เลือก เข้าถึงข้อมูลการแบ่งหน่วยเวลา r_{tpqn} แล้วตรวจสอบให้มั่นใจว่าไฟล์มีเครื่องดนตรีเพียงแค่ 1 (ซึ่งจะถูกถือว่าเป็นทำนองหลักของเพลงที่จะใช้เป็นเกนท์) วนซ้ำแต่ละคำสั่ง หากเป็นคำสั่งเปลี่ยนเทมโปให้อัปเดตเทมโปปัจจุบัน r_{bpm} หากเป็นคำสั่งกดโน้ตให้เติมหมายเลขโน้ตลงไปในสัญญาณใหม่ $T^*[n]$ หากเป็นคำสั่งปล่อยโน้ตให้เติมเลขศูนย์ คำสั่งทั้งคู่ทำให้เวลาปัจจุบัน t ที่บวกสะสมคำนวณมาจาก r_{tpqn} , r_{bpm} และ τ

2.2.2 เปิดไฟล์ WAV ที่เลือก เข้าถึงข้อมูลอัตราการซีกตัวอย่างและจำนวนช่องเสียงมาคำนวณขนาดของหน้าต่าง วนซ้ำไปแต่ละตัวอย่างในไฟล์แล้วทำการประมวลผลเมื่อหน้าต่างเต็ม โดยจะนำตัวอย่างทั้งหมดในหน้าต่างไปใช้หาความถี่มูลฐาน ณ กรอบเวลาสั้น ๆ นั้น แล้วนำความถี่มาแปลงเป็นหมายเลขโน้ต เติมเข้าไปในสัญญาณใหม่ $I[n]$ เลื่อนไปหน้าต่างต่อไป แล้ววนซ้ำจนกว่าตัวอย่างหมด สำหรับ parameter ในการคำนวณส่วนนี้ได้แก่ $N_w = 1024$ คือขนาดของหน้าต่าง และ parameter สำหรับขั้นตอนวิธี YIN ได้แก่ $p = \frac{N_w}{2}$ คือขนาดของการเติมช่องว่าง $P_T = 1$ คือเกนท์กำลังเสียงขั้นต่ำ และ $\chi_T = 0.9$ คือเกนท์ขั้นต่ำความมั่นใจในการประมาณค่าความถี่มูลฐานของขั้นตอนวิธี

2.2.3 นำสัญญาณไม่สม่ำเสมอ $T^*[n]$ มาทำการถี้อค่าอันดับที่ศูนย์ โดยจะใช้อัตราการซีกตัวอย่างของ $I[n]$ ได้สัญญาณใหม่ $T[n]$

2.2.4 นำสัญญาณ $T[n]$ และ $I[n]$ มาเปรียบเทียบจำนวนตัวอย่าง แล้วทำการเติมศูนย์ให้กับสัญญาณที่สั้นกว่า จนกว่าสัญญาณทั้งสองจะมีจำนวนตัวอย่าง N เท่ากัน

2.2.5 นำสัญญาณ $T[n]$ และ $I[n]$ มาจัดตำแหน่งเวลาของสัญญาณ $T[n]$ ให้ตรงกับ $I[n]$ โดยคำนวณเวลาเลื่อนที่เหมาะสมที่สุด τ_a

2.2.6 นำสัญญาณ $T[n]$ และ $I[n]$ มาจัดตำแหน่งช่วงกว้างของสัญญาณ $T[n]$ ให้ตรงกับ $I[n]$ โดยคำนวณช่วงกว้างเลื่อนที่เหมาะสมที่สุด Δ

2.2.7 คำนวณคะแนนความแม่นยำ โดยจะแบ่งออกเป็น 4 มิติ

มิติแรกคือ คะแนนความครอบคลุมทำนอง α_c เริ่มโดยสร้างเซต S แทนคู่อันดับตัวอย่างของ $T[n]$ และ $I[n]$ สำหรับทุกดัชนีที่สัญญาณทั้งสองไม่เป็นศูนย์ นิยามดังนี้

$$S = \{(T[i], I[i]) | (T[i], I[i]) \neq (0, 0)\} \quad (14)$$

และ m แทนจำนวนสมาชิกที่ $T[n] \neq 0$ แต่ $I[n] = 0$ นิยามดังนี้

$$m = |\{i | T[i] \neq 0 \wedge I[i] = 0\}| \quad (15)$$

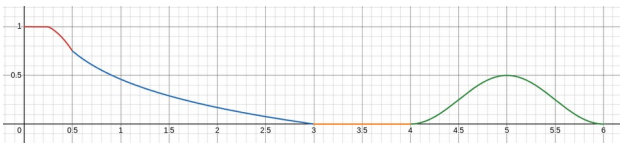
ค่าความครอบคลุมทำนอง α_c คำนวณได้ดังนี้

$$\alpha_c = \sqrt[3]{1 - \frac{m}{n}} \quad (16)$$

ข้อสังเกตคือ รากที่สามของจำนวนที่น้อยกว่า 1 จะมีค่ามากกว่าเดิม มีผลทำให้ $\frac{m}{n}$ ที่ไม่มากเกินไปจะไม่ส่งผลต่อค่าของ α_c ให้น้อยกว่า 1 มากเกินไปเช่นเดียวกัน

มิติที่สองคือ คะแนนความตรงต่อระดับเสียง α_p เริ่มต้นจากการนิยามฟังก์ชันคำนวณค่าความแม่นยำจากพหุคูณของครึ่งเสียงที่คาดเคลื่อนไป $g_k(d)$ มีนิยามดังนี้

$$g_k(d) = \begin{cases} \min\{-2d^2 + \frac{1}{2}d + 1, 1\} & , 0 \leq d < 0.5 \\ \frac{3}{4} \log_6 3 - \frac{3}{4 \ln 6} \ln d & , 0.5 \leq d < 3 \\ 0 & , 3 \leq d < 4 \\ -\frac{1}{4} \cos(\pi d) + \frac{1}{4} & , 4 \leq d < 6 \\ 0 & , otherwise \end{cases} \quad (17)$$



รูปที่ 2 กราฟของฟังก์ชัน $g_k(d)$ เมื่อแกน d แนวนอนและแกนความแม่นยำแนวตั้ง

ข้อสังเกตคือ เมื่อความคาดเคลื่อนน้อยกว่า 0.5 คะแนนจะมีแนวโน้มลดด้วยความชันที่ลดลงเป็นสมการกำลังสอง และในช่วง 0.5 ถึง 3 คะแนนจะมีแนวโน้มโค้งเว้าออกแบบซ้ำจนตัดแกน x ที่ (3,0) และในช่วง 3 ถึง 4 คะแนนกับในช่วง 6 ขึ้นไปจะมีค่าคงที่เป็น 0 ช่วงนี้พิจารณาเพียงค่าความคาดเคลื่อนที่น้อย หรือประมาณเป็นคู่หนึ่งสมบูรณ์ (พหุคูณความถี่มูลฐาน 1:1) กล่าวอีกนัยหนึ่งคือเป็นโน้ตเดียวกับโน้ตอ้างอิงหนึ่ง สำหรับช่วงสุดท้ายนั้น ช่วง 4 ถึง 6 จะมีการชดเชยคะแนนที่ควรจะเป็น 0 เนื่องจากเป็นช่วงโน้ตที่ใกล้เคียงคู่สมบูรณ์ของโน้ตอ้างอิง และใกล้เคียงมีคู่แปดสมบูรณ์ของโน้ตอ้างอิงเป็นคู่ห้าสมบูรณ์ เป็นการเพิ่มความยุติธรรมให้คนที่ออกเสียงผิดโน้ตแต่ยังคงฟังดูมีความสุนทรีทางดนตรี สุดท้ายแล้ว คำนวณ α_p ได้จากนิยามดังนี้

$$\alpha_p = med(g_k(|x-y|) | (x,y) \in S) \quad (18)$$

นั่นคือ เป็นการคำนวณค่ามัธยฐานของเซตของ g_k ของผลต่างหมายเลขโน้ตแต่ละคู่ในการวัดค่ากลาง สาเหตุที่ใช้ค่ามัธยฐานเนื่องจากชุดข้อมูลมักจะเบ้ทางใดทางหนึ่ง และมักจะมีค่านอกเกณฑ์สูง ไม่เหมาะสมกับการใช้ค่าเฉลี่ยเลขคณิต

มิติที่สามคือ คะแนนความตรงต่อเวลา α_t เริ่มต้นจากการคำนวณเวลาเลื่อนที่ดีที่สุดหน่วยวินาที

$$\Delta \tau = \frac{\tau_a}{f_s} \quad (19)$$

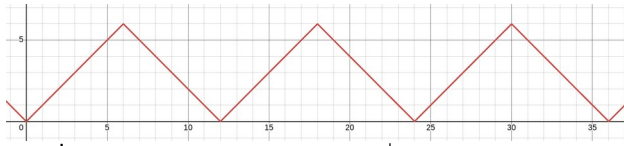
แล้วนำเวลามาคิดคะแนนดังนี้

$$\alpha_t = \begin{cases} 1 & , 0 \leq d < 0.05 \\ \frac{|\Delta \tau| - 0.05}{0.2} & , 0.05 \leq d < 0.25 \\ 0 & , otherwise \end{cases} \quad (20)$$

ข้อสังเกตคือ เวลาเลื่อนที่น้อยกว่า 0.05 วินาทีจะไม่ถูกหักคะแนน เวลาเลื่อนที่อยู่ในช่วง 0.05 ถึง 0.25 จะถูกหักคะแนนแนวโน้มลดลงด้วยความชันคงที่ และเวลาเลื่อนที่มากกว่า 0.25 จะได้ 0 คะแนน

มิติที่สี่คือ คะแนนความตรงต่อท่วงเสียง α_k เริ่มต้นจากการนิยามฟังก์ชันคำนวณผลต่างของพหุคูณของครึ่งเสียงและช่วงคู่แปดถัดไปที่ใกล้เคียงที่สุด $r_o(d)$ มีนิยามดังนี้

$$r_o(d) = \min\left\{\left|d - \left\lfloor \frac{x}{12} \right\rfloor \cdot 12\right|, \left|d - \left\lceil \frac{x}{12} \right\rceil \cdot 12\right|\right\} \quad (21)$$



รูปที่ 3 กราฟของฟังก์ชัน $r_o(d)$ เมื่อแกน d แนวนอนและแกนผลต่างระดับเสียงแนวตั้ง

แล้วนำมาคิดคะแนนดังนี้

$$\alpha_k = g_k(r_o(\Delta)) \quad (22)$$

การเรียกใช้ฟังก์ชัน r_o จะทำให้ช่วงกว้างเลื่อนที่เหมาะสมที่สุด Δ ที่ใกล้เคียงกับช่วงคู่แปดหนึ่ง ถูกขุดเซยคะแนนที่ถูกหักไปอย่างสิ้นเชิงและจะไม่ถูกหักคะแนนเลย และเป็นการเพิ่มความยุติธรรมสำหรับคนที่พิสัยเสียงร้องแคบและระดับเสียงของทำนองเพลงอยู่ภายนอกพิสัยนั้น

2.2.8 คะแนนความแม่นยำในการร้องเพลงผลลัพธ์ α คำนวณดังนี้

$$\alpha = \alpha_c \cdot \alpha_t \cdot (\alpha_p \cdot w_p + \alpha_k \cdot (1 - w_p)) \quad (23)$$

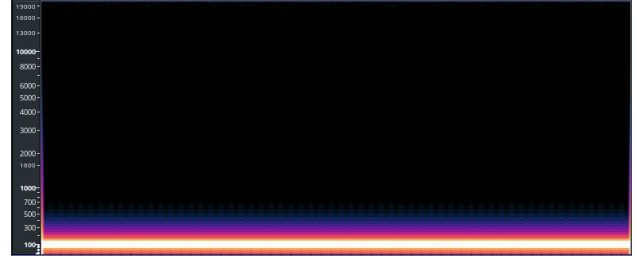
โดย $w_p = 0.75$ แทนน้ำหนักของความตรงต่อระดับเสียง เป็น parameter ที่สามารถปรับได้ตามอิสระ ข้อสังเกตคือค่าความแม่นยำ α_c และ α_t ไม่ได้ถูกนำมาเฉลี่ยแบบมีน้ำหนักเหมือน α_p และ α_k เนื่องจากคะแนนความครอบคลุมทำนองและความตรงต่อเวลาเป็นสิ่งที่ทางโครงการเห็นว่าพื้นฐานต่อการร้องเพลงมาก หากล้มเหลวที่จะทำคะแนนมิติเหล่านี้ให้ได้มากพอ คนร้องก็ควรได้รับการหักคะแนนแบบทวีคูณ แทนแบบเฉลี่ยแบบมีน้ำหนัก เพราะทำให้คะแนนมีผลมากกว่ายิ่งยวด

2.2.9 แสดงคะแนนความแม่นยำในการร้องเพลงออกมาให้ผู้ใช้งาน เห็น โดยให้ความสำคัญกับคะแนนความแม่นยำผลลัพธ์ แล้วแยกเป็นแต่ละมิติสำหรับผู้ใช้งานที่สนใจ

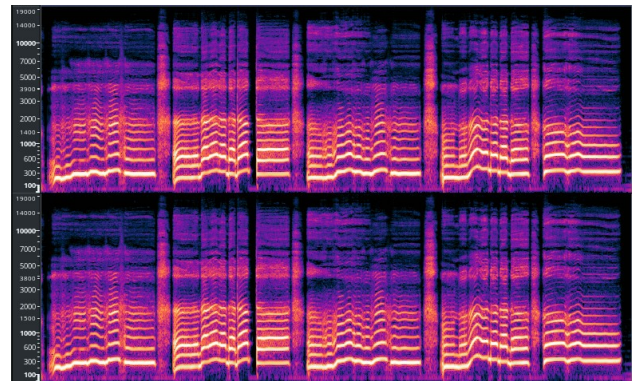
3. ผลการทดลอง

3.1 ข้อมูลนำเข้าและข้อมูลนำออก

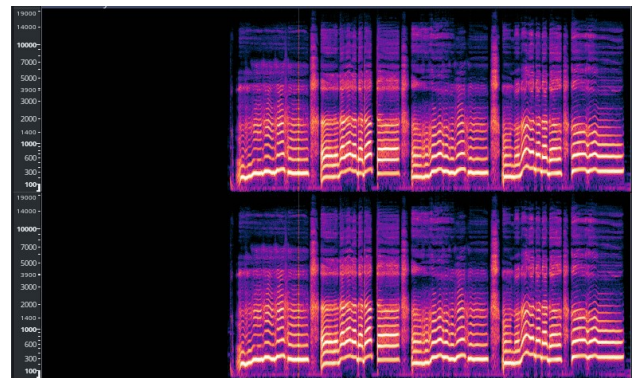
จากการทดลอง ได้นำเสียงสังเคราะห์ควบคุม 1 เสียง และเสียงบันทึกร้องเพลงของมนุษย์ 8 เสียงไว้ดังนี้



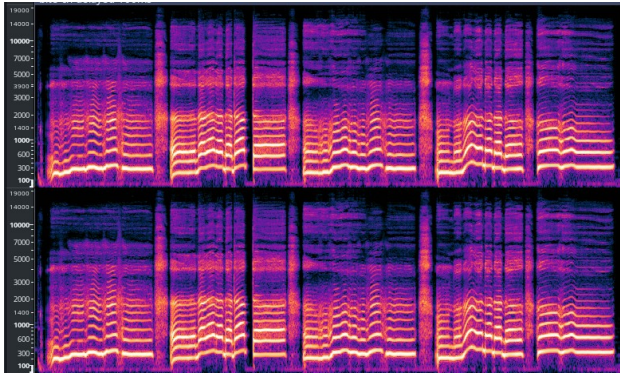
รูปที่ 4 สเปกโตรแกรมของเสียงสังเคราะห์คลื่นไซน์ความถี่ 100 เฮิรตซ์ ความยาว 4 วินาที



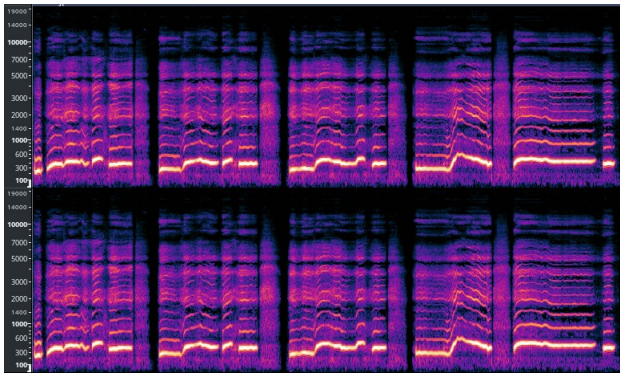
รูปที่ 5 สเปกโตรแกรมของการร้องเพลง *Sân-Z – BITE!* โดย 萧清源



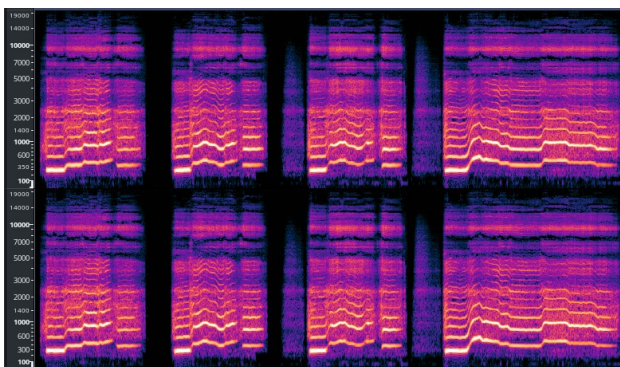
รูปที่ 6 สเปกโตรแกรมของการร้องเพลง *Sân-Z – BITE!* โดย 萧清源 พร้อมมีเวลาล่าช้าสังเคราะห์ 6.1 วินาที



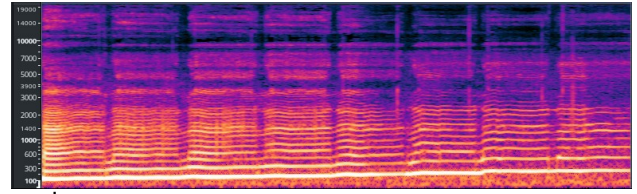
รูปที่ 7 สเปกโตรแกรมของการร้องเพลง *Sān-Z - BITE!* โดย 萧清源 พร้อมมีเวลาล่าช้าสังเคราะห์ 100 มิลลิวินาที



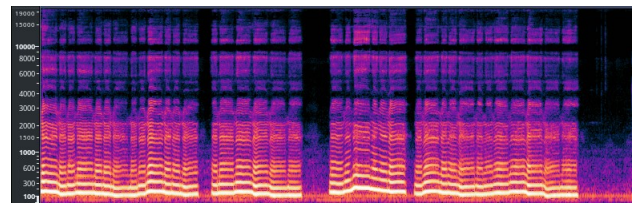
รูปที่ 8 สเปกโตรแกรมของการร้องเพลง *Sān-Z - BITE!* โดย 若山 詩音



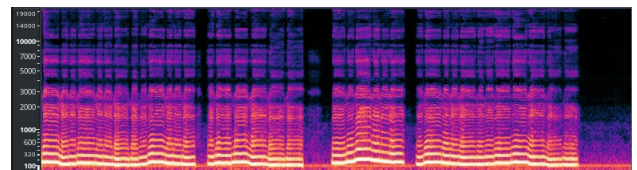
รูปที่ 9 สเปกโตรแกรมของการร้องเพลง *Sān-Z - BITE!* โดย 이지나



รูปที่ 10 สเปกโตรแกรมของการร้องทำนองบันไดกัญแจเสียงโดเมเจอร์จากโดกลางจนครบช่วงคู่แปด เมื่อแต่ละโน้ตถือค่าไว้ 500 มิลลิวินาที โดยผู้พัฒนาโครงการงาน



รูปที่ 11 สเปกโตรแกรมของการร้องเพลง *Tetris - Theme A* โดยผู้พัฒนาโครงการงาน



รูปที่ 12 สเปกโตรแกรมของการร้องเพลง *Tetris - Theme A* โดยผู้พัฒนาโครงการงานครั้งที่ 2

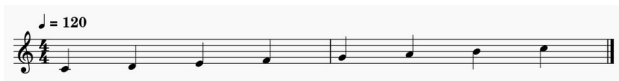
และการทดลองยังได้นำไฟล์ MIDI บันทึกทำนองเพลงไว้ดังนี้



รูปที่ 13 แผ่นถอดเสียงจากไฟล์ MIDI ของเพลง *Sān-Z - BITE!* เมื่อกัญแจเสียงคือโดชาร์ปเมเจอร์ และเทมโปคือ 165 จังหวะต่อนาที สำหรับการร้องเพลงโดย 萧清源



รูปที่ 14 แผ่นถอดเสียงจากไฟล์ MIDI ของเพลง *San-Z – BITE!* เมื่อถูกแปลงเสียงคือซอลเมเจอร์ และเทมโปคือ 150 จังหวะต่อนาที สำหรับการร้องเพลงโดย 若山 詩音 และ 이지나

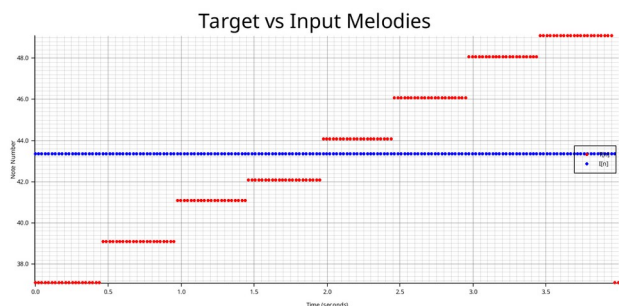


รูปที่ 15 แผ่นถอดเสียงจากไฟล์ MIDI ของทำนองบันได ถูกแปลงเสียงคือโดเมเจอร์จากโดกลางจนครบช่วงคู่แปด เมื่อเทมโปคือ 120 จังหวะต่อนาที

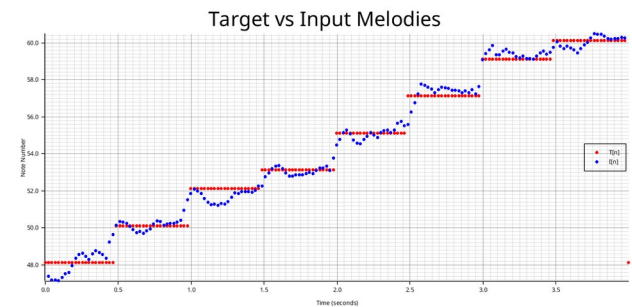


รูปที่ 16 แผ่นถอดเสียงจากไฟล์ MIDI ของเพลง *Tetris – Theme A* เมื่อถูกแปลงเสียงคือลาไมเนอร์ และเทมโปคือ 149 จังหวะต่อนาที

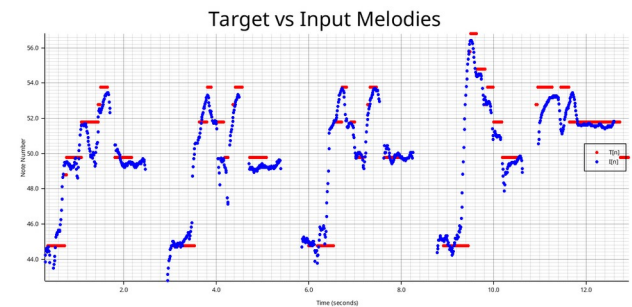
เมื่อนำไฟล์เหล่านี้มาจับคู่ให้สอดคล้องกัน แล้วนำมาประมวลผลตามการทดลอง ได้ความสัมพันธ์ระหว่างทำนองที่ดึงมาจากไฟล์ WAV (หลังจากผ่านการจัดตำแหน่งเวลาและความกว้างแล้ว) และทำนองของไฟล์ MIDI ได้ดังนี้



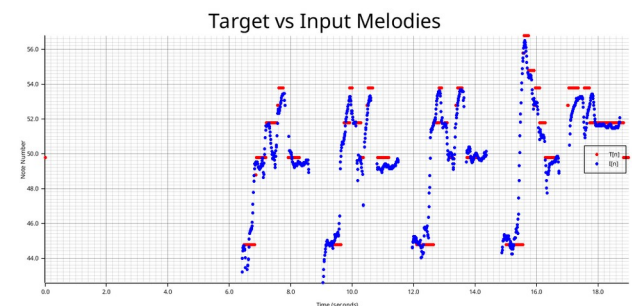
รูปที่ 17 กราฟเปรียบเทียบทำนองที่ดึงมาจากเสียงสังเคราะห์ คลื่นไซน์ความถี่ 100 เฮิรตซ์ ความยาว 4 วินาที และทำนองบันได ถูกแปลงเสียงคือโดเมเจอร์จากโดกลางจนครบช่วงคู่แปด เมื่อเทมโปคือ 120 จังหวะต่อนาที (*การทดลองที่ 1*)



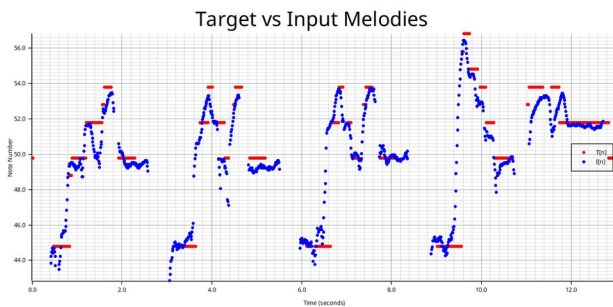
รูปที่ 18 กราฟเปรียบเทียบทำนองที่ดึงมาการร้องทำนองบันได ถูกแปลงเสียงคือโดเมเจอร์จากโดกลางจนครบช่วงคู่แปด เมื่อแต่ละโน้ตถือค่าไว้ 500 มิลลิวินาที โดยผู้พัฒนาโครงการ และทำนองบันไดถูกแปลงเสียงคือโดเมเจอร์จากโดกลางจนครบช่วงคู่แปด เมื่อเทมโปคือ 120 จังหวะต่อนาที (*การทดลองที่ 2*)



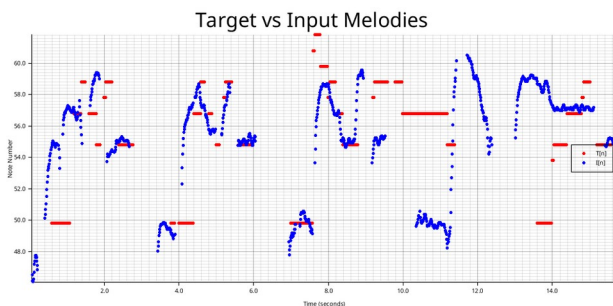
รูปที่ 19 กราฟเปรียบเทียบทำนองที่ดึงมาจากการร้องเพลง *San-Z – BITE!* โดย 萧清源 และทำนองเพลงเดียวกันเมื่อถูกแปลงเสียงคือโดชาร์ปเมเจอร์ และเทมโปคือ 165 จังหวะต่อนาที (*การทดลองที่ 3*)



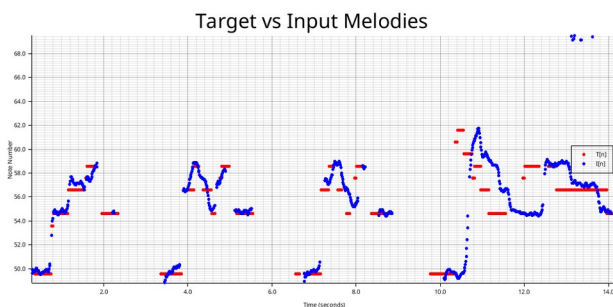
รูปที่ 20 กราฟเปรียบเทียบทำนองที่ดึงมาจากการร้องเพลง *Sân-Z – BITE!* โดย 萧清源 พร้อมมีเวลาล่าช้าสังเคราะห์ 6.1 วินาที และทำนองเพลงเดียวกันเมื่อถูกแยกเสียงคือโดชาร์ปเมเจอร์ และเทมโปคือ 165 จังหวะต่อนาที (การทดลองที่ 4)



รูปที่ 21 กราฟเปรียบเทียบทำนองที่ดึงมาจากการร้องเพลง *Sân-Z – BITE!* โดย 萧清源 พร้อมมีเวลาล่าช้าสังเคราะห์ 100 มิลลิวินาที และทำนองเพลงเดียวกันเมื่อถูกแยกเสียงคือโดชาร์ปเมเจอร์ และเทมโปคือ 165 จังหวะต่อนาที (การทดลองที่ 5)

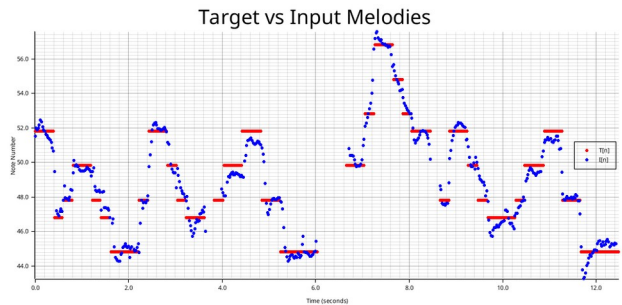


รูปที่ 22 กราฟเปรียบเทียบทำนองที่ดึงมาจากการร้องเพลง *Sân-Z – BITE!* โดย 若山 詩音 และทำนองเพลงเดียวกันเมื่อถูกแยกเสียงคือซอลเมเจอร์ และเทมโปคือ 150 จังหวะต่อนาที (การทดลองที่ 6)

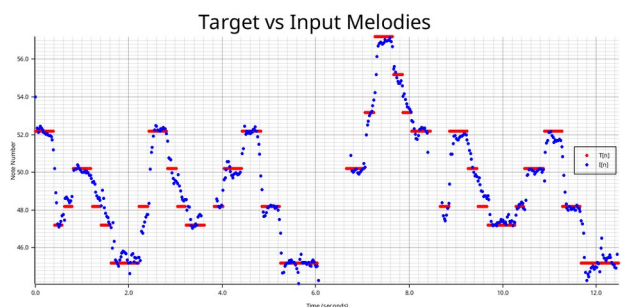


รูปที่ 23 กราฟเปรียบเทียบทำนองที่ดึงมาจากการร้องเพลง *Sân-Z – BITE!* โดย 이지나 และทำนองเพลงเดียวกันเมื่อถูกแยกเสียง

คือซอลเมเจอร์ และเทมโปคือ 150 จังหวะต่อนาที (การทดลองที่ 7)



รูปที่ 24 กราฟเปรียบเทียบทำนองที่ดึงมาจากการร้องเพลง *Tetris – Theme A* โดยผู้พัฒนาโครงการ และทำนองเพลงเดียวกันเมื่อถูกแยกเสียงคือลาไมเนอร์ และเทมโปคือ 149 จังหวะต่อวินาที (การทดลองที่ 8)



รูปที่ 25 กราฟเปรียบเทียบทำนองที่ดึงมาจากการร้องเพลง *Tetris – Theme A* โดยผู้พัฒนาโครงการครั้งที่สอง และทำนองเพลงเดียวกันเมื่อถูกแยกเสียงคือลาไมเนอร์ และเทมโปคือ 149 จังหวะต่อวินาที (การทดลองที่ 9)

และสุดท้ายแล้ว ได้คะแนนความแม่นยำของการทดลองแต่ละกรณีดังนี้

ตารางที่ 2 ความสัมพันธ์ระหว่างการทดลองแต่ละกรณีและคะแนนความแม่นยำที่ได้มาในแต่ละมิติ

# การทดลอง	α_c	α_t	α_p	α_k	α
1	99.81	100	8.32	42.67	16.87
2	99.61	100	98.65	100	98.60

3	98.46	100	71.86	56.02	66.86
4	99	0	76.23	57.02	0
5	98.47	28.33	71.67	56.02	18.91
6	96.31	0	32.8	100.0	0
7	96.07	100	70.3	87.24	71.67
8	99.03	100	88.3	100	90.35
9	99.03	100	99.18	100	98.42

3.2 การวัดประสิทธิภาพของวิธีการนำเสนอ

จากผลการทดลอง สามารถเปรียบเทียบค่าคะแนนความแม่นยำในแต่ละมิติที่ระบบที่พัฒนาขึ้นให้มา กับค่าคะแนนความแม่นยำในแต่ละมิติที่คาดหวังไว้ จากการประเมินผลเชิงความคิดเห็นของผู้พัฒนาโครงการได้ดังนี้

การทดลองที่ 1 เป็นการทดลองควบคุม ซึ่งควรจะได้คะแนนน้อยเนื่องจากการนำเสนอเสียงที่สังเคราะห์ขึ้นมาที่ไม่มี ความคล้ายกับทำนองที่ต้องการเปรียบเทียบ ปรากฏว่าได้คะแนน α_c และ α_t สูงใกล้เคียงสมบูรณ์ เป็นผลลัพธ์ที่มีเหตุผลเนื่องจากเป็น คลื่นเสียงต่อเนื่อง ทำให้ครอบคลุมทำนองทั่วถึง และเป็นคลื่น เสียงสม่ำเสมอ นั่นคือ เวลาเลื่อนที่ติตที่สุดสามารถเป็นได้ทุกค่า ซึ่ง ระบบคำนวณมาว่าควรใกล้เคียงศูนย์ ในขณะที่คะแนน α_p และ α_k ต่ำเนื่องจากเป็นคลื่นที่ระดับเสียงคงที่ ไม่ตรงกับทำนองอ้างอิง เป็นอย่างมาก เหมาะสมที่ได้ α ต่ำ

การทดลองที่ 2 8 และ 9 เป็นการทดลองที่เกิดขึ้นใน สภาพแวดล้อมควบคุม นั่นคือ ผู้พัฒนาโครงการได้บันทึกเสียง ร้องเพลงที่พยายามให้ถูกต้องตามทำนองอ้างอิงทุกโน้ต มีเสียง รบกวนต่ำถึงไม่มีเลย และเวลาเริ่มต้นทันที เพื่อที่จะพร้อมกับ จังหวะเริ่มต้นของทำนองอ้างอิง ทำให้ลดการหักคะแนนจากการ จัดตำแหน่งเวลาและความกว้าง ผลลัพธ์ที่ออกมาคือคะแนน ความแม่นยำในทุกมิติสูงใกล้เคียงสมบูรณ์ ตรงกับที่คาดหวังไว้ ข้อ สังเกตคือถึงแม้เสียงของผู้พัฒนาโครงการเป็นเสียงต่ำกว่าที่ ทำนองอ้างอิงหลายระดับ ขึ้นคู่เสียงเป็นพหุคูณของช่วงคู่แปด และระบบได้ชดเชยความแตกต่างไปอย่างถูกต้อง จากคะแนน α_k ที่สูง

การทดลองที่ 3 และ 7 เป็นการทดลองโดยใช้เสียงฮัมที่ไม่ ได้เกิดขึ้นในบริบทที่เคร่งครัด นั่นคือ เป็นการฮัมเชิงดนตรีแต่ไม่ ได้มีจุดหมายให้มีความแม่นยำตามโน้ตสูงเหมือนร้องเพลงจริง สามารถเห็นได้จากคะแนน α_p และ α_k ที่สูงแต่ไม่ใกล้เคียงสมบูรณ์ มาก แต่ก็ยังคงสูงกว่าการเปล่งเสียงของมนุษย์ที่ไม่เชิงดนตรี

การทดลองที่ 6 เป็นการทดลองโดยใช้เสียงฮัมเช่นเดียว กับการทดลองที่ 3 และ 7 แต่ในเสียงบันทึกมีการสะดุดและหยุด เปล่งเสียงในตรงกลาง ทำให้ได้คะแนน α_t น้อย และเนื่องจาก ระบบนำคะแนนมิตินี้มาคูณ ทำให้คะแนนผลลัพธ์ได้น้อยที่สุด คือ 0 สอดคล้องกับแนวคิดที่หากร้องเพลงติดขัด เสียงจังหวะ ควรจะ ถือว่าการร้องเพลงล้มเหลวและควรทำใหม่ตั้งแต่แรก

การทดลองที่ 4 และ 5 เป็นการทดลองโดยใช้เสียงบันทึก เดียวกับการทดลองครั้งที่ 3 แต่เพิ่มเวลาล่าช้าสังเคราะห์ ในการ ทดลองที่ 4 ล่าช้า 6.1 วินาที ซึ่งถือว่าล่าช้าที่ยาวด เหมาะสมที่จะ ได้รับคะแนนความตรงต่อเวลาเป็น 0 และในการทดลองที่ 3 ล่าช้า 0.1 วินาที ซึ่งมนุษย์ปกติรู้สึกว่าจะไม่ตรงจังหวะและขาด ความสุนทรีย์ทางดนตรี เหมาะสมที่จะถูกหักคะแนนความตรงต่อ เวลาเช่นเดียวกัน แต่ไม่ถูกหักอย่างมากที่สุด

โดยรวมแล้ว ระบบมีประสิทธิภาพในการคำนวณคะแนน ความแม่นยำที่ยุติธรรมได้อย่างมากพอประมาณ

4. สรุปผล

จากการทดลองการออกแบบระบบการให้คะแนนความ สามารถในการร้องเพลงถูกต้องตามโน้ตดนตรีอย่างยุติธรรม ซึ่ง ใช้ไฟล์ MIDI บันทึกข้อมูลทำนอง และไฟล์ WAV ในการบันทึก ข้อมูลเสียงร้องเพลง แล้วทำการหาความถี่มูลฐานของแต่ละ กรอบเวลา นำมาเปรียบเทียบกับหมายเลขโน้ตของทำนอง และ คำนวณเป็นค่าความแม่นยำแบบแยกแต่ละมิติในที่สุด ได้ผลการ ทดลองมาว่า โครงการสามารถคิดค้นแนวทางการคำนวณค่า ความแม่นยำขึ้นมาใหม่โดยอาศัยแนวคิดจากทฤษฎีดนตรี และได้ วิเคราะห์ว่าระบบสามารถให้คะแนนความแม่นยำในการร้องเพลง ได้แบบไม่เคร่งครัดหรือผ่อนปรนเกินไปเช่นเดียวกัน สอดคล้อง

กับจุดประสงค์ของโครงการ ซึ่งมุ่งพัฒนาระบบให้คะแนนการร้องเพลงที่ยุติธรรม และถือว่าดำเนินโครงการได้อย่างสำเร็จผล นอกจากนี้แล้ว ยังได้พัฒนาอินเทอร์เฟซผู้ใช้ที่เรียบง่ายและเบาต่อทรัพยากรของคอมพิวเตอร์ เพื่อความเข้าถึงได้สำหรับผู้ใช้งานทุกประเภท โดยอินเทอร์เฟซจะแสดงคะแนนความแม่นยำในแต่ละมิติออกมาอย่างชัดเจน รวมถึงคะแนนความแม่นยำผลลัพธ์

อย่างไรก็ตาม โครงการนี้สามารถต่อยอดและพัฒนาต่อได้ทั้งการปรับค่า Parameter ต่าง ๆ ซึ่งอาจจะได้ระบบที่ยุติธรรมมากกว่าเดิม และทางโครงการได้ขาดโอกาสการเปรียบเทียบระบบที่โครงการพัฒนากับระบบอื่น ๆ ที่มีอยู่ในโลกแห่งความเป็นจริง ซึ่งจะทำให้สามารถวัดประสิทธิภาพของวิธีการนำเสนอได้อย่างมีความน่าเชื่อถือยิ่งขึ้น

หากมีความประสงค์ในการทดสอบใช้งานระบบหรือพัฒนาต่อยอดโครงการนี้ สามารถเข้าถึงโค้ดแหล่งที่มาได้ที่ <https://github.com/krtchnt/cantometria>

5. เอกสารอ้างอิง

[1] Alain de Cheveigné and Hideki Kawahara, “YIN, a fundamental frequency estimator for speech and

music,” *The Journal of the Acoustical Society of America.*, Vol. 111(4), pp. 1917-1930, 2002.

[2] Matthew James Briggs. (27 February 2025). *Rust MIDI File Library Documentation*. [Online] Available : https://docs.rs/midi_file/latest/midi_file/

[3] Ruud van Asseldonk. (27 February 2025). *Hound Documentation*. [Online] Available : <https://docs.rs/hound/latest/hound/>

[4] Alessandro Genova. (27 February 2025). *Rust Pitch Detection Library Documentation*. [Online] Available : https://docs.rs/pitch-detection/latest/pitch_detection/

[5] Elliott Mahler. (27 February 2025). *RustFFT Documentation*. [Online] Available : <https://docs.rs/rustfft/latest/rustfft/>

[6] Hao Hou, Aaron Erhardt, Fabio A. Correa. (27 February 2025). *Plotters Documentation*. [Online] Available : <https://docs.rs/plotters/latest/plotters/>

[7] Josh McKinney, Leon Sautour, Orhun Parmaksız, Arijit Basu, Valentin271. (6 March 2025). *Ratatui Website*. [Online] Available : <https://ratatui.rs/>