

## ชื่อเรื่อง ชิงค์เสียงร้องเพลงให้พร้อมกันโดยอัตโนมัติ

โดย นางสาวมริสา สมสมัย

## บทคัดย่อ

โครงการนี้นำเสนอการออกแบบและการทำงานของระบบชิงค์เสียงร้องเพลงให้พร้อมกันโดยอัตโนมัติ ซึ่งเป็นระบบที่ช่วยในการชิงค์เสียงสองเสียงให้มีจังหวะและระยะเวลาเท่ากัน โดยใช้ต้นฉบับเสียงจากนักร้องหรือคนที่ร้องเพลงไพเราะ ถูกจังหวะ ให้เป็น Reference Voice และเสียงคนใช้งานที่อัดใหม่ให้เป็น User Voice ขั้นแรกใช้ MFCCs (Mel-Frequency Cepstral Coefficients) มาวิเคราะห์ข้อมูลเสียงเพื่อให้ได้ข้อมูลที่สำคัญทั้งของ Reference Voice และ User Voice จากนั้นใช้ DTW (Dynamic-time warping) จับคู่ความคล้ายคลึงของสองสัญญาณ ซึ่งได้ผลลัพธ์คือ path หรือ mapping function การจับคู่ของเฟรม ขั้นตอนต่อไปคือ ใช้ WORLD เพื่อแยกองค์ประกอบเสียงทั้งของ Reference Voice และ User Voice องค์ประกอบเสียงที่แยกได้คือ เวกเตอร์สเปกตรัม (sp) เวกเตอร์ความถี่/ไฮของเสียง (ap) และค่าความถี่ที่สั้นในลำคอ ( $f_0$ ) จากนั้นนำ mapping function ดัดแปลงองค์ประกอบเสียงของ User Voice ให้ได้จังหวะตรงกับ Referenece Voice หากต้องการการเปลี่ยนแปลง Pitch contour ก็ทำการดัดแปลง  $f_0$  จากนั้น นำองค์ประกอบเสียงใหม่ที่ถูกดัดแปลง มาประกอบกับเป็นเสียงร้องเพลงใหม่ การทดสอบการชิงค์เสียงแบ่งออกเป็น 2 ตอน คือ ตอนที่ 1 เป็นการทดสอบการทำงานของระบบชิงค์เสียงว่าจะสามารถชิงค์เสียง Reference Voice และเสียง User Voice ให้ได้ตรงกันหรือไม่ โดยแบ่งกรณีทดสอบเป็น 3 กรณี คือ ชิงค์เสียง User Voice ที่ร้อง ช้ากว่า เร็วกว่าและเท่ากันกับเสียง Reference Voice กรณีต่อมาคือการทดสอบร้องผิดเป็นบางคำและบางประโยค ตอนที่ 2 เป็นการทดลองดึงค่า  $f_0$  ของเสียง Reference Voice มาแทนค่า  $f_0$  ของเสียง User Voice เพื่อให้ผลลัพธ์ของการชิงค์เสียง User Voice ให้มีเสียงสูง/ต่ำตามเสียง Reference Voice ผลการทดสอบโดยรวมพบว่าระบบสามารถชิงค์เสียงร้องเพลงให้พร้อมกันได้และมีความไพเราะตามเสียงต้นฉบับ ซึ่งมีแนวโน้มที่จะนำไปสามารถนำมาประยุกต์ใช้ในอุตสาหกรรมดนตรี หรือนำไปประยุกต์ใช้ในด้านอื่นๆ ได้

**Thesis Title : Automatic Synchronization of Singing Voices**

By Miss. Marisa Somsamai

**ABSTRACT**

This project presents the design and operation of an automatic singing synchronization system. This is a system that helps to sync two sounds with equal rhythm and duration. Using the original voice from a singer or someone who sings beautiful, timely songs as a Reference Voice and a new user voice recorded as a User Voice. First, MFCCs (Mel-Frequency Cepstral Coefficients) were used to analyze the audio data to obtain information. Important for both Reference Voice and User Voice, then use DTW (Dynamic-time warping) to match the similarity of the two signals. This results in a path or mapping function to match the frames. The next step is to use WORLD to extract both the Reference Voice and User Voice audio components. Spectral vector ( $sp$ ), turbidity/clear vector ( $ap$ ) and throat oscillation frequency ( $f_0$ ). The mapping function was then used to alter the sound composition of the User Voice to match the Reference Voice if desired. Pitch contour was modified  $f_0$  and then applied new modified sound elements. accompanied by a new singing voice. The audio sync test is divided into 2 parts. Part 1 is to test the functionality of the audio sync system to see if the Reference Voice and User Voice can be synchronized or not. The test cases are divided into 3 cases, namely, the user voice that sings slower, faster, and the same as the Reference Voice. The next case is the test of wrong singing in some words and some sentences. Part 2 is an experiment to retrieve the reference sound  $f_0$ . Voice represents the  $f_0$  value of User Voice to provide the result of synchronization of User Voice with high/low tones based on Reference Voice. Overall test results showed that the system was able to synchronize singing with the original melody. which tends to be applied in the music industry or applied in other fields.