การอ่านริมฝีปากเป็นงานวิจัยที่จำเป็นสำหรับสถานการณ์ที่การสื่อสารด้วยเสียงไม่สามารถกระทำได้ ในงานวิจัยนี้จึงได้นำเสนอระบบการอ่านริมฝีปากแบบอัตโนมัติ จากคำพูดที่ได้กำหนดไว้ โดยใช้เพียงข้อมูลที่ได้เป็นภาพเคลื่อนไหวเท่านั้น ระบบนี้เริ่มต้นจากการตรวจจับใบหน้าคนในแต่ละเฟรมรูปภาพ โดยการใช้ตัวแบบของสีผิว และ การแบ่งกลุ่มด้วย FCM จากนั้นทำการตัดแยกเอาเฉพาะส่วนบริเวณที่เป็นริมฝีปากออกจากส่วนสีผิวอื่นๆ โดยการใช้การแปลงค่าสี และ การแบ่งกลุ่มด้วย FCM อีกครั้งหนึ่ง ค่าลักษณะเด่นสามารถหาได้จากบริเวณริมฝีปากที่ตัดแยกได้ ซึ่งจะใช้ในการระบุคำพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ ในงานวิจัยนี้ได้แบ่งการทดลองออกเป็นสองส่วนหลัก ได้แก่ การอ่านริมฝีปากด้วยคำ 5 คำ และการอ่านริมฝีปากด้วยคำ 10 คำ โดยในแต่ละการทดลองหลักจะถูกแบ่งออกเป็นการทดลองย่อย 3 การทดลอง การทดลองแรกได้แก่ การอ่านริมฝีปากจากผู้พูดเพียงคนเดียว การทดลองที่สองได้แก่ การอ่านริมฝีปากจากผู้พูดหลายคน และการทดลองสุดท้ายคือ การอ่านริมฝีปากจากผู้พูดที่ระบบไม่เคยเห็นมาก่อน (หรือที่เรียกว่าการทดสอบแบบบอด)

ผลการทดลองที่ได้ สำหรับการทดลองแรก ให้ความถูกต้องสูงสุดร้อยละ 92.00 สำหรับการอ่านริมฝีปากจากผู้พูดเพียงคนเดียว ให้ความถูกต้องสูงสุดร้อยละ 70.80 สำหรับการอ่านริมฝีปากจากผู้พูดหลายคน และให้ความถูกต้องสูงสุดร้อยละ 59.33 สำหรับการทดสอบแบบบอด และในการทดลองที่สอง ให้ความถูกต้องสูงสุดร้อยละ 77.00 สำหรับการอ่านริมฝีปากจากผู้พูดเพียงคนเดียว ให้ความถูกต้องสูงสุดร้อยละ 60.50 สำหรับการอ่านริมฝีปากจากผู้พูดหลายคน และให้ความถูกต้องสูงสุดร้อยละ 22.33 สำหรับการทดสอบแบบบอด จากผลการทดลองแสดงให้เห็นว่าระบบการอ่านริมฝีปากแบบอัตโนมัตินี้ใช้งานได้ดีกับการอ่านริมฝีปากจากผู้พูดเพียงคนเดียว และให้ผลที่ยอมรับได้กับการอ่านริมฝีปากจากผู้พูดหลายคน แต่ไม่เหมาะที่จะใช้งานกับผู้พูดที่ระบบไม่เคยเห็นมาก่อน เนื่องจาก การขยับปากของผู้พูดแต่ละคนค่อนข้างต่างกันแม้ว่าจะพูดคำเดียวกันก็ตาม

Lip reading is an important research in some situations which the audio can not be used. This work proposes the automatic lip reading system which recognizes the assigned words using only video of human lips (without any audio part). This system starts from the face detection in each frame of image sequences by using skin color models and FCM clustering. Then lip regions are separated by using color transformation and FCM clustering again. Static and dynamic features are extracted from lip regions. These features are used in HMMs to recognize the assigned words. In this work, the experiment is divided into two main parts. For the first part, this system is used for classifying the five assigned words. And for the second part, this system is used for classifying the ten assigned words. In each main part of the experiment, it is divided into three sub sections. For the first sub section this system is tested with single speaker (The system is trained and tested by one speaker.). For the second sub section, this system is used by grouping the data from all speakers and dividing them into the training data and the testing data. And for the last experiment, this system is trained by four speakers and tested by another speaker (Blind test).

The results of the first main experiment achieve 92.00% maximum accuracy for the single speaker, 70.80% accuracy for multi speakers and 59.33% maximum accuracy for the blind tests. The results of the second main experiment achieve 77.00% maximum accuracy for the single speaker, 60.50% accuracy for multi speakers and 22.33% maximum accuracy for the blind tests. All recognition rates show that the system works great for using with single speaker and acceptable for using with multi-speakers. But this system gives the unsatisfying results when it is used with unknown speakers because the movements of the lips from the different speakers are quite distinct although they speak the same words.