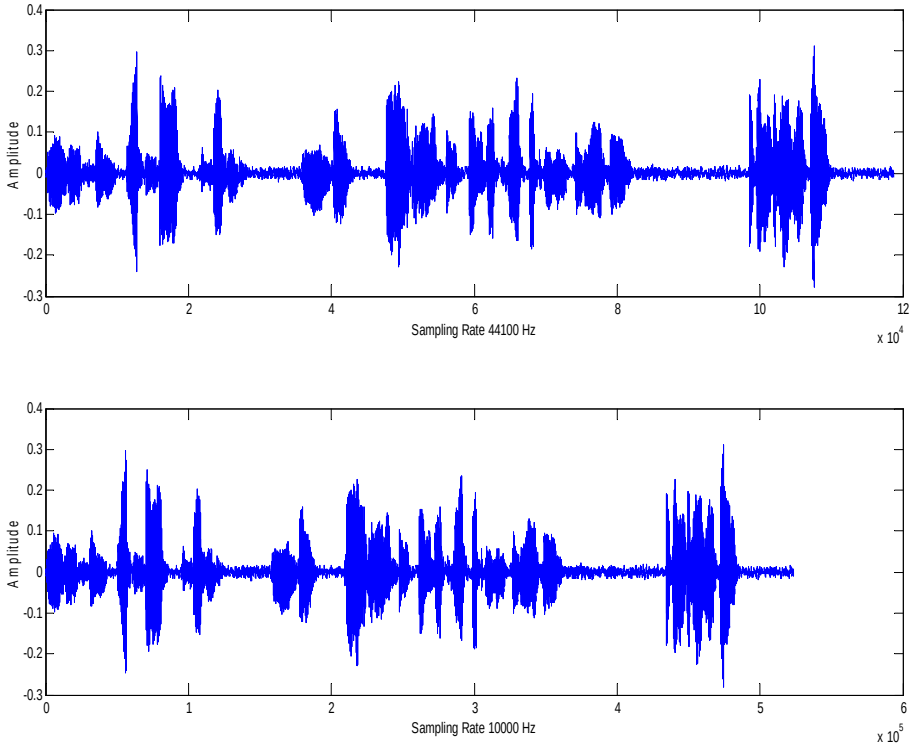
บทที่ 4 ผลการทดลอง/วิจัย

ในบทนี้จะกล่าว ถึงการทดลองและผลของการทดลองการวิเคราะห์ ML ของเสียงพูดระหว่างคนปกติและผู้ป่วยโรคซึมเศร้าด้วย MFCC โดยมีขั้นตอนดังต่อไปนี้

4.1 การทดลองสัญญาณเสียง

ที่เก็บบันทึกนั้นมีอัตราการสุ่มตัวอย่างเท่ากับ 44100 เฮิรตซ์แต่ในวิทยานิพนธ์ใช้อัตราการสุ่มตัวอย่างที่ 10000 เฮิรตซ์ จึงต้องทำการลดอัตราการสุ่มตัวอย่างลงมาให้ได้ตามต้องการจะช่วยในการวิเคราะห์ผลได้เร็วมากยิ่งขึ้น เนื่องจากข้อมูลที่เก็บมานั้นเป็นเสียงพูดจากผู้ป่วยอาการซึมเศร้าและเป็นเสียงของมนุษย์ซึ่งเสียงที่ได้นั้นไม่เกิน 5000 เฮิรตซ์ อยู่แล้วฉะนั้นอัตราการสุ่มตัวอย่างที่ 10000 เฮิรตซ์ ก็จะเหมาะสมที่สุดในการวิเคราะห์ครั้งนี้

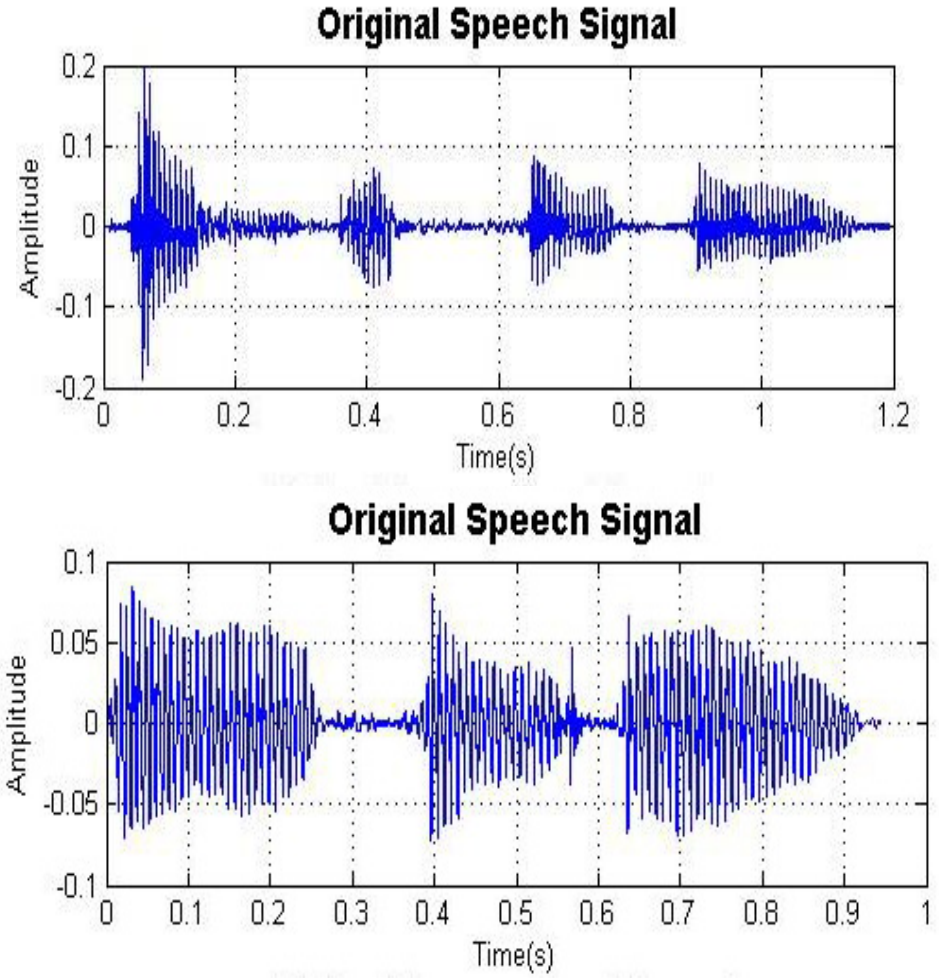


รูปที่ 4.1 แสดงการลดอัตราการสุ่มตัวอย่างจาก 44100 Hz ลงมาเป็น 10000 Hz

จากรูปที่ 4.1 จะเห็นว่าจำนวนข้อมูลที่มีอัตราการสุ่มตัวอย่างเท่ากับ 44100 เฮิรตซ์ จะมีจำนวนข้อมูลที่มากกว่า อัตราการสุ่มตัวอย่างเท่ากับ 10000 เฮิรตซ์ มากฉะนั้นจึงมีความจำเป็นต้องลดอัตราการสุ่มตัวอย่างลงมาให้เหมาะสมและช่วยในการประมวลผลได้เร็วมากยิ่งขึ้นเพราะมีจำนวนข้อมูลน้อยลงแต่ก็ต้องใช้วิธีการอื่นอีกเพื่อให้ใช้เวลาในการประมวลผลให้น้อยกว่านี้ซึ่งจะกล่าวในขั้นตอนต่อๆไป

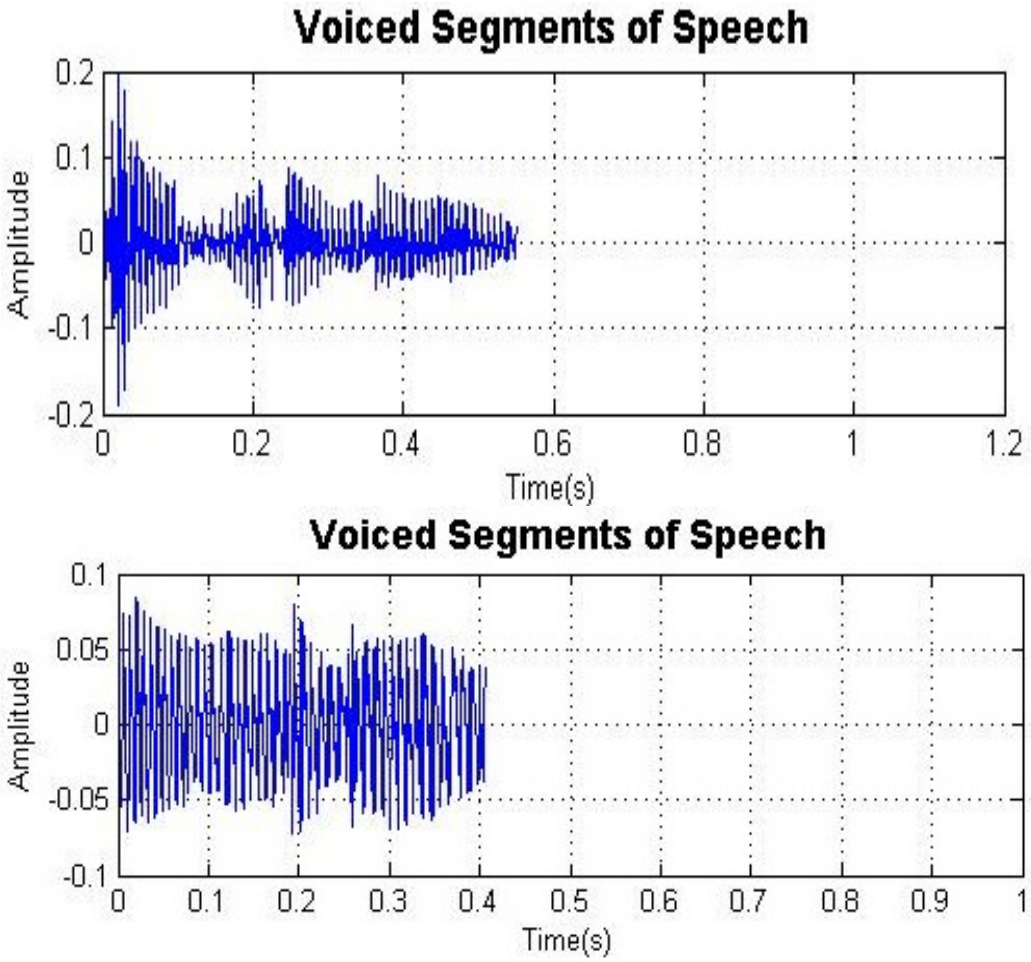
4.2 การประมวลผลสัญญาณเสียงเบื้องต้น

นำสัญญาณเสียงที่ได้มาผ่านกระบวนการประมวลผลสัญญาณเสียงเบื้องต้นคือ การปรับสัญญาณสู่แกนศูนย์ กรรมวิธีการเน้นล่วงหน้า การวางกรอบสัญญาณหน้าต่าง ดังรูปที่ 4.2 โดยใช้โปรแกรม MATLAB

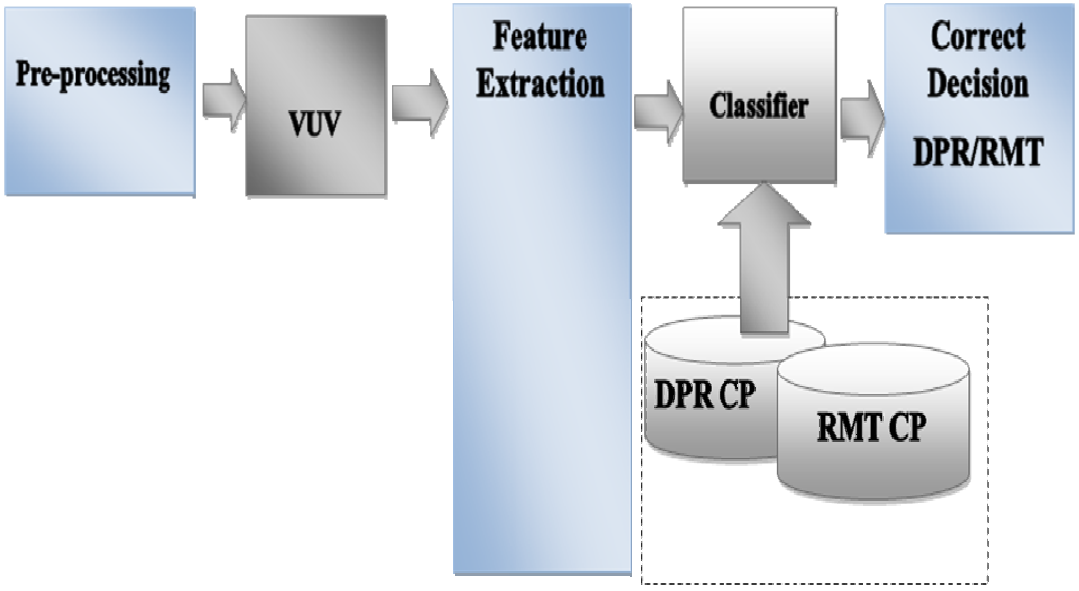


รูปที่ 4.2 สัญญาณเสียงที่บันทึกโดยใช้โปรแกรม MATLAB

จากนั้นนำสัญญาณเสียงที่ได้จากรูปที่ 4.2 มาทำการหาจุดสิ้นสุดของเสียงพูดโดยวิธีการตัดหัวท้ายคำโดยใช้ค่าพลังงาน จะเป็นการตัดเฉพาะส่วนที่เป็นส่วนของเสียงพูดและตัดส่วนที่ไม่เป็นเสียงพูด เสียงเงียบออก แสดงดังรูปที่ 4.3



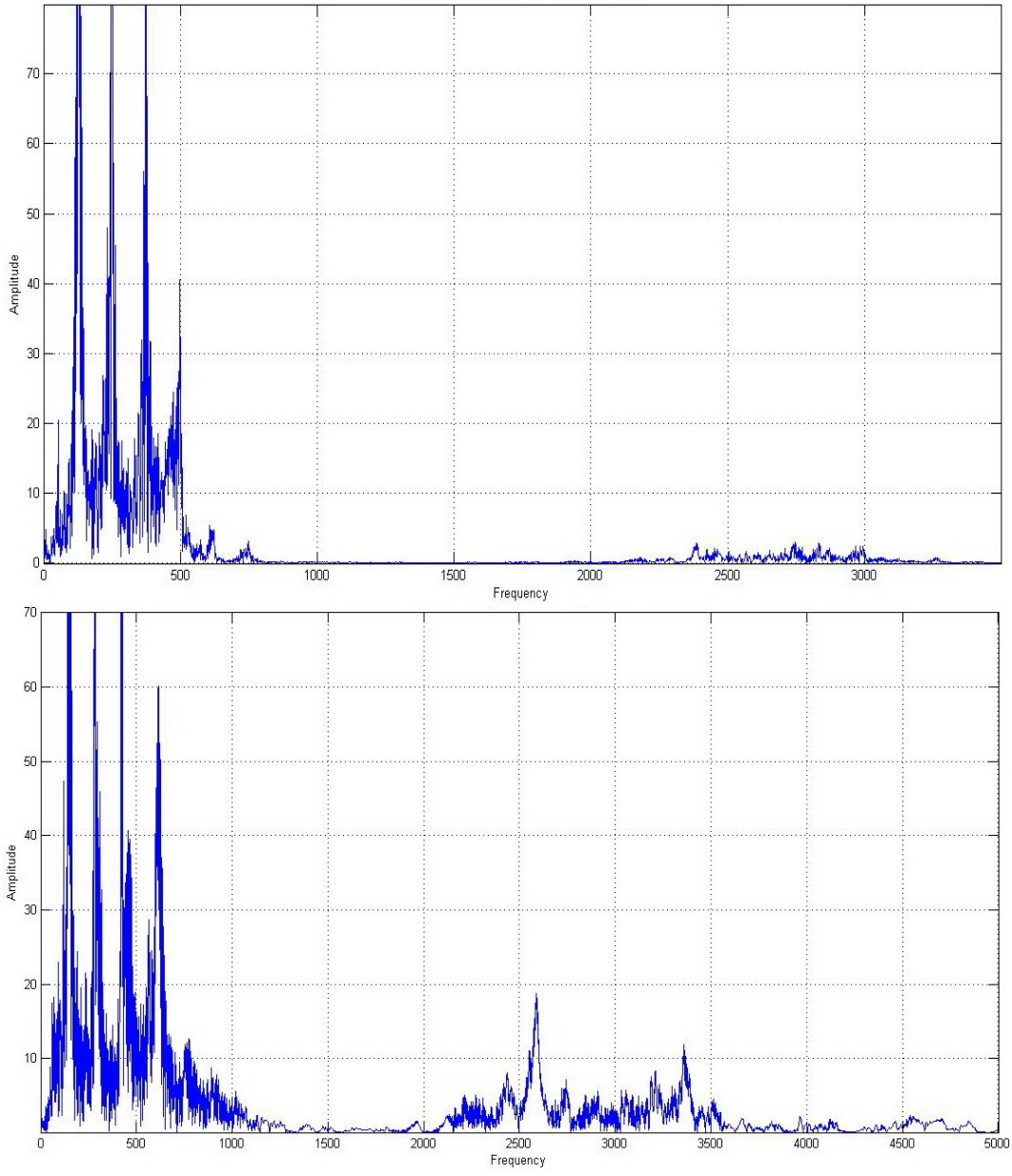
รูปที่ 4.3 สัญญาณเสียงที่ตัดส่วนที่ไม่เป็นเสียงพูดและเสียงเงียบออก



รูปที่ 4.4 แสดง Workflow สำหรับการคัดแยกตัวอย่างเสียงพูด

4.3 การสกัดค่าลักษณะสำคัญ (Feature Extraction)

นำสัญญาณเสียงที่ได้มาหาค่าสัมประสิทธิ์ MFCC จำนวน 16 ตัว โดยการแปลงสัญญาณโดยใช้สมการฟูริเยร์แบบไม่ต่อเนื่อง (DFT) เพื่อเปลี่ยนสัญญาณเสียงจากโดเมนของเวลาให้อยู่ในโดเมนความถี่ ดังรูปที่ 4.5

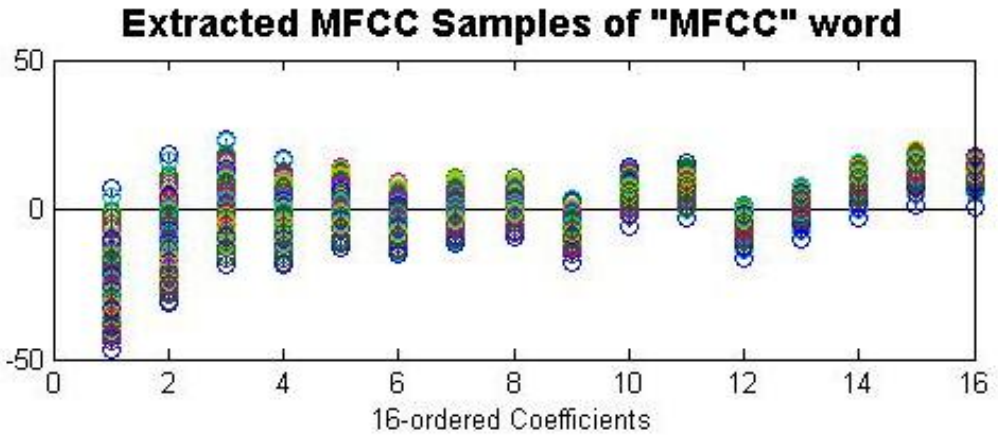


รูปที่ 4.5 แสดงสัญญาณเสียงในโดเมนของความถี่

นำสัญญาณเสียงในโดเมนของความถี่มาส่งผ่านชุดตัวกรองแบบสามเหลี่ยม โดยใช้ชุดตัวกรองแบบสามเหลี่ยมจำนวน 16 สัมประสิทธิ์ เพื่อใช้หาค่า MFCC ในขั้นต่อไป

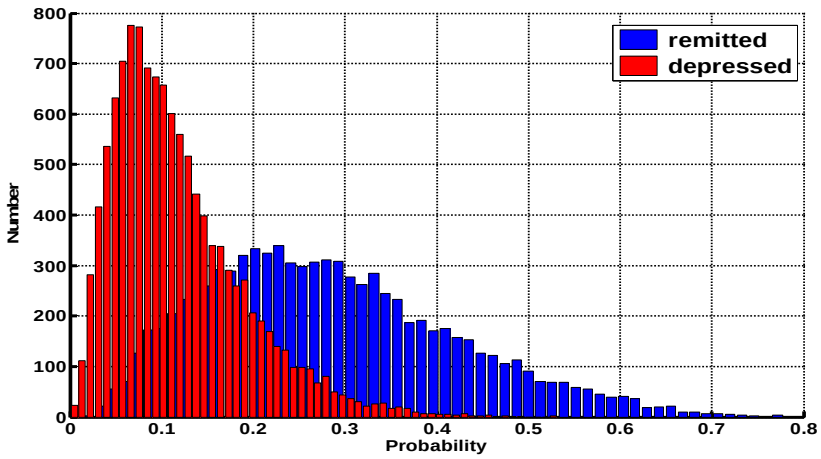
4.4 การสกัดค่าลักษณะสำคัญจากสัญญาณเสียงเพื่อให้ได้ค่า MFCC

หลังจากหาขนาดกำลังสองและผ่านการกรองด้วยชุดตัวกรองแบบสามเหลี่ยมแล้ว จะเข้าสู่วิธีการแปลงโคไซน์แบบไม่ต่อเนื่อง (DCT) จะได้ค่าสัมประสิทธิ์ MFCC จำนวน 16 สัมประสิทธิ์ ดังรูปที่ 4.6



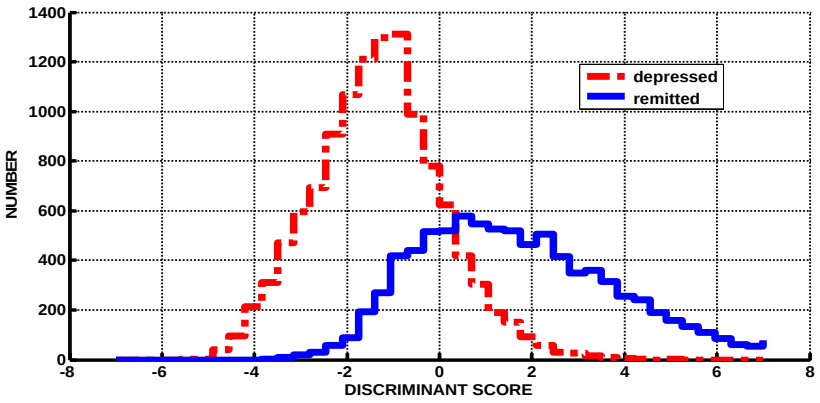
รูปที่ 4.6 ค่าสัมประสิทธิ์ MFCC

ค่าเฉลี่ย, ค่าความแปรปรวนและความน่าจะเป็นที่สกัดได้ ที่ได้รับจากการปรับรูปแบบ Gaussian Mixer กับ Vocal-tract cepstra ที่ใช้ในทั้งสองกลุ่มคือ กลุ่ม Training และกลุ่ม Testing คุณสมบัติของตัวอย่างที่จะถูกทดสอบครั้งแรกสำหรับการกระจายตัวอย่างระหว่างการพูดทั้งสองกลุ่มแจ้งให้ทราบอย่างชัดเจนจากการเปรียบเทียบการพล็อตกราฟ histograms



รูปที่ 4.7 การกระจายความน่าจะเป็นระหว่าง 2 กลุ่ม

รูปที่ 4.7 แสดง Histograms ของความน่าจะเป็นส่วนผสมที่สกัดเดิมที่เกี่ยวข้องกับการเลือก Gaussians ที่โดดเด่นเห็นได้ชัดเจน สำหรับการแยกที่สังเกตเห็นได้อย่างมีนัยสำคัญระหว่างเสียงพูดทั้งสองกลุ่ม นอกจากนี้ในการคำนวณ Score จำแนกจากการแจกแจงระหว่างกลุ่มโดยมีเงื่อนไขเช่นกัน เป็นผลจากการแปลงข้อมูลที่แตกต่างอย่างมีนัยสำคัญในกลุ่ม ค่าเฉลี่ยจะถูกกำหนดอย่างเห็นได้ชัดสำหรับแยกระยะห่างซึ่งแสดงถึงความเป็นไปได้สำหรับเราในการได้รับการจำแนกที่มีประสิทธิภาพระหว่างกลุ่ม



รูปที่ 4.8 การจำแนก Score ระหว่างกลุ่ม Remitted (สีฟ้า) และกลุ่มตัวอย่าง Depressed (เส้นประ-สีแดง)

รูปที่ 4.8 แสดง histograms ที่มีรูปทรงในระยะของ score จำแนกตามการใช้ฟังก์ชั่นพื้นฐานการจำแนกของ Fisher's ที่มีความแปรปรวนร่วมระหว่างความแปรปรวนกลุ่มเมทริกซ์ [13] เพื่อให้มีความซับซ้อนน้อยกว่าในการคำนวณของฟังก์ชันจำแนกและยังมีประสิทธิภาพสูงพอสมควรในรูปแบบคุณลักษณะประชากรที่หลากหลายที่เป็นสมมติฐานเหมาะสำหรับการศึกษาข้อมูลจึงเป็นการประมาณที่เป็นกลางของความแปรปรวนร่วมของประชากรจะถูกกำหนดสำหรับระยะห่างยกกำลังสองตัวอย่างในสูตร (D^2) ซึ่งถูกอ้างถึงก่อนในส่วนความสามารถของการดึงของคำพูดทุกตัวอย่างที่ถูกประมวลผลบนพื้นฐานของกรอบเสียงพูดความยาว 51.2 ms. ดังนั้นระยะห่างของรอยขีดที่ปรากฏในรูปที่ 3 ถูกคำนวณจากเฟรมเสียงพูดแบ่งจากฐานข้อมูลของสัญญาณเสียงพูดทั้งในการจัดหมวดหมู่ประเภทกลุ่มเสียงพูดกับที่ระบุที่ซ่อนสำหรับจำนวนของเฟรมตัวอย่างที่ใช้ในการอธิบายของแต่ละบุคคล การแจกแจงความน่าจะเป็นของน้ำหนักคำพูดที่สกัดจากทั้งสองกลุ่มนำมาพล็อตในการจำแนกรอยขีดเป็นตัวชี้วัดค่าเฉลี่ยที่แตกต่างกันอย่างมีนัยสำคัญ ตัวอย่างระยะห่างยกกำลังสอง (D^2) ระหว่างค่าเฉลี่ยสองกลุ่มตัวอย่างสามารถทำได้ตามสมการ 4.1

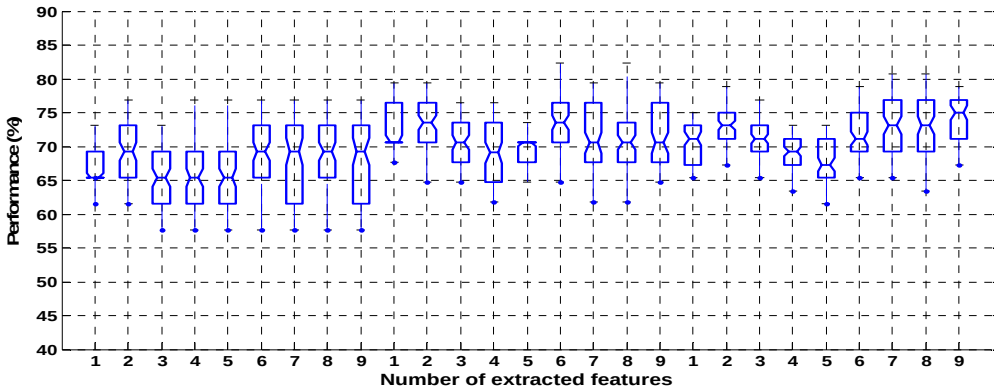
$$D^2 = (\bar{X}_{RMT} - \bar{X}_{DPR})' S^{-1} (\bar{X}_{RMT} - \bar{X}_{DPR}) \qquad (4.1)$$

$\bar{X}_{DPR}$ คือ ค่าเฉลี่ยกลุ่มตัวอย่างที่คำนวณจากการสกัดจาก Cepstral ของเสียงพูดของผู้ป่วยโรคซึมเศร้า

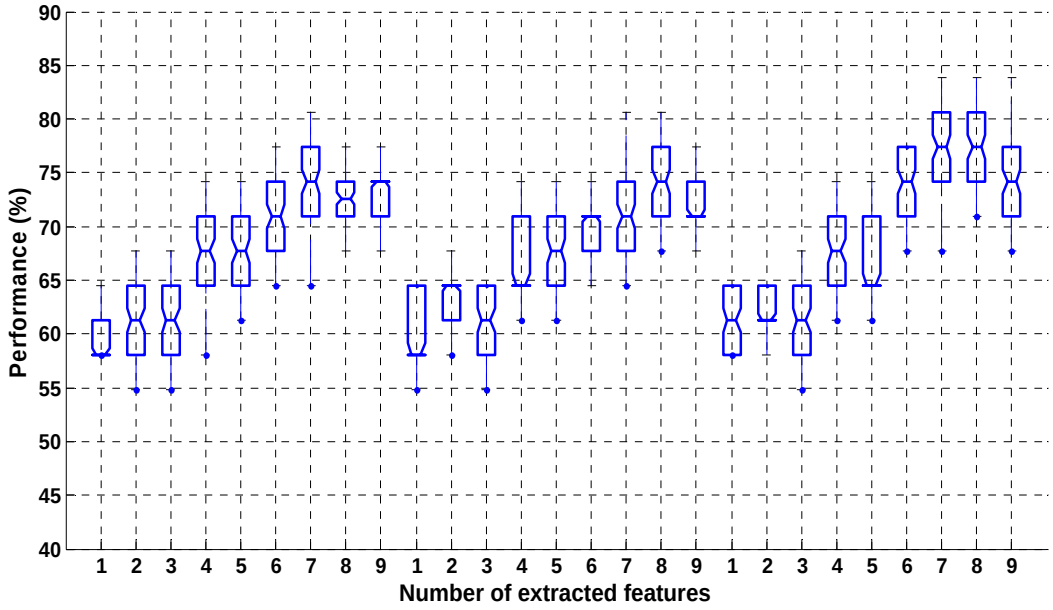
$\bar{X}_{RMT}$ คือ ค่าเฉลี่ยกลุ่มตัวอย่างของผู้ป่วยโรคซึมเศร้าที่ได้รับการรักษาจนหายเป็นปกติ

S คือ เมทริกซ์ความแปรปรวนร่วมของตัวอย่างรวมกันระหว่างสองเมทริกซ์ความแปรปรวนร่วมที่คำนวณได้จากกลุ่มตัวอย่างทั้งสองกลุ่มที่สกัดจาก Cepstral.

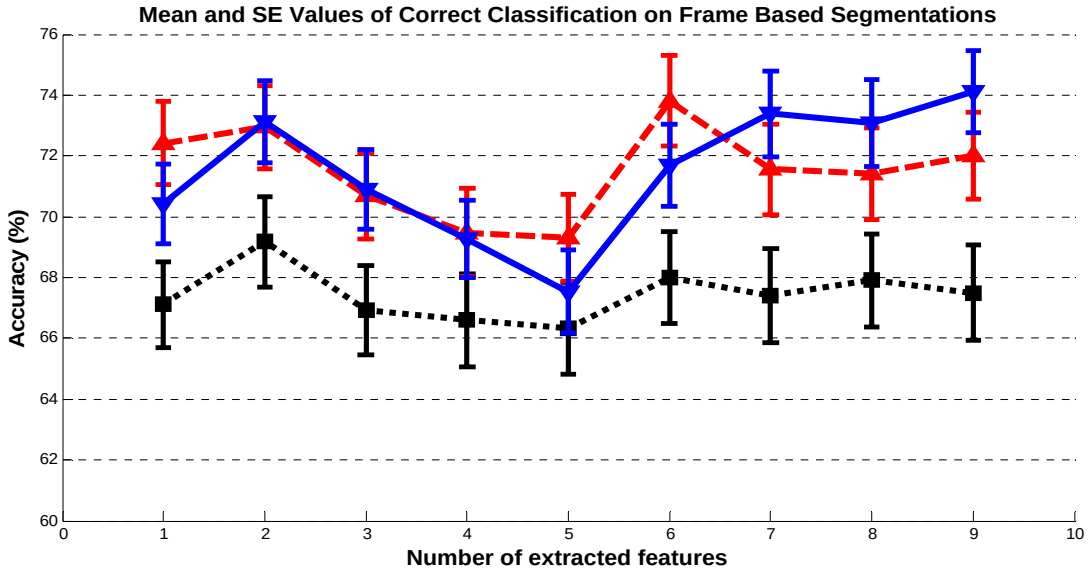
การพล็อตแสดงการเปรียบเทียบโดย Classifier ที่ใช้ได้ทั้งหมด การประเมินผลที่เพิ่มขึ้นสำหรับทุกกรณีการแบ่งส่วนคำพูดเป็น 3 ส่วนกับจำนวนของ features ที่เพิ่มขึ้นใน cross-validation เมื่อตัวอย่าง subject-based ถูกแบ่งแยกไว้แล้ว เราพบว่าจำนวน features ทั้งหมด 9 ชนิดที่คัดเลือกด้วย f-ratio มีการจัดอันดับทำให้ประสิทธิภาพการคัดแยกค่อยๆลดลง ผลจากการคัดแยกของตัวอย่างที่มีความยาวของเฟรมที่แตกต่างกัน (พล็อตในรูปที่ 4.10) แสดงให้เห็นว่าประมาณ 6 features ของ Cepstral ในการรวมกันสามารถแยกความแตกต่าง ระหว่าง Classifier กับความหลากหลายของความยาวของเฟรม แต่แนวโน้มความคล้ายคลึงกันมากขึ้นในการรวมประสิทธิภาพการทำงานที่สามารถสังเกตเห็นการรวมที่สูงขึ้นของ features ในกรณีของการคัดแยกตัวอย่าง frame-based ดูเหมือนว่าผลลัพธ์จะไม่มีความสอดคล้องกันและที่แตกต่างกันน้อยลงอย่างมีนัยสำคัญทั้ง 3 มีการแบ่งออกเป็นส่วนที่แตกต่างกันสำหรับ features เพิ่มเติมที่เกิดขึ้นในรูปแบบการรวมกันในกรณีของการคัดแยกตัวอย่างคำพูดที่มีความยาว 200 เฟรม ของตัวอย่างที่ระบุผลการดำเนินงานจะต้องได้ score เฉลี่ยสูงสุดของ 78% ในขณะที่การทดสอบ subject-based ที่ขึ้นอยู่กับตัวอย่างการพูดโดยอัตโนมัติจะดำเนินการกับ 8 features ในการรวมกันใส่เข้าไปใน Classifier เพื่อศึกษาเปรียบเทียบข้อมูลเพิ่มเติมเกี่ยวกับการพูดอัตโนมัติเผยให้เห็นว่าเมื่อเราใช้อย่างใดอย่างหนึ่งตัวอย่างที่ใช้เฟรมหรือตัวอย่าง Subject-based (Subject ที่แสดงโดยค่าเฉลี่ยของเฟรมทั้งหมดของ feature ที่เก็บรวบรวมจาก subject) ในการทดสอบสถานะของ Classifier cross-validation, การแสดงที่เพิ่มขึ้นและมีการสังเกต outperforming เห็นได้ชัดสำหรับแบ่งประเภทของการพูดกับกลุ่มตัวอย่าง 200 ตัวอย่าง/เฟรม เมื่อเทียบกับกรณีอื่น ๆ ของอัตราส่วน ตัวอย่าง/เฟรม ที่แตกต่างกันที่จำนวนที่สูงขึ้นของการรวม features



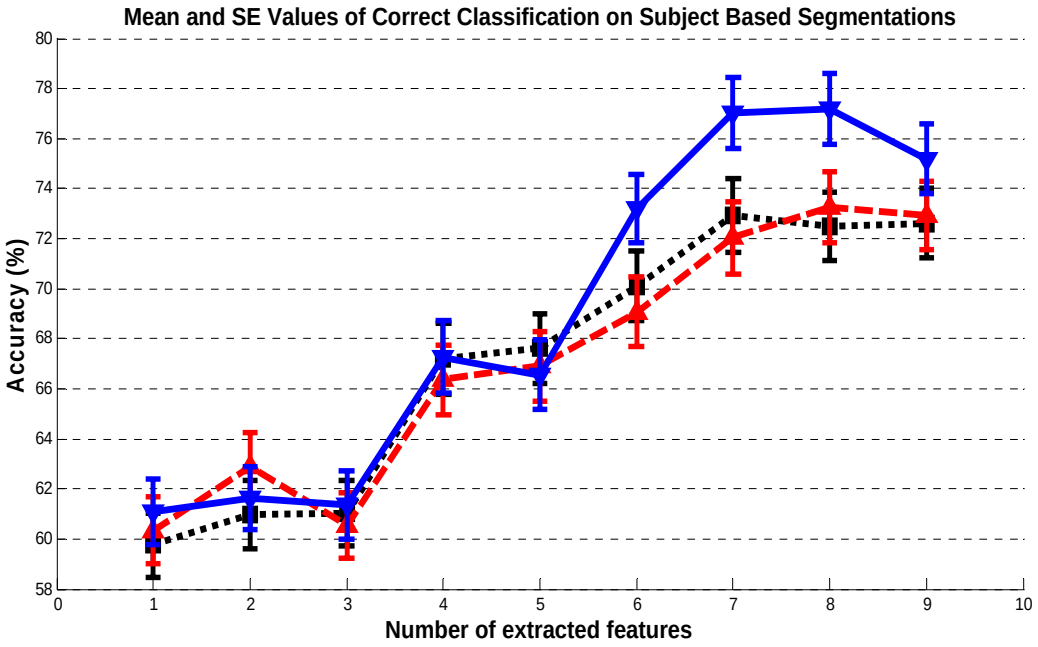
รูปที่ 4.9 ผลลัพธ์ที่ได้จากการจำแนกตามเฟรมตัวอย่างการพูดกับกลุ่มตัวอย่าง 400, 300 และ 200 / เฟรม



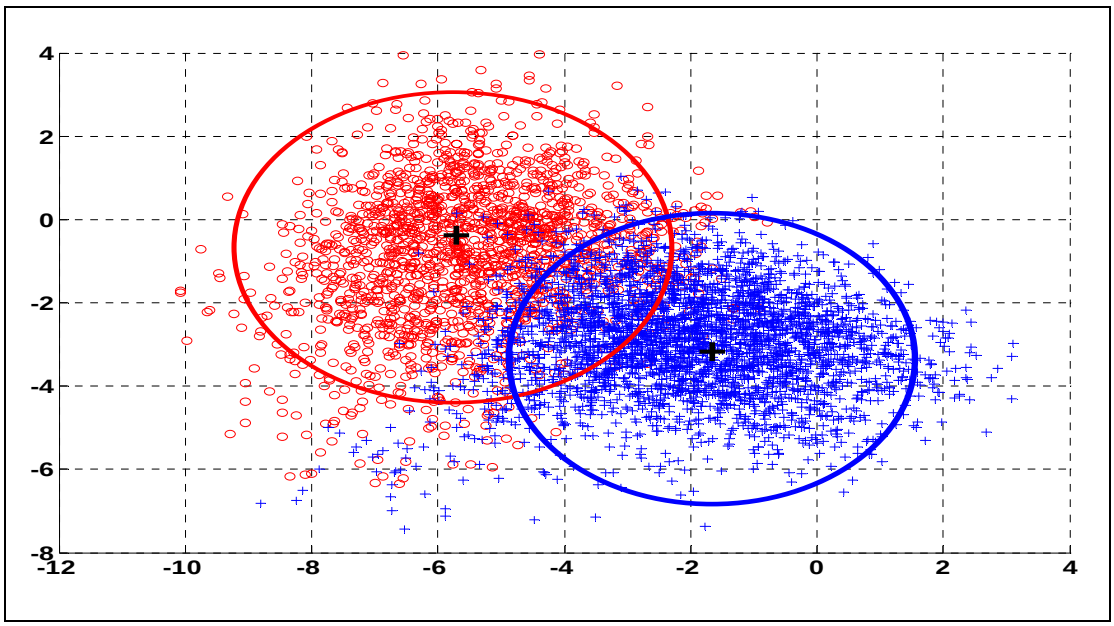
รูปที่ 4.10 ผลการแบ่งประเภทตัวอย่างเสียงพูดของฐานข้อมูลผู้ป่วยกับกลุ่มตัวอย่าง 400, 300 และ 200 / เฟรม



รูปที่ 4.11 การเปรียบเทียบความถูกต้องในค่าเฉลี่ยและค่า Standard Error จากการจำแนก segmentations ตามเฟรมที่มีความยาว 400 (จุด), 300 (เส้นประ) และ 200 (เส้นต่อเนื่อง)



รูปที่ 4.12 การเปรียบเทียบความถูกต้องในค่าเฉลี่ยและค่า Standard Error จากการจำแนก segmentations ตามเฟรม ที่มีความยาว 400 (จุด), 300 (เส้นประ) และ 200 (เส้นต่อเนื่อง)



รูปที่ 4.13 การกระจาย MFCC ของเสียงที่สกัดได้ระหว่างกลุ่ม Remitted [สีฟ้า (+)] และตัวอย่างการพูดของผู้ป่วยซึมเศร้า [สีแดง (o)]

รูปที่ 4.11 และ 4.12 แสดงให้เห็นถึงแนวโน้มของค่าเฉลี่ยและค่ามาตรฐานความผิดพลาด (S.E.) ที่ได้รับจากการทดสอบการคัดแยกที่มี 3 segmentations ตัวอย่างที่แตกต่างกับจำนวนที่เพิ่มขึ้นของการรวม features ในการคัดแยก การแสดงผลได้มาจากการคัดแยกชนิดของพารามิเตอร์ค่าเฉลี่ยและค่า SD จะแสดงในจุดสีดำ,เส้นประสีแดงและเส้นทึบสีฟ้าเมื่อใช้ 400, 300 และ 200 ตัวอย่าง/เฟรม ตามลำดับ ในการประมาณค่า ค่าเฉลี่ยและค่า SD ที่ป้อนข้อมูลสำหรับการคัดแยก ในการทดลองที่ผิดพลาดขนาดเล็กมาตรฐาน (S.E.) กำหนดเมื่อค่าเฉลี่ยของ score การคัดแยกให้ถูกต้องทั้งหมด จำนวนที่มีขนาดเล็กลงสามารถบ่งบอกถึงผลการคัดแยกโดยรวมในความหมายของสถิติที่เชื่อถือได้เกี่ยวกับการที่ช่วงความเชื่อมั่นสูงขึ้นของการคัดแยก ผลลัพธ์ทั้งหมดเหล่านี้ในที่สุดนำเราไปสู่ข้อสรุปที่ว่าลักษณะทางเดินเสียงมีการตอบสนองต่อสื่อโดยสภาวะซึมเศร้าที่แสดงอยู่ในค่าสัมประสิทธิ์ Cepstral จะสามารถประเมินและใช้เพื่อตรวจสอบอาการที่ผ่านผลเสียงขึ้นอยู่กับความสอดคล้องกันทางสถิติที่เปิดเผยเกี่ยวกับผลการวิเคราะห์ตามที่กล่าวไว้แล้ว .