ชื่อ-สกุลผู้เขียนบทความ

การศึกษาวิจัยและพัฒนา

การจัดทำบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันหรือใกล้เคียงกับเวลาออกอากาศ

Research and Development of a Near Real-time Captioning Service

ชื่อ-สกุลผู้เขียนบทความ ๑: นายโยธิน สิทธิบดีกุล
หน่วยงานที่สังกัด: องค์การกระจายเสียงและแพร่ภาพสาธารณะแห่งประเทศไทย
หมายเลขโทรศัพท์: 086-032-9809
Email: yothins@thaipbs.or.th
ที่อยู่ปัจจุบัน: 145 ถนนวิภาวดีรังสิต แขวงตลาดบางเขน เขตหลักสี่ กรุงเทพ 10210

ชื่อ-สกุลผู้เขียนบทความ ๒: นายณัฐนันท์ ทัดพิทักษ์กุล
หน่วยงานที่สังกัด: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
หมายเลขโทรศัพท์: 02-564-6900 ต่อ 72232
E-mail: nattanun.thatphithakkul@nectec.or.th
ที่อยู่ปัจจุบัน: 112 ถนนพหลโยธิน ตำบลคลองหนึ่ง อำเภอคลองหลวง จังหวัดปทุมธานี 12120

ชื่อ-สกุลผู้เขียนบทความ ๓: นางสาวอนันต์ลดา โชติมงคล
หน่วยงานที่สังกัด: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
หมายเลขโทรศัพท์: 02-564-6900 ต่อ 72715
E-mail: ananlada.chotimongkol@nectec.or.th
ที่อยู่ปัจจุบัน: 112 ถนนพหลโยธิน ตำบลคลองหนึ่ง อำเภอคลองหลวง จังหวัดปทุมธานี 12120

ชื่อ-สกุลผู้เขียนบทความ ๔: นายสุชาติ อรุณนิธิกุล
หน่วยงานที่สังกัด: องค์การกระจายเสียงและแพร่ภาพสาธารณะแห่งประเทศไทย
หมายเลขโทรศัพท์: 081-983-9964
Email: sucharta@thaipbs.or.th

ชื่อ-สกุลผู้เขียนบทความ ๕: นางกมลพรรณ พันพึ่ง
หน่วยงานที่สังกัด: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
หมายเลขโทรศัพท์: 02-564-6900 ต่อ 72042
E-mail: kamolpun.punpuing@nectec.or.th
ที่อยู่ปัจจุบัน: 112 ถนนพหลโยธิน ตำบลคลองหนึ่ง อำเภอคลองหลวง จังหวัดปทุมธานี 12120

ชื่อ-สกุลผู้เขียนบทความ ๖: นางสาวสุภาพันธุ์ เกตุคำ
หน่วยงานที่สังกัด: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
หมายเลขโทรศัพท์: 02-564-6900 ต่อ 72043
E-mail: suphaphan.katekum@nectec.or.th
ที่อยู่ปัจจุบัน: 112 ถนนพหลโยธิน ตำบลคลองหนึ่ง อำเภอคลองหลวง จังหวัดปทุมธานี 12120

การศึกษาวิจัยและพัฒนา

การจัดทำบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันหรือใกล้เคียงกับเวลาออกอากาศ

Research and Development of a Near Real-time Captioning Service

บทคัดย่อ

การจัดทำบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวหรือใกล้เคียงกับเวลาออกอากาศ เพื่อทดสอบออกอากาศโทรทัศน์แบบเสมือนจริง มีโครงสร้างของระบบออกอากาศดังนี้ ระบบถอดความเสียงแบบทันต่อเวลาเลือกใช้เทคนิคการถอดความโดยใช้ผู้ถอดความแบบแบ่งพิมพ์ซึ่งพัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (เนคเทค) ส่วนอุปกรณ์ที่ต่อเชื่อมกับระบบฮาร์ดแวร์และเครือข่ายที่มีอยู่เดิมของสถานีโทรทัศน์ ThaiPBS เลือกใช้อุปกรณ์ VT3 รุ่น Peach ร่วมกับซอฟต์แวร์ PIMmala subtitle gateway ในการทดสอบทางเทคนิคและทดสอบการยอมรับจากผู้ใช้งานที่เป็นคนพิการทางการได้ยิน จากผลการทดสอบเชิงเทคนิค 5 ครั้ง และทดสอบกลุ่มผู้ใช้งานคนพิการทางการได้ยินซึ่งเป็นกลุ่มเป้าหมายหลัก 2 ครั้ง พบว่าโครงสร้างระบบของสถานีโทรทัศน์ ThaiPBS ที่มีการปรับปรุงให้รองรับบริการคำบรรยายแทนเสียงฯ สามารถออกอากาศได้จริงในเชิงเทคนิค ด้านความถูกต้องของการถอดความรายการข่าวมีค่าความถูกต้องเฉลี่ยอยู่ที่ 80-90% จากการทดสอบกับผู้ใช้ที่เป็นคนพิการทางการได้ยินพบว่าผู้ใช้ยอมรับความถูกต้องที่ระดับนี้ได้ ด้านความทันต่อเวลาไม่สามารถแสดงข้อความที่ละตัวอักษรได้ เนื่องจากข้อจำกัดด้าน Bandwidth ที่จัดสรรไว้ให้กับบริการนี้อยู่ที่ 100 kpbs ทำให้ต้องแสดงทุก 500 msec และขนาดภาพข้อความเท่ากับ SD และในด้าน delay ของการถอดความเสียงพูดในการออกอากาศอยู่ที่ 6.6 วินาที

Abstract

To broadcasting a live television program with a real-time closed captioning service, two sub-systems must be added to the existing broadcasting system: 1) a real-time speech transcribing module and 2) a live stream text receiving module. In this study, the technique called simulated-typing developed by National Electronics and Computer Technology Center (NECTEC) is used to produce real-time transcription for Thai. The PIMmala subtitle gateway and VT3 PEACH DVB-sub generator together are then used to produce closed captioning service from a live text stream in DVB-T2 format. From five technical experiments and two user experiments, it has been shown that the television broadcasting systems developed under this study can produce and broadcast a real-time closed captioning service that is acceptable by people with hearing disability which is the targeted group of this study. The captioning

accuracies of live news programs are around 80-90%. In terms of delay, the caption appears on a television screen within 6.6 seconds after the corresponding speech. Due to bandwidth limitation, which the regulator allocated only 100 kpbs for the closed caption service, the caption has to be displayed in block of characters every 500 milliseconds which increases the delay.

คำสำคัญ

คำบรรยายแทนเสียงแบบปิด, เวลาจริง, ระบบออกอากาศ, โทรทัศน์ระบบดิจิทัล, อัตราบิต

Keywords

Closed caption, CC, real-time, broadcasting system, digital television, bit rate

ความเป็นมา

สำนักงานคณะกรรมการกิจการกระจายเสียง กิจการโทรทัศน์ และกิจการโทรคมนาคมแห่งชาติ (กสทช.) มุ่งเน้นให้เกิดการเข้าถึงหรือรับรู้ข้อมูลข่าวสารอย่างทั่วถึงและเสมอภาค และการจัดบริการตามประกาศ กสทช. เรื่อง การส่งเสริมและคุ้มครองสิทธิของคนพิการให้เข้าถึงหรือรับรู้และใช้ประโยชน์จากรายการของกิจการโทรทัศน์ พ.ศ. 2559 ข้อ 5 กำหนดหลักเกณฑ์ให้ผู้รับใบอนุญาตมีการจัดบริการล่ามภาษามือ (sign language) คำบรรยายแทนเสียง (captioning) เสียงบรรยายภาพ (audio description) หรืออื่นใด โดยระบุให้จัดบริการให้เหมาะสมกับรูปแบบของแต่ละรายการ โดยต้องมีสัดส่วนเรื่องบริการคำบรรยายแทนเสียงอย่างน้อยร้อยละ 40 ของสัดส่วนรายการที่เป็นข่าวสารหรือสาระที่เป็นประโยชน์ต่อสาธารณะ และต้องจัดให้มีเพิ่มขึ้นจนถึงร้อยละ 60 ภายในระยะเวลาที่กำหนด รวมเวลาตลอดทั้งวันเฉลี่ยแล้วต้องไม่น้อยกว่า 180 นาทีต่อวัน

ทั้งนี้ สำนักงาน กสทช. เห็นว่าประเทศไทยยังไม่เคยมีการจัดบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกัน หรือใกล้เคียงกับเวลาออกอากาศ (near real-time captioning) ซึ่งคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ หมายถึง คำบรรยายแทนเสียงของรายการโทรทัศน์ที่ผู้ผลิตจัดทำคำบรรยายแทนเสียงในขณะเดียวกันกับที่รายการกำลังออกอากาศ โดยคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ จะใช้กับรายการโทรทัศน์ที่เป็นรายการสด เช่น รายการเหตุการณ์สด รายการข่าว รายการกีฬา เป็นต้น คำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ มีความแตกต่างรวมทั้งมีความซับซ้อนเชิงเทคนิคมากกว่าคำบรรยายแทนเสียงที่ทำแบบบันทึกล่วงหน้า เนื่องจากข้อความบรรยายเสียงแบบใกล้เคียงเวลาจริง จะอยู่ในรูปแบบสายข้อมูลตัวอักษร (text stream) ที่มีการผลิตอย่างต่อเนื่องเพื่อให้ทันต่อเวลาจริงมากที่สุด ต่างจากข้อความบรรยายเสียงของรายการที่บันทึกเทปไว้ล่วงหน้า ซึ่งจะอยู่ในรูปแบบของข้อมูลที่ระบุเวลา (timecode) คือเป็นประโยค หรือวลี ที่เตรียมไว้ล่วงหน้าและ

กำกับเวลาที่จะถูกแสดงผลไว้เรียบร้อยแล้ว ดังนั้นจึงต้องมีการศึกษาและปรับปรุงระบบออกอากาศ ทั้งในส่วนของฮาร์ดแวร์ และซอฟต์แวร์เพื่อให้รับรองข้อมูลนำเข้าแบบสายข้อมูลตัวอักษร และสามารถแสดงผลคำบรรยายแทนเสียงบนจอโทรศัพท์ได้ตามมาตรฐานโทรทัศน์ระบบดิจิทัล

นอกจากนี้ ยังไม่มีการกำหนดรายละเอียด วิธีการ หรือมาตรฐานในการออกอากาศในกิจการโทรทัศน์มาก่อน สำนักงาน กสทช. จึงร่วมมือกับสถานีโทรทัศน์ไทยพีบีเอส จัดทำโครงการ "ศึกษาวิจัยและพัฒนาการจัดทำบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ" โดยมีการทบทวนความรู้และประเด็นที่เกี่ยวข้องในการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ และเทคนิค หรือมาตรฐานเทคโนโลยี ในกิจการโทรทัศน์ของต่างประเทศ 3 ประเทศ ได้แก่ สหรัฐอเมริกา สหราชอาณาจักร และญี่ปุ่น เพื่อศึกษาถึงประสบการณ์ที่มีการดำเนินงานด้านการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ มาก่อนหน้าแล้วในต่างประเทศ ซึ่งอาจนำมาประยุกต์ ออกแบบ เพื่อให้เหมาะสมสำหรับการจัดทำระบบบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ในประเทศไทย

อย่างไรก็ตาม ข้อกำหนดหรือแนวทางการจัดทำคำบรรยายแทนเสียงในโลกยังไม่มีการกำหนดหรือตกลงให้มีมาตรฐานกลางที่ใช้สำหรับทุกประเทศ มีปรากฏเพียงการตกลงของบางกลุ่มประเทศที่ใช้แนวทางร่วมกัน เช่น ประเทศในสหภาพยุโรปส่วนใหญ่จะอ้างอิงแนวทางของบรรษัทแพร่ภาพกระจายเสียงอังกฤษ (British Broadcasting Corporation--BBC) ในประเทศสหราชอาณาจักรเหมือนกัน ซึ่งโดยทั่วไปภายในแต่ละประเทศจะมีการตกลงหลักการกลางขั้นต่ำที่หน่วยงานผลิตคำบรรยายแทนเสียงจะใช้เป็นแนวทางร่วมกัน หลังจากนั้นหน่วยงานผลิตแต่ละแห่งอาจจัดทำข้อกำหนดรายละเอียดของหน่วยงานขึ้นใช้เอง ซึ่งอาจแตกต่างกันไปตามบริบทของสังคมและวัฒนธรรม หรือเทคโนโลยีที่ใช้

สำหรับแนวทางการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ในประเทศต่าง ๆ นั้น ไม่มีระบุรายละเอียดในการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ มากเท่ากับคำบรรยายแทนเสียงแบบบันทึกล่วงหน้า โดยมีการระบุถึงหลักการสำคัญเฉพาะที่เกี่ยวข้อง ได้แก่ 1) การกำหนดความแม่นยำสมบูรณ์ของเนื้อหา (accuracy) ซึ่งจะกำหนดระดับไว้ต่ำกว่าคำบรรยายแทนเสียงแบบบันทึกล่วงหน้า 2) การประสานเวลา (synchronicity) กำหนดระยะเวลาความต่างของช่วงเวลาที่เกิดขึ้นระหว่างเสียงที่เกิดขึ้นของรายการกับคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ที่แสดงบนหน้าจอโทรทัศน์ 3) การจัดทำบริการให้ครบเวลาสมบูรณ์ของทั้งรายการ (completeness) 4) การจัดวางตำแหน่ง (placement) ที่ปรากฏบนหน้าจอโทรทัศน์ โดยที่ข้อความที่แสดงไม่บดบังส่วนสำคัญของภาพ กำหนดลักษณะการแสดงผลตัวอักษรบนจอภาพให้ปรากฏเลื่อนไปอย่างต่อเนื่อง ไม่ควรเกิน 2 บรรทัด โดยอาจใช้การเปลี่ยนสีหรือสัญลักษณ์บางอย่างที่จำเป็นเท่านั้นเพื่อระบุตัวผู้พูด โดยมีเป้าหมายของบริการเพื่อให้ผู้ชมรายการโทรทัศน์สามารถเข้าใจเนื้อหาได้

การทบทวนความรู้ดังกล่าวทำให้ทราบถึงประเด็นข้อพิจารณาสำคัญของการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ว่ามีลักษณะเฉพาะและข้อพึงระวังต่างกันระหว่างการจัดทำคำบรรยายแทนเสียงในช่วงเวลา

เดียวกันฯ กับการจัดทำคำบรรยายแทนเสียงแบบบันทึกล่วงหน้า เนื่องจากธรรมชาติของการจัดทำคำบรรยายแทนเสียงแบบบันทึกล่วงหน้ามีข้อจำกัดเรื่องเวลาน้อยกว่า เอื้อให้มีการตรวจสอบความถูกต้องของเนื้อหา การประสานเวลาได้มากกว่า ส่งผลให้มีการกำหนดความแม่นยำสมบูรณ์ของเนื้อหาถึงระดับร้อยละ 99 ขณะที่การจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ มีการกำหนดความแม่นยำสมบูรณ์ของเนื้อหาในระดับที่ต่ำกว่า

นอกจากนี้ การจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ มีขั้นตอนกระบวนการจัดทำที่ซับซ้อนเกี่ยวข้องกับหลายปัจจัยมากกว่าการจัดทำคำบรรยายแทนเสียงแบบบันทึกล่วงหน้า เช่น ความคลาดเคลื่อนของการประสานเวลาที่เกิดขึ้น อาจเกี่ยวข้องกับ เทคนิควิธีการจัดทำคำบรรยายแทนเสียงว่าผลิตโดยใช้คนพิมพ์ หรือระบบรู้จำเสียงพูด ความสามารถในการพิมพ์ที่รวดเร็วและถูกต้องของคนพิมพ์ การมีข้อมูลล่วงหน้าเกี่ยวกับชื่อคนและสถานที่ ความชัดเจนของเสียงพูดและความเร็วในการพูดของผู้พูด จำนวนคนพูดในช่วงเวลาเดียวกัน ประสิทธิภาพของอุปกรณ์เชื่อมต่อการสื่อสาร ประสิทธิภาพการประมวลผลของอุปกรณ์ฮาร์ดแวร์ ความชัดเจนและความสม่ำเสมอของสัญญาณสื่อสาร

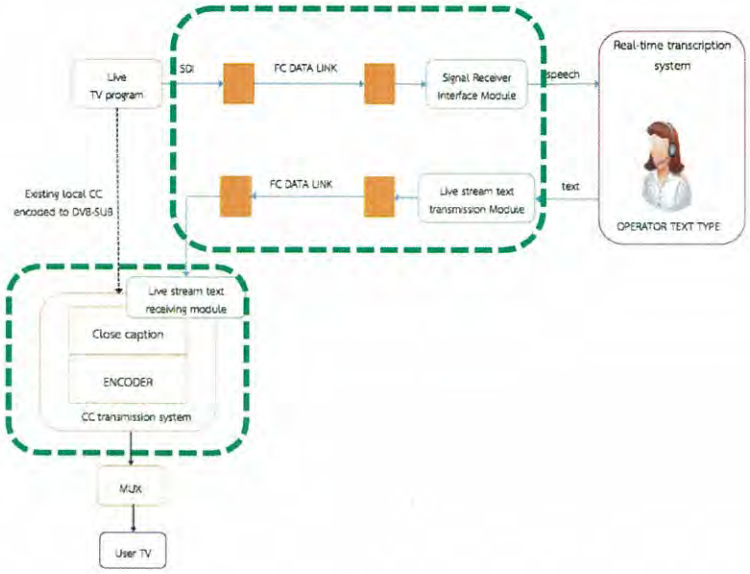
ดังนั้น สำนักงาน กสทช. และสถานีโทรทัศน์ไทยพีบีเอส จึงนำข้อมูลที่ได้จากการทบทวนความรู้ข้างต้นมาประกอบการพิจารณา ในการพัฒนาต้นแบบ อันจะส่งผลให้เกิดการกระตุ้นการจัดบริการคำบรรยายแทนเสียงในประเทศไทยให้เกิดขึ้นอย่างเป็นรูปแบบ รวมถึงจัดทำข้อเสนอแนะทั้งเชิงนโยบายและเชิงเทคนิค เพื่อใช้เป็นแนวทางในการจัดบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ที่เหมาะสมกับประเทศไทยมากที่สุด เพื่อส่งเสริมให้คนพิการ ผู้สูงอายุ และผู้ด้อยโอกาส สามารถเข้าถึงข้อมูลข่าวสารผ่านกิจการโทรทัศน์ได้อย่างเสมอภาคกับบุคคลทั่วไปในอนาคต

แนวทางการศึกษาวิจัย

การวิจัยเพื่อพัฒนาระบบบริการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันหรือใกล้เคียงกับเวลาออกอากาศ จะแบ่งออกเป็น 2 ส่วนคือ 1) ส่วนของระบบออกอากาศของสถานีโทรทัศน์ และ 2) ส่วนของระบบถอดความเสียงแบบใกล้เคียงเวลาจริง

ในส่วนของระบบออกอากาศ การศึกษาของโครงการศึกษาวิจัยฯ เป็นการพัฒนาต่อยอดจากระบบฮาร์ดแวร์และเครือข่ายที่มีอยู่เดิมของสถานีโทรทัศน์ไทยพีบีเอส โดยจะปรับปรุงหรือแก้ไขระบบอุปกรณ์เฉพาะในส่วนการออกอากาศระบบโทรทัศน์ดิจิทัล ให้สามารถบริการคำบรรยายแทนเสียงแบบใกล้เคียงเวลาจริงได้ โดยส่วนที่ต้องปรับปรุงคือ Module ที่ใช้จัดทำคำบรรยายแทนเสียงแบบปิดที่มีอยู่เดิม ซึ่งใช้กับรายการที่มีการบันทึกไว้ล่วงหน้าและมีคำบรรยายแทนเสียงที่ได้เตรียมไว้ล่วงหน้าในฐานข้อมูลแล้ว การจะให้บริการคำบรรยายแทนเสียงแบบใกล้เคียงเวลาจริงได้ จะต้องพัฒนามอดูลนี้ให้สามารถรับข้อความที่ได้จากการถอดข้อความ ที่อยู่ในรูปแบบของ Real-time text stream โดยจะต้องมีการปรับปรุงระบบส่งสัญญาณคำบรรยายแทนเสียงแบบปิด (closed captions transmission system) ซึ่งมีหน้าที่สร้างคำบรรยายแทนเสียงตามมาตรฐาน DVB-Subtitle โดยเพิ่ม

ชุดอุปกรณ์ต่อพ่วงในการรับข้อมูลข้อความ (text) แบบถ่ายทอดสด (live stream หรือ live stream text receiving module) และต้องมีการพัฒนาส่วนต่อเชื่อม (interface module) ระหว่างระบบถอดความเสียงกับระบบออกอากาศ เพื่อรับสัญญาณเสียงจากระบบออกอากาศของสถานีโทรทัศน์มายังระบบถอดความเสียงพูดแบบใกล้เคียงเวลาจริง (signal receiver interface module) และส่งข้อความที่ได้จากการถอดความกลับไปยังระบบออกอากาศของสถานีโทรทัศน์ในรูปแบบของการถ่ายทอดสด (live stream text transmission module) ดังแสดงในภาพที่ 1 โดยส่วนที่จะต้องมีการพัฒนาเพิ่มขึ้นดังที่ได้กล่าวมาแล้วแสดงอยู่ในกรอบสีเขียว



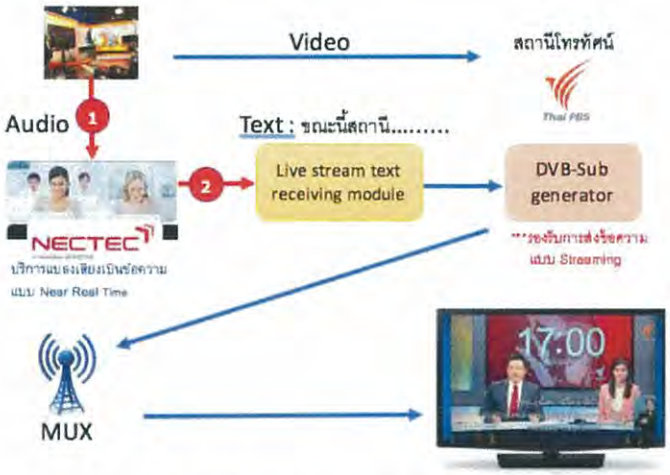
ภาพที่ 1 โครงสร้างของระบบให้บริการคำบรรยายแทนเสียงสำหรับรายการโทรทัศน์ที่ออกอากาศสด

ในส่วนของระบบถอดความเสียงแบบใกล้เคียงเวลาจริง เนื่องจากความเร็วในการพูดของคนโดยทั่วไปจะอยู่ที่ 120-160 คำต่อนาที แต่ความเร็วในการพิมพ์ของเจ้าหน้าที่พิมพ์ดีดที่มีความชำนาญจะอยู่ที่ 50-60 คำนาที ซึ่งไม่เพียงพอที่จะจัดทำคำบรรยายเสียงแบบใกล้เคียงเวลาจริง สำหรับเทคโนโลยีรู้จำเสียงพูดแบบอัตโนมัติ (Automatic Speech Recognition, ASR) ยังมีข้อจำกัดในเรื่องของความถูกต้อง เนื่องจากความหลากหลายของผู้พูด รูปแบบการพูด และเสียงรบกวน โดยเฉพาะระบบรู้จำเสียงพูดภาษาไทยยังมีความถูกต้องอยู่ในระดับ 80% (Chunwijitra, Chotimongkol, and Wutiwiwatchai, 2015, pp.2385-2389) ต้องมีการแก้ไขให้ข้อความมีความถูกต้องมากกว่า 90% เสียก่อน เพื่อให้คนพิการทางการได้ยินสามารถเข้าใจเนื้อหาได้ จากข้อจำกัดข้างต้นทีมวิจัยจึงเลือกใชเทคนิคการถอดความเสียงแบบช่วยกันพิมพ์ของเนคเทค (ห้างหุ้นส่วนจำกัด นีว่า เทคโนโลยี, 2560, น. 1-2) ซึ่งเป็นเทคนิคที่ใช้ผู้ถอดความหลายคนช่วยกันถอดความโดยใช้แป้นพิมพ์คีย์บอร์ดแบบปรกติ

โดยเจ้าหน้าที่แต่ละคนจะรับผิดชอบถอดความเสียงเฉพาะช่วงสั้น ๆ สำหรับเสียงพูดในช่วงถัดไปจะถูกถอดความโดยเจ้าหน้าที่คนอื่น เมื่อนำข้อความที่เจ้าหน้าที่แต่ละคนถอดความได้มาเรียงต่อกัน จะได้ข้อความที่สมบูรณ์ภายในเวลาจริง ระบบถอดความเสียงพูดแบบช่วยกันพิมพ์จะใช้ Protocol WebRTC (Web Real-Time Communication) ในการส่งข้อความ เนื่องจาก WebRTC เป็นโปรโตคอลแบบ real-time สามารถส่งข้อมูลออกมาทุกครั้งเมื่อมีการกดปุ่มตัวอักษรบนแป้นพิมพ์ โดยไม่ต้องรอให้มีการกด Enter หรือ Send เมื่อจบข้อความก่อน ทำให้สามารถส่งข้อความจากผู้ถอดความมายังระบบออกอากาศได้ในรูปแบบสายข้อมูลตัวอักษรที่มีการผลิตอย่างต่อเนื่อง เพื่อให้ทันต่อเวลาจริงมากที่สุด (Mozilla, 2018, online)

<u>การทดสอบการเชื่อมต่อกับอุปกรณ์</u>

การทดสอบการเชื่อมต่ออุปกรณ์เพื่อออกแบบระบบบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ มีการทดสอบใน 2 ส่วน คือ 1) Signal receiver interface module และ 2) Live stream text receiving module ดังแสดงในภาพที่ 2



ภาพที่ 2 แผนผังการทดสอบการเชื่อมต่ออุปกรณ์

การทดสอบการเชื่อมต่ออุปกรณ์ในส่วนที่ 1) signal receiver interface module โดยการส่งสัญญาณเสียงที่ดึงจากห้องส่งของสถานีโทรทัศน์ไทยพีบีเอสผ่านอินเทอร์เน็ตไปยังระบบถอดความเสียงของเนคเทค และส่งสัญญาณเสียงกลับมายังสถานีโทรทัศน์ไทยพีบีเอสอีกครั้ง พบว่าสัญญาณเสียงที่ส่งกลับมามีลักษณะเป็นเสียงสะท้อน (echo) ของเสียงที่ส่งไป แสดงว่าระยะเวลาหน่วง (delay) ที่เกิดจากการส่งสัญญาณเสียงไปและกลับผ่านอินเทอร์เน็ตน้อยกว่า 100 milliseconds ซึ่งถือว่าน้อยมาก

สำหรับการทดสอบในส่วนที่ 2) จะเป็นการทดสอบการรับข้อมูลข้อความ (text) แบบถ่ายทอดสด (live stream) ของ Live stream text receiving module รวมทั้งการแปลงข้อความที่ได้รับให้เป็นคำบรรยายแทนเสียงแบบปิดตามมาตรฐาน DVB-Subtitle (DVB-sub generator) โดยได้ทดสอบใน 2 แนวทางคือ

แนวทางที่ 1

ใช้เครื่อง Screen U8000 เป็น DVB-sub generator และใช้ซอฟต์แวร์ WinCAPS Quantum Q-Live (Q-Live) เป็น live stream text receiving module

แนวทางที่ 2

ใช้เครื่อง VT3 รุ่น Peach ทำหน้าที่เป็น DVB-Subtitle Generator และใช้ซอฟต์แวร์ PIMmala Subtitle Gateway ที่พัฒนาขึ้นใหม่โดย หจก.นีว่า เทคโนโลยี ทำหน้าที่เป็น live stream text receiving module

จากการทดสอบในแนวทางที่ 1 พบว่า Q-Live ออกแบบมาสำหรับการแสดงผลภาษาอังกฤษ หรือภาษาที่มีช่องว่างคั่นระหว่างคำเป็นหลัก ทำเมื่อนำมาใช้กับภาษาที่ไม่มีช่องว่างระหว่างคำ แต่มีช่องว่างระหว่างวลี เช่น ภาษาไทย การแสดงผลข้อความจะไม่สมบูรณ์ มีปัญหาการจัดบรรทัดเมื่อมีวลียาว ๆ ดังแสดงในภาพที่ 3 เนื่องจาก Q-Live ไม่สามารถทำ Line wrapping ได้ วลีที่ยาวจะถูกเลื่อนลงมาอยู่บรรทัดถัดไป และเกิดช่องว่างขนาดใหญ่ในบรรทัดบน เมื่อพยายามปรับค่า (setting) ให้สามารถแสดงผลทีละตัวอักษรได้ พบว่าข้อความทั้งข้อความจะเกิดการกระพริบ เนื่องจาก Q-Live มีการประมวลผลข้อความแล้วนำไป Render ภาพใหม่ทุกครั้ง เมื่อสอบถามไปยังบริษัท Screen พบว่าการใช้งานกับภาษาที่ไม่มีช่องว่างระหว่างคำเหมือนกับภาษาไทย จะใช้การแสดงผลทีละบรรทัด ซึ่งจะทำให้มี Delay time สูงขึ้น และได้รับแจ้งในเบื้องต้นว่าไม่สามารถปรับซอฟต์แวร์ให้รองรับการแสดงผลภาษาไทยได้

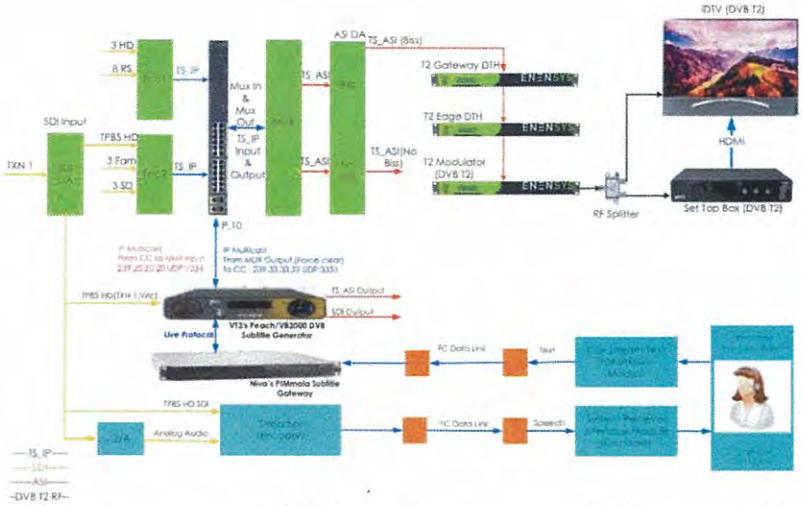
ภาพที่ 3 ตัวอย่างการแสดงผลคำบรรยายแทนเสียงบนหน้าจอโทรทัศน์ด้วยซอฟต์แวร์ Q-Live

จากการทดสอบในแนวทางที่ 2 โดยใช้ซอฟต์แวร์ PIMmala Subtitle Gateway ควบคุมการแสดงผล ปรับขนาดตัวอักษร สี ตำแหน่งสูง-ต่ำ การจัดวางกึ่งกลาง ชิดซ้ายหรือชิดขวา และรูปแบบการแสดงผลอักษร พบว่าสามารถแสดงผลแบบทยอย โดยแบ่งการแสดงผลออกเป็นส่วน ๆ แล้วทยอยแสดงอักษรข้อความในแต่ละส่วนตามลำดับ และกำหนดจังหวะเวลาที่เหมาะสมต่อระบบการส่งออกอากาศ ดังแสดงในภาพที่ 4 ทำให้ข้อความปรากฏขึ้นบนจอภาพค่อนข้างเร็ว และอ่านได้ง่ายกว่าซึ่งเหมาะกับการแสดงผลภาษาไทยซึ่งไม่มีช่องว่างระหว่างคำ

ภาพที่ 4 ตัวอย่างการแสดงผลคำบรรยายแทนเสียงบนหน้าจอโทรทัศน์ด้วยซอฟต์แวร์ PIMmala

จากผลการทดสอบเปรียบเทียบด้วยระบบอุปกรณ์และซอฟต์แวร์ในแนวทางที่ 1 และแนวทางที่ 2 ได้ผลเป็นที่ประจักษ์ สามารถสรุปได้ว่า ระบบอุปกรณ์และซอฟต์แวร์ในแนวทางที่ 2 เหมาะสมกับการแสดงผลภาษาไทยมากกว่า และยังคงมีการปรับปรุงพัฒนาซอฟต์แวร์ให้รองรับการแสดงผลภาษาไทยได้อย่างต่อเนื่อง

ดังนั้น จึงสามารถออกแบบโครงสร้างของระบบออกอากาศเพื่อให้บริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ดังภาพที่ 5

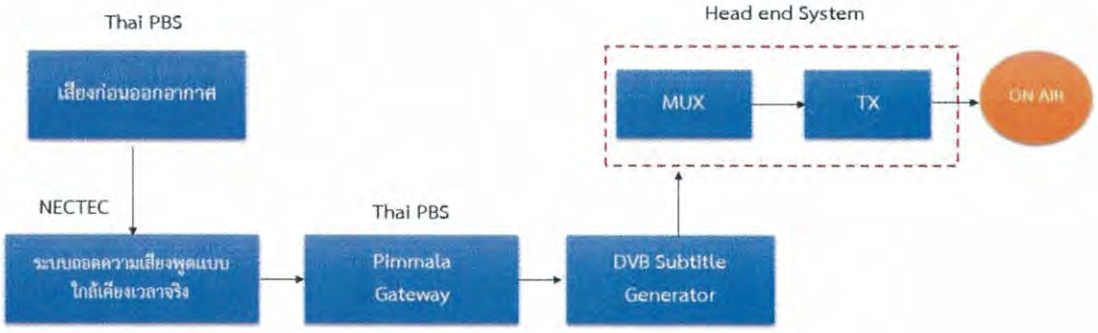


ภาพที่ 5 โครงสร้างของระบบออกอากาศเพื่อให้บริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ที่เหมาะสม

ผลการศึกษา

ผลการทดสอบเชิงเทคนิค

โครงการศึกษาวิจัยฯ ดำเนินการทดสอบระบบบริการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิด เชิงเทคนิค จำนวน 5 ครั้ง เพื่อทดลองการทำงานของส่วนที่เกี่ยวข้องในการจัดทำระบบบริการให้สามารถจัดทำผลงานคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิดได้อย่างมีประสิทธิภาพที่สุด โดยใช้รูปแบบการต่อเชื่อมมอดูลและอุปกรณ์ต่าง ๆ ในการผลิตคำบรรยายแทนเสียงแบบปิดตามมาตรฐาน DVB-Subtitle เพื่อออกอากาศในระบบออกอากาศโทรทัศน์ภาคพื้นดินระบบดิจิทัล (DVB T2) ของสถานีโทรทัศน์ ดังแสดงในภาพที่ 6



ภาพที่ 6 การเชื่อมต่ออุปกรณ์สำหรับทดสอบเชิงเทคนิค

ทดสอบการต่อเชื่อมองค์ประกอบต่าง ๆ ของระบบ ให้บริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิดพบว่า สามารถเลือกเปิด/ปิดคำบรรยายแทนเสียงภาษาไทยผ่านเมนูของเครื่องรับสัญญาณโทรทัศน์

ภาคพื้นดินฯ ด้านการแสดงผลคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิดบนหน้าจอโทรทัศน์พบว่า สามารถปรับรูปแบบตัวอักษร สีของตัวอักษร สีของพื้นหลัง และการจัดวางตำแหน่งของตัวอักษรบนหน้าจอโทรทัศน์ได้

สำหรับผลการทดสอบความเสถียรของระบบออกอากาศเมื่อมีการให้บริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิดร่วมด้วย โดยการวัดค่า bit rate ที่ใช้ในการออกอากาศพบว่า การแสดงผลคำบรรยายแทนเสียงทีละตัวอักษร (character) เพื่อให้ทันต่อเวลาจริง ทำให้ระบบต้อง refresh บ่อย และอาจทำให้เกิดปัญหาการใช้ bit rate สูงกว่าที่ได้รับการจัดสรร (bandwidth overshoot) ซึ่งจะกระทบต่อการทำงานระบบโครงข่าย (MUX) และอาจทำให้โครงข่ายล่มได้ ทั้งนี้ ปัจจุบัน กสทช. จัดสรร bandwidth สำหรับบริการคำบรรยายแทนเสียงไว้ที่ 100 kpbs (สำนักงานคณะกรรมการกิจการกระจายเสียง กิจการโทรทัศน์ และกิจการโทรคมนาคม, 2560, น.36) โดยสถานีโทรทัศน์ไทยพีบีเอสได้กำหนดค่า bit rate ที่ไว้ที่ 80 kpbs เพื่อป้องกันปัญหาไว้อีกชั้นหนึ่ง เพื่อแก้ปัญหา bandwidth overshoot จึงปรับรูปแบบการแสดงผลคำบรรยายแทนเสียงจากแบบทีละตัวอักษร เป็นใช้ความถี่ในการส่ง (refresh rate) ที่ตั้งตามเวลาแทน โดยได้ทดสอบหาค่า Refresh rate ที่มีค่าความถี่สูงสุด เพื่อให้ได้การแสดงผลทันต่อเวลา แต่ยังคงมี bit rate ที่อยู่ในกรอบ bandwidth ที่ได้จัดสรรไว้สำหรับบริการคำบรรยายแทนเสียง

จากการทดสอบพบว่า การแสดงผลคำบรรยายแทนเสียงแบบ HD 720 ความถี่ในการส่งข้อมูลสูงสุดที่ไม่ทำให้เกิดปัญหา bandwidth overshoot คือ refresh rate ที่ 500 milliseconds ส่วนการแสดงผลคำบรรยายแทนเสียงแบบ SD 625 ความถี่ในการส่งข้อมูลสูงสุดที่ไม่ทำให้เกิดปัญหา bandwidth overshoot คือ refresh rate ที่ >= 250 milliseconds ซึ่งทำให้สามารถแสดงผลคำบรรยายแทนเสียงได้ทันต่อเวลาจริงและมีความต่อเนื่องมากกว่าการแสดงผลแบบ HD

หลังจากทดสอบและปรับปรุงระบบให้บริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิด ในประเด็นต่าง ๆ ที่กล่าวข้างต้นแล้ว ได้มีการทดสอบออกอากาศบริการคำบรรยายแทนเสียงแบบเสมือนจริง กับโครงข่าย (MUX) สำรอง ในวันที่ 6 มกราคม พ.ศ. 2561 โดยทดสอบการรับชมจากเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ ประเภท STB และเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินประเภท iDTV ซึ่งสามารถเลือกเปิด/ปิดบริการคำบรรยายแทนเสียงได้จากรีโมทคอนโทรล ดังแสดงในภาพที่ 7

ภาพที่ 7 การเปิดบริการคำบรรยายแทนเสียงบนเครื่องรับสัญญาณโทรทัศน์ฯ ประเภท iDTV

การทดสอบออกอากาศบริการคำบรรยายแทนเสียงแบบเสมือนจริง ใช้ความละเอียดการแสดงผลคำบรรยายแทนเสียงแบบ SD ที่ refresh rate 250 milliseconds ตามที่ได้จากการทดสอบเชิงเทคนิคครั้งที่ 2 อย่างไรก็ดีพบกว่าเมื่อทดสอบออกอากาศผ่าน MUX ความถี่ในการแสดงผลที่ปรากฏบนจอโทรทัศน์ผ่านเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ ประเภท STB และเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินประเภท iDTV มีความถี่น้อยกว่าค่าที่ DVB subtitle generator ตั้งไว้ และมีความไม่สม่ำเสมอ ซึ่งอาจจะเนื่องมาจากความเร็วในการแปลงข้อมูล (encode) เพื่อส่งสัญญาณภาพไปออกอากาศ จึงได้ปรับมาใช้การแสดงผลแบบ SD ที่ refresh rate 500 milliseconds แทน

ค่าความถูกต้องในการถอดความรายการข่าวอยู่ที่ 80-90% เมื่อใช้ผู้ถอดความ 5 คน หากใช้ผู้ถอดความน้อยกว่านี้ค่าความถูกต้องจะลดลงค่อนข้างมาก เนื่องจากรายการข่าวมีอัตราการพูดค่อนข้างเร็ว สำหรับระยะเวลาหน่วง หรือค่า delay จะวัดจากระยะห่างระหว่างเวลาที่ผู้ชมได้ยินเสียงพูดกับเวลาที่ผู้ชมเห็นคำบรรยายแทนเสียงของประโยคนั้นปรากฏบนจอโทรทัศน์ จากการวิเคราะห์ค่า delay ของรายการทันข่าว 16.00 น. พบว่ามีค่าเฉลี่ยอยู่ที่ 6.6 วินาที

เนื่องจากคำบรรยายแทนเสียงจะปรากฏช้ากว่าเสียงพูด และต้องเผื่อเวลาให้ผู้ใช้อ่านข้อความได้ครบ ดังนั้นควรแสดงผลคำบรรยายแทนเสียงค้างไว้สักระยะหนึ่งหลังจากรายการที่มีบริการคำบรรยายแทนเสียงจบลง เมื่อสิ้นสุดการถอดความเสียงระบบถอดความเสียงแบบทันต่อเวลาจะส่งข้อความ “[สิ้นสุดการถอดความ]” เพื่อให้ผู้ใช้ทราบดังแสดงในภาพที่ 22 และส่งอักษระพิเศษ “∈” (Unicode U+22F3) เพื่อให้โปรแกรม PIMmala

subtitle gateway หยุดการแสดงผลคำบรรยายแทนเสียง คลิป https://youtu.be/fOaijkBAsZw นาทีที่ 5.13 แสดงการคงไว้ของคำบรรยายแทนเสียงหลังจบรายการระยะหนึ่ง ก่อนหยุดการแสดงผลคำบรรยายแทนเสียง

ภาพที่ 5 ข้อความ "[สิ้นสุดการถอดความ]" ปรากฏขึ้นเมื่อจบรายการที่มีบริการคำบรรยายแทนเสียง

ผลการทดสอบกับผู้ใช้

โครงการศึกษาวิจัยฯ ดำเนินการทดสอบระบบบริการจัดทำคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิด กับกลุ่มผู้ใช้กลุ่มคนพิการทางการได้ยิน จำนวน 2 ครั้ง จำนวน 233 คน เพื่อรับทราบความคิดเห็น และความต้องการเกี่ยวกับการจัดทำระบบบริการฯ ผ่านการตอบแบบสอบถาม โดยการให้ผู้ใช้ดูคลิปวิดีโอรายการโทรทัศน์ที่มีคำบรรยายแทนเสียงรูปแบบต่าง ๆ จากเครื่องรับโทรทัศน์ โดยแบ่งผู้ใช้เป็นกลุ่มย่อยจำนวน 10-15 คน ต่อเครื่องรับโทรทัศน์ 1 เครื่อง ดังแสดงในภาพที่ 9 เพื่อให้ใกล้เคียงกับการรับชมโทรทัศน์โดยทั่วไป การทดสอบโดยการใช้คลิปวิดีโอช่วยให้สามารถเปรียบเทียบคำบรรยายแทนเสียงรูปแบบต่าง ๆ โดยใช้เนื้อหารายการเดียวกันได้ และสามารถทดสอบกับผู้ใช้จำนวนมากโดยแบ่งเป็นรอบ ๆ ได้

ภาพที่ 9 การทดสอบรับชมคลิปวีดิโอรายการโทรทัศน์ที่มีคำบรรยายแทนเสียงจากเครื่องรับโทรทัศน์

โดยในการทดสอบกับผู้ใช้ครั้งที่ 1 ได้สอบถามความเห็นจากผู้ใช้ เพื่อทดสอบและประเมินปัจจัยที่มีผลต่อการใช้งานบริการคำบรรยายแทนเสียง 3 ปัจจัยคือ 1) ความเร็วในการพูดของผู้พูด 2) ความถูกต้องในการถอดความ และ 3) รูปแบบการแสดงผลคำบรรยายแทนเสียง โดยประมวลความเห็นของผู้ใช้ได้ดังนี้

ความเห็นของผู้ใช้ต่อการอ่านคำบรรยายแทนเสียงที่ผู้พูดมีอัตราความเร็วในการพูดแตกต่างกัน พบว่าความเห็นของผู้ใช้ต่อคำถามที่ว่า "อ่านคำบรรยายแทนเสียงได้ทันหรือไม่" ไม่แตกต่างกันมากนัก ทั้งนี้จากข้อกำหนดของ กสทช. ที่กำหนดให้การถอดความเพื่อจัดทำคำบรรยายแทนเสียงต้องมีข้อความครบตามที่เสียงพูด ยังไม่อนุญาตให้ใช้การสรุปความ จึงยังไม่สามารถแสดงผลข้อความให้ช้าลงหรือมีปริมาณข้อความที่น้อยลงได้ อย่างไรก็ดีจากการสอบถามเพิ่มเติมพบว่า ลักษณะการแสดงผลการไหลของตัวอักษร (text flow) เช่น จังหวะการเลื่อนตัวอักษรเพื่อขึ้นบรรทัดใหม่ ความสม่ำเสมอของตัวอักษรที่ปรากฏขึ้นใหม่ มีผลต่อความยากง่ายในการอ่านคำบรรยายแทนเสียง จึงได้มีการปรับปรุงการไหลของข้อความคำบรรยายแทนเสียงให้ตัวอักษณที่ปรากฏขึ้นใหม่มีความสม่ำเสมอมากขึ้น และมีการรอการแสดงผลเมื่อมีการเลื่อนบรรทัดเพื่อให้ผู้ใช้อ่านได้ทัน และได้มีการทดสอบกับผู้ใช้อีกครั้ง โดยกลุ่มผู้ใช้ที่เข้าร่วมการทดสอบทั้ง 2 ครั้ง จำนวน 30 คน ทั้งหมดให้ความเห็นว่าการปรับปรุงการแสดงผลในเรื่องการไหลของข้อความ ช่วยให้อ่านคำบรรยายแทนเสียงได้ง่ายขึ้น

สำหรับความเข้าใจเนื้อหาของรายการโทรทัศน์จากการอ่านบริการคำบรรยายแทนเสียงพบว่า ความถูกต้องในการตอบคำถามที่เกี่ยวข้องกับเนื้อหารายการสอดคล้องกับความเห็นของผู้ใช้ว่าอ่านคำบรรยายแทนเสียงทันหรือไม่ โดยกลุ่มผู้ใช้ที่ตอบว่าอ่านคำบรรยายแทนเสียงได้ทันจะมีคะแนนความถูกต้องโดยเฉลี่ยสูงกว่ากลุ่มผู้ใช้ที่ตอบว่าอ่านไม่ทัน

ความเห็นของผู้ใช้ต่อคุณภาพของคำบรรยายแทนเสียง ผู้ใช้ 47.1% ยอมรับข้อผิดพลาดที่สังเกตเห็นได้ เนื่องจากเห็นว่าข้อผิดพลาดไม่ร้ายแรง ไม่กระทบกับเนื้อหา ขณะที่ 51.9% ไม่ยอมรับ เนื่องจากเห็นว่าข้อผิดพลาดที่พบทำให้อ่านคำบรรยายแทนเสียงได้ยาก และมีผลต่อการเรียนรู้ภาษาเขียนของคนหูหนวก ทั้งนี้ ความถูกต้องในการถอดความเสียงแบบทันต่อเวลาสำหรับรายการข่าวอยู่ที่ 80-90%

ผลการทดสอบอุปกรณ์ภาคออกอากาศ

โครงการศึกษาวิจัยฯ ได้นำเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB (DVB-T2 Set-top box) ที่มีจำหน่ายในท้องตลาด มาทดสอบการรับสัญญาณเสมือนการแพร่ภาพออกอากาศรายการโทรทัศน์จริง และตรวจดูการแสดงผลของข้อความคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ บนหน้าจอภาพ ในวันที่ 6 มกราคม 2561 เวลา 10.00 – 20.15 น. เพื่อทดสอบใน 2 ประเด็น คือ 1) เครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB ยี่ห้อต่าง ๆ ที่มีจำหน่ายในตลาด จะสามารถรับสัญญาณคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิดได้หรือไม่ และมีความแตกต่างกันหรือไม่อย่างไร 2) หากมีการเปลี่ยนแปลงค่าพารามิเตอร์ที่ใช้ในการส่งคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ แบบปิด จะเกิดผลอย่างไรต่อเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB ที่นำมาทดสอบ เพื่อนำมาใช้ในการกำหนดค่าพารามิเตอร์ของระบบ subtitle generator ต่อไป

ทั้งนี้ ทีมวิจัยออกแบบการทดสอบโดยใช้เครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB จำนวน 8 เครื่องที่มียี่ห้อแตกต่างกัน ส่งสัญญาณคลื่นโทรทัศน์ระบบดิจิทัล ที่ทดสอบการให้บริการระบบคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ให้กับเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB ผ่านการต่อสายสัญญาณภาพและเสียงชนิด HDMI จากเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB ไปที่เครื่องรับโทรทัศน์ และบันทึกรายการโทรทัศน์ที่ทดสอบให้บริการระบบคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ลงใน flash memory ด้วยเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB ซึ่งมีคุณสมบัติสามารถบันทึกรายการโทรทัศน์ได้ ในระหว่างดำเนินการทดลอง ดังแสดงในภาพที่ 10

ภาพที่ 10 ลักษณะการใช้งานอุปกรณ์เพื่อทดสอบ
การแสดงผลของข้อความคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ บนหน้าจอภาพ

ผลที่ได้จากการทดสอบ มีดังนี้

เครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB	พารามิเตอร์ของระบบ		การแสดงผลคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ
	Refresh rate (milliseconds)	ระบบภาพ	
จำนวน 4 เครื่อง (4 ยี่ห้อ)	500	SD 625	แสดงผลสมบูรณ์เป็นปกติ
	500	HD 720	แสดงผลสมบูรณ์เป็นปกติ
	250	SD 625	แสดงผลสมบูรณ์เป็นปกติ
	250	HD 720	แสดงผลสมบูรณ์เป็นปกติ
จำนวน 4 เครื่อง (4 ยี่ห้อ)	500	SD 625	แสดงผลสมบูรณ์เป็นปกติ
	500	HD 720	ข้อความคำบรรยายแทนเสียงกระพริบ
	250	SD 625	ข้อความคำบรรยายแทนเสียงกระพริบ
	250	HD 720	ข้อความคำบรรยายแทนเสียงกระพริบ

ตารางที่ 2 สรุปผลการทดสอบเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB

ข้อสรุปที่ได้จากการทดสอบ มีดังนี้

เมื่อกำหนดค่าพารามิเตอร์ของ subtitle generator เป็นระบบ HD 720 และ MUX ที่ใช้ในการทดสอบได้จำกัด bandwidth สูงสุดไว้ที่ไม่เกิน 80 kbps และมี head room เผื่อไว้กรณีที่อาจจะเกิด overshoot ได้ถึง 100 kbps ส่งผลให้เครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB จำนวน 4 ยี่ห้อ พบปัญหาคือ การแสดงผลของข้อความครบถ้วน แต่สั่นและกระพริบเป็นจังหวะ ไม่ว่าจะเปลี่ยนค่า refresh rate อย่างไร ส่วนเครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB อีก 4 ยี่ห้อ สามารถแสดงผลสมบูรณ์เป็นปกติ แสดงให้เห็นว่า เครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB ที่ไม่ได้ออกแบบมาเพื่อรองรับปริมาณข้อมูลที่มากเพียงพอ ไม่สามารถแสดงผลข้อความคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ได้เป็นปกติ เช่น ภาพกระพริบ หรือบางครั้งข้อความขาดหายไป

เมื่อนำไฟล์วิดีโอที่บันทึกได้จากเครื่อง STB ที่แสดงผลข้อความคำบรรยายแทนเสียงกระพริบ กลับมา playback ด้วยโปรแกรม VLC ที่สามารถถอดรหัส closed caption ได้ ปรากฏว่าข้อความคำบรรยายแทนเสียงแสดงผลได้เป็นปกติ ไม่มีอาการกระพริบ

ดังนั้น จึงสามารถสรุปได้ว่า เครื่องรับสัญญาณโทรทัศน์ภาคพื้นดินฯ แบบ STB ที่มีขายอยู่ในตลาดช่วงเวลา 4 ปีที่ผ่านมาจนถึงปัจจุบัน คาดว่าสามารถรองรับอัตราข้อมูลของระบบคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ได้สูงสุดระหว่าง 80–100 Kbps หากอัตราข้อมูลมากกว่านี้ จะส่งผลให้เครื่องแปลงสัญญาณโทรทัศน์บางยี่ห้อหรือบางรุ่นไม่สามารถแสดงผลคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ตามโครงการนี้ได้

บทสรุปผลการทดสอบ และข้อเสนอแนะเชิงเทคนิค

จากผลการศึกษาพบว่า สถานีโทรทัศน์ไทยพีบีเอสสามารถดำเนินการจัดทำบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ได้จริงในเชิงเทคนิค และกลุ่มผู้ใช้งานคนพิการทางการได้ยิน ซึ่งเป็นกลุ่มเป้าหมายหลักยอมรับผลการทดสอบว่าอยู่ในระดับการใช้งานที่เข้าใจได้ ทั้งนี้ จากผลการทดสอบเชิงเทคนิค ยังพบข้อจำกัดที่ยังไม่สามารถแก้ไขได้ในปัจจุบัน ต้องการการพัฒนาเพิ่มเติมเพื่อเพิ่มคุณภาพบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ดังนี้

ด้านความถูกต้องของการถอดความ ค่าความถูกต้องในการถอดความด้วยเทคนิคการแบ่งพิมพ์ สำหรับรายการข่าวมีค่าความถูกต้องเฉลี่ยอยู่ที่ 80-90% จากการทดสอบกับผู้ใช้พบว่าผู้ใช้บางส่วนยอมรับความถูกต้องที่ระดับนี้ได้ ในขณะที่ผู้ใช้บางส่วนต้องการความถูกต้องที่สูงกว่านี้ ปัจจัยที่มีผลต่อความถูกต้องของการถอดความได้แก่ คุณภาพของเสียงพูด และคำศัพท์และชื่อเฉพาะ ปัจจัยด้านคุณภาพของเสียงอาจเป็นเรื่องที่ปรับปรุงได้ยาก เนื่องจากรายการข่าวอาจมีการสัมภาษณ์นอกสถานที่ ส่วนปัจจัยด้านคำศัพท์สามารถปรับปรุงได้ โดยให้ผู้ถอดความศึกษาสคริปต์ที่ใช้ในการอ่านข่าวก่อน ซึ่งแนวทางนี้ควรมีการพัฒนาส่วนต่อเชื่อมสำหรับดึงข้อมูลสคริปต์ข่าวเพื่อส่งต่อไปยังระบบถอดความเสียงพูดแบบทันต่อเวลา

ด้านความทันต่อเวลา การจัดสรรอัตราบิตไว้ให้กับบริการคำบรรยายแทนเสียงของประเทศไทยอยู่ที่ 100 kpbs ส่งผลให้ไม่สามารถแสดงผลคำบรรยายแทนเสียงแบบทีละตัวอักษรได้ ซึ่งการแสดงผลดังกล่าวจำเป็นต้องใช้ bandwidth เกินกว่าที่กำหนด ทั้งนี้ ค่าความถี่ที่สามารถใช้ได้ในสัดส่วนดังกล่าวคือ 500 milliseconds สำหรับการแสดงผลแบบ SD เท่านั้น หากต้องการแสดงผลแบบ HD จำเป็นต้องใช้ค่าความถี่ที่น้อยกว่านี้ การศึกษานี้จึงต้องใช้การแสดงข้อความปรากฏบนจอโทรทัศน์แบบทีละกลุ่มตัวอักษร ทำให้ค่า delay ของการแสดงผลคำบรรยายแทนเสียงเมื่อเทียบกับเสียงพูด เพิ่มขึ้นจากค่า delay ของระบบถอดความแบบทันต่อเวลาที่สามารถแสดงผลทีละตัวอักษรได้ โดยจากการทดสอบพบว่าค่า delay เฉลี่ยอยู่ที่ 6.6 วินาที ดังนั้น ควรเพิ่มการจัดสรรอัตราบิตให้กับบริการคำบรรยายแทนเสียงเป็น 200 kpbs

ด้านรูปแบบการแสดงผลคำบรรยายแทนเสียง เนื่องจากข้อจำกัดด้าน Bandwidth ที่กล่าวข้างต้น ทำให้ต้องแสดงผลคำบรรยายแทนเสียงบนจอโทรทัศน์ทีละกลุ่มตัวอักษร ซึ่งกลุ่มตัวอักษรนี้อาจจะไม่ครบเต็มคำ ซึ่งส่งผลให้ขาดความต่อเนื่องในการอ่านข้อความ การแก้ปัญหาการแสดงผลไม่ครบคำต้องอาศัยโปรแกรมตัดคำเข้ามาช่วย อย่างไรก็ดี โปรแกรมตัดคำภาษาไทยที่มีอยู่ในปัจจุบันไม่สามารถรองรับข้อความที่ยังไม่สมบูรณ์ที่ได้จากการพิมพ์ข้อความแบบต่อเนื่อง และอาจมีการแก้ไขข้อความที่พิมพ์ไปแล้ว ซึ่งจะต้องมีการพัฒนาโปรแกรมตัดคำภาษาไทยที่สามารถรองรับข้อความที่ได้จากระบบถอดความเสียงพูดแบบทันต่อเวลาขึ้นมาใหม่

ข้อเสนอแนะเชิงนโยบาย

1) สำนักงาน กสทช. ควรพิจารณาปรับนโยบายสำหรับแนวทางการบริหารจัดการอัตราบิตที่เหลืออยู่ ให้สามารถนำมาใช้ประโยชน์ได้อย่างคุ้มค่ามากขึ้นกับบริการที่เกิดขึ้นในทีวีดิจิทัล โดยจัดสรรอัตราบิตให้ถึง 200 kbps ทั้งนี้ ปัจจุบันการให้บริการโครงข่ายโทรทัศน์ภาคพื้นดินในระบบดิจิทัลของบริษัท อสมท. จำกัด (มหาชน) กรมประชาสัมพันธ์ และ ส.ส.ท. ยังคงมี bandwidth เหลืออยู่ ซึ่งสามารถนำอัตราบิตที่เหลือจากการให้บริการโครงข่ายโทรทัศน์ฯ มาใช้ประโยชน์ได้

2) ควรมีการศึกษาวิจัย เพื่อทดลองดำเนินการจริงของสถานีโทรทัศน์แบบหลายสถานีส่งสัญญาณแพร่ภาพพร้อมกัน และการทำงานของหน่วยบริการ MUX

3) ควรมีศูนย์บริการการจัดทำคำบรรยายแทนเสียงที่เป็นหน่วยงานกลาง ทำหน้าที่สนับสนุนการจัดบริการคำบรรยายแทนเสียงในช่วงเวลาเดียวกันฯ ให้กับสถานีโทรทัศน์ที่มีความต้องการ เพื่อเป็นการใช้ทรัพยากรร่วมกันหลายหน่วยงาน แทนที่หน่วยงานจะแยกจัดทำบริการเอง

4) ควรส่งเสริมการใช้งานระบบบริการคำบรรยายแทนเสียงของกลุ่มผู้ใช้คนพิการทางการได้ยิน โดยทำงานร่วมกับสถาบันการศึกษา สมาคม ชมรม หรือกลุ่มของคนพิการทางการได้ยิน ซึ่งมีสมาชิกเป็นคนหูหนวกหรือคนหูตึง ในการให้ความรู้ ความเข้าใจเรื่องการใช้งานบริการคำบรรยายแทนเสียงแบบปิดในระบบโทรทัศน์ดิจิทัล

กิตติกรรมประกาศ

ขอขอบคุณ คุณวีระ ชิดชอบ สำหรับข้อมูลทางเทคนิคของอุปกรณ์ VT3 รุ่น Peach และซอฟต์แวร์ PIMmala subtitle gateway และคำแนะนำทางเทคนิคในการทดสอบต่อเชื่อมอุปกรณ์ทั้งสองในการทดสอบในงานวิจัยนี้ และขอขอบคุณ คุณเดชาวัต จึงจตุพรชัย สำหรับคำแนะนำวิธีถอดความเสียงพูดของเจ้าหน้าที่บริการถอดความเสียงพูดในช่วงเวลาเดียวกัน

บรรณานุกรม

สำนักงานคณะกรรมการกิจการกระจายเสียง กิจการโทรทัศน์ และกิจการโทรคมนาคม, 2560, *แนวปฏิบัติทางเทคนิคสำหรับการให้บริการโทรทัศน์ภาคพื้นดินในระบบดิจิตอล (Technical Guidelines for Digital Terrestrial Television Broadcasting).* กรุงเทพมหานคร. น. 15, 36.

ห้างหุ้นส่วนจำกัด นีว่า เทคโนโลยี. (2560). *คำป้อน Subtitle.* น. 1-2. (ออนไลน์). สืบค้นจาก http://www.nivatech.co.th/download/Kumpon-Subtitle.pdf

Chunwijitra, V., Chotimongkol, V., and Wutiwiwatchai, C., (2015)., Combining multiple-type input units using recurrent neural network for LVCSR language modeling. *Proceedings of Interspeech 2015, Germany*, *16*, 2385-2389.

Mozilla. (2018, February 3). *Introduction to WebRTC protocols*. Retrieved from https://developer.mozilla.org/en-US/docs/Web/API/WebRTC_API/Protocols

คำยินยอมในการนำเสนอบทความเพื่อลงตีพิมพ์ในวารสาร กสทช.

ข้าพเจ้ายินยอมให้กองบรรณาธิการ วารสาร กสทช. มีสิทธิ์ในการกำหนดผู้กลั่นกรอง (Reviewer) โดยอิสระเพื่อพิจารณาต้นฉบับที่ข้าพเจ้า (และผู้แต่งร่วม) ส่งมาตามที่เห็นสมควร และยินยอมให้กองบรรณาธิการวารสาร กสทช. สามารถตรวจแก้ไขต้นฉบับดังกล่าวได้ตามที่เห็นสมควร และขอรับรองว่าบทความนี้ไม่เคยลงตีพิมพ์ในวารสารใดมาก่อน และจะไม่นำบทความดังกล่าวลงตีพิมพ์ในวารสารฉบับอื่น