การรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวนของการพูดคุยในภาษาไทย และในภาษาต่างประเทศ

จุฑามณี อ่อนสุวรรณ

คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์

ศูนย์แห่งความเป็นเลิศทางวิชาการด้านสารสนเทศอัจฉริยะ เทคโนโลยีเสียงพูดและภาษา และนวัตกรรมด้านบริการ มหาวิทยาลัยธรรมศาสตร์

บทคัดย่อ

งานวิจัยนี้ศึกษาการรับรู้ และความสับสนในเสียงวรรณยุกต์ภาษาไทยที่เกิดขึ้นเมื่อปราศจากเสียงรบกวน และเมื่อมีการปนเปื้อนด้วยเสียงรบกวนตามอัตราส่วนเสียงพูดต่อเสียงรบกวน (SNR) 3 ระดับ ได้แก่ 0dB –5dB และ –10dB เสียงรบกวนของการพูดคุยที่ใช้เป็นแบบผู้พูด 2 คน 2 ชนิด ได้แก่ การพูดคุยในภาษาไทย และในภาษาต่างประเทศ (อังกฤษ) อาสาสมัครชาวไทยจำนวน 35 คน ทำการทดสอบโดยใช้คำเร่งเร้าเป้าหมายจำนวน 40 คู่ ผ่านการเลือกคำที่ได้ยินจาก 2 ตัวเลือก ผลการวิจัยแสดงให้เห็นว่าเมื่อเสียงรบกวนและระดับของ SNR มีค่าต่ำลง ร้อยละของคำตอบที่ถูกต้อง และร้อยละของการฟังเข้าใจจะลดลงตามลำดับ ณ ระดับ SNR –5dB และ–10dB ชนิดภาษาของเสียงรบกวนของการพูดคุยยังส่งผลต่อร้อยละของการฟังเข้าใจอย่างมีนัยสำคัญทางสถิติ กล่าวคือ การรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวนของการพูดคุยในภาษาไทยจะมีระดับต่ำกว่าภายใต้เสียงรบกวนของการพูดคุยในภาษาอังกฤษอย่างมีนัยสำคัญ นอกจากนี้ความสับสนที่เกิดขึ้นมีลักษณะที่ไม่สมดุล และความสับสนที่เกิดขึ้นในลำดับต้นๆ คือ ความสับสนระหว่างวรรณยุกต์เอก (เป็นสามัญ) จัตวา (เป็นเอก) และจัตวา (เป็นสามัญ) ในภาพรวมวรรณยุกต์กลุ่มที่ก่อให้เกิดความสับสนได้น้อย คือ โท ตรี และสามัญ ส่วนกลุ่มที่ก่อให้เกิดความสับสนสูงคือ เอก และจัตวา เมื่อพิจารณาร่วมกับผลวิจัยจากการศึกษากับการรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวนประเภท white noise (Onsuwan et al., 2012) กล่าวได้ว่าประเภทของเสียงรบกวนมีผลอย่างมากต่อการรับรู้ และรูปแบบของความสับสนในเสียงวรรณยุกต์ภาษาไทย

คำสำคัญ การรับรู้เสียงพูด เสียงรบกวนของการพูดคุย วรรณยุกต์ภาษาไทย อิทธิพลของชนิดภาษาของเสียงรบกวน

Perception of Thai Lexical Tones in Native- and Foreign-language Babble Noise

Chutamanee Onsuwan

Faculty of Liberal Arts, Thammasat University

Center of Excellence in Intelligent Informatics, Speech & Language Technology and Service Innovation (CILS), Thammasat University

Abstract

This work investigated perception and confusion patterns of Thai lexical tones in clean versus adverse (background noise) conditions at 3 levels of signal-to-noise ratio (SNR): 0dB, –5dB, and –10dB. Two types of two-talker babble were used: foreign (English) and native (Thai) languages. Thirty-five Thai participants took a perceptual identification test consisting of 40 target pairs (in various noise conditions) with A/B forced choice. The results indicated that as SNR decreased, percent correct responses (accuracy) and percent intelligibility also decreased. At SNRs of –5dB and –10dB, language type had a statistically significant effect on percent intelligibility, such that percent intelligibility of Thai tones in Thai babble noise was significantly lower than that in the English babble noise condition. Tonal confusions were asymmetrical with the most easily confused being low tone (as mid), rising tone (as low), and rising tone (as mid). In all, the set of tones which showed a lower degree of confusability was the falling, high, and mid tones—the higher degree one included low and rising tones. These findings, taken together with a previous study that investigated the perception of Thai lexical tones in white noise (Onsuwan et al., 2012), suggest that types of (masker) noise play a major role in perception and confusion patterns of Thai lexical tones.

Keywords: babble noise, linguistic effect from babble noise, speech perception, Thai lexical tone

1. ที่มาและความสำคัญ

ในการสื่อสารในชีวิตประจำวัน เสียงพูดของมนุษย์ได้รับการปนเปื้อนจากเสียงรบกวนต่างๆ อย่างหลีกเลี่ยงได้ยาก ทั้งจากสภาพแวดล้อม และสัญญาณรบกวนในกรณีที่เป็นการสื่อสารผ่านโทรศัพท์ หรือเครื่องมือสื่อสารอื่นๆ ดังนั้น เสียงรบกวน หมายถึง เสียงที่เกิดขึ้นจากความไม่ตั้งใจ และเป็นเสียงที่ไม่พึงปรารถนา (Stocker, 2013) ในแง่ฟิสิกส์ ไม่ว่าจะเป็นเสียงรบกวนหรือเสียงพูดก็ถือได้ว่าเป็นเสียง (sound) เช่นเดียวกันเนื่องจากเป็นผลมาจากการสั่นสะเทือนของตัวกลางไม่ว่าจะเป็นน้ำ หรืออากาศ อย่างไรก็ดีกระบวนการรับรู้เสียงของมนุษย์นั้นมีความพิเศษที่ทำให้มนุษย์สามารถรับรู้โดยแยกเสียงพูดออกจากเสียงรบกวนได้ (ในระดับหนึ่ง) (Miller, 1947; Loizou, 2007)

เสียงรบกวนมีผลต่อการรับรู้เสียงพูด และอาจก่อให้เกิดความสับสน (perceptual confusion) และความผิดพลาดในการรับรู้เสียงพูด (perceptual error) งานวิจัยจำนวนมากชี้ให้เห็นว่าการรับรู้เสียงพูดของมนุษย์มีกระบวนการที่ต่างกันในการประมวลผลเสียงพูดที่ปราศจากเสียงรบกวน (clear speech) และเสียงพูดที่อยู่ในเสียงรบกวน (speech in noise) และที่น่าสนใจคือเสียงรบกวนต่างประเภทกัน (Pollack, 1975) จะส่งผลต่อกระบวนการรับรู้ของเสียงมนุษย์ไม่เหมือนกัน กล่าวคือ เสียงรบกวนประเภทพลังงาน (energetic or non-speech masking) เช่น white noise, speech-shaped noise, babble-modulated noise และ speech-weighted noise จะส่งผลต่อกระบวนการรับรู้เสียงของมนุษย์ที่แตกต่างจากเสียงรบกวนประเภทข้อมูลภาษา (informational or linguistic masking) เช่น เสียงรบกวนของการพูดคุย (babble noise) (Cutler, 2012) ทั้งนี้เนื่องมาจากเสียงรบกวนประเภทพลังงานส่งผลโดยตรงกับกระบวนการประมวลผลทางเสียงในระดับล่าง (acoustic and auditory) ซึ่งต่างจากเสียงรบกวนประเภทข้อมูลภาษาที่ส่งผลกับการประมวลผลทางเสียงในระดับสูงขึ้นไป (cognitive and linguistic) นอกจากนี้สำหรับเสียงรบกวนที่เป็นประเภทข้อมูลภาษา ความคล้ายคลึงของเสียงรบกวนกับคำหรือเสียงเป้าหมายจะมีผลต่อการรับรู้ที่เกิดขึ้นด้วย (เช่น เพศเดียวกับผู้พูดเสียงเป้าหมาย ภาษาเดียวกับเสียงเป้าหมาย) เสียงรบกวนประเภทหนึ่งๆ อาจมีผลต่อการรับรู้เสียงพูดเพียงบางเสียง ไม่ใช่กับเสียงพูดทุกเสียง (Miller, 1947) ดังนั้นจึงกล่าวได้ว่าความผิดพลาด และความสับสนในการรับรู้เสียงพูดที่เกิดจะมีความแตกต่างกันขึ้นอยู่กับประเภทของเสียงเป้าหมาย เช่น เสียงพยัญชนะ เสียงวรรณยุกต์ วลี และประโยค กล่าวคือ การรับรู้เสียงวรรณยุกต์ และเสียงสระจะมีความคงที่ (ทนต่อเสียงรบกวน) มากกว่าเสียงพยัญชนะ (Cutler, 2012)

ในแง่ภาษาศาสตร์ เมื่อเกิดความผิดพลาด และความสับสนในการรับรู้เสียง ผลที่ได้มักจะมีลักษณะไม่สมดุล (asymmetrical) และบ่อยครั้งจะมีแนวโน้ม (trend) ที่ค่อนข้างชัดเจน ดังตัวอย่างที่กล่าวว่ามีความเป็นไปได้สูงในภาษาอังกฤษที่เสียง ‘th’ จะมีการสับสนและได้ยินพลาด (misperceive) ว่าเป็นเสียง ‘f’ แต่ไม่ใช่ในทางตรงข้าม ที่เสียง ‘f’ จะมีการสับสนและได้ยิน

พลาดเป็น ‘th’ โดย จอห์นสัน (2003) ชี้แจงว่าลักษณะความสับสนที่พบนี้มีความสอดคล้องกับปรากฏการณ์การกลายเสียงในภาษาข้ามสมัย (sound change) ในภาษาอังกฤษ กล่าวคือจะพบแนวโน้มของการกลายเสียงในเชิงประวัติ จาก ‘th’ เป็น ‘f’ มากกว่าในทิศทางที่กลับกัน ดังนั้น นักสัทศาสตร์และสัทวิทยา เช่น จอห์นสัน (2003) และ เบงกี้ (2003) มีข้อสรุปที่คล้ายกันว่าข้อมูลที่ได้จากความผิดพลาด และความสับสนในการรับรู้เสียงพูดสามารถนำมาเชื่อมโยงเพื่ออธิบายรูปแบบทางระบบเสียงของภาษา (phonological pattern) ที่เป็นเฉพาะสมัย (synchronic) รวมถึงการเปลี่ยนแปลงทางระบบเสียงในภาษาข้ามสมัย (diachronic) แนวคิดในเรื่อง “บทบาทของผู้ฟังกับการกลายเสียง” (listener as a source of sound change) ได้มีการเสนอและยืนยันอย่างกว้างขวางในทางภาษาศาสตร์โดย โอฮาล่า (1974, 1981, 1995)

ดังนั้นจะเห็นได้ว่าการศึกษาการรับรู้เสียงพูดในประเด็นข้างต้นมีความสำคัญอย่างยิ่งที่จะช่วยให้เราเข้าใจกลไกในการรับรู้ และเข้าใจภาษาของมนุษย์ และอาจจะนำมาอธิบายรูปแบบ และแนวโน้มที่เกิดขึ้นเมื่อเกิดการสับสนทางเสียงในผู้ใช้ภาษาปกติ (และผู้ใช้ภาษาที่มีความบกพร่องในแง่การรับรู้เสียง) ทั้งในระดับสากล (universal) และระดับเฉพาะภาษา (language-specific) นอกจากนี้ในด้านวิทยาศาสตร์และเทคโนโลยี การศึกษาในประเด็นนี้ยังช่วยเพิ่มองค์ความรู้ด้านการรับรู้เสียงในภาษาต่างๆ และความสับสนทางเสียงที่เกิดขึ้นในสถานการณ์ที่มีเสียงรบกวนเพื่อนำไปใช้สร้างแบบจำลองในการเพิ่มสมรรถนะเสียง (speech enhancement) ในเครื่องมือ หรืออุปกรณ์ทางเทคโนโลยีทางเสียงพูด และภาษาต่อไป

ข้อมูลความผิดพลาด และความสับสนในการรับรู้เสียงพูดนั้นส่วนใหญ่ได้มาจากการทำการทดสอบด้านการรับรู้เสียง (perception experiment) ที่เป็นระบบ และมีหลายวิธีการด้วยกัน แต่วิธีการหนึ่งที่เป็นที่ยอมรับและมีการนำมาใช้อย่างต่อเนื่องคือ การทดสอบเสียงพูดในระดับอัตราส่วนเสียงพูดต่อเสียงรบกวน (signal-to-noise ratio (SNR)) ที่แตกต่างกัน และรายงานผลด้วยเมทริกซ์ความสับสนทางการรับรู้เสียง (phoneme confusion matrix) (อาทิ Miller & Nicely (1955), Benkí (2003) และ McLoughlin (2008)) ซึ่งนอกจากจะแสดงปริมาณความถูกต้องของการรับรู้เสียงแล้วยังมีการแจกแจงว่าเสียงใดมีความสับสนกับเสียงใดในปริมาณมากน้อยเพียงใด ทั้งนี้ค่า SNR (dB) ที่ต่ำลง มักจะทำให้ความถูกต้องในการรับรู้เสียงลดลง และความสับสนมักจะเพิ่มขึ้น (อัตราส่วนที่มากกว่า 1:1 หรือมากกว่า 0dB จะมีสัญญาณเสียงมากกว่าเสียงรบกวน) (Miller, 1947)

ในงานวิจัยนี้ผู้วิจัยมีความประสงค์ที่จะทำการทดสอบการรับรู้เสียงที่มีบางส่วนคล้ายคลึงกับงานที่ผู้วิจัยและคณะได้ดำเนินการไปกับเสียงพยัญชนะ (Tantibundhit & Onsuwan, 2015) เสียงสระ (Onsuwan et al., 2013) และเสียงวรรณยุกต์ (Onsuwan et al., 2012) ในภาษาไทย อย่างไรก็ดีในงานวิจัยข้างต้นคณะวิจัยได้ใช้เสียงรบกวนประเภทเดียวคือ white noise (additive white Gaussian noise (AWGN)) จึงทำให้ผู้วิจัยต้องการทำการเปรียบเทียบ และขยายผลการ

ศึกษา โดยในงานวิจัยนี้จะศึกษาการรับรู้เสียงวรรณยุกต์ และศึกษาการสับสนทางเสียงที่เกิดขึ้นภายใต้เสียงรบกวนของการพูดคุย 2 ชนิดคือ การพูดคุยในภาษาไทย (ที่มีความคล้ายคลึงกับเสียง หรือคำเป้าหมาย) และในภาษาต่างประเทศ (ภาษาอังกฤษ ที่มีความต่างจากเสียง หรือคำเป้าหมาย แต่เป็นภาษาที่คนไทยมีความคุ้นเคย)

2. วัตถุประสงค์ของการวิจัย

1) เพื่อศึกษาและวิเคราะห์การรับรู้เสียงวรรณยุกต์ (คำพยางค์เดียว) ในภาษาไทย และความสับสนทางเสียงที่เกิดขึ้นภายใต้เสียงรบกวนของการพูดคุยในภาษาไทย และในภาษาต่างประเทศ

2) เพื่อเปรียบเทียบผลวิจัยที่ได้กับผลวิจัยของการรับรู้เสียงวรรณยุกต์ในภาษาไทย และความสับสนทางเสียงที่เกิดขึ้นภายใต้เสียงรบกวนประเภท white noise ในงานวิจัยก่อนหน้านี้

3. ทฤษฎีที่เกี่ยวข้อง

งานวิจัยชิ้นนี้มีที่มาและความเกี่ยวข้องกับทฤษฎี และแนวคิดต่างๆ ได้แก่ ทฤษฎีและแนวคิดด้านการรับรู้เสียงพูด (Theories of Speech Perception) (อาทิ Strange (1995), Laas (1996), Pisoni & Remez (2004), Smith (2009), Hardcastle et al. (2010) และ Cutler (2012)) ทฤษฎีด้านกลสัทศาสตร์และการรับรู้เสียงวรรณยุกต์ (production and perception of tones) (อาทิ Laas (1996) และ Hardcastle et al. (2010)) และเมทริกซ์ความสับสนทางการรับรู้เสียง (phoneme confusion matrix) (Miller & Nicely, 1955; Benkí, 2003; McLoughlin, 2008)

ผู้วิจัยขอหยิบยกเฉพาะประเด็นที่เกี่ยวข้องโดยตรงกับงานวิจัย ดังต่อไปนี้

3.1 การรับรู้เสียงพูดในเสียงรบกวน

ตามที่ได้กล่าวไปแล้วว่า งานวิจัยจำนวนมากชี้ให้เห็นว่าการรับรู้เสียงพูดของมนุษย์มีกระบวนการที่ต่างกันเมื่อประมวลผลเสียงพูดที่ปราศจากเสียงรบกวน และเสียงพูดที่อยู่ในเสียงรบกวน (Cutler, 2012) การศึกษาและงานวิจัยในประเด็นนี้ส่วนใหญ่เป็นการศึกษาที่ทำกับภาษาในประเทศแถบตะวันตก เช่น ภาษาอังกฤษ ภาษาดัตช์ ส่วนในภาษาไทยงานวิจัยในประเด็นนี้มีจำนวนจำกัด และส่วนหนึ่งเป็นงานวิจัยที่ผู้วิจัยและคณะได้ดำเนินการไปแล้ว ซึ่งเป็นการรับรู้ในระดับหน่วยเสียงและคำ ที่ใช้เสียงรบกวนประเภทพลังงาน ได้แก่ white noise (Onsuwan et al., 2012; Onsuwan et al., 2013; Tantibundhit & Onsuwan, 2015)

การศึกษาจำนวนมากแสดงให้เห็นว่าเสียงรบกวนประเภทพลังงานจะส่งผลต่อกระบวนการรับรู้เสียงของมนุษย์ที่แตกต่างจากเสียงรบกวนประเภทข้อมูลภาษา (Brungart, 2001; Simpson & Cooke, 2005; Cutler, 2012) ทั้งนี้ไม่ว่าจะเป็นเสียงรบกวนประเภทใดก็ตามเมื่อระดับ SNR ต่ำลง ความถูกต้องในการรับรู้เสียงจะลดลง (ทั้งนี้ ค่าอัตราส่วนที่มากกว่า 1:1 หรือ

มากกว่า 0dB จะมีสัญญาณเสียงมากกว่าเสียงรบกวน) (Miller, 1947) การทดลองที่ใช้เสียงรบกวน (masker) มักเป็นการศึกษาระดับการรับรู้ของสิ่งเร้าเป้าหมาย ในระดับของเสียง คำ วลี หรือประโยคภายใต้เสียงรบกวนต่างประเภทกัน หรือเมื่อมีระดับ SNR ที่ต่างกัน โดยจะใช้ระดับการรับรู้ที่ปราศจากเสียงรบกวน (clear, clean or quiet) (ในแต่ละการทดลอง) เป็นเกณฑ์หลัก (base line) การให้คำตอบของอาสาสมัครอาจเป็นได้ในรูปแบบของตัวเลือก หรือการให้อาสาสมัครพูด หรือเขียนตอบตามที่ได้ยิน ทั้งนี้ขึ้นกับวัตถุประสงค์ของงานวิจัย

ผลการศึกษาส่วนใหญ่มีความสอดคล้องกันในข้อสรุปที่ว่าเสียงรบกวนประเภทข้อมูลภาษา (โดยเฉพาะอย่างยิ่ง เสียงของการพูดคุย) ส่งผลกระทบต่อการประมวลผลทางเสียงในระดับสูงขึ้นไป (cognitive and linguistic) ดังนั้นปัจจัย และอิทธิพลหลักๆ ที่มีการนำมาศึกษาสามารถสรุปได้ ดังนี้

1) ความคล้ายคลึงของเสียงของการพูดคุยกับเสียงเป้าหมาย (match or unmatch) เช่น เพศที่เหมือนหรือต่างของเสียงพูดคุยกับเสียงเป้าหมาย (Brungart, 2001) ชนิดภาษาของเสียงพูดคุยที่เป็นภาษาเดียวกันหรือต่างกับเสียงเป้าหมาย (Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007; Calandruccio et al., 2010; Brouwer et al., 2012) และความคล้ายคลึงของความหมายของเนื้อความ (semantic content) ของเสียงพูดคุยกับเสียงเป้าหมาย (Brouwer et al., 2012)

2) รูปแบบของเสียงของการพูดคุย เช่น การพูดคุยที่ชัดถ้อยชัดคำหรือการพูดคุยแบบสนทนาทั่วไป (Calandruccio et al., 2010) และจำนวนของผู้พูดในเสียงของการพูดคุย (Simpson & Cooke, 2005; Garcia Lecumberri & Cooke, 2006; Van Engen & Bradlow, 2007)

3) ภูมิหลังทางภาษาของผู้ฟัง (listener's language background) เช่น ความคุ้นเคย หรือความเป็นเจ้าของภาษาของผู้ฟังต่อเสียงเป้าหมาย (Garcia Lecumberri & Cooke, 2005; Cutler et al., 2008; Brouwer et al. 2012)

ในประเด็นเรื่องเพศที่เหมือนหรือต่างของเสียงพูดคุยกับเสียงเป้าหมาย บรุนการ์ท (2001) ศึกษาการรับรู้เสียงในระดับวลี (ตัวเลข และชื่อเรียกสีในภาษาอังกฤษ) โดยใช้ 1) เสียงรบกวนประเภทพลังงาน Gaussian noise (SNR ในช่วง 15dB ถึง–18 dB แต่ละช่วงห่างกัน 3dB) และ speech-shaped noise (SNR ในช่วง 0 dB ถึง–21 dB แต่ละช่วงห่างกัน 3 dB) 2) เสียงรบกวนประเภทข้อมูลภาษา ได้แก่ เสียงของการพูดคุย (SNR ในช่วง 15dB ถึง–12 dB แต่ละช่วงห่างกัน 3dB) ในส่วนของเสียงของการพูดคุยเป็นแบบที่มีเสียงพูดคนเดียว (single talker) ใน 2 รูปแบบ คือ เสียงของการพูดคุยที่เพศของผู้พูดเหมือนกับเสียงเป้าหมาย และเสียงของการพูดคุยที่เพศของผู้พูดต่างกับเสียงเป้าหมาย ผลการทดลองพบว่าความถูกต้องในการรับรู้เสียงจะลดลงเมื่อ SNR ลดลงทั้งในเสียงรบกวนประเภทพลังงาน และประเภทข้อมูลภาษา อย่างไรก็ดีรูปแบบและความมากน้อยของการถดถอยของการรับรู้เสียงจะต่างกันไปในเสียงรบกวน 2 ประเภท นอกจากนี้

พบว่าภายใต้เสียงของการพูดคุย ระดับการรับรู้จะต่ำที่สุดถ้าเสียงของการพูดคุยเป็นคนๆ เดียวกับเสียงเป้าหมาย และจะสูงที่สุดเมื่อเสียงของการพูดคุย และเสียงเป้าหมายมาจากผู้พูดต่างเพศกัน

ในประเด็นของอิทธิพลของชนิดภาษา (เสียงพูดคุยที่เป็นภาษาเดียวกับหรือต่างกับเสียงเป้าหมาย) แวน เองเกิ้น และ แบรดโล (2007) ศึกษาการรับรู้ประโยคอังกฤษของอาสาสมัครที่พูดภาษาอังกฤษในเสียงรบกวนที่เป็นเสียงของการพูดคุยในภาษาอังกฤษ และภาษาจีน ในระดับ SNR 5dB 0dB และ–5dB โดยเสียงของการพูดคุยทั้งใน 2 ภาษา มี 2 แบบด้วยกันคือ แบบที่มีผู้พูด 2 คน และที่มีผู้พูด 6 คน ผลการทดลองพบว่าภายใต้เสียงของการพูดคุยของทั้ง 2 ภาษาที่มีจำนวนผู้พูดมากกว่า (คือ 6 คน) การรับรู้ประโยคจะด้อยลงมากกว่า และพบว่าภายใต้เสียงของการพูดคุยที่มีผู้พูด 2 คนเท่านั้นที่ทำให้การรับรู้ประโยคภายใต้เสียงของการพูดคุยในภาษาอังกฤษด้อยกว่าเมื่อเทียบกับเสียงของการพูดคุยในภาษาจีน กล่าวโดยสรุปคือ การรับรู้จะมีลักษณะด้อยลงเมื่อภาษาของเสียงเป้าหมายมีความคล้ายคลึงกับเสียงรบกวนของการพูดคุย

อิทธิพลของชนิดภาษาของเสียงพูดคุยในลักษณะเดียวกันปรากฏในงานของ คาแลนดรูชิโอ และคณะ (2010) ที่ศึกษาการรับรู้ประโยคในภาษาอังกฤษของอาสาสมัครที่พูดภาษาอังกฤษภายใต้เสียงของการพูดคุยแบบมีผู้พูดสองคนใน 2 ภาษาคือ ภาษาอังกฤษ และภาษาโครเอเชียที่ระดับ SNR –3dB และ–5dB (experiment 1) ผลการทดลองพบว่าอาสาสมัครมีการรับรู้ในระดับที่ดีกว่าภายใต้เสียงพูดคุยภาษาโครเอเชีย มากกว่าเสียงพูดคุยในภาษาอังกฤษ นอกจากนี้ยังพบว่า ไม่ว่าเสียงของการพุดคุยจะเป็นการพูดคุยที่ชัดถ้อยชัดคำ หรือการพูดคุยแบบสนทนาทั่วไป ไม่ส่งผลต่อระดับการรับรู้ประโยคเป้าหมาย แต่ประโยคเป้าหมายที่พูดในรูปแบบชัดถ้อยชัดคำทำให้การรับรู้สูงกว่าประโยคเป้าหมายที่พูดในการพูดคุยแบบสนทนาทั่วไป

นอกจากนี้ บราวเวอร์และคณะ (2012) ศึกษาการรับรู้ประโยคภาษาอังกฤษของอาสาสมัครที่พูดภาษาอังกฤษเป็นภาษาเดียว (monolingual) และอาสาสมัครชาวดัตช์ที่รู้ภาษาอังกฤษในระดับดี (bilingual) ภายใต้เสียงรบกวนที่เป็นเสียงของการพูดคุยในภาษาอังกฤษ และในภาษาดัตช์ ที่ระดับ SNR –1dB –3dB และ–5dB ผลการทดลองแสดงให้เห็นถึงอิทธิพลของชนิดภาษาของเสียงพูดคุยตามที่ได้กล่าวก่อนหน้านี้ กล่าวคืออาสาสมัครที่พูดภาษาอังกฤษ และอาสาสมัครชาวดัตช์มีระดับการรับรู้ประโยคในภาษาอังกฤษที่ดีกว่าภายใต้เสียงรบกวนที่เป็นเสียงพูดคุยในภาษาดัตช์ ดังนั้นข้อสรุปที่สำคัญในงานวิจัยนี้คือความคล้ายคลึงของเสียงเป้าหมายกับเสียงรบกวนมีอิทธิพลสูงกว่าปัจจัยด้านความคุ้นเคยของภาษา เนื่องจากอาสาสมัครชาวดัตช์มีการรับรู้ที่สูงในการได้ยินประโยคภาษาอังกฤษ (ภาษาที่ 2) ที่เป็นเป้าหมาย ถึงแม้ว่าเสียงรบกวนเป็นภาษาดัตช์ (ภาษาแม่) ที่ตัวเองคุ้นเคยกว่า

ในแง่อิทธิพลของจำนวนผู้พูดในเสียงของการพูดคุย ซิมพ์สันและคุกค์ (2005) ศึกษาการรับรู้เสียงพยัญชนะในภาษาอังกฤษ (VCV) ของอาสาสมัครที่พูดภาษาอังกฤษภายใต้เสียงรบกวนประเภทพลังงาน babble-modulated noise และในเสียงรบกวนประเภทข้อมูลภาษาแบบเสียงของ

การพูดคุยในภาษาอังกฤษ ที่มี SNR ในระดับเดียวกันคือ –6dB โดยตัวแปรที่สำคัญคือจำนวนของผู้พูดในเสียงของการพุดคุย ผลการทดลองพบว่าเมื่อจำนวนของผู้พูดในเสียงของการพูดคุยเพิ่มมากขึ้นระดับการรับรู้เสียงจะด้อยลง ที่เห็นได้ชัดเจนคือการด้อยลงเมื่อจำนวนผู้พูดอยู่ระหว่าง 2 ถึง 6 คน ระดับการด้อยลงจะมีลักษณะแบบค่อยเป็นค่อยไป จากนั้น (เป็นที่น่าสนใจว่า) การด้อยลงเมื่อจำนวนผู้พูด 8 คนขึ้นไปจนถึง 128 คนจะมีผลไม่ต่างกันมากนัก และเนื่องจากงานวิจัยนี้ใช้เสียงรบกวนที่เทียบเคียงกันได้ชัดเจน ต่างกันที่อันหนึ่งเป็นประเภทพลังงาน babble-modulated noise อีกอันหนึ่งเป็นประเภทข้อมูลภาษาได้แก่ natural babble ผลการทดลองของงานนี้จึงยืนยันอิทธิพลจากข้อมูลภาษาของเสียงรบกวน และสรุปว่ารูปแบบและความมากน้อยของการถดถอยของการรับรู้ในเสียงรบกวนประเภทพลังงานจะมีลักษณะลดลงอย่างสม่ำเสมอ (gradual fall) ในขณะที่การถดถอยของการรับรู้ในเสียงรบกวนประเภทข้อมูลภาษาจะมีลักษณะที่ลดลง แต่ไม่เป็นรูปแบบชัดเจน (non-monotonic function)

ในแง่ความคุ้นเคย หรือความเป็นเจ้าของภาษาของผู้ฟังต่อเสียงเป้าหมาย ลีคัมเบร์รี่และคุค (2006) ศึกษาการรับรู้เสียงพยัญชนะภาษาอังกฤษ (VCV) ของอาสาสมัครที่พูดภาษาอังกฤษ และชาวสเปน (มีความคุ้นเคยกับภาษาอังกฤษ) โดยใช้เสียงรบกวนประเภทพลังงาน speech-shaped noise และเสียงรบกวนประเภทข้อมูลภาษาแบบเสียงรบกวนของการพูดคุย ใน 3 รูปแบบคือ เสียงของการพูดคุยในภาษาอังกฤษแบบจำนวนผู้พูด 8 คน เสียงของการพูดคุยแบบจำนวนผู้พูด 1 คนในภาษาอังกฤษ และเสียงของการพูดคุยแบบจำนวนผู้พูด 1 คน ในภาษาสเปน ทั้ง 3 รูปแบบของเสียงของการพูดคุยอยู่ที่ระดับ SNR 0dB ผลการทดลองพบว่ากลุ่มอาสาสมัครทั้ง 2 กลุ่มมีการรับรู้ที่ต่ำลงภายใต้เสียงรบกวนในทุกประเภทที่ SNR ลดลง และทำได้ดีที่สุดในเสียงของการพูดคุยแบบจำนวนผู้พูด 1 คน ด้อยที่สุดในเสียงภายใต้การพูดคุยแบบจำนวนผู้พูด 8 คน (อิทธิพลเรื่องจำนวนของผู้พูดในเสียงของการพูดคุย) อาสาสมัครที่พูดภาษาอังกฤษจะมีการรับรู้ที่ดีกว่าเมื่อเสียงของการพูดคุยเป็นภาษาสเปน (อิทธิพลของชนิดภาษาของเสียงพูดคุย) นอกจากนี้ผลจากเมทริกซ์ความสับสนทางการรับรู้เสียงสรุปได้ว่าอาสาสมัครที่ไม่ใช่เจ้าของภาษาได้รับผลกระทบจากเสียงรบกวนทั้ง 2 ประเภทสูงกว่าเจ้าของภาษา ซึ่งอาจเนื่องมาจากอยู่ในช่วงของการเรียนรู้ระบบหน่วยเสียงในภาษาอังกฤษ เช่นเดียวกันนี้ คัทเลอร์และคณะ (2008) ศึกษาการรับรู้เสียงพยัญชนะภาษาอังกฤษ (VCV) ของอาสาสมัครที่พูดภาษาอังกฤษ และชาวดัตช์ (ที่มีความคุ้นเคยกับภาษาอังกฤษ) ภายใต้เสียงของการพุดคุยในภาษาอังกฤษของผู้พูดจำนวน 8 คน SNR 0dB และพบว่ากลุ่มชาวดัตช์มีระดับการรับรู้ที่ต่ำกว่า กล่าวคือได้รับผลกระทบจากเสียงรบกวนมากกว่ากลุ่มที่มีภาษาแม่เป็นภาษาอังกฤษ

3.2 การศึกษาทางกลสัทศาสตร์ และการรับรู้เสียงวรรณยุกต์ในภาษาไทย

ภาษาไทยเป็นภาษาที่มีวรรณยุกต์ (tonal language) กล่าวคือ รูปแบบของระดับเสียง (pitch) ที่แตกต่างกันในระดับพยางค์และคำนำมาซึ่งความหมายของคำที่แตกต่างกันได้ เสียงวรรณยุกต์ในภาษาไทยมีทั้งสิ้น 5 หน่วยเสียง ได้แก่ เสียงสามัญ (กลางระดับ) (mid) เสียงเอก (ต่ำระดับ) (low) เสียงโท (สูงตก) (falling) เสียงตรี (สูงระดับหรือกลางขึ้น) (high) และเสียงจัตวา (ต่ำขึ้น) (rising) (Abramson, 1976; Tingsabadh & Abramson, 1993; ธีระพันธ์ เหลืองทองคำ, 2554)

ตัวบ่งชี้ทางกลสัทศาสตร์ (acoustic cue) ของเสียงวรรณยุกต์ในภาษาไทยมีหลายค่า แต่ค่าที่สำคัญที่สุด และมีอิทธิพลต่อการรับรู้เสียงวรรณยุกต์ในภาษาไทย ได้แก่ ค่าความถี่มูลฐาน (ของเสียงก้องในพยางค์) (fundamental frequency (F0)) ซึ่งรวมถึงความสูงของระดับเสียง รูปร่างของเส้นความถี่ ค่าพิสัย รูปแบบของช่วงเริ่มต้น และช่วงท้าย ทั้งนี้ตัวบ่งชี้ทางกลสัทศาสตร์ที่เกี่ยวข้องอื่นๆ คือ ความยาวของเสียงสระ และความดังของเสียง (Abramson, 1962, 1975) อย่างไรก็ดีค่าความถี่มูลฐานของเสียงวรรณยุกต์ในคำพูดเดี่ยว และคำพูดต่อเนื่องมีความแตกต่างกันค่อนข้างชัดเจน (Tingsabadh & Deeprasert, 1997)

ข้อมูลทางกลสัทศาสตร์ของเสียงวรรณยุกต์ในภาษาไทย ได้มีปรากฏในงานวิจัยจำนวนหนึ่ง และสามารถนำมาใช้เป็นข้อมูลอ้างอิงในการสร้างและประเมินผลการทดสอบการรับรู้ทางเสียงได้เป็นอย่างดี (Abramson, 1962, 1975; Anivan, 1985; Potisuk, et al.,1994; Tingsabadh & Deeprasert, 1997; Teeranon, 2007; Thepboriruk, 2010; ปิยฉัตร ปานโรจน์, 2534; กุสุมา นะสานี, 2545; ธีระพันธ์ เหลืองทองคำ, 2554) และเป็นที่ทราบกันดีว่ารูปแบบทางเสียงของวรรณยุกต์ในภาษาไทยได้มีการเปลี่ยนแปลงอย่างต่อเนื่อง (change in progress) จากที่ได้มีการบันทึกไว้ในการศึกษาดั้งเดิมของเอแบรมสันในปี ค.ศ. 1962 (Teeranon, 2007; Thepboriruk, 2010) อย่างไรก็ดีงานวิจัยชิ้นนี้มิได้มีจุดมุ่งหมายที่จะศึกษาประเด็นของการเปลี่ยนแปลงทางเสียงวรรณยุกต์ ดังนั้นผู้วิจัยเลือกใช้เสียงจากผู้บอกภาษาซึ่งอยู่ในวัยกลางคน และมีรูปแบบของเสียงวรรณยุกต์ที่สอดคล้องกับงานวิจัยของกาญจนา เทพโพธิรักษ์ (2010) (รายละเอียดในหัวข้อ 5.4.1)

งานวิจัยจำนวนหนึ่งศึกษาการรับรู้เสียงวรรณยุกต์ในภาษาไทย โดยสรุปได้ว่าการรับรู้เสียงวรรณยุกต์ในภาษาไทยของคนไทย (ที่มีการได้ยินปกติ) อยู่ในระดับโดดเด่นและดีมาก (robust) (Abramsom, 1962, 1972, 1975; Gandour & Dardaranarda, 1983; Burnham & Francis, 1997; Tingsabadh & Deeprasert, 1997) เช่นเดียวกับ ข้อค้นพบในการรับรู้เสียงวรรณยุกต์ในภาษาอื่นๆ เช่น ภาษาจีนกลาง (Whalen & Xu, 1992) และจากลักษณะทางกลสัทศาสตร์โดยเฉพาะความคล้ายคลึงกันของรูปร่างของค่าความถี่มูลฐาน (F0 contour) ของเสียงวรรณยุกต์ในภาษาไทยได้มีการคาดการณ์ว่าจะเกิดความสับสนสูงในการรับรู้เสียงของวรรณยุกต์สามัญ และเอก ซึ่งสอดคล้องกับผลการทดสอบการรับรู้เสียงวรรณยุกต์ของ แอแบรมสัน (1975)

แกนดอร์ และรจนา ทรรทรานนท์ (1983) และผลการทดสอบของณัฐกร ทับทอง (2001) อย่างไรก็ดีการทดสอบของ แอแบรมสัน (1972, 1975) และ แกนดอร์ และรจนา ทรรทรานนท์ (1983) ใช้คำทดสอบเพียง 5 ถึง 17 คำ และถึงแม้จะมีการรายงานผลแบบเมทริกซ์ความสับสนทางการรับรู้เสียง แต่เป็นการทดสอบที่มีขนาดเล็ก และรายการคำที่ใช้ในการทดสอบมีจำนวนน้อยจึงมีความจำกัดค่อนข้างมาก

การทดสอบการรับรู้เสียงวรรณยุกต์ของ เบิร์นแฮม และฟรานซิส (1997) เป็นลักษณะการแยกแยะความแตกต่าง (discrimination test) ของคำทดสอบ 5 คำ (จับคู่กันได้ 10 คู่ทดสอบ) โดยอาสาสมัครทั้งเด็กและผู้ใหญ่ที่ใช้ภาษาไทยเป็นภาษาแม่ และอาสาสมัครชาวออสเตรเลีย ส่วนการทดสอบของณัฐกร ทับทอง (2001) มีขนาดไม่ใหญ่นัก และเป็นการออกแบบเพื่อเน้นการนำข้อมูลไปใช้ในด้านการพัฒนาระบบการรู้จำเสียงวรรณยุกต์ในภาษาไทย (tone recognition system) ในทางวิศวกรรมไฟฟ้าและคอมพิวเตอร์

แอแบรมสัน (1975) ใช้เทคนิคการสังเคราะห์เสียงพูด และคำ (synthesized monosyllabic word) โดยมีการปรับเปลี่ยนด้านความดังของเสียง ร่วมไปกับรูปร่างของค่าความถี่มูลฐาน (F0 contour) แต่ไม่มีการใช้เสียงรบกวน และพบว่าการฟังเข้าใจ (intelligibility score) ของอาสาสมัคร 34 คน อยู่ในระดับที่สูงมากคือร้อยละ 96.1 ความผิดพลาดและความสับสนในการรับรู้เสียงวรรณยุกต์เกิดขึ้นกับวรรณยุกต์สามัญ และตรี (รับรู้วรรณยุกต์สามัญว่าเป็นตรี) ซึ่งความผิดพลาดนี้จะน้อยลงเมื่อความดังของเสียงเพิ่มขึ้น นอกจากนี้ยังมีการสับสนระหว่างวรรณยุกต์สามัญ และเอก (รับรู้วรรณยุกต์เอกว่าเป็นสามัญ)

แกนดอร์ และรจนา ทรรทรานนท์ (1983) ศึกษาความสับสนในการรับรู้เสียงวรรณยุกต์ในภาษาไทยในคนปกติ และผู้มีภาวะเสียการสื่อภาษา (aphasic patient) โดยใช้เสียงพูดจริง (ในชุด A และ B) และเสียงพูดสังเคราะห์ (ชุด C) ไม่มีการใช้เสียงรบกวน คะแนนการฟังเข้าใจของคนไทยปกติ (อาสาสมัคร 1 คน) ในชุด A อยู่ที่ ร้อยละ 93.75 ถึง 100 และในชุด B และ C อยู่ที่คะแนนเต็ม ส่วนความผิดพลาด และความสับสนในการรับรู้เสียงวรรณยุกต์เกิดขึ้นกับวรรณยุกต์สามัญ และเอก (รับรู้วรรณยุกต์สามัญว่าเป็นเอก แต่ไม่มีการรับรู้เอกเป็นสามัญ) และกับวรรณยุกต์โท และตรี (รับรู้วรรณยุกต์โทว่าเป็นตรี และตรีเป็นโท) นอกจากนี้ยังคาดว่าวรรณยุกต์จัตวาน่าจะเป็นวรรณยุกต์ที่ชัดเจน (salient) มากที่สุด การศึกษาประเด็นนี้เป็นข้อมูลเพื่อช่วยในด้านการวินิจฉัยและบำบัดฟื้นฟูผู้มีภาวะเสียการสื่อภาษา

จุฑามณี อ่อนสุวรรณ และคณะ (2012) ศึกษาความสับสนในการรับรู้เสียงวรรณยุกต์ในภาษาไทย โดยมีการสร้างและพัฒนารายการคำที่มีขนาดใหญ่และหลากหลาย (40 คู่ 80 คำ) (โดยดัดแปลงจากหลักการสร้างชุดทดสอบในภาษาอังกฤษ (House et al., 1965) และภาษาจีน (Li et al., 2000)) ใช้เสียงรบกวนแบบ (additive white Gaussian noise (AWGN)) ใน 4 ระดับ SNR ได้แก่ –6dB –12dB –18dB และ –24dB และมีการประเมินเมทริกซ์ความสับสนทางการ

รับรู้เสียง ผลการวิจัยพบว่า การฟังเข้าใจ ของอาสาสมัคร 34 คนอยู่ในระดับที่ลดลงเมื่อระดับ SNR ลดลงจาก ร้อยละ 88.5, 89.9, 88.0 และ 64.4 ตามลำดับ และพบว่าเสียงวรรณยุกต์ตรีเป็นเสียงที่มีความสับสนมากที่สุด ส่วนวรรณยุกต์เอกสับสนน้อยที่สุด และมีความสับสนแบบทิศทางเดียว (unidirectional) ในวรรณยุกต์ตรี (เป็นเอก) ส่วนความสับสน 2 ทิศทาง (ในระดับต่างๆ กัน) เกิดขึ้น ใน 9 คู่วรรณยุกต์ (จากที่เป็นไปได้ทั้งหมด 10 คู่) และคู่ที่เด่นชัดที่สุดคือระหว่างวรรณยุกต์เอกกับโท และสามัญกับตรี (อ้างอิงจาก Table 3 ใน Onsuwan et al. (2012)) นอกจากนี้คณะผู้วิจัยยังคาดว่ามีปัจจัยนอกเหนือจากความคล้ายคลึงกันทางด้านค่าความถี่มูลฐานที่ทำให้เกิดความสับสนในการรับรู้เสียงวรรณยุกต์คือ ความดัง และการเปลี่ยนแปลงทางระบบเสียงในภาษาข้ามสมัย (diachronic stability) (โดยเฉพาะของวรรณยุกต์ตรี) การศึกษานี้เป็นข้อมูลพื้นฐาน และแนวทางในการทำการทดสอบการฟังเข้าใจ (intelligibility test) ที่ใช้เสียงพูดในภาษาไทยภายใต้รูปแบบสถานการณ์เสียงต่างๆ

3.3 การรับรู้เสียงวรรณยุกต์ในภาษาจีนกลางในเสียงรบกวน

ภาษาจีนกลางเป็นภาษาที่มีวรรณยุกต์และได้มีการศึกษาในแง่มุมด้านสัทศาสตร์อย่างกว้างขวาง อย่างไรก็ดีงานวิจัยที่ศึกษาการรับรู้วรรณยุกต์ภาษาจีนกลางในเสียงรบกวนมีจำนวนจำกัด ลีและคณะ (2010) กล่าวว่าที่เป็นเช่นนี้อาจเป็นเพราะการรับรู้วรรณยุกต์ในภาษาจีนโดยทั่วไปมีความชัดเจนสูง

การศึกษาที่ใช้เสียงรบกวนส่วนใหญ่เป็นเสียงรบกวนประเภทพลังงาน เช่น งานของคองและเชง (2006) ที่ศึกษาการรับรู้วรรณยุกต์ในภาษาจีนในเสียงรบกวนประเภทพลังงาน ได้แก่ noise-excited vocoder speech และเสียงที่ผ่านการดัดแปลงในรูปแบบต่างๆ เช่น frequency-modulated และ whispered speech แต่การศึกษานี้เป็นลักษณะการวัดผลและด้านการพัฒนาระบบการรู้จำสัญญาณเสียง (speech recognition) นอกจากนี้ ลีและคณะ (2010) ศึกษาการรับรู้วรรณยุกต์ในภาษาจีนในเสียงรบกวนประเภทพลังงาน speech-shaped noise ในระดับ SNR 0dB –5dB –10 dB และ–15dB เปรียบเทียบระหว่างอาสาสมัครชาวจีน และอาสาสมัครที่ใช้ภาษาอื่นและกำลังศึกษาภาษาจีน ผลการทดลองพบว่าภายใต้เสียงรบกวนอาสาสมัครเจ้าของภาษามีการรับรู้ที่ดีกว่าและพบว่าระยะเวลาในการศึกษาภาษาจีน (ตามที่ได้ควบคุมในงานวิจัยนี้) ไม่ได้มีผลต่อระดับการรับรู้เสียงวรรณยุกต์

งานวิจัยหนึ่งชิ้นของ มิกซ์ดอร์ฟและคณะ (2005) เท่านั้นที่ศึกษาการรับรู้วรรณยุกต์ในภาษาจีนกลางของอาสาสมัครชาวจีน ที่ไม่มีเสียงรบกวน เปรียบเทียบกับภายใต้เสียงรบกวนที่เป็นเสียงของการพูดคุยที่ระดับ SNR –3dB –6dB –9dB และ–12dB รวมไปถึงเสียงรบกวนประเภทพลังงาน ได้แก่ devoiced (หนึ่งระดับ) และ amplitude-modulated noise (หนึ่งระดับ) การทดสอบการรับรู้เสียงทำไปควบคู่กับการใช้วิดีโอภาพเคลื่อนไหวของใบหน้าในขณะออกเสียง (visual cue) ผลการทดลองพบว่า ระดับความถูกต้องของการรับรู้วรรณยุกต์เมื่อไม่มีเสียงรบกวน

อยู่ที่เกือบ 100 เปอร์เซ็นต์ และมีระดับลดลงเมื่อมีเสียงรบกวนทั้งสองประเภท นอกจากนี้พบว่าภายใต้เสียงรบกวนของการพูดคุยเท่านั้นที่ visual cue จะเข้ามาเสริมการรับรู้ได้ เมื่อ SNR ต่ำลง (ไม่พบการเสริมนี้กับเสียงรบกวนประเภทพลังงาน)

จะเห็นได้ว่าการศึกษาการรับรู้เสียงวรรณยุกต์ในเสียงรบกวนทั้งในภาษาไทยและภาษาจีนมีอยู่จำกัดมาก โดยเฉพาะอย่างยิ่งการศึกษาการรับรู้ในเสียงรบกวนของการพูดคุย ดังนั้นผู้วิจัยจึงต้องการศึกษาการรับรู้เสียงวรรณยุกต์และความสับสนทางเสียงที่เกิดขึ้นภายใต้เสียงรบกวนของการพูดคุยสองชนิดคือ การพูดคุยในภาษาไทย (ที่มีความคล้ายคลึงกับเสียง หรือคำเป้าหมาย) และในภาษาต่างประเทศ (ภาษาอังกฤษ ที่มีความต่างจากเสียง หรือคำเป้าหมาย แต่เป็นภาษาที่คนไทยมีความคุ้นเคย) โดยสมมติฐานหลักของงานนี้คือ การรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวนของการพูดคุยในภาษาไทยจะมีระดับที่ต่ำกว่า ภายใต้เสียงรบกวนของการพูดคุยในภาษาอังกฤษ

4. ประโยชน์ที่จะได้รับจากการวิจัย

1) เพิ่มองค์ความรู้ด้านการรับรู้เสียงวรรณยุกต์ในภาษาไทย และความสับสนทางเสียงที่เกิดขึ้น เพื่อนำไปเปรียบเทียบกับการรับรู้เสียงประเภทอื่น หรือเสียงวรรณยุกต์ในภาษาอื่น เช่น ภาษาจีนกลาง

2) เพิ่มองค์ความรู้ด้านการรับรู้เสียงวรรณยุกต์ในภาษาไทย และความสับสนทางเสียงที่เกิดขึ้นในสถานการณ์ที่มีเสียงรบกวน เพื่อนำไปใช้สร้างแบบจำลองในการเพิ่มสมรรถนะเสียงวรรณยุกต์ของภาษาไทยในเครื่องมือ หรืออุปกรณ์ทางเทคโนโลยีทางเสียงพูด และภาษาต่อไป

3) เป็นแนวทางของการทำการทดสอบที่ใช้เสียงพูดในภาษาไทยในด้านการรับรู้เสียงพูด และการฟังเข้าใจในรูปแบบสถานการณ์เสียงต่าง ๆ ซึ่งในปัจจุบันยังมีอยู่จำกัด

5. วิธีการวิจัย

5.1 อาสาสมัคร

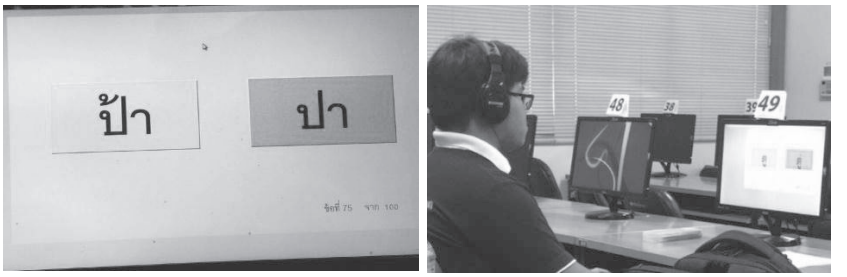
อาสาสมัครที่เข้าร่วมการทดสอบมีจำนวน 35 คน (หญิง 30 คน ชาย 5 คน) อายุระหว่าง 20-22 ปี ใช้ภาษาไทยเป็นภาษาแม่ เติบโตในสภาพแวดล้อมที่ใช้ภาษาไทยกลาง ไม่เป็นผู้พูด 2 ภาษาในภาษาต่างประเทศ และมีการใช้ภาษาและการได้ยินที่ปกติ

5.2 การเก็บรวบรวมข้อมูล

อาสาสมัครทำการทดสอบการรับรู้เสียงผ่านโปรแกรมที่สร้างขึ้นด้วย Netbeans ที่ใช้ภาษา Java (https://netbeans.org/) โดยรับฟังเสียงผ่านหูฟัง (Shure 240 และ 440) ข้อมูลการรับรู้จะเก็บไว้ในคอมพิวเตอร์ ส่วนเอกสารอื่นๆ อาสาสมัครกรอกข้อมูลบนกระดาษ

ระยะเวลาในขณะทำการทดสอบ 40-60 นาที จำนวนครั้งที่นัดอาสาสมัคร 1 ครั้ง การทดสอบทั้งหมดดำเนินการที่ห้องปฏิบัติการคอมพิวเตอร์ (ที่ปราศจากเสียงรบกวน) ที่คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์ (ศูนย์รังสิต) และปิดเครื่องปรับอากาศขณะทำการทดสอบ

ภาพที่ 1 ตัวอย่างภาพที่ปรากฏบนจอเพื่อให้อาสาสมัครเลือกตัวเลือกตามที่ได้ยิน จากตัวเลือกซ้ายหรือตัวเลือกขวา โดยใช้คีย์บอร์ด และอาสาสมัครขณะเข้าร่วมการทดสอบ



5.3 แบบทดสอบ

แบบทดสอบทั้งหมดประกอบไปด้วยคู่คำทดสอบ 700 ข้อ แบ่งเป็น 7 แบบทดสอบย่อย ในแต่ละข้ออาสาสมัครจะได้ยินคำเร่งเร้าเป้าหมาย (คำใดคำหนึ่งในแต่ละคู่) ที่มีลักษณะใดลักษณะหนึ่งดังนี้

5.3.1 คำเร่งเร้าเป้าหมายปราศจากเสียงรบกวนใด ๆ (clean speech)

จำนวน 40 ข้อ X 2 ตำแหน่ง = 80 ข้อ (หมายเหตุ 2 ตำแหน่ง หมายถึง การสลับการปรากฏของคำเร่งเร้าบนจอให้ปรากฏทั้งทางด้านซ้าย และด้านขวาของหน้าจอ)

5.3.2 คำเร่งเร้าเป้าหมายที่มีการปนเปื้อนด้วยเสียงรบกวนของการพูดคุยในภาษาไทย หรือภาษาอังกฤษ จากระดับ SNR 1 ใน 3 ระดับได้แก่ 0dB –5dB และ–10dB จำนวน 40 ข้อ X 3 ระดับเสียงรบกวน X 2 ภาษา X 2 ตำแหน่ง = 480 ข้อ (ก่อนการทดสอบจริงได้มีการทำการทดสอบนำร่องเพื่อให้แน่ใจว่าเสียงรบกวนทั้ง 3 ระดับซึ่งได้แก่ SNR 0dB –5dB และ–10dB เป็นระดับที่เหมาะสมและไม่ทำให้ผลการทดสอบมีระดับที่สูง (100%) และต่ำจนเกินไป)

5.3.3 คำเร่งเร้าที่เป็นคำลวง

คำพยางค์เดียว จำนวน 10 ข้อ X 2 ตำแหน่ง X 7 ครั้ง = 140 ข้อ ซึ่งปนด้วยสัญญาณเสียงรบกวนแบบ white noise ที่ระดับ –10dB เพื่อให้แตกต่างและเบี่ยงเบนความสนใจไปจากคำเร่งเร้าเป้าหมาย และไม่ก่อให้เกิดความเบื่อหน่าย

ในการทดสอบจะนำข้อทดสอบทั้ง 700 ข้อ มาสุ่มลำดับ และแบ่งการทดสอบเป็น 7 ส่วนย่อย ส่วนละ 100 ข้อ

5.4 การสร้างข้อมูลเสียง

การสร้างคำเร่งเร้าแบ่งเป็น 2 ส่วน ได้แก่ การสร้างคำเร่งเร้าเป้าหมาย (target stimuli) จำนวน 40 คู่ (80 คำ) และการสร้างเสียงรบกวน (masking noise) ดังรายละเอียดต่อไปนี้

5.4.1 การสร้างคำเร่งเร้า

ผู้วิจัยได้ดำเนินการตามงานวิจัยก่อนหน้านี้ (Onsuwan et al., 2012) กล่าวคือ คำเร่งเร้าเป้าหมายจำนวน 40 คู่ (80 คำ) เป็นคำคล้องจอง (rhyming word pair) (House et al., 1965) คำพยางค์เดียวชนิด CVTV ที่มีองค์ประกอบทางเสียงที่เหมือนกันยกเว้นเพียงหนึ่งองค์ประกอบในที่นี้คือเสียงวรรณยุกต์ เช่น ปู – ปู่ ตามที่แสดงในภาพที่ 3

การออกแบบสร้างคู่คำเร่งเร้าเป้าหมายเริ่มจากรวบรวมรายการคำจำนวนมากที่มีลักษณะเข้าข่ายที่จะเป็นคู่คำเร่งเร้าเป้าหมาย จากนั้นคัดเลือกคำเร่งเร้าเป้าหมายโดยให้ครอบคลุมหน่วยเสียงวรรณยุกต์ทั้งหมดในภาษาไทย (5 หน่วยเสียง) และเป็นคำที่ใช้ทั่วไป ทั้งนี้แต่ละคู่คำเร่งเร้าเป้าหมายเป็นการเปรียบเทียบวรรณยุกต์เป็นคู่ๆ ไป การจับคู่เปรียบเทียบวรรณยุกต์ที่ไม่ซ้ำกันได้คู่เปรียบเทียบวรรณยุกต์ทั้งหมด 10 คู่ ดังนี้

สามัญ – เอก สามัญ – โท สามัญ – ตรี สามัญ – จัตวา เอก – โท
เอก – ตรี เอก – จัตวา โท – ตรี โท – จัตวา ตรี – จัตวา

คู่เปรียบเทียบวรรณยุกต์หนึ่งๆ จะมีคู่คำเร่งเร้าเป้าหมายทั้งสิ้น 3 คู่ เช่น คู่ 3 คู่ คือ ปู – ปู่ ตู – ตู่ และ กอ – ก่อ เปรียบเทียบการรับรู้ และความสับสนทางเสียงที่อาจเกิดขึ้นระหว่างเสียงวรรณยุกต์สามัญ และเอก ดังนั้นเสียงวรรณยุกต์แต่ละหน่วยเสียงจะมีปรากฏในคู่คำทดสอบทั้งหมด 16 คู่ นอกจากนี้คำทั้ง 2 คำในแต่ละคู่คำทดสอบจะเป็นคำที่คาดว่าจะมีความถี่ในการปรากฏในภาษาใกล้เคียงกัน และหากเป็นคำสแลงก็จะเป็นคำสแลงทั้งสองคำเพื่อหลีกความโน้มเอียงที่ไม่ต้องการ (bias) ที่อาจเกิดขึ้น (ทั้งนี้ คำว่า เก๋ ค้า ช้า มีปรากฏคำละ 2 ครั้ง)

ค่าเฉลี่ยของค่าความถี่มูลฐานของคำเร่งเร้าเป้าหมาย (วัดจากส่วนของสระ) ของเสียงวรรณยุกต์ทั้ง 5 แสดงไว้ในภาพที่ 2 มีลักษณะโดยรวมสอดคล้องกับค่าเฉลี่ยของค่าความถี่มูลฐานของเสียงวรรณยุกต์ในภาษาไทยในสมัยปัจจุบันในงานวิจัยก่อนหน้านี้ โดยเฉพาะอย่างยิ่งในงานของกาญจนา เทพบริรักษ์ (2010)

นอกจากนี้ยังมีคำลวงซึ่งเป็นคำพยางค์เดี่ยว CVT(V)(C) 10 คู่ จำนวน 20 คำ เช่น ยา – รา ปรากฏปะปนในการทดสอบเพื่อหลีกเลี่ยงความซ้ำซากและเพื่อเบี่ยงเบนความสนใจผู้เข้าทดสอบจากคู่คำเร่งเร้าเป้าหมาย

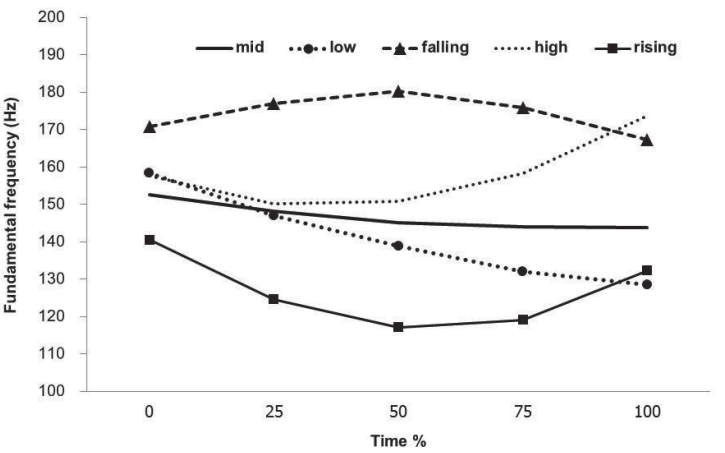
ผู้บอกภาษาเป็นชายไทยอายุ 37 ปี เป็นผู้อ่านรายการคำเสียงวรรณยุกต์ที่สร้างขึ้นในกรอบประโยคที่กำหนดขึ้น (carrier sentence) คือ “เจอ คำเป้าหมาย เออ ดีจัง” ซึ่งกรอบประโยคประกอบด้วยคำที่มีวรรณยุกต์สามัญทั้งหมด รายการคำสำหรับอ่านนี้ได้รับการอ่าน 5 ครั้งด้วยความเร็วและจังหวะที่ปกติไม่ช้าหรือเร็วจนเกินไป และบันทึกเสียงที่ความถี่ (sampling rate)

44.1kHz ด้วยโปรแกรม Adobe Audition เสียงที่ได้จากการบันทึกถูกนำมาวิเคราะห์และตัดต่อด้วยโปรแกรมวิเคราะห์เสียง Praat คำทดสอบแต่ละคำได้ผ่านการประเมิน และการคัดเลือกด้วยการฟังจากผู้ชำนาญการด้านภาษา รวมไปถึงการวิเคราะห์ด้วยคลื่นเสียง (waveform) และภาพแถบคลื่นเสียง (spectrogram) เพื่อนำไปใช้เป็นคำเร่งเร้า ทั้งนี้การตัดเสียงในแต่ละจุด (splice point) จะตัดที่จุดตัดที่ศูนย์ (zero crossing) เพื่อให้คลื่นเสียงมีความต่อเนื่องและเป็นธรรมชาติมากที่สุด คำเร่งเร้าทั้งหมดมีความยาวโดยเฉลี่ยคือ 256.4 มิลลิวินาที (millisecond) (183.5-315.7 มิลลิวินาที) (Onsuwan et al., 2012)

5.4.2 การสร้างเสียงรบกวนของการพูดคุย

ช่วงเสียงรบกวนของการพูดคุยในภาษาไทย และภาษาอังกฤษ ในแต่ละภาษามาจากเสียงของผู้พูดเพศชาย จำนวน 2 คน (เพศเดียวกับเสียงเร่งเร้าเป้าหมาย (Brungart, 2001)) จำนวนประมาณ 10-12 พยางค์ต่อคน โดยสกัดเสียงพูดจากการรายงานข่าวพยากรณ์อากาศ (รายงานเดี่ยว) ซึ่งเป็นการพูดค่อนข้างเร็วคล้ายกับการสนทนาปกติ และได้ยินชัดเจนว่าเป็นภาษาไทย และภาษาอังกฤษ (อเมริกัน)

ภาพที่ 2 ค่าเฉลี่ยของค่าความถี่มูลฐาน (F0) ของคำเร่งเร้าเป้าหมาย



ภาพที่ 3 คำเร่งเร้าเป้าหมายจำนวน 40 คู่ (80 คำ) (Onsuwan et al., 2012)

Pair no.	Thai script	Trans-cription	Trans-lation	Thai script	Trans-cription	Trans-lation
1.	ปู	pūu	crab	ปู่	pùu	grand father
2.	ปา	pāa	throw	ป้า	pâa	aunt
3.	ป่า	pàa	jungle	ป่า	pǎa	dad
4.	เป๋	pêe	ruck-sask	เป่	pěe	slanted
5.	ตู	tūu	1st person pronoun	ตู่	tùu	assume
6.	โต	tōo	huge	โต้	tôo	debate
7.	ต่อ	tɔ̀ɔ	connect	ต้อ	tɔ̂ɔ	cata ract
8.	ตี่	tìi	narrow	ตี๋	tǐi	Chinese boy
9.	กอ	kɔ̄ɔ	clump	ก่อ	kɔ̀ɔ	build
10.	เก	kēe	cripple	เก๊	kěe	fake
11.	แก่	kɛ̀ɛ	old	แก้	kɛ̂ɛ	edit
12.	เก๊	kée	fake	เก่	kě̀e	chic
13.	จอ	tɕɔ̄ɔ	screen	จ๋อ	tɕɔ̌ɔ	monkey
14.	เจ่อ	tɕɤ̀ɤ	protrude	เจ๋อ	tɕɤ̌ɤ	obtrusive
15.	พอ	pʰɔ̄ɔ	enough	พ่อ	pʰɔ̂ɔ	father
16.	แพ	pʰɛ̄ɛ	raft	แพ้	pʰɛ́ɛ	lose
17.	โพธิ์	pʰōo	pipal	โผ	pʰǒo	dash
18.	ฟี้	fʰíi	smoke	ผี	pʰǐi	ghost
19.	ท่อ	tʰɔ̂ɔ	tube	ท้อ	tʰɔ́ɔ	discourage
20.	ถ้า	tʰâa	if	ท้า	tʰáa	defy
21.	ทู่	tʰûu	blunt	ถู	tʰǔu	rub
22.	แค	kʰɛ̄ɛ	sesbania flower	แข	kʰɛ̌ɛ	moon
23.	ขี่	kʰìi	ride	คี่	kʰǐi	odd
24.	ขู่	kʰùu	threaten	คู้	kʰúu	curl
25.	ค่า	kʰâa	value	ค้า	kʰáa	trade
26.	ค้า	kʰáa	trade	ขา	kʰǎa	leg
27.	ชา	tɕʰāa	tea	ช้า	tɕʰáa	slow
28.	ฉ่า	tɕʰàa	sizzle	ช้า	tɕʰáa	slow
29.	ฉี่	tɕʰìi	piss	ชี้	tɕʰíi	indicate
30.	แช่	tɕʰɛ̂ɛ	soak	แฉ	tɕʰɛ̌ɛ	reveal
31.	โบว์	bōo	ribbon	โบ๋	bǒo	hollow
32.	บ่า	bàa	shoulder	บ้า	bâa	crazy
33.	บู่	bùu	goby	บู๊	búu	fight
34.	เบ้	bêe	twist	เบี๊	béе	servant
35.	ดือ	dɨ̂ɨ	navel	ดื้อ	dɨ́ɨ	naughty
36.	หนี้	nîi	debt	หนี	nǐi	escape
37.	น้า	náa	aunt	หนา	nǎa	thick
38.	นอ	nɔ̄ɔ	horn	หน่อ	nɔ̀ɔ	shoot
39.	มา	māa	come	ม้า	máa	horse
40.	หมี่	mìi	rice noodle	หมี	mǐi	bear

จากนั้นนำเสียงพูดของแต่ละคนมาทำการปรับค่าเสียงโดยใช้วิธี Root Mean Square (RMS) normalization (Calandruccio et al., 2010) ในลำดับต่อมานำเสียงพูดของทั้งสองคนมาซ้อนทับกันในแต่ละภาษา และทำ RMS normalization อีกครั้งเพื่อให้ได้เสียงรบกวนของการพูดคุยแบบผู้พูดสองคน (two-talker babble) ที่เกิดช่องว่างหรือช่วงเงียบ (temporal gap) น้อยที่สุด โดยเฉพาะอย่างยิ่งไม่เกิดช่วงเงียบในช่วงกลางของเสียงรบกวน (ภาพที่ 4) ทั้งนี้ความยาวของเสียงรบกวนของการพูดคุยทั้งสองภาษา อยู่ที่ประมาณ 1.65 วินาที คำสำคัญที่ปรากฏในเสียง

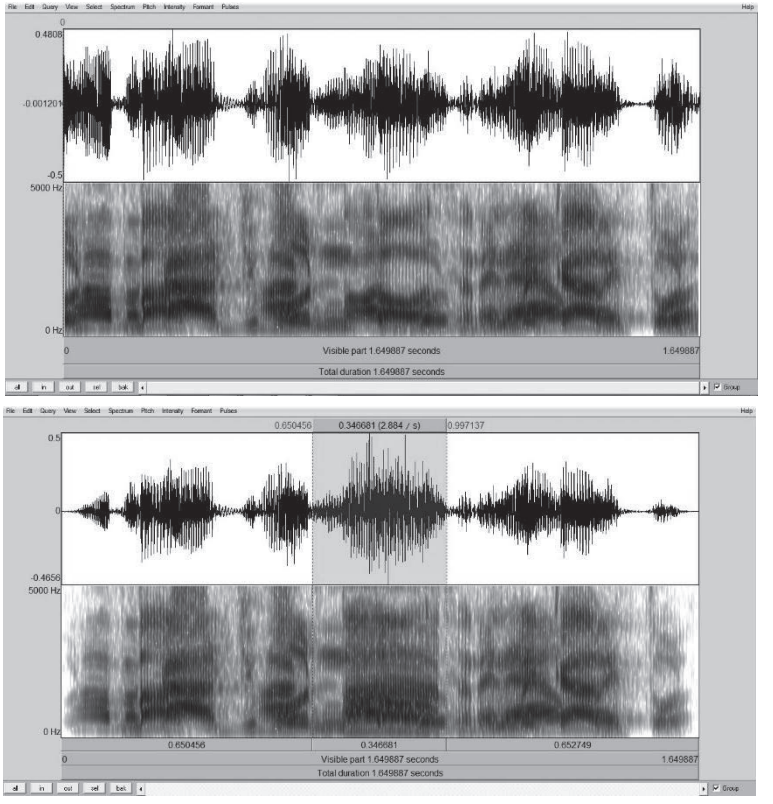
รบกวน 2 ช่วง/ภาษา สามารถเทียบเคียงกันได้ ดังนี้

ภาษาอังกฤษ: *risk, let's talk about, heavy rain*

ภาษาไทย: *อิทธิพล, ไปตรวจ, สถานการณ์ฝน*

จากนั้นนำคำเร่งเร้ามาปนเปื้อนด้วยเสียงรบกวนทั้งสองภาษา ให้มีระดับอัตราส่วนเสียงพูดต่อเสียงรบกวน SNR 3 ระดับ ได้แก่ 0dB –5dB และ –10dB โดยให้คำเร่งเร้า (ความยาวไม่เกิน 316 มิลลิวินาที) อยู่กึ่งกลางของเสียงรบกวน (ภาพที่ 4) ดังนั้นช่วงก่อนและหลังคำเร่งเร้าจะมีความยาวเสียงรบกวนประมาณ 600-700 มิลลิวินาที นอกจากนี้เพื่อไม่ให้ได้ยินเป็นเสียงที่กระตุก ในช่วงเริ่มต้นถึงประมาณ 250 มิลลิวินาทีแรกมีการทำให้เสียงเพิ่มระดับขึ้นจากเบาไปดัง (Tukey window or cosine-tapered window) และในช่วงท้าย 250 มิลลิวินาทีเสียงจะลดระดับจากดังไปเบาเช่นกัน ทั้งนี้การประมวลผลสัญญาณเสียงพูดทั้งหมดทำด้วยโปรแกรม MATLAB

ภาพที่ 4 คลื่นเสียง และแถบคลื่นเสียง (spectrogram) ของเสียงรบกวนของการพูดคุยในภาษาไทย (ผู้พูดเพศชายจำนวน 2 คน) (ภาพบน) และของคำเร่งเร้าเป้าหมาย "บ้า" ที่ปนเปื้อนด้วยเสียงรบกวนของการพูดคุยในภาษาไทยที่ SNR 0dB คำเร่งเร้าเป้าหมายซ้อนทับอยู่ในส่วนที่แรเงา (ภาพล่าง) ส่วนของสัญญาณเสียงในช่วงต้นและท้ายสุดในภาพล่างได้มีการปรับระดับเสียงจากดังไปเบา และเบาไปดังตามลำดับ



5.5 วิธีดำเนินการทดสอบการรับรู้

การทดสอบเป็นแบบ identification test แบบสองตัวเลือก (A/B forced choice) เช่นเดียวกับจุฑามณี อ่อนสุวรรณ และคณะ (2012) อาสาสมัครทุกคนจะทดสอบด้วยชุดทดสอบในแบบเดียวกันทั้งหมด กล่าวคือ แบบฝึกหัดเพื่อสร้างความคุ้นเคยจำนวน 10 ข้อ และแบบทดสอบจริงจำนวน 700 ข้อ (แบบทดสอบย่อย 7 ชุด) ตามขั้นตอนต่อไปนี้

1) ผู้วิจัยเชิญอาสาสมัครเข้าสู่ห้องทดสอบ ในแต่ละรอบจะมีอาสาสมัครเข้าร่วมไม่เกิน 5 คน และนั่งทดสอบโดยเว้นระยะห่างกันไม่ต่ำกว่า 1 เมตร

2) ผู้วิจัยชี้แจงวัตถุประสงค์ของการทดสอบและวิธีทดสอบทั้งหมดให้อาสาสมัครทราบโดยละเอียด หากอาสาสมัครไม่ยินยอมให้ทำการทดสอบ หรือไม่พร้อมทำการทดสอบ สามารถถอนตัวจากการทดสอบได้ทันที

3) ผู้วิจัยชี้แจงว่าแบบทดสอบย่อยมีจำนวน 7 ชุด ชุดละ 100 ข้อ (ชุดละไม่เกิน 8 นาที) ก่อนเริ่มแบบทดสอบย่อยชุดที่ 1 อาสาสมัครจะว่าได้ทำแบบฝึกหัด (จำนวน 10 ข้อ ไม่เกิน 3 นาที) การทำแบบทดสอบทั้งหมดจะผ่านคอมพิวเตอร์ และหูฟัง

4) อาสาสมัครนั่งทดสอบในท่าที่ผ่อนคลาย พร้อมสวมหูฟัง

5) ในแต่ละข้อ อาสาสมัครจะได้ยินคำไทยหนึ่งพยางค์ (เช่น “ค้า”) จำนวน 1 ครั้ง ผ่านหูฟังคำๆ นั้นอาจปรากฏโดยปราศจากเสียงรบกวน หรือถูกปนเปื้อนด้วยเสียงรบกวนที่ประเภทและระดับต่างกัน โดยไม่เป็นอันตรายต่อการได้ยิน

6) ตัวเลือกจำนวน 2 ตัวเลือก จะปรากฏบนหน้าจอคอมพิวเตอร์ อาสาสมัครเลือกตอบคำที่ใกล้เคียงกับที่ได้ยินมากที่สุดจากตัวเลือกโดยการกดแป้นพิมพ์ และไม่สามารถกลับมาฟังซ้ำได้อีก หากอาสาสมัครไม่สามารถเลือกคำตอบได้ ให้เดาตัวเลือกใดตัวเลือกหนึ่ง เมื่อกดคำตอบแล้ว จะได้ยินข้อต่อไป (ระยะเวลาของการทดสอบจะอยู่ในความควบคุมของอาสาสมัคร และไม่มีความเร่งรีบในการทำ)

7) อาสาสมัครทำซ้ำขั้นตอนที่ 5) และ 6) จนครบแบบทดสอบย่อยแต่ละชุด มีการพัก 5 นาที (ควบคุมอัตโนมัติ) หลังจากจบบททดสอบย่อยแต่ละชุด

5.6 การวัดผลการวิจัย

นำข้อมูลการรับรู้เสียงวรรณยุกต์ ได้แก่ คำตอบ และระยะเวลาการตอบสนอง (reaction time) มารวบรวมและคำนวณค่าต่างๆ ได้แก่ (ร้อยละ) คำตอบที่ถูกต้อง (correct response) (ร้อยละ) การฟังเข้าใจ (intelligibility score) และคำนวณเมทริกซ์ความสับสนทางการรับรู้เสียง นอกจากนี้ได้ใช้ Analysis of Variance (ANOVA) ในการวิเคราะห์ทางสถิติ

6. ผลการวิจัย

ในบทความนี้ผู้วิจัยขอรายงานในส่วนของร้อยละของคำตอบที่ถูกต้อง ร้อยละของการฟังเข้าใจ และความสับสนทางเสียง โดยจะยังไม่กล่าวถึงระยะเวลาการตอบสนอง

6.1 ร้อยละของคำตอบที่ถูกต้อง (percent correct response)

ตารางที่ 1 แสดงผลร้อยละของคำตอบที่ถูกต้องของอาสาสมัครทั้ง 35 คน เมื่อไม่มีเสียงรบกวน เมื่อมีเสียงรบกวนของการพูดคุยในภาษาอังกฤษ และเมื่อมีเสียงรบกวนของการพูดคุยในภาษาไทย ผลร้อยละแสดงให้เห็นว่าเมื่อไม่มีเสียงรบกวน ร้อยละของคำตอบที่ถูกต้องอยู่ในระดับที่สูงมากคือ ร้อยละ 96 และเมื่อมีเสียงรบกวนของการพูดคุย ร้อยละของคำตอบที่ถูกต้องจะลดลงเมื่อระดับของ SNR ต่ำลงตามลำดับ โดยเฉพาะอย่างยิ่งค่าเฉลี่ยร้อยละของคำตอบที่ถูกต้องจากทั้งสามระดับเสียงรบกวนของการพูดคุยในภาษาอังกฤษคือ ร้อยละ 88.57 ส่วนในภาษาไทยต่ำกว่าคือ ร้อยละ 84.69 นอกจากนี้ร้อยละของคำตอบที่ถูกต้องในแต่ละระดับ SNR (0dB –5dB และ –10dB) จะสูงกว่าเมื่อมีเสียงรบกวนของการพูดคุยในภาษาอังกฤษเมื่อเทียบกับเสียงรบกวนในภาษาไทย ความแตกต่างนี้จะเห็นได้ชัดเจนที่ระดับ SNR –5dB และ –10dB

ตารางที่ 1 ร้อยละของคำตอบที่ถูกต้อง (35 คน) เมื่อไม่มีเสียงรบกวน เมื่อมีเสียงรบกวนของการพูดคุยในภาษาอังกฤษ และเมื่อมีเสียงรบกวนของการพูดคุยในภาษาไทย ค่าเบี่ยงเบนมาตรฐานอยู่ภายในวงเล็บ

no noise	96.07		
	SNR level (dB)		
	0	–5	–10
English babble	93.75 (3.47)	90.29 (4.49)	81.68 (6.73)
Thai babble	92.14 (3.17)	86.68 (5.02)	75.25 (5.54)

6.2 ร้อยละของการฟังเข้าใจ (percent intelligibility score)

ร้อยละของการฟังเข้าใจจะพิจารณาคะแนนการตอบผิดควบคู่ไปด้วย โดยใช้สูตรของการแปลงจำนวนคำตอบที่ถูกต้องเป็นร้อยละของการฟังเข้าใจ ดังนี้

$$P_e = \frac{N_r - N_w}{T} \times 100\%$$

(P_e, N_r, N_w, and T หมายถึง ร้อยละการฟังเข้าใจ, จำนวนคำตอบที่ถูกต้อง, จำนวนคำตอบที่ผิด, จำนวนคำตอบทั้งหมด (ตามลำดับ)) (Voiers, 1983)

ตารางที่ 2 แสดงผลร้อยละของการฟังเข้าใจ ของอาสาสมัครทั้ง 35 คน เมื่อไม่มีเสียงรบกวน เมื่อมีเสียงรบกวนของการพูดคุยในภาษาอังกฤษ และเมื่อมีเสียงรบกวนของการพูดคุยในภาษาไทย ผลร้อยละสอดคล้องกับตารางที่ 1 และชี้ให้เห็นว่าเมื่อไม่มีเสียงรบกวนร้อยละของการฟังเข้าใจอยู่ในระดับที่สูงคือร้อยละ 92 เมื่อมีเสียงรบกวนในการพูดคุยร้อยละของการฟังเข้าใจจะลดลงเมื่อระดับของ SNR ต่ำลงตามลำดับ โดยเฉพาะอย่างยิ่งค่าเฉลี่ยร้อยละของการฟังเข้าใจจากทั้งสามระดับเสียงรบกวนของการพูดคุยในภาษาอังกฤษคือ ร้อยละ 77.14 ส่วนในภาษาไทยต่ำกว่าคือ ร้อยละ 69.38 นอกจากนี้ร้อยละของการฟังเข้าใจในแต่ละระดับ SNR จะสูงกว่าเมื่อเสียงรบกวนของการพูดคุยเป็นภาษาอังกฤษ

ตารางที่ 2 ร้อยละของการฟังเข้าใจ (35 คน) เมื่อไม่มีเสียงรบกวน เมื่อมีเสียงรบกวนของการพูดคุยในภาษาอังกฤษ และเมื่อมีเสียงรบกวนของการพูดคุยในภาษาไทย ค่าเบี่ยงเบนมาตรฐานอยู่ภายในวงเล็บ

no noise	92.14		
	SNR level		
	0	–5	–10
English babble	87.5 (6.94)	80.57 (8.98)	63.36 (13.46)
Thai babble	84.29 (6.35)	73.36 (10.04)	50.50 (11.08)

การวิเคราะห์ทางสถิติ one-way ANOVA ชี้ให้เห็นว่าชนิดของเสียงรบกวนที่ต่างกัน (การพูดคุยในภาษาอังกฤษ การพูดคุยในภาษาไทย และไม่มีเสียงรบกวน) มีผลต่อร้อยละของการฟังเข้าใจอย่างมีนัยสำคัญทางสถิติ [$F(2,102) = 97.58$, $p < 0.001$] การวิเคราะห์ทางสถิติ two-way ANOVA (ชนิดภาษาของเสียงรบกวนของการพูดคุย และระดับของเสียงรบกวน) แสดงให้เห็นว่าชนิดภาษาของเสียงรบกวนของการพูดคุยมีผลต่อร้อยละของการฟังเข้าใจอย่างมีนัยสำคัญทางสถิติ [$F(1,204) = 33.08$, $p < 0.001$] เช่นเดียวกับระดับของเสียงรบกวน [$F(2,204) = 161.08$, $p < 0.001$] นอกจากนี้พบว่าชนิดภาษาของเสียงรบกวนของการพูดคุยกับระดับของเสียงรบกวนมีปฏิสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ [$F(2,204) = 4.3$, $p < 0.05$] ทั้งนี้เมื่อเปรียบเทียบผลร้อยละของการฟังเข้าใจใน 21 คู่ ตามชนิดของเสียงรบกวน (3 ชนิด) และระดับของเสียงรบกวน (4 ระดับ แต่ที่ไม่มีเสียงรบกวนมีเพียง 1 ระดับ) พบว่ามีนัยสำคัญทางสถิติที่ระดับ 0.05 ในทุกคู่ยกเว้นใน 3 คู่ ได้แก่ 1) ไม่มีเสียงรบกวน กับเสียงรบกวนของการพูดคุยในภาษาอังกฤษระดับ SNR 0dB 2) เสียงรบกวนของการพูดคุยในภาษาไทยระดับ SNR 0dB กับเสียงรบกวนของการพูดคุยในภาษาอังกฤษระดับ SNR 0dB 3) เสียงรบกวนของการพูดคุยในภาษาไทยระดับ SNR 0dB กับ เสียงรบกวนของการพูดคุยในภาษาอังกฤษระดับ SNR –5dB ดังนั้นที่ระดับ SNR –5dB

และ–10dB เท่านั้นที่ชนิดภาษาของเสียงรบกวนของการพูดคุยมีผลต่อร้อยละของการฟังเข้าใจอย่างมีนัยสำคัญทางสถิติ กล่าวคือที่ระดับ SNR ดังกล่าวร้อยละของการฟังเข้าใจในเสียงรบกวนของการพูดคุยในภาษาอังกฤษมีค่าสูงกว่าในภาษาไทยอย่างมีนัยสำคัญทางสถิติ

6.3 ความสับสนทางเสียง

ความสับสนทางเสียงจะรายงานจากค่าร้อยละของคำตอบที่ถูกต้องที่ระดับ SNR –5dB ซึ่งเป็นระดับที่คะแนนร้อยละของการฟังเข้าใจไม่สูงใกล้คะแนนเต็ม (90-100) หรือต่ำใกล้ในระดับของการเดา (50-60)

ตารางที่ 3 และ 4 แสดงเมทริกซ์ความสับสนทางการรับรู้เสียงวรรณยุกต์ในภาษาไทยเมื่อมีเสียงรบกวนของการพูดคุยในภาษาอังกฤษ (ตารางที่ 3) และในภาษาไทย (ตารางที่ 4) ใน SNR –5dB คะแนนร้อยละคำตอบที่ถูกต้อง (ช่องตารางที่แรเงา) เรียงจากสูงที่สุดไปต่ำที่สุดในเสียงรบกวนภาษาอังกฤษ คือ โท ตรี สามัญ จัตวา และเอก ส่วนในเสียงรบกวนภาษาไทย คือ ตรี สามัญ โท เอก และจัตวา ถึงแม้ว่าลำดับในรายละเอียดมีความแตกต่างกัน แต่เมื่อพิจารณาจะพบว่ามีรูปแบบที่คล้ายคลึงกันคือ วรรณยุกต์กลุ่มที่ก่อให้เกิดความสับสนต่ำ คือ โท ตรี และสามัญ ส่วนวรรณยุกต์กลุ่มที่ก่อให้เกิดความสับสนสูงคือ เอก และจัตวา

ความสับสนที่เกิดขึ้นในแง่ของการรับรู้ทางเสียงวรรณยุกต์เกิดขึ้นในลักษณะที่ไม่สมดุลและมีความแตกต่างกันในรายละเอียดระหว่างในเสียงรบกวนภาษาอังกฤษและในภาษาไทย อย่างไรก็ดีรูปแบบที่คล้ายคลึงกันระหว่างเสียงรบกวนในภาษาอังกฤษและในภาษาไทยที่เห็นได้ชัดเจนคือ มีการสับสนแบบทิศทางเดียว (unidirectional) ระหว่างวรรณยุกต์สามัญ (เป็นตรี) เอก (เป็นตรี) จัตวา (เป็นโท) จัตวา (เป็นตรี) และจัตวา (เป็นสามัญ) ส่วนความสับสน 2 ทิศทางเกิดขึ้นระหว่างวรรณยุกต์สามัญและเอก และระหว่างเอกและจัตวา นอกจากนี้ความสับสนที่เกิดขึ้นสูงที่สุดอันดับต้นๆ ทั้งในเสียงรบกวนภาษาอังกฤษและในภาษาไทย ได้แก่ ระหว่างวรรณยุกต์เอก (เป็นสามัญ) จัตวา (เป็นเอก) และจัตวา (เป็นสามัญ)

ตารางที่ 3 เมทริกซ์ความสับสนทางการรับรู้เสียงวรรณยุกต์ในภาษาไทย (35 คน) เมื่อมีเสียงรบกวนของการพูดคุยในภาษาอังกฤษใน SNR –5dB

response / stimulus	mid	low	falling	high	rising
mid	92.86	2.86	0.89	2.86	0.54
low	10.71	79.29	2.86	3.93	3.21
falling	0.71	0.18	97.32	1.25	0.54
high	0.54	0.71	0.54	96.79	1.43
rising	4.11	6.79	1.61	2.32	85.18

ตารางที่ 4 เมทริกซ์ความสับสนทางการรับรู้เสียงวรรณยุกต์ในภาษาไทย (35 คน) เมื่อมีเสียงรบกวนของการพูดคุยในภาษาไทยใน SNR –5dB

response / stimulus	mid	low	falling	high	rising
mid	91.25	1.61	2.50	4.46	0.18
low	7.68	83.75	5.36	1.79	1.43
falling	6.96	2.68	88.75	1.43	0.18
high	0.89	0.36	3.57	94.46	0.71
rising	9.11	6.79	1.96	6.96	75.18

7. สรุปและอภิปรายผล

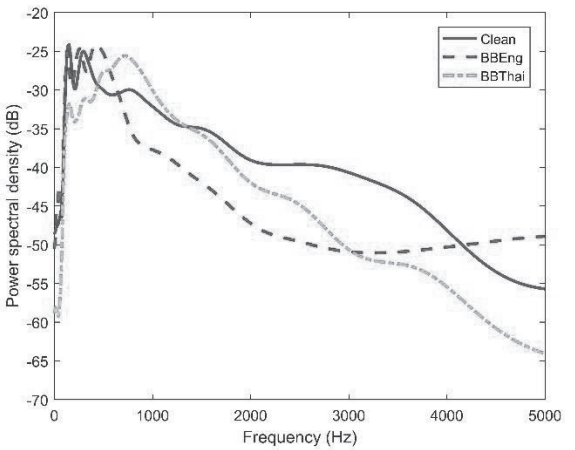
ผลการวิจัยเป็นไปตามสมมติฐานหลัก กล่าวคือ ชนิดภาษาของเสียงรบกวนของการพูดคุยมีผลต่ออัตราร้อยละของการฟังเข้าใจของเสียงวรรณยุกต์ในภาษาไทยอย่างมีนัยสำคัญทางสถิติ โดยที่การฟังเข้าใจภายใต้เสียงรบกวนของการพูดคุยในภาษาอังกฤษมีค่าสูงกว่าในภาษาไทยอย่างมีนัยสำคัญทางสถิติ ทั้งนี้ในงานวิจัยนี้อิทธิพลของชนิดภาษาของเสียงรบกวน (linguistic effect) เกิดขึ้นเมื่อ SNR ลดลงถึงระดับที่เสียงรบกวนมีความดังกว่าสัญญาณเสียงที่ระดับ SNR –5dB และ–10dB

ประเด็นที่ว่าเสียงรบกวนจำเป็นต้องมีปริมาณสูงจึงจะเห็นอิทธิพลของชนิดภาษานี้ไม่เป็นที่น่าแปลกใจ เนื่องจากงานวิจัยก่อนหน้านี้ได้รายงานผลในทำนองเดียวกัน โดยพบว่าอิทธิพลของชนิดภาษาของเสียงรบกวนจะเกิดขึ้นเมื่อ SNR มีค่าต่ำกว่า 0dB (Calandruccio et al., 2010; Brouwer et al., 2012) นอกจากนี้ผลงานวิจัยนี้ได้ขยายผลให้กับงานวิจัยก่อนหน้านี้ที่ทดสอบการรับรู้ในระดับวลี หรือประโยค เนื่องจากเป็นการทดสอบระดับหน่วยเสียง ได้แก่ เสียงวรรณยุกต์

ข้อสรุปเรื่องอิทธิพลของชนิดภาษาของเสียงรบกวนที่เป็นผลให้การรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวนของการพูดคุยในภาษาไทยมีระดับที่ต่ำกว่าภายใต้เสียงรบกวนของการพูดคุยในภาษาอังกฤษในงานนี้ สอดคล้องกับข้อสรุปจากงานวิจัยก่อนหน้านี้ที่ชี้ให้เห็นว่าอิทธิพลของชนิดภาษาเป็นผลมาจากความคล้ายคลึงทางภาษาของเสียงพูดคุย (ที่เป็นภาษาเดียวกับหรือต่างกับเสียงเป้าหมาย) ซึ่งส่งผลต่อกระบวนการการรับรู้เสียงและการประมวลผลในระดับสูง มากกว่าที่จะเป็นผลมาจากลักษณะของสัญญาณเสียง (speech signal) ซึ่งเป็นการประมวลผลทางเสียงในระดับที่ต่ำกว่า (Van Engen & Bradlow, 2007; Calandruccio et al., 2010; Brouwer et al., 2012) ในภาพที่ 5 ซึ่งเป็นภาพสเปกตรัมเฉลี่ยจากเสียงพูดจำนวนมาก (long-time average spectrum) ที่มาจากเสียงเร่งเร้าเป้าหมาย (ที่ปราศจากเสียงรบกวน) (clean) จากเสียง

รบกวนของการพูดคุยในภาษาไทย (BB Thai) และภาษาอังกฤษ (BB Eng) จะเห็นได้ว่าในช่วงความถี่ต่ำ (0-200 Hz) ซึ่งมีความเกี่ยวข้องโดยตรงกับค่าความถี่มูลฐานของเสียงวรรณยุกต์ในเสียงเร่งเร้าเป้าหมาย (ดูภาพที่ 2) นั้น เสียงเร่งเร้าเป้าหมายปกคลุมหรือซ้อนทับ (masking) กับเสียงรบกวนของการพูดคุยในภาษาอังกฤษ แต่จะไม่ซ้อนทับกับเสียงรบกวนในภาษาไทย ดังนั้นหากการรับรู้และการประมวลผลขึ้นกับลักษณะของสัญญาณเสียงเป็นหลัก การรับรู้เสียงวรรณยุกต์ในเสียงรบกวนของการพุดคุยในภาษาไทยควรที่จะดีกว่า (ถูกต้องมากกว่า) ในเสียงรบกวนภาษาอังกฤษ

ภาพที่ 5 ภาพสเปกตรัมเฉลี่ยจากเสียงพูดจำนวนมาก (long-time average spectrum) จากเสียงเร้าเป้าหมาย (ที่ปราศจากเสียงรบกวน) (clean) จากเสียงรบกวนของการพูดคุยในภาษาไทย (BB Thai) และภาษาอังกฤษ (BB Eng)



เมื่อพิจารณาเปรียบเทียบกับงานวิจัยเรื่องการรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวนประเภทพลังงานแบบ white noise (Onsuwan et al., 2012) ในภาพรวมจะพบว่าเสียงรบกวนประเภทข้อมูลภาษาจะมีผลปกคลุมเสียงเป้าหมาย (masking) ได้สูงกว่าเสียงรบกวนประเภทพลังงาน สอดคล้องกับผลการวิจัยก่อนหน้านี้ (Brungart, 2001; Simpson & Cooke, 2005; Cutler, 2012) ดังจะเห็นได้จากร้อยละการฟังเข้าใจของเสียงวรรณยุกต์ในภาษาไทย (34 คน) ในจุฑามณี อ่อนสุวรรณและคณะ (2012) ใน 4 ระดับ SNR –6dB –12dB –18dB และ–24dB ได้แก่ ร้อยละ 88.5, 89.9, 88.0 และ 64.4 ตามลำดับ โดยเฉพาะอย่างยิ่งที่ระดับ SNR ที่ใกล้เคียงกันคือ –6dB (Onsuwan et al., 2012) และ –5dB (งานวิจัยนี้) ร้อยละของการฟังเข้าใจในเสียงรบกวนแบบ white noise อยู่ที่ประมาณร้อยละ 89 ในขณะที่ในเสียงรบกวนของการพุดคุย ร้อยละของ

การฟังเข้าใจอยู่ที่ระดับที่ต่ำกว่ามากคือ ประมาณร้อยละ 81 (เสียงรบกวนในภาษาอังกฤษ) และร้อยละ 73 (เสียงรบกวนในภาษาไทย)

นอกจากนี้รูปแบบความสับสนในการรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวนประเภทพลังงาน (SNR –24dB) และประเภทข้อมูลภาษา (SNR –5dB) ยังมีความแตกต่างกันค่อนข้างชัดเจน (ข้อสังเกต ในที่นี้เปรียบเทียบรูปแบบความสับสนที่ SNR ที่แสดงความสับสนชัดเจนที่สุดจากสองงานวิจัย) ตารางที่ 5 แสดงเมทริกซ์ความสับสนทางการรับรู้เสียงวรรณยุกต์ในภาษาไทยเมื่อมีเสียงรบกวน white noise SNR –24dB จากจุฑามณี อ่อนสุวรรณ และคณะ (2012) ซึ่งพบว่าเสียงวรรณยุกต์ตรีเป็นเสียงที่มีความสับสนมากที่สุด รองลงมาคือ จัตวา สามัญ โท และเอกมีความสับสนน้อยที่สุด เมื่อพิจารณาร่วมกับผลวิจัยในงานนี้พบว่า สิ่งที่คล้ายคลึงกันคือ วรรณยุกต์จัตวาก่อให้เกิดความสับสนสูง และวรรณยุกต์สามัญก่อให้เกิดความสับสนค่อนข้างต่ำ อย่างไรก็ดีลำดับความสับสนในรายละเอียดในสองงานนี้มีความแตกต่างกัน เนื่องจากในงานนี้พบว่าวรรณยุกต์กลุ่มที่ก่อให้เกิดความสับสนต่ำ คือ โท ตรี และสามัญ ส่วนวรรณยุกต์กลุ่มที่ก่อให้เกิดความสับสนสูงคือ เอก และจัตวา

ความสับสนที่เกิดขึ้นในแง่ของการรับรู้ทางเสียงวรรณยุกต์ภายใต้เสียงรบกวนประเภทพลังงานใน จุฑามณี อ่อนสุวรรณ และคณะ (2012) และประเภทข้อมูลภาษาในงานวิจัยนี้เกิดขึ้นในลักษณะที่ไม่สมดุลเช่นเดียวกัน แต่มีความแตกต่างกันในรายละเอียด ความสับสนที่เกิดขึ้นสูงที่สุดลำดับต้นๆ ในเสียงรบกวนของการพูดคุยในภาษาอังกฤษและในภาษาไทย ได้แก่ ระหว่างวรรณยุกต์เอก (เป็นสามัญ) จัตวา (เป็นเอก) และจัตวา (เป็นสามัญ) ส่วนในเสียงรบกวนประเภทพลังงาน คือ สามัญ (เป็นตรี) ตรี (เป็นจัตวา) โท (เป็นเอก) และตรี (เป็นสามัญ)

ดังนั้นอาจกล่าวได้ว่าประเภทของเสียงรบกวนมีผลต่อรูปแบบของความสับสนในการรับรู้เสียงวรรณยุกต์ในภาษาไทย จึงเป็นที่น่าสนใจที่ควรมีการนำข้อสรุปนี้ไปขยายผล และประเมินต่อไปว่าเป็นลักษณะเฉพาะของการรับรู้เสียงในภาษาไทย หรือเกิดขึ้นในภาษาอื่น และเป็นลักษณะที่เกิดกับหน่วยเสียงประเภทอื่นๆ ในภาษาไทยหรือไม่ อย่างไร

อย่างไรก็ดี เมื่อพิจารณาความสับสนที่เกิดขึ้นในแง่ของการรับรู้ทางเสียงวรรณยุกต์ภายใต้เสียงประเภทข้อมูลภาษาในงานวิจัยนี้ ความสับสนที่เกิดขึ้นสูงที่สุดลำดับต้นๆ ในเสียงรบกวนของการพูดคุยในภาษาอังกฤษและในภาษาไทย ได้แก่ ระหว่างวรรณยุกต์เอก (เป็นสามัญ) จัตวา (เป็นเอก) และจัตวา (เป็นสามัญ) ซึ่งถือได้ว่าเป็นความสับสนที่เกิดข้ามกันระหว่างวรรณยุกต์คงระดับและเปลี่ยนระดับ และเป็นวรรณยุกต์ที่อยู่ในย่านระดับเสียงกลางและเสียงค่อนข้างต่ำ จึงเป็นไปได้ว่าย่านระดับเสียงมีบทบาทต่อการรับรู้เสียงวรรณยุกต์เหล่านี้ไม่น้อยไปกว่าลักษณะการเปลี่ยนหรือคงระดับของระดับเสียง

ตามที่ได้กล่าวไปแล้วในหัวข้อ 3.2 ว่า จากลักษณะทางกลสัทศาสตร์ (ความคล้ายคลึงกันของรูปร่างของค่าความถี่มูลฐาน) ของเสียงวรรณยุกต์ในภาษาไทยได้มีการคาดการณ์ว่าจะ

เกิดความสับสนสูงในการรับรู้เสียงของวรรณยุกต์สามัญ และเอก ผลการทดสอบการรับรู้เสียงวรรณยุกต์ของ แอแบรมสัน (1975) แกนดอร์ และรจนา ทรรทรานนท์ (1983) และผลการทดสอบของณัฐกร ทับทอง (2001) สอดคล้องกับข้อสรุปดังกล่าว และในงานวิจัยชิ้นนี้ และจุฑามณี อ่อนสุวรรณ และคณะ (2012) พบว่าในบรรดาความสับสนที่เกิดกับเสียงวรรณยุกต์คู่อื่นๆ มีความสับสนของระหว่างวรรณยุกต์ 2 เสียงนี้เช่นกัน และเป็นความสับสนแบบ 2 ทิศทาง อย่างไรก็ดีรูปแบบและลำดับที่เกิดขึ้นมีความแตกต่างกัน กล่าวคือ ความสับสนระหว่างวรรณยุกต์เอก (เป็นสามัญ) เกิดขึ้นสูงในลำดับต้นๆ ในเสียงรบกวนของการพูดคุยในภาษาอังกฤษและในภาษาไทย ในขณะที่ความสับสนระหว่างเสียงสามัญ (เป็นเอก) เกิดขึ้นสูงกว่า เอก (เป็นสามัญ) ในเสียงรบกวนประเภทพลังงาน white noise (Onsuwan et al., 2012) จากข้อสรุปที่ว่าประเภทของเสียงรบกวนมีผลต่อรูปแบบของความสับสนในการรับรู้เสียงวรรณยุกต์ในภาษาไทย ทำให้การกล่าวสรุปภาพรวมของรูปแบบความสับสนในการรับรู้เสียงวรรณยุกต์ในภาษาไทยเป็นไปได้จำกัด เนื่องจากงานวิจัยที่กล่าวถึงรูปแบบความสับสนนี้ ได้แก่ แอแบรมสัน (1975) ที่ใช้การปรับความดังของเสียงไม่ได้ใช้เสียงรบกวน ส่วน แกนดอร์ และรจนา ทรรทรานนท์ (1983) นั้นไม่ได้ใช้เสียงรบกวน

ตารางที่ 5 เมทริกซ์ความสับสนทางการรับรู้เสียงวรรณยุกต์ในภาษาไทย (34 คน) เมื่อมีเสียงรบกวน white noise (additive white Gaussian noise) SNR –24dB (จากตารางที่ 3 ใน Onsuwan et al. (2012))

response / stimulus	mid	low	falling	high	rising
mid	81.6	4.4	2.9	7.7	3.3
low	2.8	88.6	4.8	0.9	2.9
falling	5.5	6.1	83.5	3.5	1.5
high	6.1	2.4	8.8	76.5	6.3
rising	4.8	5.5	4.0	4.8	80.9

จากงานวิจัยนี้ เมื่อกล่าวถึงปัจจัยที่ส่งผลต่อรูปแบบ และลำดับของความสับสนในการรับรู้เสียงวรรณยุกต์ในภาษาไทยภายใต้เสียงรบกวน (อาจเป็นไปได้กับหน่วยเสียงประเภทอื่นๆ ด้วยเช่นกัน) ความคล้ายคลึงการทางกลสัทศาสตร์ระหว่างเสียงวรรณยุกต์นั้นเป็นเพียงส่วนหนึ่งเท่านั้น ประเภท ชนิดและระดับของเสียงรบกวน เป็นปัจจัยที่ต้องคำนึงถึงอีกด้วย

นอกจากนี้ ในประเด็นของวิธีวิจัยงานวิจัยส่วนหนึ่งใช้เสียงรบกวนของการพูดคุยโดยเฉพาะภาษาอังกฤษได้สร้างขึ้นจากการสกัดเสียงพูดจากคลังข้อมูลเสียงขนาดใหญ่ซึ่งมีอยู่แล้ว

ในภาษา เช่น TIMIT (speech) corpus (Simpson & Cooke, 2005) speech corpus for multitalker communications research (the coordinate response measure (CRM)) (Brungart, 2001) ซึ่งคลังข้อมูลเสียงในลักษณะนี้ในภาษาไทยยังมีจำกัด อย่างไรก็ดีงานวิจัยชิ้นนี้แสดงให้เห็นว่าการสร้างเสียงรบกวนของการพูดคุยขึ้นโดยการสกัดเสียงพูดจากการรายงานข่าวซึ่งเป็นข้อมูลเสียงสาธารณะที่สามารถนำมาใช้ได้ ซึ่งในงานวิจัยนี้เลือกใช้การรายงานพยากรณ์อากาศที่มีลักษณะการพูดต่อเนื่องที่เป็นธรรมชาติและที่มีลักษณะเทียบเคียงกันได้ในภาษาไทย และภาษาอังกฤษ (หัวข้อ 5.4.2) แม้จะใช้เวลาในการคัดสรร และวัดค่าทางสัญญาณเสียง แต่ได้ผลเป็นที่น่าพอใจ และสามารถนำไปประยุกต์ใช้กับงานวิจัยที่เกี่ยวข้อง

8. กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณนางสาวธนวรรณ สายใหม่ และนางสาวนันทพร สายใหม่ ที่ทำหน้าที่เป็นอย่างดีในการเขียนโปรแกรม และช่วยเก็บรวบรวมข้อมูล

เอกสารอ้างอิง/ References

ปิยฉัตร ปานโรจน์. (2534). *ลักษณะเชิงกลสัทศาสตร์ของวรรณยุกต์ในภาษาไทยกรุงเทพฯ: การแปรตามกลุ่มอายุ* (วิทยานิพนธ์ปริญญามหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย.

กุสุมา นะสานี. (2545). *วรรณยุกต์ภาษาไทยที่ออกเสียงโดยผู้พูดที่ใช้หลอดลม-หลอดอาหาร: การวิเคราะห์ทางกลสัทศาสตร์และการทดสอบการรับรู้* (วิทยานิพนธ์ปริญญามหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย.

ธีระพันธ์ เหลืองทองคำ. (2554). *เสียงภาษาไทย: การศึกษาทางกลสัทศาสตร์.* กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Abramson, A. S. (1962). The vowels and tones of standard Thai: Acoustical measurements and experiments. *Indiana University Research Center in Anthropology, Folklore and Linguistics,* Pub. 20 . Bloomington, IN: Indiana University.

_____. (1972). Publications of pierre delattre. In A. Valdman (Ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (pp. 21-30). Netherlands: Mouton.

_____. (1975). The tones of central Thai: Some perceptual experiments. In J. G.Harris, & J. R. Chamberlain (Eds.), *Studies in Thai Linguistics in Honor of William J. Gedney* (pp. 1-16). Bangkok: Central Institute of English Language.

_____. (1976). Thai tones as reference system. In T. W. Gething, J. G. Harris, & P. Kullavanijaya (Eds), *Tai linguistics in honor of Fang-Kuei Li* (pp.1–12). Bangkok: Chulalongkorn University Press.

Anivan, S. (1985). *An acoustical study of Thai tones* (Unpublished Ph.D. thesis). Macquarie University.

Benkí, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica, 60*(2), 129-157.

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America, 131*(2), 1449-1464.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101-1109.

Burnham, D., & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong* (pp. 29-47). Bangkok: Chulalongkorn University Press.

Calandruccio, L., Van Engen, K., Dhar, S., & Bradlow, A. R. (2010). The effectiveness of clear speech as a masker. *Journal of Speech, Language, and Hearing Research, 53*(6), 1458-1471.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spokenwords.* London: MIT Press.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America, 124*(2), 1264-1268.

Gandour, J., & Dardarananda, R. (1983). Identification of tonal contrasts in Thaiaphasic patients. *Brain and Language, 18*(1), 98-114.

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119*(4), 2445-2454.

Hardcastle, W. J., Laver, J., & Gibbon, F. E. (Eds.). (2010). *The Handbook of Phonetic Sciences* (2nd ed.). Malden, MA: Wiley-Blackwell.

House, A. S., Williams, C. E., Hecker, M. H., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *The Journal of the Acoustical Society of America, 37*(1), 158-166.

Johnson, K. (2003). *Acoustic and Auditory Phonetics* (2nd ed.). Oxford: Blackwell. Kong, Y. Y., & Zeng, F. G. (2006). Temporal and spectral cues in Mandarin tone recognition. *The Journal of the Acoustical Society of America, 120*(5), 2830-2840.

Lass, N. J. (1996). *Principles of experimental phonetics.* St. Louis: Mosby.

Lee, C. Y., Tao, L., & Bond, Z. S. (2010). Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Communication, 52,* 900-910.

Li, Z., Tan, E.C., Mcloughlin, I., & Teo, T. T. (2000). Proposal of standards for intelligibility tests of Chinese speech. *IEE Proceedings Vision, Image and Signal Processing, 147*(3)*,* 254-260.

Loizou, P. C. (2007). *Speech enhancement: Theory and practice.* Boca Raton, FL: CRC press.

Mcloughlin, I. (2008). Subjective intelligibility testing of Chinese speech. *IEEE Transactions on Audio, Speech, and Language Processing, 16*(1), 23–33.

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin, 44*(2), 105-129.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338-352.

Mixdorff, H., Hu, Y., & Burnham, D. (2005). Visual cues in Mandarin tone perception. In *6th Interspeech 2005 and 9th European Conference on Speech Communication and Technology (Eurospeech).* Retrieved from http://public.beuthhochschule.de/~mixdorff/thesis/files/mixdorff_hu_eurosp2005.pdf

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson, & C. Jones (Eds.), *Historical Linguistics II. Theory and Description in Phonology* (pp. 353-389). Amsterdam: North Holland.

_____. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Proceedings of the Chicago Linguistics Society 17: Papers from the Parasession on Language and Behaviour* (pp. 178-203). Chicago: Chicago Linguistic Society.

_____. (1995). The perceptual basis of some sound patterns. In B. A. Connell, & A. Arvaniti (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp. 87-92). Cambridge: Cambridge University Press.

Onsuwan, C., Tantibundhit, C., Saimai, T., Saimai, N., Chootrakool, P., & Thatphithakkul, S. (2012). Analysis of Thai tonal identification in noise. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology (SST)* (pp. 173-176). Sydney: Macquarie University.

_____. (2013). Perception of Thai distinctive vowel length in noise. In *Proceedings of Meetings on Acoustics (POMA): Proceedings of the 21st International Congress on Acoustics/ 165th Meeting of the Acoustical Society of America* (pp. 1-7). Montreal, Canada.

Pisoni, D. B., & Remez, R. E. (Eds.). (2004). *The Handbook of Speech Perception.* Malden, MA: Wiley-Blackwell.

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America, 57*(S1), S5.

Potisuk, S., Gandour, J. T., & Harper, M. P. (1994). *F_0 correlates of stress in Thai. Linguistics of the Tibeto-Burman Area, 17*(2), 1-27.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of *N. The Journal of the Acoustical Society of America, 118*(5), 2775-2778.

Smith, C. L. (2009). Experimental phonetics. In D. Eddington (Ed.), *Quantitative and Experimental Linguistics* (pp. 159-196). Muenchen: Lincom Europa.

Stocker, M. (2013). *Hear where we are: Sound, ecology, and sense of place.* New York, NY: Springer.

Strange, W. (Ed.). (1995). *Speech perception and linguistic experience.* Maryland: York Press.

Tantibundhit, C., & Onsuwan, C. (2015). Speech intelligibility tests and analysis of confusions and perceptual representations of Thai initial consonants. *Speech Communication, 72,* 109-125.

Teeranon, P. (2007). The change of standard Thai high tone: An acoustic study and a perceptual experiment. *Journal of Theoretical Linguistics, 4*(3), 1-17.

Thepboriruk, K. (2010). Bangkok Thai tones revisited. *Journal of the Southeast Asian Linguistics Society, 3*(1), 86-105.

Thubthong, N. (2001). *A study of various linguistic effects on tone recognition in Thai continuous speech* (Doctoral dissertation). Chulalongkorn University.

Tingsabadh, K., M.R., & Abramson, A. S. (1993). Thai. *Journal of the International Phonetic Association, 23*(1), 24-28.

Tingsabadh, K., M.R., & Deeprasert, D. (1997). Tones in standard Thai connected speech. *Southeast Asian Linguistic Studies in Honour of Vichin Panupong* (pp. 297-307). Bangkok: Chulalongkorn University Press.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native-and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America, 121*(1), 519-526.

Voiers, W. D. (1983). Evaluating processed speech using the diagnostic rhyme test. *Speech Technology,* 30-39.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*(1), 25-47.