

รายงานวิจัยฉบับสมบูรณ์

การประเมินการจดจำเสียงพูดภาษาไทย
ของระบบประสาทหูเทียมภายใต้สภาพแวดล้อมที่แตกต่างกัน
Evaluation of Thai Word Recognition
for Cochlear Implant Systems in Different Environments



คิริพร เดชะคีลารักษ์ คณวิศวกรรมศาสตร์
ชฎาภรณ์ เหลืองสว่าง คณแพทยศาสตร์
ศิริประภา แก้วศรี คณแพทยศาสตร์
ปฐมดาวดี ดวงตา คณแพทยศาสตร์
มหาวิทยาลัยนเรศวร

สนับสนุนโดย
งบประมาณรายได้มหาวิทยาลัยนเรศวร
ปีงบประมาณ 2559

กิตติกรรมประกาศ

คณะกรรมการเป็นอย่างสูงในความกรุณาของ ผู้ช่วยศาสตราจารย์ ดร.เขมพัฒน์ ตันติวัฒนกุล อาจารย์ประจำภาควิชาศึกษาครรเรื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัย เทคโนโลยีพระจอมเกล้าพระนครเหนือ ที่สละเวลาอันมีค่ามาเป็นที่ปรึกษาโครงการวิจัย พร้อมทั้งให้ คำแนะนำตลอดระยะเวลาในการทำโครงการวิจัยนี้ และขอขอบพระคุณทุกท่านที่เกี่ยวข้องกับ โครงการวิจัยนี้อันประกอบไปด้วย ดร.รัตตินันท์ ภูริวนิชย์กุลและอาจารย์ดา ดาวา อาจารย์ประจำ ภาควิชาบริหารศาสตร์สื่อสารมวลชนและความผิดปกติของการสื่อสารมวลชน คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล ที่ได้กรุณาให้คำแนะนำเกี่ยวกับการประเมินผู้ใช้ประสาททุ เที่ยมตลอดจนข้อมูลที่เป็นประโยชน์ต่อการดำเนินการโครงการวิจัย ผู้ช่วยศาสตราจารย์ ดร.นิพัทธ์ จันทร์มินทร์ อาจารย์ประจำภาควิชาศึกษาไฟฟ้าและคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร ที่ได้กรุณาให้ข้อมูลด้านภาษาศาสตร์ที่เป็นประโยชน์ต่อการวิเคราะห์ข้อมูลและมีส่วนร่วมในการเขียน บทความวิจัย รวมทั้งอาสาสมัครที่ได้สละเวลาอันมีค่าเพื่อเข้าร่วมทดสอบงานทำให้โครงการวิจัยนี้ สำเร็จลุล่วงได้อย่างสมบูรณ์

โครงการวิจัยนี้ได้รับเงินอุดหนุนงบประมาณรายได้ มหาวิทยาลัยนเรศวร ประจำปี งบประมาณ 2559 และได้รับความร่วมมือจากภาควิชาศึกษาไฟฟ้าและคอมพิวเตอร์ คณะ วิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวร ที่อำนวยความสะดวกและเอื้อเทือทัพยการต่างๆ ที่เกี่ยวข้อง สำหรับการทำโครงการวิจัย จนทำให้การศึกษาวิจัยสำเร็จลุล่วงสมบูรณ์ได้ ผู้วิจัยกราบขอบพระคุณ เป็นอย่างสูงไว้ ณ ที่นี่ หนึ่งอิสิ่งอื่นใดขอกราบขอบพระคุณ บิดา มารดา ของผู้วิจัยที่ให้กำลังใจและให้ การสนับสนุนในทุกๆ ด้านอย่างดีที่สุดเสมอมา

ผู้วิจัยหวังเป็นอย่างยิ่งว่าคุณค่าและคุณประโยชน์จากโครงการวิจัยนี้ จะส่วนหนึ่งต่อการ ปรับปรุงคุณภาพชีวิตของผู้ใช้ประสาททุเที่ยมหรือผู้ที่บกพร่องทางการได้ยิน รวมทั้งจะเป็นประโยชน์ ต่อผู้ที่เกี่ยวข้องและผู้ที่สนใจบ้างไม่มากก็น้อย

คณะผู้วิจัย

รายงานสรุป (Executive Summary)

ประสาทหูเทียมหรือ CI (cochlear implant) เป็นอุปกรณ์อิเล็กทรอนิกส์ที่ช่วยให้ผู้บกพร่องทางการได้ยิน (hearing-impaired listeners) ที่มีการสูญเสียการได้ยินในระดับที่ค่อนข้างรุนแรงจนถึงหูหนวก โดยประสาทหูเทียมจะไปทดแทนส่วนของประสาทหูที่ถูกทำลายและใช้การกระตุนด้วยสัญญาณไฟฟ้าไปยังประสาทรับรู้การได้ยิน ทำให้ผู้ใช้ประสาทหูเทียม (CI users) สามารถกลับมาได้ยินเสียงอีกรึว่างผู้ใช้ประสาทหูเทียมสามารถที่จะเข้าใจเสียงพูด (speech intelligibility) และสื่อสารได้ค่อนข้างดีเกื้องเห่ากับคนที่มีการได้ยินปกติ (normal-hearing listeners) ในสภาพแวดล้อมที่ปราศจากเสียงรบกวน (quiet environments) อย่างไรก็ตามความสามารถในการเข้าใจเสียงพูดของผู้ใช้ประสาทหูเทียมจะลดลงอย่างมากในสภาพแวดล้อมที่มีเสียงรบกวน (noisy environments) เนื่องจากปัจจัยที่สำคัญ 2 อย่างคือ ปัจจัยแรกเกี่ยวข้องกับตัวผู้ใช้ประสาทหูเทียม (CI user-related factors) ได้แก่ อายุ เพศ ระยะเวลาในการใช้ประสาทหูเทียม การจัดวางตำแหน่งของอิเล็กtrode อาร์เรย์ในหูชั้นใน เป็นต้น ปัจจัยที่สองเกี่ยวข้องกับตัวอุปกรณ์ประสาทหูเทียม (CI processor-related factors) ได้แก่ รูปแบบการเข้ารหัสเสียงพูด (speech coding strategies) จำนวน อิเล็กtrode ที่ใช้ เป็นต้น ซึ่งระบบประสาทหูเทียมที่ถูกพัฒนาขึ้นจากผู้ผลิตของแต่ละบริษัทจะมีรายละเอียดแตกต่างกัน

ปัจจุบันการศึกษาและงานวิจัยทางด้านนี้ ในประเทศไทยมีค่อนข้างน้อยมาก โดยทั่วไปการประเมินการจัดจำเสียงพูดเพื่อศึกษาว่าผู้ใช้ประสาทหูเทียมสามารถเข้าใจเสียงพูดได้มีประสิทธิภาพมากหรือน้อยจะใช้ภาษาอังกฤษเป็นส่วนใหญ่ ซึ่งมีภาษาอังกฤษมีความแตกต่างจากภาษาไทย ภาษาอังกฤษเป็นภาษาที่เน้นคำและไม่มีเสียงวรรณยุกต์ แต่ภาษาไทยเป็นภาษาที่มีเสียงวรรณยุกต์ (tonal language) ในคำเดียวกันแต่การออกเสียงด้วยเสียงวรรณยุกต์ต่างกัน ทำให้เกิดความหมายที่แตกต่างกันได้ ดังนั้นโครงการวิจัยนี้ต้องการประเมินการจัดจำเสียงพูดภาษาไทยของผู้ใช้ระบบประสาทหูเทียมภายใต้สภาพแวดล้อมที่มีเสียงรบกวนแตกต่างกัน โดยพิจารณาพารามิเตอร์ที่สำคัญที่เกี่ยวข้องกับการทำงานของระบบประสาทหูเทียมที่อาจส่งผลต่อความเข้าใจในเสียงพูดของผู้ใช้ประสาทหูเทียม ดังนี้ การเข้ารหัสเสียงพูดแบบ CIS (Continuous Interleaved Sampling) และ ACE (Advanced Combination Encoding) การลดเสียงรบกวนด้วยเทคนิค MBSS (multi-band spectral subtraction) และ WF (Wiener filter) ประเภทของเสียงรบกวนแบบ SSN (speech-shaped noise) และ BBN (babble noise) ที่ระดับอัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวนหรือ SNR (signal-to-noise ratio) เท่ากับ 0, 5, 10 และ 15 เดซิเบล

การประเมินการจัดจำเสียงพูดภาษาไทยหรือการเข้าใจเสียงพูดจะใช้ชุดคำพยางค์เดียว (monosyllabic words) และคำพยางค์คู่ (bisyllabic words) มาผสมกับเสียงรบกวนตามเงื่อนไขที่กำหนด การประเมินจะแบ่งออกเป็น 2 การทดลอง การทดลองแรกเพื่อศึกษาเบรียบเทียบประสิทธิภาพการเข้าใจเสียงพูดของการเข้ารหัสเสียงพูดแบบ CIS และ ACE โดยนำชุดคำห้องหมอดที่ไม่มีเสียงรบกวน (clean words) และคำที่มีเสียงรบกวน (noisy words) มาทำการเข้ารหัสเสียงจะได้เสียงจำลอง (vocoded speech) ซึ่งเป็นเสียงจำลองได้จากการประมวลผลสัญญาณของระบบ

ประชาทหุเที่ยม ลักษณะของเสียงจำลองที่ได้จะมีความแตกต่างจากเสียงพูดที่ได้ยินทั่วไปในชีวิตประจำวันและค่อนข้างฟังเข้าใจยากเล็กน้อย เสียงจำลองเหล่านี้จะถูกนำมาทดสอบกับอาสาสมัครที่มีการได้ยินปกติจำนวน 12 คน โดยก่อนทำการทดสอบจริง (actual test) อาสาสมัครจะต้องทดลองฟังเสียงจำลอง (training test) เพื่อให้เกิดความคุ้นเคยต่อเสียงจำลองก่อน เป็นเวลาประมาณ 5 นาที โดยอาสาสมัครจะฟังเสียงจากหูฟังที่ต่อเข้ากับเครื่องคอมพิวเตอร์ โดยเลือกฟังจากหูซ้ายหรือหูขวาเพียงข้างใดข้างหนึ่งเท่านั้น ในการทดสอบจริง อาสาสมัครจะได้ฟังคำในแต่ละเงื่อนไขและเขียนคำที่ได้ยินลงบนกระดาษที่ได้จัดเตรียมไว้ การนับคะแนนการจำเสียงพูดจะนับคำที่อาสาสมัครเข้าใจถูกต้องโดยคิดเป็นปอร์เซนต์ของคำที่ถูกต้อง จากผลการทดลองแสดงให้เห็นว่าผู้ฟังสามารถเข้าใจเสียงพูดที่มีการเข้ารหัสเสียงแบบ ACE ได้ต่ำกว่าแบบ CIS ในทุกเงื่อนไขที่กำหนด

การทดลองที่สองเพื่อศึกษาเบรียบเทียบประสิทธิภาพการเข้าใจเสียงพูดของการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF โดยชุดคำทั้งหมดที่มีเสียงรบกวนแบบ SSN และ BBN จะผ่านขั้นตอนการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF จะได้สัญญาณเสียงที่ถูกปรับปรุงแล้ว (enhanced words) จากนั้นเสียงที่ถูกปรับปรุงแล้วจะนำไปเข้ารหัสเสียงแบบ ACE เพื่อสร้างเป็นเสียงจำลองของระบบประชาทหุเที่ยม (vocoded speech) เสียงจำลองเหล่านี้จะถูกนำมาทดสอบกับอาสาสมัครที่มีการได้ยินปกติจำนวน 20 คน ขั้นตอนการประเมินการเข้าใจเสียงพูดจะทำเช่นเดียวกับการทดลองแรก ผลการทดลองแสดงให้เห็นว่าประสิทธิภาพของความเข้าใจเสียงพูดด้วยการลดเสียงรบกวนแบบ WF จะช่วยให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นอย่างเห็นได้ชัดในเงื่อนไขที่กำหนดเป็นส่วนใหญ่ ส่วนเทคนิคการลดเสียงรบกวนแบบ MBSS จะช่วยให้ผู้ฟังสามารถเข้าใจเสียงพูดเพียงเล็กน้อยในบางเงื่อนไขเท่านั้น เช่น เสียงรบกวนแบบ BBN ที่ระดับ SNR เท่ากับ 5 เดซิเบล เป็นต้น ผู้ฟังสามารถเข้าใจคำพยางค์คู่ได้กว่าคำพยางค์เดียวอย่างเห็นได้ชัดในทุกเงื่อนไขที่กำหนด จากการทดลองที่ได้นี้ สามารถนำไปประยุกต์ใช้เพื่อทำนายแนวโน้มของผลที่คาดว่าจะได้รับจากผู้ใช้ประชาทหุเที่ยมที่ใช้ภาษาไทยได้และนำไปสู่การพัฒนาระบบประชาทหุเที่ยมที่เหมาะสมสำหรับคนไทยต่อไป

ชื่อเรื่อง	การประเมินการจดจำเสียงพูดภาษาไทยของระบบประสาทหูเทียมภายใต้สภาวะแวดล้อมที่แตกต่างกัน Evaluation of Thai Word Recognition for Cochlear Implant Systems in Different Environments
ผู้วิจัย	ศิริพร เดชาศิลารักษ์ ชญาธาร เหลืองสว่าง ศิรประภา แก้วศรี ปฐมาวดี ดวงตา

บทคัดย่อ

ผู้ใช้ประสาทหูเทียมมักจะประสบกับปัญหาในการสื่อสารร่วมกับผู้อื่นเมื่ออยู่ในสภาพแวดล้อมของการฟังที่มีเสียงรบกวน นักวิจัยได้ให้ความสนใจและพยายามแก้ปัญหานี้ อย่างไรก็ตามงานวิจัยด้านประสาทหูเทียมโดยส่วนใหญ่ จะทำการประเมินประสิทธิภาพการเข้าใจเสียงพูดโดยใช้ภาษาอังกฤษ ในการศึกษานี้ได้ใช้เทคนิคการลดเสียงรบกวนแบบช่องทางเดียวเพื่อใช้เป็นขั้นตอนการประมวลผล ก่อนที่จะนำสัญญาณที่ได้ไปป้อนเข้าระบบประสาทหูเทียม การลดเสียงรบกวนด้วยเทคนิคการลบสเปกตรัมแบบหลายแบบด้วยตัวกรองเวเนียร์ จะถูกนำมาใช้เพื่อประเมินประสิทธิภาพการเข้าใจเสียงพูดของผู้ใช้ประสาทหูเทียม เสียงที่ใช้ทดสอบประกอบด้วยชุดคำพยางค์เดียวและคำพยางค์คู่ที่มีการผสมเสียงรบกวนสองแบบคือ babble noise และ speech-shaped noise ที่ระดับของอัตราส่วนสัญญาณเสียงพูดต่อเสียงรบกวนเท่ากับ 0, 5 และ 10 เดซิเบล สัญญาณเสียงของชุดคำที่มีเสียงรบกวนจะถูกลดเสียงรบกวนด้วยเทคนิคห้องสอง จากนั้นสัญญาณเสียงที่ผ่านการลดสัญญาณรบกวนแล้วจะถูกนำไปเข้ารหัสเสียงพูดแบบ ACE ที่มีการמודูลेटด้วยสัญญาณรบกวน เพื่อสร้างเสียงจำลองของระบบประสาทหูเทียม เสียงจำลองทั้งหมดจะถูกนำมาใช้ทดสอบกับอาสาสมัครที่มีการได้ยินปกติจำนวน 20 คน ผลการทดลองแสดงให้เห็นว่าประสิทธิภาพของความเข้าใจเสียงพูดเมื่อทำการลดเสียงรบกวนด้วยเทคนิคตัวกรองเวเนียร์ ทำให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นอย่างเห็นได้ชัดในเงื่อนไขที่กำหนดเป็นส่วนใหญ่ สำนวนเทคนิคการลบสเปกตรัมแบบหลายแบบด้วยตัวกรองเวเนียร์ ทำให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีเพียงเล็กน้อยในบางเงื่อนไขที่กำหนดเท่านั้น ผู้ฟังสามารถเข้าใจคำพยางค์คู่ได้กว่าคำพยางค์เดียวอย่างเห็นได้ชัดในทุกเงื่อนไขที่กำหนด โดยเฉพาะอย่างยิ่งในเสียงรบกวนแบบ speech-shaped noise จากผลการทดลองที่ได้ตั้งกล่าวจะเป็นประโยชน์สำหรับผู้ใช้ประสาทหูเทียมสำหรับคนไทยต่อไป

คำสำคัญ ประสาทหูเทียม, การลดสัญญาณรบกวน, การลบสเปกตรัมของสัญญาณ, ตัวกรองเวเนียร์, ความเข้าใจเสียงพูด

Abstract

Cochlear implant (CI) listeners encounter difficulties in communicating with other persons in noisy listening environments. Many researchers have attempted to solve these problems. However, most CI research has been carried out using the English language. In this study, single-channel noise reduction (NR) strategies as a pre-processing approach for the CI system were investigated in terms of Thai speech intelligibility improvement. Two NR algorithms, namely multi-band spectral subtraction (MBSS) and Wiener filter (WF) algorithms, were evaluated. Speech signals consisting of monosyllabic and bisyllabic Thai words were degraded by two noise types (speech-shaped noise and babble noise) at different SNR levels (0, 5, and 10 dB). Then the noisy words were enhanced using NR algorithms. The enhanced words were fed into the advanced combinational encoder with a noise-band vocoder to synthesize the vocoded speech. The vocoded speech was presented to twenty normal-hearing listeners. The results indicated that intelligibility was marginally improved for the MBSS algorithm and significantly improved for the WF algorithm in some conditions. The enhanced bisyllabic words had a noticeably higher intelligibility improvement than the enhanced monosyllabic words in all conditions, particularly in speech-shaped noise. Such outcomes may be beneficial to Thai-speaking CI listeners.

Key words cochlear implant, noise reduction, spectral subtraction, Wiener filter, speech intelligibility

สารบัญ

หน้า

กิจกรรมประจำ	ก
รายงานสรุป	ข
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
สารบัญภาพ	ช
สารบัญตาราง	ณ

บทที่ 1 บทนำ

1.1 ความสำคัญและที่มาของปัญหา	1
1.2 วัตถุประสงค์	3
1.3 ขอบเขตของงานวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3

บทที่ 2 การลดสัญญาณระบบในระบบประสานหูเทียม

2.1 ระบบประสานหูเทียม	5
2.2 การเข้ารหัสเสียงพูด	5
2.3 เทคนิคการลดสัญญาณรบกวน	8
2.4 สรุป	10

บทที่ 3 การประเมินประสิทธิภาพการเข้าใจเสียงพูด

3.1 การเปรียบเทียบประสิทธิภาพของการเข้ารหัสเสียงพูด	16
3.1.1 อาสาสมัคร	16
3.1.2 คำที่ใช้ทดสอบ	16
3.1.3 ขั้นตอนการประเมิน	17
3.2 การเปรียบเทียบประสิทธิภาพของเทคนิคการลดสัญญาณรบกวน	18
3.2.1 อาสาสมัคร	18
3.2.2 คำที่ใช้ทดสอบ	18
3.2.3 ขั้นตอนการประเมิน	19

บทที่ 4 ผลการทดลองและการวิเคราะห์ข้อมูลที่ได้

4.1 การเปรียบเทียบประสิทธิภาพของการเข้ารหัสเสียงพูดแบบ CIS และ ACE	20
4.2 การเปรียบเทียบประสิทธิภาพของการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF	23

สารบัญ (ต่อ)

หน้า

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	30
5.1 สรุปผลการวิจัย	30
5.2 ข้อเสนอแนะ	31

บรรณานุกรม	33
ภาคผนวก	37
ภาคผนวก ก	38
ภาคผนวก ข	39
ภาคผนวก ค	49



สารบัญภาพ

รูปที่	หน้า
2.1 บล็อกไดอะแกรมของระบบประสาทหูเทียน	6
2.2 บล็อกไดอะแกรมการสร้างเสียงพูดจำลองที่มีการเข้ารหัสแบบ ACE	7
2.3 บล็อกไดอะแกรมการสร้างเสียงพูดจำลองของระบบประสาทหูเทียนร่วมกับอัลกอริทึมของการลดสัญญาณรบกวน	12
4.1 สเปกโตรแกรมของคำพยางค์เดียว “ยาม” ที่มีการเข้ารหัสเสียงแบบ CIS และ ACE	20
4.2 สเปกโตรแกรมของคำพยางค์เดียว “เรือ” และคำพยางค์คู่ “เล็กน้อย” ที่มีการเข้ารหัสเสียงแบบ ACE	21
4.3 เปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องของเงื่อนไขที่กำหนดจำนวน 28 เงื่อนไข	22
4.4 ตัวอย่างรูปคลื่นสัญญาณของคำพยางค์เดียว “ยาม” และคำพยางค์คู่ “แน่นอน”	24
4.5 ตัวอย่างอิเล็กโทรได้แกรมของคำพยางค์เดียว “ยาม” และคำพยางค์คู่ “แน่นอน”	25
4.6 เปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องของอาสาสมัครที่มีการได้ยินปกติจำนวน 20 คน	26
4.7 ความแตกต่างสัมพัทธ์ของเปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้อง	28

สารบัญตาราง

ตารางที่	หน้า
2.1 คุณลักษณะของแบบคุณภาพที่จำนวน 22 แบบ	9
3.1 การกำหนดเงื่อนไขที่ใช้ทดสอบสำหรับการเข้ารั้วเสียง	17
3.2 การกำหนดเงื่อนไขที่ใช้ทดสอบสำหรับเทคนิคการลดเสียงรอบกวน	19



บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหา

จากรายงานขององค์การอนามัยโลก (world health organization) พบว่าในปี 2018 มีผู้ที่บกพร่องทางการได้ยิน (hearing-impaired listeners) ประมาณ 466 ล้านคน (มากกว่า 5% ของประชากรโลก) และคาดการณ์ว่าในปี 2050 จะมีผู้ที่บกพร่องทางการได้ยินมากกว่า 900 ล้านคน หรืออาจกล่าวได้ว่าในทุกๆ ประชากร 10 คนจะมีผู้ที่บกพร่องทางการได้ยิน 1 คน นอกจากนี้ยังพบว่าประชากรที่มีอายุระหว่าง 12 ถึง 35 ปีประมาณ 1.1 พันล้านคน มีความเสี่ยงต่อความบกพร่องทางการได้ยินเนื่องจากการได้ยินเสียงรบกวนเป็นเวลากวนในสถานบันเทิงต่างๆ โดยทั่วไปผู้ที่บกพร่องทางการได้ยินจะมีการสูญเสียการได้ยิน (hearing loss) ตั้งแต่ระดับน้อย (mild hearing loss) จนถึงระดับรุนแรง (profound hearing loss) เนื่องมาจากหล่ายาเหตุ เช่น พันธุกรรม ภาวะแทรกซ้อน โรคติดเชื้อบางชนิด การติดเชื้อที่หูเรื้อรัง การใช้ยาบางชนิด การได้ยินเสียงรบกวนมากเกินไปเป็นเวลากวน การมีอายุมากขึ้น เป็นต้น

สำหรับผู้ที่สูญเสียการได้ยินในระดับรุนแรงหรือผู้ที่หูหนวกโดยกำเนิดจะไม่สามารถใช้ประโยชน์จากเครื่องช่วยฟัง (hearing aids) ได้ จึงจำเป็นต้องใช้ประสาทหูเทียม (cochlear implants) เพื่อช่วยในการได้ยินและสามารถสื่อสารในชีวิตประจำวันได้ การทำงานของเครื่องช่วยฟังและประสาทหูเทียมมีความแตกต่างกัน กล่าวคือเครื่องช่วยฟังจะทำหน้าที่ในการขยายเสียงให้ดังขึ้นเท่านั้น (acoustic amplifier) แต่ประสาทหูเทียมจะไปทดแทนส่วนของประสาทหูที่ถูกทำลายไปและใช้การกระตุนด้วยสัญญาณไฟฟ้าไปยังประสาทรับรู้เสียง (electric acoustic stimulation) ทำให้ผู้ใช้ประสาทหูเทียม (cochlear implant users) สามารถยินเสียงได้

ในปี 2012 พบว่ามีผู้ใช้ประสาทหูเทียมทั่วโลกประมาณ 320,000 คนและจากข้อมูลอย่างไม่เป็นทางการของผู้ผลิตเปิดเผยว่าในช่วงต้นปี 2016 มีผู้ใช้ประสาทหูเทียมจำนวนมากกว่า 500,000 คน (Wilson, 2017) โดยทั่วไปผู้ใช้ประสาทหูเทียมส่วนใหญ่สามารถเข้าใจเสียงพูดได้มากกว่า 80% หรือเกือบทั้งหมดกับคนที่มีการได้ยินปกติ เมื่ออุย្ញในสภาพแวดล้อมที่เงียบสงบปราศจากเสียงรบกวนใดๆ (quiet environments) อย่างไรก็ตามผู้ใช้ประสาทหูเทียมจะสามารถเข้าใจเสียงพูดลดลงอย่างมากเมื่ออุย្ញในสภาพแวดล้อมที่มีเสียงรบกวน (noisy environments) (Kokkinakis, Azimi, Hu, & Friedland, 2012) ซึ่งสามารถพบรูปได้ในชีวิตประจำวัน เช่น เสียงผู้คนที่กำลังสนทนากันในร้านอาหารเสียงรถบนถนนทาง เสียงรถไฟ เสียงเครื่องบิน เป็นต้น หากการดับของเสียงรบกวนมีปริมาณมากเท่าไร ความสามารถในการเข้าใจเสียงพูด (speech intelligibility) ของผู้ใช้ประสาทหูเทียมจะลดน้อยลงอย่างมากเมื่อเทียบกับคนที่มีการได้ยินปกติ (normal-hearing listeners) (Hu, Loizou, Li, & Kasturi, 2007) ทั้งนี้เนื่องมาจากการข้อจำกัดของตัวอุปกรณ์ที่ให้รายละเอียดของข้อมูลเชิงความถี่ (frequency) เวลา (time) และแอมพลิจูด (amplitude) ได้น้อย ซึ่งส่งผลต่อการส่งข้อมูลเสียงพูดไปยังระบบประสาทการได้ยิน รวมทั้งข้อจำกัดของผู้ที่สูญเสียการได้ยินเอง ที่มีความสามารถในการ

แยกแยะเสียงพูดและเสียงรบกวนได้น้อยกว่าคนที่ได้ยินปกติ (Qin & Oxenham, 2003) ดังนั้นนักวิจัยด้านระบบประสาทหูเทียมได้พยาบยามที่จะปรับปรุงประสิทธิภาพการเข้าใจเสียงพูด (speech intelligibility performance) โดยพัฒนารูปแบบการเข้ารหัสเสียงพูด (speech coding) และอัลกอริทึมการลดเสียงรบกวน (noise reduction algorithms) เพื่อให้ผู้ใช้ประสาทหูเทียมสามารถสื่อสารและเข้าใจเสียงพูดได้ดีและมีประสิทธิภาพมากขึ้นในสภาพแวดล้อมที่มีเสียงรบกวนที่มีความหลากหลาย

อย่างไรก็ตามการศึกษาเกี่ยวกับการประเมินประสิทธิภาพการเข้าใจเสียงพูดของเทคนิคการลดเสียงรบกวนจะใช้ภาษาอังกฤษ (English language) ในการทดสอบเป็นส่วนใหญ่ สำหรับภาษาอื่น มีค่อนข้างน้อยมาก เช่น ภาษาฝรั่งเศส (French) (Kallel, Frika, Ghorbel, Ben Hamida, & Berger-Vachon, 2012) ภาษา希伯来语 (Hebrew) (Fink, Furst, & Muchnik, 2012) ภาษาดัตช์หรือเฟรนซ์ (Dutch/Flemish) (Koning, Madhu, & Wouters, 2015) และภาษาจีน (Chinese) (F. Chen, Hu, & Yuan, 2015) เป็นต้น สำหรับภาษาไทยนั้นยังไม่มีงานวิจัยใดที่ทำการศึกษาการประเมินประสิทธิภาพการเข้าใจเสียงพูดของเทคนิคการลดเสียงรบกวนสำหรับผู้ใช้ประสาทหูเทียม เนื่องจากแต่ละภาษามีเสียงที่เป็นเอกลักษณ์ ลักษณะเฉพาะของเสียง (acoustic cues) และหน่วยพื้นฐานของเสียง (phonemics) แตกต่างกัน เมื่อใช้เทคนิคการลดเสียงรบกวนเทคนิคเดียวกัน อาจทำให้ประสิทธิภาพการเข้าใจเสียงพูดมีแตกต่างกันได้

ภาษาอังกฤษและภาษาไทยเป็นภาษาที่มีความแตกต่างกันดังนี้ ภาษาอังกฤษเป็นภาษาที่ไม่มีเสียงวรรณยุกต์ (non-tonal language) ขณะที่ภาษาไทยเป็นภาษาที่มีเสียงวรรณยุกต์ (tonal language) เช่นเดียวกับภาษาในแถบเอเชีย (Asian languages) เช่น ภาษาจีน (Chinese) ภาษาเวียดนาม (Vietnamese) เป็นต้น ภาษาที่มีเสียงวรรณยุกต์จะใช้ระดับของเสียงวรรณยุกต์แสดงความหมายของคำ (lexical meaning) เสียงวรรณยุกต์แต่ละเสียงที่ใช้ในคำนั้นๆ จะให้ความหมายต่างกัน โดยทั่วไปพยางค์ไทยนั่งพยางค์ประกอบด้วยพยัญชนะต้น (initial consonant) สระ (vowel) ที่อาจเป็นสระเสียงสั้นหรือยาว พยัญชนะท้าย (final consonant) ที่อาจจะมีหรือไม่มีก็ได้ และเสียงวรรณยุกต์

เสียงวรรณยุกต์ไทยแบ่งออกเป็น 5 เสียง ดังนี้ เสียงสามัญ (middle) / / เสียงเอก (low) // เสียงโท (falling) // เสียงตรี (high) // เสียงจัตวา (rising) // ภาษาไทยจะมีทั้งคำที่เป็นคำพยางค์เดียว (monosyllabic words) และคำที่มีหลายพยางค์ (polysyllabic words) จำนวนมากมาย คำที่มีหลายพยางค์อาจเกิดจากการนำคำพยางค์เดียวมาร่วมกัน คำพยางค์เดียวที่มีเสียงวรรณยุกต์ต่างกันจะมีความหมายต่างกันด้วย เช่น ปา ป้า ป้า ป้า เป็นต้น จะเห็นได้ว่าองค์ประกอบของคำไทยต่างๆ เป็นสิ่งสำคัญที่จะนำไปสู่การการศึกษาและวิเคราะห์เกี่ยวกับการประเมินประสิทธิภาพการเข้าใจเสียงพูดของเทคนิคการลดเสียงรบกวนของผู้ใช้ประสาทหูเทียมสำหรับคนไทย

ดังนั้นโครงการวิจัยนี้ต้องการศึกษาเกี่ยวกับการทำงานระบบประสาทหูเทียมและพารามิเตอร์สำคัญที่เกี่ยวข้องกับการทำงานของระบบประสาทหูเทียมที่อาจส่งผลต่อความเข้าใจเสียงพูดภาษาไทยของผู้ใช้ประสาทหูเทียมที่อยู่ในสภาพแวดล้อมที่แตกต่างกัน เช่น รูปแบบการเข้ารหัสเสียงพูด ประเภทของเสียงรบกวน ระดับของเสียงรบกวน วิธีการลดเสียงรบกวน เป็นต้น เพื่อเป็นแนวทางในการพัฒนาและออกแบบระบบประสาทหูเทียมที่เหมาะสมต่อไป ดังนั้นในวิจัยนี้จะ

ทำการศึกษาพารามิเตอร์ที่เกี่ยวข้องกับการออกแบบภาษาที่เปลี่ยน เพื่อตรวจสอบและประเมินความสามารถในการเข้าใจเสียงพูดภาษาไทย (Wutiwiwatchai and Furui, 2009) ของผู้ใช้ภาษาที่เปลี่ยนในสภาพแวดล้อมต่างกัน

1.2 วัตถุประสงค์

- 1.2.1 เพื่อศึกษาและเข้าใจการเข้ารหัสเสียงพูดของระบบภาษาที่เปลี่ยนรูปแบบต่างๆ
- 1.2.2 เพื่อศึกษาและเข้าใจปัจจัยหรือพารามิเตอร์ที่สำคัญที่เกี่ยวข้องกับการเข้าใจเสียงพูด เช่น ประเภทของเสียงรบกวน ระดับของเสียงรบกวน วิธีการลดเสียงรบกวน เป็นต้น
- 1.2.3 พัฒนาโปรแกรมเพื่อใช้ในการสร้างเสียงจำลองที่ได้จากการประมวลผลของระบบภาษาที่เปลี่ยน
- 1.2.4 เพื่อประเมินการจัดจำเสียงพูดภาษาไทยของระบบภาษาที่เปลี่ยนในสภาพแวดล้อมที่แตกต่างกัน

1.3 ขอบเขตของงานวิจัย

คณผู้จัดได้กำหนดขอบเขตเนื้อหาของโครงการวิจัย ดังต่อไปนี้

- 1.3.1 ศึกษาแนวคิดและระบบการทำงานของระบบภาษาที่เปลี่ยน
- 1.3.2 ศึกษาวิธีการเข้ารหัสเสียงพูดของระบบภาษาที่เปลี่ยนในรูปแบบ CIS (Continuous Interleaved Sampling) และแบบ ACE (Advanced Combination Encoding)
- 1.3.3 ศึกษาวิธีการลดเสียงรบกวนแบบ speech-shaped noise (SSN) และ babble noise (BBN) ที่ระดับอัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวนหรือ SNR (signal-to-noise ratio) เท่ากับ 0, 5, 10 และ 15 เดซิเบล ด้วยเทคนิคการลบสเปกตรัมแบบหลายแบนด์หรือ MBSS (multi-band spectral subtraction) และตัวกรองเวนีเยอร์ หรือ WF (Wiener filter)
- 1.3.4 ออกแบบและพัฒนาโปรแกรมเพื่อสร้างเสียงของระบบภาษาที่เปลี่ยนด้วย MATLAB
- 1.3.5 ศึกษาวิธีการประเมินการเข้าใจเสียงพูดของเสียงจำลองโดยใช้อาสาสมัครที่มีการได้ยินปกติ
- 1.3.6 ศึกษาผลลัพธ์ที่ได้จากการทดลองอาสาสมัครที่มีการได้ยินปกติ เพื่อใช้ทำนายผลที่คาดว่าจะได้รับจากผู้ใช้ภาษาที่เปลี่ยน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 เกิดความรู้ความเข้าใจเกี่ยวกับการเข้าใจในเสียงพูดภาษาไทยของผู้ใช้ภาษาที่เปลี่ยนที่สภาพแวดล้อมที่แตกต่างกัน
- 1.4.2 ได้พัฒนาโปรแกรมการสร้างเสียงจำลองของระบบภาษาที่เปลี่ยน
- 1.4.3 องค์ความรู้ที่ได้สามารถนำไปใช้ เพื่อพัฒนาระบบภาษาที่เปลี่ยนให้เหมาะสมกับคนไทยต่อไปได้
- 1.4.4 เพื่อพัฒนาบุคลากรให้มีความรู้ความสามารถในการประมวลผลสัญญาณ

เสียงพูดมากขึ้น

- 1.4.5 เพื่อให้เกิดการศึกษาค้นคว้าเทคโนโลยีใหม่ของนิสิตและคณาจารย์ของมหาวิทยาลัย รวมทั้งความรู้ที่เป็นประโยชน์ต่อบุคลากรทางการแพทย์ ผู้ที่ปฏิบัติงานเกี่ยวกับ อุปกรณ์เครื่องช่วยในการได้ยิน รวมทั้งผู้สนใจด้านเสียงพูดและการได้ยิน 1.4.6 สามารถนำไปประยุกต์ใช้กับงานด้านอื่นที่เกี่ยวข้อง เช่น การพัฒนาหุ่นยนต์ อุปกรณ์ สื่อสาร ระบบรักษาความปลอดภัย เป็นต้น



บทที่ 2

การลดสัญญาณรบกวนในระบบประสาทหูเทียม

ในเนื้อหาบทนี้จะกล่าวถึงระบบประสาทหูเทียม การเข้ารหัสเสียงพูดของระบบประสาทหูเทียม เทคนิคการลดเสียงรบกวนในระบบประสาทหูเทียม

2.1 ระบบประสาทหูเทียม (Cochlear Implant System)

ระบบประสาทหูเทียมช่วยให้ผู้ที่สูญเสียการได้ยิน (hearing loss) ในระดับรุนแรงจนถึงไม่ได้ยินเสียงหรือหูหนวกสามารถกลับมามาได้ยินเสียงอีกรัง ผู้ใช้ประสาทหูเทียมสามารถสื่อสารกับผู้อื่นได้โดยไม่ต้องใช้การอ่านริมฝีปากช่วย (lip reading) และบางคนสามารถสื่อสารผ่านทางโทรศัพท์ได้ เช่นเดียวกับคนที่ได้ยินปกติ (P.C. Loizou, 1998) การพัฒนาระบบประสาทหูเทียมได้ประสบความสำเร็จเป็นอย่างมาก ซึ่งเกิดจากความร่วมมือกันระหว่างนักวิทยาศาสตร์และผู้ที่เกี่ยวข้องในหลายสาขาวิชา เช่น นักโสตสัมผัสร่วม (audiologists) นักแก้ไขการพูด (speech therapists) ผู้เชี่ยวชาญด้านหูคอจมูก (otolaryngologists) วิศวกรรมชีวทางการแพทย์ (biomedical engineering) เป็นต้น ทำให้เกิดมุมมองและแนวคิดในการออกแบบและพัฒนาระบบประสาทหูเทียม ที่มีความหลากหลายและก่อให้เกิดประโยชน์อย่างสูงสุดต่อผู้ใช้ประสาทหูเทียม โดยเฉพาะอย่างยิ่งการประมวลผลสัญญาณ (signal processing) มีบทบาทสำคัญอย่างยิ่งต่อการพัฒนาเทคโนโลยีที่ใช้ในระบบประสาทหูเทียมเพื่อทำการแปลงจากสัญญาณเสียงพูด (speech signals) ไปเป็นสัญญาณกระแสตุ้นทางไฟฟ้า (electrical stimuli) ที่สามารถเลียนแบบการทำงานของคอเคลียที่ปกติได้ (healthy cochlea)

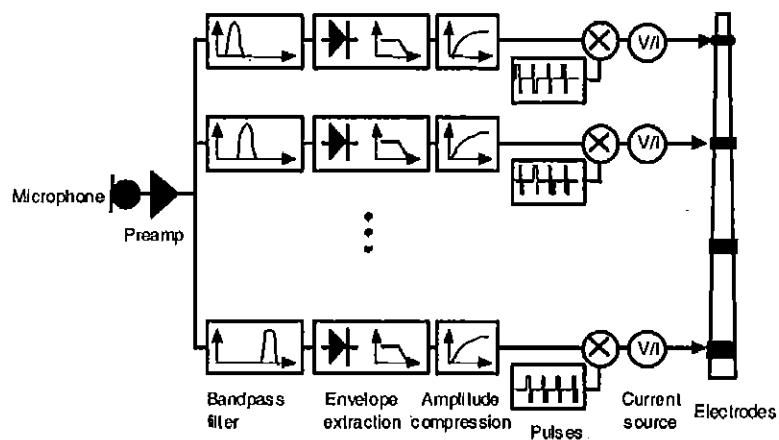
ระบบประสาทหูเทียมแบบช่องเดียว (single-channel implants) ได้ถูกพัฒนาขึ้นครั้งแรกในช่วงต้นทศวรรษ 1970 หรือช่วงระยะเวลา 40 ปีที่ผ่านมา (P.C. Loizou, 2006) แต่ยังคงมีข้อจำกัดในเรื่องการส่งข้อมูลเสียงที่แคบความถี่ (bandwidth) จำกัดไปยังคอเคลียที่เป็นส่วนสำคัญในการรับรู้เสียง ต่อมาระบบประสาทหูเทียมได้ถูกพัฒนาเป็นแบบหลายช่อง (multi-channel implants) และได้ถูกนำมาใช้จนถึงปัจจุบัน เพื่อให้สามารถส่งข้อมูลที่มีองค์ประกอบหลายแบบด้วยความถี่ได้จำนวนแบบต่ำความถี่ที่มากทำให้มีรายละเอียดของความถี่มาก (frequency resolution) ทำให้เกิดการทำงานมีความใกล้เคียงกับคอเคลียมากขึ้นด้วย ทำให้ผู้ใช้ประสาทหูเทียมสามารถเข้าใจเสียงพูดได้ดีขึ้น ในปัจจุบันบริษัทผู้ผลิตระบบประสาทหูเทียมได้พยายามที่จะพัฒนาการเข้ารหัสเสียงพูดในหลายรูปแบบมากขึ้น สำหรับระบบประสาทหูเทียมแบบหลายช่อง สำหรับบริษัทที่มีส่วนแบ่งการตลาดที่มากที่สุดของโลก ได้แก่ บริษัท Med-El (ออสเตรีย) บริษัท Nucleus (ออสเตรเลีย) และ บริษัท Clarion II (สหรัฐอเมริกา) เป็นต้น

การเข้ารหัสเสียงพูดเป็นส่วนที่สำคัญของระบบประสาทหูเทียมซึ่งทำหน้าที่หลักในการแปลงสัญญาณเสียงพูดไปเป็นสัญญาณกระแสตุ้นทางไฟฟ้า ทำให้ผู้ใช้ประสาทหูเทียมสามารถเข้าใจและสามารถสื่อสารได้ดีขึ้น รูปแบบการเข้ารหัสเสียงสามารถแบ่งออกเป็น 2 ประเภท คือ ประเภทแรกเป็น

การนำข้อมูลเออนเวโลปหรือกรอบของสัญญาณ (envelope channel) ของทุกแบบด้วยความถี่ส่งไปยัง อิเล็กโทรด เช่น CIS และ HiResolution (HiRes) เป็นต้น ประเภทที่สองเป็นการเลือกข้อมูลเออนเวโลป บางแบบด้วยความถี่เพื่อส่งไปยังอิเล็กโทรด เช่น Spectral Peak (SPEAK) และ ACE เป็นต้น การเข้ารหัสเสียงแบบ CIS จะเป็นรูปแบบที่ถูกกำหนดไว้เป็นทางเลือกหนึ่งของระบบประสาทหูเทียม สำหรับทุกบริษัทผู้ผลิต (Fan-Gang, Rebscher, Harrison, Xiaoan, & Haihong, 2008; Wilson & Dorman, 2008) แต่วิธีการออกแบบและสร้างการเข้ารหัสเสียงแบบ CIS ของแต่ละบริษัทที่ผลิตจะมีรายละเอียดแตกต่างกัน สำหรับรูปแบบการเข้ารหัสเสียงแบบนี้จะถูกกำหนดให้รูปแบบหลักของระบบประสาทหูเทียมขึ้นกับบริษัทผู้ผลิตนั้นๆ รายละเอียดเพิ่มเติมสามารถศึกษาได้จากการงานของ P.C. Loizou (2006)

โดยทั่วไปการเข้ารหัสเสียงประกอบด้วยขั้นตอนที่สำคัญดังนี้ ชุดແຕบความถี่ (bank of bandpass filter) การตรวจจับเออนเวโลปหรือกรอบของสัญญาณ (envelop detection) การบีบ ข้อมูลสัญญาณ (compression) และการ modulation (modulation) แต่ละบริษัทผู้ผลิตจะมีรายละเอียดการออกแบบแต่ละขั้นตอนแตกต่างกัน ตัวอย่างเช่น ชุดແຕบความถี่สามารถออกแบบโดยใช้ผลตอบสนองอิมพัลส์จำนวนจำกัดหรือ FIR (finite impulse response) ผลตอบสนองอิมพัลส์จำนวนไม่จำกัดหรือ IIR (infinite impulse response) การแปลงฟูเรียร์แบบเร็ว หรือ FFT (fast Fourier transform) เป็นต้น หรือการตรวจจับเออนเวโลปในแต่ละแบบด้วยความถี่ของสัญญาณอาจใช้วิธีการแปลงซีลเบิร์ต (Hilbert transform) การใช้งานเรียงกระแสแบบเต็มคลื่น (full-wave rectification) ร่วมกับตัวกรองແຕบความถี่ผ่านหรือ LPF (lowpass filter) เป็นต้น

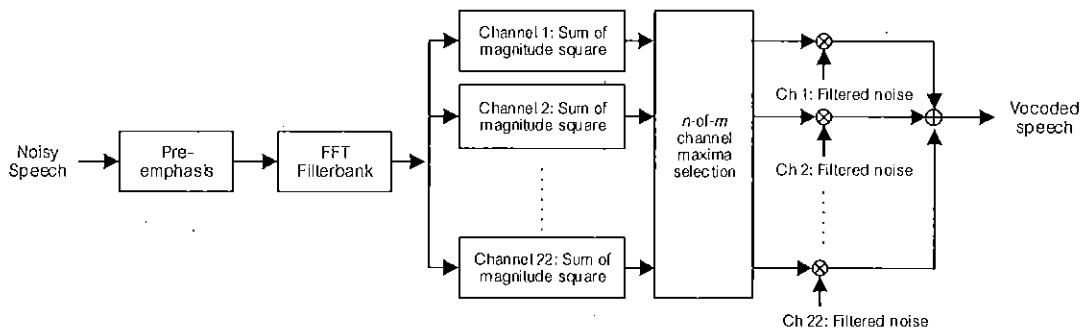
หลักการทำงานเบื้องต้นของระบบประสาทหูเทียมแสดงดังรูปที่ 2.1 (Fan-Gang et al., 2008) เริ่มจากไมโครโฟน (microphone) ทำการรับสัญญาณเสียงที่เข้ามา สัญญาณเสียงจะผ่านวงจรกรองพรี-emphasis (pre-emphasis filter) เพื่อทำให้ข้อมูลคงค่าประกอบความถี่สูงมีความชัดเจนขึ้น จากนั้นสัญญาณเสียงที่ได้จะผ่านตัวกรองແຕบความถี่ผ่านหรือ BPF (bandpass filter) เพื่อแยก



รูปที่ 2.1 บล็อกໄດอະແກມของระบบประสาทหูเทียม

องค์ประกอบความถี่ของสัญญาณออกเป็นแบบด้วยมือ (subband) โดยที่จำนวนแบบด้วยมือที่ย่ออยู่น้อยกว่าจำนวนอิเล็กโทรดที่ใช้ สัญญาณในแต่ละแบบด้วยมือถี่จะถูกแยกข้อมูลส่วนที่เป็นเสียง เล่นไปหรือกรองของสัญญาณ (envelope extraction) ออกมา ด้วยวิธีการเรียงกระแสแบบเต็มคลื่น และการปรับสัญญาณให้เรียบด้วยตัวกรองແเบนความถี่ต่ำผ่าน ขั้นตอนนี้เป็นการหาค่าสัมบูรณ์ (absolute values) ของสัญญาณในแต่ละแบบด้วยมือและนำเผยแพร่พลิจูดของเสียงไว้ในแต่ละแบบด้วยมือที่จะถูกแปลงไปเป็นเผยแพร่พลิจูดของสัญญาณทางไฟฟ้าโดยการบีบขนาดของสัญญาณแบบลอการิตึม (logarithmic compression) เนื่องจากเผยแพร่พลิจูดของสัญญาณเสียงมีขนาดใหญ่กว่า และเผยแพร่พลิจูดของสัญญาณไฟฟ้า จากนั้นแอมพลิจูดของเสียงไว้ในแต่ละแบบด้วยมือจะถูกลดลงกับสัญญาณพัลส์ (pulse) เพื่อส่งสัญญาณไฟฟ้าไปยังยังอิเล็กโทรดอาร์เรย์

โดยทั่วไปการทดสอบเพื่อศึกษาประสิทธิภาพของระบบประสาทหูเทียมจะใช้เสียงพูดจำลอง (vocoded speech) ที่ได้จากการประมาณสัญญาณเสียงผ่านขั้นตอนต่างๆ ของระบบประสาทหูเทียม (CI Simulation) ดังรูปที่ 2.2 เสียงพูดจำลองที่ได้จะถูกนำมาใช้ทดสอบระบบประสาทหูเทียมในเบื้องต้นกับคนที่มีการได้ยินปกติแทน ซึ่งเป็นขั้นตอนการประเมินประสิทธิภาพของระบบที่นิยมใช้กันอย่างกว้างขวาง เพื่อหลีกเลี่ยงปัจจัยบางอย่างที่เกี่ยวข้องกับผู้ใช้ประสาทหูเทียม ที่อาจหลอกให้เกิดความสับสน (confounding factors) และส่งผลต่อการประเมินประสิทธิภาพของระบบได้ (Fei Chen & Loizou, 2011) เช่น ระยะเวลาการใช้ประสาทหูเทียม (duration of implantation) รูปแบบของเส้นประสาทการได้ยินที่คงเหลืออยู่ ความพึงพอใจของผู้ใช้ประสาทหูเทียม เป็นต้น นอกจากนี้การหาอาสาสมัครที่มีการได้ยินปกติสามารถหาได้ยากกว่าคนใช้ประสาทหูเทียมจริง เสียงพูดจำลองที่ใช้ทดสอบจะมีลักษณะของเสียงเหมือนกับเสียงที่ผู้ใช้ประสาทหูเทียมได้ยิน แต่มีความแตกต่างจากเสียงที่คนเราได้ยินหรือคุ้นเคยในชีวิตประจำวัน ทำให้ผู้ถูกทดสอบที่ได้ยินปกติจะรู้สึกเข้าใจความหมายได้ยากในช่วงแรก แต่เมื่อได้ฟังหลายครั้ง ผู้ถูกทดสอบจะเกิดความคุ้นเคยกับเสียงจำลองจนสามารถเข้าใจความหมายได้เจ้ายิ่งขึ้น



รูปที่ 2.2 บล็อกไซด์แกรมการสร้างเสียงพูดจำลองที่มีการเข้ารหัสแบบ ACE

2.2 การเข้ารหัสเสียงพูด (Speech Coding)

ระบบประสาทหูเทียมมีการเข้ารหัสเสียงหลายรูปแบบขึ้นอยู่กับการออกแบบของบริษัทผู้ผลิต ในงานวิจัยนี้ใช้รูปแบบการเข้ารหัสเสียงของบริษัท Nucleus โดยเน้นการเข้ารหัสเสียงแบบ CIS และ ACE ที่ใช้เพียงรูปแบบข้อมูลที่เป็นเนื่องเวลารูปของสัญญาณเชิงเวลา (temporal-envelope information) สำหรับการกระตุ้นสัญญาณทางไฟฟ้าเพื่อส่งไปยังอิเล็กโตรด ซึ่งเป็นข้อมูลที่เพียงพอต่อความเข้าใจเสียงพูด (speech intelligibility) ของมนุษย์ (Fan-Gang et al., 2008)

การเข้ารหัสเสียงแบบ CIS จะทำการส่งสัญญาณพัลส์ทุกเบนด์ความถี่จำนวน 22 เบนด์ไปยังอิเล็กโตรดอาร์เรย์ แต่ละพัลส์ของสัญญาณในแต่ละเบนด์ความถี่ (interleaved pulses) จะถูกส่งไปในลักษณะที่ไม่พร้อมกัน (nonsimultaneous) เพื่อหลีกเลี่ยงความผิดเพี้ยนของสัญญาณไฟฟ้าของอิเล็กโตรดที่อยู่ติดกัน การเข้ารหัสเสียงแบบ ACE มีลักษณะคล้ายกับการเข้ารหัสเสียงแบบ CIS แต่มีความแตกต่างกันที่การเข้ารหัสเสียงแบบ ACE จะไม่ส่งสัญญาณพัลส์ของทุกเบนด์ความถี่ไปยังอิเล็กโตรดอาร์เรย์ แต่จะทำการเลือกส่งเขตของเบนด์ความถี่ที่มีพลังงานสูงสุดหรือพลังงานสูงสุดจำนวน 8-12 เบนด์จากเบนด์ความถี่ทั้งหมด 22 เบนด์ (k -of- m channel)

ในงานวิจัยนี้เลือกใช้การส่งเบนด์ความถี่ที่มีพลังงานสูงสุดจำนวน 12 เบนด์ (12-of-22 channel) ดังรูปที่ 2.2 ทำให้การเข้ารหัสเสียงแบบ ACE สามารถเพิ่มอัตราการกระตุ้นสัญญาณพัลส์สูงขึ้น (stimulation rate) ซึ่งเป็นการเพิ่มรายละเอียดของสัญญาณในเชิงเวลา (time resolution) ทำให้ข้อมูลเสียงที่สำคัญถูกส่งไปยังอิเล็กโตรดได้เพิ่มขึ้น (Nogueira, Buchner, Lenarz, & Edler, 2005) การเลือกเบนด์ความถี่ที่มีพลังงานสูงสุดซึ่งเป็นเบนด์ข้อมูลเสียงที่สำคัญมากที่สุดจำนวน 12 เบนด์และตัดเบนด์ความถี่ที่มีพลังงานน้อยกว่าซึ่งเป็นเบนด์ข้อมูลเสียงที่สำคัญน้อยกว่าที่ถูกเป็นการลดข้อมูลเสียงพูดที่ซ้ำซ้อนอย่างหนึ่ง นอกเหนือจากการลดจำนวนเบนด์ความถี่ที่ส่งไปยังอิเล็กโตรด ทำให้ช่วยลดปัญหาความผิดเพี้ยนของสัญญาณที่อาจจะเกิดจากสนามไฟฟ้าระหว่างอิเล็กโตรดที่อยู่ติดกัน (electrode interaction) และยังสามารถช่วยเพิ่มอายุการใช้งานของแบตเตอรี่ได้

การเข้ารหัสเสียงแบบ CIS และ ACE สามารถทำการปรับค่าพารามิเตอร์ต่างๆ เพื่อให้เหมาะสมกับผู้ใช้ประสาทหูเทียมแต่ละคน (Loizou et al., 2000, Kokkinakis et al., 2012) เช่น ลำดับการกระตุ้นสัญญาณไปยังอิเล็กโตรด (electrode stimulation order) อัตราการกระตุ้นสัญญาณพัลส์ การเลือกจำนวนเบนด์ความถี่ที่มีค่าพลังงานสูงสุดเพื่อส่งไปอิเล็กโตรดอาร์เรย์ (channel selection) เป็นต้น การปรับค่าพารามิเตอร์เหล่านี้จะส่งผลต่อความเข้าใจเสียงพูดของผู้ใช้ประสาทหูเทียมที่สภาพแวดล้อมที่แตกต่างกันด้วย การเข้ารหัสเสียงแบบ CIS และ ACE ของบริษัท Nucleus จะใช้การแปลง FFT ในการออกแบบชุดเบนด์ความถี่ 22 เบนด์

การสร้างเสียงพูดจำลอง (vocoded speech) ที่มีการเข้ารหัสเสียงแบบ ACE แสดงดังรูปที่ 2.2 โดยสัญญาณอินพุตเป็นสัญญาณเสียงพูดที่ไม่มีเสียงรบกวน (clean speech) หรือเสียงพูดที่มีเสียงรบกวน (noisy speech) ที่อัตราการสุ่มตัวอย่าง (sampling rate) 16 กิโลเฮิรตซ์ (kHz) สัญญาณจะถูกแบ่งออกเป็นเฟรมขนาด 128 แมมเปิลและแต่ละเฟรมมีการซ้อนทับกันที่ 75% สัญญาณเสียงทั้งหมดจะถูกประมวลผลทีละเฟรม (frame-by-frame processing) สัญญาณแต่ละเฟรมจะผ่านตัวกรองพร้อมฟัชชิสเพื่อเน้นข้อมูลเสียงพูดที่องค์ประกอบความถี่สูงและลดถอนองค์ประกอบความถี่ต่ำหลังจากนั้นเฟรมขนาด 128 แมมเปิล จะถูกแปลงด้วย FFT จะได้ความถี่ทั้งหมด 128 บิน (bins) และ

จะบินมีความกว้างของแบนด์ความถี่เท่ากับ 125 เฮิรตซ์ ในที่นี้จะใช้ 64 บินแรกเท่านั้นสำหรับการสร้างแบนด์ความถี่จำนวน 22 แบนด์ โดยทำการรวมบินของความถี่ที่อยู่ติดกันให้ได้ความถี่ตัด (cutoff frequencies) และความถี่กลาง (center frequencies) ตามมาตรฐานของบริษัท Nucleus (Cochlear, 2002) คุณลักษณะต่างๆ (characteristics) ของแบนด์ความถี่จำนวน 22 แบนด์ ในทุกของความถี่ขอบล่าง (lower frequencies) ความถี่ขอบบน (upper frequencies) ความถี่กลาง และแบนด์วิดท์ สามารถแสดงรายละเอียดดังตารางที่ 2.1

จากนั้นค่าแอมเพลจูดของทั้ง 22 แบนด์จะถูกถ่วงน้ำหนักด้วยค่าที่เหมาสม จากนั้นทำการเลือกแบนด์ที่มีค่าแอมเพลจูดสูงสุดจำนวน 12 แบนด์จากทั้งหมด 22 แบนด์ เพื่อไปทำการมอคูลেตกับสัญญาณรบกวนแบบ white (white noise) ที่แบนด์ความถี่เดียวกัน สัญญาณที่ถูกมอคูละต์ในแต่ละแบนด์ความถี่จะถูกนำมารวมกันเพื่อสร้างเป็นเสียงพูดจำลอง

ตารางที่ 2.1 คุณลักษณะของแบนด์ความถี่จำนวน 22 แบนด์

อีลีกโกรดที่	แบนด์ความถี่			
	ความถี่ขอบล่าง	ความถี่ขอบบน	ความถี่กลาง	แบนด์วิดท์
22	187.5	312.5	250	125
21	312.5	437.5	375	125
20	437.5	562.5	500	125
19	562.5	687.5	625	125
18	687.5	812.5	750	125
17	812.5	937.5	875	125
16	937.5	1062.5	1000	125
15	1062.5	1187.5	1125	125
14	1187.5	1312.5	1250	125
13	1312.5	1562.5	1437	250
12	1562.5	1812.5	1687	250
11	1812.5	2062.5	1937	250
10	2062.5	2312.5	2187	250
9	2312.5	2687.5	2500	375
8	2687.5	3062.5	2875	375
7	3062.5	3562.5	3312	500
6	3562.5	4062.5	3812	500
5	4062.5	4687.5	4375	625
4	4687.5	5312.5	5000	625
3	5312.5	6062.5	5687	750
2	6062.5	6937.5	6500	875
1	6937.5	7937.5	7437	1000

2.3 เทคนิคการลดสัญญาณรบกวน (Noise Reduction Techniques)

การพัฒนาเทคนิคการลดเสียงรบกวนได้ประสบความสำเร็จเป็นอย่างมาก ในการที่จะปรับปรุงประสิทธิภาพการเข้าใจเสียงพูดในสภาพแวดล้อมที่มีเสียงรบกวน แนวทางการพัฒนาเทคนิคการลดเสียงรบกวนสำหรับงานด้านสื่อสารการพูด (speech communication) แบ่งออกเป็นการลดเสียงรบกวนแบบใช้การบันทึกเสียงด้วยไมโครโฟนตัวเดียว (single microphone) และไมโครโฟนหลายตัว (multiple microphones) (Kokkinakis et al., 2012)

เทคนิคการลดเสียงรบกวนแบบใช้ไมโครโฟนหลายตัว (multi-microphone noise reduction) สามารถลดเสียงรบกวนได้ดีและมีประโยชน์อย่างมาก แต่เทคนิคนี้มีข้อจำกัดต่อการลดเสียงรบกวนแบบเสียงสะท้อน (reverberation) หรือในบางสถานการณ์ที่เสียงพูดและเสียงรบกวนไม่ได้แยกจากกันอย่างชัดเจนและอาจถูกบันทึกที่ไมโครโฟนตัวใดตัวหนึ่งได้ (Dawson, Mauger, & Hersbach, 2011) นอกจากนี้หากเครื่องซ่อมฟังหรือระบบประสาทหูเทียมมีการออกแบบด้วยไมโครโฟนหลายตัวอาจดูไม่สวยงามตามหลักศรีรachaสตร์ (cosmetically unappealing prosthesis) รวมทั้งอัลกอริทึมของการประมวลผลสัญญาณอาจจะค่อนข้างยุ่งยากในการหาค่าที่เหมาะสมสำหรับผู้ใช้ประสาทหูเทียมแต่ละคน ในทางตรงกันข้ามสำหรับเทคนิคการลดเสียงรบกวนแบบใช้ไมโครโฟนตัวเดียว (single-microphone noise reduction) ที่ไม่เกี่ยวข้องกับเงื่อนไขของการรับรู้ทิศทางของเสียง อาจจะมีประโยชน์ในการกำหนดเงื่อนไขของการฟังเสียงที่มีความหลากหลายมากกว่า นอกจากนี้การลดเสียงรบกวนแบบใช้ไมโครโฟนตัวเดียวจะถูกใช้เป็นส่วนมากในระบบประสาทหูเทียมแบบดังเดิม เป็นส่วนใหญ่ นอกจากนี้เทคนิคนี้สามารถนำไปประยุกต์ใช้กับเทคนิคการลดเสียงรบกวนแบบใช้ไมโครโฟนหลายตัวได้ ดังนั้นงานวิจัยนี้จึงเลือกใช้เทคนิคการลดเสียงรบกวนแบบใช้ไมโครโฟนตัวเดียว

ในช่วงแรกเทคนิคการลดเสียงรบกวนแบบใช้ไมโครโฟนตัวเดียวได้ถูกพัฒนาขึ้นมาสำหรับคนที่ได้ยินปกติมาหลายทศวรรษ สามารถจำแนกได้เป็น 4 ประเภท (P.C Loizou, 2013) คือ เทคนิคที่ใช้รูปแบบทางสถิติ (statistical-model based noise reduction) เทคนิคปริญามิย่อย (subspace-based noise reduction) เทคนิคของเวเนียร์ (Wiener-type noise reduction) และเทคนิคของการลบสเปกตรัม (spectral subtractive noise reduction) อย่างไรก็ตามในการพัฒนาเทคนิคเหล่านี้พบว่ามีความก้าวหน้าอย่างมากในเรื่องของคุณภาพของเสียง (speech quality) แต่มีความก้าวหน้าเพียงเล็กน้อยในเรื่องของการเข้าใจเสียงพูด (speech intelligibility) (P. Loizou & Gibak, 2011)

การศึกษาเปรียบเทียบการเข้าใจเสียงพูดของเทคนิคการลดเสียงรบกวนแบบใช้ไมโครโฟนตัวเดียวของ Hu and Loizou (2007) และ Li et al. (2011) โดยทดสอบกับคนที่ได้ยินปกติ พบว่าเทคนิคการลดเสียงรบกวนส่วนมากไม่ได้ทำให้ผู้ฟังเข้าใจเสียงพูดได้ดีขึ้นหรือดีขึ้นเพียงเล็กน้อยเท่านั้น สำหรับผู้ใช้ภาษาอังกฤษ (Hu & Loizou, 2007) หรือภาษาจีนและญี่ปุ่น (Li et al., 2011) มีเพียงเทคนิคของเวเนียร์เท่านั้นที่ทำให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นอย่างมีนัยสำคัญสำหรับผู้ใช้ภาษาอังกฤษกรณีที่มีเสียงรบกวนแบบ Car และผู้ใช้ภาษาญี่ปุ่นที่มีเสียงรบกวนแบบ Car และ White เท่านั้น

อย่างไรก็ตามในการศึกษาเทคนิคการลดเสียงรบกวนแบบใช้ไมโครโฟนตัวเดียวสำหรับคนที่บกพร่องทางการได้ยิน พบว่าผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นอย่างมีนัยสำคัญ เทคนิคการลดเสียง

กระบวนการแบบใช้ไมโครโฟนตัวเดียวที่นำมาประยุกต์ใช้กับระบบประสาทหูเทียม สามารถแบ่งออกเป็น 2 ประเภท (Kokkinakis et al., 2012) ดังนี้ ประเภทแรกคือการลดเสียงรบกวนในขั้นตอนการประมวลผลก่อน (preprocessing approach) เป็นการใช้นำสัญญาณเสียงที่มีเสียงรบกวน (noisy speech) มาลดทอนเสียงรบกวนด้วยเทคนิคต่างๆ ก่อน ทำให้ได้สัญญาณเสียงที่ถูกปรับปรุงแล้ว (enhanced speech) เสียงที่ได้นี้จะนำไปป้อนเข้าระบบประสาทหูเทียมต่อไป การลดเสียงรบกวนแบบนี้จะเหมือนกับการปรับปรุงเสียงพูดในอุปกรณ์การสื่อสารอื่นๆ (communication devices)

ประเภทที่สองคือการลดทอนเสียงรบกวนในขั้นตอนการตรวจจับ.envelope (envelope detection) ของแอมเพลจูดในแต่ละแบบด้วยความถี่ (envelope-based approach) เป็นการลดเสียงรบกวนที่แอมเพลจูดของเอโนเวโลปโดยตรง การลดเสียงรบกวนประเภทนี้สามารถแบ่งออกเป็น 2 รูปแบบคือ การถ่วงน้ำหนักที่แอมเพลจูดของเอโนเวโลป (envelope-weighting approach) โดยการคูณค่าน้ำหนัก (weight value) ที่แอมเพลจูดของเอโนเวโลปในแต่ละแบบด้วยความถี่เพื่อลดทอนเสียงรบกวนในแต่ละแบบด้วยน้ำหนักสามารถหาได้จากการประมาณค่าด้วยวิธีการต่างๆ และการเลือกแบบด้วยเอโนเวโลป (envelope-selection approach) (Nogueira et al., 2005) โดยพิจารณาจากแอมเพลจูดของเอโนเวโลปที่มีค่ามากที่สุดซึ่งเป็นข้อมูลที่สำคัญของเสียงพูดหรือพิจารณาจากลักษณะการได้ยินของมนุษย์ (psychoacoustic model) โดยหูของมนุษย์จะไม่สามารถได้ยินเสียงที่แอมเพลจูดของเสียงที่มีค่าต่ำกว่าค่าระดับของเสียงที่กลบ (masking threshold) ทำให้แบบด้วยความถี่ของข้อมูลส่วนนี้สามารถตัดทิ้งได้ในงานวิจัยนี้เลือกใช้การลดเสียงรบกวนแบบการประมวลผลก่อน

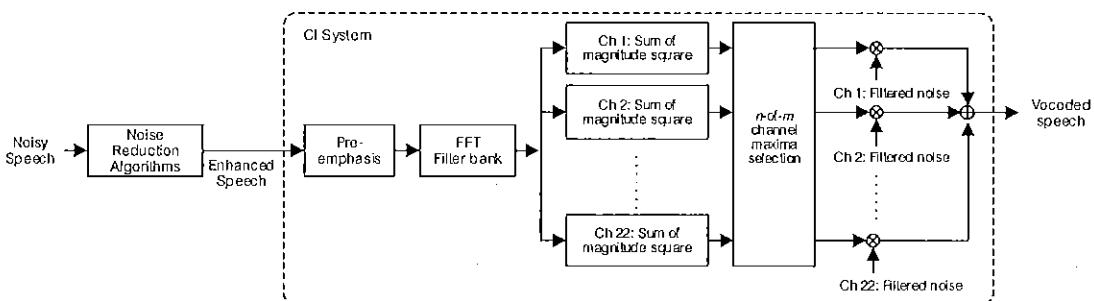
เทคนิคการลดเสียงรบกวนแบบการประมวลผลก่อนที่นำมาประยุกต์ใช้กับระบบประสาทหูเทียม ได้แก่ ปริภูมิย้อย (P. C. Loizou, Lobo, & Hu, 2005) เกนีเยอร์ (Bolner et al., 2016) และการลบสเปกตรัม (Verschuur, Lutman, & Wahat, 2006; Weiss, 1993; Wilson, 2017; Yang & Fu, 2005) โดยมีรายละเอียดดังนี้ จากงานวิจัยของ P. C. Loizou et al. (2005) แสดงให้เห็นว่าการลดเสียงรบกวนด้วยเทคนิคการลบสเปกตรัมของ Hu and Loizou (2002) ในที่มีเสียงรบกวนแบบ SSN ที่ระดับ SNR เท่ากับ 5 เดซิเบล ทำให้ผู้ใช้ประสาทหูเทียมจำนวน 14 คนสามารถเข้าใจประโยค (sentence recognition) ได้ดีขึ้นอย่างน้อยสำคัญโดยเฉลี่ย 44% งานวิจัยของ Bolner et al. (2016) พบร่วมกับการลดเสียงรบกวนด้วยเทคนิคแบบเวนีย์ร์หรือ WF ของ Scalart and Vieira (1996) ทำให้คนที่ได้ยินปกติจำนวน 10 คนสามารถเข้าใจประโยคได้ดีขึ้นในสภาพแวดล้อมที่มีเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 0 เดซิเบล แต่ไม่ได้ทำให้ผู้ฟังเข้าใจประโยคได้ดีขึ้นในที่มีเสียงรบกวนแบบ BBN นอกจากนี้จากการวิจัยของ Koning et al. (2015) แสดงให้เห็นว่าการลดทอนเสียงรบกวนด้วยเทคนิค WF ในแอมเพลจูดของเอโนเวโลปของแต่ละแบบด้วยความถี่ ทำให้ผู้ฟังเข้าใจเสียงพูดและรู้สึกว่าคุณภาพของเสียงดีขึ้นจากการทดสอบกับผู้ใช้ประสาทหูเทียมจำนวน 6 คนและคนที่ได้ยินปกติจำนวน 6 คนซึ่งเป็นคนดั้งเดิม

เทคนิคการลบสเปกตรัมได้ถูกพัฒนาหลายเวอร์ชัน (version) มานานมากกว่า 40 ปีแล้ว เทคนิคการลบสเปกตรัมบางเวอร์ชันได้ถูกนำมาประยุกต์ใช้กับระบบประสาทหูเทียมดังนี้ Wilson (2017) ได้นำเทคนิคการลบสเปกตรัมที่เรียกว่า INTEL มาประยุกต์ใช้กับระบบประสาทหูเทียมครั้งแรก เพื่อลดเสียงรบกวนแบบ SSN ที่ SNR ตั้งแต่ -10 ถึง 25 เดซิเบล สำหรับการทดสอบกับคนที่ได้ยินปกติจำนวน 10 คน และที่ SNR ตั้งแต่ -5 ถึง 25 เดซิเบล สำหรับการทดสอบกับผู้ใช้ประสาทหู

เที่ยมจำนวน 10 คน จากผลทดลองแสดงให้เห็นว่าผู้ใช้ประสาทหูเทียมสามารถเข้าใจคำพูดได้ดีขึ้น แต่ไม่มีผลต่อคนที่ได้ยินปกติ ต่อมาก็ Weiss (1993) ได้แสดงให้เห็นว่าเทคนิคการลับสเปกตรัมแบบ INTEL ทำให้เกิดความผิดพลาดในการแยกลักษณะฟอร์เม้นท์ที่สอง (formant) ในการเข้ารหัสเสียงของระบบประสาทหูเทียม การศึกษาของ Yang and Fu (2005) พบว่าเทคนิคการลับสเปกตรัมของ Gustafsson, Nordholm, and Claesson (2001) ที่ใช้ในการลดเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 0, 3, 6, 9 เดซิเบล ทำให้ผู้ใช้ประสาทหูเทียมจำนวน 7 คน สามารถรู้จำประโยค (sentence recognition) ได้ดีขึ้น

Verschuur et al. (2006) ได้ทำการทดลองเข้มเดียวกับงานวิจัยของ Yang and Fu (2005) แต่ใช้เทคนิคการลับสเปกตรัมของ Lockwood and Boudy (1992) และเพิ่มจำนวนผู้ใช้ประสาทหูเทียมเป็น 17 คน จากการทดลองพบว่าเทคนิคการลับสเปกตรัมนี้ ช่วยให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นในที่มีเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 5 และ 10 เดซิเบล ต่อมาก็ Kallel et al. (2012) ได้ใช้เทคนิค NSS (nonlinear spectral subtraction) (Berouti, Schwartz, & Makhoul, 1979) และเทคนิค MBSS (Kamath & Loizou, 2002) เพื่อลดเสียงรบกวนที่ SNR เท่ากับ -3, 0, 3, 6 เดซิเบล จากการทดลองพบว่าผู้ใช้ประสาทหูเทียมจำนวน 3 คน สามารถเข้าใจเสียงพูดได้ดีขึ้นประมาณ 4-9% และคนที่ได้ยินปกติสามารถเข้าใจเสียงพูดได้ดีขึ้นประมาณ 7-13% นอกจากนี้ยังพบว่าเทคนิค MBSS ทำให้ผู้ฟังเข้าใจเสียงพูดได้ดีกว่าเทคนิค NSS เมื่อเสียงรบกวนมาจากการทั้งแหล่งกำเนิดที่เดียวกันหรือจากหลายแหล่ง

จากเหตุผลได้กล่าวข้างต้น ในงานวิจัยนี้จึงเลือกใช้การลดเสียงรบกวนด้วยเทคนิค MBSS (Kamath & Loizou, 2002) และ WF (Scalart & Vieira, 1996) เนื่องจากเทคนิคทั้งสองมีการปรับเพื่อลดเสียงรบกวน (noise reduction) และความผิดเพี้ยนของเสียงพูด (speech distortion) ได้อย่างเหมาะสม รวมทั้งใช้การคำนวณน้อย (low computation) และเหมาะสมสำหรับการใช้งานที่เวลาจริง (real-time implementation) (Scalart & Vieira, 1996; Verschuur et al., 2006) นอกจากนี้เทคนิค WF ช่วยให้ผู้ฟังเข้าใจเสียงพูดที่ได้มาก สำหรับผู้ใช้ภาษาอังกฤษ ภาษาจีน และภาษาญี่ปุ่น (Li et al., 2011) ในส่วนของเทคนิค MBSS เป็นการลับสเปกตรัมเวอร์ชันหนึ่งที่ทำให้ผู้ใช้ประสาทหูเทียมขาดร่องเศษ (Kallel et al., 2012) และจีน (F. Chen et al., 2015) สามารถเข้าใจเสียงพูดได้ดีขึ้น



รูปที่ 2.3 บล็อกโดยรวมของการสร้างเสียงจำลองของระบบประสาทหูเทียมร่วมกับอัลกอริทึมของการลดเสียงรบกวน

ในงานวิจัยนี้ทำการศึกษาการลดเสียงรบกวนหรือการปรับปรุงเสียงพูด (speech enhancement) ของระบบประสาทหูเทียมเป็นการประมวลผลก่อน (preprocessing) ดังรูปที่ 2.3 โดยนำสัญญาณเสียงพูดที่มีเสียงรบกวน (noisy speech) ทำการลดเสียงรบกวนโดยใช้เทคนิค MBSS และ WF จะได้สัญญาณเสียงที่ถูกปรับปรุงแล้ว (enhanced speech) ซึ่งจะนำไปป้อนเข้าระบบประสาทหูเทียมและสร้างเป็นเสียงพูดจำลอง ในที่นี้จะกล่าวรายละเอียดของแต่ละเทคนิค โดยสรุปได้ดังนี้

สมมติให้ $y(n)$ เป็นสัญญาณเสียงพูดที่มีเสียงรบกวนด้วยอัตราการสุ่มตัวอย่าง 16 กิโลเอิร์ตซ์ ประกอบด้วยสัญญาณเสียงพูด $x(n)$ และเสียงรบกวน $d(n)$ ดังสมการ

$$y(n) = x(n) + d(n) \quad (2.1)$$

เมื่อทำการแปลงสัญญาณด้วย FFT จะสามารถหาสเปกตรัมกำลัง (power spectrum) ของสัญญาณเสียงพูดที่มีเสียงรบกวน $|Y(k)|^2$ โดยการประมาณค่าจากสเปกตรัมกำลังของสัญญาณเสียง $|X(k)|^2$ และเสียงรบกวน $|D(k)|^2$ ดังสมการ

$$|Y(k)|^2 \approx |X(k)|^2 + |D(k)|^2 \quad (2.2)$$

2.3.1 การลบสเปกตรัมแบบหลายแบนด์ (Multi-band spectral subtraction)

การลบสเปกตรัมแบบหลายแบนด์หรือ MBSS (Kamath & Loizou, 2002) จะเป็นเทคนิคที่แตกต่างจากการลบสเปกตรัมแบบไม่เป็นเชิงเส้นหรือ NSS (nonlinear spectral subtraction) เสิ่นอย่าง เทคนิค MBSS จะใช้ค่าแฟกเตอร์ของการลบสเปกตรัมที่ได้จากการประมาณค่าในแต่ละแบนด์ความถี่และแต่ละbin หลักการของเทคนิค MBSS จะพิจารณาว่าคุณลักษณะของสเปกตรัมของเสียงรบกวน $D(k)$ อาจจะไม่ได้ส่งผลต่อสเปกตรัมของสัญญาณเสียงพูด $X(k)$ ที่ทุกแบนด์ความถี่ แต่อาจจะส่งผลต่อแบนด์ความถี่มากน้อยแตกต่างกัน ดังนั้นการประมาณค่าสเปกตรัมของเสียงรบกวน $D(k)$ ของแต่ละแบนด์ความถี่จะแตกต่างกัน สเปกตรัมกำลังของสัญญาณเสียง $X(k)$ และค่าแฟกเตอร์น้ำหนัก δ ของแบนด์ความถี่ที่ i^{th} ประมาณได้จาก

$$|\hat{X}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \delta_i |\hat{D}_i(k)|^2, b_i \leq k \leq e_i \quad (2.3)$$

เมื่อ $|\hat{D}_i(k)|^2$ คือค่าประมาณของสเปกตรัมกำลังของเสียงรบกวน, b_i และ e_i คือบินเริ่มต้นและบินสุดท้ายของแบนด์ความถี่ที่ i^{th} ตามลำดับ, α_i และ δ_i คือค่าแฟกเตอร์การลบเกิน (over-subtraction factor) และ แฟกเตอร์น้ำหนัก (weight factor) ของแบนด์ความถี่ที่ i^{th} ตามลำดับ ค่าแฟกเตอร์การลบเกิน α_i ของแบนด์ความถี่ที่ i^{th} เป็นพิ่งขึ้นที่ขึ้นกับค่า SNR ของแบนด์ความถี่ที่ i^{th} หาได้จาก

$$\alpha_i = \begin{cases} 5, & SNR_i < -5 \\ 4 - \frac{3}{20}SNR_i, & -5 \leq SNR_i \leq 20 \\ 1, & SNR_i > 20 \end{cases} \quad (2.4)$$

$$SNR_i(\text{dB}) = 10 \log_{10} \left(\frac{\sum_{k=b_i}^{c_i} |Y_i(k)|^2}{\sum_{k=b_i}^{c_i} |\hat{D}_i(k)|^2} \right) \quad (2.5)$$

ค่าแฟกเตอร์น้ำหนัก δ_i ของแบบต์ความถี่ที่ i^{th} ขั้นอยู่กับความถี่ขอบบน (upper frequency) หรือ f_i ของแบบต์ความถี่ที่ i^{th} และความถี่ของการสุ่มตัวอย่าง (sampling frequency) หรือ F_s หากได้จาก

$$\delta_i = \begin{cases} 1, & f_i \leq 1\text{kHz} \\ 2.5, & 1\text{kHz} < f_i \leq \frac{F_s}{2} - 2\text{kHz} \\ 1.5, & f_i > \frac{F_s}{2} - 2\text{kHz} \end{cases} \quad (2.6)$$

2.3.2 ตัวกรองเวนเนียร์ (Wiener filter)

ตัวกรองเวนเนียร์หรือ WF (Scalart & Vieira, 1996) จะใช้พิงก์ชันเกน $g(k)$ ที่อยู่ในเทอมของค่า priori SNR ξ_k ซึ่งค่าของ ξ_k จะประมาณได้จากการประมาณของเฟรมก่อนที่ $m-1$ และเฟรมปัจจุบันที่ m ในส่วนที่เป็นค่าน้ำหนักร่วม $\hat{\xi}_k(m)$ ที่ประมาณได้จากการประมาณปัจจุบันที่ m ดังนี้

$$g(k) = \frac{\xi_k}{\xi_k + 1} \quad (2.7)$$

$$\hat{\xi}_k(m) = \alpha \frac{|\hat{X}_k(m-1)|^2}{|D_k(m-1)|^2} + (1-\alpha) \max \left(\frac{|Y_k(m)|^2}{|D_k(m)|^2} - 1, 0 \right) \quad (2.8)$$

เมื่อ α คือค่าคงที่ในที่นี่ใช้ $\alpha=0.98$, $\hat{X}_k(m-1)$ คือค่าประมาณสเปกตรัมของสัญญาณเสียง $X(k)$ ของเฟรมก่อนที่ $m-1$, $Y_k(m)$ คือสเปกตรัมของสัญญาณเสียงที่มีเสียงรบกวน $Y(k)$ ของเฟรมปัจจุบันที่ m , $D_k(m)$ คือสเปกตรัมของเสียงรบกวน $D(k)$ ของเฟรมปัจจุบันที่ m

2.4 สรุป

การประเมินประสิทธิภาพการเข้าใจเสียงพูดของผู้ใช้ระบบ普通话ที่ยังไม่ใช้การเสียงจำลอง (vocoded speech) ที่ได้จากการประมวลผลสัญญาณของระบบ มาใช้ในการทดสอบประสิทธิภาพของระบบเบื้องต้น ในงานวิจัยนี้เลือกใช้การเข้ารหัสเสียงแบบ CIS และ ACE ร่วมกับการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF โดยการลดเสียงรบกวนจะเป็นขั้นตอนการประมวลผล ก่อน (preprocessing) จะได้สัญญาณเสียงที่ปรับปรุงแล้ว (enhanced speech) สัญญาณที่ปรับปรุงแล้วจะเป็นสัญญาณอินพุตของระบบ普通话ที่ยังไม่สามารถเข้ารหัสเสียงแบบ ACE และสร้างเป็นเสียงจำลองต่อไป



บทที่ 3

การประเมินประสิทธิภาพการเข้าใจเสียงพูด

การประเมินประสิทธิภาพการเข้าใจเสียงพูดของผู้ใช้ประสาทหูเทียมโดยตรงนั้น ขึ้นอยู่กับปัจจัยที่สำคัญ 2 ปัจจัยคือ ปัจจัยที่เกี่ยวข้องกับผู้ใช้ประสาทหูเทียม (CI user-related factors) เช่น อายุ เพศ ระดับความเรียน ความลึกของช่องเส้นเลือดโกร径 และว่าใจฟังที่เกี่ยวข้องกับระบบประสาทหูเทียม (CI processor-related factors) ที่ใช้ในปัจจุบัน เช่น เทคนิค การเข้ารหัสเสียงพูดของระบบประสาทหูเทียมที่ถูกพัฒนาขึ้นของแต่ละบริษัทมีความแตกต่างกัน เป็นต้น ปัจจัยเหล่านี้สามารถส่งผลต่อการประเมินประสิทธิภาพการเข้าใจเสียงพูดได้พื้นฐาน ทำให้การแปลความหมายของผลที่ได้มีความคลาดเคลื่อนและขาดความน่าเชื่อถือได้ ดังนั้นการทดสอบกับคนที่ได้ยินปกติสามารถช่วยลดปัจจัยต่างๆ เหล่านี้ได้และสามารถนำไปใช้ในการทวนสอบในมูลลักษณะที่คาดว่าจะได้รับจากผู้ใช้ประสาทหูเทียมจริงได้ นอกจากนี้อาสาสมัครที่มีการได้ยินปกติมีจำนวนที่มากเพียงพอและหาได้ง่ายกว่าผู้ที่ใช้ประสาทหูเทียม

ในบทนี้จะกล่าวถึงรายละเอียดการประเมินประสิทธิภาพ (performance evaluation) ของการเข้าใจเสียงพูด (speech intelligibility) ของผู้ใช้ประสาทหูเทียม โดยทำการเปรียบเทียบประสิทธิภาพการเข้ารหัสเสียงพูด (CIS และ ACE) และเทคนิคการลดเสียงรบกวน (MBSS และ WF) โดยใช้คำทดสอบประกอบด้วยคำไทยพยางค์เดียวและพยางค์คู่ สำหรับการประเมินภายใต้สภาพแวดล้อมที่ปราศจากเสียงรบกวนและที่มีเสียงรบกวน (SSN และ BBN) ที่ระดับ SNR ต่างกัน ขั้นตอนการประเมินทั้งหมดได้ผ่านการรับรองโครงการวิจัยตามแนวทางหลักจริยธรรมการวิจัยในมนุษย์ที่เป็นมาตรฐานสากลจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ของมหาวิทยาลัย การประเมินประสิทธิภาพประกอบด้วยรายละเอียดที่เกี่ยวข้องกับอาสาสมัคร คำที่ใช้ทดสอบ และขั้นตอนการประเมิน ดังนี้

3.1 การเปรียบเทียบประสิทธิภาพของการเข้ารหัสเสียงพูด

3.1.1 อาสาสมัคร (Subjects)

อาสาสมัครเป็นนิสิตหรือบุคลากรในมหาวิทยาลัยจำนวน 12 คน โดยแบ่งเป็นเพศหญิงจำนวน 2 คนและเพศชายจำนวน 10 คน ที่มีอายุระหว่าง 22-34 ปีหรือมีอายุโดยเฉลี่ย 24 ปี อาสาสมัครทั้งหมดจะต้องเป็นผู้ที่พูดภาษาไทยมาตั้งแต่กำเนิดและมีระดับการได้ยินปกติไม่เกิน 25 เดซิเบล โดยเฉพาะที่ช่วงความถี่ตั้งแต่ 250 ถึง 8000 เฮิรตซ์ นอกจากนี้อาสาสมัครจะต้องไม่เคยผ่านการทดสอบเสียงจำลองการทำงานระบบประสาทหูเทียมมาก่อน

3.1.2 คำที่ใช้ทดสอบ (Materials)

คำไทยที่ใช้ทดสอบทั้งหมดเป็นคำที่ใช้ทดสอบในทางคลินิกสำหรับผู้ที่บกพร่องทางการได้ยิน (Kangsanarak, 1980b; Komalarajun, 1979) คำทั้งหมดประกอบด้วยชุดของคำพยางค์เดียว

(monosyllabic words) และคำพยางค์คู่ (bisyllabic words) อย่างละ 8 ชุด ชุดละ 25 คำ รวมทั้งหมด 400 คำ โดยถูกบันทึกเสียงโดยชาญไทย ที่อัตราการสูมตัวอย่างที่ 22 กิโลเฮิรตซ์และเก็บข้อมูล 1 คำต่อ 1 ไฟล์เสียง จากนั้นชุดคำทั้งหมดจะถูกผสมกับเสียงรบกวนแบบ SSN (speech-shaped noise) ที่ SNR เท่ากับ 0, 5, 10 เดซิเบล และเสียงรบกวนแบบ BBN (babble noise) ที่ SNR เท่ากับ 5, 10, 15 เดซิเบล การเลือกระดับ SNR ที่ค่าเหล่านี้เพื่อหลีกเลี่ยงกรณีที่ผู้ฟังอาจจะไม่สามารถเข้าเสียงพูดเลย (0 คะแนน) หรือสามารถเข้าใจเสียงพูดได้ถูกต้องทุกคำ (100 คะแนน) (floor and ceiling effects) (F. Chen et al., 2015) โดยเฉพาะอย่างยิ่งคำพยางค์เดียวที่ระดับ SNR มีค่าน้อย หรือคำที่มีเสียงรบกวนมาก ผู้ฟังอาจจะฟังคำและเข้าใจได้ถูกต้องน้อยมากเกินไปจนถึงไม่ถูกต้องเลย หรือคำพยางค์คู่ที่ระดับ SNR มีค่ามากหรือคำที่มีเสียงรบกวนน้อย ผู้ฟังอาจจะฟังคำและเข้าใจได้ถูกต้องทั้งหมด ทำให้ไม่สามารถเปรียบเทียบความแตกต่างของแต่ละเงื่อนไขที่ต้องการศึกษาได้ ชุดคำที่ไม่มีเสียงรบกวน (clean words) และชุดคำที่มีเสียงรบกวน (noisy words) จะถูกเข้ารหัสเสียงแบบ CIS และ ACE เพื่อสร้างเสียงพูดจำลองของระบบประสาทหูเทียม เสียงจำลองที่ได้จะถูกกำหนดด้วยเงื่อนไขทั้งหมด 28 เงื่อนไข ประกอบด้วยคำที่ไม่มีเสียงรบกวนจำนวน 4 เงื่อนไข (การเข้ารหัสเสียง 2 แบบ × ประเภทของคำ 2 แบบ) และคำที่มีเสียงรบกวนจำนวน 24 เงื่อนไข (การเข้ารหัสเสียง 2 แบบ × ประเภทของคำ 2 แบบ × เสียงรบกวน 2 แบบ × ระดับของ SNR 3 รูปแบบ) ดังตารางที่ 3.1

ตารางที่ 3.1 การกำหนดเงื่อนไขที่ใช้ทดสอบสำหรับการเข้ารหัสเสียง

คำพยางค์เดียว

การเข้ารหัสเสียงพูด (speech coding)	คำที่ไม่มีเสียงรบกวน (clean words)	คำที่มีเสียงรบกวน (noisy words)					
		SSN			BBN		
		0 dB	5 dB	10 dB	5 dB	10 dB	15 dB
CIS	1	2	3	4	5	6	7
ACE	8	9	10	11	12	13	14

คำพยางค์คู่

การเข้ารหัสเสียงพูด (speech coding)	คำที่ไม่มีเสียงรบกวน (clean words)	คำที่มีเสียงรบกวน (noisy words)					
		SSN			BBN		
		0 dB	5 dB	10 dB	5 dB	10 dB	15 dB
CIS	15	16	17	18	19	20	21
ACE	22	23	24	25	26	27	28

3.1.3 ขั้นตอนการประเมิน (Procedure)

การทดสอบจะดำเนินการในห้องที่เงียบปราศจากเสียงรบกวน อาสาสมัครจะฟังเสียงจำลองผ่านหูฟัง (headphone) ที่ต่อจากเครื่องขยายเสียงและเครื่องคอมพิวเตอร์ โดยอาสาสมัครจะต้องเลือกฟังด้วยหูข้างซ้ายหรือหูข้างขวาเพียงข้างใดข้างหนึ่งเท่านั้น ที่อาสาสมัครรู้สึกพึงพอใจและสะดวก

ในการพัฒนาที่สุด อาสาสมัครจะต้องฟังเสียงจากหูข้างนั้นตลอดการทดลองจนเสร็จสิ้น การปรับระดับความดังของเสียงจะเลือกความดังระดับที่ใช้สนทนากันในชีวิตประจำวัน

การทดลองจะแบ่งออกเป็น 2 ครั้ง แต่ละครั้งจะทำการทดสอบ 14 เสื่อนไขและแต่ละครั้ง จะต้องทำการทดลองห่างกันอย่างน้อย 7 วัน เพื่อหลีกเลี่ยงผลกระทบที่เกิดจากการเรียนรู้ของอาสาสมัคร (learning effect) ที่อาจจะส่งผลต่อปอร์เซ็นต์ของคะแนนคำที่ถูกต้องได้ ในการทดลอง แต่ละครั้ง อาสาสมัครจะต้องทดลองฟังเสียงจำลองของทุกเสื่อนไขที่กำหนดก่อน (training test) ประมาณ 5 นาที เพื่อให้เกิดความคุ้นเคยกับเสียงจำลองของระบบประสาททุกเที่ยม เนื่องจากเสียงจำลองจะไม่เหมือนกับเสียงพูดที่อาสาสมัครได้ยินตามปกติในชีวิตประจำวัน สำหรับคำที่นำมาใช้ทดลองฟังนั้นจะไม่เป็นคำที่ซ้ำกับคำที่ใช้ในการทดสอบจริง (actual test)

สำหรับการทดสอบจริง อาสาสมัครจะได้ฟังเสียงจำลองของแต่ละคำเพียงครั้งเดียวเท่านั้น และทำการเขียนสิ่งที่ได้ยินลงในกระดาษที่ได้จัดเตรียมไว้ให้ (ภาคผนวก ค) หากอาสาสมัครได้ยินไม่ชัดเจนหรือไม่แน่ใจ สามารถเดาคำนั้นได้ตามความเหมาะสม ใน การทดสอบอาสาสมัครจะไม่สามารถรู้ได้ก่อนว่าต้องฟังคำอะไรและคำนั้นอยู่ในเสื่อนไขใด อาสาสมัครแต่ละคนจะได้ฟังเสียงทดสอบแต่ละชุด คำของแต่ละเสื่อนไขและลำดับของแต่ละเสื่อนไขแตกต่างกัน โดยชุดคำและเสื่อนไขจะถูกเลือกในลักษณะแบบสุ่มเลือก นอก จากนี้ การทดลองของแต่ละเสื่อนไขในการทดสอบแต่ละครั้งจะต้องไม่ซ้ำกัน เพื่อหลีกเลี่ยงผลกระทบจากการเรียนรู้คำของอาสาสมัคร ในการทดสอบจะใช้วาประมาณ 1 ชั่วโมง ในระหว่างการทดสอบจะมีการพักทุก 20 นาทีหรือขึ้นอยู่กับความต้องการของอาสาสมัครเพื่อหลีกเลี่ยงจากความล้าในการฟังที่นาน ที่อาจส่งผลต่อประสิทธิภาพการเข้าใจเสียงพูดได้

3.2 การเปรียบเทียบประสิทธิภาพของเทคนิคการลดเสัญญาณรบกวน

การเปรียบเทียบประสิทธิภาพการเข้าใจเสียงพูดด้วยเทคนิคการลดเสียงรบกวน จะมีขั้นตอนรายละเอียดคล้ายกับการเปรียบเทียบเพื่อประสิทธิภาพการเข้าใจเสียงพูดของการเข้ารหัสเสียง แต่จะมีรายละเอียดแตกต่างกันในบางส่วนดังนี้

3.2.1 อาสาสมัคร (Subjects)

อาสาสมัครเป็นนิสิตหรือบุคลากรในมหาวิทยาลัยจำนวน 20 คน โดยแบ่งเป็นเพศหญิงจำนวน 6 คนและเพศชายจำนวน 14 คน ที่มีอายุระหว่าง 20-40 ปีหรือมีอายุโดยเฉลี่ย 26 ปี อาสาสมัครทั้งหมดจะต้องเป็นผู้พูดภาษาไทยมาตั้งแต่กำเนิดและมีระดับการได้ยินปกติไม่เกิน 25 เดซิเบล นอกจากนี้อาสาสมัครจะต้องไม่เคยผ่านการทดสอบเสียงจำลองการทำงานระบบประสาททุกเที่ยมมาก่อน

3.2.2 คำที่ใช้ทดสอบ (Materials)

คำไทยที่ใช้ทดสอบทั้งหมดจะเหมือนกับหัวข้อที่ 3.1.2 จากนั้นชุดคำทั้งหมดจะถูกผสมกับเสียงรบกวนแบบ SSN (speech-shaped noise) ที่ SNR เท่ากับ 0 และ 5 เดซิเบล เสียงรบกวนแบบ BBN (babble noise) ที่ SNR เท่ากับ 5 และ 10 เดซิเบล คำที่มีเสียงรบกวนจะถูกลดเสียงรบกวนด้วยเทคนิค MBSS และ WF จะได้สัญญาณที่ถูกปรับปรุงแล้ว (enhanced speech) จากนั้นสัญญาณเสียง

ที่ถูกปรับปรุงแล้วจะนำมาเข้ารหัสเสียงแบบ ACE เพื่อสร้างเสียงจำลองของระบบประสิทธิภาพเพื่อการลดเสียงรบกวนที่ได้จะถูกกำหนดด้วยเงื่อนไขทั้งหมด 24 เงื่อนไข ([การลดเสียงรบกวน 2 แบบ + ไม่มีการลดเสียงรบกวน] × ประเภทของคำ 2 แบบ เสียงรบกวน 2 แบบ × ระดับของ SNR 2 รูปแบบ) ดังตารางที่ 3.2

ตารางที่ 3.2 การกำหนดเงื่อนไขที่ใช้ทดสอบสำหรับเทคนิคการลดเสียงรบกวน

คำพยานค์เดียว

เทคนิคการลดเสียงรบกวน (noise reduction)	คำที่มีเสียงรบกวน (noisy words)			
	SSN		BBN	
	0 dB	5 dB	5 dB	10 dB
ไม่มีการลดเสียงรบกวน	1	2	3	4
MBSS	5	6	7	8
WF	9	10	11	12

คำพยานคู่คู่

เทคนิคการลดเสียงรบกวน (noise reduction)	คำที่มีเสียงรบกวน-(noisy words)			
	SSN		BBN	
	0 dB	5 dB	5 dB	10 dB
ไม่มีการลดเสียงรบกวน	13	14	15	16
MBSS	17	18	19	20
WF	21	22	23	24

3.2.3 ขั้นตอนการประเมิน (Procedure)

ขั้นตอนการประเมินจะเหมือนกับหัวข้อที่ 3.1.3 แต่แตกต่างกันที่การทดลองนี้จะแบ่งออกเป็น 2 ครั้ง แต่ละครั้งทำการทดสอบ 12 เงื่อนไข

บทที่ 4

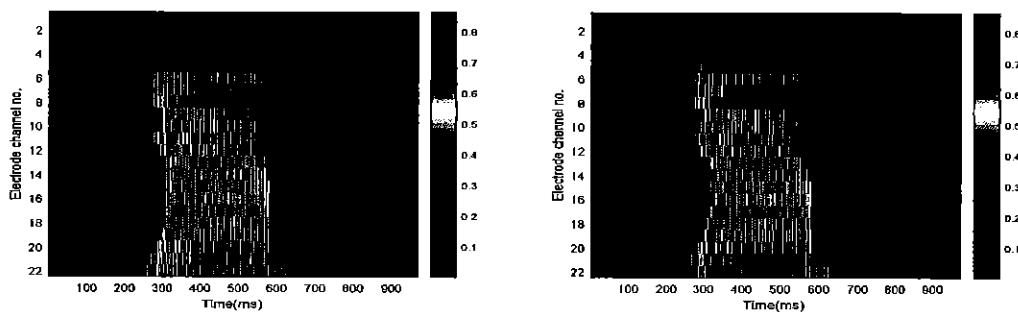
ผลการทดลองและการวิเคราะห์ข้อมูลที่ได้

ในบทนี้จะกล่าวถึงผลการทดลองที่ได้จากการเปรียบเทียบประสิทธิภาพของการเข้ารหัสเสียงแบบ CIS และ ACE การเข้ารหัสเสียงแบบ ACE ร่วมกับเทคนิคการลดเสียงรบกวนด้วย MBSS และ WF รวมทั้งการวิเคราะห์ข้อมูลจากผลการทดลองที่ได้โดยเปรียบเทียบประสิทธิภาพการเข้ารหัสเสียงพูดของแต่ละเทคนิค

4.1 การเปรียบเทียบประสิทธิภาพของการเข้ารหัสเสียงพูดแบบ CIS และ ACE

การเข้ารหัสเสียงแบบ CIS และ ACE สามารถแสดงตัวอย่างสเปกตรограм (spectrogram) ได้ดังรูปที่ 4.1 ซึ่งเป็นสเปกตรограмของคำพยางค์เดียว “ยาม” ที่เป็นสัญญาณเสียงของคำที่ปราศจากเสียงรบกวน (clean word) โดยอิเล็กโทรแคร์ล่าดับที่ 1 และ 22 แสดงแนวความถี่ขององค์ประกอบความถี่สูงสุดและองค์ประกอบความถี่ต่ำสุดตามลำดับ จากรูปจะเห็นได้ว่าการเข้ารหัสเสียงแบบ ACE จะมีรายละเอียดของสัญญาณน้อยกว่าแบบ CIS เนื่องจากการเข้ารหัสเสียงแบบ ACE จะเลือกส่งสัญญาณที่มีพลังงานสูงสุดเพียง 12 แบนด์เท่านั้น จาก 22 แบนด์ไปยังอิเล็กโทรด (12-of-22 channel) ทำให้ข้อมูลที่ไม่สำคัญบางส่วนถูกตัดทิ้งไป เช่น แบนด์ความถี่ที่ 21 และ 22 เป็นต้น

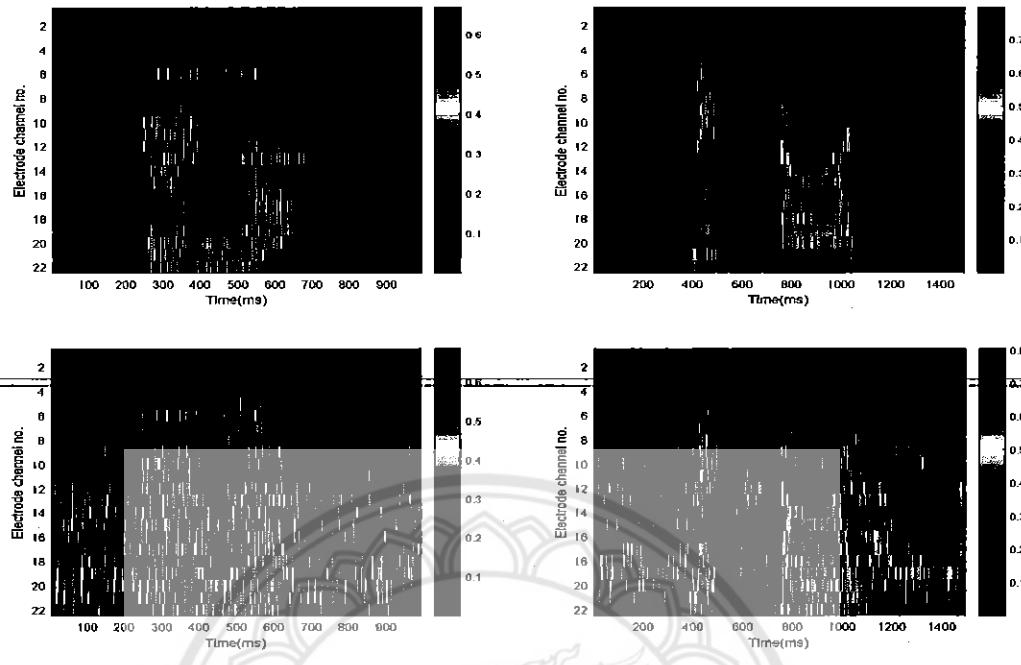
จากรูปที่ 4.2 แสดงตัวอย่างสเปกตรограмที่มีการเข้ารหัสแบบ ACE ของคำพยางค์เดียว “เรือ” และคำพยางค์คู่ “เลิกน้อย” รูปแควร์แสดงสเปกตรกรรมของคำที่ไม่มีเสียงรบกวน รูปสถาที่สองแสดงสเปกตรกรรมของคำพยางค์เดียวที่มีเสียงรบกวนแบบ SSN และคำพยางค์คู่ที่มีเสียงรบกวนแบบ BBN ที่ 5 เดซิเบล จะเห็นว่าพลังงานของเสียงรบกวนแบบ SSN จะกระจายในช่วงความถี่ที่กว้างกว่าแบบ BBN และพลังงานของเสียงรบกวนแบบ BBN จะมีค่ามากในช่วงแบนด์ความถี่ต่ำโดยเฉพาะอย่างยิ่งแบนด์ที่ 18 ถึง 20 จากสเปกตรกรรมของคำที่มีเสียงรบกวนจะเห็นว่ามีการส่งสัญญาณไปยังอิเล็กโทรดตลอดช่วงเวลา 1 วินาที เมื่อเปรียบเทียบกับคำที่ไม่มีเสียงรบกวน ทำให้สื้นเปลืองการใช้แบตเตอรี่ด้วย



(g) CIS

(h) ACE

รูปที่ 4.1 สเปกตรกรรมของคำพยางค์เดียว “ยام” ที่มีการเข้ารหัสเสียงแบบ CIS และ ACE



(ก) คำพยานค์เดียว

(ข) คำพยานค์คู่

รูปที่ 4.2 สเปกตรограмของคำพยานค์เดียว “เรือ” และคำพยานค์คู่ “เล็กน้อย” ที่มีการเข้ารหัสเสียงแบบ ACE และการแสดงสเปกตรกรรมของคำที่ปราศจากเสียงรบกวน แกวที่สองแสดงสเปกตรกรรมของคำพยานค์เดียวที่มีเสียงรบกวนแบบ SSN และคำพยานค์คู่ที่มีเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5 เดซิเบล

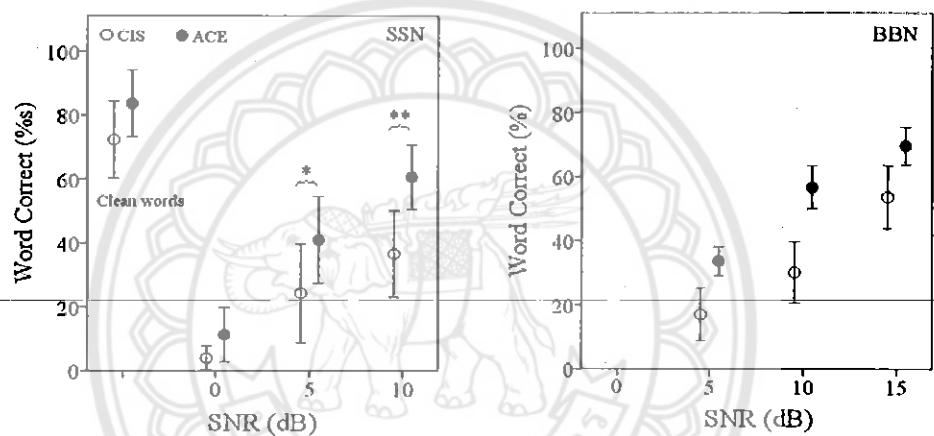
การวัดเบอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้อง สำหรับเงื่อนไขที่กำหนดเพื่อใช้ทดสอบจำนวน 28 เงื่อนไข ประกอบด้วยชุดคำพยานค์เดียวและพยานค์คู่ที่ไม่มีเสียงรบกวนและที่ผสมเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 0, 5 และ 10 เดซิเบล และผสมเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5, 10 และ 15 เดซิเบล ชุดคำทั้งหมดจะถูกเข้ารหัสเสียงแบบ CIS และ ACE ตามตารางที่ 3.1 เงื่อนไขทั้งหมดจะถูกนำมาทดสอบกับอาสาสมัครจำนวน 12 คน คะแนนคำที่ถูกต้องในแต่ละเงื่อนไขของอาสาสมัครทุกคนจะถูกนำมาเฉลี่ยเพื่อคิดเป็นเบอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้อง

ผลการทดลองได้แสดงดังรูปที่ 4.3 จากผลการทดลองแสดงให้เห็นว่าเบอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้องจะแปรผันตามกับระดับ SNR ของเสียงรบกวน นั่นคือถ้าคำที่ผสมเสียงรบกวนด้วยระดับ SNR มีค่าสูง จะให้เบอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้องสูงด้วย ในทางตรงกันข้ามถ้าระดับ SNR มีค่าต่ำ จะให้เบอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้องต่ำด้วย

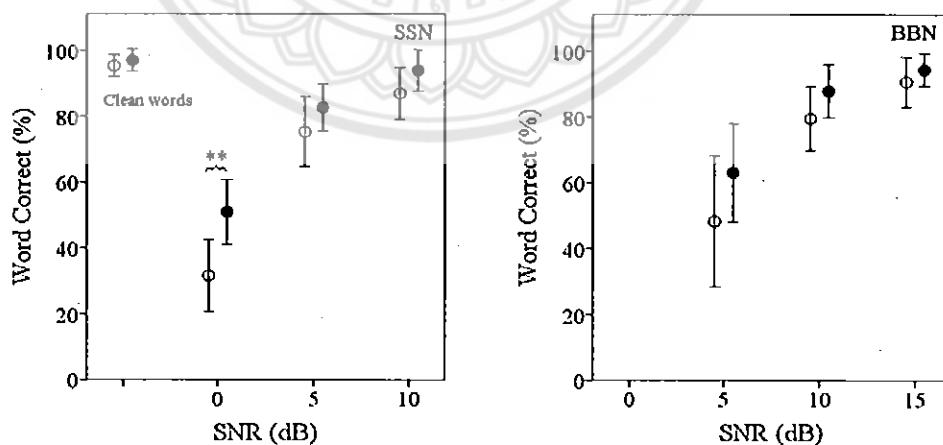
เมื่อพิจารณาตามขบวนของคำจากรูปที่ 4.3 จะเห็นได้ว่าคำพยานค์คู่มีเบอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องสูงกว่าคำพยานค์เดียวอย่างชัดเจนในทุกเงื่อนไข นอกจากนี้ยังเห็นได้ว่าเบอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องของคำพยานค์คู่ที่ระดับ SNR มีค่าสูงจะมีค่าใกล้เคียงกับคำที่ไม่มีเสียงรบกวน เมื่อเปรียบเทียบประสิทธิภาพการเข้ารหัสเสียงจะพบว่าการเข้ารหัสเสียงแบบ ACE มีเบอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องสูงกว่าการเข้ารหัสเสียงพูดแบบ CIS อย่างชัดเจนในทุกเงื่อนไขที่ทำการทดสอบ

จากการวิเคราะห์ความแปรปรวนแบบวัดซ้ำเพื่อหาความแตกต่างของเปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องของปัจจัยที่เกี่ยวข้อง 2 ปัจจัยคือรูปแบบการเข้ารหัสเสียงพูดและระดับ SNR จะพบว่ามีความแตกต่างของคะแนนเฉลี่ยอย่างมีนัยสำคัญสำหรับระดับ SNR ทุกค่าของชุดคำและเสียงรบกวนทั้งสองรูปแบบ ซึ่งแสดงให้เห็นว่าระดับของ SNR ที่มีค่าสูงขึ้นจะให้เปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องสูงขึ้นอย่างมีนัยสำคัญตัววัย

การวิเคราะห์ทางสถิติด้วย ANOVA ของความสัมพันธ์ระหว่างรูปแบบการเข้ารหัสเสียงและระดับ SNR พบว่าในกรณีของคำพยางค์เดียวที่มีเสียงรบกวนแบบ SSN มีความแตกต่างของปัจจัยเกี่ยวข้องอย่างมีนัยสำคัญดังนี้ รูปแบบการเข้ารหัสเสียง [$F(1,11)=81.93, p<0.0005$] และระดับของ SNR [$F(2,22)=116.78, p<0.0005$] รวมทั้งผลการทดสอบที่กันและกันระหว่างการเข้ารหัสเสียงและ



(ก) คำพยางค์เดียว



(ข) คำพยางค์คู่

รูปที่ 4.3 เปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องของเงื่อนไขที่กำหนดจำนวน 28 เงื่อนไข เครื่องหมาย (*) แสดงเปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องที่สูงกว่าอย่างมีนัยสำคัญที่ระดับ SNR เดียวกัน ($*p<0.05$ และ $**p<0.0005$) และเครื่องหมาย error bar แสดงค่าเบี่ยงเบนมาตรฐานของคะแนนคำที่ถูกต้อง

ระดับ SNR [$F(2,22)=4.43, p<0.05$] เมื่อเปรียบเทียบเปอร์เซ็นต์ค่าแนวเฉลี่ยของคำที่ถูกต้องของแต่ละคู่ของรูปแบบการเข้ารหัสเสียงและระดับ SNR พบร่วมกับการเข้ารหัสเสียงแบบ ACE จะให้คะแนนการเข้าใจเสียงพูดสูงกว่าแบบ CIS ที่ระดับ SNR เท่ากับ 5 และ 10 เดซิเบล เมื่อพิจารณาคำพยางค์เดียวที่มีเสียงรบกวนแบบ BBN จะพบว่ามีความแตกต่างของปัจจัยเกี่ยวข้องอย่างมีนัยสำคัญดังนี้ รูปแบบการเข้ารหัสเสียง [$F(1,11)=51.19, p<0.0005$] และระดับ SNR [$F(2,22)=103.08, p<0.0005$] แต่ไม่มีผลกระทบซึ่งกันและกันระหว่างสองปัจจัยนี้

เมื่อพิจารณากรณีคำพยางค์คู่ที่มีเสียงรบกวนแบบ SSN จะพบว่ามีความแตกต่างของปัจจัยเกี่ยวข้องอย่างมีนัยสำคัญดังนี้ รูปแบบการเข้ารหัสเสียงพูด [$F(1,11)=20.34, p<0.05$] และระดับ SNR [$F(2,22)=262.53, p<0.0005$] รวมทั้งผลการทดสอบที่แสดงถึงกันระหว่างหัวเรื่องของปัจจัย [$F(2,22)=6.75, p<0.05$] เมื่อเปรียบเทียบเปอร์เซ็นต์ค่าแนวเฉลี่ยของคำที่ถูกต้องของแต่ละคู่สำหรับรูปแบบการเข้ารหัสเสียงและระดับ SNR พบร่วมกับการเข้ารหัสเสียงแบบ ACE ให้คะแนนการเข้าใจเสียงพูดสูงกว่าแบบ CIS ที่ระดับ SNR เท่ากับ 0 เดซิเบลอย่างมีนัยสำคัญ เมื่อพิจารณาคำพยางค์คู่ที่มีเสียงรบกวนแบบ SSN พบร่วมกับการเข้ารหัสเสียงแบบ CIS ที่ระดับ SNR [$F(1,11)=12.62, p<0.05$] และระดับ SNR [$F(2,22)=278.27, p<0.0005$] แต่ไม่มีผลกระทบซึ่งกันและกันระหว่างสองปัจจัย

จากผลการทดลองที่ได้นี้พบว่ามีความสอดคล้องกับการศึกษา ก่อนหน้านี้ (Kiefer, Hohl, Sturzebecher, Pfennigdorff, & Gstoettner, 2001; Skinner et al., 2002) นั้นคือผู้ใช้ภาษาที่มีเสียงรบกวนส่วนใหญ่จะชอบรูปแบบการเข้ารหัสเสียงแบบ ACE มากกว่าแบบ CIS ยิ่งกว่านั้นจากการทดลองพบว่าการเข้ารหัสเสียงแบบ ACE จะให้คะแนนการเข้าใจเสียงพูดสูงกว่าแบบ CIS เนื่องจาก ACE จะเลือกแบบคำว่าที่มีขนาดแอมพลิจูดของเออนเวโลปมากที่สุดบางแบบด้วยความถี่เท่ากัน มาแปลงเป็นสัญญาณทางไฟฟ้าเพื่อไปกระตุนการทำงานของระบบประสาทการได้ยิน ทำให้แบบคำว่าที่มีขนาดแอมพลิจูดของเออนเวโลปน้อยกว่า ซึ่งเป็นข้อมูลที่มีความสำคัญน้อยกว่าจะถูกตัดทิ้งไปและเป็นการลดเสียงรบกวนของสัญญาณเสียงทั้งหมดด้วย (Nogueira et al., 2005)

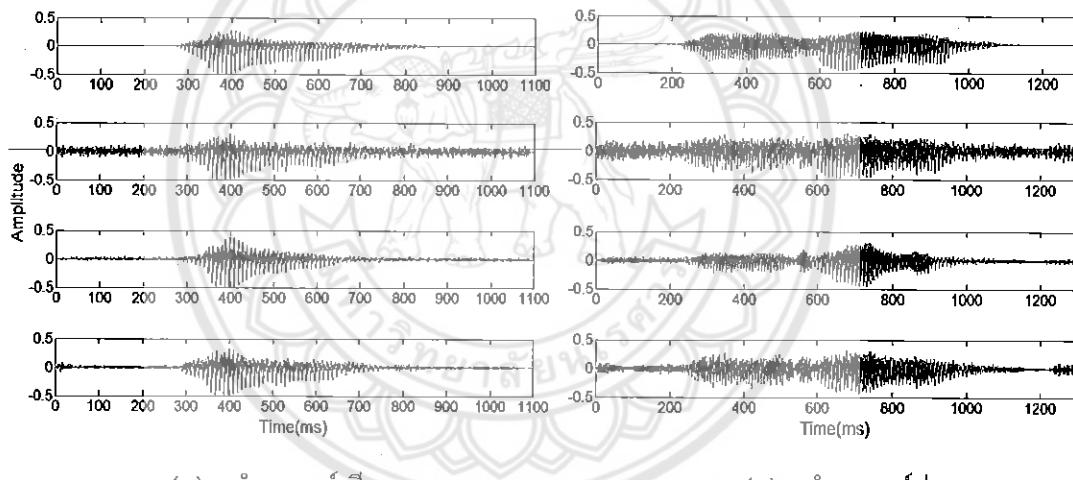
เมื่อพิจารณาประเภทของคำไทยที่ใช้ทดสอบ พบร่วมกับการศึกษาที่ได้จะสอดคล้องกับผลการทดลองของ Vaewvichit and Luangpitakchumpol (1999) ซึ่งการทดลองดังกล่าวพบว่าผู้ใช้ภาษาที่มีเสียงรบกวนของบริษัท Nucleus จำนวน 9 คน สามารถเข้าใจเสียงพูดของคำพยางค์คู่ได้ถูกต้องมากกว่าคำพยางค์เดียวในสภาพแวดล้อมที่ปราศจากเสียงรบกวน ผู้ใช้ภาษาที่มีเสียงรบกวนส่วนใหญ่ที่ใช้การฟังร่วมกับการอ่านริมฝีปาก (lip reading) สามารถเข้าใจเสียงพูดได้ดีกว่าผู้ใช้ภาษาที่มีเสียงรบกวนที่ใช้การฟังเพียงอย่างเดียวอีกด้วย

4.2 การเปรียบเทียบประสิทธิภาพการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF

ตัวอย่างรูปคลื่นของสัญญาณเสียงจำลองแสดงดังรูปที่ 4.4 และอิเล็กโทรโกราฟ (electrograms) ดังรูปที่ 4.5 ของคำพยางค์เดียว “ยาม” และคำพยางค์คู่ “แน่นอน” โดยแควรร์แสดงคำที่ไม่มีเสียงรบกวน 伟大ที่สองแสดงคำพยางค์เดียวที่มีเสียงรบกวนแบบ SSN และคำพยางค์คู่ที่มีเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5 เดซิเบล 伟大ที่สามและสี่แสดงคำที่ได้ลดเสียงรบกวนด้วยเทคนิค MBSS และ WF ตามลำดับ รูปคลื่นของสัญญาณเสียงจำลองจะแสดงแอมพลิจูดของ

สัญญาณเสียงที่เปลี่ยนแปลงตามเวลาในหน่วยมิลลิวินาที อิเล็กโทรโกรดแกรมจะแสดงแอมเพลจูดของสัญญาณไฟฟ้าในแต่ละอิเล็กโทรดที่เปลี่ยนแปลงตามเวลาในหน่วยมิลลิวินาที ซึ่งเป็นเครื่องมือที่มีประสิทธิภาพและมีประโยชน์ต่อการประเมินประสิทธิภาพของการประมวลผลสัญญาณเสียงและพารามิเตอร์ที่เกี่ยวข้องกับการกระตุนด้วยสัญญาณไฟฟ้าในระบบประสาทหูเทียม (Fan-Gang et al., 2008)

จากรูปที่ 4.4 และ 4.5 จะเห็นได้อย่างชัดเจนว่าเทคนิค MBSS และ WF สามารถลดเสียงรบกวนด้วยเทคนิค MBSS จะทำให้ข้อมูลของเสียงพูดบางส่วนหายไปด้วย ในขณะที่เทคนิค WF ยังคงให้เก็บรายละเอียดของข้อมูลเสียงพูดได้ดีกว่า โดยเฉพาะอย่างยิ่งที่องค์ประกอบความถี่ต่ำประมาณ 187 ถึง 533 เฮิรตซ์ หรือที่แบนด์ความถี่ (อิเล็กโทรด) ที่ 20 ถึง 22 ซึ่งเป็นช่วงความถี่ที่มีข้อมูลที่จำเป็นและมีผลต่อความเข้าใจเสียงพูดของมนุษย์เรา อย่างไรก็ตามเทคนิค WF ยังคงมีเสียงรบกวนหลงเหลืออยู่มากกว่าเทคนิค MBSS และจะเสียงรบกวนแบบ BBN จะยังคงเหลือมากกว่าเสียงรบกวนแบบ SSN



รูปที่ 4.4 ตัวอย่างรูปคลื่นสัญญาณของคำพยางค์เดียว “ยาม” และคำพยางค์คู่ “แป่นอน” โดยแก้กราฟแสดงคำที่ไม่มีเสียงรบกวน แตกต่างกันที่สองแสดงคำพยางค์เดียวที่มีเสียงรบกวนแบบ SSN และคำพยางค์คู่ที่มีเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5 เดซิเบล แตกต่างกันที่สามและแตกต่างกันที่สี่แสดงคำที่ลดเสียงรบกวนด้วยเทคนิค MBSS และ WF ตามลำดับ

2

RP

305

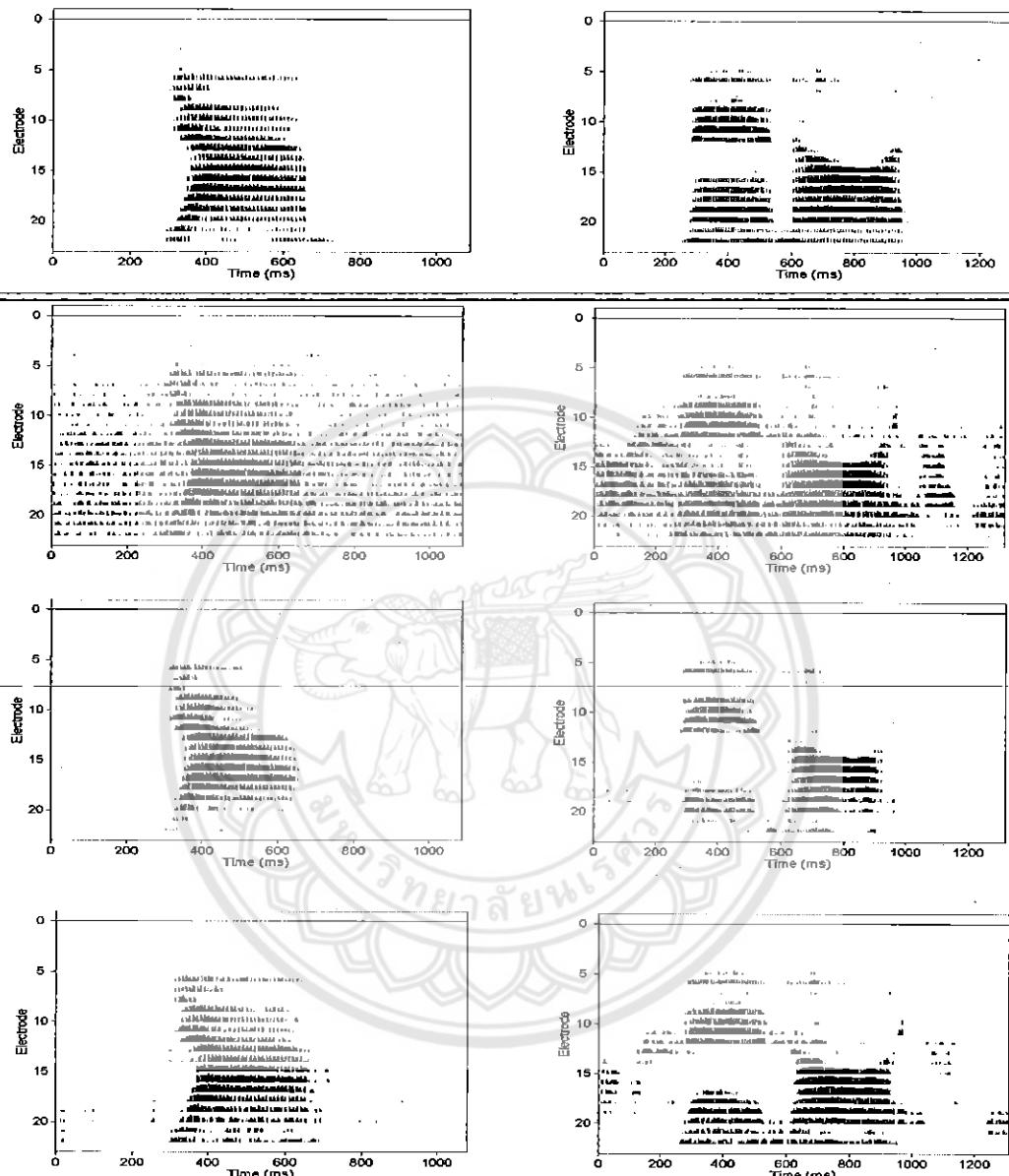
04634

2562



06 ม.ค. 2564

1034722.



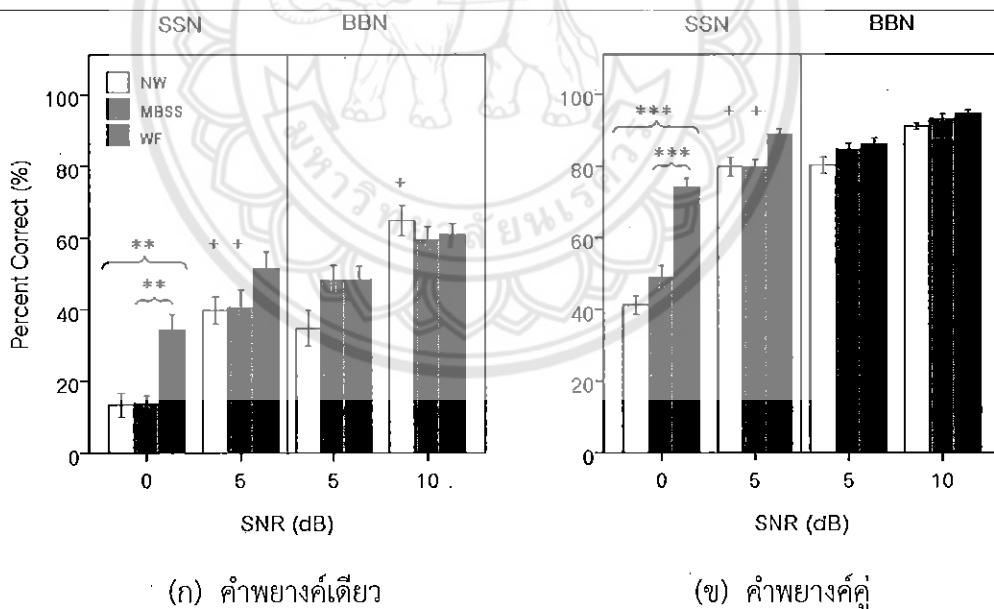
(ก) คำพยานค์เดียว

(ข) คำพยานค์คู่

รูปที่ 4.5 ตัวอย่างอิเล็ก trodo แกรมของคำพยานค์เดียว “ยาม” และคำพยานค์คู่ “แผ่นอน” โดยแก้ แรกแสดงคำที่ไม่มีเสียงรบกวน แล้วที่สองแสดงคำพยานค์เดียวที่มีเสียงรบกวนแบบ SSN และคำ พยานค์คู่ที่มีเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5 เดซิเบล แล้วที่สามและสี่แสดงคำที่ได้ลดเสียง รบกวนด้วยเทคนิค MBSS และ WF ตามลำดับ

การวัดเปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้อง สำหรับเงื่อนไขที่กำหนดเพื่อใช้ทดสอบจำนวน 24 เงื่อนไขตามตารางที่ 3.2 ประกอบด้วยชุดคำพยางค์เดียวและพยางค์คู่ที่มีเสียงรบกวน (noisy words) แบบ SSN ที่ SNR เท่ากับ 0 และ 5 เดซิเบลและเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5 และ 10 เดซิเบล ชุดคำทั้งหมดจะถูกลดเสียงรบกวนด้วยเทคนิค MBSS และ WF จากนั้นนำสัญญาณเสียงที่ได้ปรับปรุงแล้ว (enhanced words) ไปทำการเข้ารหัสเสียงแบบ ACE เงื่อนไขทั้งหมดจะถูกนำไปทดสอบกับอาสาสมัครจำนวน 20 คน คะแนนคำที่ถูกต้องในแต่ละเงื่อนไขของอาสาสมัครทุกคนจะถูกนำมาเฉลี่ยเพื่อคิดเป็นเปอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้อง

จากการทดลองที่ได้แสดงดังรูปที่ 4.6 จะเห็นว่าเปอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้องจะแปรผันตามค่าระดับ SNR ของเสียงรบกวนชนิดเดียวคันและชุดคำแบบเดียวกัน นั่นคือถ้าคำพยางค์เดียวหรือคู่ที่มีเสียงรบกวนด้วยระดับ SNR มีค่าสูง จะมีเปอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้องสูงด้วยในทางตรงกันข้ามถ้าระดับ SNR มีค่าต่ำ จะมีเปอร์เซ็นต์ค่าเฉลี่ยของคำที่ถูกต้องต่ำด้วย คำพยางค์คู่ที่มีเสียงรบกวนแบบ BBN จะมีค่าเฉลี่ยของคำที่ถูกต้องสูงขึ้นเพียงเล็กน้อยที่ SNR มีค่าเพิ่มขึ้น ทั้งคำพยางค์เดียวและพยางค์คู่ที่มีเสียงรบกวนแบบ SSN และ BBN ที่ SNR เท่ากับ 5 เดซิเบลจะมีค่าเฉลี่ยของคำที่ถูกต้องแตกต่างกันเพียงเล็กน้อย คำที่ถูกปรับปรุงแล้วจะมีค่าเฉลี่ยของคำที่ถูกต้องสูงกว่าคำที่มีเสียงรบกวนอย่างชัดเจนในเกือบทุกเงื่อนไขที่ทดสอบ การลดเสียงรบกวนด้วยเทคนิค WF จะให้



(ก) คำพยางค์เดียว

(ข) คำพยางค์คู่

รูปที่ 4.6 เปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องของอาสาสมัครที่มีการได้ยินปกติจำนวน 20 คน สำหรับคำที่มีเสียงรบกวน (NW) แบบ SSN ที่ SNR เท่ากับ 0 และ 5 เดซิเบลและแบบ BBN ที่ SNR เท่ากับ 5 และ 10 เดซิเบล เปรียบเทียบกับคำที่ถูกลดเสียงรบกวนด้วยเทคนิค MBSS และ WF เครื่องหมาย (+) หมายถึงเปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องที่มีความแตกต่างอย่างมีนัยสำคัญ เมื่อเปรียบเทียบด้วยระดับ SNR ต่างกัน เครื่องหมาย (*) หมายถึงเปอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องมีความแตกต่างอย่างมีนัยสำคัญเมื่อเปรียบเทียบด้วยระดับ SNR เดียวกัน (**p<0.05, ***p<0.0005)

ประประสิทธิภาพของการเข้าใจเสียงพูดตีก่าวเทคนิค MBSS อาย่างชัดเจนในเกือบทุกเงื่อนไขที่ทดสอบยกเว้นเงื่อนไขของคำพยานค์เดียวที่มีเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5 เดซิเบล

ในการณีคำพยานค์เดียวจะเห็นได้ว่าเทคนิค WF จะปรับปรุงประสิทธิภาพของการเข้าใจเสียงพูดได้มากกว่าเทคนิค MBSS โดยเฉพาะอย่างยิ่งในเงื่อนไขของเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 0 และ 5 เดซิเบล ทั้งสองเทคนิคจะมีคะแนนเฉลี่ยของคำที่ถูกต้องเกือบจะเท่ากันในเงื่อนไขของเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5 และ 10 เดซิเบล อย่างไรก็ตามคำที่ถูกปรับปรุงแล้วด้วยเทคนิคทั้งสองจะมีคะแนนเฉลี่ยของคำที่ถูกต้องน้อยกว่าคำที่มีเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 10 เดซิเบล ในกรณีคำพยานค์คู่จะเห็นได้ว่าคำที่ถูกปรับปรุงแล้วด้วยเทคนิค WF จะเพิ่มประสิทธิภาพของการเข้าใจเสียงพูดตีก่าวเทคนิค MBSS อาย่างน้อยได้ชัดโดยเฉพาะอย่างยิ่งในเงื่อนไขของเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 0 และ 5 เดซิเบล เทคนิค WF และ MBSS จะเพิ่มประสิทธิภาพของการเข้าใจเสียงพูดเพียงเล็กน้อยสำหรับเสียงรบกวนแบบ BBN กล่าวโดยสรุปคะแนนเฉลี่ยของคำที่ถูกต้องของคำพยานค์คู่จะสูงกว่าคำพยานค์เดียวในทุกเงื่อนไข

การวิเคราะห์ทางสถิติด้วย ANOVA ของความสัมพันธ์ระหว่างเทคนิคที่ใช้ในการลดเสียงรบกวนและระดับ SNR พบว่าในกรณีของคำพยานค์เดียวที่มีเสียงรบกวนแบบ SSN มีความแตกต่างของปัจจัยเกี่ยวข้องอย่างมีนัยสำคัญดังนี้ เทคนิคที่ใช้ในการลดเสียงรบกวน [$F(2,38)=17.60, p<0.0005$] และระดับ SNR [$F(1,19)=68.73, p<0.0005$] เมื่อเปรียบเทียบคะแนนเฉลี่ยของคำที่ถูกต้องของแต่ละคู่ของเทคนิคที่ใช้ในการลดเสียงรบกวนพบว่าเทคนิค WF มีการปรับปรุงความเข้าใจเสียงพูดสูงกว่าเทคนิค MBSS ที่ระดับ SNR เท่ากับ 0 เดซิเบล เมื่อเปรียบเทียบคะแนนเฉลี่ยของคำที่ถูกต้องของแต่ละคู่ของระดับ SNR พบว่าระดับ SNR ที่เพิ่มขึ้นของคำที่เสียงรบกวนและคำที่ถูกลดเสียงรบกวนด้วยเทคนิค MBSS จะทำให้การเข้าใจเสียงพูดสูงขึ้นด้วยอย่างมีนัยสำคัญ แต่คำที่ถูกลดเสียงรบกวนด้วยเทคนิค WF จะทำให้การเข้าใจเสียงพูดสูงขึ้นด้วยอย่างไม่มีนัยสำคัญ

กรณีคำพยานค์เดียวที่มีเสียงรบกวนแบบ BBN จะพบว่ามีความแตกต่างของปัจจัยเกี่ยวข้องอย่างมีนัยสำคัญดังนี้ ระดับ SNR [$F(1,19)=105.73, p<0.0005$] และมีผลกระบทซึ่งกันและกันระหว่างสองปัจจัย [$F(2,38)=7.95, p<0.05$] เบล เมื่อเปรียบเทียบคะแนนเฉลี่ยของคำที่ถูกต้องของแต่ละคู่ของระดับ SNR พบว่าคำที่มีเสียงรบกวนที่ SNR เท่ากับ 10 เดซิเบลจะให้การเข้าใจเสียงพูดได้ดีกว่าคำที่มีเสียงรบกวนที่ SNR เท่ากับ 5 เดซิเบล

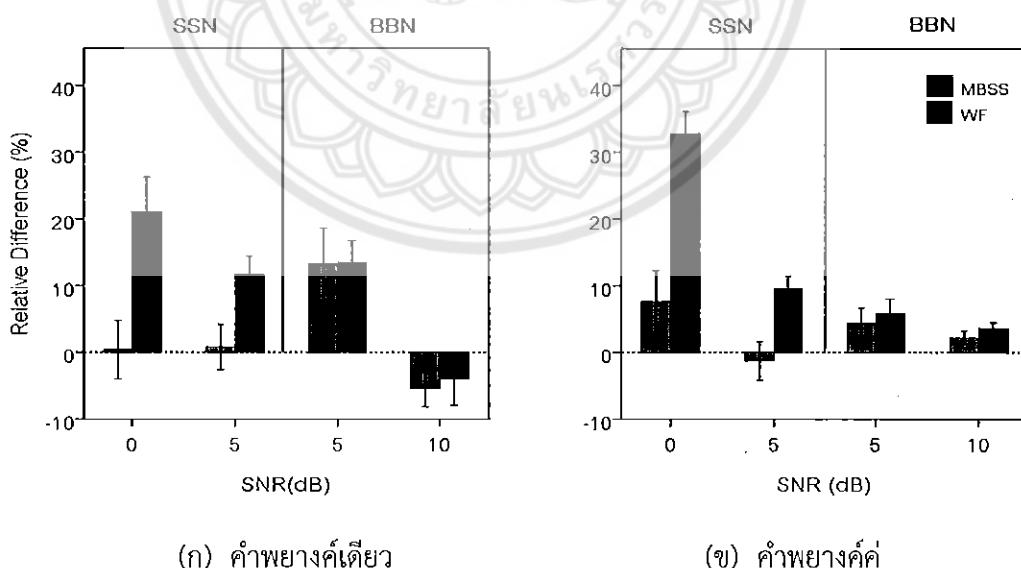
เมื่อพิจารณากรณีคำพยานค์คู่ที่มีเสียงรบกวนแบบ SSN จะพบว่ามีความแตกต่างของปัจจัยเกี่ยวข้องอย่างมีนัยสำคัญดังนี้ เทคนิคที่ใช้ในการลดเสียงรบกวน [$F(2,38)=50.28, p<0.0005$] และระดับ SNR [$F(1,19)=268.57, p<0.0005$] รวมทั้งผลกระบทซึ่งกันและกันระหว่างสองปัจจัย [$F(2,38)=12.27, p<0.0005$] เมื่อเปรียบเทียบคะแนนเฉลี่ยของคำที่ถูกต้องของแต่ละคู่สำหรับเทคนิคที่ใช้ในการลดเสียงรบกวนและระดับ SNR พบว่าให้ผลการทดลองเช่นเดียวกับคำพยานค์เดียวที่มีเสียงรบกวนแบบ SSN

การหาค่าความแตกต่างสัมพัทธ์ของเบอร์เช็นต์คะแนนเฉลี่ยของคำที่ถูกต้องระหว่างคำที่มีเสียงรบกวนและคำที่ลดเสียงรบกวนด้วยเทคนิค MBSS และ WF แสดงได้ดังรูปที่ 4.7 ค่าบกหมายถึงความเข้าใจเสียงพูดที่เพิ่มขึ้น ค่าลบหมายถึงความเข้าใจเสียงพูดที่ลดลง จากรูปจะเห็นว่า เมื่อทำการลดเสียงรบกวนด้วยทั้งสองเทคนิค ทำให้ผู้ฟังสามารถเข้าใจเสียงพูดได้เพิ่มขึ้นอย่างชัดเจนเมื่อกำหนด

เงื่อนไขเดียวกันทั้งคำพยางค์เดียวและคำพยางค์คู่ ยกเว้นกรณีคำพยางค์เดียวที่มีเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 10 เดซิเบล เมื่อทำการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF ทำผู้ฟังเข้าใจเสียงพูดได้ลดลง ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นโดยเฉลี่ยประมาณ 12% สำหรับคำที่ลดเสียงรบกวนด้วยเทคนิค WF และผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นเพียงเล็กน้อยโดยเฉลี่ยประมาณ 3% สำหรับคำที่ลดเสียงรบกวนด้วยเทคนิค MBSS

เมื่อพิจารณาโดยภาพรวมจะพบว่าผู้ฟังสามารถเข้าใจเสียงพูดที่ถูกลดเสียงรบกวนด้วยเทคนิค WF ดีกว่าเทคนิค MBSS นอกจากนี้การลดเสียงรบกวนด้วยเทคนิคทั้งสอง ทำให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นเป็นอย่างมากที่ระดับ SNR ต่ำ เช่น ระดับ SNR เท่ากับ 0 เดซิเบล เป็นต้น อย่างไรก็ตามเมื่อรายงานประสรุที่วิเคราะห์ในเสียงพูดจะแสดงที่ระดับ SNR สูงขึ้นหรือแนวโน้มของประสิทธิภาพการเข้าใจเสียงพูดจะน้อยกว่าคำที่มีเสียงรบกวนในบางเงื่อนไข เช่น คำพยางค์เดียวที่มีเสียงรบกวนแบบ BBN ที่ระดับ SNR เท่ากับ 10 เดซิเบล เช่น คำพยางค์คู่ที่มีเสียงรบกวนแบบ SSN ที่ระดับ SNR เท่ากับ 5 เดซิเบล เป็นต้น

จากการทดสอบการประเมินประสิทธิภาพของการเข้ารหัสเสียงและการลดเสียงรบกวนสามารถสรุปได้ว่าการเปลี่ยนแปลงเสียงของคำพยางค์เดียว ทำให้ผู้ฟังสามารถรู้จำคำหรือเข้าใจคำนั้นได้ไม่ยากต้อง การเปลี่ยนแปลงเสียงของคำพยางค์เดียวอาจเกิดจากรูปแบบการเข้ารหัสเสียง เทคนิคการลดเสียงรบกวนและระดับ SNR สามารถแบ่งออกเป็น 6 แบบดังนี้ การเพิ่มพยัญชนะ (consonant insertion) การลบพยัญชนะ (consonant deletion) การแทนที่พยัญชนะ



รูปที่ 4.7 ความแตกต่างสัมพัทธ์ของเบอร์เซ็นต์คะแนนเฉลี่ยของคำที่ถูกต้องระหว่างคำที่มีเสียงรบกวนและคำที่ลดเสียงรบกวนด้วยเทคนิค MBSS และ WF สำหรับเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 0, 5 เดซิเบลและเสียงรบกวนแบบ BBN ที่ SNR เท่ากับ 5, 10 เดซิเบล ตามลำดับ

(consonant replacement) การเปลี่ยนเสียงสระ (vowel change) การเปลี่ยนเสียงวรรณยุกต์ (tone change) และการเปลี่ยนแปลงแบบผสม (combined change)

จากผลการทดลองว่าการเพิ่มพยัญชนะจะเกิดขึ้นที่ตำแหน่งสุดท้ายของพยางค์ (final position) ตัวอย่างเช่น ‘ว่า’→‘ว่าง’ เป็นต้น ตรงกันข้ามกับการลบพยัญชนะและการแทนที่พยัญชนะสามารถเกิดขึ้นได้ที่ตำแหน่งแรกและสุดท้ายของพยางค์ ตัวอย่างเช่น ‘ปลา’→‘ลา’ ‘การ’→‘ก้า’ ‘แท้’→‘แท็’ ‘จับ’→‘จัด’ เป็นต้น การเปลี่ยนเสียงสระจะเกิดขึ้นได้ตอนข้างน้อยมาก ตัวอย่างเช่น ‘บุญ’→‘บัน’ เป็นต้น การเปลี่ยนเสียงวรรณยุกต์จะเกิดขึ้นจำนวนปานกลาง ตัวอย่างเช่น ‘ม้า’→‘หมา’ เป็นต้น คำบางคำอาจมีการเปลี่ยนแปลงแบบผสม ตัวอย่างเช่น ‘ลูก’→‘ลุง’ มีการเปลี่ยนเสียงวรรณยุกต์ /ท/→/สาเมัญ/ และการแทนที่พยัญชนะ /ก/→/ก/ เป็นต้น การเปลี่ยนแปลงของเสียงที่เกิดขึ้นเป็นสาเหตุทำให้เกิดความสับสนและการรับรู้คำมิดไป โดยส่วนมากจะเกิดจากการแทนที่พยัญชนะโดยเฉพาะที่ตำแหน่งแรกของพยางค์และการเปลี่ยนเสียงวรรณยุกต์ตามลำดับ

การเปลี่ยนแปลงเสียงของคำพยางค์คู่อาจเกิดการแทนที่ด้วยพยางค์ใหม่ที่พยางค์ต้น (preceding syllables) หรือพยางค์ท้าย (following syllables) ของคำ การแทนที่ด้วยพยางค์ใหม่เนื่องมาจากการเพิ่มพยัญชนะที่ตำแหน่งแรก ตำแหน่งสุดท้าย หรืออาจเป็นได้ทั้งสองตำแหน่งพร้อมกัน นอกจากนี้การเปลี่ยนแปลงเสียงของคำพยางค์คู่อาจเกิดจากการลบพยัญชนะ การเปลี่ยนเสียงวรรณยุกต์หรือมีการเปลี่ยนแปลงแบบผสม ตัวอย่างเช่น ‘รอรถ’→‘ล้อรถ’ มีการเปลี่ยนเสียงวรรณยุกต์ ‘หน้าต่าง’→‘หน้าตา’ มีการลบพยัญชนะที่ตำแหน่งสุดท้ายและเปลี่ยนเสียงวรรณยุกต์เป็นต้น ดังนั้นจะเห็นว่าเมื่อมีการเปลี่ยนแปลงของเสียงในรูปแบบใดก็ตาม จะทำให้ความหมายของคำเปลี่ยนไปด้วย

โดยทั่วไป ลักษณะของสัญญาณสำหรับคำพยางค์ไทยได้ๆ ประกอบด้วยสัญญาณของหน่วยเสียง 3 สัญญาณ คือ เสียงพยัญชนะเริ่มต้น (initial consonant) เสียงสระ (vowel) และเสียงพยัญชนะท้าย (final consonant) (Theera-Umpon, Chansareewittaya, & Auephanwiriyakul, 2011) เสียงวรรณยุกต์เป็นเสียงที่เกิดจากการเปลี่ยนแปลงรูปร่างของความถี่มูลฐาน (F0 contour) ตลอดทั้งพยางค์ หน่วยเสียงทั้งสามจะรวมกันที่ช่วงเวลาสั้นๆ ทำให้เกิดเป็นเสียงพยางค์ เสียงสระเป็นเสียงที่มีบทบาทสำคัญเนื่องจากเป็นองค์ประกอบหลักของคำและเป็นตัวกำหนดจำนวนพยางค์ของแต่ละคำ ดังนั้นจากการทดลองจะเห็นว่าผู้ฟังจะเกิดความสับสนในการรู้จำเสียงสระตอนข้างน้อยมาก ในทางตรงกันข้ามผู้ฟังจะเกิดความสับสนจากการรู้จำเสียงพยัญชนะได้มากกว่า โดยเฉพาะการเปลี่ยนแปลงของเสียงโดยการแทนที่พยัญชนะ การฟังเสียงพยัญชนะต้นจะทำให้เกิดความสับสนได้มากกว่าพยัญชนะท้ายของพยางค์ เนื่องจากผู้ฟังจำเป็นต้องรับรู้เสียงของพยัญชนะต้นที่เป็นหน่วยเสียงแรกก่อนเสมอ เมื่อสัญญาณเสียงผ่านการประมวลผลหรือมีการผสมเสียงรบกวน ผู้ฟังจะเกิดความสับสนและรู้จำคำมิดของเสียงของพยัญชนะต้นได้ยังกว่าหน่วยเสียงอื่น นอกจากนี้องค์ประกอบของหน่วยเสียงต่างๆ อาจเกิดจากการซ้อนทับกันของสัญญาณ ตัวอย่างเช่น องค์ประกอบของเสียงพยัญชนะต้นอาจจะรวมกับเสียงสระ ดังนั้นขั้นตอนการประมวลผลของสัญญาณต่างๆ เช่น การเข้ารหัสเสียงและการผสมเสียงรบกวน เป็นต้น อาจทำให้องค์ประกอบเหล่านี้เกิดการผิดเพี้ยนของสัญญาณและนำไปสู่การเปลี่ยนแปลงของเสียง

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้ได้ศึกษาเกี่ยวกับการประเมินประสิทธิภาพ (performance evaluation) ของความเข้าใจเสียงพูด (speech intelligibility) ของผู้ใช้ภาษาทุกที่ยกภาษาให้สภาพแวดล้อมที่แตกต่างกัน โดยใช้ชุดคำพยางค์เดียวและคำพยางค์คู่ที่มีเสียงรบกวนแบบ SSN (speech-shaped noise) และแบบ BBN (babble noise) ที่ระดับ SNR เท่ากับ 0, 5 และ 10 เดซิเบล ชุดคำทั้งหมดจะนำมาสร้างเป็นเสียงจำลอง (vocoded speech) และนำมาทดสอบกับอาสาสมัครที่มีการได้ยินปกติ (normal hearing) เพื่อศึกษาเปรียบเทียบประสิทธิภาพการเข้าใจเสียงพูดเมื่อมีการเข้ารหัสเสียงแบบ CIS และ ACE รวมทั้งการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF ร่วมกับการเข้ารหัสเสียงแบบ ACE สามารถสรุปผลการทดลองได้ดังนี้

การทดสอบเพื่อเปรียบเทียบประสิทธิภาพการเข้าใจเสียงพูดของชุดคำพยางค์เดียวและคำพยางค์คู่ที่มีการเข้ารหัสเสียงแบบ CIS และ ACE ทั้งที่ไม่มีเสียงรบกวน (clean words) และที่มีเสียงรบกวน (noisy words) ของอาสาสมัครจำนวน 12 คน พบว่าการเข้ารหัสเสียงแบบ ACE จะให้ประสิทธิภาพของการเข้าใจเสียงพูดได้ดีกว่าแบบ CIS ในทุกเงื่อนไขที่ทดสอบ และให้ประสิทธิภาพการเข้าใจเสียงพูดได้ดีกว่าอย่างมีนัยสำคัญทางสถิติในเงื่อนไขของคำพยางค์เดียวที่มีเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 5, 10 เดซิเบล และคำพยางค์คู่ที่มีเสียงรบกวนแบบ SSN ที่ SNR เท่ากับ 0 เดซิเบล เนื่องจากการเข้ารหัสเสียงแบบ ACE มีลักษณะที่สำคัญคือการเลือกชุดแบบเดียวที่มีพลังงานสูงสุด (maxima selection) เพียงบางส่วนเพื่อนำไปสร้างเป็นเสียงจำลอง ในงานวิจัยนี้ได้เลือกแบบเดียวที่มีพลังงานสูงสุดจำนวน 12 แบบจากทั้งหมด 22 แบบ (12-of-20 channel) ดังนั้นแบบเดียวที่คงเหลืออยู่จะถูกตัดทิ้งไป ทำให้สัญญาณเสียงโดยรวมมีปริมาณเสียงรบกวนลดลงได้ ทำให้ผู้ฟังสามารถเข้าใจเสียงพูดของการเข้ารหัสเสียงแบบ ACE ได้ดีกว่าแบบ CIS

การทดสอบเพื่อเปรียบเทียบประสิทธิภาพการเข้าใจเสียงพูดของชุดคำพยางค์เดียวและคำพยางค์คู่ที่มีเสียงรบกวน (noisy words) และที่มีการลดเสียงรบกวน (enhanced words) ด้วยเทคนิค MBSS และ WF ของอาสาสมัครจำนวน 20 คน พบว่าการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF ช่วยให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นในเงื่อนไขที่ทดสอบเป็นส่วนใหญ่และผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นอย่างมีนัยสำคัญทางสถิติด้วยเทคนิค WF ในบางเงื่อนไข การลดเสียงรบกวนด้วยเทคนิค WF ทำให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีกว่าเทคนิค MBSS ในเกือบทุกเงื่อนไขที่ทดสอบ เนื่องจากเทคนิค MBSS จะลดเสียงรบกวนได้มากกว่าเทคนิค WF แต่ขณะเดียวกันเทคนิค MBSS ได้ลดหรือตัดข้อมูลของเสียงพูดบางส่วนทิ้งไปด้วย ซึ่งเป็นข้อมูลที่จำเป็นต่อการเข้าใจเสียงพูดของมนุษย์ การลดเสียงรบกวนด้วยเทคนิค WF จะยังคงมีเสียงรบกวนเหลืออยู่อย่างเห็นได้ชัด แต่เทคนิค WF สามารถรักษาข้อมูลของเสียงพูดส่วนใหญ่ไว้ได้ดีกว่าและใกล้เคียงกับข้อมูลเสียงที่ไม่มีเสียงรบกวน (clean words) มากกว่าเทคนิค MBSS

อย่างไรก็ตามการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF จะทำให้ผู้ฟังสามารถเข้าใจเสียงพูดได้ดีขึ้นอย่างชัดเจนเมื่อเทียบกับคำที่มีเสียงรบกวนที่ระดับ SNR มีค่าต่ำ ผู้ฟังมีแนวโน้มการเข้าใจเสียงพูดได้ลดลงที่ระดับ SNR มีค่าสูง นอก จากนี้แนวโน้มการเข้าใจเสียงพูดของคำที่มีการลดเสียงรบกวนอาจจะน้อยกว่าคำที่มีเสียงรบกวนสำหรับที่ระดับ SNR มีค่าสูง เนื่องจากการลดเสียงรบกวนอาจจะน้อยกว่าคำที่มีเสียงรบกวนสำหรับที่ระดับ SNR มีค่าสูง เนื่องจากการลดเสียงรบกวนด้วยเทคนิค MBSS และ WF อาจทำให้ข้อมูลของเสียงพูดผิดเพี้ยน (distortion) ซึ่งอาจเกิดจากขั้นตอนการประมวลผลสัญญาณได้ โดยเฉพาะอย่างยิ่งเทคนิค MBSS ที่การประมาณค่าสเปกตรัมของเสียงรบกวนอาจจะมีค่ามากเกินไป ทำให้การลบเสียงรบกวนออกจากเสียงที่มีเสียงรบกวนอยู่ในปริมาณมากเกินไปและอาจส่งผลต่อการลบข้อมูลของเสียงพูดไปด้วย ลักษณะของเสียงที่ผิดเพี้ยนไปดังกล่าวอาจส่งผลต่อการเข้าใจเสียงพูดของคนเรามากกว่าคำที่มีเสียงรบกวนที่น้อยๆ ได้

จากการทดลองที่ได้จะเป็นข้อมูลที่เป็นประโยชน์ต่อผู้ใช้ประชาทุพารามีสำหรับคนไทย หรือคนต่างชาติต่างภาษา เพื่อใช้ในศึกษาแนวโน้มของประสิทธิภาพการเข้าใจเสียงพูดของการเข้ารหัสเสียงและเทคนิคการลดเสียงรบกวนของระบบประชาทุพารามี ซึ่งนำไปสู่การพัฒนารูปแบบการเข้ารหัสเสียงแบบอื่นๆ รวมทั้งเทคนิคการลดเสียงรบกวนใหม่ๆ สำหรับอุปกรณ์หูเทียม (auditory prostheses) ในอนาคต

5.2 ข้อเสนอแนะ

ในการวิจัยนี้ยังคงมีข้อจำกัดในเรื่องของชุดคำที่ใช้สำหรับทดสอบเพื่อศึกษาประสิทธิภาพการเข้าใจเสียงพูดของพารามิเตอร์ต่างๆ ที่เกี่ยวข้องกับรูปแบบการเข้ารหัสเสียงหรือเทคนิคการลดเสียงรบกวนต่างๆ เมื่อจากชุดคำที่ใช้เป็นคำที่ใช้ทดสอบกับผู้ที่บกพร่องทางการได้ยิน (hearing-impaired listeners) ที่ใช้ในทางคลินิกนานา民族มากกว่า 40 ปีแล้ว (Kangsanarak, 1980a; Komalarajun, 1979) และจำนวนคำที่ใช้มีปริมาณจำกัด ทำให้การออกแบบการทดลองค่อนข้างจำกัดในเรื่องการศึกษาพารามิเตอร์ต่างๆ ที่อาจมีผลต่อประสิทธิภาพการเข้าใจเสียงพูดของผู้ใช้ประชาทุพารามีได้ เช่นเดียวกับงานวิจัยของ Li et al. (2011) และ F. Chen et al. (2015)

ชุดคำไทยที่นำมาใช้ในการทดสอบในปัจจุบันอาจจะไม่เพียงพอและไม่มีความหลากหลายนอกเหนือจากเป็นคำพยางค์เดียวหรือคำพยางค์คู่ ความซุ่มคำอื่นๆ ที่อาจเป็นประโยชน์ต่อวิจัยหรือคลินิกด้านนี้ ความหลากหลายของชุดคำที่ใช้ทดสอบเป็นสิ่งที่สำคัญ ชุดทดสอบที่เป็นพยัญชนะ (consonants) สรร (vowels) หรือวรรณยุกต์เสียง (tones) หรือคำ(words) อาจเหมาะสมกับการทดสอบเพื่อวิเคราะห์ความสามารถในการรับรู้รายละเอียดต่างๆ ของข้อมูลเสียงพูดสำหรับผู้ที่บกพร่องทางการได้ยิน เมื่อมีการเปลี่ยนแปลงพารามิเตอร์ต่างๆ ก็สามารถที่จะเข้าใจการรับเสียงพูดและสามารถสังเกตเห็นรายละเอียดได้ง่ายขึ้น หากชุดทดสอบเป็นประโยชน์ (sentences) อาจเหมาะสมกับการทดสอบเพื่อวิเคราะห์ความแตกต่างในการรับรู้ของความหมาย (meaning) บริบทของเนื้อหา (context) หรือจังหวะน้ำเสียง (rhythms) ซึ่งใกล้เคียงกับการสื่อสารที่ใช้ในชีวิตประจำวันอย่างแท้จริง

หากประกาศจากข้อจำกัดในเรื่องของชุดคำที่ใช้ทดสอบ งานวิจัยควรศึกษาเกี่ยวกับประสิทธิภาพการเข้าใจเสียงพูดของพารามิเตอร์ต่างๆ เพิ่มเติมในเชิงของประเภทของเสียงรบกวน (noise types) ระดับของ SNR (SNR levels) และเทคนิคการลดเสียงรบกวน (noise reduction techniques) ที่มีความหลากหลายมากขึ้น พารามิเตอร์เหล่านี้ยังสามารถปรับให้เหมาะสมตามความ

พึงพอใจของผู้ใช้ประสาทหูเทียมแต่ละคนได้ รวมทั้งการทดสอบกับผู้ใช้ประสาทหูเทียมโดยตรงจะทำให้ผลการศึกษาที่ได้มีความชัดเจนและน่าเชื่อถือมากขึ้น อย่างไรก็ตามการศึกษานี้จะเป็นก้าวแรกในการนำไปสู่การศึกษากับผู้ประสาทหูเทียมสำหรับคนไทยต่อไป



បរណ្តុំក្រោម

- Berouti, M., Schwartz, R., & Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise. *ICASSP-79. IEEE International Conference on Acoustics, Speech and Signal Processing*, 208-211.
- Bolner, F., Goehring, T., Monaghan, J., van Dijk, B., Wouters, J., & Bleek, S. (2016, 20-26 March). *Speech enhancement based on neural networks applied to cochlear implant coding strategies*. Paper presented at the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China.
- Chen, F., Hu, Y., & Yuan, M. (2015). Evaluation of Noise Reduction Methods for Sentence Recognition by Mandarin-Speaking Cochlear Implant Listeners. *Ear and Hearing*, 36(1), 61-71.
- Chen, F., & Loizou, P. C. (2011). Predicting the Intelligibility of Vocoded Speech. *Ear and Hearing*, 32(3), 331-338.
- Cochlear. (2002). ACE™ and CIS DSP Strategies Software Requirements Specification, Part Number: N95287F Issue 1. Lane Cove, New South Wales, Australia.
- Dawson, P. W., Mauger, S. J., & Hersbach, A. A. (2011). Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus(R) cochlear implant recipients. *Ear Hear.*, 32(3), 382-390.
- Fan-Gang, Z., Rebscher, S., Harrison, W., Xiaoan, S., & Haihong, F. (2008). Cochlear implants: system design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering*, 115-142.
- Fink, N., Furst, M., & Muchnik, C. (2012). Improving word recognition in noise among hearing-impaired subjects with a single-channel cochlear noise-reduction algorithm. *J. Acoust. Soc. Am.*, 132(3), 1718-1731.
- Gustafsson, H., Nordholm, S. E., & Claesson, I. (2001). Spectral subtraction using reduced delay convolution and adaptive averaging. *IEEE Transactions on Speech and Audio Processing*, 9(8), 799-807.
- Hu, Y., & Loizou, P. C. (2002). A subspace approach for enhancing speech corrupted by colored noise. *IEEE Signal Processing Letters*, 9(7), 204-206.
- Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *Journal of the Acoustical Society of America*, 122(3), 1777-1786.
- Hu, Y., Loizou, P. C., Li, N., & Kasturi, K. (2007). Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. *Journal of the Acoustical Society of America*, 122(4), 128-134.

- Kallel, F., Frikha, M., Ghorbel, M., Ben Hamida, A., & Berger-Vachon, C. (2012). Dual-channel spectral subtraction algorithms based speech enhancement dedicated to a bilateral cochlear implant. *Applied Acoustics*, 73(1), 12-20.
- Kamath, S., & Loizou, P. (2002). A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vols I-IV, Proceedings*, 4164-4167.
- Kangsanarak, B. (1980a). *Development of Thai Spondee Words for Speech Audiometry*. (Master of Art (Communication disorders)), Mahidol University.
- ~~Kangsanarak, B. (1980b). *Development of Thai Spondee Words for Speech Audiometry*. (Master of Art (Communication disorders)), Mahidol University.~~
- Kiefer, J., Hohl, S., Sturzebecher, E., Pfennigdorff, T., & Gstoettner, W. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. *Audiology*, 40(1), 32-42.
- Kokkinakis, K., Azimi, B., Hu, Y., & Friedland, D. R. (2012). Single and Multiple Microphone Noise Reduction Strategies in Cochlear Implants. *Trends Amplif*, 16(2), 102-116.
- Komalarajun, S. (1979). *Development of Thai Speech Discrimination Materials*. (Master of Art (Communication disorders)), Mahidol University.
- Koning, R., Madhu, N., & Wouters, J. (2015). Ideal Time-Frequency Masking Algorithms Lead to Different Speech Intelligibility and Quality in Normal-Hearing and Cochlear Implant Listeners. *IEEE Transactions on Biomedical Engineering*, 62(1), 331-341.
- Li, J. F., Yang, L., Zhang, J. P., Yan, Y. H., Hu, Y., Akagi, M., & Loizou, P. C. (2011). Comparative intelligibility investigation of single-channel noise-reduction algorithms for Chinese, Japanese, and English. *Journal of the Acoustical Society of America*, 129(5), 3291-3301.
- Lockwood, P., & Boudy, J. (1992). Experiments with a Nonlinear Spectral Subtractor (Nss), Hidden Markov-Models and the Projection, for Robust Speech Recognition in Cars. *Speech Communication*, 11(2-3), 215-228.
- Loizou, P., & Gibak, K. (2011). Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Transactions on Audio, Speech and Language Processing*, 47-56.
- Loizou, P. C. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine*, 15(5), 101-130.

- Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. In A. R. Moller (Ed.), *Cochlear and Brainstem Implants* (Vol. 64, pp. 109-143). Basel, Switzerland: Karger.
- Loizou, P. C. (2013). *Speech Enhancement: Theory and Practice*. Boca Raton, Florida: CRC Press, Taylor & Francis Group.
- Loizou, P. C., Lobo, A., & Hu, Y. (2005). Subspace algorithms for noise reduction in cochlear implants. *Journal of the Acoustical Society of America*, 118(5), 2791-2793.
-
- Nogueira, W., Buchner, A., Lenarz, T., & Edler, B. (2005). A psychoacoustic "NoFM"-type speech coding strategy for cochlear implants. *Eurasip Journal on Applied Signal Processing*, 2005(18), 3044-3059.
-
- Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*, 114(1), 446-454.
-
- Scalart, P., & Vieira, J. (1996). Speech enhancement based on a priori signal to noise estimation. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing, Conference Proceedings, Vols 1-6*, 629-632.
-
- Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., & Holden, T. A. (2002). Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear and Hearing*, 23(3), 207-223.
- Theera-Umpon, N., Chansareewittaya, S., & Auephanwiriyakul, S. (2011). Phoneme and tonal accent recognition for Thai speech. *Expert Systems with Applications*, 38(10), 13254-13259.
- Vaewwichit, K., & Luangpitakchumpol, P. (1999). Cochlear implantation in Thailand. *J Laryngol Otol*, 113(6), 515-517.
- Verschuur, C., Lutman, M., & Wahat, N. H. (2006). Evaluation of a non-linear spectral subtraction noise suppression scheme in cochlear implant users. *Cochlear Implants Int*, 7(4), 193-196.
- Weiss, M. R. (1993). Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor. *Journal of Rehabilitation Research and Development*, 30(1), 117-128.
- Wilson, B. S. (2017). The cochlear implant and possibilities for narrowing the remaining gaps between prosthetic and normal hearing. *World J Otorhinolaryngol Head Neck Surg*, 3(4), 200-210.
- Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: A remarkable past and a brilliant future. *Hear Res*, 242(1-2), 3-21.

Yang, L. P., & Fu, Q. J. (2005). Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *Journal of the Acoustical Society of America*, 117(3), 1001-1004.





ภาคผนวก ก

ผลงานที่ได้รับการตีพิมพ์

1. Evaluation of Thai speech intelligibility based on noise reduction techniques for cochlear implants, Dachasilaruk, S. , Luangsawang, C. , Jantharamin, N., Kaewsri, S., Doungta, P. (2018) Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017.
2. Speech Intelligibility Evaluation of Sound Coding Strategies in Noisy Environments for Thai Cochlear Implant Users, Siriporn Dachasilaruk and Apichai Rungruang, Naresuan University Journal: Science and Technology, volume 26, no. 4, 2018.
3. Speech intelligibility enhancement for Thai-speaking cochlear implant listeners, Siriporn Dachasilaruk, Niphat Jantharamin, and Apichai Rungruang, Indonesian Journal of Electrical Engineering and Computer Science, vol. 13, No. 3, March 2019, pp. 866~875.



ภาคผนวก ข

ชุดคำพยางค์เดียวและคู่

ลำดับที่	คำไทย	การออกเสียง	คำแปล
1	อ่าน	an	read
2	เป็ด	pet	duck
3	เต่า	tao	turtle
4	ไกล	kai	far
5	ผ้า	pha	cloth
6	บ้าน	ban	house
7	ม้า	ma	horse
8	พาน	phan	tray
9	ลิง	ling	monkey
10	ฉิ่ง	ching	cymbal
11	แหวน	wæn	ring (noun)
12	จาน	chan	plate
13	ช้อน	chon	spoon
14	เรือ	rue	ship
15	ยุง	yung	mosquito
16	ควาย	khwai	buffalo
17	ฝน	fɔn	rain
18	หนู	nú	rat
19	เด็ก	dekk	child
20	หอย	hoi	shell
21	งู	ngu	snake
22	ฟัน	fan	tooth
23	ธง	thong	flag
24	ปืน	puen	gun
25	ยาม	yam	guard
26	รถ	rot	car
27	ตาย	tai	die
28	จริง	ching	true
29	ถ้วย	thuay	cup

ลำดับที่	คำไทย	การออกเสียง	คำแปล
30	เสื้อ	sûe	shirt
31	ท่า	thâ	posture
32	ชัด	chat	clear
33	ซ่อม	sôm	repair
34	แม่	mâe	mother
35	ผึ้ง	phuêng	bee
36	ขวด	khwuat	bottle
37	อุ่น	un	warm
38	บัว	buâa	lotus
39	ฉาบ	chap	plaster (verb)
40	ໄຟ	fai	mole
41	เจาะ	ngua	rambutan
42	ดาว	dao	star
43	เก้า	kâo	9 (number)
44	เล่น	lêñ	play
45	ห่าน	han	goose
46	ค่ำ	kâm	evening
47	อ้วน	uân	fat
48	ยอด	yot	top
49	ตาม	tam	follow
50	ร้อน	rón	hot
51	ปู	pu	crab
52	ชั้น	chan	level
53	ผัด	phat	fry
54	ดำ	dam	black
55	แผล	thaew	row
56	แสง	sâeng	light
57	จันทร์	chan	moon
58	บุญ	bun	merit
59	ฝูง	fung	flock
60	มาก	mâk	a lot
61	พบ	phop	to be found
62	น้อย	noi	little

ลำดับที่	คำไทย	การออกเสียง	คำแปล
63	แก่	kae	old
64	ง่าย	ngai	easy
65	叔	lung	uncle
66	ทอง	thong	gold
67	ฉัน	chan	I (1st person reference)
68	ฟัง	fang	listen
69	ว่าว	waw	kite
70	อิ่ม	im	full, have enough food
71	ไทย	thai	Thai
72	ซื้อ	sue	buy
73	เดา	do	slight kick
74	ลาต	lat	slop
75	ชาย	chai	light (verb)
76	แว่น	waen	glasses
77	ต้ม	tom	boil
78	ดิน	din	soil
79	จบ	chop	end, finish
80	คือ	kue	to be
81	ผ่าน	phan	pass
82	สร้อย	soi	necklace
83	งาม	ngam	beauty
84	ยิง	ying	shoot
85	ราก	rak	root
86	ชีพ	chip	life
87	ฟ้า	fa	sky
88	หอม	hom	onion
89	พระ	phra	monk
90	ผึ้น	fin	opium
91	กรรม	kam	karma, action
92	ปลา	pla	fish
93	ปาก	pak	mouth
94	อุ้ม	um	hold, carry
95	ญาติ	yat	relative

ลำดับที่	คำไทย	การออกเสียง	คำแปล
96	เร็ว	rew	fast
97	ผม	phǒm	hair
98	ไม้	mai	wood
99	หก	hok	spill
100	กล้วย	kluay	banana
101	ดื่ม	duem	drink
102	งอ	ngo	bent
103	ขุน	khun	not clear (water)
104	ข้าง	khāng	side
105	โต๊ะ	to	table
106	เบ็ด	bet	fishing rod
107	เคย	koey	use to be
108	ช้า	cha	slow
109	หลับ	lap	sleep
110	นิด	nit	little
111	สิบ	sip	10 (number)
112	แท้	thae	genuine, real
113	พูด	phút	speak
114	ฟอง	fong	bubble
115	job	chop	digging tool
116	กิน	kin	eat
117	ดู	du	see, watch
118	ขวา	khwā	right (direction)
119	ไว้	wai	keep
120	ตัว	tua	classifier for animals
121	ขาย	khai	sell
122	กับ	kap	with
123	ไฟ	fai	fire
124	หนัง	nang	skin
125	โดย	doy	by
126	ปล่อย	ploy	let something go
127	เสียง	siong	sound
128	เป็น	pen	to be

ลำดับที่	คำไทย	การออกเสียง	คำแปล
129	ได้	dâi	get
130	เข้า	khâo	get in, enter
131	ห้า	hâ	5 (number)
132	ล้าน	lan	million
133	อ่อน	on	young
134	พี	phi'	bother
135	งาน	ngan	job
136	ว่า	wâ	blame
137	แก้ว	kwaew	glass
138	กวาด	kwat	sweep
139	ค่ำ	khâm	evening
140	สั้น	sân	short
141	โจร	chon	theif
142	ข้าว	khâo	rice
143	ค้า	kha'	sell
144	เธอ	thoe	2nd person reference
145	การ	kan	noun prefix
146	สอง	söng	2 (number)
147	ลุก	luk	get up
148	หมู	mû	pork
149	นาย	nai	boss
150	กล้อง	klöng	camera
151	ทั่ว	thuâ	overall
152	มอง	mong	look
153	ปี	pi	year
154	วัด	wat	temple
155	หน้า	nâ	face
156	คน	khon	people
157	จับ	cháp	catch
158	นั่ง	nâng	sit
159	หมอ	mô	doctor
160	ลูก	luîk	child
161	เสีย	sie	spoil

ลำดับที่	คำไทย	การออกเสียง	คำแปล
162	เปล่า	plao	empty
163	หู	hū	ear
164	ดัง	dang	noisy, famous
165	กำ	kam	hold
166	ส้ม	sōm	orange
167	ก่อน	kōn	before
168	เสร็จ	set	finish
169	คุณ	khun	2 person reference
170	เนื้อ	nue	beef
171	จำ	cham	remember
172	เก็บ	kep	harvest (verb)
173	แล้ว	læw	finish
174	ตา	ta	eye
175	เลี้ยง	lieng	to give a treat
176	คอ	kho	neck
177	ใช้	chai	use
178	ยั่ง	yāng	stand
179	วัน	wan	day
180	ปิด	pit	close
181	สั่ง	sang	order
182	มัน	man	oily, it (3 rd person reference for animals and non-living things.
183	ความ	khwam	matter, noun prefix
184	บน	bon	above
185	แก้	kāe	repair
186	ใส่	sai	put on
187	นา	na	rice field
188	ดี	di	good
189	ทัน	than	catch up with
190	คืน	kuen	return

ลำดับที่	คำไทย	การออกเสียง	คำแปล
1	นกยูง	nok.yung	peacock
2	พัดลม	phat.lom	fan
3	ไฟฟ้า	fai.fa	electricity
4	เด็กชาย	dek.chai	boy
5	ช้อนส้อม	chǒn.sōm	spoon and fork
6	ของเล่น	khǒng.lēn	toy
7	รถยนต์	rot.yon	car
8	รถไฟ	rot.fai	train (noun)
9	นายก	na.yok	prime minister
10	เอาใจ	ao.chai	please (verb)
11	กำลัง	kam.lang	power
12	เร็วกว่า	rew.kwa	faster
13	ตอนเย็น	ton.yen	evening
14	ร่างกาย	rāng.kai	body
15	ด้วยกัน	duay.kan	together
16	คนสวย	khon.suay	beautiful person
17	เท่ากัน	thâo.kan	equal
18	อย่าทำ	ya.tham	Don't do it.
19	พร้อมแล้ว	phrōm.laew	ready
20	พูดเล่น	phút.lēn	chitchat
21	สงสาร	sǒng.san	pity
22	รู้แล้ว	ru.laew	already knew it.
23	เปิดไฟ	poet.fai	turn on the light
24	พ่อแม่	phô.mae	parents
25	ตื่นนอน	tuen.non	get up
26	พุคตรา/พุตรา	phut.sa	jujube
27	รถเมล์	rot.may	bus
28	ปกครอง	pok.khrong	govern
29	พี่น้อง	phi.nōng	sibling
30	แสงแดด	săeng.deat	sunlight
31	เสื้อผ้า	sue.pha^	cloth
32	สังคม	săng.khom	society
33	ต้องการ	tōng.kan	want
34	สำคัญ	săm.khan	importance

ลำดับที่	คำไทย	การออกเสียง	คำแปล
35	ตื่นนอน	tuen.non	get up
36	วุ่นวาย	wun.wai	crowded, busy
37	ชั้นล่าง	chan.lâng	basement
38	ตั้งใจ	tâng.chai	stay focused
39	เล็กน้อย	lek.nōi	little
40	น้ำร้อน	nám.rón	hot water
41	ของใช้	khöng.chai	daily stuff
42	เวลา	we.ta	time
43	หลายคน	lai.khon	many people
44	รู้เรื่อง	rú.ruéng	understand the story
45	มดแดง	mot.daeng	ant
46	น้ำตก	nam.tok	waterfall
47	ยศักดิ์	yot.sak	high position
48	ยุเหง่า	yu.yae	make one angry
49	ปวดหัว	puat.huá	headache
50	ร้องไห้	rong.hai	cry
51	น้ำส้ม	nam.som	orange juice
52	รองเท้า	rong.tao	shoes
53	จอดป้าย	chot.pai	stop at a bus stop
54	แน่นอน	nae.non	certainly
55	เคารพ	kao.rop	respect
56	ดับไฟ	dapfai	turn off the light
57	โดยมาก	doymak	mostly
58	ใช้เงิน	chai.ngoen	use the money
59	เจ้าของ	chao.khöng	owner
60	แนะนำ	nae.nam	suggestion
61	เก็บของ	kep.khöng	pack up
62	ไปด้วย	pai.duay	go together
63	งานบ้าน	ngan.ban	house chore
64	ชิงช้า	ching.cha	swing (noun)
65	นกแก้ว	nok.kwaēw	parrot
66	ม้านั่ง	ma.nâng	bench (noun)
67	พากพ้อง	phuak.phong	friends
68	ชกมวย	chok.muay	boxing

ลำดับที่	คำไทย	การออกเสียง	คำแปล
69	น้ำหวาน	nam.wan	nectar
70	ก่อนนอน	kon.non	before bed time
71	ของใคร	khong.khai	Whose stuff is this?
72	ตลาด	ta.lat	market
73	ไปนอน	pai.non	go to bed
74	ทุกเช้า	thuk.chao	every morning
75	ไม่ว่าง	mai.wang	busy
76	ถ้าหาก	lam.bak	struggle
77	เงินทอง	ngoen.thong	money and gold
78	นักเรียน	nak.rien	student
79	น่ารัก	na.rak	lovely, cute
80	เคราะห์ร้าย	khro.rai	bad luck
81	ฟ้าร้อง	fa.rong	thunder
82	หน้าต่าง	na.tang	window
83	ดอกไม้	dok.mai	flower
84	มากกว่า	mak.kwa	more than
85	เจ้าของ	chao.khong	owner
86	รอเพื่อน	ro.phuen	wait for friend
87	ดนตรี	don.tri	music
88	รอรถ	ro.rot	wait for the bus
89	ง่ายกว่า	ngai.kwa	easier
90	ภาษา	phasa	language
91	ต้นไม้	ton.mai	tree
92	ตำรวจ	tam.ruat	polic officer
93	ฟุตบอล	fut.bon	football
94	มะยม	ma.yom	star gooseberr
95	ต้อนรับ	ton.rap	welcome
96	ลูกรัก	luk.rak	my baby
97	จำเป็น	cham.pen	necessary
98	ทรงผม	sung.phom	hair style
99	บอกแล้ว	bok.laew	already told/informed
100	ไม่ควร	maikhuan	should not
101	นอนแล้ว	non.laew	went to bed

ลำดับที่	คำไทย	การออกเสียง	คำแปล
102	ตื่นนอน	tuen.non	get up
103	มะม่วง	ma.muāng	mango
104	พระพุทธ	phra.phut	buddha statue/image
105	นกเขา	nok.khǎo	dove
106	น้องชาย	nōng.chai	younger brother
107	ปิดไฟ	pit.fai	turn the light off
108	เดือดร้อน	duet.ron	in trouble
109	ฟังเพลง	fang.pheng	listen to the music
110	แม่ค้า	māe.kha	merchant
111	ฝนตก	fǒn.tok	rain
112	ซักผ้า	sak.phā	do laundry
113	ตื่นเต้น	tuen.tēn	excitement
114	กระต่าย	kra.tai	rabbit
115	น้ำชา	nam.cha	tea
116	หนักใจ	nak.chai	worry
117	ป้ายรถ	pāi.rot	bus stop

ภาคผนวก ค

รูปแบบการเก็บข้อมูลของอาสาสมัคร

ล่าดับที่ บุ: ชาย ใจา เพศ: ชาย หญิง อายุ ปี ครึ่งที่ 1 ครึ่งที่ 2

no.	Conditions						
	c1	c2	c3	c4	c5	c6	c7
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
จำนวนที่ ติดต่อ							