ระบบรู้จำทำนองเสียงพูดสำหรับเสียงพูดภาษาไทยโดยใช้โครงข่ายประสาทเทียม

นายปฐวี ชาญไววิทย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2546

ISBN 974-17-4902-3

THAI SPEECH INTONATION RECOGNITION USING ARTIFICIAL NEURAL NETWORK

Mr. Patavee Charnvivit

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Electrical Engineering

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

Academic Year 2003
ISBN 974-17-4902-3

หัวข้อวิทยานิพนธ์	ระบบรู้จำทำนองเสียงพูดสำหรับเสียงพูดภาษาไทยโดยใช้โครงข่ายประสาทเทียม
โดย	นาย ปฐวี ชาญไววิทย์
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล
อาจารย์ที่ปรึกษาร่วม	ดร.เสถียร เตรียมล้ำเลิศ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…… คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

…….. ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุดาพร ลักษณียนาวิน)

…….. อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล)

…….. อาจารย์ที่ปรึกษาร่วม
(ดร.เสถียร เตรียมล้ำเลิศ)

…….. กรรมการ
(อาจารย์สุวิทย์ นาคพีระยุทธ)

ปฐวี ชาญไววิทย์ : ระบบรู้จำทำนองเสียงพูดสำหรับเสียงพูดภาษาไทยโดยใช้โครงข่ายประสาทเทียม. (THAI SPEECH INTONATION RECOGNITION USING ARTIFICIAL NEURAL NETWORK) อาจารย์ที่ปรึกษา : รศ.ดร. สมชาย จิตะพันธ์กุล, อาจารย์ที่ปรึกษาร่วม : ดร. เสถียร เตรียมล้ำเลิศ 157 หน้า. ISBN 974-17-4902-3

ทำนองเสียงพูดภาษาไทยอาจจัดได้ว่าเป็นสารสนเทศกึ่งภาษาศาสตร์ที่เกิดขึ้นจากรูปลักษณ์ความถี่มูลฐาน ของประโยคเสียงพูด วิทยานิพนธ์นี้นำเสนอวิธีการในการรู้จำรูปแบบของทำนองเสียงพูดภาษาไทย โดยนำเสนอคอนทัวร์สำคัญสองลักษณะ ซึ่งหาได้จากลักษณะของความถี่มูลฐาน จากนั้นจึงนำคอนทัวร์ลักษณะทั้งสองประเภทนี้ไปแปลงเป็นเวกเตอร์ลักษณะ เพื่อนำไปใช้เป็นข้อมูลป้อนเข้าโครงข่ายประสาทเทียม ในแต่ละการทดลองจะฝึกฝน และทดสอบเสียงพูดของผู้ชาย และเสียงพูดของผู้หญิงแยกจากกัน การทดลองแรกจะจำแนกทำนองเสียงออกเป็น 3 ประเภท คือ ทำนองเสียงตก ทำนองเสียงขึ้น และทำนองเสียงผสม อัตราการรู้จำทำนองเสียงพูดมีค่าร้อยละ 61.6 สำหรับเสียงผู้ชาย และร้อยละ 73.7 สำหรับเสียงผู้หญิง เมื่อพิจารณาความผิดพลาดของการรู้จำเสียงพูดของแต่ละทำนองเสียงจากตารางความสับสนพบว่า ระบบรู้จำมีความสับสนระหว่างทำนองเสียงขึ้น และทำนองเสียงผสมสูง จึงได้ทำการทดลองที่สอง โดยจัดให้ทำนองเสียงผสมเป็นประเภทเดียวกับทำนองเสียงขึ้น ผลการทดลองพบว่าอัตราการรู้จำมีค่าเป็น ร้อยละ 81.7 สำหรับเสียงผู้ชาย และร้อยละ 90.8 สำหรับเสียงผู้หญิง

ภาควิชา วิศวกรรมไฟฟ้า	ลายมือชื่อนิสิต .....
สาขาวิชา วิศวกรรมไฟฟ้า	ลายมือชื่ออาจารย์ที่ปรึกษา .....
ปีการศึกษา 2546	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

4370376921 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: INTONATION RECOGNITION / THAI SPEECH / F_0 CONTOUR / FEATURE CONTOUR / NEURAL NETWORK

PATAVEE CHARNVIVIT : THAI SPEECH INTONATION RECOGNITION USING NEURAL NETWORK. THESIS ADVISOR : ASSO. PROF. SOMCHAI JITAPUNKUL, Ph.D., THESIS COADVISOR : SATIEN TRIAMLUMLERD, Ph.D. 157 pp. ISBN 974-17-4902-3.

Thai intonation can be categorized as paralinguistic information of F_0 contour of the utterance. This thesis presents a method of intonation pattern recognition of Thai utterance. Two intonation feature contours, extracted from F_0 contour, are proposed. The feature contours are converted to feature vector to be used as input of neural network recognizers. For each experiment, utterances from male and female speakers are trained and tested separately. In the first experiment, the utterances are devided into three classes of intonation pattern, the fall class, the rise class and the convolution class. The recognition rate of this experiment is 61.6% for male speakers and 73.7% for female speakers. The confusion matrices show that there is a lot of confusion between the rise class and the convolution class. So the second experiment is constructed, the number of classes of intonation is reduced to two classes. The utterances of the convolution class are re-labeled as the rise class. In the second experiment, the recognition rates are improved. The recognition rate is 81.7% for male speakers and 90.8% for female speakers.

Department Electrical Engineering Student's signature .....

Field of study Electrical Engineering Advisor's signature .....

Academic year 2003 Co-advisor's signature .....

กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ แนวคิด และวิธีการดำเนินการวิจัย รวมทั้งยังจัดหาอุปกรณ์ และแนะนำทุนวิจัยซึ่งทำให้ข้าพเจ้าได้มีโอกาสไปนำเสนอผลงานยังต่างประเทศ นับเป็นประสบการณ์ที่มีค่ายิ่งนัก และขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร. สุดาพร ลักษณียนาวิน และอาจารย์สุวิทย์ นาคพีระยุทธ ซึ่งได้กรุณาให้คำแนะนำ และคำปรึกษา อันมีค่าอย่างยิ่งต่องานวิจัย

ขอกราบขอบพระคุณ ดร. เสถียร เตรียมล้ำเลิศ ในฐานะอาจารย์ที่ปรึกษาร่วม ซึ่งได้กรุณาให้คำแนะนำต่าง ๆ ที่มีประโยชน์ต่องานวิจัย และขอกราบขอบพระคุณ ดร.รชฎ ทองประเสริฐ ในฐานะอาจารย์ที่ปรึกษาร่วมท่านแรก ซึ่งได้คอยเอาใจใส่ ให้คำปรึกษา และคำแนะนำต่าง ๆ เกี่ยวกับวิทยานิพนธ์อย่างมากมาย ข้าพเจ้ารู้สึกเสียใจอย่างสุดซึ้ง ต่อการจากไปของท่านก่อนเวลาอันควร และขอขอบคุณทุน TGIST ซึ่งได้ให้การสนับสนุนงานวิจัย

ขอกราบขอบพระคุณ ดร.ณัฐกร ทับทอง ที่กรุณาให้คำปรึกษา ให้คำแนะนำ และให้ความรู้เกี่ยวกับเทคโนโลยีด้านเสียงพูด และโครงข่ายประสาทเทียม ซึ่งมีคุณค่าอย่างยิ่งต่องานวิจัยนี้ และขอกราบขอบพระคุณ Prof. Dr. Hansjörg Mixdorff ซึ่งกรุณาให้คำแนะนำต่าง ๆ ในขณะที่ท่านได้มาทำงานวิจัยที่จุฬาลงกรณ์มหาวิทยาลัย การที่ข้าพเจ้าได้ร่วมทำงานวิจัยกับท่าน นับเป็นเกียรติอย่างยิ่ง และทำให้ข้าพเจ้าได้รับความรู้ และประสบการณ์อันมีค่าต่าง ๆ มากมาย

ขอขอบคุณ คุณเอกฤทธิ์ มณีน้อย และดร.วิศรุต อาขุบุตร ที่ได้ให้ความรู้เกี่ยวกับเทคโนโลยีการประมวลผลสัญญาณเสียงพูด และการเขียนซอฟต์แวร์ ซึ่งมีความสำคัญต่อข้าพเจ้ามาก และขอขอบคุณเพื่อน ๆ พี่ ๆ และน้อง ๆ ในห้องปฏิบัติการกรรมวิธีสัญญาณดิจิทัล ซึ่งคอยให้กำลังใจ และสร้างความสนุกสนานให้กับข้าพเจ้ามาตลอดระยะเวลาที่ทำงานอยู่ที่นี่

ขอขอบคุณคุณพิมพ์จนา ขวัญแพรแก้ว ที่คอยช่วยเหลืองานวิจัย เป็นกำลังใจ และคอยให้คำปรึกษาเสมอมา จนงานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี และขอขอบคุณผู้บันทึกเสียงทุกท่าน

สุดท้ายนี้ขอกราบขอบพระคุณคุณพ่อ คุณแม่ คุณตา คุณยาย และน้องของข้าพเจ้า ที่คอยเป็นกำลังใจ และให้การสนับสนุน ทั้งกำลังทรัพย์ และเวลา ทั้งยังคอยดูแล ห่วงใยข้าพเจ้ามาตลอด

สารบัญ

หน้า

หน้า

สารบัญตาราง

หน้า

หน้า

หน้า

สารบัญภาพ

หน้า

หน้า

หน้า

หน้า

หน้า

หน้า

หน้า

หน้า

หน้า

หน้า

หน้า

หน้า

คำจำกัดความที่ใช้ในการวิจัย

ครอสวาลิเดชันแบบ 5 โฟล	5-fold cross validation
คำสั่งการเน้นเสียง	accent command
ขนาดของคำสั่งเน้นเสียง	accent command amplitude (Aa)
ตัวกรองทางเสียง	acoustic filter
โครงข่ายประสาทเทียม	artificial neural network (ANN)
อัตสหสัมพันธ์	autocorrelation
ความถี่ฐาน	base frequency (Fb)
ความถี่ตัด	cutoff frequency (Fc)
การสกัด	extraction
คอนทัวร์ F_0	F_0 contour
คอนทัวร์ความแปรปรวนของ F_0	F_0 variation contour (FVC)
ลักษณะ	feature
คอนทัวร์ลักษณะ	feature contour
ตัวกรอง	filter
ผลต่างอันดับหนึ่ง	first difference (Δ)
แบบจำลองฟูจิซากิ	Fujisaki model
ความถี่มูลฐาน	fundamental frequency (F_0)
คาบมูลฐาน	fundamental period (T_0)
ช่องว่างระหว่างเส้นเสียง	glottis
คอนทัวร์ความถี่สูง	high frequency contour (HFC)
ทำนองเสียงพูด	intonation
สารสนเทศทางภาษาศาสตร์	linguistic information
คอนทัวร์ความถี่ต่ำ	low frequency contour (LFC)
สารสนเทศที่ไม่ใช่ภาษาศาสตร์	nonlinguistic information
สารสนเทศกึ่งภาษาศาสตร์	paralinguistic information
คำสั่งวลี	phrase command
ขนาดของคำสั่งวลี	phrase command magnitude (Ap)
เสียงซ้อน	prosody
กึ่งรายคาบ	quasi-periodic
สัญญาณรบกวนแบบสุ่ม	random noise
โครงสร้างการสั่นพ้อง	resonant structure

การสุ่มตัวอย่าง	sampling
การวิเคราะห์แบบช่วงเวลาสั้น	short-time analysis
การทดสอบ	testing
ทำนองเสียงผสม	the convolution class (intonation)
ทำนองเสียงตก	the fall class (intonation)
(วรรณยุกต์) เสียงโท	the falling tone
(วรรณยุกต์) เสียงตรี	the high tone
(วรรณยุกต์) เสียงเอก	the low tone
(วรรณยุกต์) เสียงสามัญ	the mid tone
ทำนองเสียงขึ้น	the rise class (intonation)
(วรรณยุกต์) เสียงจัตวา	the rising tone
เสียงวรรณยุกต์	tone
การฝึกฝน	training
เสียงไม่ก้อง	unvoiced sound
เสียงก้อง	voiced sound

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเทคโนโลยีทางด้านเสียงพูดได้เข้ามามีบทบาทต่อชีวิตประจำวันของเรามากขึ้นเรื่อย ๆ เทคโนโลยีทางด้านเสียงพูดสามารถแบ่งออกได้เป็นหลายประเภท เช่น การวิเคราะห์เสียงพูด (speech analysis) การสังเคราะห์เสียงพูด (speech synthesis) การรู้จำผู้พูด (speaker recognition) และการรู้จำเสียงพูด (speech recognition) สำหรับเทคโนโลยีด้านการรู้จำเสียงพูดก็ได้มีการนำมาปรับปรุงเพื่อให้สามารถประยุกต์ใช้งานในหลายรูปแบบ เช่น การสั่งงานและควบคุมคอมพิวเตอร์ (computer command and control) การพิมพ์ตามคำพูด (dictation) หรือการประยุกต์ใช้งานทางด้านโทรศัพท์ (telephony applications) (Huang, Acero, และ Hon, 2001) การประยุกต์ใช้งานกับโทรศัพท์มักจะเป็นลักษณะของการโต้ตอบทางเสียงพูดระหว่างระบบกับผู้ใช้งาน ซึ่งนอกจากระบบจะต้องรู้ว่าผู้พูดพูดคำว่าอะไรแล้วยังต้องมีความสามารถในการเข้าใจภาษาพูด (spoken language understanding) ด้วย และการที่ระบบจะเข้าใจความหมายของสิ่งที่ผู้พูดพูดจะต้องวิเคราะห์ปัจจัยหลาย ๆ ประการ เช่น โครงสร้างของบทสนทนา (dialog structure) รวมถึงไวยากรณ์ของภาษา ซึ่งก่อนจะวิเคราะห์ปัจจัยเหล่านี้ระบบจะต้องรู้ก่อนว่าผู้พูดพูดคำว่าอะไรบ้าง แต่ยังมีอีกปัจจัยหนึ่งที่ระบบสามารถนำมาช่วยในการทำความเข้าใจความหมายของการพูดได้ในเบื้องต้น โดยที่ยังไม่ต้องรู้ว่าผู้พูดพูดคำว่าอะไรบ้างนั่นคือทำนองเสียงพูด (intonation)

ในระยะเวลาไม่กี่ปีที่ผ่านมาความสำคัญของทำนองเสียงพูดได้เริ่มมีบทบาทกับศาสตร์ต่าง ๆ ที่นอกเหนือจากภาษาศาสตร์มากขึ้น ซึ่งรวมทั้งทางด้านวิศวกรรมทางเสียงพูด (speech engineering) เช่น การสังเคราะห์เสียงแบบอัตโนมัติ และการรู้จำเสียงพูด (Hirst และ Cristo, 1998) โดยมีงานของ Batliner และคนอื่น ๆ (2001) ได้กล่าวถึงการนำแบบจำลองของเสียงซ้อน (prosodic models) มาประยุกต์ใช้ในการเข้าใจเสียงพูดแบบอัตโนมัติ (automatic speech understanding : ASU) และการสังเคราะห์เสียง นอกจากนี้ยังมีงานของ Abdou และ Scordilid (2001) ซึ่งเป็นการปรับปรุงการเข้าใจเสียงพูดโดยนำตัวจำแนกประเภทของเสียงพูด (utterance type classifier) มาจำแนกระหว่างเสียงของประโยคบอกเล่าและประโยคคำถามโดยใช้ลักษณะ (features) ของความถี่มูลฐาน (fundamental frequency: F_0) และยังมีงานที่เกี่ยวข้องกับการจำแนกทำนองเสียงสำหรับภาษาอื่นที่ไม่ใช่ภาษาอังกฤษ เช่น Ishi, Minematsu, Nishide, และ Hirose (2001) ได้วิเคราะห์ลักษณะต่าง ๆ ของความถี่มูลฐาน เพื่อนำมาใช้จำแนกประเภทของทำนองเสียงในภาษาญี่ปุ่นเพื่อนำไปใช้ในระบบ CALL (Computer Aided Language Learning) ซึ่งช่วยสอนผู้เรียนเกี่ยวกับวิธีการออกเสียงเน้น (accent) และลักษณะของทำนองเสียงพูดในภาษาญี่ปุ่น รวมทั้งงานของ Nöth และคน

อื่น ๆ (2000) ได้กล่าวถึงโมดูลเสียงซ้อน (prosodic module) ในระบบ VERBMOBIL speech-to-speech translation system ซึ่งเป็นระบบแปลเสียงพูดจากภาษาเยอรมันเป็นเสียงพูดภาษาอังกฤษ และเป็นระบบเข้าใจเสียงพูดระบบแรกของโลกที่ประสบความสำเร็จในการนำลักษณะทางเสียงซ้อนของเสียงพูด (เช่น ความถี่มูลฐาน และความดังของเสียง) มาปรับปรุงสมรรถนะระบบ

การศึกษาเรื่องทำนองเสียงพูดเป็นเรื่องที่ยาก เนื่องจากทำนองเสียงพูดเป็นสิ่งที่มีลักษณะที่เป็นสากล นั่นคือลักษณะบางอย่างคล้ายกันมากในทุก ๆ ภาษา (most universal) ในขณะเดียวกันก็มีลักษณะบางอย่างที่ขึ้นกับแต่ละภาษาเป็นอย่างมากด้วย (most language specific) (Hirst และ Cristo, 1998) การออกแบบระบบรู้จำทำนองเสียงพูดภาษาไทยจึงอาจนำลักษณะของสัญญาณเสียงที่นิยมใช้ในภาษาอื่น ๆ มาใช้ได้ (ความถี่มูลฐาน พลังงาน และระยะเวลา) แต่ควรจะต้องนำมาผ่านการประมวลผลสัญญาณเบื้องต้นก่อน เนื่องจากโครงสร้างทางภาษาไม่เหมือนกัน ยกตัวอย่างเช่น ลักษณะของคอนทัวร์ความถี่มูลฐานในภาษาไทย ซึ่งนอกจากจะขึ้นกับทำนองเสียงพูดแล้วยังขึ้นกับชนิดของเสียงวรรณยุกต์ด้วย (Luksaneeyanawin, 1998) ดังนั้นการนำความถี่มูลฐาน มาใช้ในการรู้จำทำนองเสียงจึงควรกำจัดส่วนของรูปร่างของคอนทัวร์ความถี่มูลฐาน ที่ขึ้นกับชนิดของเสียงวรรณยุกต์ทิ้งไป เพื่อให้ตัวรู้จำสามารถรู้จำได้อย่างมีประสิทธิภาพ

งานวิจัยที่เกี่ยวข้องกับเทคโนโลยีทางด้านการรู้จำเสียงพูดภาษาไทยมีจำนวนมาก โดยเฉพาะงานวิจัยของห้องปฏิบัติการวิจัยกรรมวิธีสัญญาณดิจิทัล ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย เช่น การรู้จำเสียงพูดตัวเลขภาษาไทย (ระพีพัฒน์ เพ็ญศิริ, 2538; เสาวลักษณ์ อารีพงศา, 2538; วุฒิพงษ์ พรสุขจันทรา, 2539) การรู้จำเสียงสระโดด ๆ (ธีระ ภัทราพรนันท์, 2538) การรู้จำเสียงพยัญชนะเพื่อใช้ในการสะกดคำ (อุมาวสี ทาทอง, 2544) การรู้จำคำไทยหลายพยางค์ (วิศรุต อาบุบุตร, 2539 ; ชัย วุฒิวิวัฒน์ชัย, 2540) และกำลังก้าวไปสู่การรู้จำเสียงพูดภาษาไทยแบบคำพูดต่อเนื่อง เช่น การหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย (ณัฏฐชา จิตติวรางกูล, 2541) การรู้จำหน่วยเสียงสระภาษาไทย (เอกฤทธิ์ มณีน้อย, 2541) การศึกษาหน่วยเริ่มของพยางค์เชิงกลสัทศาสตร์ (วิศรุต อาบุบุตร, 2545) การศึกษาหน่วยตามของพยางค์เชิงกลสัทศาสตร์ (เอกฤทธิ์ มณีน้อย, 2547) ระบบรู้จำเสียงพูดภาษาไทยแบบคำพูดต่อเนื่องมีส่วนประกอบหนึ่งที่สำคัญ คือ ส่วนรู้จำเสียงวรรณยุกต์ของคำพูดต่อเนื่อง การประยุกต์ใช้งานระบบรู้จำทำนองเสียงพูดภาษาไทยนอกจากจะสามารถนำไปใช้ในระบบเข้าใจเสียงพูดภาษาไทยซึ่งเป็นส่วนประกอบหนึ่งของระบบรู้จำเสียงพูดแบบพูดโต้ตอบที่จะเกิดขึ้นในอนาคตแล้ว ยังสามารถนำมาใช้ปรับปรุงประสิทธิภาพของระบบรู้จำเสียงวรรณยุกต์ภาษาไทยแบบคำพูดต่อเนื่องได้อีกด้วย เนื่องจากคอนทัวร์ความถี่มูลฐานที่นำมาใช้รู้จำเสียงวรรณยุกต์ ขึ้นกับทั้งเสียงวรรณยุกต์ และทำนองเสียงตามที่ได้กล่าวไว้ข้างต้น แต่งานวิจัยการรู้จำเสียงวรรณยุกต์ภาษาไทยโดยทั่วไปมักจะคำนึงถึงแต่ผลของการลดระดับ (declination effect) ของความถี่มูลฐานในแต่ละประโยคเนื่องจากผลของทำนองเสียงตก ใน

ประโยคบอกเล่าเท่านั้น (Potisuk, Harper, และ Gandour, 1999; Thubthong, Kijsirikul, และ Luksaneeyanawin, 2001) ระบบรู้จำเสียงวรรณยุกต์แบบหลายทำนองเสียงจึงอาจนำลักษณะของระบบรู้จำทำนองเสียงมาช่วยในการจำแนกเสียงวรรณยุกต์ที่ทำนองเสียงต่าง ๆ ได้ หรืออาจนำระบบรู้จำทำนองเสียงมาใช้แยกประเภทของทำนองเสียงที่ผู้พูดพูด แล้วจึงส่งสัญญาณเสียงไปยังระบบรู้จำเสียงวรรณยุกต์ที่ได้รับการฝึกฝนสำหรับทำนองเสียงประเภทต่าง ๆ

วัตถุประสงค์ของการวิจัย

1. เพื่อออกแบบวิธีการหาค่าลักษณะจากเสียงพูด ที่เหมาะสมต่อการนำไปใช้ในการจำแนกประเภทของทำนองเสียงพูดของเสียงพูดภาษาไทย
2. เพื่อออกแบบระบบรู้จำทำนองเสียงพูดภาษาไทย โดยใช้ลักษณะที่ได้ออกแบบไว้ เพื่อทดสอบประสิทธิภาพของลักษณะ

เป้าหมาย และขอบเขตของการวิจัย

1. ได้กรรมวิธีในการหาลักษณะจากสัญญาณเสียง ที่เหมาะสมต่อการนำไปใช้ในการจำแนกประเภทของทำนองเสียงในภาษาไทย
2. อัตราการรู้จำทำนองเสียงพูดภาษาไทย ซึ่งใช้ลักษณะที่ได้ออกแบบไว้ โดยเฉลี่ยมีค่าสูงกว่าร้อยละ 80

ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ได้วิธีการหาค่าลักษณะจากเสียงพูด ที่เหมาะสมต่อการนำมาใช้จำแนกประเภทของทำนองเสียงพูดภาษาไทย
2. สามารถนำลักษณะต่าง ๆ ที่นำเสนอไปช่วยในการรู้จำเสียงวรรณยุกต์ภาษาไทยได้ (ลักษณะของสัญญาณเสียงที่นำมาใช้รู้จำเสียงวรรณยุกต์จะเปลี่ยนไปเมื่อพูดด้วยทำนองเสียงต่างกัน)
3. สามารถนำลักษณะที่ใช้ในการรู้จำทำนองเสียงพูดภาษาไทย ไปใช้กับระบบเข้าใจเสียงพูดภาษาไทยได้ในอนาคต

วิธีดำเนินการวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องกับ ทำนองเสียงพูดต่าง ๆ และงานวิจัยที่เกี่ยวข้องกับทำนองเสียงพูด ทั้งภาษาไทย และภาษาต่างประเทศ รวมทั้งวิธีการฝึกฝน และทดสอบอัตราการรู้จำโดยใช้โครงข่ายประสาทเทียม
2. เก็บรวบรวมข้อมูลเสียงพูด ที่จะนำมาใช้ในงานวิจัย
3. วิเคราะห์ลักษณะของเสียงพูดของทำนองเสียงประเภทต่าง ๆ พร้อมทั้งออกแบบวิธีการสกัดลักษณะ จากสัญญาณเสียงพูด ที่เหมาะสมต่อการนำไปใช้จำแนกประเภทของทำนองเสียงภาษาไทย
4. พัฒนาโปรแกรมที่ใช้ในการสกัดลักษณะจากเสียงพูด ตามวิธีการที่ได้ออกแบบไว้
5. พัฒนาระบบรู้จำทำนองเสียงพูดภาษาไทย
6. ทดสอบผลการรู้จำทำนองเสียงพูดภาษาไทย วิเคราะห์ผล และแก้ไขข้อบกพร่องของระบบที่พบ
7. สรุปรวบรวมผลการวิจัย พร้อมทั้งจัดทำเอกสารที่เกี่ยวข้องกับวิทยานิพนธ์

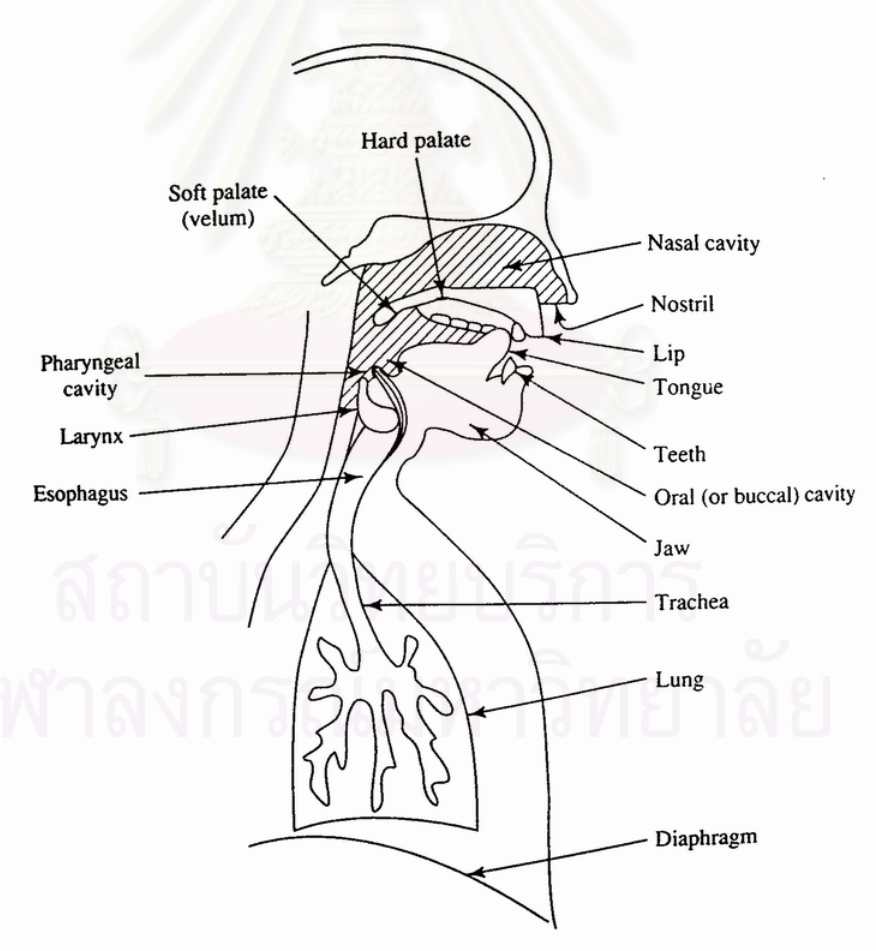
บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

บทนี้กล่าวถึงทฤษฎีที่สำคัญที่เกี่ยวข้องกับทำนองเสียงพูด โดยเริ่มจากหัวข้อ 2.1 กล่าวถึงกลไกการให้กำเนิดเสียงพูดของมนุษย์ และความหมายของความถี่มูลฐานของเสียงพูด ซึ่งเป็นลักษณะหลัก ของทำนองเสียงพูด จากนั้นในหัวข้อ 2.2 กล่าวถึงความหมายของทำนองเสียงพูด ลักษณะของทำนองเสียงพูดของบางภาษา และลักษณะของทำนองเสียงพูดภาษาไทย และในหัวข้อ 2.3 กล่าวถึงแบบจำลองฟูจิซากิ ซึ่งเป็นแบบจำลองทางคณิตศาสตร์แบบจำลองหนึ่งของทำนองเสียงพูด

2.1 เสียงพูด และความถี่มูลฐานของเสียงพูด

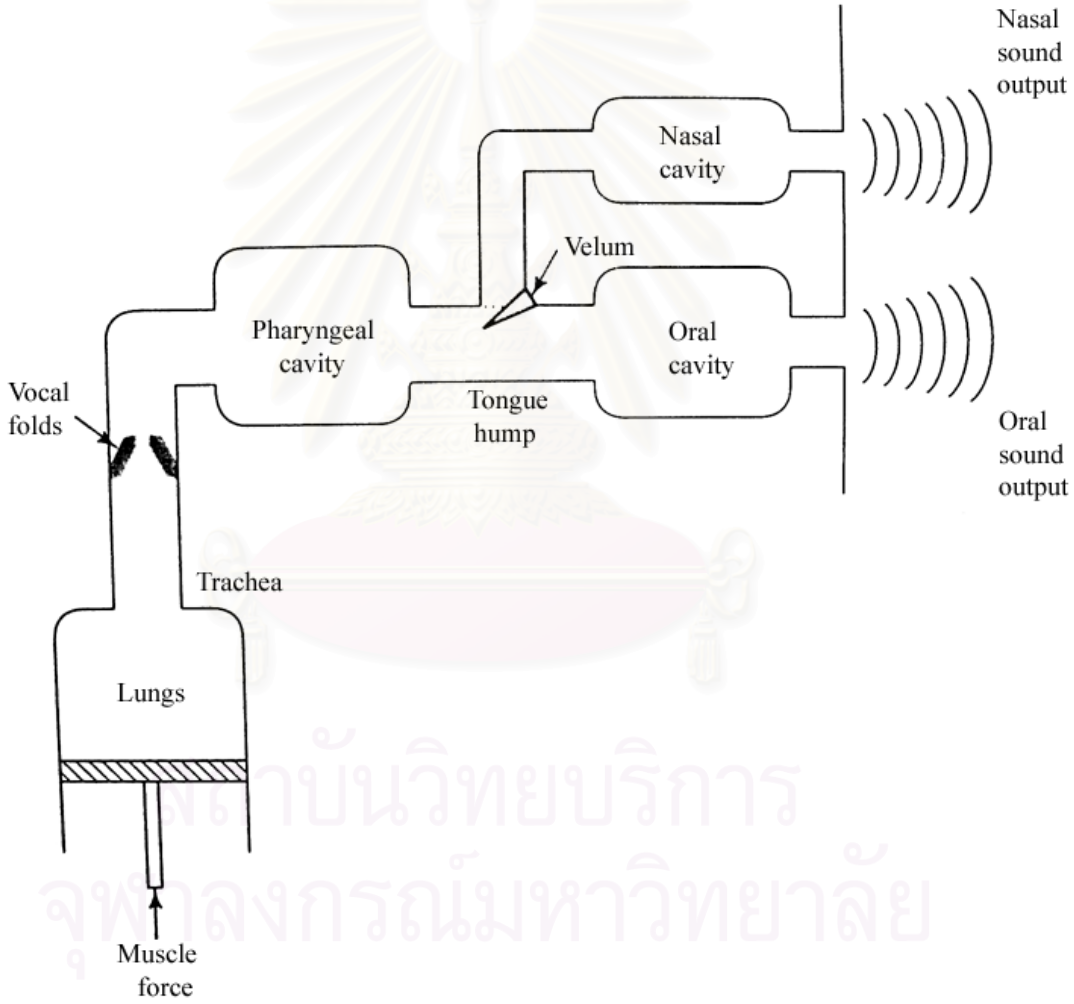
2.1.1 กลไกการกำเนิดเสียงพูด



รูปที่ 2.1 แผนภาพของกลไกการกำเนิดเสียงพูดของมนุษย์

(Deller, Proakis และ Hansen, 1993: 102)

รูปคลื่นเสียงพูดเป็นคลื่นของความดันอากาศ ที่เกิดจากการเคลื่อนไหวของโครงสร้างทางกายวิภาค ก่อให้เกิดกลไกการกำเนิดเสียงพูดของมนุษย์ รูปที่ 2.1 แสดงอวัยวะที่ทำให้เกิดเสียงพูด ซึ่งประกอบด้วยส่วนประกอบหลัก ได้แก่ ปอด (lungs) หลอดลม (trachea) กล่องเสียง (larynx) คอหอย หรือช่องคอ (pharyngeal cavity) ช่องปาก (oral หรือ buccal cavity) และ ช่องจมูก (nasal cavity) โดยมักจะเรียกช่องคอ และช่องปากว่า ช่องทางเดินเสียง (vocal tract) และเรียกช่องจมูกว่า ช่องทางเดินจมูก (nasal tract) นอกจากอวัยวะหลัก ๆ เหล่านี้ยังมีอวัยวะอื่น ๆ ซึ่งเป็นส่วนประกอบสำคัญที่ทำให้เกิดเสียงพูดอีก ได้แก่ เส้นเสียง (vocal folds หรือ vocal cords) ซึ่งอยู่ในกล่องเสียง เพดานอ่อน (soft palate หรือ velum) ลิ้น (tongue) ฟัน (teeth) และ ริมฝีปาก (lips) (Deller และคนอื่น ๆ , 1993: 101)

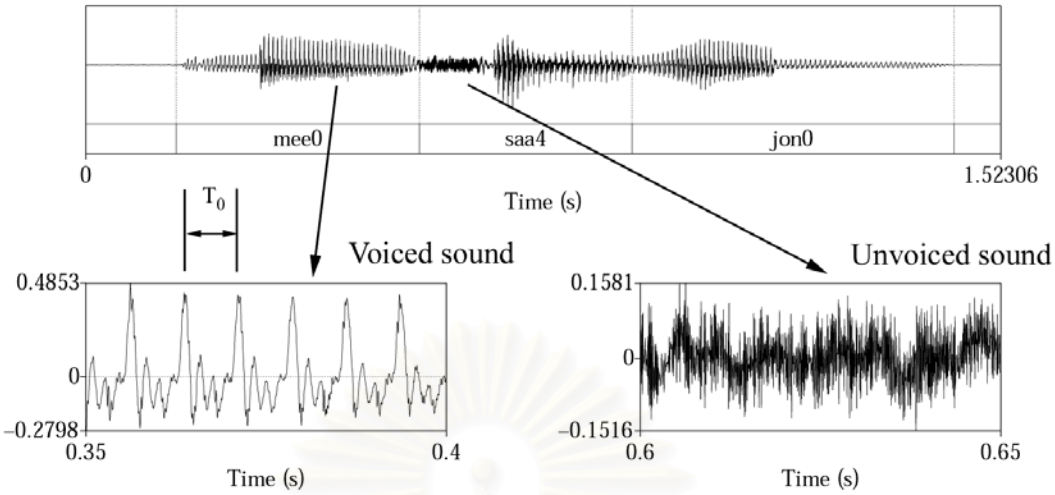


รูปที่ 2.2 แบบจำลองของกลไกการกำเนิดเสียงพูดของมนุษย์
(Deller และคนอื่น ๆ, 1993: 103)

เราสามารถแบ่งชนิดของเสียงพูดออกเป็นประเภทใหญ่ ๆ ได้ 2 ประเภท คือเสียงก้อง (voiced sound) และ เสียงไม่ก้อง (unvoiced sound)

เสียงก้อง เกิดจากการที่อากาศถูกบังคับให้ไหลผ่านช่องว่างระหว่างเส้นเสียง (glottis) และเกิดการสั่นในลักษณะของกึ่งรายคาบ (quasi-periodic) เนื่องจากความตึงของเส้นเสียง ซึ่งจะกล่าวถึงอย่างละเอียดในหัวข้อ 2.1.2 เสียงที่เกิดจากการสั่นของอากาศนี้จะผ่านขึ้นมายังช่องคอ และออกสู่ภายนอกโดยผ่านช่องปาก และ/หรือ ช่องจมูก ซึ่งถ้ามองโดยมุมมองของวิศวกร สามารถแสดงแบบจำลองกลไกการกำเนิดเสียงพูดได้ดังรูปที่ 2.2 โดยช่องว่างทั้งสามช่องนี้สามารถพิจารณาได้ว่าเป็นโครงสร้างของการสั่นพ้อง (resonant structure) ของเสียงพูดของมนุษย์ ซึ่งเปรียบเหมือนตัวกรองทางเสียง (acoustic filter) การขยับอวัยวะต่าง ๆ เช่น ลิ้น ริมฝีปากหรือ เพดานอ่อน เป็นการเปลี่ยนแปลงรูปร่างของช่องว่างเหล่านี้ จึงเปรียบเหมือนการเปลี่ยนแปลงคุณสมบัติของตัวกรอง ซึ่งทำให้สัญญาณเสียงที่ได้เปลี่ยนไป (Deller และคนอื่น ๆ, 1993: 102) เช่น เสียงสระแบบต่าง ๆ หรือเสียงพยัญชนะบางชนิด เช่น เสียง บ ม ย หรือ น ในภาษาไทย รูปคลื่นของเสียงพูดในส่วนที่เป็นเสียงก้องมีลักษณะเป็นแบบกึ่งรายคาบ ดังแสดงตัวอย่างในรูปที่ 2.3 ในส่วนที่เป็นเสียงสระเอของคำว่า “เมษายน”

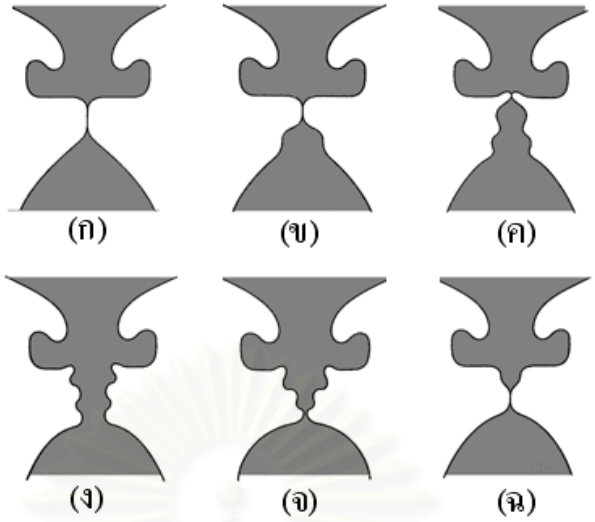
เสียงไม่ก้อง เกิดจากการที่ตำแหน่งบางตำแหน่งในช่องทางเดินเสียงถูกทำให้แคบลง เส้นเสียงจะเปิดออก (แต่ไม่กว้างเท่าเวลาหายใจ) ทำให้อากาศไหลผ่านออกมาได้สะดวกโดยที่เส้นเสียงไม่สั่น เมื่ออากาศไหลผ่านช่องทางเดินเสียงในส่วนที่ถูกทำให้แคบลง อากาศจะเกิดการปั่นป่วน (turbulence) ทำให้เกิดเสียงบางชนิด เช่น เสียงพยัญชนะ พ ท ป ส ในภาษาไทย เป็นต้น (กาญจนา นาคสกุล, 2541) รูปคลื่นของเสียงพูดในส่วนที่เป็นเสียงไม่ก้องมีลักษณะคล้ายสัญญาณรบกวนแบบสุ่ม (random noise) ดังแสดงตัวอย่างในรูปที่ 2.3 ในส่วนที่เป็นเสียงพยัญชนะ ‘ษ’ ของคำว่า “เมษายน”



รูปที่ 2.3 รูปคลื่นของเสียงพูดของคำว่า เมษายน (mee0saa4jon0) ขยายให้เห็นถึงส่วนที่เป็นเสียงก้อง ในส่วนที่เป็นเสียงสระเอของพยางค์ เม (mee0) และเสียงไม่ก้องในส่วนที่เป็นพยัญชนะ ‘ษ’ ของพยางค์ ษา (saa4)

2.1.2 ความถี่มูลฐานของเสียงพูด

เมื่อออกเสียงที่เป็นเสียงก้อง แรงจากกล้ามเนื้อช่องท้อง (abdominal muscles) จะทำให้กระบังลม (diaphragm) ยกตัวขึ้น ส่งผลให้อากาศออกจากปอดเข้าไปยังหลอดลม และเคลื่อนที่ขึ้นไปยังช่องว่างระหว่างเส้นเสียง ที่ช่องว่างระหว่างเส้นเสียงนี้อากาศจะถูกขัดจังหวะในลักษณะที่เป็นรายคาบเนื่องจากการเคลื่อนไหวของเส้นเสียง การเปิด-ปิดแบบซ้ำไปซ้ำมาของช่องว่างระหว่างเส้นเสียงนี้ขึ้นอยู่กับความดันอากาศจากหลอดลม ในขั้นแรกความดันของอากาศที่อยู่ใต้เส้นเสียงจะสูงขึ้นทำให้เส้นเสียงแยกตัวออกจากกัน (รูปที่ 2.4 (ก), (ข)) ช่องว่างระหว่างเส้นเสียงจะเริ่มเปิดออก (รูปที่ 2.4 (ค)) อากาศจะเริ่มไหลออกจากหลอดลมผ่านช่องว่างนี้ (รูปที่ 2.4 (ง)) โดยความดันอากาศจะทำให้ช่องว่างเปิดกว้างขึ้นเรื่อย ๆ ทำให้กระแสอากาศไหลออกมากขึ้น จนความดันอากาศเริ่มลดลง ช่องว่างระหว่างเส้นเสียงจะยังคงเปิดอยู่ จนกระทั่งแรงอัดเนื่องจากความยืดหยุ่นของเส้นเสียง เท่ากับแรงแยกเส้นเสียงเนื่องจากความดันอากาศ ที่ตำแหน่งนี้อากาศจะมีความเร็วสูงสุด พลังงานจลน์เนื่องจากการเคลื่อนไหวของอากาศได้เปลี่ยนไปเป็นพลังงานศักย์ยืดหยุ่นซึ่งถูกเก็บไว้ในเส้นเสียงซึ่งทำให้เส้นเสียงเริ่มจะปิดลงอีกครั้ง (รูปที่ 2.4 (จ)) ทั้งแรงที่เกิดจากการยืดหยุ่นของเส้นเสียง และแรงเบอร์นูลี (Bernoulli force) ซึ่งเกิดจากการไหลของอากาศ ดึงให้เส้นเสียงปิดลงอย่างรวดเร็ว (รูปที่ 2.4 (ฉ)) (Deller และคนอื่น ๆ, 1993: 111)



รูปที่ 2.4 ลำดับของภาพตัดขวางของกล่องเสียง แสดงให้เห็นถึงวัฏจักรการเปล่งเสียงจนครบ 1 คาบ ส่วนที่แรเงาคือช่องว่างระหว่างเส้นเสียง (ดัดแปลงจาก Vennard, 1967 อ้างถึงใน Deller และคนอื่น ๆ, 1993: 111)

ความดันอากาศด้านล่างเส้นเสียง และแรงตึงของเส้นเสียงจะทำให้เกิดการเปิด-ปิดของเส้นเสียงในลักษณะเดิมอีกครั้งซ้ำไปซ้ำมา ทำให้เกิดสัญญาณเสียงที่มีลักษณะเป็นกึ่งรายคาบดังเช่นเสียงสระเอในรูปที่ 2.3 ระยะเวลาที่เส้นเสียงเปิด-ปิด 1 รอบนี้เรียกว่า คาบมูลฐาน (fundamental period: T_0) ส่วนอัตราการเปิดปิดของเส้นเสียงเรียกว่า ความถี่มูลฐาน (fundamental frequency: F_0) หรือพิทช์ (pitch) (Deller และคนอื่น ๆ, 1993: 112)

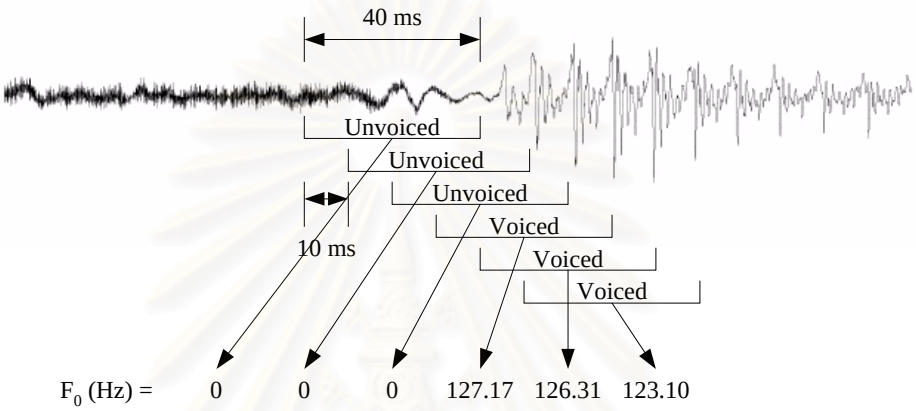
โดยจะได้ว่า

$$F_0 = \frac{1}{T_0} \tag{2.1}$$

ช่วงของค่าความถี่มูลฐานของเสียงพูดของแต่ละคนจะแตกต่างกัน โดยขึ้นกับลักษณะทางกายภาพของกล่องเสียงของคน ๆ นั้น โดยทั่วไปค่าความถี่มูลฐานของเสียงพูดของผู้ชายมีค่าระหว่าง 50 – 250 Hz ส่วนเสียงพูดของผู้หญิงจะมีค่าระหว่าง 120 – 500 Hz (Deller และคนอื่น ๆ, 1993: 114) ความถี่มูลฐานของเสียงพูดเป็นลักษณะของสัญญาณเสียง ที่บอกให้รู้ว่าเสียงนั้นเป็นเสียงสูง หรือเสียงต่ำ เสียงพูดของคน ๆ หนึ่ง ที่พูดประโยค หนึ่ง ๆ ไม่ได้มีค่าคงที่ตลอดทั้งประโยค แต่จะมีลักษณะขึ้น ๆ ลง ๆ ทำให้เสียงที่ได้มีลักษณะสูง ๆ ต่ำ ๆ ซึ่งขึ้นอยู่กับลักษณะของทำนองเสียงพูดของประโยคนั้น ในการศึกษาลักษณะของทำนองเสียงพูดแบบต่าง ๆ จึงนิยมใช้ F_0 เป็นพารามิเตอร์หลัก (Hirst และ Cristo, 1998)

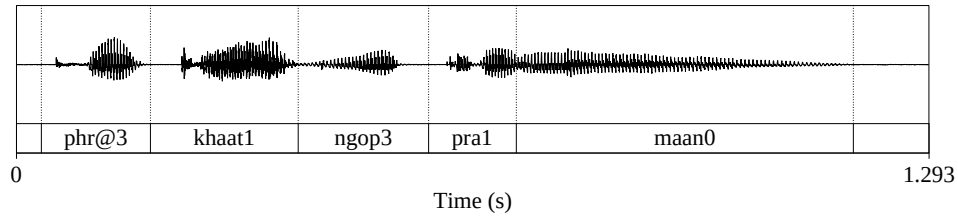
2.1.3 คอนทัวร์ F_0

วิธีการหาค่า F_0 มีลักษณะเหมือนกับการหาค่าลักษณะ (feature) อื่น ๆ ของสัญญาณเสียงพูด คือใช้วิธีการวิเคราะห์แบบช่วงเวลาสั้น (short-time analysis) โดยการแบ่งสัญญาณเสียงพูดออกเป็นอนุกรมของช่วงสั้น ๆ ตามแกนเวลา โดยเรียกแต่ละส่วนว่าเฟรม (frame) จากนั้นจึงพิจารณาว่าเสียงในเฟรมนั้นเป็นเสียงก้องหรือไม่ ถ้าเป็นเสียงก้องจึงหาค่า F_0 ของเฟรมนั้น แต่ถ้าไม่ใช่เสียงก้องให้กำหนดค่า F_0 ของเฟรมนั้นเป็น 0 Hz หรือไม่ทำการนิยามไว้ก็ได้ (Rabiner และ Schafer, 1978)

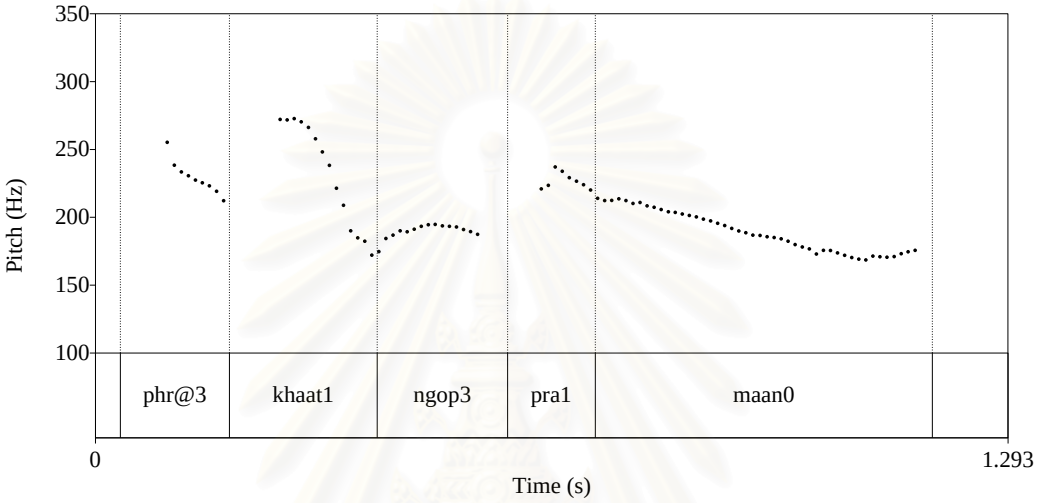


รูปที่ 2.5 การแบ่งสัญญาณเสียงพูดออกเป็นเฟรม เพื่อหา F_0 ในกรณีที่หาโดยใช้โปรแกรม Praat เสียงแต่ละเฟรมจะกว้าง 40 ms โดยจะเลื่อนตำแหน่งของเฟรมไปทีละ 10 ms

วิธีการหาคอนทัวร์ F_0 ของแต่ละเฟรมนั้นมีหลายวิธี เช่น วิธีอัตสหสัมพันธ์ (autocorrelation) วิธีสหสัมพันธ์ไขว้ (cross-correlation) วิธีฟังก์ชันผลต่างขนาดเฉลี่ย (average magnitude difference function: AMDF) และวิธีเซปสตรัม (cepstrum) ตัวอย่างของคอนทัวร์ F_0 ของเสียงพูดแสดงในรูปที่ 2.6 ซึ่งจะเห็นได้ว่าคอนทัวร์ F_0 ขาดตอน ในช่วงที่เสียงเป็นเสียงไม่ก้อง ได้แก่ เสียงพยัญชนะ “พ” ของคำว่า “เพราะ” (/phr@3/) เสียงพยัญชนะ “ข” ของคำว่า “ขาด” (/khaat1/) และเสียงพยัญชนะ “ป” ของคำว่า “ประมาณ” (/pra1/maan0/) ส่วนในช่วงที่เสียงเป็นเสียงก้อง จะเห็นว่าคอนทัวร์ F_0 มีรูปร่างต่าง ๆ กันในแต่ละพยางค์ และมีแนวโน้มที่จะลดระดับลงจากต้นวลีถึงท้ายวลี เนื่องจากผลของเสียงวรรณยุกต์ในแต่ละพยางค์ และทำนองเสียงของวลี ซึ่งจะกล่าวถึงโดยละเอียดในหัวข้อถัดไป



(ก)



(ข)

รูปที่ 2.6 (ก) รูปคลื่นเสียงพูดของวลี “เพราะขาดงบประมาณ” (/phr@3/khaat1/ngop3/pra1/maan0/)
(ข) คอนทัวร์ F_0 ของรูปคลื่นเสียงในข้อ (ก)

2.2 ทำนองเสียงพูด

2.2.1 สารสนเทศที่ได้จากเสียงพูด

เมื่อพิจารณาโดยผิวเผินแล้ว อาจจะเห็นว่า สารสนเทศที่ได้จากเสียงพูด มีเพียงสารสนเทศที่บอกว่าผู้พูด พูดคำว่าอะไรออกมาบ้าง แต่ในความเป็นจริงแล้วเสียงพูดของมนุษย์ได้ให้สารสนเทศมากมายนอกเหนือไปจากคำต่าง ๆ ที่พูดออกมา เช่น ถ้าเราฟังน้ำเสียงของผู้พูด ในบางกรณีอาจสามารถแยกแยะได้ว่าผู้พูดต้องการพูดเพื่อเล่าเรื่องให้ฟัง ต้องการสั่ง ต้องการถาม ต้องการอ้อนวอน หรือต้องการแสดงความสงสัย ในบางครั้งเราสามารถบอกได้ด้วยว่าผู้พูดมีอารมณ์เช่นใด และนอกจากนี้เรามักจะสามารถบอกได้ว่าผู้พูดเป็นผู้ชาย หรือผู้หญิง เป็นเด็ก วัยกลางคน หรือคนแก่ได้อีกด้วย สิ่งเหล่านี้ถือเป็นสารสนเทศที่ได้รับจากเสียงพูดทั้งสิ้น Fujisaki (1996b) ได้อธิบายว่าสารสนเทศที่ได้จากเสียงพูดสามารถแบ่งได้เป็น 3 ประเภทหลัก คือ สารสนเทศทางภาษาศาสตร์ สารสนเทศกึ่งภาษาศาสตร์ และ สารสนเทศที่ไม่ใช่ภาษาศาสตร์ โดยได้ให้คำจำกัดความของสารสนเทศแต่ละประเภทดังนี้

- *สารสนเทศทางภาษาศาสตร์ (linguistic information)* เป็นสารสนเทศทางสัญลักษณ์ที่สามารถแสดงให้เห็นได้ในลักษณะของเซตของสัญลักษณ์แบบดีสครีต (discrete symbol) และกฎของการนำสัญลักษณ์ต่าง ๆ เหล่านี้มาเรียงกัน สารสนเทศประเภทนี้สามารถแสดงให้เห็นอย่างชัดเจนทางภาษาเขียน หรือการสรุปใจความสำคัญจากเนื้อหา สารสนเทศทางภาษาศาสตร์ตามนิยามนี้จึงเป็นสารสนเทศ แบบดีสครีตและสามารถแยกประเภทได้ (categorical) ตัวอย่างของสารสนเทศประเภทนี้ เช่น ประเภทของการเน้นเสียงของคำในภาษาญี่ปุ่น ซึ่งเป็นข้อมูลที่มีลักษณะดีสครีต เนื่องจากสามารถเจาะจงได้ว่าเป็นเสียงวรรณยุกต์ใด หรือการเน้นเสียงประเภทไหนจากประเภทต่าง ๆ ที่เป็นไปได้ ซึ่งมีจำนวนจำกัด

- *สารสนเทศกึ่งภาษาศาสตร์ (paralinguistic information)* เป็นสารสนเทศที่ไม่สามารถสรุปความได้จากข้อความที่เขียน แต่เป็นสิ่งที่ผู้พูดต้องการที่จะเพิ่มเข้าไปเพื่อเปลี่ยนแปลง หรือเพื่อเสริมสารสนเทศทางภาษาศาสตร์ ดังจะเห็นได้ว่าข้อความที่เขียนขึ้นสามารถอ่านออกเสียงได้หลายรูปแบบเพื่อแสดงเจตนา ทัศนคติ และรูปแบบการพูด (speaking style) ต่าง ๆ ซึ่งผู้พูดเป็นผู้ควบคุม สารสนเทศกึ่งภาษาศาสตร์สามารถเป็นได้ทั้งสารสนเทศแบบดีสครีตและสารสนเทศแบบต่อเนื่อง (continuous) ตัวอย่างเช่น สารสนเทศที่บอกให้เราทราบว่าผู้พูดต้องการยืนยัน หรือผู้พูดต้องการถาม เป็นสารสนเทศที่มีลักษณะดีสครีต แต่จะเป็นสารสนเทศแบบต่อเนื่องเมื่อเราต้องการทราบว่าผู้พูดต้องการแสดงความรู้สึกแบบต่าง ๆ ด้วยระดับขั้น (degree) มากน้อยเพียงใด

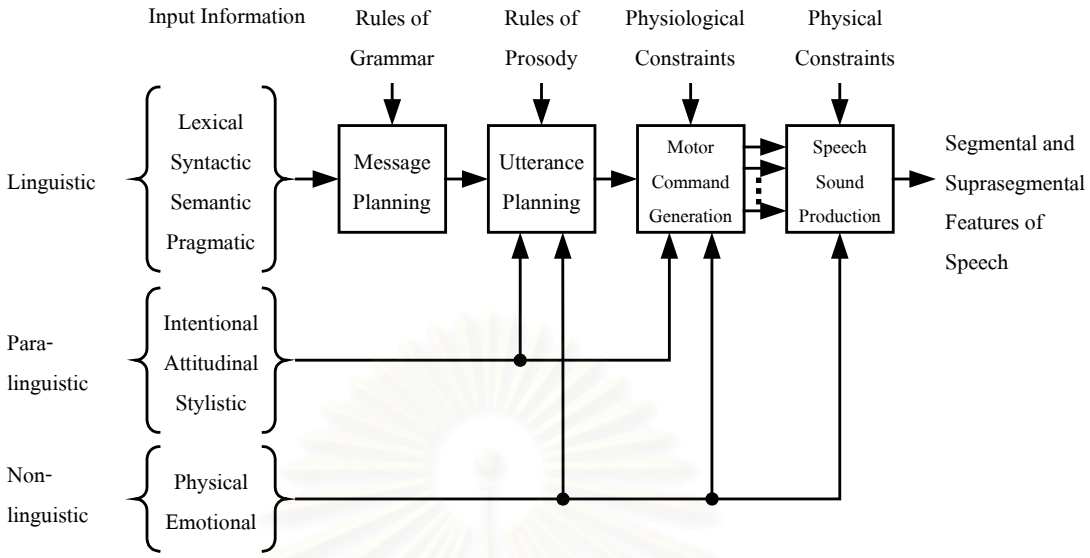
- *สารสนเทศที่ไม่ใช่ภาษาศาสตร์ (nonlinguistic information)* เป็นสารสนเทศที่เกี่ยวข้องกับปัจจัยต่าง ๆ เช่น อายุ เพศ นิสัย ลักษณะทางกายภาพ และทางอารมณ์ของผู้พูด ปัจจัยต่าง ๆ เหล่านี้ไม่ได้เกี่ยวข้องโดยตรงต่อสารสนเทศทางภาษาศาสตร์ และสารสนเทศกึ่งภาษาศาสตร์ของผู้พูด นอกจากนี้โดยทั่วไปแล้วผู้พูดจะไม่สามารถควบคุมสารสนเทศประเภทนี้ได้โดยตรง (ยกเว้นในกรณีของนักแสดงซึ่งสามารถแสดงออกด้วยอารมณ์แบบต่าง ๆ ได้) สารสนเทศประเภทนี้เหมือนกับสารสนเทศกึ่งภาษาศาสตร์ตรงที่สามารถเป็นได้ทั้งสารสนเทศแบบดีสครีต และสารสนเทศแบบต่อเนื่อง นั่นคือเราสามารถจัดระดับความมากน้อยให้กับทั้งสารสนเทศกึ่งภาษาศาสตร์ และสารสนเทศที่ไม่ใช่ภาษาศาสตร์ประเภทต่าง ๆ ได้ ซึ่งจะแตกต่างจากสารสนเทศภาษาศาสตร์ ซึ่งมีลักษณะเป็นดีสครีตโดยธรรมชาติ

ในกรณีของสารสนเทศจากคอนทัวร์ F_0 ของเสียงพูด Fujisaki (1992) ที่มีการอ้างถึงใน Mixdorff (2002: 16 - 17) ได้พิจารณาธรรมชาติของกระบวนการในการประมวลผลสารสนเทศทั้ง 3 ประเภท ของผู้พูดจนทำให้ได้คอนทัวร์ F_0 ของเสียงพูดที่มีรูปร่างแบบต่าง ๆ โดยได้เสนอแบบจำลองการประมวลผลสารสนเทศของผู้พูดดังแสดงในรูปที่ 2.7 ซึ่งเป็นการนำเอาสารสนเทศทั้ง 3 แบบไปเข้ารหัสโดยผ่านขั้นตอนต่าง ๆ ที่ซับซ้อน

จากรูปที่ 2.7 เมื่อผู้พูดต้องการพูดประโยคหนึ่ง ๆ สารสนเทศในระดับสูง (สารสนเทศขาเข้า) จะถูกนำไปเข้ารหัสจนกลายเป็นหน่วยแบบนามธรรม (abstract unit) และโครงสร้างทางภาษา ขั้นตอนนี้เรียกว่า การออกแบบข้อความ (message planning) โดยมีกฎทางไวยากรณ์ (rules of grammar) เป็นตัวกำหนดแนวทางในการออกแบบข้อความ หลังจากนั้นจะเข้าสู่ส่วนการออกแบบประโยคพูด (utterance planning) ซึ่งจะเป็นการแบ่งวลี (phrasing) การกำหนดลักษณะของการเน้นเสียงในพยางค์ต่าง ๆ (accentuation) และการกำหนดการหยุดเว้นวรรค (pausing) ขั้นตอนนี้ถือเป็นขั้นแรกที่สารสนเทศกึ่งภาษาศาสตร์ และสารสนเทศที่ไม่ใช่ภาษาศาสตร์เข้ามามีส่วนในกระบวนการผลิตเสียงพูด ซึ่งจะเป็นตัวกำหนดรูปแบบการพูด และการแบ่งประโยค ในลักษณะต่าง ๆ

หลังจากผ่านกระบวนการออกแบบประโยคพูดจะทำให้เกิดคำสั่งทางประสาทที่ทำหน้าที่ควบคุมการเคลื่อนไหว (neuro-motor commands) ซึ่งจะไปควบคุมอวัยวะต่าง ๆ ที่เป็นกลไกในการกำเนิดเสียงพูด เนื้อหาในสารสนเทศจะถูกเปลี่ยนไปเป็นสัญญาณเสียงพูด โดยจะมีข้อจำกัดทางด้านสรีระ (physiological) และด้านกายภาพ (physical) ของผู้พูดเป็นตัวกำหนดส่วนสร้างคำสั่งทางการเคลื่อนไหว (motor command generation) และส่วนให้กำเนิดเสียงพูด (speech sound production)

แบบจำลองนี้ได้แสดงให้เห็นถึงความยากในการหาความสัมพันธ์ที่ชัดเจน (clear) และเป็นหนึ่งเดียว (unique) ระหว่างลักษณะที่สังเกตได้จากสัญญาณเสียงพูด (ข้อมูลขาออก) กับการจัดโครงสร้างของประโยคของผู้พูด (ข้อมูลขาเข้า) ซึ่งเป็นกระบวนการย้อนกลับของรูปที่ 2.7



รูปที่ 2.7 กระบวนการของผู้พูดเพื่อเปลี่ยนสารสนเทศชนิดต่าง ๆ ไปเป็นลักษณะของเสียงพูด (Fujisaki, 1995)

2.2.2 ความหมายของทำนองเสียงพูด

มีผู้ให้ความหมายของทำนองเสียงพูดไว้มากมาย ตัวอย่างเช่น Botinis, Granström และ Möbius (2001) ได้นิยามความหมายของทำนองเสียงพูดไว้ว่า ทำนองเสียงพูด หมายถึง การรวมกันของลักษณะทางความสูงต่ำของเสียง (tonal feature) จนกลายเป็นหน่วยทางโครงสร้างที่ใหญ่ขึ้น ซึ่งมีความสัมพันธ์กับพารามิเตอร์ทางเสียงของความถี่มูลฐานของเสียงพูด (acoustic parameter of voice fundamental frequency) และลักษณะเฉพาะของการเปลี่ยนแปลงค่าของ F_0 ในกระบวนการทางเสียงพูด

ในบทความที่เกี่ยวกับทำนองเสียงพูดโดยทั่วไปมักจะมีความสับสนระหว่างทำนองเสียงพูด (intonation) และเสียงซ้อน (prosody) แต่ส่วนมากนิยมใช้คำว่าทำนองเสียงพูดเมื่อกล่าวถึงลักษณะทางความสูงต่ำของเสียง (F_0) เท่านั้น ในขณะที่คำว่าเสียงซ้อนมักจะใช้เมื่อหมายถึงลักษณะทางความสูงต่ำของเสียงรวมถึง ระยะเวลา (duration) และระดับความดันของเสียง (sound pressure level) ด้วย นอกจากนี้ถ้ามองในความหมายกว้าง ๆ ทำนองเสียงพูดจะรวมทั้ง ระดับเสียงสูงต่ำแบบเฉพาะที่ (local tonal distribution) และระดับเสียงสูงต่ำแบบกว้าง (global tonal distribution) แต่ถ้ามองในความหมายแคบ ๆ จะใช้คำว่าทำนองเสียงสูงต่ำ (intonation proper) แทนระดับสูงต่ำของทั้งประโยค โดยจะนำลักษณะของเสียงสูงต่ำของเสียงที่ใช้จำแนกความหมายของคำ (lexical tonal features) ไปเป็นลักษณะหนึ่งของเสียงซ้อน (prosody)

นอกจากนี้ในกรณีของภาษาไทย กาญจนา นาคสกุล (2541) ได้ให้ความหมายของทำนองเสียงพูดว่า หมายถึงระดับความสูงต่ำของเสียง ที่ปรากฏเป็นลักษณะของประโยคทั้งประโยค ทำนองเสียงไม่ใช่หน่วยเสียงซึ่งจะใช้เปลี่ยนความหมายของคำ แต่เป็นสิ่งสำคัญที่จะแสดงความหมายของประโยค ว่าเป็นประโยคบอกเล่า ประโยคคำสั่ง ประโยคคำถาม ฯลฯ การเปลี่ยนทำนองเสียงจะไม่ทำให้ความหมายของคำต่างไป แต่จะทำให้ความหมายของประโยคเปลี่ยนไป

2.2.3 ทำนองเสียงพูดของแต่ละภาษา

ทำนองเสียงพูดมีทั้งลักษณะที่เป็นสากล (universal) และขึ้นกับแต่ละภาษา (language specific) ยกตัวอย่างของความเป็นสากลของทำนองเสียงพูด ได้แก่ ภาษาโดยส่วนมาก ลักษณะเสียงพูดมีค่าพิทช์ (F_0) ที่สูงขึ้น มักจะนำไปใช้ในการทำให้เสียงพูดแตกต่างจากเสียงพูดธรรมดาที่มีค่าพิทช์ต่ำ เพื่อแสดงให้เห็นว่าผู้พูดต้องการถาม (Hirst และ Cristo, 1998)

อย่างไรก็ตามทำนองเสียงพูด ก็มีลักษณะเฉพาะที่ขึ้นกับแต่ละภาษา ด้วยเช่นกัน ตัวอย่างเช่น Hirst และ Cristo (1998) ได้รวบรวมบทความที่กล่าวถึงลักษณะของทำนองเสียงของภาษาต่าง ๆ 20 ภาษา บทความแต่ละบทความเขียนโดยผู้เชี่ยวชาญด้านทำนองเสียงของภาษานั้น ๆ ซึ่งแสดงให้เห็นถึงความแตกต่างของแต่ละภาษาได้อย่างชัดเจน

เราสามารถแบ่งภาษาได้เป็น 2 กลุ่ม คือ ภาษาที่มีเสียงวรรณยุกต์ (tonal language) กับภาษาที่ไม่มีเสียงวรรณยุกต์ (non-tonal language) ภาษาที่มีเสียงวรรณยุกต์ คือ ภาษาที่ลักษณะของความสูงต่ำของเสียงในแต่ละคำ หรือรูปร่างของคอนทัวร์ F_0 ในแต่ละคำมีผลต่อความหมายของคำนั้น เช่น ภาษาไทย ภาษาเวียดนาม ภาษาลาว หรือภาษาจีน ส่วนภาษาที่ไม่มีเสียงวรรณยุกต์ คือ ภาษาที่รูปร่างของคอนทัวร์ F_0 ไม่มีผลต่อความหมายของคำ เช่น ภาษาอังกฤษ ภาษาเยอรมัน หรือภาษาญี่ปุ่น

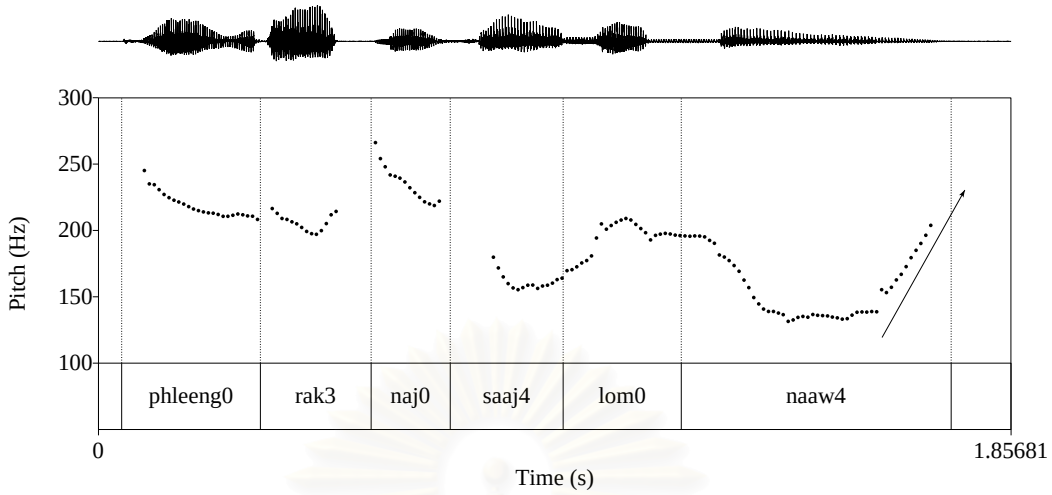
ภาษาที่ไม่มีเสียงวรรณยุกต์โดยส่วนใหญ่ สามารถแยกความแตกต่างระหว่างประโยคบอกเล่า กับประโยคคำถามได้โดยสังเกตจากลักษณะของคอนทัวร์ F_0 ซึ่งสูงขึ้นที่ท้ายประโยค เช่น ภาษาอังกฤษ (Taylor, 1992) ภาษาเยอรมัน (Mixdorff, 1998) และภาษาญี่ปุ่น (Ishi และคนอื่น ๆ, 2001) ตัวอย่างเช่น ลักษณะของทำนองเสียง 6 แบบของภาษาญี่ปุ่น ดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 ตัวอย่างของลักษณะของคอนทัวร์ F_0 ของทำนองเสียงทั้ง 6 แบบของภาษาญี่ปุ่นซึ่งไม่มีเสียงวรรณยุกต์ จะเห็นได้ว่า การชันขึ้นของ คอนทัวร์ F_0 แสดงให้เห็นถึงทำนองเสียงของประโยคคำถาม (Toki และ Murata, 1989 อ้างถึงใน Ishi และคนอื่น ๆ, 2001)

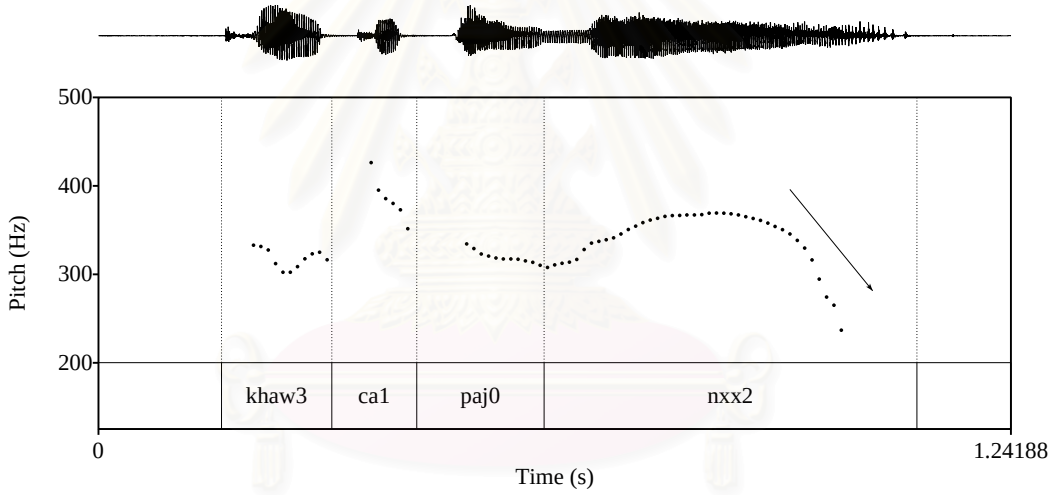
Type	Intention	Impression
Long Rise (LRs) ⤴	Question, confirmation, offer, invitation	Gentle
Short Rise (SRs) ↗	Question, confirmation, agreement	Carefree, cheerful
Long Flat (LFt) →	General answer sentences	Calm
Short Flat (SFt) ↠	General answer sentences	Carefree, cheerful
Weak Flat (WFt) ⇢	Reserved question, reserved decline	As talking to oneself
Long Fall (LFa) ⤵	Understanding, discovering, confirmation, doubt, offer	Consent, dissatisfied, disappointed

ส่วนภาษาที่มีเสียงวรรณยุกต์ เช่น ภาษาไทย และภาษาจีน (Yuan, Shih, และ Kochanski, 2002) นั้นไม่จำเป็นว่าการที่คอนทัวร์ F_0 สูงขึ้นที่ท้ายประโยค จะหมายถึงประโยคคำถามเสมอไป อาจเกิดจากการที่เสียงวรรณยุกต์ของพยางค์สุดท้ายมีรูปร่างของคอนทัวร์ F_0 ที่เพิ่มขึ้นอยู่แล้วก็ได้ ตัวอย่างเช่น คอนทัวร์ F_0 ในรูปที่ 2.8 (ก) เป็นเสียงพูดของผู้หญิงที่พูดว่า “เพลงรักในสายลมหนาว”.ซึ่งเป็นทำนองเสียงของประโยคบอกเล่า แต่คอนทัวร์ F_0 ที่ช่วงท้ายของประโยคจะมีลักษณะชันขึ้นเนื่องจากผลของเสียงวรรณยุกต์จัตวาของภาษาไทย ส่วนในรูปที่ 2.8 (ข) เป็นเสียงพูดของผู้หญิงที่พูดว่า “เค้าจะไปแน่?” ซึ่งเป็นทำนองเสียงของประโยคคำถาม แต่คอนทัวร์ F_0 ที่ช่วงท้ายของประโยคมีลักษณะตกลง เนื่องจากผลของเสียงวรรณยุกต์โทของภาษาไทย

ดังนั้นจะเห็นได้ว่า แต่ละภาษาจำเป็นจะต้องมีวิธีการรู้จำทำนองเสียงที่แตกต่างกันไป ขึ้นอยู่กับลักษณะของเสียงสูงต่ำของภาษา และการนิยามความหมายของทำนองเสียงของภาษานั้น ๆ



(ก)



(ข)

รูปที่ 2.8 ตัวอย่างของลักษณะของคอนทัวร์ F_0 ของภาษาไทยซึ่งมีเสียงวรรณยุกต์ โดยสามารถเกิดกรณี (ก) คอนทัวร์ F_0 ชันขึ้นที่ท้ายประโยค ในกรณีของทำนองเสียงของประโยคบอกเล่า และกรณี (ข) คอนทัวร์ F_0 ตกลงที่ท้ายประโยคในกรณีของทำนองเสียงของประโยคคำถาม

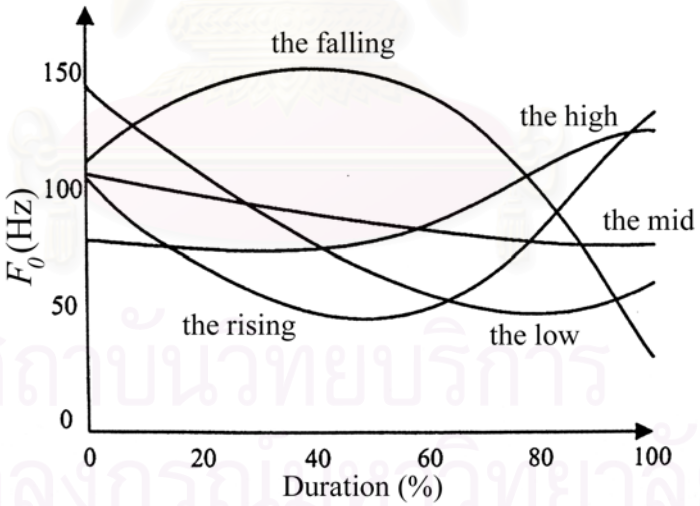
2.2.4 ลักษณะทางความสูงต่ำของเสียงพูดภาษาไทย

เราสามารถมองลักษณะทางความสูงต่ำของเสียงในภาษาไทย ว่าแบ่งได้เป็น 3 ประเภท ตามประเภทของสารสนเทศที่ให้ ในหัวข้อ 2.2.1 ได้แก่

2.2.4.1 เสียงวรรณยุกต์ (tone)

เป็นลักษณะทางความสูงต่ำของเสียงในแต่ละพยางค์ ที่ให้สารสนเทศทางภาษาศาสตร์ คือ บอกได้ว่าคำที่พูดนั้นมีความหมายว่าอย่างไร เช่น คำว่า คา ข่า ค่า ค้า และ ขา ซึ่งมีเสียงพยัญชนะต้นเป็นเสียง ‘ค’ และ มีเสียงสระ ‘อา’ เหมือนกัน แต่มีความหมายต่างกัน เนื่องจากลักษณะการออกเสียงสูงต่ำต่างกัน

เสียงวรรณยุกต์ของภาษาไทยมี 5 เสียง ได้แก่ เสียงสามัญ (the mid, แทนด้วยเลข 0) เสียงเอก (the low, แทนด้วยเลข 1) เสียงโท (the falling, แทนด้วยเลขเลข 2) เสียงตรี (the high, แทนด้วยเลข 3) และเสียงจัตวา (the rising แทนด้วยเลข 4) รูปที่ 2.9 แสดงคอนทัวร์ F_0 ของเสียงวรรณยุกต์แต่ละประเภทในภาษาไทยโดยเฉลี่ย จากผู้พูดที่เป็นผู้ชาย ซึ่งพูดคำ 1 พยางค์ครั้งละคำ



รูปที่ 2.9 คอนทัวร์ F_0 ของเสียงวรรณยุกต์ทั้ง 5 เสียงในภาษาไทยโดยเฉลี่ย พูดโดยผู้พูดที่เป็นผู้ชาย ซึ่งพูดคำ 1 พยางค์ ครั้งละคำ (Thubthong, 2001)

เมื่อพิจารณาคอนทัวร์ F_0 ของเสียงพูดต่อเนื่อง ดังที่ได้แสดงในรูปที่ 2.6 และ 2.8 โดยสังเกตจากตัวเลขแทนเสียงวรรณยุกต์ในสัญลักษณ์แทนเสียงของแต่ละพยางค์ (ดูวิธีการอ่านสัญลักษณ์แทนเสียงได้ในภาคผนวกที่ ก) จะเห็นได้ว่า คอนทัวร์ F_0 ของแต่ละพยางค์ เป็นผลมาจากเสียงวรรณยุกต์ ซึ่งมีลักษณะคล้ายกับรูปที่ 2.9 แต่จะไม่เหมือนเลยทีเดียว ซึ่งเป็นผลมาจากปัจจัยต่าง ๆ ได้แก่ โครงสร้างของพยางค์ (syllable structure) รูปร่างของคอนทัวร์ F_0 ของพยางค์ข้างเคียง (coarticulation) ทำนองเสียงพูด (intonation) การเน้นเสียง (stress) อัตราเร็วในการพูด (speaking rate) สำเนียงของภาษาท้องถิ่นของผู้พูด (dialect) เพศของผู้พูด (gender) อายุของผู้พูด (age) และอารมณ์ของผู้พูด (emotion) (Thubthong, 2001)

เมื่อพิจารณาตามลักษณะของสารสนเทศในหัวข้อ 2.2.1 จะได้ว่า เสียงวรรณยุกต์ในภาษาไทย เป็นสารสนเทศทางภาษาศาสตร์ เนื่องจากเสียงวรรณยุกต์มีผลต่อความหมายของคำ และเราสามารถบอกประเภทของเสียงวรรณยุกต์ได้จากข้อความที่เขียน

ตัวอย่างของงานวิจัย ที่เกี่ยวข้องกับการรู้จำเสียงวรรณยุกต์ภาษาไทย ทั้งแบบพยางค์เดียวและแบบเสียงพูดต่อเนื่อง ได้แก่ Thubthong (1995); Thubthong และ Kijsirikul (1999); Thubthong (2001); Charnvivit และคนอื่น ๆ (2001); Potisuk, Harper และ Gandour (1999); และ Ngarmchatetanarom และคนอื่น ๆ (2004)

2.2.4.2 ทำนองเสียงพูดภาษาไทย

งานวิจัยนี้ได้เลือกใช้ความหมายของทำนองเสียงพูดภาษาไทยตามความหมายของ กาญจนา นาคสกุล (2541) ตามที่ได้กล่าวไปแล้วในหัวข้อ 2.2.2 โดยเมื่อเทียบกับความหมายอื่น ๆ ของทำนองเสียง รวมทั้งลักษณะของสารสนเทศจากเสียงพูด จึงสรุปได้ว่า ทำนองเสียงพูดในความหมายของงานวิจัยนี้ คือ ลักษณะของคอนทัวร์ F_0 ของทั้งประโยค ซึ่งให้สารสนเทศกึ่งภาษาศาสตร์

การจัดประเภทสารสนเทศที่ได้จากทำนองเสียงพูด ว่าเป็นสารสนเทศกึ่งภาษาศาสตร์ เนื่องจากทำนองเสียงพูด เป็นสิ่งที่ผู้พูดเพิ่มเข้าไปนอกเหนือจากข้อความที่เขียน เพื่อแสดงทัศนคติต่าง ๆ โดยในงานวิจัยนี้ ได้พิจารณาทำนองเสียงพูดในแง่ของสารสนเทศที่เป็นดิสครีต โดยแบ่งทำนองเสียงพูดภาษาไทยออกเป็น 3 ประเภท ตามความหมายของ Luksaneeyanawin (1998)

Luksaneeyanawin ได้จำแนกประเภทของทำนองเสียงพูดภาษาไทยออกเป็น 3 ประเภท คือ ทำนองเสียงตก ทำนองเสียงขึ้น และทำนองเสียงแบบผสม โดยทำนองเสียงทั้งสามประเภทเกิดเมื่อผู้พูดพูดประโยคที่มีลักษณะต่าง ๆ กัน ดังนี้

- *ทำนองเสียงตก (the Fall Class)* เกิดเมื่อผู้พูดพูดประโยคบอกเล่า (statement) พูดชมเชย หรือพูดเป็นคำ ๆ (citation form) พูดแบบไม่แสดงความคิดเห็น (attitudinally unmarked) พูดแบบยอมจำนน (submissive) พูดแบบซ่อนความโกรธ (concealed anger) พูดแบบเบื่อ (bored) และพูดแบบแสดงอำนาจ (authoritative) เมื่อพิจารณารูปร่างของคอนทัวร์ F_0 แบบแคบ (ในระดับพยางค์) จะพบว่ารูปร่างของคอนทัวร์ F_0 ของแต่ละพยางค์ ขึ้นกับชนิดของเสียงวรรณยุกต์ของพยางค์นั้น ดังแสดงในรูปที่ 2.9 แต่ระดับของ F_0 ของพยางค์ที่ถัดจากพยางค์แรกจะค่อย ๆ ลดระดับลงไปเรื่อย ๆ (downdrift) จนจบประโยค ดังนั้นเมื่อพิจารณาแบบกว้าง (ในระดับประโยค) จึงเห็นได้ว่าคอนทัวร์ F_0 ของทำนองเสียงตก จะค่อย ๆ ลดระดับลง (decline) เมื่อพูดไปเรื่อย ๆ จนจบประโยค

- *ทำนองเสียงขึ้น (the Rise Class)* เกิดเมื่อผู้พูดพูดประโยคที่แสดงให้เห็นถึงการถาม (question) แสดงความไม่เห็นด้วย (disagreeable) แสดงความไม่เชื่อ (disbelieving) แสดงความไม่คาดฝัน (surprised) และพูดแบบแสดงให้เห็นว่ายังพูดไม่จบความ (unfinished) เมื่อพิจารณารูปร่างของคอนทัวร์ F_0 แบบแคบ จะพบว่ารูปร่างของคอนทัวร์ F_0 ของแต่ละพยางค์ ขึ้นกับชนิดของเสียงวรรณยุกต์ เช่นเดียวกับทำนองเสียงตก แต่ช่วงค่า F_0 จะแคบกว่าเล็กน้อย และมีระดับของ F_0 ที่สูงกว่า จึงส่งผลให้รูปร่างของคอนทัวร์มีลักษณะสูงขึ้นกว่าทำนองเสียงตกเมื่อพิจารณาในระดับประโยค

- *ทำนองเสียงผสม (the Convolution Class)* เกิดเมื่อผู้พูดพูดเพื่อต้องการเน้นหนัก (emphatic) พูดแสดงความโกรธ (anger) แสดงความเห็นด้วยอย่างมาก (very agreeable) แสดงความสนใจมาก (very interested) และแสดงความเชื่อถือมาก (very believing) ในกรณีของทำนองเสียงผสม รูปร่างของคอนทัวร์ F_0 ในแต่ละพยางค์จะขึ้นกับประเภทของเสียงวรรณยุกต์ เช่นเดียวกับทั้ง 2 ทำนองเสียง ดังที่ได้กล่าวไปแล้ว แต่ช่วงค่า F_0 ของเสียงวรรณยุกต์ทุกเสียงจะกว้างกว่า ทำนองเสียงตก และทำนองเสียงขึ้น นอกจากนี้ในกรณีของเสียงวรรณยุกต์ สามัญ เอก โท และตรี จะพบว่า F_0 มีระดับที่สูงกว่า ทำนองเสียงตก แต่เสียงวรรณยุกต์จัตวาจะมีระดับ F_0 ที่ต่ำ ซึ่งส่งผลให้รูปร่างของคอนทัวร์ F_0 ในระดับประโยค มีลักษณะสูงกว่า ทำนองเสียงตก และช่วงการแกว่งกว้างกว่าทำนองเสียงตก และทำนองเสียงขึ้น

ตัวอย่างของคอนทัวร์ F_0 ของทำนองเสียงทั้งสามแบบ ของผู้พูดชาย และผู้พูดหญิง แสดงให้เห็นในบทที่ 3 รูปที่ 3.1 – 3.8 ซึ่งจะเห็นได้ว่า ทำนองเสียงพูดส่งผลต่อลักษณะของคอนทัวร์ F_0 ในระดับประโยค เช่น ระดับความสูงของคอนทัวร์ ความลาดเอียงของคอนทัวร์ หรือ ความกว้างของช่วงการแกว่งของคอนทัวร์ ในขณะที่ประเภทของเสียงวรรณยุกต์ ส่งผลถึงลักษณะของคอนทัวร์ F_0 ในแต่ละพยางค์ โดยจะเห็นว่าคอนทัวร์ F_0 ของเสียง

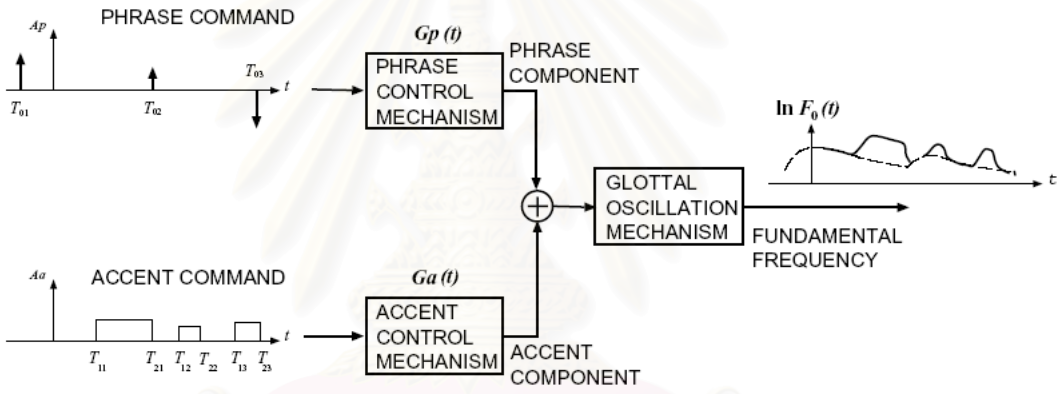
2.2.4.3 ลักษณะทางความสูงต่ำอื่น ๆ

ลักษณะทางความสูงต่ำอื่น ๆ ซึ่งถือว่าเป็นสารสนเทศที่ไม่ใช่ภาษาศาสตร์ ได้แก่ เพศ อายุ และลักษณะทางกายภาพของผู้พูด จากที่กล่าวไว้ในหัวข้อ 2.1.2 จะเห็นได้ว่าเพศของผู้พูด สามารถส่งผลถึงความสูงของ F_0 และช่วงกว้างของค่า F_0 ได้ เช่นเดียวกับทำนองเสียงพูด

2.3 แบบจำลองฟูจิซากิ (Fujisaki model)

แบบจำลองฟูจิซากิ (Fujisaki และ Hirose, 1982; Fujisaki และ Hirose, 1984) เป็นแบบจำลองทางคณิตศาสตร์ของคอนทัวร์ F_0 โดยมีพื้นฐานมาจาก กลไกการสั่นของเส้นเสียง (glottal oscillation mechanism) ซึ่งได้รับการแสดงให้เห็นแล้วว่า แบบจำลองฟูจิซากิสามารถประมาณคอนทัวร์ F_0 ได้ใกล้เคียงกับ คอนทัวร์ F_0 ที่กำหนดให้มาก

รูปที่ 2.10 แสดงแผนภาพของแบบจำลองฟูจิซากิ จะเห็นได้ว่า รูปร่างของคอนทัวร์ F_0 เป็นผลมาจากสัญญาณคำสั่งขาเข้า 2 สัญญาณ คือ คำสั่งวลี (phrase command) ซึ่งเป็นฟังก์ชันอิมพัลส์ (impulse function) และคำสั่งการเน้นเสียง (accent command) ซึ่งเป็นฟังก์ชันขั้น (stepwise function) ซึ่งจะให้ผลลัพธ์เป็นคอนทัวร์ F_0 ในสเกลลอการิทึม ดังสมการที่ 2.2



รูปที่ 2.10 แผนภาพของแบบจำลองฟูจิซากิ (Fujisaki และ Hirose, 1984)

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j [Ga(t - T_{1j}) - Ga(t - T_{2j})] \qquad (2.2)$$

โดยที่

$$Gp(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \dots t \geq 0, \\ 0, & \dots t < 0, \end{cases} \qquad (2.3)$$

และ

$$Ga(t) = \begin{cases} \min [1 - (1 + \beta t) \exp(-\beta t), \gamma], & \dots t \geq 0, \\ 0, & \dots t < 0, \end{cases} \qquad (2.4)$$

ตัวแปรต่าง ๆ ในสมการที่ 2.2 – 2.4 เรียกว่า พารามิเตอร์ฟูจิซากิ (Fujisaki parameters) มีความหมายดังนี้

Ap ขนาดของคำสั่งวลี (phrase command magnitude)

T0 เวลาเริ่มต้นของคำสั่งวลี (phrase command onset time)

α ความถี่เชิงมุมธรรมชาติของกลไกควบคุมวลี (natural angular frequency of the phrase control mechanism)

Aa ขนาดของคำสั่งเน้นเสียง (accent command amplitude)

T1 เวลาเริ่มต้นของคำสั่งเน้นเสียง (accent command onset time)

T2 เวลาสิ้นสุดของคำสั่งเน้นเสียง (accent command offset time)

β ความถี่เชิงมุมธรรมชาติของกลไกควบคุมการเน้นเสียง (natural angular frequency of the accent control mechanism)

γ ค่าเพดานขององค์ประกอบการเน้นเสียง (ceiling value)

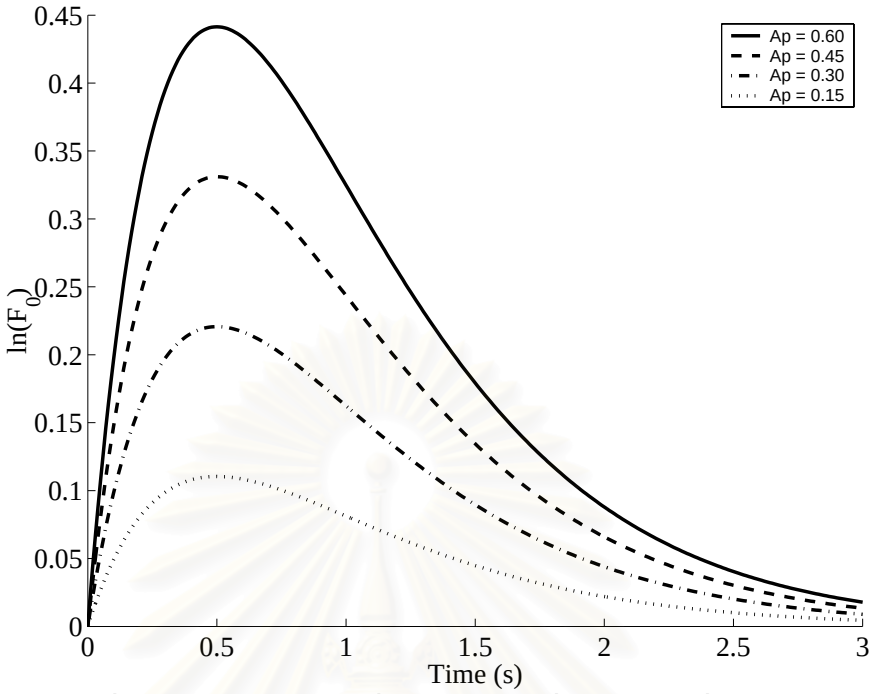
Fb ความถี่ฐาน (base frequency) เป็นค่าซึ่งขึ้นกับผู้พูดแต่ละคน

องค์ประกอบวลี (Phrase component)

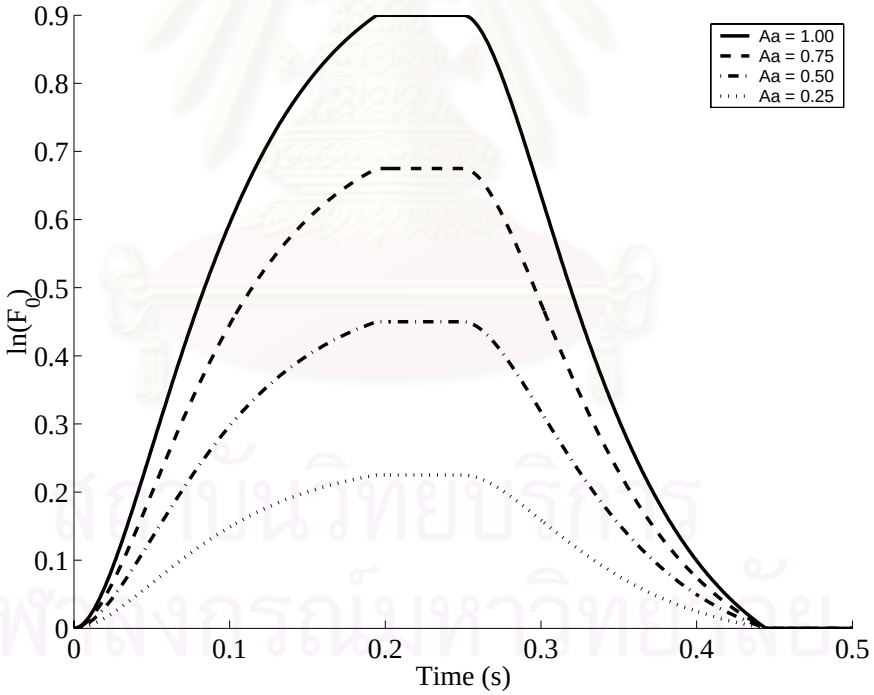
องค์ประกอบวลีแทนด้วยฟังก์ชัน Gp(t) ในสมการที่ 2.3 เป็นผลตอบอิมพัลส์ (impulse response) ของกลไกควบคุมวลี (phrase control mechanism) สัญญาณอิมพัลส์ขาเข้ากำหนดด้วยขนาด Ap และเวลาเริ่มต้น T0 โดยที่ α คือ ความถี่เชิงมุมธรรมชาติของกลไกควบคุมวลี ซึ่งสามารถกำหนดให้มีค่าคงที่ได้สำหรับแต่ละประโยค ตัวอย่างขององค์ประกอบวลีที่ค่า $\alpha = 2.0\ s^{-1}$ (Mixdorff, 1998) และค่า Ap ต่าง ๆ แสดงดังรูปที่ 2.11 ซึ่งจะเห็นได้ว่าการเพิ่มค่า Ap เป็นการเพิ่มอัตราการลดระดับของ F_0 ซึ่งสามารถใช้เป็นแบบจำลองของทำนองเสียงตกได้

องค์ประกอบการเน้นเสียง (accent component)

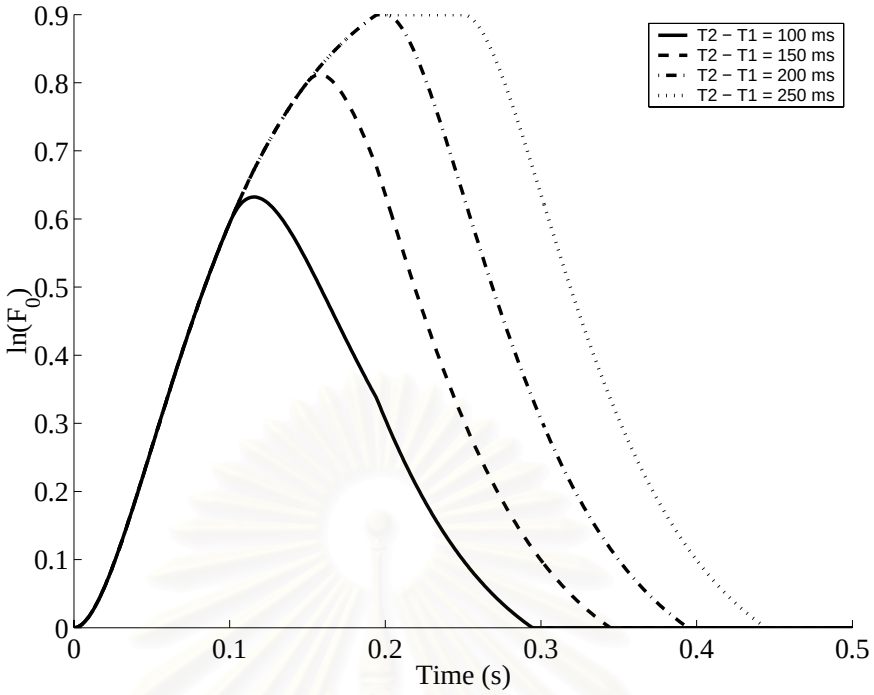
องค์ประกอบการเน้นเสียง แทนด้วยฟังก์ชัน Ga(t) ในสมการที่ 2.4 เป็นผลตอบของฟังก์ชันขั้น ของกลไกควบคุมการเน้นเสียง สัญญาณขั้นขาเข้ากำหนดได้ด้วยขนาด Aa เวลาเริ่มต้น T1 และเวลาสิ้นสุด T2 โดยที่ β คือ ความถี่เชิงมุมธรรมชาติของกลไกควบคุมการเน้นเสียง ซึ่งสามารถกำหนดให้มีค่าคงที่ได้สำหรับแต่ละประโยค ตัวอย่างขององค์ประกอบการเน้นเสียงที่ค่า Aa ต่าง ๆ โดยกำหนดให้ค่าความกว้างของคำสั่งการเน้นเสียง (T2 – T1) มีค่าคงที่เป็น 250 ms (Mixdorff, 1998) แสดงดังรูปที่ 2.12 โดยมีค่า γ (นิยมกำหนดให้มีค่า 0.9 (Mixdorff, 1998)) ทำให้ค่าขององค์ประกอบการเน้นเสียงมีค่าถึงค่าสูงสุดที่เวลาที่ไม่เป็นอนันต์ ส่วนรูปที่ 2.13 แสดงให้เห็นถึงองค์ประกอบการเน้นเสียง เมื่อเปลี่ยนค่าความกว้างของคำสั่งการเน้นเสียง



รูปที่ 2.11 องค์ประกอบวลีที่ค่า Ap ต่าง ๆ เมื่อ α มีค่าคงที่ = 2.0 s^{-1}



รูปที่ 2.12 องค์ประกอบการเน้นเสียงที่ค่า Aa ต่าง ๆ เมื่อกำหนดให้ ความกว้างของคำสั่งการเน้นเสียง (T2 – T1) มีค่าคงที่ = 250 ms และ β มีค่าคงที่ = 20 s^{-1}



รูปที่ 2.13 องค์ประกอบการเน้นเสียงที่ค่าความกว้างของคำสั่งการเน้นเสียง (T2 – T1) ต่าง ๆ เมื่อกำหนดให้ Aa มีค่าคงที่ = 1.0 และ β มีค่าคงที่ = 20 s^{-1}

แบบจำลองฟูจิซากิ ถูกนำมาใช้ครั้งแรกกับภาษาญี่ปุ่น (Fujisaki และ Hirose, 1982; Fujisaki และ Hirose, 1984) และในระยะต่อมาได้มีผู้นำแบบจำลองฟูจิซากิไปปรับใช้กับภาษาอื่น ๆ จำนวนมาก ทั้งภาษาที่มีเสียงวรรณยุกต์ และภาษาที่ไม่มีเสียงวรรณยุกต์ เช่น ภาษาอังกฤษ (Fujisaki และ Ohno, 1995) ภาษาเยอรมัน (Mixdorff, 1998) ภาษากรีก (Fujisaki, Ohno และ Yagi, 1997) ภาษาเกาหลี (Fujisaki, 1996a) ภาษาโปรตุเกส (Teixeira, Freitas และ Fujisaki, 2003) ภาษาสเปน (Fujisaki และคนอื่น ๆ, 1994) ภาษาสวีเดน (Fujisaki, Ljungqvist และ Murata, 1993) ภาษาจีน (Fujisaki, Hallé และ Lei, 1987) และภาษาเวียดนาม (Mixdorff และคนอื่น ๆ, 2003) เป็นต้น โดยในภาษาที่มีเสียงวรรณยุกต์มักจะใช้คำว่า คำสั่งเสียงวรรณยุกต์ (tone command) และองค์ประกอบเสียงวรรณยุกต์ (tone component) แทนที่จะเป็นคำสั่งการเน้นเสียง และองค์ประกอบการเน้นเสียง

สำหรับภาษาไทยนั้นมีผู้นำแบบจำลองฟูจิซากิไปใช้ในการรู้จำเสียงวรรณยุกต์ภาษาไทย (Potisuk, Harper และ Gandour, 1995; Potisuk, Harper และ Gandour, 1999; Ngarmchatetanarom และคนอื่น ๆ, 2004) นอกจากนี้ยังมีงานด้านการวิเคราะห์ลักษณะของเสียงพูดภาษาไทยซึ่งใช้แบบจำลองฟูจิซากิอีกด้วย เช่น การทดสอบการรับรู้เสียงวรรณยุกต์ และความสั้นยาวของสระภาษาไทย (Mixdorff และคนอื่น ๆ, 2002) และการวิเคราะห์จังหวะการพูดภาษาไทยเพื่อนำไปใช้ในงานด้านการสังเคราะห์เสียงพูด (Mixdorff และคนอื่น ๆ, 2003)

งานวิจัยนี้ได้เน้นหนักในเรื่องของการรู้จำทำนองเสียงพูด ซึ่งสามารถพิจารณาได้ว่าขึ้นอยู่กับองค์ประกอบวลี ความถี่ฐาน (base frequency) และ ยังรวมไปถึงขนาดขององค์ประกอบวรรณยุกต์อีกด้วย แต่ในงานวิจัยนี้ไม่ได้นำพารามิเตอร์ฟูจิซากิมาใช้โดยตรง เนื่องจากกรรมวิธีในการหาค่าพารามิเตอร์ฟูจิซากิแบบอัตโนมัติ ในปัจจุบันยังใช้วิธีที่ค่อนข้างซับซ้อน และใช้เวลามาก (Mixdorff, 2000) แต่จะนำแนวคิดการแบ่งคอนทัวร์ F_0 ออกเป็นองค์ประกอบวรรณยุกต์ และองค์ประกอบวลีมาใช้ โดยจะกล่าวถึงอย่างละเอียดในบทที่ 3

บทที่ 3

คอนทัวร์ลักษณะ และระบบรู้จำทำนองเสียงพูด

บทนี้กล่าวถึงการหาลักษณะที่เกี่ยวข้องกับทำนองเสียงพูดจากสัญญาณเสียงพูด โดยนำเสนอวิธีการหาคอนทัวร์ลักษณะจากคอนทัวร์ F_0 ซึ่งรูปร่างของคอนทัวร์ลักษณะนี้จะขึ้นกับลักษณะของทำนองเสียงพูด จากนั้นจะกล่าวถึงระบบรู้จำทำนองเสียงพูดที่ออกแบบขึ้น โดยนำคอนทัวร์ลักษณะไปแปลงเป็นเวกเตอร์ลักษณะ แล้วจึงนำไปรู้จำด้วยโครงข่ายประสาทเทียม

3.1 คอนทัวร์ F_0 ของทำนองเสียงแบบต่าง ๆ

จากทฤษฎีที่กล่าวถึงในบทที่ 2 จะเห็นได้ว่าคอนทัวร์ F_0 ประกอบด้วยสารสนเทศภาษาศาสตร์ สารสนเทศกึ่งภาษาศาสตร์ และสารสนเทศที่ไม่ใช่ภาษาศาสตร์ แต่เนื่องจากความหมายของทำนองเสียงพูดภาษาไทยที่ใช้ในงานวิจัยนี้ เป็นสารสนเทศประเภทกึ่งภาษาศาสตร์ ระบบรู้จำทำนองเสียงจึงต้องสามารถกำจัด หรือลดผลของสารสนเทศประเภทอื่นออกไปให้ได้มากที่สุด เพื่อที่จะได้ไม่เป็นภาระแก่ตัวรู้จำมากเกินไป งานวิจัยนี้จึงนำเสนอคอนทัวร์ลักษณะ (feature contours) ซึ่งเป็นคอนทัวร์ที่ได้มาจากการนำคอนทัวร์ F_0 ไปลดผลของสารสนเทศประเภทอื่นที่ไม่ใช้สารสนเทศกึ่งภาษาศาสตร์

รูปที่ 3.1 ถึงรูปที่ 3.8 แสดงให้เห็นถึงคอนทัวร์ F_0 ของผู้พูดชาย และหญิง เมื่อพูดด้วยทำนองเสียงแบบต่าง ๆ ทั้งสามแบบ จะเห็นได้ว่าสารสนเทศภาษาศาสตร์ คือ รูปร่างของคอนทัวร์ F_0 ในแต่ละพยางค์ ซึ่งเป็นตัวบอกประเภทของเสียงวรรณยุกต์ สารสนเทศกึ่งภาษาศาสตร์ คือ รูปร่างในระดับประโยคของคอนทัวร์ F_0 ซึ่งเป็นตัวบอกประเภทของทำนองเสียง สารสนเทศที่ไม่ใช่ภาษาศาสตร์ คือ ระดับและ ช่วงกว้างของคอนทัวร์ F_0 ซึ่งเป็นตัวบอกเพศของผู้พูด

จะเห็นได้ว่าลักษณะของคอนทัวร์ F_0 ที่ให้สารสนเทศกึ่งภาษาศาสตร์ และสารสนเทศที่ไม่ใช่ภาษาศาสตร์นั้นมีความสัมพันธ์กันมาก นั่นคือระดับของ F_0 ของเสียงผู้ชายที่พูดด้วยทำนองเสียงตกกับระดับ F_0 ของเสียงผู้หญิงที่พูดด้วยทำนองเสียงขึ้นนั้นใกล้เคียงกันมาก จึงเป็นการยากที่จะแบ่งสารสนเทศกึ่งภาษาศาสตร์ และสารสนเทศที่ไม่ใช้ภาษาศาสตร์ออกจากกัน คอนทัวร์ลักษณะที่นำเสนอจึงเป็นการลดผลของสารสนเทศภาษาศาสตร์เท่านั้น โดยยังคงเหลือสารสนเทศกึ่งภาษาศาสตร์ และสารสนเทศที่ไม่ใช่ภาษาศาสตร์อยู่ แต่จะทำการทดลองเสียงผู้ชาย และเสียงผู้หญิงแยกจากกัน เพื่อให้สารสนเทศที่ไม่ใช่ภาษาศาสตร์ส่งผลต่อการรู้จำทำนองเสียงน้อยที่สุด

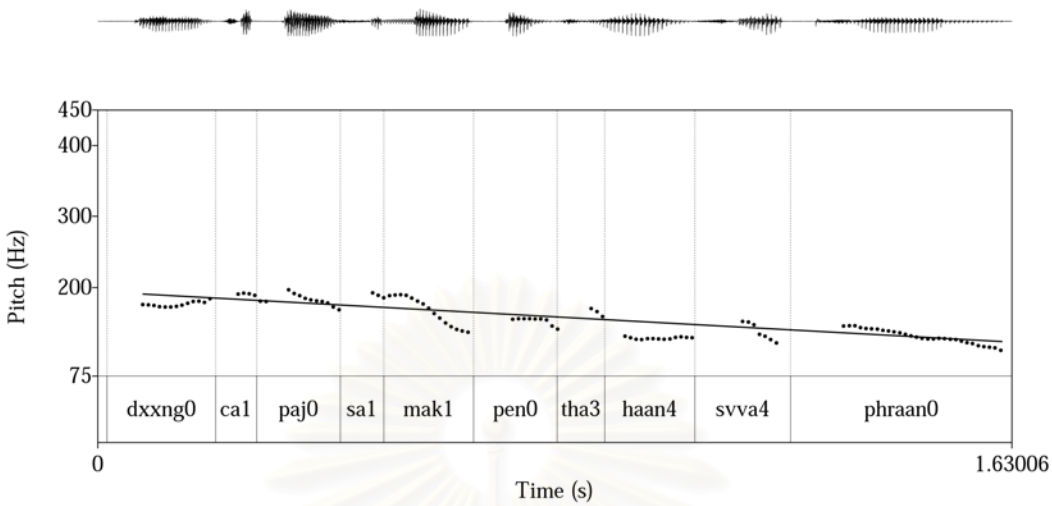
การลดผลของสารสนเทศภาษาศาสตร์ เริ่มจากการพิจารณาคอนทัวร์ F_0 ว่า เกิดจากองค์ประกอบ 3 ส่วน ตามแบบจำลองฟูจิซากิ สำหรับภาษาที่มีเสียงวรรณยุกต์ ตามที่ได้กล่าวถึงในบทที่ 2 คือ องค์ประกอบวลี องค์ประกอบเสียงวรรณยุกต์ และค่าความถี่ฐาน Fb

เมื่อพิจารณาคอนทัวร์ F_0 ออกเป็นองค์ประกอบต่าง ๆ เช่นนี้ จะเห็นได้ว่าองค์ประกอบสำคัญที่ต้องนำมาใช้ในการรู้จำทำนองเสียงคือ องค์ประกอบวลี ซึ่งสามารถนำมาใช้บอกประเภทของทำนองเสียงได้ ตัวอย่างเช่น ประโยคที่มีทำนองเสียงตกคอนทัวร์ F_0 จะมีลักษณะค่อย ๆ ตกลง ดังแสดงในรูปที่ 3.1 และ 3.2

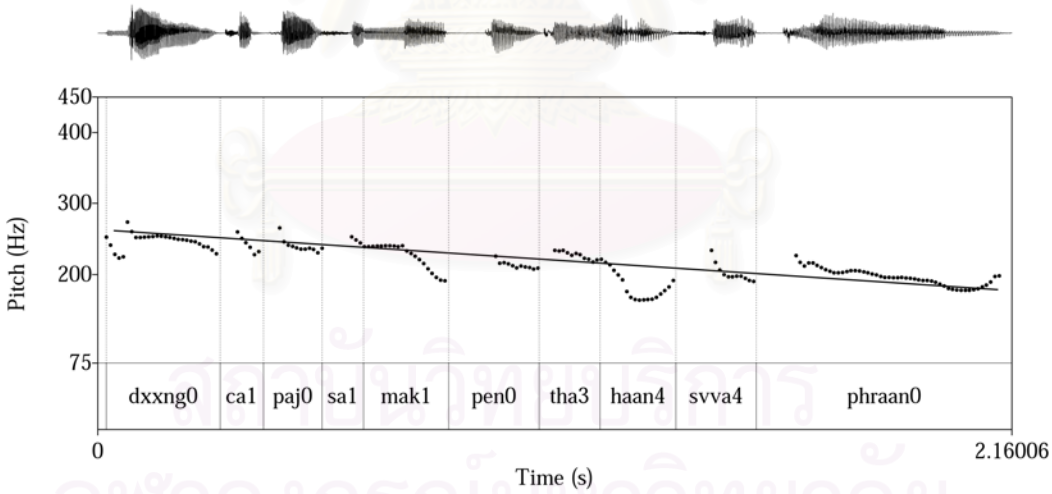
ส่วนประโยคที่มีทำนองเสียงขึ้น ซึ่งพบได้ในประโยคคำถาม และประโยคที่พูดไม่จบ จะพบคอนทัวร์ F_0 2 แบบ แบบแรก คือ คอนทัวร์ F_0 มีลักษณะเอียงขึ้น ดังแสดงในรูปที่ 3.3 และ 3.4 ส่วนแบบที่ 2 คอนทัวร์ F_0 มีลักษณะเอียงลง เช่นเดียวกับทำนองเสียงตก ดังแสดงในรูปที่ 3.5 และ 3.6 ความแตกต่างระหว่างทำนองเสียงตก และทำนองเสียงขึ้นแบบที่สอง คือ ทำนองเสียงขึ้นแบบที่สองนี้มีระดับ F_0 ที่สูงกว่า F_0 ของทำนองเสียงตก ด้วยเหตุนี้การรู้จำทำนองเสียงพูดจึงต้องรวมเอาองค์ประกอบ F_b ซึ่งแสดงถึงระดับของ F_0 เข้าไปด้วย

ส่วนในประโยคที่พูดด้วยทำนองเสียงผสมนั้น องค์ประกอบวลีจะมีลักษณะขึ้น ๆ ลง ๆ รวมทั้งระดับ F_0 ก็สูงกว่าในกรณีทำนองเสียงตกด้วยเช่นกัน ดังแสดงในรูปที่ 3.7 และ 3.8

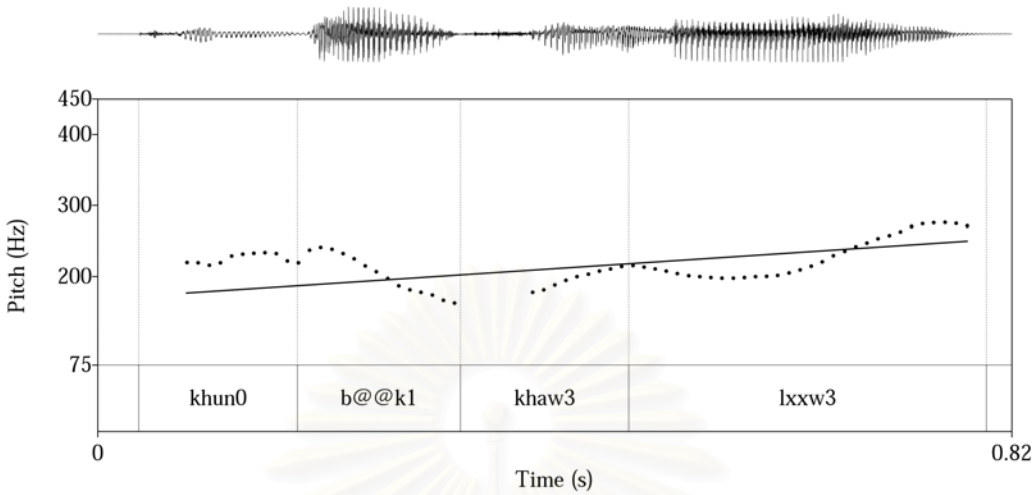
ถึงแม้ว่าองค์ประกอบวรรณยุกต์ในภาษาไทย จะเป็นองค์ประกอบที่ให้สารสนเทศประเภทภาษาศาสตร์ แต่จากรูปที่ 3.1 ถึง 3.8 จะเห็นได้ว่า ช่วงการแกว่งขององค์ประกอบวรรณยุกต์ ในกรณีของทำนองเสียงขึ้น และทำนองเสียงผสม จะกว้างกว่าในกรณีของทำนองเสียงตก แสดงว่าองค์ประกอบวรรณยุกต์ได้รับผลจากประเภทของทำนองเสียงด้วย ดังนั้นในการรู้จำทำนองเสียงจึงต้องพยายามหาลักษณะ ที่แสดงถึงช่วงกว้างของการแกว่งขององค์ประกอบวรรณยุกต์ แต่ไม่แสดงถึงรูปร่างของ F_0 อันเนื่องมาจากประเภทของเสียงวรรณยุกต์



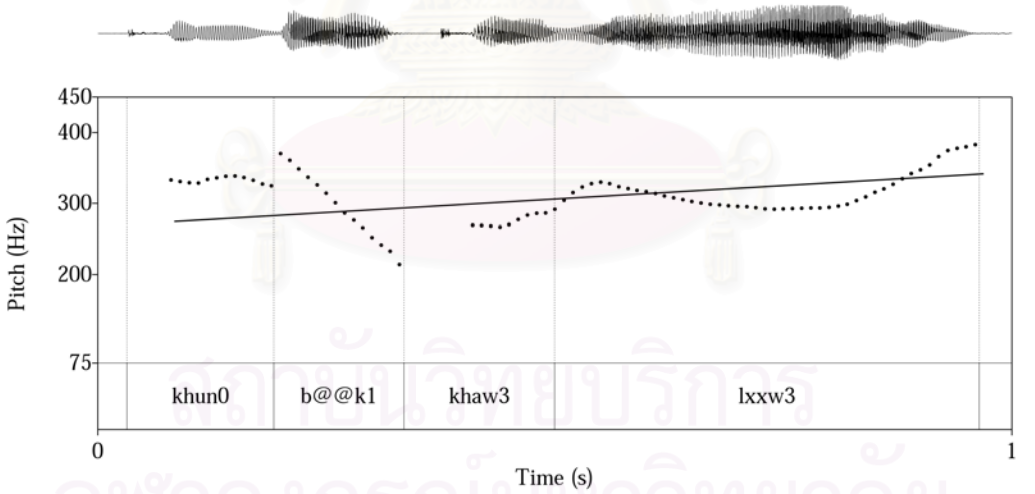
รูปที่ 3.1 ตัวอย่างของทำนองเสียงตก: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค "แดงจะไปสมัครเป็นทหารเสือพราน" เมื่อผู้พูดเป็นผู้ชาย



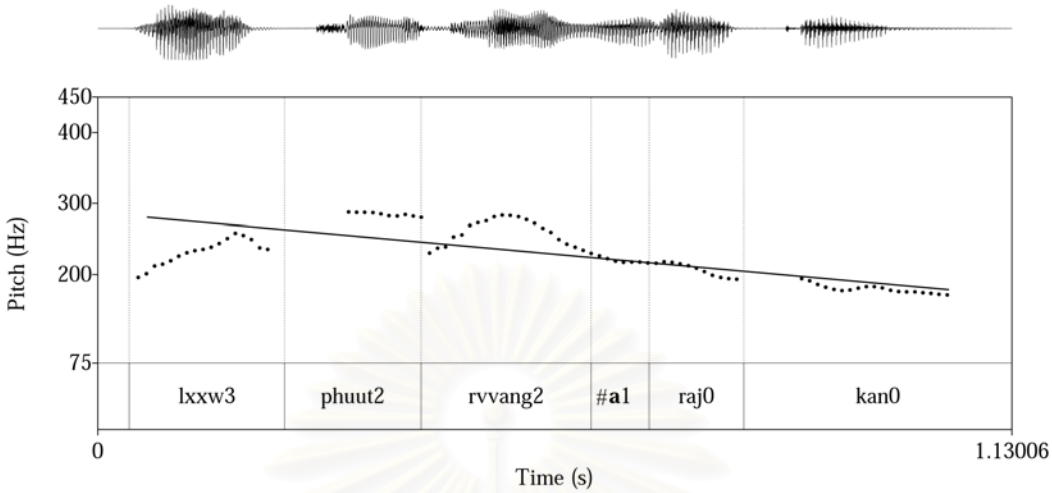
รูปที่ 3.2 ตัวอย่างของทำนองเสียงตก: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค "แดงจะไปสมัครเป็นทหารเสือพราน" เมื่อผู้พูดเป็นผู้หญิง



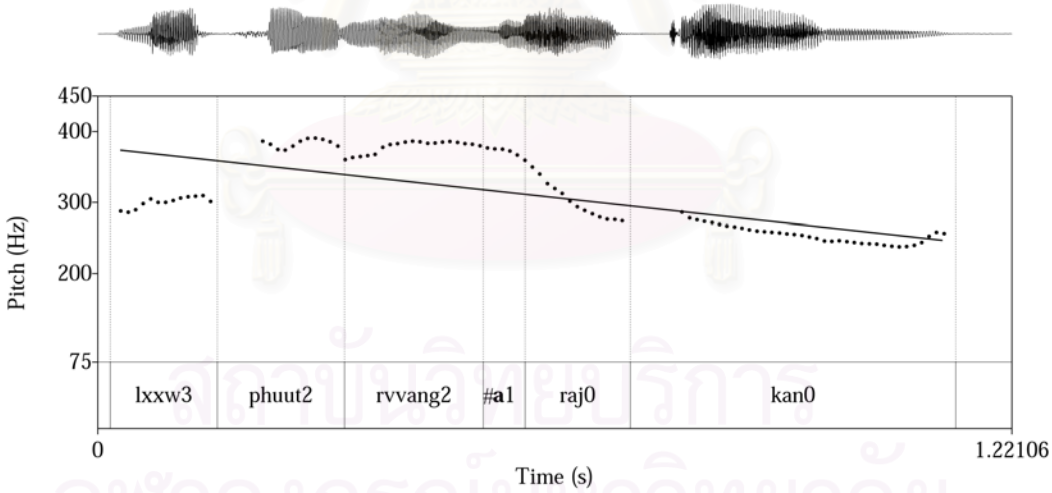
รูปที่ 3.3 ตัวอย่างของทำนองเสียงขึ้นแบบที่ 1: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค “คุณบอกเค้าแล้ว?” เมื่อผู้พูดเป็นผู้ชาย



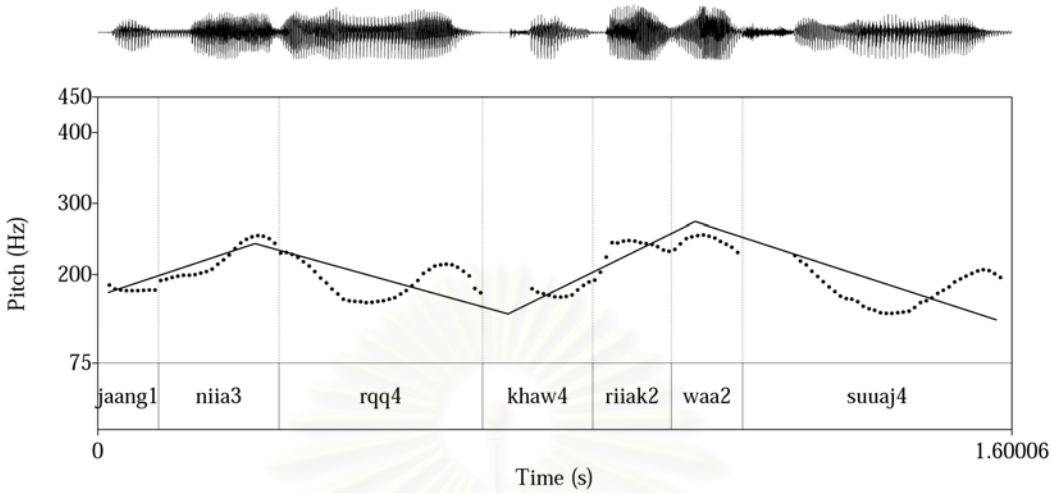
รูปที่ 3.4 ตัวอย่างของทำนองเสียงขึ้นแบบที่ 1: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค “คุณบอกเค้าแล้ว?” เมื่อผู้พูดเป็นผู้หญิง



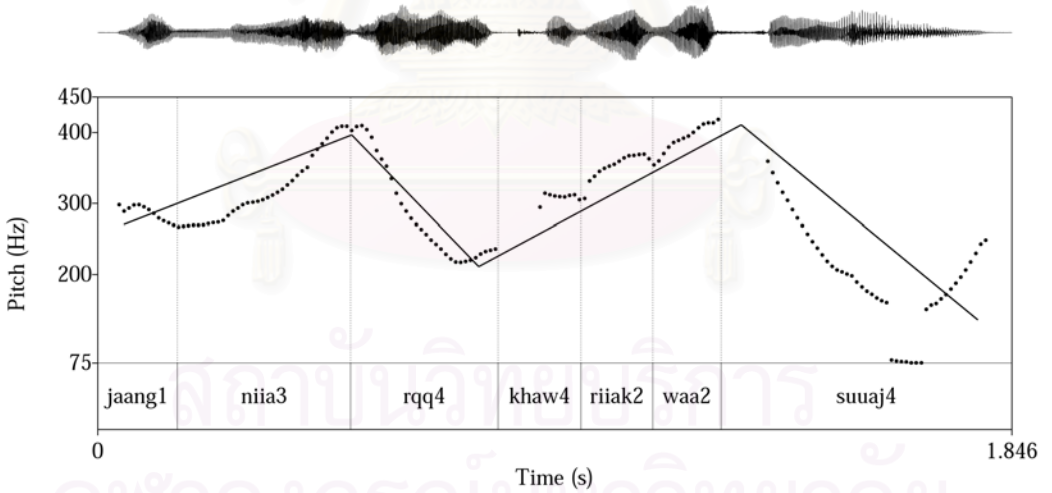
รูปที่ 3.5 ตัวอย่างของทำนองเสียงขึ้นแบบที่ 2: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค “แล้วพูดเรื่องอะไรกัน?” เมื่อผู้พูดเป็นผู้ชาย



รูปที่ 3.6 ตัวอย่างของทำนองเสียงขึ้นแบบที่ 2: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค “แล้วพูดเรื่องอะไรกัน?” เมื่อผู้พูดเป็นผู้หญิง



รูปที่ 3.7 ตัวอย่างของทำนองเสียงผสม: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค “อย่างเนี้ยเหรอเขาเรียกว่าสวย!” เมื่อผู้พูดเป็นผู้ชาย



รูปที่ 3.8 ตัวอย่างของทำนองเสียงผสม: รูปคลื่นเสียง (รูปบน) คอนทัวร์ F_0 (เส้นประในรูปล่าง) และเส้นแสดงทำนองเสียง (เส้นทึบในรูปล่าง) ของประโยค “อย่างเนี้ยเหรอเขาเรียกว่าสวย!” เมื่อผู้พูดเป็นผู้หญิง

3.2 การหาคอนทัวร์ลักษณะ

เนื่องจากงานวิจัยนี้พิจารณาคอนทัวร์ F_0 ตามแบบจำลองฟูจิซากิ วิธีการหนึ่งซึ่งสามารถนำมาใช้ในการรู้จำทำนองเสียงพูด คือ การพิจารณาแบบจำลองฟูจิซากิในกระบวนการย้อนกลับ นั่นคือการหาพารามิเตอร์ฟูจิซากิจากคอนทัวร์ F_0 ของเสียงพูดขาเข้า แล้วนำพารามิเตอร์ที่เกี่ยวข้องกับทำนองเสียงพูด นั่นคือ ค่า Fb, Ap, T0 และ Aa ไปใช้เป็นเวกเตอร์ลักษณะขาเข้าของตัวรู้จำ ตัวอย่างของงานวิจัยที่เกี่ยวข้องกับการหาค่าพารามิเตอร์ฟูจิซากิจากคอนทัวร์ F_0 แบบอัตโนมัติ ได้แก่ งานของ Mixdorff (2000) ซึ่งเน้นที่ภาษาเยอรมัน และ งานของ Mixdorff, Fujisaki, Chen และ Hu (2003) ซึ่งเน้นที่ภาษาจีนกลาง งานต่าง ๆ เหล่านี้แสดงให้เห็นถึงการพิจารณาคอนทัวร์ F_0 เป็นสัญญาณดิสครีตทางเวลา (discrete-time signal) แล้วจึงใช้ตัวกรอง (filter) เพื่อแยกคอนทัวร์ F_0 ออกเป็น องค์ประกอบความถี่ต่ำ (low frequency component) ซึ่งแทนด้วยคอนทัวร์ความถี่ต่ำ (low frequency contour: LFC) และองค์ประกอบความถี่สูง (high frequency component) ซึ่งแทนด้วยคอนทัวร์ความถี่สูง (high frequency contour: HFC) โดยพิจารณาว่าองค์ประกอบความถี่ต่ำเกิดจาก Fb และองค์ประกอบวลี ส่วนองค์ประกอบความถี่สูงเกิดจากองค์ประกอบเสียงวรรณยุกต์ จากนั้นจึงกำหนดค่าเริ่มต้นให้กับพารามิเตอร์ฟูจิซากิจากองค์ประกอบทั้งสอง แล้วใช้วิธีการค้นหาแบบปีนเขา (hill-climb search) เพื่อลดความผิดพลาดกำลังสองเฉลี่ย (mean square error) ระหว่างคอนทัวร์ F_0 ขาเข้า และคอนทัวร์ F_0 ที่ได้จากแบบจำลอง แล้วจึงผ่านขั้นตอนวิธีการปรับแต่งโดยละเอียด (fine tune) ต่าง ๆ แบบอัตโนมัติ เพื่อให้ได้คอนทัวร์ F_0 จากแบบจำลองที่ใกล้เคียงกับคอนทัวร์ F_0 ขาเข้ามากที่สุด

งานวิจัยนี้ได้เลือกใช้วิธีตามที่กล่าวข้างต้น ถึงแค่เพียงการแยกองค์ประกอบของคอนทัวร์ F_0 ออกเป็นองค์ประกอบความถี่สูง และองค์ประกอบความถี่ต่ำเท่านั้น เนื่องจากลักษณะขององค์ประกอบทั้งสอง สามารถบ่งบอกได้ถึงลักษณะของทำนองเสียงแล้ว โดยไม่จำเป็นต้องหาค่าพารามิเตอร์ฟูจิซากิ เนื่องจากการหาค่าพารามิเตอร์ฟูจิซากิแบบอัตโนมัติ ใช้วิธีที่ค่อนข้างซับซ้อน ตัวอย่างเช่น ขั้นตอนการค้นหาแบบปีนเขา ดังที่กล่าวข้างต้น นอกจากนี้ ในปัจจุบันยังไม่มีงานวิจัยที่ศึกษาความสัมพันธ์ระหว่างพารามิเตอร์ฟูจิซากิ และทำนองเสียงพูดภาษาไทยประเภทอื่น ที่ไม่ใช่ทำนองเสียงตก

หัวข้อนี้กล่าวถึงวิธีการหาคอนทัวร์ลักษณะ 2 เส้น เส้นหนึ่ง คือ คอนทัวร์ LFC ซึ่งสามารถบอกได้ถึงลักษณะการลดระดับ หรือการสูงขึ้นของ F_0 เนื่องจากทำนองเสียงได้ คอนทัวร์ LFC ที่ใช้ในงานวิจัยนี้คล้ายกับคอนทัวร์ LFC ของ Mixdorff แต่ได้ตัดกระบวนการบางอย่างที่ไม่จำเป็นต่อการรู้จำทำนองเสียงพูดออกไป เช่น ตัดขั้นตอนการทำคอนทัวร์ F_0 ให้เรียบโดยใช้ วิธีการประมาณค่าในช่วงแบบกำลังสาม (cubic interpolation) ส่วนคอนทัวร์ลักษณะอีกเส้นหนึ่ง คือ คอนทัวร์ FVC

(F_0 variation contour) ซึ่งงานวิจัยนี้ได้นำเสนอขึ้น เป็นคอนทัวร์ที่แสดงถึงความมากน้อยในการแกว่งของคอนทัวร์ F_0 ซึ่งสามารถบอกถึงลักษณะของทำนองเสียงได้เช่นกัน นอกจากนี้งานวิจัยนี้ใช้คอนทัวร์ F_0 ในสเกลเชิงเส้น (linear) แทนที่จะเป็นสเกลลอการิทึม เนื่องจากไม่ได้ต้องการหาองค์ประกอบวลี และองค์ประกอบวรรณยุกต์ที่สอดคล้องกับสมการที่ 2.2 – 2.4 แต่ต้องการเพียงพิจารณาแนวโน้มการลดระดับ หรือเพิ่มระดับ และต้องการพิจารณาความมากน้อยของลักษณะการแกว่งของคอนทัวร์ F_0 เพื่อใช้ในการรู้จำทำนองเสียงพูดเท่านั้น จากเหตุผลดังกล่าว งานวิจัยนี้จึงได้นำเสนอขั้นตอนวิธีการในการสกัดลักษณะทางทำนองเสียงจากคอนทัวร์ F_0 ดังนี้

3.2.1 การหาคอนทัวร์ F_0 จากสัญญาณเสียงพูด

งานวิจัยนี้เลือกใช้วิธีการในการหาคอนทัวร์ F_0 ที่นำเสนอโดย Boersma (1993) ซึ่งได้นำเสนออัลกอริทึมสำหรับตรวจจับความเป็นรายคาบของสัญญาณ รวมทั้งการหาคอนทัวร์ F_0 โดยวิธีอัตสหสัมพันธ์ Boersma ได้พิสูจน์ให้เห็นว่าอัลกอริทึมนี้ให้ความถูกต้องมากกว่าอัลกอริทึม อื่น ๆ ที่ใช้ในการหาคอนทัวร์ F_0 ของสัญญาณเสียงพูด นอกจากนี้ Boersma ยังได้นำ อัลกอริทึมนี้ไปใส่ไว้ในโปรแกรม Praat ซึ่งเป็นโปรแกรมฟรี ที่ใช้ในการวิเคราะห์และสังเคราะห์เสียงพูด ซึ่งสามารถดาวน์โหลดได้จาก http://www.fon.hum.uva.nl/praat/

ในการหาคอนทัวร์ F_0 โดยใช้โปรแกรม Praat ของงานวิจัยนี้ ใช้ค่าพารามิเตอร์ต่าง ๆ ตามค่าดีฟอลต์ของโปรแกรม นั่นคือ ใช้ค่าขั้นของเวลา (time step) ขั้นละ 10 มิลลิวินาที และกำหนดให้ค่า F_0 ต่ำสุด และสูงสุดที่เป็นไปได้คือ 75 Hz และ 600 Hz ตามลำดับ (Boersma, 1993)

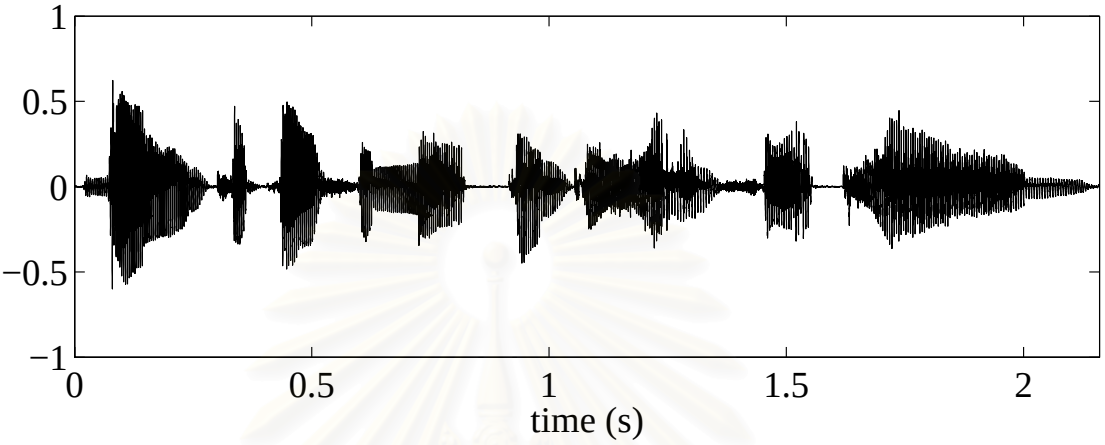
ลักษณะของคอนทัวร์ F_0 ที่หาโดยใช้โปรแกรม Praat แสดงดังรูปที่ 3.9

3.2.2 การทำให้คอนทัวร์ F_0 เรียบ

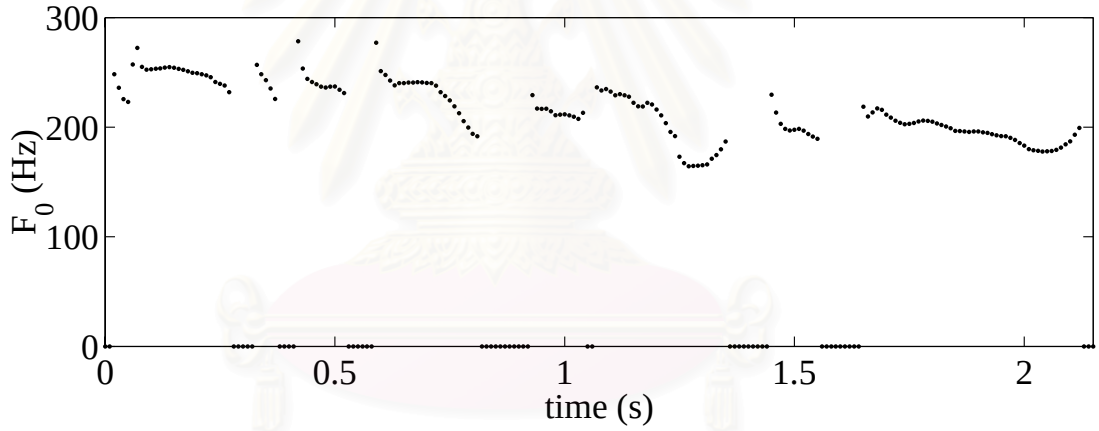
จากรูปที่ 3.9 จะเห็นได้ว่าคอนทัวร์ F_0 ที่หาได้ มีบางช่วงที่ไม่เรียบ ซึ่งเป็นผลมาจากกลไกในการให้กำเนิดเสียงพูด ทำให้สัญญาณพัลส์ที่ได้จากเส้นเสียงเกิดความไม่สม่ำเสมอที่เรียกว่าปรากฏการณ์ไมโครโพรโซดิก (microprosodic effect)

วิธีที่นิยมใช้ลดความผิดพลาดดังกล่าวที่เป็นที่นิยมคือ การนำคอนทัวร์ F_0 ไปผ่านตัวกรองมัธยฐาน (median filter) (Rabiner และ Schafer, 1978) จากการทดสอบกับคอนทัวร์ F_0 ที่ใช้ในงานวิจัยนี้ พบว่า การใช้ตัวกรองมัธยฐานขนาด 5 จุด (นำเอาค่า มัธยฐานของ F_0 ของF_0 ของเฟรมที่ต้องการหารวมทั้งเฟรมที่อยู่ข้างเคียง 4 เฟรม มาแทนที่ค่า F_0 ของเฟรมนั้น ๆ) สามารถกำจัดความไม่สม่ำเสมอของคอนทัวร์ F_0 ได้ดี ตัวอย่างของคอนทัวร์ F_0 หลังจากผ่านตัวกรองมัธยฐานมีลักษณะดังรูปที่ 3.10 โดยจะเห็นได้ว่าคอนทัวร์ F_0 เรียบขึ้นเมื่อเทียบกับรูปที่ 3.9

หลังจากนั้นจะประมาณค่า F_0 ในช่วงที่เป็นเสียงไม่ก้องด้วยเส้นตรง เพื่อให้คอนทัวร์ F_0 มีความต่อเนื่องกันทั้งประโยค และสามารถนำไปผ่านตัวกรองได้ โดยเรียกคอนทัวร์ที่ได้จากกระบวนการนี้ว่า CF_0 (connected F_0) ดังที่แสดงในรูปที่ 3.11

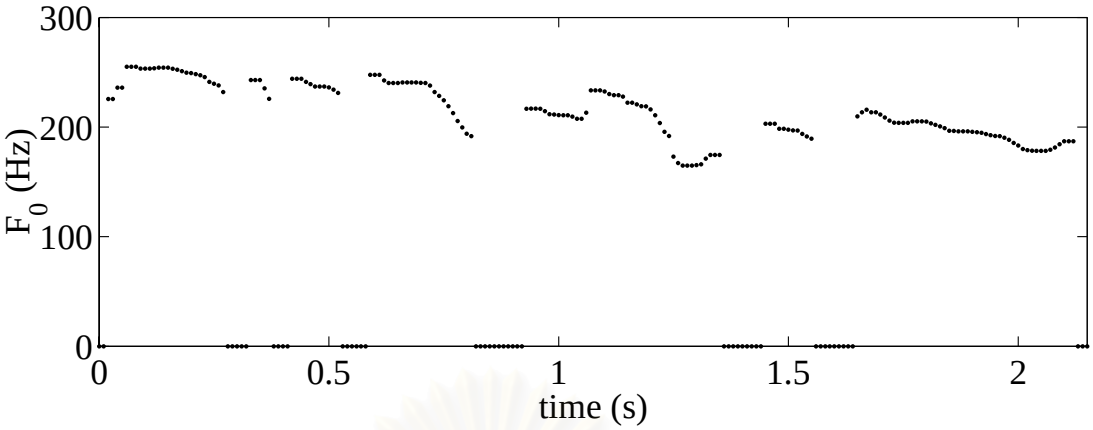


(ก)

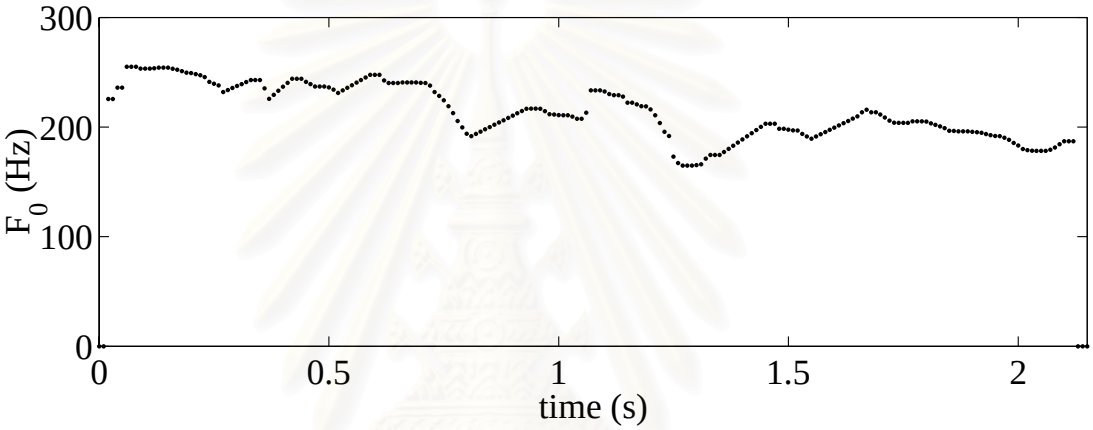


(ข)

รูปที่ 3.9 สัญญาณเสียงพูดที่ต้องการนำมาหาคอนทัวร์ลักษณะ:
(ก) รูปคลื่นของสัญญาณเสียงพูด (ข) คอนทัวร์ F_0 ที่หาโดยใช้โปรแกรม Praat



รูปที่ 3.10 คอนทัวร์ F_0 หลังจากผ่านตัวกรองมัธยฐาน

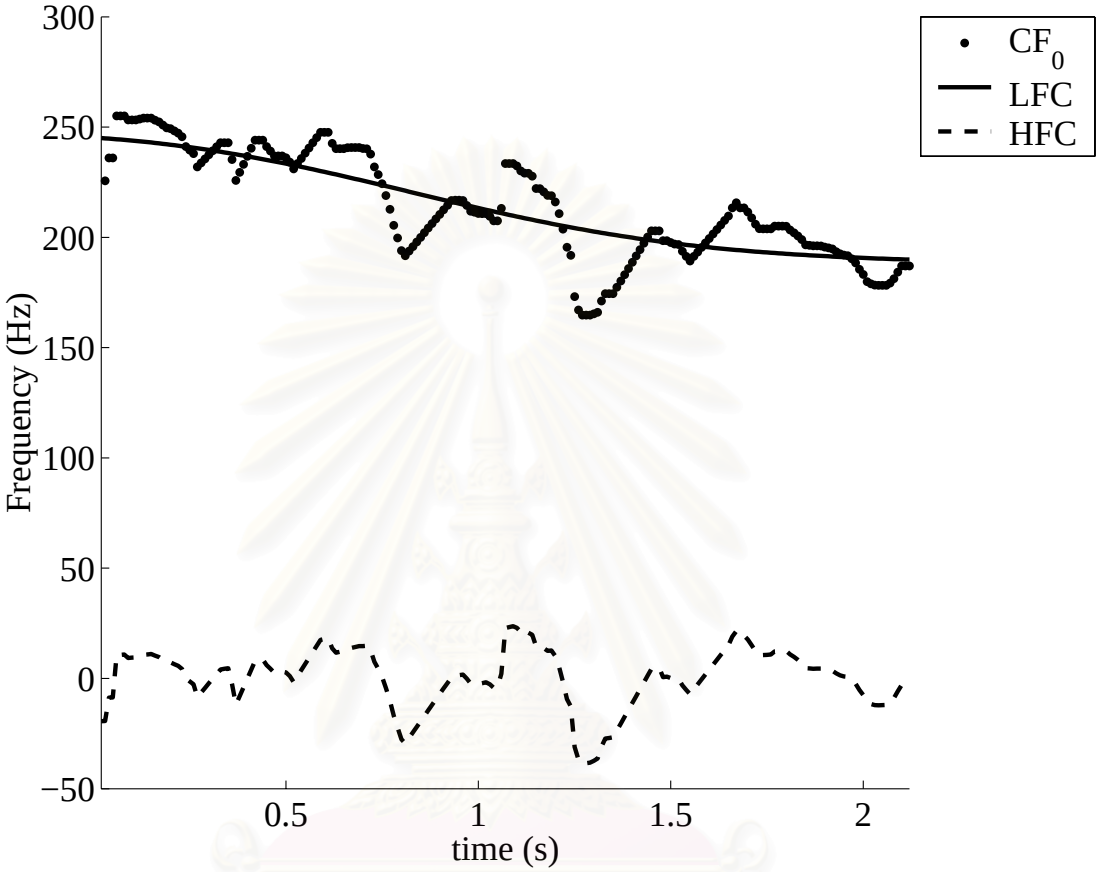


รูปที่ 3.11 คอนทัวร์ F_0 หลังจากผ่านการกำจัดช่วงที่เป็นเสียงไม่ก้อง (คอนทัวร์ CF_0)

3.2.3 การหาคอนทัวร์ LFC

สามารถหาคอนทัวร์ LFC ได้โดยการพิจารณาคอนทัวร์ CF_0 ว่าเป็นสัญญาณดิสครีตทางเวลา ในลักษณะเดียวกับ Mixdorff (2000) โดยพิจารณาว่าองค์ประกอบวรรณยุกต์เป็นองค์ประกอบที่มีการเปลี่ยนแปลงอย่างรวดเร็ว เมื่อเทียบกับองค์ประกอบวลี จึงสามารถกำจัดองค์ประกอบวรรณยุกต์ออกไปได้โดยการใช้ตัวกรองผ่านต่ำ (low-pass filter) ซึ่งจะยอมให้สัญญาณในส่วนที่มีการเปลี่ยนแปลงอย่างช้า ๆ เท่านั้นที่สามารถผ่านไปได้ งานวิจัยนี้ได้เลือกใช้ตัวกรองแบบ FIR (finite impulse response) เนื่องจากจะทำให้สัญญาณขาออกมีการเลื่อนเฟสแบบเชิงเส้น จึงสามารถชดเชยผลของการเลื่อนเฟสของสัญญาณขาออกได้โดยง่าย โดยกำหนดให้สัญญาณขาเข้าที่ช่วงเวลาก่อนเริ่มคอนทัวร์ CF_0 และสัญญาณขาเข้าที่ช่วงเวลาหลังจากคอนทัวร์ CF_0 มีค่าเท่ากับค่าเฉลี่ยของคอนทัวร์ CF_0 เพื่อให้รูปร่างของสัญญาณขาออกไม่เพี้ยนที่ปลายทั้งสองข้าง โดยเรียกสัญญาณขาออก นี้ว่าคอนทัวร์ LFC ตามที่ได้กล่าวไปแล้วโดยกำหนดให้ Fc_{LFC} คือค่าความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC ตัวอย่างของคอนทัวร์ LFC แสดงดังรูปที่ 3.12 (เส้นทึบ)

ในการทดลองหาค่า Fc_{LFC} ที่ให้อัตราการรู้จำทำนองเสียงพูดสูงที่สุด นอกจากจะใช้คอนทัวร์ LFC ที่ได้จากตัวกรองผ่านต่ำ งานวิจัยนี้ยังได้ใช้ LFC ที่เป็นเส้นตรงด้วย โดยเลือกใช้เส้นตรง 2 แบบ คือ เส้นตรงที่มีความชันเป็น 0 (เส้นแนวระดับ) และเส้นตรงที่มีความชัน โดยสามารถหาสมการเส้นตรงได้โดยใช้สมการถดถอย (regression equation) กับคอนทัวร์ CF_0



รูปที่ 3.12 คอนทัวร์ LFC และ HFC ที่หาได้จากคอนทัวร์ CF_0

3.2.4 การหาคอนทัวร์ FVC

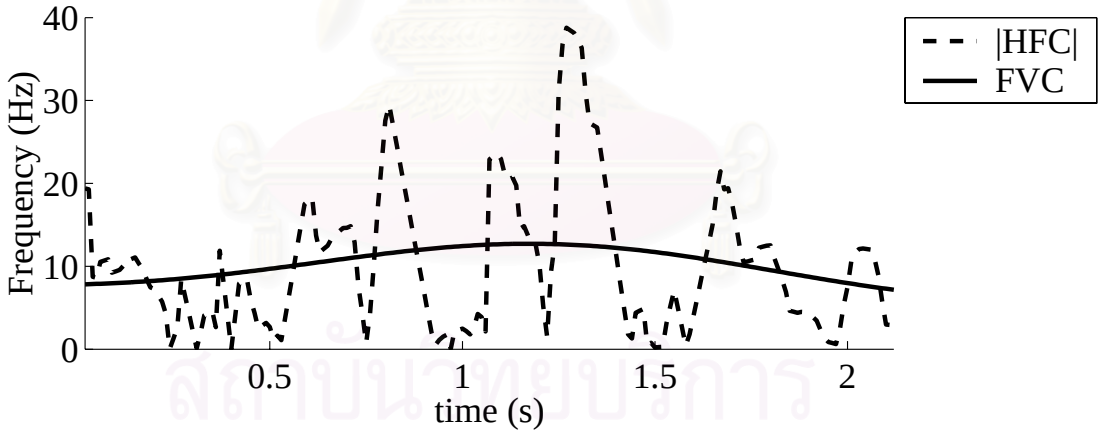
จากที่กล่าวไปแล้วข้างต้นว่า องค์ประกอบเสียงวรรณยุกต์มีลักษณะที่แสดงให้เห็นถึงประเภทของทำนองเสียงได้ นั่นก็คือช่วงกว้างของการเปลี่ยนแปลงค่าของคอนทัวร์ F_0 เราจึงไม่สามารถทิ้งองค์ประกอบเสียงวรรณยุกต์ไปได้

องค์ประกอบเสียงวรรณยุกต์เป็นองค์ประกอบที่มีการเปลี่ยนแปลงอย่างรวดเร็ว เราจึงสามารถสกัดเอาองค์ประกอบเสียงวรรณยุกต์จากคอนทัวร์ CF_0 ได้โดยใช้ตัวกรองผ่านสูง (high-pass filter) แต่เนื่องจากเรามีองค์ประกอบความถี่ต่ำ คือคอนทัวร์ LFC อยู่แล้ว เราจึงสามารถหาองค์ประกอบความถี่สูงได้โดยการนำคอนทัวร์ LFC ไปลบออกจากคอนทัวร์ CF_0 โดยเรียกคอนทัวร์ที่ได้ว่า คอนทัวร์ HFC (high frequency contour) แสดงดังรูปที่ 3.12 (เส้นประ)

สิ่งที่แสดงให้เห็นถึงลักษณะของทำนองเสียงในคอนทัวร์ HFC คือ ช่วงกว้างในการแกว่งของคอนทัวร์ HFC โดยไม่ต้องคำนึงถึงว่า เป็นการแกว่งขึ้น (HFC > 0) หรือการแกว่งลง (HFC < 0) จึงนำ HFC ไปใส่ค่าสัมบูรณ์ ลักษณะของ |HFC| แสดงดังรูปที่ 3.13

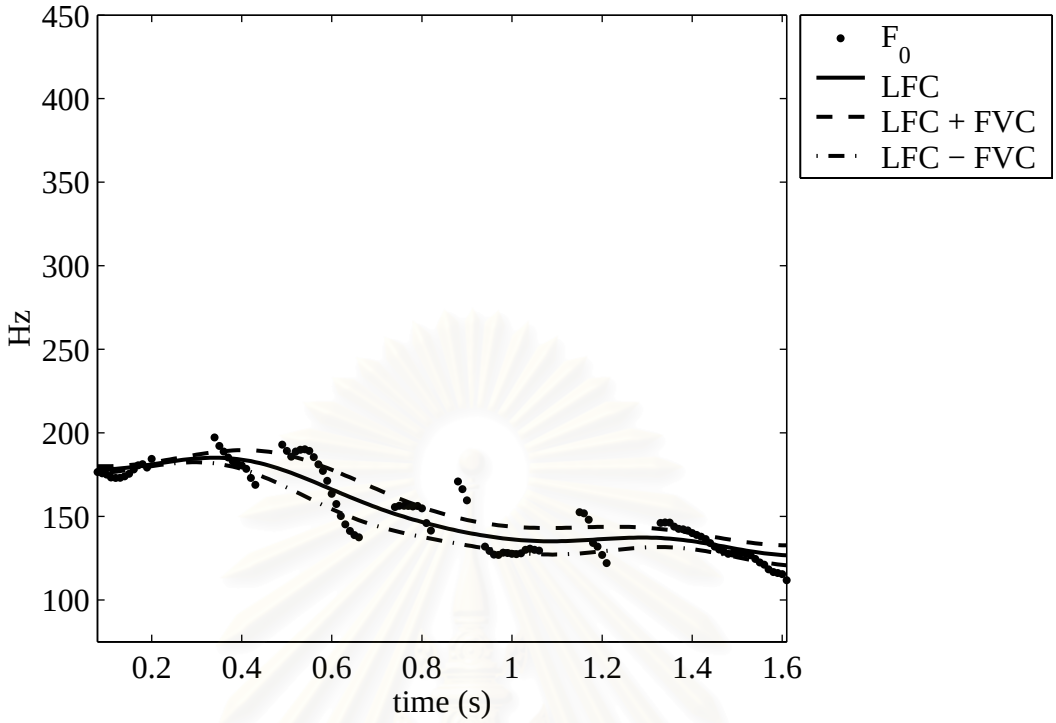
จากรูปที่ 3.13 จะเห็นว่า คอนทัวร์ |HFC| นั้นไม่เรียบ ซึ่งเป็นผลจากเสียงวรรณยุกต์ของแต่ละพยางค์ ที่ให้สารสนเทศภาษาศาสตร์ งานวิจัยนี้จึงได้นำ |HFC| ไปผ่านตัวกรองผ่านต่ำแบบ FIR แบบเดียวกับที่ใช้ในการหาคอนทัวร์ LFC เพื่อกำจัดสารสนเทศทางภาษาศาสตร์ โดยจะได้ว่าสัญญาณขาออกของตัวกรอง แสดงให้เห็นถึงความมากน้อยในการแกว่งของ HFC ซึ่งก็คือความมากน้อยในการกวัดแกว่ง หรือการเปลี่ยนแปลงค่าของคอนทัวร์ F_0 นั่นเอง จึงเรียกสัญญาณขาออกนี้ว่า คอนทัวร์ FVC (F_0 variation contour) ดังแสดงในรูปที่ 3.13 โดยกำหนดให้ความถี่ตัดของตัวกรองผ่านต่ำ ที่ใช้หาคอนทัวร์ FVC มีค่าเป็น Fc_{LFC}

คอนทัวร์ FVC เป็นคอนทัวร์ที่แสดงให้เห็นถึงความมากน้อยในการเปลี่ยนแปลงค่าของ F_0 ถ้าคอนทัวร์ FVC มีระดับที่สูง แสดงว่าคอนทัวร์ F_0 มีช่วงการแกว่งที่กว้าง ซึ่งจะพบในทำนองเสียงทำนองเสียงพูดแบบผสม แต่ถ้าคอนทัวร์ FVC มีระดับที่ต่ำ แสดงว่าคอนทัวร์ F_0 มีช่วงการแกว่งที่แคบ ซึ่งมักจะพบได้ในทำนองเสียงพูดแบบตก หรือทำนองเสียงพูดแบบขึ้น

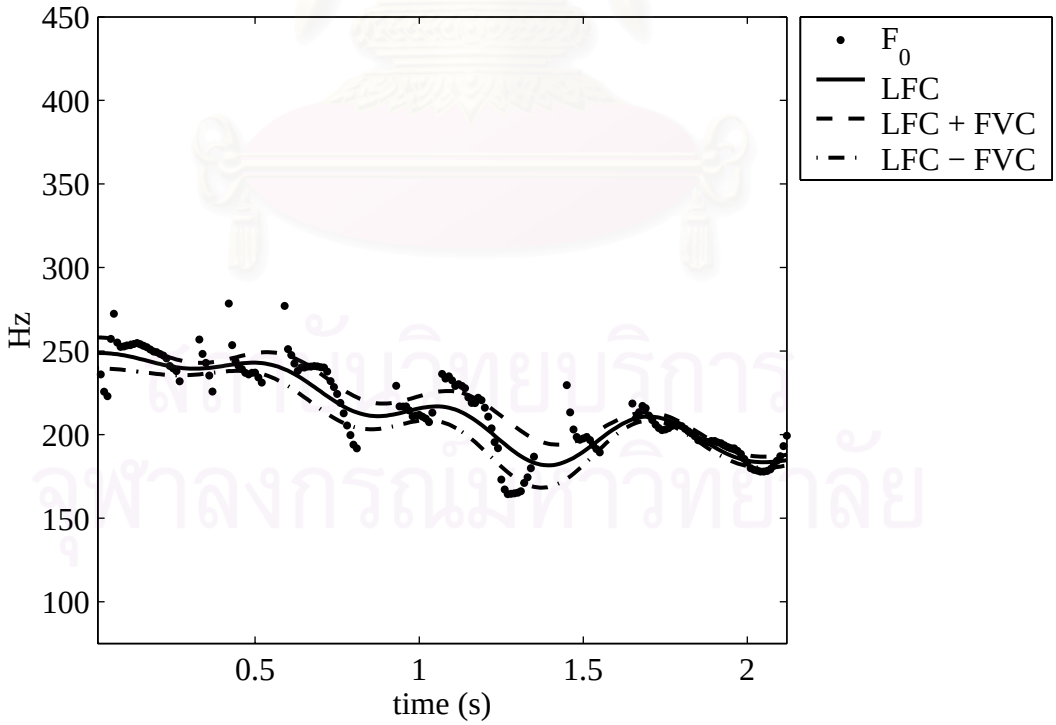


รูปที่ 3.13 คอนทัวร์ FVC ที่หาได้จากการนำ |HFC| ไปผ่านตัวกรองผ่านต่ำ

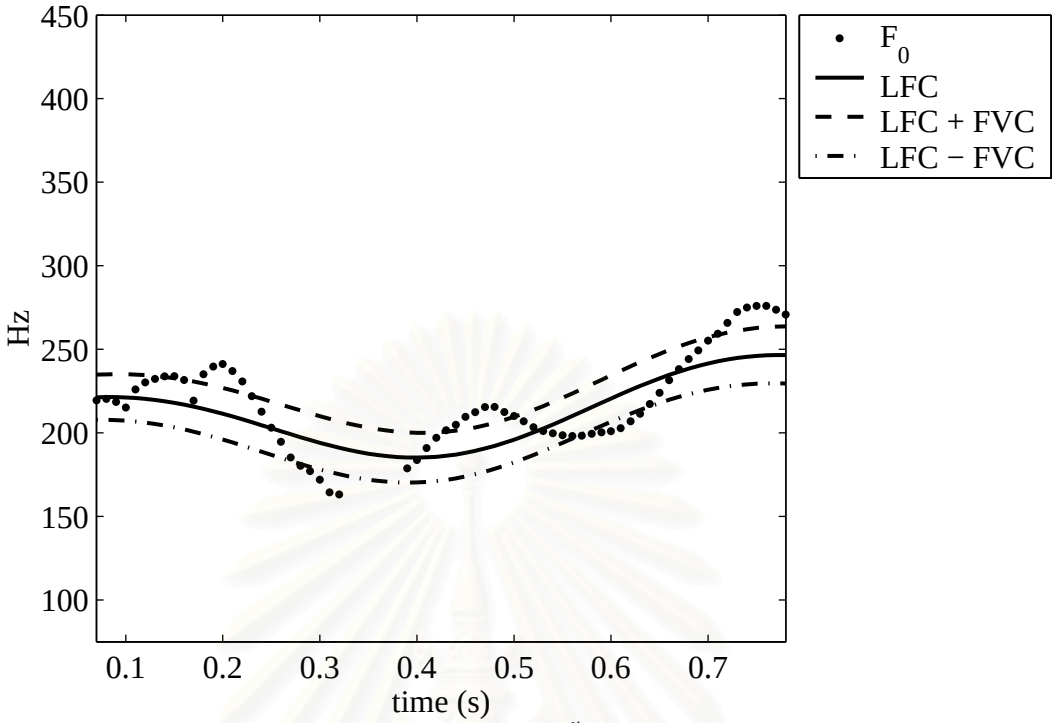
คอนทัวร์ที่นำไปเป็นคอนทัวร์ลักษณะเพื่อใช้ในการรู้จำทำนองเสียงพูด คือคอนทัวร์ LFC และคอนทัวร์ FVC และเพื่อให้ง่ายต่อการแสดงให้เห็นถึงความสัมพันธ์ระหว่าง คอนทัวร์ F_0 คอนทัวร์ LFC และคอนทัวร์ FVC จึงได้เขียนกราฟของคอนทัวร์ F_0 คอนทัวร์ LFC คอนทัวร์ LFC + FVC และคอนทัวร์ LFC – FVC ของทำนองเสียงพูดแบบต่าง ๆ ดังแสดงในรูปที่ 3.14 ถึง 3.21



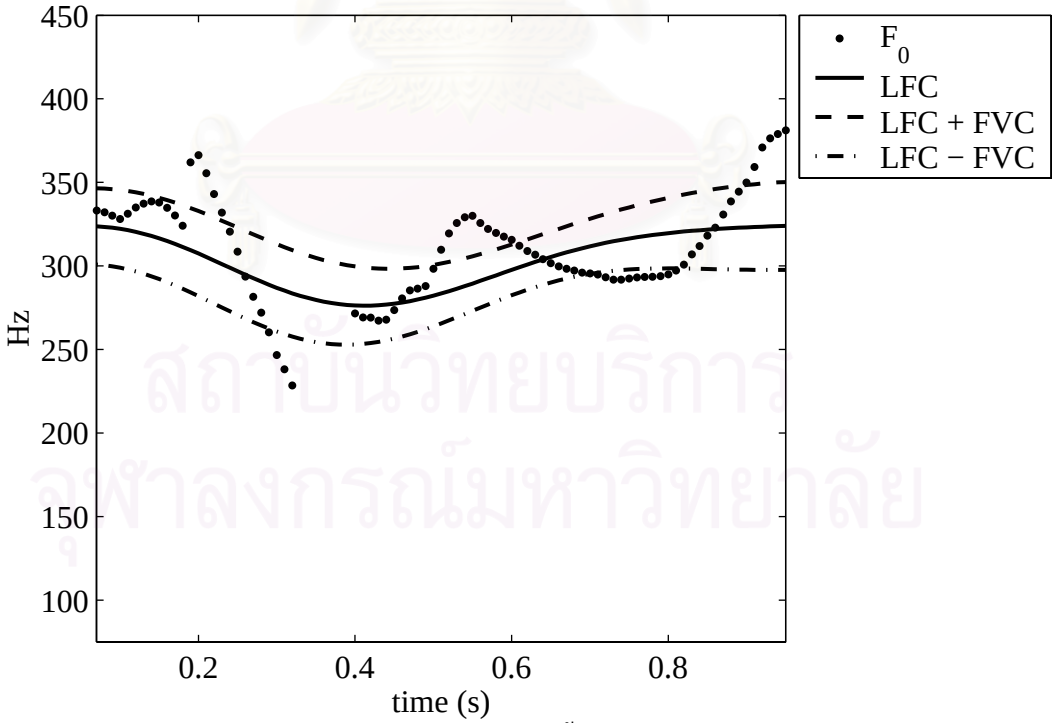
รูปที่ 3.14 คอนทัวร์ F_0 ในรูปที่ 3.1 (ทำนองเสียงตก, ผู้พูดเป็นผู้ชาย) และ คอนทัวร์ลักษณะ



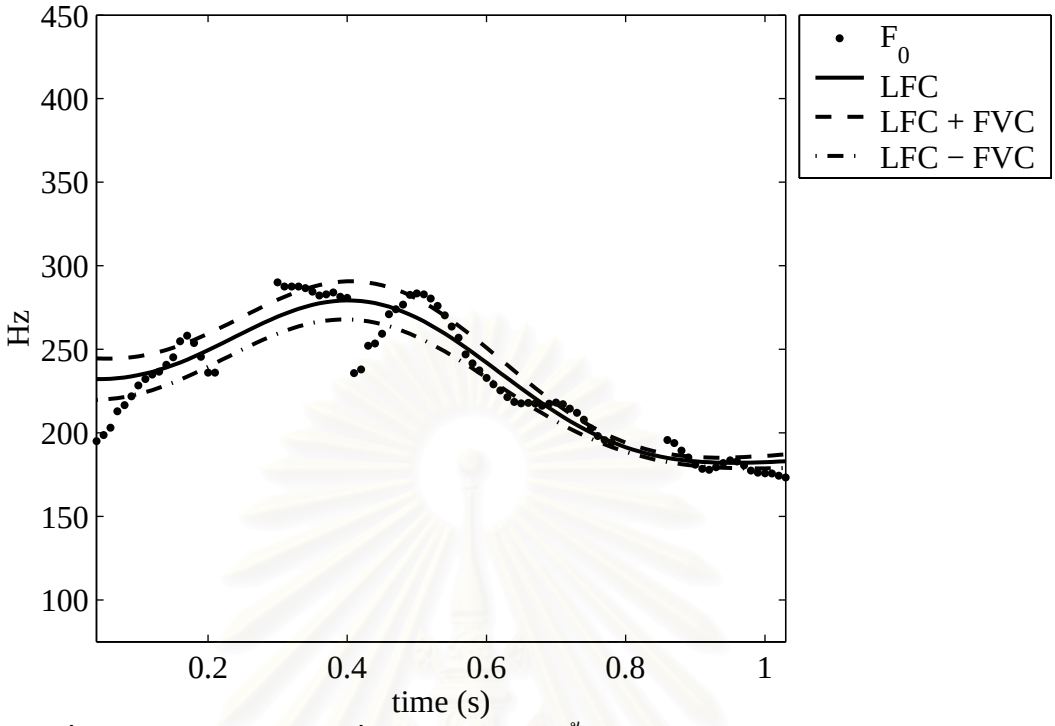
รูปที่ 3.15 คอนทัวร์ F_0 ในรูปที่ 3.2 (ทำนองเสียงตก, ผู้พูดเป็นผู้หญิง) และ คอนทัวร์ลักษณะ



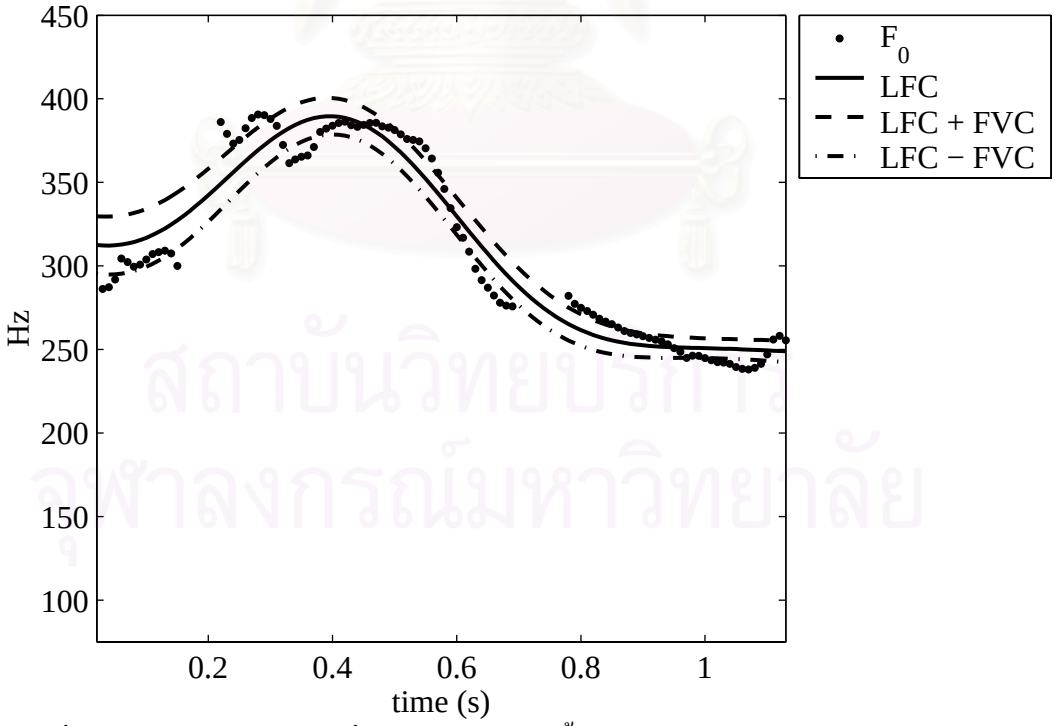
รูปที่ 3.16 คอนทัวร์ F_0 ในรูปที่ 3.3 (ทำนองเสียงขึ้น, ผู้พูดเป็นผู้ชาย) และ คอนทัวร์ลักษณะ



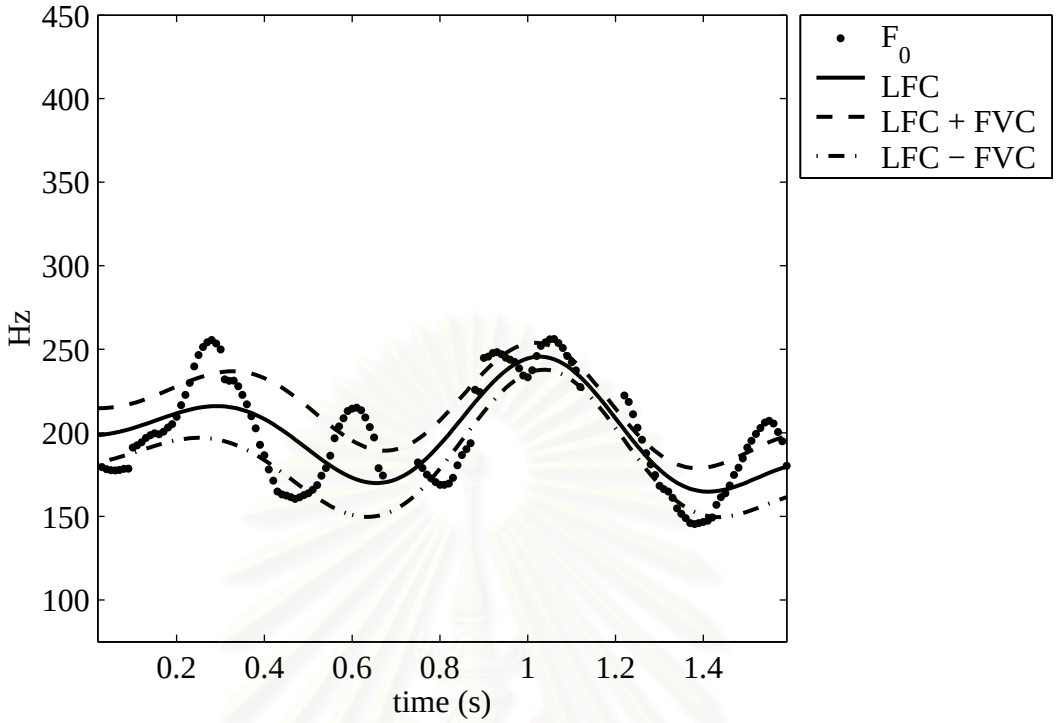
รูปที่ 3.17 คอนทัวร์ F_0 ในรูปที่ 3.4 (ทำนองเสียงขึ้น, ผู้พูดเป็นผู้หญิง) และ คอนทัวร์ลักษณะ



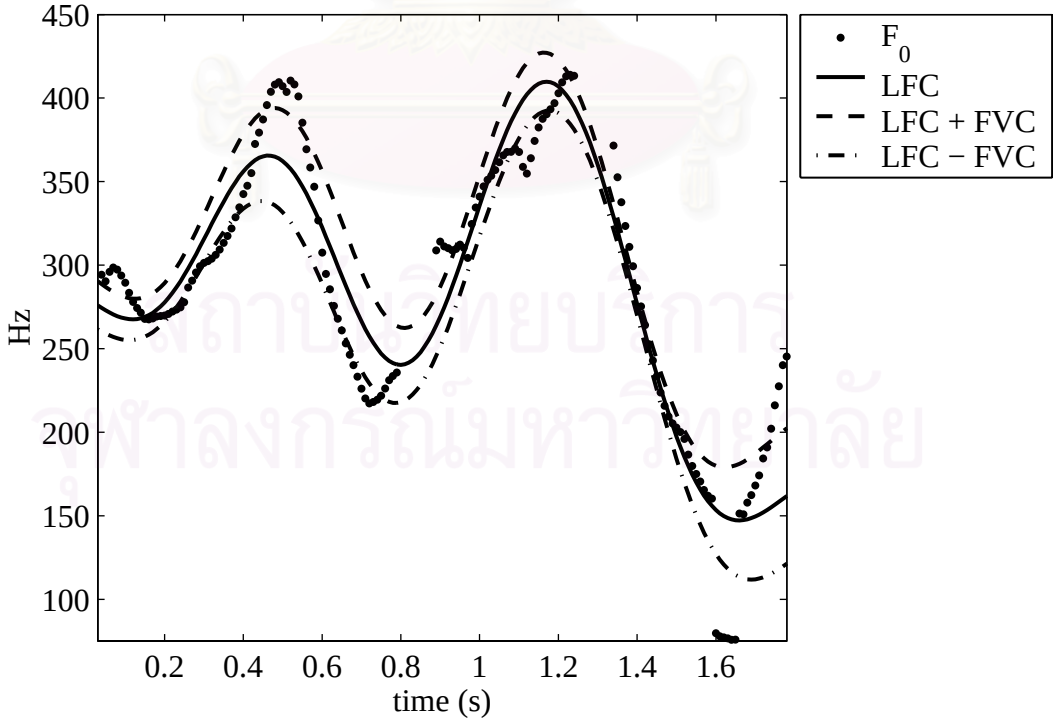
รูปที่ 3.18 คอนทัวร์ F_0 ในรูปที่ 3.5 (ทำนองเสียงขึ้น, ผู้พูดเป็นผู้ชาย) และ คอนทัวร์ลักษณะ



รูปที่ 3.19 คอนทัวร์ F_0 ในรูปที่ 3.6 (ทำนองเสียงขึ้น, ผู้พูดเป็นผู้หญิง) และ คอนทัวร์ลักษณะ



รูปที่ 3.20 คอนทัวร์ F_0 ในรูปที่ 3.7 (ทำนองเสียงผสม, ผู้พูดเป็นผู้ชาย) และคอนทัวร์ลักษณะ



รูปที่ 3.21 คอนทัวร์ F_0 ในรูปที่ 3.8 (ทำนองเสียงผสม, ผู้พูดเป็นผู้หญิง) และคอนทัวร์ลักษณะ

3.3 ระบบรู้จำทำนองเสียงพูด

3.3.1 โครงสร้างของระบบรู้จำทำนองเสียงพูด

ระบบรู้จำทำนองเสียงพูดภาษาไทยที่ออกแบบขึ้นเพื่อใช้ในการทดลองแสดงในรูปที่ 3.22 ระบบจะนำเสียงพูดที่ต้องการรู้จำทำนองเสียงไปหาค่า F_0 ในส่วน F_0 Extraction แล้วจึงนำคอนทัวร์ F_0 ที่ได้ไปหาคอนทัวร์ลักษณะ LFC และ FVC ในส่วน Feature Contours Extraction ตามวิธีที่แสดงในหัวข้อ 3.2

จากนั้นจึงสุ่มตัวอย่าง (sampling) คอนทัวร์ลักษณะทั้งสอง โดยกำหนดให้จำนวนจุดที่ใช้สุ่มตัวอย่างขึ้นกับค่าความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC และ FVC ตามทฤษฎีการสุ่มตัวอย่างของไนควิสต์ (Nyquist's sampling theory) คือสุ่มด้วยอัตรา 2 เท่าของความกว้างของช่วงความถี่ (bandwidth) ตัวอย่างเช่น ในการทดลองหนึ่งหาคอนทัวร์ LFC โดยใช้ค่าความถี่ตัดเป็น 2.5 Hz ตามทฤษฎีไนควิสต์ต้องสุ่มตัวอย่างคอนทัวร์ LFC ด้วยอัตราสุ่มอย่างต่ำ 5 Hz หรือ 5 จุดต่อวินาที และเนื่องจากค่าความยาวสูงสุดของประโยคเสียงพูดในฐานข้อมูลที่ใช้ในการทดลองคือ 3 วินาที ดังนั้นในการทดลองนี้จึงสุ่มตัวอย่าง LFC ของทุกประโยคด้วยจำนวน 15 จุด ถึงแม้ว่าประโยคนั้นจะมีความยาวน้อยกว่า 3 วินาทีก็ตาม เนื่องจากจะนำจุดที่สุ่มได้ไปใช้เป็นเวกเตอร์ลักษณะของโครงข่ายประสาทเทียม ซึ่งการฝึกฝนและทดสอบในแต่ละการทดลองจำเป็นต้องใช้เวกเตอร์ลักษณะที่มีมิติเท่ากัน

ค่าที่นำมาใช้เป็นเวกเตอร์ลักษณะ ได้จากค่าของ LFC และ FVC ที่จุดที่สุ่มตัวอย่าง และในการทดลองในหัวข้อ 4.32, 4.34, 4.42 และ 4.44 ก็ใช้ค่าผลต่างอันดับหนึ่ง (first difference) ที่จุดสุ่มตัวอย่างด้วย เนื่องจากสามารถแสดงถึงการเปลี่ยนแปลงของ LFC ได้ นิยามค่าผลต่างอันดับหนึ่งของ LFC หรือ ΔLFC มีค่าเป็น

$$\Delta LFC(k) = LFC(k+1) - LFC(k) \tag{3.1}$$

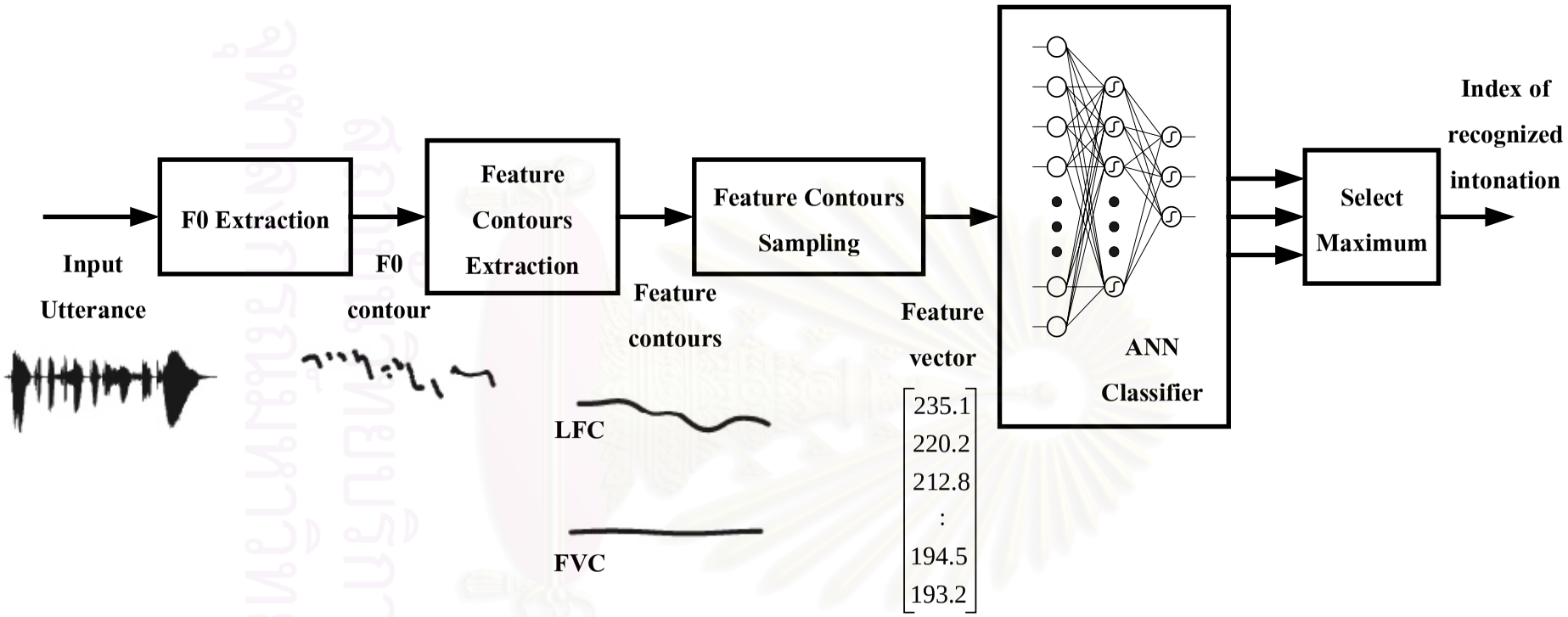
เมื่อ k คือ ดัชนี (index) ของแต่ละเฟรม ($k = 0, 1, 2, \ldots$)

ตัวอย่างของคอนทัวร์ LFC และ FVC ที่แปรค่าความถี่ตัดเป็นค่าต่าง ๆ รวมทั้งการสุ่มตัวอย่างคอนทัวร์ทั้งสองเพื่อนำมาใช้เป็นเวกเตอร์ลักษณะ แสดงดังรูปที่ 3.23 – 3.29 ประโยคที่นำมาใช้เป็นตัวอย่างนี้ คือประโยคที่ยาวประมาณ 3 วินาที ซึ่งยาวที่สุดในชุดข้อมูล เป็นประโยคที่มีทำนองเสียงแบบผสม ของผู้พูดที่เป็นผู้หญิง (ในรูปตัวอย่าง จะแสดงเฉพาะกรณีที่ความถี่ตัดของตัว

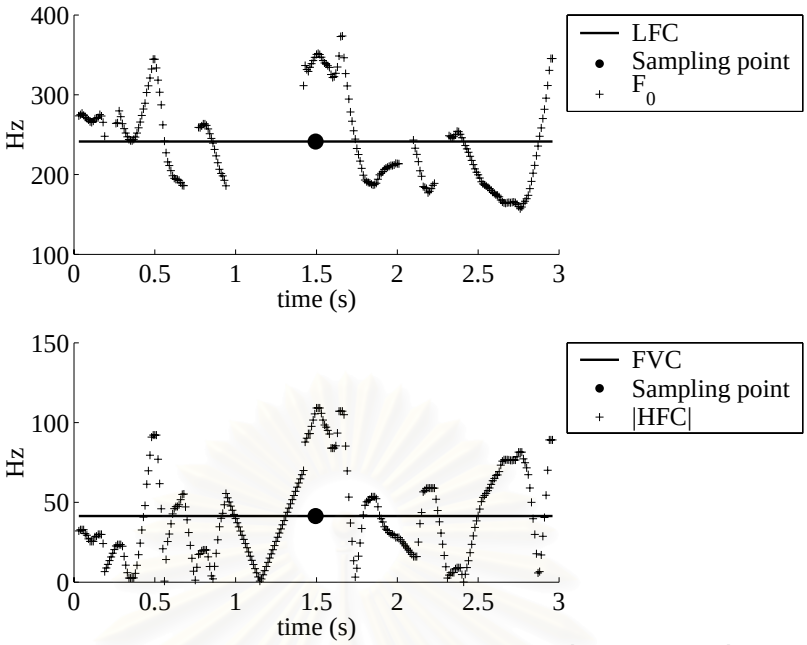
กรองในการหา LFC และ FVC มีค่าเท่ากัน แต่ในการทดลองจะจับคู่ค่าความถี่ตัดของ LFC และ FVC ให้ครบทุกแบบ เท่าที่จะเป็นไปได้ในช่วงที่กำหนด)

จากรูปที่ 3.23 – 3.29 จะเห็นได้ว่า เมื่อเพิ่มค่า Fc_{LFC} จะทำให้คอนทัวร์ LFC มีความใกล้เคียงกับคอนทัวร์ F_0 มากยิ่งขึ้น โดยเฉพาะอย่างยิ่งในรูปที่ 3.29 กรณีที่ Fc_{LFC} = 4.5 Hz จะเห็นได้ว่าสามารถพิจารณาคอนทัวร์ LFC ว่าเป็นคอนทัวร์ F_0 ที่ผ่านกระบวนการทำให้เรียบ (smoothing) ได้ การสุ่มตัวอย่างคอนทัวร์ LFC ที่มีค่าความถี่ตัดสูง ๆ จึงให้เวกเตอร์ลักษณะที่ใกล้เคียงกับการสุ่มตัวอย่างคอนทัวร์ F_0 โดยตรง ในบทที่ 4 จะมีการทดลองที่ใช้เพียงคอนทัวร์ LFC ในการรู้จำ ซึ่งการทดลองในกรณีนี้ที่ค่า Fc_{LFC} สูง ๆ สามารถประมาณได้ว่าผลการทดลองเหมือนกับกรณีที่สุ่มตัวอย่างจากคอนทัวร์ F_0 โดยตรงได้ ซึ่งช่วยให้สามารถเปรียบเทียบผลการรู้จำทำนองเสียงพูดระหว่างกรณีที่ใช้คอนทัวร์ F_0 โดยตรง กับกรณีที่ใช้คอนทัวร์ LFC และ FVC ได้

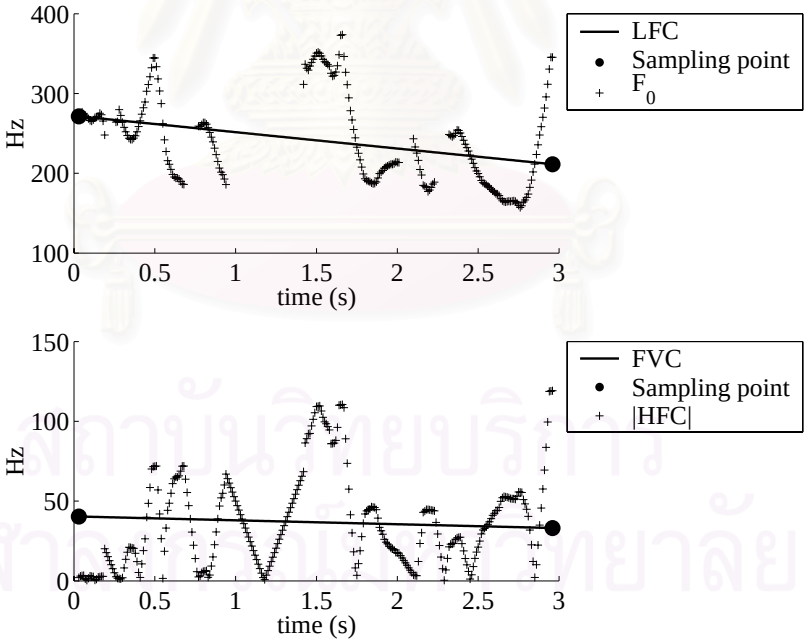
และจะเห็นได้ว่าการเพิ่ม Fc_{FVC} ก็จะทำให้คอนทัวร์ FVC มีรูปร่างใกล้เคียงกับ |HFC| มากขึ้นเช่นกัน ในขณะเดียวกันค่าของ Fc_{LFC} ก็ยังส่งผลต่อรูปร่างของคอนทัวร์ FVC ด้วย เนื่องจากคอนทัวร์ HFC ที่ใช้ในการหาคอนทัวร์ FVC มีรูปร่างที่ขึ้นกับ คอนทัวร์ LFC มาก ในการทดลองต่าง ๆ ในบทที่ 4 จึงต้องทดลองโดยจับคู่ Fc_{LFC} และ Fc_{FVC} ให้ครบทุกแบบ เท่าที่เป็นไปได้ ในช่วงค่าที่กำหนด



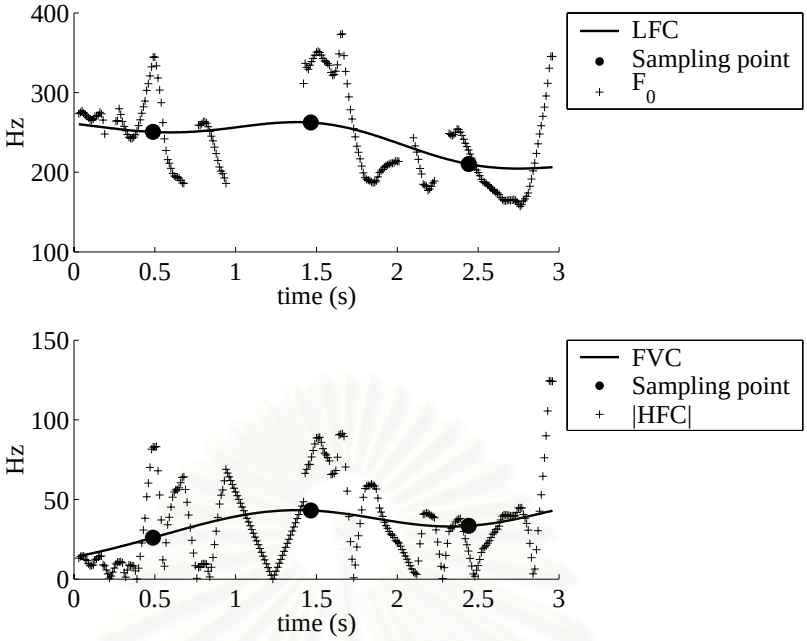
รูปที่ 3.22 แผนภาพของระบบรู้จำทำนองเสียงพูด



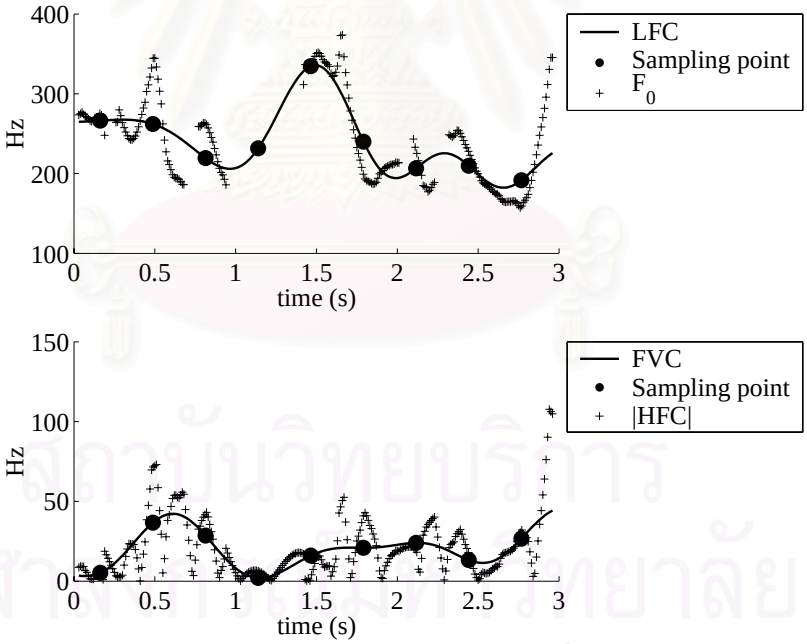
รูปที่ 3.23 ตัวอย่างคอนทัวร์ LFC และ FVC: คอนทัวร์ LFC ที่เป็นเส้นตรงที่มีความชันเป็น 0 เปรียบเทียบกับ คอนทัวร์ F_0 (บน) คอนทัวร์ FVC ที่เป็นเส้นตรงที่มีความชันเป็น 0 เปรียบเทียบกับคอนทัวร์ |HFC| (ล่าง)



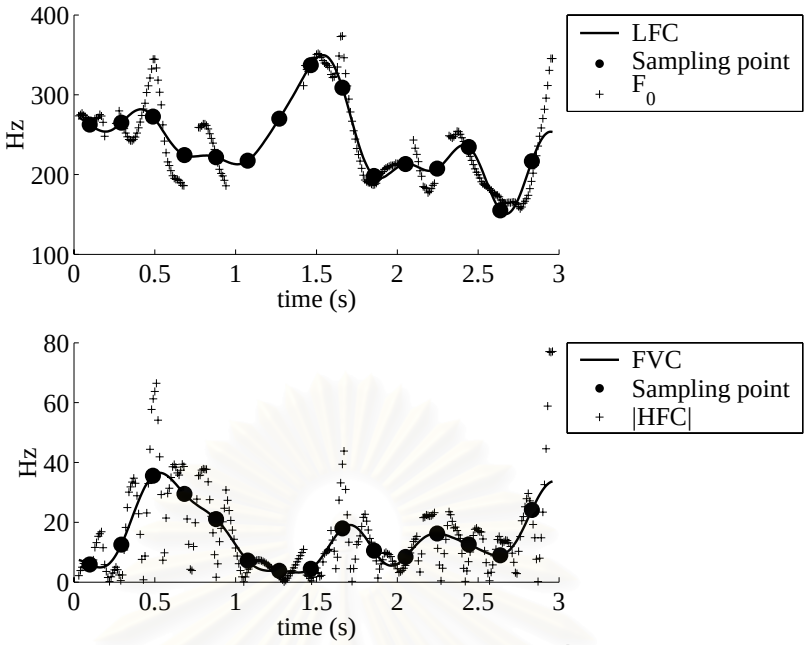
รูปที่ 3.24 ตัวอย่างคอนทัวร์ LFC และ FVC: คอนทัวร์ LFC ที่เป็นเส้นตรง เปรียบเทียบกับ คอนทัวร์ F_0 (บน) คอนทัวร์ FVC ที่เป็นเส้นตรง เปรียบเทียบกับคอนทัวร์ |HFC| (ล่าง)



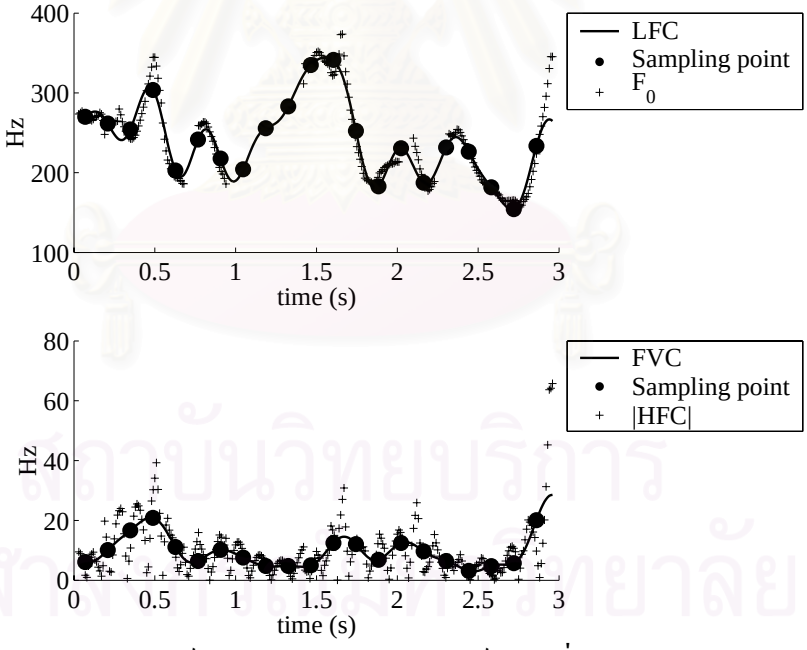
รูปที่ 3.25 ตัวอย่างคอนทัวร์ LFC และ FVC: คอนทัวร์ LFC ที่ Fc_{LFC} = 0.5 Hz เปรียบเทียบกับคอนทัวร์ F_0 (บน) คอนทัวร์ FVC ที่ Fc_{FVC} = 0.5 Hz เปรียบเทียบกับคอนทัวร์ |HFC| (ล่าง)



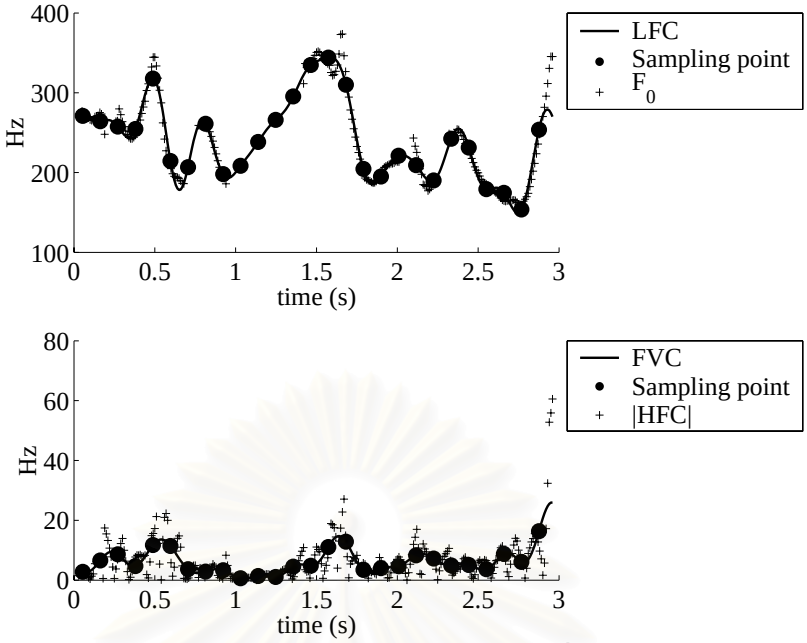
รูปที่ 3.26 ตัวอย่างคอนทัวร์ LFC และ FVC: คอนทัวร์ LFC ที่ Fc_{LFC} = 1.5 Hz เปรียบเทียบกับคอนทัวร์ F_0 (บน) คอนทัวร์ FVC ที่ Fc_{FVC} = 1.5 Hz เปรียบเทียบกับคอนทัวร์ |HFC| (ล่าง)



รูปที่ 3.27 ตัวอย่างคอนทัวร์ LFC และ FVC: คอนทัวร์ LFC ที่ Fc_{LFC} = 2.5 Hz เปรียบเทียบกับคอนทัวร์ F_0 (บน) คอนทัวร์ FVC ที่ Fc_{FVC} = 2.5 Hz เปรียบเทียบกับคอนทัวร์ |HFC| (ล่าง)



รูปที่ 3.28 ตัวอย่างคอนทัวร์ LFC และ FVC: คอนทัวร์ LFC ที่ Fc_{LFC} = 3.5 Hz เปรียบเทียบกับคอนทัวร์ F_0 (บน) คอนทัวร์ FVC ที่ Fc_{FVC} = 3.5 Hz เปรียบเทียบกับคอนทัวร์ |HFC| (ล่าง)



รูปที่ 3.29 ตัวอย่างคอนทัวร์ LFC และ FVC: คอนทัวร์ LFC ที่ Fc_{LFC} = 4.5 Hz เปรียบเทียบกับคอนทัวร์ F_0 (บน) คอนทัวร์ FVC ที่ Fc_{FVC} = 4.5 Hz เปรียบเทียบกับคอนทัวร์ |HFC| (ล่าง)

3.3.2 การฝึกฝน และทดสอบโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมที่ใช้ในการทดลองเป็นแบบ มัลติเลเยอร์เพอเซปตรอนแบบป้อนไปข้างหน้า (feed forward multi-layer perceptron) แบบ 3 ชั้น โดยมีจำนวนโนดของชั้นขาเข้าขึ้นกับจำนวนมิติของเวกเตอร์ลักษณะซึ่งขึ้นกับจำนวนจุดที่ได้จากการสุ่มตัวอย่างคอนทัวร์ ซึ่งจะเปลี่ยนค่าไปตามแต่ละการทดลอง จำนวนโนดของชั้นฮิดเดนเป็น 20 โนด จำนวนโนดของชั้นขาออกเท่ากับจำนวนของประเภทของทำนองเสียง (กำหนดให้แต่ละโนดแทนทำนองเสียงแต่ละประเภท) โดยจะมีการทดลอง 2 กรณี คือกรณีที่มีทำนองเสียง 3 ประเภท คือ ทำนองเสียงตก ทำนองเสียงขึ้น และทำนองเสียงผสมในกรณีนี้จะใช้จำนวนโนดของชั้นขาออกเป็น 3 โนด ส่วนอีกกรณี คือกรณีที่มีทำนองเสียง 2 ประเภท คือ ทำนองเสียงตก และทำนองเสียงขึ้น (รวมทำนองเสียงผสมเข้าไปเป็นประเภทเดียวกับทำนองเสียงขึ้น) ในกรณีนี้จะใช้จำนวนโนดของชั้นขาออกเป็น 2 โนด

การเลือกฟังก์ชันกระตุ้น (activation function) ของแต่ละโนดในชั้นฮิดเดน และชั้นขาเข้านั้น ในการทดลองนี้ใช้ฟังก์ชันไฮเพอร์โบลิกแทนเจนท์ (hyperbolic tangent: tanh) ซึ่งมีคุณสมบัติอสมมาตร (asymmetric) จะช่วยให้ได้โครงข่ายประสาทเทียมเรียนรู้ได้เร็วกว่า (จำนวนรอบในการฝึกฝนน้อยกว่า) การเลือกใช้ฟังก์ชันกระตุ้นแบบอื่น ๆ (Haykin, 1994) โดยฟังก์ชัน $\varphi(\upsilon)$ จะมีคุณสมบัติอสมมาตรเมื่อ $\varphi(-\upsilon) = -\varphi(\upsilon)$

การทดลองนี้ได้เลือกใช้ค่าเกนของการปรับปรุงค่าน้ำหนัก (gain of the weight updating) เป็น 1×10^{-5} ค่าโมเมนตัม (momentum) เป็น 0.95 และจำนวนครั้งในการปรับค่าน้ำหนักสูงสุดเป็น 3000 ซึ่งได้มาจากการทดสอบในเบื้องต้น ว่าสามารถทำให้โครงข่ายประสาทเทียมสามารถเรียนรู้ได้ดี

สำหรับวิธีที่ใช้ในการฝึกฝน (training) โครงข่ายประสาทเทียมนั้น การทดลองนี้ใช้วิธีการแพร่กลับ (back-propagation) การแบ่งกลุ่มข้อมูลที่ใช้ในการฝึกฝน และทดสอบ (testing) นั้น งานวิจัยนี้ใช้วิธี ครอสวาลิเดชันแบบ 5 โฟล (5-fold cross validation) โดยการแบ่งกลุ่มข้อมูลที่จะทดลองออกเป็น 5 กลุ่ม แล้วทำการทดลองฝึกฝน และทดสอบโครงข่ายประสาทเทียม 5 การทดลอง โดยในแต่ละการทดลองก็เลือกกลุ่มที่จะนำมาฝึกฝน และกลุ่มที่จะนำมาทดสอบโครงข่ายประสาทเทียม ดังแสดงในตารางที่ 3.1 หลังจากที่ทำการทดลองครบทั้ง 5 การทดลองแล้ว ผลการรู้จำที่จะนำไปใช้ จะได้มาจากการเฉลี่ยผลการรู้จำของการทดลองทั้ง 5

ตารางที่ 3.1 การแบ่งกลุ่มข้อมูลโดยใช้วิธีครอสวาลิเดชัน แบบ 5 โฟล

	กลุ่มที่ 1	**กลุ่มที่ 2**	**กลุ่มที่ 3**	**กลุ่มที่ 4**	**กลุ่มที่ 5**
การทดลองที่ 1	ทดสอบ	ฝึกฝน	ฝึกฝน	ฝึกฝน	ฝึกฝน
การทดลองที่ 2	ฝึกฝน	ทดสอบ	ฝึกฝน	ฝึกฝน	ฝึกฝน
การทดลองที่ 3	ฝึกฝน	ฝึกฝน	ทดสอบ	ฝึกฝน	ฝึกฝน
การทดลองที่ 4	ฝึกฝน	ฝึกฝน	ฝึกฝน	ทดสอบ	ฝึกฝน
การทดลองที่ 5	ฝึกฝน	ฝึกฝน	ฝึกฝน	ฝึกฝน	ทดสอบ

การแบ่งการทดลองแบบนี้มีผลดี คือ สามารถลดความแปรปรวนของผลการทดลองได้ โดยจะเห็นได้ว่าถ้าทำการทดลองเพียงครั้งเดียว ผลการทดลองที่ได้จะขึ้นกับการเลือกกลุ่มฝึกฝน และกลุ่มทดสอบมาก ถ้าเลือกข้อมูลกลุ่มที่ดีมาฝึกฝน จะทำให้ได้อัตราการรู้จำที่สูง แต่ถ้าเลือกข้อมูลที่ไม่ดีมาฝึกฝนจะทำให้ได้อัตราการรู้จำต่ำ แต่เมื่อทดลองโดยใช้วิธี 5 โฟล จะทำให้ข้อมูลทุกกลุ่มได้เป็นทั้ง กลุ่มฝึกฝน และกลุ่มทดสอบ อัตราการรู้จำที่ได้เป็นอัตราการรู้จำเฉลี่ยจากการเลือกกลุ่มทั้ง 5 แบบ จึงทำให้ผลการทดลองมีความแปรปรวนลดลง และมีความน่าเชื่อถือมากขึ้น (Schneider และ Moore, 1997) อย่างไรก็ตาม เนื่องจากการทดลองโดยวิธี 5 โฟล จะต้องฝึกฝน และทดสอบโครงข่ายประสาทเทียมถึง 5 ครั้ง จึงทำให้ใช้เวลาในการทดลองนานกว่าวิธีปกติถึง 5 เท่า

ในการทดลองการรู้จำทำนองเสียงนั้น อาจเป็นไปได้ว่า ตัวรู้จำอาจจะไม่ได้จำในสิ่งที่เราต้องการให้จำ แต่ไปจำข้อมูลอื่นแทน เช่น ตัวรู้จำอาจจะไปรู้จำรูปแบบของคอนทัวร์ F_0 ของ

ประโยคหนึ่ง ๆ จากข้อมูลที่ใช้ในการฝึกฝน แทนที่จะจำรูปแบบของทำนองเสียงของประโยคนั้น ๆ และเมื่อเรานำตัวรู้จำไปทดสอบโดยให้ตัวรู้จำรู้จำทำนองเสียงของประโยคเดียวกับประโยคที่นำไปฝึกฝน (แต่พูดด้วยผู้พูดคนละคนกัน) ตัวรู้จำอาจจะตอบทำนองเสียงถูก แต่ไม่ใช่เพราะการจำทำนองเสียงได้ แต่เป็นเพราะจำประโยคนั้นได้ ซึ่งจะมีผลให้ผลการรู้จำมีความคลาดเคลื่อนไปจากความเป็นจริง

เพื่อป้องกันปัญหาดังกล่าว ในการแบ่งกลุ่มการทดลองออกเป็น 5 กลุ่ม จึงได้พยายามจัดให้เสียงพูดที่พูดจากประโยคเดียวกัน อยู่ในกลุ่มเดียวกันมากที่สุดเท่าที่จะเป็นไปได้ ซึ่งจะทำให้โอกาสที่เสียงพูดในกลุ่มฝึกฝน และกลุ่มทดสอบเกิดจากประโยคเดียวกันเกิดขึ้นได้น้อย

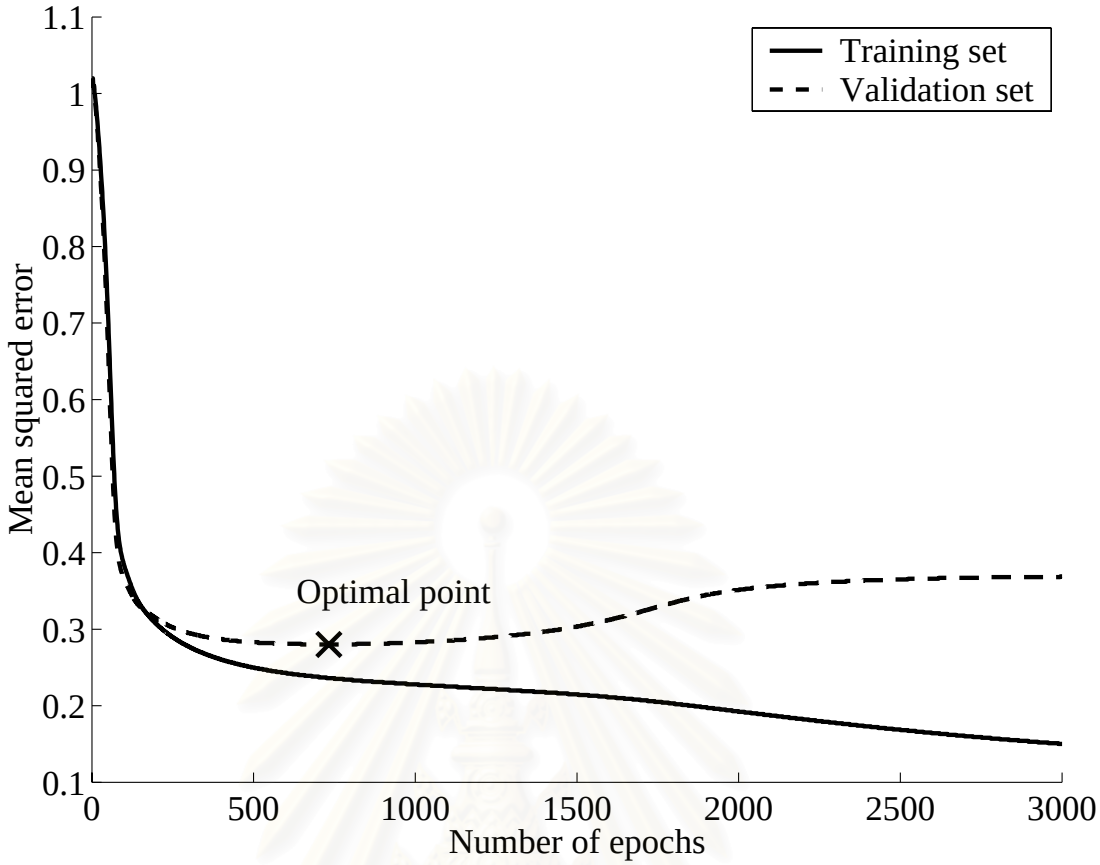
งานที่เกี่ยวข้องกับการรู้จำแบบรูป (pattern recognition) โดยทั่วไป มักจะพบปัญหาการฝึกฝนมากเกินไป (overtraining) ซึ่งเกิดจากการใช้จำนวนรอบในการฝึกฝนที่มากเกินไป จนทำให้ตัวรู้จำยึดติดกับข้อมูลกลุ่มที่นำมาฝึกฝนมาก จนขาดคุณสมบัติ *ความเป็นทั่วไป* (generalization) ซึ่งจะทำให้อัตราการรู้จำด้อยลงเมื่อนำไปทดสอบ วิธีการหนึ่งที่นิยมใช้ลดปัญหาการฝึกฝนมากเกินไป คือ วิธีครอสวาลิเดชัน (cross-validation) (Haykin, 1994)

ในการทำครอสวาลิเดชัน จะเริ่มจากการแบ่งกลุ่มข้อมูลออกเป็น 2 กลุ่ม คือกลุ่มฝึกฝน และกลุ่มทดสอบ หลังจากนั้นจะแบ่งส่วนหนึ่งของกลุ่มฝึกฝนออกมาเรียกว่า กลุ่มวาลิเดชัน (validation set) จากนั้นจึงนำกลุ่มฝึกฝนไปฝึกฝนตัวรู้จำ โดยในระหว่างการฝึกฝนแต่ละรอบ นอกจากจะหาค่าความผิดพลาดกำลังสองเฉลี่ย (mean squared error: MSE) ของกลุ่มฝึกฝนแล้ว ก็หา MSE ของ กลุ่มวาลิเดชันด้วย ตัวอย่างของค่า MSE ของกลุ่มฝึกฝน และกลุ่มวาลิเดชันเมื่อเทียบกับจำนวนรอบในการปรับค่าน้ำหนัก แสดงดังรูปที่ 3.30

จากรูปที่ 3.30 จะเห็นได้ว่า ในช่วงแรกของการฝึกฝน ค่า MSE ของกลุ่มฝึกฝน และกลุ่มวาลิเดชันต่างก็ลดด้วยกันทั้งคู่ จนกระทั่งฝึกฝนมาถึงรอบที่ 647 ค่า MSE ของกลุ่มวาลิเดชันก็เริ่มคงที่อยู่ที่ 0.28 จากนั้นพอถึงรอบที่ 819 ก็กลับมีค่าเพิ่มขึ้นแล้วก็เพิ่มขึ้นต่อไปเรื่อย ๆ ในขณะที่ค่า MSE ของกลุ่มฝึกฝนยังคงลดลงเรื่อย ๆ

การที่ค่า MSE ของกลุ่มวาลิเดชันกลับมาเพิ่มขึ้น แสดงว่าตัวรู้จำได้เริ่มเกิดผลเสียอันเกิดจากการฝึกฝนมากเกินไป ดังนั้นโครงข่ายประสาทเทียมที่จะนำมาใช้เป็นตัวรู้จำที่ดีที่สุด คือโครงข่ายในขณะที่ทำให้ได้ค่า MSE ของกลุ่มวาลิเดชันต่ำที่สุด ดังแสดงในจุด Optimal point ของรูปที่ 3.30

งานวิจัยนี้ได้กำหนดกลุ่มฝึกฝน กลุ่มวาลิเดชัน และกลุ่มทดสอบดังแสดงในตารางที่ 3.2



รูปที่ 3.30 ค่าผิดพลาดแบบกำลังสองเฉลี่ยของกลุ่มฝึกฝน และกลุ่มวาลิเดชัน ที่จำนวนรอบของการปรับค่าน้ำหนักต่าง ๆ

ตารางที่ 3.2 การแบ่งกลุ่มข้อมูลที่ใช้ในงานวิจัย

	กลุ่มที่ 1	**กลุ่มที่ 2**	**กลุ่มที่ 3**	**กลุ่มที่ 4**	**กลุ่มที่ 5**
การทดลองที่ 1	ทดสอบ	วาลิเดชัน	ฝึกฝน	ฝึกฝน	ฝึกฝน
การทดลองที่ 2	ฝึกฝน	ทดสอบ	วาลิเดชัน	ฝึกฝน	ฝึกฝน
การทดลองที่ 3	ฝึกฝน	ฝึกฝน	ทดสอบ	วาลิเดชัน	ฝึกฝน
การทดลองที่ 4	ฝึกฝน	ฝึกฝน	ฝึกฝน	ทดสอบ	วาลิเดชัน
การทดลองที่ 5	วาลิเดชัน	ฝึกฝน	ฝึกฝน	ฝึกฝน	ทดสอบ

บทที่ 4

การทดลอง และการวิเคราะห์ผลการทดลอง

บทนี้กล่าวถึง การทดลอง และการวิเคราะห์ผลการทดลองของงานวิจัยนี้ โดยจะเริ่มจากหัวข้อ 4.1 กล่าวถึงข้อมูลเสียงพูดที่ใช้ในการวิจัย หัวข้อ 4.2 กล่าวถึงลักษณะของการทดลองการรู้จำเสียงพูด รวมทั้งซอฟต์แวร์โครงข่ายประสาทเทียมที่ใช้ในการทดลอง หัวข้อ 4.3 กล่าวถึงการทดลองการรู้จำเสียงพูด ในกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท หัวข้อ 4.4 กล่าวถึงการทดลองการรู้จำเสียงพูด ในกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท และในหัวข้อที่ 4.5 แสดงรูปร่างของคอนทัวร์ LFC และ FVC โดยเฉลี่ย จากข้อมูลเสียงพูดทั้งหมดที่ใช้ในการทดลอง ของทำนองเสียงพูดแต่ละประเภท พร้อมทั้งวิเคราะห์รูปร่างของคอนทัวร์เหล่านั้น

4.1 ข้อมูลเสียงพูดที่ใช้ในการวิจัย

4.1.1 ประโยคที่ใช้

ประโยคที่ใช้ทดสอบ มีทั้งสิ้น 61 ประโยค มาจากบทสนทนา 6 บท ดังที่ได้แสดงไว้ในภาคผนวก ข แต่ละบทมีผู้พูด 2 คนพูดคุยกัน โดยในแต่ละประโยคพูดได้เขียนกำกับไว้ว่าจะต้องพูดด้วยทำนองเสียงพูดแบบใด ความยาวของแต่ละประโยคจะอยู่ระหว่าง 1 – 11 พยางค์

4.1.2 การรวบรวมข้อมูล

ข้อมูลเสียงที่ใช้มาจากผู้พูด 12 คน ผู้ชาย 6 คน ผู้หญิง 6 คน ผู้พูดจะจับคู่กันและสนทนาตามบทพูด โดยผู้พูดทุกคนจะพูดแต่ละบทสนทนา 2 ครั้ง โดยสลับบทพูดกัน เพื่อให้ผู้พูดทุกคนได้พูดครบทั้ง 61 ประโยค จึงมีประโยคสำหรับทดสอบทั้งสิ้น $12 \times 61 = 732$ ประโยค

4.1.3 การตรวจสอบทำนองเสียง

เนื่องจากผู้พูดแต่ละคนได้รับคำสั่งว่าให้พูดตามบทสนทนาให้เป็นธรรมชาติมากที่สุด จึงทำให้มีบางประโยค ที่ผู้พูดพูดด้วยทำนองเสียงที่ไม่ตรงกับทำนองเสียงที่ได้กำหนดไว้ให้ จึงได้กำหนดให้มีการนำข้อมูลเสียงทั้งหมดมาตรวจสอบประเภทของทำนองเสียงอีกครั้ง

ตัวอย่างของการตรวจสอบทำนองเสียงพูดในงานวิจัยอื่นที่เกี่ยวข้องกับการรู้จำทำนองเสียงพูด ได้แก่ งานวิจัยของ Keißling และคนอื่น ๆ (1993) ซึ่งเป็นการรู้จำทำนองเสียงพูดของภาษาเยอรมัน ซึ่งได้นำไปรวมกับระบบเข้าใจเสียงพูด EVAR (an experimental automatic information system on train table) งานวิจัยดังกล่าวกำหนดให้ประโยคที่พูดผิด (ตัวอย่างเช่น ต้องการให้พูดเป็น

ประโยค คำถาม แต่คอนทัวร์ F_0 กลับมีลักษณะลดระดับ แทนที่จะเพิ่มระดับตามลักษณะของทำนองเสียงในภาษาเยอรมัน) เป็นประโยคที่ผิดพลาด (error) และตัดประโยคเหล่านี้ทิ้งไป โดยไม่มีการนำมาฝึกฝน หรือทดสอบ

สำหรับงานวิจัยนี้พบว่า ประโยคเสียงพูดบางประโยคก็พูดด้วยทำนองเสียงที่ไม่ตรงกับที่กำหนดไว้ให้เช่นเดียวกัน และยังพบว่าประโยคเสียงพูดบางประโยคก็ยากต่อการจำแนกโดยมนุษย์ว่าเป็นทำนองเสียงประเภทใด จึงกำหนดให้มีการตรวจสอบประเภทของทำนองเสียงโดยผู้ฟัง 2 คน โดยมีวัตถุประสงค์เพื่อแบ่งกลุ่มของประโยคเสียงพูดออกเป็น ประโยคที่มีทำนองเสียงที่ชัดเจน กับประโยคที่มีทำนองเสียงที่กำกวม (ประโยคเสียงพูดที่ผู้ฟังซึ่งเป็นคนไทย ไม่สามารถบอกได้อย่างมั่นใจว่าประโยคที่ตนได้ยิน มีทำนองเสียงพูดแบบใด) โดยในงานวิจัยนี้จะนำเฉพาะประโยคที่มีทำนองเสียงที่ชัดเจนเท่านั้น มาทำการทดลอง

การตรวจสอบทำนองเสียงพูด จะทำโดยการเขียนโปรแกรมเพื่อนำเสียงทั้ง 732 ประโยค มาสลับลำดับแบบสุ่ม แล้วเปิดให้ผู้ฟังทั้ง 2 คนฟัง แล้วให้ผู้ฟังตัดสินใจว่าเสียงที่ได้ยิน เป็นทำนองเสียงพูดแบบใด โดยให้ผู้ฟังแต่ละคนตรวจสอบเสียงพูดทุกเสียงคนละ 2 รอบ ดังนั้นจะได้ว่ามีการทดสอบการฟัง 4 ครั้งสำหรับประโยคเสียงพูดแต่ละประโยค โดยในจำนวน 4 ครั้งนี้ ถึงแม้จะเป็นการตรวจสอบโดยผู้ฟัง 2 คนคนละ 2 ครั้ง แต่ก็ไม่จำเป็นว่าผู้ฟังคนเดิมฟังเสียงเดิมจะต้องเลือกว่าเป็นทำนองเสียงเดิมทุกครั้ง เนื่องจากประโยคมีจำนวนมาก และในการฟังทั้ง 2 ครั้ง ก็มีการสลับลำดับแบบสุ่มที่ไม่เหมือนกัน

ถ้ากำหนดให้ประโยคที่มีทำนองเสียงที่ชัดเจน คือประโยคที่ได้รับการเลือกให้มีทำนองเสียงประเภทเดียวกันทั้ง 4 ครั้ง จะทำให้ได้ประโยคเสียงพูดที่มีลักษณะเฉพาะของทำนองเสียงประเภทต่าง ๆ อย่างชัดเจนมาก แต่จะทำให้ข้อมูลเสียงที่จะนำไปใช้ในการทดลองมีจำนวนน้อยเกินไป ต่อการนำไปฝึกฝน และทดสอบโครงข่ายประสาทเทียม ดังแสดงในตารางที่ 4.1 (ก) และจะเป็นการจำกัดงานวิจัยมากเกินไป ดังนั้นในงานวิจัยนี้จึงได้กำหนดให้ ประโยคเสียงพูดที่มีทำนองเสียงที่ชัดเจน คือประโยคเสียงพูดที่ได้รับการเลือกจากผู้ฟัง ให้มีทำนองเสียงประเภทเดียวกัน 3 ครั้งขึ้นไป จากการฟังทั้งหมด 4 ครั้ง ซึ่งจะทำให้ได้จำนวนของประโยคเสียงพูดสำหรับแต่ละทำนองเสียง ดังแสดงในตารางที่ 4.1 (ข)

ตารางที่ 4.1 จำนวนประโยคที่ผู้ฟังสามารถบอกประเภทของทำนองเสียงได้อย่างชัดเจน แบ่งตามประเภทของทำนองเสียง และเพศของผู้พูด

(ก) กำหนดให้ประโยคเสียงพูดที่มีทำนองเสียงชัดเจนคือประโยคที่ได้รับการเลือกให้มีทำนองเสียงประเภทเดียวกันทั้ง 4 ครั้ง จากการฟังทั้งหมด 4 ครั้ง

ประเภทของทำนองเสียง	เพศของผู้พูด	
	ชาย	หญิง
ทำนองเสียงตก	59	63
ทำนองเสียงขึ้น	46	35
ทำนองเสียงผสม	46	89

(ข) กำหนดให้ประโยคเสียงพูดที่มีทำนองเสียงชัดเจน คือประโยคที่ได้รับการเลือกให้มีทำนองเสียงประเภทเดียวกัน 3 ครั้งขึ้นไป จากการฟังทั้งหมด 4 ครั้ง
(กรณีนี้เป็นกรณีที่นำไปใช้ในการทดลอง ของงานวิจัยนี้)

ประเภทของทำนองเสียง	เพศของผู้พูด	
	ชาย	หญิง
ทำนองเสียงตก	98	106
ทำนองเสียงขึ้น	91	65
ทำนองเสียงผสม	79	122

4.2 การทดลองรู้จำทำนองเสียง

การทดลองในงานวิจัยนี้จะแบ่งเป็น 2 ส่วนใหญ่ ๆ คือ การทดลองในหัวข้อ 4.3 จะแบ่งทำนองเสียงออกเป็น 3 ประเภท คือ ทำนองเสียงตก ทำนองเสียงขึ้น และทำนองเสียงผสม ส่วนการทดลองในหัวข้อ 4.4 จะแบ่งทำนองเสียงออกเป็น 2 ประเภท คือ ทำนองเสียงตก และทำนองเสียงขึ้น โดยจะจัดกลุ่มทำนองเสียงผสม เข้าไปอยู่ในประเภทเดียวกับทำนองเสียงขึ้น

การทดลองแต่ละประเภท จะเป็นลักษณะที่ขึ้นกับผู้พูด (speaker dependent) นั่นคือ ผู้พูดที่อยู่ในกลุ่มฝึกฝน และผู้พูดที่อยู่ในกลุ่มทดสอบเป็นกลุ่มเดียวกัน โดยในแต่ละการทดลองจะทดลองเสียงของผู้ชาย และเสียงของผู้หญิงแยกจากกัน เนื่องจากคอนทัวร์ F_0 ของเสียงผู้ชาย และคอนทัวร์ F_0 ของเสียงผู้หญิงมีความแตกต่างกันมาก

ซอฟต์แวร์โครงข่ายประสาทเทียมที่ใช้ในการทดลองนี้ได้มาจากชุดซอฟต์แวร์ NICO Toolkit (Neural Inference COmputation) (Ström, 1997) ซึ่งอนุญาตให้ใช้ หรือแก้ไขเพื่อการศึกษาหรืองานวิจัยได้โดยไม่เสียค่าใช้จ่าย งานวิจัยนี้จึงได้แก้ไขโปรแกรมบางส่วน เพื่อให้โปรแกรม NICO สามารถฝึกฝนโครงข่ายประสาทเทียมแบบใช้กลุ่มวาลิเดชัน ได้ดีขึ้น

4.3 การรู้จำทำนองเสียง กรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท

หัวข้อนี้จะกล่าวถึงการทดลองหาอัตราการรู้จำทำนองเสียง กรณีที่แบ่งประเภทของทำนองเสียงออกเป็น 3 ประเภท คือ ทำนองเสียงตก ทำนองเสียงขึ้น และทำนองเสียงผสม โดยในหัวข้อ 4.3.1 กล่าวถึงการรู้จำทำนองเสียงโดยใช้คอนทัวร์ LFC เพียงอย่างเดียว หัวข้อ 4.3.2 กล่าวถึงการรู้จำทำนองเสียงโดยคอนทัวร์ LFC และค่าผลต่างอันดับหนึ่งของคอนทัวร์ LFC (ΔLFC) หัวข้อ 4.3.3 กล่าวถึงการรู้จำทำนองเสียงโดยใช้คอนทัวร์ LFC และ FVC หัวข้อ 4.3.4 กล่าวถึงการรู้จำทำนองเสียงโดยใช้ คอนทัวร์ LFC คอนทัวร์ FVC และ ΔLFC และสุดท้ายหัวข้อ 4.3.5 เป็นการเปรียบเทียบผลการทดลองทั้งสี่แบบ

การรายงานผลการทดลองในแต่ละหัวข้อ จะเริ่มจากการรายงานอัตราการรู้จำทำนองเสียงประเภทต่าง ๆ และอัตราการรู้จำเฉลี่ย ที่ทุก ๆ ค่าความถี่ตัด ว่ามีค่าอยู่ในช่วงใด โดยเปรียบเทียบระหว่างผู้พูดหญิงกับผู้พูดชาย รวมทั้งวิเคราะห์ความสัมพันธ์ระหว่างอัตราการรู้จำเหล่านี้ กับค่าความถี่ตัดของตัวกรอง จากนั้นจึงพิจารณาเฉพาะกรณีของค่าความถี่ตัดที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุด โดยพิจารณาว่าตัวรู้จำ จำแนกทำนองเสียงหนึ่ง ๆ ผิดไปเป็นทำนองเสียงอื่นด้วยอัตราเท่าไร

4.3.1 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยค

การทดลองนี้ใช้ลักษณะที่เป็นพื้นฐานที่สุด คือใช้จุดที่ได้จากการสุ่มตัวอย่างคอนทัวร์ LFC รวมทั้งความยาวของประโยค (หน่วยเป็นวินาที) มาเป็นเวกเตอร์ข้อมูลขาเข้า สำหรับโครงข่ายประสาทเทียม โดยใช้ค่าความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC เป็น 0.5 , 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, และ 4.5 Hz รวมทั้งใช้ LFC ที่เป็น เส้นตรงซึ่งมีความชันเท่าใดก็ได้ (line) และเส้นตรงที่มีความชันเป็น 0 (line0)

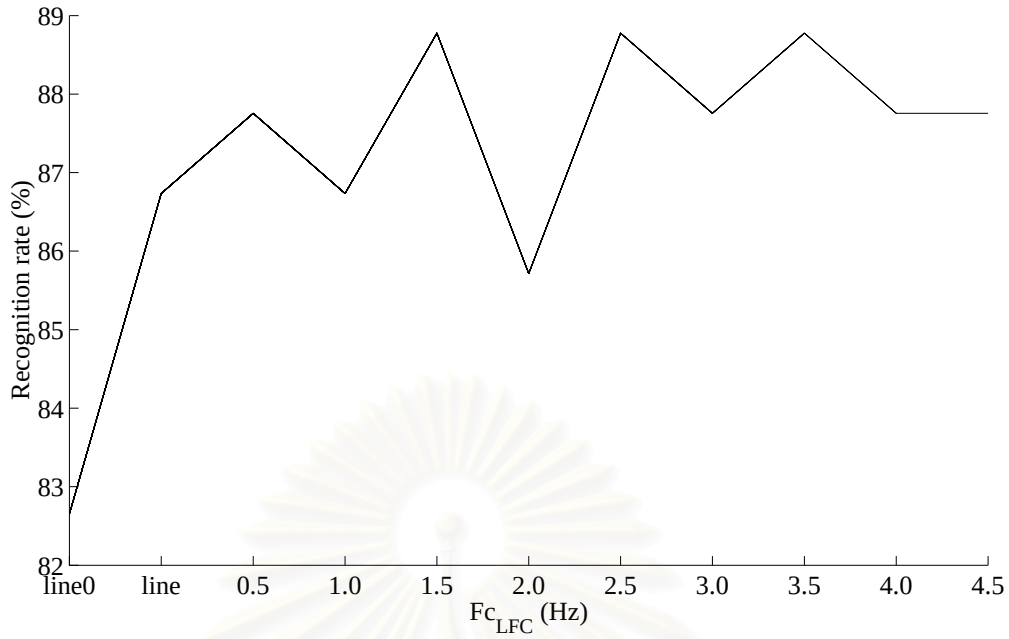
อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.1 – 4.6 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.7 และ 4.8

จากรูปที่ 4.1 – 4.6 จะเห็นได้ว่าอัตราการรู้จำของทำนองเสียงตกจะสูงกว่าทำนองเสียงขึ้นและทำนองเสียงผสม โดยในกรณีของผู้ชายอัตราการรู้จำของทำนองเสียงตกจะอยู่ระหว่างร้อยละ 82 ถึง 88 ในขณะที่ของผู้หญิงจะอยู่ระหว่างร้อยละ 91 ถึง 96 แสดงว่า ถึงแม้จะใช้เพียงองค์ประกอบ LFC และความยาวของประโยค โครงข่ายประสาทเทียมก็สามารถจำแนกความแตกต่างระหว่างทำนองเสียงตก กับทำนองเสียงประเภทอื่น ๆ ได้ดีพอสมควร

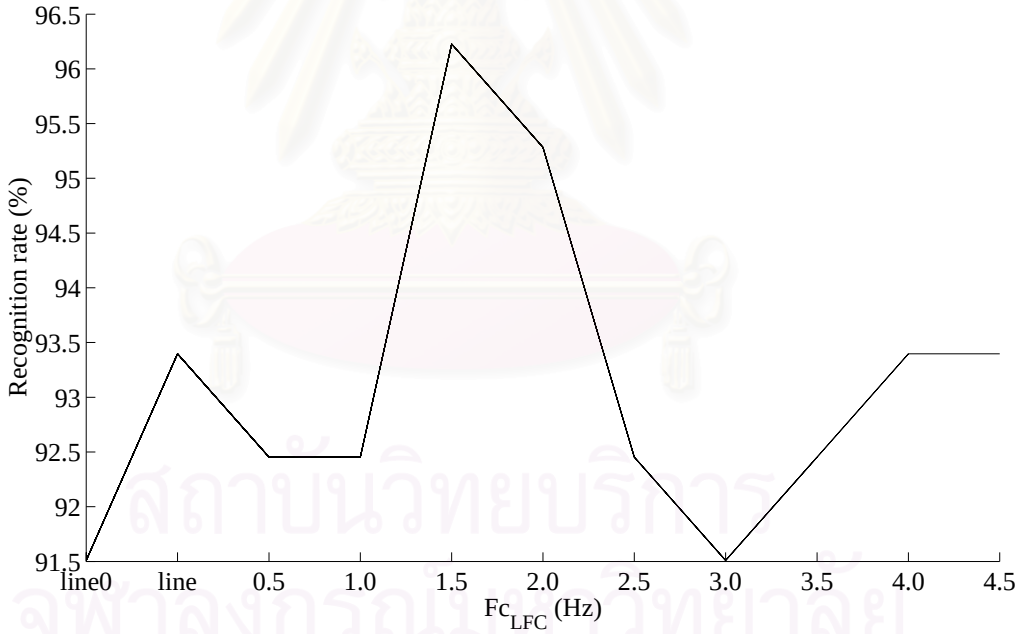
ส่วนในกรณีของทำนองเสียงขึ้นนั้น จะเห็นได้ว่าอัตราการรู้จำของผู้ชายจะสูงกว่าอัตราการรู้จำของผู้หญิง โดยอัตราการรู้จำของผู้ชายจะอยู่ระหว่างร้อยละ 44 ถึง 56 ในขณะที่อัตราการรู้จำของผู้หญิงจะอยู่ระหว่างร้อยละ 2 ถึง 20 สำหรับกรณีของทำนองเสียงผสม อัตราการรู้จำของผู้หญิงจะมากกว่าของผู้ชาย คือ อัตราการรู้จำของผู้ชายจะอยู่ระหว่างร้อยละ 16 ถึง 24 ในขณะที่อัตราการรู้จำของผู้หญิงจะอยู่ระหว่างร้อยละ 72 ถึง 84 แสดงว่าการใช้เพียงองค์ประกอบ LFC และความยาวของประโยค ยังไม่เพียงพอที่จะจำแนกประเภทของทำนองเสียงขึ้น และทำนองเสียงผสมได้ และเนื่องจากจำนวนประโยคที่นำมาทดสอบในกรณีของเสียงผู้ชาย มีประโยคที่เป็นทำนองเสียงขึ้น 91 ประโยค ในขณะที่ทำนองเสียงผสมมีเพียง 79 ประโยค จึงทำให้โครงข่ายประสาทเทียม เลือกที่จะให้ผลการรู้จำเป็นทำนองเสียงขึ้นมากกว่า ในขณะที่เสียงผู้หญิงมีประโยคที่เป็นทำนองเสียงขึ้น 65 ประโยค ในขณะที่ประโยคที่เป็นทำนองเสียงผสมมี 122 ประโยค จึงทำให้โครงข่ายประสาทเทียมเลือกที่จะให้ผลการรู้จำเป็นทำนองเสียงผสมมากกว่า

เมื่อพิจารณารูปที่ 4.7 และ 4.8 จะเห็นได้ว่า อัตราการรู้จำเฉลี่ยของทำนองเสียงทั้งสามประเภทของเสียงผู้หญิงจะมากกว่าเสียงผู้ชาย โดยอัตราการรู้จำเฉลี่ยของเสียงผู้ชายอยู่ที่ร้อยละ 56.7 ซึ่งเกิดขึ้นเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 1.5 และ 2.5 Hz ส่วนอัตราการรู้จำเฉลี่ยของเสียงผู้หญิงที่สูงที่สุดอยู่ที่ร้อยละ 70.3 ซึ่งเกิดขึ้นเมื่อใช้ LFC เป็นเส้นตรง

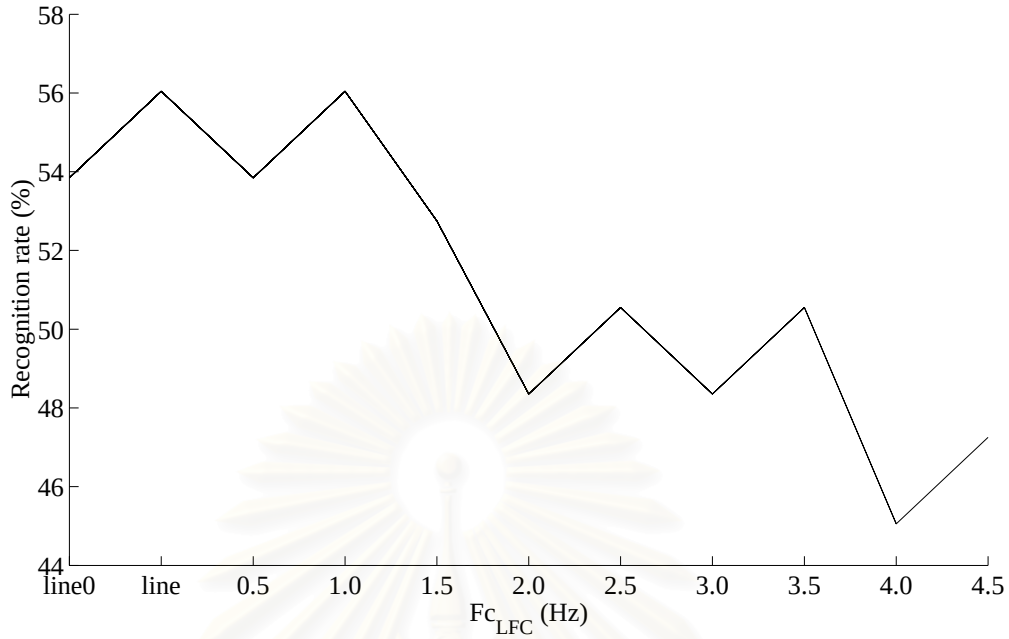
และจากรูปที่ 4.7 และ 4.8 จะเห็นได้ว่า เมื่อใช้ LFC ที่เป็นเส้นตรง อัตราการรู้จำเฉลี่ยของทั้งเสียงผู้ชาย และเสียงผู้หญิง จะสูงกว่ากรณีที่ใช้เส้นตรงที่มีความชันเป็น 0 อย่างเห็นได้ชัด แสดงว่าความชันของเส้นตรงสามารถให้ข้อมูลที่เกี่ยวข้องกับทำนองเสียงได้ นอกจากนี้เมื่อเพิ่มค่าความถี่ตัดของ LFC จะเห็นได้ว่าอัตราการรู้จำทำนองเสียงมีค่าต่ำกว่า ในกรณีที่ใช้ค่าความถี่ตัดต่ำ ๆ ซึ่งเป็นผลมาจากการใช้จำนวนจุดสุ่มตัวอย่างที่เพิ่มขึ้น ทำให้ระบบมีความซับซ้อนเพิ่มขึ้น การฝึกฝนโครงข่ายประสาทเทียมให้ได้ผลดีจึงยากขึ้น



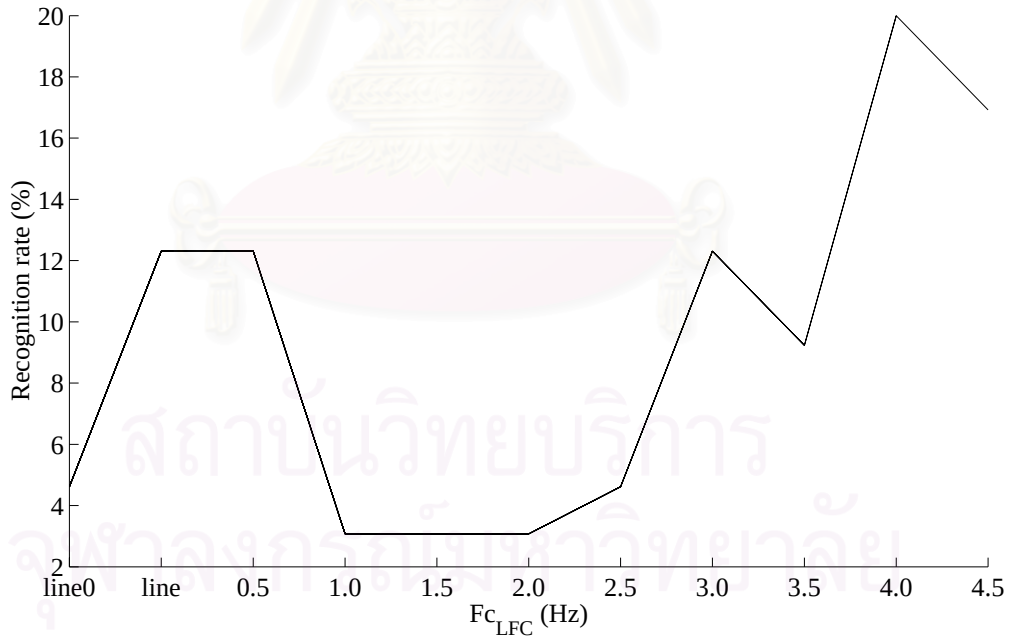
รูปที่ 4.1 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



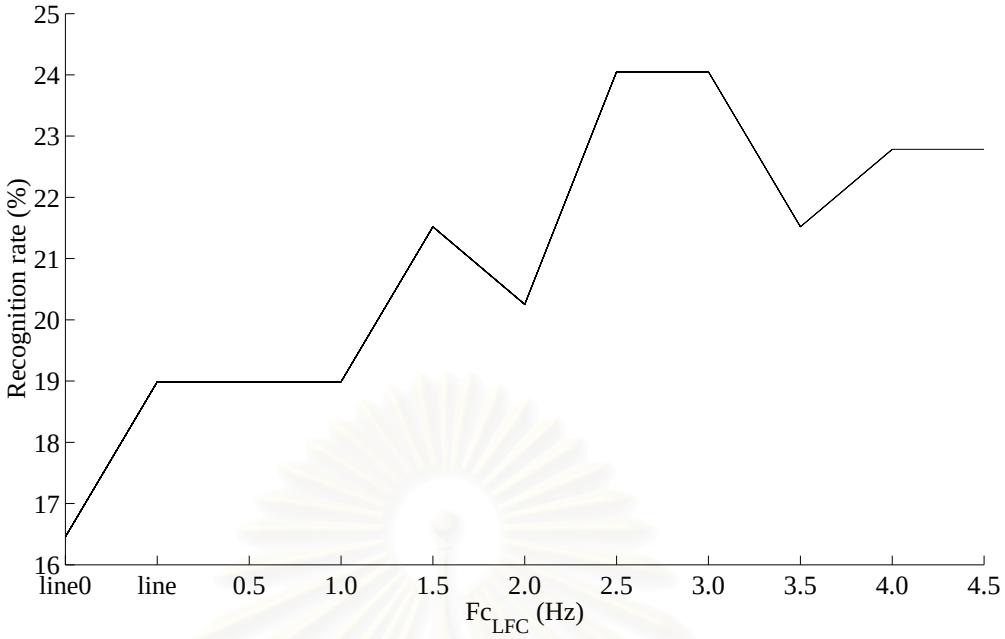
รูปที่ 4.2 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



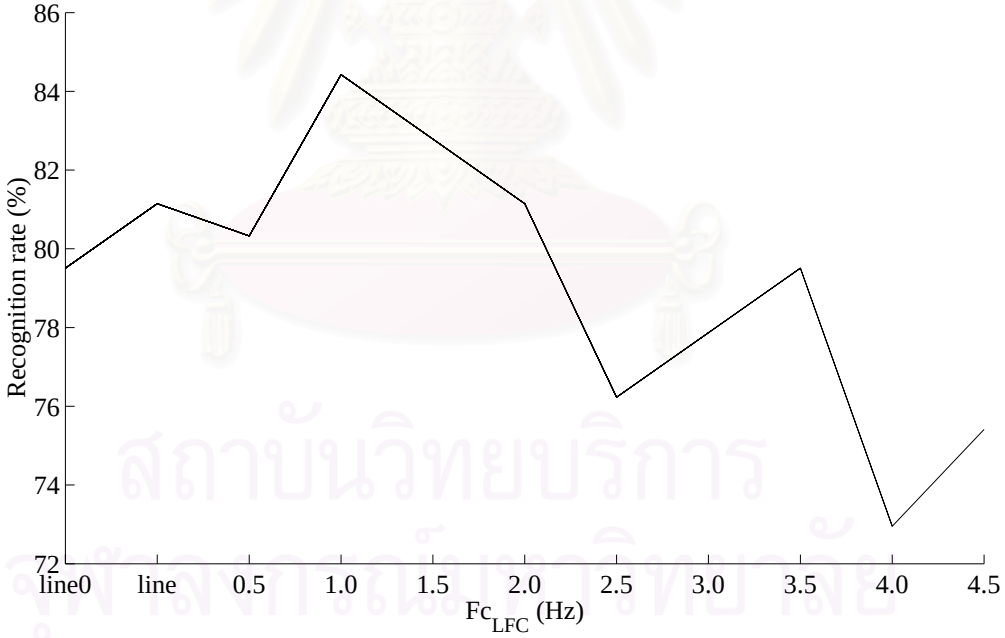
รูปที่ 4.3 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



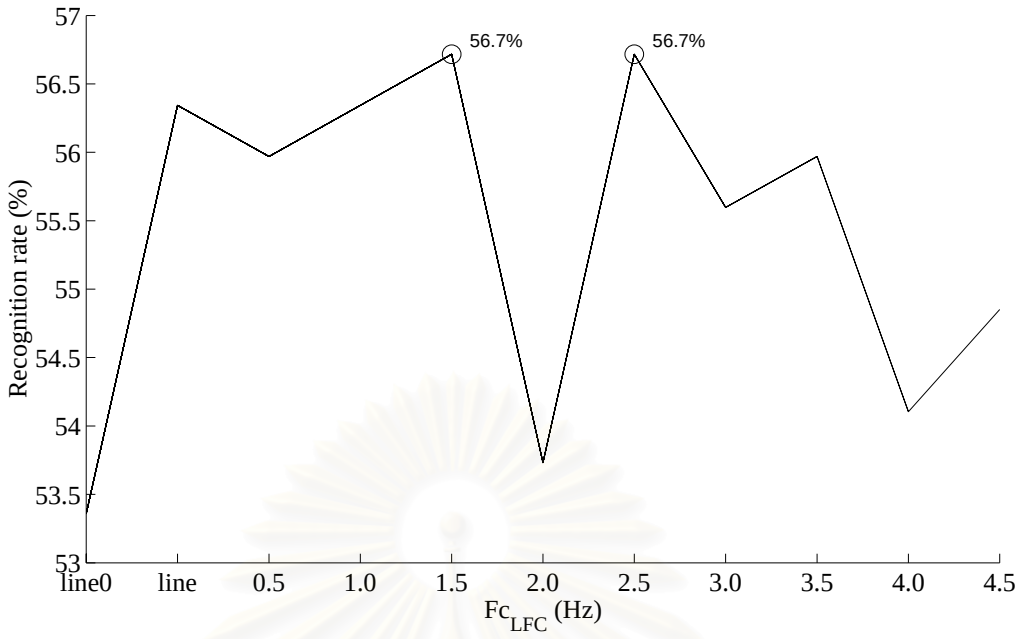
รูปที่ 4.4 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



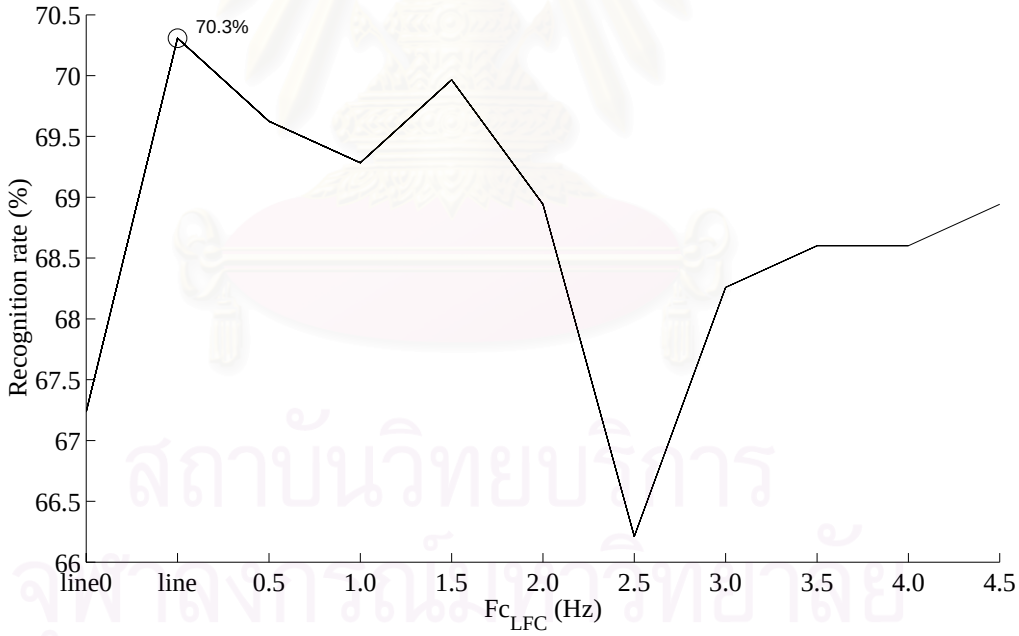
รูปที่ 4.5 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยค กรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



รูปที่ 4.6 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



รูปที่ 4.7 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



รูปที่ 4.8 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท

อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.7 และ 4.8 แสดงให้เห็นในตารางที่ 4.2 และ 4.3 ตามลำดับ

ตารางที่ 4.2 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย (จากรูปที่ 4.7)

ความถี่ตัด (Hz) หรือลักษณะของ LFC	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
อัตราการรู้จำเฉลี่ย (%)	53.4	56.3	56.0	56.3	**56.7**	53.7	**56.7**	55.6	56.0	54.1	54.9

ตารางที่ 4.3 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้หญิง (จากรูปที่ 4.8)

ความถี่ตัด (Hz) หรือลักษณะของ LFC	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
อัตราการรู้จำเฉลี่ย (%)	67.2	**70.3**	69.6	69.3	70.0	68.9	66.2	68.3	68.6	68.6	68.9

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.7 และ 4.8) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.4 – 4.6

จากตารางที่ 4.4 - 4.6 จะเห็นได้ว่าโอกาสที่โครงข่ายประสาทเทียมจะรู้จำทำนองเสียงตกผิดไปเป็นทำนองเสียงอื่นมีค่อนข้างน้อย เมื่อเทียบกับ ทำนองเสียงประเภทอื่น ๆ ในขณะที่โอกาสที่โครงข่ายประสาทเทียมจะรู้จำทำนองเสียงอื่น ๆ เป็นทำนองเสียงตกยังคงสูงอยู่

ตารางที่ 4.4 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.7) ของเสียงผู้ชายเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 1.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**88.8**	**10.2**	1.0	98
ขึ้น	**28.6**	**52.7**	18.7	91
ผสม	31.6	**46.8**	**21.5**	79
จำนวนประโยคทั้งหมด				268

ตารางที่ 4.5 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.7) ของเสียงผู้ชายเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 2.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**88.8**	**10.2**	1.0	98
ขึ้น	**29.7**	**50.5**	19.8	91
ผสม	30.4	**45.6**	**24.0**	79
จำนวนประโยคทั้งหมด				268

ตารางที่ 4.6 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.7) ของเสียงผู้หญิง เมื่อใช้ LFC เป็นเส้นตรง

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**93.4**	1.9	**4.7**	106
ขึ้น	23.1	**12.3**	**64.6**	65
ผสม	**15.6**	3.3	**81.1**	122
	จำนวนประโยคทั้งหมด			293

4.3.2 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยค

การทดลองนี้ใช้ลักษณะจากการทดลองที่ 4.3.1 คือใช้จุดที่ได้จากการสุ่มตัวอย่างคอนทัวร์ LFC และ ความยาวของประโยค มาเป็นเวกเตอร์ข้อมูลขาเข้าสำหรับโครงข่ายประสาทเทียม พร้อมทั้งได้เพิ่มค่า ΔLFC ณ ตำแหน่งที่สุ่มตัวอย่างค่า LFC เข้าไปด้วย โดยใช้ค่าความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC เป็น 0.5 , 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, และ 4.5 Hz รวมทั้งใช้ LFC ที่เป็น เส้นตรง (line)

การทดลองนี้ได้ตัดกรณีที่ใช้ LFC เป็นเส้นตรงที่มีความชันเป็น 0 ไป เนื่องจากจะทำให้ค่า ΔLFC มีค่าเป็น 0 สำหรับทุก ๆ ประโยค ซึ่งไม่มีผลต่อการรู้จำ ส่วนในกรณีที่ใช้ LFC เป็นเส้นตรงนั้น เนื่องจากค่า ΔLFC มีค่าเท่ากันทุกจุด จึงใช้ ΔLFC มิติเดียวสำหรับเวกเตอร์ลักษณะ ส่วนกรณีอื่น ๆ ใช้ค่ ΔLFC ที่ตำแหน่งเดียวกับจุดสุ่มตัวอย่างของคอนทัวร์ LFC

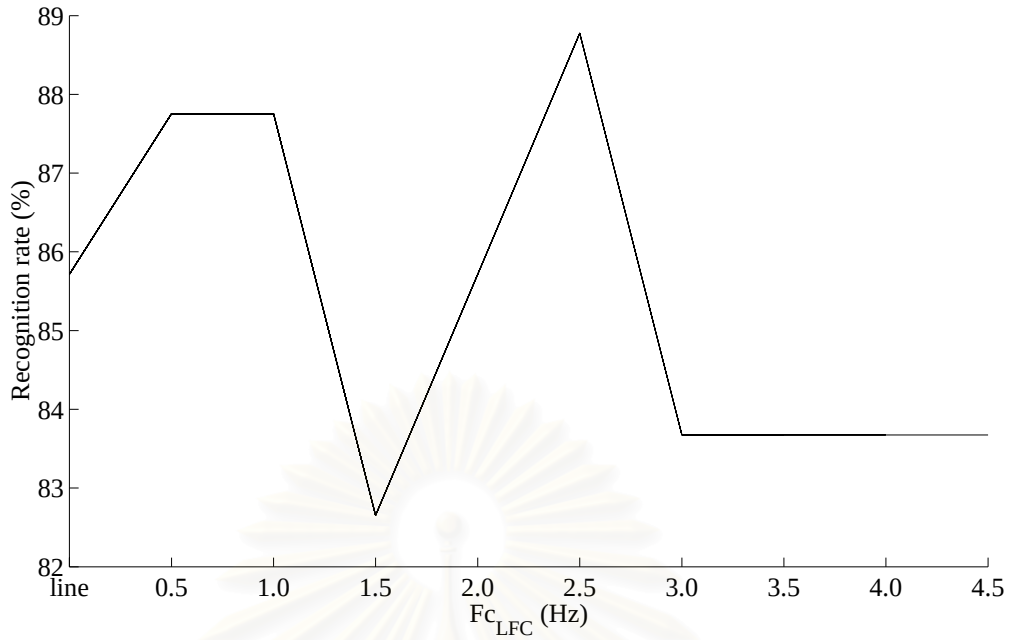
อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.9 – 4.14 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.15 และ 4.16

จากรูปที่ 4.9 – 4.14 จะเห็นได้ว่าอัตราการรู้จำของทำนองเสียงตกจะสูงกว่าทำนองเสียงขึ้นและทำนองเสียงผสม เช่นเดียวกับการทดลองในข้อ 4.3.1 โดยในกรณีของผู้ชายอัตราการ รู้จำ

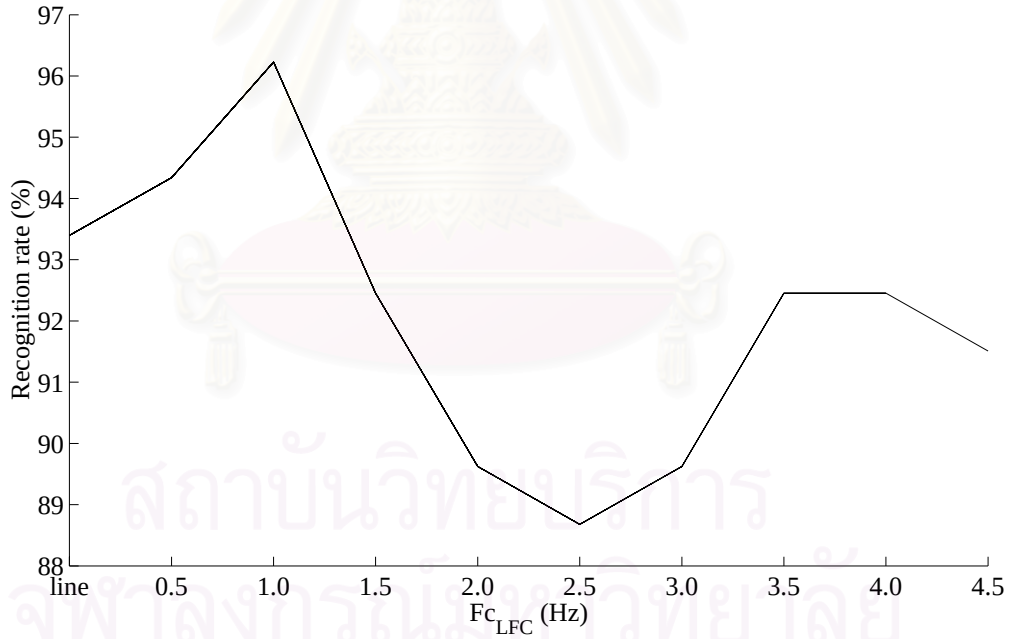
ทำนองเสียงตกจะอยู่ระหว่างร้อยละ 82 ถึง 88 ในขณะที่ของผู้หญิงจะอยู่ระหว่างร้อยละ 88 ถึง 96 ซึ่งอยู่ในช่วงที่ใกล้เคียงกับการทดลองในข้อ 4.3.1

ในกรณีของทำนองเสียงขึ้นและทำนองเสียงผสมนั้น อัตราการรู้จำยังคงมีลักษณะเดียวกับการทดลองในข้อ 4.3.1 คือ ในกรณีของทำนองเสียงขึ้น อัตราการรู้จำของเสียงผู้ชาย (ร้อยละ 38 ถึง 55) มากกว่าอัตราการรู้จำของเสียงผู้หญิง (ร้อยละ 0 ถึง 35) ส่วนในกรณีของทำนองเสียงผสม อัตราการรู้จำของเสียงผู้หญิง (ร้อยละ 68 ถึง 83) มากกว่าอัตราการรู้จำของเสียงผู้ชาย (ร้อยละ 14 ถึง 28) ด้วยเหตุผลเดียวกับข้อ 4.3.1

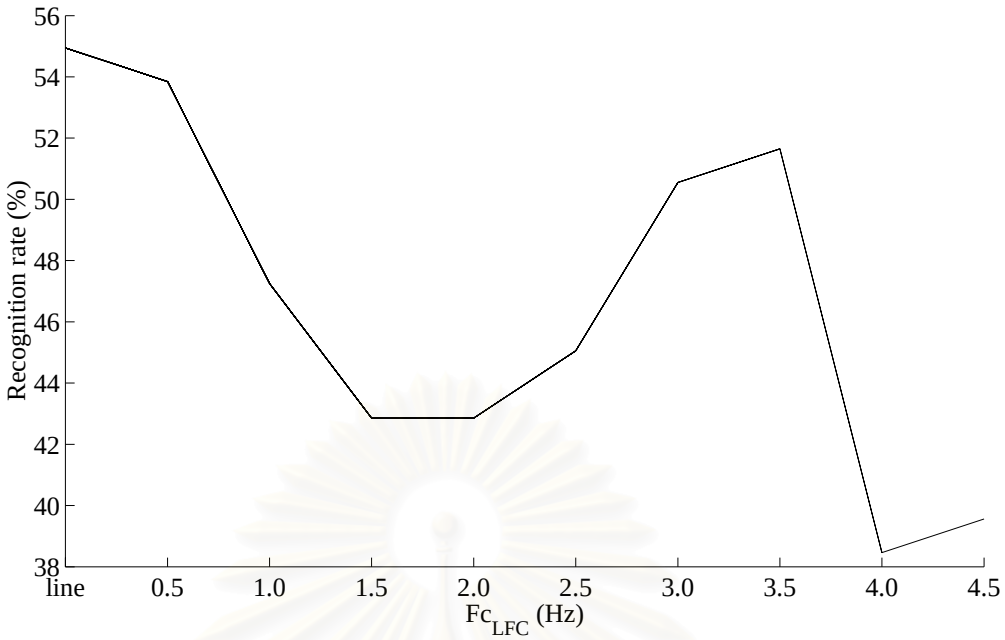
เมื่อพิจารณารูปที่ 4.15 และ 4.16 จะเห็นได้ว่า อัตราการรู้จำเฉลี่ยของทำนองเสียงทั้งสามประเภทของเสียงผู้หญิงจะมากกว่าเสียงผู้ชาย โดยอัตราการรู้จำเฉลี่ยของเสียงผู้ชายอยู่ที่ ร้อยละ 56.7 ซึ่งเกิดขึ้นเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 0.5 และ 3.5 Hz ส่วนอัตราการรู้จำเฉลี่ยของเสียงผู้หญิงที่สูงที่สุดอยู่ที่ร้อยละ 71.7 ซึ่งเกิดขึ้นเมื่อค่าความถี่ตัดของ LFC เป็น 3.0 และ 3.5 Hz



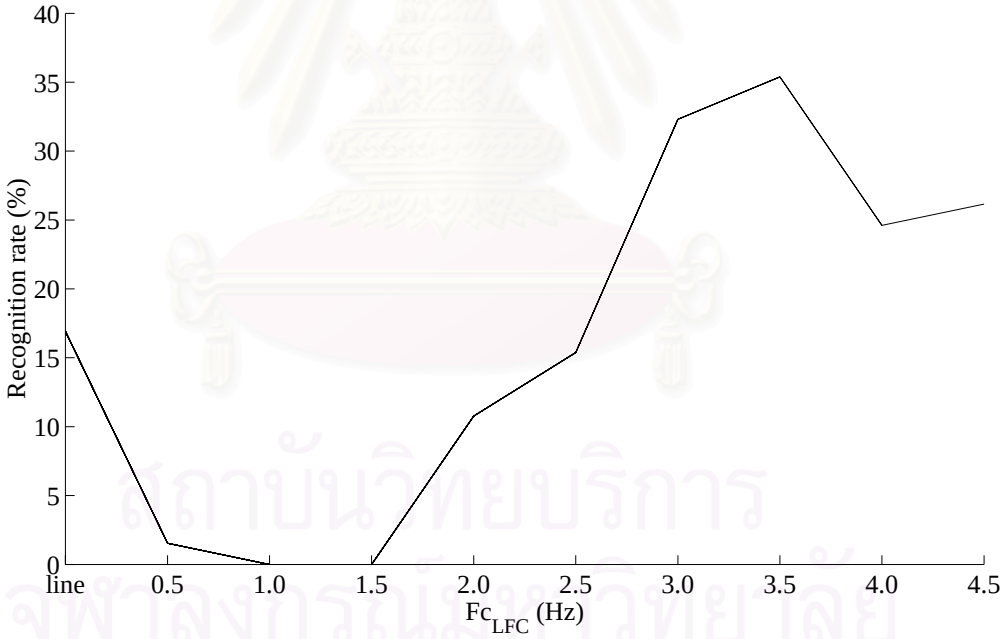
รูปที่ 4.9 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



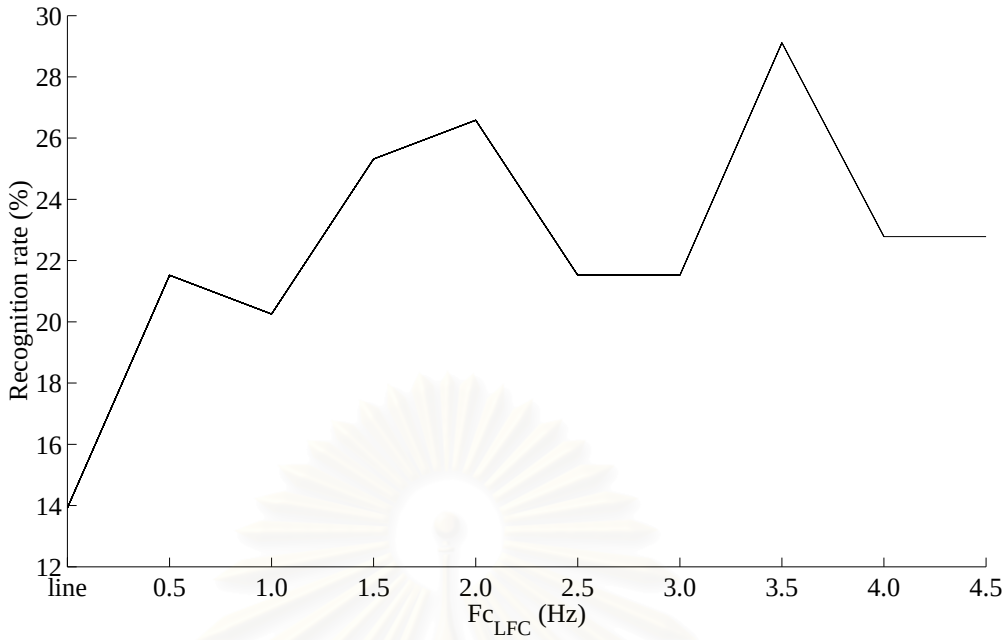
รูปที่ 4.10 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



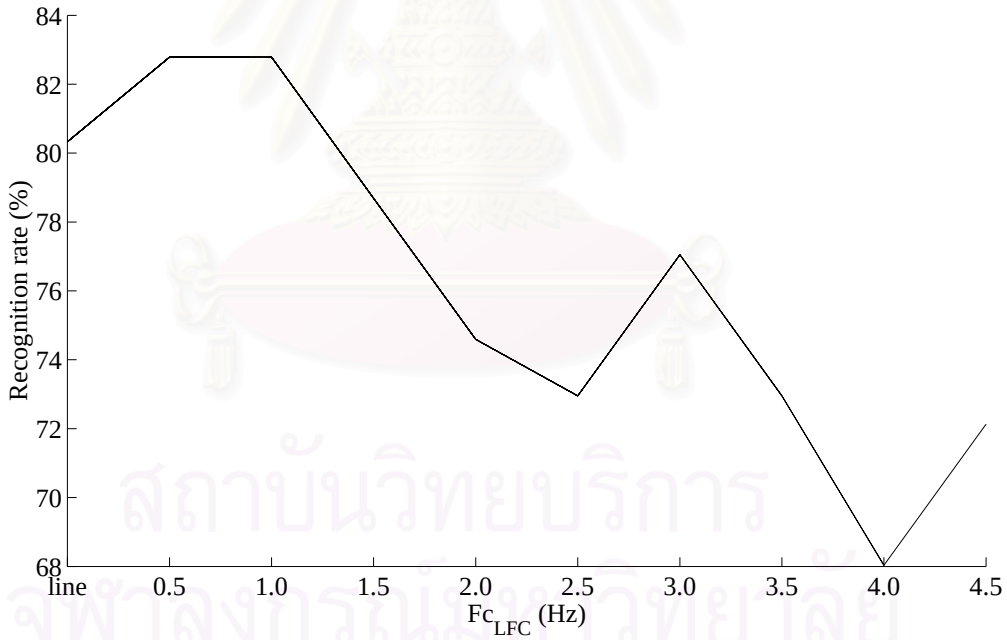
รูปที่ 4.11 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



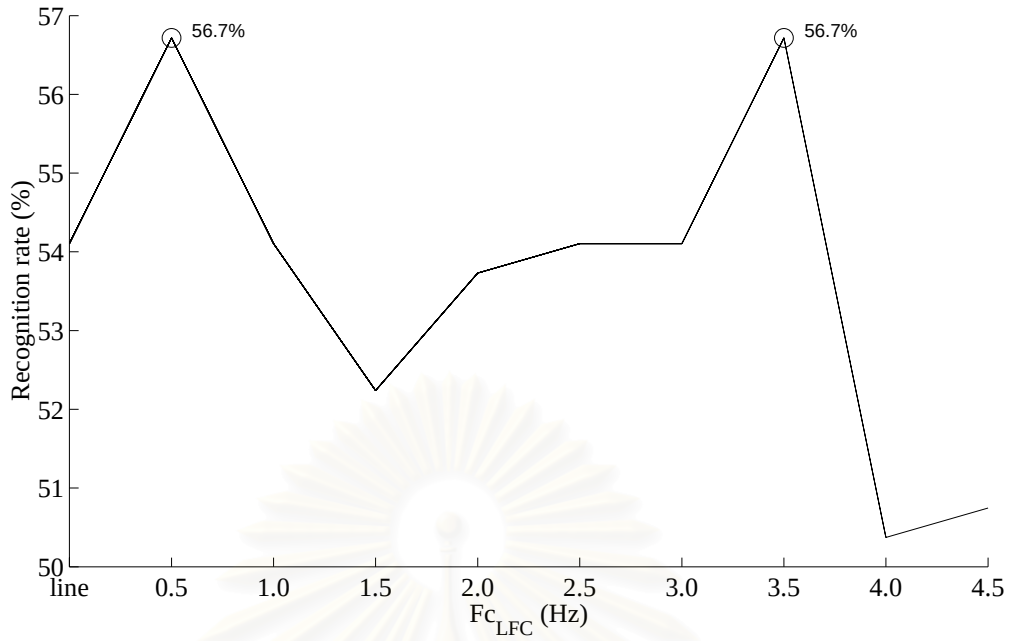
รูปที่ 4.12 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



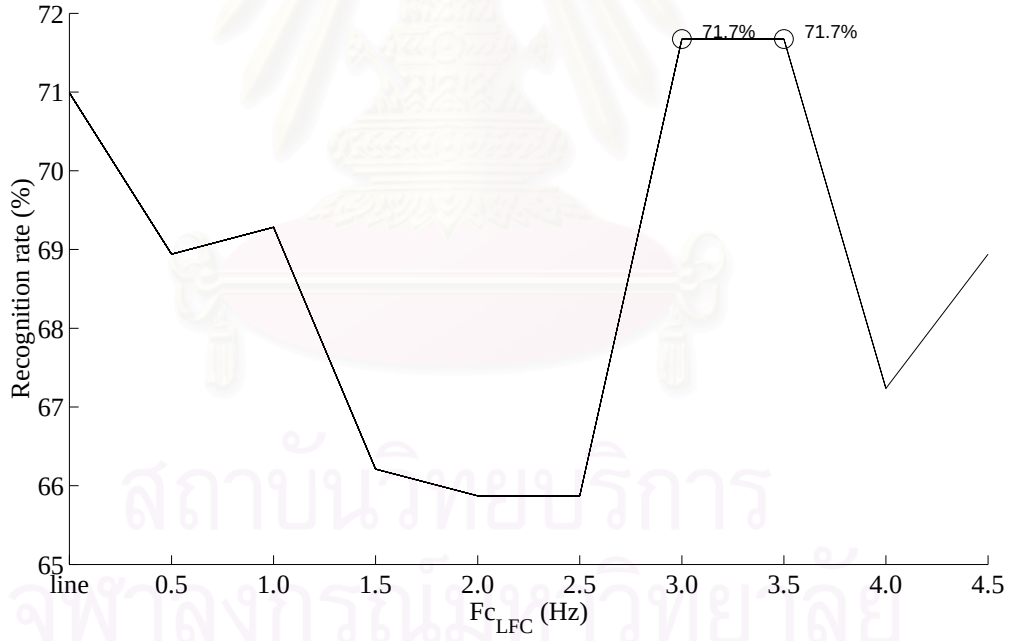
รูปที่ 4.13 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



รูปที่ 4.14 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



รูปที่ 4.15 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท



รูปที่ 4.16 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท

อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.15 และ 4.16 แสดงให้เห็นในตารางที่ 4.7 และ 4.8 ตามลำดับ

ตารางที่ 4.7 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย (จากรูปที่ 4.15)

ความถี่ตัด (Hz) หรือ ลักษณะของ LFC	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
อัตราการรู้จำเฉลี่ย (%)	54.1	**56.7**	54.1	52.2	53.7	54.1	54.1	**56.7**	50.4	50.7

ตารางที่ 4.8 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้หญิง (จากรูปที่ 4.16)

ความถี่ตัด (Hz) หรือ ลักษณะของ LFC	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
อัตราการรู้จำเฉลี่ย (%)	71.0	68.9	69.3	66.2	65.9	65.9	**71.7**	**71.7**	67.2	68.9

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.15 และ 4.16) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.9 – 4.12

จากตารางที่ 4.9 - 4.12 จะเห็นได้ว่าโอกาสที่โครงข่ายประสาทเทียมจะรู้จำทำนองเสียงตกผิดไปเป็นทำนองเสียงอื่นมีค่อนข้างน้อย เมื่อเทียบกับ ทำนองเสียงประเภทอื่น ๆ ในขณะที่โอกาสที่โครงข่ายประสาทเทียมจะรู้จำทำนองเสียงอื่น ๆ เป็นทำนองเสียงตกยังคงสูงอยู่ ซึ่งมีลักษณะเช่นเดียวกับการทดลองที่ 4.3.1

ถึงแม้ว่าอัตราการรู้จำเฉลี่ยของเสียงผู้ชายจะเท่ากับการทดลองในข้อ 4.3.1 ส่วนอัตราการรู้จำเฉลี่ยของเสียงผู้หญิงจะมากกว่าการทดลองในข้อ 4.3.1 เพียงเล็กน้อย แต่เมื่อพิจารณาผลการรู้จำในตารางที่ 4.9 – 4.12 จะพบว่า ค่าต่ำสุดของอัตราการรู้จำทำนองเสียงแต่ละประเภทสำหรับผู้พูดแต่ละเพศนั้นมีค่าเพิ่มขึ้น นั่นคือ ในกรณีของเสียงผู้ชาย ทำนองเสียงที่ให้อัตราการรู้จำต่ำสุด คือ ทำนองเสียงผสม จากตารางที่ 4.5 ในหัวข้อ 4.3.1 จะเห็นได้ว่า อัตราการรู้จำมีค่าเป็นร้อยละ 24.0 ใน

ขณะที่ในตารางที่ 4.10 อัตราการรู้จำมีค่าเพิ่มขึ้นเป็นร้อยละ 29.1 ส่วนในกรณีของเสียงผู้หญิง ทำนองเสียงที่ให้อัตราการรู้จำต่ำสุด คือ ทำนองเสียงขึ้น จาก ตารางที่ 4.6 ในหัวข้อ 4.3.1 อัตราการรู้จำมีค่าเป็นร้อยละ 12.3 ในขณะที่ในตารางที่ 4.12 อัตราการรู้จำมีค่าเพิ่มขึ้นเป็นร้อยละ 35.4 แสดงให้เห็นว่า ค่า ΔLFC มีผลทำให้ผลการรู้จำทำนองเสียงดีขึ้น

ตารางที่ 4.9 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.15) ของเสียง ผู้ชาย เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 0.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	**ประเภทของทำนองเสียงที่รู้จำได้**			**จำนวนประโยค**
	ตก	**ขึ้น**	**ผสม**	
ตก	**87.8**	**11.2**	1.0	98
ขึ้น	**28.6**	**53.8**	17.6	91
ผสม	32.9	**45.6**	**21.5**	79
	จำนวนประโยคทั้งหมด			268

ตารางที่ 4.10 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.15) ของเสียงผู้ชาย เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	**ประเภทของทำนองเสียงที่รู้จำได้**			**จำนวนประโยค**
	ตก	**ขึ้น**	**ผสม**	
ตก	**83.7**	**13.3**	3.1	98
ขึ้น	**28.6**	**51.6**	19.8	91
ผสม	25.3	**45.6**	**29.1**	79
	จำนวนประโยคทั้งหมด			268

ตารางที่ 4.11 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.16) ของเสียงผู้หญิง เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.0 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**89.6**	3.8	**6.6**	106
ขึ้น	16.9	**32.3**	**50.8**	65
ผสม	**13.1**	9.8	**77.0**	122
	จำนวนประโยคทั้งหมด			293

ตารางที่ 4.12 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.16) ของเสียงผู้หญิง เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**92.5**	**4.7**	2.8	106
ขึ้น	20.0	**35.4**	**44.6**	65
ผสม	13.1	**13.9**	**73.0**	122
	จำนวนประโยคทั้งหมด			293

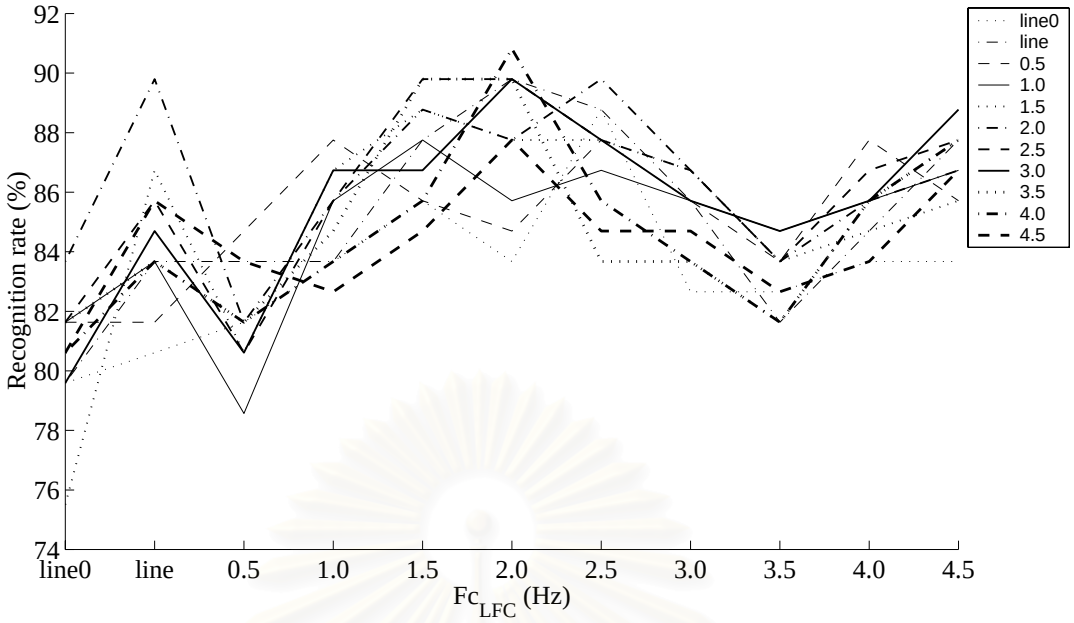
4.3.3 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยค

การทดลองนี้ใช้จุดที่ได้จากการสุ่มตัวอย่างคอนทัวร์ LFC และ FVC รวมทั้งความยาวของประโยค มาเป็นเวกเตอร์ลักษณะสำหรับโครงข่ายประสาทเทียม โดยใช้ค่าความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC และ FVC เป็น 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, และ 4.5 Hz รวมทั้งใช้ LFC ที่เป็นเส้นตรง (line) และเส้นตรงที่มีความชันเป็น 0 (line0) และเนื่องจากรูปร่างของคอนทัวร์ FVC ขึ้นกับความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC ด้วย การทดลองนี้จึงได้จับคู่ LFC และ FVC ที่ค่าความถี่ตัดต่าง ๆ จนครบทุกกรณีที่เป็นไปได้ ทำให้มีการทดลองฝึกฝน และทดสอบโครงข่ายประสาทเทียมทั้งสิ้น 121 การทดลอง ที่ค่าความถี่ตัดของ LFC และ FVC ต่าง ๆ กัน

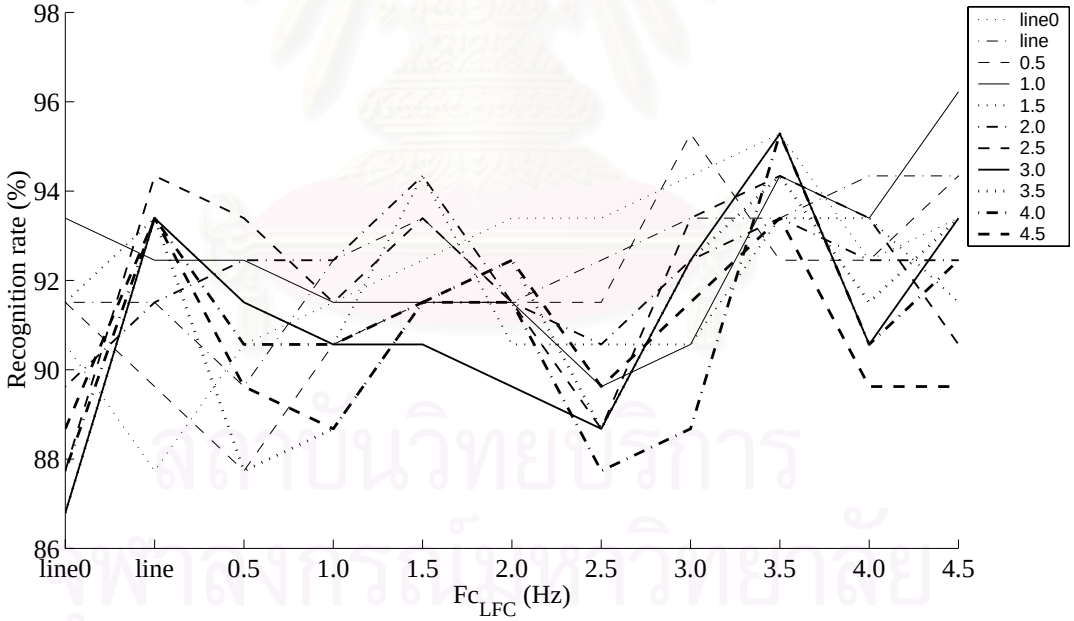
อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.17 – 4.22 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.23 และ 4.24

จากรูปที่ 4.17 – 4.22 จะเห็นได้ว่า อัตราการรู้จำของทำนองเสียงประเภทต่าง ๆ โดยส่วนใหญ่จะอยู่ในช่วงใกล้เคียงกับการทดลองในข้อ 4.31 และ 4.32 แต่จะให้อัตราการรู้จำที่สูงกว่าการทดลองก่อน ๆ เล็กน้อย ยกเว้นกรณีของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้ชาย จะเห็นได้ว่าอัตรารู้จำเพิ่มสูงขึ้นจนอยู่ในช่วงร้อยละ 15 ถึง 55 มากกว่าเดิมในข้อ 4.3.1 (ร้อยละ 16 ถึง 24) และ 4.3.2 (ร้อยละ 14 ถึง 28)

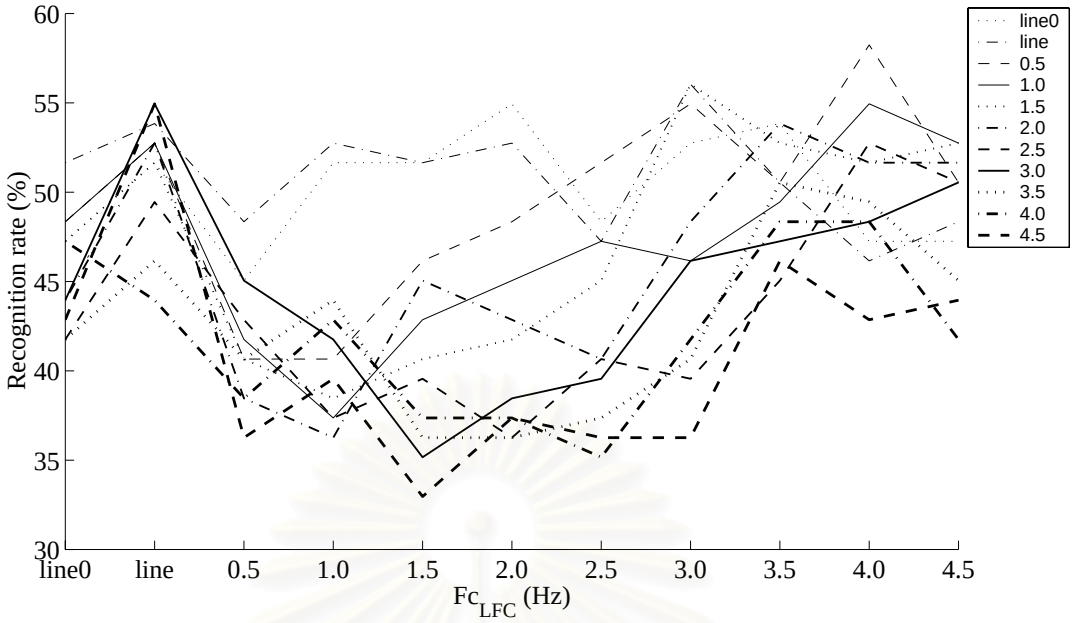
เมื่อพิจารณารูปที่ 4.23 และ 4.24 จะเห็นได้ว่า อัตราการรู้จำเฉลี่ยของทำนองเสียงทั้งสามประเภทของเสียงผู้หญิงจะมากกว่าเสียงผู้ชาย โดยอัตราการรู้จำเฉลี่ยของเสียงผู้ชายอยู่ที่ร้อยละ 61.6 ซึ่งเกิดขึ้นเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.5 Hz ของ FVC เป็น 2.0 และ 3.0 Hz ส่วนอัตราการรู้จำเฉลี่ยของเสียงผู้หญิงที่สูงที่สุดอยู่ที่ร้อยละ 73.0 ซึ่งเกิดขึ้นที่ค่าความถี่ตัดของ LFC เป็น 1.0 Hz เมื่อ FVC เป็นเส้นตรง และ ค่าความถี่ตัดของ LFC เป็น 3.0 Hz เมื่อค่าความถี่ตัดของ FVC เป็น 2.0 Hz



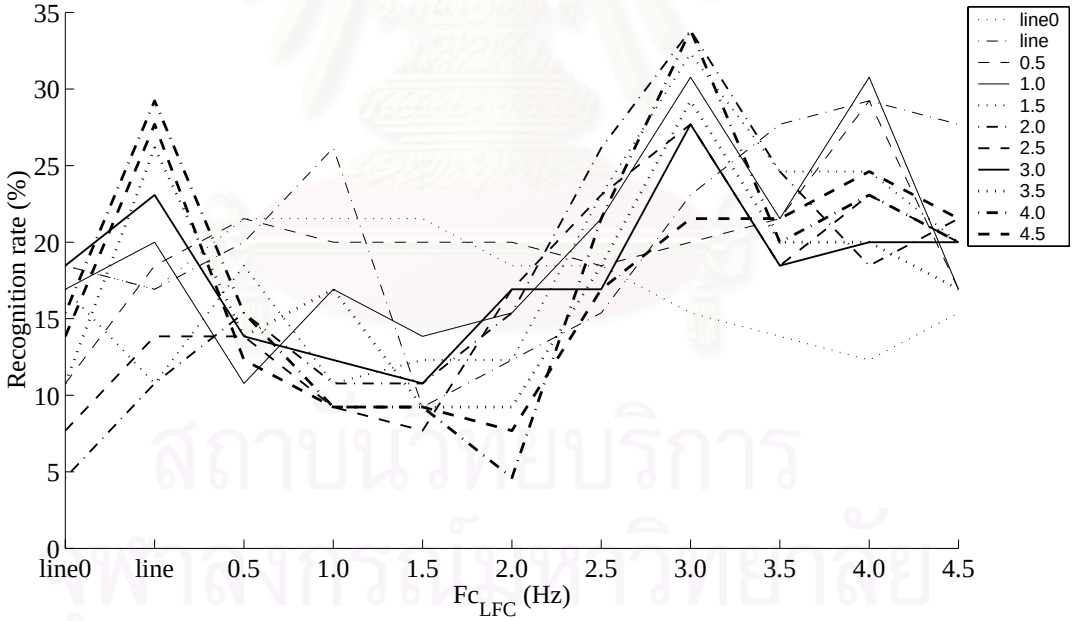
รูปที่ 4.17 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



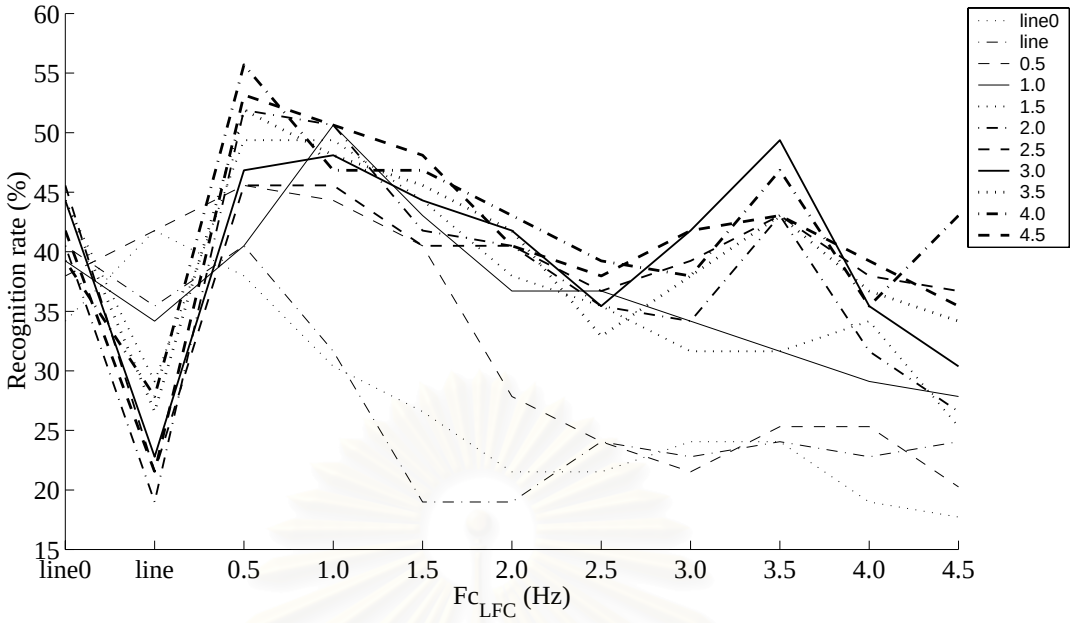
รูปที่ 4.18 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



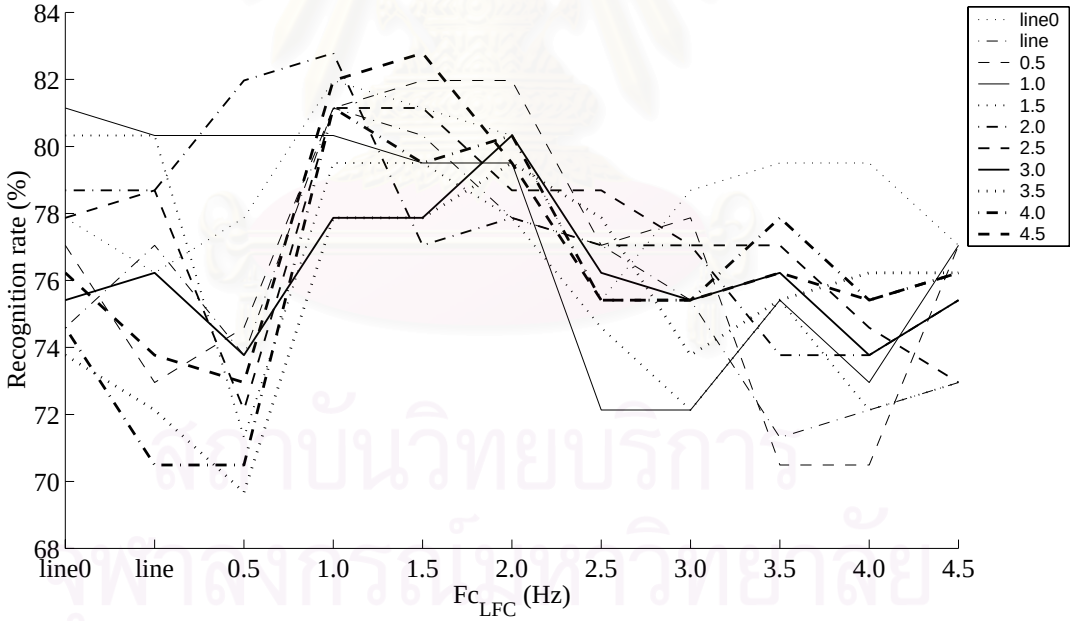
รูปที่ 4.19 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



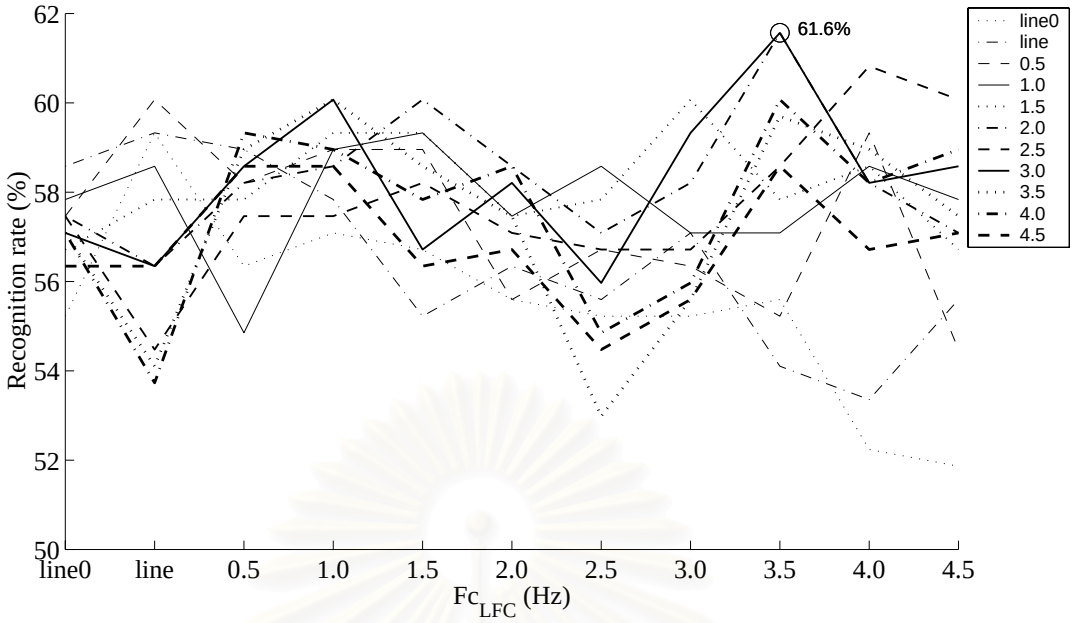
รูปที่ 4.20 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



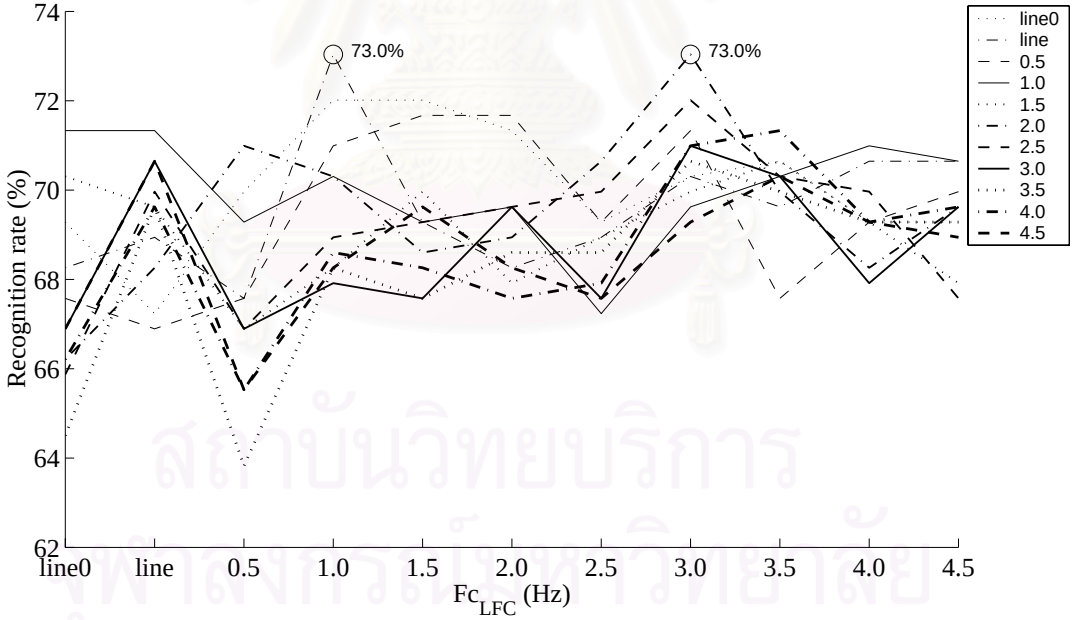
รูปที่ 4.21 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.22 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.23 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.24 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)

อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.23 และ 4.24 แสดงให้เห็นในตารางที่ 4.13 และ 4.14 ตามลำดับ

ตารางที่ 4.13 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้ชาย (จากรูปที่ 4.23)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line0	55.2	58.6	57.5	57.8	56.7	57.5	57.5	57.1	57.1	57.1	56.3
line	59.3	59.3	60.1	58.6	57.8	56.3	54.5	56.3	54.1	53.7	56.3
0.5	56.3	59.0	58.2	54.9	57.8	58.2	57.5	58.6	59.0	59.3	58.6
1.0	57.1	57.8	59.0	59.0	59.3	58.6	57.5	60.1	60.1	59.0	58.6
1.5	56.7	55.2	59.0	59.3	59.3	60.1	58.2	56.7	58.6	57.8	56.3
2.0	55.6	56.3	55.6	57.5	57.5	58.6	57.1	58.2	57.5	58.6	56.7
2.5	55.2	55.6	56.7	58.6	57.8	57.1	56.7	56.0	53.0	54.9	54.5
3.0	55.2	57.1	56.3	57.1	60.1	58.2	56.7	59.3	55.6	56.0	55.6
3.5	55.6	54.1	55.2	57.1	57.8	**61.6**	58.6	**61.6**	59.7	60.1	58.6
4.0	52.2	53.4	59.3	58.6	58.6	58.2	60.8	58.2	59.0	58.2	56.7
4.5	51.9	55.6	54.5	57.8	56.7	57.1	60.1	58.6	57.5	59.0	57.1

ตารางที่ 4.14 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้หญิง (จากรูปที่ 4.24)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line0	69.3	68.3	67.6	71.3	70.3	66.2	65.9	66.9	64.5	66.2	66.9
line	67.2	68.9	66.9	71.3	69.6	68.3	70.0	70.6	69.6	69.6	70.6
0.5	70.0	67.6	67.6	69.3	66.9	71.0	66.9	66.9	63.8	65.5	65.5
1.0	72.0	**73.0**	71.0	70.3	68.3	70.3	68.9	67.9	68.3	68.6	68.3
1.5	72.0	69.3	71.7	69.3	70.0	68.6	69.3	67.6	67.6	68.3	69.6
2.0	71.3	68.3	71.7	69.6	67.9	68.9	69.6	69.6	68.6	67.6	68.3
2.5	69.3	68.9	69.3	67.2	68.9	70.6	70.0	67.6	68.6	67.9	67.6
3.0	70.3	70.3	71.3	69.6	70.0	**73.0**	72.0	71.0	70.6	71.0	69.3
3.5	70.6	69.6	67.6	70.3	70.6	70.0	70.3	70.3	70.0	71.3	70.3
4.0	69.3	70.6	69.3	71.0	69.3	68.3	70.0	67.9	69.3	69.3	69.3
4.5	69.3	70.6	70.0	70.6	67.9	69.6	67.6	69.6	69.3	69.6	68.9

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.23 และ 4.24) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.15 – 4.18

จากตารางที่ 4.15 - 4.18 จะเห็นได้ว่ามีลักษณะคล้ายกับการทดลองในหัวข้อ 4.3.1 และ 4.3.2 คือ อัตราการรู้จำทำนองเสียงตก สูงกว่าทำนองเสียงอื่น ๆ

จากตารางที่ 4.15 และ 4.16 จะเห็นได้ว่าในกรณีของเสียงผู้ชาย โอกาสที่โครงข่ายประสาทเทียมจะรู้จำทำนองเสียงผสมเป็นทำนองเสียงผสม (ร้อยละ 43.0 ในตารางที่ 4.15 และ ร้อยละ 49.4 ในตารางที่ 4.16) จะสูงกว่าโอกาสที่จะรู้จำทำนองเสียงผสมเป็นทำนองเสียงตก (ร้อยละ 21.5 ในตารางที่ 4.15 และ ร้อยละ 25.3 ในตารางที่ 4.16) และทำนองเสียงขึ้น (ร้อยละ 35.4 ในตารางที่ 4.15 และ ร้อยละ 25.3 ในตารางที่ 4.16) ซึ่งแตกต่างจากการทดลองก่อน ๆ ดังแสดงในตารางที่ 4.4 4.5 4.9 และ 4.10 โดยการทดลองครั้งก่อน ๆ เหล่านี้ถ้าให้ประโยคขาเข้าเป็นทำนองเสียงแบบ ผสมแล้วจะมีโอกาสที่จะให้ผลการรู้จำเป็นทำนองเสียงขึ้นสูงกว่าที่จะรู้จำเป็นทำนองเสียงผสมเอง

แต่ในกรณีของเสียงผู้หญิง จะเห็นว่ายังคงมีอัตราการรู้จำทำนองเสียงขึ้นไปเป็นทำนองเสียงผสม สูงกว่า ที่จะรู้จำถูกเป็นทำนองเสียงขึ้นเองอยู่

ตารางที่ 4.15 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.23) ของเสียงผู้ชาย เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.5 Hz และค่าความถี่ตัดของ FVC เป็น 2.0 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**83.7**	**12.2**	4.1	98
ขึ้น	**25.3**	**53.8**	20.9	91
ผสม	21.5	**35.4**	**43.0**	79
จำนวนประโยคทั้งหมด				268

ตารางที่ 4.16 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.23) ของเสียงผู้ชาย เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.5 Hz และค่าความถี่ตัดของ FVC เป็น 3.0 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**84.7**	**10.2**	5.1	98
ขึ้น	**29.7**	**47.3**	23.1	91
ผสม	25.3	**25.3**	**49.4**	79
จำนวนประโยคทั้งหมด				268

ตารางที่ 4.17 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.24) ของเสียงผู้หญิง เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 1.0 Hz และใช้ FVC เป็นเส้นตรง

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**92.5**	0.0	**7.5**	106
ขึ้น	16.9	**26.2**	**56.9**	65
ผสม	**15.6**	3.3	**81.1**	122
	จำนวนประโยคทั้งหมด			293

ตารางที่ 4.18 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.24) ของเสียงผู้หญิง เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.0 Hz และค่าความถี่ตัดของ FVC เป็น 2.0 Hz

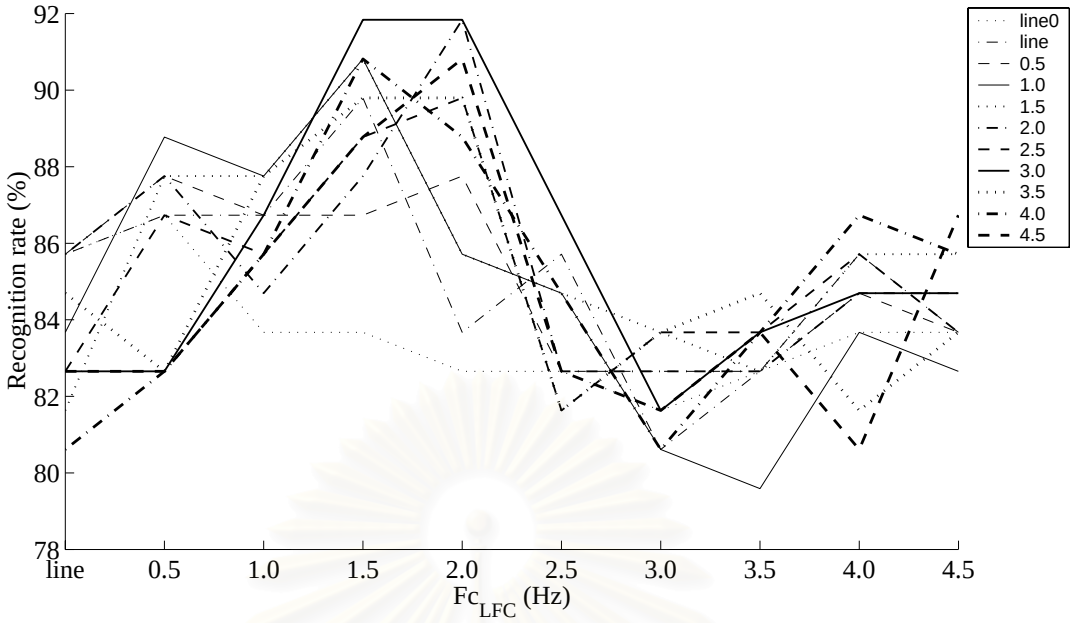
ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**92.5**	1.9	**5.7**	106
ขึ้น	20.0	**33.8**	**46.2**	65
ผสม	11.5	**11.5**	**77.0**	122
	จำนวนประโยคทั้งหมด			293

4.3.4 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยค

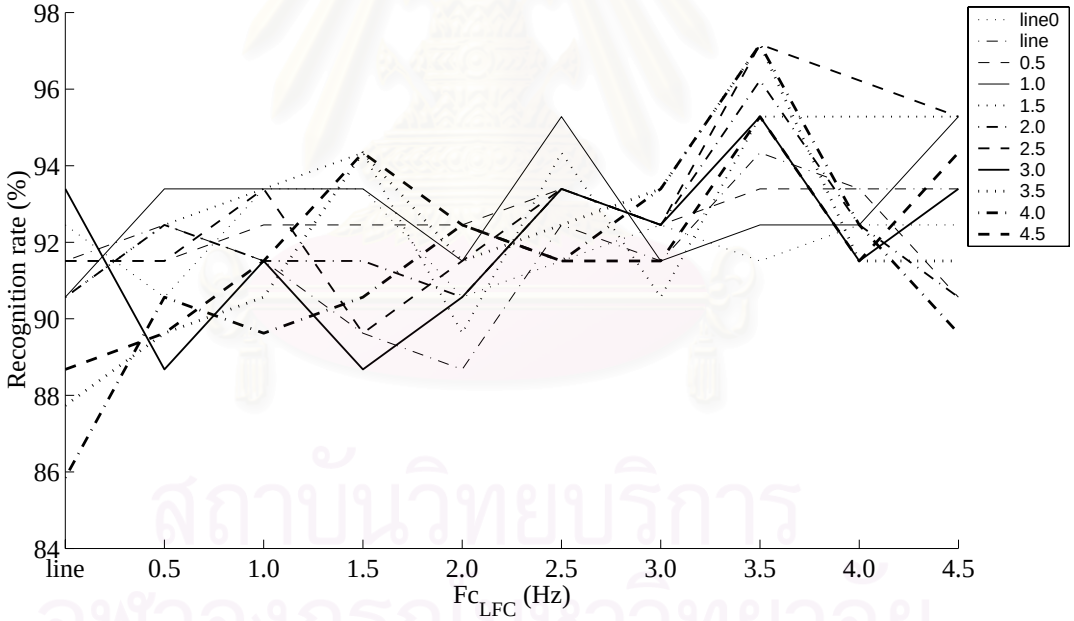
การทดลองนี้ใช้ลักษณะ LFC ΔLFC FVC และความยาวของประโยคมาเป็นเวกเตอร์ข้อมูลขาเข้าสำหรับโครงข่ายประสาทเทียม โดยใช้ความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC และ FVC เป็น 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, และ 4.5 Hz ในกรณีของ FVC ยังคงใช้ เส้นตรง และเส้นตรงที่มีความชันเป็น 0 เช่นเดียวกับการทดลองในข้อ 4.3.3 แต่ในกรณีของ คอนทัวร์ LFC นั้น ใช้เพียงเส้นตรง เนื่องจากเส้นตรงที่มีความชันเป็น 0 จะให้ค่า ΔLFC เป็น 0 เสมอ เช่นเดียวกับการทดลองในข้อ 4.3.2

อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.25 – 4.30 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.31 และ 4.32

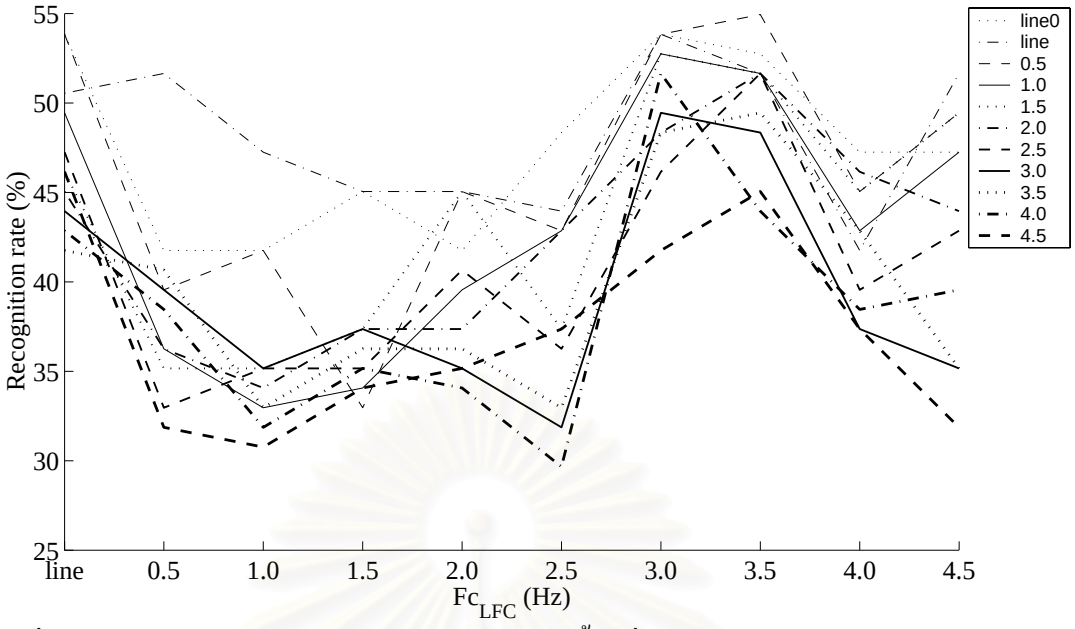
จากรูปที่ 4.25 – 4.30 จะเห็นได้ว่า อัตราการรู้จำทำนองเสียงประเภทต่าง ๆ โดยส่วนใหญ่จะมีลักษณะใกล้เคียงกับการทดลองในข้อ 4.3.3 ยกเว้นกรณีของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง จะเห็นได้ว่า อัตรารู้จำเพิ่มสูงขึ้นในช่วงที่ Fc_{LFC} มีค่าตั้งแต่ 2.5 Hz ขึ้นไป จากนั้นอัตรารู้จำจะเริ่มคงที่เมื่อ Fc_{LFC} มีค่าตั้งแต่ 3.5 Hz ขึ้นไป โดยมีค่าสูงสุดที่ร้อยละ 43.1 ที่ Fc_{LFC} = 3 Hz เมื่อ Fc_{FVC} = 1.5 Hz



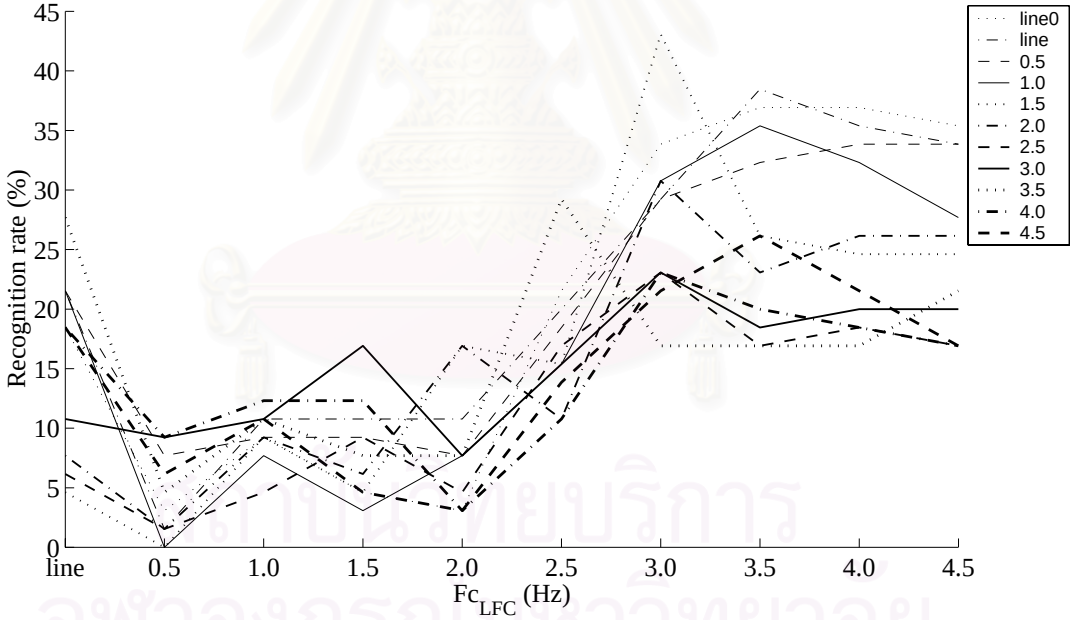
รูปที่ 4.25 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



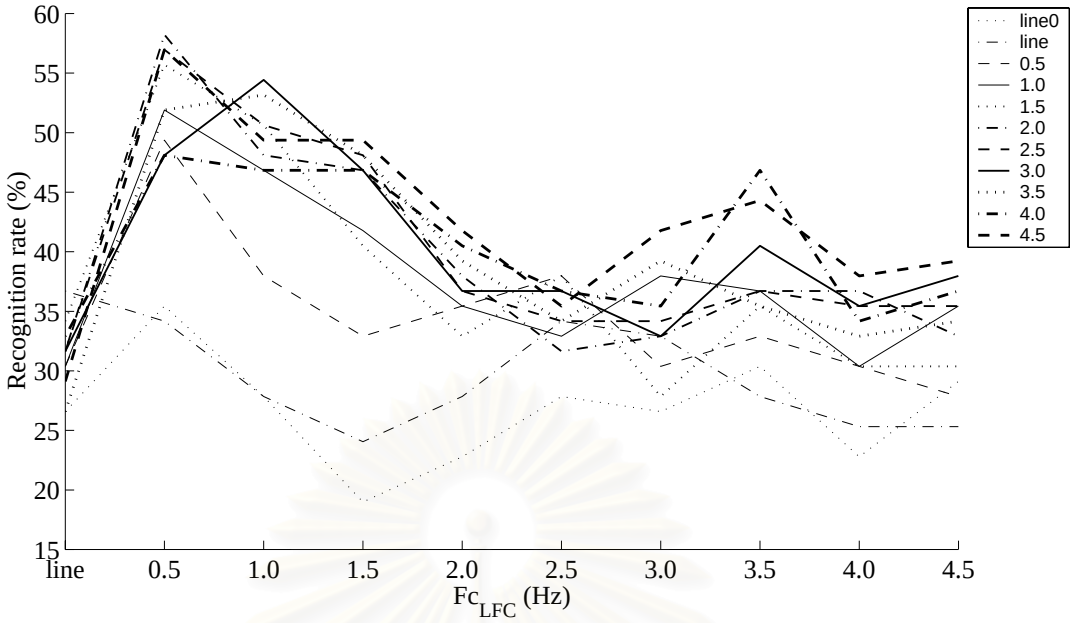
รูปที่ 4.26 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



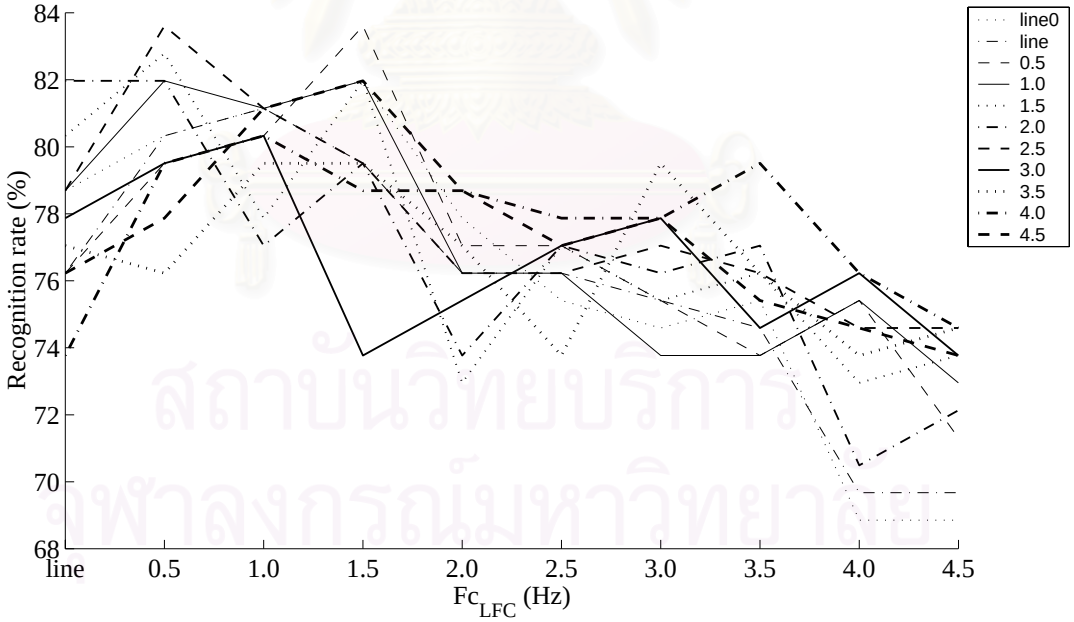
รูปที่ 4.27 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจาก คอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



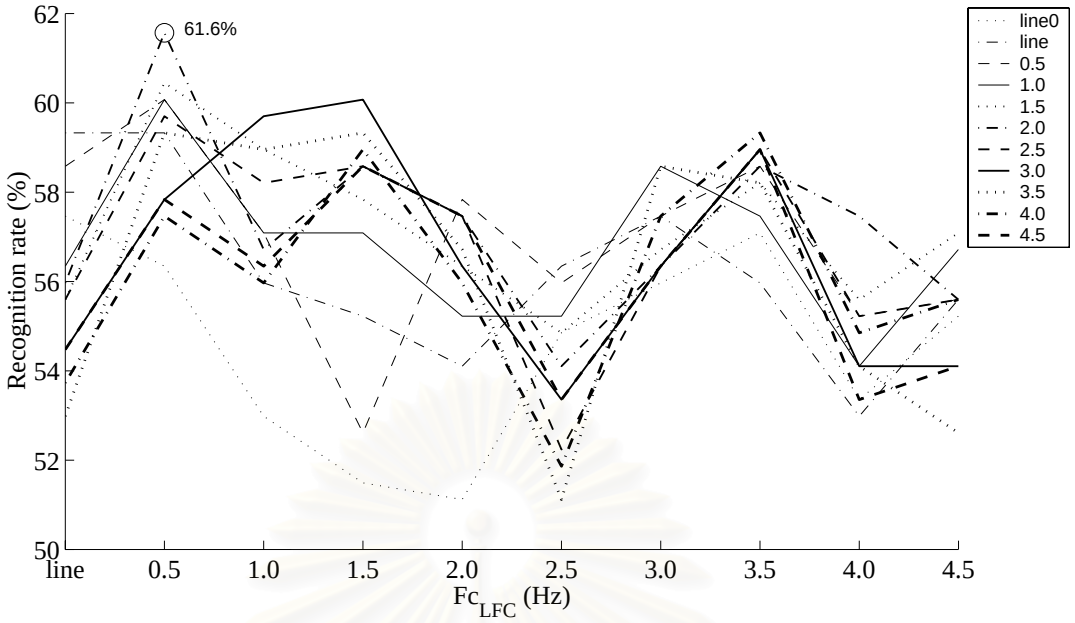
รูปที่ 4.28 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจาก คอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



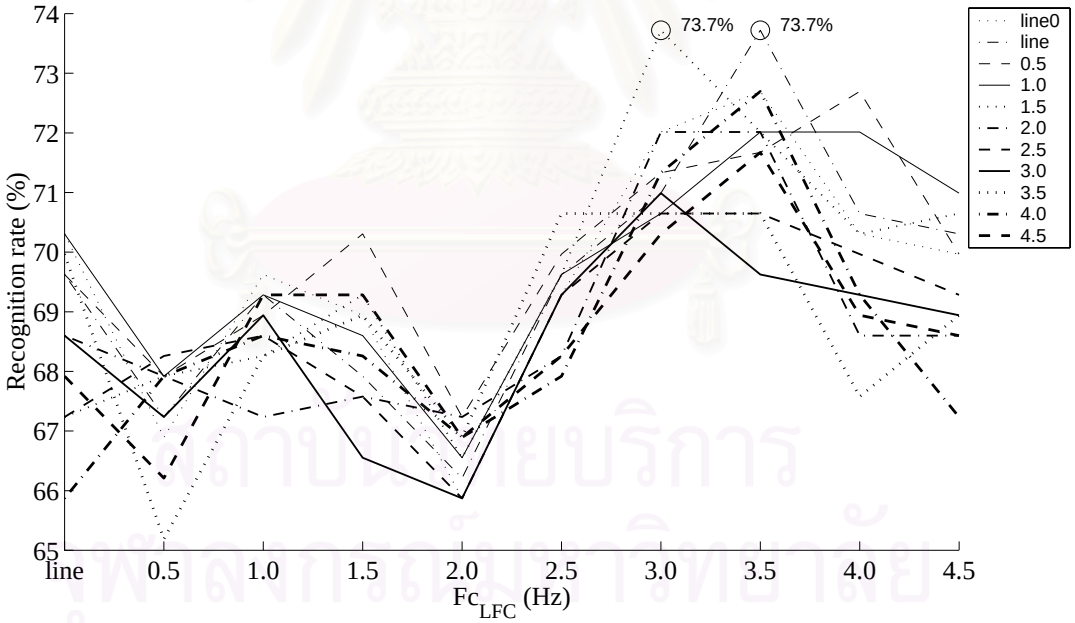
รูปที่ 4.29 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.30 อัตราการรู้จำทำนองเสียงของทำนองเสียงผสม เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.31 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยเมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.32 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยเมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)

อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.31 และ 4.32 แสดงให้เห็นในตารางที่ 4.19 และ 4.20 ตามลำดับ

ตารางที่ 4.19 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้ชาย (จากรูปที่ 4.31)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line	57.5	59.3	58.6	56.3	55.6	56.0	55.6	54.5	53.0	53.7	54.5
0.5	56.3	59.3	60.1	60.1	60.4	**61.6**	59.7	57.8	59.3	57.5	57.8
1.0	53.0	56.0	57.1	57.1	59.0	56.7	58.2	59.7	59.0	56.0	56.3
1.5	51.5	55.2	52.6	57.1	57.8	58.6	58.6	60.1	59.3	59.0	58.6
2.0	51.1	54.1	57.8	55.2	56.3	57.5	57.5	56.3	56.7	56.0	57.5
2.5	54.9	56.3	56.0	55.2	54.9	54.1	52.2	53.4	51.1	51.9	53.4
3.0	56.0	57.5	57.5	58.6	56.7	56.3	56.3	56.3	58.6	57.5	56.3
3.5	57.1	56.0	58.6	57.5	58.2	58.6	59.0	59.0	58.2	59.3	59.0
4.0	53.4	53.0	55.2	54.1	55.6	57.5	55.2	54.1	54.1	54.9	53.4
4.5	55.2	55.6	55.6	56.7	57.1	55.6	55.6	54.1	52.6	55.6	54.1

ตารางที่ 4.20 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้หญิง (จากรูปที่ 4.32)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line	70.3	69.6	69.6	70.3	67.2	68.6	67.2	68.6	70.0	65.9	67.9
0.5	66.9	67.2	67.9	67.9	67.9	67.9	68.3	67.2	65.2	67.9	66.2
1.0	69.6	69.3	68.9	69.3	68.3	67.2	68.6	68.9	68.3	68.6	69.3
1.5	68.9	67.9	70.3	68.6	69.3	67.6	67.6	66.6	68.9	68.3	69.3
2.0	65.9	66.2	67.2	66.6	66.6	67.2	65.9	65.9	66.9	66.9	66.9
2.5	69.3	69.6	70.0	69.6	69.6	68.3	69.3	69.3	70.6	67.9	68.3
3.0	72.0	71.0	71.3	70.6	**73.7**	72.0	70.6	71.0	70.6	71.3	70.3
3.5	72.7	**73.7**	71.7	72.0	72.0	72.0	70.6	69.6	70.6	72.7	71.7
4.0	70.3	70.6	72.7	72.0	70.3	68.6	70.0	69.3	67.6	69.3	68.9
4.5	70.0	70.3	70.0	71.0	70.6	68.6	69.3	68.9	68.9	67.2	68.6

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.31 และ 4.32) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.21 – 4.23

จากตารางที่ 4.21 ในกรณีของเสียงผู้ชาย จะเห็นได้ว่าอัตราการรู้จำของทำนองเสียงขึ้นมีค่าต่ำลง เมื่อเทียบกับการทดลองก่อน ๆ ในขณะที่กรณีของเสียงผู้หญิง อัตราการรู้จำของทำนองเสียงขึ้นมีค่าสูงขึ้น โดยเฉพาะในตารางที่ 4.23 ซึ่งอัตราการรู้จำทำนองเสียงขึ้นเป็นทำนองเสียงขึ้น สูงกว่าอัตราการรู้จำทำนองเสียงขึ้นเป็นทำนองเสียงประเภทอื่น ๆ ซึ่งแตกต่างจากการทดลองในหัวข้อก่อน ๆ ซึ่งโอกาสที่ตัวรู้จำจะรู้จำทำนองเสียงขึ้นผิดเป็นทำนองเสียงผสม สูงกว่าโอกาสที่จะรู้จำถูกเป็นทำนองเสียงขึ้นเองเสียอีก

ตารางที่ 4.21 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.31) ของเสียงผู้ชาย เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 0.5 Hz และค่าความถี่ตัดของ FVC เป็น 2.0 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**87.8**	**9.2**	3.1	98
ขึ้น	**34.1**	**36.3**	29.7	91
ผสม	**25.3**	16.5	**58.2**	79
	จำนวนประโยคทั้งหมด			268

ตารางที่ 4.22 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.32) ของเสียงผู้หญิง เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.5 Hz และใช้ FVC เป็นเส้นตรง

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**94.3**	**2.8**	2.8	106
ขึ้น	18.5	**38.5**	**43.1**	65
ผสม	11.5	**13.9**	**74.6**	122
	จำนวนประโยคทั้งหมด			293

ตารางที่ 4.23 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.32) ของเสียงผู้หญิง เมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.0 Hz และค่าความถี่ตัดของ FVC เป็น 1.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้			จำนวนประโยค
	ตก	ขึ้น	ผสม	
ตก	**90.6**	0.9	**8.5**	106
ขึ้น	16.9	**43.1**	**40.0**	65
ผสม	11.5	**13.1**	**75.4**	122
	จำนวนประโยคทั้งหมด			293

4.3.5 การเปรียบเทียบอัตราการรู้จำทำนองเสียง โดยใช้ลักษณะแต่ละแบบ กรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท

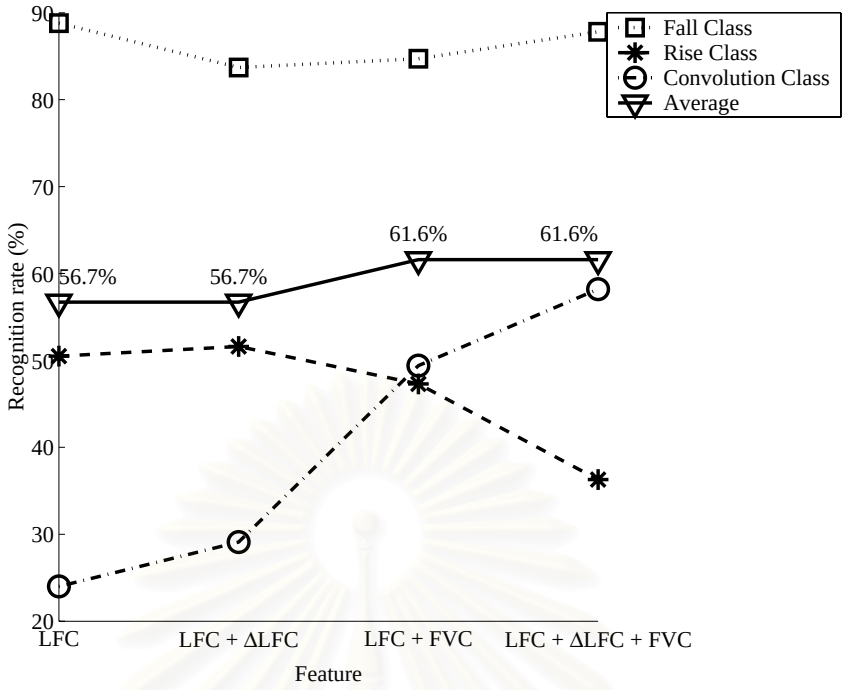
หัวข้อนี้กล่าวถึงการเปรียบเทียบอัตราการรู้จำทำนองเสียง โดยใช้ลักษณะต่าง ๆ กัน ตั้งแต่หัวข้อ 4.3.1 – 4.3.4 ซึ่งเป็นการรู้จำทำนองเสียงพูด โดยแบ่งทำนองเสียงออกเป็น 3 ประเภท คือ ทำนองเสียงตก ทำนองเสียงขึ้น และทำนองเสียงผสม

รูปที่ 4.33 และ 4.35 แสดงอัตราการรู้จำของผู้พูดที่เป็นผู้ชาย และผู้พูดที่เป็นผู้หญิง เมื่อใช้ลักษณะแบบต่าง ๆ จากการทดลองในหัวข้อ 4.3.1 – 4.3.4 โดยเลือกจากการทดลองย่อยของแต่ละหัวข้อที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุดในแต่ละเพศของผู้พูด ในกรณีที่มีการทดลองย่อยสองการทดลองที่ให้อัตราการรู้จำเฉลี่ยเท่ากัน ก็เลือกการทดลองย่อยที่ให้อัตราการรู้จำของทำนองเสียงที่มีค่าต่ำที่สุดมีค่ามากกว่า ตัวอย่างเช่น การทดลองในข้อ 4.3.4 กรณีที่ผู้พูดเป็นผู้หญิง มีการทดลองย่อยที่ให้ค่าอัตราการรู้จำเฉลี่ยสูงที่สุด (ร้อยละ 73.7) เท่ากันอยู่สองการทดลองย่อย ดังแสดงในตารางที่ 4.22 และ 4.23 ในตารางที่ 4.22 ทำนองเสียงที่มีอัตราการรู้จำต่ำที่สุด คือ ทำนองเสียงขึ้น (ร้อยละ 38.5) ส่วนในตารางที่ 4.23 ทำนองเสียงที่มีอัตราการรู้จำต่ำที่สุด คือ ทำนองเสียงขึ้น (ร้อยละ 43.1) ในกรณีนี้จึงเลือกการทดลองย่อยในตารางที่ 4.23 เนื่องจากร้อยละ 43.1 มากกว่าร้อยละ 38.5

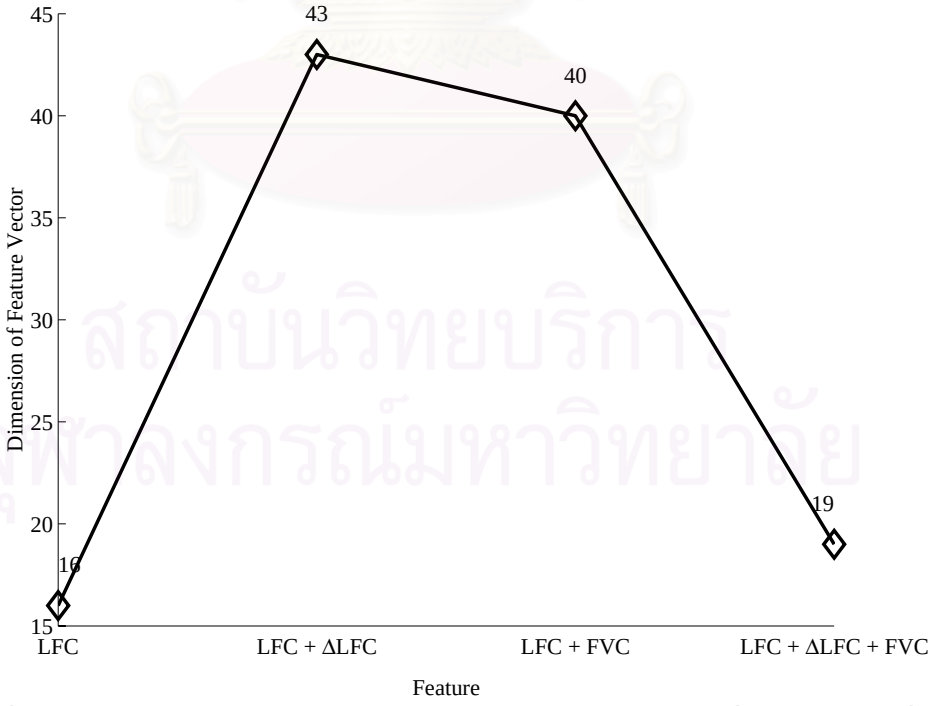
หัวข้อนี้ยังได้เปรียบเทียบจำนวนมิติของเวกเตอร์ลักษณะขาเข้าของโครงข่ายประสาทเทียมที่ใช้ในการทดลองที่เลือกมาในแต่ละหัวข้อด้วย เนื่องจากจำนวนมิติจะบอกถึงความซับซ้อนของโครงข่ายประสาทเทียม โดยโครงข่ายประสาทเทียมที่มีจำนวนมิติมากก็จะมีความ ซับซ้อนมาก และใช้เวลาในการฝึกฝนและทดสอบโครงข่ายมากขึ้นตามไปด้วย นอกจากนี้จำนวนมิติของเวกเตอร์ลักษณะขาเข้ายังแสดงถึงขนาดของพื้นที่ที่ใช้ในการเก็บข้อมูลด้วย

จากรูปที่ 4.33 และ 4.34 ซึ่งเป็นกรณีที่ผู้พูดเป็นผู้ชาย จะเห็นได้ว่าการใช้เพียงคอนทัวร์ LFC และความยาวของประโยคนั้น จะทำให้ผลการรู้จำทำนองเสียงมีค่าต่ำมาก โดยเฉพาะอย่างยิ่งการรู้จำทำนองเสียงผสม จะทำได้ไม่ดีนัก และจะเห็นได้ว่า การเพิ่ม ΔLFC หรือ FVC เข้าไป จะช่วยปรับปรุงอัตราการรู้จำได้ โดยจะเห็นได้ว่าการเพิ่ม FVC เข้าไปจะช่วยปรับปรุงอัตราการรู้จำได้ดีกว่าการเพิ่ม ΔLFC

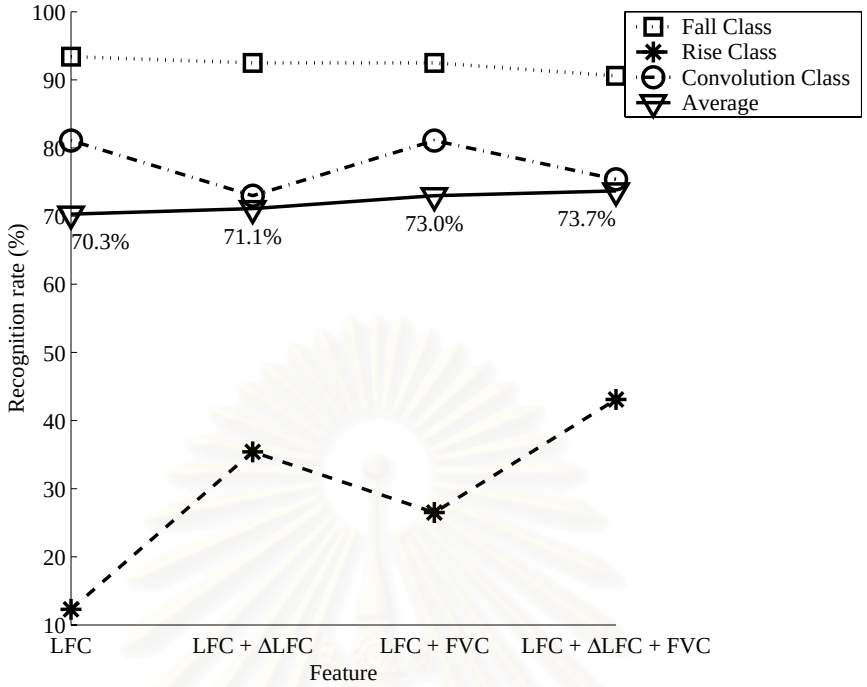
และเมื่อรวมทั้ง LFC FVC ΔLFC เข้าไปเป็นเวกเตอร์ลักษณะ จะเห็นได้ว่าอัตราการรู้จำไม่ได้เพิ่มมากไปกว่าการใช้เพียง LFC และ FVC อย่างไรก็ตาม จากรูปที่ 4.34 จะเห็นได้ว่าการใช้LFC FVC และ ΔLFC จะช่วยให้เวกเตอร์ลักษณะมีจำนวนมิติที่ลดลง กว่าการใช้เพียง LFC และ FVC เนื่องจากใช้จำนวนจุดสุ่มตัวอย่างคอนทัวร์ที่น้อยกว่า



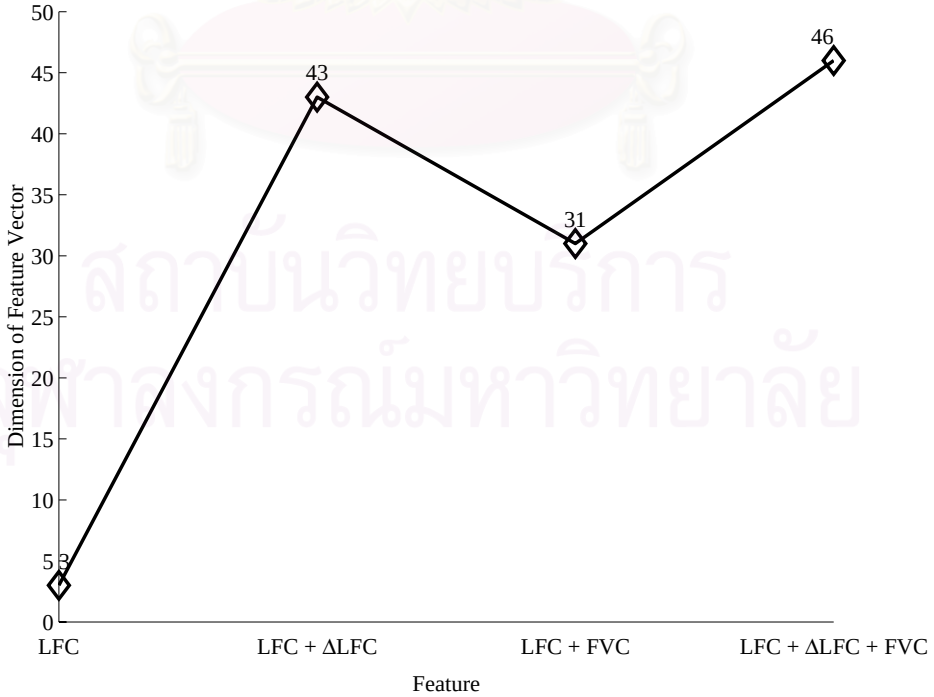
รูปที่ 4.33 อัตราการรู้จำเฉลี่ย และอัตราการรู้จำของแต่ละทำนองเสียง เลือกจากการทดลองย่อยที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุดเมื่อใช้เวกเตอร์ลักษณะแบบต่าง ๆ เมื่อผู้พูดเป็นผู้ชาย กรณีที่แบ่งทำนองเสียงเป็น 3 ประเภท (ลักษณะทุกแบบที่แสดงในกราฟ ได้รวมเอาความยาวของประโยค เข้าไปไว้ในเวกเตอร์ลักษณะด้วย)



รูปที่ 4.34 จำนวนมิติของเวกเตอร์ลักษณะขาเข้า ของการทดลองย่อยที่เลือกมาในรูปที่ 4.33



รูปที่ 4.35 อัตราการรู้จำเฉลี่ย และอัตราการรู้จำของแต่ละทำนองเสียง เลือกจากการทดลองย่อยที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุดเมื่อใช้เวกเตอร์ลักษณะแบบต่าง ๆ เมื่อผู้พูดเป็นผู้หญิง กรณีที่แบ่งทำนองเสียงเป็น 3 ประเภท (ลักษณะทุกแบบที่แสดงในกราฟ ได้รวมเอาความยาวของประโยค เข้าไปไว้ในเวกเตอร์ลักษณะด้วย)



รูปที่ 4.36 จำนวนมิติของเวกเตอร์ลักษณะขาเข้า ของการทดลองย่อยที่เลือกมาในรูปที่ 4.35

ในกรณีที่ผู้พูดเป็นผู้หญิงจะเห็นได้ว่าการใช้เพียงคอนทัวร์ LFC และความยาวของประโยคจะทำให้ได้อัตราการรู้จำเฉลี่ยที่สูงกว่าผู้พูดที่เป็นผู้ชาย แต่อย่างไรก็ตาม อัตราการรู้จำของทำนองเสียงขึ้นยังมีค่าต่ำมาก ซึ่งเมื่อพิจารณาจากตารางที่ 4.6 จะเห็นได้ว่า การที่อัตราการรู้จำทำนองเสียงขึ้นมีค่าต่ำเนื่องจากตัวรู้จำรู้จำทำนองเสียงขึ้นผิดไปเป็นทำนองเสียงผสมถึงร้อยละ 64.6 แสดงให้เห็นว่า การใช้เพียง LFC และความยาวของประโยค ถึงแม้ว่าจะทำให้ตัวรู้จำสามารถรู้จำแนกความแตกต่างระหว่างทำนองเสียงตก และทำนองเสียงผสมได้ดี แต่ตัวรู้จำก็แทบจะไม่สามารถจำแนกความแตกต่างระหว่างทำนองเสียงขึ้น และทำนองเสียงผสมได้

เมื่อเพิ่ม ΔLFC หรือ FVC เข้าไป จะเห็นได้ว่าสามารถช่วยปรับปรุงอัตราการรู้จำให้ดีขึ้นได้ เช่นเดียวกับเสียงผู้ชาย โดยจะเห็นได้ว่า FVC จะช่วยเพิ่มอัตราการรู้จำเฉลี่ยได้ดีกว่า ΔLFC เช่นเดียวกับเสียงผู้ชาย แต่ในกรณีของเสียงผู้หญิง ΔLFC สามารถเพิ่มอัตราการรู้จำทำนองเสียงขึ้นได้ดีกว่า FVC และเมื่อใช้ทั้ง LFC FVC และ ΔLFC เป็นเวกเตอร์ลักษณะจะเห็นได้ว่า อัตราการรู้จำเสียงพูดจะมีค่าสูงที่สุด โดยที่ไม่ได้เพิ่มจำนวนมิติของเวกเตอร์ลักษณะมากนัก

จะเห็นได้ว่า การรู้จำทำนองเสียงของทั้งสองเพศ ในกรณีที่แบ่งทำนองเสียงพูดออกเป็น 3 ประเภท การเพิ่ม FVC และ ΔLFC เข้าไปเป็นเวกเตอร์ลักษณะ สามารถช่วยให้อัตราการรู้จำทำนองเสียงสูงกว่าการใช้เพียง LFC และความยาวของประโยค โดยจะเห็นได้ว่าการเพิ่ม FVC จะช่วยให้อัตราการรู้จำสูงขึ้นกว่าการเพิ่ม ΔLFC ทั้งกรณีเสียงผู้ชาย และเสียงผู้หญิง ส่วนผลของการเพิ่ม ΔLFC นั้นเสียงผู้หญิงจะเห็นผลการปรับปรุงอัตราการรู้จำได้ดีกว่ากรณีของเสียงผู้ชาย

เมื่อพิจารณาผลการรู้จำโดยรวมในกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท จะเห็นได้ว่าตัวรู้จำสามารถจำแนกความแตกต่างของทำนองเสียงตก ออกจากทำนองเสียงอื่น ๆ ได้ดี แต่รู้จำทำนองเสียงขึ้นกับทำนองเสียงผสมได้ไม่ดีนัก โดยในกรณีของเสียงผู้ชายนั้นตัวรู้จำพอจะแยกความแตกต่างระหว่างทำนองเสียงขึ้น และทำนองเสียงผสมได้บ้าง แต่ในกรณีของเสียงผู้หญิงนั้น ถ้าทำนองเสียงขาเข้าเป็นทำนองเสียงขึ้น ตัวรู้จำมีโอกาสที่จะรู้จำเป็นทำนองเสียงผสมสูงกว่าโอกาสที่จะรู้จำเป็นทำนองเสียงขึ้นเองเสียอีก แสดงให้เห็นว่า ทำนองเสียงขึ้น และทำนองเสียงผสมมีคอนทัวร์ F_0 ที่ใกล้เคียงกันมาก จึงจำเป็นต้องมีการศึกษาเพิ่มเติม เพื่อนำลักษณะอื่น ๆ ของสัญญาณเสียง มาใช้ในปรับปรุงระบบ เพื่อเพิ่มประสิทธิภาพในการจำแนกความแตกต่างระหว่างทำนองเสียงทั้ง 2 ประเภทนี้

เนื่องจากการใช้คอนทัวร์ LFC และ FVC สามารถนำมาใช้จำแนกความแตกต่างระหว่างทำนองเสียงตกกับทำนองเสียงอื่น ๆ ได้ดี ทำให้สามารถนำคอนทัวร์ทั้งสองมาใช้จำแนกทำนองเสียงตก ออกจากทำนองเสียงประเภทอื่นก่อน แล้วจึงใช้ลักษณะอื่น ๆ ในการจำแนกทำนองเสียงขึ้น และทำนองเสียงผสม จึงกำหนดให้มีการทดลองในหัวข้อ 4.4 ซึ่งเป็นการทดลองที่ใช้ลักษณะต่าง ๆ ที่เหมือนกับการทดลองในหัวข้อ 4.3 นี้ทุกประการ แต่จะกำหนดให้มีทำนองเสียงเพียง 2 ประเภท โดยรวมทำนองเสียงผสมเข้าไปเป็นประเภทเดียวกับทำนองเสียงขึ้น เพื่อทดสอบประสิทธิภาพของคอนทัวร์ LFC และ FVC ในการจำแนกทำนองเสียงตกออกจากทำนองเสียงประเภทอื่น

4.4 การรู้จำทำนองเสียง กรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท

หัวข้อนี้จะกล่าวถึงการทดลองหาอัตราการรู้จำทำนองเสียง กรณีที่แบ่งประเภทของทำนองเสียงออกเป็น 2 ประเภท คือ ทำนองเสียงตก และทำนองเสียงขึ้น ส่วนทำนองเสียงผสมจะรวมไว้เป็นประเภทเดียวกับทำนองเสียงขึ้น หัวข้อ 4.4.1 กล่าวถึงการรู้จำทำนองเสียงโดยใช้คอนทัวร์ LFC เพียงอย่างเดียว หัวข้อ 4.4.2 กล่าวถึงการรู้จำทำนองเสียงโดยใช้คอนทัวร์ LFC และ ΔLFC หัวข้อ 4.4.3 กล่าวถึงการรู้จำทำนองเสียงโดยใช้คอนทัวร์ LFC และคอนทัวร์ FVC หัวข้อ 4.4.4 กล่าวถึงการรู้จำทำนองเสียงโดยใช้คอนทัวร์ LFC คอนทัวร์ FVC และ ΔLFC และสุดท้ายหัวข้อ 4.4.5 เป็นการเปรียบเทียบผลการทดลองทั้งสี่แบบ

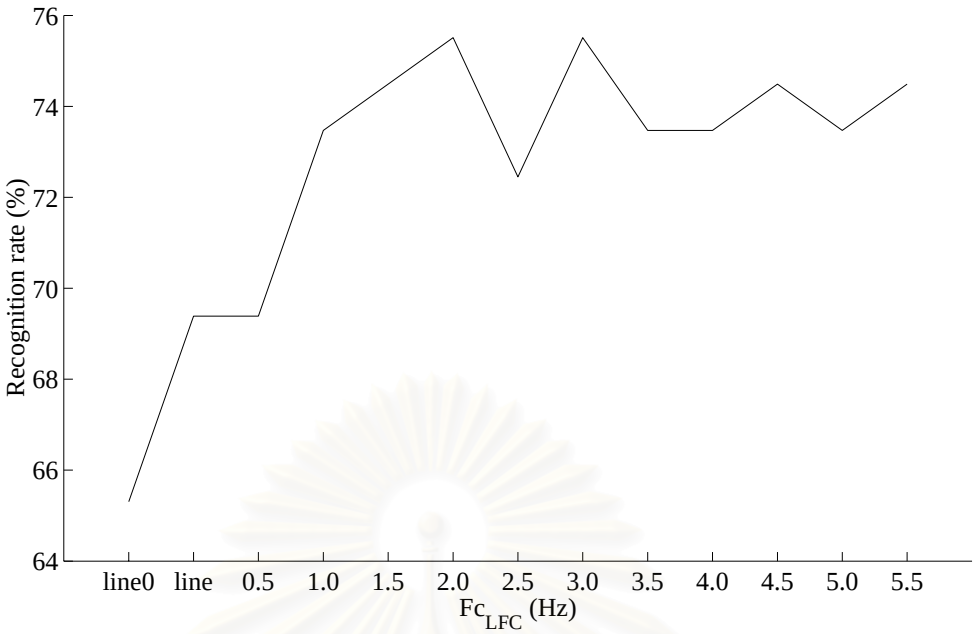
การรายงานผลการทดลองในแต่ละหัวข้อ จะเริ่มจากการรายงานอัตราการรู้จำทำนองเสียงประเภทต่าง ๆ และอัตราการรู้จำเฉลี่ย ที่ทุก ๆ ค่าความถี่ตัด ว่ามีค่าอยู่ในช่วงใด โดยเปรียบเทียบระหว่างผู้พูดหญิงกับผู้พูดชาย รวมทั้งวิเคราะห์ความสัมพันธ์ระหว่างอัตราการรู้จำเหล่านี้ กับค่าความถี่ตัดของตัวกรอง จากนั้นจึงพิจารณาเฉพาะกรณีของค่าความถี่ตัดที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุด โดยพิจารณาว่าตัวรู้จำ จำแนกทำนองเสียงหนึ่ง ๆ ผิดไปเป็นทำนองเสียงอื่นด้วยอัตราเท่าไร

4.4.1 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยค

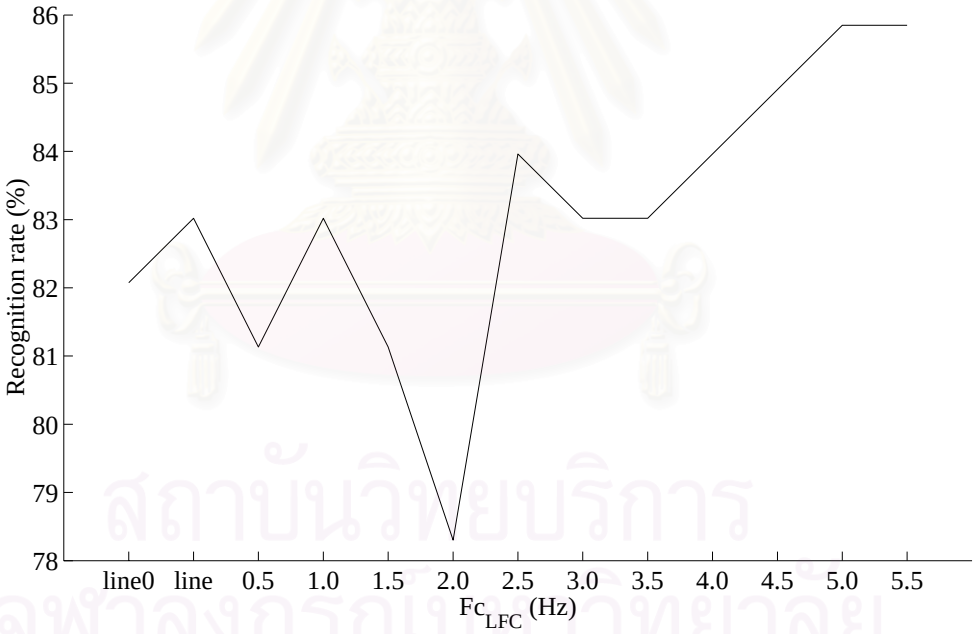
การทดลองนี้ใช้เวกเตอร์ลักษณะที่เหมือนกับการทดลองที่ 4.3.1 คือใช้เพียงค่าของ LFC ที่ได้จากการสุ่มตัวอย่าง และความยาวของประโยค ต่างกันตรงที่แบ่งทำนองเสียงเป็น 2 ประเภท นอกจากนี้ในขณะที่ทดลองพบว่า เสียงผู้หญิงที่ Fc_{LFC} = 4.5 Hz อัตราการรู้จำโดยเฉลี่ยยังคงมีแนวโน้มว่าจะเพิ่มขึ้นได้อีก การทดลองนี้จึงได้เพิ่มค่า Fc_{LFC} ไปอีกจนถึง 5.5 Hz

อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.37 – 4.40 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.41 และ 4.42 เมื่อพิจารณาจากรูปที่ 4.37 – 4.40 จะเห็นได้ว่า เมื่อรวมทำนองเสียงขึ้นและทำนองเสียงผสมเข้าเป็นกลุ่มเดียวกันแล้ว อัตราการรู้จำของทำนองเสียงต่าง ๆ จะเพิ่มสูงขึ้นมาก ถึงแม้จะใช้เพียงคอนทัวร์ LFC ก็สามารถทำให้อัตราการรู้จำในแต่ละทำนองเสียงมีค่าสูง

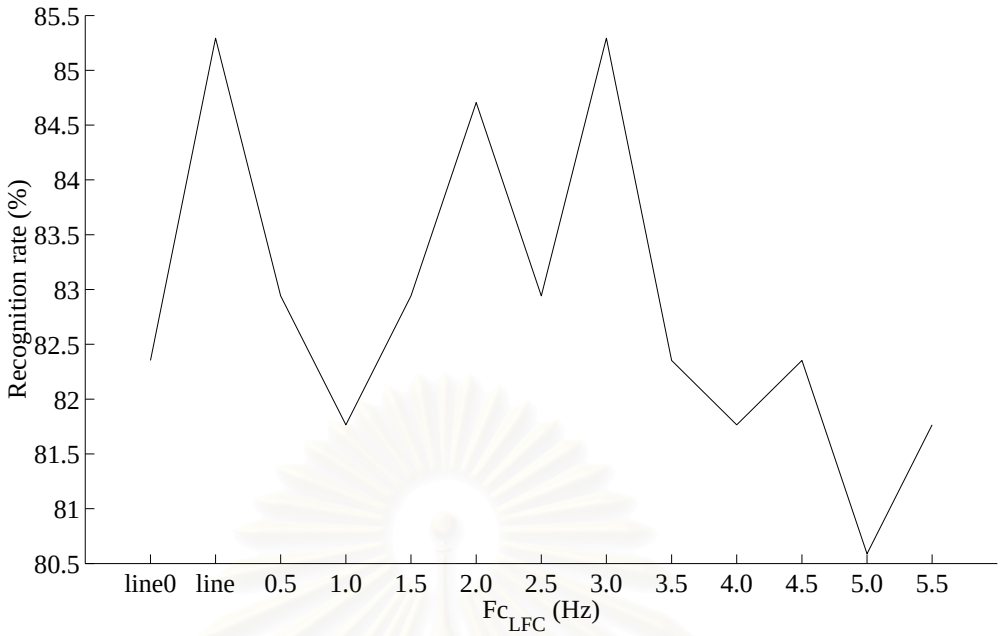
ในกรณีของทำนองเสียงตก อัตราการรู้จำของเสียงผู้ชายจะอยู่ในช่วงร้อยละ 65 ถึง 75 อัตราการรู้จำของเสียงผู้หญิงจะอยู่ในช่วงร้อยละ 78 ถึง 85 ในกรณีของทำนองเสียงขึ้น อัตราการรู้จำของเสียงผู้ชายจะอยู่ในช่วงร้อยละ 80 ถึง 85 อัตราการรู้จำของเสียงผู้หญิงจะอยู่ในช่วงร้อยละ 86 ถึง 87 เมื่อพิจารณารูปที่ 4.41 และ 4.42 จะเห็นได้ว่าอัตราการรู้จำโดยเฉลี่ยของเสียงผู้ชายมีค่าสูงสุดเป็นร้อยละ 81.7 ที่ Fc_{LFC} = 3 Hz อัตราการรู้จำโดยเฉลี่ยของเสียงผู้หญิงมีค่าสูงสุดเป็นร้อยละ 86.7 ที่ Fc_{LFC} = 5.5 Hz



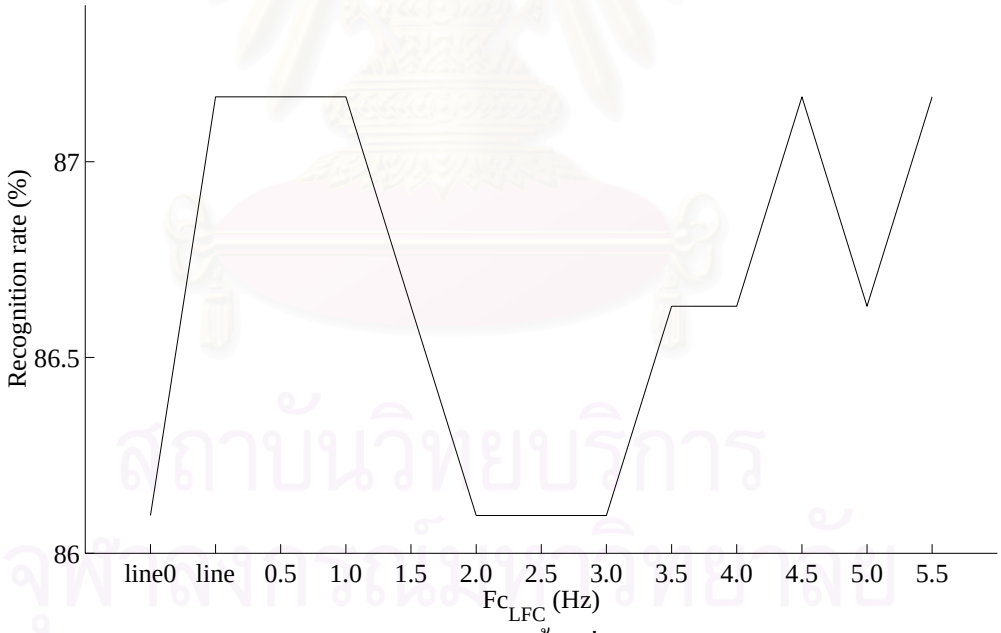
รูปที่ 4.37 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



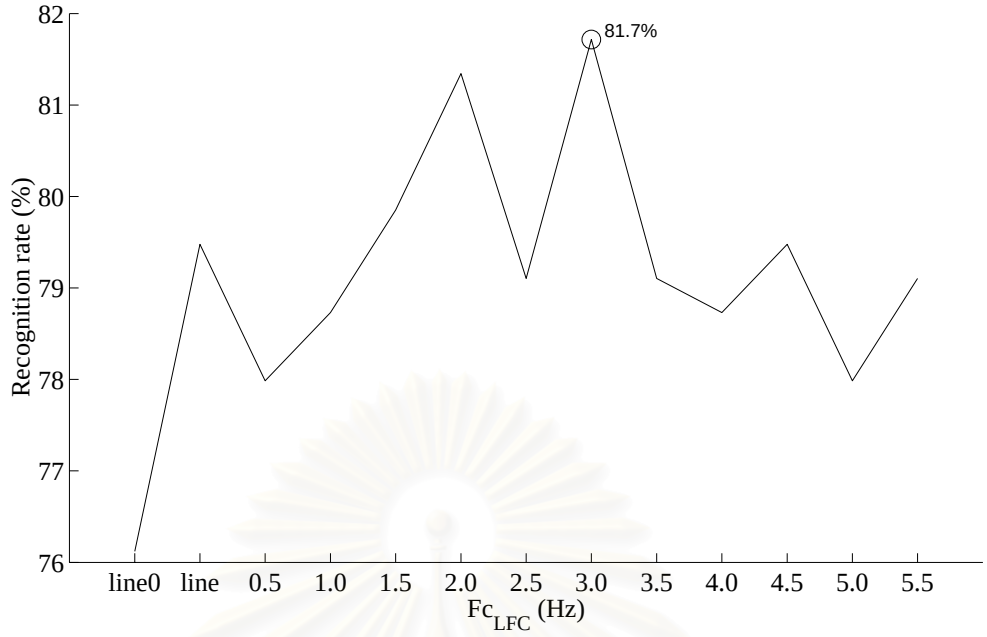
รูปที่ 4.38 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



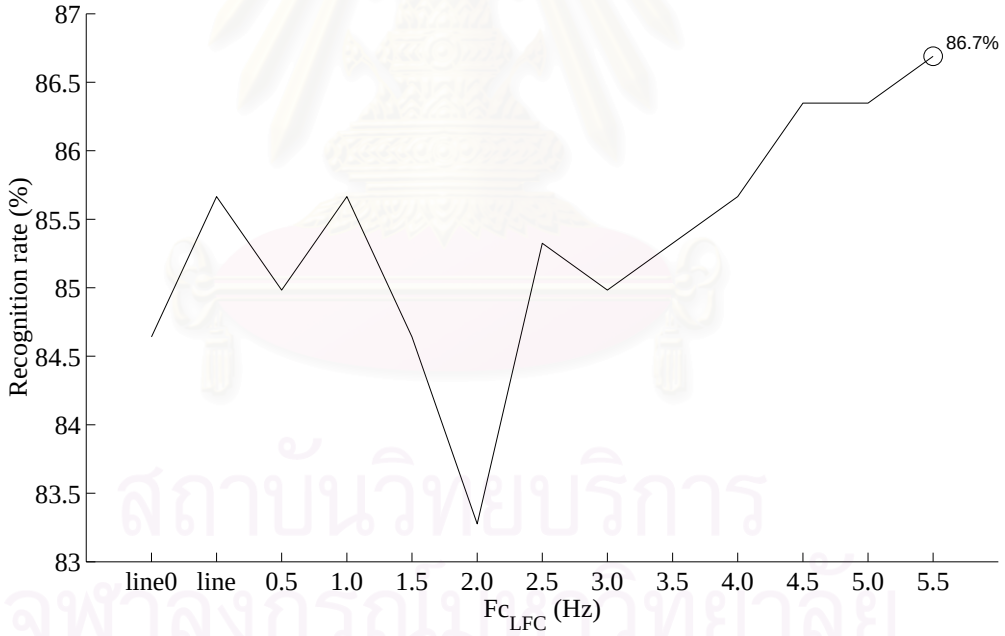
รูปที่ 4.39 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



รูปที่ 4.40 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



รูปที่ 4.41 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



รูปที่ 4.42 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท

ถึงแม้ว่าจากรูปที่ 4.42 อัตราการรู้จำโดยเฉลี่ยของเสียงผู้หญิงมีแนวโน้มว่าจะเพิ่มมากขึ้นได้อีกเมื่อเพิ่ม Fc_{LFC} แต่ในการทดลองนี้จะไม่เพิ่มค่า Fc_{LFC} ขึ้นไปอีก เนื่องจาก

- การเพิ่ม Fc_{LFC} จะยิ่งทำให้จำนวนมิติของเวกเตอร์ลักษณะเพิ่มมากขึ้น ทำให้โครงข่ายซับซ้อนขึ้น และใช้เวลามากขึ้นในการฝึกฝน และทดสอบโครงข่าย
- ถึงแม้จะเพิ่มค่า Fc_{LFC} ไปมากกว่านี้ ก็ไม่น่าจะทำให้ผลที่ได้ดีกว่าเดิมมากนัก เพราะการเพิ่มค่า Fc_{LFC} จะทำให้คอนทัวร์ LFC มีความใกล้เคียงกับคอนทัวร์ F_0 มากยิ่งขึ้น จากรูปที่ 3.29 จะเห็นได้ว่า Fc_{LFC} มีค่าเพียง 4.5 Hz ก็ทำให้คอนทัวร์ LFC เกือบจะมีรูปร่างเหมือนกับคอนทัวร์ F_0 แล้ว
- ในการทดลองในหัวข้อถัด ๆ ไป จะพบว่า การใช้คอนทัวร์ FVC หรือ ΔLFC จะทำให้อัตราการรู้จำดีขึ้นกว่านี้ โดยใช้เวกเตอร์ลักษณะที่มีจำนวนมิติน้อยลง

อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.41 และ 4.42 แสดงให้เห็นในตารางที่ 4.24 และ 4.25 ตามลำดับ

ตารางที่ 4.24 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย (จากรูปที่ 4.41)

ความถี่ตัด (Hz) หรือลักษณะของ LFC	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**	**5.5**
อัตราการรู้จำเฉลี่ย (%)	76.1	79.5	78.0	78.7	79.9	81.3	79.1	**81.7**	79.1	78.7	79.5	78.0	79.1

ตารางที่ 4.25 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้หญิง (จากรูปที่ 4.42)

ความถี่ตัด (Hz) หรือลักษณะของ LFC	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**	**5.5**
อัตราการรู้จำเฉลี่ย (%)	84.6	85.7	85.0	85.7	84.6	83.3	85.3	85.0	85.3	85.7	86.3	86.3	**86.7**

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.41 และ 4.42) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.26 – 4.27

ตารางที่ 4.26 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.41) ของเสียง ผู้ชายเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 3.0 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้		จำนวนประโยค
	ตก	ขึ้น	
ตก	**75.5**	24.5	98
ขึ้น	14.7	**85.3**	170
	จำนวนประโยคทั้งหมด		268

ตารางที่ 4.27 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.42) ของเสียง ผู้หญิงเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 5.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้		จำนวนประโยค
	ตก	ขึ้น	
ตก	**85.8**	14.2	106
ขึ้น	12.8	**87.2**	187
	จำนวนประโยคทั้งหมด		293

4.4.2 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยค

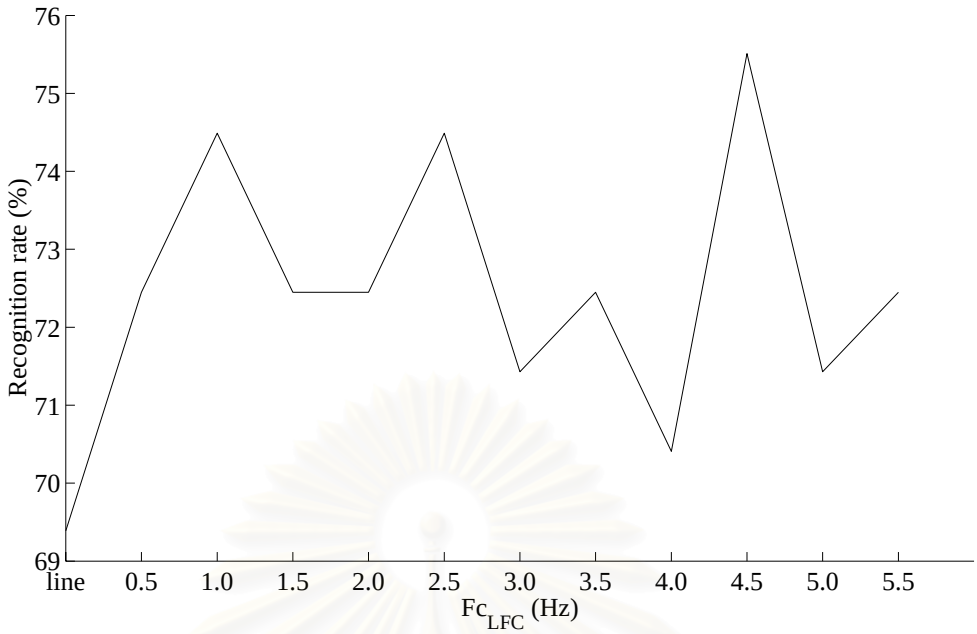
การทดลองนี้ใช้เวกเตอร์ลักษณะที่เหมือนกับการทดลองในข้อ 4.3.2 คือใช้ค่าของ LFC และ ΔLFC ที่ได้จากการสุ่มตัวอย่าง รวมทั้งความยาวของประโยค ต่างกันตรงที่แบ่งทำนองเสียงออกเป็น 2 ประเภท ในขณะที่ทดลองพบว่าเสียงผู้หญิงมีลักษณะเช่นเดียวกับการทดลองที่ 4.4.1 คือเสียงผู้หญิงที่ Fc_{LFC} = 4.5 Hz มีแนวโน้มว่าอัตราการรู้จำโดยเฉลี่ยยังสามารถเพิ่มสูงขึ้นได้อีก การทดลองนี้จึงได้เพิ่มค่า Fc_{LFC} ไปอีกจนถึง 5.5 Hz

อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.43 – 4.46 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.47 และ 4.48

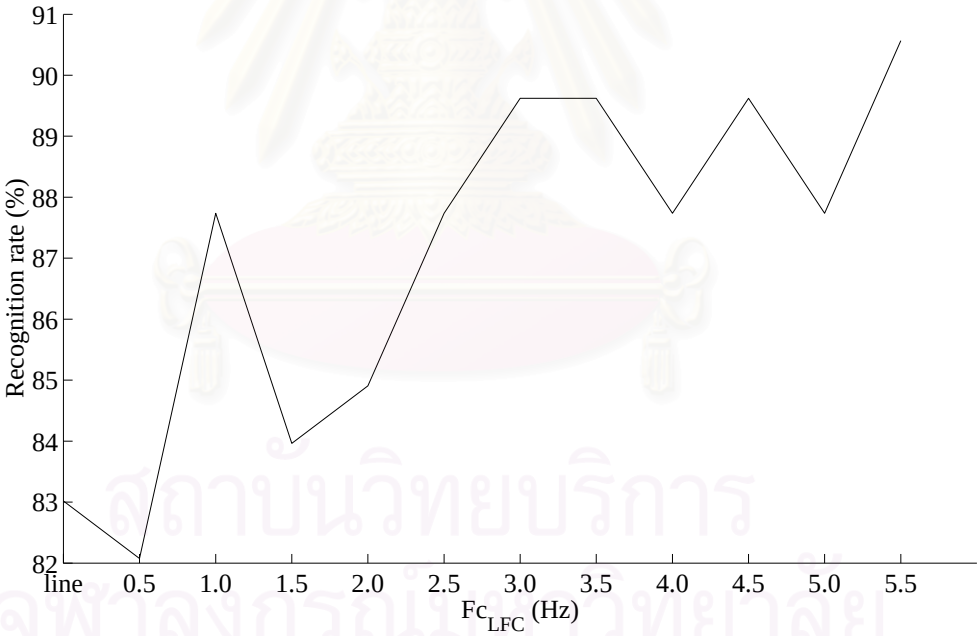
จากรูปที่ 4.43 – 4.46 จะเห็นได้ว่า ในกรณีของทำนองเสียงตก อัตราการรู้จำเสียงของผู้ชายอยู่ในช่วงร้อยละ 69 ถึง 75 ซึ่งใกล้เคียงกับการทดลองในข้อ 4.4.1 อัตราการรู้จำของเสียงผู้หญิงอยู่ในช่วงร้อยละ 82 ถึง 90 ซึ่งสูงกว่าการทดลองในข้อ 4.4.1 ส่วนในกรณีของทำนองเสียงขึ้น อัตราการรู้จำของเสียงผู้ชายอยู่ในช่วงร้อยละ 78 ถึง 84 ซึ่งต่ำกว่าอัตราการรู้จำในข้อ 4.4.1 เล็กน้อย ในขณะที่อัตราการรู้จำของเสียงผู้หญิงอยู่ในช่วงร้อยละ 86 ถึง 89 ซึ่งใกล้เคียงกับข้อ 4.4.1

เมื่อพิจารณารูปที่ 4.47 และ 4.48 จะพบว่า อัตราการรู้จำเฉลี่ยของเสียงผู้ชายมีค่าสูงสุดเป็นร้อยละ 81.0 ซึ่งต่ำกว่าข้อ 4.4.1 เล็กน้อย ส่วนในกรณีของเสียงผู้หญิงอัตราการรู้จำสูงสุดมีค่าเป็นร้อยละ 89.4 ซึ่งสูงกว่าข้อ 4.4.2 และถึงแม้ว่าจากรูปที่ 4.48 อัตราการรู้จำของเสียงผู้หญิงจะมีแนวโน้มเพิ่มมากขึ้นได้อีก แต่การทดลองนี้จะไม่เพิ่มค่า Fc_{LFC} ขึ้นไปอีก ด้วยเหตุผลเดียวกับข้อ 4.4.1

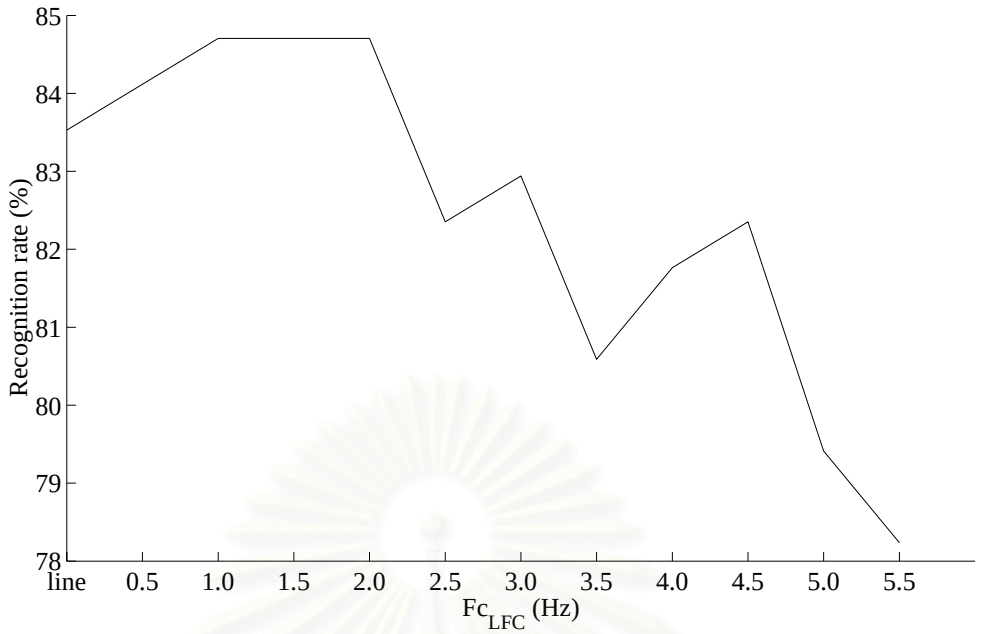
อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.47 และ 4.48 แสดงให้เห็นในตารางที่ 4.28 และ 4.29 ตามลำดับ



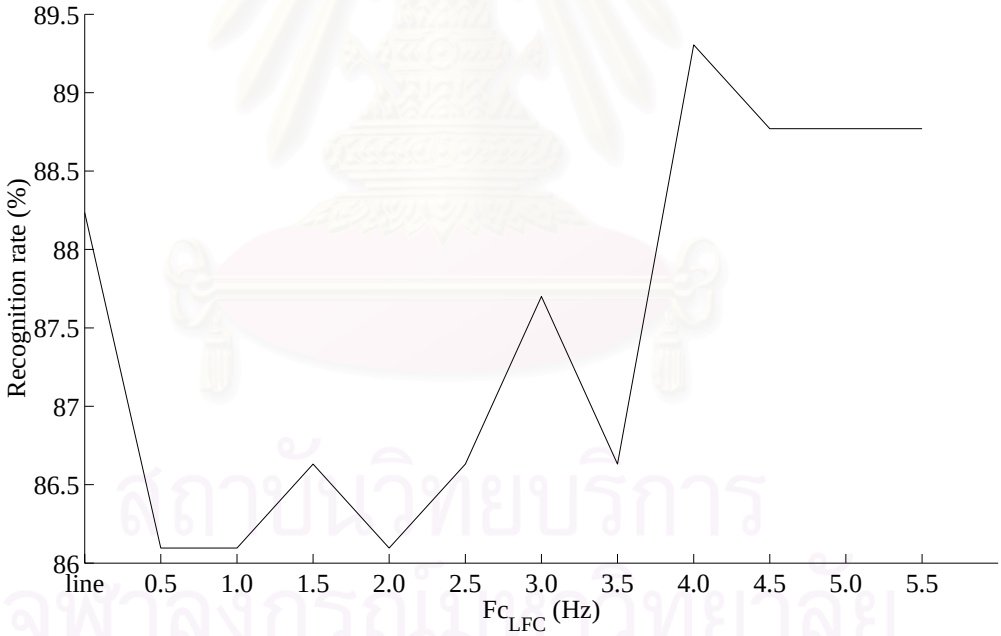
รูปที่ 4.43 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



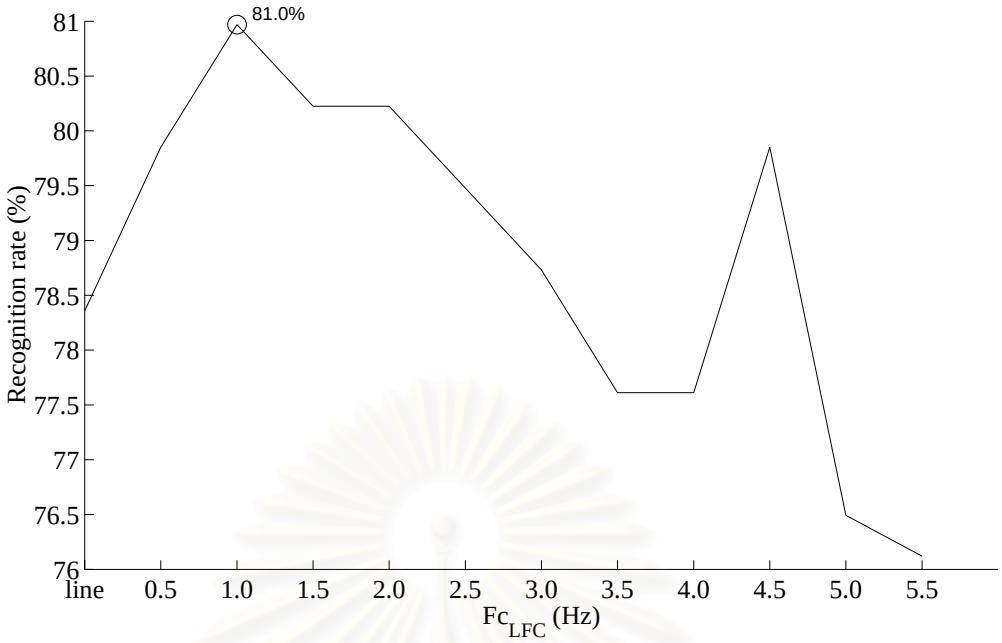
รูปที่ 4.44 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



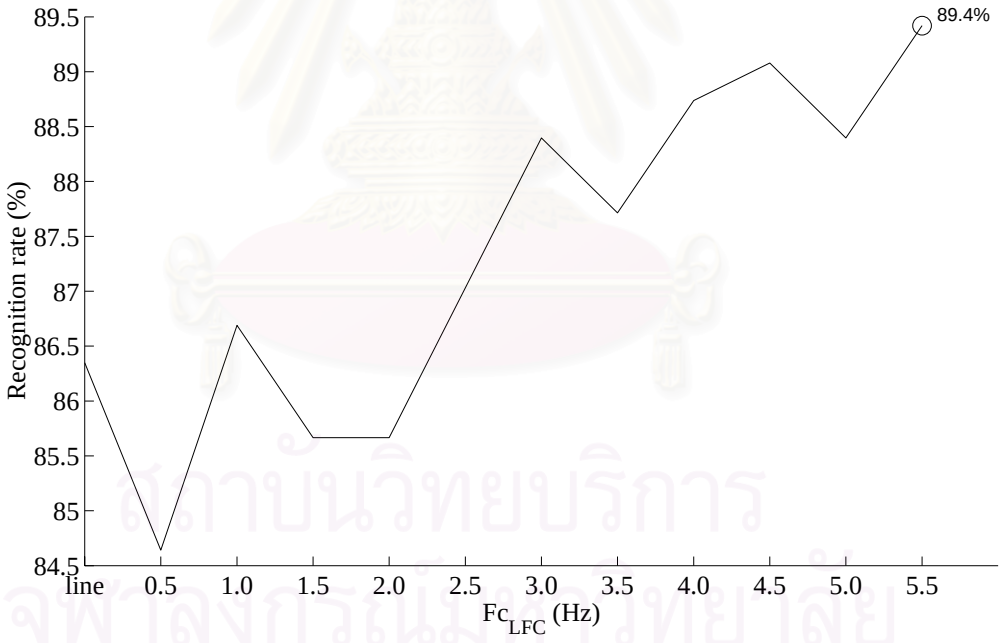
รูปที่ 4.45 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



รูปที่ 4.46 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



รูปที่ 4.47 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท



รูปที่ 4.48 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท

ตารางที่ 4.28 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย (จากรูปที่ 4.47)

ความถี่ตัด (Hz) หรือลักษณะของ LFC	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**	**5.5**
อัตราการรู้จำเฉลี่ย (%)	78.4	79.9	**81.0**	80.2	80.2	79.5	78.7	77.6	77.6	79.9	76.5	76.1

ตารางที่ 4.29 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้หญิง (จากรูปที่ 4.48)

ความถี่ตัด (Hz) หรือลักษณะของ LFC	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**	**5.5**
อัตราการรู้จำเฉลี่ย (%)	86.3	84.6	86.7	85.7	85.7	87.0	88.4	87.7	88.7	89.1	88.4	**89.4**

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.47 และ 4.48) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.30 – 4.31

ตารางที่ 4.30 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.47) ของเสียงผู้ชายเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 1.0 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	**ประเภทของทำนองเสียงที่รู้จำได้**		**จำนวนประโยค**
	ตก	**ขึ้น**	
ตก	**74.5**	25.5	98
ขึ้น	15.3	**84.7**	170
จำนวนประโยคทั้งหมด			268

ตารางที่ 4.31 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.48) ของเสียง ผู้หญิงเมื่อใช้ค่าความถี่ตัดของ LFC เป็น 5.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้		จำนวนประโยค
	ตก	ขึ้น	
ตก	**90.6**	9.4	106
ขึ้น	11.2	**88.8**	187
	จำนวนประโยคทั้งหมด		293

4.4.3 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC FVC และความยาวของประโยค

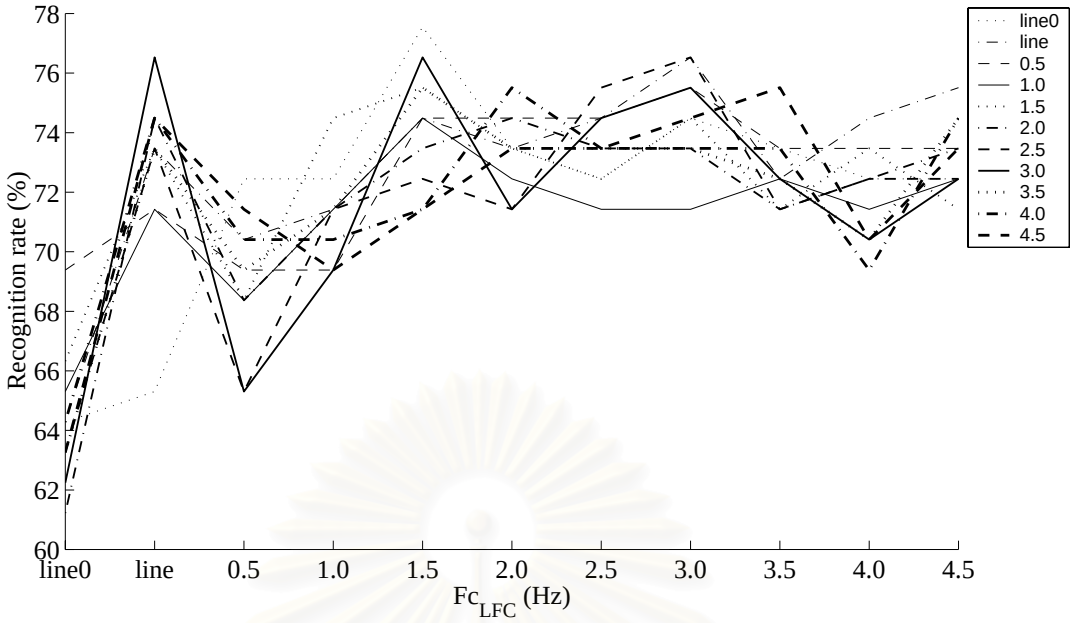
การทดลองนี้ใช้เวกเตอร์ลักษณะที่เหมือนกับการทดลองในข้อ 4.3.3 ทุกประการ ต่างกันตรงที่แบ่งทำนองเสียงออกเป็น 2 ประเภท

อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.49 – 4.52 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.53 และ 4.54

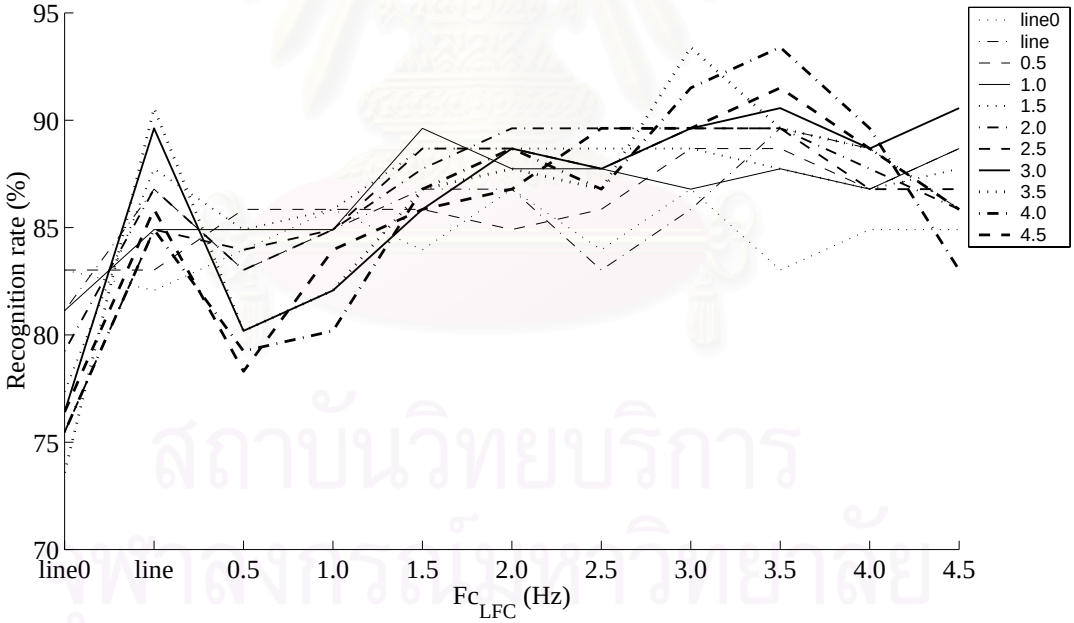
จากรูปที่ 4.49 – 4.52 จะเห็นได้ว่า ในกรณีของทำนองเสียงตก อัตราการรู้จำของเสียงผู้ชายอยู่ในช่วงร้อยละ 61 ถึง 77 อัตราการรู้จำของเสียงผู้หญิงอยู่ในช่วงร้อยละ 74 ถึง 93 ส่วนในกรณีของทำนองเสียงขึ้น อัตราการรู้จำของเสียงผู้ชายอยู่ในช่วงร้อยละ 79 ถึง 86 อัตราการรู้จำของเสียงผู้หญิงอยู่ในช่วงร้อยละ 83 ถึง 90 จะเห็นได้ว่าอัตราการรู้จำของแต่ละทำนองเสียงอยู่ในช่วงที่สูงกว่าและกว้างกว่าการทดลองในข้อ 4.4.1 และ 4.4.2

เมื่อพิจารณารูปที่ 4.53 และ 4.54 จะพบว่า อัตราการรู้จำเฉลี่ยของเสียงผู้ชายมีค่าสูงสุดเป็นร้อยละ 81.7 เสียงผู้หญิงมีค่าสูงสุดเป็นร้อยละ 89.8 ซึ่งสูงกว่าการทดลองในข้อ 4.4.1 และ 4.4.2

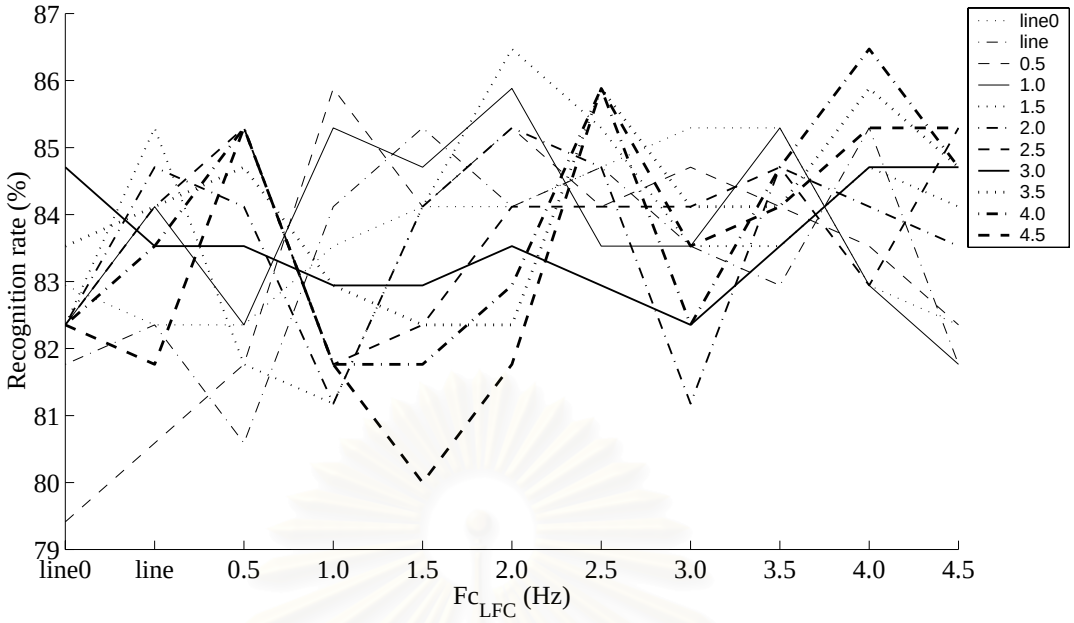
อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.53 และ 4.54 แสดงให้เห็นในตารางที่ 4.32 และ 4.33 ตามลำดับ



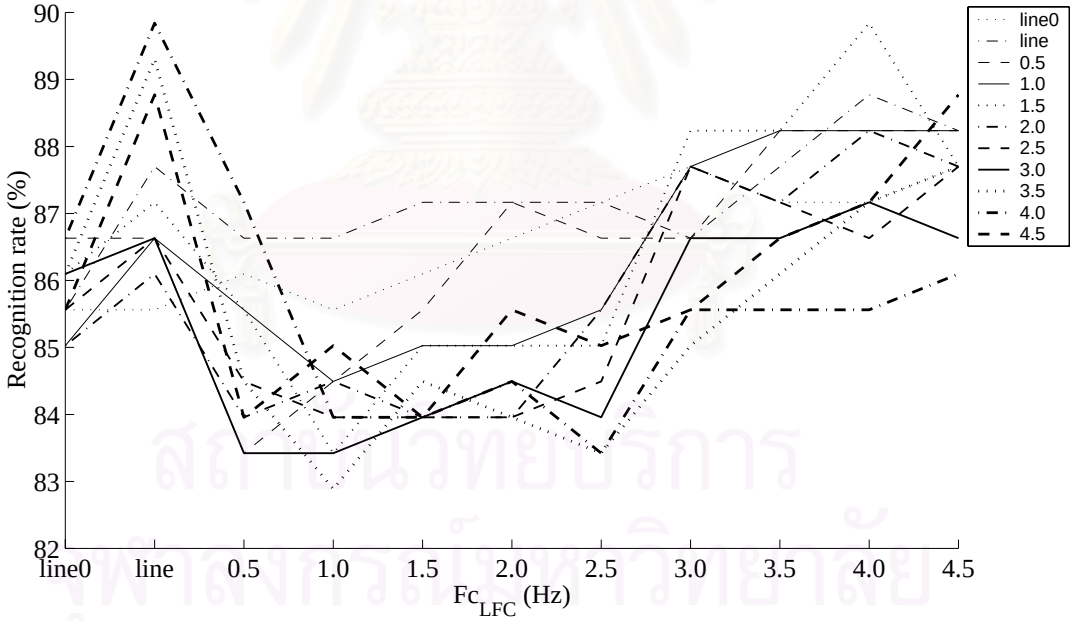
รูปที่ 4.49 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



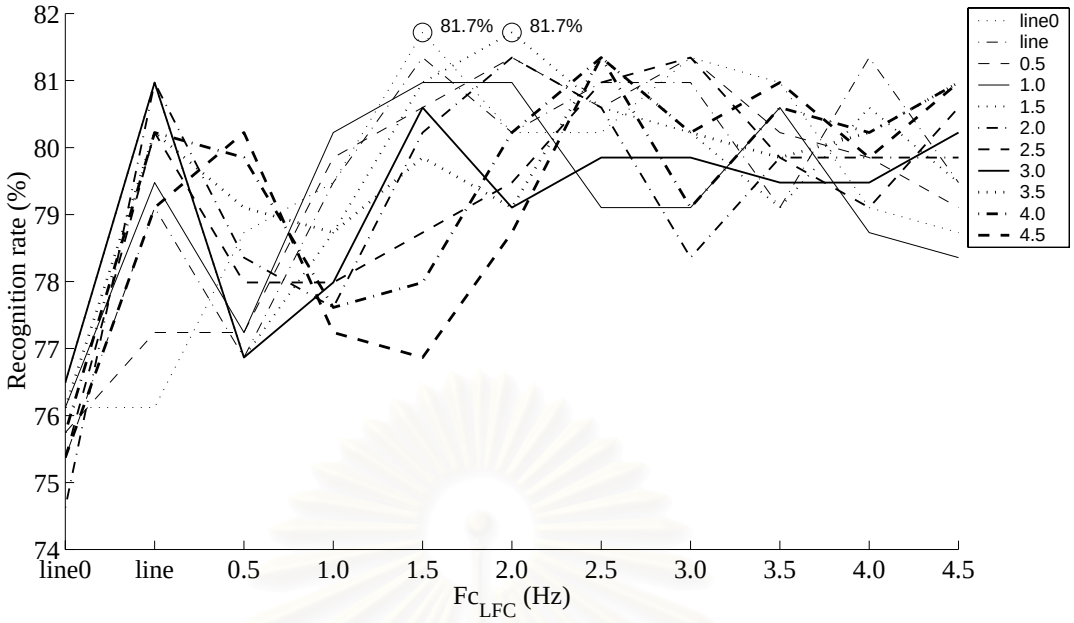
รูปที่ 4.50 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



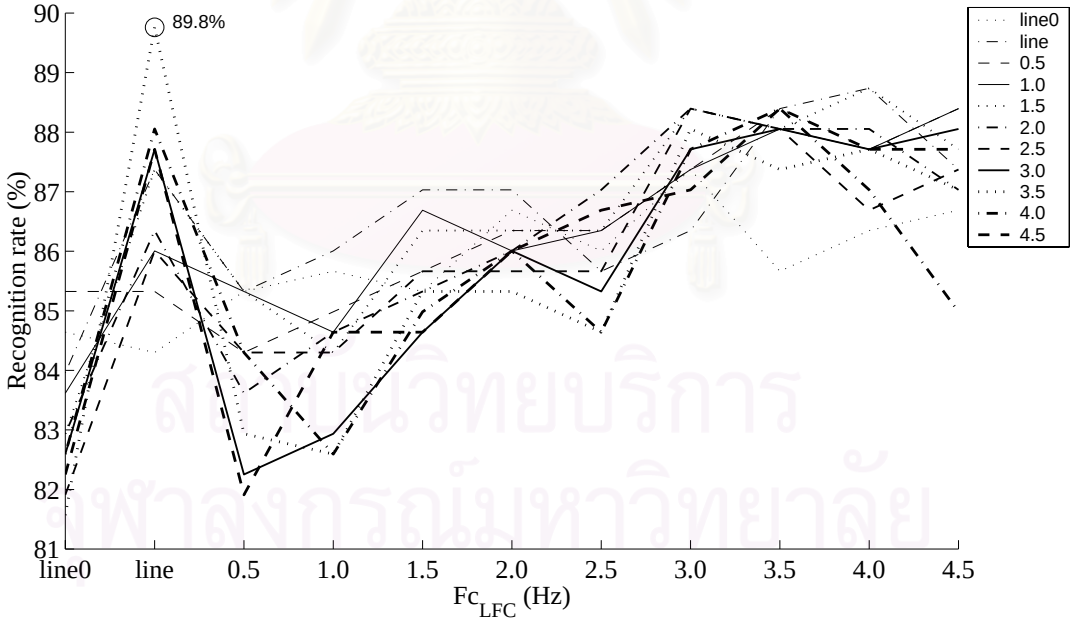
รูปที่ 4.51 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.52 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.53 อัตราการรู้จำทำนองเสียงของทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.54 อัตราการรู้จำทำนองเสียงของทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)

ตารางที่ 4.32 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้ชาย (จากรูปที่ 4.53)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line0	76.1	75.4	75.7	76.1	76.5	74.6	75.4	76.5	76.1	75.7	75.4
line	76.1	79.1	77.2	79.5	81.0	81.0	80.2	81.0	80.2	80.2	79.1
0.5	78.7	76.9	77.2	77.2	76.9	78.4	78.0	76.9	79.1	79.9	80.2
1.0	79.5	79.5	79.9	80.2	78.7	77.6	78.0	78.0	78.7	77.6	77.2
1.5	**81.7**	81.3	80.6	81.0	81.0	80.2	78.7	80.6	79.9	78.0	76.9
2.0	80.2	80.2	81.3	81.0	**81.7**	81.3	79.5	79.1	79.1	80.2	78.7
2.5	80.2	81.0	80.6	79.1	80.6	80.6	81.0	79.9	81.3	81.3	81.3
3.0	81.3	81.0	81.3	79.1	80.2	78.4	81.3	79.9	80.2	79.1	80.2
3.5	81.0	79.1	80.2	80.6	79.1	79.9	79.9	79.5	79.9	80.6	81.0
4.0	79.1	81.3	79.9	78.7	80.6	79.9	79.1	79.5	80.2	80.2	79.9
4.5	78.7	79.5	79.1	78.4	79.5	79.9	80.6	80.2	81.0	81.0	81.0

ตารางที่ 4.33 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้หญิง (จากรูปที่ 4.54)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line0	84.6	84.0	85.3	83.6	82.9	82.9	81.9	82.6	81.6	82.6	82.3
line	84.3	87.4	85.3	86.0	87.4	86.3	86.0	87.7	**89.8**	88.1	87.7
0.5	85.3	85.3	84.3	85.3	85.3	83.6	84.3	82.3	82.9	84.3	81.9
1.0	85.7	86.0	85.0	84.6	84.3	84.6	84.3	82.9	82.6	82.6	84.6
1.5	85.3	87.0	85.7	86.7	86.3	85.3	85.7	84.6	85.3	85.0	84.6
2.0	86.7	87.0	86.3	86.0	86.3	86.0	85.7	86.0	85.3	86.0	86.0
2.5	86.0	85.7	86.3	86.3	86.3	87.0	85.7	85.3	84.6	84.6	86.7
3.0	87.4	86.3	87.4	87.4	88.4	88.4	88.4	87.7	88.1	87.7	87.0
3.5	85.7	88.4	88.4	88.1	88.1	88.1	88.1	88.1	87.4	88.4	88.4
4.0	86.3	88.7	87.7	87.7	88.7	88.1	86.7	87.7	87.7	87.0	87.7
4.5	86.7	87.4	88.4	88.4	87.7	87.0	87.4	88.1	87.0	85.0	87.7

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.53 และ 4.54) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.34 – 4.36

ตารางที่ 4.34 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.53) ของเสียงผู้ชายเมื่อ Fc_{LFC} =1.5 Hz และใช้ FVC เป็นเส้นตรงที่มีความชันเป็น 0

ประเภทของทำนองเสียงที่นำมารู้จำ	**ประเภทของทำนองเสียงที่รู้จำได้**		**จำนวนประโยค**
	ตก	**ขึ้น**	
ตก	**77.6**	22.4	98
ขึ้น	15.9	**84.1**	170
จำนวนประโยคทั้งหมด			268

ตารางที่ 4.35 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.53) ของเสียงผู้ชายเมื่อ Fc_{LFC} =2.0 Hz และใช้ Fc_{FVC} = 1.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้		จำนวนประโยค
	ตก	ขึ้น	
ตก	**73.5**	26.5	98
ขึ้น	13.5	**86.5**	170
	จำนวนประโยคทั้งหมด		268

ตารางที่ 4.36 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.54) ของเสียงผู้หญิงเมื่อใช้ LFC เป็นเส้นตรง และใช้ Fc_{FVC} = 3.5 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	ประเภทของทำนองเสียงที่รู้จำได้		จำนวนประโยค
	ตก	ขึ้น	
ตก	**90.6**	9.4	106
ขึ้น	10.7	**89.3**	187
	จำนวนประโยคทั้งหมด		293

4.4.4 การรู้จำทำนองเสียง เมื่อใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยค

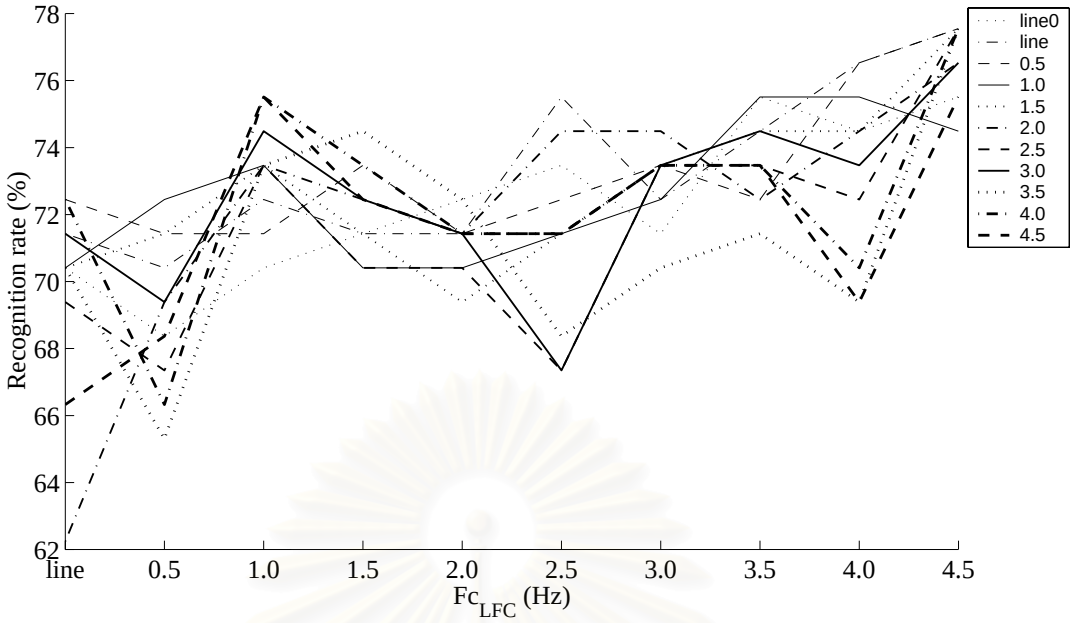
การทดลองนี้ใช้เวกเตอร์ลักษณะที่เหมือนกับการทดลองในข้อ 4.3.4 ทุกประการ ต่างกันตรงที่แบ่งทำนองเสียงออกเป็น 2 ประเภท

อัตราการรู้จำทำนองเสียง ของทำนองเสียงประเภทต่าง ๆ แยกตามเพศของผู้พูด แสดงดังรูปที่ 4.55 – 4.58 อัตราการรู้จำทำนองเสียงโดยเฉลี่ยแยกตามเพศของผู้พูด แสดงดังรูปที่ 4.59 และ 4.60

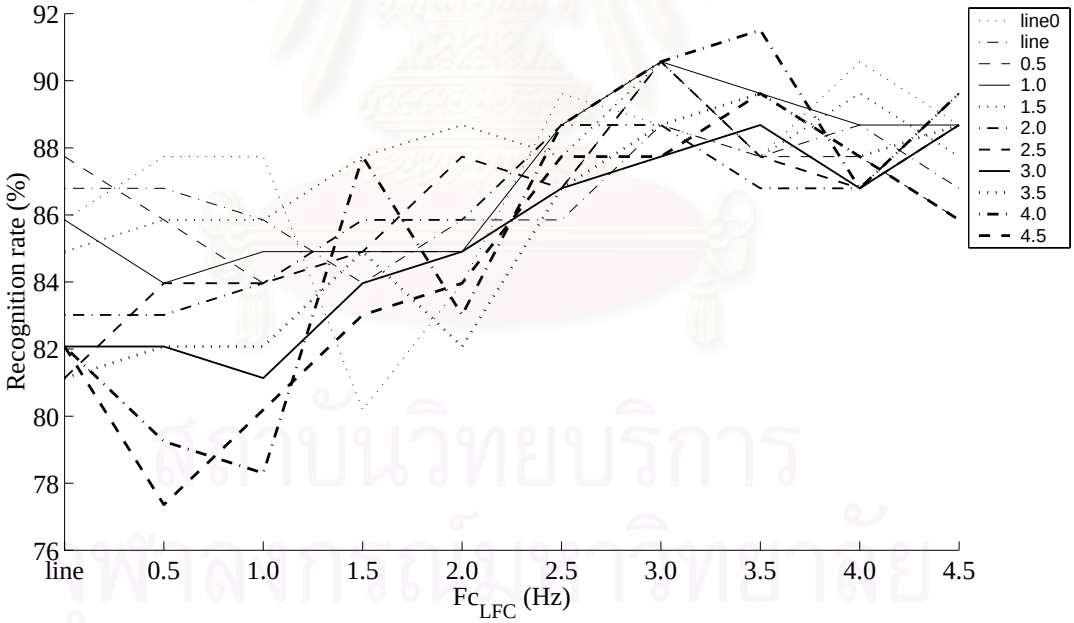
จากรูปที่ 4.55 – 4.58 จะเห็นได้ว่า ในกรณีของทำนองเสียงตก อัตราการรู้จำของเสียงผู้ชายอยู่ในช่วงร้อยละ 62 ถึง 77 อัตราการรู้จำของเสียงผู้หญิงอยู่ในช่วงร้อยละ 77 ถึง 91 ส่วนในกรณีของทำนองเสียงขึ้น อัตราการรู้จำของเสียงผู้ชายอยู่ในช่วงร้อยละ 78 ถึง 87 อัตราการรู้จำของเสียงผู้หญิงอยู่ในช่วงร้อยละ 83 ถึง 91 จะเห็นได้ว่าอัตราการรู้จำของแต่ละทำนองเสียงอยู่ในช่วงที่ใกล้เคียงกับการทดลองในข้อ 4.4.3

เมื่อพิจารณารูปที่ 4.59 และ 4.60 จะพบว่า อัตราการรู้จำเฉลี่ยของเสียงผู้ชายมีค่าสูงสุดเป็นร้อยละ 81.7 เสียงผู้หญิงมีค่าสูงสุดเป็นร้อยละ 90.8 ซึ่งใกล้เคียงกับ 4.4.3

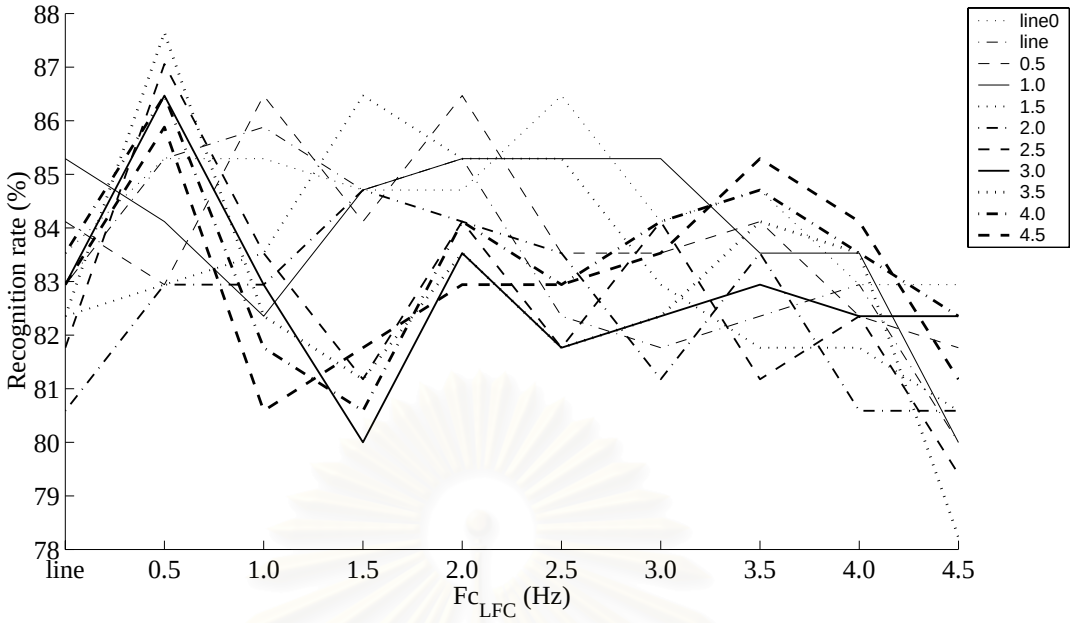
อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของเสียงผู้ชาย และเสียงผู้หญิงแสดงจากรูปที่ 4.59 และ 4.60 แสดงให้เห็นในตารางที่ 4.37 และ 4.38 ตามลำดับ



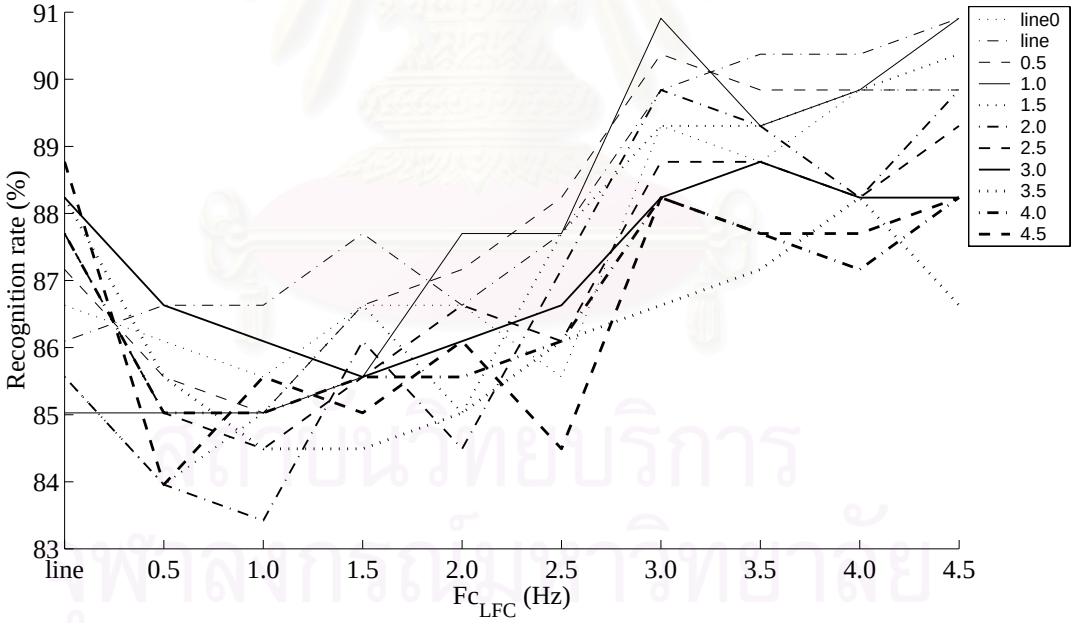
รูปที่ 4.55 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



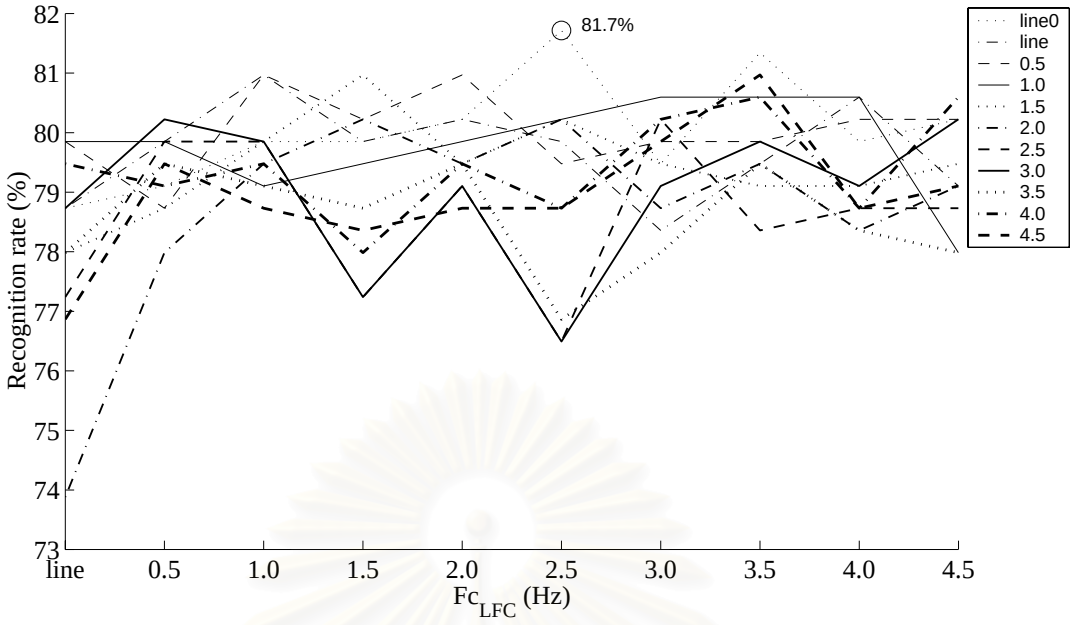
รูปที่ 4.56 อัตราการรู้จำทำนองเสียงของทำนองเสียงตก เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



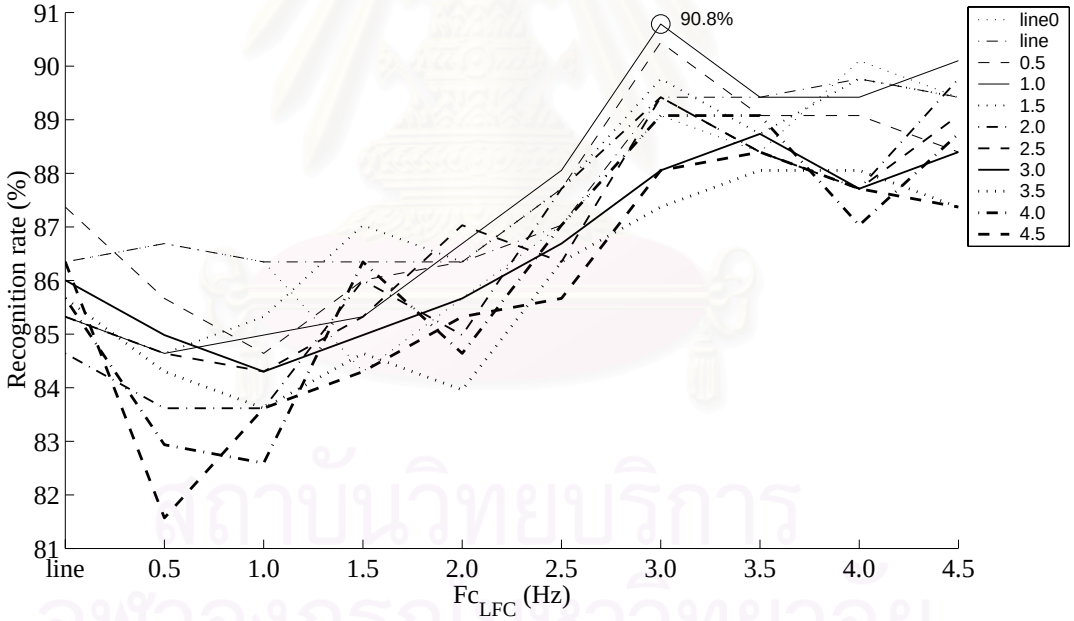
รูปที่ 4.57 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.58 อัตราการรู้จำทำนองเสียงของทำนองเสียงขึ้น เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.59 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้ชาย ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)



รูปที่ 4.60 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย เมื่อผู้พูดเป็นผู้หญิง ใช้เวกเตอร์ลักษณะจากคอนทัวร์ LFC ΔLFC คอนทัวร์ FVC และความยาวของประโยคกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท (กราฟแต่ละเส้นแสดงลักษณะที่ใช้ค่า Fc_{FVC} ต่าง ๆ กัน ดังแสดงในสัญลักษณ์ด้านขวา)

ตารางที่ 4.37 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้ชาย (จากรูปที่ 4.59)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line	78.7	78.7	79.9	79.9	78.0	73.9	77.2	78.7	78.0	79.5	76.9
0.5	79.1	79.9	78.7	79.9	78.7	78.0	79.9	80.2	79.5	79.1	79.5
1.0	79.9	81.0	81.0	79.1	79.9	79.5	79.9	79.9	79.1	79.5	78.7
1.5	79.9	79.9	80.2	79.5	81.0	80.2	77.2	77.2	78.7	78.0	78.4
2.0	80.2	80.2	81.0	79.9	79.5	79.5	79.1	79.1	79.5	79.5	78.7
2.5	**81.7**	79.9	79.5	80.2	80.2	80.2	76.5	76.5	76.9	78.7	78.7
3.0	79.5	78.4	79.9	80.6	79.5	78.7	80.2	79.1	78.0	80.2	79.9
3.5	81.3	79.5	79.9	80.6	79.1	79.5	78.4	79.9	79.5	80.6	81.0
4.0	79.9	80.6	80.2	80.6	79.1	78.4	78.7	79.1	78.4	78.7	78.7
4.5	80.2	79.1	80.2	78.0	79.5	79.1	78.7	80.2	78.0	80.6	79.1

ตารางที่ 4.38 อัตราการรู้จำทำนองเสียงโดยเฉลี่ย (%) ของเสียงผู้หญิง (จากรูปที่ 4.60)

	Fc_{FVC}(Hz)										
Fc_{LFC}(Hz)	**line0**	**line**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**
line	86.3	86.3	87.4	85.3	85.3	84.6	85.3	86.0	85.7	85.7	86.3
0.5	86.7	86.7	85.7	84.6	84.6	83.6	84.6	85.0	84.3	82.9	81.6
1.0	86.3	86.3	84.6	85.0	85.3	83.6	84.3	84.3	83.6	82.6	83.6
1.5	84.3	86.3	86.0	85.3	87.0	86.0	85.3	85.0	84.6	86.3	84.3
2.0	85.7	86.3	86.3	86.7	86.3	85.0	87.0	85.7	84.0	84.6	85.3
2.5	87.0	87.0	87.7	88.1	87.7	87.7	86.3	86.7	86.3	87.0	85.7
3.0	89.1	89.4	90.4	**90.8**	89.8	89.4	89.4	88.1	87.4	89.1	88.1
3.5	88.4	89.4	89.1	89.4	88.7	88.4	88.4	88.7	88.1	89.1	88.4
4.0	90.1	89.8	89.1	89.4	89.8	87.7	87.7	87.7	88.1	87.0	87.7
4.5	89.4	89.4	88.4	90.1	89.4	89.8	89.1	88.4	87.4	88.7	87.4

ผลการรู้จำทำนองเสียง ในกรณีที่ให้อัตราการรู้จำสูงสุด (จุดที่วงกลมในรูปที่ 4.59 และ 4.60) ของเสียงผู้ชายและเสียงผู้หญิง แสดงในตารางที่ 4.39 และ 4.40

ตารางที่ 4.39 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.59) ของเสียง ผู้ชายเมื่อ Fc_{LFC} = 2.5 Hz และใช้ FVC เป็นเส้นตรงที่มีความชันเป็น 0

ประเภทของทำนองเสียงที่นำมารู้จำ	**ประเภทของทำนองเสียงที่รู้จำได้**		**จำนวนประโยค**
	ตก	**ขึ้น**	
ตก	**73.5**	26.5	98
ขึ้น	13.5	**86.5**	170
	จำนวนประโยคทั้งหมด		268

ตารางที่ 4.40 ผลการรู้จำทำนองเสียง (%) ในกรณีที่ให้อัตราการรู้จำสูงที่สุด (จากรูปที่ 4.60) ของเสียง ผู้หญิงเมื่อ Fc_{LFC} = 3.0 Hz และใช้ Fc_{FVC} = 1.0 Hz

ประเภทของทำนองเสียงที่นำมารู้จำ	**ประเภทของทำนองเสียงที่รู้จำได้**		**จำนวนประโยค**
	ตก	**ขึ้น**	
ตก	**90.6**	9.4	106
ขึ้น	9.1	**90.9**	187
	จำนวนประโยคทั้งหมด		293

4.4.5 การเปรียบเทียบอัตราการรู้จำทำนองเสียง โดยใช้ลักษณะแต่ละแบบ กรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท

หัวข้อนี้กล่าวถึงการเปรียบเทียบอัตราการรู้จำทำนองเสียง โดยใช้ลักษณะต่าง ๆ กัน ตั้งแต่หัวข้อ 4.4.1 – 4.4.4 ซึ่งเป็นการรู้จำทำนองเสียงพูด โดยแบ่งทำนองเสียงออกเป็น 2 ประเภท คือ ทำนองเสียงตก และทำนองเสียงขึ้น

รูปที่ 4.61 และ 4.63 แสดงอัตราการรู้จำของผู้พูดที่เป็นผู้ชาย และผู้พูดที่เป็นผู้หญิง เมื่อใช้ลักษณะแบบต่าง ๆ จากการทดลองในหัวข้อ 4.4.1 – 4.4.4 โดยเลือกจากการทดลองย่อยของแต่ละหัวข้อที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุดในแต่ละเพศของผู้พูด ในกรณีที่มีการทดลองย่อยสองการทดลองที่ให้อัตราการรู้จำเฉลี่ยเท่ากัน ก็เลือกการทดลองย่อยที่ให้อัตราการรู้จำของทำนองเสียงที่มีค่าต่ำที่สุดมีค่ามากกว่า ซึ่งเป็นหลักการเดียวกันกับที่กล่าวไว้แล้วในข้อ 4.3.5

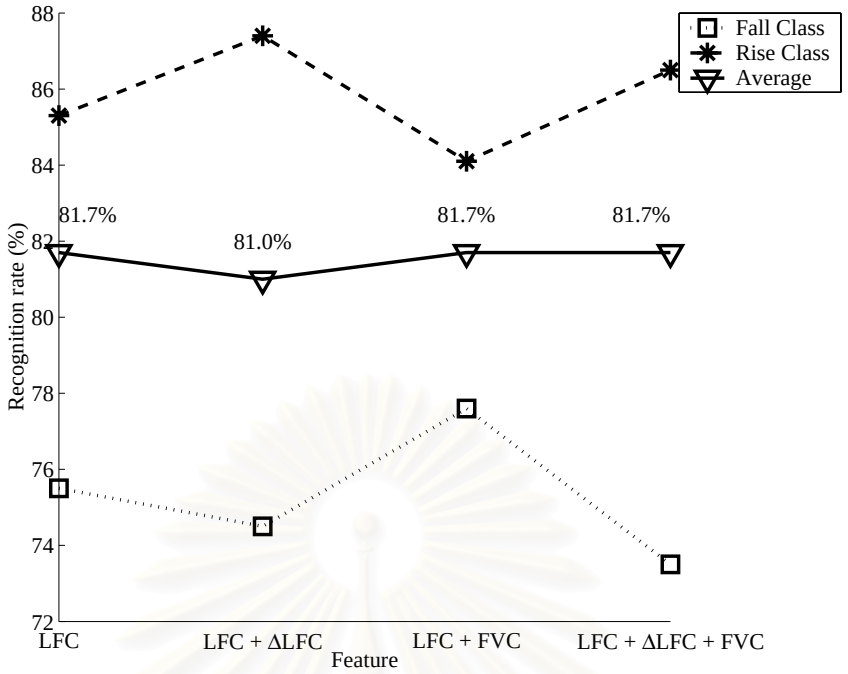
ส่วนรูปที่ 4.62 และ 4.64 แสดงจำนวนมิติของเวกเตอร์ลักษณะขาเข้า ซึ่งเป็นตัวบ่งบอกความซับซ้อนของโครงข่ายประสาทเทียม และพื้นที่ที่ใช้ในการเก็บข้อมูลในคอมพิวเตอร์ได้

จากรูปที่ 4.61 และ 4.62 ซึ่งเป็นกรณีที่ผู้พูดเป็นผู้ชาย จะเห็นได้ว่าการใช้เพียงคอนทัวร์ LFC และความยาวของประโยค ก็สามารถทำให้อัตราการรู้จำทำนองเสียงอยู่ในระดับที่สูงได้ การเพิ่ม ΔLFC เข้าไป แทบจะไม่มีผลต่ออัตราการรู้จำเลย นอกจากจะทำให้อัตราการรู้จำตกลงเล็กน้อย

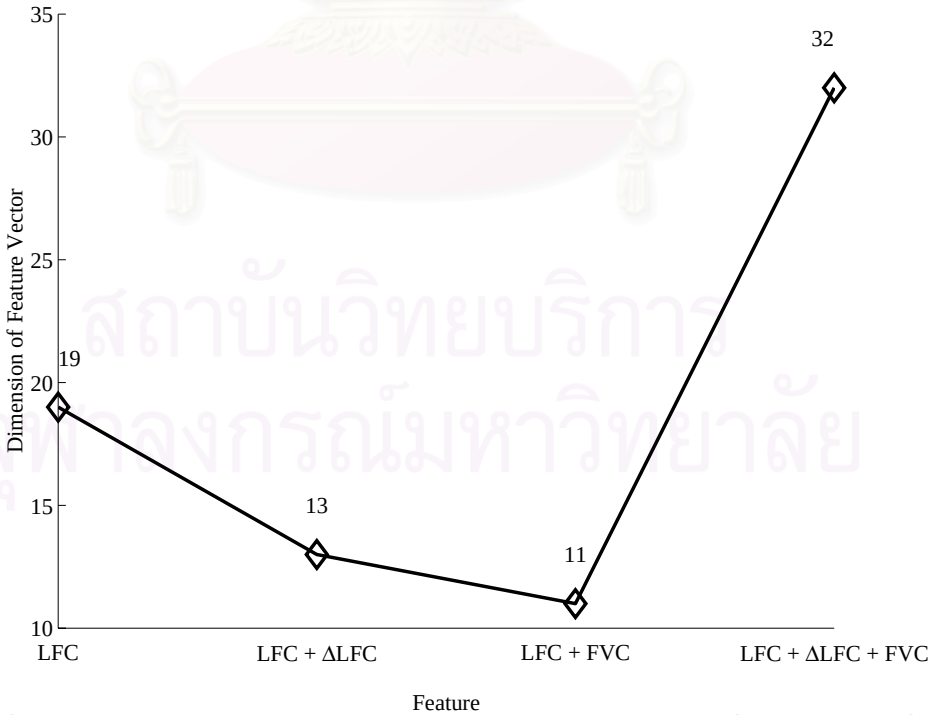
ส่วนการเพิ่ม FVC เข้าไปนั้น ก็ไม่ได้ช่วยให้อัตราการรู้จำเฉลี่ยเปลี่ยนแปลงเช่นกัน แต่เมื่อสังเกตอัตราการรู้จำของทำนองเสียงตก และทำนองเสียงขึ้นจะเห็นได้ว่าอัตราการรู้จำของทำนองเสียงทั้งสองมีค่าเข้าใกล้กันมากขึ้น แสดงให้เห็นว่า FVC สามารถช่วยลดความเอนเอียง (bias) ของตัวรู้จำได้ แต่ถ้าเพิ่มทั้ง ΔLFC และ FVC จะเห็นได้ว่าอัตราการรู้จำเฉลี่ยมีค่าเท่าเดิม แต่อัตราการรู้จำของทำนองเสียงขึ้น และทำนองเสียงตกมีค่าห่างกันมากขึ้น

และเมื่อพิจารณารูปที่ 4.62 จะเห็นได้ว่า FVC นอกจากจะช่วยลดความเอนเอียงของตัวรู้จำแล้ว ยังสามารถช่วยลดจำนวนมิติของเวกเตอร์ลักษณะขาเข้าได้อีกด้วย

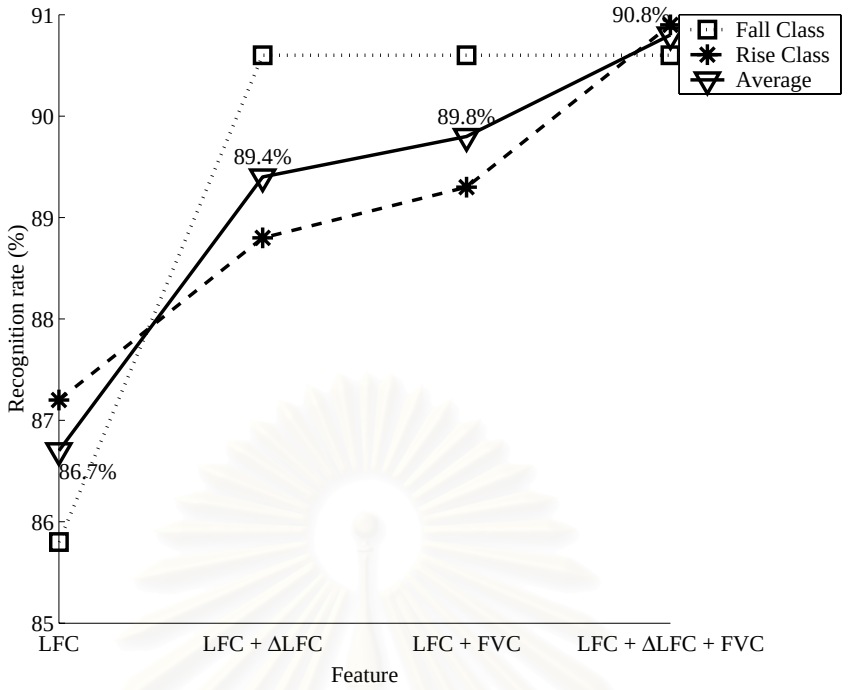
ในกรณีที่ผู้พูดเป็นผู้หญิง จะเห็นได้ว่าการใช้เพียง LFC กับความยาวของประโยคก็สามารถทำให้ได้ผลการทดลองที่ค่อนข้างสูง เช่นเดียวกบเสียงผู้ชาย แต่จะเห็นได้ว่าในกรณีของเสียงผู้หญิงนั้น ทั้ง ΔLFC และ FVC ต่างก็มีส่วนช่วยในการเพิ่มอัตราการรู้จำทำนองเสียงได้ทั้งคู่ โดยจะเห็นได้ว่าการเพิ่ม FVC เข้าไปจะให้ผลดีกว่าการเพิ่ม ΔLFC เข้าไป แต่อย่างไรก็ตามเมื่อใช้ทั้ง ΔLFC และ FVC จะเห็นได้ว่า สามารถทำให้อัตราการรู้จำทำนองเสียงมีค่าสูงสุด



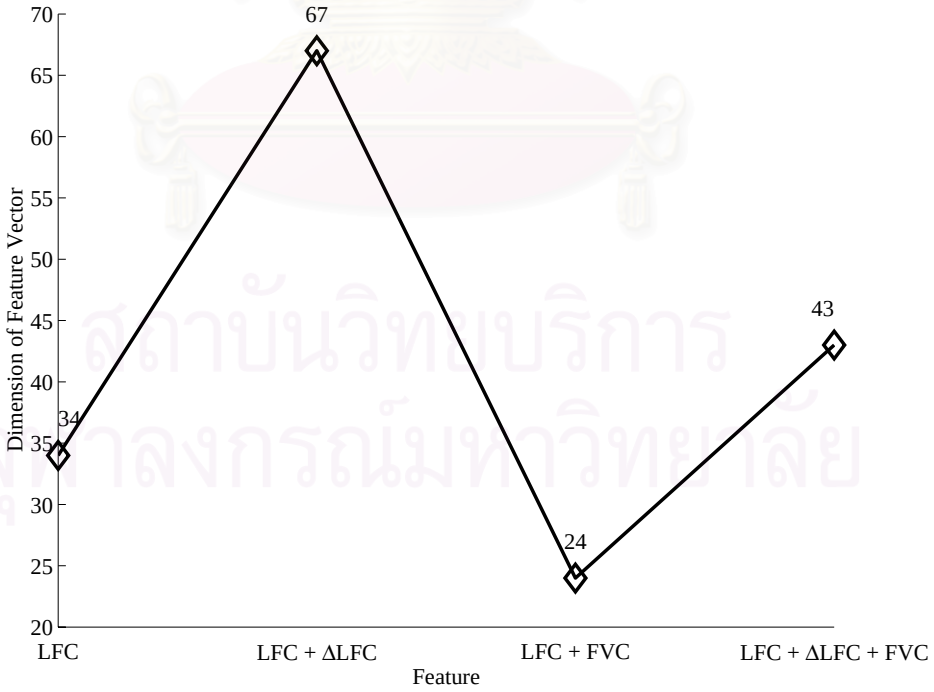
รูปที่ 4.61 อัตราการรู้จำเฉลี่ย และอัตราการรู้จำของแต่ละทำนองเสียง เลือกจากการทดลองย่อยที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุดเมื่อใช้เวกเตอร์ลักษณะแบบต่าง ๆ เมื่อผู้พูดเป็นผู้ชาย กรณีที่แบ่งทำนองเสียงเป็น 2 ประเภท (ลักษณะทุกแบบที่แสดงในกราฟ ได้รวมเอาความยาวของประโยค เข้าไปไว้ในเวกเตอร์ลักษณะด้วย)



รูปที่ 4.62 จำนวนมิติของเวกเตอร์ลักษณะขาเข้า ของการทดลองย่อยที่เลือกมาในรูปที่ 4.61



รูปที่ 4.63 อัตราการรู้จำเฉลี่ย และอัตราการรู้จำของแต่ละทำนองเสียง เลือกจากการทดลองย่อยที่ให้อัตราการรู้จำเฉลี่ยสูงที่สุดเมื่อใช้เวกเตอร์ลักษณะแบบต่าง ๆ เมื่อผู้พูดเป็นผู้หญิง กรณีที่แบ่งทำนองเสียงเป็น 2 ประเภท (ลักษณะทุกแบบที่แสดงในกราฟ ได้รวมเอาความยาวของประโยค เข้าไปไว้ในเวกเตอร์ลักษณะด้วย)



รูปที่ 4.64 จำนวนมิติของเวกเตอร์ลักษณะขาเข้า ของการทดลองย่อยที่เลือกมาในรูปที่ 4.63

เมื่อพิจารณารูปที่ 4.64 จะเห็นได้ว่าการเพิ่ม FVC ก็สามารถช่วยให้เวกเตอร์ลักษณะขาเข้าของตัวรู้จำมีมิติลดลงเช่นกัน

เมื่อพิจารณาผลการรู้จำโดยรวม กรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท จะเห็นได้ว่าอัตราการรู้จำเฉลี่ยมีค่าสูงมากทั้งผู้หญิง และผู้ชาย แสดงให้เห็นว่าลักษณะของทำนองเสียงขึ้น และทำนองเสียงผสม เมื่อพิจารณาจาก F_0 แล้ว ค่อนข้างใกล้เคียงกันมาก เมื่อรวมทำนองเสียงทั้งสองเข้าเป็นประเภทเดียวกันแล้วจึงทำให้อัตราการรู้จำมีค่าสูงมาก ใช้เพียง LFC และความยาวของประโยคก็สามารถให้อัตราการรู้จำที่สูงได้ แต่การเพิ่ม FVC เข้าไปก็สามารถช่วยปรับปรุงอัตราการรู้จำให้ดีขึ้นได้อีก และยังช่วยลดจำนวนมิติของเวกเตอร์ลักษณะขาเข้าของโครงข่ายประสาทเทียมได้อีกด้วย ส่วนการเพิ่ม ΔLFC เข้าไปนั้น ในกรณีที่เป็นเสียงผู้หญิงก็สามารถช่วยปรับปรุงอัตราการรู้จำได้เช่นกัน แต่กรณีที่เป็นเสียงผู้ชายนั้น ΔLFC ไม่ได้ช่วยปรับปรุงการรู้จำเลย และยังทำให้อัตราการรู้จำต่ำลงเล็กน้อยด้วย ซึ่งอาจเกิดจากการที่เวกเตอร์ลักษณะขาเข้ามีจำนวนมิติมากขึ้น ทำให้ฝึกฝนได้ลำบากขึ้น

4.5 รูปร่างของคอนทัวร์ LFC และ FVC โดยเฉลี่ย ของทำนองเสียงพูดแต่ละประเภท

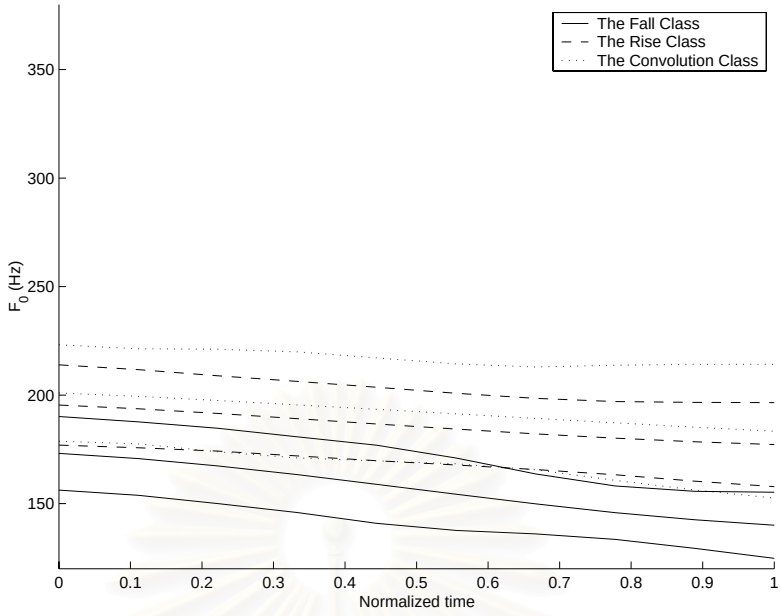
หัวข้อนี้นำเอาคอนทัวร์ LFC และ FVC ของประโยคเสียงพูดทั้งหมดที่ใช้ในการทดลอง มาหาคอนทัวร์เฉลี่ย โดยแบ่งตามประเภทของทำนองเสียงพูด และเพศของผู้พูด ค่าความถี่ตัดของตัวกรอง ได้มาจากการทดลองในข้อ 4.3 และ 4.4 ที่ให้อัตราการรู้จำสูงที่สุด โดยในแต่ละทำนองเสียงจะแสดงกราฟ 3 เส้น เส้นบน คือคอนทัวร์โดยเฉลี่ยของ LFC + FVC เส้นกลาง คือคอนทัวร์โดยเฉลี่ยของ LFC และเส้นล่าง คือคอนทัวร์โดยเฉลี่ยของ LFC – FVC ซึ่งทำให้เห็นแนวโน้มโดยเฉลี่ยของคอนทัวร์ F_0 ของทำนองเสียงประเภทต่าง ๆ ว่ามีลักษณะ เช่นใด เนื่องจากคอนทัวร์ F_0 จะแกว่งขึ้นลงไปตามคอนทัวร์ทั้งสาม ในลักษณะเดียวกับรูปที่ 3.14 ถึง 3.21 นั่นคือ ถ้ากราฟมีลักษณะลดระดับลง แสดงว่าคอนทัวร์ F_0 ก็มีการลดระดับลง ความสูงต่ำของกราฟ ก็แสดงให้เห็นถึงความสูงต่ำโดยเฉลี่ยของคอนทัวร์ F_0 เช่นเดียวกัน และนอกจากนี้ระยะห่างระหว่างเส้นกราฟทั้งสามก็คือเส้นกราฟ FVC นั่นเอง ซึ่งแสดงให้เห็นถึงช่วงกว้างในการแกว่งของคอนทัวร์ F_0

รูปที่ 4.65 และ 4.66 แสดงให้เห็นคอนทัวร์โดยเฉลี่ยของทำนองเสียงประเภทต่าง ๆ ในกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท รูปที่ 4.67 และ 4.68 แสดงให้เห็นคอนทัวร์โดยเฉลี่ยของทำนองเสียงประเภทต่าง ๆ ในกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภท

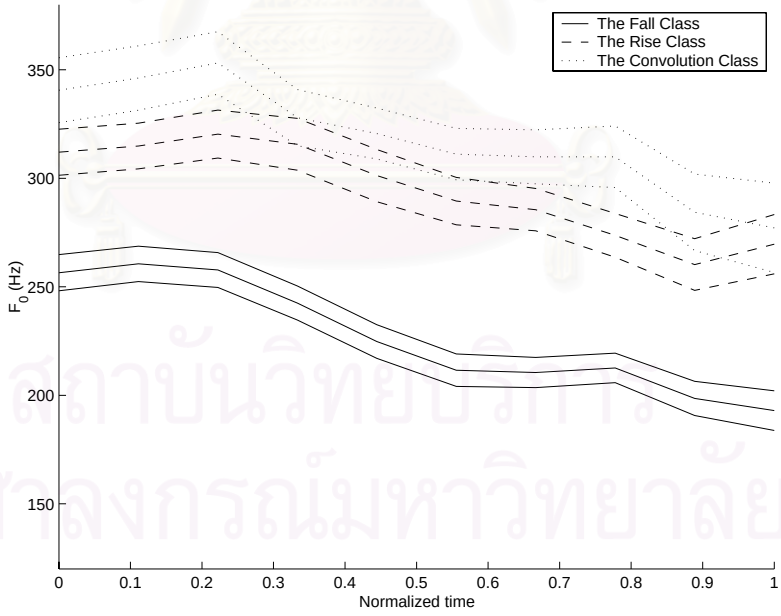
จากรูปที่ 4.65 และ 4.66 จะเห็นว่า คอนทัวร์ F_0 ของทำนองเสียงตก โดยเฉลี่ยจะแตกต่างจากทำนองเสียงขึ้น และทำนองเสียงผสมอย่างค่อนข้างชัดเจน นั่นคือ อยู่ระดับต่ำกว่า และช่วงการแกว่งของ F_0 แคบกว่าเล็กน้อย ส่วนในกรณีของทำนองเสียงขึ้น และทำนองเสียงผสมนั้นจะเห็นได้ว่าคอนทัวร์ F_0 โดยเฉลี่ยมีความใกล้เคียงกันมาก แต่ก็พอจะสามารถสังเกตเห็นความแตกต่างได้ นั่นคือ คอนทัวร์ F_0 ของทำนองเสียงผสมจะมีระดับสูงกว่า และมีช่วงการแกว่งที่กว้างกว่าทำนองเสียงขึ้นเล็กน้อย และเมื่อเปรียบเทียบระหว่างผู้พูดหญิง และผู้พูดชาย จะเห็นได้อย่างชัดเจนว่า คอนทัวร์ F_0 ของแต่ละทำนองเสียงของผู้พูดหญิง มีความแตกต่างกันมากกว่าผู้พูดชาย

จากการวิเคราะห์ลักษณะของเส้น LFC, LFC + FVC และ LFC – FVC โดยเฉลี่ย ในกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท จะเห็นได้ว่ารูปร่างของคอนทัวร์เหล่านี้ สอดคล้องกับอัตราการรู้จำทำนองเสียง นั่นคือ อัตราการรู้จำทำนองเสียงตก จะสูงกว่าอัตราการรู้จำทำนองเสียงขึ้นและทำนองเสียงผสม เนื่องจากลักษณะของคอนทัวร์ F_0 ของทำนองเสียงตกลักษณะแตกต่างจากทำนองเสียงประเภทอื่นมาก นอกจากนี้อัตราการรู้จำโดยเฉลี่ยของเสียงผู้หญิงสูงกว่าเสียงผู้ชาย ก็เนื่องมาจากลักษณะของคอนทัวร์ F_0 ของเสียงผู้หญิงในแต่ละทำนองเสียง แตกต่างกันมากกว่าเสียงผู้ชาย

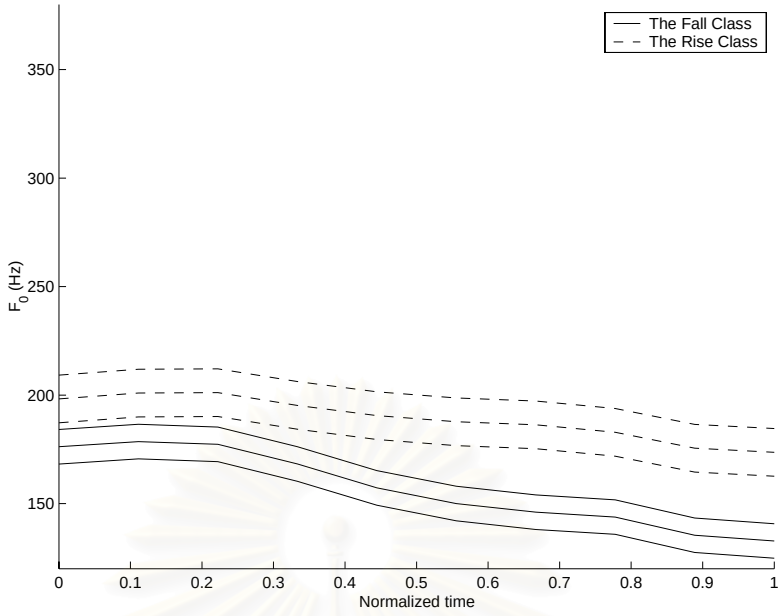
เมื่อรวมทำนองเสียงผสมไปเป็นประเภทเดียวกับทำนองเสียงขึ้น จะได้คอนทัวร์ลักษณะโดยเฉลี่ยแสดงดังรูปที่ 4.67 และ 4.68 โดยจะเห็นได้ว่าทำนองเสียงตก และทำนองเสียงขึ้นมีความแตกต่างกันอย่างชัดเจน จึงมีผลทำให้อัตราการรู้จำโดยเฉลี่ยสูงขึ้น นอกจากนี้จะเห็นได้ว่าลักษณะของคอนทัวร์ F_0 ของทำนองเสียงทั้งสองประเภทของผู้พูดหญิง มีความแตกต่างกันมากกว่าผู้พูดชาย จึงทำให้อัตราการรู้จำทำนองเสียงโดยเฉลี่ยของผู้พูดหญิง ยังคงมากกว่าผู้พูดชาย



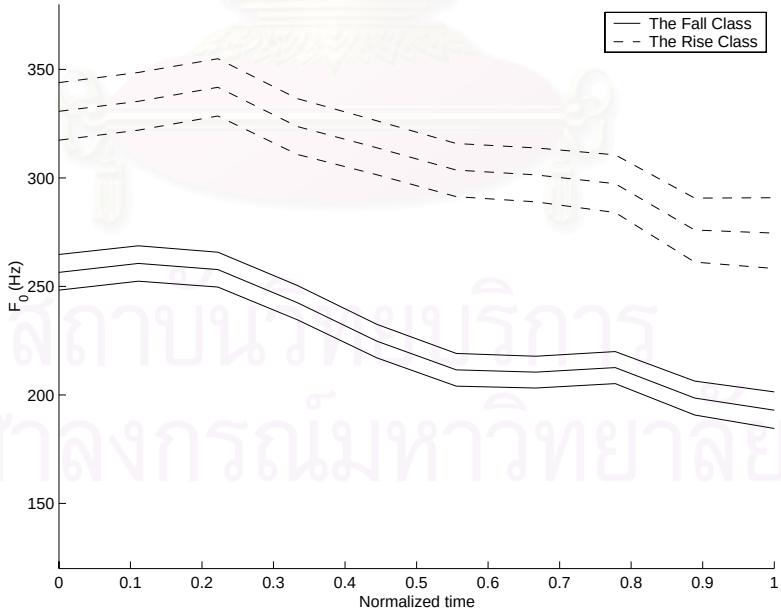
รูปที่ 4.65 คอนทัวร์ LFC + FVC (เส้นบนสุดของเส้นกราฟแต่ละแบบ) คอนทัวร์ LFC (เส้นกลางของเส้นกราฟแต่ละแบบ) และคอนทัวร์ LFC – FVC (เส้นล่างสุดของเส้นกราฟแต่ละแบบ) โดยเฉลี่ยของเสียงผู้ชาย ในกรณีที่แบ่งทำนองเสียงเป็น 3 ประเภท (Fc_{LFC} = 0.5 Hz, Fc_{FVC} = 2.0 Hz)



รูปที่ 4.66 คอนทัวร์ LFC + FVC (เส้นบนสุดของเส้นกราฟแต่ละแบบ) คอนทัวร์ LFC (เส้นกลางของเส้นกราฟแต่ละแบบ) และคอนทัวร์ LFC – FVC (เส้นล่างสุดของเส้นกราฟแต่ละแบบ) โดยเฉลี่ยของเสียงผู้หญิง ในกรณีที่แบ่งทำนองเสียงเป็น 3 ประเภท (Fc_{LFC} = 3.0 Hz, Fc_{FVC} = 1.5 Hz)



รูปที่ 4.67 คอนทัวร์ LFC + FVC (เส้นบนสุดของเส้นกราฟแต่ละแบบ) คอนทัวร์ LFC (เส้นกลางของเส้นกราฟแต่ละแบบ) และคอนทัวร์ LFC – FVC (เส้นล่างสุดของเส้นกราฟแต่ละแบบ) โดยเฉลี่ยของเสียงผู้ชาย ในกรณีที่แบ่งทำนองเสียงเป็น 2 ประเภท (Fc_{LFC} = 2.5 Hz, FVC เป็นเส้นตรงที่มีความชันเป็น 0)



รูปที่ 4.68 คอนทัวร์ LFC + FVC (เส้นบนสุดของเส้นกราฟแต่ละแบบ) คอนทัวร์ LFC (เส้นกลางของเส้นกราฟแต่ละแบบ) และคอนทัวร์ LFC – FVC (เส้นล่างสุดของเส้นกราฟแต่ละแบบ) โดยเฉลี่ยของเสียงผู้หญิง ในกรณีที่แบ่งทำนองเสียงเป็น 2 ประเภท (Fc_{LFC} = 3.0 Hz, Fc_{FVC} = 1.0 Hz)

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอวิธีการหาคอนทัวร์ลักษณะ 2 ชนิดจากคอนทัวร์ F_0 คอนทัวร์แรก คือ คอนทัวร์ LFC ซึ่งแสดงให้เห็นถึงการลดระดับ หรือการเพิ่มระดับของคอนทัวร์ F_0 คอนทัวร์ที่สอง คือคอนทัวร์ FVC ซึ่งแสดงให้เห็นถึงช่วงกว้างในการแกว่งของ F_0 โดยงานวิจัยนี้ได้นำคอนทัวร์ทั้งสอง ไปใช้ในการรู้จำทำนองเสียงพูดภาษาไทย ของเสียงพูดในลักษณะของการพูดโต้ตอบ พร้อมทั้งได้สร้างระบบรู้จำทำนองเสียงพูดจากคอนทัวร์ลักษณะดังกล่าว โดยใช้โครงข่ายประสาทเทียม เพื่อทดสอบประสิทธิภาพของคอนทัวร์ลักษณะ การรู้จำทำนองเสียงเป็นแบบขึ้นกับผู้พูด นั่นคือผู้พูดกลุ่มฝึกฝน และผู้พูดกลุ่มทดสอบเป็นกลุ่มเดียวกัน นอกจากนี้ยังเป็นการทดลองแบบขึ้นกับเพศของผู้พูดอีกด้วย นั่นคือ ทดลองเสียงของผู้ชาย และเสียงของผู้หญิง แยกจากกัน ประโยคเสียงพูดที่ใช้ในงานวิจัยนี้ จะเลือกเอาเฉพาะประโยคที่มนุษย์ฟังแล้วสามารถบอกประเภทของทำนองเสียงได้อย่างไม่กำกวมเท่านั้น โดยถือว่าประโยคที่ผู้ฟังฟังแล้วเลือกให้เป็นทำนองเสียงประเภทเดียวกัน 3 ครั้งขึ้นไปจากการฟังทั้งหมด 4 ครั้ง จึงจะถือว่าเป็นประโยคที่มีทำนองเสียงที่ไม่กำกวม

งานวิจัยนี้แบ่งการทดลองรู้จำทำนองเสียงออกเป็น 2 การทดลองใหญ่ ๆ คือ การทดลองการรู้จำทำนองเสียงในกรณีที่แบ่งทำนองเสียงเป็น 3 ประเภท คือ ทำนองเสียงตก ทำนองเสียงขึ้น และทำนองเสียงผสม กับการทดลองที่แบ่งทำนองเสียงออกเป็น 2 ประเภท คือ ทำนองเสียงตก เหมือนกับการทดลองแรก และทำนองเสียงขึ้น ซึ่งได้จากการรวมเอาทำนองเสียงขึ้น และทำนองเสียงผสมจากการทดลองแรกไว้เป็นประเภทเดียวกัน เวกเตอร์ลักษณะที่ใช้เป็นข้อมูลป้อนเข้าโครงข่ายประสาทเทียม ได้จากความยาวของประโยค และการนำคอนทัวร์ลักษณะ LFC และ FVC มาสุ่มตัวอย่าง ในการสุ่มตัวอย่างนี้จะเลือกใช้จำนวนจุดในการสุ่มตัวอย่างตามทฤษฎีการสุ่มตัวอย่างของไนควิสต์ ซึ่งขึ้นกับค่าความถี่ตัดของตัวกรองที่ใช้ในการหาคอนทัวร์ LFC และ FVC ดังนั้น การเปลี่ยนค่าความถี่ตัดก็จะเป็นการเปลี่ยนจำนวนมิติของเวกเตอร์ลักษณะด้วย ซึ่งก็จะมีผลต่อความซับซ้อน และระยะเวลาที่ใช้ในการฝึกฝนของโครงข่ายประสาทเทียม

ทั้งการทดลองที่แบ่งทำนองเสียงออกเป็น 3 ประเภท และการทดลองที่แบ่งทำนองเสียงออกเป็น 2 ประเภท จะเริ่มต้นการทดลองจากการใช้ลักษณะพื้นฐาน คือ ความยาวของประโยค และจุดสุ่มตัวอย่างจากคอนทัวร์ LFC จากนั้น จึงทดลองเพิ่ม ΔLFC และ FVC เข้าไป เพื่อสังเกตว่าคอนทัวร์ลักษณะทั้งสองมีส่วนช่วยปรับปรุงอัตราการรู้จำได้มากน้อยเพียงใด

ในการทดลองที่แบ่งทำนองเสียงออกเป็น 3 ประเภท จากการทดลองพบว่าการเพิ่มทั้ง ΔLFC และ FVC เข้าไป นอกเหนือจากลักษณะพื้นฐาน ต่างก็สามารถช่วยให้อัตราการรู้จำเพิ่มขึ้น โดยจะได้ว่า FVC ช่วยเพิ่มอัตราการรู้จำได้มากกว่า ΔLFC ทั้งในกรณีของเสียงผู้หญิง และเสียงผู้ชาย โดยในกรณีของเสียงผู้ชายพบว่า กรณีที่เพิ่มเพียง FVC กับกรณีที่เพิ่มทั้ง FVC และ ΔLFC นั้น ให้อัตราการรู้จำที่เท่ากันที่ร้อยละ 61.6 แต่การเพิ่มทั้ง FVC และ ΔLFC จะทำให้เวกเตอร์ลักษณะที่ได้ มีจำนวนมิติน้อยกว่าการเพิ่มเพียง FVC ส่วนในกรณีของเสียงผู้หญิงอัตราการรู้จำสูงสุดอยู่ที่ร้อยละ 73.7 ซึ่งเกิดขึ้นเมื่อเพิ่มทั้ง ΔLFC และ FVC

เมื่อเปรียบเทียบอัตราการรู้จำในแต่ละทำนองเสียง ของกรณีที่แบ่งทำนองเสียงออกเป็น 3 ประเภท พบว่าอัตราการรู้จำทำนองเสียงตกจะสูงกว่าทำนองเสียงประเภทอื่นอย่างเห็นได้ชัด ทั้งเสียงผู้หญิง และเสียงผู้ชาย ในทุกการทดลอง โดยพบว่าระบบรู้จำยังคงมีความสับสนระหว่างทำนองเสียงขึ้น และทำนองเสียงผสมอยู่มาก อย่างไรก็ตาม ในกรณีที่ให้อัตราการรู้จำสูงสุดของผู้หญิง และผู้ชายพบว่าระบบรู้จำสามารถจำแนกความแตกต่างระหว่างทำนองเสียงทั้ง 2 ประเภทได้บ้าง แต่ไม่มากนัก แสดงให้เห็นว่าจำเป็นต้องมีการศึกษาเพิ่มเติมเพื่อหาลักษณะจากเสียงพูด ที่สามารถแสดงให้เห็นถึงความแตกต่างระหว่างทำนองเสียงขึ้น และทำนองเสียงผสม มากกว่าคอนทัวร์ลักษณะที่นำเสนอ อย่างไรก็ตามจะเห็นได้ว่าคอนทัวร์ลักษณะทั้งสองเส้นสามารถจำแนกทำนองเสียงตก ออกจากทำนองเสียงประเภทอื่นได้ดี จึงสามารถนำคอนทัวร์ทั้งสองไปใช้ในการจำแนกทำนองเสียงตกออกจากทำนองเสียงประเภทอื่นก่อน แล้วจึงหาลักษณะจากเสียงพูดอื่น ๆ มาใช้จำแนกระหว่างทำนองเสียงขึ้น และทำนองเสียงผสม งานวิจัยนี้จึงได้เพิ่มการทดลองซึ่งแบ่งทำนองเสียงออกเป็น 2 ประเภท โดยจัดให้ทำนองเสียงผสมเป็นประเภทเดียวกับทำนองเสียงขึ้น เพื่อทดสอบประสิทธิภาพของระบบรู้จำในการจำแนกทำนองเสียงตก ออกจากทำนองเสียงประเภทอื่น ๆ

ในการทดลองที่แบ่งทำนองเสียงออกเป็น 2 ประเภท ในกรณีของเสียงผู้ชายพบว่า การใช้เพียงลักษณะพื้นฐาน คือ LFC และความยาวของประโยค ก็สามารถให้อัตราการรู้จำสูงที่สุดแล้ว คือ ร้อยละ 81.7 แต่การเพิ่มลักษณะจาก FVC เข้าไป ถึงแม้จะไม่ทำให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้น แต่ก็สามารถช่วยให้จำนวนมิติของเวกเตอร์ลักษณะลดลงได้ นอกจากนี้ยังช่วยให้ความแปรปรวนของอัตราการรู้จำของแต่ละทำนองเสียงลดลง นั่นคืออัตราการรู้จำของทำนองเสียงตก และทำนองเสียงขึ้นมีค่าใกล้เคียงกันมากขึ้น ส่วนในกรณีของเสียงผู้หญิงนั้นพบว่า การเพิ่มทั้ง ΔLFC และ FVC สามารถช่วยให้อัตราการรู้จำเฉลี่ยมีค่าสูงขึ้นได้ทั้งคู่ โดยเมื่อเพิ่มทั้ง ΔLFC และ FVC เข้าไปทำให้อัตราการรู้จำเฉลี่ยมีค่าสูงที่สุดอยู่ที่ร้อยละ 90.8

เมื่อเปรียบเทียบอัตราการรู้จำในแต่ละทำนองเสียง ของกรณีที่แบ่งทำนองเสียงออกเป็น 2 ประเภทพบว่า อัตราการรู้จำของทำนองเสียงตก และทำนองเสียงขึ้นมีค่าสูงใกล้เคียงกัน ซึ่งส่งผลให้อัตราการรู้จำเฉลี่ยมีค่าสูงขึ้นมาก เมื่อเทียบกับกรณีที่แบ่งทำนองเสียงเป็น 3 ประเภท แสดงให้เห็นว่าลักษณะของคอนทัวร์ F_0 ของทำนองเสียงขึ้น และทำนองเสียงผสมมีความใกล้เคียงกันมาก โดยลักษณะของความใกล้เคียงกันนี้สามารถสังเกตได้จากคอนทัวร์ LFC และ FVC โดยเฉลี่ย ดังแสดงในรูปที่ 4.65 ถึง 4.68

จากการเปรียบเทียบในแต่ละการทดลอง จะเห็นได้ว่า อัตราการรู้จำทำนองเสียงพูดโดยเฉลี่ยของเสียงผู้หญิง มากกว่าเสียงผู้ชาย แสดงให้เห็นว่าผู้พูดที่เป็นผู้หญิง สามารถพูดด้วยเสียงที่แสดงให้เห็นความแตกต่างระหว่างทำนองเสียงได้ดีกว่าผู้ชาย ซึ่งสามารถสังเกตได้จากคอนทัวร์ LFC และ FVC โดยเฉลี่ย ดังแสดงในรูปที่ 4.65 ถึง 4.68 เช่นกัน

5.2 ข้อเสนอแนะ

1. ระบบรู้จำทำนองเสียงพูดที่ใช้ในงานวิจัยนี้ เป็นเพียงระบบที่สร้างขึ้นเพื่อทดสอบประสิทธิภาพของคอนทัวร์ลักษณะที่หาได้จากคอนทัวร์ F_0 เท่านั้น ไม่ควรที่จะนำระบบไปใช้จริงในทางปฏิบัติในลักษณะของระบบโดด ๆ เนื่องจากจะเห็นได้ว่าระบบยังมีข้อจำกัดอยู่มาก เช่น ใช้ได้เฉพาะกับประโยคเสียงพูดที่สามารถจำแนกประเภทของทำนองเสียงได้อย่างชัดเจนโดยมนุษย์เท่านั้น

2. สิ่งที่จะนำไปใช้ได้จริงซึ่งเป็นผลจากงานวิจัยนี้ คือวิธีการสกัดลักษณะจากคอนทัวร์ F_0 ซึ่งงานวิจัยนี้ได้แสดงให้เห็นว่าการใช้ LFC และ FVC ช่วยให้สามารถจำแนกประเภททำนองเสียงได้ดีกว่าการใช้คอนทัวร์ F_0 โดยตรง (ในงานวิจัยนี้ไม่ได้ใช้คอนทัวร์ F_0 โดยตรง แต่สามารถประมาณได้ว่า การสุ่มตัวอย่างคอนทัวร์ LFC ที่มีค่า Fc_{LFC} สูง ๆ ให้ผลใกล้เคียงกับการสุ่มตัวอย่างจากคอนทัวร์ F_0 โดยตรง ดังแสดงในรูปที่ 3.29)

3. ลักษณะของการนำไปใช้งานจริงแบบหนึ่งที่แนะนำ คือ การนำไปใช้บอกความมุ่งหมายของผู้พูด (speaker intention) ในระบบโต้ตอบโดยใช้เสียง (spoken dialogue system) โดยใช้ในการตรวจจับว่าผู้พูดต้องการที่จะให้ประโยคพูดเป็นประโยคบอกเล่า หรือต้องการถาม หรือแสดงเจตนารมณ์อื่น ๆ และควรจะใช้ร่วมกับระบบรู้จำเสียงพูดด้วย ซึ่งจะสังเกตเห็นได้ว่าในภาษาไทย มีประโยคมากมายที่สามารถบอกได้ทันทีว่าเป็นประโยคคำถาม โดยดูจากคำสำคัญต่าง ๆ เช่น “ใคร?” “อะไร?” “ที่ไหน?” “อย่างไร?” “ทำไม?” “…หรือเปล่า?” “...หรือไม่?” “...เหรอ?” เป็นต้น ประโยคเหล่านี้สามารถบอกว่าเป็นประโยคคำถามหรือไม่

ได้อย่างง่ายดายโดยการนำผลการรู้จำเสียงพูดไปวิเคราะห์ทางไวยากรณ์ ประโยคประเภทนี้จึงไม่จำเป็นต้องนำมาผ่านกระบวนการวิเคราะห์ทำนองเสียงพูด

ประโยคที่จำเป็นต้องนำมาวิเคราะห์ทำนองเสียงพูด คือ ประโยคที่ผู้พูดสามารถพูดให้เป็นประโยคบอกเล่า ประโยคคำถาม หรือพูดเพื่อแสดงความตื่นเต้น หรือตกใจ ก็ได้ โดยใช้คำที่เหมือนกันทุกคำ และไม่มีคำที่แสดงถึงคำถามตามหลักไวยากรณ์ ซึ่งสามารถพบได้บ่อยในภาษาพูดทั่วไป เช่น “เค้าออกไปแล้ว” “เค้าออกไปแล้ว?” และ “เค้าออกไปแล้ว!” ตัวอย่างของการประยุกต์ใช้ในลักษณะนี้ ได้แก่ ระบบ VERBMOBIL (Nöth และคนอื่น ๆ, 2001) ที่กล่าวถึงในบทที่ 1 ซึ่งใช้ทำนองเสียงพูดในการจำแนกว่าเป็นประโยคคำถามหรือไม่ในกรณีที่ไม่มีคำไวยากรณ์

4. เนื่องจากประเภทของวรรณยุกต์ และประเภทของทำนองเสียง ต่างส่งผลต่อรูปร่างของคอนทัวร์ F_0 ด้วยกันทั้งคู่ แนวทางการประยุกต์ใช้งานอีกอย่างหนึ่งของงานวิจัยนี้ คือ การนำคอนทัวร์ลักษณะของทำนองเสียง ไปช่วยในระบบรู้จำเสียงวรรณยุกต์แบบหลายทำนองเสียง ซึ่งต่างจากระบบรู้จำเสียงวรรณยุกต์ในปัจจุบัน ซึ่งนิยมนำไปใช้กับทำนองเสียงตกเท่านั้น

5. ควรมีการศึกษาหาวิธีการนำอัตราเร็วในการพูด หรือช่วงระยะเวลาการออกเสียงในส่วนต่าง ๆ ของประโยค และระดับพลังงานของเสียงพูดมาช่วยในการจำแนกประเภทของทำนองเสียงพูดภาษาไทยด้วย นอกเหนือจากการใช้เพียงลักษณะจาก F_0 เพื่อให้ระบบรู้จำสามารถจำแนกความแตกต่างระหว่างทำนองเสียงขึ้น และทำนองเสียงคอนโวลูชันได้ดีขึ้น

รายการอ้างอิง

ภาษาไทย

กาญจนา นาคสกุล. ระบบเสียงภาษาไทย. 2,000 เล่ม. พิมพ์ครั้งที่ 4 (แก้ไขปรับปรุง). โครงการตำรา คณะอักษรศาสตร์ ลำดับที่ 38. กรุงเทพมหานคร : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2541.

ชัย วุฒิวิวัฒน์ชัย. การรู้จำเสียงคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูด โดยใช้เทคนิคแบบฟัซซีและนิวรอลเน็ตเวอร์ก. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2540.

ณัฏฐชา จิตติวรางกูล. กรรมวิธีการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2541.

ธีระ ภัทราพรนันท์. การรู้จำเสียงพูดสระภาษาไทยโดดๆ ไม่ขึ้นกับผู้พูด โดยการวัดสเปกตรัมดิสแตนซ์และใช้ไดนามิกไทม์วาร์ปปิง. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538.

ระพีพัฒน์ เพ็ญศิริ. การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นกับผู้พูดโดยการใช้ไดนามิกไทม์วาร์ปปิง. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538.

วิศรุต อาขุบุตร. ระบบรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.

วุฒิพงษ์ พรสุขจันทรา. การรู้จำเสียงตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้แอลพีซีและโครงข่ายประสาทเทียมแบบแบ็กพรอพาเกชัน. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.

เสาวลักษณ์ อารีพงศา. การรู้จำเสียงพูดตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธีฮิดเดน มาร์คอฟโมเดลและเวกเตอร์ควอนไตซ์เซชั่น. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538.

อุมาวสี ทาทอง. ระบบรู้จำคำเรียกพยัญชนะไทย: การศึกษาการวัดลักษณะสำคัญแบบต่างๆ. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2544.

เอกฤทธิ์ มณีน้อย. การรู้จำหน่วยเสียงภาษาไทยโดยใช้โครงข่ายประสาทเทียม. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2541.

ภาษาอังกฤษ

Abdou, S., and Scordilis, M. Integrating Multiple Knowledge Sources For Improved Speech Understanding. Proceedings of EUROSPEECH 2001 3 (2001) : 1783-1786.

Ahkuputra, V. An Acoustic Study of Syllable Onsets: A Basis for Thai Continuous Speech Recognition System. Doctor of Philosophy in Electrical Engineering Department of Electrical Engineering Faculty of Engineering Chulalongkorn University, 2002.

Batliner, A., Möbius, B., Möhler, G., Schweitzer, A., and Nöth, E. Prosodic models, automatic speech understanding, and speech synthesis: towards the common ground. Proceedings of EUROSPEECH 2001 4 (2001) : 2285-2288.

Boersma, P. Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound. Institute of Phonetic Sciences, University of Amsterdam, Proceedings 17 (1993) : 97-110.

Botinis, A., Granström, B., and Möbius, B. Development and paradigms in intonation research. Speech Communication 33, (2001) : 263-296.

Charnvivit, P., Jitapunkul, S., Ahkuputra, V., Maneenoi, E., Thathong, U., and Thampanitchawong, B. F0 Feature Extraction by Polynomial Regression Function for Monosyllabic Thai Tone Recognition. Proceedings of EUROSPEECH 2001 4 (2001) : 2753-2756.

Charnvivit, P., Thubthong, N., Maneenoi, E., Luksaneeyanawin, S., and Jitapunkul, S. Recognition of Intonation Patterns in Thai Utterance. Proceedings of EUROSPEECH 2003 (2003) : 137-140.

Deller, J. R., Proakis, J. G., and Hansen, J. H. L. Discrete-Time Processing of Speech Signals. New York : Macmillan Publishing Company, 1993.

Fujisaki, H. Analysis and modeling of fundamental frequency contours of Korean utterances – A preliminary study. Phonetics and Linguistics – in honour of Prof. H. B. Lee (1996a) : 640-657.

Fujisaki, H. From read speech to spontaneous speech – problems and approaches in the processing of prosody. Lecture in the opening seminar, ATR International Workshop on Computational Modeling of Prosody for Spontaneous Speech Processing, (1995).

Fujisaki, H. Prosody, Models, and Spontaneous Speech. In Y. Sagisaka, N. Campbell, and N. Higuchi (eds.), Computing Prosody, pp. 27-42. Berlin: Springer-Verlag, 1996b.

Fujisaki, H. The role of quantitative modeling in the study of intonation. Proceedings International Symposium on Japanese Prosody, (1992) : 163-174.

Fujisaki, H., Hallé, P., and Lei, H. Application of F_0 contour command-response model to Chinese tones. Reports of Autumn Meeting, Acoust. Soc. Jpn 1 (1987) : 197-198.

Fujisaki, H., and Hirose, K. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. Journal of the Acoustical Society of Japan (E) 5, 4(1984) : 233-241.

Fujisaki, H., and Hirose, K. Modeling the dynamic characteristics of voice fundamental frequency with application to analysis and synthesis of intonation. Papers for the Working Group on Intonation (1982) : 57-70.

Fujisaki, H., Ljungqvist, M., and Murata, H. Analysis and Modeling of Word Accent and Sentence Intonation in Swedish. Proc. 1993 Intern. Conf. Acoust. Speech and Signal Processing 2 (1993) : 211-214.

Fujisaki, H., and Ohno, S. Analysis and modeling of fundamental frequency contours of English utterances. Proceedings of EUROSPEECH 1995 2 (1995) : 985-988.

Fujisaki, H., Ohno, S., and Yagi, T. Analysis and modeling of fundamental frequency contours of Greek utterances. Proceedings of EUROSPEECH 1997 1 (1997) : 465-468.

Fujisaki, H., Ohno, S., Nakamura, K., Guirao, M., and Gurlekian, J. Analysis of accent and intonation in Spanish based on a quantitative model. Proceedings of ICSLP 1994 1 (1994) : 355-358.

Haykin, S. Neural Networks – A Comprehensive Foundation. United States of America : Macmillan College Publishing Company, 1994.

Hirst, D., and Cristo, A. D. A survey of intonation systems. In D. Hirst, and A. D. Cristo (eds.), Intonation systems: A Survey of Twenty Languages, pp. 1-44. United Kingdom : Cambridge University Press, 1998.

Huang, X., Acero, A., and Hon, H. W. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. United States of America : Prentice-Hall PTR, 2001.

Ishi, C. T., Minematsu, N., Nishide, R., and Hirose, K. Identification of Accent and Intonation in sentences for CALL systems. Proceedings of EUROSPEECH 2001 4 (2001) : 2455-2458.

Kießling, A., Kompe, R., Niemann, H., Nöth, E., and Batliner, A. “Roger”, “Sorry”, “I’m still listening”: Dialog Guiding Signals in Information Retrieval Dialogs. Proc. ESCA Workshop Prosody, D. House and P. Touati, Eds, Lund (1993) : 140-143.

Luksaneeyanawin, S. Intonation in Thai. In D. Hirst, and A. D. Cristo (eds.), Intonation systems: A Survey of Twenty Languages, pp. 376-394. United Kingdom : Cambridge University Press, 1998.

Maneenoi, E. An Acoustic Study of Syllable Rhymes: A Basis for Thai Continuous Speech Recognition System. Doctor of Philosophy in Electrical Engineering Department of Electrical Engineering Faculty of Engineering Chulalongkorn University, 2004.

Mixdorff, H. A Novel Approach to the Fully Automatic Extraction of Fujisaki Model Parameters. Proceedings ICASSP 2000 3 (2000) : 1281-1284.

Mixdorff, H. An Integrated Approach to Modeling German Prosody. Doktor-Ingenieur habilitatus Der Fakultät Elektrotechnik and Informationstechnik der Technischen Universität Dresden, 2002.

Mixdorff, H. Intonation Patterns of German – Model-based Quantitative Analysis and Synthesis of F_0 contours. Doktor-Ingenieurs Von der Fakultät Elektrotechnik der Technischen Universität Dresden, 1998.

Mixdorff, H., Bach, N. H., Fujisaki, H., and Luong, M. C. Quantitative Analysis and Synthesis of Syllabic Tones in Vietnamese. Proceedings of EUROSPEECH 2003 (2003) : 177-180.

Mixdorff, H., Fujisaki, H., Chen, G. P., and Hu, Y. Towards the Automatic Extraction of Fujisaki model Parameters for Mandarin. Proceedings of EUROSPEECH 2003 (2003) : 873-876.

Mixdorff, H., Luksaneeyanawin, S., Charnvivit, P., and Thubthong, N. Modeling Rhythmic Variation in Thai and its Application to Speech Synthesis. Proceedings of the ICPhS 2003 (2003).

Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H., and Charnvivit, P. Perception of Tone and Vowel Quantity in Thai. Proceedings of ICSLP 2002 (2002) : 753-756.

Ngarmchatetanarom, N., Maneenoi, E., Asdornwised, W., and Jitapunkul, S. Tone Recognition of Thai Continuous Speech Using Fujisaki’s Model. Proceedings of the 17th annual IEEE Canada Conference (2004).

Nöth, E., Batliner, A., Kießling, A., Kompe, R., and Niemannm, H. VERBMOBIL: The Use of Prosody in the Linguistic Components of a Speech Understanding System. IEEE Transaction on Speech and Audio Processing 8, 5 (September 2001) : 519-532.

Potisuk, S., Harper, M. P., and Gandour, J. Classification of Thai Tone Sequences in Syllable-Segmented Speech Using the Analysis-by-Synthesis Method. IEEE Transaction on Speech and Audio Processing 7, 1 (January 1999) : 95-102.

Potisuk, S., Harper, M. P., and Gandour, J. Speaker-Independent Automatic Classification of Thai Tones in Connected Speech by Analysis-Synthesis Method. Proc. Int. Conf. Acoustics, Speech, and Signal Processing (1995) : 632-635.

Rabiner, L. R., and Schafer, R. W. Digital Processing of Speech Signals. Eaglewood Cliffs, N.J. : Prentice-Hall, 1978.

Schneider, J. and Moore, A. W. A Locally Weighted Learning Tutorial using Vizier 1.0 [Online]. 1997. Available from: http://www-2.cs.cmu.edu/~schneide/tut5/tut5.html [2004, January 1]

Taylor, P. A. A Phonetic Model of English Intonation. Doctor of Philosophy University of Edinburgh, 1992.

Teixeira, J. P., Freitas, D., and Fujisaki, H. Prediction of Fujisaki Model's Phrase Commands. Proceedings of EUROSPEECH 2003 (2003) : 397-400.

Thubthong, N. A Study of Various Linguistic Effects on Tone Recognition in Thai Continuous Speech. Doctor of Philosophy in Computer Engineering Department of Computer Engineering Faculty of Engineering Chulalongkorn University, 2001.

Thubthong, N. A Thai tone recognition system based on phonemic distinctive features. Proc. the 2nd Symposium on Natural Language Processing (1995) : 379-386.

Thubthong, N., and Kijsirikul, B. Improving isolated Thai digit recognition using tone modeling. Proceedings of the 22nd Electrical Engineering Conference (1999) : 163-166.

Thubthong, N., Kijsirikul, B., and Luksaneeyanawin, S. Stress and Tone Recognition of Polysyllabic Words in Thai Speech. International Conference on Intelligent Technology (2001) : 356-368.

Toki, T., and Murata, M. Pronunciation & Task Listening – Innovative Workbooks in Japanese. Tokyo: Aratake Publishers, 1989.

Yuan, J., Shih, C., and Kochanski, G. P. Comparison of Declarative and Interrogative Intonation in Chinese. In B. Bel, and I. Marlien (eds.), Proceedings of the Speech Prosody 2002 Conference, pp. 711-714, 1998.

ภาคผนวก

ภาคผนวก ก

สัญลักษณ์แทนเสียงอ่านภาษาไทยที่ใช้ในงานวิจัยนี้

บางส่วนของงานวิจัยนี้ ได้ใช้สัญลักษณ์ แทนเสียงอ่าน (phonetic symbol) ของพยางค์ในภาษาไทย เช่น ประโยคที่พูดว่า “เพราะขาดงบประมาณ” สามารถเขียนโดยใช้สัญลักษณ์แทนเสียงอ่านได้เป็น /phr@3/khaat1/ngop3/pra1/maan0/ จะเห็นได้ว่าเป็นการเขียนสัญลักษณ์แทนเสียงอ่านของแต่ละพยางค์เรียงกัน

รูปแบบของสัญลักษณ์แทนเสียงอ่าน ของพยางค์ในภาษาไทย ที่ใช้ในงานวิจัยนี้มีดังนี้

$$\mathbf{C_{i1}(C_{i2})V(V)(C_f)T}$$

เมื่อ

C_{i1}	คือ เสียงพยัญชนะต้น (initial consonant)
C_{i2}	คือ เสียงพยัญชนะควบกล้ำ ในกรณีที่พยางค์นั้นเป็นพยางค์ควบกล้ำ
V	คือ เสียงสระ (vowel) โดยถ้าพยางค์นั้นเป็นสระเสียงสั้นจะใส่สัญลักษณ์แทนสระเพียงตัวเดียว แต่ถ้าพยางค์นั้นเป็นสระเสียงยาว จะใส่สองตัว
C_f	คือ เสียงพยัญชนะท้าย (final consonant) จะใส่ในกรณีที่พยางค์นั้นเป็นพยางค์ที่มีตัวสะกด
T	คือ ตัวเลขแทนเสียงวรรณเสียงวรรณยุกต์ (tone)

<u>หมายเหตุ</u> สัญลักษณ์ที่อยู่ในวงเล็บ หมายถึง หน่วยเสียงนั้นไม่จำเป็นต้องมีในทุกพยางค์ของภาษาไทย

ตารางที่ ก.1 เสียงพยัญชนะต้นในภาษาไทย (C_{i1})

พยัญชนะ	สัญลักษณ์แทนเสียงพยัญชนะ
ป	p
ต ฏ	t
จ	c
ก	k
อ	#
พ ภ ผ	ph
ท ธ ฒ ฑ ถ ฐ	th
ช ฌ ญ	ch
ข ค ฆ	kh
บ	b
ด ฎ	d
ม หม	m
น ณ หน	n
ง หง	ng
ฝ ฟ	f
ส ศ ษ ซ	s
ห ฮ	h
ร หร	r
ล ฬ หล	l
ว หว	w
ย ญ หย หญ	j

ตารางที่ ก.2 เสียงพยัญชนะควบกล้ำในภาษาไทย (C_{i2})

พยัญชนะควบกล้ำ	**สัญลักษณ์แทนเสียง**
ไม่มีตัวควบกล้ำ	ไม่ต้องใส่สัญลักษณ์
ร	r
ล	l
ว	w

ตารางที่ ก.3 เสียงพยัญชนะท้ายในภาษาไทย (C_f)

ตัวสะกด	**สัญลักษณ์แทนเสียง**
ไม่มีตัวสะกด	ไม่ต้องใส่สัญลักษณ์
แม่กง (ง)	ng
แม่กน (น ญ ณ ร ล ฬ)	n
แม่กม (ม)	m
แม่เกย (ย)	j
แม่เกอว (ว)	w
แม่กก (ก)	k
แม่กด (ด จ ฉ ช ซ ฌ ฎ ฏ ฐ ฑ ฒ ต ถ ท ธ ศ ษ ส)	t
แม่กบ (บ ป ผ ฝ พ ภ ฟ)	p

ตารางที่ ก.4 เสียงสระในภาษาไทย (V)

เสียงสระ	สัญลักษณ์แทนเสียง
อะ, อา	a, aa
อิ, อี	i, ii
อึ, อือ	v, vv
อุ, อู	u, uu
เอะ, เอ	e, ee
แอะ, แอ	x, xx
โอะ, โอ	o, oo
เอาะ, ออ	@, @@
เออะ, เออ	q, qq
เอียะ, เอีย	ia, iia
เอือะ, เอือ	va, vva
อัวะ, อัว	ua, uua

ตารางที่ ก.5 เสียงวรรณยุกต์ในภาษาไทย (T)

เสียงวรรณยุกต์	สัญลักษณ์แทนเสียง
สามัญ	0
เอก	1
โท	2
ตรี	3
จัตวา	4

ภาคผนวก ข

บทสนทนาที่ใช้ในงานวิจัย

บทสนทนาที่ใช้ในงานวิจัยนี้ ใช้บทสนทนาเดียวกับ Luksaneeyanawin, 1983 ซึ่งใช้ในการศึกษาเกี่ยวกับทำนองเสียงพูดของภาษาไทย ประกอบด้วยบทสนทนา 6 บท ดังนี้

<u>หมายเหตุ</u> fall หมายถึงทำนองเสียงตก rise หมายถึงทำนองเสียงขึ้น conv หมายถึงทำนองเสียงผสม

บทสนทนาที่ 1		
ผู้พูด ก	fall	แดงจะไปสมัครเป็นทหารเสือพราน
ผู้พูด ข	rise	เขาจะไปแน่?
ผู้พูด ก	conv fall	ดูท่าคงจะไปแน่ แข็งขันเหลือเกิน
ผู้พูด ข	fall conv	จะไปหรือไม่ไป พรุ่งนี้ก็รู้

บทสนทนาที่ 2		
ผู้พูด ก	rise rise	พรุ่งนี้สัมมนาที่กรม... เราต้องไปไหม?
ผู้พูด ข	fall conv fall	ไปก็ได้ ไม่ไปก็ได้ เถียงกันแต่เรื่องสัพเพเหระ ฉันขี้เกียจไปจะตาย

บทสนทนาที่ 3		
ผู้พูด ก	rise	เมื่อวานตู่กลับถึงบ้าน...
	fall	ไม่รู้โกรธอะไร
	fall	กระแทกนั่นกระแทกนี่ใหญ่
ผู้พูด ข	rise	ฝนมันดันตก...
	conv	รองเท้าหนังกลับคู่ใหม่เลยเสียโฉม
ผู้พูด ก	conv	อื้อฮือ!
	rise	ซื้อรองเท้าหนังกลับใส่หน้าฝน?
ผู้พูด ข	fall	เขาก็ทันสมัยของเขาเสมอ
	rise	ที่จริงแฟชั่นมันก็กลับไปกลับมานะ
	conv	ใช่
	conv	ก็กางเกงขาลีบน่ะ
	fall	กลับมาอีกแล้ว

บทสนทนาที่ 4		
ผู้พูด ก	rise	เมื่อวานพบแดงแล้ว...
	fall	ได้พูดเรื่องนั้นด้วย
ผู้พูด ข	rise	คุณบอกเขาแล้ว?
ผู้พูด ก	conv	ยัง
ผู้พูด ข	rise	ยัง?
	rise	แล้วพูดเรื่องอะไรกัน?
ผู้พูด ก	fall	เรื่องงานของป่อง
ผู้พูด ข	rise	เฮ้อ
	fall	นึกว่าจะพูดให้มันรู้แล้วรู้รอดไป

บทสนทนาที่ 5		
ผู้พูด ก	rise	ที่อังกฤษฝนมันตกเป็นหน้า ๆ...
	fall	หรือยังไง
ผู้พูด ข	rise	มันตกตลอดเวลา...
	fall	ไม่มีหน้าฝนเหมือนอย่างเราหรอก
ผู้พูด ก	rise	อะไร?
	rise	ตกมันทุกหน้า?
	rise	น้ำไม่ท่วมแย่หรือ?
ผู้พูด ข	conv	โอ๊ย!
	fall	เทศบาลเขาดี
	fall	ของเราสิ
	fall	ขายหน้าจะตาย
	conv	ตกทีไรเป็นน้ำท่วมทุกที!
	fall	ฉีกหน้านายกเทศมนตรี
ผู้พูด ก	conv	เขาจะไปเสียหน้าอะไร!
	rise	บอกว่าเครื่องสูบน้ำไม่พอ...
	fall	เพราะขาดงบประมาณ
	fall	ก็จบกัน

บทสนทนาที่ 6		
ผู้พูด ก	fall	ดูเสื้อตัวนั้นสิ
	rise	สวยดีนะ...
ผู้พูด ข	rise	คุณว่าเสื้อสวย?
	fall	ฉันว่านางแบบสวยมากกว่า
ผู้พูด ก	rise	นั่นนะซิ...
	conv	ช้วยสวยนะ
ผู้พูด ข	rise	ถ้าเลือกนางแบบสวย ๆ...
	fall	เสื้อก็ดูสวยไปด้วย
ผู้พูด ก	fall	จริงด้วย
	conv	ดูแบบนี้สิ!
	rise	อย่างเนี้ยเหรอ เขาเรียกว่าสวย
	fall	ฉันจะเป็นลม

ภาคผนวก ค

บทความทางวิชาการ

บทความทางวิชาการ ในงานประชุมวิชาการระดับนานาชาติ ซึ่งได้เขียนขึ้นในระหว่างที่ทำงานวิจัยนี้ ได้แก่

(1) Charnvivit, P., Jitapunkul, S., Ahkuputra, V., Maneenoi, E., Thathong, U., and Thampanitchawong, B. F0 Feature Extraction by Polynomial Regression Function for Monosyllabic Thai Tone Recognition. Proceedings of EUROSPEECH 2001 4 (2001) : 2753-2756.

(2) Charnvivit, P., Thubthong, N., Maneenoi, E., Luksaneeyanawin, S., and Jitapunkul, S. Recognition of Intonation Patterns in Thai Utterance. Proceedings of EUROSPEECH 2003 (2003) : 137-140.

สำเนาของบทความทั้งสองแสดงในหน้าถัดไป

F0 Feature Extraction by Polynomial Regression Function for Monosyllabic Thai Tone Recognition

Patavee Charnvivit, Somchai Jitapunkul, Visarut Ahkuputra, Ekkarit Maneenoi, Umavasee Thathong, and Boonchai Thampanitchawong

Digital Signal Processing Research Laboratory, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, THAILAND
e-mail : jsomchai@chula.ac.th

Abstract

This paper presents a monosyllabic Thai tone recognition system. The system is composed of three main processes, fundamental frequency (F0) extraction from input speech signal, analysis of F0 contour for feature extraction, and classification of each tone using the extracted features. In the F0 feature extraction, the polynomial regression functions are employed to fit the segmented F0 curve where its coefficients are used as a feature vector. In tone recognition, we used the maximum a posteriori probability classifier (MAP) to classify a tone by assuming that the feature is a multidimensional Gaussian random variable. The hypothetical words used in this paper are composed of numerical words and monosyllabic Thai words. The vocabulary set is composed of the short vowel words, the long vowel words and have the effect of initial and final consonant on the shape of F0 contour. The experimental results show that by using the system as a speaker-dependent system, the maximum recognition rate is 96.20% using three-dimension feature vector. The speaker-independent recognition rates are 79.99% for male and 82.80% for female using four-dimension feature vector.

1. Introduction

Thai is a tonal language. There are five tonemes in Thai, the mid, the low, the falling, the high and the rising. The feature of speech that was used to classify the tone is the shape of fundamental frequency (F0) contour, which shown in Figure 1. There are several parameters that also have the effect on the shape of F0 contour such as the gender and the age of speaker, the initial consonant, the final consonant and the duration of vowel (short or long). In this paper, we used the hypothetical words that consist of several effects. In our process, the F0 contour of input speech was automatically smoothed and segmented by the proposed algorithm in section 3.1. Then they were fit by the polynomial regression function, which we used its coefficients as the features of F0 contours. In the recognition process, we used the maximum a posteriori probability classifier (MAP) to classify the tones by assuming that the feature vectors are the multidimensional Gaussian random variables.

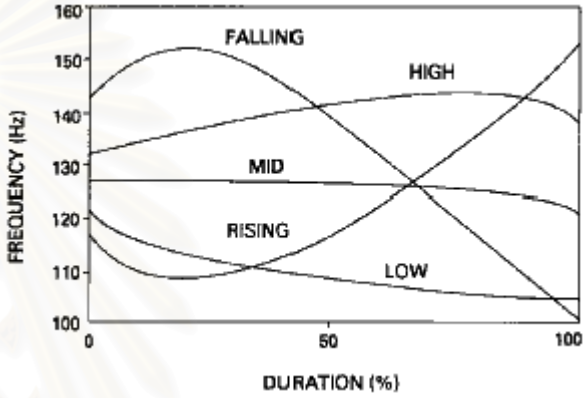


Figure 1: Average F_0 contours of the five Thai tones produced in isolation by a male speaker [3].

2. Thai Syllable Structure

The Thai syllable structure is composed of three different sound systems as follows. [2]

1. The system of consonants consists of 33 consonantal units, 21 consonants and 12 consonant clusters.
2. The system of Thai vowels consists of 18 monophthongs and 6 diphthongs. The monophthongs are qualitatively 9 different vowels, each of which has two members, short and long. Each of three different diphthongs also has 2 quantitatively different members.
3. The system of tones consists of 5 tones. There are 3 kinetic or relatively leveled tones, the high (H), the mid (M), and the low (L), and 2 dynamic or contour tones, the falling (F) and the rising (R).

The smallest construction of sounds or syllables in Thai is composed of one vowel unit or one diphthong, one two, or three consonants, and a tone. The construction can be represented with the structure as illustrated in Figure 2,

$$\mathbf{S = C_i(C_i)V^T(V)(C_f)}$$

Figure 2: Thai Syllable Structure

Where $\mathbf{C_i}$ is initial consonant, $\mathbf{C_f}$ is final consonant, $\mathbf{V}$ is vowel, and T is tone respectively.

3. F0 Feature Extraction

The F0 feature extraction process has two procedures. The first is F0 smoothing and segmentation procedure. The second is polynomial curve fitting procedure.

3.1. F0 smoothing and segmentation procedure

F0 from the F0 extraction process will be smoothed in the smoothing procedure by using median filtering. In the segmentation procedure, there is algorithm that was used to segment the smoothed F0. This algorithm will determine the beginning and the ending frame of the longest time that F0 at each frame has the value differ from the neighboring frame no more than ΔF_{max} = 17 Hz.

3.2. Polynomial regression

The objective of this procedure is to determine the coefficients b_k of a polynomial that fits the segmented F0 contour. Let $\mathbf{F} = (F_1, F_2, ..., F_L)^T$ be a sequence of segmented F0 of length L, $\hat{\mathbf{F}} = (\hat{F}_1, \hat{F}_2, ..., \hat{F}_L)^T$ be an estimated vector of $\mathbf{F}$. A d-dimension feature vector $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{d-1})^T$ is the coefficient of (d-1)-order polynomial regression function

$$\hat{F}_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + ... + \beta_{d-1} t_i^{d-1} \quad (1)$$

where $t_i = \frac{i}{L}$ is a normalized time respect to F_i. Equation (1) can be expressed in matrix form as

$$\hat{\mathbf{F}} = \mathbf{T}\boldsymbol{\beta}, \quad \begin{bmatrix} \hat{F}_1 \\ \hat{F}_2 \\ \vdots \\ \hat{F}_L \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{d-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{d-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_L & t_L^2 & \cdots & t_L^{d-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{d-1} \end{bmatrix} \quad (2)$$

A solution for $\boldsymbol{\beta}$ in a least-squares sense (minimize the Euclidean distance between vectors $\mathbf{F}$ and $\hat{\mathbf{F}}$) is obtained via forming the pseudoinverse of T, that is,

$$\boldsymbol{\beta} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{F} \quad (3)$$

However, if $\mathbf{T}^T\mathbf{T}$ is nearly singular, the numerical errors incurred in forming $\mathbf{T}^T\mathbf{T}$, and then forming the inverse, spawn a need for alternate approaches that are not plagued by numerical sensitivities. One solution, known as QR decomposition, can be used for this case.

4. Classification Algorithms

A decision rule that used in the tone recognition process is a maximum a posteriori probability classifier (MAP). Before using this classifier, the input feature vector $\boldsymbol{\beta}$ will be sent to the automatic gender identification procedure to determine the gender of speaker. This procedure will help the MAP process to select the proper model for each gender.

4.1. Automatic gender identification

The fundamental frequency of men is usually found somewhere between the two bounds 50 - 250 Hz, while for women the range usually falls somewhere in the interval 120 - 500 Hz [1]. So, the parameter that possible to be used to identify the gender is the F0 levels. The average value of the segmented F0 is used in this paper. The identification algorithm is that if the average is more than the threshold, the recognized gender is female otherwise the gender is male. Because the feature vector is composed of the coefficient of the regression polynomial that fit the segmented F0, the average value can be determined by these coefficients as shown in 5.1.1.

4.1.1. Estimating the mean of the segmented F0

The estimated F0 can be expressed as a continuous time function, that is,

$$\hat{F} = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_{d-1} t^{d-1} \quad (4)$$

where $t \in [0, \ 1]$.

The average value of this function, that used to approximate the mean of the segmented F0, can be determined by the integral from zero to one of $\hat{F}$ with respect to t as follows:

$$\begin{aligned} \hat{m} &= \int_0^1 \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_{d-1} t^{d-1} dt \\ &= \beta_0 + \frac{1}{2}\beta_1 + \frac{1}{3}\beta_2 + ... + \frac{1}{d}\beta_{d-1} \\ &= \mathbf{w}_d^T \boldsymbol{\beta} \end{aligned} \quad (5)$$

where $\mathbf{w}_d = (1, \ \frac{1}{2}, \ \frac{1}{3}, \ \ldots, \ \frac{1}{d})^T$.

4.1.2. F0 threshold

The threshold of F0, which used to classify the gender of speaker, can be determined by assuming that the probability density function (pdf) of $\hat{m}$ for each gender is Gaussian. We can apply the MAP classifier as shown in the next section but in the 1-dimension case to find this threshold.

4.2. Maximum a posteriori probability classifier (MAP)

A decision rule that used in the recognition process is a maximum a posteriori probability classifier (MAP). The MAP classifier will select tone i for the feature vector $\boldsymbol{\beta}$ of the unknown speech input if the posteriori probability

$$P(w_i \mid \boldsymbol{\beta}) > P(w_j \mid \boldsymbol{\beta}) \qquad , \forall j \neq i \quad (6)$$

where w_0, w_1, w_2, w_3 and w_4 are the class of tone M, L, F, H and R respectively. This probability can be determined from

$$P(w_i \mid \boldsymbol{\beta}) = \frac{[p(\boldsymbol{\beta} \mid w_i)P(w_i)]}{p(\boldsymbol{\beta})} \quad (7)$$

where $P(w_i)$, the priori class probability, is assumed to be equal for all tone.

The pdf of feature vectors is determined from the summation of all conditional pdf given each class, that is,

$$p(\boldsymbol{\beta}) = \sum_i p(\boldsymbol{\beta} \mid w_i) \quad (8)$$

Notice that the quantity $P(w_i)$ and $p(\boldsymbol{\beta})$ are common to all class-conditional probabilities; therefore, it represents a scaling factor that may be eliminated. Thus the decision algorithm become

select tone i if $p(\boldsymbol{\beta} \mid w_i) > p(\boldsymbol{\beta} \mid w_j)$, $\forall j \neq i$ (9)

In this paper, we assume the class-conditional pdf to be d-dimensional Gaussian pdf. Therefore,

$$p(\boldsymbol{\beta} \mid w_k) = (2\pi)^{-\frac{d}{2}} \left|\boldsymbol{\Sigma}_k\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_k)\right] \quad (10)$$

where $\boldsymbol{\beta}$ is d x 1 with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ for class k.

If we take the log function, a monotonically increasing function, to $p(\boldsymbol{\beta} \mid w_k)$ so the decision algorithm is

select tone i if $d_{ml}(\boldsymbol{\beta}, \boldsymbol{\mu}_i) < d_{ml}(\boldsymbol{\beta}, \boldsymbol{\mu}_j)$, $\forall j \neq i$ (11)

where we define the maximum likelihood distance as

$$d_{ml}(\boldsymbol{\beta}, \boldsymbol{\mu}_k) = (\boldsymbol{\beta}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_k) + \ln\left|\boldsymbol{\Sigma}_k\right| \quad (12)$$

5. Tone Recognition System

The block diagram of the tone recognition system is shown in Figure 3. The first block is the F0 extraction process which a single syllable is its input. The speech was recorded at 11025 Hz sampling frequency and 16-bit quantization level. F0 was computed in this process from 256 samples speech frame with the overlapping of ¾ frame by using the modified short-term autocorrelation with center clipping method. The second block is the F0 feature extraction process, which determines the parameters that have sufficient information to describe the shape of F0 contour by the method of polynomial regression. The final block is the tone recognition algorithm that uses the parameters obtained from the previous process to determine the best matching tone for the input speech.

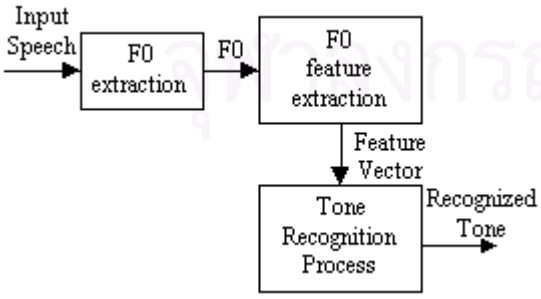


Figure 3: Block diagram of the system

6. Experiments

There are two experimental sets, the experiment on speaker-dependent tone recognition system and the experiment on speaker-independent tone recognition system. In the first one, there is no automatic gender identification system because the speaker of the training set and the testing set is the same. Both experiments use 30 hypothetical words as shown in Table 1.

Table 1: Hypothetical Words

Tone	Mid (M)	Low (L)	Falling (F)	High (H)	Rising (R)
Words	/'dqqn0/	/'paak1/	/'wing2/	/'nok3/	/'huu4/
	/'n@@n0/	/'pet1/	/'kluuaj2/	/'to3/	/'svva4/
	/'taa0/	/'kaj1/	/'som2/	/'nam3/	/'s@@ng4/
	/'mvv0/	/'hnvng1/	/'nang2/		/'saam4/
	/'thiian0/	/'sii1/	/'kxxw2/		/'suun4/
	/'kin0/	/'hok1/	/'haa2/		
	/'tiiang0/	/'cet1/	/'kaw2/		
		/'pxxt1/			

6.1. The automatic gender identification test

The average values of the segmented F0 for each gender in the training set are determined. The mean and the variance are also calculated. By assuming that the pdf of both genders are Gaussian, so we can determine the threshold for the minimum of error probability. The threshold is 161.5 Hz, which yield the accuracy of 100 % in the gender classification process of the testing set.

6.2. The speaker-dependent tone recognition test

In this case, the training set use utterances from three trials of all words in Table 1 and also using the other three trials from the same speaker as the testing set. By varying the dimension d of the feature vector from three to six, the recognition rate is shown in Figure 4. These results indicate that the recognizer has the maximum recognition rate at $d = 3$. The confusion matrix of this dimension is shown in Table 2.

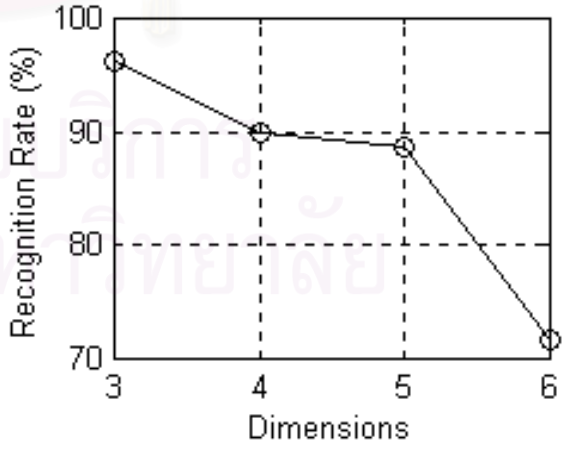


Figure 4: Recognition rate of the speaker-dependent tone recognition system

Table 2: Confusion matrix of the speaker-dependent tone recognition with $d = 3$

Desired Tone	Recognized Tone					Total	Accuracy (%)
	M	L	F	H	R		
Mid (M)	18	3	0	0	0	21	85.71
Low (L)	0	24	0	0	0	24	100.00
Falling (F)	1	0	20	0	0	21	95.24
High (H)	0	0	0	9	0	9	100.00
Rising (R)	0	0	0	0	15	15	100.00
						90	96.20

6.3. The speaker-independent tone recognition test

The training set in this experiment consists of the 30 hypothetical words similar to the previous section from four men and four women. The test set is composed of four men and women in the training set and the other four men and four women that do not have their speech in the training set. Similar to the previous section, the dimension of feature vectors is varied from three to six. Figure 5 shows the results compared between male and female speakers. These results differ from the case of speaker-dependent that for male, the recognition rate is increase when the dimension is increase until the dimension is four, for female, the recognition rate is approximately equal when the dimension is three and four. When the dimension is more than four, the recognition rate is decrease. The confusion matrices of male and female at the maximum accuracy are shown in Table 3 and 4 respectively.

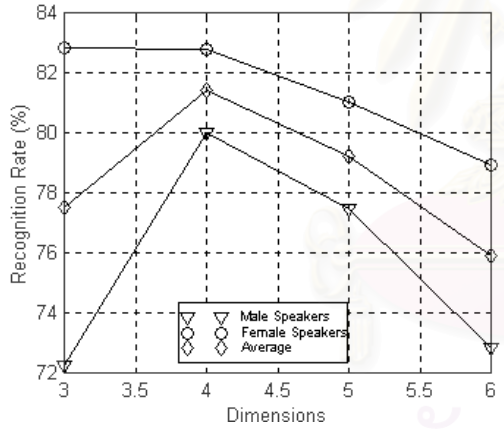


Figure 5: Recognition rate of the speaker-dependent tone recognition system

Table 3: Confusion matrix of the speaker-independent tone recognition with $d = 4$ for male speakers

Desired Tone	Recognized Tone					Total	Accuracy (%)
	M	L	F	H	R		
Mid (M)	40	6	8	2	0	56	71.43
Low (L)	17	44	1	1	1	64	68.75
Falling (F)	4	3	47	2	0	56	83.93
High (H)	1	2	1	20	0	24	83.33
Rising (R)	0	2	0	1	37	40	92.50
						240	79.99

Table 4: Confusion matrix of the speaker-independent tone recognition with $d = 4$ for female speakers

Desired Tone	Recognized Tone					Total	Accuracy (%)
	M	L	F	H	R		
Mid (M)	39	10	7	0	0	56	69.64
Low (L)	17	45	2	0	0	64	70.31
Falling (F)	8	0	48	0	0	56	85.71
High (H)	1	0	0	23	0	24	95.83
Rising (R)	0	2	0	1	37	40	92.50
						240	82.80

7. Conclusions

The system for monosyllabic Thai tone recognition has been proposed in this paper. We used the coefficient of polynomial regression function as a feature vector of the segmented F0 contour. In the training phase, the feature vector was used to determine the statistical parameters of the model for each gender. In the testing phase, the feature vector will be passed to the automatic gender identification to determine the gender of speaker and passed to the tone recognition process for each gender. The tone recognition process will determine the tone of the input speech by using the maximum a posteriori probability classifier. The results show that in the speaker-dependent system, the optimum dimension is three with recognition rate 96.2%. But in the speaker-independent system, the optimum dimension is four with recognition rate 79.99% for men and 82.80% for women.

8. Acknowledgement

The authors would like to acknowledge Digital Signal Processing Research Laboratory, Chulalongkorn University for the support of this research.

9. References

[1] Deller J. R., Proakis J. G., and Hansen J. H. L., "Discrete-Time Processing of Speech Signals", Macmillan Publishing Company, a division of Macmillan, Inc., United States of America, 1993.

[2] Maneenoi E., Jitapunkul S., Ahkuputra V., Thathong U., and Thampanitchawong B., "Thai Monophthong Recognition Using Continuous Density Hidden Markov Model and LPC Cepstral Coefficients", *Proceeding of International Conference on Spoken Language Processing (ICSLP 2000)*, Oct. 2000.

[3] Potisuk S., Harper M. P., and Gandour J., "Classification of Thai Tone Sequences in Syllable-Segmented Speech Using the Analysis-by-Synthesis Method", *IEEE Trans. On Speech and Audio Processing*, vol. 7, pp. 95-102, Jan. 1999.

[4] Schalkoff R. J., "Pattern Recognition: Statistical, Structural and Neural Approaches", John Wiley & Sons, Inc., Singapore, 1992.

[5] Thathong U., Jitapunkul S., Ahkuputra V., Maneenoi E., and Thampanitchawong B., "Classification of Thai Consonant Naming Using Thai Tone", *Proceeding of International Conference on Spoken Language Processing (ICSLP 2000)*, Oct. 2000.

[6] Thubthong, N. "A Thai Tone Recognition system based on Phonemic Distinctive Features". *Department of Computer Engineering Faculty of Engineering, Chulalongkorn University.*

Recognition of Intonation Patterns in Thai Utterance

Patavee Charnvivit[1], Nuttakorn Thubthong[2], Ekkarit Maneenoi[1], Sudaporn Luksaneeyanawin[2], and Somchai Jitapunkul[1]

[1]Digital Signal Processing Research Laboratory, Department of Electrical Engineering,
[2]Centre for Research in Speech and Language Processing, Department of Linguistics
Chulalongkorn University, Bangkok, Thailand
E-mail: patavee@chula.com Nuttakorn.T@chula.ac.th ekkarit@chula.com
Sudaporn.L@chula.ac.th jsomchai@chula.ac.th

Abstract

Thai intonation can be categorized as paralinguistic information of F0 contour of the utterance. There are three classes of intonation pattern in Thai, the Fall Class, the Rise Class, and the Convolution Class. This paper presents a method of intonation pattern recognition of Thai utterance. Two intonation feature contours, extracted from F0 contour, were proposed. The feature contours were converted to feature vector to use as input of neural network recognizer. The recognition results show that an average recognition rate is 63.4% for male speakers and 75.4% for female speakers. The recognizer can recognize the Fall Class from the others better than distinguish between the Rise Class and the Convolution Class.

1. Introduction

Like many languages, the information from F0 contour in Thai speech can be categorized into three types, linguistic, paralinguistic, and nonlinguistic information. Linguistic information in Thai is known as 5 lexical tones. Nonlinguistic information is composed of the factors such as physical and emotional states of the speaker [1]. Paralinguistic information represents different types of attitudinal meanings of the sentence. This type of information can be realized as intonation in Thai.

There are 3 classes of intonation patterns or tune in Thai continuous speech [2]. The class that the F0 at the beginning of the utterance is higher than the F0 at the end of the utterance was defined as “the Fall Class” or “the Downdrift”. This class of intonation pattern is found in utterance types of, for example, statement, citation form, submissive, concealed anger, and bored. The next class is “the Rise Class”, which can be found in sentence modes such as question, disagreeable, disbelieving, surprised, and unfinished. The last class, “the Convolution Class”, proposed in [2], has the shape of F0 contour that is the combination of Falls and Rises. This type of intonation can be found in utterances marked with emphatic, agreeable, interested, or believing attitudes.

This paper presents a method of extracting feature vectors from F0 contour. These feature vectors were used as input of neural network to classify the intonation shape of Thai utterance into three classes as mention above.

2. The Feature Contours

2.1. Feature contours extraction method

In order to extract paralinguistic information from F0 contour, we have to reduce (or eliminate) the effect of linguistic and nonlinguistic information. First of all, we view F0 contour as the superposition of three components in the same way as the Fujisaki-model of tonal language. These components are: the phrase component, which was directly affected by intonation type. The tone component, which corresponds to lexical tone type of each syllable. And the Fb, which is a constant for each speaker. So the mainly feature we have to extract is the phrase component, which may declines in declarative intonation and rises in interrogative intonation. The tone component, which seem to be directly correspond to linguistic information. It was, however, also affected by intonation type. This is because the range of F0 contour of the same tone is different in different intonation [2]. So we have to include the feature that represents the range of tone component but does not represent the shape of each tone. The Fb, the factor of F0 level, is also another feature that we have to include. This is because; the level of F0 of the utterances may different in different intonation although the same speaker spoke them.

So we present four steps to extract intonation features from speech as below:

1. Extract F0 contour from the utterance. In this work, we used PRAAT program (© P.Boersma) to do this task.

2. Apply error correction to F0 contour by using median filtering technique. Then, all unvoiced region in F0 contour will be filled by performing linear interpolation. We call the contour after passing this process as ‘connected F0’ or CF0, as can be seen in Figure 1.

3. In order to decompose the phrase component and Fb from the tone component, we use the technique, which is similar to the technique that [3] used to separate the accent component from the phrase component and Fb. That is, filter the CF0 by using FIR low-pass filter to get the slow movement component, which is the summation of the phrase component and Fb. In this work, we varied the cutoff frequency of the filter (LFC_Fc) from 0.5 to 3.0 Hz. The output of low-pass filter was called ‘low frequency contour’ or LFC. An example of the LFC compare with CF0 is shown in Figure 2.

4. Then the CF0 was subtracted by LFC to get the faster movement component, ‘high frequency contour’ (HFC), which is the tone component of F0 contour. As mention above, we do not want to get the shape of F0, which

corresponds to lexical tone. We just only want to find the feature that corresponds to the swinging range of the tone component. So the HFC is taken an absolute, then filtered by low-pass filter to get contour that is analogous to swinging range of tone component. Cutoff frequency of the filter, FVC_Fc, was varied in the range 0.5 to 2.5 Hz. This contour is called 'F0 variation contour' or FVC, as shown in Figure 3.

Now the LFC and the FVC was used as the intonation feature contours to use in our experiment. However, for easily understanding of how the feature contours relate to the raw F0, LFC, LFC + FVC, and LFC - FVC were plotted in the same graph with the raw F0 (see Figure 4 – 7).

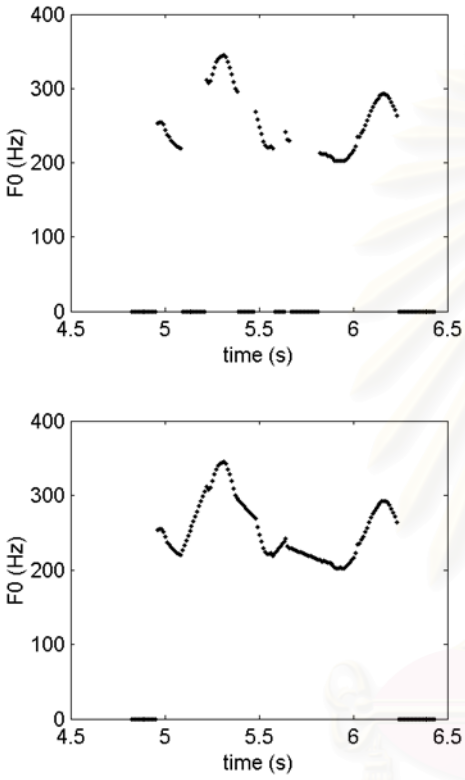


Figure 1: Top: Example of F0 contour after median filtering process *Bottom:* CF0

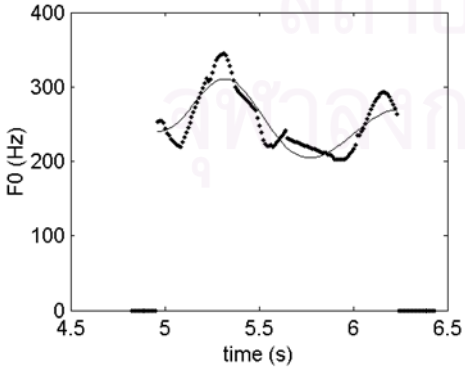


Figure 2: CF0 (dotted) and LFC (solid line) where LFC_Fc = 2.0 Hz

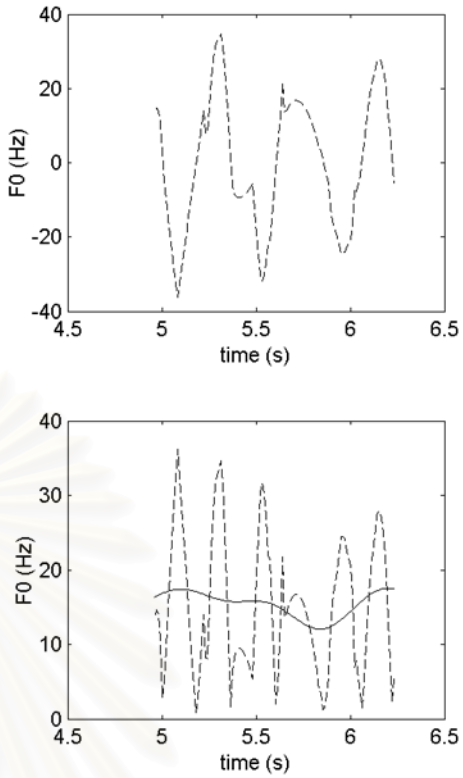


Figure 3: Top: HFC which is CF0 - LFC *Bottom:* |HFC| (dashed) and FVC (solid line) where FVC_Fc = 2.0 Hz

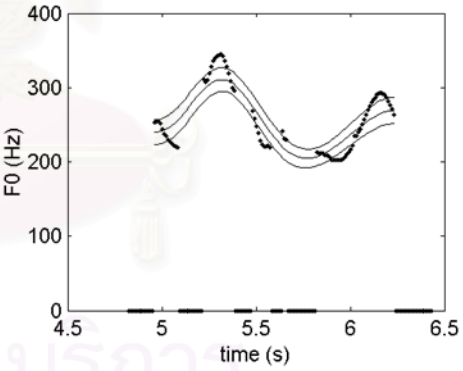


Figure 4: F0 contour (dotted) in Figure 1 compare with LFC (middle solid line), LFC + FVC (top solid line), and LFC – FVC (bottom solid line)

2.2. Examples of feature contour in different intonation class

As shown in Figure 5 – 7, the LFC is falling in the Fall Class, rising in the Rise Class, and both rising and falling in the Convolution Class. LFC + FVC and LFC – FVC are close to the LFC in the Fall Class, but they are far from the LFC in the Rise and the Convolution Class.

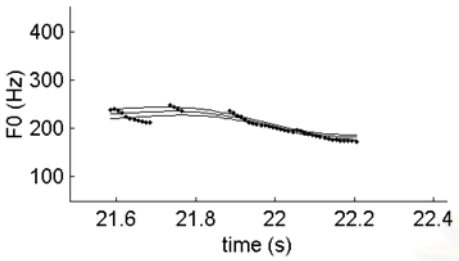


Figure 5: Example of raw F0 contour (dotted) compared with LFC (middle solid line), LFC + FVC (top solid line), and LFC – FVC (bottom solid line) of the Fall Class spoken by female speaker (LFC_Fc = 2.0 Hz, FVC_Fc = 2.0 Hz)

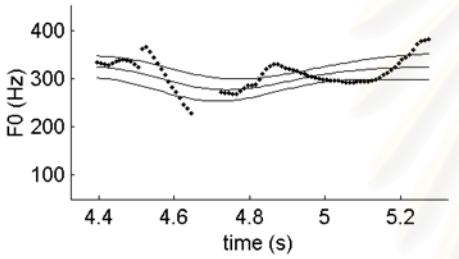


Figure 6: Example of raw F0 contour (dotted) compared with LFC (middle solid line), LFC + FVC (top solid line), and LFC – FVC (bottom solid line) of the Rise Class spoken by female speaker (LFC_Fc = 2.0 Hz, FVC_Fc = 2.0 Hz)

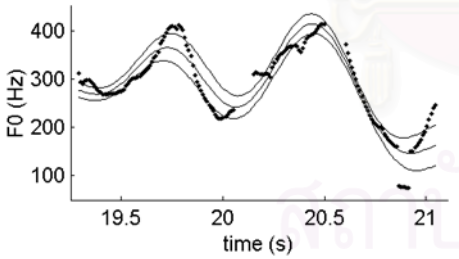


Figure 7: Example of raw F0 contour (dotted) compared with LFC (middle solid line), LFC + FVC (top solid line), and LFC – FVC (bottom solid line) of the Convolution Class spoken by female speaker (LFC_Fc = 2.0 Hz, FVC_Fc = 2.0 Hz)

3. Experiments

We used 61 sentence scripts from 6 spoken dialogues that were used in Thai intonation study [4]. There are 6 male speakers and 6 female speakers speaking these dialogues, so there are 732 sentences (366 sentences from male speakers and 366 sentences from female speakers) for our experiments. We treated the utterance from male and female speakers separately in all experiments to discard the problem of frequency scale normalization. Although each sentence script was labeled by intonation type, some speaker may speak some sentence with the different intonation from the labeled intonation type. So we have to do a perception test to determine the perceived intonation of each sentence.

3.1. Perception experiment

Two listeners were asked to listen to all 732 sentences randomly for 2 times and label the intonation type of each sentence. So, each sentence was listened for 4 times. We found that some sentences were clearly categorized the intonation type, i.e. it was labeled as the same intonation type at least 3 times. However, some sentences were ambiguously classified the intonation type, i.e. they were labeled as the same intonation type less than 3 times. So we used only the clearly sentences for the next experiment. We labeled each clearly sentence as the most frequently intonation type chosen by listeners. Table 1 lists the number of the clearly sentences of each intonation type and gender.

Table 1: The number of the sentences that were clearly classified of each intonation type and gender by listeners

Intonation type	Gender	
	Male	Female
The Fall Class	98	106
The Rise Class	91	65
The Convolution Class	79	122

3.2. Automatic recognition experiment

We used three-layer feedforward neural network as a recognizer. The number of input nodes depended on the number of dimensions of feature vectors. The feature vectors were built from the value of LFC and FVC contour at 0%, 25%, 50%, 75%, and 100% time points of the duration of the contour. So there are 10 input nodes in this experiment. The number of hidden nodes is 20. Each output node represents each intonation class, so the number of output nodes is 3. LFC_Fc was set to 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 Hz. FVC_Fc was set to 0.5, 1.0, 1.5, 2.0, and 2.5 Hz. The NNs were trained and tested for all possible combinations of FVC_Fc and LFC_Fc. Average recognition rate of each combination of male and female speakers were shown in Table 2 and 3 respectively.

Table 2: Average recognition rates (%) of male speakers of all possible combination of LFC_Fc and FVC_Fc

	LFC_Fc (Hz)					
FVC_Fc (Hz)	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**
0.5	56.7	54.9	57.5	60.1	63.4	60.1
1.0	61.9	57.5	60.8	60.8	60.8	59.0
1.5	58.6	**63.4**	62.3	61.9	61.6	59.7
2.0	61.6	61.6	60.1	60.1	62.7	58.6
2.5	62.3	59.7	62.3	61.9	60.1	57.5

Table 3: Average recognition rates (%) of female speakers of all possible combination of LFC_Fc and FVC_Fc

	LFC_Fc (Hz)					
FVC_Fc (Hz)	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**
0.5	67.2	69.3	70.3	71.3	70.0	72.0
1.0	67.9	67.9	71.3	71.3	70.0	73.4
1.5	68.6	67.6	73.4	71.0	71.7	72.4
2.0	68.6	66.2	74.4	**75.4**	68.9	73.4
2.5	65.9	67.6	71.7	73.0	72.7	72.0

Table 2 and Table 3 show that the average recognition rates of female speakers are higher than the recognition rate of male speakers. The value of LFC_Fc and FVC_Fc that yield the maximum recognition rate are also different between male and female.

Then, from Table 2 and 3, we picked the combination that the pair of LFC_Fc and the FVC_Fc give the maximum recognition rate. For male, the maximum recognition rate occurs at LFC_Fc = 1.0 Hz, and FVC_Fc = 1.5 Hz. For female, it occurs at LFC_Fc = 2.0 Hz, and FVC_Fc = 2.0 Hz. Confusion matrices of the maximum recognition rate experiments of male and female speakers were shown in Table 4 and 5, where "F Class" is the Fall class, "R Class" is the Rise Class, and "C Class" is the Convolution Class.

Table 4: Confusion matrix of the highest recognition rate of male speakers from Table 2

Input intonation class	**Recognized intonation class**			**Total utterances**
	F Class	**R Class**	**C Class**	
F Class	**77.6**	12.2	10.2	98
R Class	17.6	**56.0**	**26.4**	91
C Class	15.2	**30.4**	**54.4**	79
Total				268

Table 5: Confusion matrix of the highest recognition rate of female speakers from Table 3

Input intonation class	**Recognized intonation class**			**Total utterances**
	F Class	**R Class**	**C Class**	
F Class	**88.7**	1.9	9.4	106
R Class	15.4	**50.8**	**33.8**	65
C Class	9.0	**13.9**	**77.0**	122
Total				293

From Table 4 and 5, we found that the recognizers were easy to distinguish the Fall Class from the other classes. We also found that the recognizers have a confusion to recognize between the Rise Class and the Convolution Class.

4. Conclusions

This paper presents a Thai intonation features extracting method from F0 contour. The method uses the fact that there are three types of information in F0 contour of Thai utterance. Extracting one type of information is to eliminate or reduce the effect of the other two types. The Thai intonation can be categorized as paralinguistic information. So we have to eliminate linguistic and nonlinguistic information. We proposed a method of eliminating the linguistic information from F0 contour to get the LFC, and the FVC contour. These contours were used as feature contours to recognize intonation by neural network. The effect of nonlinguistic function was reduced by treating speech from male and female speakers separately. The results show that the maximum recognition rate of male utterances is 63.4%, which is lower than 75.4% of female utterances. The recognizers were easy to distinguish the Fall Class from the other classes, but difficult to recognize between the Rise Class and the Falling Class.

5. Acknowledgements

The authors would like to thank to Dr. Rachod Thongprasirt, and Dr. Satien Triamlamlert from National Electronics and Computer Technology Center (NECTEC) for helpful comments and suggestions. The authors would like to acknowledge the Thailand Graduate Institute of Science and Technology (TGIST) for partial financial support for the research to the first author, and Centre for Research in Speech and Language Processing, Department of Linguistics Chulalongkorn University (CRSLP) for providing facilities.

6. References

[1] Fujisaki, H., "Prosody, Models, and Spontaneous Speech". *Computing Prosody (Sagisaka, Y., Campbell, N., and Higuchi, N., eds.)*, Springer-Verlag (1996) 27-42.

[2] Luksaneeyanawin, S., "Intonation in Thai". *Intonation Systems, A Survey of Twenty Languages (Hirst, D., Cristo, A. D.)*, Cambridge University Press: 376-394, 1998.

[3] Mixdorff, H., "A Novel Approach to The Fully Automatic Extraction of Fujisaki Model Parameters", *Proceedings of 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, Istanbul, Jun. 2000, pp. 1281-1284.

[4] Luksaneeyanawin, S., "Intonation in Thai". *Ph.D. Dissertation, University of Edinburgh,* 1983.

ประวัติผู้เขียนวิทยานิพนธ์

นายปฐวี ชาญไววิทย์ เกิดวันที่ 13 มิถุนายน พ.ศ. 2522 จังหวัดกรุงเทพมหานคร เข้าศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2539 สำเร็จการศึกษาวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2542 เข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2543